

Real-time probabelistic forecasts for the 2018 – 2019 Ebola outbreak in Northern Democratic Republic of Congo

CANDIDATE NUMBER

September 19, 2019

Word counts:

Abstract: XXX

Report: XXXX

Abstract

Objective: Constructing and evaluating real-time probabilistic forecasting models for the 2018–2019 Ebola outbreak in north-eastern DRC on a national and sub-national level.

Methods: Using publicly available daily incidence data we constructed a set of semi-mechanistic models based on a branching process with a time varying reproduction number. We include models with both Poisson and negative binomial offspring distributions. To forecast the reproduction number we first estimated the 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 compared the models using proper scoring rules and checked the calibration of the forecasts using Probability Integral Transformations.

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. This model could provide calibrated forecasts for 9 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 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 northern DRC on the national level and partly at sub-national level. Such forecasts can be useful to better understand the outbreak and to support the outbreak response.

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

Contents

1	Introduction	4
1.1	2018–2019 Ebola outbreak in North East DRC	5
1.2	Mathematical Modelling of Ebola	6
2	Methods	7
2.1	Data Sources	7
2.2	Model	8
2.2.1	Offspring distribution	9
2.2.2	Reproduction number	11
2.2.3	Forecasting incidence	13
2.3	Assessing probabilistic forecasts	14
2.3.1	Calibration	14
2.3.2	Sharpness and Bias	15
2.3.3	Proper scoring rules	16
2.4	Implementation	17
3	Results	18
3.1	National level forecasts	19
3.2	Sub-national forecasts	22
4	Discussion	25
5	Conclusions	32
	Appendices	39

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 resone was pioneered durin the previous large Ebola outbreak in West Africa in 2014–2015 [3]. The models, that were developed during and after the outbreak, were used both to increase our understanding of the outbreak and to predict the spread of the outbreak. Estimates were used in the official response planning [4] and to for example estimate needed bed capacity [5]. Disease modelling has also been successfully implemented and used for many other response plans for many other infections diseases including Zika [6] and influenza [7]. To effectively use 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 be 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 probabelistic such that they assess and incorporate uncertainty into their predictions [8, 9, 10]. Without understanding the range of possible outcomes from a model and their associated probabilities it is very difficult to support effective public health action in outbreak situations. Disease transmission process are inherently random systems so quantifying all the different types of uncertaintyis key. Once a model has been forumlated, it needs to be evaluated. Evaluations for probabilistic forecasts requires us to not just assess the accuracy of point predictions, but to assess if the model correctly assess it's own uncertainty [11, 12]. A model that that successfully assesses it's own uncertainty is said to be calibrated. We should strive to create well calibrated models that can be updated in real-time and that that are interpretable and can be used in the outbreak response.

In this thesis, we will describe a framework for flexibly producing probabilistic forecasts

and assess how well the resulting models can be used in the ongoing Ebola outbreak in north-eastern DRC. We will compare different models to choose the best model and to explore the epidemiology of the current outbreak.

1.1 2018–2019 Ebola outbreak in North East DRC

Ebola virus disease(EVD) is an often fatal disease that is transmitted through close contact with bodily fluids [13]. The symptoms start 2-21 days after infection and include fever, vomiting, diarrhoea and in some cases internal and external bleeding. It has been found that health-care workers are at high risk of infection and that burial practises that include direct contact with the dead body is a source of many infections[14]. There have been multiple Ebola outbreaks since the first recorded outbreak in 1976. The case fatality rate has varied between 10-90% and apart from the large outbreak in Western Africa 2014–2015 that had over 28,000 cases all of the previous outbreaks have had less than 500 cases [13].

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 [15]. North Kivu and surrounding provinces Ituri and South Kivu constitute a populous region in north eastern DRC that borders both Rwanda and Uganda. The area has seen 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 central 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 to both the central government and international organisations. By the 27th of September the outbreak had continued spreading and the WHO assessed that the outbreak constituted a very high national and regional risk[16].

The security and distrust from the local community has made the outbreak response very challenging. This has led 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 work[17, 18].Even with 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 accross the borders to 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 have been evaluated during the outbreak. Preliminary data shows positive signs that this vaccine will provide some protection from Ebola [19]. The vaccince has 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 at the 12th of August due to findings that two of the four drugs that were included showed s clear effect of reducing mortality from Ebola, especially with early treatment [20].

On the 17th of July the Director General of the WHO declared the Ebola outbreak a Public Health Emergency of International Concern [21] and as of September 2019 the outbreak is still ongoing with new cases from 17 health zones [22].

1.2 Mathematical Modelling of Ebola

A large number of different mathematical models have been proposed for modelling and forecasting of Ebola, see e.g [3, 23]. The models can be divided into categories based on the structure of the models; mechanistic, semi-mechanistic, phenomenological and hybrid models. Mechanistic models are based on a specified mechanism for disease transmission and includes standard compartmental models while phenomenological models do not include any mechanism for disease transmission. An example of a phenomenological model would be the class of autoregressive integrated moving average models used for many applications of timeseries forecasting. A semi-mechanistic model include some mechanistic assumptions, together with some parts of the model that are based on more flexibles techniques including

non-parametric models. The aim of these models is to model some important parts of the disease transmission process, while giving enough flexibility to fit real life outbreaks. Based on the review of models for the west african ebola outbreak it was found that it was important with models that actively modeled the uncertainty in the disease transmission process [3, 24] 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 [23]. The challenge presented a range of different scenarios of an 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 comparing the models they found that for short-term forecasting the accuracy was not related to model complexity and that “light” non-parametric models were able to do well. 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 be integrated with the response surveillance.

2 Methods

2.1 Data Sources

We will use daily reports of new cases from the Ebola outbreak in north east DRC as reported by the Ministry of Health as the data source for our models. This data has been made publicly available and has been distributed via the Humanitarian Data Exchange [25]. We use data from the beginning of the outbreak until the 1st of September 2019. The

data gives the number of 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 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. This can lead to days with negative incidence, 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 are 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.

2.2 Model

In our aim to construct a flexible semi-mechanistic framework for modelling of Ebola we will consider models where the daily incidence, I_t , follows a modified branching process [26]. This class of models have been considered for infectious disease modelling previously in [27, 28, 29] and were used to model Ebola during the 2014–2016 Ebola outbreak in Western Africa [4, 30]. The main idea is that each infected person gives rise to ν new infections, where ν 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 infectious 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, λ_t , which is a product of previous incidence weighted by the discretised serial interval, w_τ , and the instantaneous reproduction number R_t . The expected value of the incidence is given by:

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

This equation, together with the probability distribution for I_t specifies our modelling framework. An individual model is then specified by the probability distribution for I_t , the serial interval and the reproduction number as a function of time. For all the models in this study we will use a gamma-distribution with mean 15.3 days and standard deviation of 9.3 days as the serial interval as fitted to data from the West-Africa 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.

2.2.1 Offspring distribution

The offspring distribution gives us the distribution of the number 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 λ is both mean and 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 sub-interval is independent.

For a number of diseases it has been shown that the offspring distribution has larger variance than that given by the Poisson distribution and has so called super-spreaders [28]. 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, μ and the dispersion parameter k . In this formulation the variance is given by $\mu(1 + \frac{1}{k})$. For $k \rightarrow \infty$ we get 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 is. For SARS it has been found that $k = 0.16$ [28]. For the Ebola outbreak in West Africa they found $k = 0.53$ using conservative assumptions and data from the transmission tree of the infections [30]. The different offspring distributions can be seen in Figure 1

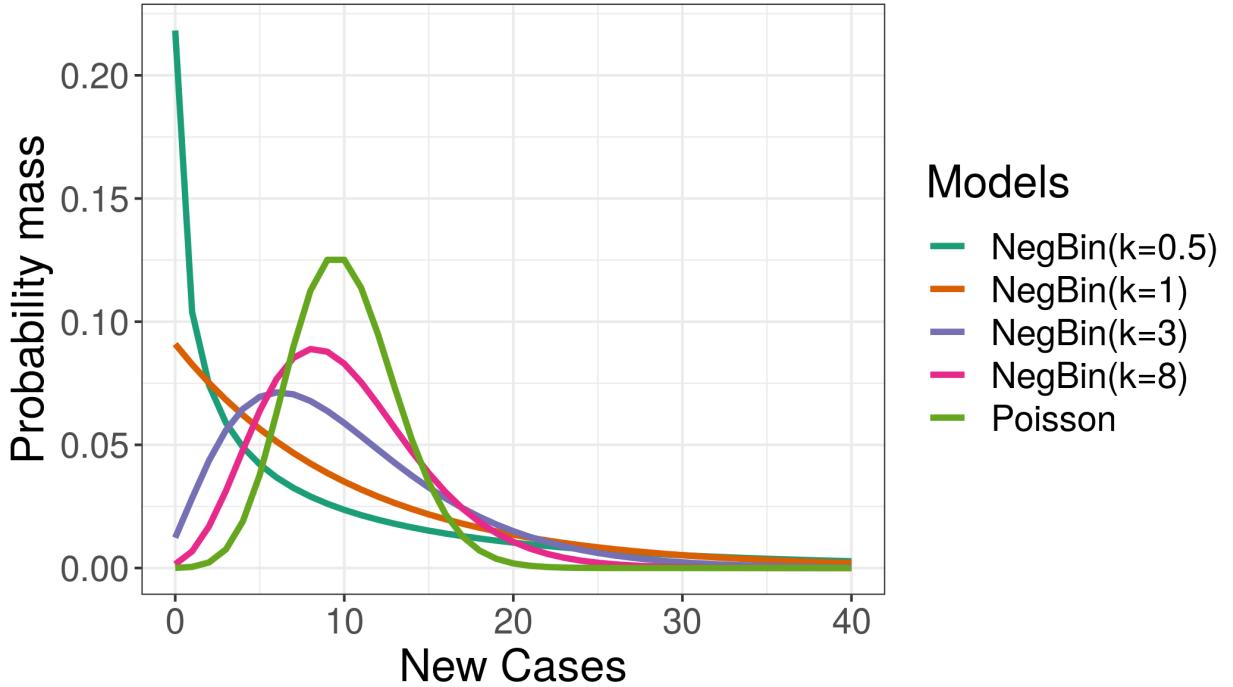


Figure 1: Poisson and negative binomial offspring distributions with varying dispersion parameter k . All the distributions have mean 10.

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 copies 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. Therefore the probability distribution for I_t is also given by either the poisson or the negative binomial distribution. We will investigate models with both offspring distributions to assess which models fit the current outbreak better.

2.2.2 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 some parametric function where we could fit the parameters to data or we can use non-parametric methods to allow a fairly unconstrained evolution with time. This free function for the reproduction number is what gives the model it's flexibility to fit a large number 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.

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 fit the values of the prepduction number prior to t using the method described below. We then use the estimated values to predict the future evolution of R_t . We will use different methods for forecasting R_t to compare how well they work.

To estimate the reproduction number from the historical incidence data we use the method developed in [27, 31]. This approaches used the same formulation as in Equation 1 to estimate the reproduction number from daily incidence and the same serial interval as in our model. If the reproduction number is estimated daily it is likely to vary too much from

day to day in a manner that it 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 uncertainty of the estimate. We use the R-package EpiEstim [32] to estimate the reproduction number using a parametric gamma distribution for the serial interval with a 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 very simple method where we assume that reproduction number remains constant from the last historical value. We use the uncertainties given by the method in [27] 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 gamma posterior distribution for R_{t-1} . We then get realisations of R_t by drawing random numbers from this posterior distribution.

The second approach to forecast the reproduction number is based on Bayesian structural time series 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:

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

We use the R-package BSTS [33] to fit the Bayesian structural time series with the following model:

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

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

δ_t is the semi-local trend that we model as an auto-regressive process of order one. This gives trend process that can oscillate around the level D with smaller random-walk

excursions. Inverse gamma-priors are used for the standard deviation parameters σ_μ and σ_δ , a Gaussian prior on D and a $N(0, 0.1)$ prior for ϕ . The ϕ 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 prior for small values of ϕ to keep the model from having 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.

2.2.3 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 application, 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+1} we first use all the data up until time t to estimate the history of the reproduction number until time t . Our probabilistic forecast will be based on sampling possible outcomes to generate a distribution of outcomes. We therefore first draw sample for R_{t+1} from the predictive distributions of the models discussed above, then we combine this with the historical incidence data to calculate the total force of infection λ_{t+1} . We then sample I_{t+1} from the specified offspring distribution with an expected value given by λ_{t+1} . If we want to forecast over multiple time-steps we follow the same procedure by sampling values for the reproduction number. When calculating λ_{t+2} we use the sampled value for I_{t+1} together with the historical data I_s for $s \leq t$. This allows us to generate probabilistic forecasts for any forecasting horizon.

2.3 Assessing probabilistic forecasts

The aim of probabilistic forecasts is to predict both the correct average value and the correct probabilities for 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. An example of such a metric is the root mean square 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 [11]. In addition we will consider proper scoring rules for comparing probability distributions. We follow the approach taken in [8], where probabilistic forecasts for the West African Ebola outbreak were assessed using similar methods.

2.3.1 Calibration

A forecasting model is calibrated if its forecasted distribution is equal to the actual distribution of observed values. This means that the model can accurately assess its own uncertainty. For example if we predict that it will rain with a 60% chance and we find that over time it does rain 60% of days when we predicted a 60% chance of rain, the model would be well calibrated. Mathematically, if we assume that 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 perfectly calibrated if $F_t = G_t$. To assess calibration we will use a randomised Probability Integral Transformation (PIT) [34]. We consider a forecasted cumulative distribution, F_t , compared to an observed value k_t . For each such pair of forecasts and observed value we calculate:

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

where ν is a standard uniform random variable. If the prediction is ideal, the u_t s 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 can reject that the model is calibrated. We implement this using the goftest [35] R-package. It is important to note that uniform PIT values is a necessary, but not sufficient condition for an ideal forecast and since we only have limited data that 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[34]. If the forecast is too dispersed the PIT values are likely to cluster in the centre of the PIT histogram, while if the forecast is under-dispersed 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

$$\text{centrality} = \frac{N(0.25 < u_t < 0.75)}{N} - 0.5,$$

where $N()$ is the number of u_t values.

When the centrality scores is less than 0, most of the PIT values or outside of 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 for each set of forecasts repeat the calculations of the PIT values 10 times and take averages to average out the effect of the randomness in the definition of u_t [8].

2.3.2 Sharpness and Bias

Sharpness is defined as the range of values in the forecast. The sharper a forecast, the more certain we are 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 y

to quantify sharpness:

$$S_t(F_t) = \frac{1}{0.675} \text{median}(|y - \text{median}(y)|),$$

the normalisation factor means that S_t is equal to the standard deviation if F_t is normal.

It is also of importance to assess the bias of the forecast. Are we more likely to predict too large or too small values? 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 the forecast is unbiased. B_t is between -1 and 1, with both extreme values signifying a completely biased forecast.

2.3.3 Proper scoring rules

Proper scoring rules have been developed to rank forecasts. To be a proper scoring rule, the scoring rule has to minimised when the forecast is ideal. These scoring rules combine calibration and sharpness and give a consistent ranking of forecasts. We will use the the continuously ranked probability score (CRPS) and the Dawid-Sebastiani score (DSS) as implemented in the scoringRules package [36] for samples from the predictive distribution.

The CRPS score is given by

$$\text{CRPS}(F_t, k_t) = \int_R (F_t(z) - \mathbb{1}_{k_t \leq z})^2 dz.$$

The CRPS can be viewed as an extension of the mean absolute error for a determinisitic forecast. THe DSS score only depends on the mean, μ_p and standard deviation, σ_p of the predictive distribution

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

The DSS allows a more 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 CRPS a kernel density estimate is used to estimate F_t , for the DSS we use the mean and standard deviation of the sample.

2.4 Implementation

The models were implemented in the R programming language [37] and are available at http://github.com/gulfa/msc_ebola. We will consider four different models to assess which model fits the data best. 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 and Poisson offspring distribution
4. Model 4 (NegBin Semi-local): Varying reproduction number and negative binomial offspring distribution

For the dispersion parameter k for the negative binomial models we fit a value for the simple negative binomial model (Model 4) by minimising the continuous ranked probability score for the one day ahead predictions for the model on the level of the whole outbreak.

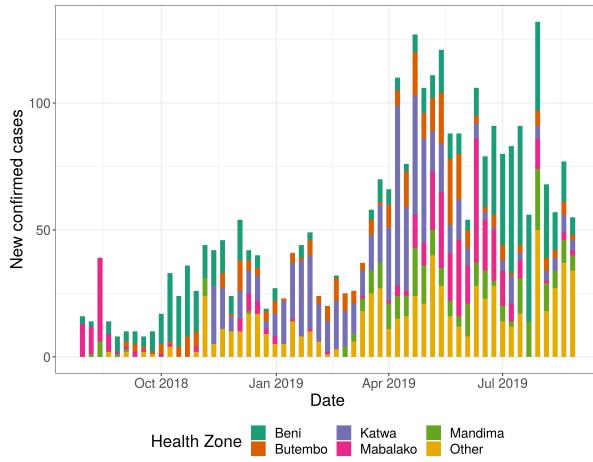
We will assess how well the models work both for the entire epidemic and for each health zone. To evaluate a model we will estimate the calibration, sharpness, bias, log score and

CRPS for forecasts up to 28 days ahead. To do this we start 17 days¹ after the first case of the outbreak in the given location and calculate the d -ahead prediction for all historically available data. For the calibration we use all forecasts for the d -ahead compared to the observed values the 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.

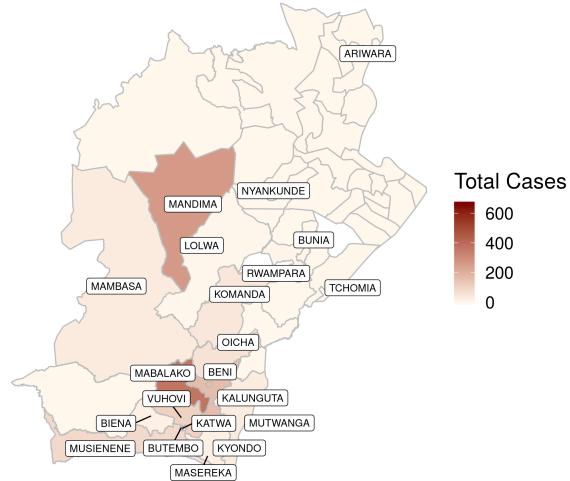
3 Results

A total of 2,926 confirmed Ebola cases and 1926 confirmed Ebola deaths have been reported in the ongoing outbreak in DRC as of the 31st of August, 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 varrying 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 varied around 1 with some waves of larger transmission. From March 2019 the number of cases per week increased significantly. The larger number of weekly cases continued in June and July even as the reproduction number decreased. From July 2019 the reproduction has varied 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 with Beni and Katwat health zones being hardet hit. From the epi-curve there seems to be some connection between large increases in incidence and introduction or reintroduction into specific health zones. We can also see that the disease is moving around between health zones since the total incidence is dominated by different health zones at different times in the outbreak.

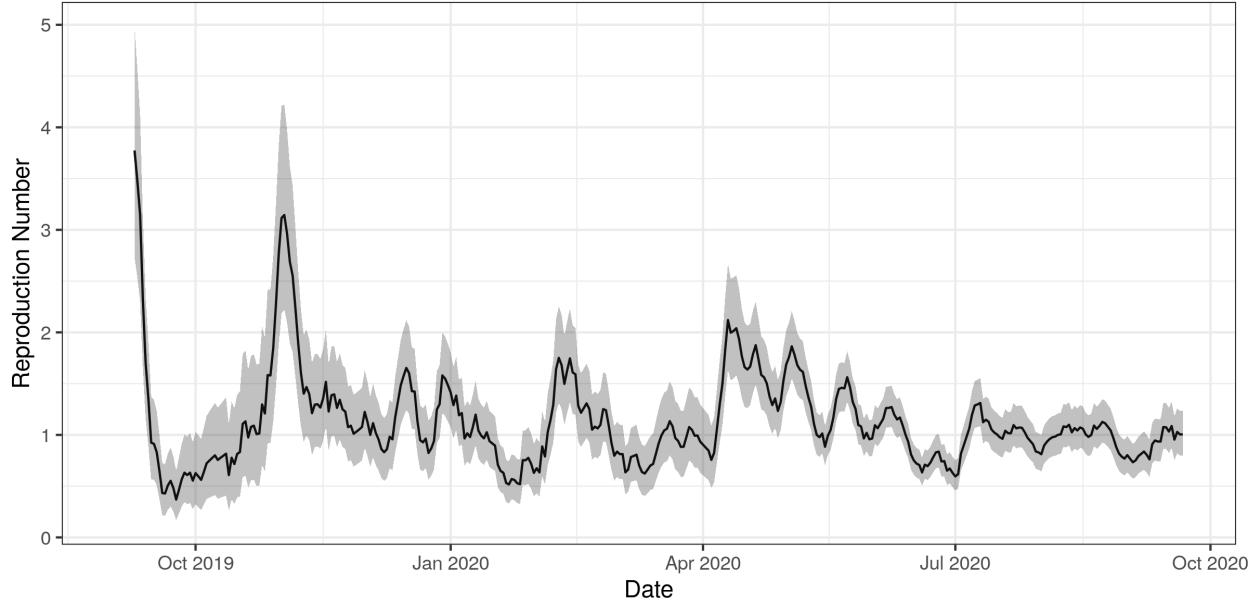
¹The 17 days are due the fact that method for estimating the reproduction number needs thi smany days to give reasonable values for the reproduction number



(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. The shaded region shows 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.

3.1 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 CRPS score for one day ahead predictions for the negative binomial models was $k \approx 8$, but a fairly large range of parameters gave a reasonable fit. We therefore chose to use $k = 8$ for the negative binomial distributions both for the national and health zone forecasts.

All four of models were successfully 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, CRPS and the centrality of PIT values as a function of the forecasting horizon in Figure 60

Model	Horizon	Sharpness	Bias	CRPS	DSS	Centrality	Calibration
NegBin Semilocal	1	4.08	-0.03	1.87	3.30	0.04	0.24
NegBin Latest	1	4.14	-0.00	1.82	3.26	0.06	0.18
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	0.15	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	0.08	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	-0.09	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 forecast, the negative binomial offspring distribution is needed for a calibrated forecast as only the two negative binomial models are calibrated with a p-value larger than 0.1. We can also see that there are no big differences for one-day

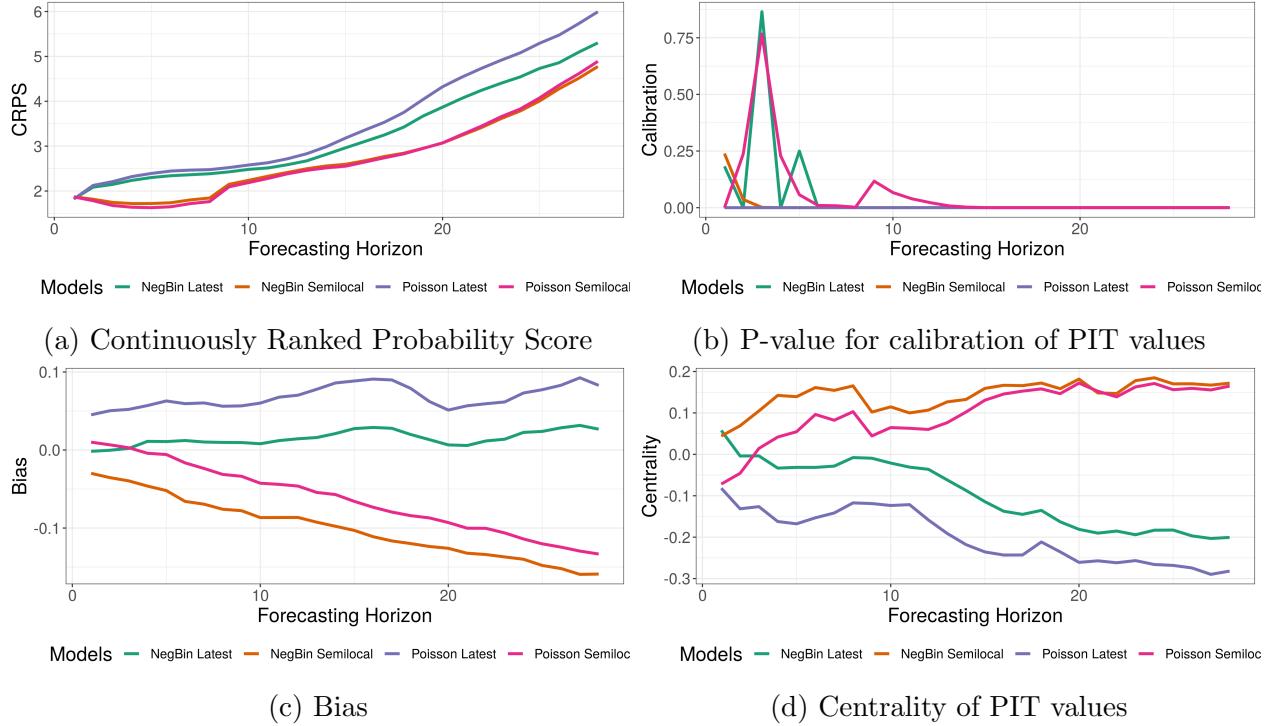


Figure 3: Evaluation of the four models as a function of forecasting horizon at the national level.

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. At very short forecasting horizons it is uncalibrated as it is underestimating its own uncertainty. This is the only model that is calibrated for longer forecasting horizons and it also has the lowest CRPS and DSS scores. The centrality of the PIT scores in Figure 3d give 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 larger uncertainty as we can also see from the sharpness figures. This is

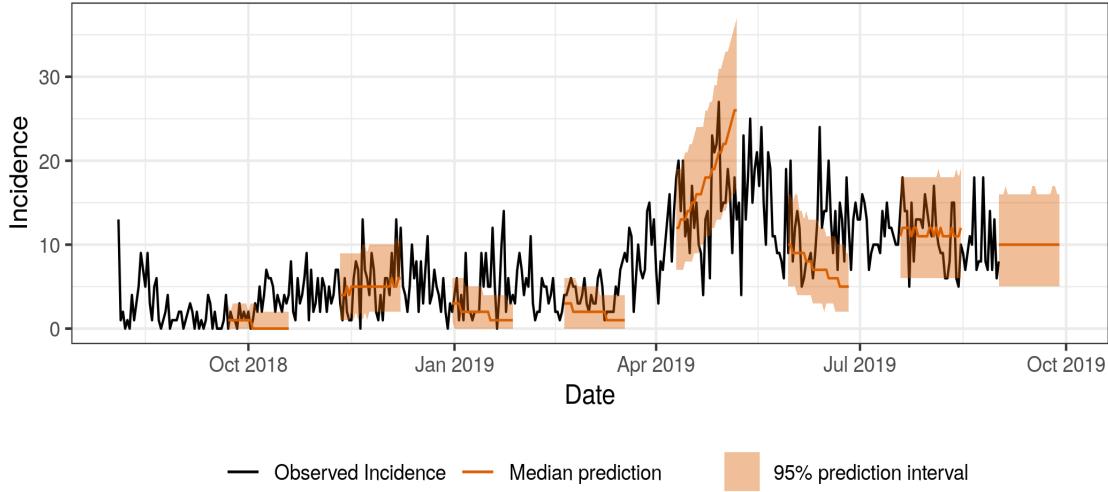
why the poisson model is prefered at larger forecasting horizons. After about one week the sharpness of the poisson semi-local model has grown to be the same as the negative binomial latest model. We can see how the semi-local models quickly become less sharp, while the sharpness for the constant models does change as much with forecasting horizon. We can also see that 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 that the amount of uncertainty often seems a bit too large for the reproduction number forecasts.

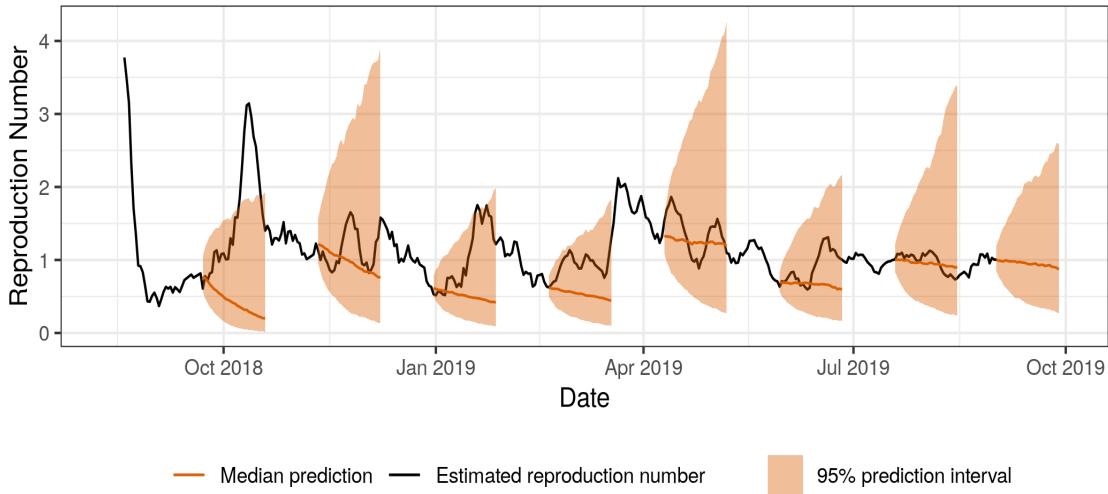
We also forecast the incidence for the 28 days following the the 1st of Sepemteber 2019 in Figure 4. Here the model predict a slighlty decreasing median reproduction number and a fairly steady incidence through the month of September. We exepct between 10–15 cases per day over the next month. If the current trend continues without any new wave of higher reproduction number there is hope that the outbreak can be brought into control if the reproduction keeps decreasing.

3.2 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 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 forecasting horizons less than 28 days.



(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

For all health zones apart from one, 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. Once our model has seen no new infections in a few weeks they will predict close to zero new cases with essentially

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
Bunia	20	NegBin Semilocal	4
Mwenga	18	All models	6
Alimbongo	13	Poisson Semilocal	5
Kayina	9	Poisson Semilocal	10
national	9	Poisson Semilocal	2942
Mutwanga	7	Poisson Semilocal	31
Rwampara	6	NegBin Latest	8
Lubero	4	NegBin Semilocal	31
Musienene	4	Poisson Semilocal	84
Butembo	3	NegBin Semilocal	279
Kyondo	3	NegBin Semilocal	22
Mabalako	3	NegBin Semilocal	371
Beni	2	NegBin Semilocal	661
Vuhovi	2	NegBin Semilocal	103
Biena	NA	No calibrated model	16
Goma	NA	No calibrated model	1
Kalunguta	NA	No calibrated model	164
Katwa	NA	No calibrated model	647
Mangurujipa	NA	No calibrated model	20
Masereka	NA	No calibrated model	50
Oicha	NA	No calibrated model	55
Pinga	NA	No calibrated model	1
Ariwara	NA	No calibrated model	1
Komanda	NA	No calibrated model	43
Mambasa	NA	No calibrated model	32
Mandima	NA	No calibrated model	264
Nyankunde	NA	No calibrated model	1
Rwampara (Bunia)	NA	No calibrated model	1

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

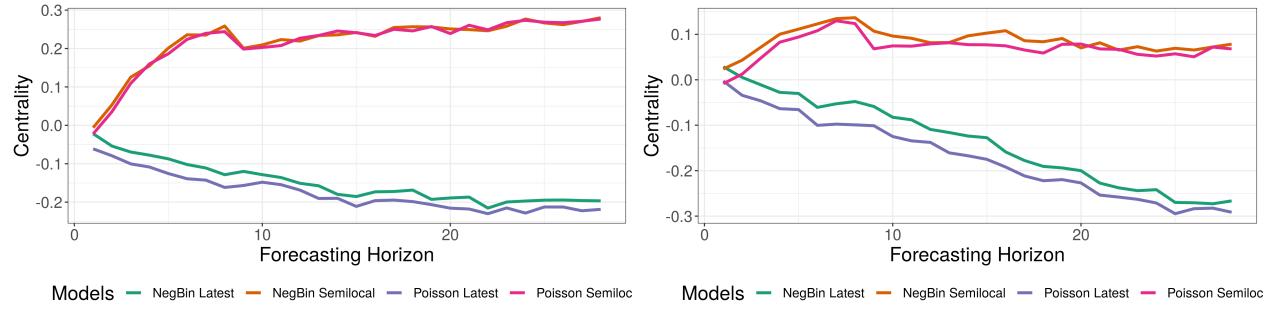
0 sharpness irrespective of the current reproduction number. This behaviour works well when there are no new imported cases as we can see in some of the health zones. For the larger health zones where there is more sustained transmission the models work for some health zones, but not as well on the national level. There are also some health zones with

many 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 models and evaluations for all the other health zones can bee 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.

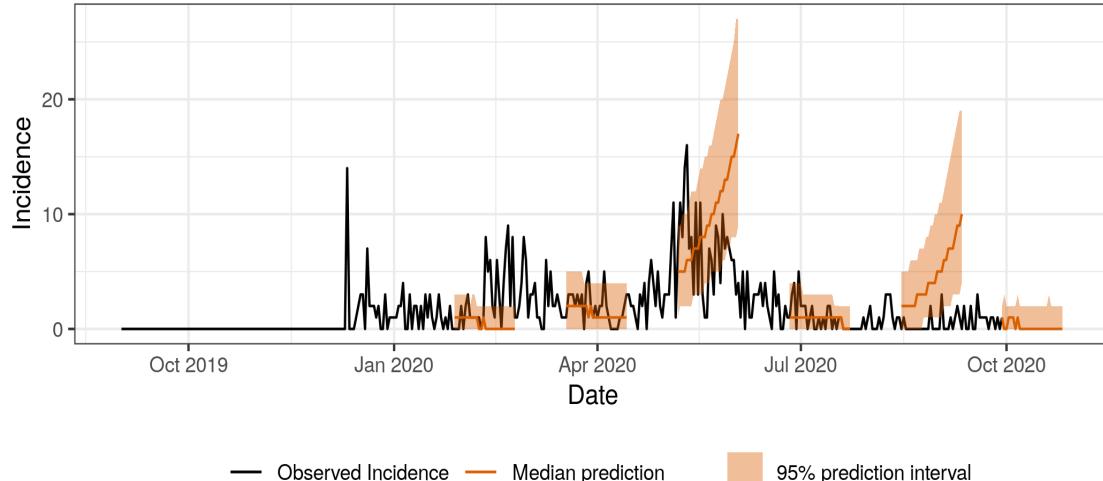
As a final check of the health zones predictions we compared the the one-day ahead predictions of the national level model with the one-day ahead predictions of a sum of the semi-local Poisson model for all the health zones in Figure 6. We see that model based on summing up all the health zones predict consistently too large incidence and would not work well for forecasting the outbreak. The combined model also does not have any forecasting horizons with calibrated forecasts and have significantly higher CRPS and DSS score compared to the model fit to national data

4 Discussion

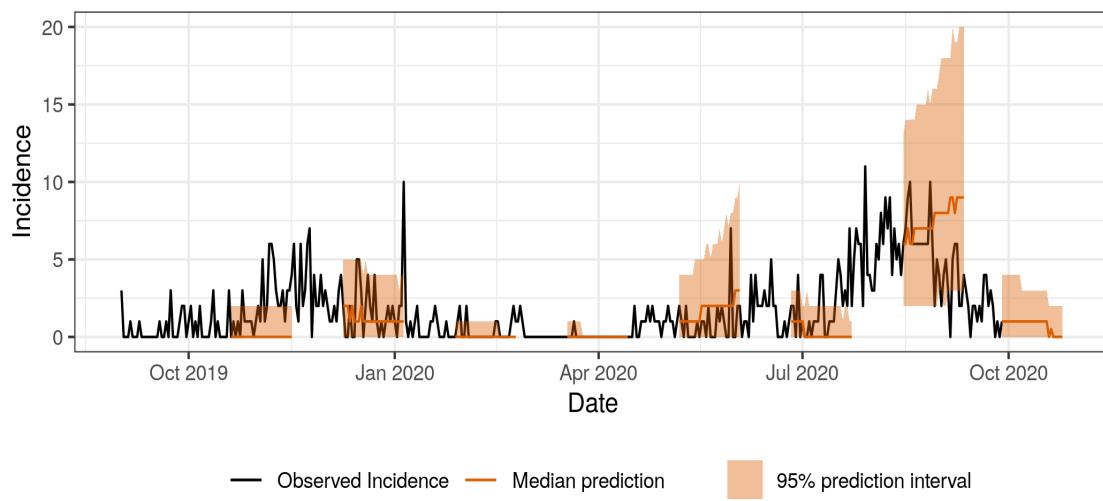
The ongoing Ebola outbreak in northern DRC is the second largest Ebola outbreak and is of significant regional and international concern. As of the 1st of September 2019, it is not under control as we can see both from the fact the reproduction number is not below one and that there are still plenty of new cases and deaths not on contact lists. In addition to a large number of confirmed Ebola cases, this outbreak has seen a high CFR of 66% compared to 51% in the west African outbreak[38]. Both the long duration, difficulty of control and high CFR are likely due to the difficult context for the outbreak response. This part of DRC has a history of violence and mistrust of the central government and international organisations. This makes all parts of the response from finding cases, tracking contacts, ensuring early treatment and safe burials much more difficult. Based on the models investigated in this paper it seem like the large increase of cases seen from March - June 2019 is over and that



(a) Centrality of PIT values for the Katwa model (b) Centrality of PIT values for the Beni model



(c) Predictions Katwa model



(d) Predictions in Beni

Figure 5: Centrality for Katwa and Beni

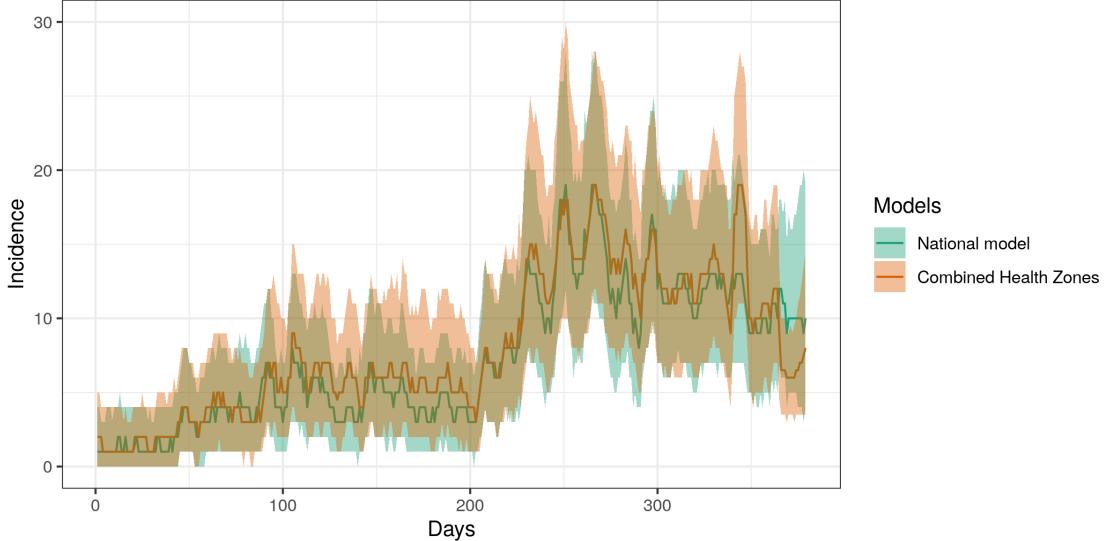


Figure 6: Median and 95% prediction intervals for the national model and for a model calculated by summing up models for all the health zones

there is hope that the outbreak can be controlled, but the large uncertainty of the situation which is reflected in the large uncertainty in the models make long term predictions very difficult if not impossible.

Compared to estimates of the reproduction number in the West-African Ebola outbreak[39] the reproduction number in the DRC outbreak has seen larger variations on short time-scales and less longer patterns. In the previous outbreak we saw in both Guinea and Liberia that the reproduction number peaked around 2 after a few months followed by a long slow decline. This outbreak has been more characterised in multiple short term peaks or waves of higher reproduction number with the peaks reaching over 3 at some points. It is possible that the different patterns from the different outbreaks can tell us about the response. The West-African response was maybe slower to get properly started, but once it did the response started working and was fairly uninterrupted. In the current outbreak the response was faster, but potentially due to the security situation the response has been interrupted and this could be why we see the short term peaks. The range of reproductive values seen is also similar to other previous Ebola outbreaks [40].

We find that on the national level the models with a semi-local trend for the reproduc-

tion number were calibrated out to about 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] where a semi-mechanistic model based on a SEIR that included reporting delays. That model was semi-mechanistic as the transmission rate was allowed to vary as a random walk. While direct comparisons are not possible due to the different outbreaks it is 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 [41, 42]. They use very different modelling techniques and a proper comparison between the models would require comparing proper scoring rules score using the same data, but in [42] they use similar forecasting horizons of around two weeks.

The best model for the national outbreak is uncalibrated when forecasting more than 9 days forward due to overestimating 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 due to the forecasting of the reproduction number. This model allows large variation in the reproduction number over fairly short time, but does not have any information about what range of reproduction numbers are actually feasible. In this outbreak we see fairly fast changes of the reproduction number, but the reproduction number almost never is larger than 3. A model for reproduction number than allows the fast variation and short peaks, but that keeps the reproduction number within a more narrow band, might give a better model.

A probabilistic modelling approach with tests for calibration and using proper scoring

rules to see which forecast is better is a crucial step to compare forecasts and to evaluate their validity. The proper scoring rules are mainly useful when comparing models, 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. Once the forecasting horizon is longer than 9 days we can not really trust that the model adequately predicts the uncertainty in the process. Even a well calibrated model can make large forecasting mistakes if the underlying process changes in a way not seen before during the time period used to test the model. Therefore it is always crucial to critically evaluate models and the metrics used to evaluate the models. We find that both the CRPS and DSS scores give similar performance, but that calibration using the PIT values was especially valuable. The introduction of the centrality measure of the PIT histograms give 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 a measure that also includes other departures from normality 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 frame does not adequately describe the transmission process. We found a relatively large value for the k parameter which gives a fairly modest increase in variance. Much smaller than found when investigating transmission chains in the west Africa ebola outbreak [30]. It is possible that variance is smaller in this outbreak, but this should be taken as fairly weak evidence of that and a proper investigation of transmission chains would be needed.

We successfully could provide forecasts for all health-zones with more than one cases, but the quality was very varying. For some health zones where there were very few cases and then 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. For almost all health zones the models with semi-local trend were better than the models with constant reproduction number.

The models investigated in this paper are all lightweight enough to easily run on a normal laptop in at timescale of minutes to allow real-time forecasts during an outbreak.

The models and forecasts in this report have multiple limitations. Firstly, the data source for the model is the cumulative number of cases by reporting day which we use to calculate the incidence. This data source has many disadvantages to for example having a line list of all cases. A line list would give a much better idea of the data of onset of disease, 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 to 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. The approaches in this report could still give reasonable forecasts.

The model in it's current form has several limitations. As previously discussed it seems to predict to much variance in the reproduction number which leads to over-dispersed predictions. There are also clear limitations on the health-zone level and that is that spread between health zones is not modelled. We only model the spread of the disease after it has been introduced to a health zone. For 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 he model structure it would be fairly straight-forward to incorporated 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 geographic 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 or if available data on inter health-zone travel could be used. Without this

spread term the model can not 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, so we need at least 17 days of data in a health zone to be able to forecast.

In addition to the spatial spread, it would be beneficial to combine the separate parts of the model together into one generative model. This means including the estimation of the reproduction number into the main model. 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, uncertainty about serial interval and could even include imputation for missing data.

In this report we used a semi-mechanistic model with fairly few assumptions about the 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 certain basic facts about how disease spread that are likely to be fairly correct for most diseases. The model used in this report 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 stay the same. The benefit of adding more structure to a model, for example an SEIR structure is that it is easier to model the results of interventions like vaccination ². One key aspect to consider when choosing models for an outbreak situation is that maybe more than one model should be used during the outbreak. At the start of the outbreak when the important mechanisms are less well known, maybe a model with less mechanistic structure is the most useful and then once we learn more about the outbreak

²Using the fact that $R_t = R_o * s(t)$ in a compartmental model of the SEIR type, it would likely be possible to model some simple interventions in the models in these papers too

we can add more structure to the model. In many forecasting challenges it has been shown that an ensemble of models perform better than individual models [43], this was also found in the RAPPID modelling competition for Ebola [23].

Theoretically it should be possible to forecast infectious disease outbreaks on the timescale of individual outbreaks [44]. 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 many additional sources of data exists that are not used in models, this includes deaths vs cases, community deaths vs 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 we could significantly increase forecasting horizon where we could provide calibrated predictions.

5 Conclusions

We found that a probabilistic forecasting model based on a branching process with a Poisson offspring distribution with a time-varying reproduction number forecasted with a semi-local trend gave well calibrated forecasts for the 2018–2019 Ebola outbreak in DRC for forecasting horizons up to 9 days. This model together with a similar model but with a negative binomial offspring distribution could also give calibrated forecasts for some of the individual health zones, but for many health zones the models had too much uncertainty.

References

- [1] H. Heesterbeek, R. M. Anderson, V. Andreasen, S. Bansal, D. D. Angelis, C. Dye, K. T. D. Eames, W. J. Edmunds, S. D. W. Frost, S. Funk, T. D. Hollingsworth, T. House, V. Isham, P. Klepac, J. Lessler, J. O. Lloyd-Smith, C. J. E. Metcalf, D. Mollison, L. Pellis, J. R. C. Pulliam, M. G. Roberts, C. Viboud, and I. N. I. I. Collaboration, “Modeling infectious disease dynamics in the complex landscape of global health,” *Science*, vol. 347, p. aaa4339, Mar. 2015.
- [2] C. Rivers, J.-P. Chretien, S. Riley, J. A. Pavlin, A. Woodward, D. Brett-Major, I. M. Berry, L. Morton, R. G. Jarman, M. Biggerstaff, M. A. Johansson, N. G. Reich, D. Meyer, M. R. Snyder, and S. Pollett, “Using outbreak science to strengthen the use of models during epidemics,” *Nature Communications*, vol. 10, pp. 1–3, July 2019.
- [3] J.-P. Chretien, S. Riley, and D. B. George, “Mathematical modeling of the West Africa Ebola epidemic,” *eLife*, vol. 4.
- [4] W. E. R. Team, “Ebola Virus Disease in West Africa The First 9 Months of the Epidemic and Forward Projections,” *New England Journal of Medicine*, vol. 371, pp. 1481–1495, Oct. 2014.
- [5] A. Camacho, A. Kucharski, Y. Aki-Sawyerr, M. A. White, S. Flasche, M. Baguelin, T. Pollington, J. R. Carney, R. Glover, E. Smout, A. Tiffany, W. J. Edmunds, and S. Funk, “Temporal Changes in Ebola Transmission in Sierra Leone and Implications for Control Requirements: a Real-time Modelling Study,” *PLoS Currents*, vol. 7, Feb. 2015.
- [6] P.-Y. Kobres, J. P. Chretien, M. A. Johansson, J. Morgan, P.-Y. Whung, H. Mukundan, S. D. Valle, B. M. Forshey, T. M. Quadelacy, 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*, p. 634832, May 2019.

- [7] J.-P. Chretien, D. George, J. Shaman, R. A. Chitale, and F. E. McKenzie, “Influenza Forecasting in Human Populations: A Scoping Review,” *PLOS ONE*, vol. 9, p. e94130, Apr. 2014.
- [8] S. Funk, A. Camacho, A. J. Kucharski, R. Lowe, R. M. Eggo, and W. J. 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*, vol. 15, p. e1006785, Feb. 2019.
- [9] W. Wei and L. Held, “Calibration tests for count data,” *TEST*, vol. 23, pp. 787–805, Dec. 2014.
- [10] T. Gneiting, “Editorial: Probabilistic forecasting,” *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 171, no. 2, pp. 319–321, 2008.
- [11] T. Gneiting, F. Balabdaoui, and A. E. Raftery, “Probabilistic forecasts, calibration and sharpness,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 69, no. 2, pp. 243–268, 2007.
- [12] C. Czado, T. Gneiting, and L. Held, “Predictive Model Assessment for Count Data,” *Biometrics*, vol. 65, no. 4, pp. 1254–1261, 2009.
- [13] World Health Organisation, “Ebola virus disease fact sheet.”
- [14] J. Brainard, L. Hooper, K. Pond, K. Edmunds, and P. R. Hunter, “Risk factors for transmission of Ebola or Marburg virus disease: a systematic review and meta-analysis,” *International Journal of Epidemiology*, vol. 45, pp. 102–116, Feb. 2016.
- [15] World Health Organization, “Ebola Outbreak in DRC 01: 7 August 2018,” Aug. 2018.

- [16] World Health Organisation, “Ebola Outbreak in DRC 09: 04 October 2018,” Oct. 2018.
- [17] World Health Organization, “Ebola Outbreak in DRC 16: 21 November 2018,” Nov. 2018.
- [18] World Health Organization, “Ebola Outbreak in DRC 38: 24 April 2019,” Apr. 2019.
- [19] W. H. 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.,” tech. rep., Apr. 2019.
- [20] 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,” Aug. 2019.
- [21] World Health Organization, “Ebola Outbreak in DRC 51: 23 July 2019,” July 2019.
- [22] World Health Organization, “Ebola Outbreak in DRC 58: 10 September 2019,” July 2019.
- [23] C. Viboud, K. Sun, R. Gaffey, M. Ajelli, L. Fumanelli, S. Merler, Q. Zhang, G. Chowell, L. Simonsen, and A. Vespignani, “The RAPIDD Ebola Forecasting Challenge: Synthesis and Lessons Learnt,” *Epidemics*, vol. 22, pp. 13–21, Mar. 2018.
- [24] A. A. King, M. Domenech de Cells, F. M. G. Magpantay, and P. Rohani, “Avoidable errors in the modelling of outbreaks of emerging pathogens, with special reference to Ebola,” *Proceedings of the Royal Society B: Biological Sciences*, vol. 282, p. 20150347, May 2015.
- [25] Humanitarian Data Exchange, “Ebola Cases and Deaths in the North Kivu Ebola Outbreak in the Democratic Republic of the Congo (DRC) - Humanitarian Data Exchange.”

- [26] C. Jacob, “Branching Processes: Their Role in Epidemiology,” *International Journal of Environmental Research and Public Health*, vol. 7, pp. 1186–1204, Mar. 2010.
- [27] A. Cori, N. M. Ferguson, C. Fraser, and S. Cauchemez, “A new framework and software to estimate time-varying reproduction numbers during epidemics,” *American Journal of Epidemiology*, vol. 178, pp. 1505–1512, Nov. 2013.
- [28] 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*, vol. 438, p. 355, Nov. 2005.
- [29] P. Nouvellet, A. Cori, T. Garske, I. M. Blake, I. Dorigatti, W. Hinsley, T. Jombart, H. L. Mills, G. Nedjati-Gilani, M. D. Van Kerkhove, C. Fraser, C. A. Donnelly, N. M. Ferguson, and S. Riley, “A simple approach to measure transmissibility and forecast incidence,” *Epidemics*, vol. 22, pp. 29–35, Mar. 2018.
- [30] International Ebola Response Team, J. Agua-Agum, A. Ariyarajah, B. Aylward, L. Bawo, P. Biliogui, I. M. Blake, R. J. Brennan, A. Cawthorne, E. Cleary, P. Clement, R. Conteh, A. Cori, F. Dafae, B. Dahl, J.-M. Dangou, B. Diallo, C. A. Donnelly, I. Dorigatti, C. Dye, T. Eckmanns, M. Fallah, N. M. Ferguson, L. Fiebig, C. Fraser, T. Garske, L. Gonzalez, E. Hamblion, N. Hamid, S. Hersey, W. Hinsley, A. Jambei, T. Jombart, D. Kargbo, S. Keita, M. Kinzer, F. K. George, B. Godefroy, G. Gutierrez, N. Kannan-garage, H. L. Mills, T. Moller, S. Meijers, Y. Mohamed, O. Morgan, G. Nedjati-Gilani, E. Newton, P. Nouvellet, T. Nyenswah, W. Pereira, D. Perkins, S. Riley, G. Rodier, M. Rondy, M. Sagrado, C. Savulescu, I. J. Schafer, D. Schumacher, T. Seyler, A. Shah, M. D. Van Kerkhove, C. S. Wesseh, and Z. Yoti, “Exposure Patterns Driving Ebola Transmission in West Africa: A Retrospective Observational Study,” *PLoS medicine*, vol. 13, p. e1002170, Nov. 2016.

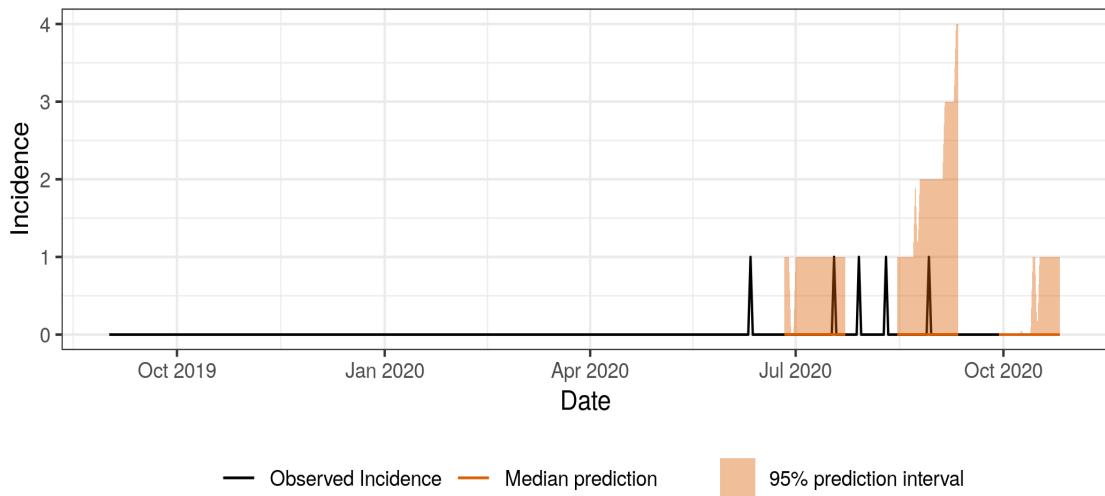
- [31] R. N. Thompson, J. E. Stockwin, R. D. van Gaalen, J. A. Polonsky, Z. N. Kamvar, P. A. Demarsh, E. Dahlqvist, S. Li, E. Miguel, T. Jombart, J. Lessler, S. Cauchemez, and A. Cori, “Improved inference of time-varying reproduction numbers during infectious disease outbreaks,” *Epidemics*, p. 100356, Aug. 2019.
- [32] A. Cori, *EpiEstim: EpiEstim: a package to estimate time varying reproduction numbers from epidemic curves*. 2013.
- [33] S. L. Scott, *bsts: Bayesian Structural Time Series*. 2019.
- [34] C. Czado, T. Gneiting, and L. Held, “Predictive Model Assessment for Count Data,” *Biometrics*, vol. 65, no. 4, pp. 1254–1261, 2009.
- [35] J. Faraway, G. Marsaglia, J. Marsaglia, and A. Baddeley, *goftest: Classical Goodness-of-Fit Tests for Univariate Distributions*. 2017.
- [36] A. Jordan, F. Krueger, and S. Lerch, “Evaluating Probabilistic Forecasts with scoringRules,” *Journal of Statistical Software*, 2018.
- [37] R Core Team, *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, 2018.
- [38] A. M. Rojek, A. Salam, R. J. Ragotte, E. Liddiard, A. Elhussain, A. Carlqvist, M. Butler, N. Kayem, L. Castle, L. Odondi, K. Stepniewska, and P. 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*, vol. 0, July 2019.
- [39] “West African Ebola Epidemic after One Year Slowing but Not Yet under Control,” *New England Journal of Medicine*, vol. 372, pp. 584–587, Feb. 2015.
- [40] J. Legrand, R. F. Grais, P. Y. Boelle, A. J. Valleron, and A. Flahault, “Understanding the dynamics of Ebola epidemics,” *Epidemiology & Infection*, vol. 135, pp. 610–621, May 2007.

- [41] J. D. Kelly, J. Park, R. J. Harrigan, N. A. Hoff, S. D. Lee, R. Wannier, B. Selo, M. Mossoko, B. Njoloko, E. Okitolonda-Wemakoy, P. Mbala-Kingebeni, G. W. Rutherford, T. B. Smith, S. Ahuka-Mundeke, J. J. Muyembe-Tamfum, A. W. Rimoin, and F. P. 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*, p. 100354, July 2019.
- [42] A. R. Akhmetzhanov, H. Lee, S.-m. Jung, T. Kayano, B. Yuan, and H. Nishiura, “Analyzing and forecasting the Ebola incidence in North Kivu, the Democratic Republic of the Congo from 201819 in real time,” *Epidemics*, vol. 27, pp. 123–131, June 2019.
- [43] T. G. Dietterich, “Ensemble Methods in Machine Learning,” in *Multiple Classifier Systems*, Lecture Notes in Computer Science, pp. 1–15, Springer Berlin Heidelberg, 2000.
- [44] S. V. Scarpino and G. Petri, “On the predictability of infectious disease outbreaks,” *Nature Communications*, vol. 10, pp. 1–8, Feb. 2019.

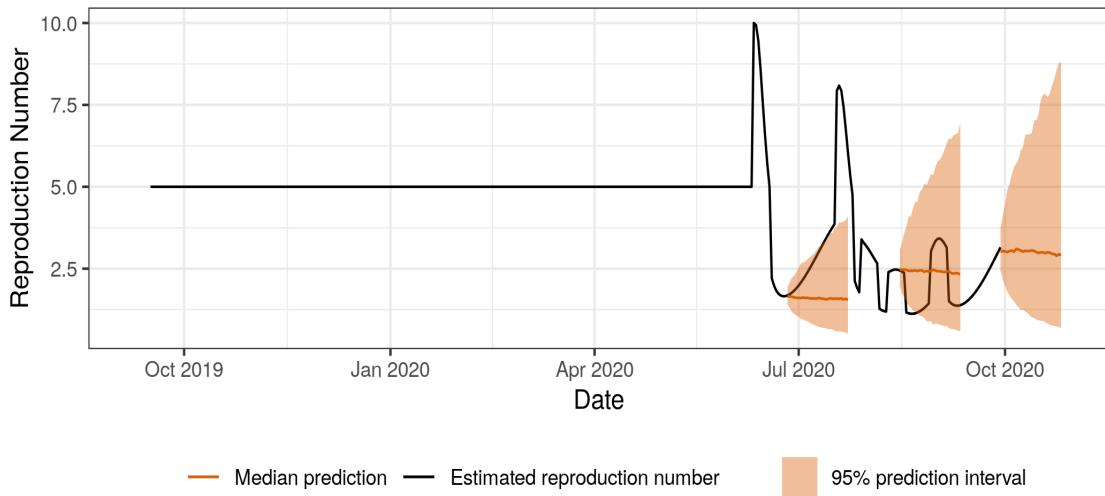
Appendices

This appendix shows the best fitting model and evaluation score for each health zone.

A Alimbongo



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



(b) Forecasted and predicted reproduction numbers for the semilocal poisson model

Figure 7: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the semilocal poisson model for Alimbongo.

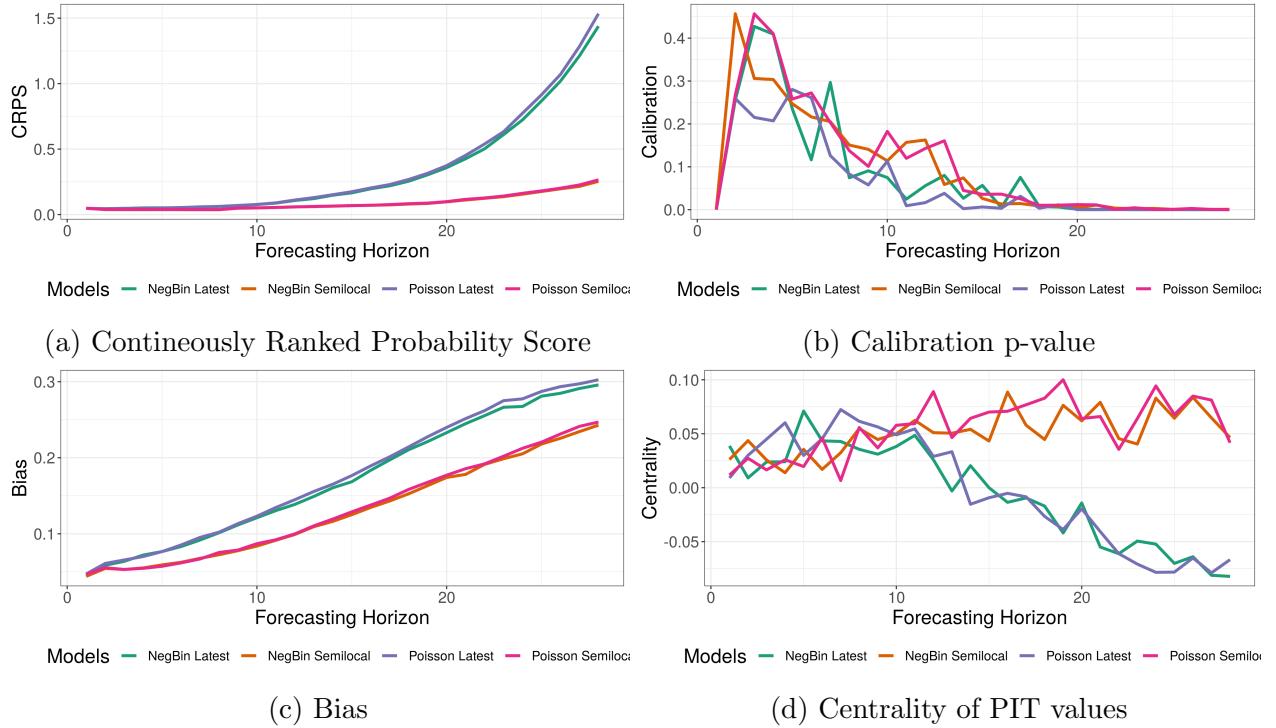
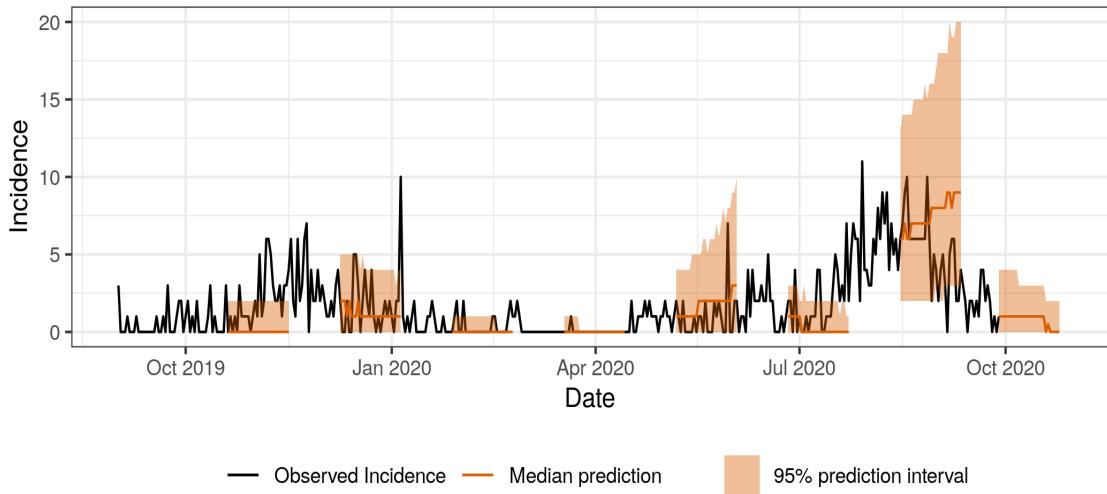
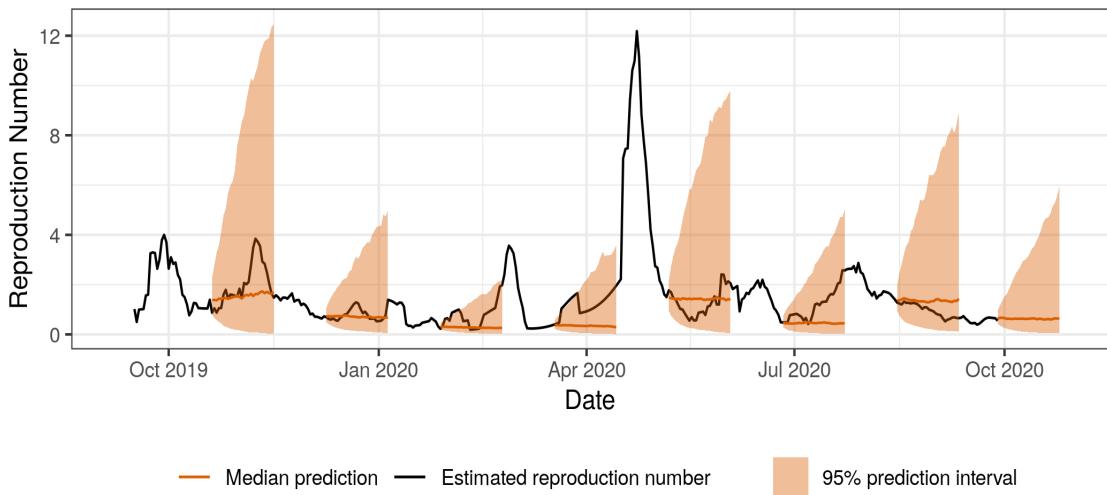


Figure 8: Scores for Alimbongo as a function of the forecasting horizon.

B Beni



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



(b) Forecasted and predicted reproduction numbers for the semilocal poisson model

Figure 9: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the semilocal poisson model for Beni.

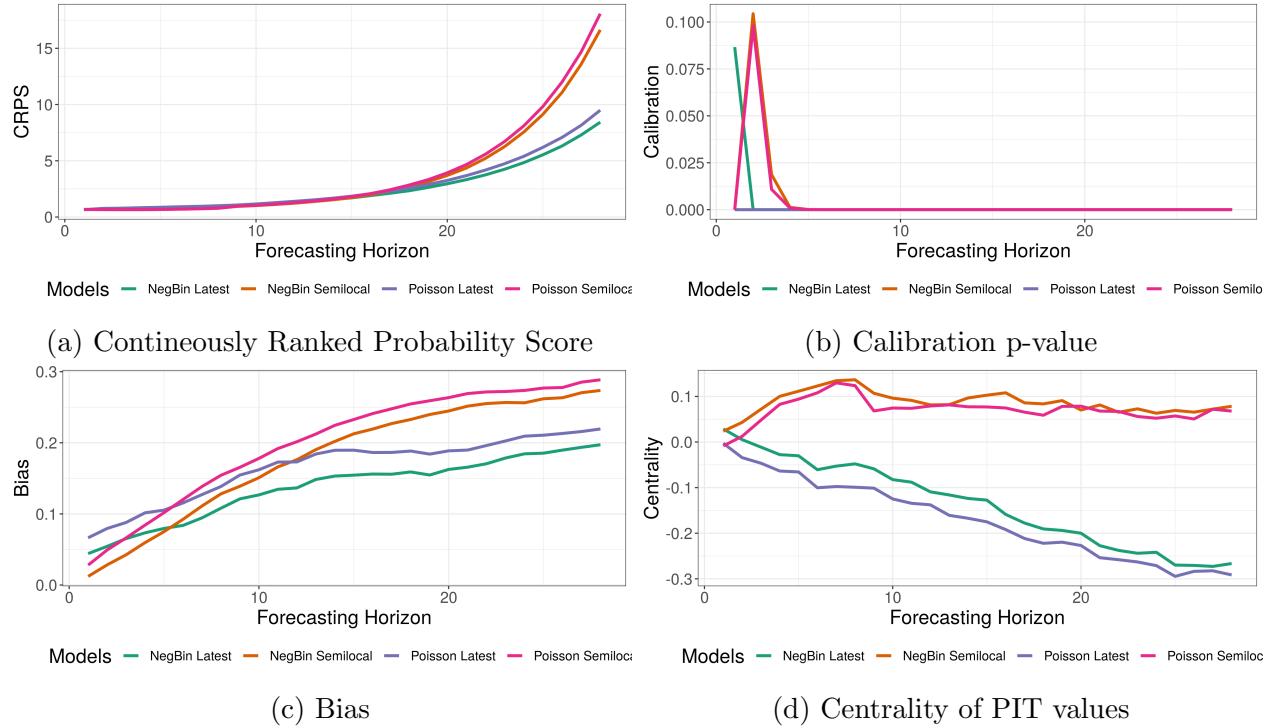
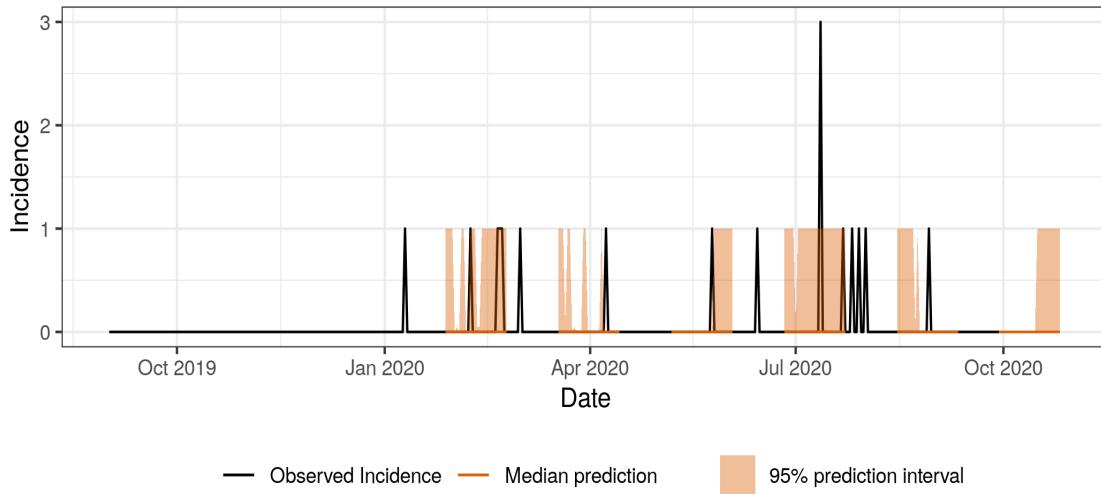
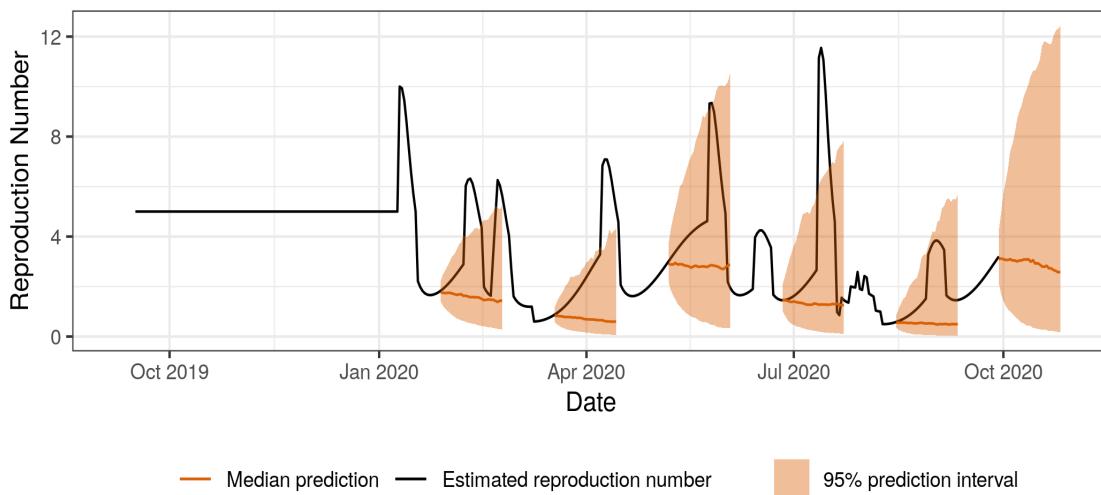


Figure 10: Scores for Beni as a function of the forecasting horizon.

C Biena



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



(b) Forecasted and predicted reproduction numbers for the semilocal poisson model

Figure 11: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the semilocal poisson model for Biena.

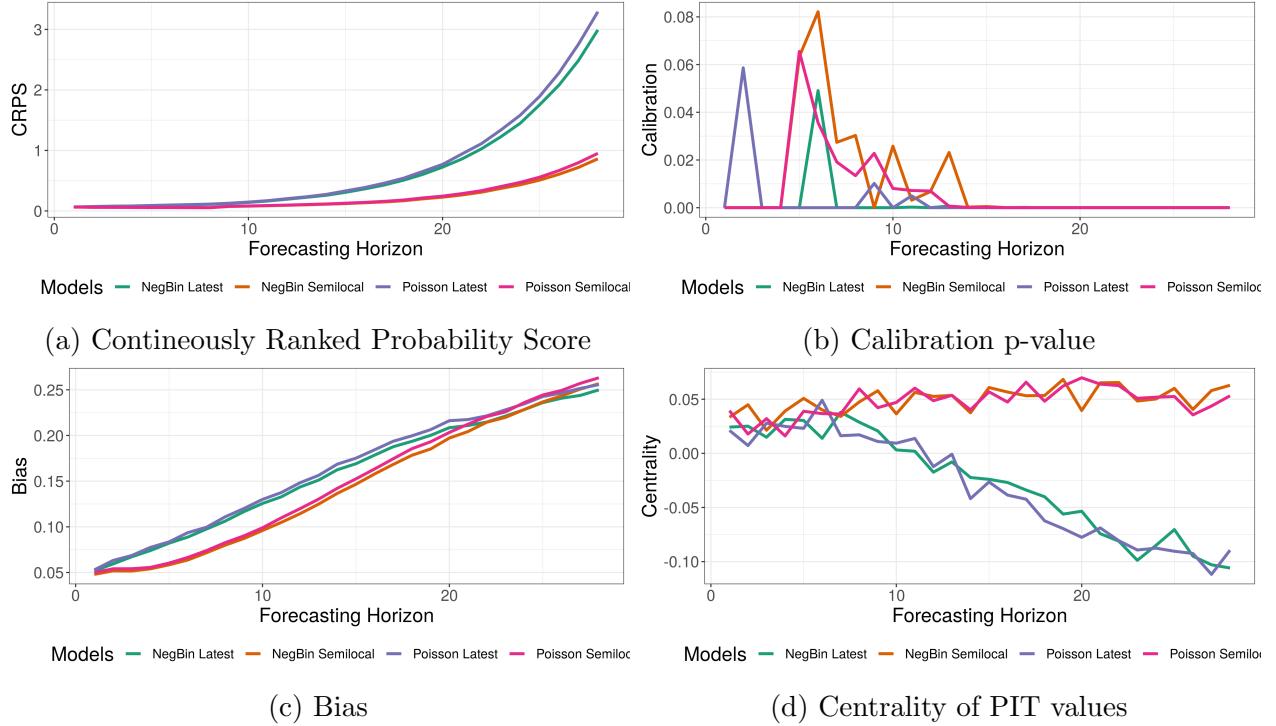
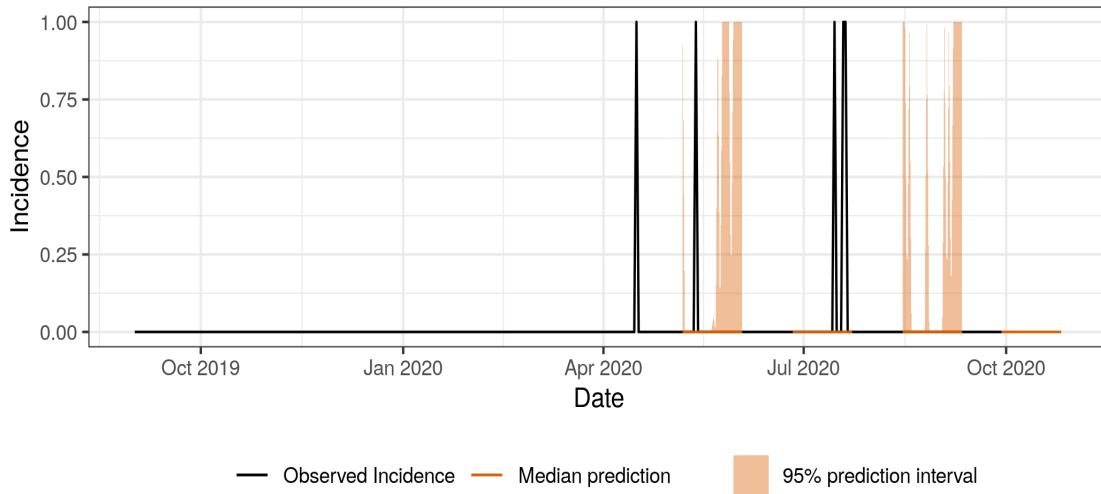
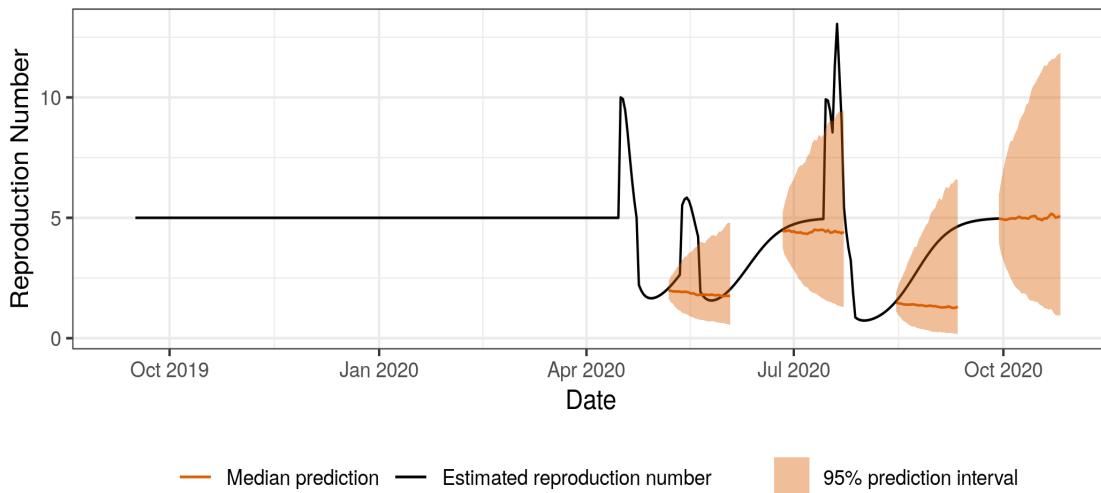


Figure 12: Scores for Biena as a function of the forecasting horizon.

D Bunia



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



(b) Forecasted and predicted reproduction numbers for the semilocal poisson model

Figure 13: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the semilocal poisson model for Bunia.

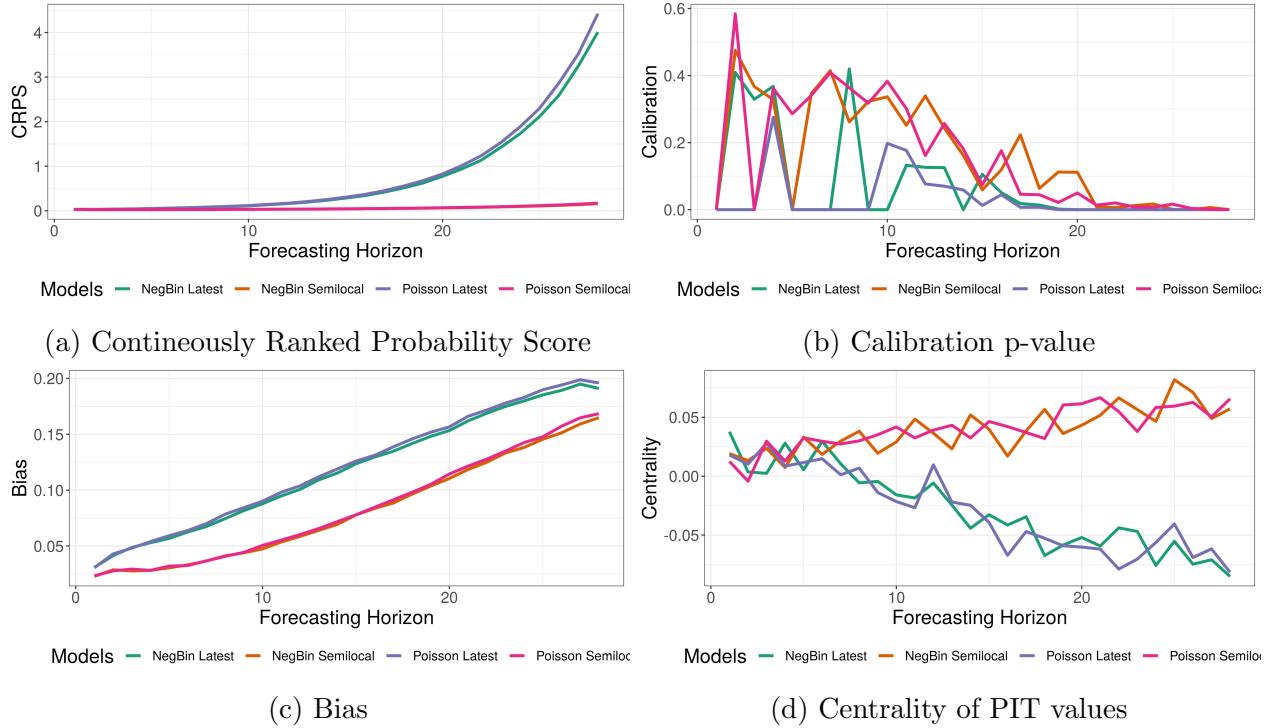
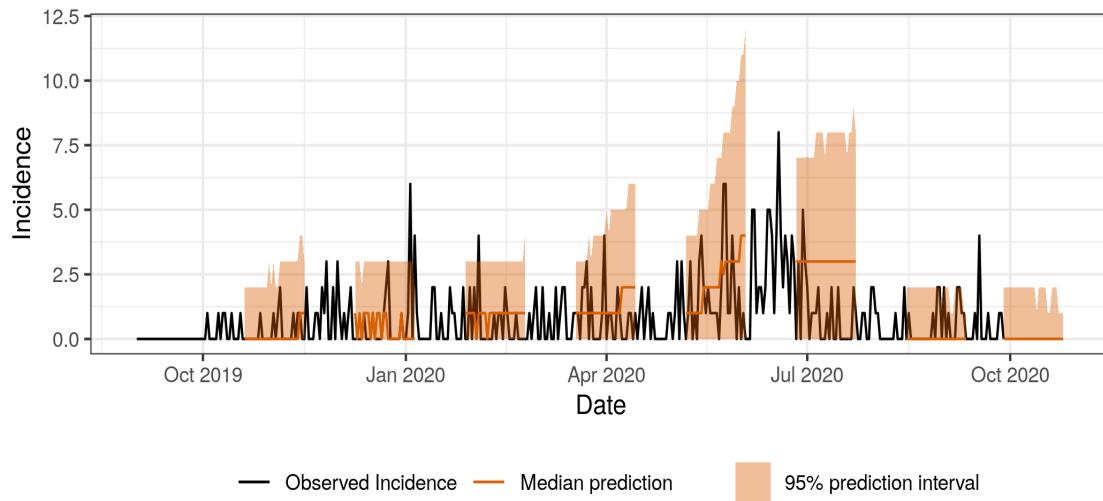
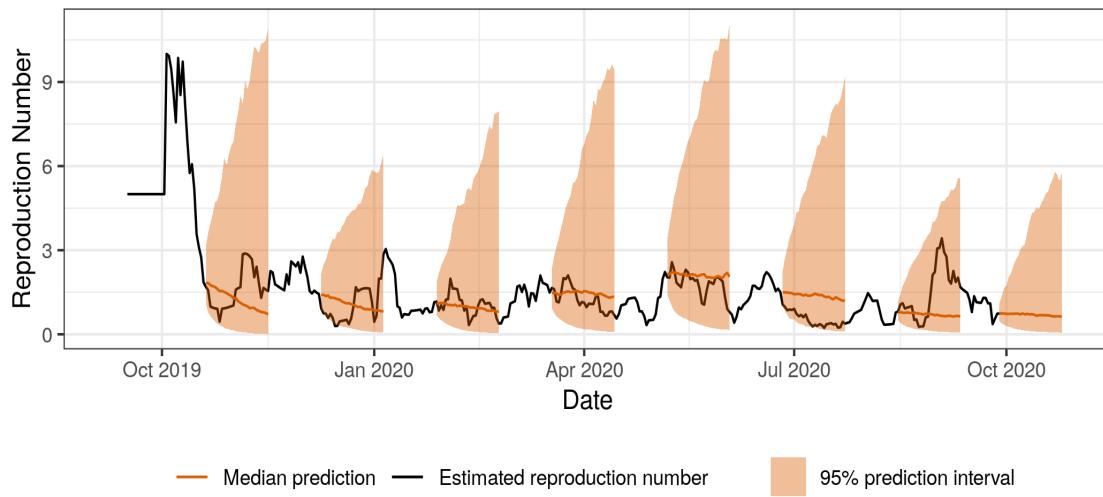


Figure 14: Scores for Bunia as a function of the forecasting horizon.

E Butembo



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



(b) Forecasted and predicted reproduction numbers for the semilocal poisson model

Figure 15: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the semilocal poisson model for Butembo.

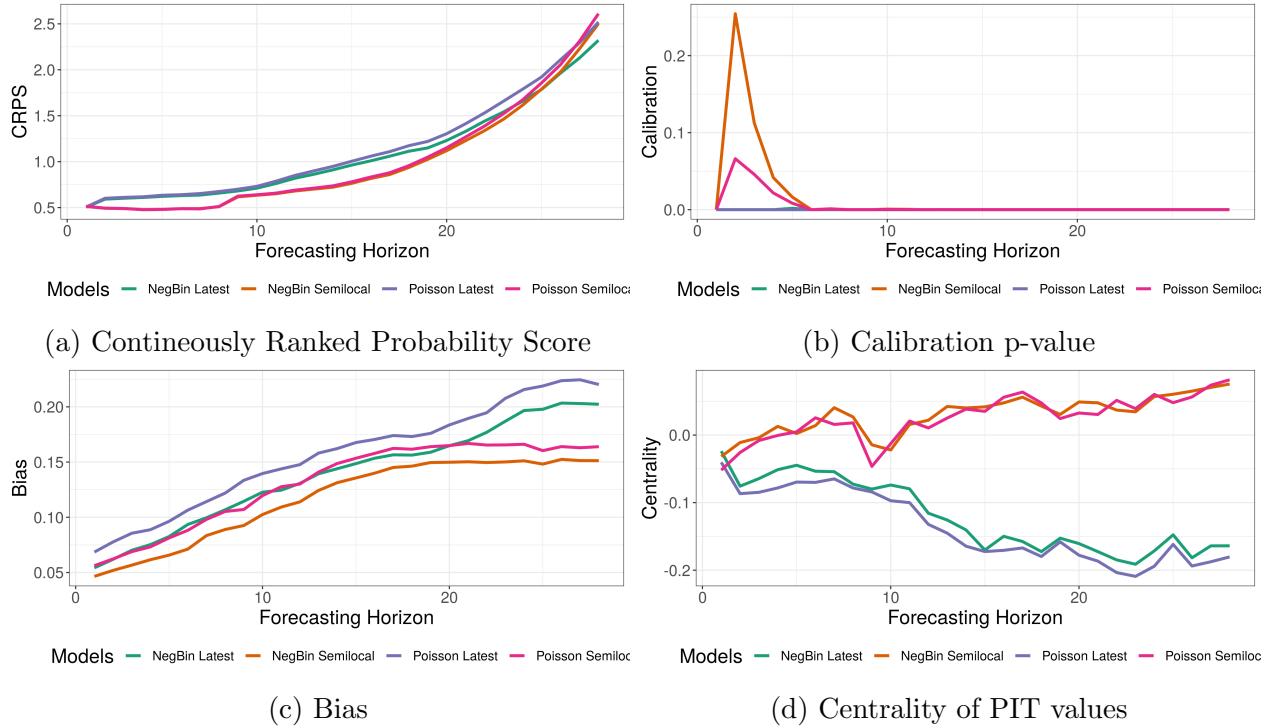
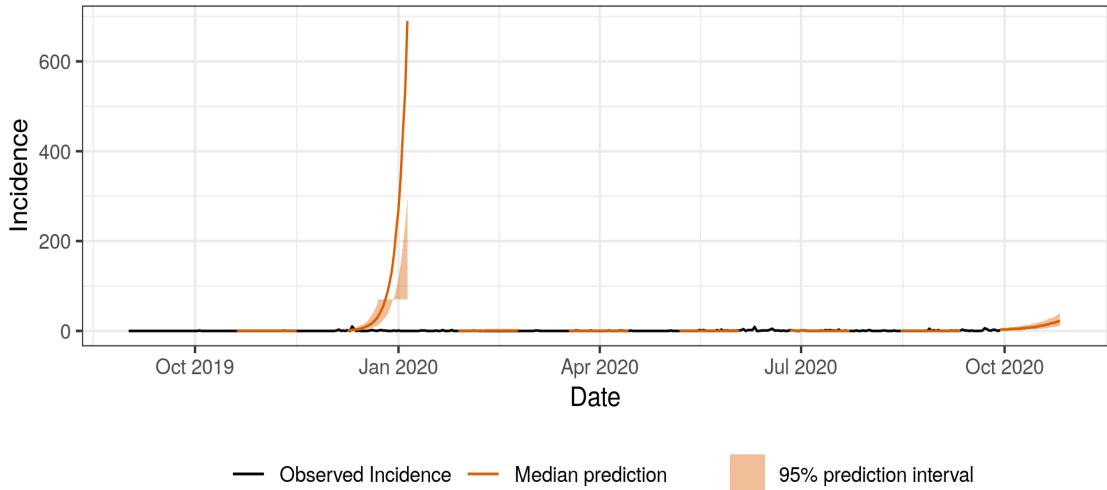
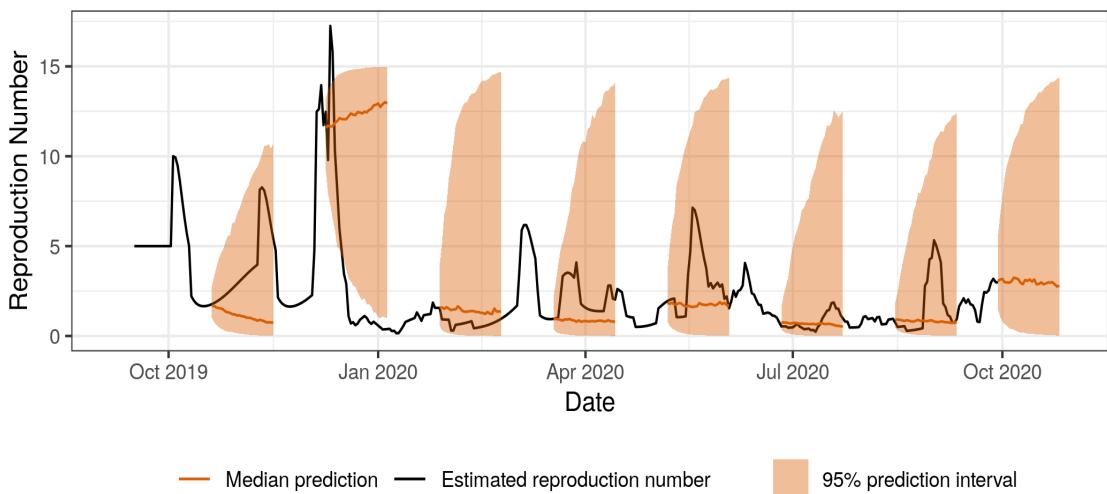


Figure 16: Scores for Butembo as a function of the forecasting horizon.

F Kalunguta



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



(b) Forecasted and predicted reproduction numbers for the semilocal poisson model

Figure 17: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the semilocal poisson model for Kalunguta.

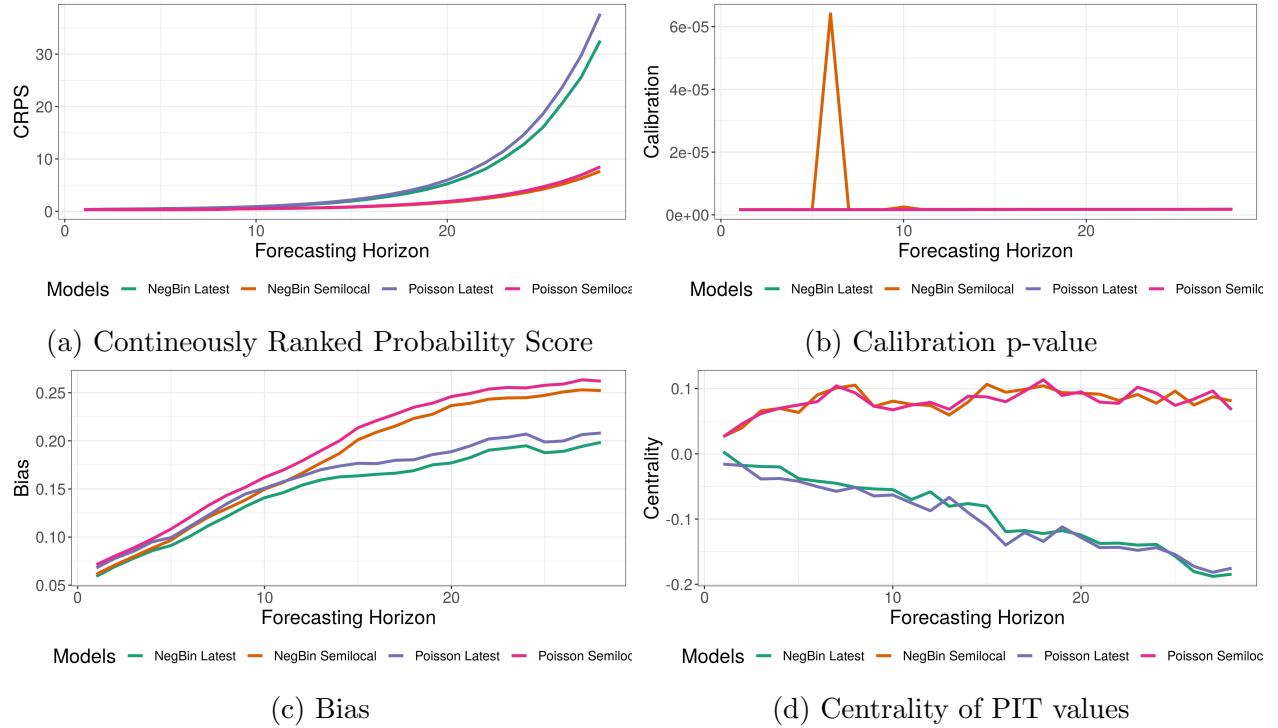
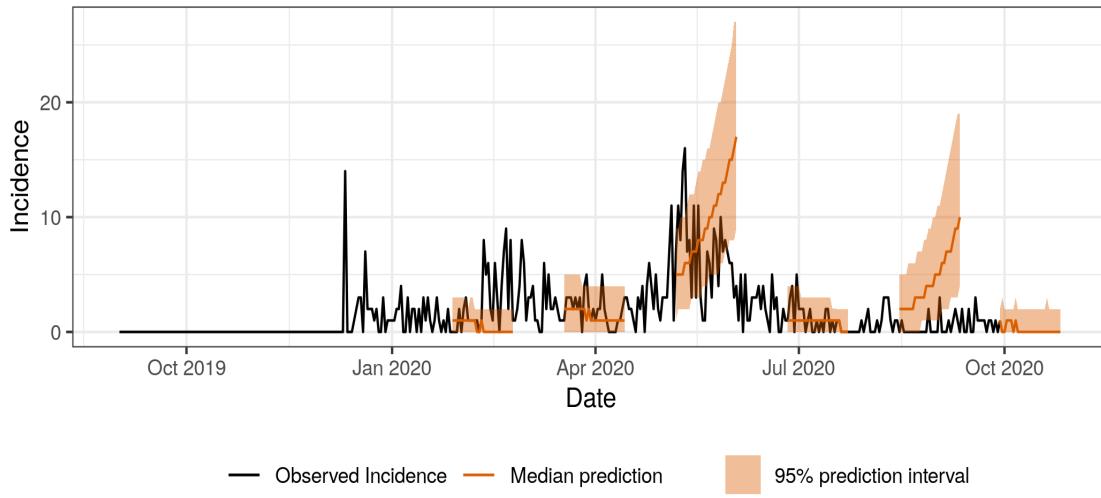
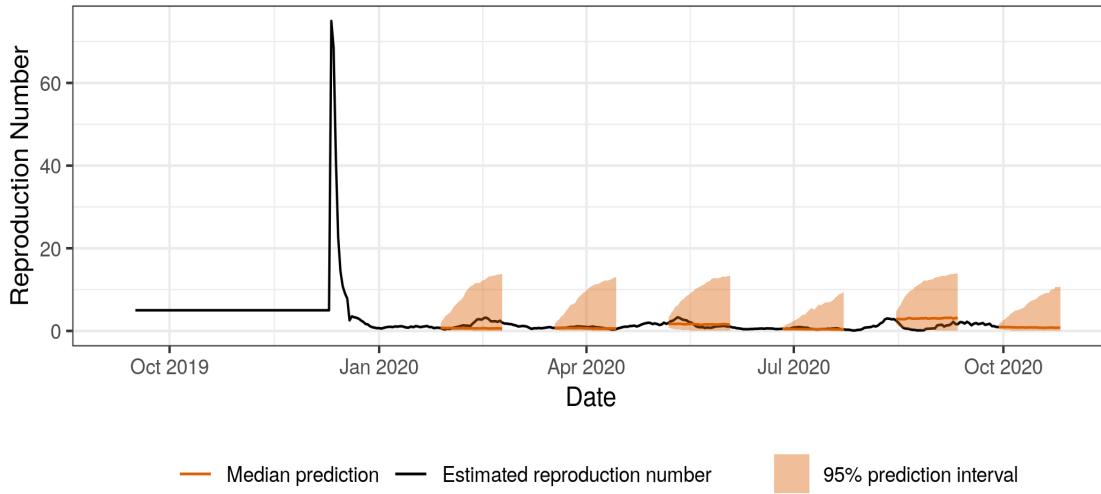


Figure 18: Scores for Kalunguta as a function of the forecasting horizon.

G Katwa



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



(b) Forecasted and predicted reproduction numbers for the semilocal poisson model

Figure 19: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the semilocal poisson model for Katwa.

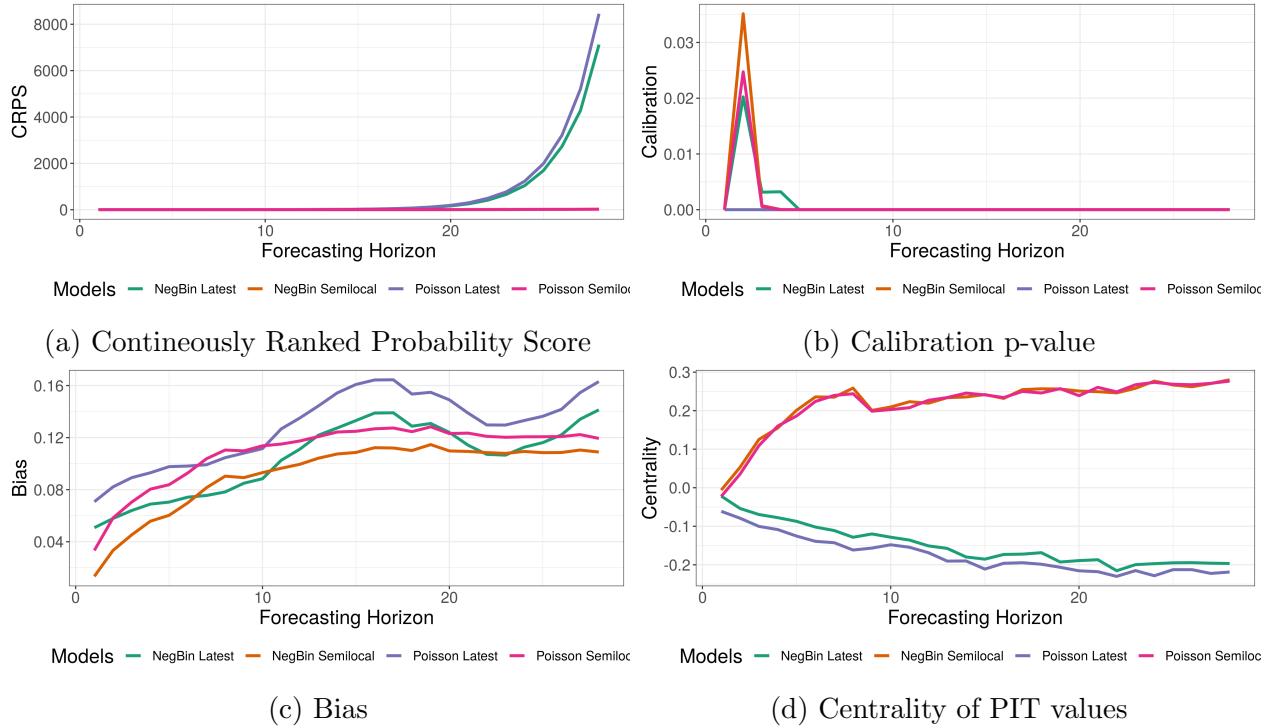
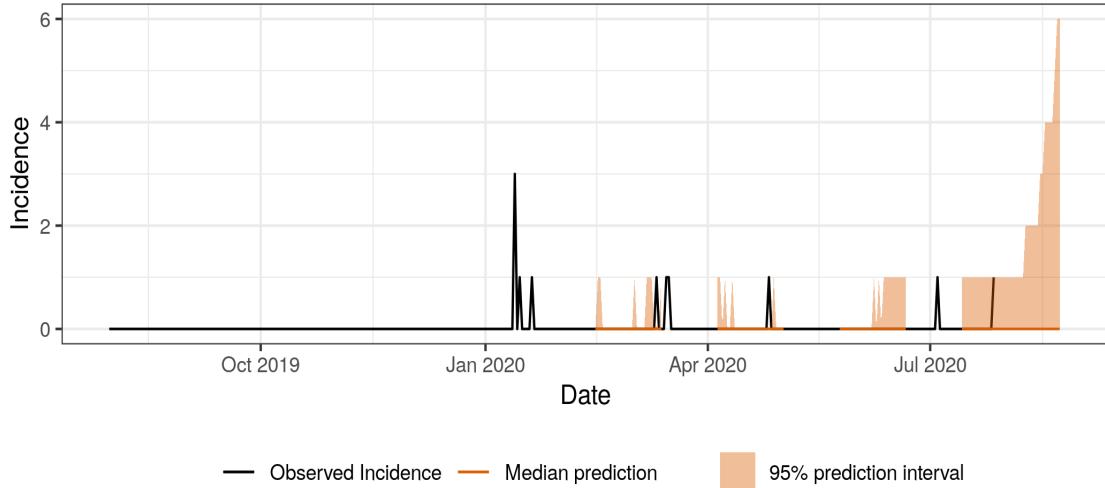
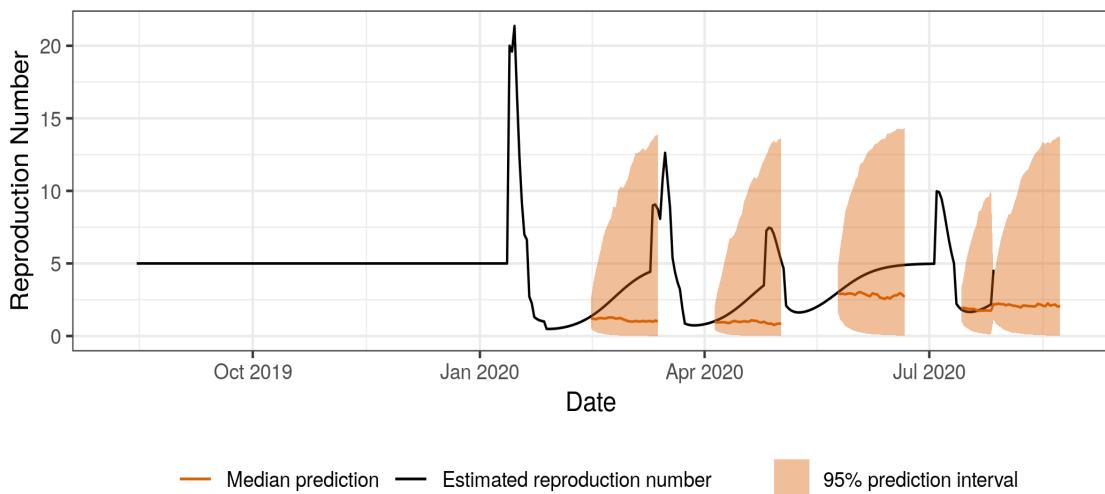


Figure 20: Scores for Katwa as a function of the forecasting horizon.

H Kayina



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



(b) Forecasted and predicted reproduction numbers for the semilocal poisson model

Figure 21: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the semilocal poisson model for Kayina.

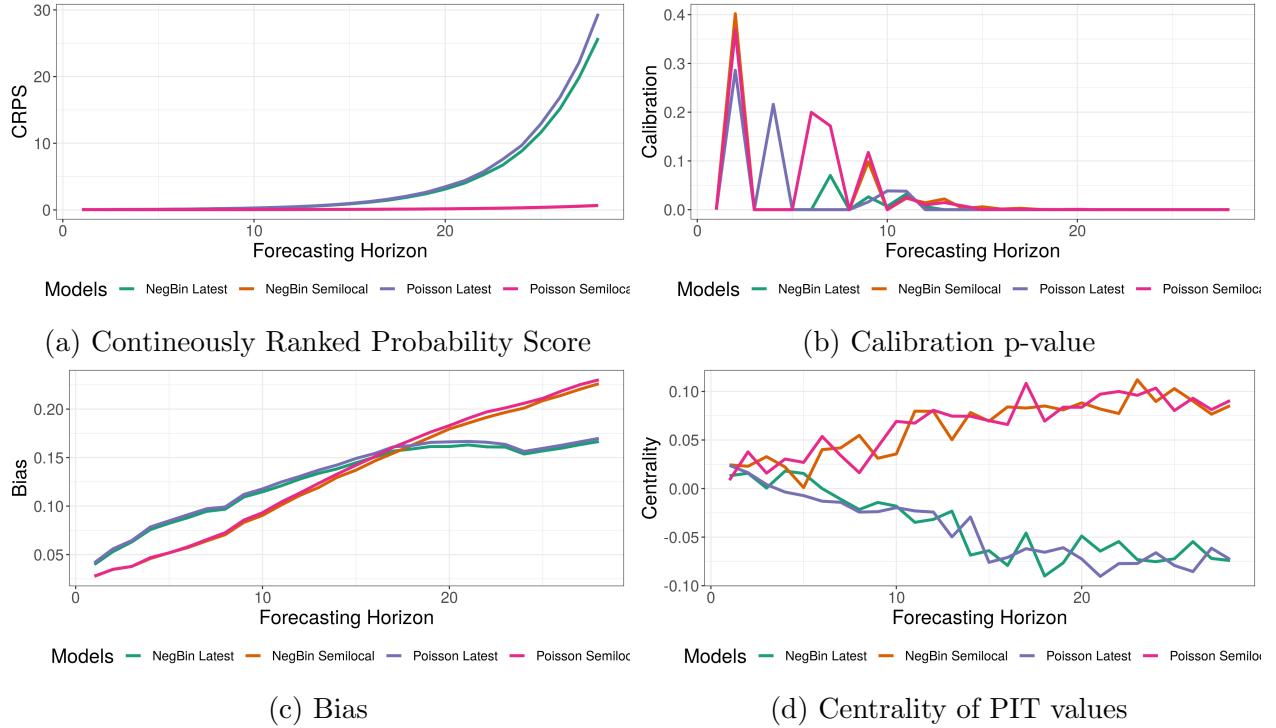
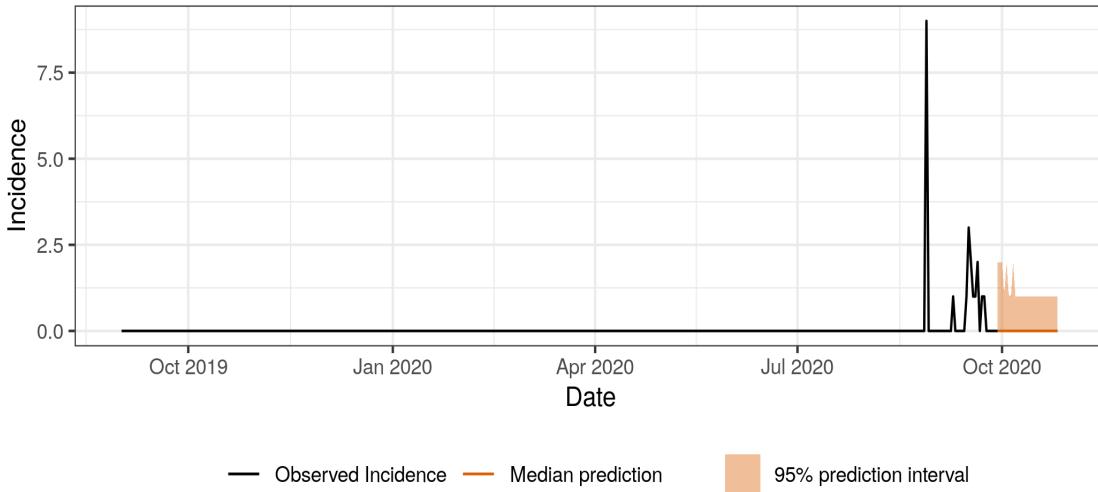
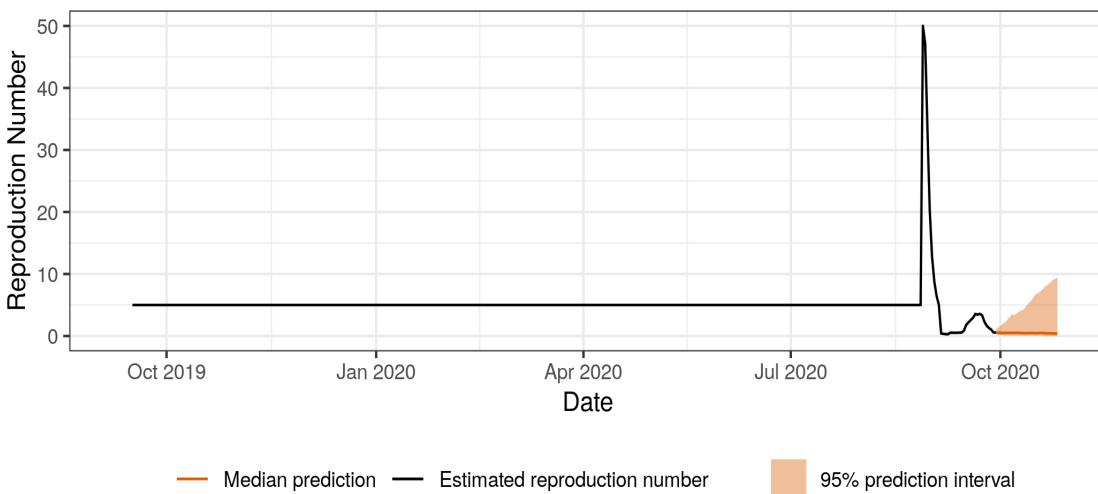


Figure 22: Scores for Kayina as a function of the forecasting horizon.

I Kayna



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



(b) Forecasted and predicted reproduction numbers for the semilocal poisson model

Figure 23: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the semilocal poisson model for Kayna.

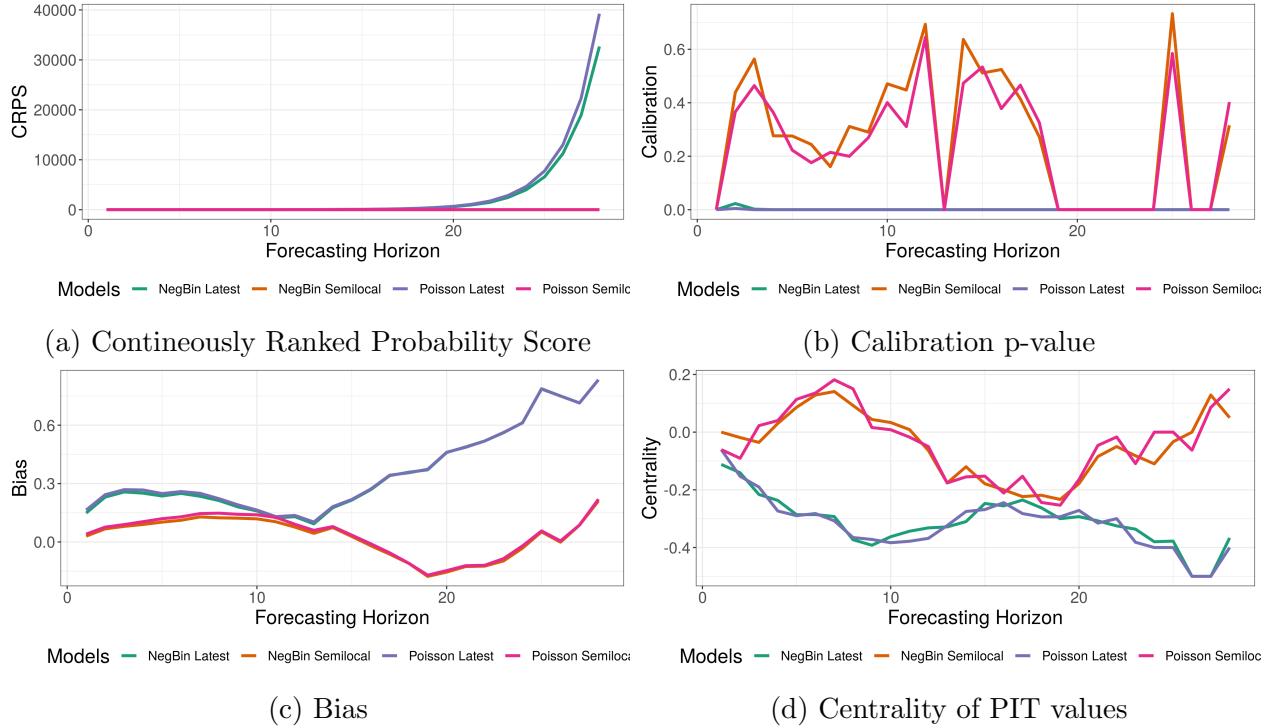
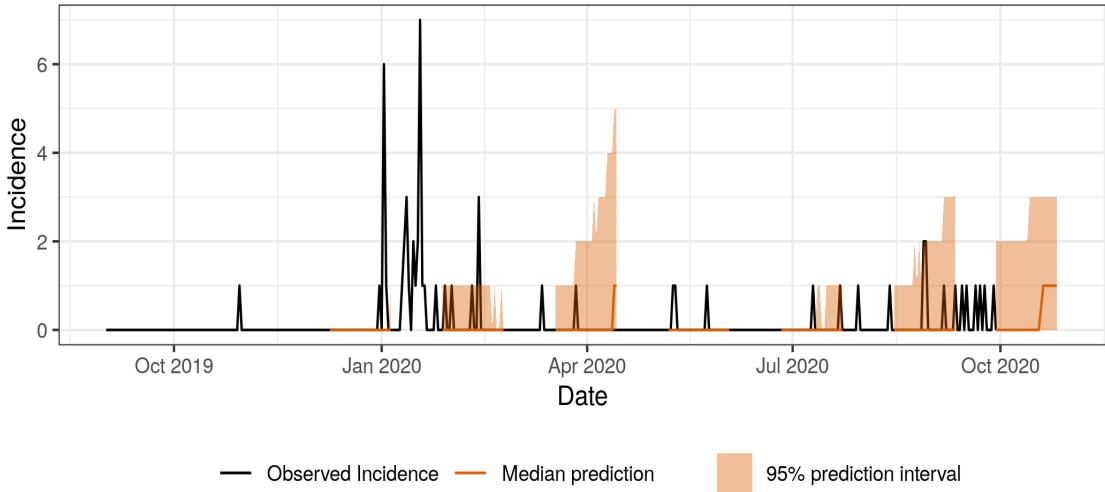
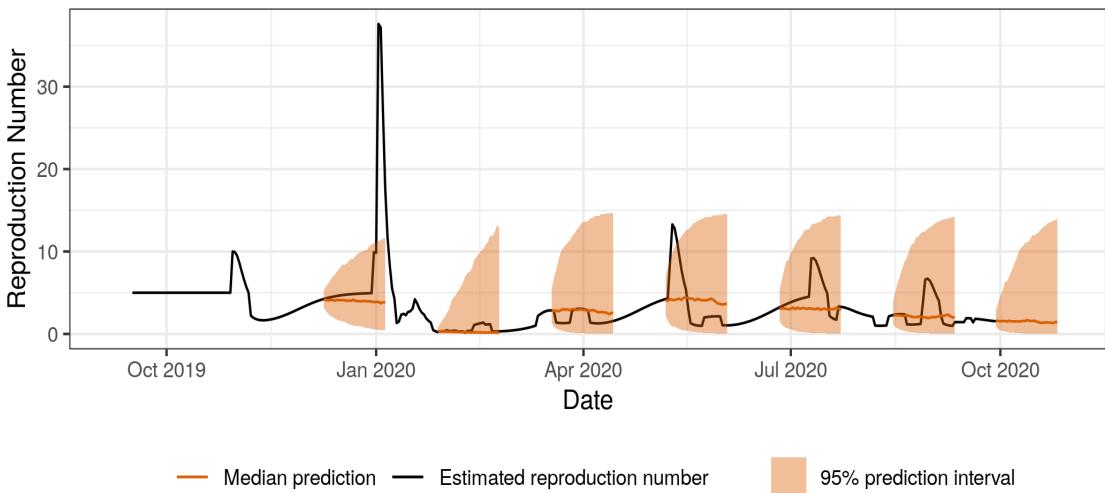


Figure 24: Scores for Kayna as a function of the forecasting horizon.

J Komanda



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



(b) Forecasted and predicted reproduction numbers for the semilocal poisson model

Figure 25: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the semilocal poisson model for Komanda.

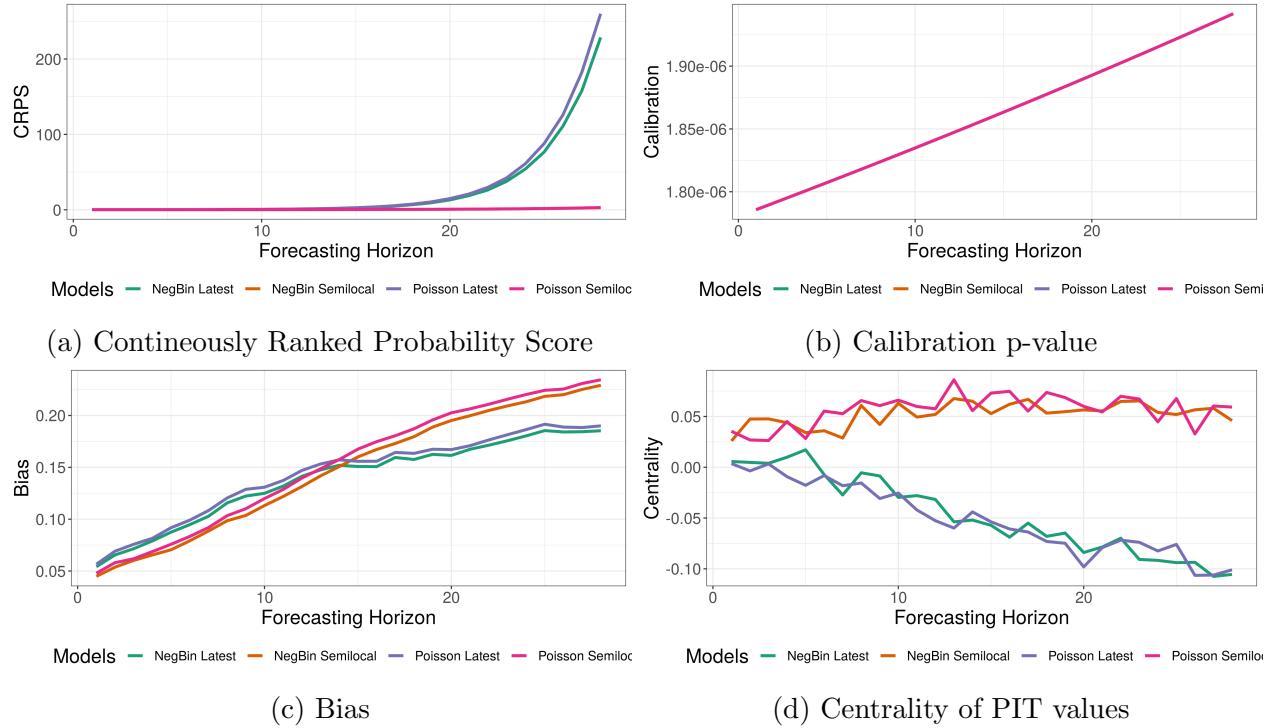
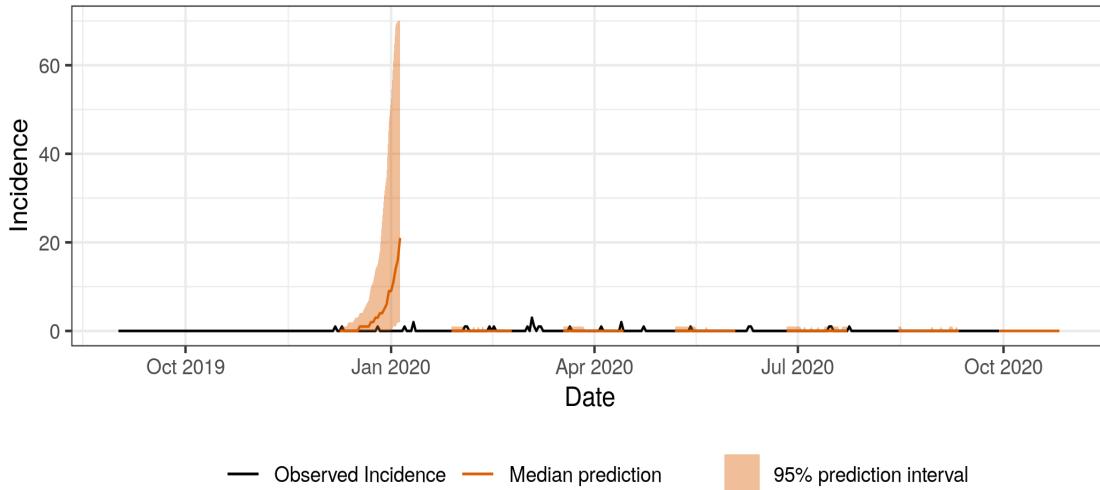
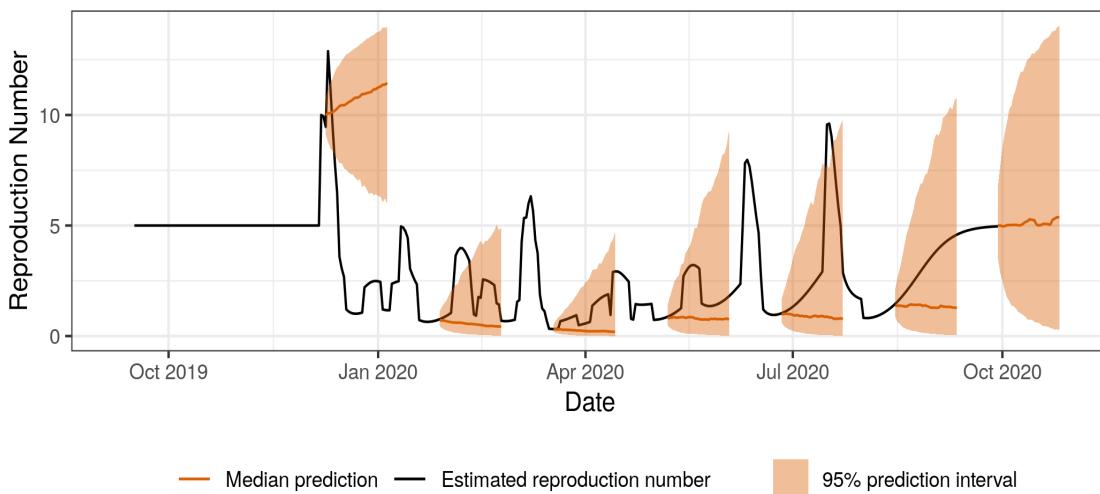


Figure 26: Scores for Komanda as a function of the forecasting horizon.

K Kyondo



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



(b) Forecasted and predicted reproduction numbers for the semilocal poisson model

Figure 27: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the semilocal poisson model for Kyondo.

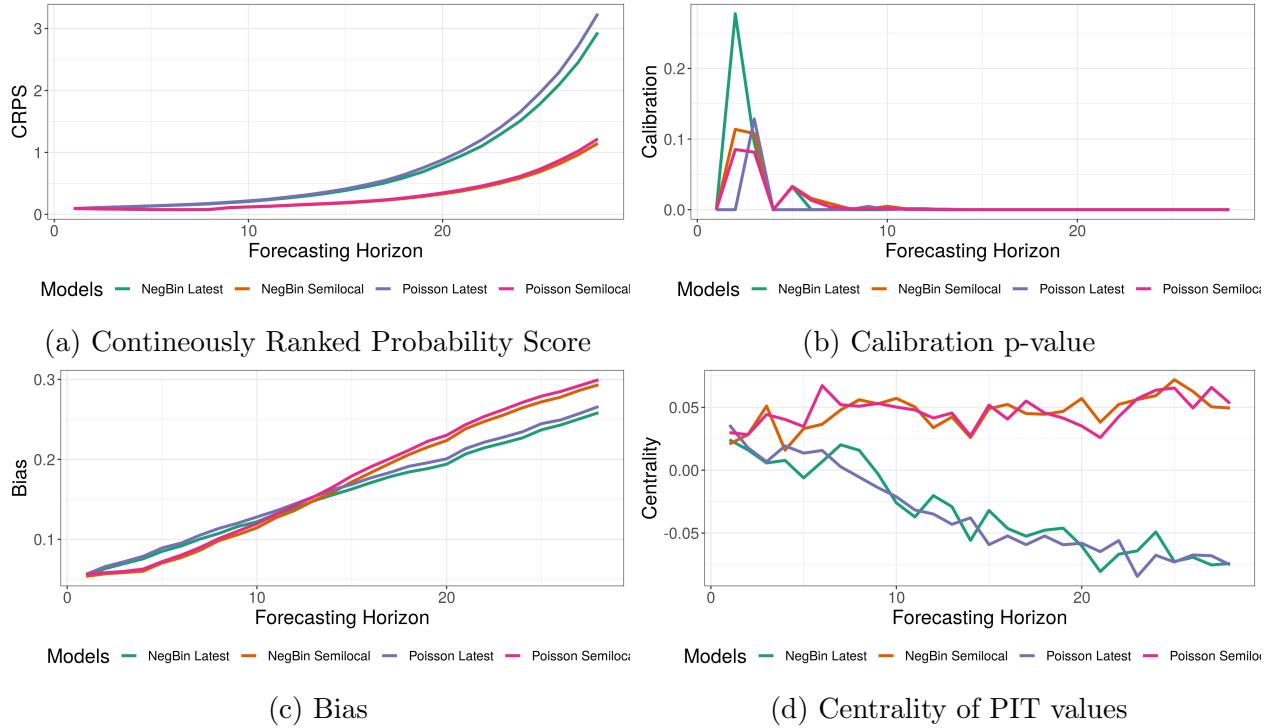
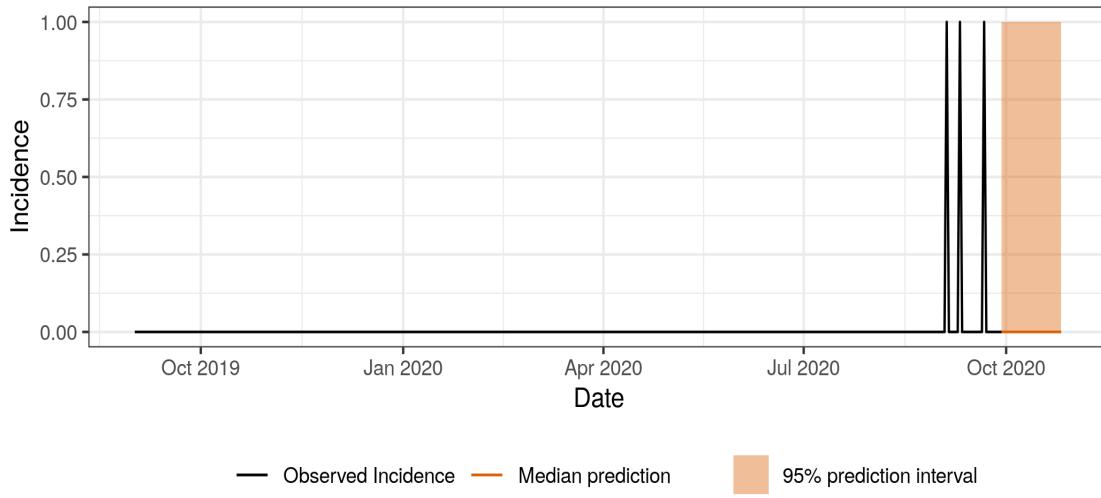
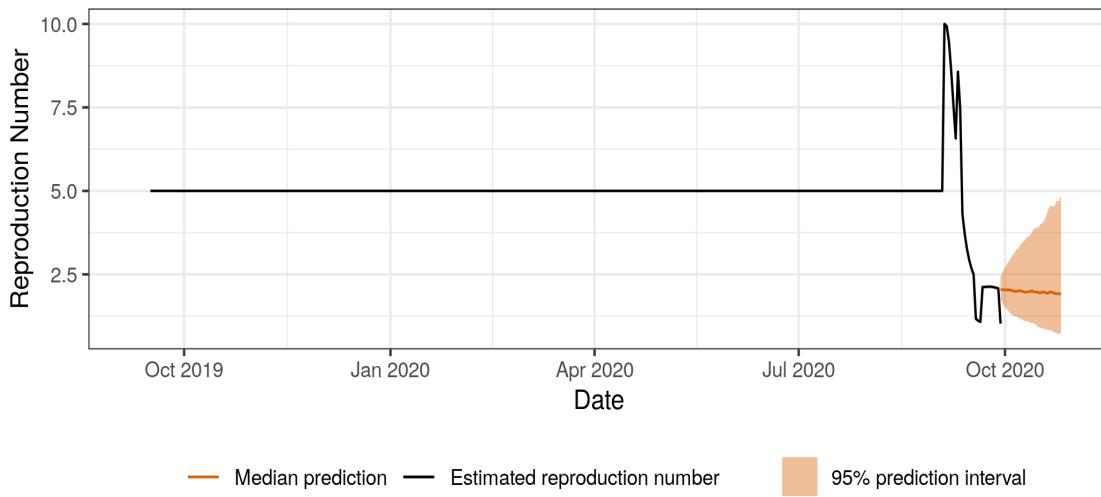


Figure 28: Scores for Kyondo as a function of the forecasting horizon.

L Lolwa



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



(b) Forecasted and predicted reproduction numbers for the semilocal poisson model

Figure 29: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the semilocal poisson model for Lolwa.

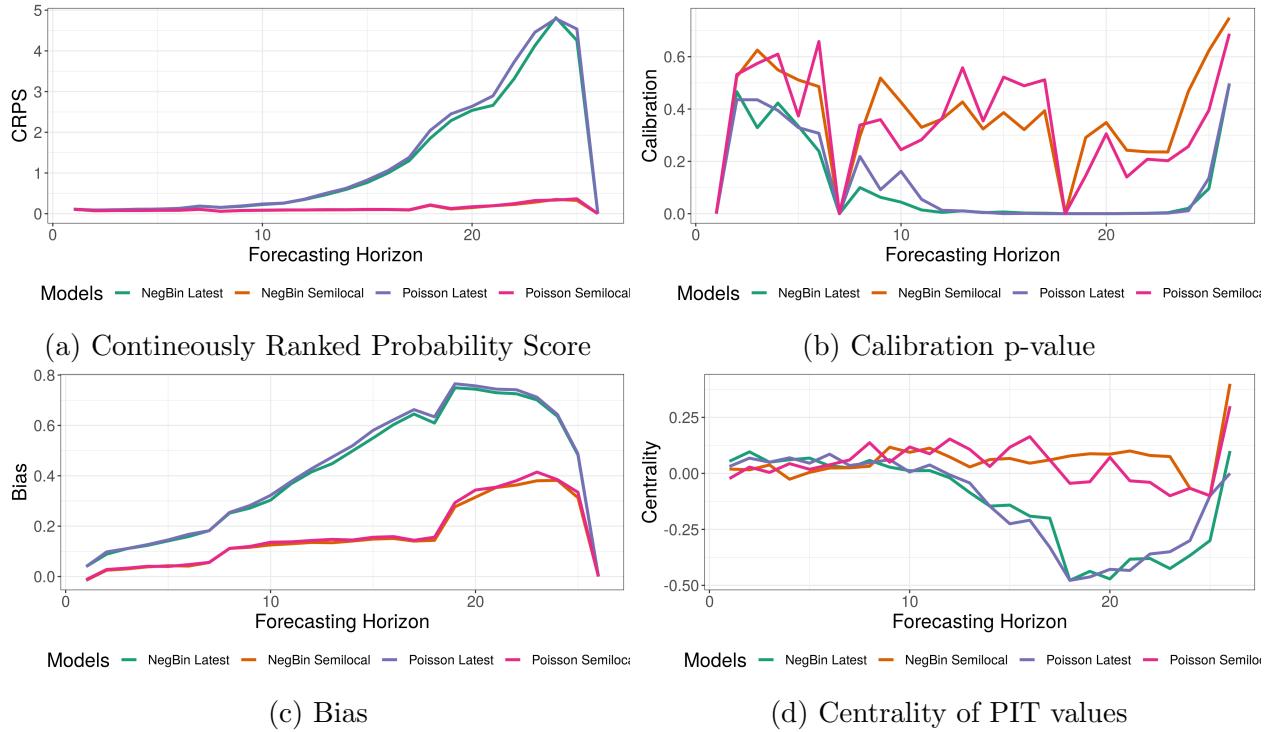
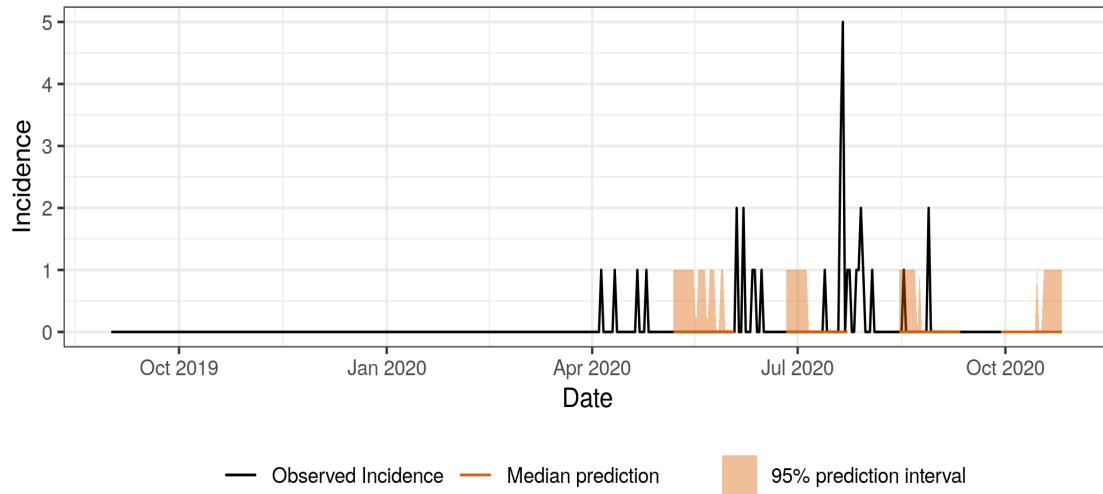
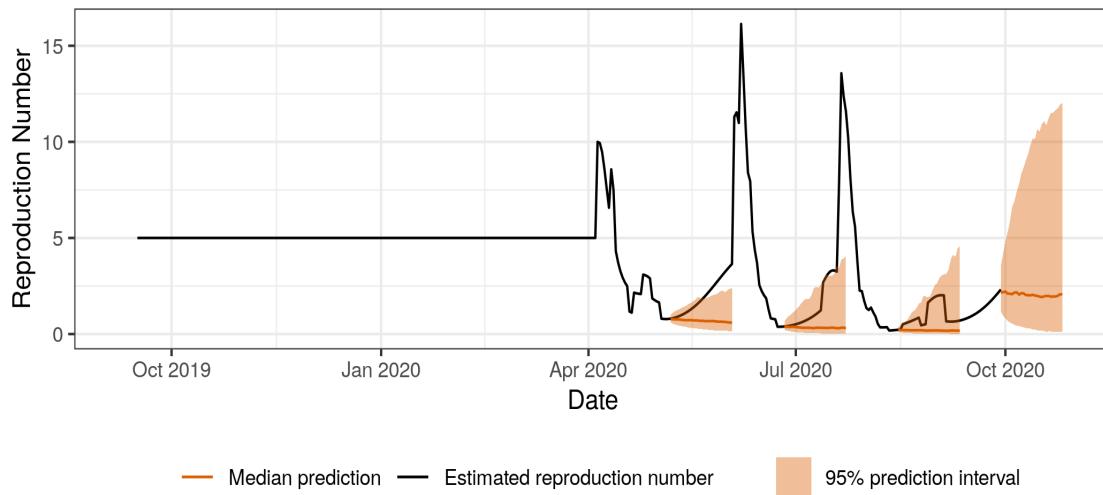


Figure 30: Scores for Lolwa as a function of the forecasting horizon.

M Lubero



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



(b) Forecasted and predicted reproduction numbers for the semilocal poisson model

Figure 31: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the semilocal poisson model for Lubero.

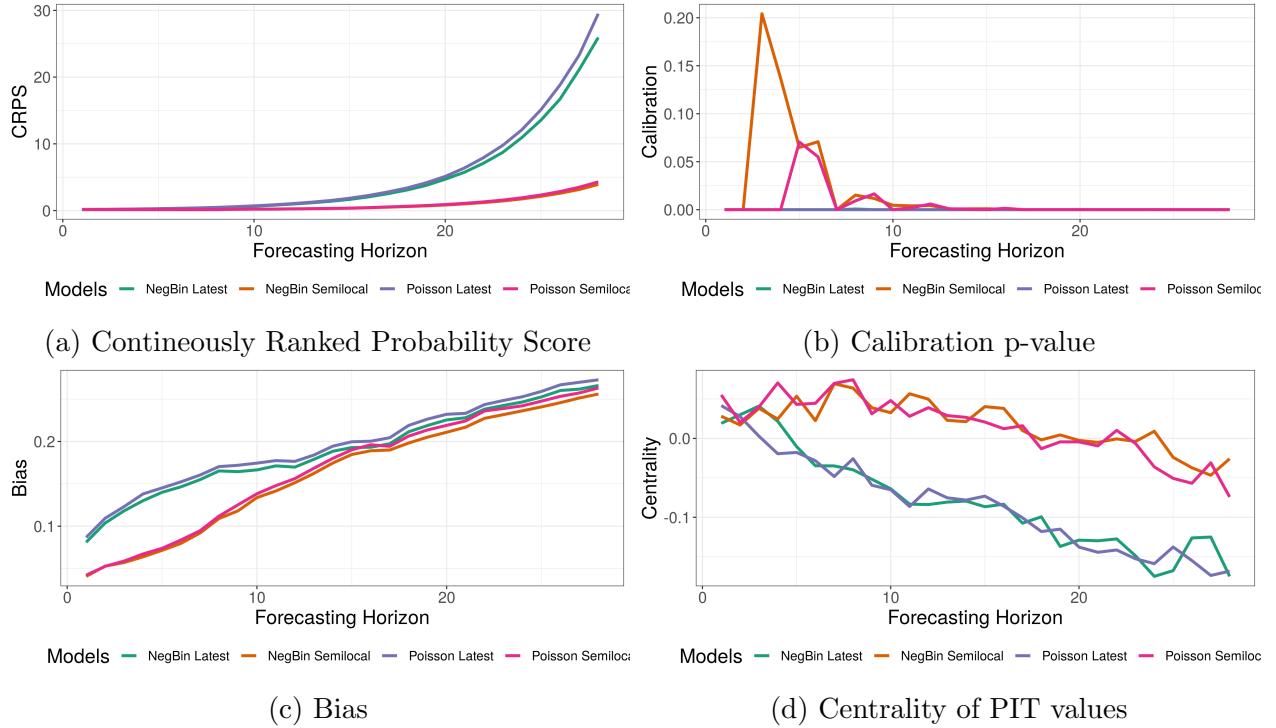
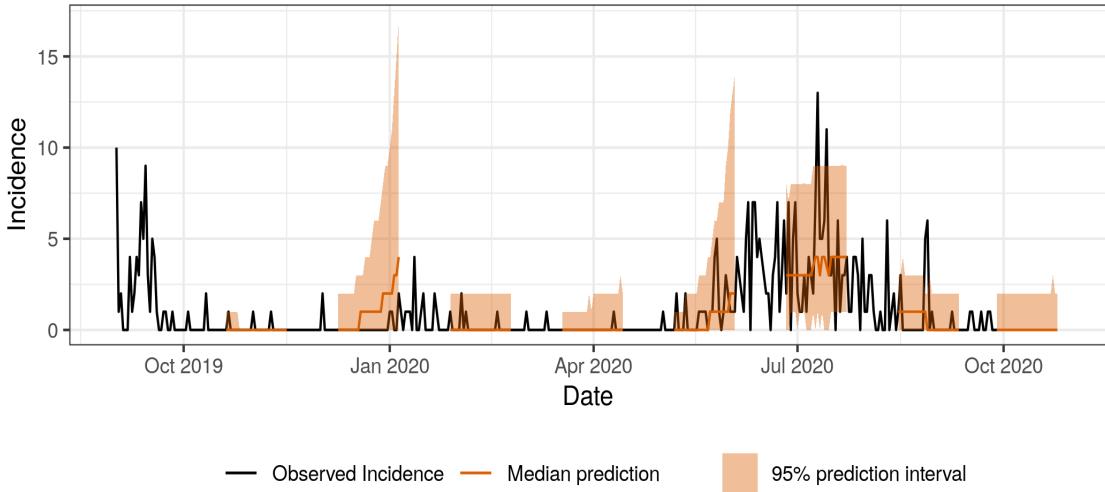
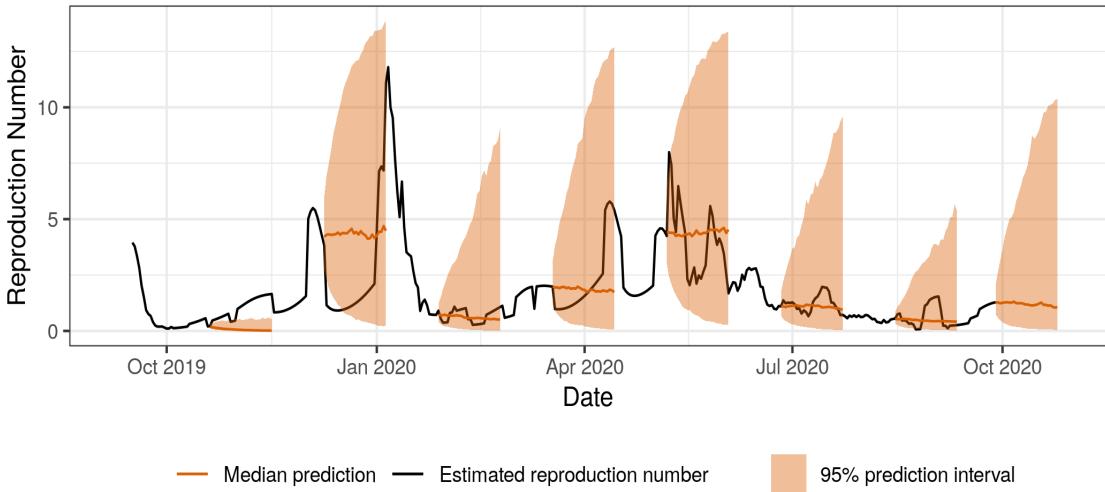


Figure 32: Scores for Lubero as a function of the forecasting horizon.

N Mabalako



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



(b) Forecasted and predicted reproduction numbers for the semilocal poisson model

Figure 33: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the semilocal poisson model for Mabalako.

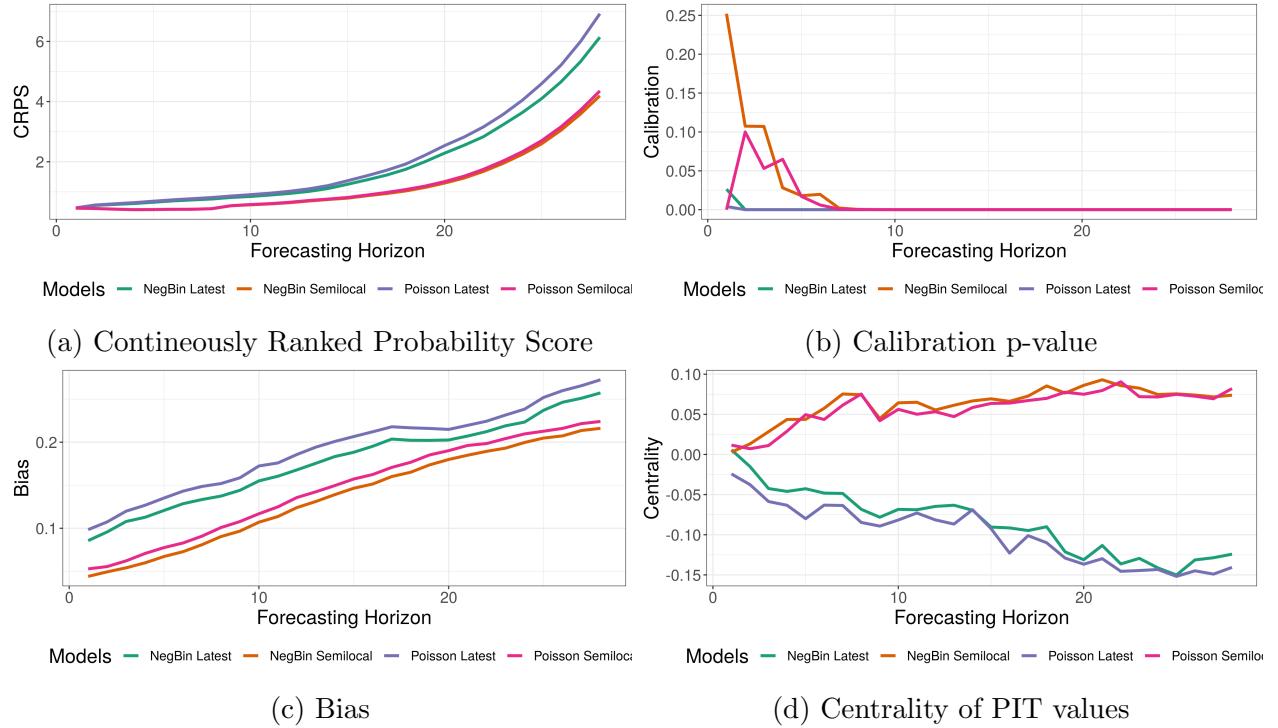
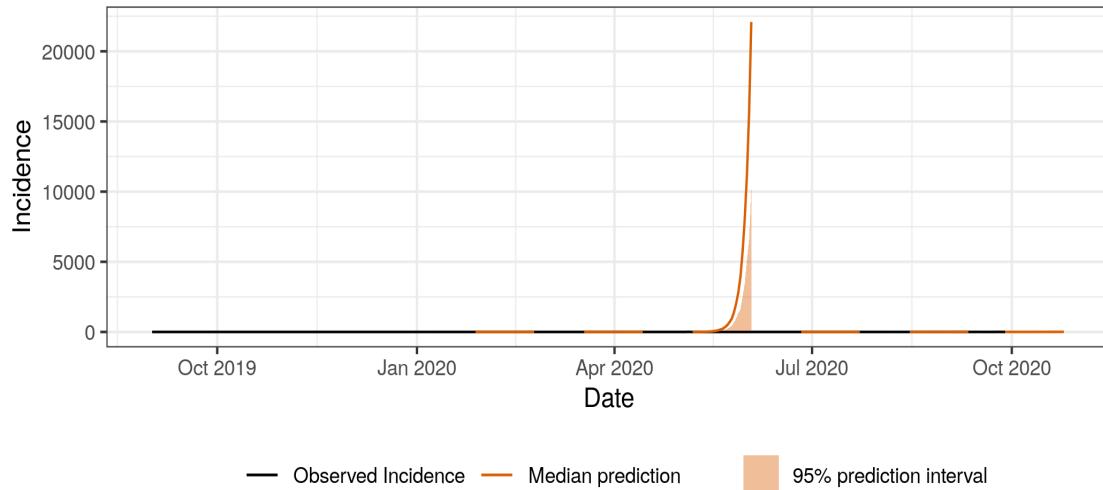
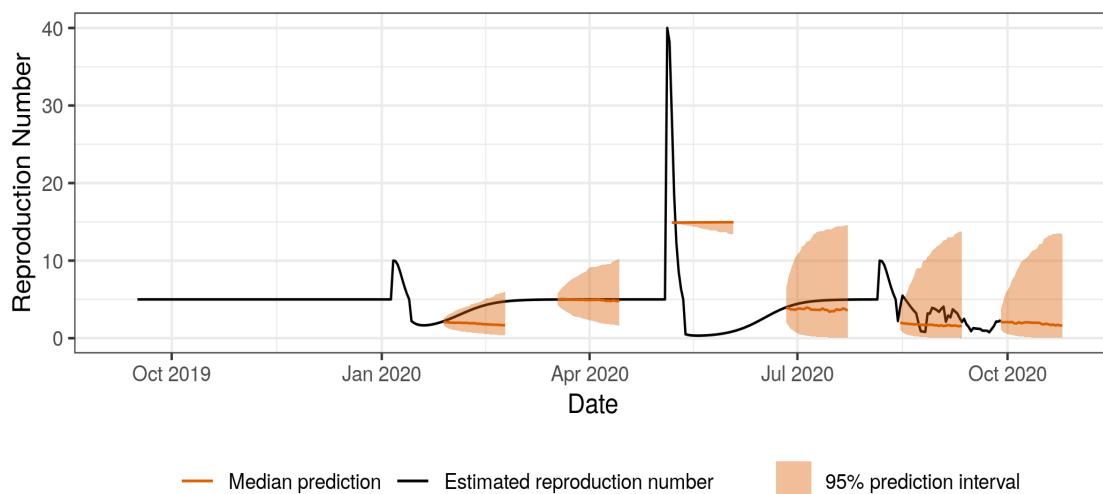


Figure 34: Scores for Mabalako as a function of the forecasting horizon.

O Mambasa



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



(b) Forecasted and predicted reproduction numbers for the semilocal poisson model

Figure 35: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the semilocal poisson model for Mambasa.

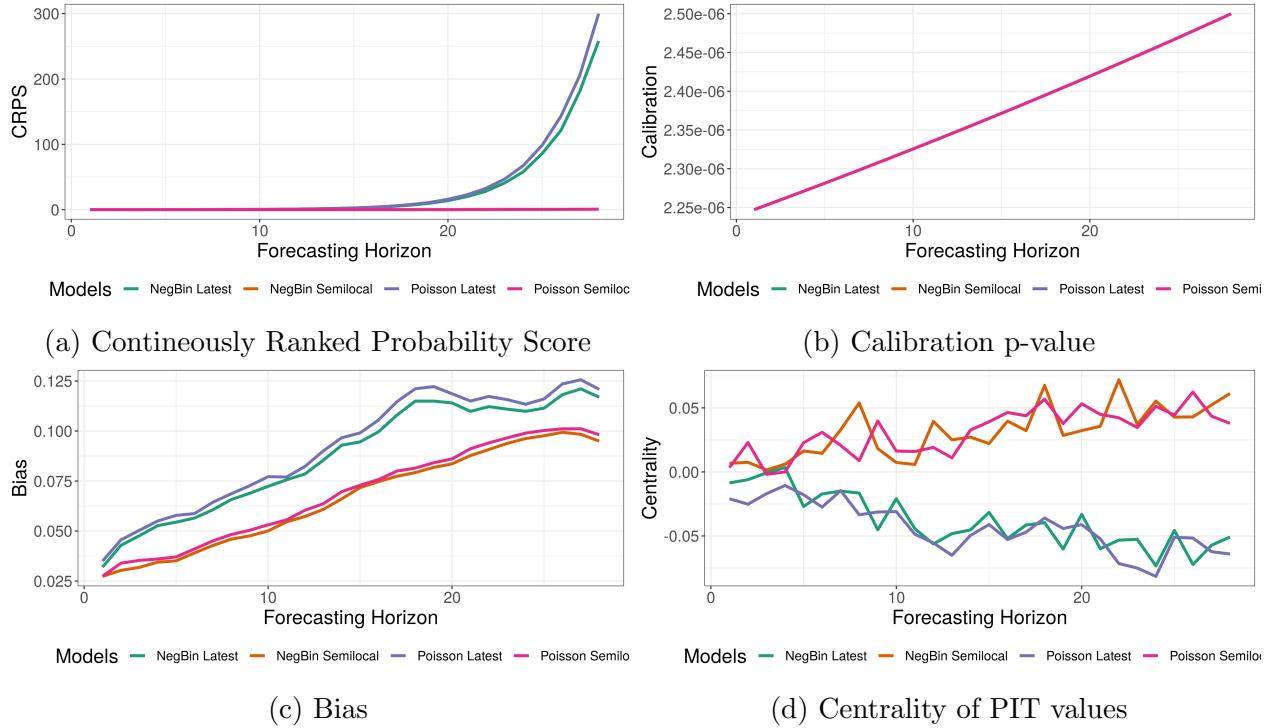
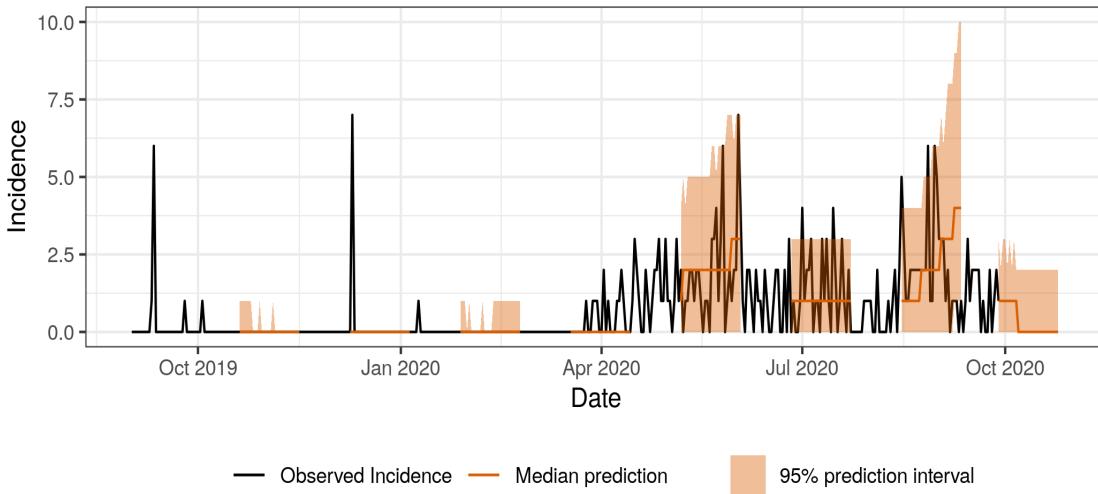
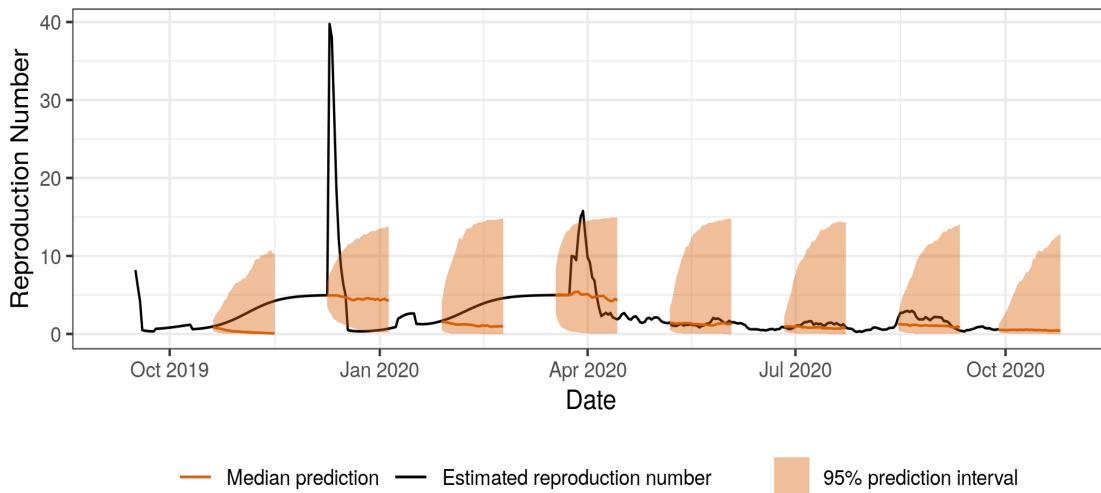


Figure 36: Scores for Mambasa as a function of the forecasting horizon.

P Mandima



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



(b) Forecasted and predicted reproduction numbers for the semilocal poisson model

Figure 37: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the semilocal poisson model for Mandima.

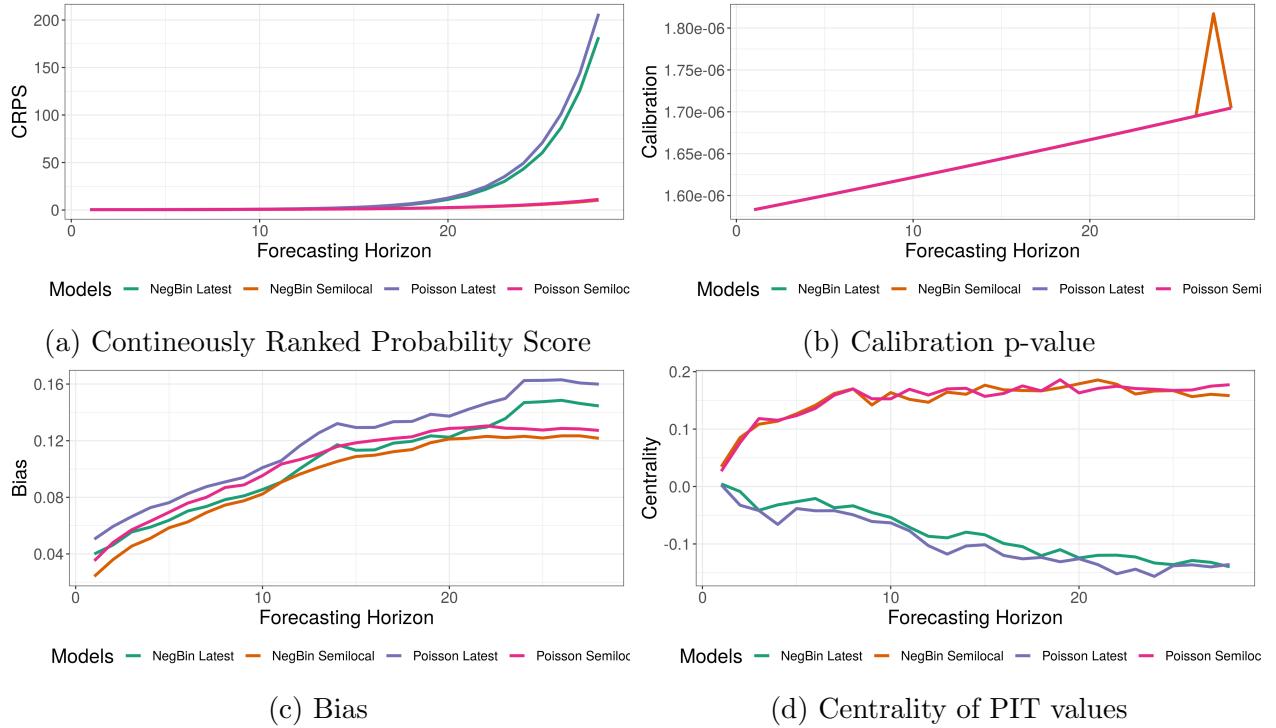
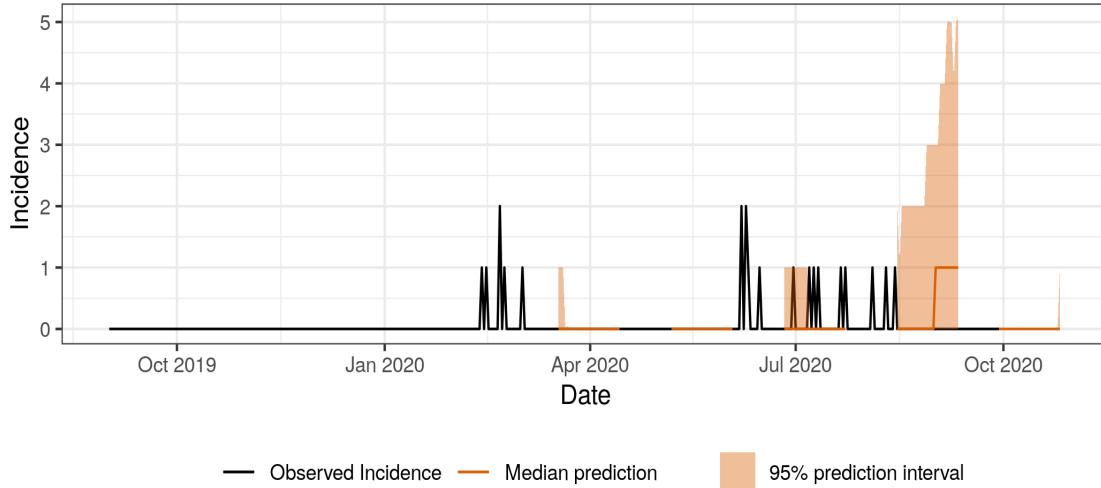
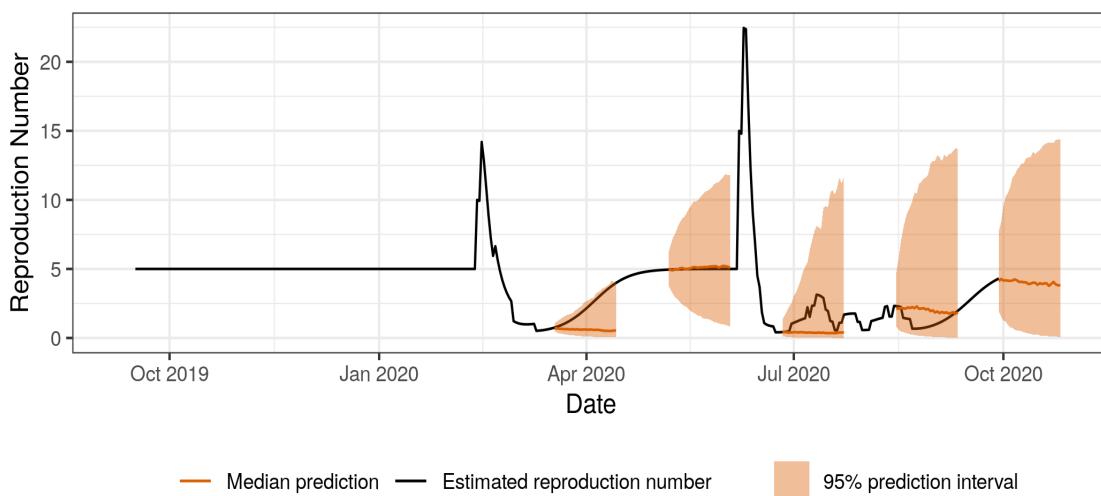


Figure 38: Scores for Mandima as a function of the forecasting horizon.

Q Mangurujipa



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



(b) Forecasted and predicted reproduction numbers for the semilocal poisson model

Figure 39: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the semilocal poisson model for Mangurujipa.

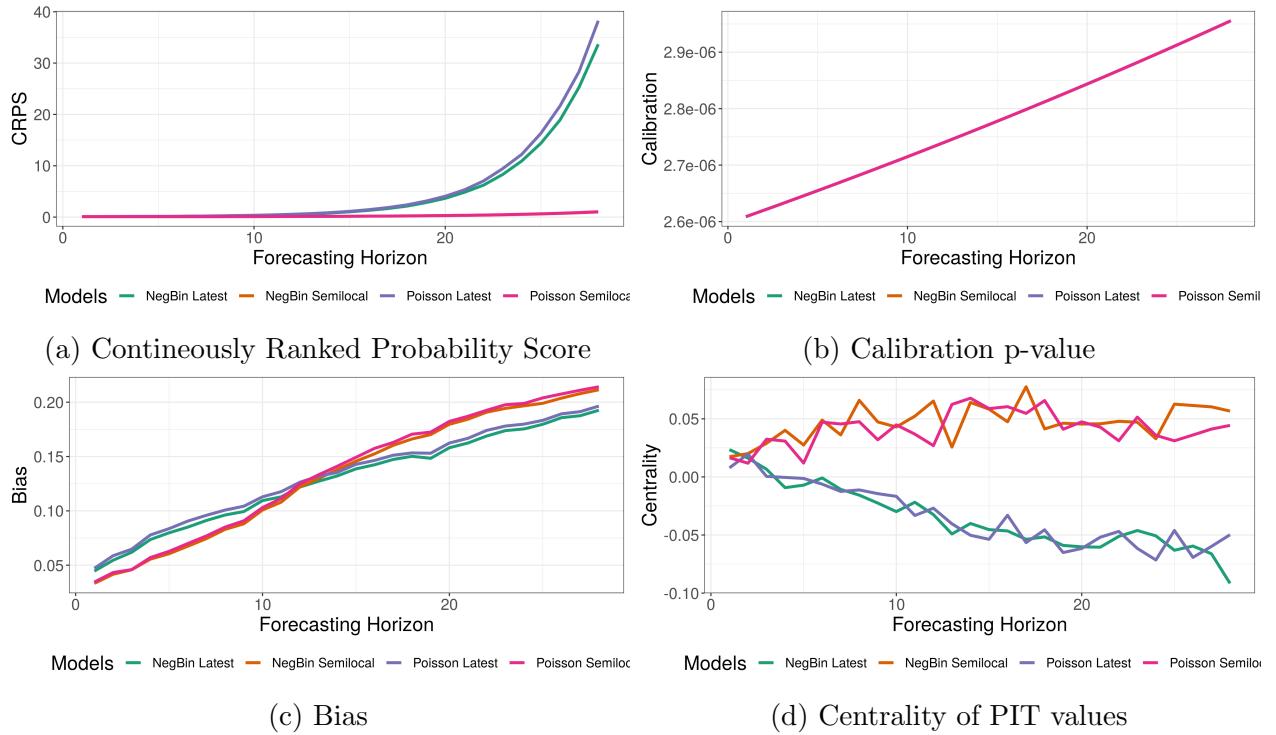
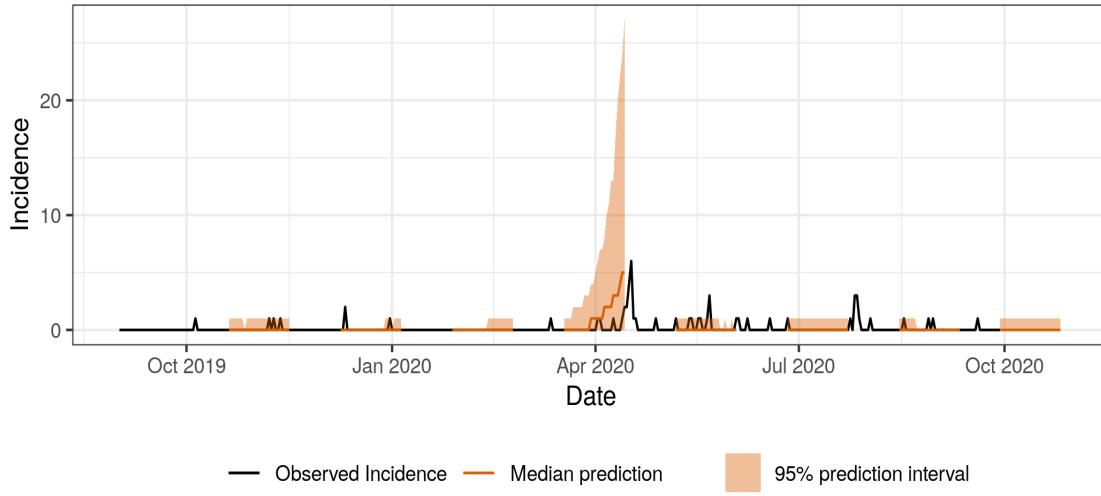
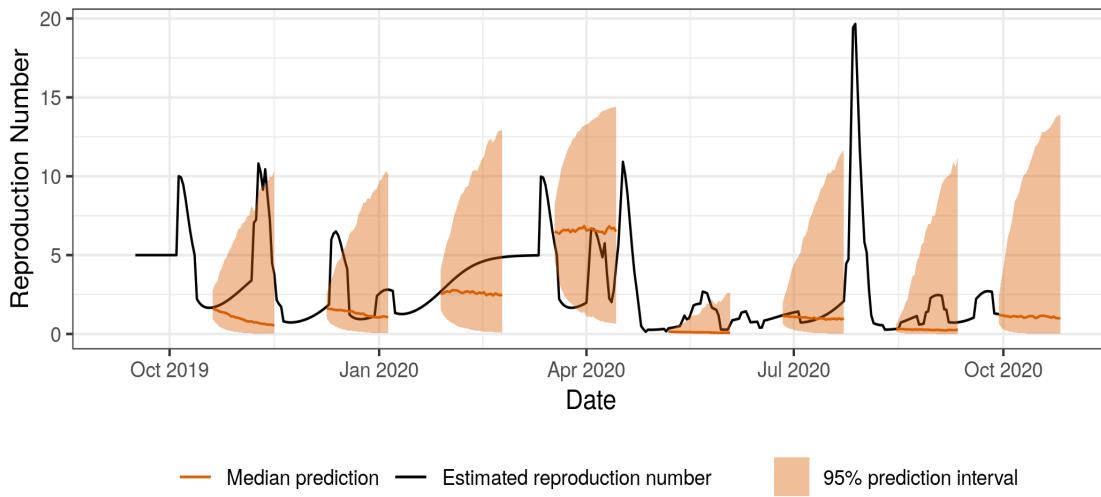


Figure 40: Scores for Mangurujipa as a function of the forecasting horizon.

R Masereka



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



(b) Forecasted and predicted reproduction numbers for the semilocal poisson model

Figure 41: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the semilocal poisson model for Masereka.

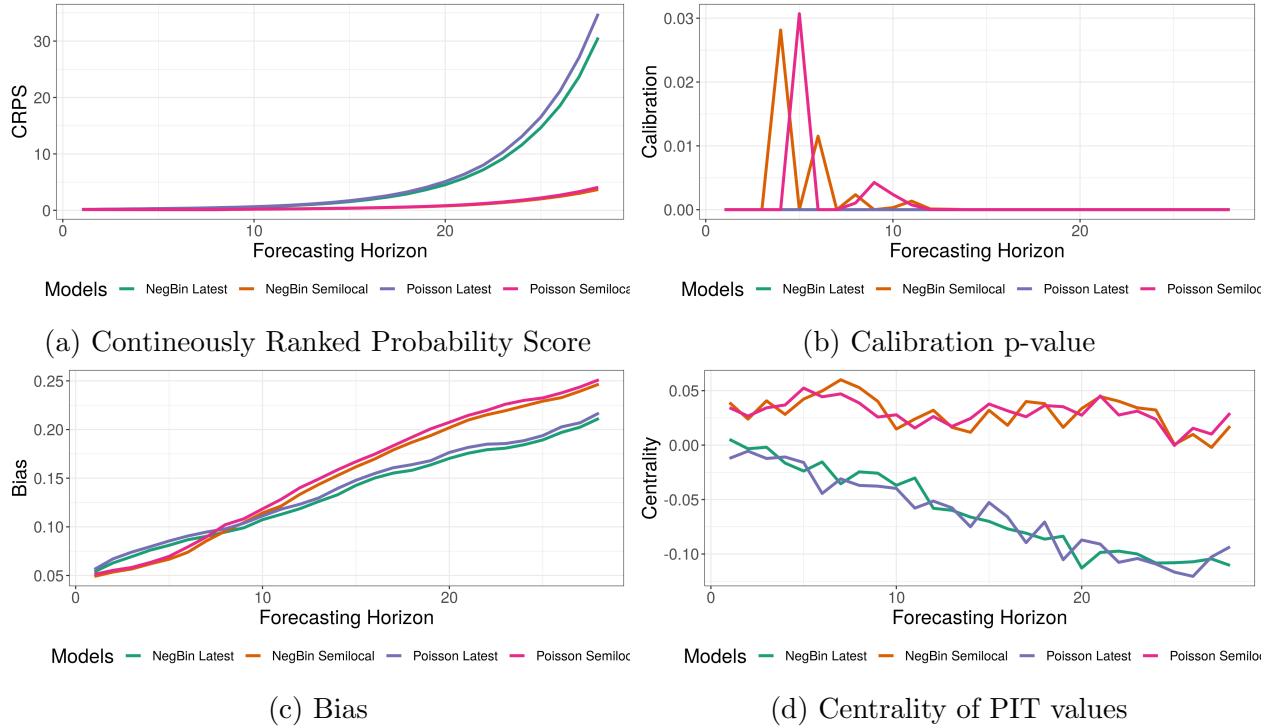
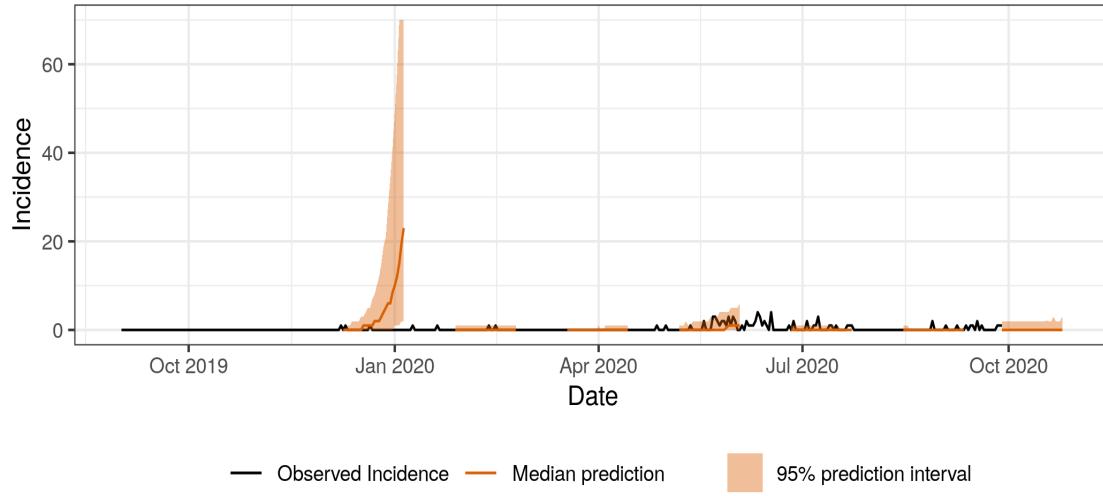
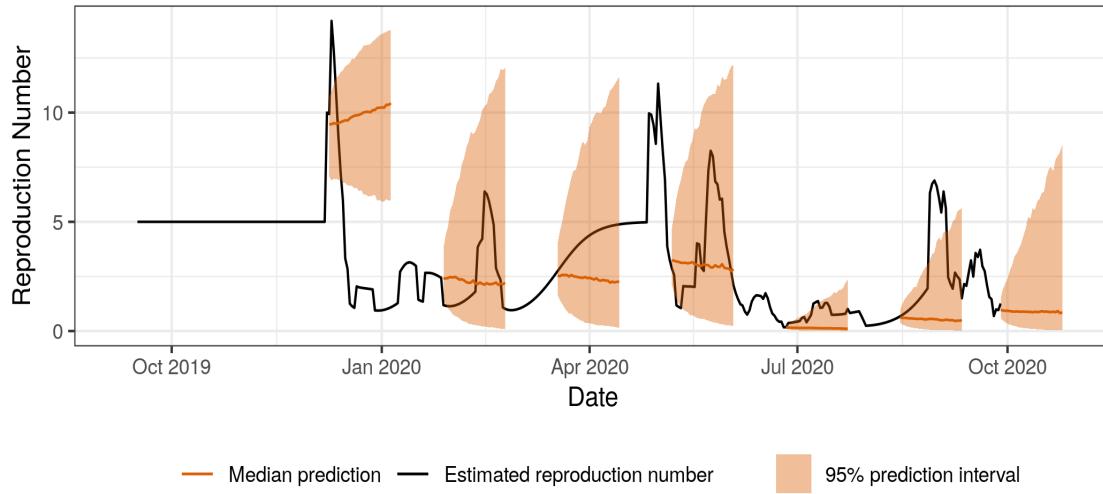


Figure 42: Scores for Masereka as a function of the forecasting horizon.

S Musienene



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



(b) Forecasted and predicted reproduction numbers for the semilocal poisson model

Figure 43: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the semilocal poisson model for Musienene.

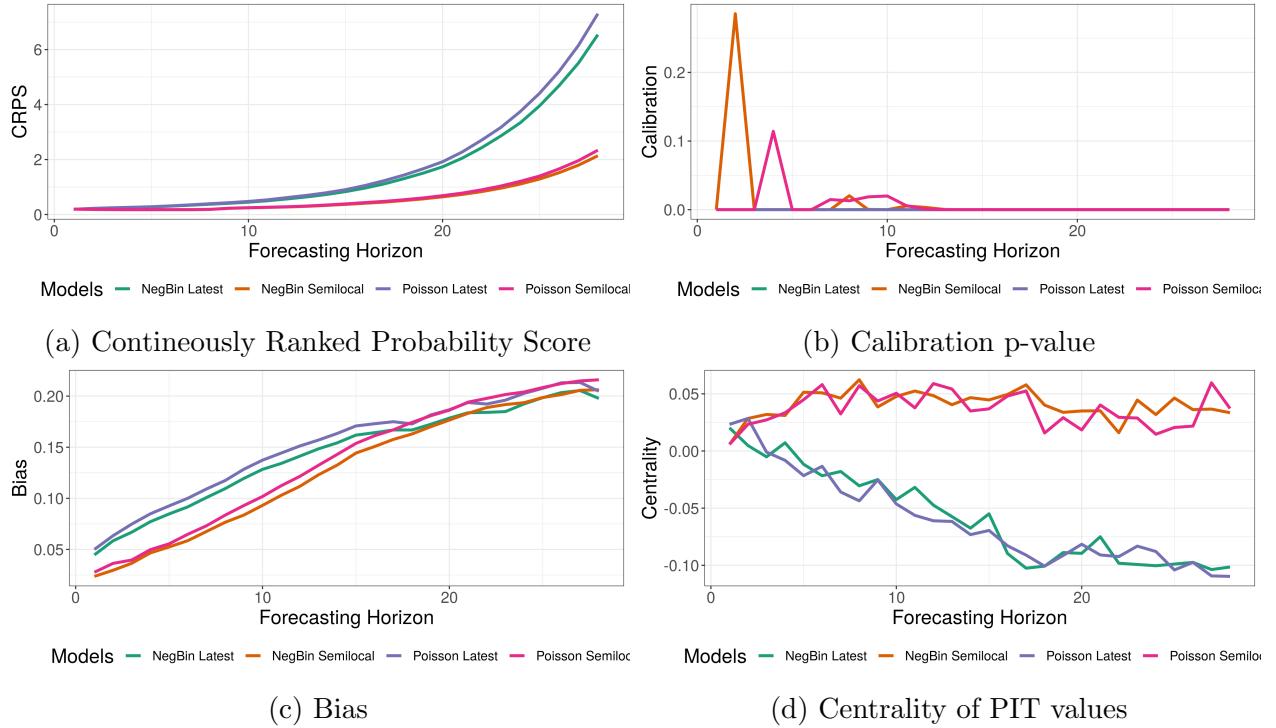
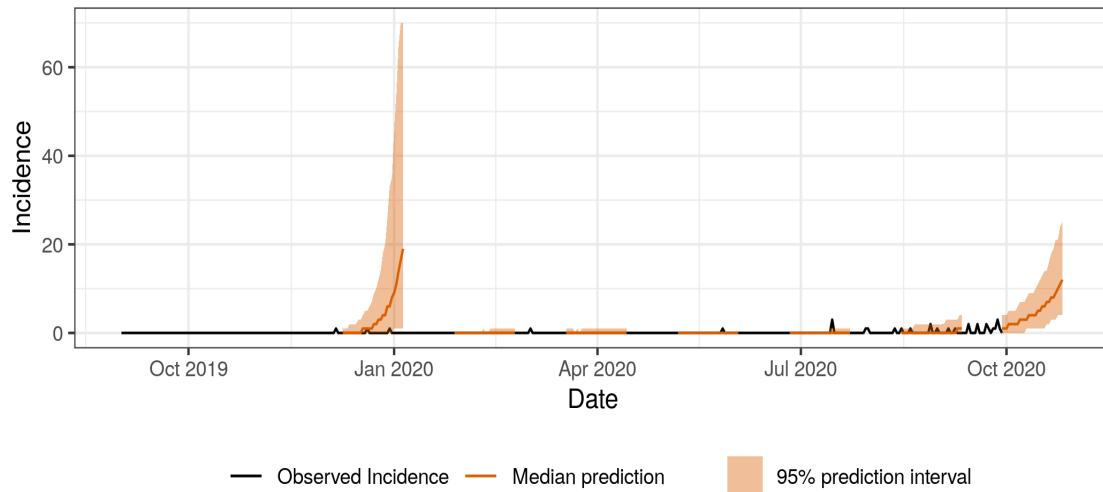
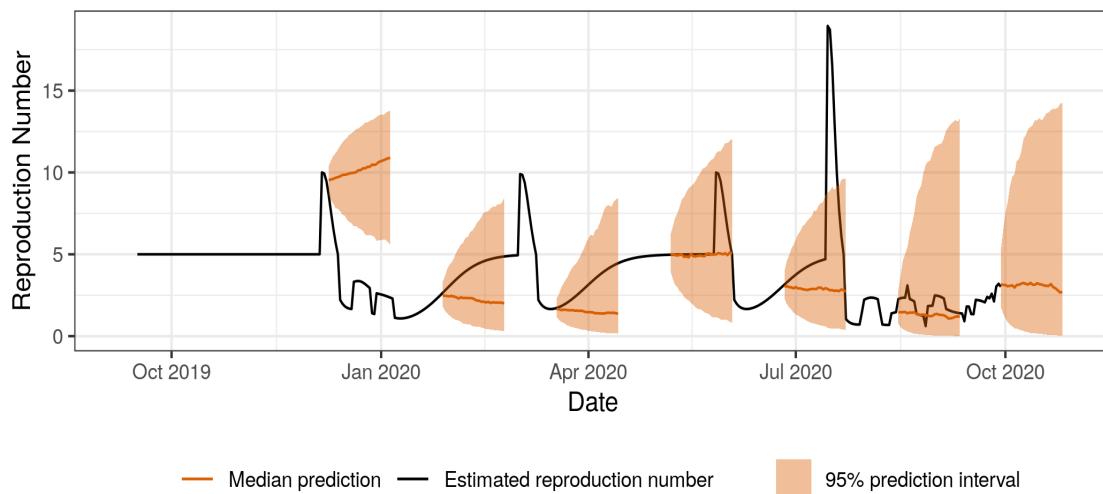


Figure 44: Scores for Musienene as a function of the forecasting horizon.

T Mutwanga



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



(b) Forecasted and predicted reproduction numbers for the semilocal poisson model

Figure 45: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the semilocal poisson model for Mutwanga.

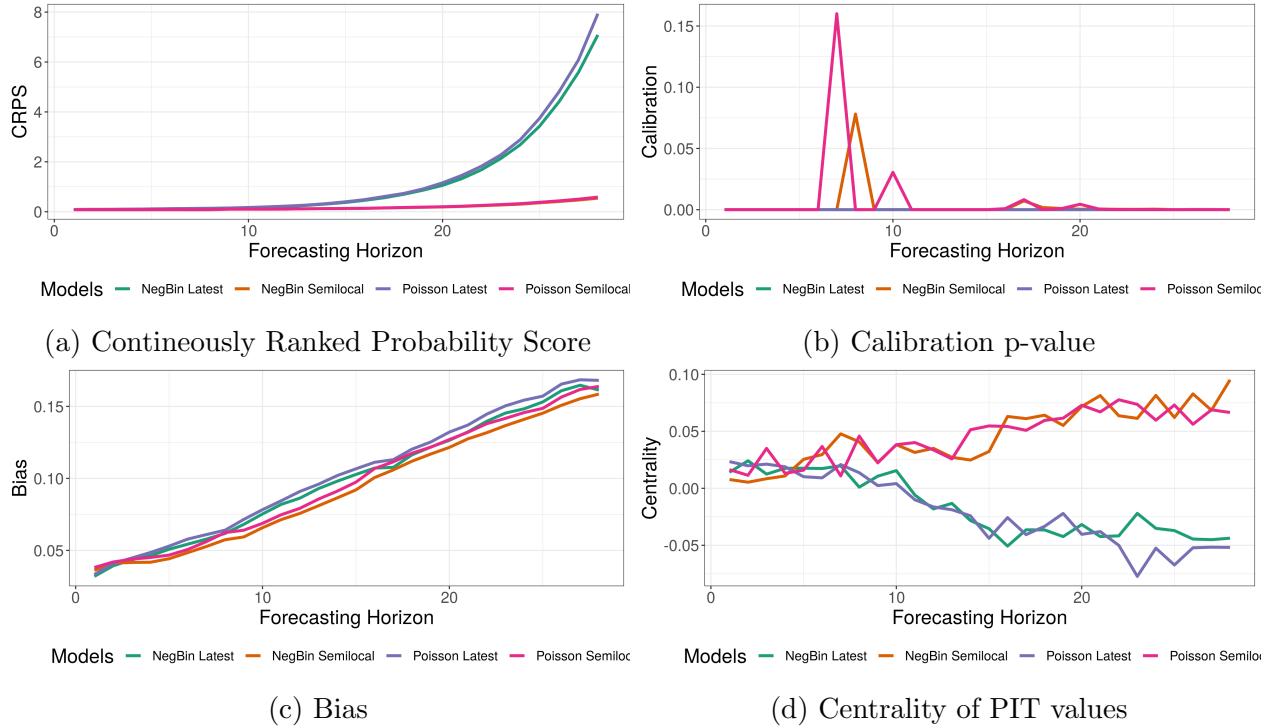
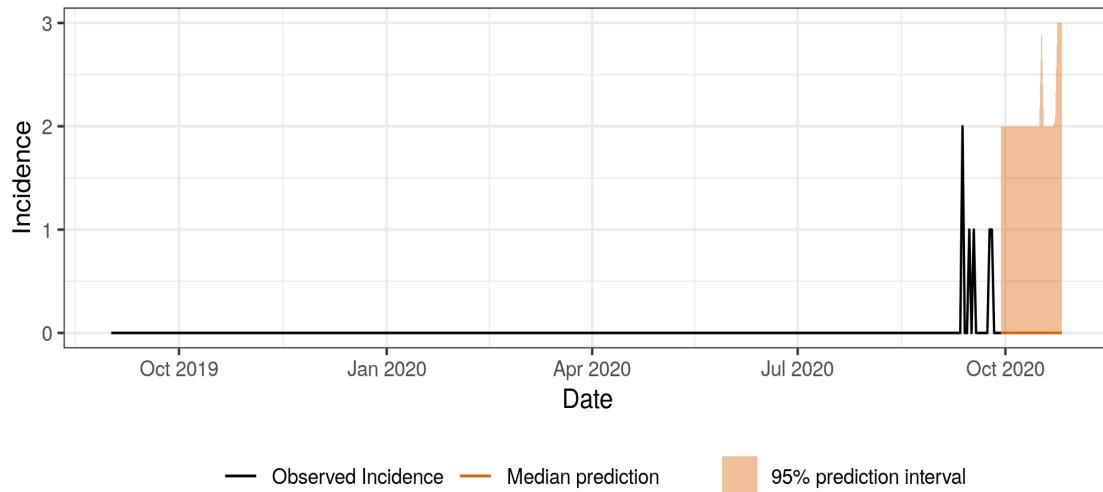
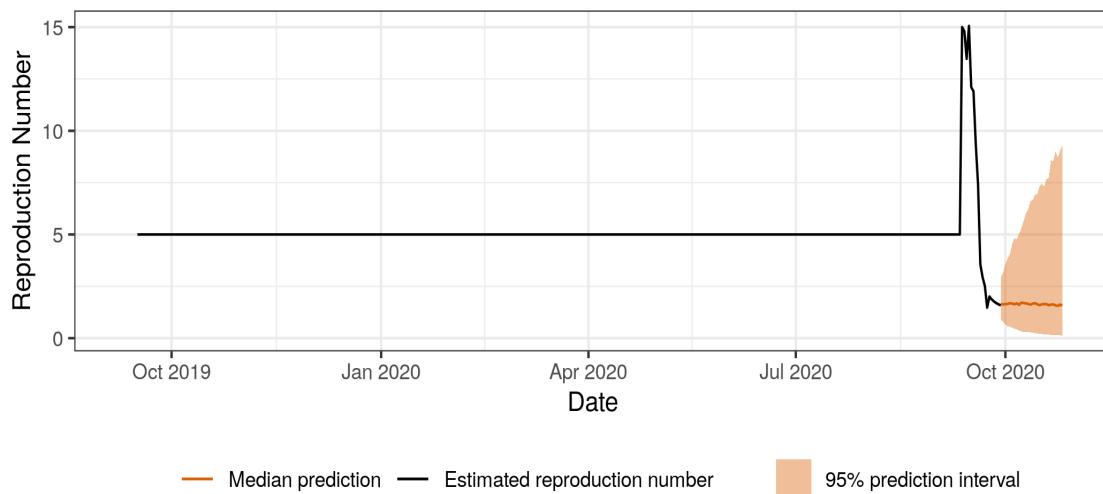


Figure 46: Scores for Mutwanga as a function of the forecasting horizon.

U Mwenga



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



(b) Forecasted and predicted reproduction numbers for the semilocal poisson model

Figure 47: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the semilocal poisson model for Mwenga.

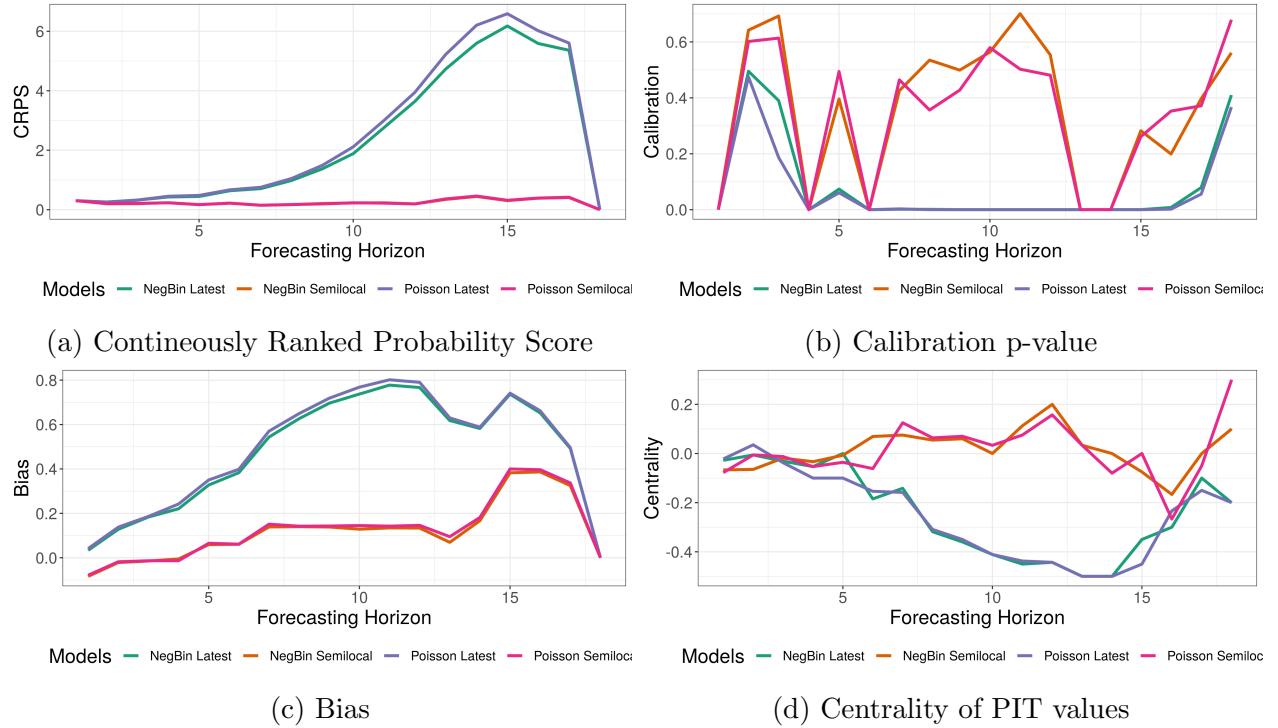
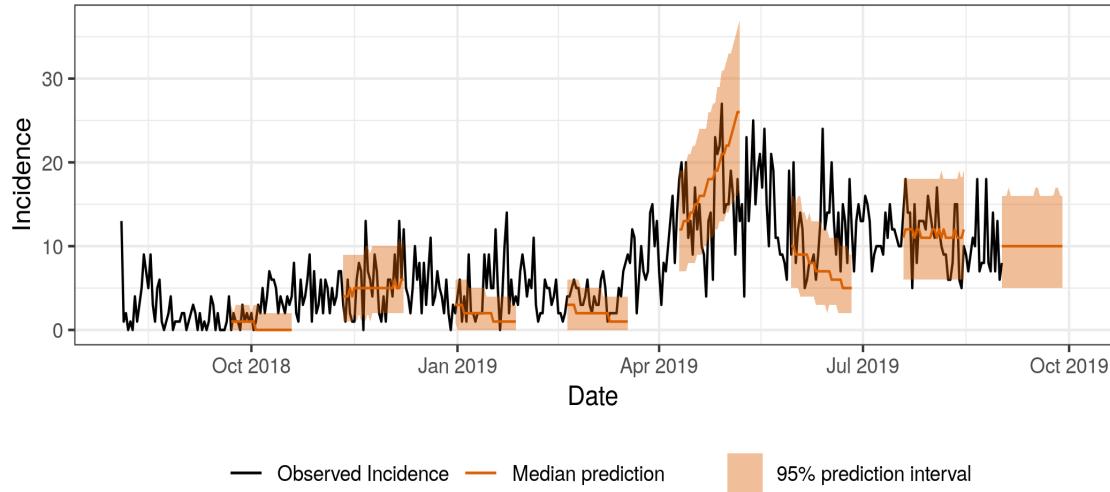
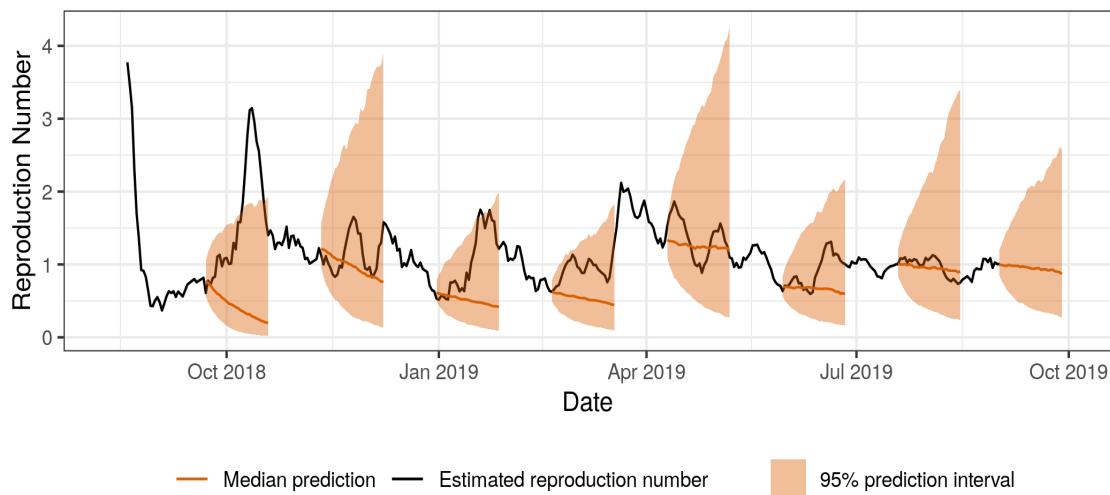


Figure 48: Scores for Mwenga as a function of the forecasting horizon.

V national



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



(b) Forecasted and predicted reproduction numbers for the semilocal poisson model

Figure 49: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the semilocal poisson model for national.

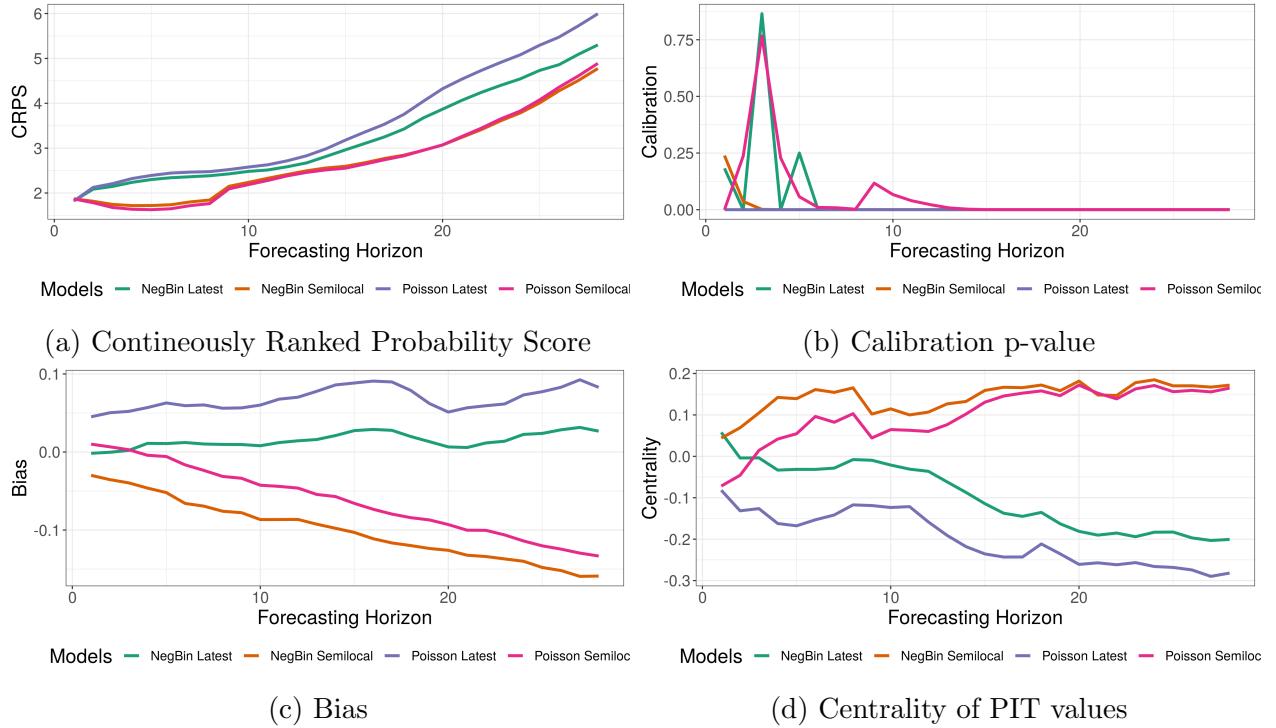
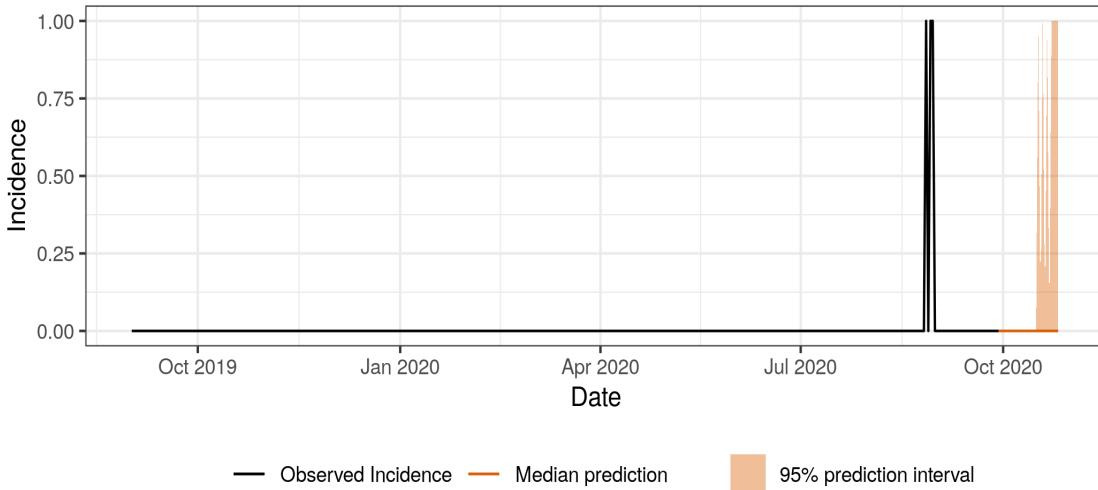
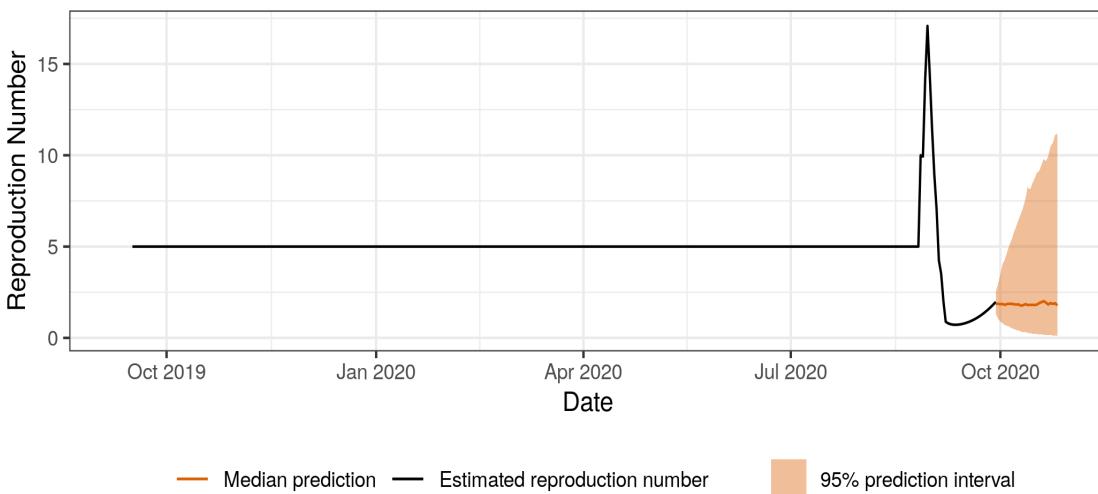


Figure 50: Scores for national as a function of the forecasting horizon.

W Nyiragongo



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



(b) Forecasted and predicted reproduction numbers for the semilocal poisson model

Figure 51: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the semilocal poisson model for Nyiragongo.

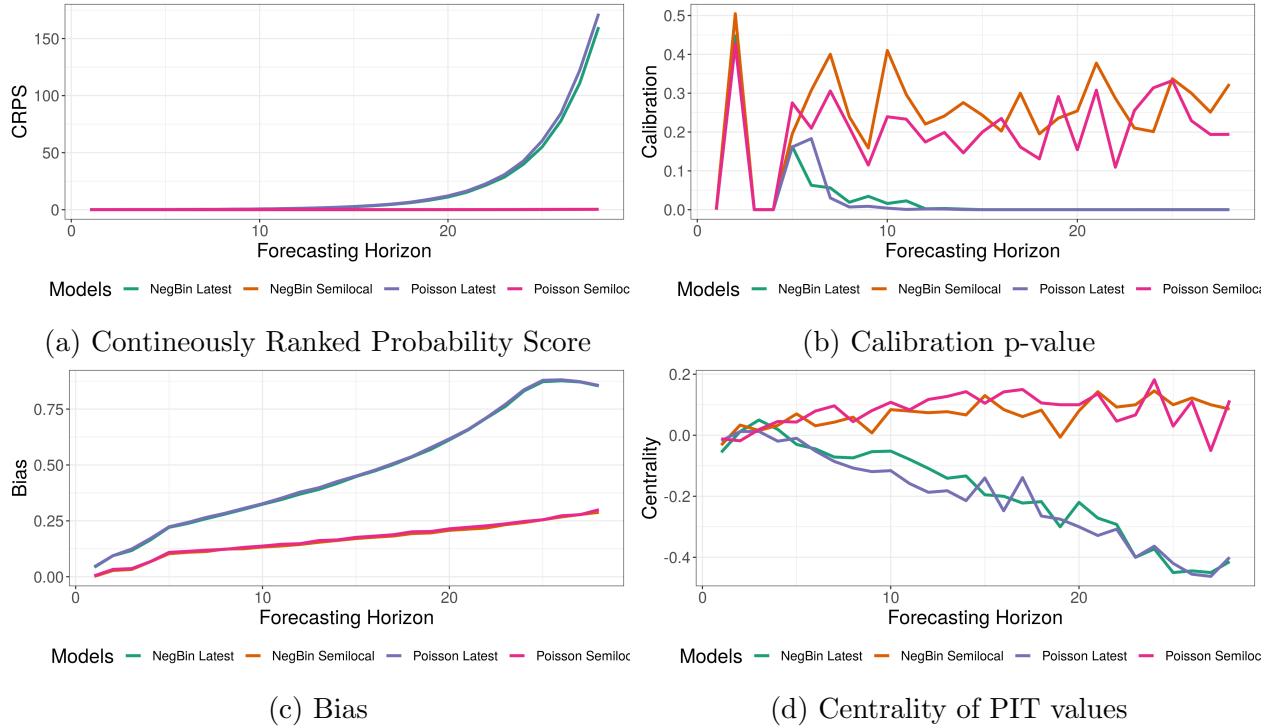
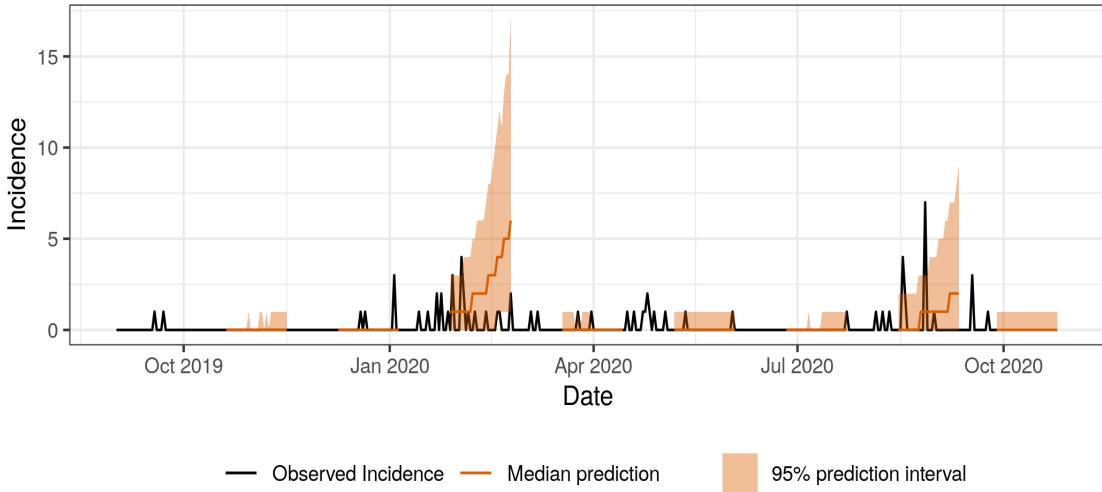
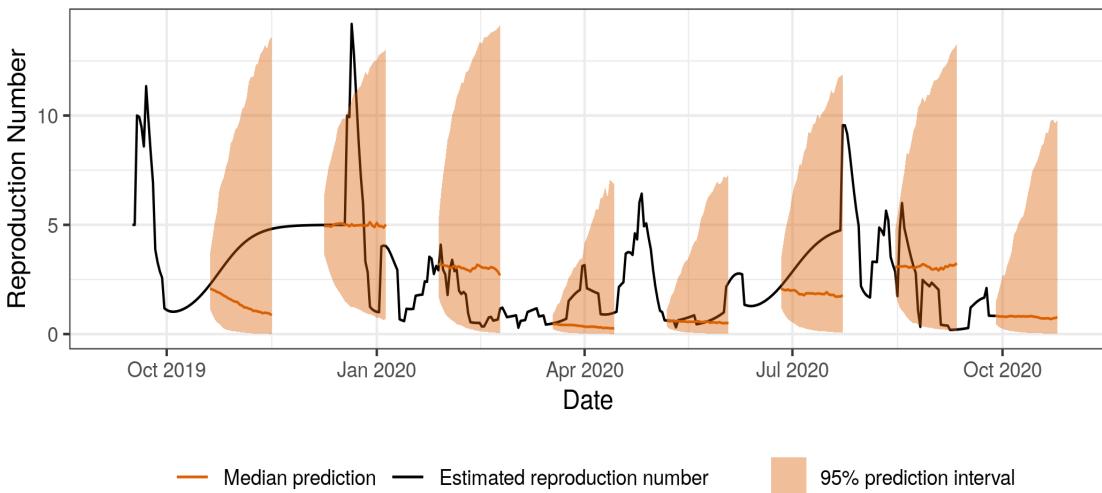


Figure 52: Scores for Nyiragongo as a function of the forecasting horizon.

X Oicha



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



(b) Forecasted and predicted reproduction numbers for the semilocal poisson model

Figure 53: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the semilocal poisson model for Oicha.

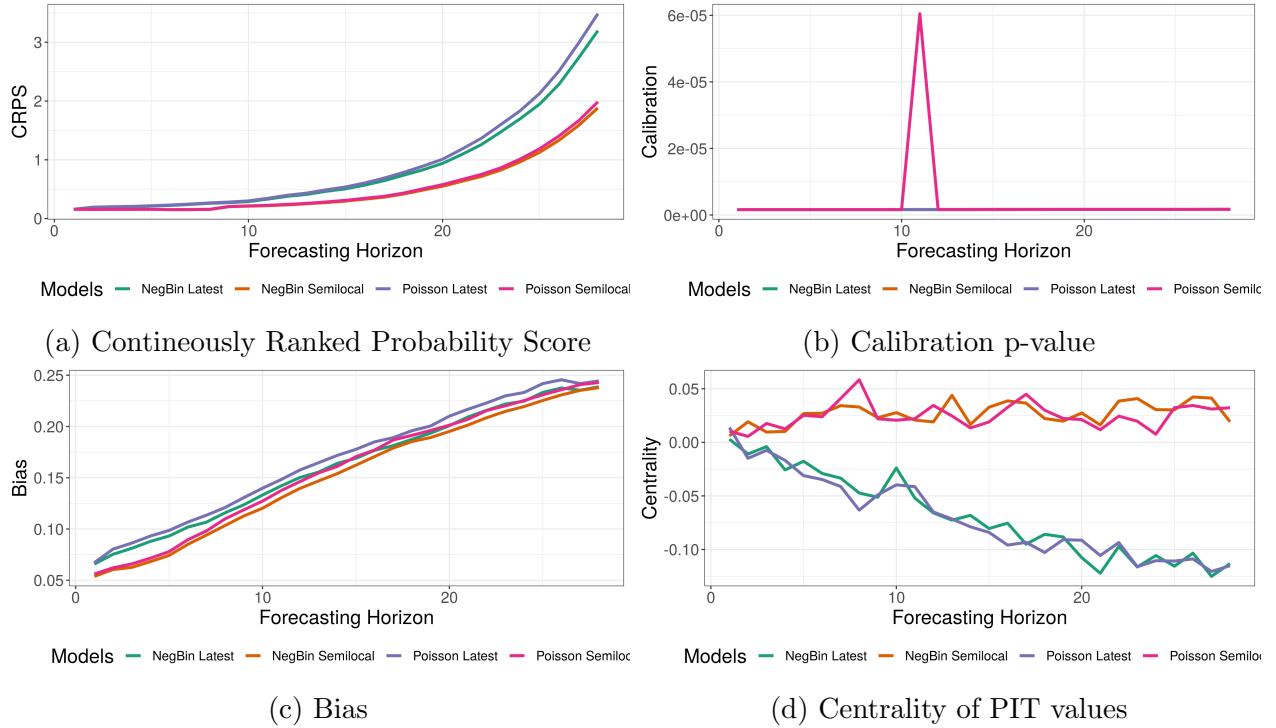
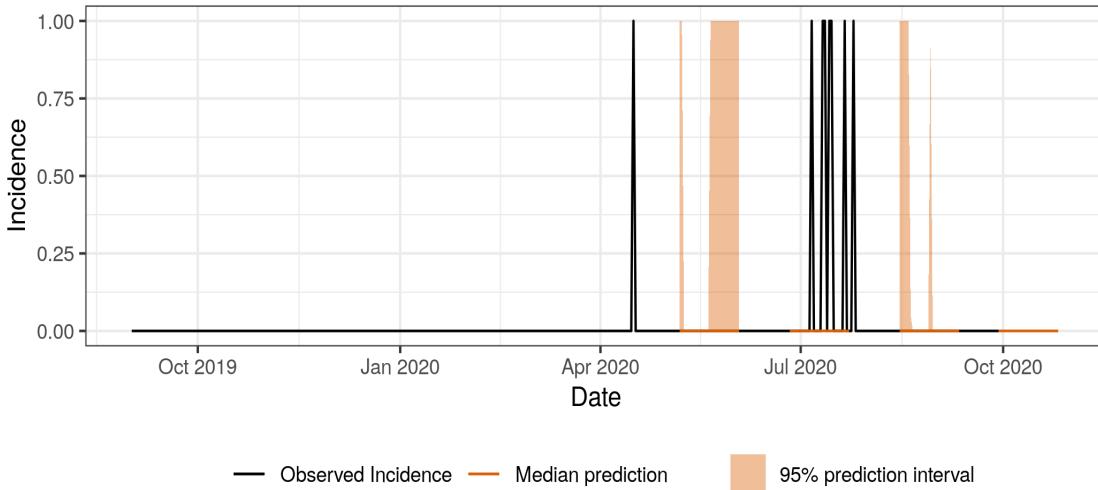
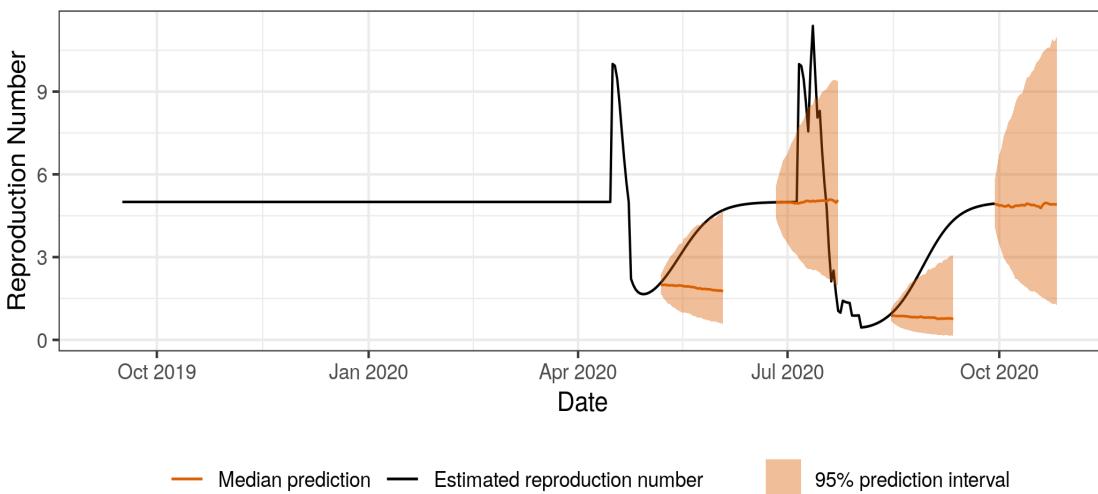


Figure 54: Scores for Oicha as a function of the forecasting horizon.

Y Rwampara



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



(b) Forecasted and predicted reproduction numbers for the semilocal poisson model

Figure 55: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the semilocal poisson model for Rwampara.

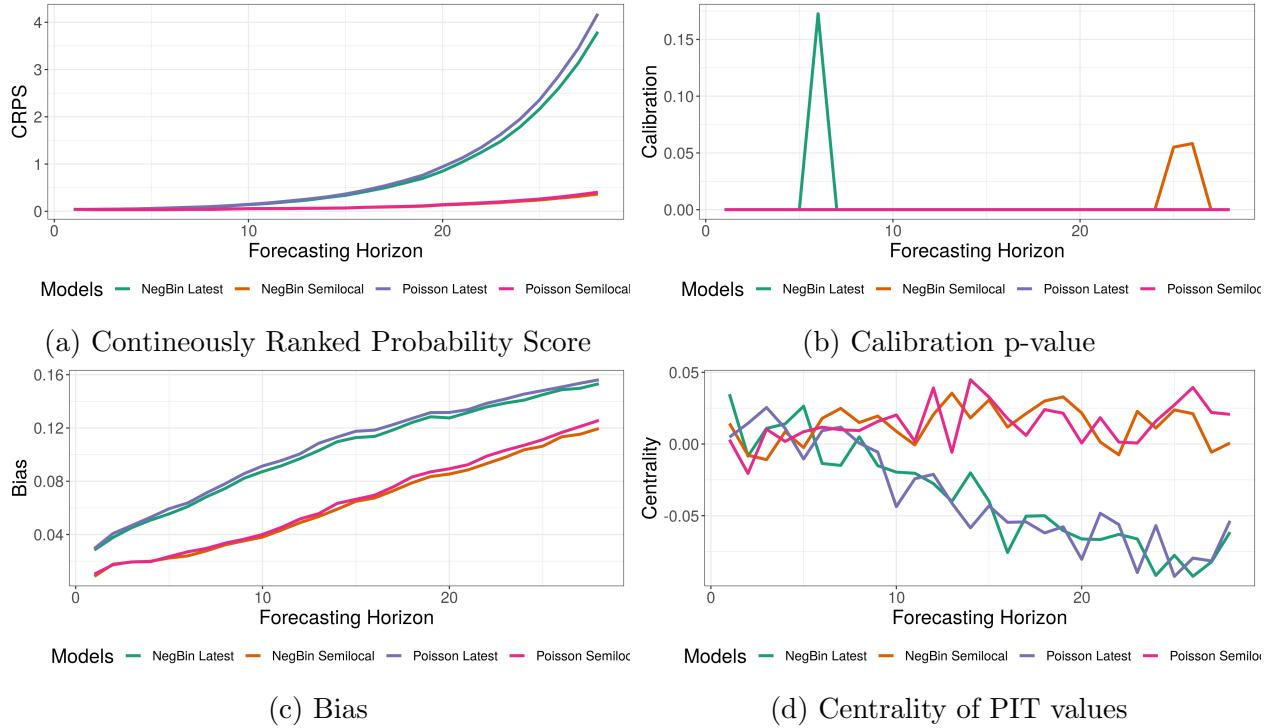
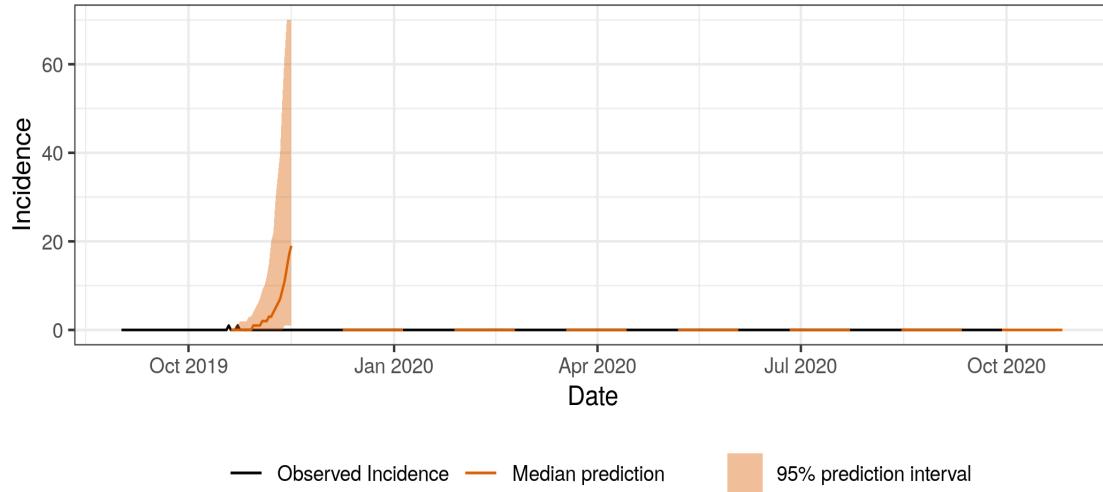
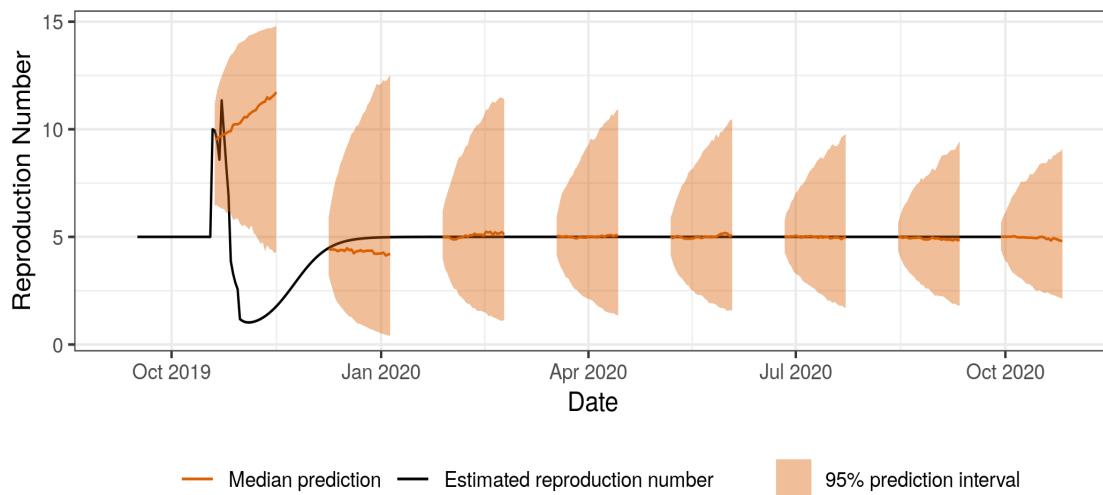


Figure 56: Scores for Rwampara as a function of the forecasting horizon.

Z Tchomia



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



(b) Forecasted and predicted reproduction numbers for the semilocal poisson model

Figure 57: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the semilocal poisson model for Tchomia.

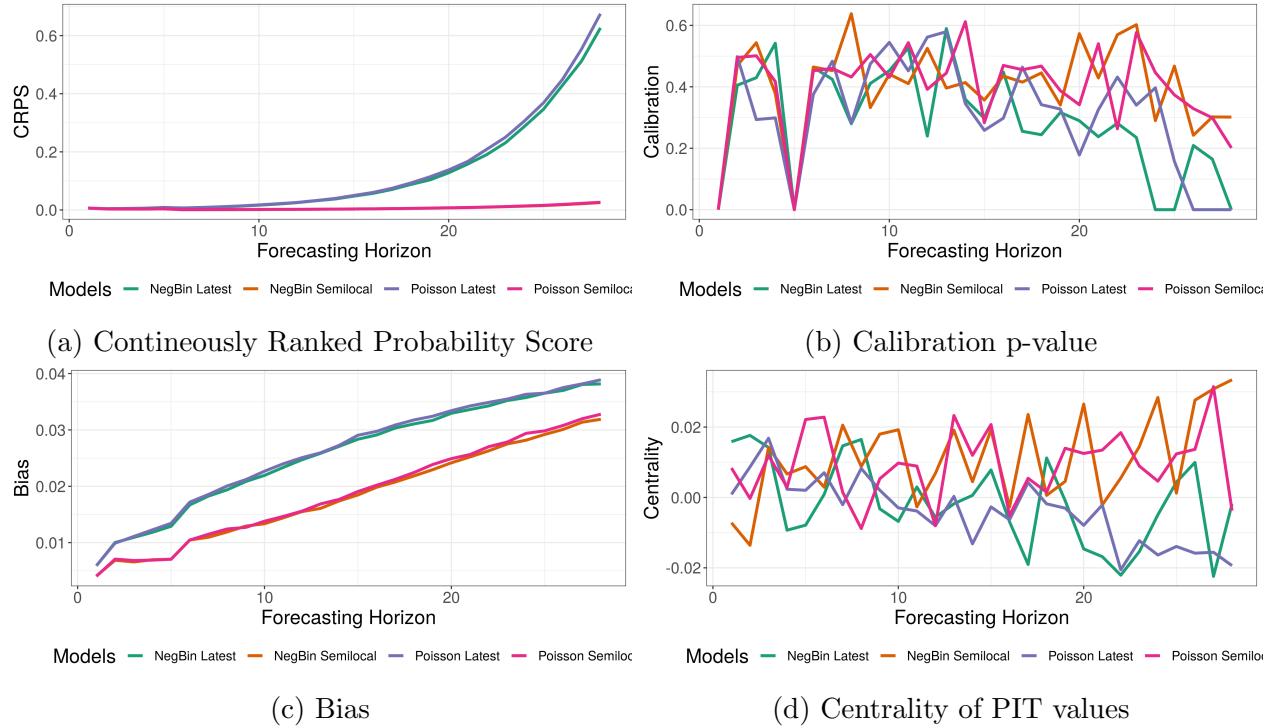
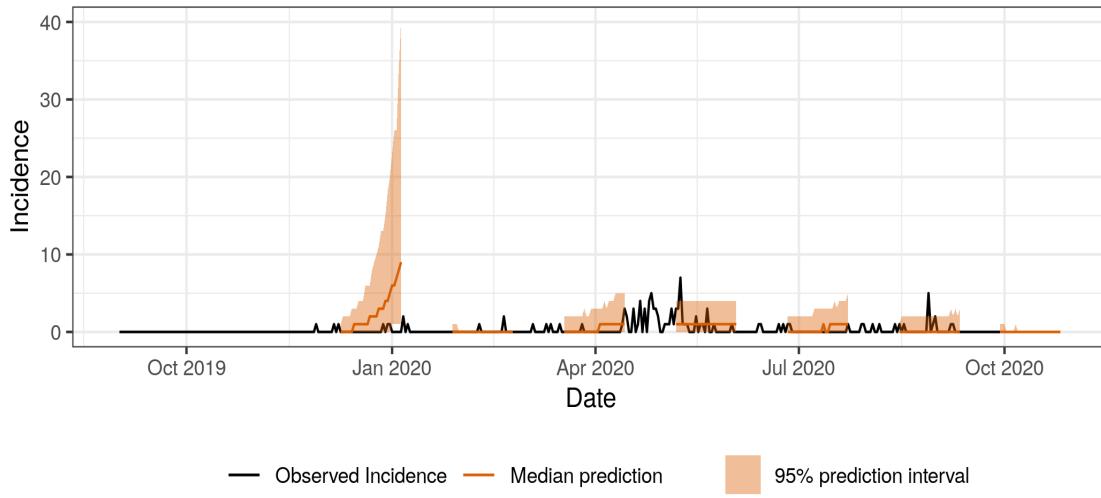
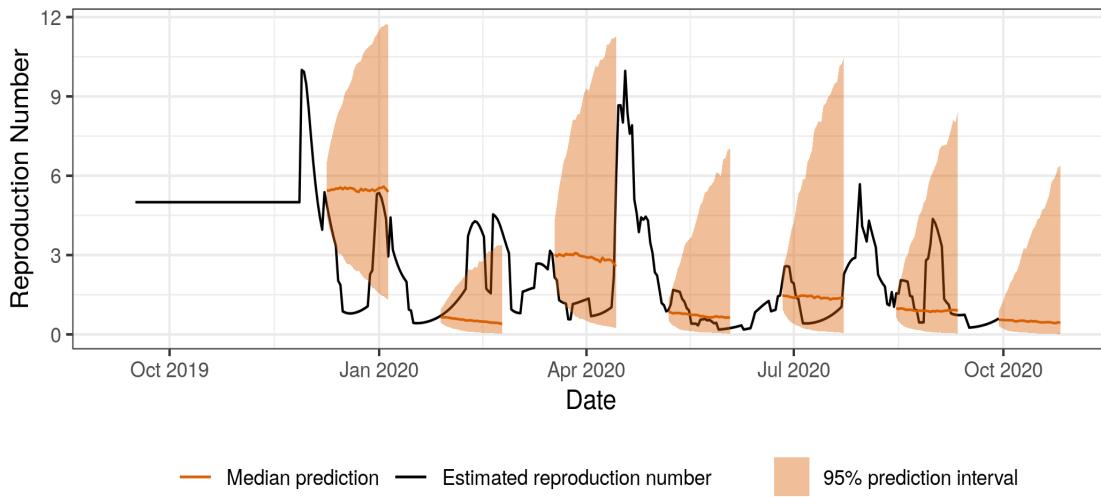


Figure 58: Scores for Tchomia as a function of the forecasting horizon.

Vuhovi



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



(b) Forecasted and predicted reproduction numbers for the semilocal poisson model

Figure 59: Median forecast with 95 % prediction intervals and observed values for incidence and reproduction number for the semilocal poisson model for Vuhovi.

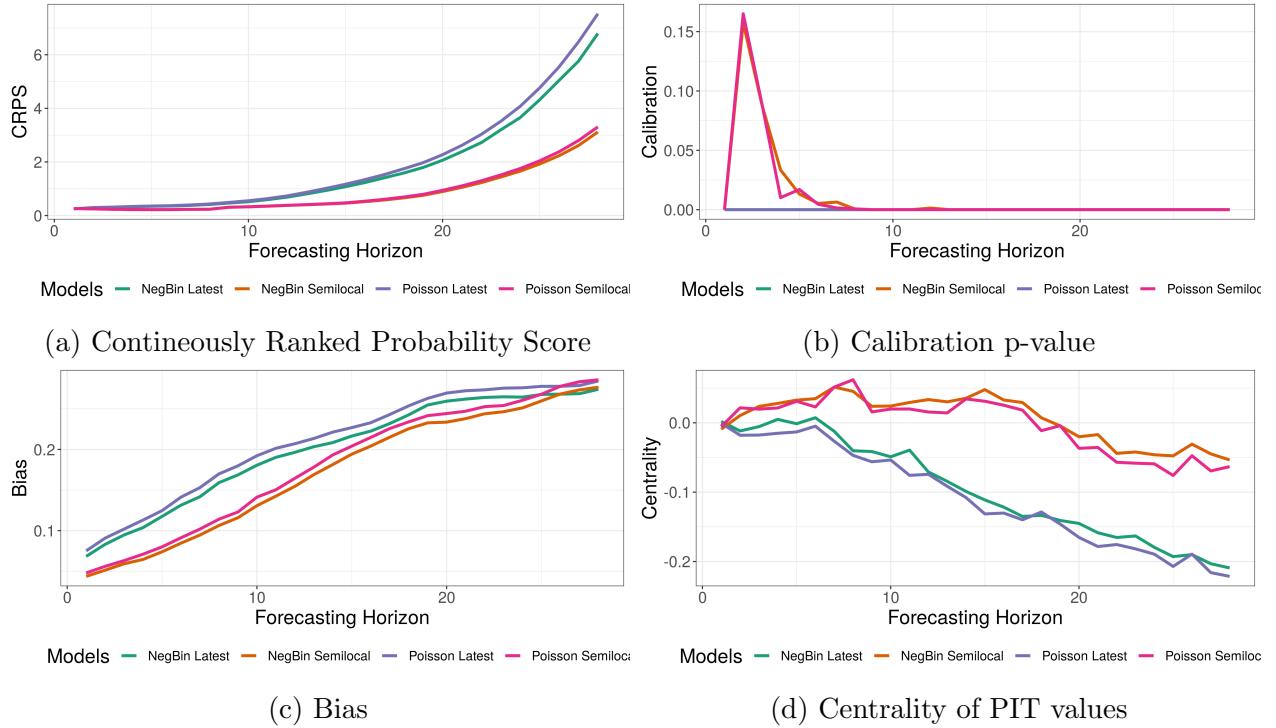


Figure 60: Scores for Vuhovi as a function of the forecasting horizon.