Hierarchical Bayesian Model for Correcting Reporting Delays in Dengue Counts

Mikee T. Demecillo[1] and Bernadette F. Tubo[2]

[1,2] *Mindanao State University – Iligan Institute of Technology, Department of Mathematics and Statistics, 9200 Iligan City, Philippines*

[1] *mikee.demecillo@g.msuiit.edu.ph*

[2] *bernadette.tubo@g.msuiit.edu.ph*

Abstract—The ability to produce real-time surveillance and accurate estimation of dengue cases is essential to gain situational awareness in identifying trends and outbreaks and gaining effective control measures. Reporting delays between disease onset and case reporting hamper our understanding of the dynamics of a sudden increase or decrease in disease cases which vary over time. In this study, a Bayesian inference for latent gaussian models implemented in R-INLA is presented as decision support with a flexible temporal nowcasting model to correct reporting delays for weekly dengue surveillance data in Northern Mindanao from 2009-2010. Moreover, it aims to quantify all the associated uncertainty in correcting the missing value. The statistical challenge lies in predicting counts of the run-off triangle based on observed counts $n_{t,d}$. The posterior predictive model on the given temporal dataset identifies the fact that there is an increasing number of dengue cases (supporting the actual scenario) as compared to the current reported cases that appears to be decreasing already. This suggests that the model was able to give a reasonable estimate of the true number of cases in the presence of delayed reports. This work provides a template for nowcasting in aid to dengue control and proper decision making by concerned authorities.

Keywords— Bayesian Hierarchical Model, Latent Gaussian Model, Count Data, Correcting Delay

I. Introduction

Epidemiological surveillance is the systematic collection, analysis, and dissemination of health data for public health purposes [8]. One of the functions of infectious disease surveillance is detecting outbreaks and initiating timely intervention. Count data arise in many applications from this field that may not fully represent the quantity of interest. Frequently, detecting an early increase in diseases is particularly problematic since reports must be reviewed and acted on as they accumulate, without the opportunity to correct errors or adjust for delays in reporting and other artifacts of the reporting system [6].

Infectious disease control requires effective and prompt responses to unanticipated increases in disease burden. Inability to produce a timely and accurate estimation of contagious disease burden can significantly impact public health control programs [4]. In common with tropical countries, one of the diseases that suffers from this type of problem is a viral infection spread by mosquitoes, Dengue. A key challenge for dengue surveillance is the time between the onset of adverse health events and its report and the time between the report and the identification of trends or outbreaks. This results in an initial underestimate of the actual burden and a late, or no, response. It impedes the response time to severe outbreaks and puts lives at risk.

The problem of occurred-but-not-yet-reported cases during outbreaks is well known from the HIV/AIDS outbreak, and different statistical approaches have been proposed to handle delayed reporting [16]. A standard reference is Lawless [9]. However, a more flexible Bayesian nowcasting approach has been developed by Höhle and an der Heiden [7]. The framework was used to correct delays, allowing temporal variation of the total number of cases N_t and the delay mechanism. In the following, the delay adjustment approach can be refered as the *nowcast* and define the *reporting delay* as the time between disease onset and official

case reporting by a health authority. In addition, Bastos et al. proposed a bayesian hierarchical modeling approach for correcting reporting delay and associated uncertainty using a lognormal survival model. They use integrated nested Laplace approximations (INLA) to apply this framework to dengue fever in Rio de Janeiro and to spatiotemporal Severe Acute Respiratory Infection (SARI) data in the state of Paraná (Brazil). The marginal distribution of each count $n_{t,d,s}$ is negative binomial with mean $\lambda_{t,d,s}$ and dispersion parameters ϕ to allow spatiotemporal variation in counts, as well as covariates.

Motivated by the works of Bastos [2], this paper presents a decision-support tool that considers the Bayesian Hierarchical Model for modeling count data with time-delays, which is sufficiently flexible to be used for a range of disease applications. The proposed model is a latent Gaussian model, which can readily implement using R-INLA that approximated marginal posterior distributions of the latent fields and the hyperparameters. This study also consider the effects of delays that occur in reporting events such as cases of a reportable disease like dengue cases in Northern Mindanao from 2009 to 2010 and in estimating the number of events that have occured but not yet been reported (OBNR) events for temporal model.

II. Related Literatures

Accounting for systematic delays in reporting has a historical precedent in actuarial sciences in modelling claims reserves [15]. It has been addressed for health outcomes in HIV/AIDS [3], mortality reporting [10], and chronic diseases, e.g., cancer registries [12]. Historically, the task of correcting the delayed reporting has been separated from the task of modeling or forecasting the incidence of the total count [3]. However, this ignores the joint uncertainty in the incidence of the total count and the presence of delay. For example, suppose that at time t the number of cases reported in the first week $n_{t,1}$ is usually low. This could either be because a low proportion of N_t was reported in the first week, or because N_t was itself unusually low, or both. Differentiating between these cases is vital for reliable prediction, so we focus on an approach that jointly models the delay mechanism and the total count.

Methods have also been developed in infectious disease outbreaks. The methods have been developed to nowcast (i.e., estimate in real-time) the current number of infected cases. Höhle and Heiden (2014) forecasted the daily hospitalized number of cases of hemolytic uremic syndrome. They consider distribution of the counts $n_{t,d}$ conditional on the totals N_t. The framework is then hierarchical where the N_t are assumed to be distributed as Poisson or Negative Binomial. Then, $n_{t,k}|N_t$ is multinomial with some probability vector of size D that needs to be estimated. The framework was used in a Bayesian nowcasting model to correct delays in the reporting of Shiga toxin-producing Escherichia coli in Germany[7]. The model allows for smooth changes in the temporal variation of the total number of cases N_t and the delay mechanism by characterizing the multinomial probability vector as a function of time. Then, Noufaily et al. (2016) proposed a method for detecting infectious disease outbreaks from laboratory data with reporting delays [13] based on a quasi-Poisson algorithm [14].

In 1993, Mark T., developed the so-called chain-ladder technique as a distribution-free method to estimate missing delayed counts [11]. The chain ladder method is probably the most popular method for estimating IBNR claims reserves. Later, Renshaw (1998) showed that the underlying model for the chain-ladder technique is a generalised linear model for $n_{t,d}$, where the mean is characterized as $\mathbb{E}[n_{t,d}] = \mu + \alpha + \beta_d$. It is shown that this enables the method to process negative incremental claims [15].

Salmon et al. (2015) showed that the conditional multinomial approach could motivate the chain ladder framework [18]. Assume first that the total counts N_t arise from a negative binomial distribution with some mean λ_t and dispersion parameter ϕ; $N_t|\lambda_t, \phi \sim NB(\lambda_t, \phi)$. This is a common assumption when modelling disease count data, where the negative binomial extends the Poisson to allow for overdispersion in data where the amount of susceptible population is unknown, a common problem in observational surveillance data. Second, assuming the counts in each row are conditionally multinomial, $n_t \sim MN(\pi_t, N_t)$; then, it can be shown that the marginal distribution of each $n_{t,d}$ is a negative binomial with mean $\pi_{t,d}\lambda_t$ and dispersion parameter ϕ. In this way, the chain ladder method, which directly models the marginals as negative binomial, can be justified from the conditional multinomial approach (noting, however, that $\phi_{t,d}$ and λ_t cannot be separated.

The field of infectious disease modelling has a rich literature of methodologies but has not previously focused on the challenge of estimating spatiotemporal reporting delays in real-time pub-

lic health control applications. Bastos et al. (2019) developed an integrated dengue monitoring system in Rio de Janeiro, Brazil, which corrects reporting delays using a lognormal survival model [2]. It assumes that the distribution of the counts $n_{t,d,s}$ follows a conditionally independent negative binomial function with the mean $\lambda_{t,d,s}$ and the scale parameter ϕ. They consider approximate Bayesian Inference in a popular subset of structured additive regression models, latent Gaussian models, where the latent fields is Gaussian, controlled by a few hyperparameters and with non-Gaussian response variables. The implementation of the model is fast due to the use of the integrated nested Laplace approximation. They compared the nonspatial version model (temporal) with dependency structure in both time and delay; the spatial version model (spatiotemporal) assumes spatial variability and dependence on borrowing information across the spatial units. The model has been further developed and is being used as a decision-making tool by Brazilian authorities as warning systems, infoDengue [23] and infoGripe [24].

This paper will be dealing with correcting time delays motivated by Bastos et al.; using bayesian inference for the latent Gaussian model, readily implemented in the INLA package. This study considers the posterior marginal of the latent gaussian fields and the posterior marginal of the hyperparameters investigated. The assessment of its usefulness is through implementing the model in the R software.

III. Method

a. Data Description

Dengue is a mosquito-borne infectious disease that causes a substantial economic and public health burden in tropical countries. For example, Dengue imposes a substantial burden in the Philippines; almost ten times higher than estimated for rabies, about twice the burden of intestinal fluke infections, and about 10% of the burden of tuberculosis [22].

Data on reported dengue cases were collected from Gregorio T. Lluch Memorial Hospital, Pala-o Iligan City. The period of the data obtained is from January 2009 to December 2010 of Northern Mindanao dengue cases. It is recorded that there was a significant number of dengue cases during the first six months of 2010. This paper aims to model exhibiting periods of very low activity but also sharp increases and decrease. To show this, 77 weeks from 2009 to 2010 are use, 52 weeks in 2009 and the 1st to 25th weeks (June 2010), for it displays an immediate increase of dengue cases in Northern Mindanao.

Table 1: Time-Delay Data of Dengue Cases in Northern Mindanao

		Delay 0	Delay 1	Delay 2	Delay 3	Delay 4	Delay 5	Delay 6	Delay 7	Delay 8
2009	Week 1	0	9	7	1	0	0	0	0	0
	Week 2	3	8	4	1	1	1	0	0	0
	Week 3	6	4	3	2	3	0	0	0	0
	Week 4	4	11	5	4	0	0	0	2	0
	Week 5	2	8	5	1	0	0	0	1	0
	⋮									
2010	Week 73	0	22	55	28	3	4	5	2	3
	Week 74	10	86	77	10	7	9	2	6	0
	Week 75	26	69	31	17	5	4	9	2	0
	Week 76	34	38	80	27	20	12	4	0	0
	Week 77	4	111	48	51	13	6	0	0	0

Table 2: Data structure for Dengue Cases with missing values

		Delay 0	Delay 1	Delay 2	Delay 3	Delay 4	Delay 5	Delay 6	Delay 7	Delay 8
2009	Week 1	0	9	7	1	0	0	0	0	0
	Week 2	3	8	4	1	1	1	0	0	0
	Week 3	6	4	3	2	3	0	0	0	0
	⋮									
2010	Week 70	1	28	16	2	2	5	7	4	NA
	Week 71	12	19	5	9	20	14	5	NA	NA
	Week 72	0	6	15	35	19	5	NA	NA	NA
	Week 73	0	22	55	28	3	NA	NA	NA	NA
	Week 74	10	86	77	10	NA	NA	NA	NA	NA
	Week 75	26	69	31	NA	NA	NA	NA	NA	NA
	Week 76	34	38	NA	NA	NA	NA	NA	NA	NA
	Week 77	4	NA	NA	NA	NA	NA	NA	NA	NA

Then, train the data or create a run-off triangle data frame illustrated in Table 2. The Table shows the dengue cases' data structure with the missing values (grey) as the delayed counts. Figure 1 presents the method and is utilized to achieve the objective of the study.

b. Latent Gaussian Method

Let $n_{t,d}$ be the notified number of cases in week *t* delayed in *d* weeks, where $t = 1,2,\ldots,T$ and $d = 0,1,2,\ldots,D$. *T* is the last time step for which data is available, and *D* is the maximum acceptable delay, which for disease applications is potentially infinite, but for simplicity, we assume that *D* is bounded. Note that if $t+d > T$, then $n_{t,d}$ is occur-but-not-yet reported, hence unknown.

The INLA framework was design to deal with latent Gaussian models, where we let the counts $n_{t,d}$ follows a conditionally independent Negative binomial likelihood distribution, with mean $\lambda_{t,d}$ and scale parameter ϕ, i.e

1. Observations

$$\mathbf{n_{t,d}} \sim \text{NegBin}(\lambda_{\mathbf{t,d}}, \phi), \qquad \lambda_{t,d} > 0, \quad \phi > 0. \tag{1}$$

The parameterization used here is such that $\mathbb{E}[n_{t,d}] = \lambda_{t,d}$ and $\mathbb{V}[n_{t,d}] = \lambda_{t,d}(1+\lambda_{t,d}/\phi)$. Bayesian approach is used so that the predictive distribution of $n_{t,d}$ for any *t* and *d* (given the data) are readily available, as well as all the associated uncertainty in their estimation.

2. Latent Gaussian Field

The parameter $\lambda_{t,d}$ is linked to a structured additive predictor through a log link function, the logarithm of their mean, $\lambda_{t,d}$, $log(\lambda_{t,d})$. This is to capture the structured temporal variability in $n_{t,d}$, which is characterized as follows:

$$log(\lambda_{t,d}) = \mu + \alpha_t + \beta_d + \gamma_{t,d} + \eta_{w(t)} + X'_{t,d}\delta,$$

- μ is the overall mean count at the log-scale. A fixed effect μ was set an improper prior proportional to one.
- the random effects α_t captures the mean temporal evolution of the count-generating process.
- β_d capture the mean structure of the delay mechanism. These can be modelled using random walks, in the simplest case, first-order ones, i.e,

  $$\alpha_t \sim N(\alpha_{t-1}, \sigma^2_\alpha), \qquad t = 2, 3, \ldots, T$$

  and

  $$\beta_d \sim N(\beta_{d-1}, \sigma^2_\beta), \qquad d = 1, 2, \ldots, D$$

  where half normal HN(τ^2) prior distributions are assumed for σ_α and β_d. These are distribution on $[0, \infty)$ where parameter τ controls the variance. Thinking about α_t and β_d as unknown functions in time and delay, τ controls the "wiggliness" of these functions-the smaller it is, the less wiggly (or in some sence "smooth") the functions will be (i.e, the smaller the first-order differences will be).
- The time-delay interaction term $\gamma_{t,d}$ is modelled as

  $$\gamma_{t,d} \sim N(\gamma_{t-1,d}, \sigma^2_\gamma)$$

  so that there is an independent realisation of a random walk order 1, for each delay column. This term is important, as it allows for changes in the delay mechanism over time.
- Lastly, $\eta_{w(t)}$, where $w(t) = 1, \ldots, 52$ is the week index, is a seasonal component defined as a second-order random effect,

  $$\eta_w \sim N(2\eta_{w-1} - \eta_{w-2}, \sigma^2_\eta),$$

  constrained in such a way that week 1 and week 52 are joined.
- $X'_{t,d}$ is a matrix of temporal and delay-related covariates with associated vector of parameters δ.

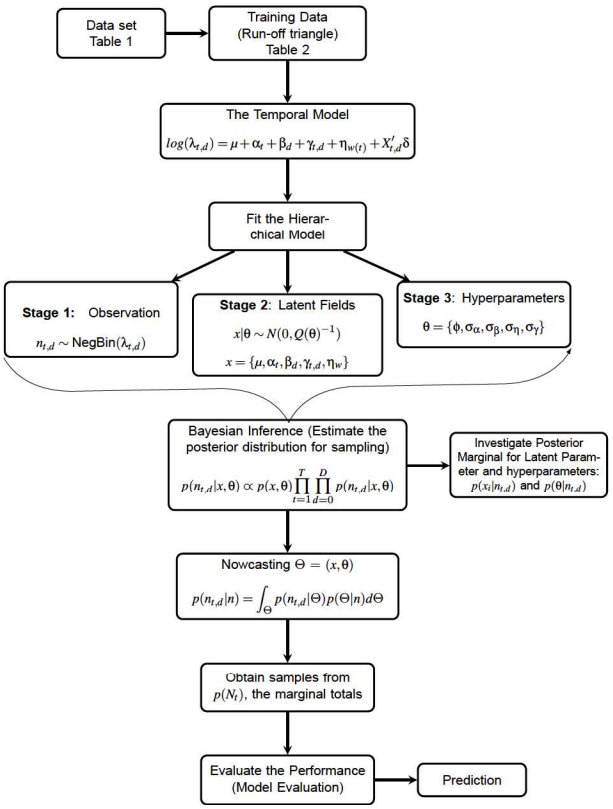

Fig. 1: Schematic Diagram of the Proposed Hierarchical Bayesian Temporal Model to Correct Delays

All of the components $\alpha_t, \beta_d, \eta_{w(t)}$, and$\gamma_{t,d}$ are constrained to sum to zero, to allow identifiability of the intercept μ.

3. Hyperparameters

The Hierarchical model is then completed with an approximate prior distribution for the hyperparameters of the model $\theta = (\phi, \sigma^2_\alpha, \sigma^2_\beta, \sigma^2_\gamma, \sigma^2_\eta)$, where $\phi \sim Gamma(1, 0.1), \sigma^2_\alpha \sim HN(\tau = 0.1)$, $\sigma^2_\beta \sim HN(\tau = 1), \sigma^2_\eta \sim HN(\tau = 1), \quad \sigma^2_\gamma \sim HN(\tau = 0.1)$.
Assume an exponential $Exp(0.1)$ prior distribution for ϕ with mean 10 and standard deviation 10. This is a weakly informative prior that places more probability over smaller values of ϕ and thus assumes the preference of the negative binomial to the Poisson. A gamma prior is then set to ϕ.

c. Integrated Nested Laplace Approximation

This section presents the Bayesian inference for the latent Gaussian model to obtain a posterior marginal distribution.

The joint posterior distribution for $\Theta = (\mu, \alpha_t, \beta_d, \gamma_{t,d}, \eta_w, \sigma^2_\alpha, \sigma^2_\beta, \sigma^2_\eta, \sigma^2_\gamma, \phi)$, given all the

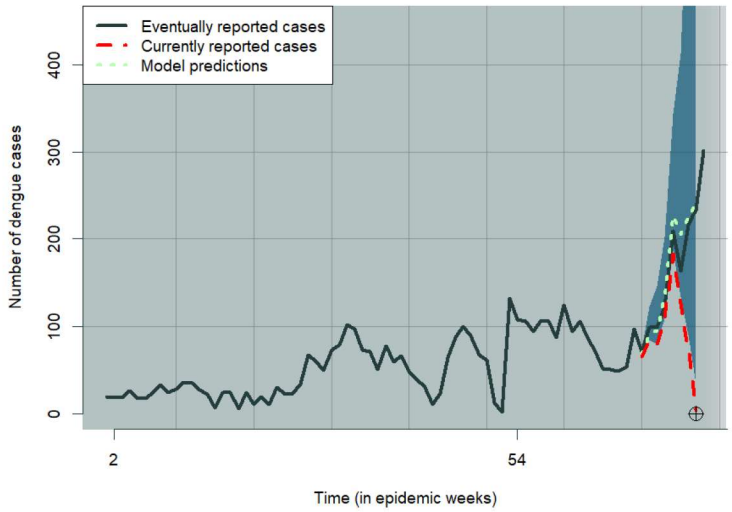

Fig. 2: Time Series of Reported of Dengue Cases in Northern Mindanao from 2009 to 2010

observed data $\mathbf{n} = n_{t,d}$ is given by

$$p(\Theta|n) \propto p(\Theta) \prod_{t=1}^{T} \prod_{d=0}^{D} p(n_{t,d}|\Theta)$$

where $p(n_{t,d}|\Theta)$ is negative binomial density function (1), and $p(\Theta)$ is the joint prior distribution given by the product of the prior distributions for $\phi, \sigma_\alpha^2, \sigma_\beta^2, \sigma_\gamma^2, \sigma_\eta^2$ and the random effects distribution. To investigate the posterior marginal distribution of hyperparameters and latent Gaussian models, set the latent Gaussian vector x, $\mathbf{x} = \{\mu, \alpha_\mathbf{t}, \beta_\mathbf{d}, \gamma_\mathbf{t,d}, \eta_\mathbf{w}\}$ and hyperparameter vector θ, $\theta = (\phi, \sigma_\alpha^2, \sigma_\beta^2, \sigma_\gamma^2, \sigma_\eta^2)$. Now, compute from

$$p(x, \theta|n) \propto p(x, \theta) \prod_{t=1}^{T} \prod_{d=0}^{D} p(n_{t,d}|x, \theta),$$

the posterior marginal $p(x_i|n_{t,d})$, for some i; and $p(\theta_i|n_{t,d})$, for some i.

d. Nowcasting

In any given step T, there are a number of occurred-but-not-yet-reported (missing) values $n_{t,d}, t = T - D + 1, \ldots, T$; $d = 1, \ldots, D$ (see the grey cells in Table 2), as well as the marginal totals $N_{T-D+1}, \ldots, N_T$. Of primary interest is of course N_T, which needs to be nowcasts; however, hindcasts of $N_{T-D+1}, \ldots, N_T - 1$ may also be of interest, especially if one wants to quantify the rate of increase or decrease in the counts.

From a Bayesian perspective, this is a prediction problem where all the missing $n_{t,d}$ can be estimated from the posterior predictive distribution

$$p(n_{t,d}|n) = \int_\Theta p(n_{t,d}|\Theta) p(\Theta|n) d\Theta,$$

where $\mathbf{n}$ denotes all the data used to fit the model.

IV. Results and Discussions

a. Time Series of Reported Dengue Cases

The available data consist of weekly counts of the number of dengue cases in Northern Mindanao for the time period January 2009 to June 2010, along with the associated delay information.

In Figure 2, the solid dark blue line shows the eventually reported number of dengue cases per week. The red dashed line shows the current reported number of cases from the 18th to the 25th week of 2010 (circled cross). The green dotted line shows the model estimates for this period, along with 95% prediction intervals in blue.

The model used for estimation is the one given in Equation (1) with D = 8 and $X_{t,d} = 0$ as no covariated information was available. The model was used to correct the total number of cases N_t in that particular week, but also for the the 7 weeks preceding it, as shown in Figure 2 (green dotted line), along with 95% prediction intervals.

The plot indicates that the predictions identify the fact that there was an increase in the number of dengue cases. Furthermore, it can be observed how similar the model prediction (green dotted line) and the eventually reported cases (dark blue line) increase the same way up to the 25th week of 2010.

Suppose the current number of cases is considered, a public health decision maker could take wrong actions because it appears to be decreasing in June 2010. However, it can be observed that the eventually reported cases are corrected for delays and occasionally for misclassification using laboratory confirmation tests or other things that might cause delayed reports.

b. Model Evaluation

A series of checks are performed to ensure that the model provides a good fit for the data. These are how much the prediction deviated from the actual value, checking of predictive samples of the total, the sample means and sample variance of the total N_t and the data's temporal dependence.

1. Predicted Totals against Observed (sorted) values

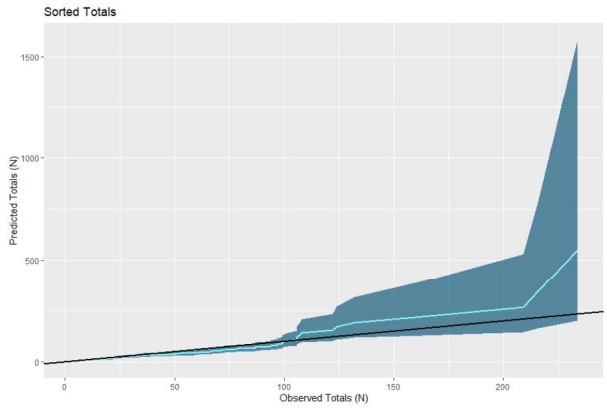

Fig. 3: Predicted Totals Plotted against the respective observed (sorted) values

Figure 3 shows the predicted N_t defined as the means of these distributions, plotted against the observed N_t sorted in ascending order. The black line represents the $x = y$, showing how much the prediction deviated from the actual value. The blue line represents where the points would fall if all predicted values perfectly matched the observed ones and 95% prediction intervals were also added. Figure 3 indicates that the model estimates for the dengue cases in Northern Mindanao capture the rank of the observed values very well, bearing in mind that eight weeks of the 77 values are based on data the model has not seen.

2. Sample Mean and Sample Variance of the total N_t

Secondly, to check how well the predictive samples of the totals are captured. Figure 4 illustrates the Predictive distributions for the sample mean and variance of the totals N_t.

Vertical lines indicate the observed values, whereas the quoted probabilities indicate the tail areas of the observed values (values less than 0.025 or over 0.975 suggest that the model does not represent the observed value well.) Therefore, since the sample mean is 0.6932, and the sample variance is 0.1618, Figure 4 indicates that the sample mean and variance are well-captured (like, are not extreme concerning the distribution).

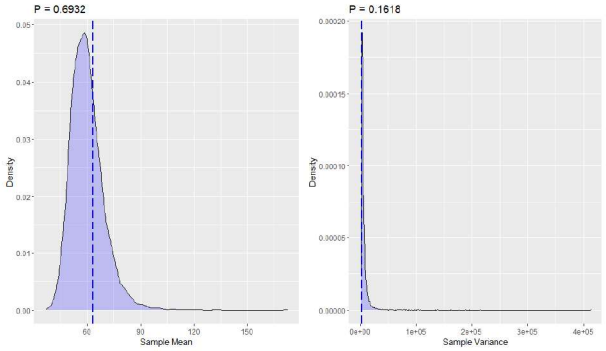

Fig. 4: Sample Mean and Sample Variance

3. Sample Autocorrelation of the Totals N_t

Temporal dependence is essential in being able to detect outbreaks. One of the most important things to consider is the dependence between various points in the time-series data. To check whether temporal dependence in N_t is well captured, Figure 5 shows the sample autocorrelation in the N_t for the eight lags.

Figure 5 shows that the model captures the temporal dependence in N_t effectively because none of the observed values (vertical lines) are extreme regarding the respective predictive distributions. Hence, our model's estimate in dengue cases is widely believed to depend highly on the recent behavior of the number of cases. This implies that it can be used for behavior prediction, as it provides a hint as to show how the number of dengue cases will behave shortly.

c. Estimate and Rolling Prediction

1. Overall Temporal, Delay and Seasonal Variation

Figure 6 shows the estimates of three components, namely, α_t the overall temporal evolution in the counts, β_d the delay structure, and $\eta_{w(t)}$ the seasonal variability.

$$log(\lambda_{t,d}) = \mu + \alpha_t + \beta_d + \gamma_{t,d} + \eta_{w(t)}, \qquad (2)$$

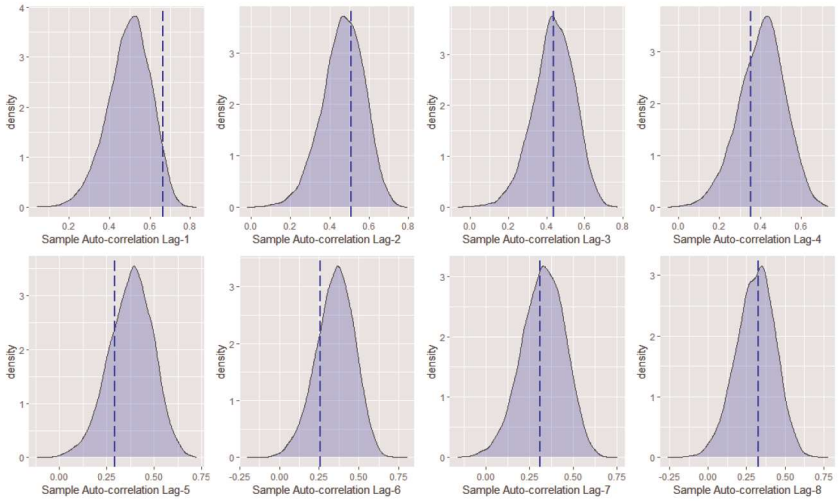

Fig. 5: Predictive distribution for the sample autocorrelation of the totals

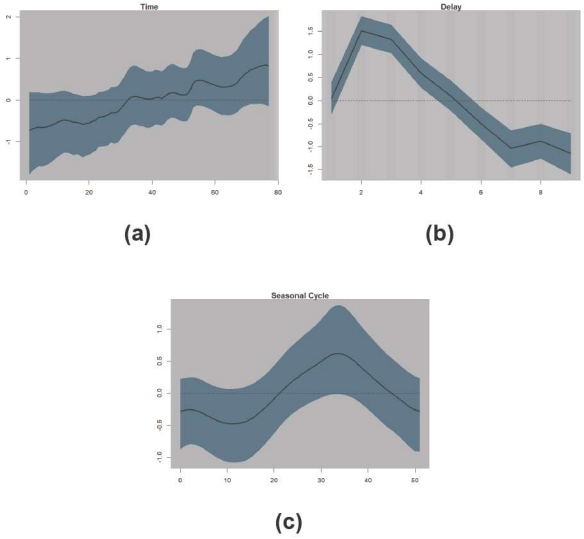

Fig. 6: Estimates of the overall temporal variation, overall delay structure and seasonal variability

Figure 6(a) suggests that the overall temporal effect slightly increases at first, then continue in increasing smoothly, perhaps reflecting an increase in the susceptible population. In Figure 6(b), the delay structure is increasing at first, then decreases; as would be expected-the more time goes by, the more cases are being reported. However, in Figure 2, the eventually reported dengue cases already display no sign of seasonality; no wonder that in Figure 6(c), no seasonal component was captured by the model.

d. Nowcasting at different time frames

This section shows the weekly rolling predictions as of the 25th week (June 14 to 20) of 2010 to the 36th week (August 16 to 22) of 2010. The black line indicates the eventually reported number of cases; the red dashed line shows the number of currently reported cases, and the green dotted line shows model prediction along with 95% prediction intervals. The circled cross symbol indicates the week $T = 77, 78, \ldots, 88$. This period was chosen specifically to test the ability of the model to capture the an outbreak as well as the relatively sharp decline in the eventually reported number of cases (black line).

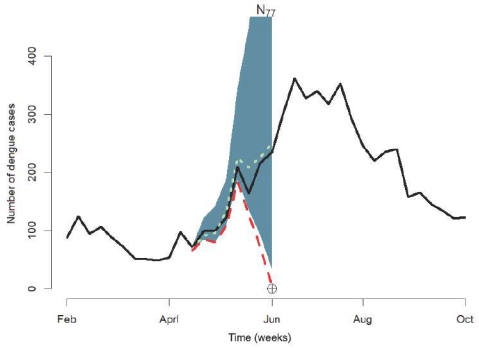

Fig. 7: N_{77} (77th week prediction)

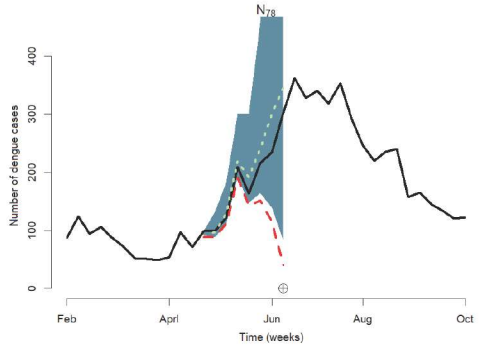

Fig. 8: N_{78} (78th week prediction)

Figures 7, 8, and 9 show the time series of the 77th-week to 78th-week prediction of dengue

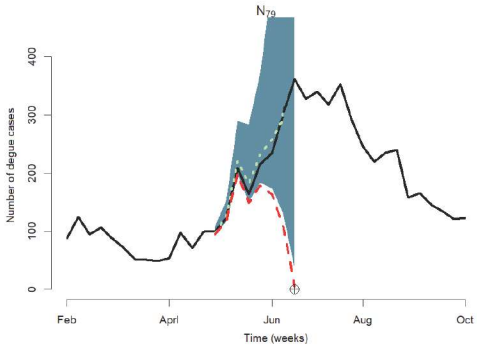

Fig. 9: N_{79} (79th week prediction)

cases in Northern Mindanao. It can be observed that the currently reported cases in the 25th week to the 27th week of 2010 appear to be decreasing (red dashed line). However, without delay, calculations show that the number of dengue cases is increasing (solid black line). Based on the currently reported cases, the result could lead to wrong actions or decisions, making the public believe that no preventive measures are to be taken since dengue cases are declining. On the other hand, the model (green dotted line) closely captures the increasing trend of the eventually reported dengue cases. Thus, it suffices to show that the one to three weeks ahead prediction gives a good result in predicting the number of dengue cases.

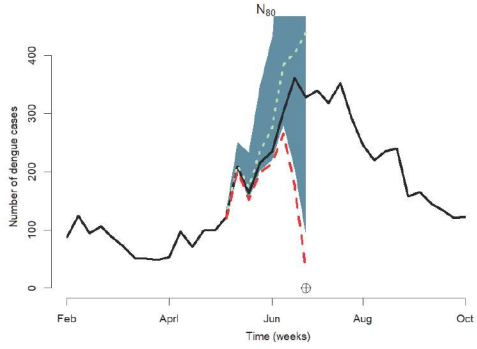

Fig. 10: N_{80} (80th week prediction)

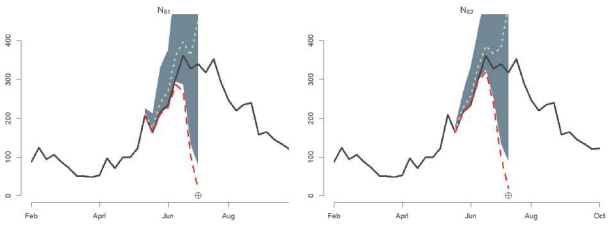

Fig. 11: N_{81} and N_{82} (83th and 81st week prediction)

Figure 10 and Figure 11 show the time series of the 80th to the 82nd-week prediction of dengue

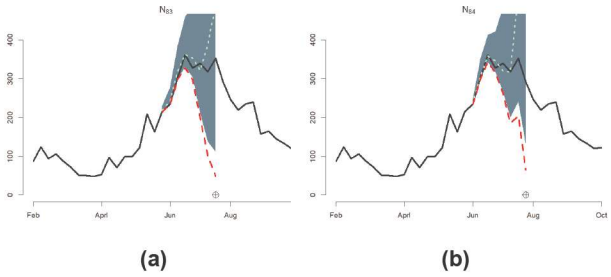

Fig. 12: N_{83} and N_{84} (83th and 84st week prediction)

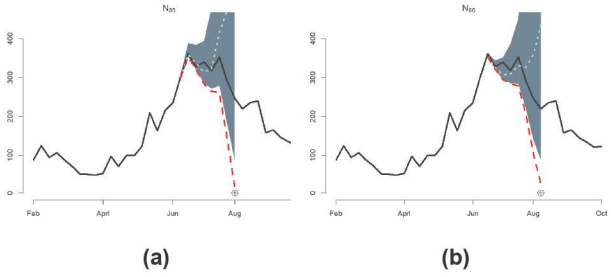

Fig. 13: N_{85} and N_{86} (85th and 86st week prediction)

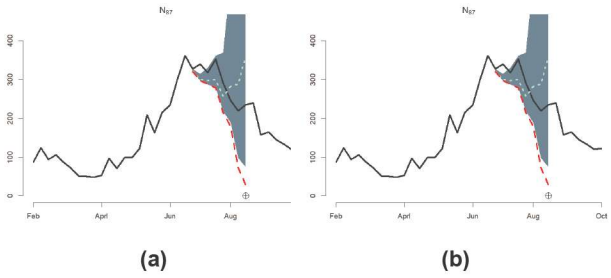

Fig. 14: N_{87} and N_{88} (87th and 88st week prediction)

cases in Northern Mindanao. The result indicates that the currently reported cases in the 28th to 30th week of 2010 appear to be decreasing (red dashed line). Nevertheless, the calculation shows that the number of dengue cases is much higher (solid black line) than the currently reported cases when there is no delay.

Furthermore, the model (green dotted line) in Figure 10 and Figure 11 (b) predicts the increasing trend but misses some slight decrease in the eventually reported cases; the estimate predicted more dengue cases than the actual counts, especially in Figure 11 (a). However, the model prediction from N_{85} to N_{88} does not capture the eventually reported number of cases as it decreases its counts.

In summary, Figure 7, 8, and 9 show that in N_{77}, N_{78}, and N_{79} weeks the model (green dotted line) captures the increase in the eventually reported number of cases. However, in Figure 10 to 11 show that in N_{80} to N_{82}, the model predicted more than the eventually reported number of dengue cases. Figure 12, 13, and 14 shows that in N_{83} to

N_{88} illustrations fail to capture the decreasing trend as the model prediction of the number of dengue cases appears to increase. But since, most of the eventually reported counts are within the 95% prediction intervals, particularly for time T (indicated by the circled cross), which is the most important value.

V. Conclusions

Implementing the Bayesian Hierarchical Approach is a flexible model for time-delayed structured applied to a disease count data. This enables the estimation of the missing (observable) data as $n_{t,d} \sim NegBin(\lambda_{t,d}, \phi)$ to perform nowcasting. To study and illustrate the model's performance, we used dengue data in Northern Mindanao, and this demonstrated the desired flexibility and complexity of the framework. Temporal dependence is significant in detecting outbreaks; although there are no covariates in the model, this is a trivial task in the R-INLA implementation. Implementing the models in the Bayesian framework is extremely fast because we use the Laplace approximation (INLA) to compute samples from the (marginal) posteriors. Surveillance and warning systems relying on reported incidence to assess risk can be misinformed if the delay is not somehow corrected; thus, accurate estimates are of utmost importance. Therefore, this method can be a helpful decision-making tool in surveillance systems. Forecasting using the predictive distribution of counts implies that the predictions from the proposed models can be readily utilized in a decision-theoretic framework for issuing warnings.

VI. Acknowledgment

I express my profound thanks to God, to my adviser, DOST-ASTHRDP and to family in supporting me greatly and were always willing to help me. To God be all the glory.

References

[1] Agresti, A. (2018). *An Introduction to categorical data analysis*. John Wiley & Sons.

[2] Bastos LS,. Economou T, Gomes MFC, et al. *A modelling approach for correcting reporting delays in disease surveillance data.* Statistics in Medicine. 2019;38:4363–4377. https://doi.org/10.1002/sim.8303

[3] Brookmeyer R, Damiano A. *Statistical methods for short-term projections of AIDS incidence*. Stat Med. 1989;8(1):23–34

[4] Chawarat R., et al. (2020) *Bayesian spatiotemporal modeling with sliding windows to correct reporting delays for real-time dengue surveillance in Thailand* International Journal of Health Geographics.

[5] Dobson, A., Barnett, A. (2008) *An Introduction to Generalized Linear Models.* Taylor & Francis Group

[6] Farrington C.P., et.al. *A statistical algorithm for the early detection of outbreaks of infectious disease.* J Royal Stat Soc, Ser (Stat Soc). 1996;159(3):547-563.

[7] Hohle M, an der Heiden M. (2011).*Bayesian nowcasting during the STEC O104:H4 outbreak in Germany*. Biometrics. 2014;70(4):993-1002. https://doi.org/10.1111/biom.12194

[8] Klaucke DN, Buehler JW, et.al. [1988] Guidelines for evaluating surveillance systems. Morb Mortal Wkly Rep. 37(Suppl5):1-18.

[9] Lawless, J. F. (1994). Adjustments for reporting delays and the prediction of occurred but not reported events. The Canadian Journal of Statistics, 22(1), 15–31.

[10] Lin H, Yip PS, Huggins RM. *A double-nonparametric procedure for estimating the number of delay-reported cases.* Stat Med. 2008;27(17):3325–39.

[11] Mack T. (1993) *Distribution-free calculation of the standard error of chain ladder reserve estimates. ASTIN Bulletin.* 23(2):213-225.

[12] Midthune DN, Fay MP, Clegg LX, Feuer EJ. Modeling reporting delays and reporting corrections in cancer registry data. J Am Stat Assoc. 2005;100(469):61–70.

[13] Noufaily A, Farrington P, Garthwaite P, et.al. Detection of infectious disease outbreaks from laboratory data with reporting delays. J Am Stat Assoc. 2016;111(514):488–99

[14] Noufaily A, Enki DG, Farrington P, Garthwaite P, Andrews N, Charlett A. *An improved algorithm for outbreak detection in multiple surveillance systems.* Stat Med. 2013;32(7):1206–22.

[15] Renshaw AE, Verrall RJ. *A stochastic model underlying the chain-ladder technique.* Br Actuar J. 1998;4(04):903-923.

[16] Rosinska, et al. (2018) *Potential Adjustment methodology for missing data and reporting delay in the HIV surveillance system*, European Union/European Economic Area, 2015. Eurosurveillance, 23. https://doi.org/10.2807/1560-7917.ES.2018.23.23.1700359

[17] Rue H, et al. *Bayesian computing with INLA: a review.* (2017) Annu Rev Stat Appl.

[18] Salmon M, et al. *Bayesian outbreak detection in the presence of reporting delays.* Biometrical Journal. (2015) ;57(6):1051-1067. https://onlinelibrary.wiley.com/doi/abs/10.1002/bimj.201400159

[19] Sara M. and Andrea R. (2019) *Integrated Nested Laplace Aprroximation (INLA)*

[20] Shoemaker, J. S., Painter, I. S., & Weir, B. S. (1999). *Bayesian statistics in genetics: A guide for the uninitiated.* Trends in Genetics 15(9), 354-358.

[21] Willink, R., & Lira, I. (2005). *A United interpretation of different uncertainty intervals.* Measurement, 38(1), 61-66.

[22] Undurraga, E., et al. (2017) *Disease Burden of Dengue in the Philippines: Adjusting for Underreporting by Comparing Active and Passive Dengue Surveillance in Punta Princesa, Cebu City*. The American Society of Tropical Medicine and Hygiene

[23] https://info.dengue.mat.br

[24] http://info.gripe.fiocruz.br