A Bayesian Hierarchical Model for COVID-19 Cases in Mindanao Philippines[1]

Jejemae D. Nacion[2] and Bernadette F. Tubo
Department of Mathematics and Statistics
MSU-Iligan Institute of Technology Iligan City, Philippines

ABSTRACT

A Bayesian hierarchical modelling approach is utilized to nowcast COVID-19 cases in Mindanao, Philippines for the year 2020 to 2021. A spatio-temporal model is considered and the proposed methodology explores the possibility of a flexible way of correcting the time and space delayed reports of the COVID-19 cases for a duration of 4 weeks for the 27 provinces in Mindanao via a Bayesian approach. The goal of the modelling approach is to include parameters that will correct reporting delays in the dataset and derive a model using the Integrated Nested Laplace Approximation (INLA). The study shows that the proposed model was able to capture the increasing trend of the COVID-19 disease counts, that is, the prediction counts derived are closer to the true count compared to the currently reported counts of COVID-19 cases which showed a decreasing behavior. The ability of the proposed model to nowcast statistically significant estimates, particularly, for epidemic counts of COVD-19 in the presence of report delays may aid health authorities to have effective control measures and issuance of warnings to the public.

Keywords: *Bayesian inference, spatio-temporal model, reporting delay, nowcasting*

I. INTRODUCTION

Epidemiological surveillance is the ongoing and systematic collection, analysis, and interpretation of health data in the process of describing and monitoring a health event (Klaucke et al. 1988). Timeliness, which relates to the speed or delay between actions in a monitoring system, is one attribute of effective surveillance. Reporting delays are well-known issues that breach timeliness. Because of defects or "lags" in the data collection method, the available count data, for a time, represents the truth less accurately (Swaan et al. 2018). The associated bias from delayed reports affects parameter estimates, predictions, and statistical inferences. This added uncertainty could reduce the confidence of the policymakers and warning systems in the public health decision-making process (Stoner and Economou 2020).

In the Philippines, the currently experienced COVID-19 disease highlighted the problem of disease surveillance. COVID-19, given its complexity and behavior, exposed the problem of delayed reporting on disease occurrences. Reporting delay is affected by conflicting factors due to the disease incidence such as (1) a prolonged interval between the time an individual recognizes symptoms and is able to seek care and receive confirmatory testing, (2) administrative backlogs and delays in the acquisition, processing, and ultimate reporting of information, and (3) the length of time necessary to conduct a full case investigation (Kline et al. 2021). However, significant choices should be made continuously notwithstanding the way that the latest data is likely incomplete. Hence, on that account, the methodology is needed to help provide a clearer picture to decision-makers in the face of the uncertainty from delays in reporting.

Studies related to reporting delays have been introduced in the past by the authors Brookmeyer and Gail 1988 and Kalbfleisch and Lawless 1989 who both dealt with back-calculation and initiating events (events in the past) for AIDS incidence. Lawless in 1994 also dealt with the estimation allowing random temporal fluctuations in reporting delay. In recent studies, van de Kassteele et al. 2019, Stoner and Economou 2020, McGough et al. 2020, and Kline et al. 2022 enhance the model on its flexibility and interpretability, and extended prior works within a Bayesian framework. Also, Rotejanaprasert et al. 2020 incorporate spatial dependence into temporal models using a Bayesian framework with sliding

[1] Presented at the 15th National Convention on Statistics, 03-5 October, 2022, Crowne Plaza Manila Galleria, Ortigas Center, Quezon City and published in this issue as a *Technical Paper*.
[2] Corresponding Author: *jejemae.nacion@g.msuiit.edu.ph*

windows. As such, in the case of spatio-temporal models, the joint distribution would describe the behavior of the process at all spatial locations and at all times. In the study of Bastos et al. 2019, a Bayesian hierarchical modelling approach was used to correct reporting delays and quantify the associated uncertainty in the missing values. The author's approach is illustrated by dengue fever incidence data in Rio de Janeiro and severe acute respiratory infection data in the state of Paraná, Brazil

Motivated by the works of Bastos et al. 2019, this paper considers the Bayesian hierarchical approach for correcting report delays, suitable for a wide range of spatio-temporal count data and applies it to counts of COVID-19 cases in the provinces of Mindanao

II. REVIEW OF RELATED LITERATURES

The problem of *occurred-but-not-yet-reported* cases is well known from the HIV/AIDS outbreak. Different statistical approaches have been proposed in the past to handle delayed reporting. A standard reference is based on the work of Lawless (1994). Moreover, some early contributions in application to the estimation of HIV/AIDS incidences were Lagakos, et al. (1998) who developed nonparametric methods, Kalbfleisch and Lawless (1989), and Harris (1990), who both considered Poisson processes, and Kalbfleisch and Lawless (1991), who defined the regression models that enabled easy tests and covariate effect estimation using right truncated data. The estimation procedure "back-calculation" by Brookmeyer and Gail (1988) and Bacchetti, et al. (1993) applied to AIDS incidence data in the United States, refers to the reconstruction of the past history of first events (onset date) that must have occurred to give rise to the observed pattern of second event cases (date report confirmed), under the assumption of a known delay distribution. Furthermore, statistical approaches are proposed not only to epidemiological data but also in actuarial science like Renshaw and Verall (1998). In recent literature, like Höhle and van der Hieden (2014), van de Kassteele, et al. (2019), and Bastos, et al. (2019), instead of back-calculation, '*nowcasting*' is often used for estimating the current number of events using only the available partial information reported.

These various approaches may be broadly classified in two groups: one which models the delay counts $(n_{t,d})$ jointly but also conditionally on the total (N_t), in conjunction with a separate model for the total, whose works includes the work of Salmon, et al. (2015), Höhle and van der Heiden (2014), and Stoner and Economou (2019). Other studies like in the works of Bastos, et al. (2019) and McGough, et al. (2020) where the delayed counts are modelled marginally without explicitly modelling using historical information on the totals.

Bastos, et al. (2019) and Rotejanaprasert, et al. (2020) applied this second approach to spatio-temporal SARI data from Brazil and to dengue fever data from Thailand, respectively. It was a generalization of older chain-ladder approaches where they extended the model with negative binomial marginals to allow for spatio-temporal variation in the counts, as well as covariate effects. The approach is quite flexible, as it can potentially incorporate a wide variety of temporal, spatial and spatio-temporal structures.

Finally, as for the approximation process, Bastos, et al. 2019 performed a comparison between the nonspatial version of the model when implemented using both Markov Chain Monte Carlo (MCMC) and Integrated Nested Laplace Approximation in R (R-INLA). Approximations involved in using the Integrated Nested Laplace Approximation (INLA) approach gained a significant increase in computational speed. Accordingly, it is a reasonable compromise to the gain in computational speed with the R-INLA model taking a matter of seconds compared to hours of MCMC.

III. SPATIO-TEMPORAL MODEL

In a spatio-temporal model, each of the different spatial regions or locations can independently execute the model. In fact, it would be more logical to include all of the data into one analysis by expanding the model to account for regional variation in both the counts' emergence and the delay mechanism. This enables both inference on how the delay mechanism differs across the various areas

and the pooling of information to improve estimation in spatial locations with less data. Therefore, spatial (Gaussian) random effects are included in the proposed model. Taking into account spatial variance, where $s = S$ designates a place or region in some spatial domain s, the model is given by

$$n_{t,d,s} = \text{NegBin}\big(\lambda_{t,d,s}, \phi\big),\ \lambda_{t,d,s} > 0,\ \phi > 0 \tag{1}$$

where $n_{t,d,s}$ is the number of occurrences in spatial location s and time point t, reported with delay d time points. The mean is then modelled as

$$log\big(\lambda_{t,d,s}\big) = \mu + \alpha_t + \beta_d + \gamma_{t,d} + \eta_{t,d} + \psi_s + \beta_{d,s} + \boldsymbol{X'_{t,d,s}\delta} \tag{2}$$

with $\boldsymbol{X'_{t,d,s}}$ being a model matrix that may also include spatially varying covariates. The quantities α_t and β_d are interpreted as the overall temporal and delay evolution across space, respectively. The component $\beta_{d,s}$ captures the way in which the delay structure varies across space, whereas ψ_s describes the overall spatial variability and dependence in the counts. This particular formulation is motivated by the application to COVID-19 data, where the spatial region is fairly small so the temporal effects α_t are not assumed to vary with space. Given the implementation of the model in R-INLA, various possible choices exist for the specific formulation of $\beta_{d,s}$ and ψ_s. The space-time or space-delay interactions can range in complexity, from spatially and temporally unstructured Gaussian processes to non-separable formulations (Knorr-Held (2000) and Blangiardo, et al. (2013)). The spatial effect ψ_s can be defined by an intrinsic autoregressive (IAR) process (Besag, et al., 1991) if the data are counts in areal units to allow similar temporal variation in neighboring areas. Equally, ψ_s can be defined by a stationary Gaussian process if the data are counts in point locations, for example, so that spatial dependence decreases exponentially with distance.

In the application of the model, where space is divided into a number of administrative areas, the model uses the type I space-time interaction as proposed by Knorr-Held (2000). This is a formulation where

$$\beta_{d,s} \sim N\big(\beta_{d-1,s}, \omega_\beta^2\big) \tag{3}$$

is an independent first-order random walk for each area s, and where $\psi_s = \psi_s^{IAR} + \psi_s^{ind}$, that is, the sum of a spatially structured IAR process:

$$\psi_s^{IAR} | \psi_{s' \neq s}^{IAR} \sim N\left(\frac{\sum_{s' \neq s} \omega_{s,s'} \psi_{s'}^{IAR}}{\sum_{s' \neq s} \omega_{s,s'}}, \frac{\sigma_{IAR}^2}{\sum_{s' \neq s} \omega_{s,s'}}\right) \tag{4}$$

and spatially unstructured random effects $\psi_s^{ind} \sim N(0, \sigma_{ind}^2)$. Here, σ_{IAR}^2 controls the strength of spatial dependence and σ_{ind}^2 is the variance of the spatially unstructured effects.

Construct of a Parameterized Prior Distribution

Let $n_{t,d,s}$ be the notified number of cases in week t delayed in d weeks occur in region s, where $t = 1,2,\dots,T$, $d = 0,1,2,\dots,D$, and $s = 1,2,\dots,27$ for the provinces of Mindanao. Note that if $t + d > T$, then $n_{t,d,s}$ is unknown.

Assume a negative binomial likelihood as follows:

$$n_{t,d,s} \sim N\text{egBin}\big(\lambda_{t,d,s}, \phi\big),$$

for any $t = 1,2,\dots,T$, $d = 0,1,2,\dots,D$, and $s = 1,2,\dots,27$. A gamma prior is set to ϕ, and the rate $\lambda_{t,d,r}$ is given by

$$\ln\big(\lambda_{t,d}\big) = \mu + \alpha_t + \beta_d + \gamma_{t,d} + \beta'_{d,s} + \Psi_s^{(IAR)} + \Psi_s^{(ind)}.$$

A fixed effect μ was set as an improper prior proportional to one. The random effects, $\{\alpha_t\}, \{\beta_d\}, \{\gamma_{t,d}\}$ were set with different random walk priors, $\{\beta'_{d,s}\}$ is an independent Gaussian space-delay random effect, and the sum $\Psi_s^{(IAR)} + \Psi_s^{(ind)}$ is model as a Besag-York-Mollier (*bym)* model random effect, all implemented in the INLA package. The hyperparameters, all random effects standard deviations were assumed to be half normal, or truncated normal at $(0, \infty)$, with a distinct standard deviation τ for each random effect, denoted as $HN(\tau)$. Table 1 summarizes all priors and hyperpriors for the COVID-19 model.

Table 1. Prior distributions of the parameters.

Parameter	Distribution	In INLA
ϕ	$\phi \sim Gamma(1,0.1)$	$e^{\phi} \sim loggamma(1.0,0.1)$
μ	$p(\mu) \propto 1$	*default*
$\alpha_t \mid \alpha_{t-1}, \sigma_\alpha^2$	$\alpha_t - \alpha_{t-1} \mid \sigma_\alpha^2 \sim N(0, \sigma_\alpha^2)$	*1st order random walk (rw1)*
σ_α^2	$\sigma_\alpha^2 \sim HN(\tau = 0.1)$	*Half normal sd(0.1)*
$\beta_d \mid \beta_{d-1}, \sigma_\beta^2$	$\beta_d - \beta_{d-1} \mid \sigma_\beta^2 \sim N(0, \sigma_\beta^2)$	*1st order random walk (rw1)*
σ_β^2	$\sigma_\beta^2 \sim HN(\tau = 0.1)$	*Half normal sd(0.1)*
$\gamma_{d,t} \mid \gamma_{d,t-1}, \sigma_\gamma^2$	$\gamma_{d,t} - \gamma_{d,t-1} \mid \sigma_\gamma^2 \sim N(0, \sigma_\gamma^2)$	*1st order random walk (rw1)*
σ_γ^2	$\sigma_\gamma^2 \sim HN(\tau = 0.1)$	*Half normal sd(0.1)*
$\beta'_{d,s} \mid \sigma_{\beta'}^2,$	$\beta'_{d,s} \mid \sigma_{\beta'}^2, \sim N\left(0, \sigma_{\beta'}^2,\right)$	*Independent gaussian (iid)*
$\sigma_{\beta'}^2,$	$\sigma_{\beta'}^2, \sim HN(\tau = 0.1)$	*Half normal sd(0.1)*
$\Psi_s \mid \sigma_{IAR}^2, \sigma_{ind}^2$	$\Psi_s = (\Psi_s^{IAR} + \Psi_s^{ind}, \Psi_s^{ind})$	*bym model*
σ_{IAR}^2	$\sigma_{IAR}^2 \sim HN(\tau = 0.1)$	*Half normal sd(0.1)*
σ_{ind}^2	$\sigma_{ind}^2 \sim HN(\tau = 0.1)$	*Half normal sd(0.1)*

By construction in INLA, the *bym* model is a representation of an IAR model added by an unstructured independent random effect.

Parameter Estimates

In this paper, a Bayesian framework is used in to consider the unknown parameters as random variables. Each parameter is given a probability distribution and is approximated by statistical models, as shown in Table 2.

Table 2. Random Effects Model descriptions.

Name	Model
Time	RW1 model
Delay	RW1 model
Time-Delay	RW1 model
Space	*bym* model
Space-Delay	IID model

The model described in Section 3.2 is considered for parameter estimation, namely, $n_{t,d,s} \sim NegBin(\lambda_{t,d,s}, \phi)$ and

$$log(\lambda_{t,d,s}) = \mu + \alpha_t + \beta_d + \gamma_{t,d} + \beta_{d,s} + \psi_s^{IAR} + \psi_s^{ind} \tag{5}$$

with $t = 1, \cdots, 66\ (weeks)$, $d = 0, \cdots, D\ (maximumdelay)$, and $s = 1, \cdots, 27\ (provinces)$. The quantities are defined as follows:

- μ is the log-scale overall mean count and 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 using random walks, in the simplest case, first-order ones.
- $\gamma_{t,d}$ is the time-delay interaction term.
- $\beta_{d,s} \sim N\left(\beta_{d-1,s}, \omega_\beta^2\right)$ allow for unstructured spatio-delay variability.
- ψ_s^{IAR} is spatially structured according to an IAR process with a neighboring structure defined by a 27 × 27 adjacency matrix **W,** where $w_{i,j} = 1$ if the province *i* is an administrative neighbor of province *j* and $w_{i,j} = 0$, otherwise.
- $\psi_s^{ind} \sim N\left(0, \sigma_\psi^2\right)$ captures spatially unstructured variability.

When constructing a spatio-temporal statistical model, the Bayesian approach requires us to assign prior probability distributions (a mathematical way to reflect our prior belief) to all the unknown parameters. This gives rise to several advantages when analyzing spatio-temporal data, impacting on every aspect of statistical analysis from model building, parameter estimation, and interpretation to model evaluation.

Nowcasting

Nowcasting is defined as the process of predicting the present, the very recent past, and the very near future using time series data known to be incomplete (Blangiardo, et al. 2013). At any given time step T, there are a number of *occurred-but-not-yet-reported* (missing) values $n_{t,d,s},\ t = T - D + 1, \cdots, T; d = 1, \cdots, D; s = 1, \cdots, 27$, as well as the marginal totals $N_T - D + 1, \cdots, N_T$. The total, N_T is obviously of primary interest and must be nowcast; however, hindcasts of $N_T - D + 1, \cdots, N_{T-1}$ may also be of interest, particularly if one wishes to quantify the rate of increase or decrease in the counts. The posterior predictive distribution given by

$$p\left(n_{t,d,s}|\mathbf{n}\right) = \int_{\theta} p\left(n_{t,d}|\theta\right) p(\theta|\mathbf{n}) d\theta, \tag{6}$$

where **n** stands for all the data used to fit the model, can be used to estimate all the missing $n_{t,d,s}$ in this prediction problem. However, analytical solutions are not possible, but Monte Carlo simulations can be used to approximation. In reality, simulation of values from a negative binomial $p\left(n_{t,d}|\theta\right)$ for each sample with posterior, $p(\theta|\mathbf{n})$ can generate a rough sample from the predictive distribution, $p(n_{t,d}|\boldsymbol{n})$. Equivalent samples can be produced from $p(N_t)$, the marginal totals, once posterior predictive samples of $N_{t,d}$ are available. As shown in small-area estimation, for instance in the work of (Vandendijck et al. 2016), samples from an approximation of the joint posterior distribution can be acquired from R-INLA using the *inla.posterior.sample() function*.

IV. MODEL ESTIMATION

The goal of this study is to use the proposed model to correct reporting delays by taking into consideration the changes in the delay mechanism's spatial and temporal characteristics. Without the use of covariates, the proposed delay model was applied, and the predicted cases were calculated using the predictive posterior distribution. The model can be used to directly *nowcast* or correct reports from previous days for the delay. The delay parameter *d* is a stand-alone variable in the model, allowing it to predict the missing cells directly, by setting it to the necessary value in the data vector used for predictions.

The data used in this study consist of weekly counts of COVID-19 reports aggregated at the provincial level. Data were extracted from the Department of Health (DOH) Data Drop starting from January 1, 2020 and ending on April 30, 2021 (70 weeks) for the whole island of Mindanao, Philippines. The island of Mindanao is divided into 449 municipalities, and each municipality belongs to one of the 27 provinces. The goal is to use the proposed model to correct reporting delays across the provincial

level, considering spatial variability in the delay mechanism and the disease process, as well as allowing for spatial dependence in neighboring regions.

Figure 1 shows the total number of COVID-19 cases from the month of March 2020 up to April 2021. The black solid line shows the actual number of cases, and the solid red line shows the reported cases with delay for the last 4 weeks up to and including the 14th pandemic week ending on April 30, 2021.

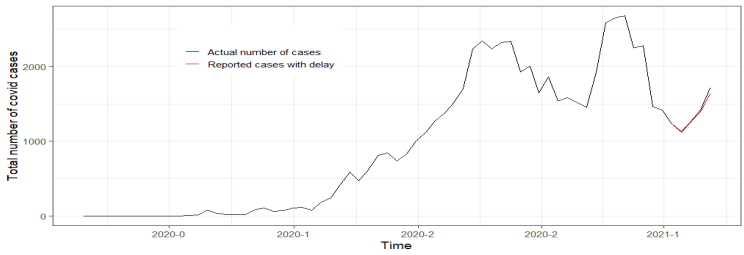

Figure 1. Total number of COVID-19 Cases in Mindanao from March 2020- April 2021.

The term $\boldsymbol{X}'_{\boldsymbol{t,d,s}}\boldsymbol{\delta}$ in Equation 2 is set to zero since no covariate information is available in the data set. Hence, the final spatio-temporal model is given by,

$$log(\lambda_{t,d,s}) = \mu + \alpha_t + \beta_d + \gamma_{t,d} + \beta_{d,s} + \psi_s^{IAR} + \psi_s^{ind}. \quad (7)$$

The model assumes that the delay structure varies across provinces, through $\beta_{d,s}$ while the overall temporal evolution of the disease counts α_t is the same across the provinces. This is because the provinces are close to each other, and we would not expect the disease transmission to vary considerably across space. Similarly, the interaction term $\gamma_{t,d}$ is spatially constant. The term $\psi_s^{IAR} + \psi_s^{ind}$ captures overall similarity in disease counts across the provinces; however, it also allows for some provinces to be different (on average), if there is such evidence in the data.

Table 3 presents the posterior estimates with their precision for the random effects of the spatio-temporal model estimated by INLA. It shows the various random effects associated with its posterior mean $(mean_p)$, posterior standard error (SD_p), and uncertainty in the form 95% credible intervals with the lower and upper limits, which are 25% and 97.5%, respectively. Lastly, the posterior median estimates are in the 50% column.

Table 3. Random effects precision of posterior estimates for the spatio-temporal model.

	$mean_p$	SD_p	2.5%	50%	97.5%
α_t	9.201	1.475	7.207	8.888	12.877
β_d	0.699	0.580	0.038	0.541	2.112
$\gamma_{t,d}$	9.955	1.278	8.382	9.655	13.189
$\beta_{d,s}$	4.066	1.510	2.516	3.637	8.057
ψ_s^{IAR}	36.595	27.930	10.116	28.525	110.587
ψ_s^{ind}	7.862	0.831	5.987	7.986	9.070

Figure 3 shows the posterior mean with 95% credible intervals for time random effects, α_t. The time series on the weekly COVID-19 starts from the 1st epidemic week of the year 2020 and ends at the 14th epidemic week of the year 2021.

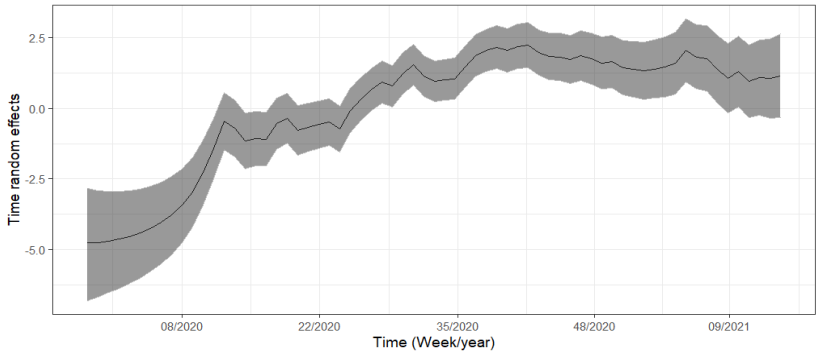

Figure 2. Posterior mean with 95% credible intervals for time on the weekly COVID-19 cases.

As shown in the figure, the COVID-19 count has a sudden increase from March to April, or about the 10th to 13th week of 2020, and a gradual increase and slight fluctuations from about the 14th week of 2020 up to the 14th week of 2021. Further, it can be observed that the number of COVID-19 cases increases as the time approaches 2021 compared to the year 2020. Moreover, the overall temporal effects do not follow any seasonal pattern.

Posterior mean estimates of the delay mechanism, which is different across provinces, $\beta_d + \beta_{d,s}$, are shown in Figure 3.

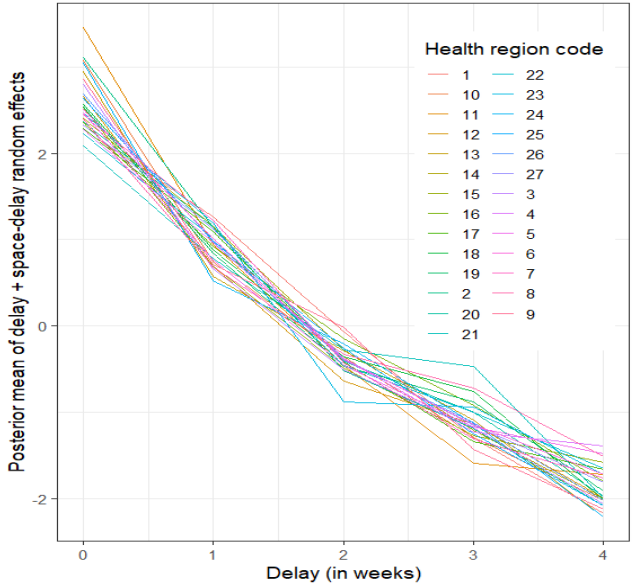

Figure 3. Posterior mean of the space-delay random effects by province.

Observe that the mean reporting count decreases with delay (in weeks), on the average. However, there is considerable variability across the regions, particularly during the second and third weeks of delay. This reflects the fact that delays are likely related to several factors such as between-region differences, improvements in reporting efficiency over time, and/or weekly cycles which vary considerably in space.

Moreover, posterior mean estimates of the time-delay interaction term $\lambda_{t,d}$, where $d = 0,1,2,3,4$ weeks are shown in Figure 4. For $d = 0$ (no delay), the number of COVID-19 cases associated with the temporal evolution clearly increases. Additionally, the temporal evolution for $d = 0$ (no delay) and $d = 1$ (1-week delay) is negative in the first quarter of each week, indicating that quicker notifications or reports of cases when they are discovered may result from potential epidemic awareness. For $d = 2,3,4$, respectively, the mean lies near zero in the first quarter of each week but fluctuates in later weeks up to the 14th week of 2021. It does not show an increasing trend of delayed reports, but an inconsistent delay. This is probably due to fewer improvements in reporting efficiency over time. The delay random effect shows the importance of the delays as the number of weeks increases, and the delay should not be neglected since it has a significant effect on the real-time case notification.

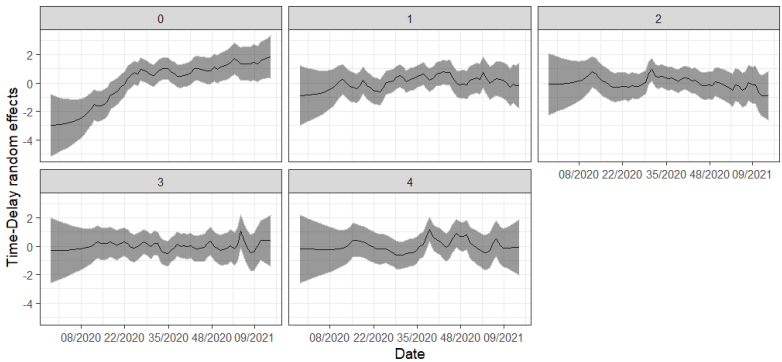

Figure 4. Posterior mean of the time-delay random effects $\lambda_{t,d}$ for $d=0,1,2,3,4$

Figure 5 shows the posterior mean estimates of the overall spatial variability term, $\Psi_s^{(IAR)} + \Psi_s^{(ind)}$. The plot indicates some variability in the number of COVID-19 reports across the provinces, but also a similarity in neighbouring regions. This may be reflecting the unobserved factors relating to the susceptible population (including population size). The delay can vary from place to place, being susceptible to the adherence of health care providers to the notification protocol as well as the access of patients to health care and health system shortcomings.

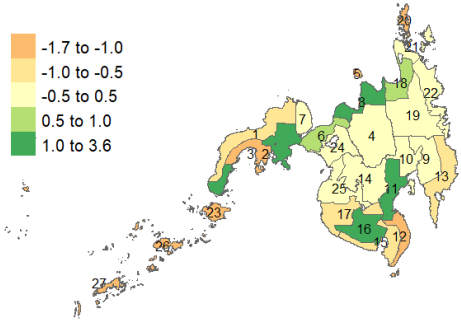

Figure 5. Posterior mean of the COVID-19 spatial random effect ψ_s.

In order to assess whether the spatial correlation was adequately captured, consider the measure or statistic R computed as follows:

$$R = \frac{var(\psi_s^{IAR})}{var(\psi_s^{IAR} + \psi_s^{ind})}.$$

The statistics R quantifies the contribution of the structured random effect ψ_s^{IAR} to the total variance of the spatial effect $\psi_s^{IAR} + \psi_s^{ind}$. Based on Figure 6, the value of R is on the average equal to 0.5, indicating that structured and unstructured spatial effects contribute roughly equally. It is to note that if value of R is close to zero, this indicates that there is little spatial correlation, whereas a value greater than 1 indicates that the structured random effects are capturing most of the variability (a nonzero correlation).

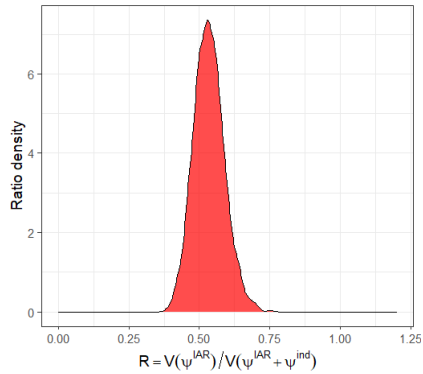

Figure 6. Posterior distribution of the statistic R.

Figure 7 (7.A and 7.B) shows the time series plot of reported COVID-19 cases, as well as predictions or *nowcast values* in the whole island of Mindanao as the maximum possible delay, which is set to 4 weeks. It is to note that Figures 8.A and 8.B differ only on the time scale, where A starts from January 2020, whereas B starts from January 2021. It also depicts the estimated mean as well as the 95% prediction intervals (dotted black line and shaded region) of the corresponding predictive distribution from Equation 5, that is, $n_{t,d,s} \sim NegBin(\lambda_{t,d,s}, \phi)$, where $log(\lambda_{t,d,s}) = \mu + \alpha_t + \beta_d + \gamma_{t,d} + \beta_{d,s} + \psi_s^{IAR} + \psi_s^{ind}$. The plot also shows the weekly time series of the eventually reported COVID-19 cases from the first epidemic week of 2020 to the 14th epidemic week of 2021 (solid black line in Figure 7.A) and COVID-19 cases from the first epidemic week of 2021 to the 14th epidemic week of 2021 (solid black line in Figure 7.B). Finally, the plot also shows the currently reported number of COVID-19 cases for the last 4 weeks, up to and including the 14th epidemic week ending on April 2, 2017 (dashed red line).

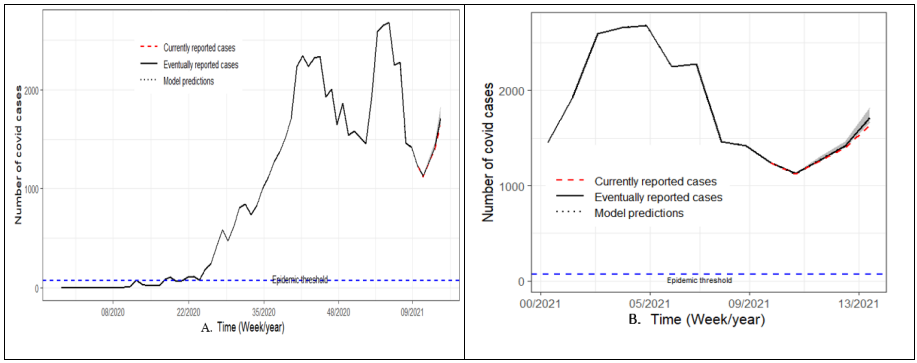

Figure 7. Time series plot of the aggregated COVID-19 cases in Mindanao, Philippines.

It can be observed in week 14 that the currently reported number of cases is $n_{14,0}$ = 1633, and the eventually reported count is $\sum_{d=0}^{4} n_{14,d} = 1710$, while the predicted value for N_t is 1708, also shown in Table 4. Hence, the proposed model is able to capture the increasing trend of the disease counts, and the predictions are much closer to the true value compared to the currently reported counts (which actually indicate a decline).

Table 4. COVID-19 cases for the last 4 weeks ending on April 30, 2021 (14th week).

2021 epidemic week	currently reported cases	eventually reported (observed)	predicted cases	RMSE (observed vs predicted)
10th week	1235	1235	1235	2.0
11th week	1120	1129	1123	
12th week	1260	1266	1276	
13th week	1397	1417	1419	
14th week	1633	1710	1708	

Furthermore, the root mean square error (RMSE) is computed to measure the difference between predicted values and observed values. Statistically, the larger the difference indicates a larger gap implying a poor model fit. Consequently, a smaller RMSE implies a better model. Based on the result shown in Table 4, the RMSE is equal to 2.0, on the average, which is low and it suggests a more accurate prediction.

Finally, nowcasting estimation is of utmost importance in order to provide accurate and reliable estimations to avoid misclassification of warning issuance. These disease-specific quantities provide a tool for setting goals for reporting delays, not only for outbreak control but also for evaluation of individual-based interventions with other aims, such as partially reducing infections or completely stopping transmission.

V. CONCLUSION

In disease surveillance, the spatial and temporal components in a pandemic or disease outbreaks are essential so that the strength, direction, and trend of the disease transmission are considered. The necessity in a disease outbreak lies in predicting or *nowcasting* the total number of disease cases, $n_{t,d,s}$, to aid health authorities to have effective control measures and issuance of warnings to the public. The Bayesian hierarchical framework is implemented in R-INLA to explore the possibility of a flexible way of correcting delayed reports considering time and space, that is, a spatio-temporal model. Results show that the proposed model was able to capture the increasing trend of COVID-19 disease counts in the presence of delayed reports.

VI. ACKNOWLEDGEMENT

The author is grateful to the DOST-SEI ASTHRDP for the financial support during her MS Statistics degree program at MSU-IIT.

VII. LITERATURE CITED

Bacchetti, P., M.R. Segal, and N.P. Jewell. 1993. “Backcalculation of HIV Infection Rates.” *Statistical Science* 8(2): 82–101.

Bastos, L. S., Economou, T., Gomes, M. F., Villela, D. A., Coelho, F. C., Cruz, O. G., Bailey, T. and Codeço, C. T. 2019. “A Modelling Approach for Correcting Reporting Delays in Disease Surveillance Data.” *Statistics in Medicine* 38(22): 4363–77.

Besag, J., J. York, and A. Mollié. 1991. "Bayesian Image Restoration, with Two Applications in Spatial Statistics." *Annals of the Institute of Statistical Mathematics* 43(1): 1–20.

Blangiardo, M., M. Cameletti, G. Baio, and H. Rue. 2013. "Spatial and Spatio-Temporal Models with R-INLA." *Spatial and Spatio-Temporal Epidemiology* 4: 33–49.

Brookmeyer, R. and M.H. Gail. 1988. "A Method for Obtaining Short-Term Projections and Lower Bounds on the Size of the Aids Epidemic." *Journal of the American Statistical Association* 83(402): 301–8.

Harris, J.E. 1990. "Reporting Delays and the Incidence of AIDS." *Journal of the American Statistical Association* 85(412): 915–24.

Höhle, M., and M. van der Heiden. 2014. "Bayesian Nowcasting during the STEC O104:H4 Outbreak in Germany, 2011." *Biometrics* 70(4): 993–1002.

Kalbfleisch, J. D., and J. F. Lawless. 1989. "Inference Based on Retrospective Ascertainment: An Analysis of the Data on Transfusion-Related AIDS." *Journal of the American Statistical Association* 84(406): 360–72.

_____. 1991. "Regression Models for Right Truncated Data with Applications to Aids Incubation Times and Reporting Lags." *Statistica Sinica* 1(1): 19–32.

Klaucke, D., Buehler, J., Thacker, S, Parrish, R. G., and Trowbridge, F. 1988 "Guidelines for Evaluating Surveillance Systems." *MMWR supplements*.

Kline, D., Hyder, A., Liu, E., Rayo, M., Malloy, S., and Root, E. 2022 "A Bayesian Spatio-Temporal Nowcasting Model for Public Health Decision-Making and Surveillance." https://arxiv.org/abs/2102.04544v1 (February 16, 2022).

Knorr-Held, L. 2000. "Bayesian Modelling of Inseparable Space-Time Variation in Disease Risk." *Statistics in Medicine* 19(17–18): 2555–67.

Lagakos, S., L. Barraj, and V. Gruttola. 1988. "Nonparametric Analysis of Truncated Survival Data, with Application to AIDS." *Biometrika, 75(3), 515-523.*

Lawless, J. F. 1994. "Adjustments for Reporting Delays and the Prediction of Occurred but Not Reported Events." *The Canadian Journal of Statistics / La Revue Canadienne de Statistique* 22(1): 15–31.

McGough, S.F., M.A. Johansson, M. Lipsitch, and N.A. Menzies. 2020. "Nowcasting by Bayesian Smoothing: A Flexible, Generalizable Model for Real-Time Epidemic Tracking." *PLOS Computational Biology* 16(4): e1007735.

Renshaw, A. E., and R. J. Verrall. 1998. "A Stochastic Model Underlying the Chain-Ladder Technique." *British Actuarial Journal* 4(4): 903–23.

Rotejanaprasert, C., N. Ekapirat, D. Areechokchai, and R.J. Maude. 2020. "Bayesian Spatiotemporal Modeling with Sliding Windows to Correct Reporting Delays for Real-Time Dengue Surveillance in Thailand." *International Journal of Health Geographics* 19(1): 4.

Stoner, O. and T. Economou. 2020. "Multivariate Hierarchical Frameworks for Modeling Delayed Reporting in Count Data." *Biometrics* 76(3): 789–98.

Swaan, C., A. van den Broek, M. Kretzschmar, and J H. Richardus. 2018. "Timeliness of Notification Systems for Infectious Diseases: A Systematic Literature Review." *PLOS ONE* 13(6): e0198845.

Vandendijck, Y., Faes, C, Kirby, R.S., Lawson, A, and Hens, N. 2016. "Model-Based Inference for Small Area Estimation with Sampling Weights." *Spatial Statistics* 18: 455–73.

van de Kassteele, J., P.H.C. Eilers, and J. Wallinga. 2019. "Nowcasting the Number of New Symptomatic Cases During Infectious Disease Outbreaks Using Constrained P-Spline Smoothing." *Epidemiology (Cambridge, Mass.)* 30(5): 737–45.