A Proposed Two-stage Bayesian Modeling for Calibration of Epidemiological Forecasts with Applications on Dengue Cases in Butuan City

Hernan G. Pantolla*[1,2] and Alex C. Gonzaga[2]

[1]Mathematics and Statistics Unit, Kolehiyo ng Lungsod ng Dasmariñas, Cavite, Philippines
[2]Department of Physical Sciences and Mathematics, University of the Philippines Manila, Manila, Philippines

ABSTRACT

This paper presents a comparative analysis of modeling approaches for a *high-frequency* time series. The moving average of 567 morbidity weeks of dengue cases in Butuan City, a Highly Urbanized City (HUC), was used as the response variable, with the aggregated rainfall, mean maximum temperature, mean minimum temperature, and mean relative humidity as the set of regressors. The last 24 morbidity weeks were set to be used for validation of predictive accuracy.

Some pieces of literature support the robustness of the use of Bayesian methods in drawing inferences, modeling, and predicting epidemiological data. Hence, candidate Bayesian Econometric models were applied following appropriate assumptions. The applicability of Bayesian Vector Autoregression (BVAR) for variable selection and lag inclusion purposes was empirically supported. The BVAR results show that the dependent variable was mostly sensitive only to the variabilities in both the (a) direct effects and (b) lags of the cases themselves, and rainfall. The generated lags as included regressors were used in a separate model using the Bayesian Metropolis-Hastings (BMH) Algorithm. For comparison, a Frequentist Vector Autoregression (FVAR) Model as the baseline model, and BMH Algorithm were applied, too. Predictions comparison shows that the variable and lag selection process of BVAR combined with the BMH Algorithm (BVAR-BMH) simulation resulted in promising gains in predictive accuracy against straightforwardly using FVAR, BVAR, or BMH algorithm for the original set of variables. The promising gains in predictive accuracy may be used in anticipatory actions for dengue epidemiological surveillance for the specified HUC, or other locations.

INTRODUCTION

Among vector-borne infectious diseases, dengue is one of the fastest spreading (Leung et al. 2023). Dengue epidemics frequently overwhelm the health system and cause massive morbidity and mortality in endemic populations in the absence of an effective early warning system. Globally, numerous applications of predictive models have been published. The summarized review by Leung et al. (2023), among others, shows that in modeling dengue outbreaks, (1) climate parameters were combined with climate change factors, (2) only a few of the models accounted for reporting time lags, (3) most predictive models applied machine learning algorithms, (4) some used regression-based approaches, (5) others used time series/autoregressive models, and (6) some models did not have validation reports or the metrics to measure the performance of the applied models or algorithms are insufficient. These findings provide channels for developing novel frameworks for predictive modeling for dengue using environmental factors given the potential and known effects of climate change.

In a review published by Brady et al. (2015) with multiple papers and reports by the World Health Organization (WHO), they summarized that disease outbreaks frequently spread rapidly, are difficult to predict, and impose a tremendous burden on society due to a dearth of response capabilities (Garg et al. 2008; Grais et al. 2007; Najera 1999; WHO 2014). Due to the apparent impacts of disease outbreaks on broader control efforts, research agendas and subsequent policy guidelines have heavily emphasized methods to predict outbreaks (early warning), to identify these while they are occurring (early detection), to respond appropriately to these (outbreak response protocols), and to better plan for future outbreak occurrences such as effective healthcare, surveillance, and control resource allocation among others (WHO 2009; WHO 2005; WHO 1999; Farrar et al. 2007; Hutwagner et al. 2003).

Objectives of the study
Primarily, this paper would aim to contribute to the literature by producing a calibrated prediction model for epidemiologic data. It will be done by using a combination of predictive models with regressors that will be identified as contributory to the spread of dengue in the selected study setting. Moreover, while there are other socioeconomic and behavioral factors, among others, that may have effects on the spread of dengue, the study will focus on using meteorological parameters and their lags. The paper, thus, may also contribute to the body of knowledge about the potential impacts of the variabilities in the meteorological parameters on a smaller geographical aggregate on the spread of communicable diseases at present and in the long run, particularly vector-borne. This may be beneficial to the local government of concern given that an effective predictive model for dengue is critical for surveillance, and cost-efficient anticipatory plans of action, especially for developing countries.

Meteorological parameters and uncertain impacts on the spread of dengue
The study will use local meteorological parameters as the initial set of regressors to gain insights into the spatiotemporal effects of the local climate on a granular geographical location. The paper of Murray, Quam, and Wilder-Smith (2013) summarizes that it is scientifically acknowledged that temperature affects adult vector survival, virus replication, and infectious period durations (Wilder-Smith and Gubler 2008; Reiter 2001; Gubler et al, 2001; Patz 2001). The extent to which the spread of *Aedes* mosquitoes is influenced by the environment, climate, or meteorological factors may provide useful data for predictive models. It has been demonstrated that the weather can predict dengue activity (Patz 2001; Earnest, Tan, and Wilder-Smith 2011; Wu et al. 2007; Hii et al. 2009). The Intergovernmental

Panel on Climate Change (IPCC) predicts an increase in global mean temperatures (IPCC 2007), too. This could create climatic and environmental conditions conducive to the spread of *Aedes* species into regions where they are not endemic. *Aedes aegypti* and *Aedes albopictus* could become established or reestablished soon due to the climatic suitability of numerous non-endemic regions and their climatic similarity to endemic regions (Reiter 2010).

Murray, Quam, and Wilder-Smith (2013) added that the rise in global temperature over the past four decades is associated with an increase in the probability of dengue outbreaks (Banu et al. 2011). Some research on climate change and dengue suggests an increase in transmission due to climate change-induced increases in temperature, humidity, and rainfall (Hii et al. 2009; Souza, Silva, and Silva 2010; Hales et al. 2002). This lends credence to the theory that observable environmental changes, such as an increase in global average temperature and humidity, increase the likelihood of dengue outbreaks (Russel et al. 2009; Van Kleef, Bambrick, and Hales 2011). Temperature increases may result in increased vector survival and/or migration into regions outside the tropics that were formerly vector-free (Hales et al. 2002).

Inclusion of the effects of lags
The lags of climatic variables on the incidence of dengue could be explained by climatic factors that indirectly influence the incidence of dengue (Naish et al. 2014). This is due to their influence on the dynamics of the life cycle of both vectors and viruses. This results in a cumulative time lag, commencing with mosquito hatching, larval and pupal development, adult emergence and virus amplification, human incubation, and a dengue outbreak (Gharbi et al. 2011; McMichael et al. 1996). The lag between climate data and incidence data varies according to the lag between the biological cycle or mosquito life stage and clinical symptoms (Naish et al. 2014). Literature suggests that the lag will be shorter for minimum temperatures typically associated with adult mosquito mortality and prolonged for high relative humidity, both of which are associated with adult survival and hatching (Naish et al. 2014). This study demonstrates that the same holds for the delays in the minimum temperature at 2 weeks, whereas the effects of relative humidity and its lags are not significant at all. On the other hand, this study is generally consistent with the notion that the average temperature is implicated in all biological cycles of *Aedes aegypti* that influence the incidence of dengue fever more slowly (CDC n. d.; Gharbi et al. 2011; Dhiman et al. 2011). This study also contributes to the literature on the effects of temperature as evidence was discovered supporting the benefits of separating maximum temperature, minimum temperature, and their delays as factors in modeling dengue-confirmed cases. Moreover, the inclusion of lags, analysis of extreme climatic events, distinctions between seasonal and long-term trends, nonlinear effects, and threshold effects in the associations are highly recommended by Naish et al. (2014) for better decision-making leading to policies related to dengue.

Additionally, the selection of a base period for climate data is essential. In the same location, the relationship between climate and dengue can vary considerably between decades (Naish et al. (2014). Variations may be attributable to alterations in socioeconomics, demography, and urbanization. This study concurs with the recommendation that long-term climate baseline data be used to calculate a mean that is unaffected by climate variability when modeling climate-based diseases (Wilder-Simith and Gubler 2008; Hales et al. 2002). However, limitations in collected data specifically in the number of cases are a prevalent concern in the Philippines. Thus, the longest time series of available data is often used for spatiotemporal modeling.

MATERIALS and METHODS

Study Setting

This study focuses on Butuan City. Particularly, the aggregated dengue cases in the city are the targeted data to be studied. The city is categorized as a Highly Urbanized City (HUC). An HUC, as also disseminated by the Department of Health (DOH), requires a minimum population of 200,000 and an income that is at least 50 million Philippine pesos, which is around US $938,000 at the current exchange rate based on the 1991 constant price (DOH, n. d.). According to the May 2020 Census of Population and Housing of the Philippine Statistics Authority, the city has a population of 372,910 (PSA 2021). Figure 1 shows the location of the selected study setting.
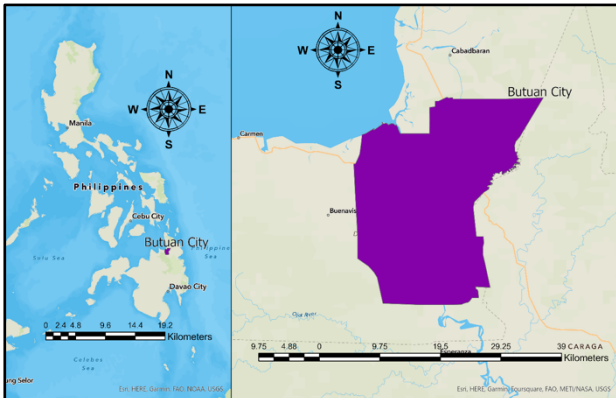

Figure 1: Map of the location of Butuan City

The (1) availability of *high-frequency* data as an alternative to using the recommended *long* time series data of reported dengue cases as well as the (2) common factors related to *globalization*, including travel and trade, patterns of modern human habitation, and vector-favorable climatic conditions (Wilder-Smith and Gubler 2008; Badurdeen et al. 2013), and (3) increased *mobility* of both the vector and human populations that may be the most crucial factor in explaining the recent rise in dengue transmission conditions in the modern era (Wilder-Smith and Gubler 2008; Badurdeen et al. 2013) are the major reasons on why the location has deemed a match as the study area. The city is densely populated and has an airport, and seaports, among others. It is also the regional center for commerce, industry, and government in the Caraga Region and a strategic trading hub due to its connectivity to other major cities (City Government of Butuan 2022). The mobility of vector and human populations makes Butuan City as an HUC a good candidate as a study area. The city also has a Philippine Atmospheric, Geophysical and Astronomical Services Administration (PAGASA) local weather station. Thus, local weather parameters, as supported by the literature as variables that can be used as predictors, can be collected, and used as regressors for the dengue cases of the city.

Data and Sources

The latest available daily data that was requested from the DOH Epidemiology Bureau through the Electronic Freedom of Information domain was from January 1, 2010, to December 31, 2020. The data covers until 2020 only to exclude the possibility of an underestimated number of cases due to the effects of the COVID-19 pandemic. The data was converted into weekly aggregates. To account for potential delayed reporting caused by holidays, suspensions of work, and other deterrents, a four-week moving average of the cases was used instead. This step was done to account for potential reporting biases which may result in misinterpretations, false alerts, or delays in the detection of outbreaks (Buckingham-Jeffery et al. 2017). In addition, the local weather parameters of (1) rainfall aggregate, (2) maximum temperature mean, (3) minimum temperature means, and (4)

relative humidity mean were used as predictors. These weather parameters that were requested from the PAGASA, were also initially in daily frequency before being adjusted to weekly frequency. Consistent with the reported cases, the four-week moving averages of the weather parameters were used for modeling. Figure 2 shows the complete time series plots of all the variables used in this study.
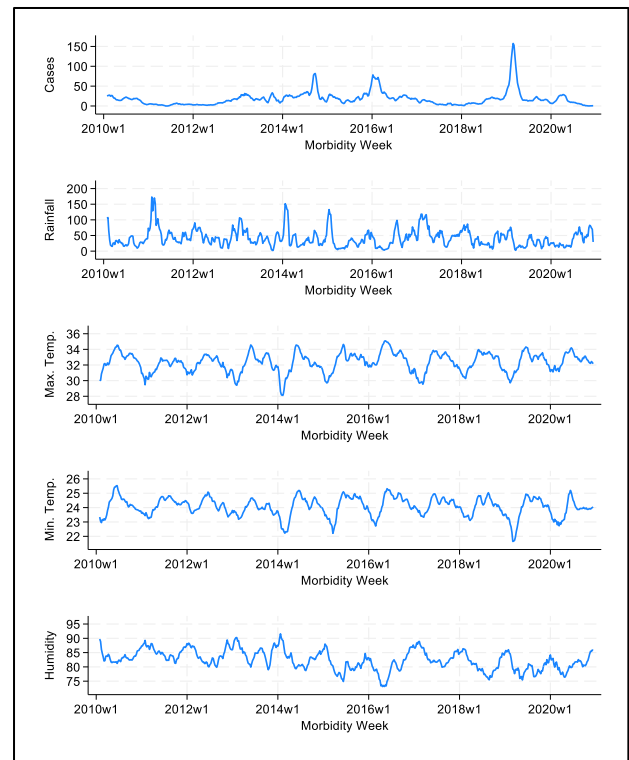

Figure 2: Five-week moving average time series plots of the variables used in this study

In Figure 2, there is a remarkable spike in 2019 in the reported cases of dengue. A surge in cases led to the DOH declaring a national dengue outbreak back then (ReliefWeb 2019). On the other hand, both total rainfall and mean humidity do not display any noticeable pattern whereas the expected seasonality on both the mean maximum temperature (Max. Temp.) and the mean minimum temperature (Min. Temp.) is observable.

METHODS

TBayesian Analysis can be applied in defining and generating inferences such as identifying models and parameters (Hobbs 1997). In the field of economics, Bayesian Econometrics are presently being used for prediction and decision analysis aside from statistical inferencing (Basturk et al. 2014). Bayesian approaches to epidemiologic data analysis are a powerful instrument for interpreting study results and evaluating hypotheses regarding exposure-disease relationships as well (Dunson 2001). In addition, Bayesian methods can predict the lag effects, which is crucial in this study that would include lags of meteorological parameters, as well as minimizing the variance with limited sample points.

In relation, this study would attempt to apply the appropriate procedures under Bayesian Econometrics to empirically identify the effects of common meteorological parameters on the spread of dengue cases in the selected study setting. In particular, the paper aims to apply Bayesian Econometrics to generate inferences on how the cases of dengue in Butuan City respond to the meteorological parameters through the lens of the

changing climate. Further, the effects of the lags of the cases and in the meteorological parameters shall be accounted for. The lags between climate variables and the number of reported dengue cases could be explained by climate factors that do not directly influence the dengue cases on record (Naish et al. 2014). This would also include the effects of the lags caused by the developmental stage of the mosquitos as a vector until its full development and potential to spread the virus (Gharbi et al. 2011; Naish et al. 2014).

Procedures

The 567 total weekly data will be divided into a training set and a test set. The first 543 entries will be used as the training set while the remaining 24 will be the test set to validate possible gains in predictive accuracy. Customary to any econometric Simultaneous Equation Models (SEM), all the variables that will be entered in the modeling procedure will first be tested for the level(s) of stationarity. The appropriate econometric procedure under the Bayesian framework will also be dependent on the level(s) of stationarity of the variables. Next, an appropriate Bayesian econometric model will be applied. Thereafter, diagnostic test(s) will be performed to determine the fitness of the model given the robustness property of a Bayesian framework. One-step-ahead predictions will be generated for both the Frequentist and Bayesian econometric models. In hopes of assessing if further gains in predictive accuracy could be achieved, another Bayesian system that will use the variables with lags that are influential to the variations in the response variable identified in the initial Bayesian model shall be performed. The predictions from the three models will be compared using predictive accuracy measures.

Stationarity of the variables

It is imperative to determine the stationarity of each variable to apply the appropriate econometric model, whether under the Frequentist or Bayesian approach. Table 1 shows the stationarity of the variables. Using the Augmented Dickey-Fuller Test proposed by Dickey and Fuller (1979), it was determined that all the variables are stationary at level.

Table 1: Results of the Augmented Dickey-Fuller Test.

Variable	p-value for Stationarity Test at Level	Integration
Cases	0.01241	$I(0)$
Rainfall	<0.0001	$I(0)$
Tmax	<0.0001	$I(0)$
Tmin	<0.0001	$I(0)$
Humid	0.0057	$I(0)$

Based on Table 1, all the variables are stationary at level or similarly, integrated of order 0. All the variables are $I(0)$ in other words. Hence, the (*Frequentist*) Vector Autoregressive (VAR) Model is a candidate. However, to apply a more probabilistic and simulation-based approach, the application of the Bayesian Vector Autoregressive (BVAR) Model is proposed. The BVAR framework is also a type of SEM. The study of Sugita (2014) states that in theory, in the cases of multivariate VAR models, overparameterization may be a cause of concern such as inconclusive inference. This could lead to inaccurate predictions. Herein, the Bayesian alternative for VAR has gained popularity progressively as the prior distributions may be constrained. A correctly specified one-step-ahead prediction tends to provide more accurate predictions, too for longer horizons and more lags added in the model for the BVAR Model, as compared to the customary VAR Model (Sugita 2014).

Optimal lags

Table 2 shows the result of the optimal lag selection of a customary VAR system at the 0.05 level of significance for easier reproducibility. As Bayesian models would entail a huge deal of computational cost during simulations, the maximum number of lags was set to 5. Results of the Likelihood Ratio (LR), Final Prediction Error (FPE), Akaike Information Criterion (AIC), Hanna-Quinn Information Criterion (HQIC), and Schwarz-Bayesian Information Criterion (SBIC) are shown. To avoid underestimation since the SBIC tends to suggest the least number of optimal lags (Pesaran, Shin, and Smith 2001), three criteria particularly LR, FPE, and AIC chose 5, the optimal lag of 5 was selected and will be used for the Bayesian modeling purposes.

Table 2: Selection of the optimal number of lags

Lag	LR	FPE	AIC	HQIC	SBIC
0	-	99743.2000	25.6997	25.7148	25.7383
1	6682.30	0.7320	13.8774	13.9678	14.1090
2	731.79	0.2171	12.6621	12.8279	13.0866*
3	75.690	0.2074	12.6163	12.8574	13.2337
4	114.610	0.1849	12.5012	12.8176*	13.3115
5	71.651*	0.1779*	12.4626*	12.8543	13.4659

FVAR Framework

An unrestricted VAR with p lags written as $VAR(p)$ is a Seemingly Unrelated Regression model with similar regressors in every equation. In the works of Lutkepohl (2005) and Greene (2008), running a linear regression in every equation generates the MLE of the coefficients. Thereafter, the estimates of coefficients can be calculated using the residuals which are eventually used in estimating the cross-equation residual variance-covariance matrix $\underline{\Sigma}$ (StataCorp 2021a).

In addition, Lutkepohl (2005), defines $VAR(p)$ with exogenous variables as

$$\underline{y}_t = \underline{A}\underline{Y}_{t-1} + \underline{B}_0\underline{x}_t + \underline{u}_{t-1} \qquad (1)$$

where $\underline{y}_t$ is the $K \times 1$ vector of endogenous variables, $\underline{A}$ is the $K \times Kp$ matrix of coefficients, $\underline{B}_0$ is the $K \times M$ matrix of coefficients, $\underline{x}_t$ is the $M \times 1$ vector of exogenous variables, $\underline{u}_t$ is the $K \times 1$ vector of white noise innovations, and $\underline{Y}_t$ is the $Kp \times 1$ matrix given by $\underline{Y}_t = \left(\underline{y}_t \dots \underline{y}_{t-p+1}\right)'$. For easier manipulation, Equation 1 can be rewritten into $\underline{Y}_t = \underline{B}\underline{Z} + \underline{U}$ where $Y = (\underline{y}_1 \dots, \underline{y}_T)$ and is $K \times T$, $\underline{B} = (\underline{A}, \underline{B}_0)$ and is $K \times (Kp + M)$, $Z = \left(\begin{smallmatrix} Y_0, \dots, Y_{T-1} \\ \underline{x}_1, \dots, \underline{x}_T \end{smallmatrix}\right)$ and is $(Kp + M) \times T$; and $\underline{U} = (\underline{x}_1, \dots, \underline{x}_t)$ and is $K \times T$. The constant terms in the model are incorporated in $\underline{x}_t$, and is empty whenever if there are no exogenous variables nor constant terms. Considering the error terms matrix denoted by $\hat{\underline{U}}$ that is obtained through $\underline{Y} - \hat{\underline{B}}\underline{Z}$, where $\hat{\underline{B}}$ denotes the matrix of coefficients estimates, the estimator $\underline{\Sigma}$ is $\hat{\underline{\Sigma}} = (\hat{\underline{U}}'\hat{\underline{U}})/\hat{T}$ where $\hat{T} = T$ is the default maximum likelihood divisor. Herein, the *gamma* matrix in the VAR framework that is stored in this process is given by $\sum_{t=1}^{T}(1, \underline{Y}'_T)(1, \underline{Y}'_T)'/T$ which is a $(Kp \times 1) \times (Kp \times 1)$ matrix (StataCorp 2021a).

BVAR Framework

Extending the $VAR(p)$ form from Equation 1 according to the works of Lutkepohl (2005; in StataCorp 2021b) provides $y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + C_1 x_t + u_t; u_t \sim N(0, \Sigma)$ where, $A_l = (a_{ij}^l)$ are unknown endogenous factors lag coefficients ($l = 1, \dots, p$) which are $K \times K$ matrices, $\underline{C} = (a_{is})$ as exogenous factors that is a $K \times m$ matrix; and $\underline{u}_t$ is a $K \times 1$ vector of residuals with a $K \times K$ covariance matrix $\underline{\Sigma}^*$. In simplified form, the extended $VAR(p)$ can be expressed as

$$\underline{Y}^* = \underline{X}\underline{B} + \underline{V}, \tag{2}$$

compactly, where $\underline{Y}^* = \left(\underline{y}'_1 \dots \underline{y}'_T\right)'$, $\underline{X} =$
$$\begin{pmatrix} \underline{y}'_0 \underline{y}'_{-1} & \cdots & \underline{y}'_{1-p} \underline{x}'_1 \\ & \vdots & \\ \underline{y}'_{T-1} \underline{y}'_{T-2} & \cdots & \underline{y}'_{T-p} \underline{x}'_T \end{pmatrix}, \underline{B} = (\underline{A}_1 \dots \underline{A}'_p \underline{C}')'$$, and $\underline{V} =$
$(\underline{u}'_1 \dots \underline{u}'_T)'$. A $T \times k$ matrix $\underline{Y}$, a $T \times (Kp + M)$ matrix $\underline{X}$, a $(Kp + M) \times K$ matrix $\underline{B}$ of all coefficients, and a $T \times K$ matrix $\underline{V}$ are makes up the components of Equation 2. Moreover, the OLS estimates of $\underline{B}$ and Σ are $\hat{\underline{B}} = (\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y}$ and

$$\Sigma_{OLS} = \hat{\underline{V}}'\hat{\underline{V}}/[T - (Kp + m + 1)], \hat{\underline{V}} = \underline{Y} - \underline{X}\hat{\underline{B}}. \tag{3}$$

In vector form, the matrix equation in Equation 3 can be expressed as $\underline{y} = \underline{X}^*\underline{\beta} + \underline{v}$. Here, $\underline{y}$ is a $KT \times 1$ vector while the $KT \times K(Kp + M)$-dimension matrix $\underline{X}^* = I_K \otimes \underline{X}$ applies the Kronecker product component signified by $\otimes$ that has a $K \times K$ dimension (that is, I_K), a $K(Kp + M) \times 1$ vector $\underline{\beta}$ representing all coefficients, and the residuals vector $\underline{v}$ and covariance matrix $\underline{\Sigma}^* = \underline{\Sigma} \otimes I_T$ that has dimensions of $KT \times 1$ and $KT \times KT$, respectively (StataCorp 2021b).

The specification of BVAR Model necessitates the distinction of an appropriate prior for the vector of coefficients $\underline{\beta}$. The Minnesota prior is selected in this study since using this prior enforces the hypothesis that each variable in the model follow a random walk process (Kuschnig and Vashold 2021; in StataCorp 2021b). The variables in this study follow a random walk process given the application of moving average.

Choice of Prior for the BVAR Model
Early works on BVAR (Doan et al., 1984; Litterman 1986; in StataCorp 2021b) simplified the model prior through the assumption of fixed-error covariance matrix that is known. The estimated $\hat{\underline{\Sigma}}$ replaces the covariance $\underline{\Sigma}$ under the Multivariate Normal (MVN) likelihood. Typically, the $\underline{\Sigma}$ of choice is a diagonal matrix of estimated variances produced by an Autoregressive (AR) Model that is run separately to each endogenous variable. The prior covariance for the coefficients of the exogenous factors is then generated from $\underline{\Sigma}$ as specified in the Original *Minnesota* prior with a known *fixed* prior error covariance. Since this paper intends to eventually propose a combination of methods that may generate more precise predictions, this will also adapt the Minnesota prior proposed by Litterman (1986) since linear regression based BVAR models typically use this prior, too (Ciccarelli and Rebucci 2003).

The covariance matrix of the error vectors $\underline{v}_t$ is considered known, that is, $\underline{\Sigma} = \underline{\Sigma}_0$ in the Original Minnesota prior in a BVAR system (Litterman 1980, 1986). As such, $\underline{v} \sim N(0, \underline{\Sigma}_0 \otimes I_T)$. The seminal version of this prior utilized a diagonal matrix that contains covariance of errors produced from distinct K number of AR models on covariance estimate diagonal given by $\underline{\Sigma}_0 = \hat{\underline{\Sigma}}_{diag} = diag(\hat{\sigma}_1^2, \dots, \hat{\sigma}_K^2)$. The proposal of Litterman is to estimate the VAR model using the equation instead of a system of equations. In essence, the calculational cost is reduced which poses critical challenge during those years (StataCorp 2021b).
The Minnesota prior for the vector of coefficient denoted by $\underline{\beta}$ is a prior that is MVN in nature and is given by $\underline{\beta} \sim N(\underline{\beta}_0, \underline{\Omega}_0)$ where $\underline{\beta}_0$ is a $KT \times 1$ vector whereas the matrix $\underline{\Omega}_0$ is $KT \times KT$.

These parameters are characterized in a manner that accounts for a $VAR(p)$ model that has a special time-series structure. This framework defines a regression vector of lag coefficients $\underline{\beta}$, consisted of a_{ij}^l (where $l = 1, \dots, p$ and $i, j = 1, \dots, K$) and the coefficients c_{is} (where $i = 1, \dots, K$ and $s = 1, \dots, M$) of exogenous variables. The expected value of each coefficient is zero under the Minnesota prior assumption excluding the self-variables first-lag coefficients that is given by $E(a_{ij}^l) = \delta_{1l}\delta_{1j}$ and $E(c_{is}) = 0$ where $\delta_{1j} = 1$ whenever $i = j$ and 0 otherwise. Hence, the prior mean vector $\underline{\beta}_0$ vector of dimension $K(Kp + M) \times 1$ takes values of either 0s or 1s. Here, the self-variables first-lag coefficients are all 1s (StataCorp 2021b).

Metropolis Hastings Sampling under the Bayesian Inference
For convenience, this paper will refer to the Metropolis-Hastings sampling in the Bayesian system as Bayesian Metropolis-Hastings (BMH) algorithm. The BMH algorithm repeatedly samples from a probability distribution through simulation by using the full joint density function. Then, independent proposal distributions are generated for every variable of interest (Yildirm 2012).

Consider a vector of d scalar model of parameters $\underline{\theta}$. Consider also T_0 to be the burn-in periods with T being the MCMC iterations. The sum of all iterations is $T_{tot} = T_0 + (T - 1) \times t^i + 1$ with t^i denoting the thinning interval that is customarily equal to 1. The BMH algorithm starts with initializing the sample value for every random variable which is typically sampled from the prior distribution of the variable. The algorithm follows a loop that generates a candidate sample at $t = 0$ with $\underline{\theta}_0 = \underline{\theta}_0^J$ setting up at $k - 1$ adaptation counter with the initial feasible state $\underline{\theta}_0^f$ from the proposal distribution $\underline{\theta}_* = \underline{\theta}_{t-1} + e, e \sim N(0, p_k^2 \Sigma_k)$; calculate the acceptance probability $\min\{[p(\underline{\theta}_*|\underline{y})/p(\underline{\theta}_{t-1}|\underline{y})], 1\} = \alpha(\underline{\theta}_*|\underline{\theta}_{t-1})$ with $p(\underline{\theta}|\underline{y}) = f(\underline{y}|\underline{\theta})p(\underline{\theta})$ as the posterior distribution of $\underline{\theta}$, corresponding to the likelihood function $f(\underline{y}|\underline{\theta})$ and prior $p(\underline{\theta})$; and set the candidate sample to $\underline{\theta}_t = \theta_*$ whenever $u < \alpha(\theta_*|\theta_{t-1}), u \sim \text{Uniform}(0,1)$ is drawn, or $\underline{\theta}_t = \underline{\theta}_{t-1}$ (Yildirm 2012; StataCorp 2021c). The most frequently applied BMH algorithm is the "Random-walk Metropolis-Hastings algorithm" (Yildirm 2012). This study will also use the same algorithm with respect to the intended purposes i.e., prediction.

RESULTS and DISCUSSION

The BVAR with Minnesota prior output restricted to the variables and lags contributory to the dependent variable is given in Table 2. Monte Carlo Markov Chain (MCMC) runs of 12,500 samples having the first 2,500 simulations used as burn-ins show the effects of predefined five lags of each regressor through the mean of coefficients. The model has a high average model efficiency of 99.27%. The prior variances were set to the customary default values seen in StataCorp (2021a). The standard deviation (SD), Monte Carlo Standard Error (MCSE), and 95% Credible Interval are also provided in Table 2. A set of dynamic predictions was generated for the BVAR Model. Separately, predictions for a Frequentist VAR model were also done for comparative purposes.

Table 3: Results of the BVAR Model with cases as the response variable

Variable	Lag	Mean	SD	MCSE	Median	95% Credible Interval	
						LL	UL
Cases	L1.	1.4445	0.0264	0.0003	1.4445	1.3931	1.4967
	L2.	-0.3233	0.0364	0.0004	-0.3234	-0.3940	-0.2513
	L3.	-0.1584	0.0265	0.0003	-0.1582	-0.2103	-0.1065
	L4.	-0.0361	0.0200	0.0002	-0.0358	-0.0752	0.0028
	L5.	0.0306	0.0146	0.0001	0.0306	0.0015	0.0592
Rainfall	L1.	0.0003	0.0065	0.0001	0.0003	-0.0124	0.0131
	L2.	<0.0001	0.0050	<0.0001	<0.0001	-0.0097	0.0096
	L3.	-0.0009	0.0035	<0.0001	-0.0009	-0.0077	0.0060
	L4.	-0.0005	0.0027	0.0000	-0.0005	-0.0058	0.0048
	L5.	-0.0003	0.0022	0.0000	-0.0003	-0.0046	0.0040
Tmax	L1.	-0.4373	0.2584	0.0026	-0.4360	-0.9441	0.0567
	L2.	0.0774	0.2129	0.0021	0.0789	-0.3433	0.4907
	L3.	0.0484	0.1487	0.0015	0.0500	-0.2481	0.3414
	L4.	0.0351	0.1118	0.0011	0.0348	-0.1855	0.2544
	L5.	0.0038	0.0891	0.0009	0.0045	-0.1731	0.1794
Tmin	L1.	0.4398	0.4840	0.0050	0.4412	-0.5111	1.3899
	L2.	0.2075	0.4279	0.0043	0.2066	-0.6347	1.0333
	L3.	-0.1736	0.2993	0.0030	-0.1705	-0.7700	0.4072
	L4.	-0.0183	0.2273	0.0023	-0.0209	-0.4544	0.4334
	L5.	-0.0311	0.1814	0.0018	-0.0312	-0.3815	0.3314
Humid	L1.	0.0282	0.0842	0.0008	0.0287	-0.1390	0.1914
	L2.	-0.0052	0.0732	0.0007	-0.0048	-0.1495	0.1385
	L3.	-0.0233	0.0503	0.0005	-0.0237	-0.1212	0.0747
	L4.	-0.0051	0.0394	0.0004	-0.0051	-0.0820	0.0719
	L5.	-0.0014	0.0313	0.0003	-0.0019	-0.0616	0.0601

The stability of the BVAR is shown in Table 3. The probability that the eigenvalues are found inside the unit circle is 50.71%. As reported, the posterior summaries for the eigenvalue moduli lie inside the unit circle as all the mean values are below 1 and therefore, the BVAR model is deemed *stable*.

Table 4: Stability of the BVAR Model

Eigenvalue Modulus	Mean	SD	MCSE	Median	95% Credible Interval	
					LL	UL
1	~1.000	0.0004	<0.0001	~1.000	0.9991	1.0008
2	0.8976	0.0222	0.0002	0.8976	0.8551	0.9421
3	0.8791	0.0236	0.0002	0.8804	0.8295	0.9207
4	0.8510	0.0278	0.0003	0.8519	0.7952	0.9011
5	0.8157	0.0286	0.0003	0.8163	0.7574	0.8687
6	0.7939	0.0270	0.0003	0.7952	0.7382	0.8434
7	0.7671	0.0326	0.0003	0.7694	0.6942	0.8230
8	0.7400	0.0393	0.0004	0.7449	0.6482	0.8032
9	0.6808	0.0680	0.0007	0.6909	0.5447	0.7850
10	0.5519	0.0695	0.0007	0.5535	0.4300	0.6829
11	0.4664	0.0388	0.0004	0.4608	0.4044	0.5646
12	0.4487	0.0268	0.0003	0.4488	0.3952	0.5025
13	0.4316	0.0264	0.0003	0.4317	0.3803	0.4849
14	0.4188	0.0245	0.0002	0.4191	0.3698	0.4660
15	0.4087	0.0244	0.0002	0.4093	0.3596	0.4551
16	0.3965	0.0250	0.0003	0.3974	0.3459	0.4430
17	0.3863	0.0249	0.0002	0.3872	0.3360	0.4331
18	0.3762	0.0259	0.0003	0.3774	0.3222	0.4237
19	0.3637	0.0278	0.0003	0.3652	0.3061	0.4146
20	0.3526	0.0289	0.0003	0.3541	0.2920	0.4050
21	0.3395	0.0330	0.0003	0.3426	0.2666	0.3961
22	0.3209	0.0380	0.0004	0.3242	0.2359	0.3847
23	0.3017	0.0439	0.0004	0.3066	0.1995	0.3720
24	0.2731	0.0600	0.0006	0.2837	0.1243	0.3623
25	0.1955	0.0929	0.0009	0.2100	0.0122	0.3406

Now the sensitivity of the moving average of the reported cases to the regressors is not clearly identifiable in Table 2. Thus, graphs for visual assessment are provided. The graphs of Impulse Response Function (IRF), Orthogonalized IRF (OIRF), and Cumulative OIRF (COIRF) are provided in Figure 3. An

IRF is typically utilized for a quick visual summarization of a VAR Model. The IRF shows how a shock in a regressor also known as *impulse variable* influences the response variable. However, the non-independence or impulse variables is not accounted for in an ordinary IRF. The dynamics between the

regressors are shown by an OIRF. An alternative method, moreover, is to present the dynamics in the accumulated effects of the shocks over time (StataCorp, 2021). Hence, the Forecast Error Variance Decomposition (FEVD) assesses the contribution of the variability of the shocks in the regressors to the predictive errors in the response variable (StataCorp, 2021a) was included.
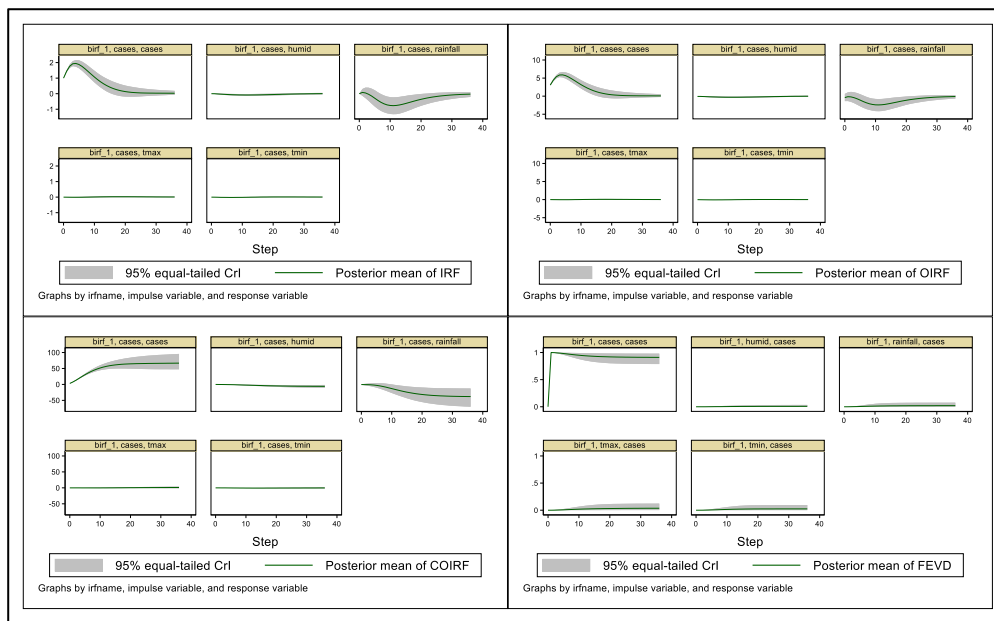

Figure 3: IRF, OIRF, COIRF, and FEVD graphs of the BVAR Model

As shown in Figure 3, the IRF, OIRF, and COIRF graphs, only the shocks in the lags of the moving averages of the cases itself and rainfall, consistently, have influential effects on the response variable. On the other hand, the FEVD reveals that only the lags in the response variable itself have long-term impacts on its predictive errors as the other regressors appear to have negligible effects. Hence, each of the first, second, third, fourth, and fifth lags of the moving averages of the reported totals of dengue cases as well as of the rainfall, were generated. These eight variables were used in a separate model using an MNMN Bayesian MH Algorithm with inverse-gamma hyperprior distribution. Table 5 shows the results of the Bayesian MH Algorithm using Gibbs Sampling employing the Laplace-Metropolis estimator.

Table 5: Coefficients of the Bayesian MH model

| Regressor | Mean | SD | MCSE | Median | 95% Credible Interval | |
					LL	UL
First lag of cases	1.7166	0.0423	0.0013	1.7176	1.6325	1.7994
Second lag of cases	-0.6642	0.0847	0.0027	-0.6641	-0.8225	-0.4961
Third lag of cases	-0.1705	0.0892	0.0027	-0.1714	-0.3478	-0.0009
Fourth lag of cases	-0.0189	0.0869	0.0028	-0.0154	-0.1986	0.1584
Fifth lag of cases	0.1122	0.0429	0.0013	0.1094	0.0264	0.1969
First lag of rainfall	0.0090	0.0090	0.0003	0.0090	-0.0090	0.0268
Second lag of rainfall	0.0045	0.0142	0.0004	0.0047	-0.0231	0.0329
Third lag of rainfall	-0.0086	0.0150	0.0004	-0.0082	-0.0381	0.0197
Fourth lag of rainfall	-0.0011	0.0143	0.0005	-0.0009	-0.0284	0.0276
Fifth lag of rainfall	0.0053	0.0094	0.0003	0.0053	-0.0131	0.0243
Variance Parameter	**7.3802**	**0.4382**	**0.0139**	**7.3671**	**6.5613**	**8.2618**

Table 5 shows that for the moving averages of reported dengue cases, the first and fifth lags could increase the moving average of the present number of cases while the other lags could decrease. On the other hand, the first, second, and fifth lags of the rainfall could increase the response variable, while the others could decrease it. This estimation has an average efficiency of 99.89% that is closer to 1, with the mean of the variance parameter at 7.3802, which is fairly acceptable for the linear combination of the variables.
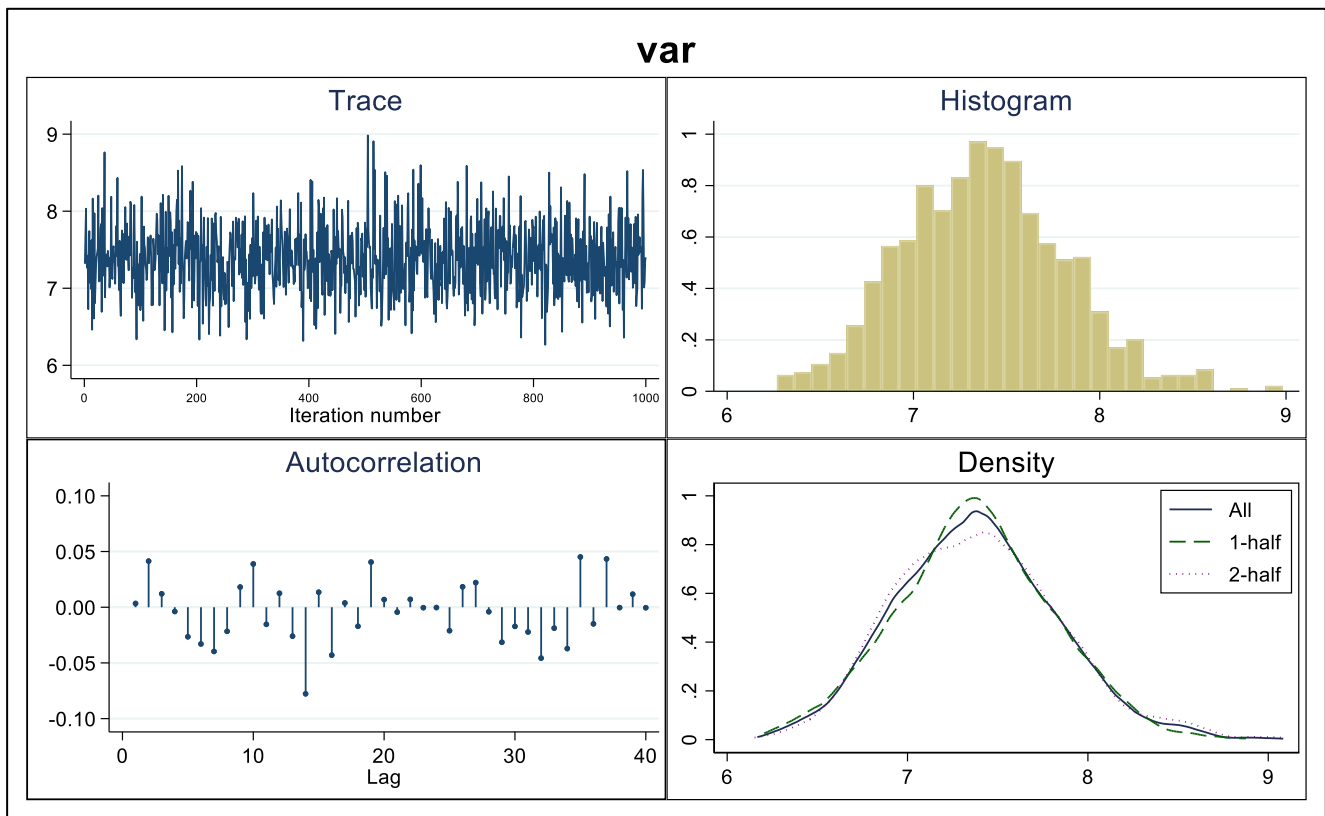
Figure 4: Diagnostics of the Bayesian MH Model

In Figure 4, the trace plots show good coverage of the marginal distribution. The histogram and kernel density plots take the shape of the expected inverse-gamma distribution, and the autocorrelation plot gives no specific pattern signifying anomalies. Therefore, the variance parameter has no apparent problems based on the visualizations provided.

Table 6 shows the 24-point prediction of each of the three models against the actual out-of-sample moving average of

reported dengue cases test set from Butuan City. The Root Mean Squared Error (RMSE) was used to determine the most accurate predictive model among the three. This measure of predictive accuracy for $r = 1, 2, , \ldots, R$ test set entries is given by $RMSE = \sqrt{\Sigma_{r=1}^{R} MSE_r}$ where $MSE = (A_r - P_r)^2$ in which, A_r and P_r respectively, denotes the actual value, and predicted value of each model.

Table 6: Prediction from each model compared to the reported cases

Week Number	Period	Cases	BVAR-BMH Prediction	MSE	BVAR Prediction	MSE	VAR Prediction	MSE	BMH Prediction	MSE
1	2020w27	9.4	8.95	0.20	8.90	0.25	8.17	1.50	19.68	105.70
2	2020w28	8.6	9.88	1.65	9.02	0.17	7.75	0.72	18.00	88.36
3	2020w29	8.8	8.18	0.38	9.32	0.27	8.5	0.09	19.33	110.79
4	2020w30	7.4	9.2	3.25	9.89	6.22	9.54	4.59	20.18	163.23
5	2020w31	6.8	6.67	0.02	10.59	14.34	10.94	17.17	21.52	216.73
6	2020w32	6	6.49	0.24	11.39	29.05	12.31	39.77	22.39	268.64
7	2020w33	6.2	5.64	0.32	12.22	36.23	13.47	52.85	20.66	209.22
8	2020w34	5.8	6.97	1.36	13.03	52.34	14.34	72.87	19.83	196.72
9	2020w35	4.8	6.19	1.93	13.88	82.37	14.96	103.21	17.94	172.70
10	2020w36	5	4.52	0.23	14.61	92.42	15.34	106.87	17.73	162.01
11	2020w37	3.6	5.27	2.78	15.29	136.74	15.59	143.69	16.79	173.89
12	2020w38	2.2	3.09	0.80	15.91	187.96	15.77	184.15	21.19	360.61
13	2020w39	1.8	1.46	0.11	16.41	213.48	15.94	200.01	20.53	350.64
14	2020w40	2	1.49	0.26	16.85	220.47	16.14	199.94	21.48	379.41
15	2020w41	0.8	2.88	4.32	17.18	268.42	16.38	242.65	21.48	427.49
16	2020w42	0.8	0.84	0.00	17.45	277.28	16.64	250.93	20.43	385.40
17	2020w43	0.6	1.10	0.25	17.7	292.56	16.92	266.29	20.98	415.44
18	2020w44	0.2	0.85	0.43	17.96	315.35	17.19	288.61	20.52	412.73
19	2020w45	0.2	0.53	0.11	18.14	321.69	17.43	296.95	19.02	354.30
20	2020w46	0.2	0.95	0.56	18.24	325.35	17.64	304.03	18.48	334.30
21	2020w47	0.4	1.14	0.55	18.30	320.5	17.79	302.58	19.41	361.27
22	2020w48	0.4	1.25	0.72	18.26	318.97	17.91	306.46	20.19	391.84
23	2020w49	0.6	0.77	0.03	18.22	310.45	17.97	301.87	20.94	413.67
24	2020w50	0.6	1.20	0.36	18.18	309.21	18.01	303.04	22.96	500.15
			RMSE	**4.57**	**RMSE**	**64.28**	**RMSE**	**63.17**	**RMSE**	**83.40**

As seen in Table 6, the BMH Algorithm, where the regressors plugged in were empirically proven by the BVAR Model, hereinafter referred to as BVAR-BMH to have inferential impacts on the current moving average of the number of reported cases in Butuan City, has the least RMSE at 4.57. It is followed by the RMSE of FVAR at 63.17 while the RMSE of the BVAR is at 64.28. The highest RMSE was obtained through BMH alone using the original set of regressors. Thus, the attempted calibration of prediction via the BVAR-BMH combination showed promising gains in accuracy. Figure 5 presents the visualization of the predictions generated from the four methods against the reported number of cases. On the other hand, Appendix 1 was also provided to visualize to comparison of predictive accuracies.
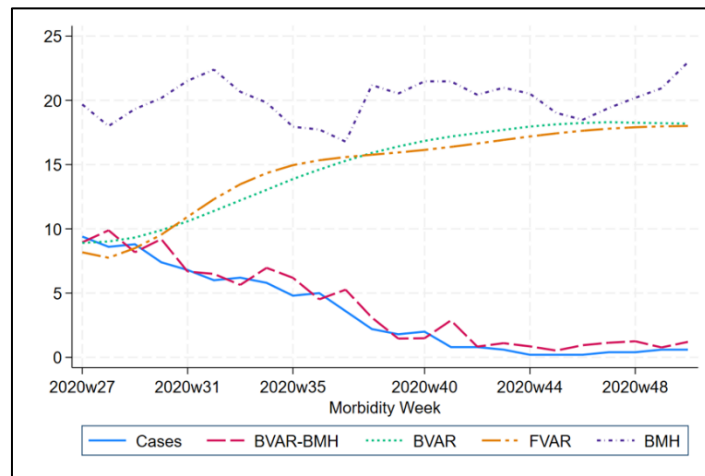

Figure 5: Comparison of the predictions from the four models

This paper, in addition, caters to a few summaries given by Leung et al. (2023) in relation to dengue outbreak modeling works by other authors. This study (1) adhered to using climate factors, (2) plugged in the effects of lags and provided inferential proof of their inclusion, (3) deviated from some studies applying machine learning algorithms i.e., implementing two methods in Bayesian econometrics, (4) being consistent with implementing regression-based modeling, (5) applying autoregressive technique, and (6) including a validation report such as the one visualized in Figure 5 to determine if gains in predictive accuracy were obtained.

The results further provide useful and could possibly be a piece of critical jumpstart information on possible uses of the Bayesian framework in modeling epidemiological data in the context of a changing climate. Hobbs (1997) stated that predictions *and updates* about climate change can be made using Bayesian analysis. Moreover, using traditional frequentist approaches as baseline modes may provide predictions that are too far from the actual data. Appendix 2 visualizes the graphs of (1) the Poisson Autoregressive Model (PAR Pred) with 4 lags which has an R-squared of less than 70% and RMSE of 15.82 but may be less sensitive to shocks in any of the variables, and (2) OLS Model with an R-Squared of less than 15%, no added lags, and RMSE of 123.11.

This study also shows empirically that at least for Butuan City, increases in the reported cases of dengue are primarily sensitive only to the sudden spikes in the number of cases themselves, and total rainfall from specific number of weeks prior. Such results could be crucial for further spatiotemporal studies. Specifically, while this paper supports the use of meteorological parameters including rainfall, the results suggests that the pattern of increased dengue cases in the selected study setting digresses from the common knowledge based on literature that temperature and humidity (Hii et al. 2009; Souza, Silva, and Silva 2010; Hales et al. 2002; Russel et al. 2009; Van Kleef, Bambrick, and Hales 2011) could increase dengue transmission in general.

CONCLUSION AND RECOMMENDATIONS

This study modeled the high frequency weekly aggregated moving average of dengue surveillance data from Butuan City using meteorological parameters as regressors. Necessarily, the BVAR Model was applied, and a stable model was produced. The first five lags for each of the response variable and total rainfall were generated to run an independent BMH Model with certain specifications. Simulation shows that gains in predictive accuracy in the introduced BVAR-BMH econometric procedures are encouraging.

The findings in the study provide crucial insights on (1) using high-frequency data as an alternative in the absence of long-term data for modeling whenever possible, (2) the advantages in accounting for possible effects of lags in epidemiological surveillance, and (3) using Bayesian approaches, specifically a combination thereof, in drawing inferences in the lens of modeling dengue, may produce calibrated predictions. The results of the proposed spatiotemporal model for the study set also show that the changes in the response variable could be mostly sensitive to both the past five reporting weeks of the cases and rainfall totals. Particularly, the results of the BVAR Models in Table 3 show that the present value of the dependent variable may be primarily affected by the sudden rise in the cases from one week prior and the increase in total rainfall from the past two weeks. It is therefore recommended that the dengue surveillance in the location be more stringent, especially after heavy rainfall, typhoons, and expected increases in rainfall amount during wet seasons. This can be valuable information, further, given that in the context of a changing climate, the variability in the spread of vector-borne diseases including dengue, and the unpredictability of the amount of rainfall or the lack of such posits major uncertainties and concerns. It is also strongly recommended that regressors and lags selection procedure be included in modeling epidemiological data when using meteorological parameters. The results may also be used by the local government of Butuan City, and others, in developing a localized early warning system for the surveillance of dengue in the said location. Lastly, statisticians and experts may develop statistical packages for free and trusted software for ease of the proposed modeling procedure.

REFERENCES

Andrieu, Christophe, and Johannes Thoms. "A Tutorial on Adaptive MCMC." *Statistics and Computing* 18, no. 4 (2008). https://doi.org/10.1007/s11222-008-9110-y.

Badurdeen, Shiraz, David Benitez Valladares, Jeremy Farrar, Ernesto Gozzer, Axel Kroeger, Novia Kuswara, Silvia Runge Ranzinger, et al. "Sharing Experiences: Towards an Evidence Based Model of Dengue Surveillance and Outbreak Response in Latin America and Asia." *BMC Public Health* 13, no. 1 (2013). https://doi.org/10.1186/1471-2458-13-607.

BaŞtürk, Nalan, Cem Çakmaklı, S. Pınar Ceyhan, and Herman K. Dijk. "On the Rise of Bayesian Econometrics after Cowles Foundation Monographs 10, 14." *OEconomia*, no. 4-3 (2014): 381–447. https://doi.org/10.4000/oeconomia.913.

Buckingham-Jeffery, Elizabeth, Roger Morbey, Thomas House, Alex J. Elliot, Sally Harcourt, and Gillian E. Smith. "Correcting for Day of the Week and Public Holiday Effects: Improving a National Daily Syndromic Surveillance Service for Detecting Public Health Threats." *BMC Public Health* 17, no. 1 (2017). https://doi.org/10.1186/s12889-017-4372-y.

City Government of Butuan. "Butuan City." City Government of Butuan. Accessed 2023. https://butuan.gov.ph/.

Dickey, D. A., and W. A. Fuller. "Distribution of the Estimators for Autoregressive Time Series With a Unit Root." JSTOR. JSTOR, 1979. https://www.jstor.org/stable/pdf/2286348.pdf.

Doan, Thomas, Robert Litterman, and Christopher Sims. "Forecasting and Conditional Projection Using Realistic Prior Distributions." *Econometric Reviews* 3, no. 1 (1984): 1–100. https://doi.org/10.1080/07474938408800053.

Dunson, D. B. "Commentary: Practical Advantages of Bayesian Analysis of Epidemiologic Data." *American Journal of Epidemiology* 153, no. 12 (2001): 1222–26. https://doi.org/10.1093/aje/153.12.1222.

Foster, D P, and S Hart. https://arxiv.org/pdf/2210.07169.pdf, 2021.

Gelman, A., W. R. Gilks, and G. O. Roberts. "Weak Convergence and Optimal Scaling of Random Walk Metropolis Algorithms." *The Annals of Applied Probability* 7, no. 1 (1997). https://doi.org/10.1214/aoap/1034625254.

Gharbi, Myriam, Philippe Quenel, Joël Gustave, Sylvie Cassadou, Guy La Ruche, Laurent Girdary, and Laurence Marrama. "Time Series Analysis of Dengue Incidence in Guadeloupe, French West Indies: Forecasting Models Using Climate Variables as Predictors." *BMC Infectious Diseases* 11, no. 1 (2011). https://doi.org/10.1186/1471-2334-11-166.

Gilks, W. R., S. Richardson, D. J. Spiegelhalter, and A E Raftery. "Hypothesis Testing and Model Selection." Essay. In *Markov Chain Monte Carlo in Practice*, 163–87. Boca Raton, FL: Chapman & Hall, 1996.

Giordani, Paolo, and Robert Kohn. "Adaptive Independent Metropolis–Hastings by Fast Estimation of Mixtures of Normals." *Journal of Computational and Graphical Statistics* 19, no. 2 (2010): 243–59. https://doi.org/10.1198/jcgs.2009.07174.

Haario, Heikki, Eero Saksman, and Johanna Tamminen. "An Adaptive Metropolis Algorithm." *Bernoulli* 7, no. 2 (2001). https://doi.org/10.2307/3318737.

Hobbs, Benjamin F., Philip T. Chao, and Boddu N. Venkatesh. "Using Decision Analysis to Include Climate Change in Water Resources Decision Making - Climatic Change." SpringerLink. Kluwer Academic Publishers, 1997. https://link.springer.com/article/10.1023/A:1005376622183.

Kuschnig, N, and L Vashold. "Hierarchical Bayesian Vector Autoregression [R Package Bvar Version 1.0.4]." The Comprehensive R Archive Network. Comprehensive R Archive Network (CRAN), 2023. https://cran.r-project.org/web/packages/BVAR/index.html.

Leung, Xing Yu, Rakibul M. Islam, Mohammadmehdi Adhami, Dragan Ilic, Lara McDonald, Shanika Palawaththa, Basia Diug, Saif U. Munshi, and Md Nazmul Karim. "A Systematic Review of Dengue Outbreak Prediction Models: Current Scenario and Future Directions." *PLOS Neglected Tropical Diseases* 17, no. 2 (2023). https://doi.org/10.1371/journal.pntd.0010631.

Litterman, R B. *A Bayesian Procedure for Forecasting with Vector Auto-Regression. Working Paper*. Cambridge, MA: Massachusetts Institute of Technology, Department of Economics., 1980.

Litterman, Robert B. "Forecasting with Bayesian Vector Autoregressions: Five Years of Experience." *Journal of Business & Economic Statistics* 4, no. 1 (1986): 25. https://doi.org/10.2307/1391384.

Lütkepohl, Helmut. *New Introduction to Multiple Time Series Analysis*. Berlin: Springer, 2005.

Naish, Suchithra, Pat Dale, John S Mackenzie, John McBride, Kerrie Mengersen, and Shilu Tong. "Climate Change and Dengue: A Critical and Systematic Review of Quantitative Modelling Approaches." *BMC Infectious Diseases* 14, no. 1 (2014). https://doi.org/10.1186/1471-2334-14-167.

Pesaran MH, Shin Y, Smith RJ. Bounds testing approaches to the analysis of level relationships. Journal of Applied Econometrics. 2001;16(3):289–326. doi:10.1002/jae.616

PSA. "2020 Census of Population and Housing (2020 CPH) Population Counts Declared Official by the President." Philippine Statistics Authority. Philippine Statistics Authority, July 7, 2021. https://psa.gov.ph/content/2020-census-population-and-housing-2020-cph-population-counts-declared-official-president.

ReliefWeb. "Philippines: Dengue Outbreak - Jul 2019." ReliefWeb, July 2019. https://reliefweb.int/disaster/ep-2019-000084-phl.

Roberts, Gareth O., and Jeffrey S. Rosenthal. "Optimal Scaling for Various Metropolis-Hastings Algorithms." *Statistical Science* 16, no. 4 (2001). https://doi.org/10.1214/ss/1015346320.

StataCorp. "Bayes: Var — Bayesian Vector Autoregressive Models." Texas, 2021a. https://www.stata.com/manuals/bayesbayesvar.pdf.

StataCorp. "bayesmh — Bayesian models using Metropolis–Hastings algorithm." Texas, 2021b. https://www.stata.com/manuals/bayesbayesmh.pdf.

Sugita K. Forecasting with Bayesian vector autoregressive models: Comparison of direct and iterated multistep methods. Asian Journal of Economics and Banking. 2022;6(2):142–54. doi:10.1108/ajeb-04-2022-0044

Wilder-Smith, Annelies, and Duane J. Gubler. "Geographic Expansion of Dengue: The Impact of International Travel." *Medical Clinics of North America* 92, no. 6 (2008): 1377–90. https://doi.org/10.1016/j.mcna.2008.07.002.

Yildirim I. Bayesian Inference: Metropolis-Hastings Sampling [Internet]. Lecture presented at; Rochester; 2012. Available from: https://nlp.jbnu.ac.kr/PGM/slides_other/MetropolisHastingsSampling.pdf

Yin, Shi, Chao Ren, Yuan Shi, Junyi Hua, Hsiang-Yu Yuan, and Lin-Wei Tian. "A Systematic Review on Modeling Methods and Influential Factors for Mapping Dengue-Related Risk in Urban Settings." *International Journal of Environmental Research and Public Health* 19, no. 22 (2022): 15265. https://doi.org/10.3390/ijerph192215265.
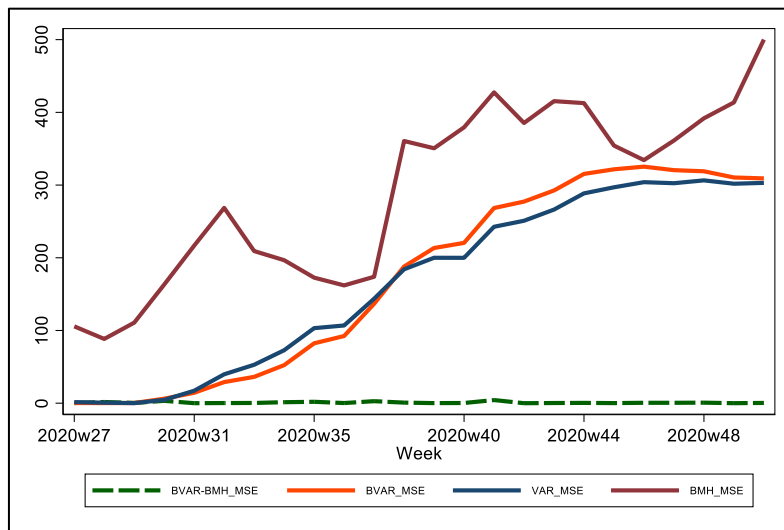
Appendix 1: Comparison of the MSE from the four methods.
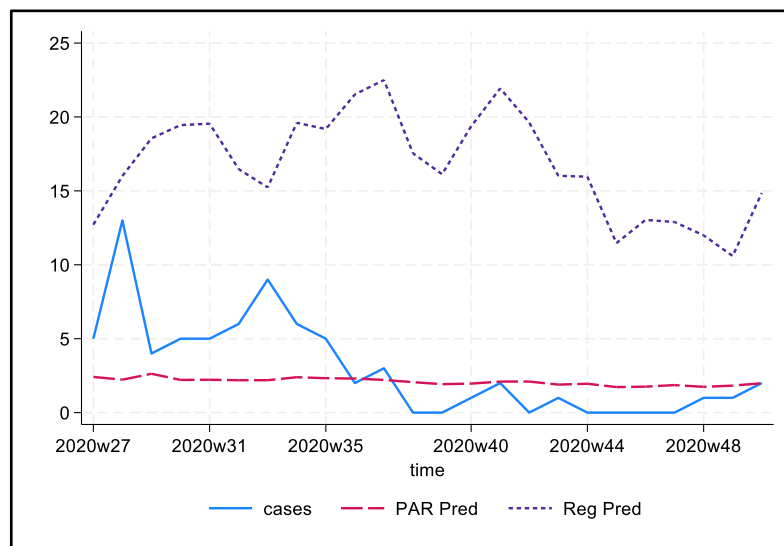
Appendix 2: Comparison of the predictions from customary frequentist model