Performance Evaluation and Comparison of Integer-Valued Time Series Models for Measles Outbreak Detection in Cavite

Vio Jianu C. Mojica
Department of Physical Sciences and Mathematics,
University of the Philippines Manila
Mathematics and Statistics Department,
De La Salle University

Frumencio F. Co
Mathematics and Statistics Department,
De La Salle University

An ideal outbreak detection algorithm must be able to generate alarms early into an outbreak while providing optimal sensitivity and specificity so as to mitigate mortality and other potential costs of investigation and response to these events. One particular disease of interest is measles, which is a highly contagious disease that exhibited periodic outbreaks in the Philippines. The performance of the NGINAR(1) and ZINGINAR(1) models for measles outbreak detection was examined through the use of simulated datasets and an actual application to reported measles cases in the Cavite province from 2010 to 2017. The models were evaluated based on their goodness-of-fit as well as the sensitivity, specificity, and timeliness of the detection thresholds they have generated. Comparisons were done against ARIMA models and the popular Poisson INAR(1) model. Results show that INAR models have considerably higher probabilities of detection than ARIMA models, particularly for outbreaks of small magnitudes. The Poisson INAR(1) generates the most alarms and thus, has the highest sensitivity metrics. The NGINAR(1) and ZINGINAR(1) models, however, have lower false positive rates with outbreak detection capabilities comparable to the Poisson INAR(1). The NGINAR(1) model may be chosen as the best model considering its simplicity and its balance of sensitivity, specificity, and timeliness which is optimal for a disease such as measles.

Keywords: NGINAR(1), ZINGINAR(1), measles, outbreak detection, Cavite

1. Introduction

An outbreak is an excess from normal levels of disease cases in a population. However, not all exceedances are sufficient to label an event as an outbreak. Determining whether an event is an outbreak or not includes a subjective element – it depends not only on the magnitude of the excess, but also on other factors such as whether the impact of not investigating an event is greater than the cost of investigation (Dato, Shephard, and Wagner, 2006). Also, definitions of outbreaks vary for different diseases. The World Health Organization (WHO) states that diseases with different incidence rates or immunization program objectives would have different thresholds for outbreaks (WHO, 1999). Outbreaks, when not detected and controlled early, can lead not only to increased morbidity and mortality, but also to incurred costs from activities in response to these events. Thus, outbreak detection must not only be timely but also efficient in the sense that false alarms are minimized, so as to mitigate the cost of investigation and response.

As an example, an outbreak of measles was recently declared on February 2019 in several regions of the Philippines, with the most cases and deaths from Metro Manila and Region IV-A (CALABARZON) – the region comprised of the Cavite, Laguna, Batangas, Rizal, and Quezon provinces (DOH, 2019a). Naturally, due to the increase in cases and deaths, the Department of Health (DOH), along with other government agencies have launched several programs and activities in response to the outbreak. This includes the deployment of vaccination centers in public places malls and fast food chains and massive information dissemination and immunization campaign (DOH-CALABARZON, 2019a; DOH-CALABARZON, 2019b; DOH, 2019b).

With the knowledge of the possible negative impacts of a disease outbreak, detecting them in their early stages is ideal. Due to the time dependent nature of data for outbreak detection (e.g., weekly or monthly reported cases of a certain disease), along with the emphasis on timeliness of detection, some common algorithms for outbreak detection are purely temporal. The problem of outbreak detection can be translated to a signal detection problem in which an alarm will be made once a monitored statistic exceeds a set threshold. The use of time series models is one class of statistical methodology used for this purpose. It has some merits of its own in that models have no assumption of temporal independence, instead, observations are assumed to exhibit a degree of autocorrelation which might be a more appropriate assumption for a disease incidence series. Also, one important objective of time series modelling is forecasting of future values. Forecasting would be useful for prospective planning and control for disease surveillance as this enables concerned institutions or individuals an idea about the future behavior of diseases.

As early as the paper by Serfling (1963) on detecting influenza outbreaks, forecast limits have been used to define outbreak detection thresholds. Allard

(1998) suggested the use of the upper forecast limits provided by the autoregressive integrated moving average (ARIMA) model as a threshold for signaling an outbreak for some infectious diseases. However, ARIMA models have an assumption of normality which do not reflect the integer-valued nature of disease incidence that involves the use of reported cases (count data). Furthermore, counts may exhibit skewed distributions, especially for rare diseases in which daily or weekly reported cases may exhibit an excess of zeros. Also, ARIMA models were observed to adjust to local peaks in the series and thus, cannot detect outbreaks of small magnitudes (Reis and Mandl, 2003). Cardinal, Roy, and Lambert (1999) proposed the use of integer valued autoregressive (INAR) models which were shown to provide smaller forecast errors and non-negative integer-valued forecast limits. Paman et al. (2017) have successfully fitted Poisson INAR models on daily reported cases of measles in Metro Manila and compared its performance with ARIMA, in terms of sensitivity, false positive rate, and relative delay in detecting outbreaks. Their results show that INAR models performed favorably as compared to ARIMA. However, it was also shown that the Poisson INAR models might not be appropriate for the counts of measles as it exhibits empirical overdispersion.

Ristić, Bakouch, and Nastić (2009) proposed the new geometric first-order integer-valued autoregressive (NGINAR(1)) model as an alternative way for modelling count data. In modelling monthly counts of sex offense, the NGINAR(1) model was shown to perform favorably compared to various first-order INAR models with different distributional assumptions in terms of the Akaike Information Criterion (AIC) and Residual Mean Squares (RMS). Ristić, Bourguignon, and Nastić (2018) proposed the use of a zero-inflated NGINAR(1) model or ZINGINAR(1) model with zero-inflated geometric marginals as an alternative when dealing with data that have more zeroes than expected. However, the adequacy of both the NGINAR(1) and ZINGINAR(1) models for outbreak detection has not yet been evaluated and thus, it is of interest to determine if there are any practical advantages of using these models as compared to other existing models.

This research was conducted to provide alternative algorithms which may be optimal for measles outbreak detection. Hence, it is aimed to examine how the conceptual advantages of the NGINAR(1) and ZINGINAR(1) such as considering overdispersion and zero-inflation translate to the problem of outbreak detection. Particularly, the specific objectives of the study include: (1) to illustrate the use of thresholds derived from the Poisson INAR(1), NGINAR(1), and ZINGINAR(1) models for outbreak detection; (2) to demonstrate the applicability of the detection algorithms to an actual measles dataset from Cavite; and (3) to evaluate and compare the performance of these models with ARIMA models.

The study focuses only on the comparison of time series models, with emphasis on the integer-valued models, for outbreak detection. Furthermore, to maintain the simplicity of model fitting and interpretability in the context of

disease surveillance, more complex developments in modelling discrete-valued time series such as combined INAR(p) models and INAR models with random coefficient (RC) thinning were not included. Weiss (2018) posits that while these models exhibit some interesting properties, the data generating mechanisms are artificial, and not easily interpretable. Lastly, results may only be generalized to measles incidence exhibiting a similar behavior to that of Cavite (i.e., similar baseline series and outbreak characteristics).

2. Measles Case Definition and Investigation

This section summarizes standard case definitions, symptoms, and detection thresholds for measles which have been documented in the Manual of Procedures for the Philippine Integrated Disease Surveillance and Response (PIDSR) prepared by the National Epidemiology Center (NEC) of the DOH (DOH-NEC, 2008; DOH-NEC, 2014).

The DOH defines measles as an acute highly communicable illness characterized by a prodrome of fever, conjunctivitis, cough, coryza, and Koplik spots followed by maculopapular rash on the third to seventh day. Transmission of the disease is through direct contact or articles freshly soiled with nasal or throat secretions of infected persons. Incubation period ranges from 7 to 21 days from exposure to onset of fever and usually 14 days until rash appears.

They also classify cases of measles as suspected, laboratory-confirmed, or epidemiologically-linked. Suspected cases of measles are defined by the appearance of one or more of the prodromes of the disease. A suspected case is reclassified as a laboratory-confirmed case after successful laboratory diagnosis while an epidemiologically-linked case of measles is defined as a suspected measles case, which had contact with another epidemiologically-linked case or a laboratory confirmed case 7 to 21 days before the onset of rash while the latter was infectious at the time of contact. Lastly, cases will be discarded or reclassified when a suspected case is not serologically confirmed or confirmed for another disease such as rubella or dengue.

Thresholds for action are set by the DOH to determine if disease levels warrant investigation or urgent response. Alert thresholds serve as an early warning threshold while epidemic thresholds warrant immediate response. Alert thresholds are computed by taking weekly or monthly averages of a particular disease during the past three to five years and adding one standard deviation, while for epidemic thresholds, two standard deviations are added. For measles, a suggested threshold of one suspected case is enough to trigger an alert. A confirmed outbreak is when the number of cases exceeds the epidemic threshold.

3. Definition of Outbreak Periods

While there are several methods proposed for outbreak detection, the issue of having a standardized evaluation procedure for these detection algorithms has

been less addressed. Watkins, Eagleson, Hall, Dailey, and Plant (2006) provides a review of 63 studies on their approaches to the evaluation of outbreak detection algorithms. One main issue identified is the definition of an outbreak criterion which will be used as gold standard to evaluate sensitivity, specificity, and timeliness of the algorithms.

The review by Watkins et al. (2006) provides different approaches in defining outbreak periods. One approach involves no outbreak criterion, relying only on descriptions of the outbreak detection algorithm such as the number of alarms and time of first alarm before peak number of cases. Alternatively, some of the studies included in their review define an outbreak criterion through an arbitrary threshold or a fixed number of cases; this approach has the advantage of simplicity. However, they have also noted that this approach may be improved by considering expert opinion and the characteristics of the disease in defining thresholds, thus, capturing the subjectivity of the definition of outbreaks.

A simulation approach was also discussed in their review in which purely simulated datasets or real datasets with simulated outbreaks are used. In this approach, the researcher can freely define the start and end of outbreaks, at the cost of authenticity. Watkins et al. (2006) suggests using a combination of these approaches to maximize the advantages of each.

For this study, a combination of these approaches was applied to a measles incidence time series dataset from Cavite to define outbreak periods. The data consists of weekly counts of measles in Cavite from January 1, 2010 to December 31, 2017 (i.e., 417 weeks). The counts consist of laboratory-confirmed and epidemiologically-linked cases. Data was requested from the Provincial Epidemiology and Surveillance Unit (PESU) of the Cavite province.

To retrospectively define which events are considered as outbreaks, as well as to determine when these periods start and end, some thresholds and padding measures were employed. The procedures for identifying outbreak periods from real data were adapted from Rolfhamre and Ekdahl (2006), while the procedure for setting the beginning and end of an outbreak period was based on Paman et al. (2017). The following steps were carried out:

1. The series was first inspected and periods with unexpected peaks in measles incidence were identified subjectively and marked as suspected outbreaks.
2. Since not all exceedances could be considered outbreaks, associated reports of an outbreak were investigated for the events identified. Only the suspected outbreaks with an associated report were considered as true outbreaks.
3. The start of an outbreak will be marked as the first of two consecutive weeks with more than three reported measles cases. The end of the outbreak will be marked as the last of two consecutive weeks with three or less reported measles cases. The threshold of more than three reported measles cases was based on the series mean.

Considering that the true start and end of an outbreak cannot be exactly identified, an adjustment of three weeks before the start, and after the end of an outbreak. These weeks will be the basis for the definition of the start and end of an outbreak to be used throughout the study. A three-week padding was based on an upper bound of the number of days from exposure to rash onset.

4. Simulation Framework

While the usage of real datasets has the advantage of authenticity, it has the disadvantage of not knowing the exact start and end of an outbreak as well as the fact that algorithms presented will only be applied to a single dataset. This leads to the inability to test the generalizability of results and to generate empirical distributions for the performance metrics. To address this, simulated datasets similar to the Cavite measles series were generated. This simulation procedure was carried out in a similar manner with Paman et al. (2017). The procedure was done as follows:

1. The original series was used as a reference dataset to estimate mean, trend, and seasonality parameters of measles. In carrying this out, identified outbreak periods were first removed from the reference series to capture the behavior of the series without the aberrations. A Serfling model was then fitted to the remaining observations.

   The Serfling model (Serfling, 1963) is a cyclic regression model, which assumes that data is sinusoidal with a period of one year along with a secular trend. Variations of this model have been used in some recent studies to simulate datasets of a desired level of mean, trend and seasonality (Paman et al., 2017; Bédubourg and Le Strat, 2017). The Serfling model fitted was of the form

   $$\hat{Y}_t = \beta_0 + \beta_1 t + a\cos{+}\frac{2\pi t}{52} + b\sin{+}\frac{2\pi t}{52} + \varepsilon_t, \qquad (1)$$

   where β_0 is the constant term, β_1 is the coefficient for the linear trend, a and b are the amplitudes of the Fourier terms, and ε_t constitutes the random error. The notations $\cos^+$ and $\sin^+$ denote the cosine and sine functions, respectively, with non-negative range. A nonlinear (weighted) least squares estimation procedure was employed to obtain the parameter estimates. This was carried out through the *nls*() function in R.

2. The random error component of the Serfling model is usually assumed to follow a pre-defined parametric distribution or an empirical distribution derived from the residuals of the Serfling model. For this study, the parameter estimates from the Serfling model were first used to generate a baseline curve for the simulated datasets. A random error term was then added to the fitted values. The error terms were derived through resampling with replacement

from the empirical distribution of the residuals from fitting the model to the original dataset. Values were then rounded off to the nearest integers to generate the final simulated baseline series.

3. Locally weighted regression (LOESS) was employed to simulate outbreaks. It is a non-parametric regression approach used when linear or polynomial regression surfaces cannot replicate the data (Cleveland and Devlin, 1988). Its equation is given by

$$y_i = g(x_i) + \varepsilon_i, \tag{2}$$

where y_i are the values of the response variable, x_i are the values from the matrix $\boldsymbol{X}$ of p predictors for individual i, and ε_i are random errors. The regression function $g(x_i)$ is approximated locally by fitting a regression surface to a certain window of data points.

To replicate the behavior of the Cavite series during outbreaks, the $\hat{Y}_t$ values from the simulated baseline series from (1) were subtracted from the actual values on the outbreak periods. The residuals were then fitted with a LOESS curve and added to the simulated baseline data. This was carried out using the *loess*() function in R.

The procedure was done to generate 1000 datasets with similar characteristics to the original series. with A burn-in period of 5000 simulations. The entire simulation procedure was carried out in R.

5. First-order Integer-Valued Autoregressive (INAR(1)) Models

5.1. The Poisson INAR(1) model

Integer-valued autoregressive (INAR) models were first introduced by McKenzie (1985) and Al-Osh and Alzaid (1987) with their INAR(1) model. This model provides a counterpart to the continuous AR(1) model, $X_t = \alpha \cdot X_{t-1} + \varepsilon_t$, where $|\alpha| < 1$, which does not preserve the integer-valued nature of count time series due to the multiplication of α even if the innovations ε_t. The INAR(1) model addresses this issue through the use of the binomial thinning operator '$\circ$' (Steutel and Van Harn, 1979) instead of multiplication. The model is given by

$$X_t = \alpha \circ X_{t-1} + \varepsilon_t, \tag{3}$$

where $\alpha \in [0,1)$, and the sequence of innovations $\{\varepsilon_t\}$ is composed of independent and identically distributed (i.i.d) random variables which are non-negative and integer-valued. The term $\alpha \circ X$ denotes the binomial thinning operation expressed as

$$\alpha \circ X = \sum_{i=0}^{X} Y_i, \tag{4}$$

where $\{Y_i\}$ is a sequence of i.i.d Bernoulli random variables with $P(Y_i = 1) = \alpha$ independent of X. The Y_i are referred to as the counting series. When used to model disease incidence, the number of new cases of a disease at time t, given by X_i is represented as the sum of (i) the number of new cases transmitted from those which developed during time $(t-2, t-1)$, where each of the latter successfully transmits the disease to one individual with probability α, and (ii) the number of cases generated from independent sources, given by ε_t. The innovations are commonly assumed to be Poisson, and due to the additive property and invariance with respect to the binomial thinning of the Poisson distribution, the observations X_t also follow a Poisson distribution. This particular specification is commonly denoted as the Poisson INAR(1) model.

The autocorrelation function (ACF) of the INAR(1) process is $\rho(k) = \alpha^k$, exponentially decaying for higher values of k. The partial autocorrelation function (PACF) for the process is given by $\rho_{part}(1) = \alpha$ and $\rho_{part}(k) = 0$ for $k > 1$. This suggests that the ACF and PACF of an INAR(1) process is similar to that of a real-valued AR(1) process.

5.2. INAR(1) models for counts with overdispersion and zero-inflation

Overdispersion is defined as the presence of greater variability in observations than expected from a given model while zero-inflation is defined as an excess of zeros as compared to what is expected from a given distributional assumption. Count time series, especially daily and weekly reported cases of infectious diseases, almost certainly exhibits both characteristics.

As discussed in Section 5.1, a Poisson distribution is commonly assumed for INAR models, comparable to the normal distribution for real-valued time series. One of its properties is that its variance (σ^2) is equal to its mean (μ). Weiss (2018) defines the Poisson index of dispersion as

$$I = \frac{\sigma^2}{\mu}, \tag{5}$$

which is used for non-negative integer-valued random variables with unlimited range ($\mathbb{N}_0$). An index of dispersion $I > 1$ denotes an overdispersed distribution.

For zero-inflation, Weiss (2018) defines the zero index

$$I_{zero} = 1 + \frac{\ln p_o}{\mu}, \tag{6}$$

where $p_0 = P(X = 0)$. The zero index takes the value of 0 for the Poisson distribution, thus indicates zero-inflation. The empirical index of dispersion $\hat{I} = \frac{S^2}{\overline{X}}$, where S^2 is the sample variance $\overline{X}$ and is the sample mean, is asymptotically unbiased for I

and thus, may be a good initial indicator of overdispersion with respect to Poisson (Schweer and Weiss, 2014).

To accommodate overdispersion, one alternative model proposed is the NGINAR(1) model introduced by Ristić et al. (2009). This model presents two main differences from the Poisson INAR(1) model in terms of construction: (i) the use of negative binomial thinning operator and (ii) the assumption of geometric marginal distribution for the observations X_t.

The NGINAR(1) process is given by

$$X_t = \alpha * X_{t-1} + \varepsilon_t, \tag{7}$$

where $\{X_t\}$ is a stationary process with Geometric $\left(\frac{1}{1+\mu}\right)$ marginals, the sequence of innovations $\{\varepsilon_t\}$ is an i.i.d. non-negative integer-valued process, and $\alpha \in [0,1)$. The term $\alpha * X$ denotes the negative binomial thinning operation expressed as

$$\alpha * X = \sum_{i=0}^{X} W_i, \tag{8}$$

where the counting process $\{W_i\}$ is a sequence of i.i.d. random variables with Geometric$\left(\frac{1}{1+\alpha}\right)$ distribution. The distribution of the innovations as derived by Ristić et al. (2009) is given by

$$P(\varepsilon_t = l) = \left(1 - \frac{\alpha\mu}{\mu - \alpha}\right)\frac{\mu^l}{(1+\mu)^{l+1}} + \frac{\alpha\mu}{\mu-\alpha}\cdot\frac{\alpha^l}{(1+\alpha)^{l+1}}, \tag{9}$$

which is a mixture of two geometric random variables. It must be noted that this probability is defined only for $\alpha \in \left[0, \frac{\mu}{(1+\mu)}\right]$, otherwise, probabilities generated might not all be non-negative.

The NGINAR(1) also has an AR(1)-like autocorrelation structure, with an ACF given by $\rho(k) = \alpha^k$. The model gives an alternative to the Poisson INAR(1) model by accommodating overdispersion and using the negative binomial thinning to allow the counting process to take on values other than 0 or 1. An interpretation of the NGINAR(1) process in terms of modelling disease incidence is similar to the INAR(1) model but with cases developing during time $(t-2, t-1)$ able to transmit to more than one individual with exponentially decaying probability.

The ZINGINAR(1) model was developed by Ristić et al. (2018) as an alternative to NGINAR(1) for data with an excess of zeros than expected from an NGINAR(1) process. It is given by:

$$Z_t = \begin{cases} 0, & \text{with probability } \pi \\ X_t, & \text{with probability } 1-\pi \end{cases} \quad (10)$$

where X_t is the NGINAR(1) process given in (7). Its ACF is given by

$$\rho(k) = \frac{(1-\pi)(1+\mu)\alpha^k}{1+\mu(1+\pi)}. \quad (11)$$

5.3. Parameter estimation and forecasting

For this study, Poisson INAR(1), NGINAR(1), ZINGINAR(1) models were fitted for each of the 1000 simulated datasets as well as the original dataset. Two ARIMA models were also included in the comparison: a real-valued AR(1) model and an ARIMA model identified through the *auto.arima()* function from the *forecast* package in R. The AR and MA orders for the latter are identified based on the Akaike Information Criterion (AIC) (Hyndman and Khandakar, 2008). These five constitute the models used for comparisons.

For parameter estimation, Yule-Walker, least squares, and maximum likelihood (ML) estimators were defined for Poisson INAR(1), NGINAR(1), and ZINGINAR(1) models on Weiss (2018), Ristić et al. (2009), and Ristić et al. (2018), respectively. However, ML estimators were employed in accordance to the results of the Monte Carlo simulations from Ristić et al. (2009) and Ristić et al. (2018). In these simulations, ML or conditional ML estimates were shown to exhibit the best parameter recovery in terms of unbiasedness and standard errors albeit by small margins as compared to Yule-Walker and least squares estimators. However, ML estimators for the INAR models do not have a closed form expression. Thus, ML estimation was carried out using numerical methods, through the *optim()* function in R. A quasi-Newton method for bound constrained optimization called the L-BFGS-B method proposed by Byrd, Lu, Nocedal, and Zhu (1995). This was chosen since the parameters for all models are bounded (e.g. no parameter can take on negative values).

The log-likelihood functions for the Poisson INAR(1), NGINAR(1) and ZINGINAR(1) models were first defined as the objective function and Yule-Walker estimates were set as initial values. The log-likelihood functions for each model were obtained from their respective conditional one-step ahead probability mass functions (pmf) presented in Freeland and McCabe (2004), Awale, Ramanathan, and Kale (2017), and Ristić et al. (2018), respectively. ML estimators based on the full likelihood function was defined for Poisson INAR(1) and NGINAR(1) model, while a conditional ML estimator was defined for the ZINGINAR(1) model. The conditional one-step ahead pmf for each model is presented below:

$$P\left(X_{t+1}=y \mid X_t=x\right)=\sum_{s=0}^{\min(y,x)}\left[\binom{x}{s}\alpha^s(1-\alpha)^{x-s}\right]\left[\frac{1}{y-s!}e^{-\lambda}\lambda^{y-s}\right]. \tag{12}$$

$$P\left(X_{t+1}=y \mid X_t=x\right)=\sum_{r=0}^{y}P(\alpha * x=r)P(\varepsilon_t=y-r), \text{where} \tag{13}$$

$$P(\alpha * x=r)=P\left(\sum_{i=1}^{x}W_i=r\right)=\binom{r+x-1}{r}p^x\left(1-p\right)^r,$$

$$P(Z_{t+1}=z \mid Z_t=u)$$

$$=\begin{cases}\pi I_{\{z=0\}}+\pi(1-\pi)\dfrac{P(X_{t+1}=Z}{P(Z_{\ t}=0)}+(1-\pi)^2\dfrac{P(X_{t+1}=z.X_t=0}{P(Z_t=0)}, & z\geq 0, u=0\\ \pi+(1-\pi)P(X_{t+1}=0 \mid X_t=u), & z=0, u>0\\ (1-\pi)P(X_{t+1}=z \mid X_t=u), & z=0, u>0\end{cases}$$

For the ZINGINAR(1) model, the optimization algorithm does not converge with Yule-Walker estimates as initial values. Accordingly, a grid search algorithm was initially employed to locate local peaks in the log-likelihood function. For the grid search algorithm, a total of 1620 parameter combinations were examined to determine which maximizes the log-likelihood function. For the parameter α, the values examined were from 0.1 to 0.9, by 0.1 units; for μ: 1 to 20 by 1 unit; and for π: 0.1 to 0.9 by 0.1 units. The optimal combination of parameters was used as initial values for the *optim()* procedure instead of the Yule-Walker estimates. The success of the *optim()* procedure is dependent on the initial values used; the grid search algorithm was done in order to utilize initial values which are sufficiently close to the optimal ML parameter estimates.

Forecasting was also carried out through an R program by calculating one-step ahead conditional means $E[X_{t+1}|X_t]$. Outbreak detection thresholds were generated by taking the upper 95% forecast limit for each model. Both the point forecasts and forecast limits were derived from the one-step ahead conditional distributions for Poisson INAR(1), NGINAR(1), and ZINGINAR(1) as shown in (12), (13), and (14), respectively.

6. Performance Evaluation of Outbreak Detection Algorithms

Forecasting accuracy and model fit were first measured by the in-sample Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the Akaike Information Criterion (AIC). These quantities were computed for each iteration of model fitting for each simulated series. The outbreak detection thresholds

generated were evaluated for their accuracy and timeliness in detecting outbreaks using the performance metrics described below. To simplify the definition of the performance metrics, we denote the number of outbreak weeks with an alarm as True Positives (*TP*), the number of non-outbreak weeks without an alarm as True Negatives (*TN*), the number of non-outbreak weeks with an alarm as False Positive (*FP*), and the number of outbreak weeks without an alarm as False Negative (*FN*):

- *Sensitivity* (Se). This metric is computed as $Se=TP/(TP+FN)$. This measures how often the model correctly identifies outbreak weeks among all alarms generated.
- *False Positive Rate* (FPR). This is defined as the proportion of weeks with an alarm in the absence of an outbreak. This is computed as $FPR=FP/(FP+TN)$.
- *Specificity* (Sp). This is defined as the proportion of non-outbreak weeks correctly identified by the model. This is computed as $Sp=TN/(TN+FP)$.
- *Probability of Detection* (POD). This is an event-based sensitivity-like metric computed for each outbreak identified. If an alarm is generated at least once within an outbreak period, then it is considered detected.
- *Probability of detection during the first week* (POD1wk). This is another sensitivity-like metric which also assesses a model's ability to make timely detections. It is defined as the proportion of correctly identified outbreaks on its first week.
- *Average time before detection* (ATBD). This is a measure of timeliness computed as the average number of weeks between the start of an outbreak and its detection.
- *Positive predictive value* (PPV), $PPV = \frac{TP}{TP + FP}$.
- *Negative predictive value* (NPV), $NPV = \frac{TN}{TN + FN}$.

All measures were computed for each of the 1000 simulated datasets by fitting the five models and computing each measure (i.e., MAE, RMSE, AIC, Se, Sp, POD,...) for each repetition. The POD and POD1wk measures were adapted from Bédubourg and Le Strat (2017).

To compare the performance of the five fitted models, the average for each performance metric was taken and used as the basis for comparison. In this way, event-based measures such as POD and POD1wk were computed as proportions of the 1000 simulations where each model has captured the event of interest. Also, the average time before detection (ATBD), as measured in weeks, was computed

conditional to the algorithm successfully detecting the outbreak. For comparisons, models with generate low false positive rates (FPR) and low delays in detection, as well as high sensitivity, specificity, and probability of detection (POD, and POD1wk) are considered to perform better. Furthermore, themes for comparison include (i) comparing INAR and ARIMA models in general and (ii) comparing the three INAR models.

The procedure for performance evaluation and comparison for the original dataset is similar to the simulated dataset. The performance measures for forecast accuracy and model fit, as well as the outbreak detection accuracy and timeliness were computed for a single dataset. Event-based measures were instead recorded as binary variables indicating whether the event of interest was captured by a model (i.e., POD = 1 if an outbreak is detected, and 0 if not). Additionally, a comparison of forecast and model fit during outbreak periods was included for the original dataset, so as to determine if models considering overdispersion better capture the variation in the series.

7. Results and Discussion

7.1. Preliminary results

An initial investigation of reported outbreaks in Cavite suggested that there have been two outbreaks in the province during the covered period. The first outbreak was reported around mid-February 2010 (ABS-CBN News, 2010), while the second outbreak was reported around late January 2014 (Manila Bulletin, 2014). Ylade (2018) reported that the latter outbreak started in 2013 for some regions of the Philippines including CALABARZON. Thus, this outbreak might already be on its course as of January 2014. Using the outbreak definition criteria discussed in Section 3, the first outbreak was found to have an average of 4.45 cases per week while the second outbreak has an average of 18.26 cases per week. There are also 189 zeroes in the series.

The non-outbreak series has a sample mean of 0.88 cases per week and exhibit an empirical index of dispersion $\hat{I} = 1.89$. A method of moments estimate for the zero index can be obtained as $\hat{I}_{zero} = 1 + \frac{\ln\left(\frac{1}{n}\sum I_{\{X_t=0\}}\right)}{\overline{x}} = 0.2378$. These indices indicate that the measles series is likely to be from an overdispersed and zero-inflated distribution with respect to the Poisson distribution. A plot of the data and the identified outbreak periods are shown in Figure 1.

Regarding the outbreak periods defined, the approaches discussed by Watkins et al. were considered. The threshold of more than three cases based on the series mean was found to be similar to the threshold utilized by the WHO Region of the Americas for a defined geographical area as of 1999 (WHO, 1999). This is a time

when the region was in a measles elimination phase, slightly before achieving sustained elimination of endemic measles in 2002 (Bellini and Rota, 2011). This phase is similar to the scenario in the Philippines in the covered period (2010-2017).

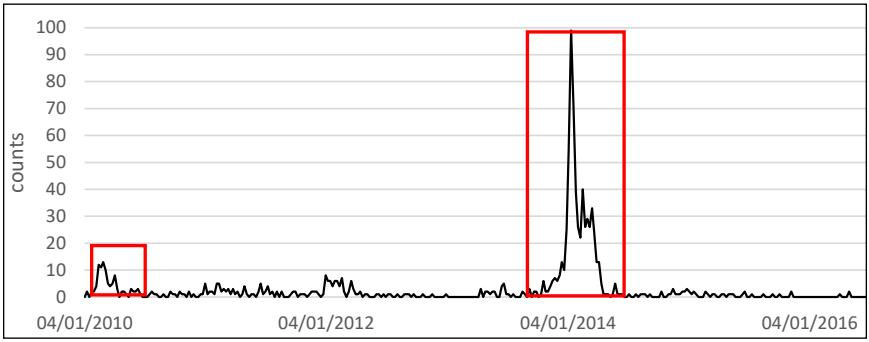

Figure 1. Reported measles cases in Cavite from January 1, 2010 to December 31, 2017

A slight peak at the beginning of 2012 was not identified as an outbreak based on the criteria set in the methodology section since no reports were tied to this suspected outbreak. The two identified outbreaks also satisfy the estimated inter-epidemic interval of 4-8 years in the vaccination era stated by Cutts, Henderson, Clements, Chen, and Patriarca (1991), and WHO (1999). Relating to Watkins et al. (2006), the outbreak definition used for the study has its strengths in that it maximizes the simplicity of defining outbreak periods (i.e., setting a defined threshold of three cases), while considering subjectivity and properties of the disease (i.e., setting padding measures to consider the early stages of measles, and examining official reports to consider if these events are enough to merit attention from the public health sector). Furthermore, a simulation approach was also used to determine if results are generalizable for similar series.

The baseline series is characterized by the fitted Serfling model. Parameter estimates from this model are shown in Table 1. The Serfling parameters suggest that the measles baseline series, which pertain to the normal level of the disease, exhibits a sustained low incidence with a faint decreasing trend and slight annual peaks even in the absence of the outbreaks.

Table 1. Parameter estimates from the Fitted Serfling Model

Parameter	Estimate
β_0	1.6472
β_1	-0.0034
a	0.3045
b	0.2005

LOESS curves were fitted to the observations in the outbreak periods. This was done to replicate the behavior of measles cases during the two defined outbreaks. The fitted values were added to the baseline series to simulate the outbreaks. A total of 1000 simulations were generated. Figure 2 shows the plot of a portion of one simulated series.

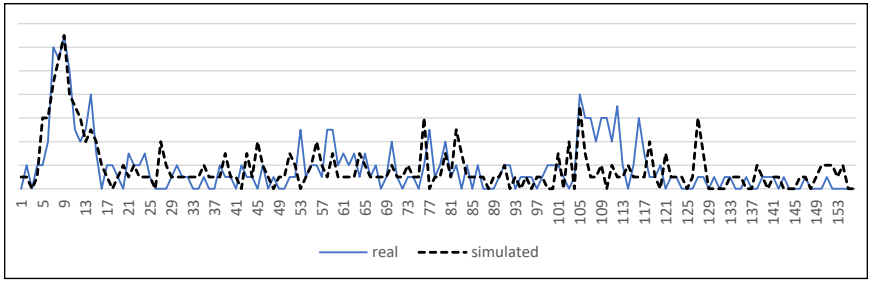

Figure 2. A portion of one of the simulated series (broken lines) plotted against the original series (solid line). The first outbreak is shown on the left side (peak at week 9).

7.2 Performance evaluation with simulated data

7.2.1 Evaluation of point forecasts

For the point forecasts, one-step ahead conditional means were used, thus, even INAR models do not generate integer-valued forecasts. However, only the ARIMA models generated by the *auto.arima()* procedure generate negative forecasts which are not desired when forecasting counts.

In comparing MAE and RMSE for the 1000 simulated series, the difference between INAR models and ARIMA models are evident. The distributions of MAE and RMSE are shown in Figure 3. ARIMA models have generally lower MAE and RMSE as compared to the INAR models. These results did not exactly follow that of Cardinal et al. (1999), where INAR models provided lower forecast errors than real-valued AR counterparts in modelling meningococcal infections. This difference emphasizes that the type of disease, and objective of modelling, must be considered in choosing a model. Negative forecasts from the Auto-ARIMA model diminishes this perceived advantage in terms of MAE and RMSE. Also, it must be noted that this difference between ARIMA and INAR models, when put into context, is not extreme. For example, the MAE of NGINAR(1) and AR(1) models have a discrepancy of only 0.2405 cases, on average.

Besides MAE and RMSE, the AIC was also used for comparison. Hyndman (2013) suggests that AIC is optimal for forecasting and that while it is computed in-sample, minimizing it is equivalent to minimizing the MSE of out-of-sample one-step ahead forecasts, making it a useful criterion in selecting a model. The

AIC was also stated to be comparable across different likelihoods if computed from the same data. The distribution of AIC for each model is shown in Figure 3. Based on AIC, the NGINAR(1) and ZINGINAR(1) models were found to provide a substantial improvement as compared to both the Poisson INAR(1) and the ARIMA models.

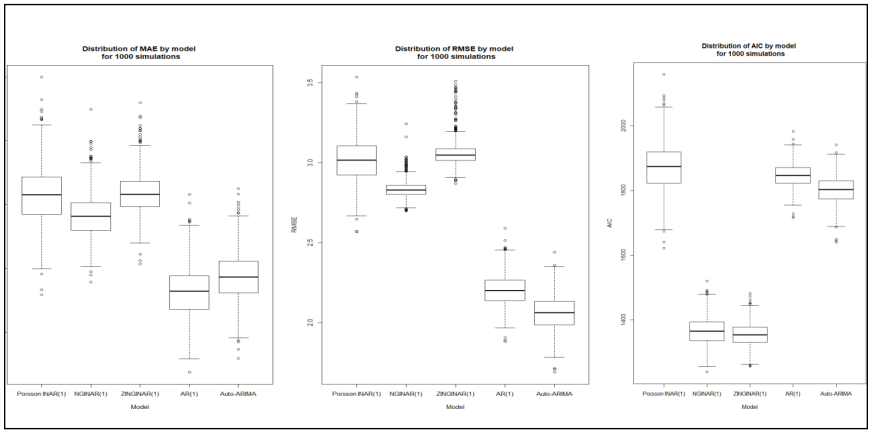

Figure 3. (From left to right) Distribution of MAE, RMSE, and AIC by model for 1000 simulations. Models compared (from left to right) are the Poisson INAR(1), NGINAR(1), ZINGINAR(1), AR(1), and Auto-ARIMA models.

7.2.2 Evaluation of outbreak detection thresholds

In evaluating the outbreak detection thresholds derived from each of the five models, two characteristics are focused on – accuracy, as measured by sensitivity and specificity metrics, and timeliness. Presented in Table 2 are averages for some performance metrics reflecting the accuracy of outbreak detection for each model. It was shown that Poisson INAR(1) generates an alarm considerably more often than the NGINAR(1) and ZINGINAR(1) model and thus, has the highest mean sensitivity rate for all models including the ARIMA models. The trend in sensitivity is similar to that of false positive rates and opposite of specificities due to the natural trade-off between them. This particular trend is expected since the Poisson INAR(1) model tends to underestimate the variance from a series that exhibits empirical overdispersion with respect to the Poisson assumption. The considerable difference in the sensitivity of the Poisson INAR(1) model suggests that this underestimation of the variance is substantial.

For the simulated datasets, POD1 and POD2 were calculated as the proportion of times when the models generated an alarm at any point over the course of the first and second outbreak, respectively. For the first outbreak, it must be noted that only Poisson INAR(1) never failed to detect the first outbreak. In this aspect, the advantage of INAR models as compared to ARIMA is clear, as both the AR(1) and Auto-ARIMA algorithms exhibited less than 60% POD1.

Table 2. Outbreak detection accuracy measures for the Poisson INAR(1), NGINAR(1), ZINGINAR(1), AR(1) and Auto-ARIMA models (averages from simulated datasets)

Measures	Poisson INAR(1)	NGI-NAR(1)	ZINGI-NAR(1)	AR(1)	Auto-ARIMA
alarms	55.93	28.73	19.09	14.07	15.63
Se	0.4539	0.2014	0.1738	0.1436	0.1526
Sp	0.9132	0.9507	0.9732	0.9826	0.9796
FPR	0.0866	0.0492	0.0267	0.0174	0.0204
POD1	1.0000	0.9090	0.7930	0.5450	0.5890
POD2	1.0000	1.0000	1.0000	1.0000	1.0000
PPV	0.4424	0.3830	0.5035	0.5627	0.5322
NPV	0.9181	0.8887	0.8876	0.8849	0.8857

In comparing the detection capabilities of the five models, the sensitivity rate (Se) may be deemed less important than POD, particularly when the latter is used in conjunction with timeliness measures. While more alarms lead to higher sensitivity rates, the Poisson INAR(1) model, which generated the most alarms, only provided a Se of 45.39%, understating the results from POD1 and POD2 which show that the Poisson INAR(1) captures both outbreaks for all 1000 simulations. As pointed out by Reis and Mandl (2003), one characteristic of time series models is that they tend to adjust to local trends in the data. This leads to less alarms over the course of an outbreak since cases from the past week are considered in determining thresholds for the following week. The advantage of this effect is that in a particularly severe epidemic, the algorithm would not generate an alarm every week. On the other hand, the disadvantage is that these models may not be able to detect slowly spreading outbreaks. However, similar to Paman et al. (2017), the INAR models fitted in this study were able to detect both the small and large outbreaks, and at the same time, not generating alarms every week, hence the low Se's. In this sense, the NGINAR(1) model is particularly interesting as it exhibits a POD of 90.9% for the first outbreak while exhibiting about half the FPR of the Poisson INAR(1) model.

Regarding the timeliness of detection, the Poisson INAR(1) can be clearly seen to be more advantageous than the other INAR models and the ARIMA models. This is a by-product of its high sensitivity as compared to all other models. Measures for timeliness are shown in Table 3.

Table 3. Outbreak detection timeliness measures for the Poisson INAR(1), NGINAR(1), ZINGINAR(1), AR(1) and Auto-ARIMA models (averages from simulated datasets)

Measures	Poisson INAR(1)	NGINAR(1)	ZINGINAR(1)	AR(1)	Auto-ARIMA
POD1wk1	0	0	0	0	0
POD1wk2	1	0.975	0.975	0.965	0.967
ATBD1 (wks)	1.793	3.581	3.334	4.481	4.750
ATBD2 (wks)	0	0.135	0.151	0.231	0.206

Looking at each outbreak separately, the results show that none of the models have successfully detected the smaller first outbreak immediately on the first week. However, average time before detection for the first outbreak (ATBD1) is 1.793 for the Poisson INAR(1) model – better than all other models. Also, it must be noted that the ATBDs are computed only for simulations where a model successfully detects an outbreak. As such, while the ZINGINAR(1) model may have a slightly lower ATBD1 than the NGINAR(1), it must be noted that this is conditional to the model detecting the outbreak, and the model's POD1 is only at 79.3%. Furthermore, ARIMA models provided yet longer delays in detection as compared to the INAR models in times when they successfully detect the first outbreak which is below 60%.

For the more apparent second outbreak, while the Poisson INAR(1) model provided a 100% detection on the first week, the other models have all provided more than 95% POD1wk2, with INAR models still having some advantage over ARIMA models. Further probing into the delays in detecting the second outbreak show that if the NGINAR(1), ZINGINAR(1), AR(1), and Auto-ARIMA models did not detect the outbreak during the first week, then the average delays are 5.40, 6.04, 6.60, and 6.24 weeks, respectively.

Overall, the NGINAR(1) model may be considered optimal for outbreak detection from the simulation results since it detects both outbreaks with POD at above 90% with half as many alarms generated as the Poisson INAR(1), mitigating FPR. This result is particularly interesting since the probability of detection and the probability of early detection (POD1wk) do not vastly differ from the Poisson INAR(1) even with about half the number of alarms and sensitivity rate. The other three remaining models had PODs less than 80% and the ARIMA models were shown to generally perform poorly in both accuracy and timeliness in detecting outbreaks.

7.3 Performance evaluation with original data

7.3.1 Evaluation of point forecasts

For the original dataset, the parameter estimates for the fitted models were also generated through ML estimation and are summarized on Table 5. For the model generated through the *auto.arima()* function, the best ARIMA model was selected based on the criteria of lowest AIC (Hyndman and Khandakar, 2008). From the results of Table 4, it must be noted that the model identified by the *auto.arima()* procedure was an ARIMA(3,0,5) model with zero mean. This model is considerably less parsimonious as compared to the other four models due to the number of its parameters. Also, the ZINGINAR(1) model has a parameter estimate for π at 0.0789, which suggests that the measles series exhibits an excess in zeroes of only about 8% as compared to a geometric model. In terms of model parsimony, these two models are at a disadvantage. Simplicity is particularly important when dealing with data from various fields in the sense that simpler models are easier to explain and implement.

Table 4. Parameter estimates for the Poisson INAR(1), NGINAR(1), ZINGINAR(1), AR(1) and Auto-ARIMA models from the Cavite weekly measles dataset

Poisson INAR(1)		NGINAR(1)		ZINGINAR(1)		AR(1)		Auto-ARIMA (3,0,5) with zero mean			
α	0.6191	α	0.6921	α	0.7266	ϕ_1	0.8693	ϕ_1	-0.5785	θ_1	1.7563
λ	0.9344	μ	2.2517	μ	2.6579	μ	2.3963	ϕ_2	0.6364	θ_2	1.0066
				π	0.0789			ϕ_3	0.6657	θ_3	-0.2688
										θ_4	-0.5545
										θ_5	-0.1660

Table 5. MAE, RMSE, and AIC for the actual dataset

Measures	Poisson INAR(1)	NGINAR(1)	ZINGINAR(1)	AR(1)	Auto-ARIMA
MAE	1.6295	1.5344	1.5835	1.5109	1.6685
RMSE	4.4133	4.1798	4.3682	3.9267	3.5460
AIC	2089.7984	1292.2915	1310.2644	2330.6342	2258.6459

From Table 5, the discrepancy in terms of MAE and RMSE between the INAR models and ARIMA models are less pronounced for the actual dataset as compared to the simulated datasets. Also, similar to the results from the simulated datasets, the INAR models had comparable MAE and RMSE, with NGINAR(1) having the slight advantage over the other two.

The ARIMA (3,0,5) model generated through the *auto.arima()* function generates negative forecasts which do not reflect the non-negative nature of the counts being modelled. The NGINAR(1) and ZINGINAR(1) models provide

considerably lower AICs as compared to the Poisson INAR(1) and ARIMA models. For the original dataset, the NGINAR(1) model provided better AICs than the ZINGINAR(1), different from the results from the simulations.

Using the same models and parameter estimates, it is also of interest to look into the forecasts during outbreak periods are also of interest. MAE and RMSE for both outbreak periods are summarized in Table 6. For these comparisons, the best ARIMA model provided the best forecasts during both outbreak periods based on both MAE and RMSE. However, this model still has the limitation of giving negative forecasts and thus, further comparison among the remaining models must be done.

Among INAR models, the NGINAR(1) is consistent in providing the best fit. In comparing the Poisson INAR(1) with the NGINAR(1) and ZINGINAR(1) models, it could be observed that the latter two have comparable MAE and RMSE, relatively lower than the former. These results show that the models considering overdispersion better forecast the peaks in the series.

Table 6. MAE and RMSE for the actual dataset on outbreak periods

	Poisson INAR(1)	NGINAR(1)	ZINGI-NAR(1)	AR(1)	Auto-ARIMA
MAE (outbreak 1)	2.1476	2.1001	2.0513	2.1463	2.0352
RMSE (outbreak 1)	2.8929	2.8424	2.8486	2.8280	2.6641
MAE (outbreak 2)	7.4790	7.0281	7.1631	7.4301	6.9706
RMSE (outbreak 2)	14.6789	13.8758	14.1408	12.9237	11.1752

7.3.2 Evaluation of outbreak detection thresholds

As observed from the results from the simulations, one general characteristic of the detection thresholds obtained from INAR models is that they generate more alarms than the ARIMA models. More alarms relate to higher sensitivity, as well as false positive rates. Among INAR models, the order would be Poisson INAR(1), NGINAR(1), and then ZINGINAR(1) in terms of number of alarms and sensitivity as seen in Table 7.

This trend can be explained by wider, more variable thresholds expected from the geometric (overdispersed) INAR models, which is shown in Figure 4. This finding is parallel to that of Cardinal et al. (1999), where they observed that forecast limits based on the empirical distribution of their meningococcal data were generally wider as that of the Poisson. This suggests an underestimation of the variance by the Poisson INAR model.

For a single dataset, POD1 and POD2 correspond to binary metrics indicating whether the outbreaks were successfully detected (✓) by the model. Both outbreaks were successfully detected for all the models being compared.

For measures of timeliness provided in Table 8, the advantage of INAR models is again clear since the ARIMA models were not able to detect any of the

two outbreaks during the first week while all INAR models successfully detected the second outbreak in the first week. Furthermore, ARIMA models exhibited an ATBD2 of 10 weeks or about 2.5 months which is not desirable, particularly since this outbreak is characterized by an abrupt increase and higher number of cases. Poisson INAR(1) provides a slight advantage over all other models in that it can detect the smaller first outbreak a week earlier.

Overall, for the original dataset, the ZINGINAR(1) model can be considered the optimal model among the five models compared while the NGINAR(1) model provided performance metrics almost similar with the former. This is explained first by the advantage of INAR models over ARIMA models in both accuracy and timeliness. The ZINGINAR(1) model also provided comparable detection accuracy and timeliness as compared with the Poisson INAR(1) and NGINAR(1) model with the least number of alarms generated, and thus, the lowest FPR.

Table 7. Outbreak detection accuracy measures for the Poisson INAR(1), NGINAR(1), ZINGINAR(1), AR(1) and Auto-ARIMA models

Measures	Poisson INAR(1)	NGINAR(1)	ZINGINAR(1)	AR(1)	Auto-ARIMA
alarms	37	20	14	6	5
Se	0.4074	0.2037	0.1667	0.1111	0.0926
Sp	0.9586	0.9751	0.9862	1	1
FPR	0.0413	0.0248	0.0138	0	0
POD1	√	√	√	√	√
POD2	√	√	√	√	√
PPV	0.5946	0.5500	0.6429	1.0000	1.0000
NPV	0.9156	0.8914	0.8881	0.8829	0.8808

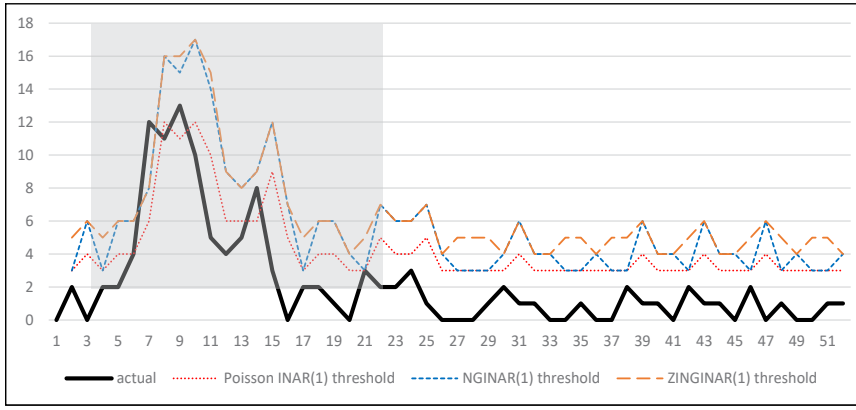

Figure 4. A portion of the real series (solid line) and the detection thresholds generated by the Poisson INAR(1) (dotted line), NGINAR(1) (short broken line), and the ZINGINAR(1) (long broken line) models. The first outbreak is enclosed within the shaded region.

Table 8. Outbreak detection timeliness measures for the Poisson INAR(1), NGINAR(1), ZINGINAR(1), AR(1) and Auto-ARIMA models from the actual dataset

	Poisson INAR(1)	NGINAR(1)	ZINGI-NAR(1)	AR(1)	Auto-ARI-MA
POD1wk1	0	0	0	0	0
POD1wk2	1	1	1	0	0
ATBD1 (wks)	3	4	4	4	4
ATBD2 (wks)	0	0	0	10	10

8. Conclusion

This study has demonstrated the applicability of INAR models, most notably, the newer NGINAR(1) and ZINGINAR(1) models for outbreak detection. The rationale of the study was to determine what improvements might these assumptions make in modelling and detecting outbreaks of measles. While the ZINGINAR(1) model has the added benefit of modelling excess zeroes as well, results show that the number of zeroes were not much in excess of what is expected from the NGINAR(1) process with the parameter hovering around 0.09 for simulations and the real dataset.

For the purpose of measles outbreak detection, the advantage of INAR models over ARIMA models is evident, mainly since the latter generally provide lower PODs. However, choosing an optimal model entails several factors. First of them is the outbreak detection capabilities of the model as measured through the performance metrics obtained in the results. High sensitivity is desired as well as low false positive rates in order to mitigate costs incurred in investigating false positives. Another one is the simplicity of the model – a model which effectively explains measles incidence and detects outbreaks with a smaller number of parameters is preferred over models with more parameters but provides similar or worse performance. A simpler model also means that its structure can be explained more easily to any decision-maker who would potentially use it. Lastly, the nature of the disease and the objectives of public health agencies is another factor which influences choosing the best model. Leong et al. (2015) suggests that diseases such as measles require systems which provide a balance between sensitivity and specificity as it is not a particularly deadly disease but still is of public health concern.

With these factors in mind, the NGINAR(1) model may be considered the best among the five models compared based from both the simulations and the original dataset as it has comparable outbreak detection capabilities to the Poisson INAR(1) while having lower false positive rates. It is also simpler than the ZINGINAR(1) model. However, this changes if measles elimination is of utmost importance, to the point that investigating any rise in measles incidence must be done regardless of the cost, a sensitive model might be more appropriate such as the Poisson INAR(1).

9. Recommendations

This study had shown the applicability and the performance of NGINAR(1) and ZINGINAR(1) for outbreak detection. These models are relatively new and thus, several directions for future research may be pursued. First, different specifications of the INAR model may be considered in comparisons. Also, the NGINAR(1) and ZINGINAR(1) models may be extended to accommodate higher-order autocorrelation structures. The development of multivariate NGINAR(1) and ZINGINAR(1) models which may be used in conjunction of supplementary data such as vaccine coverage and syndromic data may also be explored. Accommodating intervention terms analogous to ARIMA intervention models may also be of interest to consider the effect supplemental immunization activities (SIA) on measles incidence.

With the recent rise in measles incidence in the Philippines, examining the applicability of the models to the most recent figures will be relevant. Lastly, the results of the study are only applicable to Cavite or a setting with similar measles incidence patterns as Cavite. Thus, a more complex simulation framework or a larger collection of disease series may be considered to test performance of the models in detecting outbreaks for alternative settings and also, other diseases However, these recommendations may be limited by the non-availability of appropriate data.

Acknowledgement

The researchers would like to acknowledge the Provincial Epidemiology and Surveillance Unit of the Cavite province for providing the data used in this study.

REFERENCES

ABS-CBN NEWS. 2010, February 23, 6 more areas found to have measles outbreak – DOH. Retrieved from *ABS-CBN News*: https://news.abs-cbn.com/nation/02/22/10/6-more-areas-found-have-measles-outbreak-doh

ALLARD, R., 1998, Use of time series analysis in infectious disease surveillance, *Bulletin of the World Health Organization 76*(4), 327-333.

AL-OSH, M.A., & ALZAID, A.A. 1987, First-order integer-valued autoregressive (INAR(1)) process, *Journal of Time Series Analysis 8*(3). 261-275.

AWALE, M., RAMANATHAN, T.V., & KALE, M., 2017, Coherent forecasting in integer-valued AR(1) models with geometric marginals, *Journal of Data Science15*(1), 95-114.

BÉDUBOURG, G., & LE STRAT, Y., 2017, Evaluation and comparison of statistical methods for early temporal detection of outbreaks: A simulation-based study, *PloS one, 12*(7), e0181227.

BELLINI, W. J., & ROTA, P. A., 2011, Biological feasibility of measles eradication. *Virus research 162*(1-2), 72-79.

BYRD, R. H., LU, P., NOCEDAL, J., & ZHU, C., 1995, A limited memory algorithm for bound constrained optimization, *SIAM Journal on Scientific Computing*, *16*(5), 1190-1208.

CARDINAL, M., ROY, R., & LAMBERT, J., 1999, On the application of integer-valued time series models for the analysis of disease incidence. *Statistics in Medicine, 18*(15), 2025-2039.

CLEVELAND, W. S., & DEVLIN, S. J., 1988, Locally weighted regression: an approach to regression analysis by local fitting, *Journal of American statistical association, 83*(403), 596-610.

CUTTS, F. T., HENDERSON, R. H., CLEMENTS, C. J., CHEN, R. T., & PATRIARCA, P. A., 1991, Principles of measles control, *Bulletin of the World Health Organization*, *69*(1), 1.

DATO, V., SHEPHARD, R., and WAGNER, M. M., 2006, Outbreaks and Investigations. In Wagner, M. M., Moore, A. W., & Aryel, R. M., Eds.), *Handbook of Biosurveillance* (13-26). Burlington, MA: Elsevier Academic Press.

DEPARTMENT OF HEALTH, 2019a, *DOH Expands Measles Outbreak Declaration to Other Regions*, Retrieved from https://www.doh.gov.ph/node/16647.

DEPARTMENT OF HEALTH, 2019b, *DOH and other Government Agencies Working Together in Response to Measles Outbreak,* Retrieved from https://www.doh.gov.ph/press-release/DOH-and-other-Government-Agencies-Working-together-in-Response-to-Measles-Outbreak.

DEPARTMENT OF HEALTH - REGIONAL OFFICE IV-A, 2019a, *DOH CALABARZON starts measles vaccination in fast food chains in Laguna,* Retrieved from http://ro4a.doh.gov.ph/140-featured-article/421-doh-calabarzon-starts-measles-vaccination-in-fast-food-chains-in-laguna

DEPARTMENT OF HEALTH - REGIONAL OFFICE IV-A, 2019b, *DOH CALABARZON launches vaccination centers in Ayala Malls,* Retrieved from http://ro4a.doh.gov.ph/140-featured-article/446-doh-calabarzon-launches-vaccination-centers-in-ayala-malls

FREELAND, R. K., & MCCABE, B. P., 2004, Forecasting discrete valued low count time series, *International Journal of Forecasting, 20*(3), 427-434.

HYNDMAN, R., 2008, Automatic time series forecasting: The forecast package for R, *Journal of Statistical Software\v*, *27*(3), 1-22.

HYNDMAN, R., 2013, July 3, Facts and fallacies of the AIC [Blog post], Retrieved from https://robjhyndman.com/hyndsight/aic/

LEONG, R. N. F., CO, F. F., & TAN, D. S. Y., 2015, Some Zero Inflated Poisson-Based Combined Exponentially Weighted Moving Average Control Charts for Disease Surveillance, *The Philippine Statistician, 64*(2), 17-42

MANILA BULLETIN, 2014, January 31, *3 die in Cavite measles outbreak,* Retrieved from Yahoo! News*:* https://ph.news.yahoo.com/amphtml/3-die-cavite-measles-outbreak-205259594.html

MCKENZIE, E., 1985, Some simple models for discrete variate time series 1, *JAWRA Journal of the American Water Resources Association, 21*(4), 645-650.

NATIONAL EPIDEMIOLOGY CENTER, 2008, *Manual of Procedures for the Philippine Integrated Disease Surveillance and Response*, Philippines: National Epidemiology Center, Department of Health.

NATIONAL EPIDEMIOLOGY CENTER, 2014, *Manual of Procedures for the Philippine Integrated Disease Surveillance and Response,* Philippines: National Epidemiology Center, Department of Health.

PAMAN, J. M. J., SANTIAGO, F. N. M., MOJICA, V. J. C., CO, F. F., & LEONG, R. N. F., 2017, Measles Outbreak Detection in Metro Manila: Comparisons between ARIMA and INAR Models, *The Philippine Statistician, 66*(2), 71-91.

REIS, B. Y., & MANDL, K. D., 2003, Time series modeling for syndromic surveillance,*BMC Medical Informatics and Decision Making, 3*(1), 2.

RISTIĆ, M. M., BAKOUCH, H. S., & NASTIĆ, A. S., 2009, A new geometric first-order integer-valued autoregressive (NGINAR (1)) process, *Journal of Statistical Planning and Inference, 139*(7), 2218-2226.

RISTIĆ, M. M., BOURGUIGNON, M., & NASTIĆ, A. S., 2018, Zero-Inflated NGINAR (1) process, *Communications in Statistics-Theory and Methods*, 1-16.

ROLFHAMRE, P., & EKDAHL, K., 2006, An evaluation and comparison of three commonly used statistical models for automatic detection of outbreaks in epidemiological data of communicable diseases, *Epidemiology & Infection, 134*(4), 863-871.

SCHWEER, S., & WEISS, C. H., 2014, Compound Poisson INAR (1) processes: stochastic properties and testing for overdispersion, *Computational Statistics & Data Analysis, 77*, 267-284.

SERFLING, R. E., 1963, Methods for current statistical analysis of excess pneumonia-influenza deaths, *Public health reports, 78*(6), 494.

STEUTEL, F. W., & VAN HARN, K., 1979, Discrete analogues of self-decomposability and stability, *The Annals of Probability*, 893-899.

WATKINS, R. E., EAGLESON, S., HALL, R. G., DAILEY, L., & PLANT, A. J., 2006, Approaches to the evaluation of outbreak detection methods. *BMC public health*, *6*(1), 263.

WEISS, C. H., 2018, *An Introduction to Discrete-valued Time Series*. John Wiley & Sons.

WORLD HEALTH ORGANIZATION, 1999, *WHO Guidelines for Epidemic Preparedness and Response to Measles Outbreaks.* Geneva, Switzerland.

YLADE, M. C., 2018, Epidemiology of Measles in the Philippines, *Acta Medica Philippina*, *52*(4), 380.