Measles Outbreak Detection in Metro Manila: Comparisons between ARIMA and INAR Models

Joshua Mari J. Paman, Frank Niccolo M. Santiago, Vio Jianu C. Mojica, Frumencio F. Co, and Robert Neil F. Leong
Mathematics Department, De La Salle University

It is the goal of many developing countries to stop the spread of diseases. Part of this effort is to conduct ongoing surveillance of disease transmission to foresee future epidemics. However, in the Philippines, there is a lack of an automated method in determining their presence. This paper presents a comparison between an integer-valued autoregressive (INAR) model and the more commonly known autoregressive integrated moving average(ARIMA) models in detecting the presence of disease outbreaks. Daily measles reports spanning from January 1, 2010 to January 14, 2015 were obtained from the Department of Health and were used to motivate this study. Synthetic datasets were generated using a modified Serfling model. Similarity tests using a dynamic time warping algorithm were conducted to ensure that simulated datasets observe similar behavior with the original set. False positive rates, sensitivity rates, and delay in detection were then evaluated between the two models. The results gathered show that an INAR model performs favorably compared to an ARIMA model, posting higher sensitivity rates, similar lag times, and equivalent false positive rates for three-day signal events.

Keywords: measles, biosurveillance, integer-valued autoregressive model, Serfling model, dynamic time warping

1. Introduction

One of the major difficulties faced by developing countries is the ongoing battle against diseases, both of the communicable and noncommunicable forms. The World Health Organization (WHO) has identified communicable diseases as the major cause of death in many developing countries in Southeast Asia (Gupta and Guin, 2010). Further, most incidences of diseases such as measles and dengue occur in the said region, comprising 64% and 52% of the total worldwide cases, respectively.

In comparison, more developed countries have been able to significantly reduce, if not entirely eradicate, the impact of some diseases. For example, the world has not seen another case of smallpox since it was successfully eradicated in 1980 (WHO, n.d.). The Americas, Europe, and even some parts of Asia have already eliminated the spread of malaria through collaborative efforts between developed nations (Tanner and de Savigny, 2008). At present, the world has turned its attention to eradicating other contagious and infectious diseases, such as measles. WHO considers this particular disease as the leading cause of death among young children in the world, and despite the presence and easy access to vaccines, an estimated 16 deaths occur every hour globally because of the disease (WHO, 2015). As a means of combatting this threat, the United Nations (UN) has made it its goal to end epidemics of several communicable diseases as part of its objective to ensure healthy lives and promote well-being for individuals at all ages. This is in pursuance of the 2030 Agenda for Sustainable Development (UN, 2015).

An outbreak, as defined by the Department of Health (DOH) is indicated by a sudden increase of disease occurrence in a local area (DOH, 2008). Currently, DOH has yet to develop a quantitative method in determining if a disease outbreak has already occurred in a given area, and are limited to arbitrarily attributing an outbreak based on the occurrence of consecutive high daily counts. Specifically, the Philippine Integrated Disease Surveillance and Response (PIDSR) System has thresholds for an initial alert and an indication of an epidemic based on the average weekly or monthly cases of a particular disease in the past three to five years. For an alert threshold, one standard deviation is added to the average, while two standard deviations are added for an epidemic threshold. In the case of measles, a suggested threshold of one suspect case—a case that is yet to be laboratory-confirmed—in the past week is enough to merit an alert, while a confirmed outbreak from DOH is required to consider it an epidemic. However, the current methods may be inefficient, which may limit their ability to foresee unwanted outbreaks and take precautionary measures ahead of time. Because of measles' high occurrence as a communicable disease, as well as the large attention given by both local and international health agencies in its eradication, it was chosen to be the data used in this study.

A sudden emergence of studies on disease surveillance methods developed in recent years, seeing the growing interest for accurate and early detection of outbreaks in the area of public health. According to a review by Buckeridge (2006) on recent studies tackling outbreak detection, most researches employ the use of temporal detection algorithms as opposed to spatial detection algorithms, and were designed to give results available daily as opposed to weekly and monthly reports. Further, the study also revealed a more frequent use of regression modelling, which may or may not include the use of time series approaches such as an

autoregressive integrated moving average (ARIMA) model and statistical process control methods such as exponentially weighted moving average (EWMA) and cumulative sum (CUSUM).

An ARIMA model is one of the more commonly used time series models due to its simplicity, and has also been extensively used for studies on detecting disease outbreaks. Its preference can be attributed to the fact that it can adjust to localized trends inherently present in the data (Reis and Mandl, 2003). However, the model's tendency to adjust to these trends also makes it prone to adjust to long periods of outbreaks and, as a consequence, fail in effectively detecting them. Further, an ARIMA model has an underlying assumption of normality which, in consequence, enforces the distribution of the cases to be symmetric about the mean. This will be problematic for rare diseases whose distribution means are skewed to the right, or for commonly-occurring diseases whose means are skewed to the left (Cardinal et al., 1999).

In these aforementioned instances, a model suited for a continuous variable is used; intuitively, a model designed for integer-valued observations is more suitable for daily counts of reported cases. Unkel et al. (2011) furthers this claim and posits that these models may be more effective in the asymmetric distribution scenarios previously described; thus, suggests the use of an integer-valued autoregressive model (INAR) in outbreak detection. In a study on meningococcal infection in the Montreal-Centre region, significant INAR (5) and AR (5) models were found to fit the data, but it was shown that the INAR model accounted for the asymmetric distribution of the cases better, provided a relatively smaller forecast error, and gave forecast limits that were both non-negative and integer-valued (Cardinal et al., 1999).

Based on these observations, this research is conducted with the use of the aforementioned time series models as possible alternatives and supplements to the current methods employed in disease surveillance in the hope of foreseeing potential disease outbreaks and carrying out preventive measures in the soonest possible time. For this study, three objectives were considered: (1) to discuss and illustrate the use of an INAR model in outbreak detection; (2) to determine whether an INAR time series model would perform significantly better in disease outbreak detection than a model based on ARIMA; and (3) to compare the results obtained from the simulated datasets and the original dataset.

2. Theoretical Framework

2.1. Serfling method

The Serfling method was a model originally conceived as a method of detecting influenza outbreaks (Serfling, 1963). The cyclic regression model assumes that the data is sinusoidal and having a period of one year, and is given by

$$y_t = \alpha + \beta t + \varphi sin\left(\frac{2\pi(t-\delta)}{365.25}\right) + \epsilon_t \quad (1)$$

where α is a constant term, β is the trend, t is the time period, φ is the amplitude, δ is a constant that aligns the sine function to the highest point in the set, and ϵ_t is a random error term (Moore et al., 2003).

2.2. Dynamic time warping

Dynamic Time Warping (DTW) is a commonly used technique in determining the optimal match between two time-dependent sequences under certain restrictions (Müller, 2007). This is done by 'warping' the two sequences in a non-linear manner and matching each individual observation of one sequence to another sequence by minimizing a predefined cost function.

Suppose there were two sequences $A = (a_1, a_2, \ldots, a_I)$ and $B = (b_1, b_2, \ldots, b_J)$ defined in a feature space F, where $a_i, b_j, \in F$. A local cost measure or local distance measure $c(a, b)$ is applied, where a smaller value of the cost measure indicates a strong similarity between a and b, and larger value indicating otherwise. A warping path is then formulated, and is defined as a sequence $w = (w_1, \ldots, w_N)$, where $w_n = c(a_n, b_n)$. The resulting warping path must satisfy two conditions: (1) the first and last observations of each sequence must be aligned; and (2) none of the observations are omitted in the warping process, i.e., all elements in one sequence are aligned to at least one element in the other sequence.

The total cost of the warping path p in terms of the previously defined local cost measure c can be expressed as

$$c_p(A,B) = \sum_{n=1}^{N} c\left(a_{in}, b_{jn}\right) \quad (2)$$

It is important to note that warping paths are not unique, and have different total costs. Of these, the warping path with the minimum total cost among all possible paths is called the optimal warping path between A and B, p_{opt}.

Constraints can also be placed when using DTW in order to better align the optimal warping path. One of these is the use of a Sakoe-Chiba band (Sakoe and Chiba, 1978), which serves as a window that restricts the warping path to an almost diagonal form. The band width used for most studies is 5% of the total observations.

2.3. Local regression

Local regression (LOESS) is a form of non-parametric regression that attempts to fit groups of data points locally by way of a multivariate smoothing

procedure (Cleveland and Devlin, 1988). Unlike other regression methods, LOESS does not require a global function to fit all the observations, and instead makes estimates within segments of the data. It also serves as an alternative when there is a complex deterministic trend present in the dataset which cannot be interpreted by linear regression.

Let y_i be measurements of the dependent variable, and x_{ij} measurements of the independent variables, where $i = 1, 2,\ldots, n$ and $j = 1, 2,\ldots, p$. The model is written in the form

$$y_i = g(x_{ij}) + \varepsilon_i, \tag{3}$$

where $g(x_{ij})$ is a regression function and ε_i is a random error term with mean 0 and variance σ^2. The aforementioned function is locally approximated from a parametric function.

2.4. Autoregressive integrated moving average model

ARIMA is a generalized form of the Autoregressive Moving Average (ARMA) model, but with the added consideration of a differencing factor. It is formulated as

$$\Phi_P(B^s)\ \varphi_P(B)(1- B^s)^D\,(1- B)^d\,Z_t = \theta_0 + \Theta_Q(B^s)\ \theta_q(B)a_t, \tag{4}$$

where $\varphi_P(B)$ is the nonseasonal AR operator, $(1- B)^d$ is the ordinary differencing factor, d is the order of ordinary differencing, θ_0 is the mean of the process, $\theta_q(B)$ is the nonseasonal MA operator, $\Phi_P(B^s)$ is the seasonal AR operator, S denotes the seasonal period, $(1- B^s)^D$ is the seasonal differencing factor, D is the order of seasonal differencing, $\Theta_Q(B^s)$ is the seasonal MA operator, and a_t is the normally-distributed random shock with mean 0 and constant variance s^2. This model is often denoted by

$$ARIMA\ (p,d,q)\ \text{x}\ (P,D,Q)s,1, \tag{5}$$

where 1 (or 0) indicates a nonzero (or zero) mean (Wei, 2005).

ARIMA fitting can be done using an automated ARIMA modelling procedure (Hyndman and Khandakar, 2008). For the nonseasonal differencing parameter, a Unit Root Test for stationarity, namely the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test was used. The KPSS test differentiates itself from other stationarity tests such as the Dickey-Fuller tests in that it has a null hypothesis that the series is stationary around a deterministic trend (Kwiatkowski et al., 1992).

For the AR and MA components, the Akaike Information Criterion (AIC) was computed for possible p, q, P and Q values in a stepwise manner, and the combination with the lowest AIC was chosen. Once found, another test was conducted to determine whether or not the constant term should be dropped. It

was then tested for the best fit ARIMA model with weekly seasonal differencing and compared whether it is better than the ARIMA model without weekly seasonal differencing.

2.5. Integer-valued autoregressive model

An INAR model is appropriate when considering time series in the form of count data, and can be considered as the discrete counterpart of the AR process. Unlike the ARIMA model, an INAR model specifies the properties inherent in an integer-valued time series. Specifically, the range of values of the ARIMA parameters assume that the observations can take on any real value, i.e. it can generate forecasts that are negative, when in reality disease observations are restricted to only nonnegative integers. Furthermore, an ARIMA model assumes that the data follows a normal distribution. However, depending on the disease in question, the normality assumption or by extension an assumption of symmetry in the distribution of the data may not always be satisfied. If the disease is rare, the distribution would be skewed to the right as the mean is closer to zero. Conversely, if the disease occurs frequently, the distribution would be skewed to the left. INAR, on the other hand, does not impose any restrictions regarding skewness on data and, instead, follows the distribution of its error term.

The INAR model of order p is defined as

$$x_t = \sum_{i=1}^{p} \alpha_i o x_{t-i} + \epsilon_t, \tag{6}$$

where x_t is a sequence of nonnegative integer-valued variables, ϵ_t is a sequence of iid nonnegative integer-valued random variables with finite mean μ_ϵ and variance σ_ϵ^2 and α_i is a sequence of parameters such that $\alpha_1, \alpha_2,..., \alpha_{p-1} \in [0,1]$ and $\alpha_p \in [0,1]$, where $i = 1, 2,\ldots, p$ (Cardinal et al., 1999). An INAR model of order p is interpreted in the same manner as an AR(p) model. Unlike the random shock variable of an ARIMA model, ϵ_t must have a non-zero mean. This prevents the variable from taking on zero values and negates the need for a constant term in the model. In most cases of INAR, the random shock is assumed to follow a Poisson distribution with parameter l (Bu et al., 2008).

Meanwhile, the convolution operator (Steuteland van Harn, 1979) $\alpha_i o x_{t-i}$, also known as a binomial thinning operator, can be expressed as

$$\alpha \text{ o } X = \sum_{k=1}^{X} y_k \tag{7}$$

based on iid Bernoulli random variables with probability $P(y_k=1)=\alpha$ and X is defined as a non-negative discrete random variable (Unkel et al., 2012). Consequently, this means that $\alpha o X$ follows a binomial distribution, where X is the number of trials and α is the probability of success. It is assumed that the

convolution terms are mutually independent. Moreover, these convolution terms are also independent of the random shock term.

For an INAR process to be stationary, it must satisfy the following condition:

$$\sum_{i=1}^{p} \alpha_i < 1. \tag{8}$$

This leads to some interesting properties comparable to an AR process. Firstly, the autocorrelations of a stationary INAR process would follow the recursive equation that governs AR, i.e.,

$$\rho k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \ldots + \phi_p \rho_{k-p}, \; k > 0, \tag{9}$$

which indicates that the ACF of INAR also decays exponentially (Cardinal et al, 1999). Secondly, for lags greater than p, the PACF would take on zero values. Lastly, the expression

$$E(X_t) = \frac{\mu_\varepsilon}{1 - \sum_{i=1}^{p} \alpha_i} \tag{10}$$

holds true under a weak stationary assumption regardless of the distribution that the random shocks take on (Alzaid and Al-Osh, 1990).

Estimation methods have been developed to obtain estimates of the model parameters α and λ. One of which is the Conditional Least Squares (CLS) method, which is shown to produce strongly consistent and asymptotically normally distributed estimators (Bu et al., 2006; Cardinal et al., 1999). Furthermore, the CLS procedure is also computationally efficient to use. The procedure minimizes the equation

$$Q(\theta) = \sum_{t=p+1}^{T} \left[X_t - \alpha_1 X_{t-1} - \alpha_2 X_{t-2} - \ldots - \alpha_p X_{t-p} - \lambda \right]^2 \tag{11}$$

with respect to λ, α_1, ..., α_p,, which can then be used to solve the estimates by setting

$$\frac{\partial Q(\theta)}{\partial \theta} = 0. \tag{12}$$

2.7. Measles

DOH defines measles as an *acute highly communicable viral illness caused by the measles virus in the genus Morbillivirus of the family Paramyxovirus*. It is characterized by a prodrome of fever, conjunctivitis, cough, coryza, and small spots with white or bluish white centers on an erythematous base on the buccal mucosa known as Koplik spots followed by maculopapular rash on the third to

the seventh day beginning on the face then becoming generalized. The disease is transmitted by direct contact with nasal or throat secretion from infected individuals. Incubation period for measles ranges from one to three weeks from exposure to onset of fever and usually two weeks until rashes begin to appear.

A person is suspected to have measles if he/she exhibits three symptoms: (1) he/she has a fever of 38°C or more; (2) he/she has had a rash for the past three days or more; and (3) currently has a cough, a stuffy nose (coryza), or redness in the eyes (conjunctivitis). A suspected case is then classified to either a (1) laboratory confirmed case, (2) epidemiologically-linked case; or a (3) discarded case (DOH, 2008).

3. Methodology

3.1. Data

For this study, daily counts of measles cases in the National Capital Region (NCR) provided by the regional office of DOH were used. The dataset spans from January 1, 2010 to January 14, 2015. The daily counts of cases are combinations of both laboratory confirmed cases and epidemiologically-linked cases.

In determining an outbreak, an onset is marked when five consecutive days of at least 10 reported cases occur. The ending is then marked when there are five consecutive days of less than 10 reported cases. Using these definitions, the first and second outbreaks (denoted as outbreak 1 and outbreak 2, respectively) were identified to have occurred during the periods December 3, 2010 up to June 9, 2011 and August 1, 2013 up to May 20, 2014, respectively.

To ensure that the true beginning and true ending of an outbreak is captured, a padding of 90 days was added to the beginning and a padding of 45 days was added to the end. This is done because the true beginning and true ending of an outbreak is not known and an estimate is instead made.

Furthermore, non-outbreak data was examined for possible overdispersion. Overdispersion is the presence of greater variability within a data set as compared to the expected variance from an assumed distribution. The index of dispersion by Fisher (1950) was used; it is defined as the ratio of the variance and the mean, which is given by:

$$I_d = \frac{\sigma^2}{\mu}, \qquad \hat{I}_d = \frac{S^2}{\bar{X}}, \tag{13}$$

Non-outbreak data exhibited $\hat{I}_d$ = 1.5218. This suggests possible overdispersion which is expected in most count data. There are various models and ongoing works addressing overdispersion in modeling count data, however, these are not within the scope of the study.

3.2. Analysis

3.2.1. Simulation

The dataset is first cleaned and removed of outbreaks. This is to make sure that the simulated data follows the behavior of the background disease process free of outbreaks, i.e., the behavior on a normal period. For non-outbreak simulations, a model used for simulating data based on the Serfling method was formulated. The resulting modified Serfling model is given by

$$\hat{Y}_t = \text{round}(\beta_0 + \tau_i + \beta_1 t + \sum_{i=0}^{364} \varphi_i sin^{+}\left(\frac{2\pi(t-i)}{365}\right) + \varepsilon_t), \tag{14}$$

where β_0 is a constant, β_1 is the trend, φ_i is the amplitude, t is the day, τ_i is the day-of-week effect where $i = mod\ (t, 7) + 1$, $sin^{+}(x)$ is the sine function that only takes on nonnegative values, ε_t is the error term that follows an empirical model centered at t. The sum of the terms is then rounded off to the nearest integer. The addition of a day-of-week effect term in the model is justified by Fricker (2013) as it accounts for the weekly trend in the data.

Simulated datasets were generated from (14) and DTW was applied to assess the similarity of the simulated data with the actual data. Because outbreaks were removed from the original data, the simulated datasets were divided in three parts and each subsequence was correspondingly compared with the original dataset using DTW. As a threshold, at least 90% of the observations in the simulated dataset must fall within the Sakoe-Chiba band. The simulation and assessment procedure is repeated until 250 similar datasets have been obtained.

For outbreak simulations, the computed $\hat{Y}_t$ from the non-outbreak simulation were subtracted from the outbreak values in the original dataset. The resulting residuals were then fit in a LOESS line, and then used to fill the gaps between the non-outbreak simulated data.

3.2.2. Models

ARIMA fitting was carried out using the aforementioned automated procedure formulated by Hyndman and Khandakar (2008). INAR fitting, on the other hand, was carried out using a recursive significance test on the α parameters. Simply put, p is increased starting from INAR(1) until such time that α_p is no longer significant, at which case INAR($p-1$) is taken to be the best fit model. For this study, however, model fitting was limited to INAR (3) due to parsimony and computational complexity. CLS estimation was used to compute for the parameter values.

After the best fit model has been obtained for both ARIMA and INAR, the resulting models were then compared on how well they could detect disease

outbreaks in the data. Sensitivity and specificity were measured for each model relative to the aforementioned padding measures. Sensitivity is a measure of the proportion of periods correctly identified by the model as those wherein outbreaks are indeed present. On the other hand, specificity is a measure of periods correctly identified by the model as non-outbreak periods. This was done in order to measure the model's ability to identify an outbreak from a non-outbreak period. It is also of importance to determine between the models which signals first or the relative delay that the models detect the presence of an outbreak. Relative delay was measured as the number of days by which the model detects outbreaks earlier or later as compared to the actual occurrence of the identified start of the outbreak. In all criteria, both models were tested for signal events of one day of high counts, two consecutive days of high counts, and three consecutive days of high counts. Comparisons between models will be done using paired t-tests. All statistical procedures and macros to implement INAR fitting were carried out using SAS 9.3 and all subsequent analyses were done at a 0.01 significance level. SAS macros written for this study are available upon request.

4. Results and Discussion

4.1. Data simulation

There were three outbreaks in the dataset: in early 2010, in early 2011, and in late 2013. This is confirmed by the study conducted by Angkico et al. (2014) who also detected outbreaks in those periods of time. Figure 1 presents a plot of the original data. For the study, however, only the latter two outbreaks were simulated since the dataset does not cover when the first outbreak began. The two outbreaks considered possess interesting characteristics: one exhibits a slow-rising behavior, while the other exhibits a sudden spike in reported cases.

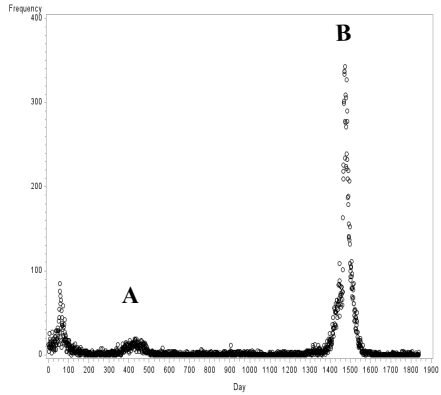


Figure 1. Graph of Measles Reports from January 1, 2010 to January 14, 2015; Labeled are Outbreaks 1 (A) and 2 (B)

The average number of daily cases on a non-outbreak period is only 1.13. During the first identified outbreak period, a mean number of 7.35 measles cases were reported daily, while during the second identified outbreak period, a mean number of 50.54 measles cases were reported daily.

After the outbreaks were determined, the data was cleaned of the outbreak periods. A modified Serfling model was fitted, which was then used to simulate non-outbreak datasets. Parameter estimates of the constant term, trend coefficient, and amplitude for selected days of the fitted modified Serfling model are presented in Table 1. The full table of parameter estimates are available upon request.

Table 1. Partial List of Modified Serfling Model Parameters

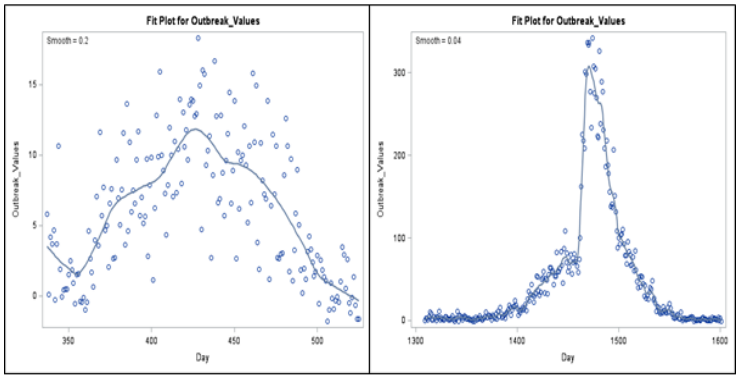
Parameter	Estimate
$\hat{\beta}_0$	4.0309
$\hat{\beta}_1$	-0.00058
$\hat{\varphi}$_236	0.0876
$\hat{\varphi}$_237	54.8588
$\hat{\varphi}$_238	-135.7234
$\hat{\varphi}$_239	91.1728
$\hat{\varphi}$_240	0

Table 2 provides the day-of-week effect values computed from the simulations done using the modified Serfling model. A visible trend can be inferred: reports rise on Mondays, then gradually decrease over the week and then spike again on Friday before decreasing considerably over the weekend.

Table 2. Estimated day-of-week effects

Day	DOW effect
Monday $(\hat{\tau}_1)$	0.1710
Tuesday $(\hat{\tau}_2)$	0.1271
Wednesday $(\hat{\tau}_3)$	0.0748
Thursday $(\hat{\tau}_4)$	0.0497
Friday $(\hat{\tau}_5)$	0.0983
Saturday $(\hat{\tau}_6)$	-0.2150
Sunday $(\hat{\tau}_7)$	-0.3087

Figure 2 shows a sample LOESS line fitting of the simulated outbreaks for the two types of outbreaks. It is seen that they behave similar to the trends observed from the outbreaks in the original data. It is noted that log transformation of the residuals for outbreak 2 was needed to obtain a proper fit in the LOESS line.



A **B**

Figure 2. Sample LOESS lines of simulated outbreak 1 (A) and 2 (B)

4.2. Performance evaluation with simulated data

4.2.1. Model Fitting

Of the 250 datasets that met the DTW similarity requirement, only 159 satisfied the white noise assumption of the ARIMA model. Models which failed to meet this assumption were dropped from further analysis as they are inadequate for further testing. Table 3 shows a summary of the fitted ARIMA models and whether or not they satisfied the white noise assumption. Again, full table is available upon request.

Table 3. Partial list of ARIMA models tested for white noise assumption

ARIMA Model	White Noise Test		
	Passed	Failed	Total
ARIMA (0,1,1) x $(0,0,0)_7$,0	46	6	52
ARIMA (0,1,1) x $(1,0,1)_7$,0	41	6	47
ARIMA (1,1,1) x $(0,0,0)_7$,0	14	0	14
…	…	…	…
ARIMA (4,1,4) x $(1,0,1)_7$,1	0	1	1
Total	159	91	250

Table 4 presents the frequencies of INAR models that the procedures have generated. From there, it is seen that an *INAR*(3) is the most commonly fitted model, accounting for more than half of the datasets, while an ARIMA (0,1,1) x $(0,0,0)_7$,0 was the best for continuous models as seen in the previous table. The former can be attributed to the fact that the fitting procedure was capped off at p=3, which may indicate that the datasets would fit higher order INAR models. However, higher order processes may no longer be useful as excess coefficients would prove too tedious to estimate and would only further complicate the model.

Table 4. Frequency of INAR models fitted

INAR Model	Frequency	Percent
INAR (3)	171	68.40
INAR (2)	59	23.60
INAR (1)	20	8.00

4.2.2. False positives

In Table 5, it is seen that INAR presented a higher false positive count compared to ARIMA. Approximately, INAR made one day signals an average of 82 times during the non-outbreak period, while ARIMA signaled an average of 48 times. Higher counts were also observed for the two-day and three-day signal events under INAR. Consequently, this meant that INAR would be expected to have a higher false positive rate than ARIMA; this is also evident in Table 5.

Table 5. False Positive Counts and Rates

Model and Signal Events	Mean Count	Mean Rate
INAR one-day signal	82.1572	0.0447
ARIMA one-day signal	47.6289	0.0259
INAR two-day signal	3.1006	0.0017
ARIMA two-day signal	1.4465	0.0008
INAR three-day signal	0.1195	0.00007
ARIMA three-day signal	0.0818	0.00004

By way of a paired *t*-test presented in Table 6, INAR false positive alarms were significantly higher than ARIMA, with the exception of a three-day signal event wherein the difference was not significant. These empirical results suggest that for a three-day signal event, the performance of ARIMA and INAR are similar in terms of effective false positive rate.

Table 6. Paired t-test of Differences in False Positive Counts and Rates

Comparison	Difference	DF	t Value
One-day signal false positive count	34.5283	158	51.83*
Two-day signal false positive count	1.6541	158	11.24*
Three-day signal false positive count	0.0377	158	1.06
One-day signal false positive rate	0.0188	158	51.95*
Two-day signal false positive rate	0.0009	158	11.25*
Three-day signal false positive rate	0.00002	158	1.06

**significant at the 0.01 level*

4.2.3. Relative Sensitivity

In terms of outbreak detection, Table 7 shows that INAR is relatively more sensitive than ARIMA on all signal events for both the first and second outbreaks. Succeeding paired t-tests, as presented in Table 8, indicate that their differences are indeed significant for all accounts.

Table 7. Outbreak Signal Event Mean Counts and Sensitivity Rate

Model and Outbreak Signal Event	Mean Count	Mean Sensitivity
INAR one-day signal (outbreak 1)	109.1132	0.5773
ARIMA one-day signal (outbreak 1)	50.7233	0.2684
INAR two-day signal (outbreak 1)	72.0063	0.3810
ARIMA two-day signal (outbreak 1)	20.9560	0.1109
INAR three-day signal (outbreak 1)	51.5723	0.2729
ARIMA three-day signal (outbreak 1)	10.0440	0.0531
INAR one-day signal (outbreak 2)	204.6855	0.6986
ARIMA one-day signal (outbreak 2)	116.0755	0.3962
INAR two-day signal (outbreak 2)	174.7484	0.5964
ARIMA two-day signal (outbreak 2)	95.2767	0.3252
INAR three-day signal (outbreak 2)	162.7799	0.5556
ARIMA three-day signal (outbreak 2)	86.2390	0.2943

Table 8. Paired t-tests on Sensitivity Rates

Comparison	Difference	DF	t Value
One-day signal (outbreak 1)	0.3089	158	54.61*
Two-day signal (outbreak 1)	0.2701	158	40.94*
Three-day signal (outbreak1)	0.2197	158	32.68*
One-day signal (outbreak 2)	0.3024	158	83.14*
Two-day signal (outbreak 2)	0.2712	158	62.84*
Three-day signal (outbreak 2)	0.2612	158	54.31*

**significant at the 0.01 level*

4.2.4. Relative delay

In terms of relative delay, it was found that the two models do not differ significantly, with the exception of one-day signal event for the second outbreak. As can be seen in Table 9, INAR detects the outbreak approximately 0.3711 days earlier than ARIMA. It is also important to note that for the three-day signal event in the second outbreak, eight ARIMA models were unable to detect the outbreak.

Table 9. Paired t-test on Delay Differences between Models

Comparison	Difference	DF	t Value
One-day signal delay (outbreak 1)	-0.1132	158	-1.96
Two-day signal delay (outbreak 1)	0.1006	158	0.15
Three-day signal delay (outbreak 1)	1.4901	150	1.02
One-day signal delay (outbreak 2)	-0.3711	158	-3.23*
Two-day signal delay (outbreak 2)	-0.8428	158	-0.94
Three-day signal delay (outbreak 2)	-4.3333	158	-2.00

**significant at the 0.01 level*

4.3. Performance Evaluation with Original Data

4.3.1. Model fitting

Based on model fitting, it was found that the original dataset fit an ARIMA $(0,1,1) \times (0,0,0)_7,0$ and an *INAR* (3). Recalling the results from the simulated data, it can be seen that similar results were obtained as most datasets were inclined to fit in models such as those in the original one. The parameter estimate for the best fit ARIMA model is $\hat{\theta}_1 = 0.9093$, while the parameter estimates for the best fit INAR model are $\hat{\alpha}_1 = 0.1363$, $\hat{\alpha}_2 = 0.1335$, $\hat{\alpha}_3 = 0.1321$, with an estimated random shock mean $\hat{\lambda} = 0.6728$.

4.3.2.Model performance on signaling events

The INAR model fitted in the original dataset registered higher false positives than the ARIMA model fitted for both one-day signal events and two-day signal events. As seen in Table 10, INAR signaled 70 times in one-day signal events, more than twice the 30 alerts made by the ARIMA model. Similar with the results from the synthetic data, INAR and ARIMA performed similarly for three-day signal events, not triggering alarms during non-outbreak periods. Evaluating their relative sensitivities in detecting high-count days, INAR detected a larger portion of the outbreaks compared to the ARIMA model for all signal events and in both outbreaks, as shown in Table 11. Meanwhile with respect to relative delay, results show that INAR reported earlier than ARIMA for all signal events, except for one-day signal event during the first outbreak and three-day signal event during the second outbreak. As shown in Table 12, it takes ARIMA almost three weeks later to detect the first outbreak compared to INAR for three-day signal events. Differences in the second outbreak, however, are small.

Table 10. False Positive Counts and Rates for the Original Dataset

Model and Signal Events	Count	Rate
INAR one-day signal	70	0.0587
ARIMA one-day signal	30	0.0178
INAR two-day signal	3	0.0025
ARIMA two-day signal	1	0.00059
INAR three-day signal	0	0
ARIMA three-day signal	0	0

Table 11.Outbreak signal event counts and sensitivity rate for the original dataset

Model and Outbreak Signal Event	Count	Sensitivity
INAR one-day signal (outbreak 1)	78	0.4127
ARIMA one-day signal (outbreak 1)	41	0.2169
INAR two-day signal (outbreak 1)	37	0.1958
ARIMA two-day signal (outbreak 1)	8	0.0423
INAR three-day signal (outbreak 1)	15	0.0794
ARIMA three-day signal (outbreak 1)	1	0.0053
INAR one-day signal (outbreak 2)	183	0.6246
ARIMA one-day signal (outbreak 2)	87	0.2969
INAR two-day signal (outbreak 2)	156	0.5324
ARIMA two-day signal (outbreak 2)	60	0.2048
INAR three-day signal (outbreak 2)	143	0.4881
ARIMA three-day signal (outbreak 2)	44	0.1502

Table 12. Signal Event Delay Differences for the Original Dataset

Signal Event Delay	Difference	Signal Event Delay	Difference
One-day signal delay (outbreak 1)	0	One-day signal delay (outbreak 2)	-3
Two-day signal delay (outbreak 1)	-9	Two-day signal delay (outbreak 2)	-3
Three-day signal delay (outbreak 1)	-20	Three-day signal delay (outbreak 2)	0

4.4. Discussion

4.4.1. Evaluation of simulation procedure

The results from the simulation procedure presented interesting trends in the data. The values fitted for the day-of-week effect matches the description provided by Fricker (2013). As expected, more reports are made on weekdays than on weekends. However, the pattern can be extended to real world incidence reporting. Reports rise on Mondays because it occurs right after a weekend and on Fridays because it is before the start of the weekend. On weekends, it is expected that less staff would be on-hand to receive and respond to these reports since it is also expected that less people would arrive during those days. The presence of this seasonal trend may suggest that the simulated data behaves similarly to the original dataset, furthering the efficiency of the simulation procedures used. Further, past studies using INAR, specifically Cardinal et al. (1999), claimed that a seasonal trend was not present based on differencing; however, it may be possible that a weekly trend existed if the day-of-week effect was taken into account. In terms of model fitting, the most commonly fitted INAR and ARIMA models for the simulated datasets matched the best fit INAR and ARIMA models for the original dataset. This could indicate their strong similarity, which makes the use of the algorithm more attractive as a simulation procedure.

4.4.2. Comparison using simulated datasets

Model fitting was also found to be more difficult with ARIMA since 91 of the models failed to meet the white noise assumption. This may indicate that an ARIMA model is not a viable choice as there is the likelihood that no adequate model can be fitted to the disease data.

In terms of the three criteria, INAR had more false positives than ARIMA, except for three-day signal events. On the other hand, INAR was more sensitive than ARIMA in all signal events and both outbreaks. These results present an interesting trend when compared with each other. One would assume that the higher sensitivity of INAR made it more prone to trigger during a non-outbreak, but as was presented in the previous section, their differences in false positive rates was not significant for three-day signals despite INAR's substantial gain in

sensitivity. This may suggest, at least based on the empirical evidence gathered, that INAR's higher sensitivity does not solely depend on its increased false positive rate. Further, this also suggests that when detecting outbreaks using three-day signals, INAR performs better than ARIMA at a comparable error rate. The relative delay, meanwhile, shows that INAR and ARIMA trigger at the same time, except for one-day signal events in the first outbreak where INAR triggers approximately a third of the day earlier. While this may not be statistically large, it may be advantageous in the practical setting, particularly for highly virulent diseases where alerts made hours earlier can make a crucial difference.

Further, the results showed that eight ARIMA models were unable to detect the first outbreak. This interesting result is best explained by what Reis and Mandl (2003) mentioned when discussing ARIMA's limitations. Because ARIMA can adjust to localized trends in the data, it was unable to identify the outbreak that was occurring over a long period of time. This is also supported by the fact that the first outbreak is known to be a gradually rising outbreak that does not necessarily spike like its other counterpart, making it more difficult to detect. In summary, these results suggest that INAR and ARIMA would detect the outbreaks approximately at the same time, provided that the latter detects the outbreak. This also furthers the notion that INAR can detect sustained shifts; ARIMA, on the other hand, cannot.

4.4.3. Comparison using the original dataset

The advantage of INAR can also be seen when implemented with the original daily counts. As Cardinal et al. (1999) highlighted, INAR is advantageous compared to ARIMA in generating limits because they are restricted to integer values. ARIMA, on the other hand, can select any real value when selecting limits which may prove difficult when discerning outbreaks. For instance, this can be the case when the upper limit is only a few tenths or hundredths above or below an observation, and rounding measures may give conflicting results.

The same trends as the ones discussed in the previous section were found in terms of false positive and relative sensitivity rates of the two models. INAR still had a higher false positive rate than ARIMA, except for three-day signal events. Also, INAR had a higher sensitivity rate than ARIMA for all instances. In comparison with the results presented in Table 10, ARIMA's three-day signal events performed worse compared to INAR's. Relative delay was also found to be relatively not that large between the models. Further, the results summarized in Table 11 still align with the notion Reis and Mandl (2003) suggested. In detecting the first outbreak, ARIMA and INAR alert at the same time for one-day signal events, but drastically fall behind the two-day and three-day signal events, triggering 9 days and 20 days later, respectively. On the other hand, the small difference in delay for the second outbreak is explained by the rapid increase in reported cases during that time; thus, making it easier to detect.

5. Conclusion

INAR showed promise as a candidate model for disease outbreak detection. Its nature of being an integer-valued process and its independence from a normality assumption made it a better alternative than the more commonly-used ARIMA models particularly for modeling disease counts. In theory, INAR would be the better choice, and it was shown in this paper that this advantage holds true in application as well.

The results of this study confirmed the assumption that INAR model performs better in detecting outbreaks than ARIMA. Beyond the theoretical benefits of using INAR, it was also shown empirically that it performed equally or even better in practice. From the test of false positive rates, it was shown that INAR had a significantly higher false positive rate compared to ARIMA based on one-day and two-day signal events. However, it was also found that INAR and ARIMA performed roughly the same when testing for three-day signals. On the other hand, it was determined that INAR was more sensitive than ARIMA in detecting outbreaks, having a significantly higher sensitivity rate than the latter. Further, between these processes, it was computed that on the average, INAR detected outbreaks around the same time as ARIMA, provided that the ARIMA model detects their presence. The proposed simulation model and procedures were also found to be promising in terms of producing simulated data comparable to the original dataset. This is evidenced by some results generated from appropriate tests and the similarities in preferred ARIMA and INAR models.

Based on these results, policymakers and decision makers for relevant government agencies may consider looking into the use of INAR and the algorithm formulated in this study as a means of making improvements in disease surveillance. Specifically, it may be beneficial to further refine and adopt the proposed system as a replacement to the current guidelines being followed by government agencies, and likewise improve the implementation of preventive and remedial measures against the spread of contagious diseases.

Directions for future research include taking overdispersion from count data into account by using models which are designed specifically for modeling overdispersed data.

ACKNOWLEDGMENT

The researchers would like to express their gratitude to the Department of Health's Regional Epidemiology and Surveillance Unit for providing the necessary data used in this study.

REFERENCES

ALZAID, A. A., and AL-OSH, M.,1990, An integer-valued pth-order autoregressive structure (INAR(p)) process, *Journal of Applied Probability*, 27(2): 314-324.

ANGKICO, L. B., DIAZ, P. A., LEONG, R. F., and CO, F. F., 2014, Biosurveillance of measles using control charts: A case study using National Capital Region laboratory confirmed measles counts from January 2009 to January 2014,*The Philippine Statistician*, 63(2): 31-49.

BU, R., 2006, Essays in financial econometrics and time series analysis (Unpublished doctoral dissertation). University of Liverpool, UK.

BU, R., MCCABE, B., and HADRI, K., 2008, Maximum likelihood estimation of higher-order integer valued autoregressive, *Journal of Time Series*, 29(6): 973-994.

BUCKERIDGE, D., 2007, Outbreak detection through automated surveillance: a review of the determinants of detection, *Journal of Biomedical Informatics*, 40(4):370-379.

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

CLEVELAND, W. S., and DEVLIN, S. J., 1988, Locally weighted regression: An approach to regression analysis by local fitting, *Journal of the American Statistical Association*, 83(403):596-610.

DEPARTMENT OF HEALTH, 2008, Manual of procedures for the Philippine integrated disease surveillance and response, Philippines.

FISHER., R.A., 1950, The significance of deviations from expectation in a Poisson series, Biometrics, 6:17-24

FRICKER, R. D., 2013, *Introduction to statistical methods for biosurveillance: With an emphasis on syndromic surveillance*, Cambridge: Cambridge University Press.

GUPTA, I. and GUIN, P., 2010, Communicable disease in the South-East Asia region of the World Health Organization: towards a more effective response, Bulletin of the World Health Organization, 88, 199-205. Available at: http://www.who.int/bulletin/volumes/88/3/09-065540/en/

HYNDMAN, R., and KHANDAKAR, Y., 2008, Automatic time series forecasting: The forecast package of R, *Journal of Statistical Software*, 27(3).

KWIATKOWSKI, D., PHILIPS, P., SCHMIDT, P., and SHIN, Y., 1992, Testing the null hypothesis of stationarity against the alternative of a unit root" How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54:159-178.

MOORE, A., COOPER, G., TSUI, R., and WAGNER, M., 2002, Summary of Biosurveillance-Relevant Statistical and Data Mining Technology, Technical Report, Pittsburgh, PA: Auton (Autnomous Systems) Lab, Carnegie Mellon University.

MÜLLER, M., 2007, *Information Retrieval for Music and Motion*, Springer Science and Business Media.

REIS, B. Y., and MANDL, K. D., 2003, Time series modeling for syndromic surveillance, *BMC Medical Informatics and Decision Making*, 3(1), 2. doi:10.1186/1472-6947-3-2

SAKOE, H., and CHIBA, S., 1978, Dynamic programming algorithm optimization for spoken word recognition, IEEE Transaction on Acoustics, Speech, and Signal Processing, ASSP-26(1):43-49.

SERFLING, R., 1963, Methods for current statistical analysis of excess pneumonia-influenza deaths, *Public Health Reports*, 78(6):494-506.

STEUTEL, F. W., andVAN HARN, K., 1979, Discrete analogues of self-decomposability and stability, *The Annals of Probability*, 7(5):893-899.

TANNER, M. andDE SAVIGNY, D., 2008, Malaria eradication back on the table. Bulletin of the World Health Organization, 86, 81-160. Available at: http://www.who.int/ bulletin/ volumes/86/2/07-050633/en/

UNITED NATIONS, 2015, The 2030 Agenda for Sustainable Development. Available at: https://sustainabledevelopment.un.org/post2015/transformingourworld

UNKEL, S., FARRINGTON, C. P., GARTHWAITE, P. H., ROBERTSON, C., and ANDREWS, N., 2012, Statistical methods for the prospective detection of infectious disease outbreaks: a review, *Journal of the Royal Statistical Society: Series A (Statistics in Society),* 175(1):49–82.

WEI, W. W. S.,2005, *Time Series Analysis: Univariate and Multivariate* Methods (2nd ed.). Boston: Pearson.

WORLD HEALTH ORGANIZATION, 2015, Measles. Available at: http://www.who.int/ mediacentre/factsheets/fs286/en/

WORLD HEALTH ORGANIZATION, n.d., Smallpox. Available at: http://www.who.int/ csr/disease/smallpox/en/