
ABSTRACT
For time-series analysis of disease counts, this paper proposes a method that identifies the shortest period of measurement, while not significantly decreasing prediction performance. There is a body of literature in statistics that shows how auto-correlation can identify the best period of measurement in order to improve the performance of a time-series prediction; therefore, period of measurement plays an important role in improving the performance of disease-count predictions. However, from the operational perspective in public health surveillance, there is a length limitation in which the period of measurements can offer meaningful and valuable predictions. The proposed method is an attempt to shorten the period of prediction but not significantly decreasing the prediction performance. Our study applies change-point analysis on auto-correlations of different periods of measurement in order to identify the shortest period that has a similar time-series prediction performance to the best prediction. Our method uses Q-Score as performance indicator. The evaluation is conducted against artificial neural networks and autoregressive, integrated-moving average as time-series analysis methods. The data used in this evaluation contains disease counts from 2007 to 2017 in northern Nevada. The disease counts, including: Chlamydia, Salmonella, Respiratory syncytial virus (RSV), Gonorrhea, Meningitis Viral and Influenza A, were predicted. Auto-correlation cannot guarantee the best performance for prediction of disease counts. However, the proposed method adopting change-point analysis suggests a period of measurement that ensures an operationally acceptable prediction period and a performance not significantly different than the best prediction.

Key words: Time Series Analysis; Prediction; Auto-correlation; Disease Counts; Public Health Surveillance.

INTRODUCTION
In a time-series prediction when a population is monitored, a measurement is the records of equally spaced disease counts over time. The length of these measurements is called period of measurement [1]. While time-series predictions have been widely used in public health surveillance, there is a body of literature in statistics that shows how auto-correlation suggesting the best period
of measurement can improve the performance of a time-series prediction [2–4]. Auto-correlation is a method in statistics that is often used to identify the best period of measurement for time-series analysis [2–4]. This method selects the period of measurement in which the auto-correlation is maximized; therefore, we can expect better prediction performance [4].

The period of measurement selection implies the interval for prediction. Thus, from the operational perspective in public health surveillance, for the prediction to be meaningful and valuable, there is a limitation on its length. For example, let us attempt to predict influenza A for next year: even though the eight-week period of measurement may generate the best performance for the prediction, it also produces eight weeks by eight weeks of predicted values. This is too long to provide any value for practitioners.

In response to the operational concern discussed above, the present work aims to identify the shortest period of measurement while not significantly decreasing performance. In order to do so, the method applies change-point analysis on auto-correlations of different periods of measurement. The objective is to identify the shortest period of measurement that has an auto-correlation value similar to the maximum of auto-correlations.

**BACKGROUND AND SIGNIFICANCE**

**Public Health Surveillance**

The initial target of public health surveillance was on infectious diseases; however, with the recent advancements in analytics, the data from surveillance systems is increasingly being used to predict future trends in a wide range of non-infectious disease distributions [5,6]. For instance, the Centres for Disease Control and Prevention organized a challenge: predict the 2013–2014 United States influenza season [7]. The ability to accurately forecast various diseases could facilitate key preparedness actions, such as: development and use of medical countermeasures, communication strategies, and healthcare resource management [8]. To achieve this goal, different statistical methods have been utilized to forecast disease counts, and time-series prediction is a method often seen in related literature [1,9–11] wherein the analysis predicts disease counts by modelling historical surveillance data [1,12].
Time Series Prediction in Public Health Surveillance

Prior work in time-series prediction of public health surveillance has been heavily relied on aberrancy-detection algorithms that are used to detect temporal changes in the data which may be indicative of a disease outbreak [13]. The Centers for Disease Control and Prevention’s Early Aberration Reporting Systems uses C algorithms. In terms of prediction capabilities, while C1 only supports Moving Average with 7-Day Window, C2 and C3 offer Moving Average with 7-Day Window and 2-Day Guardband. Similar to C1-C3, Reis [14], GScan [15] and Brillman [16] do not have long-term predictive features that allow for public health authorities to achieve annual planning.

These algorithms are primarily designed based on conventional hypothesis testing for the existence of disease outbreak. Aberrancy-detection algorithms detect only changes in static disease activity at a given time when the outbreak occurs and only notice the direction of changes in disease trends at a single time point [17]. However, in contexts where the prediction is for an annual disease count rather than the testing of a disease outbreak, Autoregressive Integrated Moving Average (ARIMA) models and machine learning can address the limitation of aberrancy-detection algorithms [18].

ARIMA models are commonly used in public health surveillance [19]. The ARIMA models have been built on three basic ideas: (1) the present value of time-series is a linear function of its past values and random noise in the autoregressive (AR) model [20], (2) the present value of time-series is a linear function of its present and past values of residuals in the moving average model [21], and (3) the AR moving average model [22] takes both AR and moving average models into account and considers both the historical values and residuals. The ARIMA model generally fits the time-series data based on the Rojas et al. (2008) AR moving average model. It also includes a differencing process which effectively transforms the non-stationary data required for the above-mentioned models into a stationary one used in ARIMA [19]. The ARIMA models have been widely used for time-series prediction in public health surveillance [12,23], including: the hemorrhagic fever with renal syndrome [24,25], dengue fever [26], tuberculosis [27] and mental health [28].

There is a growing body of literature [29–33] that uses machine learning approaches such as Artificial Neural Networks (ANN) for time-series prediction in public health surveillance. ANN is inspired by how biological nervous systems, such as the brain, process information. It is composed of a large number of highly interconnected processing elements (similar to neurons) working in unison to recognize patterns in data. In addition, ANNs, like people, learn by example.
The ability of ANN to recognize patterns in data allows for better predictions and provides assistance for public health surveillance because it is able to self-organize and self-learn processes [34]. Public health surveillance utilizes ANN to forecast diseases distributions, while Guan et al. (2004) use ANN to forecast incidents of hepatitis. Mehra et al. (2016) also use ANN in order to predict pre-planting risk of Stagonospora nodorum blotch in winter wheat.

Given the focus of this study to forecast the disease counts and the limitation of Aberrancy-detection algorithms to only detect the disease outbreak, the present paper only adopts ARIMA and machine learning.

Role of Period of Measurement for Time-Series Prediction in Public Health Surveillance

As discussed above, there has been a body of work predicting diseases for public health surveillance through the utilization of time-series methods such as ARIMA and machine learning. However, it is necessary to recognize that measurement periods play a significant role in the performance of time-series, in that time-series prediction methods may show different performance for the same population when predicting with different measurement periods [37–39]. For better surveillance of a disease, it is crucial to identify the period of measurement in which the time-series methods demonstrate the best performance for prediction in a particular population.

The performance indicators for time-series such as Q-Score [40] can be used to identify the period of measurement that generates the best performance. However, they are computationally expensive to run across multiple time-series analysis for different periods of measurement and compare the performance using the indicator. As such, the literature in this areas has suggested auto-correlation as one of the most commonly used algorithms to identify the best period of measurement in time-series [4]. Auto-correlation refers to the correlation of a time-series with its own past and future values [2]. The main objective for this method is to have an auto-correlation sequence of a periodic signal with the same cyclic characteristics as the signal itself, allowing auto-correlation to help verify the presence of cycles and determine their durations [3]. Therefore, the overall goal is to find the period of measurement that maximizes the auto-correlation wherein we can expect a better performance prediction [4].
While auto-correlation may suggest a period of measurement that operationally is too long to be meaningful, the current study aims to use change-point analysis in order to identify the shortest period of measurement with a similar auto-correlation value to the maximum auto-correlation. Therefore, we do not expect to see a significant drop in the performance prediction.

MATERIALS AND METHODS

Change-Point Analysis (CPA)

Change-point analysis is exclusively designed to detect subtle changes and characterize changing trends in a time-series [17,41]. In this paper, we used the Pruned Exact Linear Time (PELT) change-point analysis method suggested by Killick et al. (2012). This method is based on the CPA method of Jackson et al. (2005), but it incorporates a pruning step that reduces the computational cost of the method and does not affect the exactness of the resulting segmentation. Despite many of CPA methods that can only detect the most significant change-point, PELT can identify multiple change-points. Given the computational performance advantage of PELT [44], the current study has adopted this method. In addition, this study uses the R package for Change Point Analysis [45], which implements PELT. In this algorithm, a change-point is defined as the point that characterizes changing trends. In other words, the value for the change-point is significantly different than the point value immediately before the change-point.

Proposed method

Our method sorts the auto-correlations based on their period of measurements wherein the auto-correlation related to the shortest period of measurement is first while the auto-correlation for the longest period of measurement is last. After conducting CPA using the PELT algorithm on auto-correlations, our method indicates the immediate ascending change-point before the highest auto-correlations. We call this point as ascending change-point (ACP). The auto-correlation of the ACP is the auto-correlation related to the shortest period of measurement with similar performance to the highest auto-correlation. Since ACP indicates the closet ascending change point to the highest auto-correlations, so there will be no ascending change point between ACP and the highest auto-
correlations. This would result in similar performance between period of measurement associated with ACP and the period of measurement for the highest auto-correlations. This will be the shortest period of measurement with similar performance to the highest auto-correlations, because we skip all periods of measurements between ACP and the highest auto-correlations. As such ACP is the shortest period of measurement that has similar performance to the highest auto-correlations.

If the immediate change-point before the highest auto-correlations is descending, it indicates that there is no available period of measurement that is shorter than the highest auto-correlations and has similar performance to the highest auto-correlation. Therefore, the highest auto-correlations indicates the aimed period of measurement.

If there is no change-point before the highest auto-correlation, we consider the first point as the immediate change-point prior to the highest auto-correlation.

Evaluation of Proposed Method

Figure 1 presents the evaluation of the proposed method.

Data description

We used the notifiable disease case counts by epidemiological week from 2007 to 2017 in Washoe, Clark and Carson Counties in northern Nevada. The data included: Chlamydia, Salmonella, Respiratory syncytial virus (RSV), Gonorrhea, Meningitis Viral and Influenza A. The data was patient de-identified and included all-age case reporting.
Training and Test Datasets

The above-mentioned data was divided into train and test datasets with the ratio of 10:1. The time-series analysis was trained using the dataset created from the above-mentioned data from 2007 to 2016 and tested on the data for 2017. The performance was then reported.

As the data was reported on a weekly basis, the minimum period of measurement was one week. However, the study evaluated periods of measurement from 1 to 8 weeks. Depending on the period of measurement, the training set of the period of measurement could be nine years or one year. Figure 2 presents the training and testing sets in regard to periods measurement.

Time Series Analysis

Given the purpose of this study to forecast the disease counts and the shortfall of aberrancy-detection algorithms that can only detect disease outbreaks, the present paper only adopts ARIMA and machine learning. Provided the growing body of literature in ANN for public health surveillance [34,36], the model selected for machine learning was ANN. Depending on the learning structure, there are many different types of ANNs. In this study, we adopted a multilayer perceptron-based
ANN [46] as it has shown the most suitable ANNs for our data structure in our preliminary analysis.

**Performance Indicator: Q-Score**

The performance of time-series analysis was measured using an indicator called Q-Score proposed by Ghil et al. (2011). This indicator treats the data as continuous, which refers to the possibility of having any positive number as a predicted value or an observed value in the testing set. Formally, for each disease under the evaluation, we consider the prediction values of \( P(t) \in \mathfrak{C} \) and the observation values of \( O(t) \in \mathfrak{C} \) with integer time \( 1 \leq t \leq 52 \) counting weeks within a year. The overall error of the prediction is quantified by the total squared discrepancy between the prediction values and observed values for the testing set; See Equation 1.

\[
R_{\text{prediction}} = \sum_{t} [P(t) - O(t)]^2
\]

Equation 1

In order to evaluate the performance of prediction, we compare the time-series analysis under the evaluation with the unskilled prediction that predicts constant historic average count. Specifically, we define this with Equation 2.

\[
U(t) = \hat{P} = \frac{1}{n} \sum_{t=1}^{n} P(t)
\]

Equation 2

Finally, the performance indicator Q-Score is defined as the quadratic errors of the prediction under evaluation and the unskilled prediction presenting constant average. Therefore, Q-Score is defined as presented in Equation 3.

\[
Q-Score = \frac{R_{\text{prediction}}}{R_{\text{Unskilled Prediction}}} = \frac{\sum_{t=1}^{n} [P(t) - O(t)]^2}{\sum_{t=1}^{n} [U(t) - O(t)]^2}
\]

Equation 3

Q-Score may take positive values. It takes \( Q-Score = 1 \), if the time series prediction under the evaluation generates similar results than the unskilled prediction producing constant average. A desired time series analysis produces \( Q-Score < 1 \). Therefore, the aim is to minimize the
The Q-Score for each period of measurement was calculated for both ARIMA and ANN. Then a CPA was conducted to see if the suggested period of measurement generated similar performance with the best performance prediction generating the smallest Q-Score when utilizing ARIMA and ANN.

RESULTS

Figure 3 depicts the evaluation of the proposed method for Chlamydia, Salmonella, RSV, Gonorrhea, Meningitis, Viral and Influenza A. The results show that the proposed method suggests period of measurements not longer than 3 weeks, which is operationally acceptable. As discussed below, it was validated against the performance of ANN and ARIMA measured by Q-Score.

Figure 3.a presents the evaluation of the proposed method for Chlamydia. The biggest AC is for the 4-week period of measurement. However, the immediate ACP is on the 2-week PM. Therefore, the auto-correlations are similar in 2- to 4- week periods of measurements. As such the proposed method suggests that 2-week PM gives the similarly good performance than the best performance. The best performance measured by Q-Score occurs in 7-week PM for ANN and in 5-week PM for ARIMA. While there is no ACP, the DCP is on 2-Week PM. As such performance of ANN and ARIMA stays similar after for 2-week PM or longer. Although the 7-week PM for ANN and 5-week PM for ARIMA are providing the best performance i.e. smallest Q-Score, the results shows the 2-week PM indicated by the proposed method in this paper generates a similar performance.
The biggest AC: *, The best Performance measured by Q-Score for ANN and ARIMA: +

<table>
<thead>
<tr>
<th>PM (Week)</th>
<th>Q-Score of ANN</th>
<th>Q-Score of ARIMA</th>
<th>AC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.63</td>
<td>0.55</td>
<td>0.94</td>
</tr>
<tr>
<td>2</td>
<td>0.07 (DCP)</td>
<td>0.08 (DCP)</td>
<td>0.95 (ACP)</td>
</tr>
<tr>
<td>3</td>
<td>0.06</td>
<td>0.08</td>
<td>0.95</td>
</tr>
<tr>
<td>4</td>
<td>0.02</td>
<td>0.02</td>
<td>1.04*</td>
</tr>
<tr>
<td>5</td>
<td>0.03</td>
<td>0.01*</td>
<td>1.03</td>
</tr>
<tr>
<td>6</td>
<td>0.02</td>
<td>0.02</td>
<td>1.03</td>
</tr>
<tr>
<td>7</td>
<td>0*</td>
<td>0.01</td>
<td>1.03</td>
</tr>
<tr>
<td>8</td>
<td>0.01</td>
<td>0.01</td>
<td>1.02</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PM (Week)</th>
<th>Q-Score of ANN</th>
<th>Q-Score of ARIMA</th>
<th>AC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.23</td>
<td>0.59</td>
<td>0.56</td>
</tr>
<tr>
<td>2</td>
<td>1.31</td>
<td>0.57</td>
<td>0.59</td>
</tr>
<tr>
<td>3</td>
<td>0.91 (DCP)</td>
<td>0.38 (DCP)</td>
<td>0.83 (ACP)</td>
</tr>
<tr>
<td>4</td>
<td>0.89</td>
<td>0.38</td>
<td>0.85</td>
</tr>
<tr>
<td>5</td>
<td>0.89</td>
<td>0.36</td>
<td>0.86*</td>
</tr>
<tr>
<td>6</td>
<td>0.86</td>
<td>0.36</td>
<td>0.84</td>
</tr>
<tr>
<td>7</td>
<td>0.85</td>
<td>0.35</td>
<td>0.84</td>
</tr>
<tr>
<td>8</td>
<td>0.82*</td>
<td>0.32*</td>
<td>0.85</td>
</tr>
</tbody>
</table>

(a) Chlamydia
(b) Salmonella
(c) Respiratory syncytial virus (RSV)

(d) Gonorrhea
<table>
<thead>
<tr>
<th>PM (Week)</th>
<th>Q-Score of ANN</th>
<th>Q-Score of ARIMA</th>
<th>AC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.91</td>
<td>1.17</td>
<td>0.63</td>
</tr>
<tr>
<td>2</td>
<td>0.17</td>
<td>0.39</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>(DCP)</td>
<td>(DCP)</td>
<td>(ACP)</td>
</tr>
<tr>
<td>3</td>
<td>0.99</td>
<td>0.34</td>
<td>0.47</td>
</tr>
<tr>
<td>4</td>
<td>1.01</td>
<td>0.82</td>
<td>0.43</td>
</tr>
<tr>
<td>5</td>
<td>1.04</td>
<td>0.87</td>
<td>0.41</td>
</tr>
<tr>
<td>6</td>
<td>1.05</td>
<td>0.89</td>
<td>0.41</td>
</tr>
<tr>
<td>7</td>
<td>1.07</td>
<td>0.93</td>
<td>0.38</td>
</tr>
<tr>
<td>8</td>
<td>1.07</td>
<td>0.94</td>
<td>0.37</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PM (Week)</th>
<th>Q-Score of ANN</th>
<th>Q-Score of ARIMA</th>
<th>AC</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.01</td>
<td>0.07</td>
<td>0.63</td>
</tr>
<tr>
<td>2</td>
<td>0.03</td>
<td>0.09</td>
<td>0.92</td>
</tr>
<tr>
<td></td>
<td>(DCP)</td>
<td>(DCP)</td>
<td>(ACP)</td>
</tr>
<tr>
<td>3</td>
<td>0.02</td>
<td>0.10</td>
<td>0.47</td>
</tr>
<tr>
<td>4</td>
<td>0.03</td>
<td>0.1</td>
<td>0.43</td>
</tr>
<tr>
<td>5</td>
<td>0.19</td>
<td>0.1</td>
<td>0.41</td>
</tr>
<tr>
<td></td>
<td>(ACP)</td>
<td>(ACP)</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.2</td>
<td>0.13</td>
<td>0.41</td>
</tr>
<tr>
<td>7</td>
<td>0.2</td>
<td>0.49</td>
<td>0.38</td>
</tr>
<tr>
<td></td>
<td>(ACP)</td>
<td>(ACP)</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.22</td>
<td>0.94</td>
<td>0.37</td>
</tr>
</tbody>
</table>

(e) Meningitis, Viral
(f) Influenza A

Figure 3 Results

Figure 3.b demonstrates while our proposed method suggests the 3-week PM for Salmonella, the best performance occurs in the 8-week PM for both ANN and ARIMA. However, the results of CPs on Q-Scores shows that 3-week PM generates similar performance than the best Q-scores for ANN and ARIMA. Similarly, the results for RSV and Gonorrhea also validate the proposed method; See Figure 3.c and Figure 3.d.

Figure 3.e demonstrates an interesting example for Meningitis, Viral. According to Figure 3.e, 2-week PM is suggested by the proposed method, while the greatest as well as the ACP for AC occurs for the 2-weeks PM. For ANN, the best performance measured by Q-Score occurs for 3-week PM, however the 2-week has DCP for the Q-Scores of ANN. Therefore, 3-week PM generates similar performance as 2-week PM suggested by the proposed method. For ARIMA, the best performance occurs for 2-week PM which has the DCP as well. Therefore, the proposed method is also validated for Meningitis, Viral.

Influenza A has attracted a great attention from time series analysis in public health. The biggest AC occurs in 2-week PM, but the best performance is in 1-week PM for both ANN and ARIMA. However, there is no change point till 5-week PM for ANN and 7-week PM for ARIMA wherein ACP occurs. Therefore, we can assume that in both ANN and ARIMA the performance of 1-week PM presenting the best Q-Score is similar to 2-week PM suggested by the proposed method because of the biggest AC with DCP in 2-week PM. The proposed method also improves the prediction of Influenza A.
DISCUSSION

Following the extensive use of time-series predictions in public health surveillance, auto-correlation is commonly used in statistics to identify the best period of measurement and improve the performance of predictions [2–4]. However, the forecast needs to address the operational perspective as well as offer meaningful and valuable predictions. For that reason, practitioners in public health surveillance may choose a shorter period of measurements wherein the forecast results may not be as accurate as that of an analysis of longer periods of measurements.

This study proposed a method that runs CPA on auto-correlations and identifies the shortest period of measurement with a performance prediction similar to the best performance prediction. Our method was evaluated against ANN and ARIMA methods for a time-series analysis of disease counts in Cark, Carson and Washoe Counties in northern Nevada, between 2007 and 2017, including: Chlamydia, Salmonella, RSV, Gonorrhea, Meningitis, Viral and Influenza A.

Unfortunately, auto-correlation cannot guarantee the best performance for disease prediction. For example, Figure 3.a shows that for Chlamydia, the greatest auto-correlation occurs in the 4-week period of measurement while the best performance of ANN can be achieved in the 7-week period and the best performance of ARIMA is in 5-week PM. This was also the case for RSV, Gonorrhea, Meningitis, Viral and Influenza A. However, the proposed method adopting change-point analysis suggests the shortest period of measurement (to satisfy operational perspective) that ensures acceptable performance predictions similar to the best Q-Scores.

Time series prediction is an important tool for public health and clinical medicine to identify seasonal periods of changes in relative risk for disease activity. Observed values that exceed predicted parameters do not necessarily reflect a “failed” prediction but rather, a pattern of reported activity that was not observed in previous data. This is an important adjunct to other methodologies for aberration detection such as the aforementioned Early Aberration Reporting System. Predictions offer value to the unaware practitioner by offering a “most likely” hypothesis for expected disease activity, which may carry implications for proactive education and disease control policies.

Although the current study evaluated the proposed method on a variety of diseases, the data was limited to Nevada. Therefore, expanding the datasets and re-evaluating the method on a wider
range of diseases from various geographical locations and higher sample sizes would provide better understanding about the performance prediction of this method.

The evaluation of the proposed method against only ARIMA and ANN suggests a limitation that can be addressed in future studies by applying more time series prediction methods. The method uses auto-correlation; however, Fourier Transforms also has been used in literature to identify period of measurement [47]. This opens an avenue of research that can compare the performance of AC and Fourier Transforms adopted in the method proposed in this study.

**CONCLUSION**

Through the adoption of auto-correlation and change-point analysis, this paper proposes a novel method for identifying period of measurement that can improve the performance of time-series predictions for disease counts. Our method implements a practical perspective in which the aim is to determine the shortest period of measurement that achieves a better prediction performance. Our method was evaluated against ANN and ARIMA analyses for the Nevada disease counts between 2007 and 2017, of Chlamydia, Salmonella, RSV, Gonorrhea, Meningitis, Viral and Influenza A. Future work includes enhancing the evaluation of the method by using more diverse datasets as well as evaluating the utilization of Fourier Transforms instead of AC used in the current study.
REFERENCES


29 Fisichella M, Stewart A, Denecke K, *et al.* Unsupervised public health event detection for


