Timely detection of Pertussis outbreaks in Iran: The comparison performance of Wavelet-based outbreak detector, Exponential weighted moving average, and Poisson regression-based methods

Yousef Alimohamadi
  Tehran University of Medical Sciences

Seyed Mohsen Zahraei
  Iranian Ministry of Health and Medical education

Manoochehr Karami
  Hamadan University of Medical Sciences School of Public Health

Mehdi Yaseri
  Tehran University of Medical Sciences

Mojtaba Lotfizad
  Tarbiat Modares University Faculty of Engineering

Kourosh Holakouie-Naieni (✉ holakoik@hotmail.com)
  Tehran University of Medical Sciences

Abstract

Background

Early detection of outbreaks is very important for surveillance systems. Due to the importance of the subject and lack of similar studies in Iran, the aim of this study was to determine the performance of the Wavelet-Based Outbreak detection method)WOD(in detecting outbreaks and to compare its performance with Poisson regression-based model and Exponential weighted moving average (EWMA) using data of simulated pertussis outbreaks in Iran.

Methods

The data on suspected cases of pertussis from 25th February 2012 to 23rd March 2018 in Iran was used. The performance of the WOD (Daubechies 10 and Haar wavelets), Poisson regression-based method, and EWMA Compared in terms of timeliness and detection of outbreak days using the simulation of different outbreaks (literature-based and researcher-made outbreaks). The sensitivity, specificity, false alarm and false negative rate, positive and negative likelihood ratios, under ROC areas and median timeliness were used to assess the performance of the methods.

Results

In a literature-based outbreak simulation, the highest and lowest sensitivity, false negative in the detection of injected outbreaks were seen in Daubechies 10 (db10), with sensitivity 0.59 (0.56-0.62), and Haar wavelets with 0.57 (0.54-0.60). In the researcher-made outbreaks, the EWMA (K=0.5) with sensitivity 0.92 (0.90-0.94) had the best performance. About timeliness, the WOD methods showed the best performance in the early warning of the outbreak in both simulation approaches.

Conclusions

Performance of the WOD in the early alarming outbreaks was appropriate. However, it's better as the method was used along with other methods in public health surveillance systems.

Background

Outbreaks of infectious diseases are one of the main public health challenges(1). Early detection of outbreaks, timely response, and control of these aberrations are very important for public health surveillance systems(2). The main purpose of the public health surveillance system is an ongoing data collection, analysis, interpretation and final dissemination in public health practice. No doubt, the surveillance system will contribute to reducing the morbidity and mortality due to health-related events(3). Additionally, utilizing a suitable method in a surveillance system for early detection of naturally occurring or bioterrorism-related outbreaks have a very important role in reducing the time between outbreak occurrence and detection (4). Due to many restrictions of the traditional surveillance system in early detection of the outbreak, the syndrome surveillance system is recommended to be used in such

conditions. The syndrome surveillance includes collection, analysis, interpretation, and dissemination of health-related data defined in public health sectors (5−7). The objective of the syndrome surveillance system is to provide an early warning (system) for public health threats in near-real-time(8). The Important feature of the syndrome surveillance system is early warning or detection of health-related aberration or outbreak which leads to the reduction of morbidity and mortality of affected people (9). Two main tools; temporal and spatial models are used in the syndrome surveillance system to detect outbreaks as early as possible(10, 11). Different methods, including Cumulative Sum (CUSUM), EWMA, Shewhart chart and time-series models, are the main tools used by syndrome surveillance to detect outbreaks(12). It is worth mentioning that all these methods follow a two-step procedure for the detection of outbreaks. Firstly, determine the level of alarm threshold using the baseline values or a non-outbreak period. Secondly, different algorithms are used for aberration detection using the defined alarm threshold. The above-mentioned methods have two main problems, the first one is related to the baseline period, especially when the real-world data set is used, because it may include the outbreak period. The second problem is related to the nature of the surveillance system dataset. In most conditions, the surveillance data set is non-stationary and noisy(13, 14). The non-stationary of the data makes the mean and the variance of the data not stable over time leading to changes in time series behavior. This phenomenon can increase the risk of a false alarm rate. Limitations such as the effect of non-stationary data on the performance of the syndrome surveillance system necessitate using an appropriate and proper method to overcome the problem. Some researchers used the wavelet-based outbreak detector (WOD) method in the detection of outbreaks (14−16). Previous literature indicated that the number of studies that used this method is very few. Based on our knowledge, there was no evidence that WOD was ever used in any study conducted in Iran. Due to importance of timely detection in surveillance system, the aim of this study was to determine the performance of WOD method in detecting outbreaks and compare its performance with the Poisson regression-based model and EWMA, using data of simulated pertussis outbreaks in Iran

Methods

The data set used

The data used in this study included suspected daily cases of pertussis from 25th February 2012 to 23rd March 2018. The data is nationally collected from the national registry at the department of vaccine-preventable diseases, in the Iranian Ministry of Health (Fig. 1).

Outbreak Simulation

According to the information provided by the national health authorities, no outbreak detected an understudied period. So, simulation outbreaks were used to assess the performance of the WOD, also EWMA and poison regression-based methods in outbreak detection. Two approaches were considered in the simulated outbreaks. Firstly, reported daily pertussis outbreaks in the literature were used as a source for data. (17−24) Then, the extracted outbreaks were injected into the real number of reported pertussis suspected cases in Iran and considered as a gold standard. The second approach focused on the outbreaks created by the researcher; which differ in types, duration, and sizes. In this study, three types of

outbreaks were simulated and injected into real data. The three types of outbreaks are exponential (2,4,8,16,8,4,2), Linear (2,4,6,8,6,4,2) and uniform (6,6,6,6,6,6) increase cases over time. Overall, a total of 560 outbreak days were injected in the real dataset including 10 exponential, 9 linear, and 9 uniform outbreaks. The duration of these outbreaks was between 1 to 5 weeks. The number of injected cases was the utmost 3δ of real data means and the interval between each outbreak was 2 months (Figs. 2 and 3).

Methods Of Outbreak Detection

Wavelet-based outbreak detector (WOD)

In this study, we used a discrete wavelet transform (DWT) model introduced by Aradhye et al (25), and Shemuli et al (14) referred to Multi-scale statistical process control (MSSPC). In this method, the first step e understudy time series decomposed using the desired wavelet (in the current study, db10 and Haar wavelets were used). The result of this decomposition is the production of approximate and detailed coefficients at the first level. In the next step, the approximation coefficient of the first level was decomposed for producing the approximation and details coefficients of level 2. The decomposition of series continued up to 5 levels. In the next step, the Shewhart control chart was applied in monitoring all details coefficients and the last approximation coefficient (level 5). If the values of the coefficients were within the upper and lower limits of the Shewhart chart, the values were converted to zero and the values outside the range were kept. Then, the time series were reconstructed and monitored by the Shewhart control chart to detect the simulated outbreaks. The chart was used in statistical process control had upper and lower control limits calculated from $\mu \pm k\sigma$; where μ is the mean of the understudy time series, k is the fixed-parameter which ranges from 0 to 3 that considered 0.5 to 2 in the study. σ is the standard deviation of the understudy time series. The DWT has different wavelets. More details about the type of wavelets are described in other sources (14, 25−27).

The Exponentially Weighted Moving Average (EWMA)

The statistic of EWMA in a day of t was defined as follows:

$$EWMA_t = \lambda Y_t + (1 − \lambda) EWMA_{t-1} \quad (12)$$

Where, Y_t equals the number of suspected cases of pertussis in a day of t and λ is the weighting parameter which includes $0 < \lambda \leq 1$. A value of $\lambda = 1$ gives more weight to newest data and small value of λ (closer to 0) gives more weight to older data(28). In this study, the amount of λ determined 0.46 and 0.3 for literature-based and researcher made outbreaks respectively.

The upper control limit or the alarm threshold level by using this method calculated as follows:

$$Upper\ Control\ Limit = EWMA_0 + k \times \sigma_{EWMA} \quad (28−30)$$

Where k is specified in a way that results in the desired confidence interval. In the current study, k = 0.5, 1, 1.5,2 and 3 were considered. σ_{EWMA} is the standard deviation of the calculated statistics of EWMA at

times t to t_n and $EWMA_0$ is the mean of non-outbreak days. If the calculated EWMA statistics were more than the upper control limit, it is considered as an alarm for outbreak or aberration. To remove explainable patterns and create a normal distribution, Moving Average (MA):13 were used. It means the mean of the previous 13 data was replaced instead of each data.

Poisson Regression-based Method

The Upper control limits of Poisson regression

To determine the upper control limit, the expected mean of suspected pertussis cases was estimated as follows:

$$\log(\lambda_j) = \alpha_i + \beta X$$

Where log λ_j is the expected mean of occurred cases in time j, X is another effective factor, such as day, month and so on which are effective factors on the expected mean of pertussis cases and β is Coefficient of X. After the estimation of this parameter, the alarm threshold limit estimated as follows:

$$\lambda_{ij}^X \left(1 + \frac{2}{3} Z\alpha/2 \times \sqrt{\frac{1}{\lambda_{ij}} + Vt}\right)^{3/2}$$

Where Vt is the variance of estimated mean and its equal to:

Vt: var (α) + t^2 var (β) + 2t cov(α,β)(31)

The $Z_{0.90}$, $Z_{0.95}$, $Z_{0.99}$ were used to calculate different thresholds.
Measures Of The Algorithm's Performance

The performance of these methods was measured using the sensitivity, specificity; false alarm rate, likelihood ratio and area under the receiver operating characteristics (ROC) curve (AUC). The total number of outbreak-days was considered as the gold standard method to calculate the appropriate measures in order to evaluate the performance of different algorithms.

Used Software's

All analyses performed using MATLAB R2018a, STATA 15 (StataCorp LLC) and Excel 2010.

Results

Literature-based outbreaks simulation

Sensitivity, specificity, positive and negative likelihood ratio under ROC areas

According to this approach, the most sensitive and lowest false negative in the detection of injected outbreaks were seen in Daubechies 10 (db10) with sensitivity 0.59 (0.56−0.62), and Haar wavelets with 0.57 (0.54−0.60) and K = 0.5 for controls limit in Shewhart control charts respectively. In terms of specificity, the EWMA with k = 1.5, 2 and 3 and Haar wavelet with k = 1.5 and 2 had the optimum specificity (100%). So, this algorithm had the lowest false alarm in the detection of outbreaks. The most positive likelihood ratio (LR+) was seen in EWMA with k = 1.5 (LR+: 129.3), and Haar wavelet with k = 1.5 (LR+: 28.65) respectively. According to under ROC areas, the EWMA algorithm with K = 0.5 and 1 had the best performance in the detection of both outbreaks and non-outbreak days Under ROC areas 0.92 and 0.90 respectively. The size of under curve areas is between 0 to1 and more of this area (near to 1) refers to better performance of the algorithm (Table 1 and Fig. 4).

Timeliness of methods

In terms of timeliness, all algorithms with less amount of k (0.5, 1) had at least an alarm on the first day of outbreaks. But, the WOD methods showed the best performance in the early warning of the outbreak. The Haar and the db10 wavelet with K; 0.5 had the best performance in generating an alarm on the first day of outbreaks with 7 on injected outbreaks. As a result of an increase in the amount of K, the probability of generation alarm on the first day of outbreaks decreased due to an increase in the level of alarm thresholds. The median, minimum and maximum timeliness (according to day) in Haar (k; 0.5) and db10 (k: 0.5) wavelets based method was 2 (1 to 14) and 2 (1-14) respectively. This amount was less than the median timeliness of other algorithms. It means that 50% of induced alarms occurred on the second day of the outbreak. The WOD generated an alarm in all 19 outbreaks. This was the best performance among all algorithms. More information showed in Table 3.

Table 1

the performance of understudy methods in detection of Literature-based outbreaks simulation

algorithm	Sensitivity	Specificity	false Alarm	False-negative	LR+	LR-	Under the ROC area
EWMA							
K = 0.5	0.49 (0.46–0.52)	0.86 (0.84–0.88)	0.14 (0.12–0.16)	0.51 (0.48–0.54)	3.60	0.59	0.68
K = 1	0.34 (0.32–0.37)	0.95 (0.94–0.96)	0.05 (0.04–0.06)	0.66 (0.63–0.68)	7.05	0.69	0.65
K = 1.5	0.24 (0.21–0.26)	1.00 (1.00–1.00)	0.00 (0.00–0.00)	0.76 (0.74–0.79)	129.35	0.76	0.62
K = 2	0.18 (0.15–0.20)	1.00 (1.001.00)	0.00 (0.00–0.00)	0.82 (0.80–0.85)	-	0.82	0.59
K = 3	0.09 (0.07–0.10)	1.00 (1.00–1.00)	0.00 (0.00–0.00)	0.91 (0.90–0.93)	-	0.91	0.54
Poisson Regression							
Z 1-α/2 = 90	0.20 (0.18–0.22)	0.95 (0.93–0.96)	0.05 (0.04–0.07)	0.80 (0.78–0.82)	3.72	0.85	0.57
Z 1-α/2 = 95	0.16 (0.13–0.18)	0.96 (0.95–0.97)	0.04 (0.03–0.05)	0.84 (0.82–0.87)	4.20	0.88	0.56
Z 1-α/2 = 99	0.12 (0.10–0.14)	0.98 (0.97–0.98)	0.02 (0.02–0.03)	0.88 (0.86–0.90)	4.92	0.90	0.55
WOD(Haar)							
K = 0.5	0.57 (0.54–0.60)	0.52 (0.49–0.55)	0.48 (0.45–0.51)	0.43 (0.40–0.46)	1.19	0.82	0.55
K = 1	0.23 (0.21–0.26)	0.92 (0.90–0.94)	0.08 (0.06–0.10)	0.77 (0.74–0.79)	2.88	0.83	0.58
K = 1.5	0.08 (0.06–0.10)	1.00 (0.99-1.00)	0.00 (0.00-0.01)	0.92 (0.90–0.94)	28.65	0.92	0.54

algorithm	Sensitivity	Specificity	false Alarm	False-negative	LR+	LR-	Under the ROC area
K = 2	0.02 (0.01– 0.03)	1.00 (1.00– 1.00)	0.00 (0.00– 0.00)	0.98 (0.97– 0.99)	21.01	0.98	0.51
Db10							
K = 0.5	0.59 (0.56– 0.62)	0.44 (0.41– 0.47)	0.56 (0.53– 0.59)	0.41 (0.38– 0.44)	1.06	0.93	0.52
K = 1	0.26 (0.24– 0.29)	0.88 (0.86– 0.90)	0.12 (0.10– 0.14)	0.74 (0.71– 0.76)	2.22	0.84	0.57
K = 1.5	0.15 (0.13– 0.17)	0.97 (0.96– 0.98)	0.03 (0.02– 0.04)	0.85 (0.83– 0.87)	4.92	0.88	0.56
K = 2	0.10 (0.08– 0.12)	0.99 (0.98– 0.99)	0.01 (0.01– 0.02)	0.90 (0.88– 0.92)	7.13	0.91	0.54

Table 2
the performance of under study methods in detection of researcher-made outbreaks.

algorithm	Sensitivity	Specificity	False Alarm Rate	False Negative	LR+	LR-	Under ROC area
EWMA							
K = 0.5	0.92 (0.90–0.94)	0.92 (0.90–0.93)	0.08 (0.07–0.10)	0.08 (0.06–0.10)	10.87	0.09	0.92
K = 1	0.83 (0.80–0.86)	0.97 (0.96–0.98)	0.03 (0.02–0.04)	0.17 (0.14–0.20)	29.84	0.17	0.90
K = 1.5	0.69 (0.65–0.73)	0.99 (0.99-1.00)	0.01 (0.00-0.01)	0.31 (0.27–0.35)	114.25	0.31	0.84
K = 2	0.49 (0.45–0.53)	1.00 (1.00–1.00)	0.00 (0.00–0.00)	0.51 (0.47–0.55)	269.19	0.51	0.74
K = 3	0.07 (0.05–0.09)	1.00 (1.00–1.00)	0.00 (0.00–0.00)	0.93 (0.91–0.95)	-	0.93	0.53
Poisson Regression							
Z 1-α/2 = 90	0.58 (0.54–0.62)	0.96 (0.95–0.97)	0.04 (0.03–0.05)	0.42 (0.38–0.46)	15.62	0.43	0.78
Z 1-α/2 = 95	0.46 (0.42–0.50)	0.98 (0.97–0.98)	0.02 (0.02–0.03)	0.54 (0.50–0.58)	19.52	0.55	0.72
Z 1-α/2 = 99	0.28 (0.24–0.32)	0.99 (0.98–0.99)	0.01 (0.01–0.02)	0.72 (0.68–0.76)	21.14	0.73	0.63
WOD(Haar)							
K = 0.5	0.79 (0.76–0.83)	0.51 (0.48–0.53)	0.49 (0.47–0.52)	0.21 (0.17–0.24)	1.61	0.41	0.65
K = 1	0.50 (0.46–0.54)	0.84 (0.82–0.85)	0.16 (0.15–0.18)	0.50 (0.46–0.54)	3.02	0.60	0.67
K = 1.5	0.14 (0.11–0.17)	0.99 (0.98–0.99)	0.01 (0.01–0.02)	0.86 (0.83–0.89)	10.37	0.87	0.56

algorithm	Sensitivity	Specificity	False Alarm Rate	False Negative	LR+	LR-	Under ROC area
K = 2	0.03 (0.01−0.04)	1.00 (0.99-1.00)	0.00 (0.00-0.01)	0.97 (0.96−0.99)	7.90	0.97	0.51
Db10							
K = 0.5	0.73 (0.69−0.76)	0.46 (0.43−0.48)	0.54 (0.52−0.57)	0.27 (0.24−0.31)	1.34	0.60	0.60
K = 1	0.54 (0.50−0.59)	0.81 (0.79−0.83)	0.19 (0.17−0.21)	0.46 (0.41−0.50)	2.85	0.56	0.68
K = 1.5	0.31 (0.27−0.35)	0.95 (0.94−0.96)	0.05 (0.04−0.06)	0.69 (0.65−0.73)	5.93	0.73	0.63
K = 2	0.16 (0.13−0.19)	0.98 (0.97−0.98)	0.02 (0.02−0.03)	0.84 (0.81−0.87)	6.44	0.87	0.57

Table 3-the performance of understudy methods in early detection or timelines of Literature-based outbreaks simulation

	Presence alarm in outbreaks(n=19)	Minimum timeliness	Maximum timeliness	Median timeliness	Alarm in first day(n=19)
EWMA					
K=0.5	18 (0.95)	1	35	4	4(0.21)
K=1	15 (0.79)	1	44	10	5(0.26)
K=1.5	12 (0.63)	3	44	14	0(0.00)
K=2	7 (0.37)	3	83	35	0(0.00)
K=3	5 (0.26)	18	83	68	0(0.00)
Poisson Regression					
Z 1-α/2=90	17(0.89)	1	36	5	2(0.11)
Z 1-α/2=95	17(0.89)	1	44	11	1(0.05)
Z 1-α/2=99	16(0.84)	2	44	16	1(0.05)
WOD					(0.00)
K=0.5	19(1.00)	1	14	2	7(0.37)
K=1	15(0.79)	2	35	8	0(0.00)
K=1.5	7(0.37)	2	56	13	0 (0.00)
K=2	5(0.26)	3	83	66	0(0.00)
Db10					
K=0.5	19(1.00)	1	14	2	7(0.37)
K=1	16(0.84)	1	35	10	1(0.05)
K=1.5	9(0.47)	3	35	13	0(0.00)
K=2	7(0.37)	3	74	35	0(0.00

The Researcher-made Outbreaks

Sensitivity, specificity, positive and negative likelihood ratio and Under ROC areas

According to this approach, the highest sensitivity and lowest false negative in the detection of the outbreaks were seen in the EWMA (K = 0.5) with sensitivity equal to 0.92 (0.90−0.94). Furthermore, the EWMA k = 2 and 3, Haar wavelet with k = 2 had the optimum specificity (100%). So this algorithm had the

lowest false alarm in the detection of outbreaks. The most positive likelihood ratio (LR+) was seen in the EWMA with k = 2 (LR+:269.5), and EWMA with k = 1.5 (LR+:114.25) respectively. Also, under ROC areas of the EWMA algorithm with K = 0.5 and 1 had the best performance in the detection of outbreaks and non-outbreak days with 0.92 and 0.90 respectively (Table 2 and Fig. 5).

Timeliness Of Methods

In terms of timeliness, all algorithms with less than 1 had at least an alarm on the first day of the outbreaks. However, the WOD methods showed the best performance in the early warning of the outbreak. The Haar wavelet with k: 0.5 and 1 also, the db10 wavelet with K; 0.5 had the best performance in generating an alarm in the first day of outbreaks with 18,16 and 14 alarm in the first day of injected outbreaks respectively. According to the results, with an increase in K, the probability of generating an alarm on the first day of outbreaks decreased due to the increase in the level of alarm thresholds. The median, the minimum, and maximum timeliness (according to day) in Haar(k;0.5) and db10 (k:0.5) wavelets based methods was 1 (1 to 3) and 1 (1−4) respectively. It was clear that this amount was less than the median timeliness of other algorithms. It means that 50% of the induced alarms by these methods occurred on the first day of the outbreak. Also, the poison regression with $Z_{1-\alpha/2}$=90 generates alarms in all 27 simulated outbreaks and this performance was better than other methods (Table 4).

Table 4

the performance of understudy methods in early detection or timelines of researcher-made outbreaks

Method	Presence alarm in outbreaks(n = 27)	Minimum timeliness	Maximum timeliness	Median timeliness	Alarm in the first day (n = 27)
EWMA					
K = 0.5	26(0.96)	1	4	3	5 (0.19)
K = 1	24(0.89)	2	5	4	0 (0.00)
K = 1.5	24(0.89)	1	11	4	1 (0.04)
K = 2	20(0.74)	4	11	6	0 (0.00)
K = 3	6(0.22)	6	28	14	0 (0.00)
Poisson Regression					
Z 1-α/2 = 90	27(1.00)	1	5	3	7 (0.26)
Z 1-α/2 = 95	26(0.96)	1	6	3	7(0.26)
Z 1-α/2 = 99	23(0.85)	1	6	3	4(0.15)
WOD					
K = 0.5	26(0.96)	1	3	1	18(0.67)
K = 1	26(0.96)	1	12	1	16(0.59)
K = 1.5	13(0.48)	1	21	7	2(0.07)
K = 2	5(0.19)	2	22	12	0(0.00)
Db10					
K = 0.5	26 (0.96)	1	4	1	14(0.52)
K = 1	21(0.78)	1	5	2	7(0.26)
K = 1.5	18(0.67)	1	19	4	5(0.19)
K = 2	14(0.52)	1	19	8	1(0.04)

Discussion

Due to the increased number of reported outbreaks in the world(32) The timely detection is very important for surveillance systems. So, using appropriate algorithms or methods has an important role in the early detection in any surveillance system. In the current study, the performance of three outbreaks detection

method including WOD, EWMA, and Poisson regression-based methods was assessed in the detection of simulated outbreaks in pertussis cases in Iran. The performance of algorithms, such as the Cusum and EWMA in detecting alarm of outbreaks were used in different studies to address different infectious diseases (33–37). The use of wavelet-based outbreak detector methods in the detection of outbreaks in surveillance systems is very rare. So, the assessment of the performance of this method can be very important as new in surveillance systems. There are three main approaches to assess outbreaks detection which include real data testing, fully synthetic and semi-synthetic simulations respectively. The use of real data provides the most valid results. But, such information is usually unavailable or inaccessible(38). For that reason, the majority of studies use synthetic simulation and semi-synthetic simulation to assess the performance of outbreaks detection methods (39–42). One of the advantages of using this approach is that it provides a real gold standard method that can be used to calculate and evaluate indices(3).

In this study, two approaches were used to apply the outbreaks simulation. The first was based on the reported pertussis outbreaks in literature, and the second was researcher-made outbreaks that used an actual daily number of pertussis cases. The performance of understudy methods was measured by ability in the detection of outbreaks and non-outbreak days and early warning of outbreaks or timeliness. The timeliness defined the time interval between the first day of the outbreak and the first real alarm produced by the methods in this study.

According to our results, the highest sensitivity in detecting Literature-based outbreaks was related to Wavelet-based methods. But, in the researcher made outbreaks, the highest was in the EWMA algorithms with the lower value of K in threshold determination. In both simulation approaches, the most specificity in the detection of non-outbreaks days was related to the wavelet-based method and the EWMA with a high value of K. The Poisson regression had the moderate performance in two approaches. Overall, by considering the low threshold values, the sensitivity of methods increased but the specificity decreased and vice versa. So, increasing the sensitivity of outbreaks detection methods, by considering the low level of the threshold can lead to a decrease in the specificity of related methods. In comparison with Under ROC areas, the most under curve areas were related to the EWMA with lower values of K. It means that the performance of EWMA in detecting outbreaks and non-outbreaks days was better than other methods. According to other study results, the discrete wavelet transform based model compared to the autoregressive (AR) model had similar sensitivity and specificity in the detection of outbreaks (16). The different results may be due to studies on different diseases or different datasets. According to our results, the performance of understudy methods in different approaches to outbreaks simulation was absolutely different. It's worth mentioning that these results were consistent with other studies conducted in this field (42, 43). The performance of outbreak detection methods was affected by many factors including the type of outbreaks, duration, and magnitude. According to the results of another study, the ability of outbreak detection algorithms was higher with larger and lower outbreaks thresholds (40, 44). So, the performance of a method in different conditions with different properties may be different. Additionally, outbreak detection algorithms may have different sensitivity and specificity in different outbreaks setting(1). In terms of early warning of outbreaks or timeliness, the WOD methods had the best

performance as an early warning method for outbreaks in both outbreak simulation approaches. It is consistent with other study(16). Early warning and detection of outbreaks and aberrations of infectious diseases are very important to surveillance systems which provides an opportunity to stop the spread of outbreaks from one region to another and can prevent the epidemic to become a global pandemic threat(45). There are many factors that can affect the early detection of outbreaks. One of these factors is the improvement of the outbreak alarm threshold(46). The lower level of alarm threshold can lead to early warning outbreak detection. But, the increase in early alarm can make a false alarm rate increase, and this should be considered in the outbreak detection. So, in the determination of the outbreak alarm threshold, some factors such as case fatality, contagiousness, confirmation costs, etc, must be considered.

Finally, when we compare these methods, the use of the WOD method can work as an early warning method for outbreaks detection. Thus, it can be applied without any presumption on the data set, so that the researcher can use it as an early warning tool to identify outbreaks of other infectious diseases. According to the results of the study, the ability of understudy methods in the detection of outbreak days with k = 2 and 3 is not more than fifty percent. This performance may be affected by the low incidence of pertussis in Iran, and it may be changed according to the type of infectious disease. So, the effect of the nature of the outbreaks on the performance of outbreaks detection methods must be considered. Since the outbreak alarm leads to the formation of the outbreak investigation team, hence the accuracy of alarm is very important. Therefore, it is recommended to use combined algorithms rather than choosing a single one(47). The main limitations of this study were; firstly, we used simulation outbreaks, which could differ from the real outbreaks. So to increase the validity of these methods, we tried to simulate different outbreaks, with different capacities. Secondly, the number of reported outbreaks in daily cases was few; in addition, the reported outbreaks in the literature might have different patterns compared to Iran's setting. For a better assessment of methods, we used two approaches of simulation. Finally, it is recommended that the understudy methods, especially the wavelet-based outbreak detection should be applied to other infectious diseases or data sets with different incidence and patterns.

Conclusions

According to the results of the study, the wavelet-based outbreak detector had the appropriate timeliness in outbreaks detection. But due to the importance of the problem and effect of the nature of outbreaks such as duration, size, and type on the performance of outbreak detection method, it's better that the method was used along with others in public health surveillance systems.

Abbreviations

WOD
Wavelet-Based Outbreak detection
EWMA
Exponential weighted moving average

Db

Daubechies

CUSUM

Cumulative Sum

MA

Moving Average

ROC

Receiver operating characteristic

LR

likelihood ratio

Declarations

Ethics approval and consent to participate

The study was approved by the ethical committee of the University, with the ID: IR.TUMS.SPH.REC.1397.276. The consent to participate not applicable.

Consent for publication

Not applicable

Availability of data and materials

Apply to the availability of these data are not publicly available. Data are however available from the authors upon reasonable request and with permission of Committee. A person who wants to access the raw data should contact with the corresponding author.

Competing interests

The authors declare that they have no competing interests.

Funding

Authors' contributions

YA data analysis, and interpretation of data and wrote the manuscript development; SMZ contributed to the data analysis and the study concept and design, and provided supervision, data extractions and provided expert insight; MK, MY and ML contributed to the study design and the data analysis, the study quality evaluation, manuscript preparation, and KHN provided supervision, Data analysis, provided expert insight and wrote the manuscript development. The author read and approved the final manuscript.

Acknowledgements

Author details

[1] Department of Epidemiology and Biostatistics, School of Public Health, Tehran University of Medical Sciences, Tehran, Iran. [2] Center for Communicable Diseases Control, Ministry of Health and Medical Education, Tehran, Iran. [3] Research Center for Health Sciences, Hamadan University of Medical Sciences, Hamadan, Iran. [4] Assistant Professor at the School of Electrical & Computer Engineering, Tarbiat Modares University, Tehran, Iran.

References

1. Kuang J, Yang WZ, Zhou DL, Li ZJ, Lan YJ. Epidemic features affecting the performance of outbreak detection algorithms. BMC Public Health. 2012;12(1):418.
2. De Vries DH, Rwemisisi JT, Musinguzi LK, Benoni TE, Muhangi D, de Groot M, et al. The first mile: community experience of outbreak control during an Ebola outbreak in Luwero District, Uganda. BMC public health. 2016;16(1):161.
3. Bédubourg G, Le Strat Y. Evaluation and comparison of statistical methods for early temporal detection of outbreaks: A simulation-based study. PloS one. 2017;12(7):e0181227.
4. Watkins RE, Eagleson S, Hall RG, Dailey L, Plant AJ. Approaches to the evaluation of outbreak detection methods. BMC public health. 2006;6(1):263.
5. Bjelkmar P, Hansen A, Schönning C, Bergström J, Löfdahl M, Lebbad M, et al. Early outbreak detection by linking health advice line calls to water distribution areas retrospectively demonstrated in a large waterborne outbreak of cryptosporidiosis in Sweden. BMC public health. 2017;17(1):328.
6. Karami M, Soori H, Mehrabi Y, Haghdoost AA, Gouya MM. Real time detection of a measles outbreak using the exponentially weighted moving average: does it work? Journal of research in health sciences. 2012;12(1):25-30.
7. Craig AT, Joshua CA, Sio AR, Donoghoe M, Betz-Stablein B, Bainivalu N, et al. Epidemic surveillance in a low resource setting: lessons from an evaluation of the Solomon Islands syndromic surveillance system, 2017. BMC public health. 2018;18(1):1395.
8. Colón-González FJ, Lake IR, Morbey RA, Elliot AJ, Pebody R, Smith GE. A methodological framework for the evaluation of syndromic surveillance systems: a case study of England. BMC public health. 2018;18(1):544.

9. Buehler JW, Whitney EA, Smith D, Prietula MJ, Stanton SH, Isakov AP. Situational uses of syndromic surveillance. Biosecurity and bioterrorism: biodefense strategy, practice, and science. 2009;7(2):165-77.

10. Buckeridge DL, Burkom H, Campbell M, Hogan WR, Moore AW. Algorithms for rapid outbreak detection: a research synthesis. Journal of biomedical informatics. 2005;38(2):99-113.

11. Chen D, Cunningham J, Moore K, Tian J. Spatial and temporal aberration detection methods for disease outbreaks in syndromic surveillance systems. Annals of GIS. 2011;17(4):211-20.

12. Solgi M, Karami M, Poorolajal J. Timely detection of influenza outbreaks in Iran: Evaluating the performance of the exponentially weighted moving average. Journal of infection and public health. 2018;11(3):389-92.

13. Lu H-M, Zeng D, Chen H. Prospective infectious disease outbreak detection using Markov switching models. IEEE Transactions on Knowledge and Data Engineering. 2009;22(4):565-77.

14. Dillard BL, Shmueli G. Wavelet-Based Monitoring for Disease Outbreaks and Bioterrorism: Methods and Challenges. InterStat. 2010;3.

15. Goldenberg A, Shmueli G, Caruana RA, Fienberg SE. Early statistical detection of anthrax outbreaks by tracking over-the-counter medication sales. Proceedings of the National Academy of Sciences. 2002;99(8):5237-40.

16. Zhang J, Tsui F-C, Wagner MM, Hogan WR, editors. Detection of outbreaks from time series data using wavelet transform. AMIA Annual Symposium Proceedings; 2003: American Medical Informatics Association.

17. Eshofonie AO, Lin H, Valcin RP, Martin LR, Grunenwald PE. An outbreak of pertussis in rural Texas: an example of the resurgence of the disease in the United States. Journal of community health. 2015;40(1):88-91.

18. Liu X, Wang Z, Zhang J, Li F, Luan Y, Li H, et al. Pertussis Outbreak in a primary school in China: infection and transmission of the macrolide-resistant Bordetella pertussis. The Pediatric infectious disease journal. 2018;37(6):e145-e8.

19. Theodoridou M, Hadjipanagis A, Persianis N, Makri S, Hadjichristodoulou C. Pertussis outbreak detected by active surveillance in Cyprus in 2003. Eurosurveillance. 2007;12(5):11-2.

20. Santiyán AM, Estrems RF, Lara JC, Enguídanos JA, Coito JN, Cifre AS. Early intervention in pertussis outbreak with high attack rate in cohort of adolescents with complete acellular pertussis vaccination in Valencia, Spain, April to May 2015. Eurosurveillance. 2015;20(27):21183.

21. Horby P, Macintyre C, McIntyre P, Gilbert G, Staff M, Hanlon M, et al. A boarding school outbreak of pertussis in adolescents: value of laboratory diagnostic methods. Epidemiology & Infection. 2005;133(2):229-36.

22. Bassinet L, Matrat M, Njamkepo E, Aberrane S, Housset B, Guiso N. Nosocomial pertussis outbreak among adult patients and healthcare workers. Infection Control & Hospital Epidemiology. 2004;25(11):995-7.

23. Borchardt S, Polyak G, Dworkin M. Parental attitude towards mass antimicrobial prophylaxis during a school-associated pertussis outbreak. Epidemiology & Infection. 2007;135(1):11-6.

24. Choe YJ, Kim JW, Park Y-J, Jung C, Bae G-R. Burden of pertussis is underestimated in South Korea: a result from an active sentinel surveillance system. Japanese journal of infectious diseases. 2014;67(3):230-2.

25. Aradhye HB, Bakshi BR, Strauss RA, Davis JF. Multiscale SPC using wavelets: theoretical analysis and properties. AIChE Journal. 2003;49(4):939-58.

26. Chaovalit P, Gangopadhyay A, Karabatis G, Chen Z. Discrete wavelet transform-based time series analysis and mining. ACM Computing Surveys (CSUR). 2011;43(2):6.

27. Alzaq H, Üstündağ BB. A Comparative Performance of Discrete Wavelet Transform Implementations Using Multiplierless. Wavelet Theory and Its Applications. 2018:111.

28. Lucas JM, Saccucci MS. Exponentially weighted moving average control schemes: properties and enhancements. Technometrics. 1990;32(1):1-12.

29. Karami M, Soori H, Mehrabi Y, Haghdoost AA, Gouya MM. Real time detection of a measles outbreak using the exponentially weighted moving average: does it work? J Res Health Sci. 2012;12(1):25-30.

30. Faryadres M, Karami M, Moghimbeigi A, Esmailnasab N, Pazhouhi K. Levels of alarm thresholds of meningitis outbreaks in Hamadan Province, west of Iran. J Res Health Sci. 2015;15(1):62-5.

31. Brookmeyer R, Stroup DF. Monitoring the health of populations: statistical principles and methods for public health surveillance: Oxford University Press; 2004.

32. Houlihan CF, Whitworth JA. Outbreak science: recent progress in the detection and response to outbreaks of infectious diseases. Clinical Medicine. 2019;19(2):140-4.

33. Han SW, Tsui KL, Ariyajunya B, Kim SB. A comparison of CUSUM, EWMA, and temporal scan statistics for detection of increases in Poisson rates. Quality and Reliability Engineering International. 2010;26(3):279-89.

34. Sparks RS, Keighley T, Muscatello D. Improving EWMA plans for detecting unusual increases in Poisson counts. Advances in Decision Sciences. 2009;2009.

35. Steiner SH, Grant K, Coory M, Kelly HA. Detecting the start of an influenza outbreak using exponentially weighted moving average charts. BMC medical informatics and decision making. 2010;10(1):37.

36. Sparks R, Patrick E. Detection of multiple outbreaks using spatio-temporal EWMA-ordered statistics. Communications in Statistics-Simulation and Computation. 2014;43(10):2678-701.

37. Karami M, Ghalandari M, Poorolajal J, Faradmal J. Early detection of meningitis outbreaks: Application of limited-baseline data. Iranian journal of public health. 2017;46(10):1366.

38. Karami M. Validity of evaluation approaches for outbreak detection methods in syndromic surveillance systems. Iranian journal of public health. 2012;41(11):102-3.

39. Hutwagner L, Browne T, Seeman GM, Fleischauer AT. Comparing aberration detection methods with simulated data. Emerging infectious diseases. 2005;11(2):314.

40. Jackson ML, Baer A, Painter I, Duchin J. A simulation study comparing aberration detection algorithms for syndromic surveillance. BMC medical informatics and decision making. 2007;7(1):6.

41. Hutwagner L, Thompson W, Seeman G, Treadwell T. A simulation model for assessing aberration detection methods used in public health surveillance for systems with limited baselines. Statistics in medicine. 2005;24(4):543-50.

42. Fricker Jr RD, Hegler BL, Dunfee DA. Comparing syndromic surveillance detection methods: EARS'versus a CUSUM-based methodology. Statistics in Medicine. 2008;27(17):3407-29.

43. Mathes RW, Lall R, Levin-Rector A, Sell J, Paladini M, Konty KJ, et al. Evaluating and implementing temporal, spatial, and spatio-temporal methods for outbreak detection in a local syndromic surveillance system. PloS one. 2017;12(9):e0184419.

44. Buckeridge DL. Outbreak detection through automated surveillance: a review of the determinants of detection. Journal of biomedical informatics. 2007;40(4):370-9.

45. Smolinski MS, Crawley AW, Olsen JM. Finding outbreaks faster. Health security. 2017;15(2):215-20.

46. Lewis R, Nathan N, Diarra L, Belanger F, Paquet C. Timely detection of meningococcal meningitis epidemics in Africa. The Lancet. 2001;358(9278):287-93.

47. Yahav I, Lotze T, Shmueli G. Algorithm Combination For Improved Performance In Biosurveillance. Infectious Disease Informatics and Biosurveillance: Springer; 2011. p. 173-89.
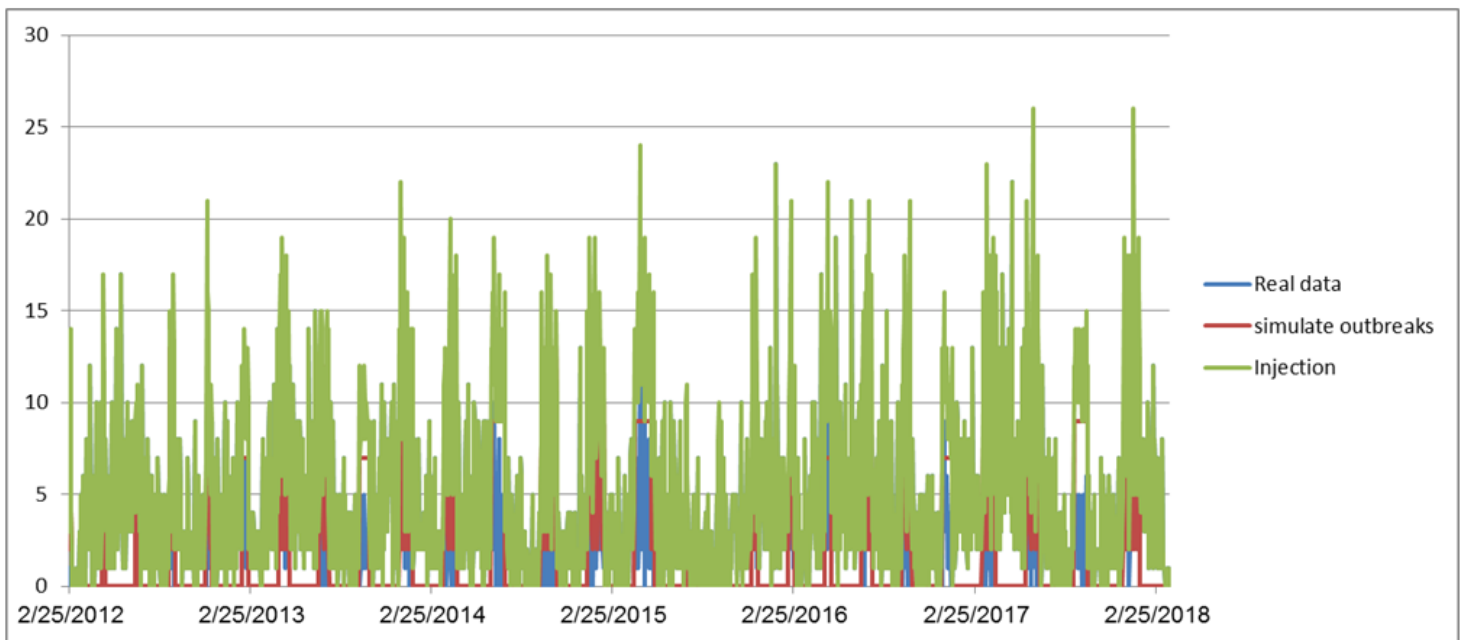
Figures
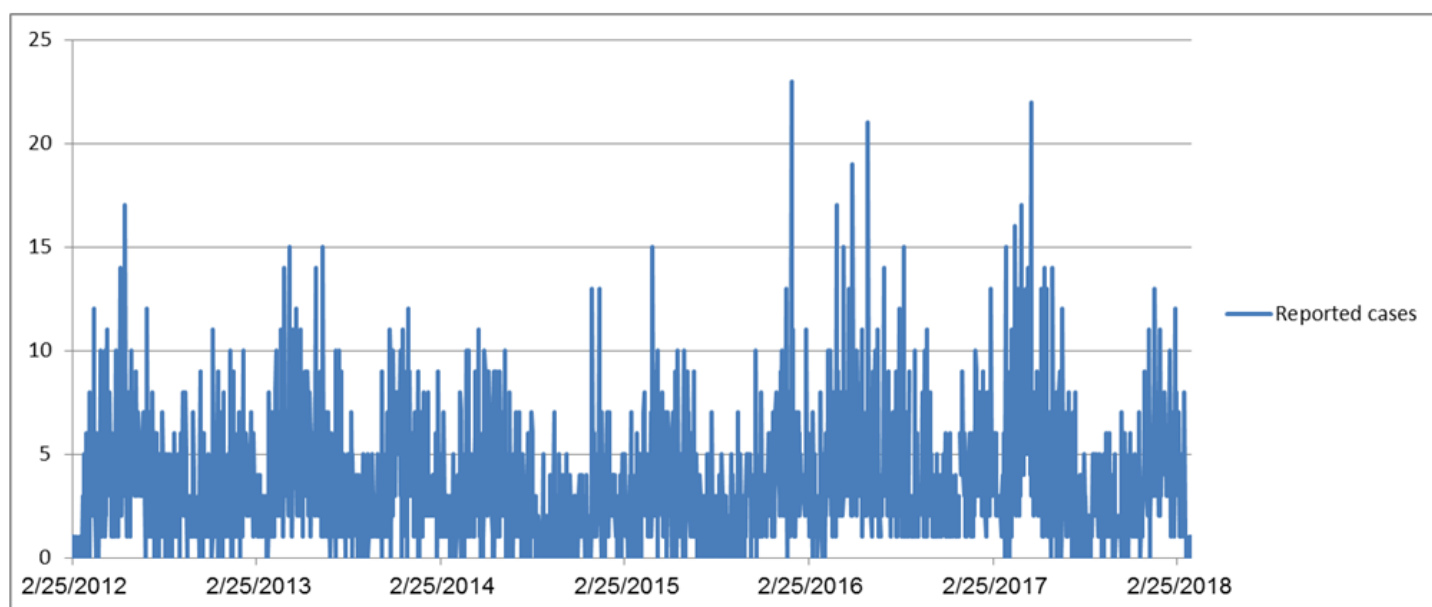
Figure 1

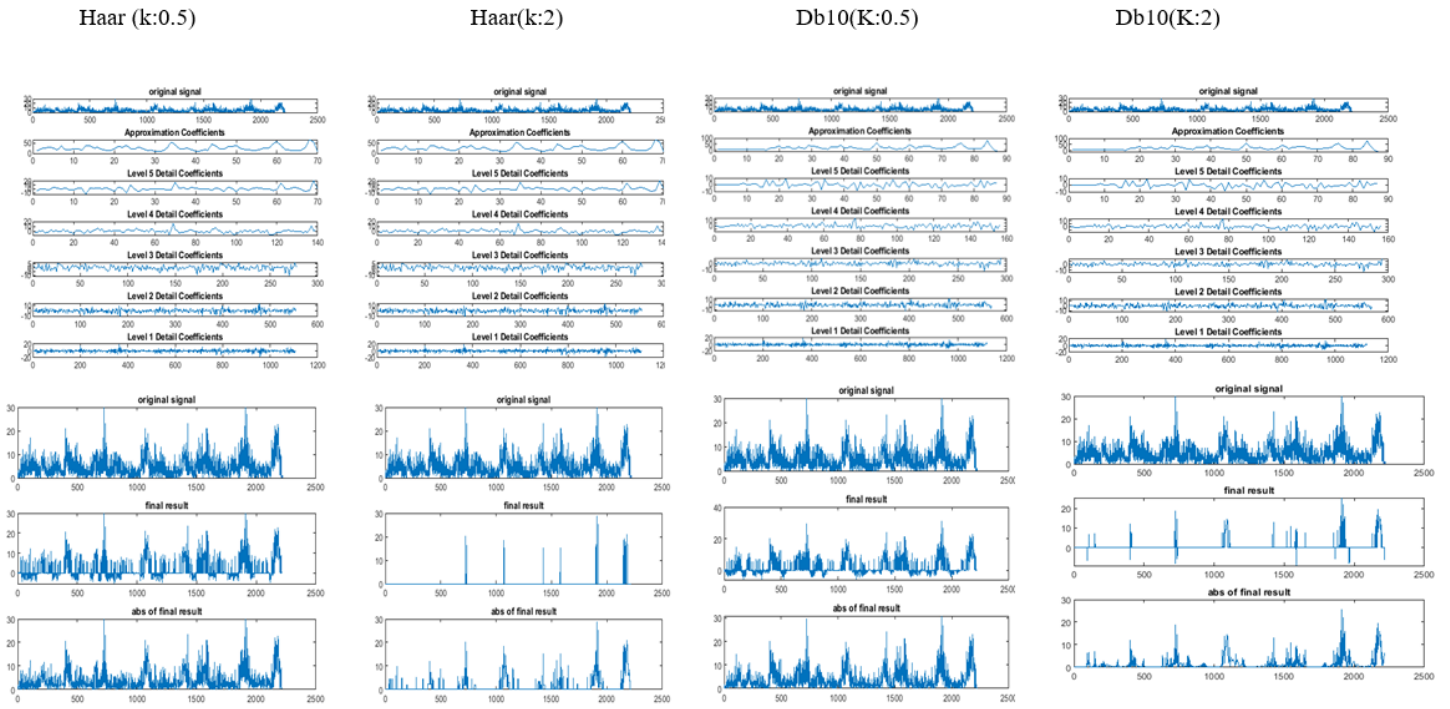Line plot of reported suspected cases of pertussis (values) from 25 February 2012 to 23 March 2018.
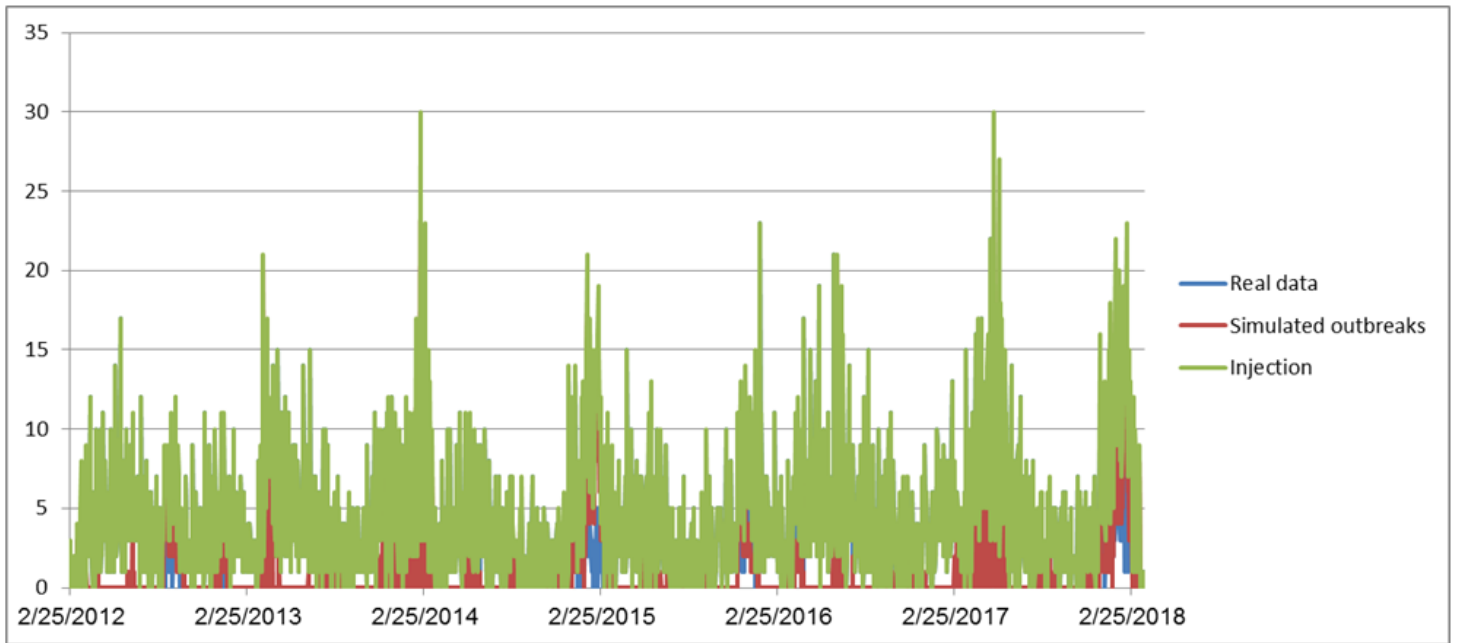
Figure 2

The Literature-based outbreaks simulation

Figure 3

The researcher-made outbreaks simulation

Haar (k:0.5)　　　　　Haar(k:2)　　　　　Db10(K:0.5)　　　　　Db10(K:2)

Figure 4

The result of the wavelet outbreaks detector in the detection of literature-based outbreaks. The above row indicates the levels of decomposition using discrete Wavelet transform and the under rows indicated the original signal (above) detected alarm (middle) and Absolute value of detected alarms (bottom).

Haar (k:0.5)　　　　　Haar(k:2)　　　　　Db10(K:0.5)　　　　　Db10(K:2)
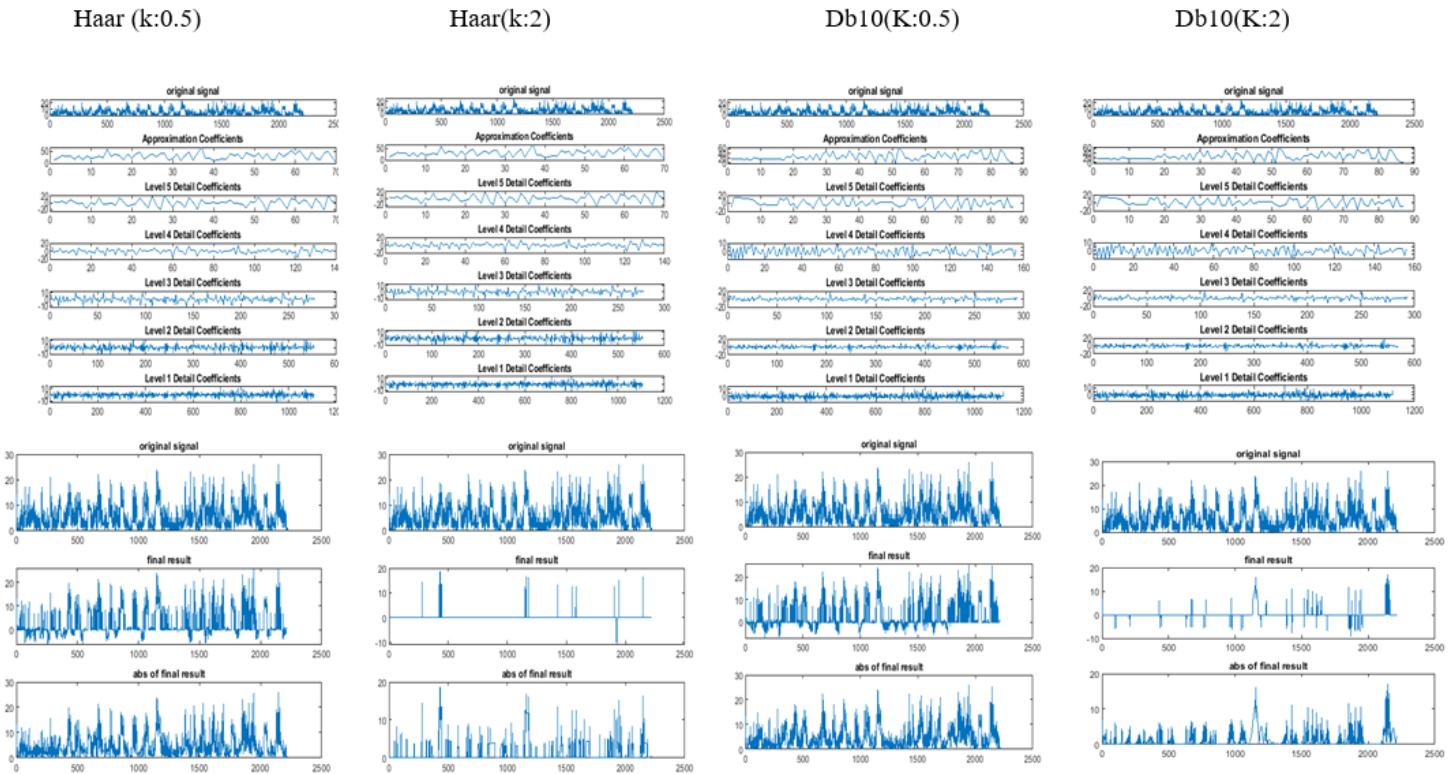
Figure 5

The result of the wavelet outbreaks detector in the detection of researcher-made outbreaks-the above row indicates the levels of decomposition using discrete Wavelet transform and the under rows indicated the original signal(above) detected alarm(middle) and Absolute value of detected alarms(bottom).