ORIGINAL RESEARCH PAPER

Statistics

SYNDROMIC SURVEILLANCE OF DENGUE USING THE EXPONENTIALLY WEIGHTED MOVING AVERAGE IN THE DISTRICT OF WESTERN MAHARASHTRA

KEY WORDS: EWMA model, Sensitivity, Specificity, PPV, NPV, FNR and FPR

Mr. Sanjay Mallinath Mulaje*

M.Sc. (Statistics), Assistant Professor, Dr. V.M. G.M. College, Solapur, Maharashtra, India. *Corresponding Author

Dr. Digambar Mahadev Zombade

Ph.D. (Statistics), Professor, Walchand College of Arts and Commerce, Solapur, Maharashtra, India.

ABSTRACT

Background: Public health officials require the immediate identification of dengue outbreaks. This study sought to evaluate the efficacy of the Exponentially Weighted Moving Average (EWMA) in identifying dengue outbreaks. **Methods:** The EWMA algorithms were utilized on daily dengue case counts to identify actual outbreaks in western Maharashtra, India, from January 2018 to December 2022. The EWMA model values were calculated for λ ranging from 0.1 to 0.9, and the Upper Control Limit (UCL) was determined for each EWMA model with a fixed k of 2 and σ equal to 1. The efficacy of EWMA models was evaluated through Sensitivity, Specificity, Positive Predictive Value, Negative Predictive Value, False Positive Rate, and False Negative Rate. **Results:** Of the various algorithms, EWMA with λ= 0.9 exhibited the most effective performance in detecting outbreaks from 2018 to 2022. The EWMA 9 (λ = 0.9) attains the optimal equilibrium among EWMA 1 to EWMA 9, exhibiting a sensitivity of 80.2%, specificity of 100%, positive predictive value of 100%, and negative predictive value of 90.9%. **Conclusion:** Using different EWMA settings (λ and σ) showed that models with higher smoothing parameters EWMA 9, had improved sensitivity and specificity.

INTRODUCTION:

India has seen a steady increase in dengue cases over the past ten years. Both dengue and malaria are dangerous diseases spread by mosquitoes that are most common in tropical or subtropical regions of the world. These conditions constitute serious health issues in India, accounting for 74.37% of the country's total burden of vector-borne diseases[1]. Dengue is driven by complex interactions among the virus, host, and carrier that are influenced by meteorological factors[2]. A serious viral disease, dengue can cause life-threatening complications. In India, as in numerous other global regions, advancements in dengue prevention and control, overseen by the National Vector-Borne Disease Control Programme (NVBDCP), have been obstructed by inadequate funding and staffing, restricted availability of rapid diagnostic tests at the point of care, and ineffective mosquito population control strategies. A coordinated initiative by the government to establish enduring surveillance systems in India, improve the diagnosis of dengue and associated viruses, innovate in mosquito vector control, and facilitate the development and assessment of dengue vaccines could substantially transform the global strategy for dengue management. This special edition on 'Dengue in India' aims to convene Indian researchers to examine contemporary studies regarding the dissemination and effects of dengue in India, encompassing the disease's overall impact, the varieties of dengue viruses present, the proliferation of mosquito vectors, and the economic ramifications of dengue in India, in addition to evaluating advancements in dengue vaccine development and the implementation of novel strategies being formulated in India[3].

Rapid responses to health incidents, especially those impacting global public health, are a primary concern for public health officials. This focus encompasses the primary methods and algorithms for monitoring public health across temporal and spatial dimensions. Syndromic surveillance systems can utilize three strategies to evaluate the effectiveness of outbreak detection algorithms: real data testing, fully synthetic simulation, and semi-synthetic simulation. Kelly J. Henning (2004) suggested that real data testing provides the highest reliability in comparison to alternative assessment methods[4]. The efficient identification of dengue epidemics is essential for public health monitoring systems. Moreover, limited research has investigated the application of statistical techniques in monitoring dengue, specifically in the Western District of Maharashtra, India.

This study assessed the efficacy of EWMA in detecting dengue epidemics in the Western District of Maharashtra, India, from January 2018 to December 2022.

MATERIAL AND METHODS

The researcher obtained data on all registered cases of dengue in the district of western Maharashtra, India, from January 2018 to December 2022 from primary health centers. In addition, the gold standard of dengue outbreak occurrence was also obtained from the District Health Office. We enrolled aggregate data of 256 dengue cases. EWMA was applied to reported dengue cases to detect outbreaks during the study period in the district of western Maharashtra, India.

Outbreak Detection Method:

The EWMA statistics is defined by the following recursive equation,

$$EWMA_t = \lambda * Y_t + (1-\lambda) * EWMA_{t-1}. \qquad (1)$$

Where Y_t equals the number of reported cases of dengue at period t, the λ is the weighting parameter that has been considered as 0.1 for EWMA 1, and 0.2 for EWMA 2. So on, an estimation of the upper control limit of EWMA was described elsewhere [5].

- The Upper Control limit (UCL) is a threshold value in the EWMA model that determines when there will be the existence of dengue outbreak
- The EWMA upper control limit (UCL) given by

$$UCL = \mu + k\sigma\sqrt{\frac{\lambda}{(2-\lambda)}} \qquad (2)$$

where μ= process mean, λ= control factor (parameter), σ= process SD, k= control limit factor

Using UCL values, no. of days having excess dengue cases were determined by comparing with official values. The term Model validity refers to the extent to which the model accurately measures infectious disease outbreaks. Validity has six measurements - Sensitivity, Specificity, False positive rate (FPR), False negative rate (FNR), Positive predictive value

(PPV), and Negative predictive value (NPV) [6].

Table 1. Performance Measurement Using The 2 X 2 Table

Model outbreak	Official outbreak		Total
	Yes	No	
Yes	a	b	a + b
No	c	d	c + d
Total	a + c	b + d	a + b +c +d

From **Table 1**, the performance indicators were defined as below:-

Sensitivity: The probability of alarm given that an outbreak occurs i.e. it measures the true positive rate of outbreak detection by the model.

Sensitivity (true positive) = [a / (a + c)] *100

Specificity: The probability of no alarm given that an outbreak does not occur i.e. it measures the true negative rate of no outbreak.

Specificity (true negative) = [d / (b + d)] *100

False Positive Rate (FPR): It is defined as the proportion of days corresponding to an alarm in the absence of an outbreak.
FPR= [b/ (b + d)] *100

False Negative Rate (FNR): It is defined as the proportion of days corresponding to no alarm in the presence of an outbreak.

FNR= [c/ (a + c)] *100

Positive Predictive Value (PPV): The probability that an alarm is truly an outbreak

PPV= [a/ (a + b)] *100

Negative Predictive Value (NPV): The probability of no outbreak given that no alarm is generated.
NPV= [d/ (c + d)] *100

The EWMA model values were compared with officially declared outbreak values. The performance measurement indices Sensitivity, Specificity, Positive predictive value (PPV), Negative predictive value, (NPV), False positive rate (FPR), and False negative rate (FNR) were computed. These indices were expressed in percentages.

RESULTS:

Table 2. Sensitivity, Specificity, PPV, And NPV Of EWMA Model For The Period 2018 To 2022

Model	Parameter	K	SD	Sensitivity	Specificity	PPV	NPV	FPR	FNR
EWMA 1	λ=0.1	K=2	σ =1	65.1 %	67.05 %	50.0 %	79.2 %	32.9 %	34.9 %
EWMA 2	λ=0.2	K=2	σ= 1	72.1 %	69.4 %	54.4 %	83.1 %	30.1 %	27.9 %
EWMA 3	λ=0.3	K=2	σ =1	75.6 %	75.9 %	61.3 %	86%	24.1 %	24.4 %
EWMA 4	λ=0.4	K=2	σ= 1	79.1 %	84.1 %	71.6 %	88.8 %	15.9 %	20.9 %
EWMA 5	λ=0.5	K=2	σ =1	81.4 %	87.6 %	76.9 %	90.3 %	12.4 %	18.6 %
EWMA 6	λ=0.6	K=2	σ= 1	81.4 %	94.1 %	87.5 %	90.9 %	5.9%	18.6 %
EWMA 7	λ=0.7	K=2	σ= 1	80.2 %	95.3 %	89.6 %	90.5 %	4.7%	19.8 %
EWMA 8	λ=0.8	K=2	σ= 1	80.2 %	99.4 %	98.6 %	90.9 %	0.58 %	19.7 %
EWMA 9	λ=0.9	K=2	σ= 1	80.2 %	100 %	100 %	90.9 %	0%	19.8 %

Table 2 assesses the efficacy of nine distinct EWMA model algorithms in forecasting dengue outbreaks, analyzed from 2018 to 2022. The key parameters of each EWMA variant differ, specifically the smoothing constant (λ) and the standard deviation (σ). The differences influence their sensitivity and selectivity in detecting data changes. Table 2 presents the EWMA models 1 through 9, featuring smoothing parameter λ values ranging from 0.1 to 0.9. Moderate sensitivity is observed, ranging from 65.1% to 81.4%. There was an increasing trend in sensitivity from λ=0.1 to 0.6, followed by a slight decrease in sensitivity values. The same result was noted for specificity, ranging from 67% to 100%. A trend of increasing specificity was noted from λ= 0.1 to 0.9. Maximum specificity was recorded at λ=0.9 (100%). As λ increases, both the positive predictive value (PPV) and negative predictive value (NPV) demonstrate consistent improvement. The analysis indicates that the EWMA 9 model (λ = 0.9) demonstrates the most favourable balance among the EWMA1 to EWMA9 models, exhibiting a sensitivity of 80.2%, specificity of 100%, positive predictive value (PPV) of 100%, and negative predictive value (NPV) of 90.9%.

Early detection is prioritized in most dengue surveillance efforts, as timely intervention is essential. Therefore, the EWMA 9 is typically preferred as it balances sensitivity and specificity while maintaining responsiveness to early indicators of outbreaks. The EWMA 9 model prioritizes the reduction of false alarms. Table 2 presents the False Positive Rate (FPR) and False Negative Rate (FNR) for nine EWMA algorithms utilized in dengue surveillance data spanning from 2018 to 2022. This study evaluates the performance of nine EWMA models, characterized by their smoothing factor (λ), control limit (K), and standard deviation (σ). A low false positive rate minimizes false alarms, while a low false negative rate guarantees accurate detection of dengue outbreaks. The values of FPR and FNR exhibit a decreasing trend as λ increases from 0.1 to 0.9. In comparing the EWMA 1 and EWMA 9 models, the EWMA 9 model exhibits no false detection of outbreaks, while it fails to identify 18.6% of actual outbreaks during the study period. Upon evaluation of the outcomes, it is observed that the EWMA 9 (λ = 0.9, σ = 1) is the most balanced model, effectively minimizing false alerts while successfully identifying the majority of true outbreaks. The Area Under the Curve (AUC) of the EWMA for dengue outbreaks from 2018 to 2022, as derived from the ROC curve, was 92.3%.

DISCUSSION:

Dengue, a vector-borne disease, is a prominent climate-sensitive illness in Maharashtra, posing substantial public health challenges. The National Centre for Vector Borne Diseases Control report on the Dengue situation in India from 2019 to 2024 indicates that Maharashtra consistently reported high dengue-related mortality, with particularly elevated figures noted in 2023 [2]. Recent dengue outbreaks underscore the necessity for timely predictions to assist health authorities in preparing and responding effectively, thereby aiming to mitigate the spread and impact of dengue.

This study focuses on the early detection of dengue outbreaks to facilitate timely public health interventions. Syndromic surveillance was adopted in place of traditional methods, such as laboratory tests and epidemiological surveillance, as it utilizes real-time health data to provide immediate analysis and feedback to authorities involved in investigating and managing potential outbreaks [8]. Syndromic surveillance employs techniques to monitor health indicators at both individual and population levels, which can be detected prior to a definitive diagnosis [11]. Epidemiological surveillance refers to the continuous and systematic collection, analysis, and interpretation of health data aimed at describing and monitoring health events [5]. Laboratory testing serves to establish a diagnosis and contributes to the understanding of disease dynamics [4].

In dengue research, the statistical methods employed include Poisson regression, Negative Binomial Regression, Autoregressive Integrated Moving Average (ARIMA), Generalized Additive Modelling (GAM), Cumulative Sum (CUSUM), and Exponentially Weighted Moving Average (EWMA) [15][3]. In 2000, a study by Chen P et al. introduced an effective two-step statistical model combining GAM and EWMA to predict dengue outbreaks, utilizing data from 2012 to 2017 in Singapore. They additionally determined that the EWMA chart is more efficacious than the CUSUM chart for detecting moderate and substantial shifts in a monitored process, which is vital for identifying dengue outbreaks. In light of the favorable results associated with EWMA for the early detection of dengue outbreaks, as established by prior research, we undertook a study to examine the efficacy of integrating EWMA with syndromic surveillance for analyzing dengue outbreaks in Maharashtra, India.

Abat C et al. (2016) [1] identified low specificity as a limitation of the syndromic surveillance system. To tackle this limitation, our study design included six EWMA variants to assess the effectiveness of various parameter settings in identifying dengue outbreaks. Through the manipulation of the smoothing parameter (λ) and standard deviation (σ), our objective was to investigate the influence of various configurations of the EWMA model on the sensitivity and specificity associated with the detection of dengue outbreaks. The findings of our study indicate that EWMA 3(λ =0.3, σ = 1) strikes an optimal balance by effectively reducing false alarms (FPR) and enhancing the detection of dengue outbreaks (low FNR). No single algorithm has been proven to consistently identify all possible outbreaks. Consequently, assessing the effectiveness of various algorithms is crucial for comprehending their advantages and drawbacks [12].

The findings of our study indicate that utilizing EWMA in syndromic surveillance holds significant potential. To gain a deeper insight into the relative performance of EWMA, future studies could explore comparisons with other algorithms. The findings indicate that the EWMA model demonstrates effectiveness when utilized with the retrospective data. Nonetheless, given that retrospective analysis might not completely capture the model's performance in real-time, high-pressure scenarios, it would be beneficial to evaluate the EWMA model's capacity to provide dependable results in a real-time context for continuous dengue monitoring. This study focuses on the western district of Maharashtra, which may limit the applicability of the findings to other areas that have different epidemiological and environmental factors. Future studies may implement the EWMA model in the context of syndromic surveillance for dengue epidemics to enhance its validation and applicability.

CONCLUSION:

This study illustrates the potential of the Exponentially Weighted Moving Average (EWMA) statistical model in identifying dengue outbreaks, especially in the western Maharashtra district of India. Nonetheless, this resulted in a trade-off regarding sensitivity, with increased specificity occasionally leading to a greater false negative rate. In particular, although the EWMA 9 model achieved the most favourable equilibrium between the false positive rate and false negative rate, none of the EWMA models demonstrated the capability to reliably identify all outbreaks with effectiveness. This highlights the necessity of exploring various methods to enhance the detection of dengue epidemics.

Financial support and sponsorship: Nil

Conflicts of interest: There are no conflicts of interest.

Table 3. Area Under The Curve And Its Significance Of Different EWMA Models For The Period 2018 To 2022

Model	Area	Std. Error	Asymptotic Significance	Asymptotic 95%vConfidence interval	
				Lower Bound	Upper Bound
EWMA 1	0.720	0.032	0.000	0.658	0.782
EWMA 2	0.809	0.027	0.000	0.757	0.862
EWMA 3	0.872	0.022	0.000	0.829	0.915
EWMA 4	0.913	0.018	0.000	0.878	0.948
EWMA 5	0.938	0.014	0.000	0.910	0.967
EWMA 6	0.959	0.011	0.000	0.938	0.981
EWMA 7	0.973	0.008	0.000	0.956	0.989
EWMA 8	0.982	0.007	0.000	0.969	0.995
EWMA 9	0.989	0.004	0.000	0.981	0.998

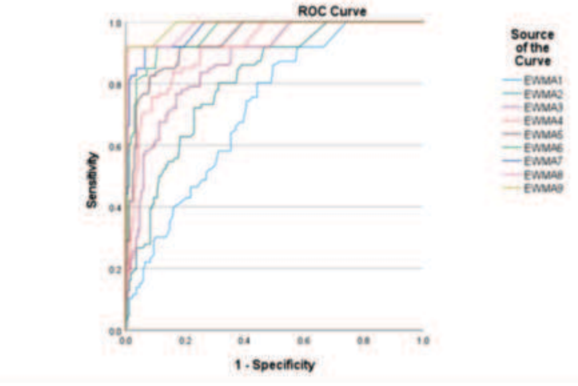

Fig. 1 The area under the curve for different EWMA models

REFERENCES:

[1] Abat C, Chaudet H, Rolain JM, Colson P, Raoult D. Traditional and syndromic surveillance of infectious diseases and pathogens. Int J Infect Dis. 2016 Jul;48:22-8. doi: 10.1016/j.ijid.2016.04.021. Epub 2016 Apr 30. PMID: 27143522; PMCID: PMC7110877.

[2] Centre Vector Borne Diseases Control report. Dengue situation in India. Dengue Cases and Deaths in the Country since 2019. https://ncvbdc.mohfw. gov.in/.

[3] Chen P, Fu X, Ma S, Xu HY, Zhang W, Xiao G, Siow Mong Goh R, Xu G, Ching Ng L. Early dengue outbreak detection modelling based on dengue incidences in Singapore during 2012 to 2017. Stat Med. 2020 Jul 10;39(15):2101-2114. doi: 10.1002/sim.8535. Epub 2020 Mar 30. PMID: 32232863; PMCID: PMC7318238.

[4] Dengue in Pune City, India (2017–2019): a comprehensive analysis. Front. Public Health, 19 September 2024. Sec. Infectious Diseases: Epidemiology andPrevention Volume 12 - 2024 | https://doi.org/10.3389/fpubh.2024. 1354510.

[5] Douglas N. Klaucke, M.D.* James W. Buehler, M.D.* Stephen B. Thacker, M.D.** R. Gibson Parrish, M.D.** Frederick L. Trowbridge, M.D.*** Ruth L. Berkelman, M.D.* and the Surveillance Coordination Group. Guidelines for evaluating.

[6] J.M. Lucas, M.S. Saccucci. Exponentially weighted moving average control schemes: properties and enhancements Technometrics, 32 (1) (1990), pp. 1-12.

[7] K. Park. Park's Textbook of Preventive and Social Medicine, Jabalpur, M/s Banarsidas Bhanot Publishers, 2023, Twenty-seventh Edition, pg. 154-155.

[8] Kelly J. Henning. Overview of Syndromic Surveillance. What is Syndromic Surveillance? MMWR September 24, 2004 / 53(Suppl);5-11.

[9] M. Karami. Validity of evaluation approaches for outbreak detection methods in syndromic surveillance systems, Iran J Public Health, 41 (11) (2012), pp. 102-103.

[10] Mandal, B., & Mondal, S. (2024). Unveiling spatio-temporal mysteries: A quest to decode India's Dengue and Malaria trend (2003-2022). *Spatial and Spatio-TemporalEpidemiology*, *51*, 100690. https://doi.org/10.1016/j.sste. 2024.100690.

[11] Mandl KD, Overhage JM, Wagner MM, Lober WB, Sebastiani P, Mostashari F, Pavlin JA, Gesteland PH, Treadwell T, Koski E, Hutwagner L, Buckeridge DL, Aller RD, Grannis S. Implementing syndromic surveillance: a practical guide informed by the early experience. J Am Med Inform Assoc. 2004 Mar-Apr;11(2):141-50. doi: 10.1197/jamia.M1356. Epub 2003 Nov 21. PMID: 14633933; PMCID: PMC353021.

[12] Manoochehr Solgia, Manoochehr Karamia, Jalal Poorolajal. Timely detection of influenza outbreaks in Iran: Evaluating the performance of the exponentially weighted moving average. Journal of Infection and Public Health 11 (2018) 389–392.

[13] Ministry of health and family welfare government of India, State action plan on Climate change and human health, Maharashtra.

[14] Mutheneni SR, Morse AP, Caminade C, Upadhyayula SM. Dengue burden in India: recent trends and importance of climatic parameters. Emerg Microbes Infect. 2017 Aug 9;6(8):e70. doi: 10.1038/emi.2017.57. PMID: 28790459; PMCID: PMC5583666.

[15] Salim, N.A.M., Wah, Y.B., Reeves, C. et al. Prediction of dengue outbreak in Selangor Malaysia using machine learning techniques. Sci Rep 11, 939 (2021). https://doi.org/10.1038/s41598-020-79193-2.

[16] Shet A, Kang G. Dengue in India: Towards a better understanding of priorities and progress. Int J Infect Dis. 2019 Jul;84S:S1-S3. doi: 10.1016/j.ijid.2019.03. 008. Epub 2019 Mar 12. PMID: 30877040.