

An Improved EWMA-Based Method for Outbreak Detection in Multiple Regions

Sesha K. Dassanayaka* and Joshua French

Mathematical and Statistical Sciences, University of Colorado Denver, Denver, CO, USA

Objective

To propose a computationally simple and a fast algorithm to detect disease outbreaks in multiple regions

Introduction

Emerging disease clusters must be detected in a timely manner so that necessary remedial action can be taken to prevent the spread of an outbreak. The Exponentially Weighted Moving Average method (EWMA) is a particularly popular method, and has been utilized for disease surveillance in the United States [1].

A spatio-temporal EWMA statistic is proposed for on-line disease surveillance over multiple geographic regions. To capture spatial association, disease counts of neighboring regions are pooled together, similar to a method originally proposed by Raubertas [2] for a different control chart. Also to increase statistical power in testing multiple EWMA statistics simultaneously, false discovery rate (FDR) is used instead of the traditional family-wise error rate (FWER).

Methods

First, an acceptable false alarm rate is set by the user to define the false discovery rate. Then, at each time t , disease counts for each of the m regions $Y_{1t}, Y_{2t}, \dots, Y_{mt}$ are collected; the weighted counts of immediate neighbors are pooled to form regional neighborhoods with counts $Y'_{1t}, Y'_{2t}, \dots, Y'_{mt}$. Then, the corresponding EWMA statistics for the regional neighborhoods $E'_{1t}, E'_{2t}, \dots, E'_{mt}$ are calculated.

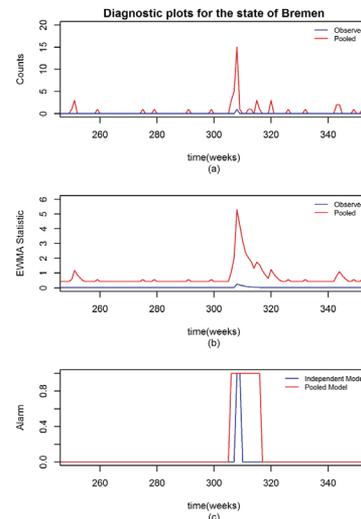
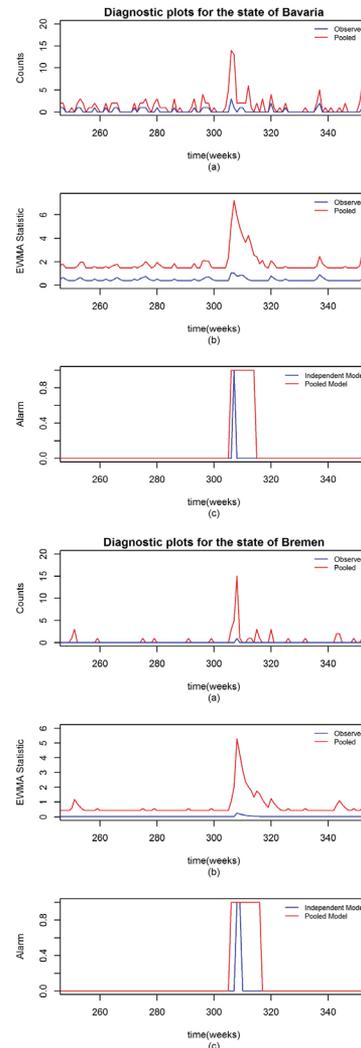
To construct empirical in-control distributions for each region, B bootstrap samples are drawn with replacement, respecting spatial order, from an initial time period with no outbreaks. For the bootstrap samples, the corresponding EWMA statistics are computed for each region to determine the empirical in-control distributions from which the corresponding p-values $p'_{1t}, p'_{2t}, \dots, p'_{mt}$ are calculated. Finally, a state-of-the-art multiple comparison procedure is used to determine the alarms with the pooled model. This model is compared to a baseline model using independent regional counts using a standard multiple comparison procedure. Simulation studies provide strong evidence that the pooled model using the more powerful and current multiple testing procedure detects outbreaks faster than the independent model using the standard multiple testing procedure.

Results

The proposed method was applied to a data set of Salmonella Newport cases reported weekly from 16 German federal states between years 2004-2014. The first two years of data (2004-2005) were used to estimate the in-control distribution in each state since there were no unusually high disease counts reported from any of the states during this period. Plots of the disease counts (a), the EWMA statistics (b), and the corresponding alarms (c) are shown for two states Bavaria (Figure 1) and Bremen (Figure 2); the blue lines depict the independent model and the red lines show the pooled model. Both plots illustrate the rapid detection ability of the proposed method.

Conclusions

The proposed method of pooling regional neighborhood counts increases the speed of detection compared to the baseline model using independent regional counts. More statistical power can be gained using a more innovative multiple testing procedure.



Keywords

Spatio-temporal EWMA charts; biosurveillance; prospective disease surveillance; control charts; false discovery rate

References

1. Elbert Y, Burkom, H. S. Development and evaluation of a data-adaptive alerting algorithm for univariate temporal biosurveillance data. *Statistics in Medicine*. 2009 November; 28(26)
2. Raubertas R.F. An analysis of disease surveillance data that uses the geographical locations of the reporting units. *Statistics in Medicine*. 1989 March; 8(3)

*Sesha K. Dassanayaka

E-mail: sesha.dassanayaka@ucdenver.edu

