



## NRC Publications Archive Archives des publications du CNRC

### **Detecting outbreaks in time-series data with RecentMax**

Carter, Dave; Martin, Joel D.

For the publisher's version, please access the DOI link below./ Pour consulter la version de l'éditeur, utilisez le lien DOI ci-dessous.

#### **Publisher's version / Version de l'éditeur:**

<http://doi.org/10.5210/ojphi.v7i1.5779>

*Online Journal of Public Health Informatics, 7, 1, 2015*

#### **NRC Publications Record / Notice d'Archives des publications de CNRC:**

<http://nparc.cisti-icist.nrc-cnrc.gc.ca/eng/view/object/?id=26726661-96b6-4580-8d40-2327b5070c>

<http://nparc.cisti-icist.nrc-cnrc.gc.ca/fra/voir/objet/?id=26726661-96b6-4580-8d40-2327b5070cd1>

Access and use of this website and the material on it are subject to the Terms and Conditions set forth at

<http://nparc.cisti-icist.nrc-cnrc.gc.ca/eng/copyright>

READ THESE TERMS AND CONDITIONS CAREFULLY BEFORE USING THIS WEBSITE.

L'accès à ce site Web et l'utilisation de son contenu sont assujettis aux conditions présentées dans le site

<http://nparc.cisti-icist.nrc-cnrc.gc.ca/fra/droits>

LISEZ CES CONDITIONS ATTENTIVEMENT AVANT D'UTILISER CE SITE WEB.

**Questions?** Contact the NRC Publications Archive team at

PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca. If you wish to email the authors directly, please see the first page of the publication for their contact information.

**Vous avez des questions?** Nous pouvons vous aider. Pour communiquer directement avec un auteur, consultez la première page de la revue dans laquelle son article a été publié afin de trouver ses coordonnées. Si vous n'arrivez pas à les repérer, communiquez avec nous à PublicationsArchive-ArchivesPublications@nrc-cnrc.gc.ca.



# Detecting Outbreaks in Time-Series Data with RecentMax

Dave Carter\* and Joel D. Martin

National Research Council Canada, Ottawa, ON, Canada

## Objective

To develop an algorithm for detecting outbreaks of typical transmissible diseases in time series data that offers better sensitivity and specificity than the CDC EARS C1/C2/C3 algorithms while offering much better noise handling performance.

## Introduction

We implemented the CDC EARS algorithms in our DADAR (Data Analysis, Detection, and Response) situational awareness platform. We encountered some skepticism among some of our partners about the efficacy of these algorithms for more than the simplest tracking of seasonal flu.

We analyzed several flu outbreaks observed in our data, including the H1N1 outbreaks in 2009, and noted that, using the C1 algorithm, even with our adjustable alerting thresholds, there was an uncomfortable number of false alarms in the noisy steady-state data, when the number of reported cases of flu-like symptoms was less than five per day.

We developed an algorithm, RecentMax, that could offer better performance in analyzing our flu data.

## Methods

We developed the RecentMax algorithm based on a simple observation: that a good alert for an outbreak tends to be on days (or, more generally, in time slices) where the number of cases today is noticeably larger than on any recent day. For example, intuitively, if the number of cases of flu today is more than twice the greatest number of cases observed on any single day in the last three weeks, it seems reasonable to generate an alert.

RecentMax takes two parameters: a window of recent history (e.g., ten days), and a threshold factor (e.g. 1.5 $\times$ ), and generates alerts when the data observed for the current time slice is greater than the maximum observed in the history by the threshold factor, subject to the confidence interval (a tuneable probability of false positive threshold, typically set to 95% or higher, assuming a gaussian distribution around the threshold value).

## Results

Our RecentMax algorithm does a commendable job alerting at the onset of the observed outbreaks.

Furthermore, RecentMax does a better job handling the noisy steady-state behaviour outside of outbreaks, offering intuitive alerts when the number of reported cases starts increasing meaningfully. The H1N1 outbreak of the fall of 2009 is illustrated using real-world data in the accompanying graphs; red bars highlight days on which CDC EARS C1 would generate alerts (at 95% and 99% confidence intervals) and days on which RecentMax would generate alerts given a 10-day history window and a threshold factor of 1.5 $\times$ .

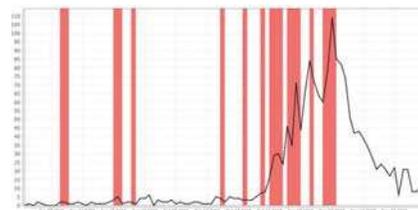
An interesting side-benefit of RecentMax is that, as an outbreak spreads and the number of cases increases somewhat linearly, the algorithm is less likely to generate alerts; that is to say, once an outbreak has been observed and many alerts have been generated (and presumably validated), the algorithm tends to stop alerting (unless the number of cases is increasing exponentially). Thus, once public health

personnel have been notified of an outbreak in its early phases, the algorithm tends to not keep alerting, thus not telling recipients what they already know.

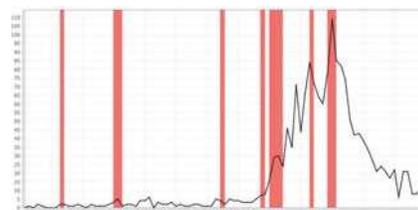
RecentMax avoids problems occasionally observed in algorithms built around a decay series. RecentMax examines the recent history without providing extra weight to more recent data, and is thus unaffected by noise in the immediately preceding time slice that can confuse some algorithms unduly (e.g., when data increases steadily for six consecutive days and then drops to zero on the seventh as the facility is closed for a holiday).

## Conclusions

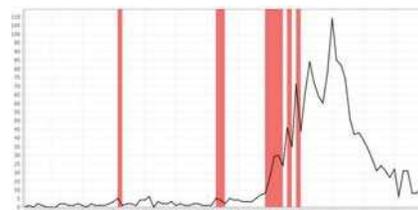
RecentMax offers a compelling alternative to the CDC EARS algorithms for detecting outbreaks of transmissible disease in time-series data.



Flu outbreak with alerts generated by CDC EARS C1 at 95% confidence threshold



2009 H1N1 outbreak with alerts generated by CDC EARS C1 at 99% confidence threshold



2009 H1N1 outbreak with alerts generated by RecentMax

## Keywords

aberration detection; algorithm; outbreak detection; surveillance; situational awareness

\*Dave Carter

E-mail: david.carter@nrc-nrc.gc.ca

