Author's response to reviews

Title: Recursive Least Squares Background Prediction of Univariate Syndromic Surveillance Data

Authors:

Amir-Homayoon Najmi (najmi@jhuapl.edu)
Howard Burkom (Howard.Burkom@jhuapl.edu)

Version: 3 Date: 11 September 2008

Author's response to reviews: see over
MS: 6964558602091996
Recursive Least Squares Background Prediction of Univariate Syndromic Surveillance Data
Amir-Homayoon Najmi and Howard Burkom

To: BioMed Central

I would like to submit the revised manuscript as well as replies to comments by both reviewers.

Yours sincerely,

Dr. A. H. Najmi
M.A. (Cantab.), D.Phil. (Oxon.)
najmi@jhuapl.edu
Tel: 443-778 3320
Response to Reviewer’s Report
MS: 6964558602091996

Reviewer: Karen Olson

Minor Essential Revisions:

- Spelling on figure 10 has been corrected and.
- The term “significant” has now been omitted because it has statistical implications. I would like to explain further what has been done here. A method for background estimation has been introduced. A comparison has been made by injecting a known multiplicative signal into 15 syndromic time series (each time series is approximately 4 years long). The performance (embodied in figure 10) shows the probability of detection for a fixed false alarm rate of 1 per week, which is an operationally important value. We feel that figure 10 indicates that the new method is superior to the conventional method (specially at lower signal values). A test of significance requires far more data sets than we have been able to process for this small scale study and so we have to postpone computation of the distribution functions of the detection probabilities to the time when we have a large number of realizations for each syndrome (we have studied 15 syndromes).

Discretionary Revisions:

- the phrase mentioned by the reviewer has been rewritten.
- The correct indices for j should be 1,…,7, as the reviewer had noticed. This was a mistake and has been corrected.

A. H. Najmi, M. A. (Cantab.), D.Phil. (Oxon.)
najmi@jhuapl.edu
Response to Reviewer’s Report

Reviewer: Rich Tsui

Revisions:

We would like to answer the questions in the reverse order (numbers correspond to the reviewer’s numbering of the revision/questions):

6. We are a little confused as to what the reviewer means here. The methods section only describes the methods. The results sections describes the results in the form of a detection performance which is the only evaluation we have conducted. We do not think we need to add any statements about the evaluation since it has been explained in the results section.

5. The method relies on the specific shape of the curves in figure 2. In fact our original method consisted of fitting 3 straight lines that best matched the shape of these curves (three lines with different gradients for the three distinct sections). This, however, is time consuming since two program loops must search for the position of the lines. In practice, we have found that the middle 50% (which is used in the “equalization” method) suffices. It may be ad-hoc, but it works just as well as the first method!

4. Each of the 7 lines is used to construct a line fitting the middle 50% of each data. There are then 7 straight lines, which when extended to the extremes of the x-axis produce the 7 left hand points, and the 7 right hand points. The median of the 7 left hand points is then connected to the median of the 7 right hand points: This is L[m].

3. We wanted our method to be insensitive to the nearest few days of the day whose background is being estimated, so that the initial rise of the type of signal we are considering would not affect the estimate. We also wanted the method to be consistent with the W2, which has a seven day delay built into it. It is possible to do better (not by much though!) using fewer days, but then that would raise questions about the comparisons with the W2 method. Thus we settled on a 7-days-ahead prediction.

2. Standard ROC curves as described in most Communications and Signal Processing textbooks [e.g. Reference 9 of our paper] plot Pd vs Pfa (or Pfa vs Pd). Alternatively, they plot Pd and Pfa vs threshold. Given the context of our data
(which has a lot in common with what we have been doing for the US Navy signal processing tasks in the last 25 years) we transform the Pfa to FAR (False Alarm Rate). This has immediate operational significance, instead of Pfa which would have to be transformed to FAR to have real meaning in a real system.

1. In the present paper we are estimating the background, and then compare how well we have estimated the background by performing a threshold detector based on a multiplicative signal model. Our results indicated that our background estimation improves the false alarm probabilities (as a function of threshold) which show as “better” ROC curves (Pd vs FAR). We are not presenting a new detection algorithm; merely a background estimation method. The detection method is based on a threshold detector in both cases on the day of the maximum signal amplitude (3 days from the onset in our simulations). Timeliness is not, therefore, of relevance in the present context. If one is interested in the performance at an earlier date, then the problem is equivalent to one with a lower value of the alpha parameter, which is exactly what we have done! We have performed the detection and the comparison at 4 different signal maxima: 25%, 50%, 75% and 100%. The Results section has been expanded and the paragraph containing equation (7) describes the injection method more clearly than before, and footnote (3) has also been added as further clarification.

A. H. Najmi, M. A. (Cantab.), D.Phil. (Oxon.)
najmi@jhuapl.edu