Author's response to reviews

**Title:** Syndromic Surveillance: STL for Modeling, Visualizing, and Monitoring Disease Counts

**Authors:**

Ryan P Hafen (rhafen@purdue.edu)
David E Anderson (danders2@xula.edu)
William S Cleveland (wsc@purdue.edu)
Ross Maciejewski (rmacieje@purdue.edu)
David S Ebert (ebertd@purdue.edu)
Ahmad Abusalah (aabusala@purdue.edu)
Mohamed Yakout (myakout@cs.purdue.edu)
Mourad Ouzzani (mourad@cs.purdue.edu)
Shaun J Grannis (sgrannis@regenstrief.org)

**Version:** 2  **Date:** 6 March 2009

**Author's response to reviews:** see over
Response to Reviewer Comments for
Syndromic Surveillance: STL for Modeling, Visualizing, and Monitoring Disease Counts

We would like to thank the reviewers for their time and helpful comments. We have addressed the reviewers concerns interlacing our responses to their questions/comments in the red text below.

Referee 1

• Are limitations of the work clearly stated?
  Yes, fairly well in the Discussion. However, the C1, C2, and C3 methods and GLM methods are not the only reasonable options to which the STL method might be compared. For example, BioSense uses a GLM-type model with weekly updates of coefficient estimates which somewhat mitigates issues such as non-constant seasonal effects. Perhaps the Discussion could mention other options to which STL might be compared. I like the STL method, but don’t feel so ready to declare it a winner simply because other options for comparison could have been considered. If the authors agree that other options might be competitive, perhaps add some type of “limitations of the work” to the Discussion section. This is a good point and we have added some discussion to this effect.

• Minor essential revisions:
  1. Clearly define the C1, C2, and C3 methods - the description on p. 11 is too brief.
     We have added the descriptions.

  2. On pages 7 and 11, the issue re pooling across years-perhaps cite Burr et al.[10], which showed that a hierarchical model that allows for varying seasonal effects across years led to simulated
data that was more appropriate for comparing algorithms that the “one-season-fits-all” model used by all other papers that I’m aware of (including [16]).

We added a cite on page 5, where we first discuss pooling.

3. Is the STL acronym ever defined? Also, loess is a good smoother, among many. Is there any compelling reason to prefer loess over, say, lokerns (in R) in this context? I note the “blending” required near the ends of the data with loess. Also, how were the bandwidths chosen? Trial and error and visual inspection of resulting plots? Some “magic” 39 day, 90 day, and 1000 day bandwidths are mentioned.

We updated the abstract and intro with the STL acronym explained.

The lokerns function in R yields extremely similar results as loess when span is forced to be constant. A simulation study revealed that lokerns also has a problem with the variance increasing at the endpoints of the data, regardless of whether the fixed-bandwidth or adaptive bandwidth is used. The variance increases even more drastically than for loess. Thus using lokerns would not mitigate the blending issue.

We point out at the top of page 6 that the parameters were chosen through “extensive use of visualization and numerical methods of model checking.” We also mention in the discussion that others should perform their own model validation and parameter selection. There is nothing magic about the parameters that we report other than they worked well for our data.

4. Page 3: should “seasonal ARIMA” and “naive Bayesian classifiers” be defined? Consider audience, but ARIMA is never defined.

We modified the paper to define the ARIMA acronym, but did not describe the method in detail since it is only mentioned and not used in our paper. We removed “naive Bayesian classifier”.
It is not important for the readers to understand how CoCo works, but just to know that we used it to classify the chief complaints.

5. STL method description: our library doesn’t have J. of Official statistics. Is there another published description? Aim to make the page 5 and 6 description clear enough that readers could implement their own STL-like approach.

There is no other write-up. A complete description of how the method works would not be feasible in this paper. We have made well-documented R code available, and have modified the paper to point this out.

6. Page 6: “The systematic components $T_t$, $S_t$, and $D_t$ are taken to be fixed...” That is not quite what you mean? Obviously each are time dependent, but presumably you mean fixed for a given value of $t$. This is slightly confusing because of course one must estimate $T_t$, $S_t$, and $D_t$, so some of the observed variability in $N_t$ arises because you therefore deal with ESTIMATES of $N_t$.

Not sure how best to clarify this, but perhaps a sentence to that effect could be added. In practice, we always then have ESTIMATED residuals, and keeping that fact in mind is sometimes useful.

This has been clarified.

7. Page 6: quantiles- 1 - $P(y_t, \lambda_t)$ - I think you really mean $P(Y\geq y)$ and note this implies using “$\geq$”, not “$>$” here - so I think you have a small issue in notation.

Fixed

8. Lognormal: write the density and define parameters? I suggest this because there is not a standard notation/parameterization for lognormal, so you should define the parameterization that you’re using.

Done

9. Rho is 97th quantile, but then you let $\rho = 0.01$ and I think you mean the 99th quantile. Check for consistency of meaning of $\rho$.

Emphasize: 3% FAR PER DAY! This is a high rate per day isn’t
it? Why chosen? And why then use 0.01 in the study of GLM versus STL actual versus nominal FAPs. Consider using terms such as “actual” and “nominal” in this context because that’s pretty standard terminology.

3% daily is a high rate, but it is not uncommon to see this in the literature. For example, see [24], in which this rate is used for a wide-scale comparison of methods. Since the thresholds are set empirically, we cannot allow the rate to get too small or the threshold values begin to be questionable due to some outliers in the data. We have changed the actual vs. nominal FAR study to 0.03 as well to avoid confusion.

10. Page 9. why repeat Eq. (2)?
We have removed the second equation.

11. Typos:
- Page 4: “The square root...” not “Square root...”
  Fixed
- Page 4: Fix this: “This is not surprising because the square root of a Poisson....” You don’t really mean what this sentence says - I think you’re trying to say that the sq. root scale stabilizes variance, and you will empirically check whether the resulting ESTIMATED residuals from fitting Eq. (1) are approximately Gaussian. However, the residual mean is nearly 0, which is “small,” so the sentence needs to be clarified.
  This has been clarified. This sentence is not referring to the N_t, but to the raw counts being Poisson, which after transformation behave like Gaussian, so that therefore the residuals are Gaussian too.
- Page 7: “Because 0 is: fixed” fix this and what is N in the 0.87N expression? And use, “The parameter f controls the...
  These have been fixed. N should have been O.
- Page 10: “Table ?” - just above Table 2 - I guess you mean “Table 2”
• Page 11 - some repetition here - why say “the two dangers ....” again (first said this on page 3)
  We have removed the repetition.

• Figures: some misleading labels-such as sq. root of counts when plotted values are negative
  We have changed labels to clarify. The square root daily count is the units.

• References: inconsistent format-all caps for article titles or not?
  Fixed

• Discretionary Revisions:
  As mentioned in ...6 above, I like STL, but am not convinced that it would “win” compared to other methods not considered, such as frequent updating of fitted GLM-type models, or perhaps other alternatives. I suspect the authors like STL, as do I, but I am not convinced that the reader should come away with “Use STL” just yet. Aren’t there other reasonable alternatives that could also beat C1, C2, C3 a
  We have added to the discussion. We do not want to give the message “use STL only”, but we certainly do want to give the message “use STL” - especially because of its visualization capabilities. Only further use and study will determine its merit.

Referee 2

Minor essential revisions:

1. My main comment relates to the lack of practical implementation detail provided (that analyses were performed using R) given the detailed process of model fitting. If the authors wish the work to be considered by the managers of public health systems (last sentence of Conclusion), more implementation detail would be useful including specifying the R packages or original (annotated) R code used for model fitting and monitoring processes, which can be provided as an additional file.
The stl and loess functions we use are included in the base distribution of R. We have created an R package that demonstrates the methods in this paper and we will provide it as an additional file.

Some additional minor revisions are also suggested:

1. The authors include elements of discussion throughout the results section which works well. Are there any recommendations for setting up fitting processes to help ensure the components model different variation in the data (para 2 page 6) if this might be a specific issue with the method? Some additional discussion which integrates this work with related surveillance literature would be useful, and acknowledgement of the study limitations could be expanded.

2. Page 10 paragraph 3 clarify 10% outperformance.
   Done.

3. Suggest include the results from Table 2 in an additional column in Table 1 as STL(90) or similar, which will allow the presentation format to be the same as the other methods for easier comparison and timeliness data to be included.
   Done.

4. I assume the daily refitting process is quite fast?
   It is very fast – only a fraction of a second.

5. Text corrections: missing table numbers, “of of”, possibly revise the term “presentation” on page 4 para 2, potential subheading inserted after Figure 2, page 9 formula repeated from page 6.
   These have been addressed.