Reviewer’s report

Title: Prediction of Dengue Outbreaks Based on Disease Surveillance, Meteorological and Socio-Economic Data

Version: 1 Date: 11 Sep 2017

Reviewer: Oliver J Brady

Reviewer's report:

In this manuscript, Jain and colleagues extensively detail the evaluation of a range of statistical models for dengue early warning in Thailand using historical data from the Thailand ministry of Health. The analysis and results are extensively described, but I have concerns about the readability of the article for wider public health audiences as much of the paper is dedicated to technical evaluations and little text is dedicated to how such models may be used to control dengue outbreaks (how much lead time could they give, what is the incremental value of collecting extra meteorological data in real time?, etc). I also have some concerns about the approach used, particularly with regards to overfitting highly flexible models to the training dataset and the resultant poor out of sample predictive value. There are definitely some novel approaches taken in this analysis, including the use of DLN models of recent cases, but currently I feel that this paper falls short of either a practical public health solution or a detailed methods evaluation.

Major comments:

Background: existing EWS literature not reviewed and not clear how this study aims to move this forward

Methods: Many methods to protect against over-fitting of highly flexible machine-learning type approaches have not been employed here. In particular I saw no mention of:

* Checking for co-linearity of covariates
* Adjustment of critical p values- given the large number of hypothesis tests with numerous lags, etc p < 0.05 does not seem appropriate
* AIC used for model selection (a more complex model will always explain more deviance, but is it more parsimonious?)

There is clearly a need for this as out of sample predictive performance was considerably worse than within sample and this should be discussed.

Discussion: no limitations of the data, approach or utility of the final product are discussed

Most forecasting approaches compare new models against a null model of seasonal averages, e.g. monthly mean of the last 5 years, how well to each of these models do against this null model in out of
sample validation?

Minor comments:
Background p1: is all this detail needed?

Methods: fig 3 and 4- remove the phrase "seems like".

Methods- define "S" in eqn. 4

The manuscript could use some language proof editing to correct a small number of grammatical and spelling issues.

Results: The results section could be considerably shortened with the over description of graphs and tables removed and only a few key findings discussed (the reader should be able to interpret the former from appropriate use and presentation of figures).

Figure 11 seems to suggest that models a and b were evaluated over different (longer) time periods than those that included dengue case terms (c, d, and e). It would seem more logical to compare them only over the same time periods to reduce potential bias.

Are the methods appropriate and well described?
If not, please specify what is required in your comments to the authors.

Yes

Does the work include the necessary controls?
If not, please specify which controls are required in your comments to the authors.

Yes

Are the conclusions drawn adequately supported by the data shown?
If not, please explain in your comments to the authors.

Yes

Are you able to assess any statistics in the manuscript or would you recommend an additional statistical review?
If an additional statistical review is recommended, please specify what aspects require further assessment in your comments to the editors.

I am able to assess the statistics
Quality of written English
Please indicate the quality of language in the manuscript:

Needs some language corrections before being published

Declaration of competing interests
Please complete a declaration of competing interests, considering the following questions:

1. Have you in the past five years received reimbursements, fees, funding, or salary from an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

2. Do you hold any stocks or shares in an organisation that may in any way gain or lose financially from the publication of this manuscript, either now or in the future?

3. Do you hold or are you currently applying for any patents relating to the content of the manuscript?

4. Have you received reimbursements, fees, funding, or salary from an organisation that holds or has applied for patents relating to the content of the manuscript?

5. Do you have any other financial competing interests?

6. Do you have any non-financial competing interests in relation to this paper?

If you can answer no to all of the above, write 'I declare that I have no competing interests' below. If your reply is yes to any, please give details below.

I declare that I have no competing interests

I agree to the open peer review policy of the journal. I understand that my name will be included on my report to the authors and, if the manuscript is accepted for publication, my named report including any attachments I upload will be posted on the website along with the authors' responses. I agree for my report to be made available under an Open Access Creative Commons CC-BY license (http://creativecommons.org/licenses/by/4.0/). I understand that any comments which I do not wish to be included in my named report can be included as confidential comments to the editors, which will not be published.

I agree to the open peer review policy of the journal