COMMENTS FOR THE AUTHOR:

Editor Comments:

I'm afraid the quality of the English used throughout your manuscript does not currently meet our requirements, as there are several spelling and grammatical errors throughout.

Answer: Thank you for the comment. We asked a native English speaking colleague to help us copyedit the paper.

After the list of abbreviations, please add the heading "Declarations".

Answer: Thank you for the comment. We added the heading “Declarations”

Reviewer reports:

Oliver J Brady (Reviewer 1): 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 colleting 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.

Your references have not been included correctly in the text (they are currently appearing as ?). Could you please ensure this is corrected.

There are some language errors in the methods section (at the beginning). We would recommend that you revise this.