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.

**Response to reviewers’ comments**

We are very grateful for the reviews provided by the editors and each of the external reviewers of this manuscript. The comments are encouraging and the reviewers appear to share our judgement that this study and its results are important. Please see below, in blue, our detailed response to comments.

**Reviewer comment:** Background: existing EWS literature not reviewed and not clears how this study aims to move this forward.

Authors’ Response: Thank you for the comment. We did some more research behind the factors that create the possibilities of dengue outbreaks and how addressing them is significant to create an effective EWS. As a result, a dependency chart for the occurrence of dengue is added as Fig 1. We also detail how this study aims to move forward the development of more advanced EWS.   
  
**Reviewer comment:** 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:

Authors’ Response: Thank you for the comment. We checked for the co-linearity of covariates. The following paragraph was added.

High correlation among predictor variables may give rise to singularity problems when fitting a statistical model. However, for GAMs, checking for collinearity is not sufficient. Since, we are now fitting smooth functions; it should be determined whether the smooth function of one variable can be produced using a combination of the smooths of the other terms in the model. This is called checking for concurvity. We performed the concurvity check for all the predictor variables.

* Adjustment of critical p values- given the large number of hypothesis tests with numerous lags, etc p < 0.05 does not seem appropriate.

Authors’ Response: Thank you for the comment. In Generative Additive models, the p values are of the null hypothesis of a zero effect of the indicated spline. We used p <0.05 because that is typically used for the threshold of statistical significance. We understand the relative weakness of using P≈0.05 as a threshold and the proposals that arise from time to time to lower it to 0.005 or below that. However, due to lack of a common consensus we refrained from using it. However, in light of your comments authors agreed that results for which 0.005 < p < 0.05 would instead be called “suggestive” rather than statistically significant. But during analyses, gladly we did not find such cases. All the p-values were observed to be much lesser than 0.005. However, for the sake of readability we continue to use the conventional p-value (0.05) in this paper. If the reviewer thinks we should lower it to say 0.005 or some other appropriate value, we would be glad to do so. Again, many thanks for the feedback.

* AIC used for model selection (a more complex model will always explain more deviance, but is it more parsimonious?)

Authors’ Response: Thank you for the comment. We compared the models using change in AIC. The results are shown in Table 1.

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

Authors’ Response: Thank you for the comment. We agree that out of sample predictive performance was worse than within the sample. But we beg to differ that this was the case of overfitting. As shown in the paper, the Thailand had an unexpected dengue outbreak in year 2015. The year saw rapid increase in dengue incidents without a parallel in last 20 years. Thus, the characters of the validation dataset found less representation in the training dataset. However, the binary system of classifying months into outbreaks and non-outbreaks as proposed in the paper worked quite well in this case. We can of-course improve on accuracy by adding more data and collecting a more diverse data. Collection of diverse data such as House index (HI), Breteau index (BI), Container index (CI) and integrating the information from social media platforms to track the dengue incidences in real time will likely better the prediction. We also aim to develop customized models for each individual district that includes demographic data, data from government surveys and abovementioned additional features at a more granular level.

**Reviewer comment:** 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).

Authors’ Response: Thank you for the comment. We removed Fig 7, parts of Fig 10 and Fig 11 to combine their relevant parts into a separate Figure. The result is shown in new Fig 9. The Fig 11 b representing the model (F) was included in the list of plots shown in new Fig 11. Fig 14 a and b are removed and only 14c is kept to show the evaluation on external dataset. The change is reflecting in Fig 12 in which we compared the model against both constant outbreak threshold and moving outbreak threshold. Fig 8b was replaced with Fig 16 and Fig 9 got removed.

TL; DR: Fig 7, Fig 9, Fig 14(a-b), Fig 12 and Fig 16 were removed. Fig 8b and Fig 10 were updated. These changes changed the numbering order.   
  
**Reviewer comment:** 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.

Authors’ Response: Thank you for the comment. We did as suggested. The prediction performance is compared over the same time periods (months 23-60) to reduce the potential bias. This period is chosen because the final model uses lagged dengue data of past 23rd month (as explained later).

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

Authors’ Response: Thank you for the comment. We have added the limitations of data and approach in the Conclusion.

**Reviewer comment:** 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?

Authors’ Response: Thank you for the comment. We added a “Seasonal Naïve Model” in which forecast is set to the dengue count data of same month in the previous year. That is considered as base model. Also, we added a binary classification of months as outbreaks (or epidemic), or non-outbreaks with respect to the constant outbreak threshold and WHO’s moving outbreak threshold.

**Reviewer comment:** Minor comments: Background p1: is all this detail needed?

Authors’ Response: Thank you for the comment. We have shortened the p1 and kept only the content which is relevant to the paper.

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

Authors’ Response: Thank you for the comment. We removed “seems like”.   
  
**Reviewer comment:** Methods- define "S" in eqn. 4

Authors’ Response: Thank you for the comment. We defined “S” at suitable place. S represents the count data of dengue incidence occurred in the surrounding districts.   
  
**Reviewer comment:** The manuscript could use some language proof editing to correct a small number of grammatical and spelling issues.

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

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

Answer: Thank you for the comment. We have corrected it.

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

Answer: Thank you for the comment. We have revised it.