Reviewer's report

Title: Spatiotemporal clustering, climate periodicity, and social-ecological risk factors for dengue during an outbreak in Machala, Ecuador, in 2010

Version: 3
Date: 26 August 2014
Reviewer: Alex Cook

Reviewer's report:

The authors present some interesting analyses of spatial and temporal dengue case data in a small city in Ecuador. I like what they've done and recommend it be published, though I have some suggestions for how they could improve the work.

Minor

2.6 Authors listed alphabetically: alas, it seems that something went wrong with the ordering, for MJB comes before EBA.

6.1 I think this could do with some references to support some of the claims made.

6.16 'An EWS', and henceforth

6.23 GIS 'based' maps. Generally there are quite a lot of little typos like this in the m/s. A proof read by a native anglophone might pick these up.

7.23 'skill' is almost certainly the wrong word!

9.2 Better to present 95%CIs rather than SEs. Saves the reader a little mental arithmetic.

10.1-3 Readers won't be able to reproduce this without clarification.

10.13 This is a little unorthodox and might need a stronger justification.

11.12 Surely you can only test whether they were random and try to show they are not random? I don't see how you can get a sampling distribution under the hypothesis that they are non-random.

12.1 'each hypothesis was represented as a suite of variables'---I don't understand this at all, sorry. An hypothesis is usually something like 'this variable is not associated with the outcome', or 'none of this collection of variables is associated with the outcome'. How can a suite of variables by itself be an hypothesis?

12.7-8 There is no need to put the "(GLM...logit)" part in. You've already said it's a logistic regression.

12.12 Since you don't present AICs, I think you can drop line 12, and then simplify line 13. I'm pretty bad at algebra but what I got was AICc=[2nk]/[n-k-1] -2log(L). By the way, L should really be the maximum of L over the parameter space. It is the optimum of the log likelihood function, not the entire function,
which goes into AIC and friends.

13.1-2 Please clarify, what exactly is a 'significant temporal scale' or 'significant period'?

13.17 What is the [0,1] doing here?

13.21-23 I feel you need to clarify what this actually *means*. It may be ok when you address the earlier point on the same issue.

14.4-9 This really should go in the results or figure captions. Methods should just contain something like 'We identified phase angles between pairs of variables, corresponding to GIVE INTERPRETATION, using HOW YOU DID IT, for different time points and periodicities.'

14.14 'greater than 155 cases' This is an odd choice of phrase/number. What is the significance of 155 cases and why is it interesting that 42 neighbourhoods fell above that threshold?

14.18 I don't think I understand this sentence, sorry.

14.21 You claim that the top ranked model is a 'better fit' than the model with everything in it. This cannot be! The model with everything in it must by definition fit as well or better than a model with fewer things in it. Whether the global model is better or more predictive is a different kettle of fish.

15.9-11 Better not to use computerese like BI_12 or popdens.

15.12-21 I thought this paragraph is not going to help people without a background in signal processing understand what you've done. If precise interpretations could be provided, it would help. So, a 'strong and significant signal for the 1-2 year periodic band for dengue incidence' means what? That if dengue incidence is high now, it will be higher than expected for the next 1-2 years? Or 1-2 years from now it will be high again? How do you reconcile the '1-2 year' or '0.5-1 year' periodic bands which you identify in lines 14 and 15, with the '1 year' and '2 year' bands you talk about in lines 18 and 19? If these refer to the same thing, then shouldn't the intervals be retained? And if you have period bands at 0.5-1 year and 1-2 years then is this the same as 0.5-2 years, or is there some additional nuance to interpretation that I'm missing?

16.2-4 Missing main clause here

16.4-16 Again I think most readers will need more help with this. Your work is interesting but if the reader can't understand, it's wasted.

17.17-9 Question: to what extent are cases in Machala infected at home vs elsewhere in the community, eg at work or school? Is this known (or estimable)?

19.5 'most important' in what sense?

19.10 omit first comma

20.23-21.6 This part seems out of place. First, I don't think they do actually test hypotheses using AIC or variants. Second, the AIC itself is also arbitrary so it's a bit odd to state that p-values are arbitrary and imply that AIC isn't. In any case, the thresholds for AIC were determined from 'standard' alpha levels from
hypothesis testing (a challenge left to the reader is to consider nested hypotheses, for which the LRT and AIC can both be used, and work out where the delta_AIC <2 threshold comes from). I think you could drop this part without any real loss to your paper.

21.8 I'm pretty sure the AIC/glmmulti combination also leads to estimates that are biased away from 0. You'd need to introduce some kind of penalty to the objective function when estimating parameters to get around that---eg with lasso or a prior.

References: ref 38 is missing a journal. I suggest the authors go through each reference carefully to check for other errors.

Table 1: I'd omit the computerese '4pplbedrm' etc---it's no longer the 1980s! Population density is missing a number

Table 2: Again, omit computerese. What is the outcome variable here? How to interpret the estimate? Suggest omit intercept, make the column be OR or similar effect size with 95% CI in brackets, omit the Std [SE?], possibly omit VIF unless you think it's important. Caption needs more detail.

Fig 1: change avg to mean to avoid abbreviation? Can some of the figures be merged into more multipanel figures?

Fig 3: what are the dark dashed lines?

Fig 7-9: I don't understand the years *at all*. Years are intervals in time, but they've been assigned to specific points on the x axes. So what are the points in between? Say for 2003 and 2004, half way between these is what?

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

'I declare that I have no competing interests'