To-do List

In this section, we put editor or reviewer comments in italics following a number. Our to-do items follow bullet points and are in plain-text.

Associate Editor:

This is a very detailed and well written manuscript, which develops a bivariate spatiotemporal model for ozone and PM10 measurements from 24 stations across Mexico City. The manuscript is impressively thorough in considering how the model can be used to predict pollution alerts based on the current policies for phase alerts. The model is relatively complex, and there should be more demonstration as to whether this complexity improves prognostic performance. In particular,

1. *Have the authors investigated what advantage, in terms of prediction, there is in modelling both PM10 and O3 jointly? Intuitively, there may be limited predictive advantage if data on both processes are relatively complete.*

   As per reviewer 2's comments, what does the spatial aspect of the model add if no out of sample (new site) predictions are made? Section 3.3 (Model Selection) could be extended to address these points.

   - I can fit combinations of spatial/non-spatial and joint/independent models to explore the benefit. I think this will require fitting three more models.
   - I'm hoping the spatial model has slightly better predictive performance because then we can justify the model without the goal of spatial interpolation. It may be better because it captures spatial structure. If it does have better predictive performance, then that is probably sufficient justification. We'll see....

     But if it doesn't, then I think we drop space. I'm not sure I want to dig into spatial prediction in this paper. I think it would be interesting, but I'm not sure that we can do it concisely. They have already asked us to shorten the paper, and I think this would require important details. That our current paper doesn't need to address.

2. *I would recommend that to do this the paper should be shortened elsewhere; Section 1 in particular is a little too long and could use some judicious shortening and editing.*

   - Trim the intro and perhaps other sections too.

3. *Spell out IW and IG terms for distributions.*

   - Do it

4. *I agree with reviewer 2 that the authors should provide details of code and where others can access it, to allow partial reproducibility. Are the data themselves available to allow full reproducibility?*

- Check to see if we can share the data
- Post the data
- Comment code

Reviewer 1

This paper employs a bivariate spatiotemporal model to model hourly ozone and PM10 in Mexico City. The predictions from this model were then used to determine Mexico City's pollution emergency phases and also to assess compliance with Mexican ambient air quality standards(MAAQS).

The paper appears to be a good application paper which is generally very pleasant to read. It has a clear structure and the contents are well detailed and explained, and the results are clearly presented. I encourage to publish it after addressing a few minor comments as below.

1. *Page 5, I am also interested in more details about the missing data, e.g. the proportion of it, whether both pollutants were missing at the same stations quite often.*

    - I'll figure this out.
    - I think we need to be careful about this so that we don't open this up to a big discussion about data.

2. *Page 7, as stated, the equation (7) uses covariates from the previous hour. In this case, whether the (c) in Figure 2 should plot the correlation with lag-1 for covariate-outcome, because this plot is used to inform modelling decision.*

    - I can update that plot and references to it in the text.

3. *Page 8, line 33, it will help if a few explanations of choosing those prior distributions, e.g. informative or not? Especially for the elements in matrix A_ψ, line 39 specifies a_{12} as a Gaussian distribution which is much different with the IG for diagonal elements a_{11} and a_{22}. More words on this are needed.*

    - Explain more about this prior selection

Reviewer 2

This paper proposes an explicit space-time bivariate model for hourly pollution levels in Mexico City. It then assesses probabilities of compliance with respect city and national level air pollution standards in Mexico City and Mexico. The paper has been written very well with a great deal of attention devoted to details with respect to legislative compliance. The space-time multivariate models have been motivated well using exploratory data analysis and the

validation and other predictive results have been written up to a very high level of satisfaction of a statistics referee. Hence this manuscript will potentially be a great contribution to Series A. I have found only a few minor issues with the manuscript.

1. *No spatial interpolation: I was surprised to find that there is almost no spatial interpolation mentioned in the paper. The prediction equations (4) detailed predictions only at the data sites indexed by i' not in new ones, say i'. The authors' model is fully spatial and the covariates (RH and TMP) can be outputted to a fine grid (of few meters) hence I believe the models can be easily extended to perform out of sample predictions. Agree that, at least superfluously, the authors can argue that they do not require those out of sample predictions. However, it can also be argued that the chief reason for proposing a space-time model is to facilitate out of sample predictions otherwise, one can simply propose a 24 dimensional model for the observations from 24 sites! Somewhat consequently, the manuscript fails to tackle the issue of providing predictive spatial patterns in compliance in areas other than the 24 sites. I believe,this opportunity has been missed here.*

    - I'm not sure what they mean by outputting this onto a fine grid since this isn't part of the model and this is only observed at monitoring stations.
    - We can certainly mention that spatial interpolation would allow us to assess compliance at unmonitored locations; however it would require a different model than what we are using. I think there are two approaches
        - Do what they ask for and lengthen the paper by two or three pages. We'll need to add several plots for this. In addition, we'll have to discuss how we do prediction another way.
        - I'd prefer to just acknowledge that this is possible, very briefly discuss what we would need to do this, and mention it in the future work. Along with this, I think we can mention that we have been asked to shorten the paper and don't know if we could do that while adding such a significant inferential goal. This is also something that we considered doing; however, we decided against it. I wonder if that is worth mentioning in the response to the reviewers.

2. *Further clarification regarding lagged observations: In model 1, generic lagged observations have been used without explicitly stating how far into the past the lags can go. This is a bit important since if the lags go beyond the beginning of time 1, then would those lags be treated as fixed input into the model? For the other lags, the distribution of Lit will contribute towards the likelihood function. The computational details in the Appendices A and B do not clarify this.*

    - I think I can do this relatively briefly.

3. *Software implementation: The authors are to be commended for working out the full conditional distributions in Appendices A and B. However, journals nowadays require verifiable code hence it would be worthwhile to publish implementable code, preferably in R. Could a publicly available package such as spBayes or STAN implement the*

model and the prediction procedure? Perhaps, I am running a bit into the editor's realm but this issue would need addressing as it would help 'operationalize' the model as noted in Section 5, Conclusions and Future Work. Perhaps, spTimer can be used as well, although univariate modelling can only be performed there. It may interersting to compare the model validation results from univariate and bivariate models, especially in the light of authors results that the more complicated but realistic heteroscedastic models are not able to outperform the humble homoscedastic models! Overall, this is a very interesting paper that makes a novel contribution to both statistics and society. Further slight improvements should make it a worthwhile Series A article.

- Do they want us to use a standard package or just provide code?
- I bet you could do this in Stan (it would probably be slow – and I think you might need to trick Stan to use IAR models), but I'd rather just provide our code. It's in R, using Rcpp.
- I think we should add a line explaining that, although these models are homoscedastic on the sqrt and log scale, they are not actually homoscedastic on the true scale of the data.
- I think we will address their model validation concerns by discussing the first reviewers concern.
