Response to the reviewers, JRSS-SA-0135, "Pollution State Modeling for Mexico City"

We very much appreciate the very positive words of the reviewers and the associate editor, e.g., "I encourage to publish it after addressing a few minor comments" and "...this manuscript will potentially be a great contribution to Series A" and "Overall, this is a very interesting paper that makes a novel contribution to both statistics and society." We have addressed all of the comments below and also in the text and we hope you will now accept the paper for publication.

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.

   - We appreciate these comments. While the model is not terribly complex - linear, with two covariates, some autoregressive structure and some joint spatial dependence - it is fair to ask whether adding the spatial modeling shows benefit relative to the models with only the two predictors and autoregegressive structure and, if so, is there further benefit to making the two processes dependent. Our specification makes it particularly easy to study both questions. For the first, we only need to remove the ψ's from the model. For the second, we only need to set $a_{12}^{(\psi)} = 0$. We have now added these comparisons in addition to the temporal lag comparisons as Tables 2 and 1, respectively, in the supplementary material (in order to shorten the text). Remarkably, there is very little out-of-sample difference in predictive performance across all of the models. Nonetheless, since the dependence spatial model is always preferred to the independence spatial model and since the dependence coefficient, $a_{12}^{(\phi)}$ is significantly positive, we present results for the former.

- We again emphasize that the intent of this paper is to model and predict pollution alerts and emergencies. Since the prescription for such declarations is entirely based upon what is observed at the collection of 24 monitoring stations across the five subregions, there is no interpolation goal here. If the goal was to develop exposure surfaces over Mexico City say in different time windows, we would adopt a geostatistical model which is intended for spatial prediction.

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.*

    - We agree that the Introduction was too long and that, in places the text was too detailed. After some careful surgery we have shortened the entire manuscript by three pages.

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

    - We now clarify the IW and IG abbreviations with their first usage.

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?*

    - We have attached provided a link to a GitHub directory file that contains all code used for this analysis. In this directory, we provide a readme file that explains what each program does and give the link to where these data are publicly available.

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.*

   - The missing data issue is a bit of a challenge here. We received the file for 2017 with no missingness and no information regarding missingness. In the linked GitHub directory we include a readme file which provides the link to where these data are publicly available. The data is collected at minute scale and averaged to hourly scale. Imputation for missingness is as described in the text, by using hourly data from the nearest station. Again, we work at hourly scales in order to align with the decision-making for declaration of alerts.
   - Furthermore, in principle, if we could attempt model-based imputation, i.e., treating the data as missing in the model fitting and did iterative updating through Gibbs sampling looping of missing data given parameters and observed data followed by parameters given missing data and observed data, under the autoregressive specifications we employ, it is very likely that such imputation would differ little from the imputation which was carried out.

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.*

   - We completely agree and now done in Figure 2.

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.*

   - We have added some words to the text regarding prior specification. However, it is important to note that all of our priors are customary choices for convenience of conjugacy and that all of our conjugate priors are very weak. With regard to the entries in the matrix A, we recall that it is a Cholesky decomposition of a covariance matrix. Therefore, the diagonal elements must be positive while the off-diagonal term can be positive or negative. This accounts for the Inverse Gamma priors for the former and the normal prior for the latter. We now note this in the text as well.

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.*

   - We must emphasize that the intent of this paper is to model and predict pollution alerts and emergencies. Since the prescription for such declarations is entirely based upon what is observed at the collection of 24 monitoring stations across the five subregions, there is no interpolation goal here. If the goal was to develop exposure surfaces over Mexico City say in different time windows, we would adopt a geostatistical model which is intended for spatial prediction.
   - It is fair to ask about the performance of a 24 dimensional model for the observations for the 24 sites. In fact, that is what we supply. In this regard, the issue becomes whether adding the spatial modeling shows benefit relative to the models with only the two predictors and autoregegressive structure and, if so, is there further benefit to making the two processes dependent. Our specification makes it particularly easy to study both questions. For the first, we only need to remove the ψ's from the model. For the second, we only need to set $a_{12}^{(}\phi) = 0$. We have now added these comparisons in addition to the temporal lag comparisons

as Tables 2 and 1, respectively, in the supplementary material (in order to shorten the text). Remarkably, there is very little out-of-sample difference in predictive performance across all of the models. Nonetheless, since the dependence spatial model is always preferred to the independence spatial model and since the dependence coefficient, $a_{12}^{(\phi)}$ is significantly positive, we present results for the former. several plots for this. In addition, we'll have to discuss how we do prediction another way. 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.*

   - We appreciate this comment and have edited the text accordingly. Typically, with auto-regressive modeling, we either (i) start the time series with say a distribution for the missing initial observations and then multiply by the likelihood for the observed data given these observations or (ii) we start the model after a sufficient time period so that all of the needed lagged data has been observed, treating the lagged data as fixed. Here, since have the data for December 2016, we know the lags and can start at the first hour of January 1, 2017, treating the lags as fixed.

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.

- We appreciate this comment and agree that reproducibility is important. To this end,we have created a git repository `https://github.com/philawhite/Pollution_state_modeling_code` with all code used in the analysis. In addition, we include a readme file that explains what each program does and how the data are obtained. We would be happy to submit this to the journal too but had some issues with this. We have even included code for heteroscedastic models that are referenced in the paper.
- We also recognize that spBayes and spTimer cannot fit the models that we propose. We didn't explore using Stan for this analysis; however, there is a case study that may be applied to this setting `http://mc-stan.org/users/documentation/case-studies/icar_stan.html`.
