Responses to Referee 3 (R3) Comments on acp-2018-1296 “Evaluating wildfire emissions projection methods in comparisons of simulated and observed air quality”

R3 Interactive Comment: The manuscript discussed the model evaluation of wildfire emissions air quality estimates with observations and compared two different approaches with traditionally used NEI emissions inventory in the PGM modeling. The manuscript is in the scope of the ACP. The novelty of this manuscript is evaluating the dynamical and statistical down scaling approaches that estimates the wild fire emissions. However the manuscript need to address few major comments and strengthen their findings with some more clarifications. Should also consider reorganizing the results such that it is easy for the reader to follow and understand.

Major Comments:
1) The author used dynamical and statistical approaches to estimate wildfire emissions, a brief discussion of how these wildfire emissions vary when compared to NEI will be useful to understand the differences in the emissions among three approaches. Comment on the uncertainty in wildfire emissions estimates from all three approaches. 2) It is important to mention how the dynamical and statistical meteorological variables performed when compared with meteorological observations. Discussing this is critical to understand the uncertainties/differences among these two approaches. 3) Some of the information presented is irrelevant to the manuscript key focus, so worth moving them to supplementary doc or removing them (mentioned in specific comments)
Response: We thank the reviewer for the positive comment on the overall scope of the paper, and address the major comments below.

1) The comparison of the dynamical and statistical approaches to estimating wildfire emissions is the subject of a paper that is already published, the emissions evaluation of Shankar et al. (2018), which also compares their effects on the underlying annual area burned (AAB) projections described in Prestemon et al. (2016) upon which these wildfire emissions estimates are based. The Discussion section of Shankar et al. (2018) specifically addresses the uncertainties in these two approaches. Both of these studies are extensively cited in this paper, and therefore these comparisons were not repeated here, but only referenced in the context of explaining the air quality model results. In this current work we do discuss the uncertainty in the NEI wildfire emissions estimates due to the use of satellite data in estimating daily area burned, given the large number of small fires occurring in the Southeast, and cite the relevant reference (Soja et al. (2009). See also our response to Reviewer 1 (in reference to p. 12, lines 14-16). The use of SMARTFIRE, which ingests MODIS fire detects in the estimates of *daily* areas burned is the main distinction between the 2010 NEI wildfire emissions estimation approach and the two downscaling approaches, which disaggregate statistical estimates of *annual* area burned to daily values using a stochastic model.
2) Meteorological differences in the two downscaling approaches are discussed extensively in the Discussion section of Shankar et al. (2018). It should be noted that these differences directly contribute only to the AAB estimates underlying the wildfire emissions estimates in each case, and affect the air quality simulations described in the current work only indirectly. In this paper we use the same daily meteorological inputs (from the Weather Research and Forecasting model) across all three model simulations for the 2010 simulations. They are therefore not a source of intermodel differences

between the statistical and dynamical downscaling approaches, and thus a detailed evaluation of the meteorology inputs is not presented here.

3) Our responses to specific comments are provided below each.

R3 (in reference to p. 1, line 30): How does (sic) these compensating errors differ among these three approaches? Or were they (sic) same among all three approaches? Include a brief comment.

Response: These compensating errors occur in all three modeling approaches presented, and the intermodel differences in PM are presented within the body of the manuscript. The range of their differences is stated in the previous sentence. To suggest that there are differences among the simulations, we have reworded this sentence in the abstract.

Revised text, p. 1, line 31-32: "…acceptable $PM_{2.5}$ performance to varying degrees (MFB between -14% and 51%) in all simulations."

R3 (in reference to P.3, lines 8 -22): This paragraph can be trimmed as the main focus of this study is wild fire emissions model performance evaluation. It is not necessary to explain the approach used in Prestemon et al., 2016 here again.

Response: We have shortened this text as suggested.

Revised text, p. 3, lines 9-17: "Prestemon et al. (2016) estimated annual areas burned (AAB) over 13 states in the U.S. Southeast using county-level projections of climate, population and income, and land use based on the Intergovernmental Panel on Climate Change emissions scenarios (Nakicenovic and Steward, 2000). Their projected AAB show a small increase (4%), from 2011 to 2060 due to the combined influences of these climate and socioeconomic factors. Shankar et al. (2018) leveraged these AAB projections to estimate wildfire emissions over a Southeastern modeling grid at 12-km x 12-km spatial resolution suitable for air quality impact assessments, and projected daily wildfire emissions in selected years from 2011 to 2060. The study also compared their wildfire emissions projections to those using 19-year historical mean AAB, and found them to be lower (by 13% - 62%) than those based on the historical AAB due to the offsetting influences of socioeconomics, which decreased AAB, and climate variability, which increased or decreased AAB, in the selected years."

R3 (in reference to p.3, line 32): Specify the meteorological variables explicitly here to educate the reader what variables are considered to calculate AAB estimates.

Response: This line refers to all the prognostic meteorological variables used in an air quality simulation in addition to those used in the AAB projections. We have revised this line to include the four meteorological inputs used in the statistical models of Prestemon et al. (2016).

Revised text, p.3, line 27-29: "…meteorological variables needed in air quality modeling in addition to those used in Prestemon et al. (2016), which are monthly average daily maximum and minimum temperature, monthly average potential evapotranspiration, and monthly average precipitation. This allows for a consistent…"

R3 (in reference to p.4 line 20): Based on the small differences seen between the approaches in the results can the authors comment or tie these differences to the methodology used to estimate wild fire emissions in the three approaches.

Response: This is done later in the Discussion section at some length after the results are presented. It would be premature in this introductory section to discuss specific results of the methodological differences except to comment, as we have done here, in a very general sense on what might be expected.

R3 (in reference to p.4, line 30): Looks like the authors used old version of CMAQ model. With the latest updates in recent versions does any of these results or model predictions change significantly? Can the authors include one or two sentences about that here.
Response: The improvements expected in the model performance are discussed later in the manuscript in Sections 3.2.3 and Section 4 (Conclusions) in the context of model biases in individual PM species with respect to observations, and the relevant publications of model enhancements in later CMAQ versions are cited. We refer the reviewer to p. 18, lines 9-19 of the manuscript regarding the organic aerosol bias, and p. 18, lines 20-24 and p. 20, lines 19-21 regarding the organic aerosol model enhancements in later CMAQ versions that could improve the performance of organic carbon. Nitrate performance would improve once current efforts by the EPA to do a comprehensive evaluation of nitrate performance are completed, according to personal communication with Dr. Kelly (EPA - OAQPS).

R3 (in reference to p.5, line 5): The manuscript did not provide the spatial plots of the emission totals from these three approaches, including spatial plots will help readers understand how these three emission inventories differ spatially in the modeling domain.
Response: We include spatial plots in Figure 5 and supplementary figure S11. These are not presented as domain total emissions, but as maximum absolute differences between the statistical and dynamical d-s column emissions over the whole fire season, specifically to understand the spatial distributions of the ozone and $PM_{2.5}$ concentration differences. This seems a more useful approach than examining domain total emissions as the domain-wide differences in the ambient concentrations are quite small, especially for ozone.

R3 (in reference to p. 7, lines 13-14): The authors mentioned that they made several updates to the [AMET] tool since the initial distribution. Are there any significant updates that the modeling community should know? If so please specify them here.
Response: The enhancements to the AMET tool were not made by us, but by the AMET developers at the EPA. AMET version 1.2.1 was released to the community by the Community Modeling and Analysis System (CMAS) Center with updated documentation in July 2013. This version did contain several updates to the meteorological and air quality analysis scripts since the initial release of AMET 1.0, and added data from several additional years of observations including network data from Canada and Europe. AMET Version 1.2.2 that was used in our analyses and cited in the manuscript addressed minor bugs in version 1.2.1. Two versions of AMET have been released to the community since that time, along with user documentation.

R3 (in reference to p.7, line 30): The authors used Boylan and Russell (2006) performance goals and criteria, but did not consider recent Emery et al., 2017 performance criteria and goals, is there any specific reason? It is important to compare with most recent criteria, update the manuscript to include Emery et al., 2017 criteria and goals.
Response: We have revised the text to reflect this newer work, in p. 4, lines 23-24, and in the revised text and tables below.
Revised Text, p. 4, lines 23-24: "...we hypothesize that they will yield results within published criteria (Boylan and Russell, 2006; Emery et al, 2017) for acceptable AQ model performance with respect to observations,..."

Revised text, p. 7, lines 25-26: "We also apply the recently recommended ozone performance metrics of Emery et al. (2017) for monthly averaged NME, NMB, and correlation coefficient r, shown in Table 2 along with…"
Revised text, p. 8, lines 1-5: "…and the ozone model performance falls outside the performance criterion of NME ≤ 25% in all months, although it meets the criterion for NMB (≤ ±15%) in December – April and October, and for the correlation coefficient r (> 0.5) in every month. Although Table 2 shows that the observed monthly averaged ozone was below the recommended 40 ppb cutoff for applying these criteria in all months, we consider all of them because of the sporadic nature of wildfire impacts on air quality, especially since these are monthly averaged mixing ratios."
p. 13, lines 6-7: "… respectively) in all months. It meets the stricter criteria of Emery et al. (2017) for NME and NMB (< 50% and < ± 30%, respectively) in April – September (Table 4)."
p. 13, line 17: "…(MFE ≤ 50% and MFB ≤ ± 30%) of Boylan and Russell (2006) in April, June, August and September, and meets their performance criteria…"
p. 15, lines 4-7: "…in Tables 4, and S1-S3 in the Supplement. Total $PM_{2.5}$ does meet the criteria for acceptable PM performance of Boylan and Russell (2006) throughout the year and across all cases for MFB and MFE, and meets the performance criteria of Emery et al. (2017) for NME, NMB and r (< 50%, < ± 30%, and > 0.4, respectively) in April – September."
p. 19, line 17-19: "…and symmetry) were proposed (Yu et al., 2006), but more nuanced and species-specific performance criteria and goals have been recommended by Emery et al. (2017), and provide the guidance for our model performance analysis in the case of NO_3."
In addition, Tables S1 and S2 in the original Supplement have been relabeled as Tables 2 and 4 and moved to the manuscript. Tables 2 and 3 have been relabeled as Tables 3 and 5 respectively, and Tables S3, S4 and S5 have been relabeled as Tables S1, S2 and S3 respectively; all text referring to these tables in the manuscript has been changed accordingly. All of these tables have been modified to include the correlation coefficient r for each modeled case.

R3 (in reference to P.8, line 4): Since Fig 2 is not showing any differences between three approaches, suggest moving it to supplementary information
Response: We agree that the differences in monthly averaged ozone among the three simulations are minute. We have moved the figure to the Supplement as Figure S1.
Revised text, p. 8, lines 5-6: "…and MFE in the supplemental Fig. S1"
We have relabeled the figures in both the manuscript and Supplement accordingly, along with all references to these figures in the text. We have renumbered all supplemental figures accordingly.

R3 (in reference to p. 8, line 30): Why these four sites had some of the largest differences, explain it here.
Response: Our practice throughout the Results and the Discussion sections of the manuscript is to describe the results in some detail before exploring the reason for the results. We therefore did not explain in this sentence why these four sites show higher ozone for the statistical d-s simulation. The explanation for that feature is provided later in the same paragraph on p. 9, lines 8-18, and also on p. 11 in "Ozone Modeling Uncertainties"; the main reason is that the larger AAB for this case drives its higher emissions of ozone precursors. The spatial patterns of the maximum intermodel differences in VOC and NO_x emissions between the statistical d-s and dynamical d-s cases also show in Fig. 4 (previously Fig. 5) that peak differences in the wildfire emissions of these ozone precursors occur at the KY-OH and MO-IL borders. They are also

consistent with the spatial distributions of AAB for each downscaled case provided in Shankar et al. (2018), and cited here.

R3 (in reference to p. 10, line 12): Explain it here why there is better agreement between dynamical and NEI than statistical and NEI.

Response: Again, this seems to be an issue of style rather than substance. Throughout the manuscript we have described a result fully before discussing the possible reasons for that result. The full explanation of the closer alignment of the dynamical d-s with the NEI, and the larger difference between the statistical d-s and the NEI simulations is provided in Section 3.1.3. To address the review comment we have added a sentence about the forthcoming explanation at the end of this subsection, on p. 10, line 14-15.
Revised text, p. 10, lines 14-15: "…The reasons underlying these intermodel biases are explored in Section 3.1.3 (Discussion)."

R3 (in reference to p. 14, line 18): Discuss why the errors and bias for OC are comparably high at rural sites than urban sites.
Response: Beginning in Section 3.2.3 (Discussion), p. 18, line 7, we discuss rural vs. urban OC performance, and point out that residential combustion emissions as the most likely cause of being degraded in the rural sites. We also cite the relevant paper that investigated the quality of these emissions in the NEI (Napelenok et al., 2014) in support of this argument. To address the reviewer's comment, we have added a sentence on p. 14, line 26 about the forthcoming explanatory text.
Revised text, p. 14, line 26: "…IMPROVE (rural) sites; possible reasons for this urban-rural bias in OC are examined in Section 3.2.3 (Discussion)".

R3 (in reference to p. 19, line 29): Can larger AAB estimates in statistical approach be considered as one of the uncertainty/limitation with this approach?
Response: The conclusions of Shankar et al. (2018), which compared the two downscaling approaches to estimate wildfire emissions in detail, were that both downscaling approaches have uncertainties. The AAB could well be overestimated in the statistical d-s approach, given the consistent overpredictions of the ambient concentrations in this case, but we also explain in the penultimate paragraph under Conclusions that the dynamical d-s and the NEI also suffer from uncertainties for other reasons (overbias in precipitation in WRF, undercounting of small fires under canopy in the NEI's SMARTFIRE tool), which would result in an underestimation of their emissions.
R3 (in reference to p. 20, line 27): The authors did not clearly point out superior candidate for estimating wild fire emissions but it is worth commenting on it.
Response: Possible improvements to each approach are provided in Conclusions, starting on p. 21, line 2. It is possible that using a more recent version of the WRF model than was used in these dynamical d-s wildfire emissions estimates will align the model results more closely between the two downscaling approaches. We have added text to address this comment.
Revised text, p. 21, line 17-27: : "In correcting the biases in the two downscaling approaches a clear "winner" could emerge in the resulting air quality model performance, but it bears remembering that each approach has inherent advantages and disadvantages. The greater computational efficiency of statistical downscaling when using an ensemble of climate models allows for a richer dataset of inputs to estimate the AAB used to calculate wildfire emissions,

even if it lacks the accuracy of mesoscale meteorological modeling. It also allows for greater flexibility in selecting the years for the air quality simulations, which are themselves a source of uncertainty in the predictions, given the year-to- year variability in the quality of the emissions inventories and downscaled meteorological inputs. On the other hand, a later WRF model version would reduce uncertainties in the predicted precipitation in the dynamical d-s method. This would allow a consistent and reliable set of meteorological inputs to better estimate AAB and wildfire emissions, and simulate air quality in a given year. The disadvantage in this case is perhaps a smaller number of climate models and model years for the downscaling due to the greater computational burden imposed by a dynamic meteorological model simulation."

R3 (in reference to Table 2 and 3): Better to include even NEI benchmark results in the tables
Response: The purpose of these tables (relabeled Tables 3 and 5) is to illuminate the relationship between the large intermodel differences in the two downscaling methods seen at particular locations and dates in the domain, and the area burned either in, or upwind of, that grid cell. This is the main driver of the emissions estimates in these two methods. As our focus is not on improving the NEI emissions, we are not including the NEI in these tables.

R3 (in reference to Figures 3, 4, 8, and 9): The statistical metrics box font should be increased, not clearly visible.
Response: As there will likely be changes to plots required during the production editing of the paper once it is accepted, we will address this issue at that time as required.

R3 (in reference to Figures 8, and 9): Explicitly mention what top, middle, bottom panels refer to.
Response: We have addressed this comment in the figure captions.
Revised caption, Figure 8 (relabeled Figure 7): "Comparisons of each pair of wildfire emissions methods for $PM_{2.5}$ ($\mu g\ m^{-3}$) predicted at grid cells containing both Interagency Monitoring of PROtected Visual Environments (IMPROVE) monitors and wildfires in 2010. Top: Statistical d-s vs. NEI benchmark; middle: dynamical d-s vs. NEI benchmark; bottom: statistical d-s vs. dynamical d-s. Top and middle right: September; bottom right: October."
Revised caption, Figure 9 (relabeled Figure 8): "Seasonal comparisons of each pair of wildfire emissions methods for $PM_{2.5}$ ($\mu g\ m^{-3}$) predicted at grid cells containing Interagency Monitoring of PROtected Visual Environments (IMPROVE) monitors and wildfires in 2010. Top: Statistical d-s vs. NEI benchmark; middle: dynamical d-s vs. NEI benchmark; bottom: statistical d-s vs. dynamical d-s."
For consistency we have revised the captions on all multi-panel scatter plots in the manuscript and supplement to state which pairs of model simulations are being compared in the top, middle and bottom panels.

Literature added to the References:
Emery, C., Liu, Z., Russell, A. G., Odman, M. T., Yarwood, G., and Kumar, N.: Recommendations on statistics and benchmarks to assess photochemical model performance. J. Air Waste Manag. Assoc., 67, 582-598, doi: 10.1080/10962247.2016.1265027, 2017.