Responses to the Comments of the Anonymous Referee #1

We very much appreciate the constructive comments and suggestions from this reviewer. Our point-by-point responses to the reviewer's comments are provided as follows (the reviewer's comments are marked in Italic font):

This study tries to quantify the impact of biomass burning (fire) and other anthropogenic (non-fire) sources to the occurrence of low visibility days (LVDs due to PM2.5) in several cities across the Southeast Asia. This is an extension of their work in Lee et al., 2017 by improving the WRF-Chem model components. Regional air quality degradation is assessed using simulated PM2.5 and ozone, derived AQI, and mortality calculations. They identify that the inclusion of measured anthropogenic dust component to the model increases performance of the model. They also assessed the performance of some machine learning algorithms to predict the occurrence of LVDs.

Generally, the study is of importance, and relevance to ACP. It can be published with a major revision.

First, the novelty of the work (if any) should be mentioned in the manuscript, in the introduction.

Studies of Southeast Asia air quality using high-resolution models with interactive chemistry and meteorology combining with observations, even for specific cases rather than decadal-scale analysis, are still rare. Our previous study using WRF coupled with a simplified tracer model for $PM_{2.5}$ provided arguably the first such quantitative analysis, which demonstrates that biomass burning aerosols contributed to up to 40-60% of haze events in the major cities of Southeast Asia during 2003-2014 (Lee et al., 2017). In this study, we have further the depth of the analysis by applying a more sophisticated regional weather-chemistry model of WRF-Chem to quantitatively address the impacts of fire and non-fire aerosols on air quality and visibility degradation over Southeast Asia. We have also used available in-situ measurements to evaluate and correct model for providing a better base for further improvement of particularly emissions over the region. Beyond the traditional process models such as WRF-Chem, we have also experimented using machine learning algorithms to identify suitable conditions for hazes based on historical data and hence to forecast the likelihood of the occurrence of such events.

To address the reviewer's point, we have further emphasized the uniqueness of our study in the introduction section of the revised manuscript, by clearly indicating the new methods and approaches adopted in our study.

Authors mention that the underestimation of particulate matter in the model could be due to horizontal resolution or missing anthropogenic dust. Have you considered any other aspects of the model before making such a statement? how about the simulated boundary layer mixing of tracers? why ozone is overestimated in the model?

We have actually used the measured particulate composition data to correct modeled biases due to missing organic matter (residual) besides anthropogenic dust component (Snider et al., 2016) (Fig. S1 in the revised version; also see response to a later comment). Although this was mentioned in the original manuscript, it may have been unclear. We have revised the text accordingly to emphasize the importance of applying the correction to the modeled $PM_{2.5}$ concentration using the measured values of organic matter residuals.

We adopted the Mellor-Yamada-Nakanishi-Niino level 2.5 (MYNN) (Nakanishi and Niino, 2009) as the planetary boundary scheme in this study. The WRF model also has a reasonably fine vertical resolution for the PBL by using a vertical coordinate that is stretched to have higher resolutions inside PBL (e.g., having an average depth of ~30 m near the surface). With four to five model layers within the PBL, the model should be able to reasonably simulate the mixing of tracers in the boundary layer. We have added description of the PBL scheme in the revised manuscript as: "The Mellor-Yamada-Nakanishi-Niino level 2.5 (MYNN) (Nakanishi and Niino, 2009) is chosen as the planetary boundary scheme in this study. By using a vertical coordinate that is stretched to have higher resolutions inside the planetary boundary layer, the model has about 4-5 vertical layers inside the planetary boundary layer with a vertical resolution of ~30 m near the surface."

We have noticed that NO_x emission is higher in REAS emission inventory compared with other emission inventories and studies (Kurokawa et al., 2013). The boundary condition of background ozone in the default WRF-Chem configuration also appears to be somewhat high (30 ppbv) for our domain. Both could lead to the overestimated ozone in the model. We have added corresponding discussion in Sect. 3.1 in the revised manuscript.

Have you tried the simulations using any other emission inventories? This is very important.

We agree with the reviewer that using different emission inventories in the model would very likely lead to different results as indicated in our previous study (Lee et al., 2017), where we used two different biomass burning inventories in the simulations and derived different results for given cases; however, such differences did not substantially influence our major conclusion. In this study, we have actually compared the differences between the two available emission inventories for WRF-Chem for the targeted domain, the REAS and EDGAR inventories, in a pair of one-year simulations comparing 2006 REAS against EDGAR emissions. The results are shown in Table R1 (Table S3 in revised manuscript). It is quite clear that the differences regarding aerosols are quite limited. After considering the high spatiotemporal resolution of REAS emission inventory and the comparison results, we decided to use REAS in our study. Besides our analysis, Kurokawa et al. (2013) have also documented the comparison of REAS with other emission inventories in Southeast Asia.

In the revised manuscript, we have added that "We have compared the modeled results using REAS versus EDGAR emission inventories in one-year paired simulations: the differences between these two model runs are rather limited regarding aerosol-related variables (Table S3). After considering high spatiotemporal resolution of REAS emission inventory and the comparison results, we decided to use REAS in this study. In addition, a detail comparison of REAS with other emission inventories in Southeast Asia was also presented by Kurokawa et al. (2013)."

Table R1. Mean annual emissions and modeled concentration of BC, OC, SO_2, CO and NO_2 from 2006 REAS and EDGAR emission inventories in the simulated domain.

	REAS		EDGAR	
	Emissions (Tg/year)	Modeled (ug/m^3 or ppmv)	Emissions (Tg/year)	Modeled (ug/m^3 or ppmv)
OC	0.12	0.1131	0.15	0.1487
BC	0.036	0.0311	0.065	0.0643
SO_2	0.43	1.03×10^{-4}	0.65	2.01×10^{-4}
NO_2	0.3	4.94×10^{-4}	0.205	4.83×10^{-4}
CO	3.53	8.10×10^{-2}	7.48	8.72×10^{-2}

Model evaluation should be conducted in a much better way before making conclusions. Spatiotemporal distribution of each species should be evaluated thoroughly, in the context of all the modeling components. PM2.5 (its components and extinction values) should be assessed, not just PM10 (there are some measurements available).

We appreciate the reviewer's suggestion. In the revised manuscript, we have modified many presentations of the results in Section 3.1. Nevertheless, a fundamental issue in evaluating model for Southeast Asia domain is the lack of observations. As we described in the manuscript, $PM_{2.5}$ observations in this region are very limited. Even in Singapore, observed $PM_{2.5}$ data are only available after 2014 for the general public and research community to access. In most other Southeast Asian counties, even PM_{10} measurement data are hard to find, especially for the time periods before 2008. We are fortunate to be able to obtain some chemical species data from WMO and long-term AQI data from the Malaysian government. In addition, $PM_{2.5}$ component data from SPARTAN filtered samples (operated after 2013) have also been used, e.g., in Fig. S1 of the revised version.

Have you assessed the importance of organic matter in PM2.5 over these regions? the 'residual matter' in Snider et al., 2016 is mainly organic, please refer to that paper; so, the statements such as "including the in situ anthropogenic dust improved the ..." should be revised (since you are adding dust and organics).

We really appreciate the reviewer for raising this issue. Indeed, the residual matters that have actually been used in the study to correct modeled $PM_{2.5}$ concentration are mostly organic carbon, though this was not made clear in the original manuscript. We have made our best effort to clearly indicate this fact in the revised manuscript.

Clearly quantify and describe the uncertainty in your estimates of LVDs etc. (for fire and *non-fire related) derived using model values. An entire section should be devoted to uncertainty analysis.*

We appreciate the reviewer's suggestion. Since a full-scale forward-integrating uncertainty analysis based on WRF-Chem model would extremely expensive computationally, we have adopted a method for dichotomous (yes or no LVDs) cases and then give a contingency table as below to address model evaluation and to quantify model performance.

		Observed LVD	
		yes	no
Modeled LVD	yes	***hits***	***false alarms***
	no	***misses***	***correct negatives***

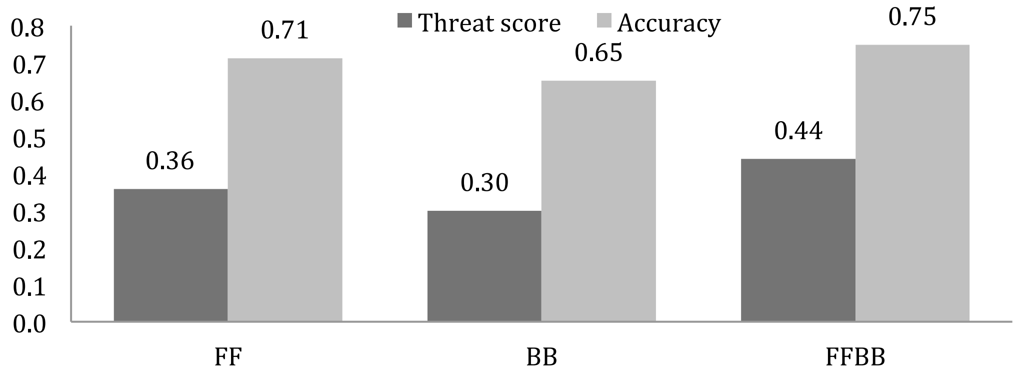
We have estimated *accuracy* based on the Eq. (1):

$$Accuracy = \frac{hits + correct\ negatives}{hits + misses + false\ alarms + correct\ negatives} \qquad (1)$$

Accuracy here is also called fraction correct, which is easy to evaluate model prediction. However, it can be misleading for some cases since it is heavily influenced by the most common category, usually "no event" in the case of LVD. Hence, we have provided *threat score* in this study as well. Based on the equation of threat score (or critical success index), we can measure the fraction of observed and/or modeled LVDs that were correctly predicted. Threat score also can be referred as the *accuracy* when correct negatives have been removed from consideration, that is, threat score only concerns modeled LVDs that count.

$$Threat\ Score = \frac{hits}{hits + misses + false\ alarms} \qquad (2)$$

The figure below shows the mean value of accuracy and threat score of modeled LVDs among 50 ASEAN cities in three experiments: FF, BB, and FFBB. Since the category of correct negatives is heavily counted in the accuracy, the values are also twice as high as the threat scores. Basically, BB has the lowest threat score while FFBB has the highest score as expected.



The above discussion has been added in Sect. S1 in the supplementary and introduced in the manuscript, Sect. 3.1.

Section 3 should be improved for a better reading, by excluding unnecessary statistical details, and by describing the figures and findings in a more clear and concise way. (abstract and conclusion sections should also be revised).

Based on the reviewer's suggestion, we have removed statistical details (i.e., mostly the standard deviations) in the text (the numbers are still presented in corresponding tables). The structure of the manuscript has been rearranged as well. We have made the manuscript more concise, including the abstract.

Separate section 3.2 into two; first, describe 4 selected cities and your conclusions; then, the entire region.

We have separated Section 3.2 into Sections 3.2 and 3.3 in the revised version. As the reviewer suggested, Section 3.2 now describes results of the 3 selected cities and Section 3.3 discusses those for the entire ASEAN cities.

Section 3.4 is too vague, are you really assessing the impact of aerosols on regional climate? need a better analysis; descriptions are loose; need to cite relevant works throughout the discussion.

We agree with the reviewer that this section diffuses the focus of the paper. We have moved it to supplementary material with a rewriting.

Provide a brief description of machine learning algorithms in the introduction itself (and your motivation for doing this); also, describe it in the method section. Section 4.2 should *be described in an entirely separate section.*

Based on the reviewer's suggestion, we have added the motivation of applying machine learning techniques to predict the occurrence of LVDs in the introduction section. We would like to keep the description of each algorithm in the machine learning section to maintain the flow of discussion. Sections 4.1 and 4.2 have been separated into two individual sections in the revised version.

Line 501-503: vague arguments; Line 569-570: describe

Lines 501-503: "Applying inverse modeling, for example, could optimize the emission inventories and hence improve the model performance" has been removed in the revised version.

We have rewrite Line 569-570 to: "Nevertheless, previous works by Reid et al. (2012) and Lee et al. (2017) also suggested the relationships between fire hotspot appearance and certain weather phenomena particularly precipitation. Therefore, we are surprised that precipitation in the fire regions does not appear to be a significant feature for predicting Singapore haze compared with other features in our current analysis."

Reducing the number of figures and tables in the main manuscript (without losing much information) would be helpful; even figure captions are too lengthy.

The reviewer's point has been well received. We have shortened the paper in the revised manuscript. Table 1 has been removed. Table 3, Fig. 7 and Fig. 9 have been moved to the supplementary material. We also have made the captions more concise.

Kurokawa, J., Ohara, T., Morikawa, T., Hanayama, S., Janssens-Maenhout, G., Fukui, T., Kawashima, K., and Akimoto, H.: Emissions of air pollutants and greenhouse gases over Asian regions during 2000–2008: Regional Emission inventory in ASia (REAS) version 2, Atmos. Chem. Phys., 13, 11019-11058, 10.5194/acp-13-11019-2013, 2013.

Lee, H. H., Bar-Or, R. Z., and Wang, C.: Biomass burning aerosols and the low-visibility events in Southeast Asia, Atmos. Chem. Phys., 17, 965-980, 10.5194/acp-17-965-2017, 2017.

Nakanishi, M., and Niino, H.: Development of an Improved Turbulence Closure Model for the Atmospheric Boundary Layer, Journal of the Meteorological Society of Japan. Ser. II, 87, 895-912, 10.2151/jmsj.87.895, 2009.

Reid, J. S., Xian, P., Hyer, E. J., Flatau, M. K., Ramirez, E. M., Turk, F. J., Sampson, C. R., Zhang, C., Fukada, E. M., and Maloney, E. D.: Multi-scale meteorological conceptual analysis of observed active fire hotspot activity and smoke optical depth in the Maritime Continent, Atmos. Chem. Phys., 12, 2117-2147, 10.5194/acp-12-2117-2012, 2012.