Response to Referee #2: We would like to thank the referee for the careful review throughout the paper and the useful comments.

Our Reply follows (*the reviewer's comments are in italics and blue*)

General Comments

This paper presents the development of an improved machine learning based air quality nowcasting system. Instead of using all possible related features in the model training and predicting, they selected those general important and effective features. The feature selection is done using a computationally efficient ensemble method. Their nowcasting system is tested on the PM2.5 forecast on a national scale and validated to be superior than a CTM model and conventional MLs. Generally speaking, the paper is clearly written and well structure, their results are scientifically solid. I recommend accepting it after a minor revision. I also have questions and comments for the author that could help to improve their manuscript.

Major comments

They have tested their regional feature selection-based ML nowcast system at a national scale and using several common ML models (RF, GB and MLP), which makes their results very sound. However, it is only tested at a 2019 winter season. I understand winter is the most severe polluted season there. Their system should be able to operate in a rolling forecast way. If extra training and testing are conducted at a less-polluted period/season, this study would be an excellent paper.

Reply: We thank the reviewer for the very comment and agree with the point. We made extra forecast in April 2020 which is a less-polluted month. The new experimental results reproduce the advantages of RFSML and we believe this experiment enriches the results of the original paper.

Remarks concerning the rolling forecast tests are now added in page 6, line 5-6 "***Our RFSML system can of course operate in a rolling way, additional forecasts in a less-polluted and emergency period 2020 April are performed with the models are trained using the recent two-year data similarly.***", and in page 19-20, line 9-11 and line 1-2 by saying "***To further confirm the predicting capability in a rolling way, we make forecasts over a less polluted month April 2020. Specific results can be found in Supplemental Material Tables S1. Steady improvement of predicting performance is still achieved by RFSML. Time series as given in Figure S22 show similar result as main text that RFSML has better predict ability than standard machine learning. As is illustrated in Figure S23-24, RFSML has both lower RMSE and MAE than***

standard machine learning, which implies the advantage of RFSML."

Table S1. Summary of prediction performance in the time period of April, 2020.

Region	*Metric*	*Predicting horizon*			
		6		*18*	
		standardML	*RFSML*	*standardML*	*RFSML*
NCP	*RMSE*	*17.71*	*12.2*	*22.11*	*16.71*
	MAE	*14.06*	*9.3*	*17.86*	*13.19*
	R	*0.71*	*0.83*	*0.5*	*0.69*
PRD	*RMSE*	*10.7*	*7.78*	*13.17*	*11.1*
	MAE	*8.51*	*5.74*	*10.38*	*8.39*
	R	*0.83*	*0.9*	*0.7*	*0.77*
SCB	*RMSE*	*13.29*	*10.37*	*17.02*	*13.51*
	MAE	*10.13*	*7.63*	*13.11*	*10.2*
	R	*0.72*	*0.81*	*0.53*	*0.66*
YRD	*RMSE*	*14.08*	*10.43*	*18.67*	*14.41*
	MAE	*11.27*	*8.09*	*14.76*	*11.48*
	R	*0.75*	*0.87*	*0.51*	*0.74*
FWP	*RMSE*	*16.26*	*13.24*	*19.8*	*16.44*
	MAE	*12.69*	*10.14*	*15.65*	*12.97*
	R	*0.66*	*0.73*	*0.47*	*0.6*
REST	*RMSE*	*21.59*	*17.89*	*26.01*	*22.25*
	MAE	*14.29*	*10.5*	*17.48*	*13.62*
	R	*0.68*	*0.79*	*0.48*	*0.66*

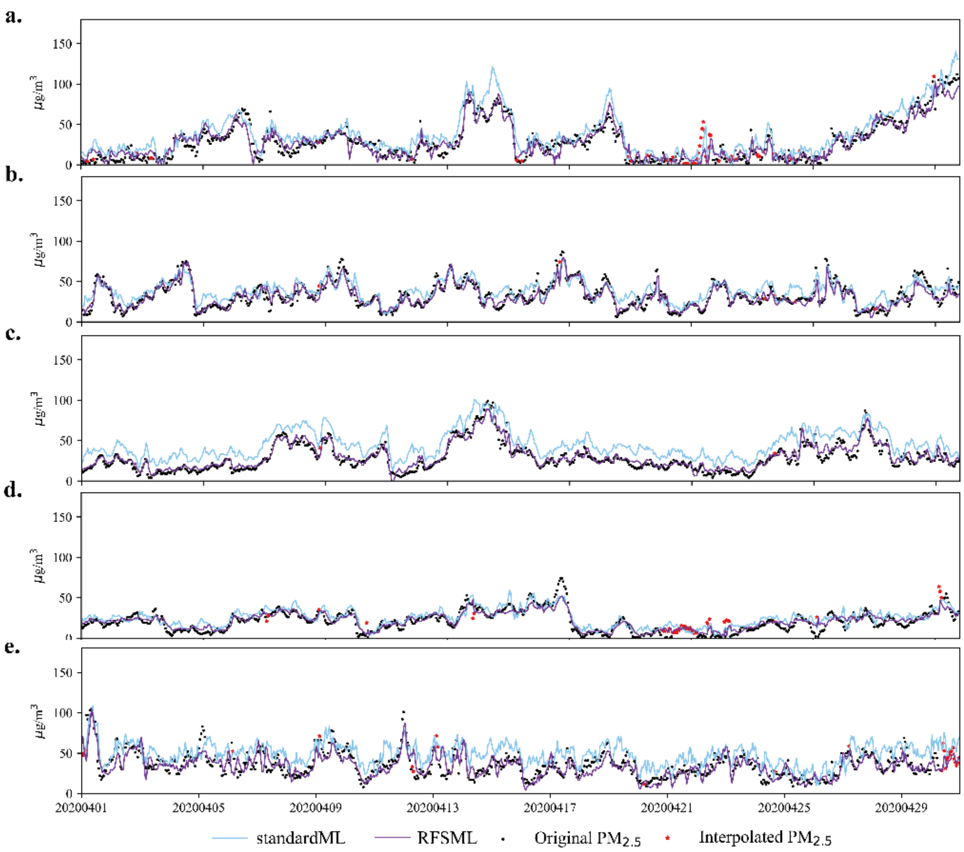

Figure S22. Time series of a prediction horizon of 6 hours in five mega-city cluster regions. The black dots and red pentacles represent original and interpolated PM2.5 respectively. The solid lines with light sky blue and dark violet represent prediction of standard machine learning system and RFSML respectively. Panel a, b, c, d and e represent a random site in NCP, YRD, PRD, SCB and FWP respectively.

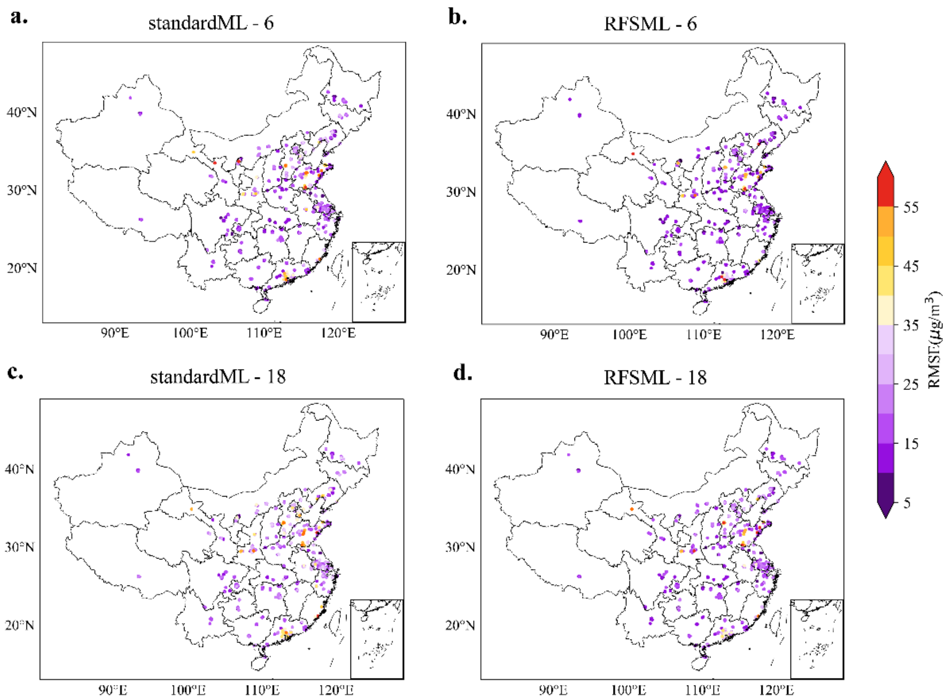

Figure S23. Spatial distribution of RMSE in a predicting horizon of 6 and 18 hours. Panel a and c are results of standard machine learning system while panel b and d are results of RFSML. The cooler the color tone, the lower the RMSE, thus the better predicting performance.

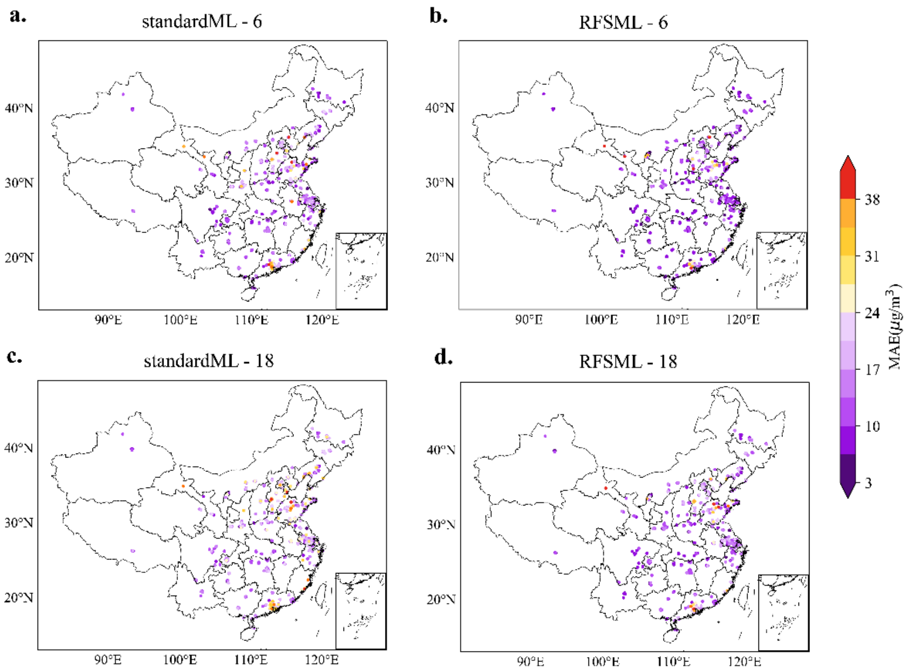

Figure S24. Spatial distribution of MAE in a predicting horizon of 6 and 18 hours. Panel a and c are results of standard machine learning system while panel b and d are results of RFSML. The cooler the color tone, the lower the MAE, thus the better predicting performance.

They should also explain the current machine learning model cannot fully replace model-based air quality forecasting systems, as ML models could not be trained and operated without inputs from the historical measurements. While for many rural regions, they are unavailable. The authors should explain this point clear.

Reply: We agree with referee that the ML cannot fully replace the current “causation” model that is a parameterization of physical rules in nature, while ML is purely based on data correlations. On the other hand, the current RFSML can indeed provide forecast at single stations instead of a full gridded one.

Actually we are exploring for a full prediction that covers the whole model domain from this current work RFSML. The basic ideas of is to fuse the high-quality RFSML prediction and the gridded CTM prediction with larger uncertainty using Bayesian Theory. The diagram can be found in the Figure below. Here the blue lines represent the high-quality forecast available at several single stations, and the model is trained using the observations marked by black dots;

the blue face here denotes the chemical transport model (CTM) giving the gridded forecast which is however usually biased. The RFSML and CTM prediction can be considered as two estimates of future situation, and each of them has the weakness and advantage. Bayesian theory will be used to fuse them together, and resulting a gridded and less-biased forecast like the brown face. That work will be soon submitted as a companion paper with this RFSML work.

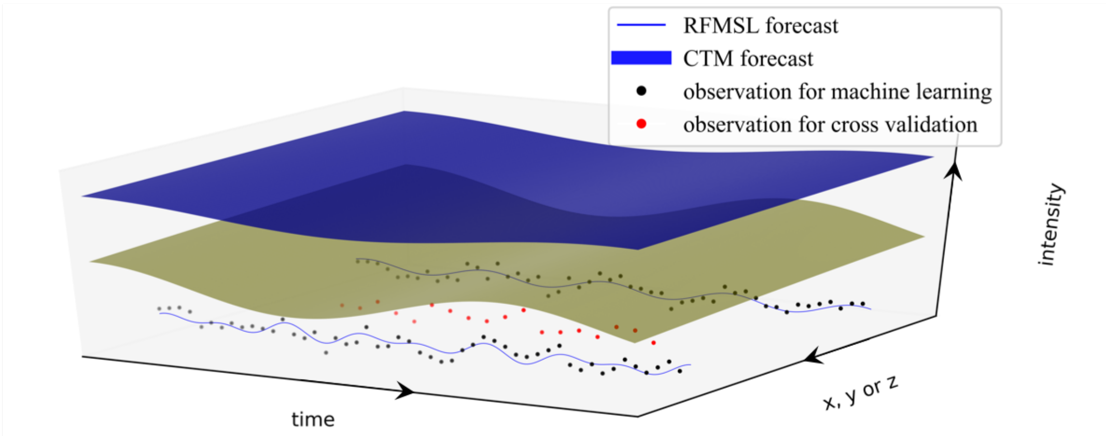

Figure Diagram of a gridded prediction from the RFSML prediction at single stations.

Remarks are now added in the Conclusion and future by saying "***Meanwhile, RFSML provides only predictions over the air quality monitoring sites where historical data is available for machine learning model training, instead of a grided forecast. A Bayesian theory -based prediction fusion is being explored now to extend the RFSML forecast available at single stations to a gridded one.***" in page 20, line 18-20.

Minor comments

Page 6, Table 1: esolution to resolution

Reply: Corrected.

Page 12, line 5: computational complexity?

Reply: Corrected.

Page 17, line 1: and at a predicting horizon?

Reply: Corrected.