Anonymous Referee #4

General comments: This paper presents one of the first estimates of benzene and toluene emissions in China using atmospheric measurement data from a rural site in the region and inverse modeling. This paper conveys important information about the toxic pollutants in China. In particular, this paper suggests the need for more measurements to better constrain these emissions.

Specific comments:

Although the result is promising, the authors need to address a few important issues before publication. First, with regard to transport model evaluation, I don't see enough description or result on this. In terms of the methodology, it should be addressed in Section 2.2, but there is no information about potential model biases. The simple question I can ask is how reliable the NCEP meteorology is in this region of China.

This is not about random errors we consider in the inversion. If the transport model is biased, then we get the biased mean estimate for the posterior emissions. The authors need to evaluate the transport model only for less than a month (pretty short period compared to other many studies). I suggest that the authors evaluate the transport model for potential biases (e.g., wind bias, planetary boundary layer height).

Response: It is almost impossible to quantify model biases, unless large-scale atmospheric experiment is carried out to quantify the biases of a model for a specific region and for a specific species. For example, some independent measurement data like radiosondes not used in the assimilation can be used to validate the model simulation. However, this is out of scope of the present study. As indicated by the reviewer, the meteorology datasets may perform different simulation (and inversion) results. Thus we tested the dataset from the European Centre for Medium-Range Weather Forecasts (ECMWF). The a posteriori emissions for PRD are very close when using these two meteo. dataset, e.g. for benzene, 4.0 Gg/month from inversion using CFSR and 4.2 Gg/month from inversion using ECMWF. Although Fang et al. (2014) shows ECMWF data set performed slightly better than the NCEP CFSR dataset for the SF6 simulations in East Asia for Hateruma, Gosan and Cape Ochiishi stations, we found CFSR dataset performed slightly better than the ECMWF dataset in the benzene simulations at the Heshan site. Thus, the CFSR dataset was used in our paper.

Table. Performance of different meteorology datasets for the benzene simulation

A priori benzene emissions	Meteo. datasets	Performance		
		E	B	r^2
MEIC v1.2	CFSR	1.53	0.96	0.20
	ECMWF	1.67	1.09	0.17
	CFSR performed better than ECMWF?	**yes**	**yes**	**yes**
RCP 2.6	CFSR	1.40	0.78	0.20
	ECMWF	1.54	0.85	0.13
	CFSR performed better than ECMWF?	**yes**	**yes**	**yes**

With respect to observation–model mismatch errors, just using the RMSE value of model minus observation is a convenient way, but not a correct way. I assume that "model" in L22 of P24845 means the model prediction for mixing ratios based on the a priori emission. The way inversion works is that the difference between predictions and measurements are minimized in a statistically consistent way with the two error covariance matrices. Thus if the error covariance is not properly defined, then the results are not reliable. What fraction of the observation-model mismatch is from the transport error? Was the background error included in the total mismatch error? If so, what fraction? After inversion, the RMSE is usually reduced compared to that before inversion. If I read the description in Section 2.3 as it is, the authors assign all of the average difference between "model" and observation to the random errors (i.e., the mismatch error). However, the difference between (prior) "model" and observation is also due to the lower or higher emission inventory (used as the a priori emission) than the true emission. Since the detail is not provided here, this is how I interpret L21 - L22). This important issue should be addressed.

Response: Yes, observation–model mismatch errors were defined as RMSE value of model minus observation. It cannot be determined that "What fraction of the observation-model mismatch is from the transport error". The background was zero, since the 20 days backward simulation could cover all emission sources from which benzene and toluene can be transported to the Heshan site. Thus, the background and its error are zero.

I also wonder how the authors evaluate the uncertainty of the OH simulations from GEOS-Chem. Ideally, this uncertainty should be explicitly estimated and included in the observation–model mismatch errors

Now, with respect to background, the authors set the background to zero. However, the initial values at the domain boundary, which represents the atmospheric background, are not zero. I agree that these background concentrations (which contribute to the receptor) will dissipate in 10 or 20 days. However, what is not clear is how large the modeling domain is. If the authors used 20-day backward simulations to remove the effect of the background, then the back trajectories would have gone outside China. I wonder if the authors used a large enough domain such that the background effect is negligible. From the figures in the paper, the domain is very small, all regional.

Response: It is right that "the back trajectories would have gone outside China". The modeling domain is globe. Thus, domain is large enough such that the background effect is negligible. The boundary condition is not the space domain but in the time domain. Therefore if the lifetime is << 20 days, then the background is probably zero or at least very close to it.

Minor comments:

P24841, L25: too much use of 'crucial'

Response: In our revised manuscript, 'crucial' has been changed to 'important'.

P24842, L2 - L3: sudden jump in the logic. Why spatial and temporal distribution all of sudden? Is this related to the main conclusion of this study?

Response: it looks not "sudden jump in the logic". In this paragraph, we first introduce the health of benzene and toluene on health and air quality, and then we summary emission information is important to evaluate these impacts.

L11: typo in "the most the most"

Response: We have incorporated this change in our revised manuscript.

P24843, L15: From the perspective of inversion, one day in December would not make a big difference for the less than a month-long dataset. It is weak reasoning that the authors say "Data from 1 December 2010 were not used, since we focused on mixing ratios and emissions in November". There should be a better explanation to not to use the December data (e.g., poor measurement accuracy). The measurements are not enough here.

Response: We agree with that "one day in December would not make a big difference for the less than a month-long dataset". However, for the sake of accuracy, "Data from 1 December 2010 were not used, since we focused on mixing ratios and emissions in November". If we used the data from December as well, the reviewers and readers would argue that the estimated emissions are not only for November but Nov and Dec combined.

Also, in the abstract it says "a rural site". Is it only one site? In Section 2.1, it says "atmospheric measurements of benzene and toluene at two sites were used," It looks like the Mt. TMS site was not used for inversion. The authors need to be clear about this at the beginning of this measurement section.

Response: At the very beginning of the measurement section, we present that "In this study, atmospheric measurements of benzene and toluene at two sites were used, the Heshan site (used for the inversion) and the Mt. Tai Mo Shan (TMS) site (used for validation)".

P24848, L18 - L20. How small? Is there a reference or supporting material?

Response: Systematic errors were not analyzed in our paper and in our revised manuscript, we have deleted this sentence.

Reference:

Fang, X., Thompson, R. L., Saito, T., Yokouchi, Y., Kim, J., Li, S., Kim, K. R., Park, S., Graziosi, F., and Stohl, A.: Sulfur hexafluoride (SF6) emissions in East Asia determined by inverse modeling, Atmos. Chem. Phys., 14, 4779-4791, 10.5194/acp-14-4779-2014, 2014.