We are grateful for the detailed revision of our manuscript and appreciate the valuable comments and suggestions that greatly helped us improve our work. We will address all the issues point by point.

Answers to Referee #2's general comments:

General comments:

This paper proposes a new method to use trajectories to evaluate emission inventories. Unfortunately, there are 2 problems: 1. the paper does not describe more recent developments which have been found to be useful, 2. the method is not convincing either in its formulation or its results. I will take each issue in turn:

1) ***See for example Stohl et al., ACP 2009 "An analytical inversion method for determining regional and global emissions of greenhouse gases: Sensitivity studies and application to halocarbons" and references.***

We thank the referee for the valuable suggestions. More related recent developments have been added in the Introduction section, to provide more useful information.

2 a) I'm not sure that Eq. 3 is valid. The emissions are in units of concentration / hr which does not make much sense.

The reason why we did not convert the units into one that is conventional for inventories is that we did not want to introduce extra uncertainty into the calculation. The conversion would require planetary boundary layer height (PBLH) data across the domain, which would bring up PBLH modeling errors. However, we agree that such a unit is not typical emissions and makes our result hard to compare to those in other studies.

Accordingly, we improved our algorism in the following way:

Instead of using the measured volume concentrations as C_l in eq.5 ($\overline{E}_{mn} = \overline{\sum_{l=1}^{M} W_{mn} \cdot C_l / T_l}^{mn}$), we replace C_l by a column mass concentration in units of $\mu g \cdot m^{-2}$. This would result in a derived $\overline{E}_{mn}$ in the unit of $\mu g \cdot m^{-2} \cdot hr^{-1}$. To bring it to unison with typical inventories, the emission field is multiplying by the area (A_{mn}) of the respective grids (m,n) , yielding a concentration field in the unit of $\mu g \cdot hr^{-1} \cdot cell^{-1}$:

$$\overline{E}_{mn} = \overline{\sum_{l=1}^{M} W_{mn} \cdot C_l \cdot A_{mn} / T_l}^{mn} \quad (1).$$

During daytime, in the convective boundary layer (CBL), we can assume that pollutants are well mixed and concentrations are the same throughout the CBL. After sunset until the early morning, under the stable boundary layer (SBL), pollutant concentrations are assumed to decrease exponentially with height. Typically, the quasi-stationary CBL has been build up in the late morning and collapses late in the afternoon, during sunset. Since in our observation period, the sunrise time lies between 4:55 a.m. and 05:11 a.m., and the sunset time varies from 07:40 p.m. to 07:09 p.m., the daytime period was defined as 9 a.m. to 6 p.m.

For CO, the original instrument measurement outputs are given as mixing ratio (γ_{CO}) in *ppbv*, thus the column concentration in the CBL can be calculated as is shown in eq.(2):

$$C_l = \frac{\gamma_{CO} \cdot M_{CO} \cdot P}{R \cdot T} \cdot PBLH \quad (2),$$

where M_{CO} is the molar concentration of CO, P and T respectively stand for the atmospheric pressure and temperature and R is the ideal gas constant. In the SBL, C_l is calculated as:

$$C_l = \frac{\gamma_{CO} \cdot M_{CO} \cdot P}{R \cdot T} \cdot \int_{z=0}^{\infty} e^{-\frac{z}{PBLH}} dz \,. \quad (3)$$

BC measurements are given as mass concentrations (m_{BC}) in $\mu g \cdot m^{-3}$, thus the column mass in the CBL and SBL can be respectively calculated according to eq. (4)-(5).

$$C_l = m_{BC} \cdot PBLH \quad (4)$$

$$C_l = m_{BC} \cdot \int_{z=0}^{\infty} e^{-\frac{z}{PBLH}} dz \quad (5)$$

BC and CO emissions mostly come from sources within the boundary layer, to which our derived results apply. Direct comparisons with the INTEX-B inventory or other inventories can then be made by converting the unit from $Ton \cdot year^{-1} \cdot cell^{-1}$ to $\mu g \cdot hr^{-1} \cdot cell^{-1}$ ($1 Ton \cdot year^{-1} \cdot cell^{-1} = 10^{12}/365/24$ $\mu g \cdot hr^{-1} \cdot cell^{-1}$).

b) The residence time based on a single trajectory seems to be rather arbitrary, and I cannot see how this equation improves on Ashbaugh et al.'s work.

The residence time in our calculation is not a substitute for that in Ashbaugh et al.'s work. In Ashbaugh et al. (1983), which was later named the Potential Source Contribution Function (PSCF), the residence time associated with concentrations at the receptor point above a certain threshold is divided by the residence time of all trajectories in a certain grid:

$$PSCF(m,n) = \frac{t(m,n)}{T(m,n)}, \quad (6)$$

where *T(m,n)* is the total number of endpoints falling with in the grid (m,n), and *t(m,n)* is the subset of endpoints that are on trajectories associated with concentrations above a certain criterion. Such a calculation results in a probability distribution of possible source regions.

In our work, two kinds of residence times are used. For the retrieval of emissions, a PBL residence time is calculated for each single trajectory to decide how long that certain air parcel has been staying in the boundary layer to take part in the mixing and transport process. The calculation of the trajectory residence time is demonstrated in Sect. 2.4.3 in the revised manuscript and will be further discussed in ***2d)***. For the calculation of the regional emission contributions to the receptor site, the total residence time of all trajectories is used, which is similar to *T(m,n)* in eq. (6). However, instead of accounting for all trajectory endpoints, we only consider those before the trajectory exceeded the PBL height.

In comparison with the PSCF technique, our work shows advantages in two different aspects:

1. In our calculations, we do not use the whole trajectory, but only that part before the air parcel rises above the PBL height. If the trajectory passes over a certain grid while it is in the free troposphere, the pollutant concentrations in that very air parcel is hardly influenced by emissions in the boundary layer. Hence concentrations at the receptor site also cannot reflect the emissions for that grid;
2. For air quality modeling, emission inventories are needed. The PSCF technique can only derive a probability distribution of source regions, whereas our work can provide quantified emission strength. After making the adjustments discussed in ***2a)***, the results are directly comparable with other inventories.

c) Eq. 5 introduces a fudge factor that lets you tweak the results so that they look like the input that you are comparing them too. By the time you get to Eq. 9, it seems you would be better off doing a sum of concentrations in each "cluster" and comparing those values, without any recourse to trajectories (the residence times might well cancel each other out mostly if you do the substitutions).

The weighting field is not used for fudging or tweaking, it is only applied to make amendments for the fact that we associate the same concentration for the whole trajectory, which is also done in CFA techniques. If the a priori field is chosen inappropriately, it might have influences on the final result, this mostly occurs when the weighting factor shows very high false signals, which can cancel out the influence of C_l and T_l on the derived emission.

We agree that in the former manuscript, it might have been inappropriate to use the INTEX-B

inventories at the same time as weighting factor and as the reference for comparing the results. Instead we are using an average distribution of MODIS Aerosol Optical Depth (AOD) in 2009 as weighting factor. An average distribution of OMI tropospheric NO_2 column is also used to make comparisons between different weighting factors. Results show very little difference between the emissions derived without any weighting factor and those calculated using AOD and NO_2 column concentration as a priori fields.

d) Basing the analysis on a single trajectory and calculating a life time based on when it exceeds an arbitrarily chosen boundary layer height seems rather perilous. I would recommend doing true particle trajectories (hundreds / thousands of particle releases), use WRF PBL heights, calculate residence times by the method of Ashbaugh.

As already explained in ***2b),*** the residence time in our retrieval is a different concept from that use by Ashbaugh, hence it is not replaceable.

We agree that the calculation of the residence time might be influenced by our simple assumption of the PBLH. However, accurate PBLH data on all the grids and in 1 hour resolution are not available. Any guess of the temporal variation of the PBLH field can pose uncertainties.

In the former WRF simulations (using WRF version 2), the PBLH during nighttime could not be well described by the model (fig. 1, blue line), which is the main reason we chose not to use the modeled PBLH. As you can see, the PBLH during the night reaches a very low level (~30m) and shows no temporal variations. This phenomenon has been observed for all the modeled days, which indicates that this PBL scheme is weak in modeling the night time boundary layer. According to our method, this would cut off almost all trajectories as soon as it goes into nighttime.

The PBL physics were parameterized using the Yonsei University (YSU) scheme, which in many assessments have been proven to be the best performing PBL parameterization scheme, especially during summer (Hu et al., 2010, García-Díez et al., 2011). We repeated the modeling work using WRF version 3 and improvements can be noted (fig. 1, red line). However, the nocturnal boundary layer still seems to be rather low. Wu et al. (2011) has evaluated 4 PBL schemes in WRF 3.2, results show that all PBL schemes generally do well in unstable and weak stable boundary layers, but poor in stable boundary layer. In their evaluation, the YSU scheme does overall slightly better than other three, for which reason we decided to keep using this scheme. However, we have kept in mind that the low stable boundary layers may result in some bias.

Since the new meteorology modeling results can provide a more reasonable variation of the PBLH, the WRF PBLH field is employed to determine the residence time. As you can see in Figure 2a), in the former manuscript, the daytime PBLH was assumed to be 1000m and the nighttime PBLH was set at 300m. The residence time is the time difference between the starting time and the time point at

which the trajectory exceeds the PBLH. In our new calculations, we applied WRF PBLH values instead of fixed assumed ones. The PBLH values associated with each time point t_i and grid point (m,n) over which trajectory l passes is sorted out and then used to determine when the trajectory flow exceeds the boundary layer height.

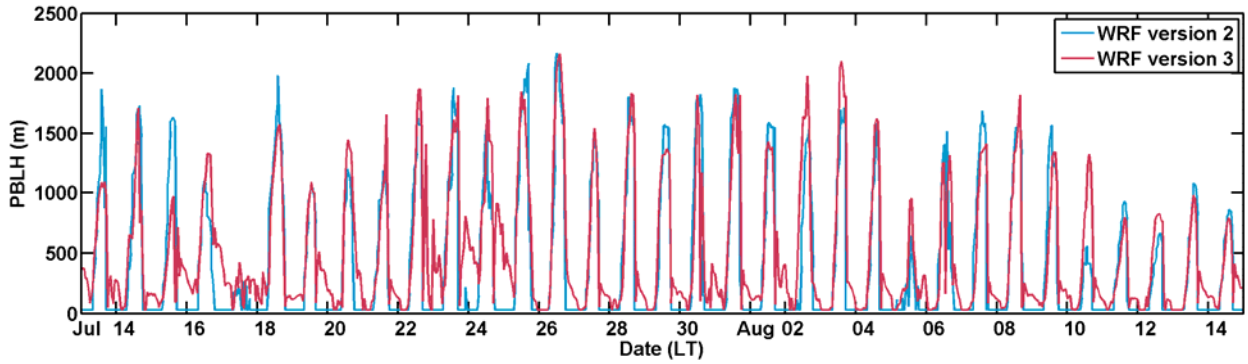

Figure 1. Modeling results of the PBLH at Wuqing.

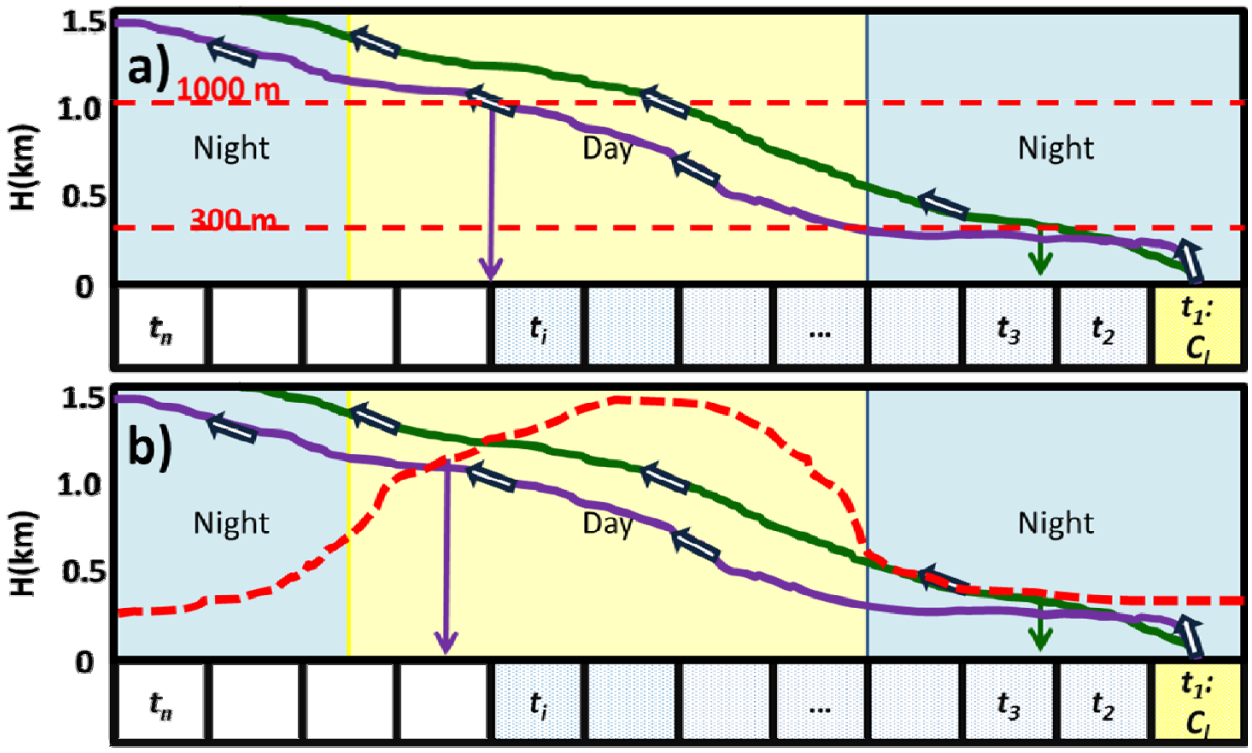

Figure 2. a) Schematic showing the determination of the PBL residence time of trajectories (Figure 2 in ACPD paper), b) New determination method of the PBL residence time.

Reference

García-Díez, M., Fernández, J., Fita, L., and Yagüe, C.: WRF skill over Europe with 3 PBL schemes during the year 2001, 3rd International Meeting on Meteorology and Climatology of the Mediterranean, 2011.

Hu, X.-M., Nielsen-Gammon, J. W., and Zhang, F.: Evaluation of Three Planetary Boundary Layer Schemes in the WRF Model, Journal of Applied Meteorology and Climatology, 49, 1831-1844, 10.1175/2010jamc2432.1, 2010.

Wu, W., Liu, Y., Vandenberghe, F., Bourgeois, A., Grim, J., Warner, T., Knievel, J., Dudhia, J., Bruyere, C., Stauffer, D., Padovani, M., Luft, G., and Fling, K.: Evaluating PBL Schemes in WRF3.2, WRF workshop, 2011.

e) Fig. 4 shows the results of the method. However, because the domain is much larger than the different source regions considered in the text, it is very difficult to see what is going on. Overall, it seems that the study would have been better served by starting out with PSCF or CFA or some of the more recent developments. If the results from these were found to have specific problems, then the paper could show results from the modified method explaining how it improves on the previous methods.

We don't think it necessary to repeat the work of PSCF or CFA, because those methods do not retrieve emissions. PSCF is used to detect possible source regions, giving results in the form of probability distributions. CFA yields a concentration distribution, which also is to show where possible source regions lie and how the concentration of substances might be distributed. Both techniques have been validated, and they have their advantages. However, what we try to do in our work is to provide a method to derive actual emissions, which can be compared to other inventories and be eventually used in models.

Answers to Referee #2's specific comments:

1) "energy statistics" – could you please be more specific about what you mean by this? It seems you are talking about energy consumption data. The introduction has some inaccuracies in the description and could use some more recent references, including review papers. For example, there are recent papers that evaluate the types of method that this paper would expand upon: Kabashnikov et al., Atmospheric Environment 2011 and Scheifinger et al., Atmospheric Environment 2007.

Thank you for your suggestions. We agree with the referee and have added more summaries on recent references to the introduction.

By "energy statistics" we do refer to the statistical data on energy consumption, e.g. the statistics on energy use for power plants, industries, various production sectors and residential combustion, etc. The term "energy statistics" is widely used, but to make it clear, we will add the above details to the text.

2) Pg 31139, line 2-9: The description of bottom-up and top-down approaches is incorrect.

The "bottom-up approach of deriving emission inventories integrates the emissions from all known types of sources using fuel consumption data and emissions factors (Wang et al., 2012). That is also what we wanted to express in the manuscript, maybe we did not make ourselves clear enough.

The "top-down" method was initially based on the thought that given the total emissions, we can disaggregate it spatially with the help of certain indicators (e.g. Population, roads, land-use, etc.). However, the total emission is often an unknown factor itself. With the rapid development of chemistry transport models, inverse modeling techniques, which use atmospheric observations (e.g. Satellite measurements or in-situ measurements) and a priori emissions as constraints to derive optimized emissions, have been developed and also included into the "top-down" method category (Brioude et al., 2011; Lee et al., 2011).

To make ourselves clear, Lines 2-9 on Page 31139 have been rephrased as:

"*The most common approaches of deriving emissions can be divided into "bottom-up" and "top-down" methods. "Bottom-up" methods integrate the emissions from all known source types using energy statistics (e.g. statistical data on energy consumption by power plants, industries, various production sectors and residential combustion, etc.) and emissions factors (Wang et al., 2012), which bear large uncertainties and cannot be easily updated (Zhang et al., 2009). Emissions may have large spatial and temporal variations, due to new policies and the rapid development, especially in developing countries such as China. "Top-down" retrieving techniques were initially based on the thought of spatially disaggregating the known total emission with the aid of statistical indicators (e.g. population, land-use, etc.). More recent developments in the "top-down" method resort to inverse modelling approaches, incorporating atmospheric observations (i.e. satellite observations or in-situ measurements) and chemistry transport models together with prior emissions as constraints to derive optimized emissions (Lin et al., 2007;Brioude et al., 2011; Lee et al., 2011).*"

Reference:

Brioude, J., Kim, S. W., Angevine, W. M., Frost, G. J., Lee, S. H., McKeen, S. A., Trainer, M., Fehsenfeld, F. C., Holloway, J. S., Ryerson, T. B., Williams, E. J., Petron, G., and Fast, J. D.: Top-down estimate of anthropogenic emission inventories and their interannual variability in Houston using a mesoscale inverse modeling technique, J. Geophys. Res., 116, D20305, 10.1029/2011jd016215, 2011.

Lee, C., Martin, R. V., van Donkelaar, A., Lee, H., Dickerson, R. R., Hains, J. C., Krotkov, N., Richter, A., Vinnikov, K., and Schwab, J. J.: SO2 emissions and lifetimes: Estimates from inverse modeling using in situ and global, space-based (SCIAMACHY and OMI) observations, J. Geophys.

Res., 116, D06304, 10.1029/2010jd014758, 2011.
Wang, S. W., Zhang, Q., Streets, D. G., He, K. B., Martin, R. V., Lamsal, L. N., Chen, D., Lei, Y., and Lu, Z.: Growth in NOx emissions from power plants in China: bottom-up estimates and satellite observations, Atmos. Chem. Phys. Discuss., 12, 45-91, 10.5194/acpd-12-45-2012, 2012.

3) Pg 31139, line 23-28: My impression was that PSCF is the name that Zeng & Hopke (1989) gave to Ashbaugh's method.

We agree with the reviewer and the according lines were rephrased as:

"*Ashbaugh (1983) and Ashbaugh et al. (1985) developed a statistical method, which was later on termed as the Potential Source Contribution Function (PSCF) analysis by Zeng and Hopke (1989). It is defined as:*

$$PSCF(m,n) = \frac{t(m,n)}{T(m,n)}, \qquad (1)$$

where T(m,n) is the total number of endpoints falling with in the grid (m,n), and t(m,n) is the subset of endpoints that are on trajectories associated with concentrations above a certain criterion."

4) Pg 31140, line 10: Seibert's method, CFA, is a variation of PSCF.

Although both the PSCF and the CFA methods are based on trajectories and their residence times, Seibert's method is rather distinct from that of Ashbaugh's in its concept, which was also stated by Scheifinger and Kaiser (2007). They share a similar form, and it seems that the only difference is that CFA methods are weighting trajectory residence times with measured concentrations. However, there are two major distinctions:

1. PSCF only considers the residence time of trajectories that are associated with concentrations at the receptor site above a certain level, the information of the other trajectories are lost in the process (Stohl et al., 1996). CFA methods take into account all the trajectories and weight them with the concentrations, which provides are more samples on each grid for statistical analyses.
2. PSCF methods derive probability fields of source locations, whereas CFA methods retrieve possible emission field distributions.

Reference:

Scheifinger, H., and Kaiser, A.: Validation of trajectory statistical methods, Atmospheric Environment, 41, 8846-8856, 10.1016/j.atmosenv.2007.08.034, 2007.

Stohl, A.: Trajectory statistics-A new method to establish source-receptor relationships of air pollutants and its application to the transport of particulate sulfate in Europe, Atmospheric Environment, 30, 579-587, 10.1016/1352-2310(95)00314-2, 1996.

5) Pg 31140, line 14: The method of Stohl et al., 1996 is mainly a development to make use of multiple sites.

Both the CFA method and the development by Stohl et al. (1996) can make use of multiple sites, we don't think that this is the main point of their improvement. Stohl et al. developed the CFA method into a redistributed CFA approach, which mainly aims at extracting more information from the data. Initial CFA results goes into the calculations as a weighting field to redistribute concentrations along the trajectories. "Hot spots" of pollution sources are located using iterative calculations. This approach compensates for the fact that CFA attributes the concentrations evenly to along the trajectories, which is their main development.

6) The "clustering" is not really a clustering but a categorization by average direction. Using a clustering algorithm would give you clusters that reflect the dominant flow types based on the data itself. This would be a big improvement on what is described in the paper.

The "clustering" was mainly used to show the coverage of the trajectories and the frequencies of the flow from different directions. However, we agree with the reviewer that using clustering algorithms can help us better reflect the flow types. In the revised manuscript, the following clustering results of the HYSPLIT clustering tool will be provided and discussed in Sect. 3.1.

"Figure 3 a1-a4 shows the 4 groups of trajectory clusters, while a2-a4 displays the frequency distribution of various trajectory heights with their travel time. During the measurement period, the cluster groups in a1-a3 show similar occurrence frequencies, varying from 31.3%-38.9%, while the group from the north-eastern direction shows the lowest occurrence frequency (5.4%). Trajectories from the eastern (E) sector (group b: 38.9 %) reveal a circular flow that dominantly travels on lower altitudes and have longer residence times. They originate from Shandong Province, eastern Hebei or are maritime flows from the Bohai Sea, which travel over the industrial area and the inner city of Tianjin. The group from the southern (S) direction (group b: 34.1%) travels mostly from the southern inland to our receptor site and flows also stay on a relatively lower altitude before 30 hours of travel time, afterwards, flows tend to gain height. Group c has an occurrence frequency of 31.3%, including short circular flows from all directions and long straight flows from the north-western direction. The mean flow path is also circular and comes from the NW, traveling altitude increases rapidly with time, indicating short residence times within the boundary layer. The cluster from the

north-eastern direction mainly grouped trajectories together that descend to rapidly to ground level 6-10 hours before arriving at the receptor point. Over 90% of those trajectories occurred during 23rd -25th July, in which period Wuqing experienced several precipitation episodes. "

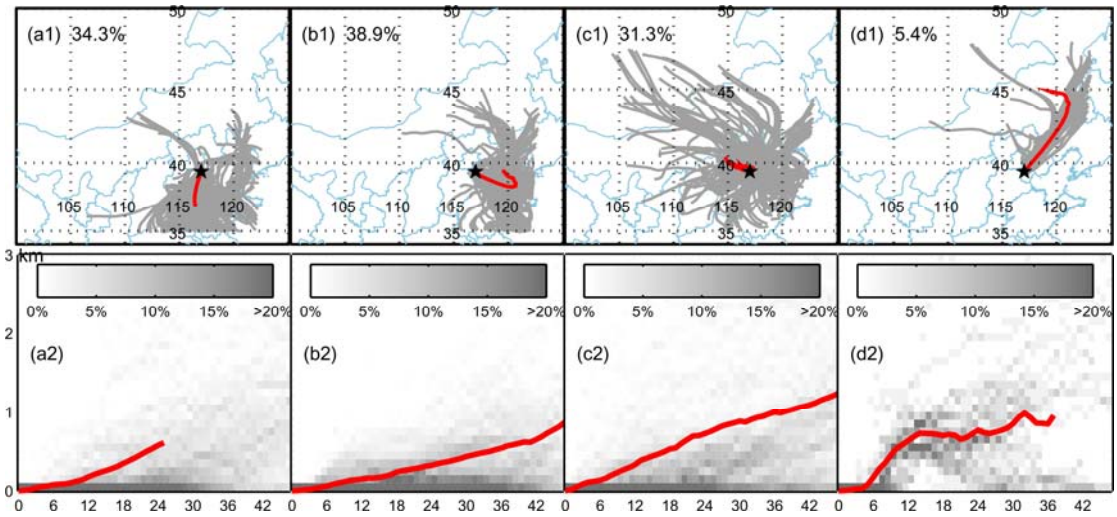

Figure 3. Cluster frequency analyses of trajectories during 13Jul. -14 Aug 2009 in the North China Plain. Upper panels (a1-d1) show the trajectories in each cluster and the mean path of the clusters, while lower panels (a2-d2) display the trajectory heights vs. travel time and the according frequency distribution of all trajectories in each cluster.

7) The choice of WRF domains seems a bit odd – the nested domain is nearly as big as the first domain. In the text you should include the input global data (GFS from NCEP at 0.5 degree resolution?)

The WRF model is initialized and driven by NCEP FNL (Final) Operational Global Analysis data at 1.0 degree resolution, this information was added to Sect. 2.2. In the WRF simulations, the outer domain was only applied to provide a better boundary condition and to downsize the horizontal resolution with a 1 to 3 ratio. In the trajectory analyses, only the inner domain was considered. To spare computational time, the outer domain was chosen to be not far larger than the inner one.
However, we agree that the unconventional outer domain setup of ours may not be able to provide the inner domain with optimal boundary values. To avoid any possible errors it might bring, we altered the size of the outer domain to 2 times the size of the inner domain (fig. 4).

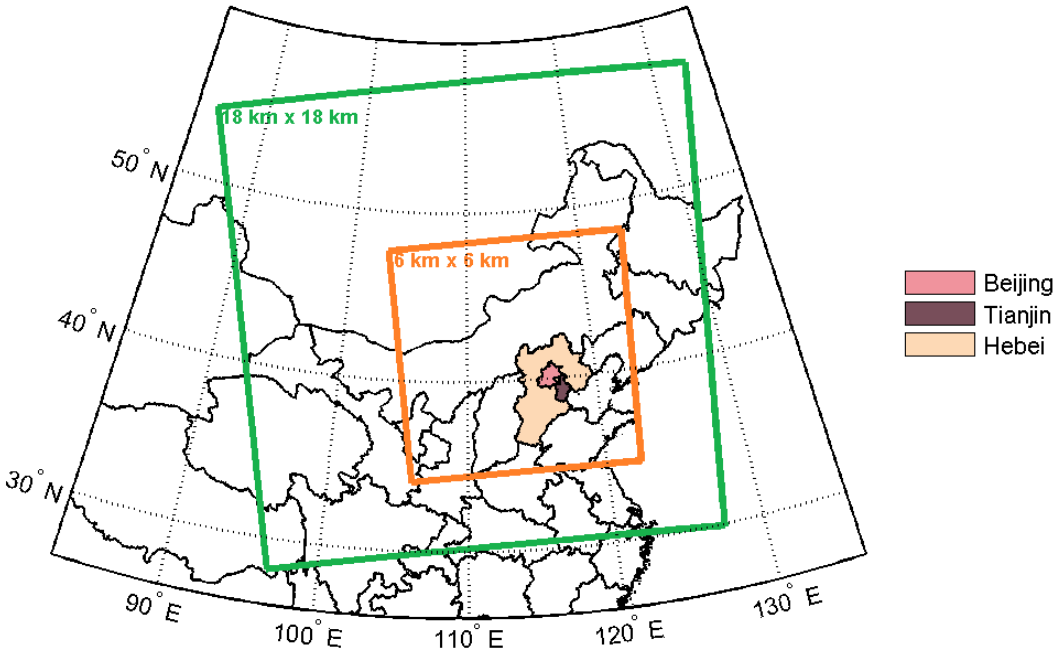

Figure 4. Model domains for the ARW meteorology simulations