General comments

Yang Wang et al. presented a comprehensive study of the temporal and spatial distribution of aerosols and trace gases in the central-western North China Plain. The manuscript is well structured and the results show good correlation with results of other instruments.
However, many results were filtered based on a cloud classification scheme which performs somehow unreliable to me. Especially low and high aerosol loads seem to be mixed up with clouds. To a certain points, MAX-DOAS profiles should be able to retrieve these differences but do not show larger deviations e.g. on 11/05 and 19/05. Since this classification has a large impact on the complete discussion, I would suggest to add a small estimate of the impact of wrongly classified/filtered scenarios on your results.
Furthermore, an additional analysis of NO2 retrieved in a different fitting window might help to clarify if horizontal inhomogeneties were present.

Specific comments

P2, L14-19: Please add a reference to Fig 1. in the Introduction.
P2, L32: Please add the full name for the abbreviation East-Aire.
P3, L9: MAX-DOAS algorithms are not only based on OE. Iterative approaches like Newton-Gauß or Levenberg-Marquardt are in use. I would rather call these algorithms "inversion algorithms" or "inversion based algorithms".
P5, L17: Wang et al 2018 seems to be the wrong reference?
P5, L26: "and for the MAD-CAT campagin" → "e.g. the MAD-CAT campaign"
P5, L30: You state detection limits but not how they were calculated. Please add the missing information.
Section 2.2.3: The uncertainties of the individiual parameters were given as percentages but where do they come from and how were they calculated? From a previous study or from not shown sensitivity tests? Please give some information.
P6, L11-12: When there is a sunphotometer measuring routinely. Why not using the exact SSA and asym parameter closest in time rather than averaged values? For which wavelengths are the averaged quantities? How did you convert to the proper wavelengths or did you assumed no wavelength dependency?
P6, L17: How were these wavelengths chosen? 354 for HONO is the mid of the fitting window but what about the other wavelengths? Was there are reason for not choosing the fitting window mid wavelengths?
P6, L21: Why was the upper grid limit chosen to be at 3km? Typical altitudes from other studies are usually at 4km.
P6, L25: Covariances of 100% of the surface value for all altitudes? Is this correct? The commonly used approach is a fixed percentage of the a priori profile of the individual altitudes.
Table 2: Why is there a different SZA limit for SO2 compared to the other trace gases?
P6, L30: Here, you write R but in the Figure its R^2. Which one was given?
Fig 4:

Please change the colors for either low or high aerosols in this and similar plots because it is hard to distinguish between both markers.

1. The aerosol retrieval shows similar profile shapes from 6 to 13 but the cloud classification finds different cloudy conditions, sometimes with thick clouds. How is it possible that the aerosol retrieval is not affected by thick clouds?
2. The a priori profiles for aerosols and SO2 are not even close to the the retrieved profiles. How can you be sure that you do not over- or underestimate the retrieved profiles due to in inaccurate a priori profile?
3. I do not understand why the degrees of freedom for aerosols are larger than any of the

trace gas ds. This is unexpected for me. Could you please explain where these larger differences in ds for the invididual retrievals come from?

P7, L15: Why is the near-surface extinction trusted under partially cloudy conditions but the AOD not? I would also assumed that the near-surface extinction is inaccurate when broken clouds led to a contaminateion of some elevation angles only. The profile will be smoothed due to the a priori smoothing and the retrieval for all altitudes should be affected.

P7, L25-28: The numbers are averages over the full time period? Please call it a total average then or specify what these numbers exactly mean.

P8, L3-5: 5-times the word "also" in three lines. Maybe you can reformulate these two sentences.

P10, L6-7: I am confused of what you wrote in Section 2.2.3 and what you wrote in lines 6-7 (see comment P6, L11-12). Please explain your approach in greater detail.

P10, L10: A parameter of 1 means an Angström Exponent of 1? Please replace the word "parameter". Furthermore, was the exponent of 1 used for all data or was the sunphotometer's Angtröm exponent closest in time applied for the conversion?

P11, L28-29: In addition to your reason, a general lower sensitivity for higher altitudes might be another reason. Furthermore, the limitation to 3km might also be an issue.

P12, L1-2: Generally, I would not assume that an elevated layer within the lowest kilometre can not be resolved by a MAX-DOAS profiling algorithm. For higher altitudes, this might be an issue. Is it possible to add a brief synthetic test on 3 or 4 elevated layers in different altitudes? Just to see the retrieval response. Because an elevated layer could also be possible when die underlying aerosol profile is box-like due to an oscillation around this box-features.

Fig 6:

1. It would be interesting to know if these outiers for high visibilitymeter AE correspond to certain geometries, time or weather conditions?
2. **P12, L11:** When this behaviour for different cloudy conditions can be attributed to clouds, why is there a different correlation for NO2 and SO2 for high aerosols and cloudy sky? NO2 has a better correlation for cloudy sky while SO2 has a better correlation for high aerosols. Do you expect large inaccuracies in the classification scheme for clouds and aerosols?

P12, L12-14: A limited vertical sensitivity for near-surface trace gas concentrations might also be important when the trace gas is concentrated in a shallow layer much smaller than the grid step width of the profiling algorithm, especially when using a coarse grid steps width of 200m.

Section 4.2:

1. What is the vertical resolution of the Lidar?
2. Was AK smoothing applied for the Lidar measurements (with the assumption that the vertical resolution is higher than that for MAX-DOAS)? Please add also the Lidar AOD to the AOD sub-figure.
3. Why are high near-surface values not similarly found by both instruments? The sensitivity for near-surface values should be the highest for both instruments and different air masses and clouds are not expected to be that important for lower altitudes. E.g. MAX-DOAS found larger extinctions at 15:00 (2016/5/16) while the Lidar found larger values in the evening of 2016/5/17.

P13, L8: How did you calculate these "combined profiles"? Linear interpolation between lowest air-craft and surface value? Please add some information.

Fig 8:

1. Please add times for MAX-DOAS measurements and the overpasses. It would also be nice to have some color-coding or different grey-scales for the green surface values to identify the surface value changes throughout the measurement period.
2. Why do the SO2 curves always agree better than NO2 even though the degrees of freedom are much lower for SO2? Did you try to retrieve NO2 also in another fitting window (> 400nm)? It would be interesting to see if the results differ strongly. That would support the argument of horizontal inhomogeneties.

P13, L24-25: Please add the time when the photos were taken on the individual days.

Section 5.1: Why does the cloud classification show highly variable results between 6 and 14 BT (11/05) but the aerosol profiles do not differ strongly? This indicates that either the profiles or the classification is inaccurate. In addition, around noon two days later (13/05), thick clouds were found but the aerosol retrieval does not show these clouds. This is surprising.
P15, L21-22: Trajectories ending at different altitudes were used, but how? You do not give information about that. Did you just average all maps for Figure 11?
Section 5.2.1: It would be interesting to see how theses maps change when the individual lifetimes are considered. Are CTM calculations of lifetimes at these days available?
Fig 12: Please add the data of the valdiating instruments, when possible.
P17, L14-15: How can I differ between the original data and the interpolated data in Fig 13? Please use other marker styles.
P17, L16: Highest values for southeast winds with southerly trajectories instead of north-westerly?
P18, L18-19: of on the order of hours --> in the order of hours
Fig 14c: The red dots are extremely small. Please increase the size of these dots.
P31: Wang, F et al.: Please change 2048 to the proper year.