Reviewer 1

Based on a large suite of ambient data, this paper presents an analysis of four different chemical assays that quantify, what the authors refer to as, oxidative potential (OP). It is based on a data set from a multi-investigator field campaign in Beijing involving summer and winter sampling periods. Overall, the topic of particle OP is of current interest as a new measure that potentially better links aerosols to adverse health. The paper mainly repeats various analysis approaches (with a few tweaks) done in many other studies and seems to largely support earlier findings, from what I can tell, since what the actual new findings are is not really clear. The paper should be substantially edited before considering publication. The following are major issues:

We thank the reviewers for their comments. Please see below point-by-point answers, addressing the reviewer's comments. All of our answers are given in blue below and we indicate the lines changed in the revised manuscript.

What are the major findings of this paper? The Abstract provides very little insight on results.

We have now amended the abstract and conclusions section to further emphasise the key findings and conclusions of the paper.

The data interpretation is often sloppy and statements are made that are either speculation (see below) or illogical. Tightening up the paper would also lead to a much more concise and readable paper.

See below more specific comments.

This paper is about assays, but few details on the assay (ie, how specifically the measurement was made) are given. Most details are in the supplement, yet this information is critical to the data interpretation. This is a major lapse since how the assay was conducted largely determines everything else in the paper. How can comparisons be made between these results and others if one doesn't even know if the assays are comparable? The most obvious is the AA assay; the authors seem to have developed their own protocol, different from previous methods, but go on anyway to compare their AA results to other published work. Specifically, for the AA assay they do not extract in synthetic lung fluid (this is not even mentioned in the paper), yet they compare their data throughout with previous AA assay results that do. What type of aerosol species are included in the assays, are these assays measuring water soluble or all aerosol species?

Due to the length of the manuscript, we felt it necessary to include the specific chemical details of the assays in the supplementary material. The methods used are well documented in the previous literature and easily accessible in detail in the supplement.

We have now specified in the main manuscript that we use an AA-only based assay (line 180-183) that does not contain additional synthetic lung fluid components, and that filter samples are extracted at pH 7 before reaction at pH 2, i.e. extraction is performed at physiological neutral pH conditions. Furthermore, we do not compare quantitative values as stated in lines 266-268, but compare components that influence aerosol OP as measured by AA in different studies. In fact, we observe that this AA-based assay is in relatively good agreement with previous studies in terms of correlation between assay response and specific particle components, as discussed in Section 3.2

The reason for using these assays is to better link aerosols to adverse health, but there is no discussion in the Introduction/Background of the current knowledge on this matter. A number of the assays have been tested in health studies, less is known about some of the others. For example, what is the logic

of detailed investigation of measurement that shows no evidence of being linked to health? Is the argument here that we don't know which assays are linked to health so these four were simply chosen? State exactly why these assays were investigated.

The main aspect here is to identify how the response of four of the most widely-used OP and ROS assays are linked to other atmospheric components and processes using one of the most comprehensive atmospheric datasets acquired in recent years, during the APHH-Beijing campaign. Such comprehensive comparisons are sparse in the literature, and this campaign provided a particularly unique opportunity to correlate aerosol OP and particle-bound ROS with a uniquely comprehensive dataset. This novelty has now been highlighted further in both the abstract and conclusion in the main manuscript. Such studies constitute an essential step in terms of understanding assay response, as a well-constrained understanding of aerosol chemical influences on these assays allow better understanding of their response and thus a firm foundation to determine the health-relevance of such measurements.

We have now added more information regarding the links between aerosol OP and toxicity in air pollution epidemiology; please see Section S2 of the Electronic Supplementary Information.

The term OP in this paper is used in a very broad sense. Assays that measure very different physiological processes related to ROS are all grouped as simply OP. More precise terminology would allow more detailed conclusion.

This distinction is made in the introduction (line 103-108), but has now also been elaborated on in Section 2.2.2, as well as in Section S2 of the ESI. We have further clarified definitions in lines 242-250.

It would seem better to separate out the assays that measure exogenous ROS (DCFH) and the assays that measure species that can form ROS in vivo (DTT, AA). As a guide maybe refer to the figure in Lakey et al, Sci Reports 2016. As an example why this may matter, maybe the lung lining fluid has sufficient antioxidants present to suppress all the ROS on the particles, (one might want to ponder the difference in concentrations of ROS on the particle, ng/m3 based on the DCFH assay, and typical O3 concentrations, ug/m3, aren't both are exogenous ROS). But the ROS generated from aerosol components through interactions with physiological species is a different mechanism to produce ROS that may involve catalytic reactions (eg, Fenton reaction).

We now more clearly distinguish the different sensitivities of different assays lines (101-105). We refer to the assays specifically as $DCFH_v$, AAv, EPR_v and DTT_v for *e.g.* volume-normalised data throughout the manuscript to distinguish the assay response, due to the differing sensitivities of each assay to different chemical components present in PM. We have now discussed these definitions in lines 242-350. We do in places refer to total OP of PM to summarise all of the assay responses (e.g. all show a stronger correlation in winter compared to summer). Additionally, we discuss the limited knowledge available in the literature about compound-specific reactivity of the different assays in lines 83-105, and some of the broader implications of antioxidant-oxidant balance in the biological context in Section S2 of the ESI.

Furthermore, can exogenous ROS species be translocated to other organs in the body such as is known for species than form ROS in vivo (eg, metal nanoparticles)? In my view, greater insights would be possible if the authors separated out these different processes (and hence assays) that can lead to oxidative stress.

We would like to emphasise that this study is not a biomedical or epidemiological study, and thus we cannot address potential translocations of particles within the body; please see Section S2 of the ESI for further information. See comment above regarding the separation of assays in the manuscript.

Specific Comments. The abstract is not informative since it contains little actual results. Most of the discussion is on what was done, whereas more emphasis could be placed on findings. For example, what exactly is the new results from this extensive research?

Line 64 defines OP: The capability of PM to produce ROS with subsequent depletion of anti-oxidants upon inhalation is defined as oxidative potential (OP) (Bates et al., 2019). By this definition is DCFH assay a measure OP since it does not produce ROS, as far as I know?

As discussed above, we have now further clarified that DCFH predominantly measures particle-bound ROS (lines 101-105).

Line 174, typo analyze?

We consistently use British English throughout the manuscript.

No detail is provided within the paper on the assay methods, instead it is given in the supplement, yet this is critical information needed in the interpretation of the results.

As discussed above, we have now elaborated Section 2.2.2 (lines 164-175), and due to the length of the manuscript, detailed assay protocols are easily accessible in the supplementary material.

Please discuss limitations in measuring ROS with the DCFH assay using a filter that measures ROS on the particle (note the key word reactive).

This is a valid point, and it is certainly a limitation of filter-based measurements for particle-bound ROS and OP. We cannot rule out underestimating both particle-bound ROS and OP using offline filter-based measurements, but this offline method allows unparalleled comparison with other aerosol composition measurements. Additionally, with reference to Figure 2, we observe substantial variability in the mass-normalised $DCFH_m$ values in both summer and winter, implying that we are capturing a variability of particle-bound ROS as measured by DCFH. These may well be longer-lived components, and at present it is difficult to estimate quantitatively the degradation of short-lived ROS species prior to analysis. We have added additional discussion in 187-190 in the revised manuscript.

For what reason was just the AA picked to be shown in Fig 1 and the other assays shown in the supplement? (Same for later on in the paper).

Figure 1 is only an example of the daily variation of the assays, and given the length of the manuscript and the inclusion of several other figures, we only show AA here as an example and the other three assays are contained in the ESI. DTT DCFH and EPR figures are still contained in the supplement. We have added all four assays to Figure 6, and added DTT data to Figure 7. Additionally, all tables in the manuscript and figures in the ESI contain data measured using all four assays where they are not already presented in the main text.

If the assays are so highly correlated with mass (eg, AAv and DCFH), does that mean that the assays are not that useful? Why not just use mass to link to adverse health or particle toxicity, it is much easier to measure?

AA_v and EPR_v are only significantly correlated with $PM_{2.5}$ mass concentrations in winter, and the in summer show a poor correlation, as evidenced in Figure 2 and discussed at length in lines 279-290. Poorer correlations are also observed for DDT_v and $DCFH_v$ *vs*. $PM_{2.5}$ mass in summer compared to winter. As we also state in lines 295-297 the difference in the strength of correlations between volume-normalised assays and overall mass in the summer and winter indicates if anything that PM mass is not always a good indicator for predicting the oxidising properties of particles, and that either source related, composition related or atmospheric changes alter the oxidising properties of particles, and

hence acellular methods and indeed relationship between assay response and composition changes are required to determine the oxidative potential of ambient particles in Beijing.

Line 289, given the high correlation between AAv and PM2.5 mass, why is it surprising or meaningful for AAm to have an inverse relation with mass since AAm= AAv/PM2.5 mass? This is totally expected and not really informative.

We believe this is not the case. If indeed a linear relationship was observed between volume-normalised OP measurements and $PM_{2.5}$ mass concentration (which is not always the case, see above) then normalising per unit mass we would see relatively constant mass normalised value on each day (as the OP would simply scale with $PM_{2.5}$ mass concentration). However, as observed in Figure 3 and discussed in lines 303-323 in the revised manuscript, substantial variability of the mass normalised values is observed for each assay, and is the case with AA and DTT, days with higher PM concentrations tend to have lower intrinsic OP_m values compared to lower PM mass days. If the above comment was the case, we would not observe this and would observe relatively constant OP_m between low and high mass days.

Lines 294-307 on possible reasons a large fraction of mass may be OP inactive on high mass days. What about the effects of particle age (oxidation, or other chemical aging processes). High PM2.5 mass could be fresher emissions? An example is PAHs to quinones or nitro-PAHs.

This is an interesting point but we do not have the required additional information to comment further on the ageing of the particles and to evaluate this argument, but certainly this should be considered in future studies as the reasons behind observations displayed in Figure 3 were unclear.

Lines 311-312. Not sure of the relevance of PM10 discussion since it was a PM2.5 measurement of DCFH.

This is a valid point and these lines have now been removed from the revised manuscript.

Line 329, are there any quinones that are semi-volatile, please list them. There are semi volatile PAHs, but when they are oxidized does the volatility change?

There are a range of smaller quinones that have been detected in both the gas and particle phase, and have been found to have a temperature-dependent concentration in the particle phase (see e.g. Dalgado-Saborit el al., Atmospheric Environment 77 (2013) 974-982.) We have clarified this further in the revised manuscript and added this reference (lines 350-354).

Line 332, how specifically does boundary layer height affect the assay results?

Boundary layer data was unfortunately not available at the time of writing and we can thus not comment further on its effects.

Line 415-430 and on. The metal ions were not measured so how can the statement be made that the list of metals correlated to AA and DTT are related to redox reactions or on the role of Fe in various reactions. The logic does not follow. The authors are equating all chemical forms of these metals in the particles to the just the ion forms.

As discussed in lines 454-463, we do not comment on the speciation or redox-state of the metals we observe a correlation with as that data was not available at the time of writing. However, we observe correlations with a range of metals with AA and DTT and qualitatively compare these observations with previous studies. Future work should certainly explore the role of metal speciation and redox-state further and not only water-soluble vs. total metal content; we partially discuss this in Section S2 of the ESI.

Line 51-545. If one does a more complete aerosol chemical analysis what is the point of the assay. Why not just use the chemical species in the health/toxicity studies?

A complete chemical analysis of aerosol is challenging, and requires a breadth of expertise and instrumentation to perform, in addition to the large expense often associated with full chemical speciation. From an oxidative potential perspective, and demonstrated in this paper, a broad range of known chemical components contribute to the response of acellular assays, in particular AA and DTT. Thus, the assays encompass a wide range of chemical components, from the perspective of their oxidative potential, and thus provide a relatively simple metric to describe a chemically complex process. This point has now been emphasised in the revised manuscript (lines 715-718, Section S2 of the ESI).

Line 671 states in the conclusions: At present no single assay is completely representative of the totality of OP effects present in atmospheric PM. What is the basis for this statement? How do the findings of this paper support this statement?

This statement has now been amended and clarified in the revised manuscript (see lines 715-718).

The last line of the conclusions; again, what is the basis for this bold statement, how do the findings of this paper support this statement?

We do not think that our statement at the end of the conclusion section is particularly bold but we have reworded in the revised manuscript (see lines 715-718).

Reviewer 2

Based on the following four acellular assays: ascorbic acid (AA), dithiothreitol (DTT), 2-7-dichlorofluoroscin/hydrogen peroxidase (DCFH) and electron paramagnetic resonance spectroscopy (EPR), the authors of this study compared the oxidative potential (OP) and reactive oxygen species (ROS) production of PM2.5 in Beijing summer and winter. Furthermore, the authors also analysed the correlation of PM2.5 OP or ROS formation with different composition of PM2.5 and concentrations of some trace gases. Overall the topic of this study is interesting. Whereas the written of the manuscript needs revision. If the authors fully address the following concerns in a revised manuscript, this work may be publishable in Atmos. Chem. Phys.

1. The manuscript title highlights the research focus of this study to be the influence of atmospheric conditions and particle composition on OP of PM2.5. The beginning of the abstract also indicates that there exists uncertainty of the atmospheric conditions and specific chemical components of PM2.5 driving the OP. However, the abstract did not show any new results from this study that decrease this uncertainty. A specific, quantitative, or conclusive information on the influence of which atmospheric condition and different particle components on the OP of investigated Beijing PM2.5 is lack. Therefore, a more informative abstract is needed.

We have now amended the abstract, see discussion above.

2. The motivation for using the selected four assay methods rather than other assays in this study is not well depicted. For instance, whether the AA, DTT, DCFH, and EPR assay results have closer association with adverse health effects of PM2.5? This context should be introduced.

These four acellular methods are amongst the most commonly applied in previous studies, and provide information on particle-bound ROS (DCFH), superoxide production upon aqueous particle suspension (EPR) and catalytic redox chemistry (AA/DTT), thus provide a broad assessment of the oxidising properties of particles. Discussion added to the revised manuscript (see lines 103-108, Section S2 of the ESI).

3. As the authors indicated in line 80-81, different acellular assays all have differing sensitivities to specific particle components that may contribute to aerosol OP. Therefore, it is not surprising to see the various performance of different assays in testing Beijing PM2.5 (e.g., results in Figure 2). Moreover, it is reasonable to see the various correlations among different assays. The unclear thing is that why the combined application of the selected four assays has advantageous in providing new information than using individual assays?

Although previous studies have demonstrated that different assays have differing sensitivities, the role of aerosol composition in promoting these assay responses is unclear. The APHH campaign provides a unique opportunity to compare these commonly applied assay responses to a comprehensive dataset. Using all four assays provides a broad assessment of the oxidising properties of PM2.5, and correlating them to an extensive composition dataset provides a unique opportunity to explore which chemical components in PM2.5 drive the assay responses. We have added additional discussion in lines 103-108 and 33-37 to clarify this in the revised manuscript.

4. Line 309-310: why the mass fraction of organic peroxides in PM2.5 increase in winter? How can you justify?

We have now amended this statement, to clarify that that the concentration of particle-bound ROS is more abundant in PM2.5 in winter compared to summer, as we cannot definitively say the sole cause of the observation is due to organic peroxides.

5. The authors referred elemental carbon (EC) to be non-redox-active. However, many studies found that EC or black carbon can produce •OH in water. Thus, it is necessary to double check this interpretation.

This is a valid point and this statement has now been deleted from the revised manuscript.

6. For the EPR analysis, the authors used Tempone-H as spin trap to measure the production of O2 - . Whereas, this probe can also react with •OH and other radicals. Moreover, TemponeH is sensitive to the pH of solution samples. Have the authors measured the pH of $PM_{2.5}$ extracts? What is the relative fraction of O2 - among all the detectable radicals?

We agree with the reviewer that Tempone-H can react with peroxynitrite, peroxyl radicals and other radicals, although this occurs at more than an order of magnitude lower rate than that for superoxide (Dikalov et al. 1997). In relation to hydroxyl radicals, in the past we have performed experiments with Tempone-H using the Fe-H_2O_2 Fenton reaction and $^{\cdot}OH$ generators such as menadione, where we find that high concentrations of these agents are need to induce notable EPR signals. When working with ambient aerosol samples, we find that the EPR-Tempone-H signal can be attenuated by use of superoxide dismutase (SOD), but whereas $^{\cdot}OH$ scavengers such as mannitol have only a marginal effect. Assessing the relative fraction of $O_2^{\cdot-}$ in the sample is complicated by the slow reaction kinetics of the radical scavengers that have high specificity for superoxide. However, we have shown that SOD attenuates the Tempone-H-EPR signal of diesel exhaust particulates (an archetypal urban air particulate standard reference material) to the same extent as it does the signal from the superoxide generating agent pyrogallol, suggesting the majority of the signal from this particle is due to superoxide (Miller et al. 2009). We have added text to the EPR methods section in the revised supplementary material manuscript to highlight these limitations.

We thank the reviewer for their suggestion to test the pH of our particulate suspensions. Due to restrictions on our lab access and ability to receive particulate samples from other institutions, this is not something we are able to check at the current time, but we will do so in future experiments.

7. Carefully check the type setting of the whole manuscript. For examples, proper use superscript or subscript for PM2.5 and NH4 + etc.

We have amended any errors in the revised manuscript.

Revised Submission Replies Reviewer #2

Review of "Atmospheric conditions and composition that influence PM2.5 oxidative potential in Beijing, China" by Steven J. Campbell et al., (MS No.: acp-2020-1024).

Overall the manuscript is improved significantly. I suggest a minor revision before the acceptance of it to be published in Atmos. Chem. Phys.

We thank the reviewer for their comments. Please see below point-by point responses in blue.

1. Line 33 in page 1: the '(APHH-Beijing)' can be after the 'campaign'.

This has now been amended in the manuscript.

2. Line 42 in page 2: why the number of '107' needs to be highlighted? I did not see the importance of this number to the abstract and the manuscript.

We emphasise this number as it is one of the most comprehensive aerosol composition datasets to date, and therefore highlight the total number of additional composition measurements that we correlate with the assay responses.

3. Line 46 in page 2: the 'SOA' should be defined.

This has now been defined in the manuscript.

4. Line 82 in page 3: change the 'electron paramagnetic spectroscopy (EPR)' to 'electron paramagnetic resonance (EPR) spectroscopy'.

This has now been amended in the manuscript.

5. Line 94 in page 3: suggest to add some more recent publications to the citation.

The following reference has been added: Chen, Q., Sun, H., Wang, M., Wang, Y., Zhang, L. and Han, Y.: Environmentally Persistent Free Radical (EPFR) Formation by Visible-Light Illumination of the Organic Matter in Atmospheric Particles, Environ. Sci. Technol., 53(17), 10053–10061, doi:10.1021/acs.est.9b02327, 2019.

6. Line 121 in page 4: you may need to cite the work of '(Shi et al., 2019)' to provide more background information about the APHH campaign.

This citation has been added to the manuscript.

7. Line 125 in page 4: delete the first 'datasets', which is surplus.

This has now been amended in the manuscript.

8. Line 128 in page 4: I am still confused by the number '107'.

See discussion above (comment 2)

9. Line 196 in page 7: add 'in this study' after the 'source apportionment'.

This has now been amended in the manuscript.

10. Line 200-201 in page 7: the phrase 'PMF would not ultimately give useful models' is confusing. It is clear that a model will not give another model.

To clarify – we could not obtain measurement uncertainty values from our collaborators for the majority of chemical speciation data, which is a required element for the available software (EPA PMF

5.0 and SoFi) to calculate the "reliability" of a particular feature in terms of its contribution to a calculated factor (i.e. to a source), which in reality we consider to be a slightly artificially imposed limitation, especially in exploratory studies. Moreover, the definitive calculation of mass balance using this dataset was challenging, as it was clear from our univariate analyses that the chemical speciation was definitely not properly accounting for the total PM masses (partially illustrated in Figure 4 and S13 in the manuscript). We also had access to physical measurements which could directly influence the OP, which PMF modelling does not include. A "useful" model (in this case of the oxidative potential assay response) is one that gives functional and interpretable information on the phenomena being modelled, and as we were missing the required input for PMF, PMF models would not have been "useful" had they been produced for this study despite their convention in this field. Although it has a history of poor acceptance in the atmospheric chemistry literature, we used PCA for the main OP modelling, as we sought to derive the most important measurements contributing to assay response, and PCA gives ready interpretability in this sense. However, we did not use it for source apportionment, which is the primary criticism of its use.

11. Line 264 in page 9: the 'PM2.5 OPv' looks strange.

We have now amended this in the manuscript to just OP_v

12. Line 265 in page 9: the unit format of 'nM [DHA] m-3 ' is different from the one in Figure 1. Keep them to be uniform in the manuscript.

This has now been amended in the manuscript.

13. Line 279-297 in page 10: change the 'Figure 2B, 2C and 2D' to 'Figure 2b, 2c, and 2d' and the same for other figures. Recent studies found that peroxide-containing highly oxygenated organic compounds (HOMs) associate with the radical formation by PM2.5 in water (Chowdhury et al., Environ. Sci. Technol., 53, 23, 13949-13958, 2019; Tong et al., Environ. Sci. Technol., 53, 21, 12506-12518, 2019; Wei et al., Environ. Sci. Technol., 55, 1, 260- 270, 2021). Thus, what is the potential contribution of HOM to the observed superoxide radicals in Figure 2d?

HOMs were not measured during this campaign, but could help to explain observed variability in the EPR signal which could not be explained with the available additional composition measurements. Additional discussion of this and citations have been added at line 521 where we discuss correlations between aerosol composition and mass normalised EPR.

14. Line 374: why the summer data points are n=33? Because it is shown that n=34 for summer in line 141.

This was a typo and has now been corrected in the manuscript.

15. Line 409: add a full stop after the 'secondary organic aerosol'.

This has now been amended in the manuscript.

16. Line 509: suggest to cite: Tong et al., Atmos. Chem. Phys., 16, 1761-1771, 2016.

This citation has now been added to the manuscript.

17. Line 567-568: there are different meanings of the 'models' here. I suggest not use 'assay' rather than 'model' for describing AA and DTT.

This has now been amended in the manuscript.