General comments

This is an interesting study which addresses important topics within the scope of ACP. I believe it should be published in the journal after revision. The work is of high scientific quality but the presentation could be improved and further discussion is needed regarding the very large differences in results between the different source apportionment methods used in the study.

Specific comments

- It would be good to add the years covered in the study (2017–2020) in the title

- The abstract is a bit too long. The first part is unnecessary; I suggest starting with "We analyzed long-term measurements…" (line 46) and skipping the part before that. Also, the last sentence (lines 79-80) "FF combustion was the prevailing…" can be removed since the same information is already given on lines 70-72.

- Lines 60–62, Unclear sentence: "Primary biological aerosol particles (PBAP) tracers (various sugars and sugar-alcohols) were elevated in the NH-season but evolved differently, whereas cellulose was completely decoupled from the other PBAP tracers." What did the PBAP evolve differently from?

- Lines 91–92: You mention increasing sea salt aerosol in the Arctic; is this a general increase of SSA in the Arctic or just at the Zeppelin site?

- Line101–102: The statement (and references) about the knowledge about Arctic INP and CCN being rudimentary seems a bit outdated. This is a very active research field and newer references may be more appropriate, such as Creamean et al., 2022 (Nat Commun 13, 3537, 2022, https://doi.org/10.1038/s41467-022-31182-x)

- Regarding Sec 2.2 Sampling, handling, and storage of ambient aerosol filter samples:

  Lines 187–188: you state that you collected aerosol particles "for one week". Do you mean that you took weekly samples (i.e. collected the aerosol particles for one week for each sample)? The LHS-analysis presented in Table S3 seems to have varying sample length from 7, 10, 26, 20, 8, 18, 8, 8, 10, 18, 16, 10 days. I assume that the longer time periods are pooled but according to lines 251–252 you "pooled two consecutive samples" for six of the months (including December) but this is not clear enough (e.g. for Dec. it seems like only three days were included for 2018 and N/A for eight days in 2017).

  You have estimated the positive sampling artifacts for OC (by the QBQ method). Is there any risk of non-negligible negative sampling artifacts in your measurements?

- Regarding the "Selection criteria for samples subject to radiocarbon analysis" (lines 249–251): The first sentence of Sec2.5.1 is unclear. Are the samples "chosen" those from the time periods given in Tables S3–S4? If so, it would be good to refer to one of these tables to specify which samples were used. If the time periods are different please provide a table in the Supplement specifying which samples were used for radiocarbon analysis.
  How were the specific weeks selected? What was the actual selection criteria used to decide which samples were used for radiocarbon analysis for each month?
  The statement about the "data capture range of 20% in February to 67% in September" is also unclear; what does it mean? In Table S3, June 2017 seems to be covered to a larger extent than September (25/30 days compared to 20/30 for Sep)? What about the "data capture" for December? From Table S3, it seems like no values are given for the December week in 2017 and the LHS only includes 3 days of December 2018.

- Sec 2.6: Please specify the MAC values used to convert the absorption data to eBC concentrations in this study. Were there significant differences in MAC values between different seasons or years? How well correlated were the absorption data with the EC measurements? This information can be provided in the Supplement.

- Sec 2.8: It is interesting that you changed the LHS S-A compared to Yttri et al., 2011a. Instead of using levoglucosan to estimate TCbb, OCbb and ECbb, you now use ^{14}C-EC. Considering the uncertain atmospheric lifetime of LG, this seems like a good idea but please explain your reasoning in the manuscript. A comparison of the bb-contributions according to your revised LHS S-A and a similar analysis based on the LG-based methodology (from Yttri et al., 2011a) should be included in this article. It would be interesting to see how different the source apportionment becomes when using ^{14}C data instead of levoglucosan data. A comparison of the two approaches is valuable when using S-A data from earlier studies.

- Results and discussion, Sec 3.1; lines 367–369: The very high CA levels in 2020 compared to the other years need to be discussed, not just stated. What is the reason for the large differences?

- Sec 3.2.2, lines 433–438: It is interesting that your results show that non-fossil EC dominates in the H-season and that the contribution is much higher than the one found for 2012–2013 by Winiger et al. Do you have any indications that the relative emissions of fossil and non-fossil EC may have changed substantially during the time period 2012–2020? You mention "differences in sample preparation and in ^{14}C analytical protocol" between the studies but do not provide details about that. Are these differences substantial? Please provide information if you think they are large enough to possibly explain the differences in the results.

- Sec 3.2.3. It is interesting that your PMF-results for the fossil – non-fossil contributions to eBC differ so much from previous studies, with a very dominant contribution from FF in your study. This is also in contrast to your LHS-results based on ^{14}C-measurements, which indicate that non-fossil EC dominates. The comparison presented in Table 4 is especially important since it shows results from the same time-periods using the two different S-A methods (+ the FLEXPART-model based S-A). The documentation of the PMF-method Platt et al. (2023) is still "in preparation", which it was already in 2021. Since the PMF-based source apportionment in your study gives such different results, you should provide a more detailed description of the methodology (possibly in the supplementary material).

- One of the most interesting and important results of this study is the completely different results for the BB contribution to EC/eBC from the LHS and PMF S-A. This indicates serious problems with either one or both of the S-A approaches. This should be highlighted more clearly in the abstract and a more thorough discussion of these differences and possible explanations is needed before this manuscript can be published in ACP. Could there be problems with the ^{14}C-measurements and/or the EC-OC split in the TOA, leading to some non-fossil OC being detected as EC? Are there any potential problems with the eBC-PMF that could misrepresent eBC_{BB} as eBC_{FF}? If no explanations can be found, it is also an important result – that it is currently not possible to determine the relative contributions of FF and BB to BC/EC with any meaningful accuracy at the Zeppelin site (and presumably also not at other sites where long-range transport is the main source of BC).

- Sec 3.2.4 (Lines 506–516). This whole (short) section about Seasonal footprints for eBC_{BB} and eBC_{FF} is confusing and does not contain substantial information. It should either be completely rewritten or removed from the manuscript. Removal may be the best choice. Additionally, Fig. 6 is difficult to understand. What are these plots really showing? What is meant by "and all footprints for this season"? Why do you use "relative" plots here instead of showing the footprints directly?

- Sec 3.3 (BSOA tracers) and 3.4 up to 3.4.2 (PBAP) are interesting and well-written. The only objection (apart from a few minor typos) is that Sec 3.4.3 (Source apportionment by LHS) should be put in another Section (probably a new Sec 3.5) since it is not only about PBAP.

- Sec 3.4.3 (to be changed to Sec 3.5) Lines 728–731: The sentence about the Zwaaftink(2022)-based tests is unclear. "We increased the PBAP fraction to 11%..." sounds odd; do you mean that the LHS-based PBAP fraction increased to 11% when you used the alternative OC-to-PBAP ER from Zwaaftink et al.? Please specify clearly that the Zwaaftink-based PBAP-estimate is based on another set of tracers than the ones used in the LHS here. As far as I understand, Zwaaftink et al. used the sum of mannitol, arabitol, glucose and trehalose (and no cellulose at all!) while you use mannitol and arabitol (for fungal spores) and cellulose (for debris) in the LHS. Does the 11% that you "increased the PBAP fraction to" exclude the cellulose tracer – or did you include it together with the mannitol+arabitol+glucose+trehalose?

- Carbonate carbon. In Sec 3.5.3 (Episode 3) you mention the presence of carbonate carbon (CC) during the episode. How frequently do you observe CC in the TOA at Zeppelin? It would be interesting to see the amount of CC in Table 1.

- Sec. 4 Implications: You suggest/encourage trend studies and a pan-Arctic investigation of the spatial variability of eBC_{BB} and eBC_{FF} based on the aethalometer-based PMF approach used in your study. This could be interesting but considering the widely different results from the PMF S-A and the tracer-based LHS S-A, do you really trust that the eBC PMF gives correct (reliable/useful) results for the source apportionment? If so, does it mean that you think that the tracer-based S-A is incorrect? To me, your results indicate that we should be careful with using aethalometer-based source apportionment data and not trust that they give an accurate split between BB and FF. I would really like it if you could prove me wrong on this point!

Technical corrections

The manuscript could benefit from thorough language editing. Some parts of the text are well-written and most of it is understandable, but it would be easier and more pleasant to read if the text was revised from a language perspective. According to the Author contribution statement, "All co-authors contributed to writing, reviewing, and editing the final article". Since there are almost 20 authors, some of whom seem to have English as their primary language, I am surprised about the language and the editing of some parts of the manuscript. Did all co-authors really check the entire manuscript before submission? Although ACP may include language editing of accepted manuscripts, it would be helpful if all co-authors reviewed the language and structure of the text before resubmission.

I will not go through every part that I found awkwardly formulated or correct all minor language mistakes (I think that is the job of the author-team) but will only mention some examples and some of the unclear parts.

- The authors use the word "prevail" (and prevailing) frequently and in a strange way. It is used more than ten times in the manuscript and I think incorrectly in most (or all) cases.

  One example (lines 653–655): "Mannitol and arabitol were highly correlated in the NH-season (r2 = 0.983) when levels were elevated, and mannitol to arabitol ratio variability minor, suggesting one common source prevailing." – I think you mean that there was one common dominant source?

Similar cases are found on lines 70, 79, 142, 442, 461, 465, 514, 699, 760, 820 – in most or perhaps all of these cases I think you mean dominant or primary rather than prevailing.

- Abstract, line56–60: The sentence “Intrusions of warm air masses from Siberia…” is too long and needs to be restructured.

Introduction

- Line 97: It might be better to write “Some PBAP are efficient ice nucleating particles” – the reference given (Tobo et al., 2019) is perhaps not the best regarding general INP properties for PBAP since the article focus on dust as a source of INP?
- Line 106–108: The sentence starting with “Overall, elucidating developments in local aerosol emissions and or formation,” is a bit awkward and (at least) the “or” should be removed.
- Line 108–112: The sentence “Meanwhile, understanding and validating these changes…” needs revision – perhaps it is best to split into at least two sentences. Replace “excepting” with “with the exception of” and add “impact on” before “climate and albedo”.
- Line 117: “mostly only” – either “mostly” or “only”
- Line 122: AH abbreviation not explained the first time it is used
- Lines137–143; The long sentence “Moschos et al. (2022) applied…” ends in a confusing way – it is unclear which of the six factors are “prevailing in winter and summer, respectively” and also which of them have “equally large contributions”. This sentence needs to be reformulated in a clearer way.
- The map shown in the right panel of Figure 1 is not of high enough resolution. E.g. the Arctic circle line is hardly visible.

Experimental

- Line 290–292: Consider referring to Table S3 here to provide information about which 13(?) samples were used. Also, clarify what happened to the 05.12.2017–13.12.2017 sample.
- Table S1: Typo in the notes a “0” instead of a “+”:
  TC × F14C-TC = OC × F14C-OC 0 EC × F14C-EC – change 0 to +
- Table S3 and S4: Extend the table headers to explain what N/A means. Also, consider adding information about the start and stop hours of the sampling.
- Lines 319–341: Please add a reference or URL for the ECLIPSEv6b emission data. Also, clarify whether the ECLIPSE emissions were estimated for the actual modelled years (2017–2020) or if they were the same for all years?
  How were the temporal variations of the anthropogenic emissions handled? Did you include some temperature dependent emissions from the residential sector? What about seasonal variations for other sectors?
  Line 331: Clarify what is meant by “in-land waters”. Does this refer to lakes and rivers? Did you not include ocean-going shipping?

Results and discussion

- Table 1: Add an explanation of OC_B in the Table header and specify which months are included for the H-S and NH-S in the table notes
- Figure 3. The figure caption is not completely correct: b) 2-MT is not shown only 2-MET and 2-MT/2-MET; c) the ratio mannitol/arabitol is not shown
- Line 348: Table 2 incorrect, it should be Table 4

- Figure 5. I suggest that you start the Figure caption with “Source apportionment using the LHS approach (Sect. 2.7)” instead of having this at the end. Also specify that the relative fractions are given in percent.
- Line 356: Table S6 should be Tables S5 and S6
- Line 357: Table 6S should be Table S5
- The first sentence of Sec. 3.1 (lines 361–363) needs to be rephrased for clarity.
- Line 365: “second lowest only to levels observed in Antarctica” – rephrase this sentence to improve readability
- Table 2. Specify in the header that all components refer to concentrations in PM_{10}. I also think it would be good to add a column with total PM_{10} concentrations to the Table.
- Line 376: “EC dropped by a factor of two” – should be “EC increased by a factor of two”
- Lines 393–406: This last part of Sec. 3.1 is a bit disorganised and would benefit from a restructuring or rewriting
- Lines 396–397: “, as seen for eBC_{FF}.” what do you mean by that?
- Lines 397–399: the sentence about Marine MSA is confusing. What do you mean by “for calculation”? What calculation did you do for MSA?
- Lines 399–400: The sentence about marine heterogenic polymer-gels seems misplaced here. It is really not part of your results and here it just makes the text more disorganised and confusing.
- Lines 426–427: “but the very high L/M ratios (occasionally > 40) in spring argued to be emissions from crops residue burning in Asia were not observed.” – please rephrase this sentence
- Lines 431–433: the sentence starting with “61 ± 15% of EC, “ is incomplete.
- Lines 444–445: “was unprecedented compared to Winiger et al. (2015)” – what do you mean by this?
- Sec 3.2.3 (Line 451): Add “and FLEXPART modelling” to the Section title
- Lines 458–463: The three sentences starting with “By the crude, but still realistic, assumption” and ending by “emphasizing RWC as a larger source of eBC than WF” need to be reformulated – they are unclear and the language needs improvements.
- Lines 482–483: Confusing sentence – what do you mean by “contributing 64% in 2020”?
- Lines 484–486: There is a discussion of the degradation of levoglucosan during LRT as a possible explanation of the differences in results but would a degradation not rather decrease the estimated BB-fraction (and thus you would have expected it to be lower rather than higher than found by the PMF and FLEXPART)?
- Lines 651–653: the sentence “The annual mean mannitol to arabitol ratio…” is a bit confusing. Either remove the numerical value for Zeppelin (and according to Table S6 it is 1.1±0.5 rather than 1.1±0.2) or add the corresponding value for Birkenes. The end of the sentence “and fungal spores (Bauer et al., 2008)” is unclear.
- Lines 675–676: “no correlations…were obtained” but you give a r^2-value of 0.423 for the NH-season; this is not a very high correlation but I would not consider it as <u>no</u> correlation?
- Lines 707 and 711: the “(here ng C m^{-3})” seem unnecessary and just confusing?
- Lines 714–717: The sentence about the plant debris contribution at low OC levels should be rephrased for clarity. The “and were correlated” part of the sentence is not totally clear about what was correlated but I assume it must be the plant debris contribution and total OC? I think the part “and elevated cellulose levels” is redundant since cellulose is the tracer used to determine the debris contribution. Is it really correct to say that the “plant debris drives observed OC levels” when the contributions are only 5 to 12%?
- Lines750–751 and 795–797: Clarify how the Episode 1 OC-level (2172 ngC/m3 for 5 days) can “explain” 17% of the “annual OC loading” (for 2020) while the Episode 3 OC-level (818 ngC/m3 for 8 days) “explain” 16% of the annual OC loading (also 2020)? If I understand this correctly the

integrated OC level for the Episode 1 is ca 66% larger than that for Episode 3, that is a much larger difference than 17 to 16%?

- Figure 7. No data are given for T<0 so this can be removed from panel B. What does T average represent in panel B? Specify this in the Figure caption.
- Lines 770–772: The sentence about “nutrient-bearing aerosol from Boreal WF” is out of place here! If it is relevant to include in the article it should probably be moved to the introduction.
- Line 784: “corresponding to 13% of their annual loading” – is this for both OC and EC?
- Table 4: Rephrase “Means are based on identical time stamps” – I guess you mean that the BB and FF fractions in the table are for the same time periods for all three methods.
- Table S8: The information presented in this table is unclear, in regards to the data provided. The header should be rephrased to better convey what is shown in the table. The last column “Mean conc. and percentile levoglucosan (ng m^{-3}) (percentile)” is particularly confusing. It would be helpful to specify what is being shown and why it is relevant, especially given the unusual selection of samples. If necessary, add a short text section could be added before the table to explain why this information is being investigated.