Reviewer #4

In this updated version of the manuscript the authors have taken into account most of the comments raised by reviewers, with evident improvements.

Despite the general quality of the manuscript has greatly improved, I still have a few comments that the authors can find below.

R1 - To compare the two series measured at the two considered sites (Ny-Ålesund and GSRS site) the authors apply the Wilcoxon test. From what I know, the use of the Wilcoxon signed-rank test does not appear appropriate in your case, since the two datasets were collected over different time periods and with different temporal resolutions. The signed-rank test requires paired observations (i.e., measurements taken simultaneously or under directly comparable conditions). You also report this at line 245 (matched or dependent observations). In your context, the samples should be treated as independent, and a test designed for unpaired data (e.g., the Mann–Whitney U test) would be more suitable. But maybe this observation is only due to a lack of details in your manuscript. I suggest to apply a different test or to provide further information.

A1 – *We see the raised point, and we thank the Reviewer for this suggestion. After a thorough evaluation, this comparison was excluded, and the Authors decided to focus only on GSRS data, as kindly suggested. The whole manuscript was then revised accordingly.*

R2 - Another thing that I don't really understand about the application of this test is found in line 277, where you report that the test showed a good agreement between the two series except for some sporadic peaks. The test is applied to whole data-series, not to single data points. How is it possible that the test is working good in general but with the exception of these sporadic peaks? Please provide some more detail about this.

A2 – *We agree with the Reviewer. The test was applied to whole data-series, not to single data points. We missed to well-explain this part. What the Authors intended to say here was that the Wilcoxon test showed a general good agreement for most of the investigated species. However, from a visual comparison, sporadic peaks showed minor discrepancies between the two series. In any case, as previous declared, the Authors agreed to focus only on GSRS data, removing the GSRS-NyÅ series comparison.*

R3 - In my first review I was wondering whether focusing only on GSRS would have been better to build a more concise and direct manuscript. The authors replied that the comparison was important as it serves to evaluate the local variability of the considered variables. According to this I still don't completely understand the reason for this comparison. The authors made a comparison (even if maybe

not with the best statistical test); they found that for some variables there is no difference, for some variables there is some difference; in light of this they decide to focus only on GSRS site. This is not very logical to me. The test highlights that for some species there is a significant difference, but the authors do not discuss this. They just say that according to the results they decide to focus only on GSRS. So I don't understand how the evaluation of local variability was carried out. The test shows that for some species there is a significant local variability, but this is never discussed across the manuscript. Maybe I am missing something, but according to this, I recommend again not to include Ny-Alesund data in the manuscript and focus only on GSRS.

A3 – *The Authors thanks the Reviewer for raising this point and agreed to not include the NyÅ-GSRS comparison.*

R4 - A lot of information frequently cited in the main text is still present only in the supplementary. I suggest to include in Table 1 not only seasonal mean values, but also yearly mean values. This would require to add only a few additional rows to the table and make reading easier. Please provide p-values for the determined correlation coefficients (lines 350-352).

A4 – *According to the Reviewer's suggestion, the Authors have added yearly mean values to Table 1 for completeness. However, due to the dynamic conditions at the GSRS site, such as accumulation, melting, and wind drift, these yearly averages may be particularly subjected to spikes. The "Total" reported in Table 1 slightly changes compared to the former Table 1, as the Authors noticed a copy-and-paste mistake from the original table in excel. Previous "Total" was indeed referred to the sum of the most (>1%) and less (<1%) abundant species in the surface snow.*

Additionally, the Authors have updated the following statement "The results indicated generally weak positive correlations ($\rho < 0.5$) for the variables considered." at lines 303-304 by adding "($\rho < 0.5$, p-values < 0.05)", as kindly suggested. P-values were not repeated for all the variables as they were all < 0.05, and this was stated before citing variables one-by-one.

R5 - In the end the correlation study between glaciochemistry and meteorological variables did not bring interesting results. From what I see the composition of snow does not seem to be correlated significantly with meteorological data. According to this I would smooth a bit the conclusions, removing all the discussion related to the Arctic Amplification (lines 640-643). In the end your interpretation shows a link between sea ice dynamics and snow composition, but it does not highlight any correlation between climatology and snow chemistry (also because the investigated time interval is too short).

A5 – *The Authors agreed with this suggestion and smoothed the conclusions accordingly. The discussion related to the Arctic Amplification was thus removed.*

According to the above, I think that once the authors will adjust their manuscript considering these comments, the manuscript will be ready for publication.