Reviewer #1 (Comments to the Author):

Title: On the dynamics of ozone depletion events at Villum Research Station in the High Arctic

This manuscript presents a statistical analysis of near surface ozone observations over a 23 year period at Villum Research Station in the high Arctic, utilizing local meteorological observations, backward air mass trajectory modeling, and statistical analysis to elucidate mechanisms controlling observed ODEs. The dataset and analysis are interesting, and the majority of the discussion section is really well done. However the data analysis suffers from some major issues that need to be rectified.

> We thank the reviewer for their comments and suggestions. We have addressed each comment below with review comments in black, author response in blue, and additions to the original text in red. We have indented the author's response for clarity. Lines numbers given in the author's response refer to lines in the revised manuscript.

The way this paper is written suggests a fundamental misunderstanding of the role of sea ice regions in halogen activation and ozone depletion chemistry. **Sea ice has snow on it!** I'm sure the authors are aware of this fact, but the analysis and discussion give the impression that they believe snow only exists on land. The physical surface of the sea ice itself does not have a pH conducive to halogen activation chemistry (Abbatt et al 2012, Wren et al 2013, Pratt et al 2013). It is the snow in sea ice regions that drives the halogen chemistry. Your analysis and discussion of snow vs sea ice needs to be completely reworked to reflect the complexity of sea ice regions. An analysis of the surface temperature along the back trajectory would potentially help with determining the potential for halogen activation along the back trajectory.

> The authors are fully aware that sea ice has snow on top of it. The satellite products used in this study (National Oceanic and Atmospheric Association/National Environmental Satellite, Data, and Information Service (NOAA/NESDIS) Interactive Multisensor Snow and Ice Mapping System (IMS)) provides information for each grid cell about the underlying surface, where each grid cell can belong to only one of five categories (0: Outside the coverage area, 1: Sea, 2: Land (without snow), 3: Sea ice, 4: Snow covered land). The product does not indicate if the sea ice is covered by snow or not. This is due to the similar spectral signatures of sea ice and snow which makes differentiation difficult, although IMS integrates different data sources, ancillary data, and advanced algorithms for surface mapping, they do not always provide clear delineation between sea ice and snow (U. S. National Ice Center, 2008). In essence, the data source we used cannot definitively discern if there is snow on top of sea ice or not. Therefore, we chose to keep the original labels from the satellite products in our analysis to remain true to the original data product. We do admit this issue with snow vs sea ice detection was not described adequately in the Methods section. We have amended the text to indicate this. We have also addressed this shortcoming in the satellite product in the Results and in the Discussion sections.
>
> Lines 195-213: For each trajectory, a surface-type footprint analysis was performed. The underlying surface types used for the surface footprint type analysis were produced by the National Oceanic and Atmospheric Association/National Environmental Satellite, Data, and Information Service (NOAA/NESDIS) Interactive Multisensor Snow and Ice Mapping System (IMS) developed under the direction of the Interactive Processing Branch (IPB) of the Satellite Services Division (SSD). The altitude at each step along the trajectory was compared to the height of the mixed layer. Steps were classified as being above the mixed layer (AML) if the

trajectory altitude was above this height. If the trajectory altitude was below this height, then the underlying surface type (land without snow, sea, sea ice, or snow on land) was recorded using a polar stereographic map of the Northern Hemisphere classified into 1024×1024 24 km grid cells. It is important to note that grid cells classified as sea ice likely contain snow on the surface, although the satellite products used in this study does not differentiate between bare sea ice and snow-covered sea ice, likely due to the similar spectral signatures between sea ice and snow (U. S. National Ice Center, 2008). We opted to keep the original labels from the satellite product for this analysis, as we cannot make any definitive statements about the presence of snow on top of sea ice. The reader should keep this in mind when interpreting the results. The time spent over different surfaces is expressed as a percentage of the total trajectory length.

Lines 380-381: It should be kept in mind that the air mass history variable, time spent over sea ice, does not give information about the presence of snow cover and only if the underlying surface was classified as sea ice or not.

Lines 502-504: The air mass history variable, time spent over sea ice, does not give information about the presence of snow cover and only if the underlying surface was classified as sea ice or not.

Lines 895-897: It should be noted that the snowpack on top of sea ice is the likely source of these halogens, given that the surface of sea ice is not conducive for halogen activation (Abbatt et al., 2012), although the satellite product used in this study cannot differentiate between snow covered sea ice and bare sea ice (see Methods).

An analysis of the temperature along the trajectories would indeed be an interesting analysis method. We have extensively examined the air mass history and source regions for ODEs within the mixed layer, which agree with previous literature indicating the central Arctic Ocean is the primary source region for ODEs and enhanced halogen levels (Ahmed et al., 2023; Begoin et al., 2010; Bognar et al., 2020; Bottenheim and Chan, 2006; Bougoudis et al., 2020; Oltmans et al., 2012; Seo et al., 2020). Halogen activation requires a frozen, acidic heterogeneous surface (Burd et al., 2017; Jeong et al., 2022; Sander et al., 2006) and given that the temperature in the springtime central Arctic Ocean is usually below freezing, such an analysis would likely give very similar results to the source region analysis already performed in this study thus likely not yielding new information. We feel we have adequately and comprehensively analyzed the source regions of ODEs observed at Villum with our current analysis utilizing frequency maps for trajectory steps below the mixed layer, time over different surface types, their geographic distribution, and their temporal dependences, all of which agrees well with previous studies.

The selection of the time period for further analysis seems arbitrary, as ODEs don't necessarily follow a clear Mar-May pattern as seen in Fig 2, particularly at these high latitudes. The paper would be strengthened if the time period analyzed were empirically defined utilizing the first to last ODE day. You could choose the earliest and latest over the whole study period to have a consistent time frame across years. It might end up being March to May still but at least you would have a better justification for the choice.

We selected the March-May period for an in-depth analysis as this is the main occurrence of ODEs throughout the Arctic as demonstrated by numerous previous studies (Barrie et al., 1988; Bottenheim and Chan, 2006; Simpson et al., 2007; Whaley et al., 2023) and from analyzing the results displayed in Fig. 2, which shows ODEs *are* mainly confined to the spring (March, April, and May) season. Furthermore, from Fig. 2, it is evident that no observations below 10 ppbv occurred in February (likely due to the absence of sunlight) and only a few occurred during the first part of June. Indeed, the ODE frequency for June is 2.37 % which is slightly less than March (3.88 %) and ODEs in June only occurring during the first few days of the month (Fig. 2a). Different environmental conditions during the summer compared to spring also contributed to this decision. June (and other summer months) also regularly experiences temperatures near or above freezing, therefore, the mechanisms behind ozone depletion during the summer are likely different from the spring as halogen propagation needs an acidic frozen, heterogeneous surface (Burd et al., 2017; Jeong et al., 2022; Sander et al., 2006). During the summer months, there is limited transport of ozone and its precursors from the mid-latitudes (diminished sources), the low absolute humidity and low NO_x levels limit in situ photochemical production, and increased areas of open water and bare land increases dry deposition compared to ice covered surfaces (increased sinks) (AMAP, 2015; Barten et al., 2021, 2023). For these reasons, we explicitly selected the months of March to May for a comprehensive and systematic analysis of ODEs and excluded June due to the low frequency of ODEs and different environmental conditions affecting ozone variability. Therefore, our selection is not arbitrary but based on the main occurrence of ODEs across the Arctic as noted in the numerous previous studies and the initial results from this study. If we understand the reviewer's suggestion correctly, using the first and last ODE day to define the ODE season would result in losing the monthly information provided in this study as the analysis would be limited to only ODEs vs Non-ODEs over this period and would not examine any temporal dependencies. Using the first and last ODE days and analyzing each month individually (as done in this study) would only result in a few additional ODEs for the first few days of June. We are also planning a separate publication which analyzes the dynamics of ozone during the summer months (June, July, and August) as described below.

Ozone seems to be persistently below background through the summer months, this is an interesting finding that merits more discussion/analysis. In my view this is a big missed opportunity by the authors particularly given the low number of ozone observations at this latitude and the discussion of the potential role of iodine motivated by the MOSAIC papers (e.g. Benavent et al 2023).

We agree the dynamics of ozone during the summertime is an interesting topic that warrants further discussion. The focus of this paper was confined to the springtime ODEs for reasons described above. Therefore, we were already planning a follow up publication on the dynamics of ozone in the High Arctic summertime. This paper will examine the observations below 10 ppbv in the first weeks of June to determine if they are caused by the same mechanisms as ozone depletion in spring, the role of IO in summertime ozone destruction, the role of entrainment from aloft on ozone levels (preliminary findings indicate that subsidence of dry, ozone rich air from above the mixed layer contributes to enhanced ozone levels). This work is currently under preparation and will be submitted to ACP in due time.

The description and utility of a SHAP value needs to be in the main text as the whole ML discussion relies on the reader having an understanding of those values and being able to interpret them. Additionally, Section 3.4 needs to be revised for clarity, I've read it a few times and I'm not entirely sure what I am supposed to be taking away from this section, especially figure 10. Maybe folks with a background in machine learning will find value here, but the broader community I think is going to be lost.

We have moved the description of the ML methodology including the description of the SHAP methodology to the main text for readers unfamiliar with these concepts. We originally included them in the SI for brevity.

Sect. 3.4 of the Results section describes the results of the ML model and we discuss these results in the Discussion Sect 4.2. We begin by highlighting why we utilized ML and its added benefits over statistical analysis, which were both performed in this study. We then detail the accuracy and applicability of our ML model through an evaluation of its predictive performance using robust and comprehensive evaluation metrics for a classification ML model. The most important features in each month are then described (which is not evident using classical statistical analysis). The relationships between the input features and their contribution to the model prediction are then analyzed. From Fig. 10, the relationship between the input features and their contribution to the model output is displayed, this gives information about how certain levels (and threshold ranges) of the input features affects the model's prediction of an observation being an ODE or not. For example, this is especially evident for solar radiation (Fig. 10d), which shows that after the 112 to 153 W m^{-2} bin range solar radiation starts to make a positive contribution to the model prediction an ODE (i.e., the model is more likely to predict a positive label or an ODE) and below this range it makes a negative contribution (the model is more likely to predict a negative label or Non-ODE). Such a threshold range is not evident from Fig. 4e. Overall, this section shows the ML brings added value to our analysis, our ML model is robust and accurate, the input features that are most important to modeling ODEs, and how the features affect the model prediction.

We have added the following lines to make this clearer in the text:

Lines 566-582: The evaluation metrics of the ML for all spring months combined and individual months are displayed in Table 1. We use three common metrics for evaluating a binary classification ML model: accuracy, recall, and AUC ROC (Area Under Curve Receiver Operating Characteristics). Briefly, accuracy is the fraction of correctly predicted observations regardless of label (ODE vs Non-ODE), recall is the fraction of ODEs correctly predicted and AUC ROC evaluates how well a model can discriminate between positive and negative labels across all decision thresholds for binary classification. In general, the ML model can accurately reproduce ODEs over all spring months combined as evidenced by how all three metrics are close to unity (their maximum value). However, when evaluating the results on an individual monthly basis, there is an increase in the recall metric and decrease in the accuracy and AUC ROC (see Sect. 2.6 for a detailed description of the evaluation metrics) from March to May (Table 1), which is likely connected to the increasing frequency of ODEs from March to May. With increased ODE occurrence, the recall metrics would increase as positive labels (ODEs) are more likely to be identified when they occur more often and the accuracy and AUC ROC metrics would decrease with the increased occurrence of positive labels due to a concurrent increase in number of incorrectly labeled ODEs. The ML model is also free from over-fitting given the close agreement between the train and test sets. Overall, this ML model is sufficiently accurate, robust, and suitable for the investigation of ODEs.

Caption of Table 1: The accuracy gives an overview of the model performance for both labels (ODEs vs Non-ODEs), recall gives the model performance only for positive labels (ODEs), and AUC ROC evaluates the model performance over different decision thresholds, together, these three metrics give a comprehensive view of the model's performance. The three metrics range from 0 (worst) to 1 (best).

Lines 593-605: The SHAP approach is designed to estimate the importance of each input variable to the model output based on coalitional game theory (Molnar, 2022) (see Sect. 2.6 for a more detailed description). SHAP values represent the marginal contribution of each input variable to the model output, or in other words: how each observation for each variable affects the model's prediction. SHAP values can be positive or negative, with positive values indicating a variable is more likely to contribute to an observation being predicted as an ODE while negative values mean a variable is more likely to contribute to an observation being labeled as a Non-ODE. SHAP can produce both local and global explanations. The global importance gives an overview of the most important variables to the model output. The local importance of each observation can give information about the relationship between the SHAP and input values (positive or negative relationship, linear or non-linear), or in other words how does the model output vary over the range of input values.

Minor points:

Line 284: Given that high wind speed enhances vertical mixing it is not surprising that ozone would not be depleted during those conditions.

As noted in the literature, both low and high wind speed can have an effect on ozone variability (Blechschmidt et al., 2016; Choi et al., 2012; Jones et al., 2009; Zhao et al., 2016), therefore while maybe not surprising this observation is of importance. We have noted the dual effect of wind speed on ODEs in the Results and Discussion and we will add this insight to our discussion.

Lines 841-846: High wind speeds can also enhance vertical mixing of ozone enriched air masses from aloft, which could mask the contribution of halogen activation from blowing snow. Only during May does high wind speeds regularly make a positive contribution to the model output, and the magnitude of this contribution is small (Fig. 10b). Overall, the rare occurrence of high wind speeds (Fig. S4b) hinders any definitive conclusions about their effect on ODEs.

Abbatt, J. P. D., Thomas, J. L., Abrahamsson, K., Boxe, C., Granfors, A., Jones, A. E., King, M. D., Saiz-Lopez, A., Shepson, P. B., Sodeau, J., Toohey, D. W., Toubin, C., von Glasow, R., Wren, S. N., and Yang, X.: Halogen activation via interactions with environmental ice and snow in the polar lower troposphere and other regions, Atmos. Chem. Phys., 12, 6237–6271, https://doi.org/10.5194/acp-12-6237-2012, 2012.

Ahmed, S., Thomas, J. L., Angot, H., Dommergue, A., Archer, S. D., Bariteau, L., Beck, I., Benavent, N., Blechschmidt, A.-M., Blomquist, B., Boyer, M., Christensen, J. H., Dahlke, S., Dastoor, A., Helmig, D., Howard, D., Jacobi, H.-W., Jokinen, T., Lapere, R., Laurila, T., Quéléver, L. L. J., Richter, A., Ryjkov, A., Mahajan, A. S., Marelle, L., Pfaffhuber, K. A., Posman, K., Rinke, A., Saiz-Lopez, A., Schmale, J., Skov, H., Steffen, A., Stupple, G., Stutz, J., Travnikov, O., and Zilker, B.: Modelling the coupled mercury-halogen-ozone cycle in the central Arctic during spring, Elementa: Science of the Anthropocene, 11, 00129, https://doi.org/10.1525/elementa.2022.00129, 2023.

AMAP: AMAP Assessment 2015: Black carbon and ozone as Arctic climate forcers., Arctic Monitoring and Assessment Programme (AMAP), 116, 2015.

Barrie, L. A., Bottenheim, J. W., Schnell, R. C., Crutzen, P. J., and Rasmussen, R. A.: Ozone destruction and photochemical reactions at polar sunrise in the lower Arctic atmosphere, Nature, 334, 138–141, https://doi.org/10.1038/334138a0, 1988.

Barten, J. G. M., Ganzeveld, L. N., Steeneveld, G.-J., and Krol, M. C.: Role of oceanic ozone deposition in explaining temporal variability in surface ozone at High Arctic sites, Atmos. Chem. Phys., 21, 10229–10248, https://doi.org/10.5194/acp-21-10229-2021, 2021.

Barten, J. G. M., Ganzeveld, L. N., Steeneveld, G.-J., Blomquist, B. W., Angot, H., Archer, S. D., Bariteau, L., Beck, I., Boyer, M., von der Gathen, P., Helmig, D., Howard, D., Hueber, J., Jacobi, H.-W., Jokinen, T., Laurila, T., Posman, K. M., Quéléver, L., Schmale, J., Shupe, M. D., and Krol, M. C.: Low ozone dry deposition rates to sea ice during the MOSAiC field campaign: Implications for the Arctic boundary layer ozone budget, Elementa: Science of the Anthropocene, 11, 00086, https://doi.org/10.1525/elementa.2022.00086, 2023.

Begoin, M., Richter, A., Weber, M., Kaleschke, L., Tian-Kunze, X., Stohl, A., Theys, N., and Burrows, J. P.: Satellite observations of long range transport of a large BrO plume in the Arctic, Atmospheric Chemistry and Physics, 10, 6515–6526, https://doi.org/10.5194/acp-10-6515-2010, 2010.

Blechschmidt, A.-M., Richter, A., Burrows, J. P., Kaleschke, L., Strong, K., Theys, N., Weber, M., Zhao, X., and Zien, A.: An exemplary case of a bromine explosion event linked to cyclone development in the Arctic, Atmos. Chem. Phys., 16, 1773–1788, https://doi.org/10.5194/acp-16-1773-2016, 2016.

Bognar, K., Zhao, X., Strong, K., Chang, R. Y.-W., Frieß, U., Hayes, P. L., McClure-Begley, A., Morris, S., Tremblay, S., and Vicente-Luis, A.: Measurements of Tropospheric Bromine Monoxide Over Four Halogen Activation Seasons in the Canadian High Arctic, J. Geophys. Res. Atmos., 125, e2020JD033015, https://doi.org/10.1029/2020jd033015, 2020.

Bottenheim, J. W. and Chan, E.: A trajectory study into the origin of spring time Arctic boundary layer ozone depletion, J. Geophys. Res. Atmos., 111, https://doi.org/10.1029/2006JD007055, 2006.

Bougoudis, I., Blechschmidt, A.-M., Richter, A., Seo, S., Burrows, J. P., Theys, N., and Rinke, A.: Long-term time series of Arctic tropospheric BrO derived from UV–VIS satellite remote sensing and its relation to first-year sea ice, Atmos. Chem. Phys., 20, 11869–11892, https://doi.org/10.5194/acp-20-11869-2020, 2020.

Burd, J. A., Peterson, P. K., Nghiem, S. V., Perovich, D. K., and Simpson, W. R.: Snowmelt onset hinders bromine monoxide heterogeneous recycling in the Arctic, J. Geophys. Res. Atmos., 122, 8297–8309, https://doi.org/10.1002/2017jd026906, 2017.

Choi, S., Wang, Y., Salawitch, R. J., Canty, T., Joiner, J., Zeng, T., Kurosu, T. P., Chance, K., Richter, A., Huey, L. G., Liao, J., Neuman, J. A., Nowak, J. B., Dibb, J. E., Weinheimer, A. J., Diskin, G., Ryerson, T. B., da Silva, A., Curry, J., Kinnison, D., Tilmes, S., and Levelt, P. F.: Analysis of satellite-derived Arctic tropospheric BrO columns in conjunction with aircraft measurements during ARCTAS and ARCPAC, Atmospheric Chemistry and Physics, 12, 1255–1285, https://doi.org/10.5194/acp-12-1255-2012, 2012.

Jeong, D., McNamara, S. M., Barget, A. J., Raso, A. R. W., Upchurch, L. M., Thanekar, S., Quinn, P. K., Simpson, W. R., Fuentes, J. D., Shepson, P. B., and Pratt, K. A.: Multiphase Reactive Bromine Chemistry during Late Spring in the Arctic: Measurements of Gases, Particles, and Snow, ACS Earth Space Chem., 6, 2877–2887, https://doi.org/10.1021/acsearthspacechem.2c00189, 2022.

Jones, A. E., Anderson, P. S., Begoin, M., Brough, N., Hutterli, M. A., Marshall, G. J., Richter, A., Roscoe, H. K., and Wolff, E. W.: BrO, blizzards, and drivers of polar tropospheric ozone depletion events, Atmos. Chem. Phys., 9, 4639–4652, https://doi.org/10.5194/acp-9-4639-2009, 2009.

Molnar, C.: Interpretable Machine Learning: A Guide for Making Black Box Models Explainable, 2nd ed., 2022.

Oltmans, S. J., Johnson, B. J., and Harris, J. M.: Springtime boundary layer ozone depletion at Barrow, Alaska: Meteorological influence, year-to-year variation, and long-term change, J. Geophys. Res. Atmos., 117, https://doi.org/10.1029/2011JD016889, 2012.

Sander, R., Burrows, J., and Kaleschke, L.: Carbonate precipitation in brine – a potential trigger for tropospheric ozone depletion events, Atmospheric Chemistry and Physics, 6, 4653–4658, https://doi.org/10.5194/acp-6-4653-2006, 2006.

Seo, S., Richter, A., Blechschmidt, A.-M., Bougoudis, I., and Burrows, J. P.: Spatial distribution of enhanced BrO and its relation to meteorological parameters in Arctic and Antarctic sea ice regions, Atmos. Chem. Phys., 20, 12285–12312, https://doi.org/10.5194/acp-20-12285-2020, 2020.

Simpson, W. R., von Glasow, R., Riedel, K., Anderson, P., Ariya, P., Bottenheim, J., Burrows, J., Carpenter, L. J., Frieß, U., Goodsite, M. E., Heard, D., Hutterli, M., Jacobi, H. W., Kaleschke, L., Neff, B., Plane, J., Platt, U., Richter, A., Roscoe, H., Sander, R., Shepson, P., Sodeau, J., Steffen, A., Wagner, T., and Wolff, E.: Halogens and their role in polar boundary-layer ozone depletion, Atmos. Chem. Phys., 7, 4375–4418, https://doi.org/10.5194/acp-7-4375-2007, 2007.

U. S. National Ice Center: IMS Daily Northern Hemisphere Snow and Ice Analysis at 1 km, 4 km, and 24 km Resolutions, Version 1, , https://doi.org/10.7265/N52R3PMC, 2008.

Whaley, C. H., Law, K. S., Hjorth, J. L., Skov, H., Arnold, S. R., Langner, J., Pernov, J. B., Bergeron, G., Bourgeois, I., Christensen, J. H., Chien, R.-Y., Deushi, M., Dong, X., Effertz, P., Faluvegi, G., Flanner, M., Fu, J. S., Gauss, M., Huey, G., Im, U., Kivi, R., Marelle, L., Onishi, T., Oshima, N., Petropavlovskikh, I., Peischl, J., Plummer, D. A., Pozzoli, L., Raut, J.-C., Ryerson, T., Skeie, R., Solberg, S., Thomas, M. A., Thompson, C., Tsigaridis, K., Tsyro, S., Turnock, S. T., von Salzen, K., and Tarasick, D. W.: Arctic tropospheric ozone: assessment of current knowledge and model performance, Atmos. Chem. Phys., 23, 637–661, https://doi.org/10.5194/acp-23-637-2023, 2023.

Zhao, X., Strong, K., Adams, C., Schofield, R., Yang, X., Richter, A., Friess, U., Blechschmidt, A.-M., and Koo, J.-H.: A case study of a transported bromine explosion event in the Canadian high arctic, J. Geophys. Res. Atmos., 121, 457–477, https://doi.org/10.1002/2015JD023711, 2016.

Reviewer #2 (Comments to the Author):

Title: On the dynamics of ozone depletion events at Villum Research Station in the High Arctic

This paper describes a study on the parameters impacting ODEs in the high Artic. It uses a comprehensive set of atmospheric parameters and sea ice conditions together with back trajectory analyses to investigate the sources of ODEs. Apart from a statistical analysis, a machine-learning algorithm is used to identify the most important parameters affecting ozone.

The paper is very well written and addresses important processes in a part of the Earth's atmosphere that is most vulnerable to climate change. It is therefore of high scientific relevance and fits well into the scope of ACP. The way the data analysis is performed and the results are discussed are appropriate given the complexity and multiphase nature of halogen release and ozone depletion events in Polar Regions. There are, however, a few points that should be addressed before final publication:

> We thank the reviewer for their comments and suggestions. We have addressed each comment below with review comments in black, author response in blue, and additions to the original text in red. We have indented the author's response for clarity. Line numbers given in the author's response refer to lines in the revised manuscript.

The authors do not pay appropriate credit to former studies on tropospheric ozone depletion which apply very similar methods as the present study. In particular, Frieß et al. [2023] presents statistical analysis of multi-decadal ozone (and BrO) observations based on back-trajectories and sea ice data, just as this present study, but for the Antarctic. The paper would benefit a lot from a discussion on the possible similarities and differences between the drivers of ODEs in both hemispheres via comparison of the results from both studies.

> We thank the reviewer for bringing this valuable reference to our attention. While Frieß et al. (2023) uses a similar methodology, they do not explicitly investigate the relationship between formally defined ozone depletion events and meteorological/air mass history variables but rather focus on vertically resolved BrO mixing ratios. Nevertheless, Frieß et al. (2023) is a valuable resource, one which the authors were not aware of, therefore, we have cited Frieß et al. (2023) on lines 84, 182-184, 192-194, 802-804, 837-839, 868, 894-895, and 932-933.

I am not an expert in machine learning and I have to admit that I was quite lost while reading Section 3.4. I could imagine that other experts in atmospheric physics and chemistry, but not in machine learning, would experience the same. I therefore feel that Sections 2.6 (ML methods) and 3.4 (ML results) require substantial revision as discussed in more detail in the specific comments below.

> We originally moved the two pages of text describing the ML methods to the SI to reduce the overall length of the article. We admit this was not the correct decision in hindsight. Therefore, we have moved the entire description of the ML methods to the main text.
>
> We think that the science in these sections is sound so what needs to be improved is the readability and clarity. We have thus clarified/simplified parts of the text to make it more understandable by readers inexperienced with ML
>
> We have added the following lines to make this clearer in the text:
>
> Lines 261-265: Cross validation involves splitting the training data in 10 equally sized folds (or groups), training the model using nine folds and testing the model using the remaining fold. This was repeated 10 times to use each fold as a test set once. The final evaluation metrics were

averaged using the arithmetic mean to select the optimal hyperparameters and make an overall evaluation of the model performance.

Lines 566-582: The evaluation metrics of the ML for all spring months combined and individual months are displayed in Table 1. We use three common metrics for evaluating a binary classification ML model: accuracy, recall, and AUC ROC (Area Under Curve Receiver Operating Characteristics). Briefly, accuracy is the fraction of correctly predicted observations regardless of label (ODE vs Non-ODE), recall is the fraction of ODEs correctly predicted and AUC ROC evaluates how well a model can discriminate between positive and negative labels across all decision thresholds for binary classification. In general, the ML model can accurately reproduce ODEs over all spring months combined as evidenced by how all three metrics are close to unity (their maximum value). However, when evaluating the results on an individual monthly basis, there is an increase in the recall metric and decrease in the accuracy and AUC ROC (see Sect. 2.6 for a detailed description of the evaluation metrics) from March to May (Table 1), which is likely connected to the increasing frequency of ODEs from March to May. With increased ODE occurrence, the recall metrics would increase as positive labels (ODEs) are more likely to be identified when they occur more often and the accuracy and AUC ROC metrics would decrease with the increased occurrence of positive labels due to a concurrent increase in number of incorrectly labeled ODEs. The ML model is also free from over-fitting given the close agreement between the train and test sets. Overall, this ML model is sufficiently accurate, robust, and suitable for the investigation of ODEs.

Caption of Table 1: The accuracy gives an overview of the model performance for both labels (ODEs vs Non-ODEs), recall gives the model performance only for positive labels (ODEs), and AUC ROC evaluates the model performance over different decision thresholds, together, these three metrics give a comprehensive view of the model's performance. The three metrics range from 0 (worst) to 1 (best).

Lines 594-605: The SHAP approach is designed to estimate the importance of each input variable to the model output based on coalitional game theory (Molnar, 2022) (see Sect. 2.6 for a more detailed description). SHAP values represent the marginal contribution of each input variable to the model output, or in other words: how important each variable is to the model for making a prediction. SHAP values can be positive or negative, with positive values indicating a variable is more likely to contribute to an observation being predicted as an ODE while negative values mean a variable is more likely to contribute to an observation being labeled as a Non-ODE. SHAP can produce both local and global explanations. The global importance gives an overview of the most important variables to the model output. The local importance of each observation can give information about the relationship between the SHAP and input values (positive or negative relationship, linear or non-linear), or in other words how does the model output vary over the range of input values.

The abstract is quite short. It would be important to provide some more specific information on this study (e.g., measurement site, observation period, etc.).

The Atmospheric Chemistry and Physics guidelines on manuscript formatting (https://www.atmospheric-chemistry-and-physics.net/submission.html#getready) specify a maximum abstract length of 250 words, which the abstract is currently at. The measurement site is indicated on lines 12 and 13 and in the title and we have added the years on line 13.

Specific Comments

L17: It is not clear what you mean with "increasing". Is this a seasonal tendency or an increase over the years?

> Our intent was to describe the results of our trend analysis for these two parameters. We have added the words. We have amended this sentence.
>
> Line 17-18: Positive trends in ODE frequency and duration are observed during May (low confidence) and April (high confidence), respectively.
>
> We have also changed the language throughout the manuscript to indicate the direction of trends as positive or negative where appropriate instead of increasing or decreasing.

L64: Please explain what you mean with "relative rate principle"

> Relative rate principle is a standard method for investigating atmospheric chemistry and is used widely in laboratory studies of unknown reaction rate constants by reacting with a compound with a known rate constant with another compound with unknown rate constant (Finlayson and Pitts, 1986). An expression is obtained independent of the mutual reactant concentration and a fitted line is obtained consisting of ratio between the known rate constant and unknown rate constant and thus the unknown rate constant can be calculated from the slope and known constant; therefore, the name “relative rate principle”. We have earlier used this approach to demonstrate that Br is reacting with O_3 and gaseous elemental mercury (GEM) rather than Cl (Skov et al., 2004, 2020).

Section 2.4: It should be pointed out that the back-trajectory analysis applied here is very similar to the methods by Frieß et al. [2023].

> We have added a sentence which shows that this methodology has been applied to previous studies including Frieß et al. (2023).
>
> Lines 192-194: This methodology has been utilized by previous studies to systematically analyze the geographic origns of air masses (Dall’Osto et al., 2017, 2018; Frieß et al., 2023; Heslin-Rees et al., 2020; Pernov et al., 2022).

Sections 2.6: The description of the machine learning algorithm is far too short. I think the reader should be able to get at least a basic understanding of the model without going through the detailed description in the supplemental material. See also my comments to Section 3.4 below.

> We admit the Methods section describing the ML algorithm is short. This is due to the majority of the text being moved to the supplement in an effort to shorten the length of the paper. We have moved the entire section describing the ML methodology to the main text as outlined above and below.

L267: It is not clear to me what you mean with "monthly hours within the same bin".

> The word “monthly” should not have been included in this sentence and we have removed it. We thank the reviewer for this good catch.

Section 3.4: This section is hard to understand for readers inexperienced in machine learning. I do not have any clue what to learn from Table 1, except that high numbers are good. What is a "cross validation score"? What is "Aera Under Curve Receiver Operating Characteristics"? What does "Recall" mean? The following discussion is mainly based on SHAP values. The explanation of this parameter should therefore be moved from the Supplemental to Section 2.6.

We have now moved the entire ML methodology section to the main text to aid inexperienced readers.

The next sections of Sect 3.4 describe the efficacy of our ML model using common evaluation metrics for classification tasks, the most important variables to the model output, and how they affect the ML model's prediction of ODEs using the SHAP methodology. Detailed descriptions of the evaluation metrics and other ML terms are now included in the methods section. We show that our model is capable of reproducing ODEs and it gives physically meaningful results. The input variable importance is not revealed from the statistical analysis demonstrating the added value of ML. The relationships between ambient and SHAP values demonstrates how individual observations contributes to the model's prediction of ODEs, which largely agrees with the statistical analysis but reveals intricacies that are not borne out of the statistical analysis (e.g., threshold values for positive prediction of ODEs) as the SHAP methodology accounts for dependencies between variables in the model something the statistical analysis does not.

We have added the following lines to make this clearer in the text:

Lines 261-265: Cross validation involves splitting the training data in 10 equally sized folds (or groups), training the model using nine folds and testing the model using the remaining fold. This was repeated 10 times to use each fold as a test set once. The final evaluation metrics were averaged using the arithmetic mean to select the optimal hyperparameters and make an overall evaluation of the model performance.

Lines 566-582: The evaluation metrics of the ML for all spring months combined and individual months are displayed in Table 1. We use three common metrics for evaluating a binary classification ML model: accuracy, recall, and AUC ROC (Area Under Curve Receiver Operating Characteristics). Briefly, accuracy is the fraction of correctly predicted observations regardless of label (ODE vs Non-ODE), recall is the fraction of ODEs correctly predicted and AUC ROC evaluates how well a model can discriminate between positive and negative labels across all decision thresholds for binary classification. In general, the ML model can accurately reproduce ODEs over all spring months combined as evidenced by how all three metrics are close to unity (their maximum value). However, when evaluating the results on an individual monthly basis, there is an increase in the recall metric and decrease in the accuracy and AUC ROC (see Sect. 2.6 for a detailed description of the evaluation metrics) from March to May (Table 1), which is likely connected to the increasing frequency of ODEs from March to May. With increased ODE occurrence, the recall metrics would increase as positive labels (ODEs) are more likely to be identified when they occur more often and the accuracy and AUC ROC metrics would decrease with the increased occurrence of positive labels. The ML model is also free from over-fitting given the close agreement between the train and test sets. Overall, this ML model is sufficiently accurate, robust, and suitable for the investigation of ODEs.

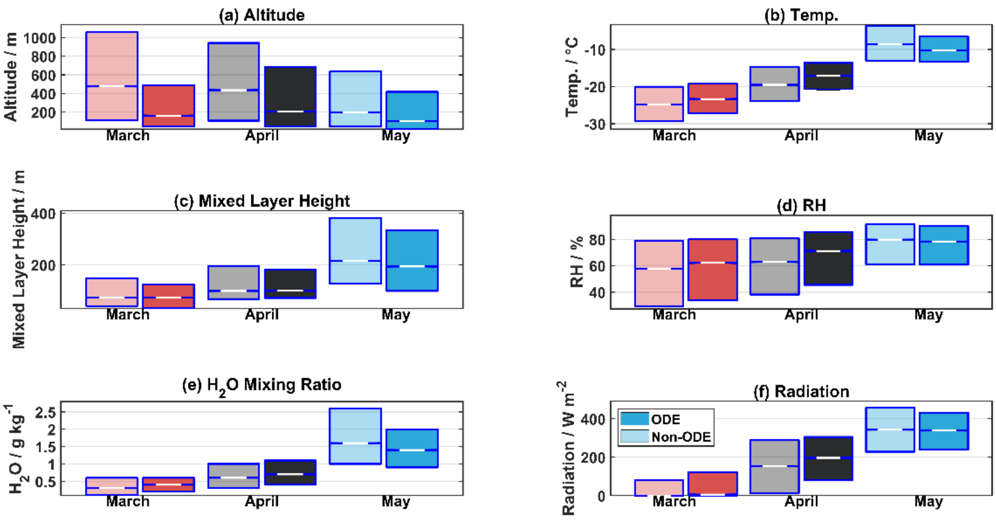
Caption of Table 1: The accuracy gives an overview of the model performance for both labels (ODEs vs Non-ODEs), recall gives the model performance only for positive labels (ODEs), and AUC ROC evaluates the model performance over different decision thresholds, together, these three metrics give a comprehensive view of the model's performance. The three metrics range from 0 (worst) to 1 (best).

Lines 594-605: The SHAP approach is designed to estimate the importance of each input variable to the model output based on coalitional game theory (Molnar, 2022) (see Sect. 2.6 for a more detailed description). SHAP values represents the marginal contribution of each input variable to the model output, or in other words: how important each variable is to the model for making a prediction. SHAP values can be positive or negative, with positive values indicating a variable is more likely to contribute to an observation being predicted as an ODE while negative values mean a variable is more likely to contribute to an observation being labeled as a Non-ODE. SHAP can produce both local and global explanations. The global importance gives an overview of the most important variables to the model output. The local importance of each observation can give information about the relationship between the SHAP and input values (positive or negative relationship, linear or non-linear), or in other words how does the model output vary over the range of input values.

L615ff: You state that in situ radiation measurements would not be available for the entire measurement period, and would also not be indicative for the radiation along the trajectory. Is there any reason why you do not use radiation along the trajectory, which is part of the Hysplit model output?

An analysis of the radiation along the trajectories would indeed be an interesting analysis method. The radiation from HYSPLIT (produced from the NCEP/NCAR reanalysis data) is for the Earth's surface and is not output for an air mass' vertical position (Kalnay et al., 1996), which could affect its interpretation. The in situ meteorological variables give interpretable results which are in line with the current theory of ODEs. We produced boxplots of the meteorological variables from HYSPLIT, similar to Figure 5 in the main text, for all trajectory steps. This shows that the HYSPLIT meteorological variables display similar distributions to the in situ variables and therefore would not add new results and therefore were not included in this manuscript. We also decided to not use the meteorological data output by HYSPLIT since this would necessitate an investigation of all meteorological variables from HYSPLIT. We felt this would substantially increase the length of the manuscript. We have added text in the "Summary and Outlook" section stating that this would be an avenue for further research.

Lines 1039-1040: Analyzing meteorological conditions along the trajectory path (e.g., temperature and radiation) would help extrapolate the observations from individual stations to the larger Arctic region.



L622ff: It would be worth mentioning here that an important process that promotes bromine release at lower temperatures is carbonate precipitation from the sea ice, which reduces its buffer capacity and facilitates acidification [Sander et al., 2006]

We have added a sentence mentioning this process and reference in this paragraph.

Lines 775-777: Cold temperatures also facilitate calcium carbonate precipitation from sea ice which acidifies and lowers the buffering capacity of the salty sea ice surface thus promoting halogen release (Sander et al., 2006).

L649: It is not true that a relationship between RH and ODEs has not been reported before - see Frieß et al. [2023].

Frieß et al. (2023) reports Pearson correlations coefficients between surface measurements of RH and O_3, however, they do not explicitly investigate the relationship between RH and *ODEs* and while they observe negative correlations they do not discuss these correlations between RH and O_3 in the text.

We have added "in the Arctic" in this sentence to indicate the author's intent this was exclusively for the Arctic. We have added a sentence after this one showing that the relationship between RH and ozone has been explored in Antarctica.

Lines 802-804: However, the relationship between RH and ozone has been explored in Antarctica by Frieß et al. (2023), who showed negative correlations at Neumayer and Arrival Heights, supporting observations made in this study.

L825: This is not a new finding. Replace "Our results show..." with "Our results confirm...".

We have removed the phrase "Our results show that" from Line 845.

Technical Comments

L40: "long range" -> "long-range"

We have made this change to the manuscript.

L120: What is the meaning of "i.d."?

"i.d." stands for "inner diameter". We have added this abbreviation to Line 124.

L148: Maybe the term "accept" would be more appropriate than "require" here.

We feel the current wording is more reflective of the input data requirements for machine learning algorithms and will keep it as is.

L152: Start a new sentence after "horizontal plane".

Starting a new sentence after "horizontal plane" would result in the next sentence being too short "This includes both direct and diffuse radiation". To address the reviewer's comment while avoiding this short sentence, we have rearranged the text to:

Lines 155-157: ERA5 output of "shortwave solar radiation downwards" was used, which is the amount of shortwave downwelling solar radiation (including both direct and diffuse radiation) that reaches the Earth's surface on a horizontal plane.

L179: Either state "below mixed layer HEIGHT" or "within the mixed layer".

We have added the word “height” to line 191.

L192: "A trend analysis of trends...": please rewrite.

We have removed the words “of trends” from this sentence.

L295: Remove "For temperatures" at the beginning of the sentence.

In each paragraph of this section, we present the results for each variable separately. We have adopted a coherent, standardized structure for starting paragraphs in this section, with the variable name at the beginning to allow the reader to easily grasp the description of each variable. We discuss each variable in the same order throughout the manuscript. This will facilitate readers to easily access the pertinent information for each variable. We have adopted the same structure in the Discussion section to remain consistent. For these reasons, we have opted to keep the original text on line 403 of the revised manuscript.

Figure 10: It seems that the y-axis scale refers to the lines (SHAP values), but the histograms have different units. So probably a second y-axis on the right needs to be added for the histograms.

To satisfy the reviewers comment, we have replaced Figure 10, S9, and S10 with the ones showing the relative frequency for the histograms on the right axis. The relative frequency is proportional to the histogram count and gives a more intuitive indication of the data distribution. We originally omitted the y-axis labels for the histograms for clarity as mentioned in the caption of Fig. 10.

L572: What do you mean with "SS"? Define acronym/abbreviation.

The acronym “SS” is defined as “statistically significant” on lines 216-217 in Sect. 2.5 “Trend Analysis” of the Methods.

Reviewer’s References

Bognar, K., Zhao, X., Strong, K., Chang, R. Y.-W., Frieß, U., Hayes, P. L., McClure-Begley, A., Morris, S., Tremblay, S., and Vicente-Luis, A.: Measurements of Tropospheric Bromine Monoxide Over Four Halogen Activation Seasons in the Canadian High Arctic, Journal of Geophysical Research: Atmospheres, 125, e2020JD033015, https://doi.org/https://doi.org/10.1029/2020JD033015, 2020.

Frieß, U., Kreher, K., Querel, R., Schmithüsen, H., Smale, D., Weller, R., and Platt, U.: Source mechanisms andtransport patterns of tropospheric bromine monoxide: findings from long-term multi-axis differential optical absorption spectroscopy measurements at two Antarctic stations, Atmospheric Chemistry and Physics, 23, 3207–3232, https://doi.org/10.5194/acp-23-3207-2023, 2023.

Sander, R., Burrows, J., and Kaleschke, L.: Carbonate precipitation in brine - a potential trigger for tropospheric ozone depletion events, Atmos. Chem. Phys., 6, 4653–4658, https://doi.org/10.5194/acp-6-4653-2006, 2006.

References from Author’s Reply

Dall'Osto, M., Beddows, D. C. S., Tunved, P., Krejci, R., Ström, J., Hansson, H. C., Yoon, Y. J., Park, K.-T., Becagli, S., Udisti, R., Onasch, T., O´Dowd, C. D., Simó, R., and Harrison, R. M.: Arctic sea ice melt leads to atmospheric new particle formation, Sci. Rep., 7, 3318, https://doi.org/10.1038/s41598-017-03328-1, 2017.

Dall'Osto, M., Geels, C., Beddows, D. C. S., Boertmann, D., Lange, R., Nojgaard, J. K., Harrison, R. M., Simo, R., Skov, H., and Massling, A.: Regions of open water and melting sea ice drive new particle formation in North East Greenland, Sci. Rep., 8, https://doi.org/10.1038/s41598-018-24426-8, 2018.

Finlayson, B. J. and Pitts, J. N.: Atmospheric chemistry: Fundamentals and experimental techniques, John Wiley and Sons Inc.,New York, NY, United States, 1986.

Frieß, U., Kreher, K., Querel, R., Schmithüsen, H., Smale, D., Weller, R., and Platt, U.: Source mechanisms and transport patterns of tropospheric bromine monoxide: findings from long-term multi-axis differential optical absorption spectroscopy measurements at two Antarctic stations, Atmospheric Chemistry and Physics, 23, 3207–3232, https://doi.org/10.5194/acp-23-3207-2023, 2023.

Heslin-Rees, D., Burgos, M., Hansson, H. C., Krejci, R., Ström, J., Tunved, P., and Zieger, P.: From a polar to a marine environment: has the changing Arctic led to a shift in aerosol light scattering properties?, Atmos. Chem. Phys., 20, 13671–13686, https://doi.org/10.5194/acp-20-13671-2020, 2020.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project, Bulletin of the American Meteorological Society, 77, 437–472, https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.

Molnar, C.: Interpretable Machine Learning: A Guide for Making Black Box Models Explainable, 2nd ed., 2022.

Pernov, J. B., Beddows, D., Thomas, D. C., Dall´Osto, M., Harrison, R. M., Schmale, J., Skov, H., and Massling, A.: Increased aerosol concentrations in the High Arctic attributable to changing atmospheric transport patterns, npj Clim Atmos Sci, 5, 1–13, https://doi.org/10.1038/s41612-022-00286-y, 2022.

Sander, R., Burrows, J., and Kaleschke, L.: Carbonate precipitation in brine – a potential trigger for tropospheric ozone depletion events, Atmospheric Chemistry and Physics, 6, 4653–4658, https://doi.org/10.5194/acp-6-4653-2006, 2006.

Skov, H., Christensen, J. H., Goodsite, M. E., Heidam, N. Z., Jensen, B., Wahlin, P., and Geernaert, G.: Fate of elemental mercury in the arctic during atmospheric mercury depletion episodes and the load of atmospheric mercury to the arctic, Environ. Sci. Technol., 38, 2373–2382, https://doi.org/10.1021/es030080h, 2004.

Skov, H., Hjorth, J., Nordstrøm, C., Jensen, B., Christoffersen, C., Bech Poulsen, M., Baldtzer Liisberg, J., Beddows, D., Dall'Osto, M., and Christensen, J. H.: Variability in gaseous elemental mercury at Villum Research Station, Station Nord, in North Greenland from 1999 to 2017, Atmos. Chem. Phys., 20, 13253–13265, https://doi.org/10.5194/acp-20-13253-2020, 2020.

Reviewer #3 (Comments to the Author):

Title: On the dynamics of ozone depletion events at Villum Research Station in the High Arctic

In this study, Pernov et al. investigate ozone and ozone depletion events (ODE) over a long time period (1996-2019) at the Villum Research Station located northeast of Greenland. A statistical analysis and machine learning (ML) approach is used to analyze the relation of ozone and meteorological variables as well as back-trajectories to study air mass history and surface properties. ODE frequency and duration were found to be highest in May, declining in April and March. Sunny and calm conditions connected with northerly winds seem to favor ODEs in Villum. The ML model revealed that radiation, time over sea ice, and temperature seem to be the most important variables for modeled ODEs during spring time.

To my knowledge, there has been no study which applied an ML approach to investigate ODEs at a specific location. This approach adds some further information regarding the interaction between variables and indicates threshold values for some variables that contribute to ODEs. However, since ML is still a fairly new method and probably relatively unknown to some in this community, I suggest, to include large parts of the ML description from the supplements into the main text, especially the explanation of SHAP values (see below in 'Specific comments).

> We originally moved the two pages of text describing the ML methods to the SI to reduce the overall length of the article. We admit this was not the correct decision in hindsight. Therefore, we have moved the entire description of the ML methods to the main text.

Further, I am uncertain about why one week was chosen for back-trajectories, as the ODEs at Villum are usually limited to a few hours, with only a few exceptional cases extending to several days. In addition, it was found that ODEs mainly occur under calm and stable meteorological conditions, which would suggest only minor transport of air masses during ODEs. This long time span could bias the analysis, particularly when examining the time above the mixing layer, which occurs towards the later part of the trajectory (see below in 'General comments' and 'Specific comments'). But overall, this paper is a pleasure to read, particularly the results and discussion parts are very well-executed. Therefore, I recommend publication in ACP with minor revision.

> We thank the reviewer for their comments and suggestions. We have addressed each comment below with review comments in black, author response in blue, and additions to the original text in red. We have indented the author's response for clarity. Lines numbers given in the author's response refer to lines in the revised manuscript.

General comments:

Why were 7-days backward trajectorys chosen? On page 27, lines 809-815, you already list the problems of these long backward trajectories (uncertainties, distortion due to the predominant time over the mixed layer further back along the trajectory). Furthermore, it is relatively unrealistic that air masses from 7 days before have a direct influence on ODEs in Villum, especially when they seem to occur mainly during calm and stable conditions and are therefore less affected by transport of air masses. Accordingly, it is also quite unlikely that the Chukchi Sea and the Beaufort Sea are relevant source regions for the "average" ODE in Villum. This could be the case in situations with a lot of transport (e.g. cyclone), but this seems to be the exception here.

> We selected a back trajectory length of one week to fully capture the air mass history of ODEs. The motivation for this selection was based on the longest ODE duration observed during the

study period (~6.5 days) and avoiding uncertainty associated with long trajectory calculations (~10 days) (Stohl, 1998). ODE air masses can extent over great distances in the Arctic (Halfacre et al., 2014) and satellite studies have shown that enhancements in reactive halogens (e.g., BrO) is a widespread phenomenon in the Arctic covering several million square kilometers including both sea-ice covered surfaces and continental regions (Bougoudis et al., 2020; Platt and Wagner, 1998; Richter et al., 1998; Schönhardt et al., 2012; Skov et al., 2004). Additionally, reactive halogens can be recycled on aerosol particles which can persist over large spatial scales (Peterson et al., 2017), therefore a longer trajectory time is warranted. Emission of reactive halogen species is ultimately a surface related process (snow on land, snow on sea ice, frost flowers, sea salt emission, refreezing leads, amongst others) and these halogen recycling can be sustained on aerosol particles aloft, selecting a longer trajectory time allows us to fully capture all of these processes. ODEs longer than one week have been observed at the Arctic land-based station, Alert, which observed an ODE of 9 days (Strong et al., 2002). Depletion of ozone has been shown to persist for several weeks in the central Arctic Ocean (Bottenheim et al., 2009). Other studies have used shorter (Bognar et al., 2020; Frieß et al., 2023) or longer (Begoin et al., 2010; Bottenheim and Chan, 2006; Simpson et al., 2018) trajectory lengths than one week, therefore a trajectory length of one week is a compromise between short and long trajectory lengths. We have added text to motivate our select of back trajectory length.

Lines 176-184: The trajectory length was chosen to avoid the uncertainty associated with extremely long trajectory calculations, while capturing the entire geographic extent of ODE air masses. This trajectory length of one week roughly corresponds to the longest observed ODE at Villum during the study period (~6.5 days, Sect. 3.1) and is shorter than the longest observed ODE at a land-based station (9 days at Alert by Strong et al. (2002)). Previous studies have shown that ODE air masses can extent over great distances in the Arctic (Halfacre et al., 2014; Peterson et al., 2017), therefore we selected a trajectory length of one week to fully investigate the air mass history of ODEs. Other studies have used shorter (Bognar et al., 2020; Frieß et al., 2023) or longer (Bottenheim and Chan, 2006; Begoin et al., 2010; Simpson et al., 2018) trajectory lengths than one week.

I would suggest to shorten the "Summary and Outlook" section, especially the last four paragraphs. In general, all of the topics mentioned in these four paragraphs are relevant and related to tropospheric ozone in the Arctic, but in some cases they are not directly related to what you did in your study (e.g. radiative forcing, AMDEs, cloud cover, etc.) so they come a bit out of the blue and lack context.

We purposefully called the last section "Summary and Outlook" over "Conclusions" so we could provide a detailed summary of our results and conclusions from this study as well as give an outlook on the effects of a changing climate on the occurrence of ODEs. Given there are significant changes in ODE frequency and ozone mixing ratios observed around the Arctic (Law et al., 2023; Tarasick and Bottenheim, 2002), a discussion about the possible environmental conditions that could affect these changes in future is highly relevant to the Arctic atmospheric chemistry community and ACP readers. The threshold ranges revealed by the ML model is highly pertinent to such a discussion as we identify thresholds for the influence of ambient meteorological parameters on the likelihood of an observation to be predicted as an ODE. This is especially pertinent given the rapid pace of change in meteorological conditions in the Arctic (Rantanen et al., 2022). Although these threshold ranges are site-specific to Villum, it is our hope that future studies will incorporate such a methodology to identify threshold range (or lack thereof) at other High Arctic sites with long-term in situ meteorological and ozone mixing ratio data.

While we aspired to investigate all aspects related to springtime ODEs, certain topics are inevitably outside the scope of the study such as radiative forcing although they remain important and relevant topics related to Arctic tropospheric ozone (as the reviewer noted). A discussion of these topics is pertinent for the larger Arctic atmospheric chemistry community, who would already be familiar with these concepts.

After review of the "Summary and Outlook" section, we have decided each paragraph adds value to the overall discussion and would foster future studies of Pan-Arctic ODEs and how climate change would affect them. Therefore, we opted not to remove them.

Specific comments:

Page 4 Line 132: It would be more coherent to use consistent units for the uncertainties, either % or ppbv.

We use common notation of uncertainty following EN norms. The uncertainty close to detection limit has to be indicated in absolute values as relative uncertainty (%) goes to infinity close to zero. At mixing ratios much greater than the detection limit and within the calibration range of the instrument the relative uncertainty is used. In this area the absolute uncertainty is a function of the mixing ratio. Thus, we keep the notation of uncertainty and which we have applied in earlier papers in e.g. Skov et al. (2020).

Page 6 Chapter 2.6: Include parts of the Supplement in here: missing data imputation, machine learning, model, ML explain ability approach

We have now moved the entire description of the ML methods from the SI to the main text.

Page 9 Lines 265-268: This sentence is very long and hard to read. I would suggest to split it up in several sentences.

We have split this sentence into two.

Page 10 Line 325 and following: Maybe include a 'snow on land' to every 'snow' in the text, to make clear that no snow on sea ice is analyzed.

We have made this change throughout the manuscript where appropriate.

Page 11 Figure 4: I suggest to only have 2 images per row, to make the individual plots bigger. Even when zooming in, the numbers on the bars are very hard to read.

We have rearranged the subpanels of Fig. 4 so that there are only two subpanels per row.

Page 13 Line 366: Is it really a ODE source region and not just a origin of the air masses? (see above General comments)

We have changed the text on these specific lines and throughout the manuscript to indicate we are referring to ODE air masses to be more precise.

Lines 475-477: During March, the main source regions for ODE air masses appear to be the Chukchi Sea while for Non-ODE air masses the main source region is Greenland with a minor contribution from the central Arctic Ocean (Fig. 6a and d).

Page 13, line 367: Perhaps it should be emphasized that Greenland plays a more important role for the Non-ODE source region (due to the higher trajectory frequencies) compared to the Arctic Ocean.

We have changed the text to indicate Greenland is the main source region and the central Arctic Ocean makes a minor contribution.

Lines 475-477: During March, the main source regions for ODE air masses appear to be the Chukchi Sea while for Non-ODE air masses the main source region is Greenland with a minor contribution from the central Arctic Ocean (Fig. 6a and d).

Page 15, Line 415: Maybe list different surface types here and mention already that land without snow does not play a role. This came as a bit of a surprise further down in the text.

We have added the surface types "(sea, sea ice, or snow on land)" to lines 528-529.

Page 16 Lines 428/429: Shouldn't it be '… start to descend earlier ...'?

This sentence refers to a comparison of Non-ODE air masses during May to those during March/April. During May, Non-ODE air masses begin their descent later along the trajectory (further back in time) compared to March/April. May Non-ODE air masses begin their descent on average at approx. 50 hours back along the trajectory while those during March/April begin their descent closer to the start of the trajectory. In other words, May Non-ODE air masses spend more time within the mixed layer compared to March/April Non-ODE air masses. To alleviate any confusion, we have added "compared to March and April" on line 541.

Page 17, Line 455: What is meant by 'model performance' here? I only see an increase in the 'Recall' variable from March to May.

We were referring to the Recall metric here, which gives the most informative measure of model performance although all three metrics are complementary. We admit the other two metrics do not increase from March to May and this was not explained in the text properly.

We have added the following lines to make this clearer in the text:

Lines 566-582: The evaluation metrics of the ML for all spring months combined and individual months are displayed in Table 1. We use three common metrics for evaluating a binary classification ML model: accuracy, recall, and AUC ROC (Area Under Curve Receiver Operating Characteristics). Briefly, accuracy is the fraction of correctly predicted observations regardless of label (ODE vs Non-ODE), recall is the fraction of ODEs correctly predicted and AUC ROC evaluates how well a model can discriminate between positive and negative labels across all decision thresholds for binary classification. In general, the ML model can accurately reproduce ODEs over all spring months combined as evidenced by how all three metrics are close to unity (their maximum value). However, when evaluating the results on an individual monthly basis, there is an increase in the recall metric and decrease in the accuracy and AUC ROC (see Sect. 2.6 for a detailed description of the evaluation metrics) from March to May (Table 1), which is likely connected to the increasing frequency of ODEs from March to May. With increased ODE occurrence, the recall metrics would increase as positive labels (ODEs) are more likely to be identified when they occur more often and the accuracy and AUC ROC metrics would decrease with the increased occurrence of positive labels due to a concurrent increase in number of incorrectly labeled ODEs. The ML model is also free from over-fitting given the close agreement between the train and test sets. Overall, this ML model is sufficiently accurate, robust, and suitable for the investigation of ODEs.

Caption of Table 1: The accuracy gives an overview of the model performance for both labels (ODEs vs Non-ODEs), recall gives the model performance only for positive labels (ODEs), and AUC ROC evaluates the model performance over different decision thresholds, together, these three metrics give a comprehensive view of the model's performance. The three metrics range from 0 (worst) to 1 (best).

Page 17, Line 458: The difference in the train and test data set and how the model is trained should be explained more detailed in Chapter 2.6.

We have added the following sentence in Sect. 2.6 about the differences between the train and test set.

Line 259-262: The purpose of the training set is for the model to learn how to model the data and the test set is used to evaluate the model's performance on unseen data.

We describe how the ML model learns (i.e., is trained on the data) in the previous paragraph.

Page 17 Line 466: Are the mean SHAP values meant by 'The mean ...'? Should be specified.

Yes, this is our intention here. We have added "SHAP values" to line 606.

Page 18 Line 485: The relationship between SHAP and ambient values and the information its results provide for this study should be explained more detailed (maybe with an example).

We have added the following text in Sect. 3.4 describing the SHAP methodology in plain terms.

Lines 594-605: The SHAP approach is designed to estimate the importance of each input variable to the model output based on coalitional game theory (Molnar, 2022) (see Sect. 2.6 for a more detailed description). SHAP values represent the marginal contribution of each input variable to the model output, or in other words: how important each variable is to the model for making a prediction. SHAP values can be positive or negative, with positive values indicating a variable is more likely to contribute to an observation being predicted as an ODE while negative values mean a variable is more likely to contribute to an observation being labeled as a Non-ODE. The SHAP methodology can produce both local and global explanations. The global importance gives an overview of the most important variables to the model output. The local importance of each observation can give information about the relationship between the SHAP and input values (positive or negative relationship, linear or non-linear), or in other words how does the model output vary over the range of input values.

Page 18 Line 497: Does 'negative effect on model prediction of ODEs' mean the model predicts ODEs wrong when RH is below average?

It does not, "negative effect on model prediction of ODEs" for below average RH values indicates that the model is more likely to predict a Non-ODE rather than an ODE. We have added text to indicate this.

Lines 638-639: (i.e., the model is more likely to predict a Non-ODE)

Page 19 Line 505: Maybe 'after this bin' should be replaced with 'towards lower temperatures'

We have made this change.

Page 20 Figure 10: I suggest to include a legend as it was done in Figure 4. An explanation of what the lines represent in the images should be included in the figure description.

We have changed the caption of Fig. 10 to indicate what the lines and bars represent and how the legend is the same for both.

Caption of Fig. 10: The relationships between SHAP and ambient values for (**a**) RH, (**b**) wind speed, (**c**), temperature, (**d**) radiation, (**e**) pressure, (**f**) wind direction, time air masses spent over (**g**) sea ice, (**h**) snow on land, and (**i**) time above the mixed layer for each month. Fifteen equally spaced bins were calculated for each variable, and the median of the SHAP values was computed for each bin, as represented by the colored lines. The value listed on the x-axis is the midpoint of each bin. The colored bars represent a histogram of the ambient values for each month. The relative frequency of each histogram bin for each variable is displayed on the right axis. The legend is the same for the colored lines and bars.

Page 22 Lines 598-600: Have there been any investigations into halogen release during ODEs in Villum (generally or specifically for this study)?

Measurements of halogen were not available for inclusion in this study so unfortunately there have not been.

Page 22 Lines 606-609: I suggest to make 2 sentences out of this long one.

We have split this sentence into two on lines 755-757.

Page 22 Line 616/617: Might it be possible to use the ERA 5 solar radiation to investigate solar radiation along the trajectory path?

The high resolution (0.25°) ERA5 data of surface solar radiation downwards was only extracted for the ERA5 grid cell containing the location of Villum Research Station and not for the entire Arctic region. This is due to the very large size of highly resolved gridded datasets for the entire Arctic over several decades. This reanalysis product is only representative of the surface and is not vertically resolved (similar to the solar radiation product from HYSPLIT and NCEP/NCAR (Kalnay et al., 1996)). ERA5 on pressure levels does not include solar radiation (https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-pressure-levels?tab=overview). Therefore, we are unable to perform this analysis for this manuscript.

Page 23 Line 662: Maybe include a rough location of the buoys, so one can assume where northerly/easterly/westerly is located.

We have added the text "from the Beaufort Sea" to indicate the approx. location of the buoys on Line 815.

Page 24 Line 673: Which relationship is meant here?

We have added text to indicate the relationship we are referring to is for wind speeds and ozone mixing ratios and normalized ODE hours.

Lines 825-829: Our statistical analysis revealed no relationship between wind speeds and ozone mixing ratios/normalized ODE hours during March, a tendency for high normalized ODE hours with higher wind speeds during April (although little effect on ozone mixing ratios), and two modes during May (one at low and one at high wind speeds) (Fig. 4c).

Page 25 Line 734: What is meant by 'higher values' here?

We have changed the text to indicate that ODEs experience higher values of time over sea ice compared to Non-ODEs.

Lines 898-899: The amount of time spent over sea ice increases from early to late spring (Fig. S4f) and ODE air masses experience higher values of time over sea ice during each spring month compared to Non-ODEs (Fig. 5f).

Page 25, Lines 748/749: Are ODEs meant here and not air masses? If the air masses have low ozone levels, these can only be observed in Villum.

We are referring to the ODE air mass source regions of in these lines. We have changed the text to indicate this.

Lines 911-915: During April, ODE air mass source regions are located over the Beaufort and Chukchi Seas but also over the central Arctic Ocean, which represents a mix of FYI and multi-year sea ice (MYI). During May, ODE air mass source regions are in closer proximity to Villum, mainly arriving from the central Arctic Ocean, which contains the highest concentration of MYI.

We have changed the text throughout the manuscript to indicate we are referring to ODE air masses or ODE air mass source regions where appropriate.

Page 26, Lines 772-777: Split this into two sentences.

We have made this change.

Page 26, Lines 774/775 & 782: The acidity as an additional factor for ODEs comes a bit out of the blue here. I would suggest to include some sentences about acidity and its impact on halogens/ODEs in the Introduction or exclude the acidity part from the text.

We have added text in the introduction highlight the role of acidic surfaces in halogen propagation.

Lines 59-60: These reactions require the presence of a frozen, heterogenous surface aided by high acidity (Sander et al., 2006; Simpson et al., 2007, 2015).

Lines 93-95: This is likely connected to the need for an acidic, frozen heterogeneous surface (sea ice, snowpack, blowing snow, and aerosols) required for halogen propagation (Burd et al., 2017; Jeong et al., 2022), although other studies have not found such evidence (Halfacre et al., 2014; Jacobi et al., 2010).

We have added this reference in other locations throughout the manuscript where appropriate.

Page 27 Line 812: Have you tried what happens if you take shorter (e.g. 3 days) back-trajectories?

For this study, we have not extensively tested the sensitivity of the trajectory length, accuracy of the mixed layer height output from HYSPLIT, nor the starting altitude at the receptor location. Using a shorter back trajectory length would result in the source regions being closer to Villum due to the limited geographical extent of shorter trajectories. This would likely affect the results for March which had the furthest ODE air mass origins compared to April and May. The distribution of the air mass history variables would also change which can be discerned from Fig. 8, which shows the different surface types encountered by air masses as a function of time. This would likely result in the total percentage of time above the mixed layer being reduced. As mentioned in our reply in the general comments, we selected one week to fully capture the air mass history of ODEs. The accuracy of the mixed layer height is likely influenced by the presences of temperature inversions in the Arctic which are difficult to capture in meteorological reanalysis data (Gryning et al., 2023; Xi et al., 2024). The starting altitude, in combination with the accuracy of the mixed layer height, would affect the proportion of time air masses spent within or above the mixed layer, although this would have the largest influence closer to the measurement location. Starting trajectories at a shorter altitude could result in trajectories intercepting the surface would which affect their accuracy (Stohl, 1998) and starting them too high would result in trajectories not being representative of the measurement site. A previous study at Villum tested the effect of the starting altitude but on a short term campaign basis and found similar results for 20 vs 50 m although trajectories starting at 20 m often intercepted the surface (Pernov et al., 2021) thus a higher starting of 50 m was selected for that study and therefore this one. These topics are worth investigating, therefore, we have added text highlighting avenues for future studies.

Lines 985-988: Proper representation of air mass history therefore is an important aspect of evaluating ODEs and other atmospheric phenomena and future studies should evaluate this in more detail including the effects of varying trajectory lengths, the accuracy of the mixed layer height from HYSPLIT, and starting altitude at the receptor location.

Page 28 Line 835: ‘high time’ → long time?

We have made this change.

Page 28 Lines 840-842: see Page 13 Line 366

We have added text to make these two sections congruent.

Lines 1010-1011: During March, sea ice (likely FYI) in the Chukchi Sea is the main source region for ODE air masses.

Supplement:

S1 Machine learning modeling methodology

Second paragraph:

‘We imputed missing data using the median value for the hour of the day for that day

of the year.’ This sentence is very hard to follow, maybe describe it with an example.

We have added an example of this procedure.

Lines 234-235: For instance, if a value is missing for hour 12 on the 90th day of the year then this value was imputed using the median of all values from hour 12 on the 90th day of the year from the entire dataset.

Fourth paragraph:

‘Tuning was performed for 1000 trials and the best parameters were selected.’ Is parameters or hyperparameters meant here? If parameters, maybe shortly explain the difference.

We have changed “parameters” to “hyperparameters” on line 266.

Figure S3 description Lines 3/4: Which blue bars?

We have changed the text to show black bars represent not SS trends.

Caption of Fig. S3: The black bars represent trends that are not significant on the 95th % confidence level.

Technical corrections:

Page 4 Line 120: please define i.d.

We have defined “i.d.” as “inner diameter” on Line 124.

Page 6 Line 203: please define RH (first mention)

We have defined “RH” as “relative humidity” on Line 150.

Page 7 Line 234: please define CA (first mention)

We have written out the country name "Canada" throughout the text.

Page 8 Line 246: please define CL (first mention)

We have defined "CL" as "confidence level" on Line 136.

Page 8 Line 248: CL instead of Cl

"CI" on this line represents "confidence intervals" which we have defined on line 351 and in the caption of figure 3 on line 366.

Page 8 Line 258: (d) bold

We have made this change

Page 9 Line 291: ODE instead of ODEs

We have made this change

Page 23 Line 660: please define AK (first mention)

We have written out the state name "Alaska" throughout the text.

Supplement:

Figure S19 → Figure S9

We have made this change

Begoin, M., Richter, A., Weber, M., Kaleschke, L., Tian-Kunze, X., Stohl, A., Theys, N., and Burrows, J. P.: Satellite observations of long range transport of a large BrO plume in the Arctic, Atmospheric Chemistry and Physics, 10, 6515–6526, https://doi.org/10.5194/acp-10-6515-2010, 2010.

Bognar, K., Zhao, X., Strong, K., Chang, R. Y.-W., Frieß, U., Hayes, P. L., McClure-Begley, A., Morris, S., Tremblay, S., and Vicente-Luis, A.: Measurements of Tropospheric Bromine Monoxide Over Four Halogen Activation Seasons in the Canadian High Arctic, J. Geophys. Res. Atmos., 125, e2020JD033015, https://doi.org/10.1029/2020jd033015, 2020.

Bottenheim, J. W. and Chan, E.: A trajectory study into the origin of spring time Arctic boundary layer ozone depletion, J. Geophys. Res. Atmos., 111, https://doi.org/10.1029/2006JD007055, 2006.

Bottenheim, J. W., Netcheva, S., Morin, S., and Nghiem, S. V.: Ozone in the boundary layer air over the Arctic Ocean: measurements during the TARA transpolar drift 2006–2008, Atmos. Chem. Phys., 9, 4545–4557, https://doi.org/10.5194/acp-9-4545-2009, 2009.

Bougoudis, I., Blechschmidt, A.-M., Richter, A., Seo, S., Burrows, J. P., Theys, N., and Rinke, A.: Long-term time series of Arctic tropospheric BrO derived from UV–VIS satellite remote sensing and its relation to first-year sea ice, Atmos. Chem. Phys., 20, 11869–11892, https://doi.org/10.5194/acp-20-11869-2020, 2020.

Burd, J. A., Peterson, P. K., Nghiem, S. V., Perovich, D. K., and Simpson, W. R.: Snowmelt onset hinders bromine monoxide heterogeneous recycling in the Arctic, J. Geophys. Res. Atmos., 122, 8297–8309, https://doi.org/10.1002/2017jd026906, 2017.

Frieß, U., Kreher, K., Querel, R., Schmithüsen, H., Smale, D., Weller, R., and Platt, U.: Source mechanisms and transport patterns of tropospheric bromine monoxide: findings from long-term multi-axis differential optical absorption spectroscopy measurements at two Antarctic stations, Atmospheric Chemistry and Physics, 23, 3207–3232, https://doi.org/10.5194/acp-23-3207-2023, 2023.

Gryning, S.-E., Batchvarova, E., Floors, R., Münkel, C., Sørensen, L. L., and Skov, H.: Observed aerosol-layer depth at Station Nord in the high Arctic, International Journal of Climatology, 43, 3247–3263, https://doi.org/10.1002/joc.8027, 2023.

Halfacre, J. W., Knepp, T. N., Shepson, P. B., Thompson, C. R., Pratt, K. A., Li, B., Peterson, P. K., Walsh, S. J., Simpson, W. R., Matrai, P. A., Bottenheim, J. W., Netcheva, S., Perovich, D. K., and Richter, A.: Temporal and spatial characteristics of ozone depletion events from measurements in the Arctic, Atmos. Chem. Phys., 14, 4875–4894, https://doi.org/10.5194/acp-14-4875-2014, 2014.

Jacobi, H.-W., Morin, S., and Bottenheim, J. W.: Observation of widespread depletion of ozone in the springtime boundary layer of the central Arctic linked to mesoscale synoptic conditions, J. Geophys. Res. Atmos., 115, https://doi.org/10.1029/2010JD013940, 2010.

Jeong, D., McNamara, S. M., Barget, A. J., Raso, A. R. W., Upchurch, L. M., Thanekar, S., Quinn, P. K., Simpson, W. R., Fuentes, J. D., Shepson, P. B., and Pratt, K. A.: Multiphase Reactive Bromine Chemistry during Late Spring in the Arctic: Measurements of Gases, Particles, and Snow, ACS Earth Space Chem., 6, 2877–2887, https://doi.org/10.1021/acsearthspacechem.2c00189, 2022.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K. C., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-Year Reanalysis Project, Bulletin of the American Meteorological Society, 77, 437–472, https://doi.org/10.1175/1520-0477(1996)077<0437:TNYRP>2.0.CO;2, 1996.

Law, K. S., Hjorth, J. L., Pernov, J. B., Whaley, C. H., Skov, H., Collaud Coen, M., Langner, J., Arnold, S. R., Tarasick, D., Christensen, J., Deushi, M., Effertz, P., Faluvegi, G., Gauss, M., Im, U., Oshima, N., Petropavlovskikh, I., Plummer, D., Tsigaridis, K., Tsyro, S., Solberg, S., and Turnock, S.: Arctic Tropospheric Ozone Trends, Geophys. Res. Lett., 50, e2023GL103096, https://doi.org/10.1029/2023GL103096, 2023.

Lundberg, S. M. and Lee, S.-I.: A Unified Approach to Interpreting Model Predictions, in: Adv Neural Inf Process Syst, https://doi.org/10.48550/arXiv.1705.07874, 2017.

Molnar, C.: Interpretable Machine Learning: A Guide for Making Black Box Models Explainable, 2nd ed., 2022.

Pernov, J. B., Jensen, B., Massling, A., Thomas, D. C., and Skov, H.: Dynamics of gaseous oxidized mercury at Villum Research Station during the High Arctic summer, Atmos. Chem. Phys., 21, 13287–13309, https://doi.org/10.5194/acp-21-13287-2021, 2021.

Peterson, P. K., Pöhler, D., Sihler, H., Zielcke, J., General, S., Frieß, U., Platt, U., Simpson, W. R., Nghiem, S. V., Shepson, P. B., Stirm, B. H., Dhaniyala, S., Wagner, T., Caulton, D. R., Fuentes, J. D., and Pratt, K. A.: Observations of bromine monoxide transport in the Arctic sustained on aerosol particles, Atmos. Chem. Phys., 17, 7567–7579, https://doi.org/10.5194/acp-17-7567-2017, 2017.

Platt, U. and Wagner, T.: Satellite mapping of enhanced BrO concentrations in the troposphere, Nature, 395, 486–490, https://doi.org/10.1038/26723, 1998.

Rantanen, M., Karpechko, A. Y., Lipponen, A., Nordling, K., Hyvärinen, O., Ruosteenoja, K., Vihma, T., and Laaksonen, A.: The Arctic has warmed nearly four times faster than the globe since 1979, Commun. Earth Environ., 3, 1–10, https://doi.org/10.1038/s43247-022-00498-3, 2022.

Richter, A., Wittrock, F., Eisinger, M., and Burrows, J. P.: GOME observations of tropospheric BrO in northern hemispheric spring and summer 1997, Geophysical Research Letters, 25, 2683–2686, https://doi.org/10.1029/98GL52016, 1998.

Sander, R., Burrows, J., and Kaleschke, L.: Carbonate precipitation in brine – a potential trigger for tropospheric ozone depletion events, Atmospheric Chemistry and Physics, 6, 4653–4658, https://doi.org/10.5194/acp-6-4653-2006, 2006.

Schönhardt, A., Begoin, M., Richter, A., Wittrock, F., Kaleschke, L., Gómez Martín, J. C., and Burrows, J. P.: Simultaneous satellite observations of IO and BrO over Antarctica, Atmospheric Chemistry and Physics, 12, 6565–6580, https://doi.org/10.5194/acp-12-6565-2012, 2012.

Simpson, W. R., von Glasow, R., Riedel, K., Anderson, P., Ariya, P., Bottenheim, J., Burrows, J., Carpenter, L. J., Frieß, U., Goodsite, M. E., Heard, D., Hutterli, M., Jacobi, H. W., Kaleschke, L., Neff, B., Plane, J., Platt, U., Richter, A., Roscoe, H., Sander, R., Shepson, P., Sodeau, J., Steffen, A., Wagner, T., and Wolff, E.: Halogens and their role in polar boundary-layer ozone depletion, Atmos. Chem. Phys., 7, 4375–4418, https://doi.org/10.5194/acp-7-4375-2007, 2007.

Simpson, W. R., Brown, S. S., Saiz-Lopez, A., Thornton, J. A., and Glasow, R.: Tropospheric halogen chemistry: sources, cycling, and impacts, Chem Rev, 115, 4035–62, https://doi.org/10.1021/cr5006638, 2015.

Simpson, W. R., Frieß, U., Thomas, J. L., Lampel, J., and Platt, U.: Polar Nighttime Chemistry Produces Intense Reactive Bromine Events, Geophysical Research Letters, 45, 9987–9994, https://doi.org/10.1029/2018GL079444, 2018.

Skov, H., Christensen, J. H., Goodsite, M. E., Heidam, N. Z., Jensen, B., Wahlin, P., and Geernaert, G.: Fate of elemental mercury in the arctic during atmospheric mercury depletion episodes and the load of atmospheric mercury to the arctic, Environ. Sci. Technol., 38, 2373–2382, https://doi.org/10.1021/es030080h, 2004.

Skov, H., Hjorth, J., Nordstrøm, C., Jensen, B., Christoffersen, C., Bech Poulsen, M., Baldtzer Liisberg, J., Beddows, D., Dall'Osto, M., and Christensen, J. H.: Variability in gaseous elemental mercury at Villum Research Station, Station Nord, in North Greenland from 1999 to 2017, Atmos. Chem. Phys., 20, 13253–13265, https://doi.org/10.5194/acp-20-13253-2020, 2020.

Stohl, A.: Computation, accuracy and applications of trajectories - A review and bibliography, Atmos. Environ., 32, 947–966, https://doi.org/10.1016/s1352-2310(97)00457-3, 1998.

Strong, C., Fuentes, J. D., Davis, R. E., and Bottenheim, J. W.: Thermodynamic attributes of Arctic boundary layer ozone depletion, Atmospheric Environment, 36, 2641–2652, https://doi.org/10.1016/S1352-2310(02)00114-0, 2002.

Tarasick, D. W. and Bottenheim, J. W.: Surface ozone depletion episodes in the Arctic and Antarctic from historical ozonesonde records, Atmos. Chem. Phys., 2, 197–205, https://doi.org/10.5194/acp-2-197-2002, 2002.

Xi, X., Yang, Q., Liu, C., Shupe, M. D., Han, B., Peng, S., Zhou, S., and Chen, D.: Evaluation of the Planetary Boundary Layer Height From ERA5 Reanalysis With MOSAiC Observations Over the Arctic Ocean, Journal of Geophysical Research: Atmospheres, 129, e2024JD040779, https://doi.org/10.1029/2024JD040779, 2024.