We would like to thank Andreas Stohl for taking the time to read and provide comments on our paper. We believe the changes made in response to their comments have strengthened the manuscript.

See below for our response (in red) to their comments (in black).

Response to Andreas Stohl

This paper describes an interesting analysis of the SO2-to-sulfate oxidation in models relative to those inferred from observations after the Holuhraun eruption. The paper shows that gas-phase oxidation rates in the models are all slower than the observed rates, which is an important result. The main result of the study is presented in Figure 6. However, I have a few concerns about this figure, as detailed below in my major comments below. Most importantly, I am not convinced that a robust separation between gas- and aqueous-phase oxidation is possible based on the available observation data, mostly for two reasons: 1) the mono- and bi-exponential fits are very similar, and it is not so clear that the bi-exponential fit is SIGNIFICANTLY better than the mono-exponential fit; 2) the attribution of the two e-folding times obtained by the fit to gas- and aqueous-phase oxidation seems quite a stretch. I think this interpretation needs independent support before the paper can be published. A few other points also need to be addressed, as outlined below.

Major:

The trajectory analysis is somewhat problematic. First of all, how are the 27 members of the trajectory ensembles (line 174) different from each other? This is not explained in the text. Second, all EMEP stations are located in the atmospheric boundary layer, where air mass trajectories are not well representing the properties of the flow, due to turbulence. This will likely affect the quality of the attribution of events to Holuhraun (or not). Third, the definition of "vicinity" of the Holuhraun eruption is highly subjective. Depending, e.g., on the transport time and distance, trajectory errors will likely be very much case-dependent, and a single "vicinity area" might not be appropriate for all cases (e.g., stations closer to Holuhraun will have a greater chance of hitting the defined vicinity area.

Thank you for raising your concerns regarding the trajectory analysis. We hope the following changes reassure you that our methodology is sound.

Firstly, a more detailed description on the trajectory analysis has been provided. This includes a clarification on how the 27 ensemble members differ due to small perturbations in the input meteorology data (i.e. offset by a fixed grid factor, a maximum of 1.0° of latitude/longitude in the horizontal and 0.01 sigma units in the vertical, and so all possible offsets result in the 27 members).

Secondly, we acknowledge there is often greater uncertainty in using single-particle trajectories as opposed to dispersion modelling. Nevertheless, single-particle trajectory frameworks like ours have been widely used in many previous studies to characterise long-range transport (e.g. Nieminen et al., 2015; Räty at al., 2023; Väänänen et al., 2013). Plus, our goal is to inter-compare models consistently rather than through a rigorous dispersion exercise, and so believe our trajectory framework is sufficient to achieve this. We have added additional comments to the manuscript discussing this.

In addition, the reviewer is right in that these uncertainties may affect the quality of attribution. However, the Holuhraun emissions were so substantial within the region that it would be extremely unlikely that most identified events were misattributed to Holuhraun. Moreover, sulphurous surface concentrations are rather low over Europe in recent times and spikes in SO_2 are a rare occurrence.

Finally, we agree that the "vicinity" of the Holuhraun eruption is subjective. To address this, we have taken a more quantitative approach. We now define multiple spherical bounding areas with radii increasing with a station's increasing distance from Holuhraun. These radii values are based on the positional error of a trajectory being approximately 10–30% of the total distance travelled (Stohl, 1998). A special case is made for the Irafoss station due to the typical spatial resolution of the trajectories being greater than the error estimated from 10-30% of the distance. Consequently, we define the bounding radius here using local wind speeds following the methodology set out in Hughes et al. 2012.

The comparison between models and IASI data is not fully convincing. It seems model output is shown irrespective of whether IASI retrievals are available for a location or not. IASI retrievals can easily miss volcanic SO2, e.g., underneath clouds. Thus, models should only be sampled in pixels where IASI SO2 retrievals are actually made. The authors write that models often have larger plume areas than the IASI retrievals, which can be attributed to clouds affecting IASI. Still, it appears that many models actually have often smaller plume areas than IASI. This would even be worse when cloud screening is applied.

We have listened to your suggestion and agree the better approach is to only sample models where successful IASI SO_2 retrievals have been made. Subsequently, our comparison with IASI retrievals is now only made on grid cells within the observed plume extent. On redoing the analysis, the modelled SO_2 plume heights now overestimate observations whilst no substantial changes in the SO_2 mass burden comparison is found.

On the comment "many models actually have often smaller plume areas than IASI", we acknowledge that this is true across certain periods, yet overall we observe that the modelled plume extents are larger than those observed, particularly in October. We have improved Fig. 2 and the animation by explicitly distinguishing between regions inside and outside the observed plume extent in the model simulations (coloured vs hatched areas) to improve the evidence for this statement.

Figure 5: Since the conversion rate of SO2 to sulfate is shown to be uncertain, I am wondering why Figure 5 does not also show a comparison for total sulphur (SO2 + sulfate). This should provide the most robust comparison between the models and the observations.

A total sulphur comparison has been added to the figure and the discussion extended.

Figure 6: This is the core result of the paper and quite interesting. However, I am not at all convinced that the bi-exponential fit is any better than the mono-exponential fit. That the bi-exponential fit is better (line 417) is a trivial result. But is it really SIGNIFICANTLY better? The two e-folding times obtained are interpreted as gas-phase and aqueous-phase e-folding times. But I am concerned that the fit is not stable enough to reliably distinguish between the two. Furthermore, how do you know which e-folding time is which? The data per se do not give any information on the two processes, but the authors immediately jump to the conclusion that these are gas- and aqueous phase e-folding times. What is the evidence for this?

We thank the reviewer for raising their concerns with the biexponential fitting. A similar comment was made from another reviewer. In hindsight, we acknowledge that attempting to separate the in-

plume SO_2 into its gaseous and aqueous pathways was too much of stretch for our dataset. We now only fit to an exponential with a single decay constant. The derived gaseous-phase and aqueous-phase oxidation rate constants are now replaced with a single value generalising them. However, we believe that our efforts to explore the complexity of volcanic SO_2 oxidation is still worth mentioning and a suggested direction for future works has been included.

Figure 6: The aqueous-phase oxidation occurs only in clouds, so is a single e-folding time even appropriate to characterize this oxidation? This must be highly variable, depending on the time the SO2 spends in a cloud.

See previous comment.

Figure 6: All events are exclusively attributed to Holuhraun. However, there are likely always (perhaps minor) contributions from other sources. How might these affect the results, especially far away from the volcano, where SO2/sulfate ratios are low and even relatively small anthropogenic SO2 emissions could affect the ratio substantially.

Our fitting considers both the uncertainty in the ratio and plume age, with the latter being by far the dominant source. Generally, the plume age error increases with increasing plume age. Consequently, ratios sampled in a mature plume have less influence on the fitting parameters than those sampled in a young plume. As you say, the ratios "far away from the volcano" are low and so possibly are affected by small anthropogenic emissions, whilst the ratios sampled closer to the eruption are larger and less likely to be significantly affected. However, these "far away" ratios have a limited influence on the fitting due to their associated large plume age errors, whereas the ratios with a low likelihood of anthropogenic impact have small plume age errors. Hence, the ratios that could be potentially affected by small anthropogenic SO_2 emissions substantially have a relatively minor influence on the overall oxidation rate constant and near-vent ratios derived.

In addition, a major goal of the EMEP network is to gather observations at locations where significant sources of local pollution are minimised (Tørseth et al., 2012). As such, anthropogenic SO_2 contributions in the observations used here should be minimal.

Line 435: A modelled event is considered successful if both SO2 and sulfate concentrations are within a factor 5 of the observations. Doesn't this introduce a bias in the analysis? You show that modelled oxidation rates are too slow – in this case one would expect the model to often substantially overestimate observed SO2 concentrations. But large overestimations would be substantially removed from the analysis, which would lead to biased results.

This is a good point and is seemingly a flaw in our analysis. As such, we now no longer only fit model output to the events captured within a factor of 5. All models are fitted to all observed events attributed to Holuhraun preventing this bias from impacting our results. The new analysis without the bias does not change remarkably.

Line 161 and Table 2: Why are ERA-Interim reanalyses used? These are superseded since quite a few years already by ERA5 reanalyses with better resolution, and which should have better quality!

The experiment was initialised prior to the public release of ERA5 (2019) which is why it is not included. As to why it has taken so long to reach the results stage, that is an unfortunate multitude of disruptions including COVID, a change of experiment lead, and submission errors.

Lines 279-280: How do you know that varying IASI SO2 burdens are due to changing IASI retrieval coverage and plumes passing in and out of the region, and not due to variations in emission flux? I don't think there is good enough data to prove that the emission flux was really constant.

We agree that there is not sufficient data to suggest the volcanic SO_2 emission flux was constant during September and October 2014 and that the coverage of IASI retrievals was too scarce to establish definitive conclusions. Rather, it is more likely variation existed (e.g. Thordarson and Hartley, 2015) and so contributes to the mass burden variation shown. Our comment has been amended to reflect this.

Lines 370-371: Why is a poor performance of concentration ratios expected? The two species are not simulated independently, so a plume in one species should always correlate with a plume in the other species.

Agreed, we have removed this comment.

Table 3: Why does OsloCTM3 not have the data required for filling Table 3? This should be basic model output (SO2 concentration fields) that is also needed for all other analyses?

The SO_2 columns loads and mass burdens for OsloCTM3 have now been included and the discussion amended accordingly. Unfortunately, from the diagnostics made available in the submission for this experiment it is not possible to derive an SO_2 plume height for OsloCTM3.

Line 55: word aerosol is duplicated in text

Corrected, thank you.

Hughes, E. J., Sparling, L. C., Carn, S. A., and Krueger, A. J.: Using horizontal transport characteristics to infer an emission height time series of volcanic SO_2, J. Geophys. Res., 117, D18307, doi:10.1029/2012JD017957, 2012.

Nieminen, T., Yli-Juuti, T., Manninen, H. E., Petäjä, T., Kerminen, V.-M., and Kulmala, M.: Technical note: New particle formation event forecasts during PEGASOS–Zeppelin Northern mission 2013 in Hyytiälä, Finland, Atmos. Chem. Phys., 15, 12385–12396, doi:10.5194/acp-15-12385-2015, 2015.

Räty, M., Sogacheva, L., Keskinen, H.-M., Kerminen, V.-M., Nieminen, T., Petäjä, T., Ezhova, E., and Kulmala, M.: Dynamics of aerosol, humidity, and clouds in air masses travelling over Fennoscandian boreal forests, Atmos. Chem. Phys., 23, 3779–3798, doi:10.5194/acp-23-3779-2023, 2023.

Stohl, A.: Computation, accuracy and applications of trajectories—A review and bibliography, Atmos. Environ., 32, 947-966, 1998.

Thordarson, T., and Hartley M.: Atmospheric sulfur loading by the ongoing Nornahraun eruption, North Iceland, Geophys. Res., Abstracts, 17(EGU2015-10708), 2015.

Tørseth, K., Aas, W., Breivik, K., Fjæraa, A. M., Fiebig, M., Hjellbrekke, A. G., Lund Myhre, C., Solberg, S., and Yttri, K. E.: Introduction to the European Monitoring and Evaluation Programme (EMEP) and observed atmospheric composition change during 1972–2009, Atmos. Chem. Phys., 12, 5447–5481, doi:10.5194/acp-12-5447-2012, 2012.

Väänänen, R., Kyrö, E.-M., Nieminen, T., Kivekäs, N., Junninen, H., Virkkula, A., Dal Maso, M., Lihavainen, H., Viisanen, Y., Svenningsson, B., Holst, T., Arneth, A., Aalto, P. P., Kulmala, M., and Kerminen, V.-M.: Analysis of particle size distribution changes between three measurement sites in northern Scandinavia, Atmos. Chem. Phys., 13, 11887–11903, doi:10.5194/acp-13-11887-2013, 2013.