Reply to review comments

We thank the reviewers for the thoughtful comments, time and efforts spent on reviewing the manuscript. We carefully considered the comments and revised the manuscript accordingly. Below please find our point-by-point replies (colored in blue). Besides, a revised manuscript with tracked changes was uploaded.

Reviewer #1

General comments

The manuscript presents an interesting method to estimate the SO_2 emission rates from volcanic eruptions, based on model simulations and SO_2 index from satellite data. While the paper presents a useful technique which are illustrated with an interesting case study of the Nabro eruption, the description of the method is not clear, in particular how satellite data is used and how uncertainties are addressed, also the results are not sufficiently validated, and references and comparisons to some other key publications on the Nabro event are lacking. The manuscript has potential for publication after being revised, with particular focus on the comments below.

We revised the method description to make the proposed inversion approach better understandable. We also extended the discussion and compared our results with those presented in the suggested key publications. For details please see responses to specific comments below.

Specific comments

1. "Ensemble-simulations":

- The authors refer to the term "ensemble simulations" throughout the manuscript, which is not explained until section 3.1. It needs to be clarified in the very beginning of the paper.

Following other review comments, the term "ensemble simulations" has been removed or rephrased throughout the manuscript in order to avoid unnecessary confusion. The changes are found in the revised manuscript: page 1, title, page 2, line 9, page 4, line 12, page 5, line 25, page 6, line 4, line 8, line 26, page 11, line 7, page 12, line 4, line 5, page 17, line 22, which correspond to those on page 9103, title, 9104, line 9, page 9106, line 1, page 9107, line 9, line 11, line 15, line 20, page 9111, line 9, page 9112, lines 1-3, page 9117, line 7 in the original GMDD paper.

- In the introduction it is stated that "The fine temporal and spatial discretization of this domain creates a need for large-scale ensemble simulations". This is unclear. First, what is the argument for using a very finely discretized emission domain, and second, why does this create a need for ensemble?

The volcanic emissions in the Nabro case study are strongly time- and altitude-dependent. We discretized the emission domain as finely as technically feasible in order to reveal local details of the SO_2 emissions at high temporal and spatial resolution. This way, we expect to obtain more reliable simulation results. However, the fine discretization increases demands on computing capabilities. The iterative estimation procedure of the emission rates requires a large number of unit simulations. Besides, a study based on a sufficiently fine discretization can provide us valuable reference information for the further development of an adaptive discretization strategy that may help to reduce the computational burden in future work. To clarify we modified the text on page 9107, lines 6-12 (GMDD paper), which is found on page 5, lines 23-29, page 6, lines 1-5 in the revised manuscript.

- In section 3.1 the authors finally explain that they refer to the set of all unit simulations in the inversion procedure as an "ensemble simulation". However, is this really characterized as ensemble simulations? Ensemble dispersion modelling implies, as explained by Galmarini et al. (2004) (doi:10.1016/j.atmosenv.2004.05.030), variations in the meteorological drivers and/or source parameters, or the approach to dispersion modelling by using different models. I do not see that these simulations fall under either of these categories. They are simply "sensitivity"-simulations, "scenario"-simulations, "unit"-simulations, or sometimes described as "source-receptor" relationships. I do not see how they are true ensemble simulations. Please justify the use of this term, or consider changing to another wording.

We agree and removed or rephrased the term "ensemble simulations" used for describing our inversion approach. Please see reply to comment 1. "Ensemble-simulations".

2. Methodology description (Section 3):

It was sometimes quite hard to follow and understand the methodology, mainly the first section 3.1. In particular I did not fully understand

- the difference or similarity between the terms "ensemble simulations" , "forward simulations" and "model forecasts", which seem to be used interchangeably? Also, in the abstract you point out two types of simulations; "ensemble", and "final transport simulations". A more consistent wording is needed throughout the paper.

Following earlier comments we made corresponding changes regarding the term "ensemble simulations". For consistency, we replaced the term "final transport simulations" on page 9104, line 10, line 21, page 9111, line 13, page 9112, lines 14-15, line 25 (GMDD paper) with "final forward simulations" on page 2, line 10, line 21, page 11, lines 18-19, page 12, line 25, page 13, lines 6-7 in the revised manuscript. The term "model forecasts" refers to the results of the "final forward simulations".

- the justification for using 12 h accumulated model and observations data?

The AIRS instrument requires 12 h to gather observations on a global scale. At mid and low latitudes there are typically two satellite overpasses per day (at 01:30 and 13:30 local time). An accumulation time interval shorter than 12 h may lead to time periods in the CSI analysis during which the satellite observations do not cover the volcanic plume at all. Therefore 12 h is a reasonable minimum time period for this analysis. To clarify we added text on page 13, lines 16-20 in the revised manuscript.

- why the importance weights need to be estimated iteratively? You say this is needed in order to obtain reliable results, please elaborate. Why do the unit simulations have to be re-run for each iteration (alas section 4.1)?

As an initial guess for the first iteration, the simulated air parcels are distributed uniformly in the considered time- and altitude-dependent initialization domain. Correspondingly, the importance weights for all 15840 subdomains are 1/15840. In order to find more realistic importance weights that reflect the relative distribution of emissions in the subdomains, unit simulations then have to be performed to estimate importance weights in an iterative scheme. Changes in the importance weights indicate how many air parcels should be reassigned to each subdomain and considered as new initial conditions for the next iteration.

In our case, after 1-2 iterations we already obtained rather stable importance weights that led to good simulation results. Nevertheless, in order to establish a robust computational procedure, we defined a stopping criterion for the iterative update process. Namely, the iterative procedure will stop when the change of the updated importance weights is sufficiently small. Here, the change of the importance weights is evaluated according to a given mathematical measure that is based on the Frobenius norm. In this study, by choosing a threshold of 1% for the relative difference of the updated important weights, the stopping condition was fulfilled after 3 iterations (see page 16, lines 16-17 in the revised manuscript).

To clarify we added text on page 12, lines 15-22 in the revised manuscript.

- In section 1 it is stated that the algorithm does not require explicit source-receptor relationships. But this is not clarified any further in the methodology section. Please explain.

By saying source-receptor relationships we mean the full source-receptor matrix computed based on the forward model and the discretization of the unknown quantity on a regular grid (cf. Seibert (2000) and related works), which is used for the solution of the arising linear inverse problem. In our approach this source-receptor matrix is not explicitly calculated. More details have been addressed on page 9106, lines 18-27 in the GMDD paper.

- what is "new" compared to the methods by Stohl et al (2011) and Flemming and Innes (2013), and what are the benefits of your method compared to those previously published methods? This should be clearly stated already in the Introduction.

In the work of Stohl et al. (2011) and Flemming and Inness (2013), the forward models FLEXPART and ECMWF's IFS were applied, respectively. In this study, we used our recently developed Lagrangian transport model Massive-Parallel Trajectory Calculations (MPTRAC) to perform forward simulations at large scale. The work of Stohl et al. (2011) extended the inversion algorithm initially presented in Seibert (2000). Compared to that approach our inversion approach requires no a-priori information on the emissions and does not require the calculation of the full source-receptor matrix. Compared to the work of Flemming and Inness (2013), we considered a much finer discretization for the unknown time- and altitude-dependent emission function (250 m in altitude and 1 h in time in our case, about 2–3 km and more than 6 h in the case of Flemming and Inness (2013), and 19 vertical layers stacked up to 12.3 km altitude and 3 h time intervals in the case of Stohl et al. (2011)). We did not find a need to solve the ill-posed inverse problem by means of a Tikhonov or smoothing constraint. Furthermore, the Critical Success Index (CSI) was used here for the first time to evaluate the goodness-of-fit of the forward simulations and to estimate the importance weights of the time- and altitude-dependent SO_2 emission distribution. This way, we were able to provide relative distributions of the emissions in a two-dimensional view (in time and altitude) and its local details at relatively high (or even unprecedented) temporal and spatial resolution.

We modified the text on page 9107, lines 12-15 (GMDD paper) that emphasizes the advantages of our inversion approach. Improved text can be found on page 6, lines 9-21 in the revised manuscript.

3. Adequate referencing:

The authors should refer to and compare their results to the following two publications which also reported on SO_2-inversions and satellite-derived height estimates for the Nabro eruption:

- Theys et al. (2013) Volcanic SO_2 fluxes derived from satellite data: a survey using OMI , GOME-2 , IASI and MODIS, Atmospheric Chemistry and Physics, doi:10.5194/acp-13-5945-2013.
- Clarisse, L. et al. (2014) The 2011 Nabro eruption, a SO_2 plume height analysis using IASI measurements, Atmos. Chem. Phys., 14, 3095-3111, doi:10.5194/acp-14-3095-2014

We added these two references in the introduction section on page 4, lines 21-22 in the revised manuscript. We also discussed and compared the results from these two publications and our work in the revised manuscript. We improved the text in the introduction, in Sect. 2.3 (validation data sets), in Sect. 4.5 (validation of emission time series) and in Sect. 4.6 (final forward simulations), accordingly.

4. Altitude sensitivity:

How is averaging kernels handled in your approach (see explanations and Figure 1 of the above mentioned Theys et al. (2013) paper)? You say in section 2.2 that the AIRS SI is most sensitive to SO_2 layers at about 8 to 13 km altitude which reflects the infra-red kernel. Do you take this altitude-sensitivity into account when you compare model and observation data? I.e. do you count the model values below 8 km in the same way as those above 8 km, or are they weighted by the averaging kernel so that low altitude model values count less since the satellite data is less sensitive at these heights? In other words, the AIRS SI data might not contain information to constrain the emissions below 3-5 km altitude. How do you deal with this?

The SO_2 kernel functions of the AIRS channels used to calculate the SI depend on atmospheric conditions and altitude (e.g., Hoffmann et al. (2016, Fig. 1)). However, variations in the UT/LS region where most of the Nabro emissions occurred are not too large. Hence, we did not consider this dependency in our analysis. However, the consideration of the AIRS kernel functions in the CSI analysis will be an important aspect in future work.

To clarify we added the aforementioned text to the Outlook section on page 26, lines 23-28 in the revised manuscript.

5. Uncertainties:

How do you deal with uncertainties in the satellite SI index, and also uncertainties related to the unit simulations in particular to errors in the meteorological driving data? At longer forecast times it is likely that the errors in the meteorological data are more important.

Uncertainties in the meteorological data are an important source of error. The topic is addressed in a recent study by Hoffmann et al. (2016), wherein four different meteorological products have been tested for the MPTRAC simulations. The study shows that spatial and temporal resolution of the meteorological data is an important factor influencing the quality of the simulations. Here we used the ERA-Interim reanalysis, which provides good accuracy for the simulations. Better results were only obtained with ECMWF operational analysis, which poses higher demands regarding the memory available on the computing system. This work aims to introduce an inversion for SO_2 transport simulations. A more detailed, quantitative study of the errors resulting from the uncertainties of different meteorological data will be considered in future work. To clarify we added the aforementioned text to the outlook on page 26, lines 28-29, page 27, lines 1-4 in the revised manuscript.

We also added more detailed information on the measurement uncertainties of the AIRS SO_2 index on page 9, lines 12-13 in the revised manuscript. The precision of the SI is quite good (0.14 K at 250 K scene temperature) and we concluded that measurement noise has only minor effects on the inversion results presented here.

6. Loss of SO_2:

Do the MPTRAC SO_2 simulations take into account decay of SO_2 by for example reaction with OH? This would be important particularly on the >2 days time scales.

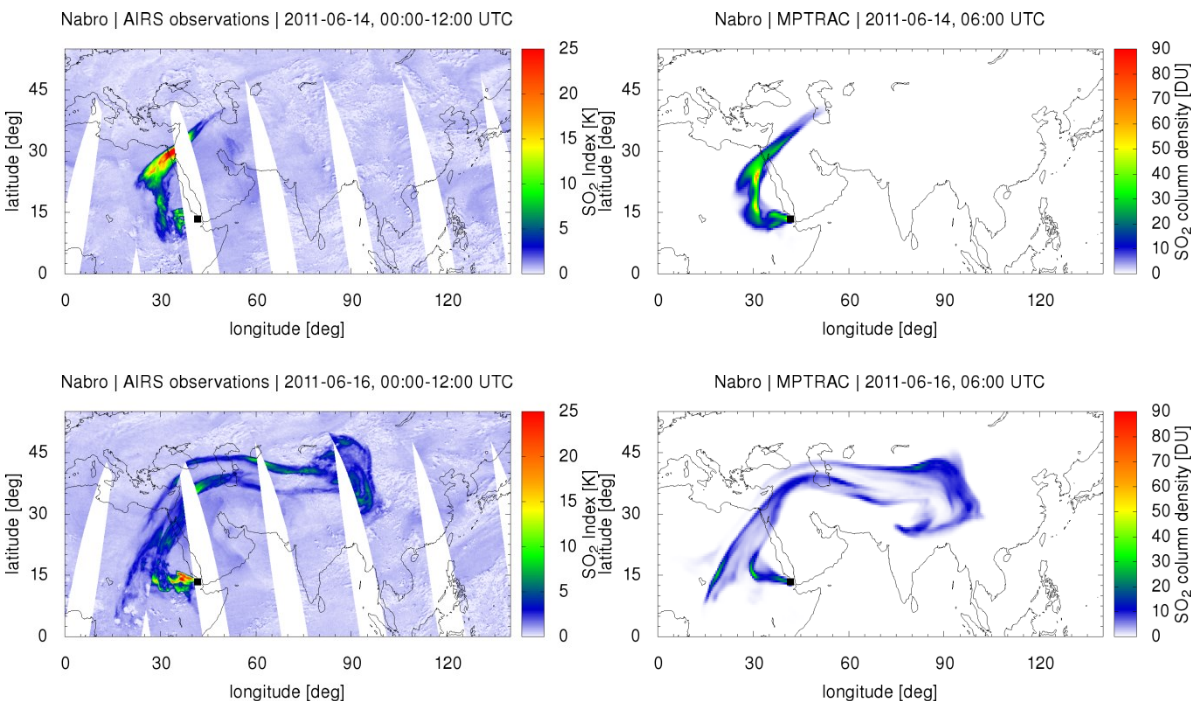
The version of MPTRAC used in this study does not consider any loss processes of SO_2. Hoffmann et al. (2016) considered a newer version of MPTRAC, which takes into account chemical decomposition of SO_2 based on a simple exponential decay model. Our analysis reveals that the simulation results by means of the two different versions of MPTRAC are rather similar. In particular, the relative distributions and patterns of the SO_2 plume remain the same. Only the total amount of SO_2 is changing slowly over time. Note that typical e-folding lifetimes of SO_2 in the UT/LS region are about 7-14 days, which is about the entire time period covered by the simulations.

7. "Validation":

The authors refer in several places to validation of their algorithm by comparisons to the AIRS satellite observations. However, this is not an independent set of data since these data were used to reconstruct the altitude-dependent SO_2 emission time series. This should be highlighted, and perhaps a better word is "evaluate". A better and independent dataset for validation would be IASI. In the above mentioned two papers (Theys et al (2013) and Clarisse et al (2014) there are many sources of data which you could use for validation. I consider this aspect the most important which needs extensive improvements. A thoughtful validation in lines with that presented in Theys et al. (2013) and Clarisse et al. (2014) is needed.

To extend the discussion we added in Sect. 4.6 in the revised paper: "*Our simulation by means of the product rule and AIRS satellite observations yields similar relative horizontal distributions of SO_2 on 15 June 2011 compared with IASI satellite data and FLEXPART model output as reported by Theys et al. (2013, Fig. 10a). Simulation results for other days, e.g., for 16, 18 and 20 June 2011 are also similar to the GOME-2 satellite retrievals reported by Theys et al. (2013, Fig. 10b – Fig. 10d). Our simulations (Fig. 12 and Fig. 14) show more realistic transport patterns on 14 and 16 June 2011 than the FLEXPART model outputs based on the IASI data (Theys et al., 2013, Fig. 12). Besides, the SO_2 distributions on 16 June 2011 and 18 June 2011 in China are not well captured by the FLEXPART model outputs based on the GOME-2 data (Theys et al., 2013, Fig. 10b and Fig. 10c), but by our simulations (Fig. 14 and Fig. 15). Furthermore, the SO2 transport patterns of our simulations are in good agreement with IASI observations that were extensively studied in the context of the Nabro eruption (Clarisse et al., 2014, Figs. 6–10).*"

Compared to the work of Theys et al. (2013, Fig. 12), we got the impression that our simulation may actually show even more realistic transport patterns, in particular on 14 June and 16 June, see figures below (left: AIRS observations, right: MPTRAC simulation results). However, this finding is based on visual inspection rather than a quantitative comparison.



To further validate the reconstructed SO_2 emissions, we added the following text on page 23, lines 16-18: "*Clarisse et al. (2014) also reported that the early Nabro plume mostly raised to altitudes between 15 and 17 km, which agrees well with our reconstructed emission time* *series (cf. Fig. 5).".* We also added text in Sect. 4.6*: "Furthermore, the SO_2 transport patterns of our simulations are in good agreement with IASI observations that were extensively studied in the context of the Nabro eruption (Clarisse et al. (2014), Figs. 6 –10).*"

We updated Fig. 11-17 in the revised manuscript by showing results for 13-16 June, 18 June, 20 June and 24 June) in order to allow for a direct comparison with the results presented by Theys et al (2013). We also added a brief introduction and overview of GOME-2 and IASI satellite measurements to Sect. 2.3 (validation data sets) in the revised paper.

8. Resolution:

In Section 4.1 you specify that you use a 1 h time step and 250 m altitude step leading to 15 840 emission domains. You do not specify how much AIRS satellite data you have, the temporal resolution and number of SI values. Do the AIRS satellite data contain enough information to constrain the emissions at this high resolution? Would the emission subdomain at 16.25 km altitude be sufficiently different from 16.5 km altitude? This also relates to the resolution of the meteorological data (both vertically, and the 3-6 hourly temporal resolution) you used for the unit simulations. Please elaborate

In this study, we tried to discretize the emission domain as finely as possible as permitted by the computing resources. This way, it is possible to reveal more local details of the emissions in case that high resolved meteorological and satellite data are available. It can also provide us information for the further development of an adaptive strategy for discretizing the emission domain, which will be considered in future work. During the time period 12 to 18 June AIRS detected volcanic SO_2 in nearly 75.000 satellite footprints, which means that the inversion is constrained by a large number of individual satellite observations. We added text on page 17 lines 14-16 in the revised manuscript: "*During this time period AIRS detected volcanic SO_2 in nearly 75.000 satellite footprints. Hence, the inversion of volcanic SO_2 emissions is constrained by a large number of satellite observations.*"

9. Section 4.3:

It is not clear to me how you obtain 0.1052 kg m^{-1} s^{-1} as the equal emission rate. Since you are using "binary" satellite data, i.e. the SO_2 index data, how does the inversion itself produce quantitative emission rates? Or are you distributing the 1.5x10*9 kg estimate from Clarisse et al 2012 over the entire emission domain? In that case I get 1.2e9 kg / (475200 sec*30000 m) = 0.0842 kg$m^{-1}s^{-1}$. Also later you say "since the total amount of emitted SO_2 is fixed" while this is not stated before. Please clarify.

The method does not provide an estimate of the total amount of SO_2 released, but tries to optimize its relative distributions in height and time. The manuscript was revised to make this more clear.

If we distribute the 1.5×10^9 kg total mass estimate from Clarisse et al. (2012) over the entire emission domain, the emission rate by assuming air parcels are equally distributed over the domain is 1.5×10^9 kg / (475200 sec × 30000 m) ≈ 0.1052 kg m^{-1} s^{-1}.

10. Section 4.4:

You should highlight that the comparison of the SO_2 emission rates with the aerosol observations (e.g. CALIOP) is not a direct comparison as one is gas and the other aerosol which would be either sulfate (converted from the SO_2) or ash. In Clarisse et al. (2014) this is nicely explained.

To clarify this we added text in Sect. 4.5 (validation of emission time series) on page 22, line 28, page 23, lines 1-4 in the revised manuscript. Besides, we replaced the word "measurements" with "aerosol observations" on page 23, lines 4-5 in the revised manuscript. Please note that the CALIOP and MIPAS observations used here most likely indicate the presence of sulfate aerosols rather than volcanic ash (Vernier et al., 2013; Griessbach et al., 2015).

Reviewer #2

Overview:

The paper describes a method to construct emission height and rate of SO_2 emitted from volcanic eruptions. The method juxtaposes large-scale ensemble simulations of a lagrangigan trajectory model and satellite retrieved SO_2 indizies (AIRS) to obtain these parameters in an iterative way. The method is applied to the 2011 Nabro eruption. The final forward model simulation using these parameters is evaluated with AIRS SO_2 and compared with imagery from MVIRI IR and WV and aerosol profiles from CALIPO and MIPAS.

General remarks:

Estimating volcanic emissions from satellite retrievals of ash or SO_2 is an important scientific task. The estimates are usually rather uncertain because of the limitations of the satellite retrievals and uncertainties in the transport simulation. Different approaches have been applied in the past and the one presented in this paper might be an interesting new approach.

However, the paper in its present form cannot convince the reader of the merits of the method and, more importantly, the validity of the results because of the following main points.

We carefully revised the manuscript according to your specific suggestions below.

(i) A major omission is that the results (emission parameters) are not compared with other studies presenting SO_2 emission (flux and height) estimate for the Nabro such as Theys et al. (2013).

We added two more references (Theys et al., 2013; Clarisse et al., 2014) in the introduction section to address their contributions to the study of the Nabro volcanic eruption. Discussions on the comparison of the SO_2 inversions and satellite/derived height estimates have been added and can also be found in our responses to the comments of Reviewer 1 (7. Validation). Correspondingly, we revised the introduction section and discussions in Sect. 2.3, Sect. 4.5 and Sect. 4.6 in the revised manuscript.

(ii) Although SO_2 and ash plumes sometimes coincide, but they do often not so. The evaluation of the SO_2 emission heights with aerosol retrievals (CALIPO and MIPAS) as well as the imagery is therefore questionable. Another SO_2 retrieval (IASI, GOME-2, OMI) would have been the best choice for the validation of the results with independent observations.

We extended the discussion and compared our results with the work of Theys et al. (2013) and Clarisse et al. (2014), which provide extensive studies of SO_2 retrievals and inverse modeling based on nadir satellite instruments such as IASI and GOME-2. The main motivation for a comparison with CALIOP and MIPAS is that these instruments provide accurate altitude information due to their measurement geometries (lidar and limb sounding), which is not directly provided by the nadir instruments. In the revised manuscript we tried to make clear that the comparison with CALIOP and MIPAS is only indirect (SO_2 versus sulfate aerosols). We would like to stress that both CALIOP and MIPAS have capabilities to distinguish between sulfate aerosols and volcanic ash. For the Nabro case study sulfate aerosols (formed by decomposition of volcanic SO_2) have been identified. We added text on page 22, line 28, page 23, lines 1-4, accordingly.

(iii) The choice of the AIRS SO_2 index (SI) data for emission parameter estimate needs to be motivated as it might not be the most suited data set for the inversion.

The method itself is designed to be independent of the choice of meteorological data and satellite data (from the implementation point of view), although the quality of final forward simulation results may be different due to the nature of the different data sets. In this study, we used AIRS satellite observations and ERA-Interim meteorological data as input to the inversion procedure. An intercomparison of the simulation results by using our method using other meteorological and satellite data products is beyond the scope of this work.

However, note that the AIRS SO_2 index used in this study provides significant improvements compared to the NASA operational data product (Hoffmann et al., 2014). The noise level of the AIRS and IASI radiance measurements is comparable. AIRS provides observations for 2.9 million footprints per day whereas IASI provides observations only for 1.3 million footprints per day. We were also interested in exploring the capabilities of the new AIRS data product for the estimation of SO_2 emissions rather than using existing data sets.

(iv) The basic methodology needs to better be explained and case specific fine tuning should better be avoided. It requires clarification what additional information – apart from the AIRS SI and the ERA-interim meteorological data – was fed into the inversion approach. It appears that the method does not actually provide a quantification of the emission rates. Also, the sensitivity to ad-hoc choices such threshold for plume presence in model and observations are not sufficiently discussed.

The method does not provide an estimate of the total amount of SO_2 released, but tries to optimize its relative distributions in height and time. The total mass of SO_2 for all air parcels needs to be specified as a priori information. The manuscript was revised to make this more clear (e.g., on page 17, lines 25-27, page 20, lines 1-2 in the revised manuscript).

We found that fine-tuning for the Nabro case study improves the simulation results. The sensitivity study on the weight-updating schemes in Sect. 4.4 in the revised manuscript illustrates that different choices of the split point introduce some differences (cf. Figure 8 in the revised manuscript) in the estimated relative distribution of the time- and altitude-dependent SO_2 emissions. We concluded that the split point is a parameter that should be optimized for each case and added corresponding text on page 22, lines 8-11 in the revised manuscript.

(v) The model does not seem to include any SO_2 loss processes (chemical conversion, deposition). The literature suggest a lifetime of about one to two a weeks. This will have an impact on emission parameter estimates.

We revised the manuscript to indicate that in the version of MPTRAC used in this study we did not consider loss processes of SO_2. Hoffmann et al. (2016) used a new version of MPTRAC, which took into account decomposition of SO_2. We found that the reconstructed emission data as well as the simulation results by means of the two different versions of MPTRAC are very similar for the time period covered by this study. In particular, the simulated distributions' patterns of SO_2 do not change, only the total mass is affected. Nevertheless, considering SO_2 loss in the future will likely provide an option for further optimization. We added text on page 27, lines 4-8 in the revised manuscript: "*Furthermore, the version of MPTRAC used in this study did not consider loss processes of SO_2. Hoffmann et al. (2016) used a newer version of MPTRAC, which takes into account loss processes of SO_2. Although the simulation results by means of the two different versions of MPTRAC are rather similar, a precise quantitative analysis considering the SO_2 loss will be subject of future efforts.*"

A positive aspect of the paper is the methodology the authors apply to evaluate the match between model and observations by using contingency tables. However, only the binary match (yes/no) w.r.t location seems to be tested. The approach should be developed further as the evaluation of volcanic plumes simulations can often not be done justice with simpler approaches.

The CSI is a frequently used measure to validate simulations of volcanic eruption events (e.g., Stunder et al., 2007; Webley et al., 2009; Harvey and Dacre, 2016). We pointed this out on page 13, lines 11-13 in the revised manuscript. Nevertheless, further statistical methods and additional quality measures can provide valuable information in future work. This aspect is also discussed by Harvey and Dacre (2016).

Specific remarks:

Abstract

Please give numbers how much the CSI improved from the constant scenario to your best estimate (max 32.3% - >41.2 and average 8.1%. ->16.6 %,) We added this on page 2, lines 23-27 in the revised manuscript.

P 9105

L 10: Using the SO_2 as proxy for ash and vice versa (as done in this paper) has to be done with caution. There are also examples of the separation of the two (Moxnes, et al. 2014 for Grimsvoetn). This point is of great importance for this paper as ash observations are used to validate the SO_2 emissions.

To clarify we added the reference and modified the text accordingly on page 3, lines 14-16: "*In practice, the presence of volcanic SO_2 can often be considered as a good proxy for the presence of volcanic ash (Sears et al., 2013), although in some cases different transport directions of SO_2 and ash were also observed because of different injection altitudes and vertical wind shear (Moxnes et al., 2014).*"

To clarify we also added text on page 22, line 28, page 23, lines 1-4: "*We would like to stress that both CALIOP and MIPAS have capabilities to distinguish between sulfate aerosols and volcanic ash (Vernier et al., 2013; Griessbach et al., 2015). For the Nabro case study sulfate aerosols (formed by decomposition of volcanic SO_2) have been identified. The comparison with CALIOP and MIPAS is only indirect (SO_2 versus sulfate aerosols).*"

L 17: Your collaboration in research activities is not of importance for the paper. Please omit.

We rephrased the text on page 3, lines 24-27, page 4, lines 1-2 in the revised manuscript: "*In order to further improve the quality of available satellite data, e. g., to perform more effectual suppression of interfering background signals, new detection algorithms for volcanic emissions for European Space Agency (ESA) and National Aeronautics and Space Administration (NASA) satellite experiments have been developed and are used in this study (Griessbach et al., 2012, 2014; Hoffmann et al., 2014; Griessbach et al., 2015).*"

L 20: Please discuss the limitation of the satellite observations of volcanic SO_2 in more detail. Mention the consequences for the source inversions but also for the evaluation of the forward model runs.

To clarify we added the text on page 9, lines 7-13 in the revised manuscript: "*The AIRS data product provides SO_2 indices for atmospheric columns, i.e., no vertical profile information on the SO_2 distributions is directly available. However, radiative transfer calculations showed (Hoffmann and Alexander, 2009; Hoffmann et al., 2016) that the SI of Hoffmann et al. (2014) is most sensitive to SO_2 layers at about 8 to 13 km altitude. Besides, nearly global coverage can only be achieved every 12 hours and there is information lacking for uncovered regions between the satellite scans. Note that the AIRS data product considered here has low noise, i.e., about 0.14 K at 250 K scene temperature.*"

In general, the solution quality of an inverse problem depends on the mathematical evaluation procedure, and also on how complete and how noisy the observational data are. For example, our reconstructed time- and altitude dependent emissions match well qualitatively with those obtained by applying a simple backward trajectory approach (Hoffmann et al., 2016) for the Nabro case study. A more detailed quantitative study of the uncertainties of the inverse solutions against data uncertainties is valuable, which we would like to consider in future work. Here the focus is on introducing the method.

L 25: Please mentioned also the NAME model, which is used by the UK met-office for ash plume forecasts.

We cited the corresponding paper on page 4, lines 9-10 in the revised manuscript.

P 9106

L 11: This does not make sense. Inverse techniques also use satellite data. You should have lists for the used observations and the applied techniques.

We modified the text on page 4, lines 21-26 and on page 5, lines 11-15 in the revised manuscript.

L 22: Please explain "Tikhonov-type regularization" or provide reference.

Corresponding reference has been added. We modified the text on page on page 5, lines 5-6 in the revised manuscript: "*A Tikhonov-type regularization method (Tikhonov and Arsenin, 1977, Seibert et al., 2000) was used to resolve the ill-posedness of the inverse problem.*".

L22: "objective function" , please clarify

In the field of mathematical optimization, an objective function, a loss function or a cost function is often defined for a minimization problem in order to find an "optimal" solution. To clarify we modified the text on page 5, lines 6-8 in the revised manuscript: "*The objective function defined for the minimization problem quantifies the misfit between model values and observations, but also enforces smoothness of the solution.*"

P 9107

L 1: Why only "nadir" and not limb sounders, the latter could provide better profiles

Although we considered the nadir satellite observations for the inversion of SO_2 emission, we have to mention that our method is in general also applicable to other sort of measurements. Hence we removed the word "nadir" on page 5, line 18 to give a more general message.

L2: Please explain the main idea of "sequential importance resampling"

Sequential importance resampling is a special type of particle filter (Del Moral, 1996) that is used to estimate the posterior density of state variables given indirect observations. The method approximates the probability density by a weighted set of samples. Here we infer the probability density of "hidden" variables (i.e. the SO_2 emissions at the volcano) based on indirect observations (AIRS detections of the SO_2 plume). The method provides the relative distribution of the SO_2 emissions. To clarify we added explanation on page 11, lines 10-17 in the revised manuscript.

L8: please explain the typical resolution of the discretization

For the case study of Eyjafjallajökull eruption, Stohl et al. (2011) used 19 vertical layers stacked up to 12.3 km altitude and 3-h time intervals for the discretization of the emission variable. In the work of Flemming and Inness (2013), different vertical heights between 2 and 15 km in intervals of about 2–3 km and more than 6 h time intervals were considered. In the current study, we considered a much finer discretization for the unknown time- and altitude-dependent emission variable (250 m in altitude and 1 hour in time). To clarify we added explanation on page 6, lines 12-15 in the revised paper.

L8: please explain why this needs massive parallel computing, what is your definition of "massive"

We refer to "massive parallel" simulations because a large number of processors is used to perform a set of coordinated computations simultaneously. The volcanic emissions in the Nabro case study are strongly time- and altitude- dependent. We tried to discretize the emission domain as finely as possible. This way, we would like to reveal the local details of the emission in a relatively small temporal and spatial resolution. For each small discretized subdomain, we need to perform a unit simulation. In this specific study 15840 MPTRAC simulations for all subdomains have been performed in parallel. These calculations are repeated multiple times.

L8: please give approximate number of calculation required

During the inversion we performed 15840 unit simulations in parallel in each iteration. We added this information "(> 10000)" in the text on page 6, line 4 in the revised manuscript.

L 14: Your method seems to have communalities with Flemming and Inness (2013) as they also use an ensemble of test tracers plumes and their match with observations to determine the emission parameters.

The work of Flemming and Inness (2013) used ECMWF's IFS as the forward model. In the current study, we used the Lagrangian transport model MPTRAC to perform the forward simulations. Compared to the work of Flemming and Inness (2013), we considered a much finer discretization for the unknown time- and altitude-dependent emission source term (250 m in altitude and 1 hour in time). The Critical Success Index (CSI) was used to evaluate the goodness-of-fit of the forward simulations and to estimate the importance weights of the time- and altitude-dependent SO_2 emission. This way, we provided relative distributions of the emission in a two-dimensional view (both in time and altitude) and its local details in a relatively high temporal and spatial resolution can be revealed. To clarify we added explanation on page 6, lines 9-21 in the revised paper.

L18: please motivate the choice of the AIRS SO_2 data. There other data sets e.g. from UV instruments such as OMI, GOME-2 or IR like IASI

Discussion and comparison of the SO_2 inversions with other satellite data have been added in the introduction and result sections. Please find more details in our response to the comments of Reviewer 1's (7. Validation).

L 20: please mention your evaluation with CALIPO and MIPAS aerosols

We added this on page 6, line 29, page 7, line 1 in the revised manuscript: "*Firstly, the reconstructed altitude-resolved time series of volcanic emissions are discussed and validated with MVIRI infrared imagery and CALIOP and MIPAS aerosol measurements.*"

P 9108

L 6ff: please provide reference for mid-point method, and the approach to simulate diffusion (Markov model is a very general term). Why do you distinguish between "atmospheric diffusion" and "turbulent diffusion" ?

We added a reference on page 7, line 11 for the midpoint method. We modified the text on page 7, lines 14-15 in the revised manuscript: "*Diffusion and subgrid-scale wind fluctuations are simulated following the approach of the FLEXPART model (Stohl et al., 2005; Hoffmann et al., 2016).*".

L 8: Please explain what chemical conversion processes or removal process of SO_2 are considered. If not this may have important consequences for your results.

The version of MPTRAC used in this study does not consider the loss of SO_2. Hoffman et al. (2016) considers a new version of MPTRAC, which takes into account loss processes of SO_2. The simulations with the two versions of MPTRAC show rather similar results, especially the relative distribution of SO_2 both in time and space.

L 9-13: Why is the detail on the parallelisation of importance here? Most of the atmospheric models require high amount of parallelism. Perhaps omit.

An efficient parallelisation was considered as a main feature in the design and development of the MPTRAC model. We would prefer to keep this short statement in the paper.

P 9109

L 1: Is the 6 h time resolution of the ERA-interim data good enough for trajectory calculations. Is this a limitation of your modelling?

In a recent study of Hoffmann et al. (2016), four different meteorological products have been compared for the Nabro simulations. Simulations with MERRA (3 h resolution) showed slightly lower performance than ERA-Interim (6 h resolution). However, the MERRA data had also slightly lower spatial resolution than the ERA-Interim data. This indicates that both spatial and temporal resolution of the meteorological data sets play an important role regarding the accuracy of simulations.

L 20: What is the possible range of SI, please explain how a quantitative information can be obtained from this index.

The maximum range of SI found in the data is about 0 to 50 K, but most observations are located between 0 and 20 K. Hoffmann et al. (2014) performed radiative transfer calculations for SO_2 layers located at different heights that can be used to correlate the AIRS SO_2 index with the SO_2 column density.

P 9110

L2: How is the limitation to this height range affecting your results?

Since we assume a pre-given amount of total SO_2 emission based on the study of Clarisse et al. (2012), namely $1.5x10^9$ kg, the emission rates for the equal-probability strategy will by definition vary with the chosen size of time- and altitude dependent initialization domain. In general, if the height range is chosen sufficiently large for the individual eruption case, the choice of the initialization domain will not anymore affect our results in the cases of mean rule or product rule.

Section 2.3

The evaluation of the final SO_2 forecast should use observations that represents the model result, i.e. SO_2. Therefore please evaluate with SO_2 retrieval from IASI or UV instruments such as GOME-2, OMI etc. You may use the CALIPO data or the imagery as secondary test for the injection height but you can not rely on them entirely.

We added a more detailed discussion for the validation of both emission reconstruction and forward simulation results by comparing with the work of Theys et al. (2013) and Clarisse et al. (2014). Detailed response can be found in our response to Reviewer 1's comments (7. Validation).

P 9111

L11: It is not clear how this quantification of the emission flux is achieved if only the match in location is testes with CSI

To clarify we modified the text on page 11, lines 7-17 in the revised manuscript: "*In this study, an inversion approach based on the concept of sequential importance sampling (Gordon et al., 1993) in combination with different resampling strategies is proposed to iteratively estimate the volcanic SO_2 emission rates the relative distribution of the volcanic SO_2 emissions. Sequential importance resampling is a special type of a particle filter (Del Moral , 1996) that is used to estimate the posterior density of the state variables given the observation variables. In our case, it aims to approximate the probability density by a weighted set of N samples, namely to infer the probability densities of "'hidden"' variables (the SO_2 emissions at the volcano) based on indirect observations (AIRS detections of the SO_2 plume). The SO_2 emission rates can then be calculated indirectly by assuming that the total SO_2 mass is known a-priori.*".

L25: What is the strength of the emission pulse for each of unit simulation?

In general, the number of air parcels used for the unit simulations is independent of the number of air parcels used for the final forward simulations, because only a relative distribution (relative weights) of the SO_2 emissions is the output of the inversion. For the unit simulations, we tested with numbers of air parcels ranging from a few hundreds to a few thousands, all yielded similar results for a final forward simulation with two million air parcels.

P 9112

L6: why "hidden"

In the terminology of particle filtering or sequential importance resampling the term "hidden" refers to the variables that are not directly (easily) observable. As emission rates at the volcano are usually not directly (or easily) observable, we refer to a "hidden" initialization.

Please see reply to comment on page 9107, line 2.

L12: Please make clear at what point the observations are used

To clarify we modified the text on page 12, lines 12-15: "*This way, the task of reconstructing the altitude-resolved time series of the volcanic emissions from satellite observations mathematically turns into the task of iteratively estimating the importance weight matrix W.*".

P 9113

L 9: It is not clear how you obtain the threshold of 0.1%. Is the SO_2 mass the mass of the ensemble unit simulations? What is the correspondence to the observed values. Or is it just a match between yes/no etc.

This threshold is considered as a method parameter. It defines the minimal number of air parcels that are needed to trigger a forecast event in a grid box. It needs to be adjusted to the number of air parcels used in the simulations. Based on a given total mass of all air parcels, it can also be converted into an equivalent mass per grid box or a column density. The threshold value of 0.1% used here was chosen empirically to optimize the results in the CSI analysis.

L 14: 4 DU is already a strong volcanic SO_2 signal. I could imagine that the choice of this threshold is important for the final outcome of your simulation. Please clarify.

The choice of the threshold is important as it determines the area fraction of the SO_2 plume that is validated by the CSI approach. As shown by Hoffmann et al. (2016), the threshold should be adjusted for different volcanic eruption events to cope with different sensitivity of the AIRS instrument and to obtain optimal results in the CSI analysis. For the Nabro case study we tested several values and found that 4 DU is a proper threshold value to be used.

P 9115

L 10: The choice of the split point seems important for the results. Is it just a heuristic choice. Please explain in more detail. How universal is a split point of 48 h. Please mention your sensitivity study in section 4.5

Section 3.3 provides a motivation for the product rule and for introducing the split point (cf. page 9114, line 18 – page 9115, line 13 GMDD paper). To further clarify we added text on page 22, lines 8-11: "*Note that the choice of the split point might be different for each particular volcanic eruption. A suitable value for the Nabro case study is 48 h. Nevertheless, our sensitivity analysis shows that the forward simulation results do not vary much with small changes (± 12h) of the chosen split point.*"

L 17: subdomains of the emission column?

We mean subdomains of the time- and altitude-dependent initialization domain. To clarify this we modified the text on page 16, lines 6-8 in the revised manuscript: "*Furthermore, the numbers of SO_2 air parcels in all subdomains (i.e. the discretized grid boxes of the initialization domain along the time axis and the altitude axis) are scaled linearly with the corresponding importance weights.*"

P 9116

L 11: How does the number of Clarisse et al. (2012) compare with your estimate.

In this study we used the total emission estimate of 1.5×10^9 kg by Clarisse et al., (2012) as an input parameter. All schemes (equal-probability strategy, mean rule, and product rule) yield information on the relative distribution of the emission rates in the considered time- and altitude-dependent initialization domain, but do not alter the total emissions.

P 9117

L 4: the AIRS data are only available for the respective overpasses. Please specify which orbits (times) have been used for the for the emission update. How does the temporal resolution of the data impacts the results.

To clarify we modified the text on page 17, lines 18-21 in the revised manuscript: "*For the reconstruction of the SO_2 emission rates we use the AIRS satellite data for 13-22 June 2011, which are measured at nearly fixed local times of 01:30 and 13:30.*".

In general, from the viewpoint of inverse problems, an increase of temporal resolution of the data will improve the accuracy of the emission estimates. For example, the "jumps" the emission time series in Fig. 5 are likely due to the limited time resolution of the AIRS observations.

P 9117

L 15: Does this simulation use a constant and uniform (in the vertical) emission flux? If yes say so.

A temporal and spatial initialization domain for the volcanic emissions is selected and finely discretized for the numerical computation. The first scheme, the equal-probability strategy, assumes a constant and uniform rate in the initialization domain. To clarify this, on page 18, lines 5-6 in the revised manuscript we added: "*, which leads to constant and vertically uniform emission rates for the simulation.*"

P 9118

L 1: Please say that you only test the match in space and not the SO_2 total column value.

To clarify we modified the text on page 18, lines 15-17 in the revised manuscript: "*Since the AIRS satellite data used here lack vertical information, only horizontally projected simulation* *results are used to test the data match in grid boxes. SO_2 column densities are not compared directly.*".

P 9119

L 3: But this all depends on the arbitrary choice of your split point.

Please see reply to comment on page 9115, line 10. The choice is considered optimal and not arbitrary. A sensitivity test is reported in Sect. 4.4 in the revised manuscript.

L 14: It is not clear where this number comes from. How did you obtain the total SO_2 burden? Please clarify.

To clarify this, on page 20, lines 1-3 we rephrased: "*By assuming a total mass of 1.5×10^9 kg for the entire initialization domain, the equal-probability strategy (first guess) considers an equal weight of wij = 1/15840 that leads to an equal emission rate of approximately 0.1052 kg m^{-1} s^{-1}.*"

L 20: Is the total emission obtained with product and mean rule the same or not. Please give numbers. The plot suggest different total emissions for the two cases.

The same total emissions of 1.5×10^9 kg are considered in this study for all schemes (equal-probability strategy, mean rule and product rule). The obtained maximum emission rates with respect to the mean rule and the product rule differ, but the integrated emissions in the considered initialization domain are the same.

P 9120

L 5: Why are they an underestimation?

We added text on page 20, lines 21-25: "*Since the total emission considered in this study (1.5×10^9 kg) is the same for all emission reconstruction schemes, and the mean rule yields some local emissions for unlikely cases (for instance at altitudes above 20 km), the emissions for more likely cases (e. g., on 13 June 2011, 00:00 UTC at 16.5 km altitude) are underestimated.*"

L 23 Figure 7 is not clear. Why is the diagram above the imagery? Perhaps two panels are better.

We put the WV images on top and IR images on bottom of the diagram by purpose. To add the reason for this, we added the following to describe Fig. 9 in more detail on page 22, lines 14-20:

"*From MVIRI WV and IR measurements (Fig. 9, top and bottom panel) we derived time series information (Fig. 9, middle panel) of the eruption history. The WV channel gives information on the high altitude eruption phase, because this channel is sensitive for altitudes down to the middle troposphere (around 6 km) where also the AIRS SO_2 channel is optically thick. In contrast, the IR channel reaches down to the ground and gives also information on low altitude plumes (e.g., on 17 June 2011).*"

P 9122

This whole discussions should perhaps be moved upward. In it is present from there are no real conclusions, which split point is best. Is it case specific etc.?

We moved the section of sensitivity study upward (Sect. 4.5 → Sect. 4.4).

We extended the discussion on page 22, lines 8-11 in the revised manuscript: "*Note that the choice of the split point might be different for each particular volcanic eruption. A suitable value for the Nabro case study is 48 h. Nevertheless, our sensitivity analysis shows that the forward simulation results do not vary much with small perturbation (± 12 h) of the chosen split point.*"

L 25 Please discuss which one is better. Are there any recommendation for the split point choice – or not. If not it is perhaps not necessary to include this section in the paper.

The section of sensitivity analysis aims to state that a case-specific split point needs to be chosen to apply the product rule. Nevertheless, the forward simulation results actually do not vary much with small changes of the chosen split point.

P 9123

L4: "equal-probability strategy" explain that this is the uniform and constant emission scenario.

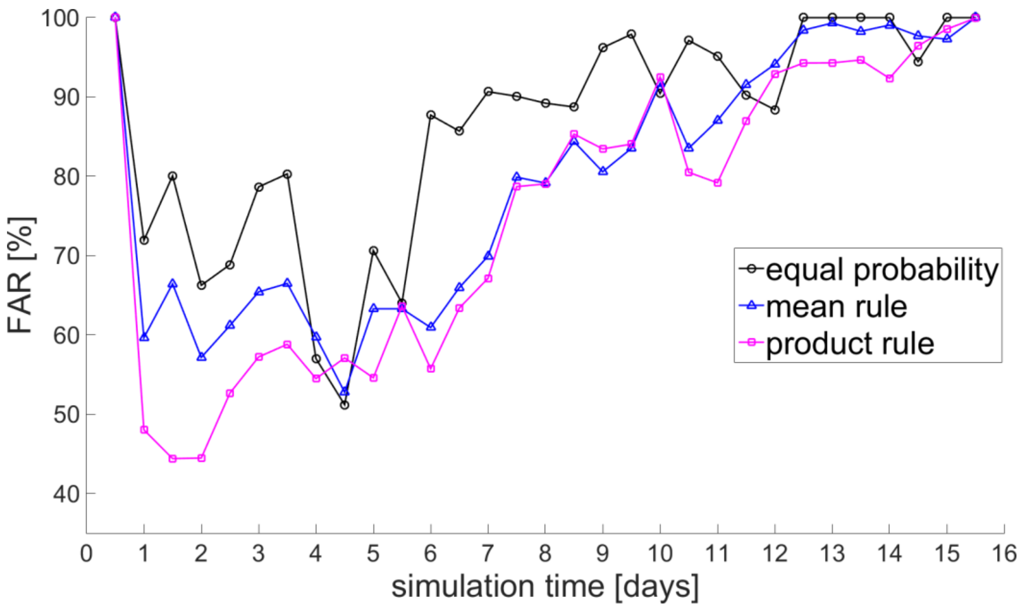
To clarify this, on page 24, lines 5-6 we added: "*Note that the equal-probability strategy assumes a constant emission rate in the entire time- and altitude-dependent initialization domain.*"

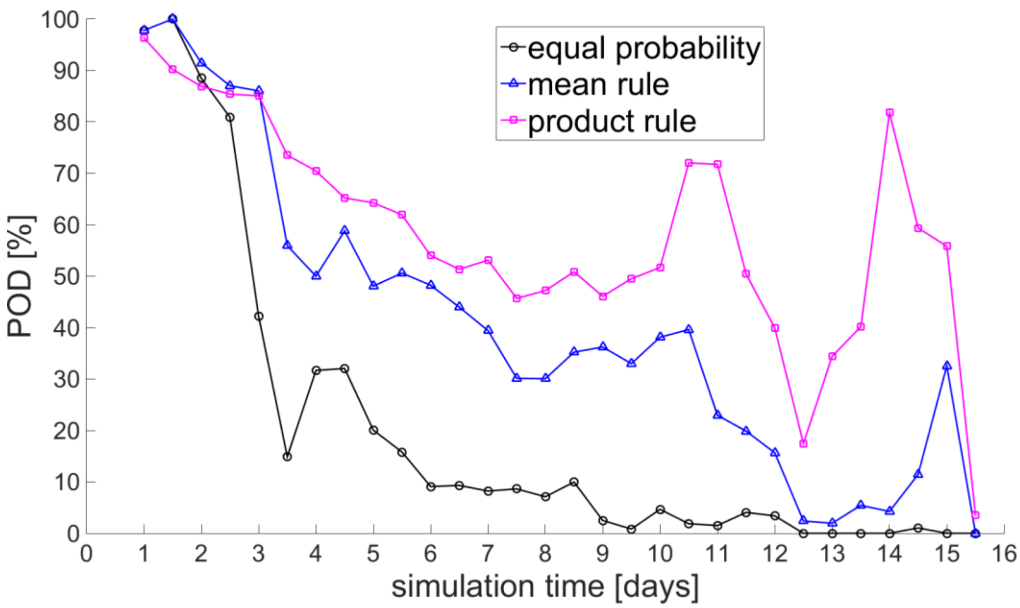
L 4: Do all emission scenarios have the same total or do they differ? Please clarify.

All three emission scenarios consider the same total mass according to Clarisse et al. (2012). On page 17, lines 25-27 in the revised manuscript we modified: *"The sum of these parcels then hold the total Nabro emission mass, which is estimated as 1.5×10^9 kg according to the work of Clarisse et al. (2012).*"

L 7: Are the number of "false alarms" (model = yes, obs = false) and "misses" " (model= no, obs = yes) more or less the same in the three model runs. Or does one type dominate? This would be important additional information. Please try to show also time series of these components of the CSI.

Below we compare the time series of "false alarms" and "Probability Of Detection" during 12 h time intervals obtained by applying the equal-probability strategy, the mean rule, and the product rule. The plots also confirm the message of the CSI plot, namely the use of product rule yields the best simulation results of the three cases. We added them into Fig. 10 in the revised manuscript and added corresponding text on page 14, lines 9-16, page 24, line 2, lines 15-17.





L 16: On average 16% hit is perhaps not so overwhelmingly high.

The application of the product rule provides the best simulation results of all three cases. Its maximum and mean CSI values are 52.4 and 21.4 %, respectively. The CSI values reported here are comparable to those in other studies on transport simulations for volcanic eruptions. For example, Stunder et al. (2007), Webley et al. (2009), and Harvey and Dacre (2016) mostly present CSI values between 10% and 50% for their simulations.

P 9124

L 8: Please clarify how the total emissions are assed and what the input data are and what your ad-hoc choices are.

In this study, the number of air parcels and the total mass of the SO_2 emissions are considered as inputs to the simulation system. Based on the information of the relative distribution of the

SO_2 total emissions in the time- and altitude-dependent initialization domain, which is estimated by the proposed inversion algorithm, the local SO_2 emission rates can be obtained (cf. Figure 5). To clarify we added the aforementioned text on page 25, lines 20-24.

L 9: Please say more clearly what the "equal probability assumption" is.

To clarify this we added text on page 24, lines 5-6 in the revised manuscript: "*Note that the equal-probability strategy assumes a constant emission rate in the entire time- and altitude-dependent initialization domain.*"

L 12: Make clear that this evaluation with the imagery is only qualitative.

We replace the word "assessed" with "qualitatively assessed" on page 25, line 27 in the revised manuscript.

L 16: Please give numbers how much the CSI improved for the three scenarios

To clarify we added text on page 26, lines 5-8 in the revised manuscript: "*The mean and maximum CSI values obtained by using the equal-probability strategy are 8.1% and 32.3%, respectively. The mean rule yields a mean CSI value of 16.6% and a maximum of 41.2%. The product rule leads to an improvement of the mean CSI value to 21.4% and of the maximum CSI value to 52.4%.*"