Manuel Marcelino Titos Luzón
Post Doctoral researcher, Icelandic Meteo Office, Iceland
manuel@vedur.is

Editor-in-Chief
Natural Hazards and Earth System Sciences

Dear editor;

I am pleased to submit a reply letter (extended in compliance with the reviewers' comments) of the article entitled "Assessing potential impact of explosive volcanic eruptions from Jan Mayen Island (Norway) on aviation in the North Atlantic" (MS No.: nhess-2021-264) to be considered for publication in NHESS.

The revision process and the thoughtful comments given by the reviewers have improved the quality of this manuscript. From its previous version, the manuscript has changed extensively (please find attached a latexdiff version). We have tried to motivate the originality of the work as well as clearly order the objectives that led us to tackle it. Several new references have been included supporting our ideas. Besides the modifications suggested by the reviewers, some others have been carried out in order to speed up the reading and the understanding of the manuscript. Amongst them, the authors would like to stress**:** 1) the usage of Flight Level (FL) instead of feet for the specific height at which the model results are calculated and shown. Thus, the references to 5000 and 25000 feet along the text have been changed to FL050 and FL250, 2) almost all the images have been modified from scratch following the comments of the reviewers making them more didactic and useful. Some of them have been merged into a single one in order to facilitate the comparison of results between eruptive classes, 3) Caption have been also modified making them self explicative, 4) considering that the results for the different airports in the U.K (London) do not show significant differences in the expected impact, the references to London Luton airport have been omitted (also figures), keeping only those to London Heathrow since it is the airport with the highest number of passengers in the country. As a complement, the same analysis for Edinburgh airport has been included in the new version, 5) Discussion section has been grammarly revised, 6) The isolines at some images have been plotted until 10% to avoid irregular and uninformative curves, which has had an impact on also in some conclusions.

Finally, to make clearer the interest of the approach presented (produce a long-term hazard assessment based on the information available from the past eruptive history of the volcano) and to avoid misunderstandings, we modified the title of the manuscript. The new title is**:** **Long-term hazard assessment of explosive eruptions at Jan Mayen Island (Norway) and implications on air traffic in the North Atlantic.**
We hope that this new version of the manuscript meets the expectations of the reviewers and the journal itself.

The authors confirm that neither the manuscript nor any parts of its contents are currently under consideration or published in another journal.

ANSWER TO THE REVIEWER'S COMMENTS

In the following, we have provided detailed answers to the comments of the reviewers. The original texts from the reviewers are in normal font. Our answers are in bold font. We would like to take this opportunity to thank the reviewers for their valuable comments and for their time and resources.

Answer to comments of Reviewer#1

We would like to thank Reviewer#1 for the careful reading of this manuscript and the thoughtful comments that have improved the quality of this manuscript. As the reviewer suggested, we have modified the manuscript and, added additional references. Furthermore, below are those comments that need more clarification.

Since it is a forward uncertainty analysis the results depend completely on the prior assumptions made by the authors, given very limited information about prior eruptions, in terms of type, magnitude and impacts. While the authors give some background to their choice of prior distributions the subjectivity inherent in such choices is rather glossed over. Indeed, it is in some cases presented in a rather pseudo-scientific way to appear more persuasive – for example the mention of the Akaike information criterion to choose between distributional forms in representing the distribution of total erupted volumes when the primary influence has to be subjective choice of probabilities for the different types of eruption and their range of magnitudes.

It is therefore somewhat ironic that the authors suggest that their analysis is unbiased (L353). This is clearly not the case, it is biased by all their prior assumptions. The results are the logical consequence of those assumptions (given enough samples) so it is not clear what they mean by unbiased (probably best to avoid it).

We thank the reviewer for this comment that gives us the opportunity to make some basic points of our work clearer.
Firstly, we did not mean at all to be "persuasive" (not to mention "pseudo-scientific", which in these days might indeed be taken as offensive when used by a scientist to describe other scientists' work). Rather we used the Akaike Information Criterion for exactly its goal, that is: objectively select, among six different possible statistical models (the 6 we used are among the most common ones to describe a continuous random variable), the one reaching the best balance between model simplicity and goodness-of-fit to the available data.
That said, it is true that we may have overstating by saying our analysis is unbiased. Following the reviewer's comment, we are going to smooth that part in the text, acknowledging that there is a degree in subjectivity in the selection of the types of PDFs used to model the eruptive size (we used 6 different ones as mentioned above, and chose the best among those by AIC) and those used to describe the variability in Eruption Source Parameters (ESPs).

In any case, in the manuscript we are going to keep the message that our analysis is "less biased" than what is commonly done in other hazard assessments in volcanology (and specifically tephra dispersal hazard assessment), where usually a specific sharp scenario is assumed (no uncertainty on the ESPs, so-called One Eruption Scenario in Bonadonna, 2006) or, less frequently, only the variability within a specific scenario is explored (Eruption Range Scenario in Bonadonna, 2006). The latter case is exactly the one used in the paper by Prata et al 2019, pointed by the reviewer: in that study, the authors assumed "The eruption plume height was set to 15,000 masl with a duration of 16 hr to ensure significant ash dispersal at cruise altitude (approximately 10 km)" and then explored 600 simulations by sampling the variability with a uniform distribution ("All model parameters considered in the LHS analysis were sampled from a uniform distribution. By sampling from a uniform distribution, it is assumed that all values between their specified ranges are equally likely.", sic). Further, the ranges over which the uniform distributions were sampled were based on the available literature "The ranges selected for each ESP were made based on typical ranges reported in the literature", sic).
In this view, we believe that our approach is "similarly biased" to the type of work by Prata et al (2019) by assuming ranges on the ESPs from the literature, and "less biased" than that in the sense that we fully explore the eruptive size variability (by simulating all the scenarios producing tephra, not only a specific one), and take into account their relative frequency as it appears from the geological and historical record at Jan Mayen.

The other major conceptual issue with the paper is the value of such a forward uncertainty analysis to inferences about impacts on air traffic. The results presented in the paper are integrated over the whole distribution of events simulated. They therefore clearly do not apply to any single event, and will not apply to the next event that might actually cause disruption. So certainly the areas indicated as at risk in the maps presented could be affected, and the maps could be used to guide an initial response of areas to avoid lacking further information. But application of such an approach to the next event would require Bayesian updating given information about that event (see e.g. Harvey and Dacre, 2016). Initially this might perhaps be to identify a sub-set of “similar” simulations but later perhaps using more explicit data assimilation – and about the specific meteorological conditions that might be quite different to those associated with a “similar” event run with reanalysis data, particularly in respect of wind fields relative to the integral of reanalysis sampling. I think this should be discussed further in the paper.

We thank the reviewer as this comment is very important for us because it demonstrates that our message on the main purpose of our analysis is not clear enough to the reader. Here we get the chance to clarify the general approach adopted in volcanology regarding hazard assessment, and the text in the manuscript will be changed accordingly.
In volcanology, one of the main challenge is actually to know which will be the next event and, in the case of poorly characterized volcanoes, as Jan Mayen, it is very difficult to estimate the absolute probability associated to each potential scenario. So,

it's difficult for us to understand what the reviewer means with "will not apply to the next event". As explained in the text, our goal is to produce a long term hazard assessment based on the information available from the past eruptive history of the volcano. Such a product is surely useful for long term planning.
When a volcano begins showing some restless conditions we might ask ourselves ""which will be the impact of the next eruption?" but more data (mainly from the monitoring) would be indicative of the potential expected eruption. For example the location of the potential vent, the volume of intruded magma, the presence of water bodies in the area of intrusion, might suggest which eruptive scenario is more likely than others. When eventually an eruption starts the amount of data available and the direct observations of the ongoing phenomena will be the primary input to dispersal models to answer a question like "what will be the impact of the current eruption, given the current meteorological conditions and the observed plume height?". In this sense the first case is what we call long-term hazard assessment, the second is a short-term hazard assessment, that eventually would become a more deterministic forecasting approach in the third case.
As we mentioned above, this paper aims to perform a long-term hazard assessment for Jan Mayen volcano (as stated at line 6-8 of the Abstract), which is currently in a repose phase, but is considered active and an eruption in the future cannot be excluded. For this reason in our study, the application of the dispersal model has been done by using primarily information on what this volcano featured in the past (in terms of size of eruption, volumes, plume heights, grain size distribution), as no more constraints on the next eruptive scenario can be done at this stage. The results presented here are indicative of the potential impact of an eruption in Jan Mayen in light of what is known about its previous eruptive activity. The maps produced could be used for land-use planning, definition of mitigation actions, identification of vulnerable infrastructure, but not for emergency response, which is out of the scope of our work. Whenever an eruption will start at Jan Mayen (as at any other volcano) the direct observations, current meteorological data will be used as long as data assimilation procedures and probabilistic forecasting products.
We hope we clarified this point and we will revise the text in this sense, however, in order to avoid misunderstandings, we decided to change the title of the manuscript as: "Long-term hazard assessment of explosive eruptions at Jan Mayen Island (Norway) and implications on air-traffic in the North Atlantic"

There is one technical point that I did not follow. On L377 it is stated that The probability of each combination is weighted in accord with the associated magnitude. It is not clear why this weighting should be applied. You have already sampled from the distribution of magnitudes which gives the (assumed) probability of such an event occurring, so why weight by magnitude?

We would like to thank Reviewer#1 for this comment. When sampling from a fixed probability distribution of magnitudes, the combinations are inherently being weighted by the associated magnitude (or assumed probability). This sentence has

been removed in the new version of the manuscript, since in section 3.2, line 140, we describe the sampling process and the combination of parameters clearly. In addition, we have modified the manuscript (line 361) and emphasized that:

The total erupted volume, expressed as DRE is computed uniformly within a range of values (10^8-$10^{8.9}$ m^3). In a second step, we weight each total eruption volume based on the Weibull distribution function previously defined (Figure 2). Doing that the unlikely events are properly represented.

Some minor points

Section 3.4 Location differences – is this a matter of resolution and limitations of the reanalysis rather than lack of real differences (what does such a wind profile really mean at reanalysis scales??).

We thank the Reviewer#1 for this comment. For the sake of clarity, we have modified the last sentence of this section. In the new version, we highlight that given the current limitations in the resolution of both the grid and meteorological data, the wind profiles analysis carried out over the different points of the Jan Mayen island (NE-SW) do not show significant differences. Therefore, we conclude that the location of the vent is a marginal parameter in our dispersion problem, since such location will not affect the final results. However, the location of the vent is a high impact parameter if other analyzes are performed.

The sentence has been corrected as:

Considering the current limitation of both grid resolution and meteorological data resolution, the location of potential JM vents does not influence the ash dispersal pattern. As a result, we will not consider the uncertainty on the vent location and assume a fixed vent at the middle of the island.

L85. Delete “occurred”

We thank the Reviewer#1 for this suggestion. The correction has been addressed.

L228. Predictions made without uncertainty quantification (UQ) are usually not trustworthy and inaccurate. I think this could be better worded. Normally for a forward uncertainty analysis the “best guess” prediction without uncertainty estimation would be within the modal range of the uncertainty analysis. Thus if it is not trustworthy and inaccurate so too are the equivalent ensemble members close to it … which is not really what you meant.

We thank the Reviewer#1 for this suggestion. We agree with the reviewer, inaccurate is not the correct word.

This sentence has been corrected as:

Predictions made without uncertainty quantification (UQ) do not quantify how constrained the prediction is, whereas ensembles members gives such an idea.

L276. This is due to the fact that the height of the eruptive column for medium eruptive class eruptions does not exceed 11 km (see section 3.1)

This a clear example of how inference depends on prior assumptions in a forward uncertainty analysis

We thank the reviewer#1 for this comment that gives us the opportunity to make some basic points of our work clearer. Our work addresses a comprehensive long-term Probabilistic Volcanic Hazard Assessment (PVHA) focused on the potential impact of airborne tephra concentration at different flight levels. To do that, we propose two differentiated types of eruptions depending on their magnitude, Medium and Large. In this sense, medium-sized eruptions have been defined to get volcanic plumes reaching up to 11 km height. Therefore, the analysis and the conclusions obtained are closely related to the characteristics of the modeled eruptive scenarios. However, within large eruption scenarios, the dispersion and concentration of tephra associated with plumes above 11 km are also addressed and, differently from the reviewer, we don't see any bias in that (columns higher than 11 km belong to the Large Magnitude scenario).

L389. This result, that we quantify at each point of the target domain, allows integrating hazard in quantitative risk analysis, through fragility curves. In this view, it represents the most complete way to quantify hazard.

This could be discussed later but is not really relevant here since you do not apply any such fragility curves (or mention evaluating their uncertainties at additional computational expense….)

The reviewer is correct here as we mention the possibility of extending the usage of our results towards a risk evaluation, without really doing it as it is beyond the scope of this manuscript. The sentence has been changed and now it is: "This result, that we calculate at each point of the target domain, could be eventually used as input for risk analysis like for producing fragility curves, tolerance analysis and in general investigation of impact on infrastructure. In this view, it represents the most complete way to quantify hazard."

L330. Finally we want to highlight the robustness of our PVHA in terms of uncertainty quantification, that should be routinely considered in all this kind of studies.

What do you mean by robust exactly? Why should your subjective prior assumptions be considered robust?

We thank the reviewer#1 as this comment gives us the opportunity to clarify the idea behind the robustness concept. Based on the arguments given within the framework of unbiased/less biased issue at the beginning of this reply, we think that while our work keeps the message that our analysis is less biased than what is commonly done in other hazard assessments in volcanology where usually a specific sharp scenario is assumed or, only the variability within a specific scenario is addressed, we fully explore the eruptive size variability (by simulating all the scenarios producing tephra, not only a specific one), and take into account their relative frequency as it appears from the geological and historical record. As a result, we are considering a large number of potential scenarios, while quantifying the uncertainty associated with these ESPs. In this sense, we think that our method, in terms of UQ (uncertainty quantification) is really robust and should be routinely considered in all these kinds of studies.

Answer to comments of Reviewer#2

We would like to thank Reviewer#2 for the careful reading of this manuscript and the thoughtful comments that have improved its quality. In the new version, the manuscript has been widely modified. As we pointed out at the beginning of this document, besides the modifications suggested by the reviewer, some others have been carried out in order to speed up the reading and the understanding of the manuscript. The answers to the questions raised by the reviewer have been carried out taking into account these modifications. Furthermore, below are those comments that need more clarification.

Summary

This paper deals with an unexpected eruption scenario, i.e. the resumption of volcanic (explosive) activity at Jan Mayen Island in the North Atlantic, a location unknown to most but whose eruptive historical record would suggest focusing on in the future.

The topic is certainly interesting and the possibility that an eruption from this so distant island could affect busy flight airspaces is largely proved. I think that the paper is an important contribution to the hazard assessment to air traffic in North Europe. The manuscript shows that much work has been done and that a lot of original data are presented.

The topic of the paper is addressed clearly and the introduction section gives the context to the research well enough and the motivation for doing the work. I have also suggested to add some references in some points. The analytical methods are very rigorous with regard to the probabilistic hazard assessment approach, allowing to compare in detail the different tephra dispersal simulations. The same thing unfortunately cannot be said for the description of the eruption scenarios, which is crucial for the right comprehension of the obtained results. This made it difficult for me to review this article due to several ambiguities in the selection of eruption categories. Moreover, several imprecisions are reported throughout the paper (both in the text and in the figures) that in my opinion need correcting.

However, my main criticism concerns the assumption of the representivity of a few eruptions to infer on the features of the studied explosive eruption classes. I think that the possible limits of such a characterization could be better discussed (much more and earlier than in the few lines in the conclusion section), reporting for example also the potential and limits of considering only a few eruptions, to validate and make the study more “robust”. See my comments in detail.

Finally, I suggest having the paper read by a mother tongue reader because although English is not my first language, I found several inaccuracies.

Overall, my view is that the paper may be deemed appropriate for possible publication in NHESS after reviewing the main critical issues. In the following, I set out some of my considerations and possible suggestions to improve the paper.

General comments

The paper needs to improve some descriptions to avoid misunderstandings and also to be more readable.

C1: The main problem is the first line of section 3.1 Eruption scenarios: “The possible eruptive scenarios at JM are based on 5 historical and prehistorical known eruptions”. The Cumbre eruption at La Palma is an excellent case of comparison with past eruptions which now doesn’t always seem coincident for the style, while, for example, we have to wait until the end of the eruption for some ESP such as the duration comparison. For example, at lines 101-102 you write that “Based on historical occurrence, this scenario can last for about 35-40 hours”: if the possible eruptive scenarios are 5, how many small eruptions are there? I think you should discuss these poor statistics much better because few eruptions may not be representative of the whole eruptive story in terms of ESP (plume height, duration, intensity and so on) and forcing/constraining the assessment toward a prevalent eruption type.

We thank the Reviewer#2 for this comment that gives us the opportunity to clarify some key points of our work.

Firstly, according to Gjerløw et al., (2016), Jan Mayen volcano tephrochronology reveals at least 8 eruptive periods over the last 600 years, 5 of them concentrated in the last 200 years, however, its Holocene eruptive history is basically unknown. It comprises only a very few distal sediment cores as well as lava flows and tephra deposits from eruptions on the ice-free parts of the Beerenberg volcano. Then, due to the lack of a complete geological record (composed of chronological and statistical data) and considering both the natural characteristics of the volcanic system itself (underlying geological framework, presence of ice / water, geothermal activity, vents location, etc.) as well as the existing historical reconstruction works (Larsen et al., 2017; Gjerløw et al., 2016), we described the conditional probability of different eruptive magnitudes based on fixed mass ranges and historical relative frequency using a Weibull Probability Density Function (Figure 2).

We fully agree with the reviewer that the recent eruption of Cumbre Vieja, La Palma, it is an excellent case of comparison with past eruptions which now doesn’t always seem coincident for the style. However, considering other eruptive scenarios would

also bias the results. In this sense, as we have said before, we agreed to categorize the different eruptive types based on the volume of tephra emitted in DRE (Dense Rock Equivalent) based on historical studies. This information is detailed in both Section 2 and 3.1.

C2: You often talk about estimation of "different flight levels" but actually, in the paper you always show only 2 levels (5000 and 25000 feet). Maybe can you explain this apparent incongruity once and for all? Can you explain on what basis you select the 2 flight levels of 5000 and 25000 feet? You should also indicate which are the most important flight levels to consider at that latitude and both close to and distant from the airports.

All references to "at different flight levels" along the manuscript have been changed to "at FL050 and FL250". Regarding the choice of these two specific flight levels, they correspond to cruise altitude and maximum risk actions altitudes like takeoff or landing. In this sense, the following description has been included in the manuscript, section 4 lines 212-215:

"Although our method allows analyzing any desired FL, in this work, only FL050 and FL250 will be analyzed. Such FLs were motivated considering cruise (FL250) and maximum risk actions altitudes like takeoff or landing (FL050). Finally, three selected ash concentration thresholds (0.2, 2, and 4 mg/m^3) were selected based on the impacts of volcanic ash on jet engines summarized in Figure 7 and the considerations included in the Volcanic Ash Contingency Plan published by the International Civil Aviation Organization (ICAO and VACP, 2016). "

C3: Medium and large are terms mainly used for category, class, eruption, size and so on. They are sometimes in capital letters, sometimes not. It would be better to make them uniform, or use them always in capitals, for example before "category". When possible, avoid different terminology, such as medium/large size eruption class, Medium size eruptive class, Medium ERUPTIVE class ERUPTIONS (?).

The correction has been addressed. We made these terms uniform along the manuscript.

C4: The moderate eruptions at line 103 seem to correspond to the future "Medium" class. You should describe these two terms (Medium and Large) better in a table to give more clarity on the terms about medium, moderate, large (eruption or magnitude size eruption, or class or category) and to better correlate the Table 1 with the two study cases (Medium and Large), or alternatively improve Table 1.

It is not clear even if the eruption scenario is coincident with the different eruptive magnitudes, namely if Table 1 correlates with Figure 2, because if so the terminology should be made uniform (e.g. small eruption with scenario effusive category and Moderate with Medium).

The term Medium has been included and correlated in Table 1 and Section 3.1.

C5: Also, the geographic framework is often not too clear and the reader (especially if not familiar with this area of the world) needs to consult internet to find some simple but lacking information crucial to follow properly the obtained results and the related discussion. For example, can you add a table with the airport locations, azimuth and especially distance from Jan Mayen?

In the new version of the manuscript a new table with the airport locations, azimuth and especially distance from Jan Mayen (complementing the visualization already done in Figure 1 where the airports considered in this work are plotted using black circles) has been added.

C6: The height of the volcano is important to better evaluate the scenario of the volcanic plumes (3-11 km or more than 10 km) and compare them with the flight levels. Are these heights all above sea level?

For the sake of clarity we have added a reference explaining that all these heights are above sea level (a.s.l) along the manuscript.

C7: At 5000 feet, the output maps in the different type of presented data seem to assign more hazard to the Medium category than to the Large one, while the opposite pattern (and much more clearly) occurs at 25000 feet. Can you explain the first, apparent contrasting result (the larger the eruption, the lower the hazard)? Can it be justified by the modelling used for the pulsating eruptions? If so, do you not think that the proposed methodology could have altered the magnitude (in terms of tephra emission and dispersal) of the Medium category?

As Section 3.1 and Table 1 explain, both eruption classes (Medium and Large) are mainly characterized by duration, plume height and total volume emitted. Then, considering that Medium eruption class generates volcanic plumes between 3 and 11 km above sea level (a.s.l) lasting between 4 and 40 days, both concentration and persistence probability maps should show greater exceedance probabilities at lower heights than Large eruption class. As the reviewer has pointed out, this can be easily seen in Figures 9, and 11-12.

From Figure 9, we can see that although at FL050 the area affected is approximately the same in both eruptive sizes, the probability of reaching ash concentrations at the given thresholds is greater in medium size eruptions. This is mainly due to the sustained duration of the pulses injecting ash at this height and nearby regions for long periods of time. The same conclusions can be obtained from Figures 11-12. Given the sustained ash emission into the atmosphere in Medium eruptive size, both the probability and the area affected increase when measuring persistence for concentrations of 2 mg/m3.

Regarding FL250, from Figure D1 it can be seen that the probability of reaching ash concentrations at the given thresholds increases in Large eruption, as well as the area affected and the persistence. Unlike Large eruptions where the plume height is obtained from the Mastin et al. (2009) formula from mass flow rate, in Medium size eruptions the height plumes are sampled from a triangular distribution centered around 6 km a.v.l. Therefore, many of the proposed scenarios do not inject ashes at FL250, affecting the final result.

Finally, answering the question if the proposed methodology could have altered the modelled impact of the Medium class, we want to highlight that this class includes different eruptive styles, so impact couldn't result simply in the rescaling of the Small and Large classes on the basis of the magnitude only. The choice of the classes and eruptive styles is based on previously published works supported by field analysis is explained in Section 2.

C8: Concerning some figures and partially the text, when you write "probability of reaching or exceeding ash concentration above …… at some time during the eruption up to 48 hours after its end", do you mean "at some time from the onset of the eruption up to 48 hours after its end"?

For the sake of clarity this sentence has been modified along with the manuscript. In the new version, we use "at some time from the onset of the eruption up to 48 hours after its end"

C9: Finally, Figures should be better presented so that they can be self-readable.

In the new version of the manuscript, figure captions are brief, but self-explicative.

In the following, my detailed comments and technical corrections on the text and figures.

Specific comments and technical corrections

Introduction

C10: Lines 22-23 - The eruption at Grímsvötn (Iceland, 2011) does not seem correlated to any of the following references. If so, please add at least one reference about this eruption. "economic, 2010" is not correct.

Correction performed

C11: Line 23 – Replace "remainder" with "reminder".

Correction performed

C12: Line 25 – I am not sure that "covid-19" is the right way to write this pandemic; probably it is better "COVID-19" in capital letters or "Covid-19".

Correction performed

C13: Line 27 – Add a space after "flights" and remove a space after "(".

Correction performed

C14: Line 32 – According to the standard guidelines, "(Gjerløw et al., 2016)" becomes "Gjerløw et al. (2016)". Please check this point carefully because eventually there are many other similar mistakes, e.g., line 44 "(Sandri et al., 2006)", line 62 "(Kandilarov et al., 2012)", and also in the caption of Figure 1 where some parentheses are not correctly inserted.

Correction performed.

C15: Lines 49-50 – “A novel strategy has been developed to treat and describe the styles of pulsating eruptions, characterized by a series of discrete short-lived events followed by occasional interruption of the tephra emission.”: this is very interesting and also very common although not always properly considered. Can you correlate some real deposits or eruption observations of Jan Mayen to this eruption style?

We thank the reviewer for the interesting comment. We think it should be possible to correlate tephra deposits at Jan Mayen to this eruption style by analyzing volcanic products in the field but we believe that a discussion about this point would go well beyond the scope of this paper. Moreover the mapped deposits in Jan Mayen are pretty poor and this would require a much deeper investigation. However, we have included the reference (Gjerløw et al., 2015) addressing the interpretation of the deposits, field reports and historical sources of the Eggøya explosive eruption (1732) (emplacing approx 0.3–0.4 km^3 (VEI 4) of tephra up to distances of at least 100 km from Jan Mayen and covering a minimum area of around 500 km2 within the 2-cm isopach) and presenting Surtseyan eruption with activity shifting between tephra jetting, continuous uprush and more magmatic phases as possible eruption scenario.

C16: Lines 58-59 – Replace “;” with “,” and “volume and address” with “volume, addresses”, and add “s” to “describe”. Alternatively, rewrite better the whole sentence “Section 3….. pulsating eruptions”.

Correction performed

Jan Mayen Volcanism

C17: Lines 66-67 – Insert the reference(s) which shows “at least five eruptive periods”.

Correction performed

C18: Line 70 – Can you give a rough estimation of how many km are these “Distal records”? From tens/hundreds to thousands?

According to the sentence: “Distal records as trachytic tephra found in Ireland (Hunt ,2004) and basaltic tephra found in older sediment-records in the North-Atlantic (Lacasse and Garbe-Schönberg, 2001; Brendryen et al., 2010; Voelker and Haflidason, 2015) or in Greenland ice-cores (Abbott and Davies, 2012) have shown the potential for explosive ash-forming eruptions whose size, frequency, and potential impact are, however, uncertain”, and based on table 2, these distal records correspond to hundred to thousand of km from Jan Mayen.

C19: Line 76 - Can “Beerenberg central volcano” and “Midt- and Sør-Jan volcanic ridge” be located on the map in Figure 1? Due to their importance in the subsection, you should eventually add an inset with a small-scale map showing these important volcanic structures. At the same time, to get a picture of the position of the study area it is necessary to go to Figure 4, while it should be useful to have e geographic outline in the first figure of the paper.

Following other suggestions an improved version of Figure 4 in Section 2 has been included in order to show Beerenberg central volcano and Eggoya crater as well as a general overview of the location of the Jan Mayen island and the study area.

Methodology

C20: Line 97 – This is the first time you present "JM". It is easy understand what you mean, but you should define it before and afterwards use the acronym when possible.

Correction performed

C21: Line 99 – Write "sub-Plinian" with the lowercase initial according to all the other citations.

Correction performed

C22: Line 100 –You are using eruption magnitude and VEI parameters without defining them somewhere (reference, table, other?). Especially the magnitude is a parameter not well known even among volcanologists and similar.

Correction performed. In the new version of the manuscript, references describing both magnitude and VEI terms have been included.

C23: Line 103 – Probably you should use the term "Surtseyan" with the capital (or not) letter everywhere.

Correction performed

C24: Line 122-123 – You are reporting a sub-Plinian type I eruption… I can imagine that someone defined at least two different types of sub-Plinian eruptions, but you should cite them and eventually define or characterize some features of this type (if necessary also in Table 1).

Line 122 and other lines – You reported wrongly the hyphen between the two values (108.7Ë˜109).

Correction performed

C25: Line 129 – Can you add a reference or more discussion about the methodology based on the "representative eruptive scenario"?

Some references addressing such methodology have been included in the new version.

C26: Line 140 – weighting or weighing?

Weighting. This correction have been fixed in the manuscript

C27: Line 143 – In Section 3.1 you have selected the possible eruption scenarios, not categories.

Correction performed

C28: Line 147 – Please, make uniform Section (capital or lowercase initial letter) everywhere throughout the paper.

Correction performed

C29: Line 150-151 –Can you give a short description and/or a reference about the “Akaike Information Criteria”?

2 references have been included.

C30: line 155 – You are talking about “Medium and Large classes”.. but in section 3.1 and Table 1 you define Moderate (“subaerial, sub-glacial and even surtseyan eruptions”) and Large (“expected to be initially subglacial and include moderate to sub-Plinian eruptions”). There is a little bit of confusion, are they classes, categories, or eruption scenarios as maybe provided by Table 1? See general comments.

Possible relative eruption scenarios on JM Island are based on the volume of tephra emitted in DRE. These scenarios can be splitted in Small, Medium, Large and subPlinian/Plinian (http://icelandicvolcanoes.is/?volcano=BEE; Gjerløw et al., 2016), however, according to the geological record (extending beyond Holocene), subPlinian/Plinian events are highly unlikely (1%). Because of this, they are not included in table 1. Each of these eruptive scenarios is based on different eruptive dynamics. Thus, for example, small magnitude scenarios are mostly associated with small lava flows or small scoria cones, while medium ones are characterized by effusive and/or Vulcanian to violent Strombolian or Surtseyan eruptions.

C31: Line 156-157 – What do you mean for “densities in the range of 250 and 350 kg/m3 and diameters between 100 and 250 μm”? Not clear to what you refer.

We mean densities and diameters of the particle aggregates being sampled in these ranges. These particle aggregates are being used to complement the Total Grain Size Distribution of the eruptive scenarios.

C32: Line 158 – Pulsating eruptions: according to Table 1, they belong only to Surtseyan eruptions. Can you remind the reader of this also in this subsection? So that they do not include the Large category but only the (Moderate scenario)/Medium category, right?

Correction performed

C33: Line 169 – Is the column height above the hypothetical vent or above sea level? How high is Jan Mayen island or the area supposed for the possible vent opening?

In this work, column or plume height are considered above sea level. When referring to Surtseyan (Medium) scenarios, we are supposing shallow sea eruptions. In the case of Large class, Beerenberg volcano has an altitude of 2277 m. We added the term above sea level (a.s.l) along the manuscript for the sake of clarify.

C34: Line 170 – Mastin et al. is not reported in the reference list.

Correction performed

C35: Lines180-180 – The wind patterns are overlapping each other at both sites. Which is the resolution of the wind data (i.e. the computational grid)? Is this a similarity real or artificial because the 2 sites result in the same cell?

The wind patterns were analyzed using ERA5 dataset. ERA5 dataset provides hourly estimates of a large number of atmospheric, land and oceanic climate variables. The data cover the Earth on a 30km (0.25° x 0.25° resolution) grid and resolve the atmosphere using 137 levels from the surface up to a height of 80 km. Considering that these 2 points are 55 km far away, we can conclude that this similarity is real because the 2 sites are in different cells.

Results

C36: In this section, you have only a subsection (4.1 Hazard maps and uncertainty quantification), so probably you can join the two titles or use only the second one.

Correction performed. Both sections have been merged into one single section titled: "Results: Hazard maps and uncertainty quantification"

C37: Lines 203 and 206 – Add "Island" after "in Jan Mayen", while if you mean the volcano, change to "at Jan Mayen".

Correction performed.

C38: Line 211 – How do you choose the selected ash concentration thresholds (0.2, 2, and 4 mg/m3)? By following international standard/guidelines on aviation safety or other considerations? Please specify.

Yes, these 3 thresholds were selected based on the aviation safety considerations included in the Volcanic Ash Contingency Plan published by the International Civil Aviation Organization. The reference of this contingency plan has been also included in the new version of the manuscript (Section 4, line 212)

C39: Lines 224-225 – You don't report that Figure 9 is referred to Large eruptions and Figure 10 to the Medium eruptions. The different flight levels are only 2 (5000 and 25000 feet).

Correction performed. The sentence was modified: " Figures 9 and D1 in the appendix show the probability of reaching or exceeding ash concentration above 0.2 mg/m^3, 2 mg/m^3, and 4 mg/m^3 at FL050 and FL250 at some time from the onset of the eruption up to 48 hours after its end, for Medium and Large eruption classes respectively."

C40: Line 228 - Does "It can..." refer to "predictions" (if so, it should be "They can") or "uncertainty quantification"?

Correction performed. For the sake of clarity "It can…" was modified by "Such UQ…".

Discussion

C41: Line 255 –At least in the first map at small scale (Figure 4), please write down the name of some places/countries (Jan Mayen, Iceland, Faroe Island, London, airports etc.).

Correction performed.

C42: Line 259 – Replace “ten” with “10”.

Correction performed.

C43: Line 260 – The description of the “probability of exceeding the threshold at any airport …” does not seem to fit with the Large eruptions (approximately 3 days) but also with the Medium eruptions (approximately 10 days). Can you check and eventually describe better?

Correction performed. The sentence was modified: ”The probability at any airport is neglectable during the first hours (approx. 10 and 15 hours for Mediumand Large classes respectively) and then increases until stabilizing after several days (approx. 7 and 5 days for Medium and Large classes respectively)”.

C44: Lines 263-265 – Sorry, I do not understand clearly to which figures you relate the sentences: “We can also highlight that after 48 hours since the beginning of the eruption, only medium eruption class exceeds probabilities above 5% to reach the threshold of 2 mg/m3. No airport shows exceedance probabilities for this critical threshold in ash concentration above 25%.” Can you refer clearly?

This sentence corresponds to Figure 13. In fact, the entire paragraph from line 265 to line 275, corresponds to the analysis of Figure 13.

C45: Line 275 – If you had defined 25000 feet in m as done for the 5000 feet level, it is easier for the reader to understand where an eruption plume of 11 km (or more) occurs.

Correction performed. The altitude in km has been defined in section 4, line 218, just when FLs in the results are presented.

C46: Line 284 – Insert “occur” after “when large size eruptions”.

Correction performed.

C47: Lines 285-286 – Why duplicate the same phrase: “For medium-size eruptive class, only polar routes above 25000 feet would be threatened. Then, we can conclude that for medium-size eruption class, only polar routes above 25000 feet would be threatened”? Eventually modify the first one and delete the latter one.

Correction performed

C48: Line 287 – “similar” or “similarly”?

The sentence was modified: “in a similar way to..”

C49: Line 290 – Add “s” after “represent”.

Correction performed

C50: Lines 300-304 – Not clear 1) if this sentence relates to a specific figure(s), and 2) if you find or not some differences depending on the eruption category; if so, please separate the description because it seems that you are talking about a unique, large/medium eruption category (that is, there is no difference between Large and Medium).

The sentencewas modified for the sake of clarify. The information was splitted to show better the difference between Large and Medium eruption. We wrote: "However, when analyzing the spatial pattern for long-term persistence (more than 12 hours), we find some differences depending on the eruption class. For persistence above 12 hours, at FL050, an ash cloud has 2% to 10% probability to reach latitudes as low as 65° N and 62° N for Large and Medium eruption respectively. Such southernmost latitude increases for longer persistence values, meaning that (obviously) only closer to the source we may get long-persisting clouds"

C51: Line 307 – To give more clarity, I would delete the bracket before London and use a dot before, or a comma followed by "while" or similar word.

Correction performed

C52: Lines 310-313 – The statement that "The sustained injection of tephra into the atmosphere related with a series of discrete short-lived events increases the probability of prolonged persistence scenarios" is important as well crucial, because it probably explains also that medium eruptions reach probabilities higher than large ones even in most (or all) of the previous plots. How can you support this? A pulsating event should disperse in the atmosphere much earlier than a continuous injection event of tephra, because single tephra pulses are not able to continuously "feed" a plume in the atmosphere as well and they singularly result dispersed more proximally than a continuous tephra emission.

First, Medium-sized eruptions can reach even greater mass flow rates (MFR) than Large eruptions, as long as the duration and eruption volume conditions provide it. This feature, coupled to the use of a mass fraction of 0.8 for Medium eruptions versus [0.05, 0.1] for Large one allows a greater injection of tephra into the atmosphere for the first case, thus increasing the persistence conditions for given concentration level.

Second, it is important to note that the duration of Medium-size eruptions (4-40 days) are much longer than Large-size ones (1-5 days), which, added to the aforementioned features, directly affects the persistence conditions. In addition, the rest period between pulses is computed considering the difference between the total duration of the pulses and the total duration sampled for the eruptive scenario (Figure 3), finding scenarios where there is no rest period between pulses, and having pulses with different duration and size, also directly affecting to the persistence conditions.

For these reasons and based on the results obtained, we conclude that the sustained injection of tephra into the atmosphere related with a series of discrete short-lived events increases the probability of prolonged persistence scenarios.

We hope now this point is clearer.

Conclusions

C53: Lines 314-315 – I think that you should have discussed the problem for the representivity of your data before describing the results i.e. in the methodology section, and discussing them there.

Correction performed. The problem of data representivity has been mentioned in the methodology section (lines 97-98) and discussion section (lines 261-264)

C54: Line 325 – Specify which are the low flight levels for better clarity. See also general comments.

Correction performed.

Appendixes

C55: Line 336 – Write Fall3d in capital letters.

Correction performed.

C56: Line 339 and 340 – large and medium are categories or classes?

Line 340 – “Since medium eruptive classes”: whatever you decide to use, I think medium can be related to a single “class” (or “category”).

We decided to use class or classes along the manuscript for the sake of clarity.

C57: Line 360 – The steps start in the section without describing what you are talking about. Please insert a short phrase preceding the workflow.

Correction performed.

C58: Line 366 – Add a “kg/s”. The upper size of the Large class is lower than the one for the Medium class… is that an error? Maybe 1.39 is elevated to the 105 and not 106? Should they be consistent with those in table B1 or not?

Correction performed. Regarding the limits of the mass flow rate ranges in both classes, the values are correct. They were obtained in accordance with the data proposed in this work. Large eruptions may have a lower mass flow rate than Medium size eruptions mainly due to the mass fraction of tephra in the volcanic plume and the duration of the eruption.

These mass flow rates (MFR) ranges are computed from the total erupted volume (TEV) in the PDF of Figure 3. Thus, for example, the TEV for Medium-size eruptions will vary between 10^8 and $10^{8.7}$$m^3$, while for Large ones, this value will vary between $10^{8.7}$ and $10^{8.9}$ m^3. To obtain the mean MFR range we need to compute the total erupted mass (TEM). This value corresponds to TEM = mass_fraction * density_average * TEV, where the density and mass fraction values are shown in Table B1, in the appendix, for both sizes.

Once the TEM is obtained, MFR is computed as MFR = TEM / eruption_duration. Thus, since the mass fraction in Medium category is much higher than in Large sizes, 0.8 against [0.05, 0.1] respectively, in the case of a medium eruption of short duration (4 days) and large TVE, the upper limit of the MFR may be greater than the upper limit of a Large one.

C59: Line 374 – Add a dot at the end.

Correction performed.

C60: Line 402 – Correct the space in "(X,Y ,Z)" between Y and ",".

Correction performed.

C61: **Competing interests**: change "uthor Giovanni Macedonio" to "author Giovanni Macedonio".

Correction performed.

References

C62: Check if "economics, 2010" is correctly reported or if it is better to write "Oxford Economics, 2010".

Correction performed.

C63: Budnitz et al. is not completed in the references list.

Correction performed.

Figure Captions

C64: Figure 1 – You are using the abbreviation or acronym "JMMC" without having written it in full here (in the caption). It is impossible to understand how the Jan Mayen volcano is related to the Jan Mayen island. You should insert at least an inset of the island showing the study area or how the volcano is located with respect to the island and how much the occupied area is approximately. This is also useful for better appreciating the related location of the wind pattern points.

Correction performed. Given the geological completeness of Fig 1, we considered it necessary to describe the geographical framework of the Island. Instead of modifying this image, Figure 4 has been improved and placed in section 2 in order to identify the location of both the island and those important volcanic structures such as Beerenberg and Eggoya (also described in this section). Following the reviewer's suggestions, some geographic references and scale bar to help the reader to follow the manuscript have been included.

C65: Figure 2 – In the caption you write "High" instead of "Large".

Correction performed.

C66: Figure 7 – I don't see the "Volcanic ash safety implications regions", probably I don't understand the colors or the shape of the boxes? Also, the meaning of the oblique dashed ellipses should be detailed. Can you add in the x-axis also the dashed threshold value of 4 mg/m3?

The figure comes from the Rolls Royce analysis on volcanic ash impact on RR engines and published as such. What they found is that the key factor is the dose, i.e. for how long an aircraft would fly within a specific concentration of ash. These concentrations are those used throughout the paper to analyze the potential impact on aviation. The ellipses represent events that have been documented and for which it is possible to prove the effects on engine functionality when exposed to such doses.

C67: Figure 8 - Change 1,5 to 1.5. At the first citation of this figure, the airports are not still cited.

Correction performed.

C68: Figures 9 and 10 – It seems that the Medium category has lower exceedance probability than the Large one: does it depend on the modeling of the pulsating (i.e. Medium) eruptions? Does it make sense? Can you discuss this result better?

Figures 13 and 14 – Similarly to the previous case, the Large category seem less impacting than the Medium one. Can you explain and describe better if true? The titles should report the term "map" as in the other similar (8-12) Figures (and Figures D3 and D4 as well).

Figures 9, 11 and 12 as well as their analysis have been addressed above in this text. Given the sustained character of the ash injection into the atmosphere in Medium eruptive sizes, both the probability and the area affected increase when measuring persistence for concentrations of 2 mg/m^3. As a result, we conclude that eruption source parameters associated with Medium class and pulsating dynamics have an important impact on the results obtained.

C69: In the title of the Figures where it occurs, use always the same format for mg/m3 – mg m-3.

Correction performed.

C70: In the Figure captions where they occur, it should be reported that the isolines of the maps are the expected arrival times in hours or probability percentages.

Correction performed.

C71: In the captions of all the figures where it appears "(Large)" and "(Medium)", adding category "(Large category)" and "(Medium category)" or other.

Correction performed.

C72: In the lowest isolines of some figures, some "spots" are related to small closed isoline.. if true, can you describe them in the text/caption?

We tried to soft the isolines and even reduce the resolution between grid-point. Unfortunately, these spots still appear in the figures. In the new version of the manuscript small closed lines have been removed using only values where the lines are well defined.

C73: Figures 15, 16, 17 – Don't you think that using the same range of values in the y-axis could give an immediate comparison between Large and Medium eruptions?

Correction performed.

C74: Figure 16 – Cut "of" after "above".

Correction performed.

C75: Figures 13, 14, D3 and D4 – You lost "18 hours" in the caption.

Correction performed.

C76: Figure D2 – You lost "feet" after 25000.

Correction performed.

Table Captions

C77: Table 1 – In the row "Eruption type", "volcanian" is erroneously written, moreover you should make it uniform with the other citations in the text with capital (or not) initial letter. Also, you should cancel "eruption" after Surtseyan to make it uniform with the other descriptions.

Correction performed.

C78: Table B1 – There is a "tephra" to correct.

Correction performed.