Response to the reviewers — Article GMD-2022-246

Title: *Reconstructing tephra fall deposits via ensemble-based data assimilation techniques*

We thank the reviewers for their constructive comments, which have allowed us to improve the quality of the manuscript. We have addressed the comments and incorporated your suggestions in the revised manuscript. The only exception is the Point P 3.3 (see justification below in reply to P 3.3). In summary, the revised manuscript includes new comparisons with traditional Ensemble Kalman Filter (EnKF) methodologies and with the first guess. In addition, a new strategy was adopted to define two independent datasets of observations (i.e. the assimilation and validation datasets). The Appendix was removed and the discussion there was incorporated in the main body of the revised manuscript to produce a self-contained work. In the following we provide a detailed response addressing your comments point by point. Our responses are written following each comment. All page and reference numbers in our response are based on the original manuscript. The line and reference numbers mentioned in the reviewers' comments are kept intact and are based on the original manuscript.

Text modified or added to the manuscript is given in this format: added text. Removed text is given in this format: removed text.

We hope that you find the following response satisfactory.

Sincerely,
L. Mingari, A. Costa, G. Macedonio, and A. Folch

General response

As suggested by Reviewer 3, a single dataset was generated by merging the full dataset of available observations (i.e. the previous assimilation/validation datasets). Subsequently, the complete dataset was split into two datasets by randomly selecting data in order to have a better coverage of the deposit region. For example, in Fig. 1 we split the full dataset into an assimilation dataset and a validation dataset with 60% and 40% of the observations, respectively. In principle, this should lead to a better deposit reconstruction since the assimilated data is homogeneously distributed. However, this procedure also has its drawbacks as the assimilation and the validation datasets are more correlated and the validation of the methodologies turns out to be more challenging now. This point is discussed in detail in the revised version of the manuscript.

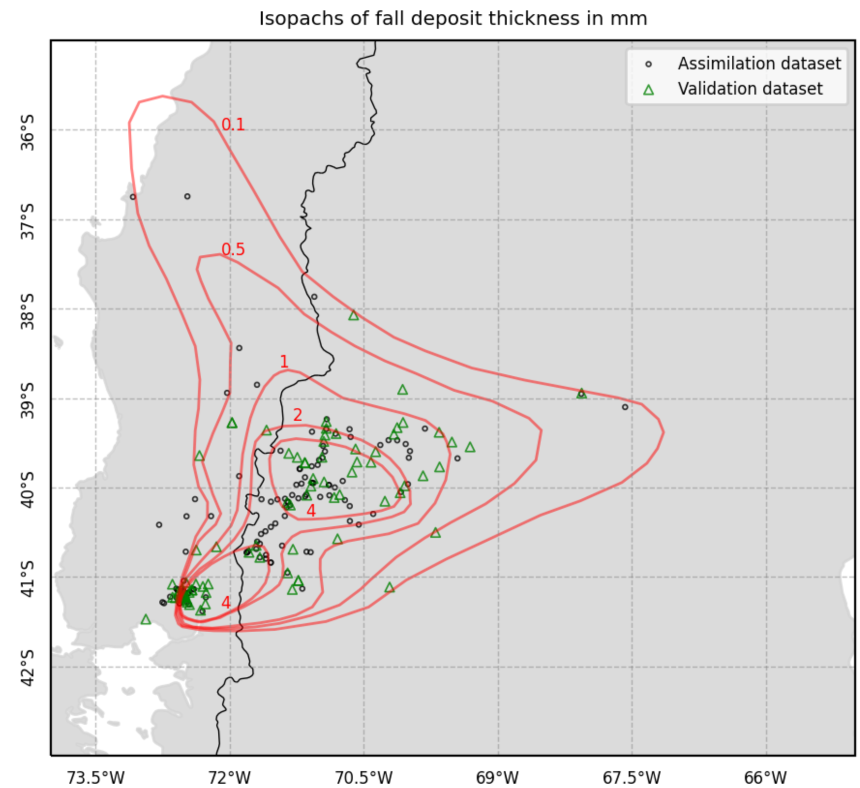

Figure 1: Location of sampling sites corresponding to the full (merged) dataset of observations. The dataset was split into two datasets by taking data randomly: an assimilation dataset (60%) and a validation dataset (40%). The isopachs of fall deposit thickness in mm used for validation is also shown with added labels.

On the other hand, following the reviewers' recommendation, we have included the results of the classical ensemble Kalman Filter (EnKF) in order to highlight the problems that arise when dealing with the assimilation of volcanological data using traditional approaches. As an example, the figures below compare the analysis of the GNC method (Fig. 2) with the results of the EnKF method (Fig. 3).

The EnKF method leads to a very noisy deposit field with strong oscillations and negative

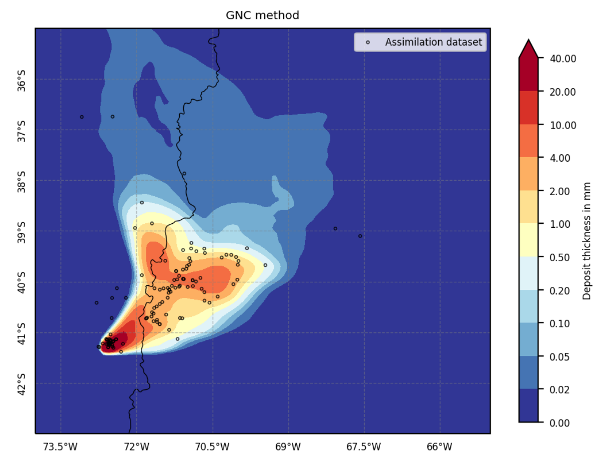

Figure 2: Deposit reconstruction according the GNC method.

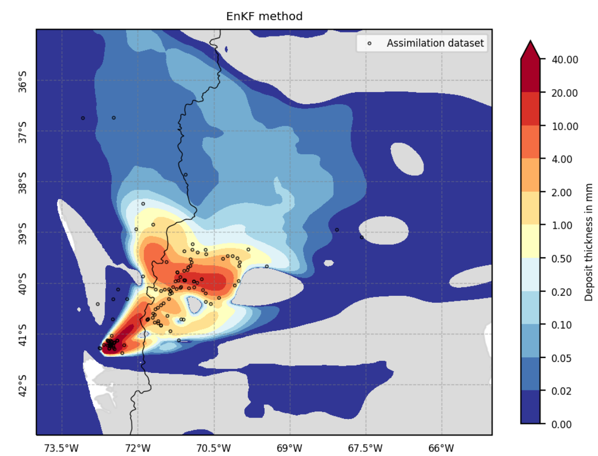

Figure 3: Deposit reconstruction according the Ensemble Kalman Filter (EnKF) method.

deposit thickness values (regions with no filled contours). In regions where no observation data is available, the EnKF method yields nonphysical results (i.e. negative deposit thickness). Comparing the analysis in Fig. 3 with the isopachs of fall deposit thickness (in mm) reported by Romero et al. (2016), shown in Fig. 1 with red contours, we can conclude that the EnKF approach produces artificial features in regions where no data is assimilated. This is discussed in detail in the revised version of the manuscript in order to highlight the importance of the methodology presented in this work.

Reviewer 1

General comments

Reviewer Point P 1.1 — My biggest concern is that I feel that a comparison against a "control" Kalman filter methodology is necessary to highlight the strength of the new methods tested. As the main aim of the paper is to show that the two methods used are better-suited for tephra, I think that it is necessary to show that a "normal" Kalman filter methodology is problematic, or at least that it leads to worse results.

> **Reply**: Done. Please, see the General response section above.

Reviewer Point P 1.2 — A smaller point is that I believe that the discussion section could be expanded a bit by including a paragraph that discusses uses of these methodologies beyond tephra, to highlight the strength of the methodologies in other settings.

> **Reply**: As suggested by Reviewer 1, we will include a discussion about the potential use of the proposed methodology in a broader context beyond tephra dispersal.

Specific comments

Reviewer Point P 1.3 — Abstract: I would change to "volcanic ash/tephra transport" as "modelling volcanic plumes" brings to mind a different kind of modelling

> **Reply**: Done. We changed modelling volcanic plumes by modelling the atmospheric transport of volcanic ash/tephra and gases

Reviewer Point P 1.4 — Line 21–23: This sentences seems a bit out of place in this paragraph as it focuses on the hazard. It would perhaps feel more natural as the first sentence of the following paragraph.

> **Reply**: Done

Reviewer Point P 1.5 — Line 29: well-recognised

> **Reply**: Done

Reviewer Point P 1.6 — Line 33: I would suggest adding a paragraph break here as it gets a bit too long

> **Reply**: Done

Reviewer Point P 1.7 — Figure 2: The font size in the figure is too small

> **Reply**: Font size was increased

Reviewer Point P 1.8 — Lines 300: I would change to "further understand the similarities"

> **Reply**: Done. Replaced dig further into the similarities by further understand the similarities

Reviewer Point P 1.9 — Figure 5: Too much white space and too small font size here. I would suggest using the y label only on panels a,e,i,m, x label only on the last row, legend only in one of the panels, panel letters within the panel (for example upper right corner), coordinates within the panel (for example middle right) and the just eliminating the white space and increasing the font size.

> **Reply**: Figure presentation was improved as suggested by Reviewer 1

Reviewer Point P 1.10 — Lines 379: Is there a reference for this statement?

Reply: This is shown in the revised version of this paper. For example, see Fig. 3 above

Reviewer Point P 1.11 — Appendix A: I feel that the appendix is very short and could be incorporated in the main discussion

Reply: Appendix was removed and the discussion there was incorporated in the main body of the revised paper

Reviewer Point P 1.12 — Figure 8: GIG

Reply: Corrected

Reviewer Point P 1.13 — Figure 9: What does the dot-dash line indicate?

Reply: The dash-dotted line indicates the mass emission rate ($\times 10^6$) in $kg\ s^{-1}$. Details were added in the caption.

Reviewer 2

Specific comments

Reviewer Point P 2.1 — L45: Did you assimilate the observation in the prior ensemble mean by using the traditional ensemble Kalman filters? Comparison of which can present the advantage of using GNC and GIG more clearly

Reply: Done. Please, see the General response section above.

Reviewer Point P 2.2 — L235: I cannot follow the analysis of observation error, which seems to be irrelevant to this work. What are these 7 groups referred to?

Reply: Assimilation methods require observation data and the corresponding errors. Consequently, the assumptions made to establish the observation errors are critical for this work. The strategy adopted in this work to provide reasonable error estimates is based on a clustering algorithm, i.e., observational data is organized into groups with similar characteristics. An absolute and relative error is assigned for each group or cluster. A more detailed explanation of the strategy used to estimate errors has been incorporated in the supplementary material.

Reviewer Point P 2.3 — L315: In Fig. 6, there are more than 4 (2) data points laying outside the 1:10 ratio band

Reply: Thanks for noticing this mistake. It has been corrected in the revised manuscript

Reviewer Point P 2.4 — L345: In Fig. 8 there are 5 counter lines. What does the first isopach denote? I also suggest to show the prior ensemble mean in Fig. 8, thus we can see how the (a) GNC and (b) GIG (typo in the title) improve the simulation of Tephra fall deposit.

Reply: We have added labels to contours to indicate the deposit thickness for each isopach (e.g. the outer contour denotes 0.1-mm thickness), as shown in Fig. 1 above. As suggested by Reviewer 2, the plot for the prior ensemble mean is also shown in the multi-panel figure. Finally, the typo in the title for the GIG plot has been corrected.

Reviewer 3

General comments

Reviewer Point P 3.1 — The methodology of validation is well-posed in the sense that validation is done against non-assimilated observations. However, the validation stations and the assimilated measurements (Figure 2) are located in quite different areas. Moreover, many validation stations are out of the regions where tephra deposits were significant. How valid are the validation measurements? I would recommand to build a single dataset with all the available data and then to split it homogeneously (using some random procedure for instance) into a assimilation and a validation one, unless there are good arguments not to do so.

Reply: As suggested by Reviewer 3, a single dataset was constructed by merging the previous assimilation/validation datasets. Then, the merged dataset was split into two dataset by randomly selecting data: an assimilation dataset (60% of the full dataset) and a validation dataset (40% of the full dataset). The advantage of this procedure is that we have better coverage of the deposit area and more homogeneous data. However, there is now a higher correlation between the assimilation/validation datasets. See also the General response section and Fig. 1 above.

Reviewer Point P 3.2 — In order to strenghten the results of the article, some plots may be added in the figures: the prior should be added to Figure 6 and 7, and a panel showing the prior should be added to Figure 8, in order to discuss the benefit of assimilation. It may be interesting, also, to plot the error values (any relevant indicator: bias, rmse or wrmse) of the analyses and the of prior at the different validation stations on a map; this could be an addition to Figure 8 for instance. In this way we could see how much and where error reduction occurs after assimilation.

Reply: Thanks for this comment. We fully agree with the reviewer's suggestions. We have added the prior to Figs. 6–8. In addition, we'll try to add errors to plots (error bars in log-log plots can be cluttered though) and other relevant indicators (this has not been done yet).

Reviewer Point P 3.3 — In its present form, the connection of Section 4 with the purpose of the manuscript is not clear. Section 4 addresses the estimation of the source term, and it seems out of the main scope of the article. It is an interesting topic, that would deserve a complete study by itself, if a more clear connection to the present manuscript is not provided.

Reply: In Section 4, we present a simple application of the GNC method and we think that it is not worth making a new article focused exclusively on this topic. In addition, in Section 4 we use this inversion technique to estimate the erupted volume corresponding to the first and second phases of the Calbuco eruption and show that are in a very good agreement with observations. In consequence, we decided to include this section to further support the GNC method and show that this method leads to physically consistent estimations.

Specific comments

Reviewer Point P 3.4 — lines 168-169: why only considering the GIG equations that assume a Gamma distribution for the prior and Inverse-Gamma for the observations? Arguments should be provided at this point, even though some results later (Figure 4) give some for the prior.

Reply: We provide additional arguments at this points to justify the use of the GIG equations and we refer the reader to Figures 4-5 and Section 3.4 to further support this decision.

Reviewer Point P 3.5 — Section 3.4: it is important indeed to evaluate the probability law of the prior ensemble. It would also be important to evaluate the dispersion of the ensemble and the prior error variance, and particularly to assess it compared to the observation error variance, in order to quantify the relative weight of observations and of prior in the assimilation process.

Reply: OK. We can easily address this by including error bars for observations and first guess in the new 6-panels figure replacing Figs. 6–7 (see reply to P 3.7).

Reviewer Point P 3.6 — Figure 2 (upper-right panel): are the isotachs from Van Eaton et al (2016) or from Romero et al (2016), as in Table 2?

Reply: These isopachs are from Van Eaton et al. (2016). However, these contours are never used in this work and were removed from the revised paper. Only contours from Romero et al. (2016) are shown now, as in Fig. 1 above.

Reviewer Point P 3.7 — Figure 6 and 7 may be merged into a single 4-panels figure, unless adding the prior in this figures (see General comments) makes it difficult.

Reply: As suggested by Reviewer 3, figures 6 and 7 were merged into a single 4-panels figure (actually a 6-panels figure after adding the prior).

Reviewer Point P 3.8 — Line 376: please rephrase "based on the Gaussian hypothesis", since Gaussian hypothesis is not formally a condition for applying EnKF. In case of non-Gaussian errors, the EnKF can apply, but it provides a result (analysis) that is not optimal.

Reply: We understand that EnKF can be applied to a wide variety of situations with non-Gaussian errors and still obtain reasonably good state estimates. However, the original Kalman filter equations (Kalman, 1960) and EnKF make the assumption that all probability distributions involved are Gaussian. In consequence, I see no reason to rephrase this sentence.

Reviewer Point P 3.9 — Line 405, it is stated that "These reasons make the GIG method a better candidate for implementation in VATD models". However, "The GIG method is a sequential assimilation procedure proposed by Bishop (2016), in which single observations are sequentially assimilated" (line 402), and from this sequential aspect we may assume that paralleization is not possible. So what is the potential impact of this sequential aspect on runtime and on operations? Are some parallelization strategies possible? This deserves to be discussed when addressing operational implementation.

Reply: Exactly. Parallelization is not possible for this method since it involves a loop with dependencies inside it and I don't see an obvious parallelization strategy possible here. In fact, the execution time of the GIG method code was significantly longer than the execution times for the GNC and EnKF codes. For that reason we are talking about a possible 'candidate'. But we're not entirely sure if it's suitable for operational forecasting. Still, we suppose that computation times involved in VATD solvers could be dominant in most of the applications. The GIG performance problem is discussed in the revised manuscript when addressing operational implementation.