

Reply to Reviewer 1

We thank Reviewer 1 for the thorough review of our manuscript and their thoughtful comments focusing on improving the scientific quality of the revised manuscript and its readability to reach a broader audience.

The reviewer addresses the review by separating their comments in general and specific comments. The reply will be structured in the same manner and replying point-by-point. We present the plain text in the interactive discussion and the formatted reply as an attachment. Firstly, the reviewer comments will be copied in **bold, blue color** and will be followed by the replies of the authors, for the sake of the answer legibility. All the manuscript editions from the Reviewer 1 suggestions are incorporated into the attached newer version of the manuscript (marked **in blue** in the text), providing in the revised document their exact location in terms of manuscript lines, and copied after the answer in grey color when possible.

General comments

- 1. Overall manuscript is very technical and mathematically heavy, made it quite difficult to fully understand. I will suggest rephrasing some of your text to increase the readability for more audiences.**

The authors acknowledge the technicality of the manuscript, although we find that the new formulation of matrix factorization through the horseshoe prior indeed demands the thorough and detailed mathematical description to properly present the work frame to advanced users. However, we agree on the need to present this work in a more accessible way to better display the advantages and limitations of the method for end-users. We aim to clarify and land the meaning of these expressions through Figure S1 (a) and (b) and through some more grounded explanations on the Methodology

Lines 153-155: "... The idea behind this prior is that species with very small contributions to a factor are shrunk toward zero through an automatic shrinkage mechanism, whereas species with substantial support from the data are largely unaffected. ..."

Line 220: "... The parameters of the model, primarily \mathbf{F} and \mathbf{G} but also all the other defined parameters (τ , λ , α , β) are sampled from their posterior distributions, ..."

Lines 221-224: "... In the Hamiltonian analogy, the evolution of these parameters across samples is computed as the trajectory of a fictitious particle. This particle moves through the parameter space driven by random momentum in all directions. This approach avoids the random-walk behavior of simpler sampling methods and enables faster convergence. ..."

Lines 224-226: "... The trajectory is hence simulated using a discretized approximation, and candidate positions are accepted or rejected according to the Metropolis criterion (Metropolis et al. 1953). Accepted positions correspond to plausible parameter values given both the model assumptions and the data. ..."

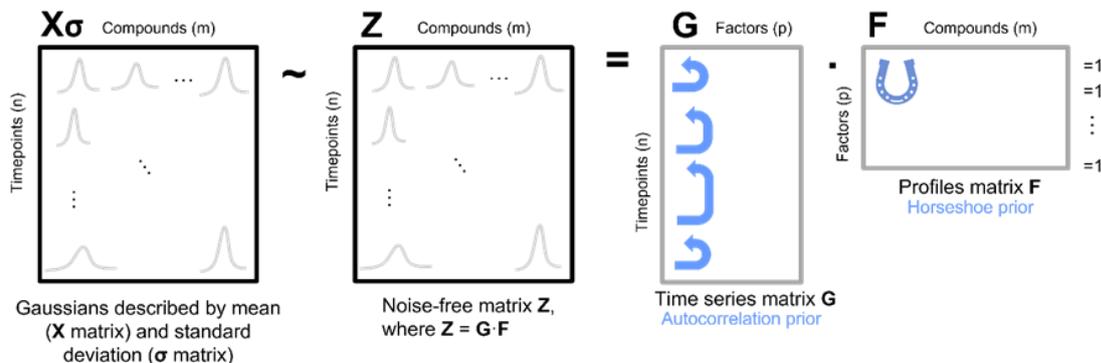
and Discussion sections.

Lines 609-611: "... Firstly, the use of the simplistic toy dataset highlighted the added value of the sparsity introduction through the horseshoe prior in the totally constrained experiment. In this controlled setting, the ground truth structure is well defined, allowing the effect of sparsity to be clearly isolated and the method performance validated. ..."

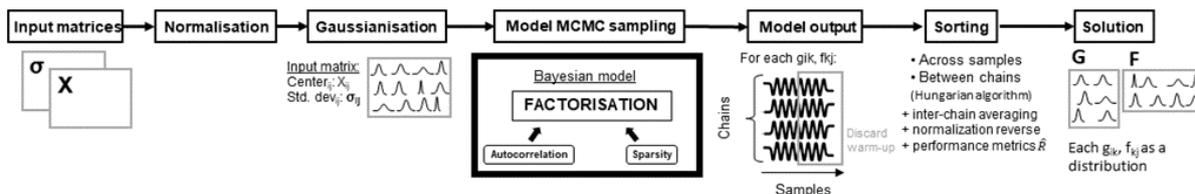
Lines 620-623: "... Its application also proved advantageous for the real-world dataset, despite not being able to be compared to the truth. In this case, improvements are assessed through increased profile distinctness and internal consistency rather than absolute accuracy. ..."

Lines 645-647: "... The BAMF+HS method, contrarily, acts globally, shrinking those species with lowest signals in favour of the matrix factorisation, hence no user intervention is needed. This makes the approach more objective but also less targeted, returning the factorization optimisation agency to the model. ..."

Figure S1, which was already in the submission version of the manuscript, shows the Bayesian work frame for aerosol source apportionment through matrix factorization, where each of the \mathbf{X} , \mathbf{Z} matrices elements are distributions. It also schematises how the autocorrelation and the horseshoe priors are applied into the \mathbf{G} , \mathbf{F} matrices.



The newly created figure, which will be shown in the newer version of the manuscript as Figure S1 (b) shows the source apportionment workflow for BAMF+HS. It portrays the previous and posterior processes to the MCMC sampling, and both starting and end points. Concepts as chains and samples, mentioned in the text, are schematized in the workflow as well.



This newly modified Figure S1 has been included in the supplementary with the following caption:

(Lines S22-S25): **Figure S1. (a) Bayesian matrix factorisation for aerosol source apportionment sketch with autocorrelation and horseshoe priors. (b) Workflow diagram showing the aerosol source apportionment stages with the BAMF+HS model. This figure portrays the previous and posterior processes to the MCMC sampling, and both starting and end points. Concepts as chains and samples, mentioned in the text, are schematised in the workflow as well.**

Moreover, some adaptations on the text have been carried out in order to make the text more accessible (see specific comment 6).

- 2. Given the complexity of this approach, a schematic flow diagram will be appreciated to demonstrate the workflow of BAMF+HS and when and where this approach is the best applicable (i.e., roadmap)**

The Figure S1 (b) (lines S22-S25), added in the revised version of the manuscript, aims to depict the workflow of BAMF+HS procedure.

- 3. Explain what the $F \rho$ or F contr. ρ are in your tables/figures**

The description of the metrics to assess the quality of the solutions in comparison to truth are described in the revised version of the manuscript as follows:

(Lines 284-286): The similarity to truth, when available, is tackled by comparing the median ratio between modelled G and truth (G/G_0), the Pearson correlation for the G matrix ($G r$), and the Spearman correlation for F amongst models ($F \rho$).

- 4. Just out of curiosity, how does BAMF+HS will perform on online Xact dataset?**

The BAMF+HS represents a propitious method for element datasets, as shown for the offline synthetic and real-world datasets. Its sparsity enforcement enables minimizing non-significant contributions of certain species in some factors, hence, these should contribute fully to their relevant factors. The more accurate profile also promotes more accurate time series, obtaining a better solution overall.

Even if the usage of the BAMF+HS for Xact is highly recommended, the authors want to make a note regarding its logistics. The use of Bayesian inference for filter-based datasets, usually of lower time resolution is already computationally-intensive and demands high-performance computer resources. The Xact source apportionment (typically with 1 hour time resolution) would therefore require substantially greater computational effort, which may represent a practical limitation for routine analyses or long campaigns. Nevertheless, compared to other monitoring instruments such as an ACSM, it would require less computational resources given the same measurement span and resolution due to its lower number of species (~20 for Xact, ~100 for ACSM). Fortunately, ongoing efforts are being directed toward code optimization and the transformation of the methodology into more efficient computational frameworks capable of handling larger datasets.

5. Together with the flow diagram/roadmap, provide some limitations and recommendations when and where your approach can be used.

The recommendations and limitations of the presented method are not shown in the flow diagram in the revised version of the manuscript since they require a higher level of detail and contextual explanation that cannot be conveyed through a schematic representation alone. The purpose of the flow diagram in Figure S1(b) shows only the ensemble source apportionment process steps.

Instead, the conclusion section provides a thorough, summary evaluation of the proposed work and explains its highlights and limitations. This section explicitly outlines the conditions under which the method performs well and indicates the problems stemming from certain kinds of datasets, as shown below:

(Lines 672-675): “BAMF+HS has been shown to be advantageous to introduce sparsity in factor profiles for offline datasets and to not deprecate the solution for the more complex ACSM-like datasets. “

(Lines 683-685): “The BAMF+HS model does not create sparsity in ACSM-like datasets, which are originally, indeed, non-sparse. BAMF+HS is more unstable than BAMF for these more complex datasets as a result, the higher chain divergence during HMC sampling as suggested by the metric. However, the effects of the horseshoe prior do not affect the overall performance of BAMF or its autocorrelation accuracy.”

The manuscript highlights that BAMF+HS is particularly well suited for introducing sparsity in factor profiles for offline datasets, where sparsity is physically meaningful and improves factor interpretability. For more complex ACSM-like datasets, the BAMF+HS is not as fruitful due to convergence issues, but importantly, the manuscript clarifies that these effects do not compromise the overall performance or its ability to accurately reproduce autocorrelation structures.

6. Captions for figures and tables needs to be more informative and self-explanatory without reading the main text.

Some changes were made in the revised version of the manuscript to make some captions more self-explanatory:

(Line 854): “**Table 1. Models used in the current study and their priors on the G, F matrices.**”

(Line 857): “**Table 2. Profile sparsity metrics for the truth of the synthetic datasets.**”

(Line 867): “**Figure 1. F matrix components distributions for CMB and CMB+HS (solid lines) compared to truth (markers).**”

(Line 870-871): “**Figure 2. Source apportionment results for the toy dataset obtained using PMF, BAMF, and BAMF+HS, compared against the true solution (black bars). (a) Factor time series. (b) Factor profiles.**”

(Line 880-882): “**Figure 4. Profile components distribution for PMF, BAMF, BAMF+HS (solid colored lines) in comparison to the truth (markers) on the real-world filters dataset. Rows represent the species of the source apportionment and columns represent sources.**”

(Lines 774-776): “Figure 7. European cities synthetic datasets summary statistics; from top to bottom, median ratio time series with truth (G/G_0), Pearson correlation coefficient of G with truth (G_r), Spearman correlation coefficient of F with truth (F_ρ). The axis of the G/G_0 plot is in logarithmic scale.”

Specific comments

7. Line 25: what is Toy?

The toy dataset has been here described as a highly simplified dataset as follows:

(Lines 29-30): “evaluated using three synthetic datasets designed to reflect increasing levels of data complexity (Toy, representing a highly simplified dataset; Offline, representing a filter dataset; and Online, representing an ACSM-like dataset), and a real-world multi-site filter dataset.”

A further, more detailed explanation of the toy dataset can be found in section 2.3.1.

8. Line 28: define ACSM

ACSM was directly presented as an acronym in the submitted version of the manuscript. This is changed in the revised version of the manuscript:

(Lines 32-33): “However, its application to higher-complexity Aerosol Chemical Speciation Monitor (ACSM) datasets ...”

9. Line 39: Be more careful in referring OP as toxicity. I will be more conservative on this by refereeing it to one of the health metrics.

The authors acknowledge the too direct association between oxidative potential and toxicity and have rephrased that sentence as follows:

(Lines 44-45): “Moreover, because some proxies for aerosol toxicity, among them oxidative potential, are highly dependent on its sources (Daellenbach et al. 2020), ...”

10. Line 43: first time introducing PMF needs to be spelled out. Also, PMF is one of the RMs to conduct source apportionment analysis instead of the only approach to do source apportionment. Please rephrase. Also, SA is not just identification, but also quantification sources. You will need to make that clear.

The PMF was mistakenly introduced in that sentence, which should not contain any PMF reference. We also add the quantification aspect in the reformulation of this definition, which results as follows:

(Lines 49-50: “Source apportionment is the process of identifying and quantifying **PMF** sources by using information about their chemical composition, and is commonly conducted through receptor models (RMs)...”

11. Line 46: decomposes -> deconvolute.

This change was adopted into the manuscript as follows:

(Lines 51-53): “*Positive Matrix Factorisation model (PMF, Paatero and Tapper, 1994), which **deconvolutes** the input chemical composition into the product of composition and time series matrices.*”

12. Line 50: I will avoid using mathematical terms like $\mathbb{R}^{n \cdot m}$ to accomodate wider audiences, suggesting spelling it out.

This suggestion has been applied throughout the manuscript as follows:

(Lines 57-59): “where **X** is the input matrix, a **$n \cdot m$ matrix of n timepoints and m species**, which is decomposed into **G** and **F**, **matrices of dimensions $n \cdot p$ and $p \cdot m$, respectively**, where p is the number of factors, and **E** is the residuals matrix **of dimensions $n \cdot m$...**”

(Line 117): “... **Z** (**matrix of dimensions $n \cdot m$**) ...”

(Line 118-119): “... **σ** (**positive matrix of dimensions $n \cdot m$**) ...”

(Line 137-138): “... **α** (**positive vector or dimension p**) and **β** (**positive vector or dimension p**), ...”

(Line 141): “... represents positive real numbers.”

(Line 158): “... where **μ** (**matrix of dimensions $n \cdot m$**) represents...”

(Line 161): “... **τ** (**vector of dimension p**) ...”

(Line 170-171): “...Here, **λ** is a model parameter of dimensions **$n \cdot m$**”

13. Line 53: some introduction about unconstrained PMF or constrained PMF is necessary.

A clarification is added between brackets, the reader can amplify this concept through the cited literature.

(Lines 62-63): “In such cases, guiding the model by introducing a priori **knowledge (common practice known as constraining the models)** has been proven beneficial...”

14. Line 55: disentanglement to “identification”

The purpose of using “disentanglement” here is to refer to the extraction of an additional source initially merged to another, which does not have the same meaning as “identification” in this context. Hence, we prefer to replace “disentanglement” with “deconvolution” as suggested by Reviewer 3 as follows:

(Lines 63-64): “been proven beneficial for the source [deconvolution](#) (Lingwall and Christensen et al. 2007, Belis et al. 2014, Dinh et al. 2025).”

and also later in the text:

(Lines 200): “This swap should allow \mathbf{F} to retain the \mathbf{X} matrix mass and could potentially help [deconvoluting](#) profiles ...”

15. Line 57: CMB does not 100% equal to fully constrained PMF. Also, I don't know why you introduce CMB here. Perhaps you will need a few sentences in this paragraph to introduce the limitations of PMF or CMB in general.

The introduction of CMB in this paragraph responds to the need to introduce this model since a “Bayesian CMB” is later described and applied to the toy dataset. However, we agree that the phrasing might be misleading, so we rephrase to:

(Lines 66-67): “A [very strongly-constrained RM](#) is the [Chemical Mass Balance model \(CMB\)](#), which factorises the initial matrices with a totally fixed \mathbf{G} or \mathbf{F} . “

Lines 61-66 already show the limitations of both unconstrained and constrained PMF.

16. Line 70: approach -> conduct

This change was applied to the revised version of the manuscript as follows:

(Line 70): “... a Bayesian RM to [conduct](#) multi-site source apportionment ... “

17. Line 82: overlapping emissions -> mixed emission sources

This change was applied to the revised version of the manuscript as follows:

(Lines 81-82): “... of obtaining chemically distinct and interpretable source profiles from complex [and mixed emission sources](#) ...”

18. Line 83: slight F differences -> slight differences of F

This change was applied to the revised version of the manuscript as follows:

(Lines 96-97): ... “it has been shown in Rusanen et al. (2024) that in BAMF, [slight differences of F](#) can severely compromise...”

19. Line 84-83: use even simpler language to briefly explain sparsity and why it makes sense to enforce it in PMF analyses. Also, change elements to variables since not all the variables of F is element.

If possible, we would like to maintain the formal definition of sparsity as is and later insist on the potential benefits of adopting it in receptor models. We added an extra sentence to the text which highlights the mechanism for which sparsity would improve source decoupling:

(Lines 100-103): ... “could be favourable for this problem. The accomplishment of sparse source fingerprints could represent “cleaner” emission sources **with less mixing among resolved factor profiles, since substituting non-significant contributions in a factor by zeros might allow allocating more importance to the actually relevant contributions of species in factors.** ...”

20. Line 86-87: Change it to: “The accomplishment of sparse source fingerprints could represent “cleaner” emission sources without mixing among resolved factor profiles.”

The suggested change was implemented in the revised version of the manuscript.

(Lines 100-101): “...The accomplishment of sparse source fingerprints could represent “cleaner” emission sources **with less mixing among resolved factor profiles, since ...**”

21. Line 102: What is N? please introduce it in the text

The N mentioned in many equations refers to the normal distribution. A note is made below equation (2) to explicit it.

(Line 122): “...where represents the normal distribution”

22. Line 222: avoiding using hence twice in one sentence

The sentence needed to be rephrased to avoid repetition:

(Line 269): “The factor ordering in the matrices is random in the model results, **hence, the solution factors must be sorted.** ...”

23. Line 256: OK, now I understand what is the toy dataset. Is it more appropriate to use dummy instead of toy? It doesn't make much sense to me when I first saw it in the beginning of the manuscript without context.

The term “*toy dataset*” is widely used in the literature to denote a highly simplified, synthetic dataset intended solely for methodological illustration and validation. An example is the article from Piironen and Vektari (2017) from which the horseshoe prior is adopted. For this reason, we have retained the term to ensure consistency with common practice and reader expectations.

24. Table 3 : For the “toy” dataset, the BAMF or BAMF+HS in general is worse than PMF results, could this be a major flaw of the BAMF? How can this be addressed?

The results in Table 3 show that even if $G R^2$ is slightly higher for PMF than for BAMF or BAMF+HS, all other F, G metrics show better solution representation for BAMF and BAMF+HS. The profiles ρ and R^2 are better for both Bayesian models than those for PMF and the BAMF+HS achieves the highest sparsity and the most similar sparsity compared to the truth. With that, the BAMF+HS presents the most comparable solution to truth in terms of F and G. Further discussion on this comparison can be found on comment #2 from Reviewer 2.

25. Figure 1: it’s a bit confusing for me with the y-axis. They are not real m/z, right? Also, what is the unit of the y axis? Have you done some repeats of CMB or CMB+HS, and the y-axis are the frequency of the iterations end up with these concentrations? It’s not clear from your text and figure captions. Please clarify.

The several plots in Figure 1 represent the distributions of each of the species (rows) of three factors (columns). Hence, the y axis are unitless as m/zs are unitless in source apportionment, since they show normalized profiles. These are not real m/zs but a synthetic mimicking of how a very simplistic ACSM-like dataset would look like. The species for such a dataset have been called m/zs instead of species since the used time series were extracted from modelled OA sources time series in Jiang et al. 2019.

The distributions do not account for repetition of the CMB or CMB+HS runs, but for the Bayesian model sampling. Each sample represents a draw and fit for all parameters (F, G, sigma). Hence, the shown distributions show the variability across the drawn samples, and depicts the variability of m/z contributions to each factor instead of providing a point estimate (as a mean across samples would) with an uncertainty (as the standard deviation across samples would). Similarly, the standard deviation across the samples, estimaThe authors acknowledge this is a paradigm change with respect to PMF, which generates point estimates and consequently hope the additional explanations and figures provided in this manuscript iteration will help clarify this new framework.

26. Section C.1 of SI: there are inconsistencies in BMF or BAMF, BAMF-GS or BAMF+GS.

Some inconsistencies were detected and repaired in the revised version of the manuscript:

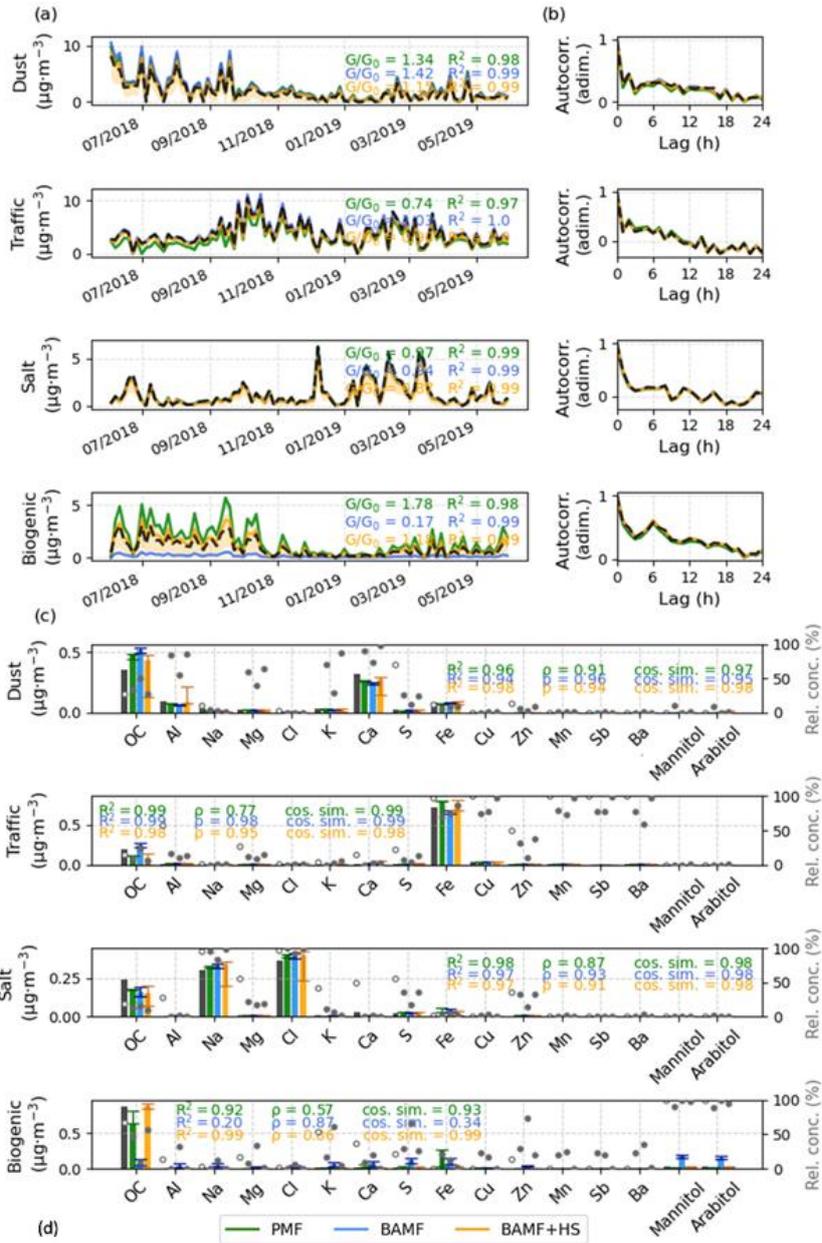
(Lines S68-S69): “Figure S9. Filters real-world filters dataset profiles components distributions for (a) PMF, BAMF, BAMF+HS; (b) BAMF, BAMF-AR1, BAMF-AR1+HS, BAMF-GS.”

(Lines 120-121): “...Figure S7 shows the time series, profiles and F components distributions for BAMF, BAMF-AR1, BAMF-GS, BAMF+HS. ... ”

(Lines 136-137): “...Regarding the non-horseshoed models, the sum of the correlation with truth G for all factors was very similar for BAMF, BAMF-AR1, BAMF-GS (3.93, 3.92, 3.93, respectively) but the Spearman ... ”

27. Figure 3: You will need a legend for which color is which...

The legend for that Figure was missing and the newer version of the Figure 3 (below) does contain a legend.



Factor	Model	F contr. R^2	F contr. ρ	F Gini	F Gini ratio (*)	Truth zeros sum
Dust	PMF	0.95	0.84	0.79	0.90	0.0035
	BAMF	0.77	0.90	0.80	1.08	0.0025
	BAMF+HS	0.94	0.90	0.79	1.07	0.0011
Traffic	PMF	0.99	0.90	0.89	1.19	0.0000
	BAMF	0.93	0.94	0.85	0.99	0.0005
	BAMF+HS	0.995	0.93	0.88	1.02	0.0004
Salt	PMF	0.95	0.84	0.79	0.90	0.0012
	BAMF	0.95	0.88	0.80	1.01	0.0118
	BAMF+HS	0.98	0.87	0.82	1.04	0.0031
Biogenic	PMF	0.87	0.43	0.85	1.08	0.0035
	BAMF	0.53	0.51	0.50	0.56	0.0118
	BAMF+HS	0.96	0.84	0.90	1.01	0.0007
Σ_k (*)	PMF	3.80	3.04	3.34	0.43	0.0198
	BAMF	3.18	3.23	2.95	0.53	0.0172
	BAMF+HS	3.88	3.54	3.39	0.15	0.0031

Figure 3. Synthetic offline dataset source apportionment results for PMF, BAMF, and BAMF+HS models. (a) Time Series. (b) Autocorrelation. (c) Profiles. (d) Table with additional metrics for comparison to truth. Bold numbers reflect the highest value amongst models. F contr. represents here the percentage of each factor into a given species. The sum row reflects the overall performance of the model for all sources for each statistic metric except for the ones marked with (*), in which the difference to 1 in absolute value is summed up.

28. Figure 8: I'm still confused about what you are showing here. Is it the autocorrelation of each model of each source? Is it the model vs truth for each source for each model? Or is it the correlation of the autocorrelation between model vs truth? If it's the third one, what does it mean?

Figure 8 shows the autocorrelation from truth and the model outputs correlate (Pearson coefficient of determination) for each model and source, which is what the reviewer described as the third possibility. The intention of this plot is to compare the autocorrelation and how well the solutions autocorrelation mimics the original one. The explanation of this figure has been rewritten for further clarity as:

(Lines 565-566): Figure 8 shows the autocorrelation from truth and the model outputs correlate (Pearson coefficient of determination) for each model and source in the 4 cities.

PMF shows a much autocorrelation solution-to-truth R compared to all other models, highlighting the power of the autocorrelation prior applied to all these Bayesian models. Regarding models, the comparison shows:

- The use of the horseshoe sparsity prior (BAMF+HS model) does not deprecate the autocorrelation quality in comparison to the model without the sparsity prior (BAMF).
- The two autocorrelation formulations (BAMF, BAMF-AR1) employed in the manuscript show little performance differences, and if any, BAMF seems to reproduce better the autocorrelation of the truth.
- BAMF-GS shows a slightly better autocorrelation reproduction with respect to truth, but the differences with BAMF are scarce.

References

Jiang, J., Aksoyoglu, S., El-Haddad, I., Ciarelli, G., Denier van der Gon, H. A., Canonaco, F., ... & Prévôt, A. S. (2019). Sources of organic aerosols in Europe: A modelling study using CAMx with modified volatility basis set scheme. *Atmospheric Chemistry and Physics Discussions*, 2019, 1-35.

Reply to Reviewer 2

We thank Reviewer 2 for the thorough review of our manuscript and their thoughtful comments focusing on improving the scientific quality of the revised manuscript and its readability to reach a broader audience.

The reviewer addresses the review by separating their comments in general and specific comments. The reply will be structured in the same manner and replying point-by-point. We present the plain text in the interactive discussion and the formatted reply as an attachment. Firstly, the reviewer comments will be copied in **bold, blue color** and will be followed by the replies of the authors, for the sake of the answer legibility. All the manuscript editions from the Reviewer 2 suggestions are incorporated into the attached newer version of the manuscript (marked **in blue** in the text), providing in the revised document their exact location in terms of manuscript lines, and copied after the answer in grey color when possible.

Major revisions

1. **Line 22: A question that arises is why not remove non-essential species directly from PMF? Since the authors later discuss that sparsity could also be introduced into PMF, it would be helpful to explain why this was not tested, or discuss how that would be expected to compare. This is more of a suggestion than a requirement.**

As mentioned in the text and pointed out by the reviewer, the sparsity property can also be used in the PMF framework to force sparse \mathbf{F} , \mathbf{G} solutions. This is enforced through pulling equations or constraints (Paatero et al. 2003). However, these approaches are targeted to shrink down certain \mathbf{G} , \mathbf{F} matrix components specified beforehand by the user, which are assessed to be zero upon *a priori* knowledge. The approach in BAMF+HS is substantially different, since it introduces an untargeted sparsifying mechanism without user intervention. Hence, the BAMF+HS approach results are more objective than sparsity-enforced PMF, although BAMF+HS can also make use of hard-coded constraints as shown in Rusanen et al. (2024) in case the *a priori* knowledge enforcing is desired. Nevertheless, to the authors' knowledge, there is not any mechanism to introduce untargeted sparsity in PMF.

Because of this paradigm shift, a direct comparison between sparse PMF and BAMF+HS is not straightforward, since sparse PMF requires the user to explicitly define which and how many elements of the \mathbf{F} and \mathbf{G} matrices should be driven toward zero. In contrast, BAMF+HS introduces sparsity in a data-driven and untargeted manner, allowing the model to automatically identify and suppress unnecessary components without requiring prior specification of their location or number. As a result, BAMF+HS shifts the burden of sparsity selection from the user to the inference framework itself, leading to solutions that are less subjective and more reproducible across different applications. This discussion is explored in the text in the discussion section, in lines 637-644 of the revised text.

2. **Further justification is needed for why PMF outperforms BAMF and BAMF+HS in the toy (simple) dataset.**

This concern is also raised by Reviewer 1 in comment #24. The authors have expanded the comparison for the reviewers to acknowledge better the differences between models.

By looking at Table 3(a), one can see that PMF has a better correlation \mathbf{X} vs. \mathbf{Z} , while the median weighted error ($|\mathbf{Z}-\mathbf{X}|/\sigma$) is lower for BAMF and BAMF+HS than for PMF, and the maximum weighted error is lower for PMF. This might entail that even if the median weighted errors are, in median, better characterized by BAMF, BAMF+HS, these might generate some outliers which PMF does not generate. Hence, one can't say that BAMF, or BAMF+HS are straight-forwardly worse.

Table 3(b) further illustrates the differences among the models in terms of the accuracy of both the \mathbf{F} and \mathbf{G} matrices. For the \mathbf{G} matrix, PMF shows the largest deviation from the ideal value $G/G_0=1$, accumulating the greatest error (2.64), followed by BAMF (2.10), while BAMF+HS exhibits the smallest deviation (0.81), indicating the highest overall accuracy. In terms of the squared Pearson correlation, the models yield broadly comparable results, with the exception of Factor 3, for which PMF achieves a higher value (0.74) than BAMF (0.50) and BAMF+HS (0.48).

Regarding the \mathbf{F} matrix, both BAMF and BAMF+HS outperform PMF in terms of ρ and R^2 . Notably, BAMF+HS exhibits substantially higher sparsity than either PMF or BAMF, with across-factors summed sparsity values of 1.29, 0.98, and 0.96, respectively. Consistently, the Gini ratios of the inferred solutions relative to the truth are markedly closer to unity for BAMF+HS (range 0.40–0.93) than for PMF (0.45–0.64). Furthermore, the total contribution assigned to components that are zero in the truth is globally lower for BAMF+HS than for the other models. Taken together, these results indicate that BAMF+HS not only promotes sparsity, but does so in a chemically consistent manner, leading to a more accurate mass apportionment across factors, despite a slightly reduced time-series correlation for the third factor.

The revised manuscript has been modified to highlight the BAMF+HS benefits:

(Lines 426-446): "... In the next evaluation step, we test the various models assuming no prior knowledge. Figure 2 shows the results of PMF, BAMF, and BAMF+HS models on the toy dataset and Table 3 shows their factorisation **performance** and comparison to truth **metrics**. In terms of factorisation, median relative errors are better for BAMF+HS and BAMF than for PMF, but their maximum errors are higher and the Pearson coefficients slightly lower, all this entailing comparable factorisation performances. All models generally adapt well to the truth features, but they present non-negligible differences. PMF results better resemble the truth in terms of $\mathbf{G} R^2$, but it is the model whose G/G_0 differs from 1 the most, accumulating the greatest error (2.64), followed by BAMF (2.10), while BAMF+HS exhibits the smallest deviation (0.81), indicating the highest overall accuracy. In terms of profiles, the BAMF+HS model is the closest to the truth both in terms of ρ and R^2 , especially for the second and third factors for which the sparsity introduction results are advantageous with respect to BAMF results. Consistently, the Gini ratios of the inferred solutions relative to the truth are markedly closer to unity for BAMF+HS (range 0.40–0.93) than for PMF (0.45–0.64). The sparsity effects can also be seen in Figure S5, in

which the horseshoe shrinkage is evident for the low m/z s allowing in turn the larger m/z s to retain more mass, hence resembling better the truth profiles. Taken together, these results indicate that BAMF+HS not only promotes sparsity, but does so in a chemically consistent manner, leading to a more accurate mass apportionment across factors, despite a slightly reduced time-series correlation for the third factor. However, the BAMF+HS could not shrink down the lowest signals in Factor 1, likely because their contribution estimated by the mass balance and the autocorrelation restrictions of this model made it unclear for the horseshoe to shrink them down completely. With this result, this toy dataset depicts the capacities and limitations of the horseshoe implementation on BAMF: it is capable to sparsify effectively only the signals which are close enough to zero as given by the restrictions of the BAMF model. ...”

3. **Line 103-105: Please emphasize that in the case of PMF it's the opposite; normalization takes place after resolving the factors. Also in Line 216: \mathbf{X} and σ are not normalized in PMF. Please clarify how these differences might affect the comparison and whether they could bias the results.**

The following sentence has been added to the text to make this distinction more apparent:

(Lines 128-132): “... It is worth noting that PMF applies the normalisation of profiles after a \mathbf{F} , \mathbf{G} solution is found, not as a model prior as done in BAMF. The PMF generates mass-loaded \mathbf{F} , \mathbf{G} solution matrices, which are reweighted to provide a normalised \mathbf{F} and a mass-loaded \mathbf{G} . In the Bayesian models used in this study, the normalisation of \mathbf{F} is inherent to the model by design. The different formulations eventually provide normalised \mathbf{F} and mass-weighted \mathbf{G} , with unlikely affectations due to the normalisation procedure. ...”.

Similarly, we clarify that in PMF neither the data matrix \mathbf{X} nor the uncertainty matrix σ are normalized, the Bayesian models normalise these matrices before running the factorisation algorithm. Since the input quantities are equal but only divided by a certain factor $f_{norm} = \sum_{i,j} X_{ij} / (n \cdot m)$, the authors do not expect the scaling to imply substantial differences in the factorised matrices, since the normalisation does not alter the underlying chemical species or time series relative weights, only its total mass accountancy.

4. **Lines 212-213: Could this create bias when comparing PMF and BAMF?**

The authors assume that the reviewer refer to the following remark done in the lines 211-214 of the submitted version of the manuscript:

(Lines 211-214): “The number of runs may seem compromising the PMF quality in comparison to the 6000 samples per chain used in Bayesian models. However, this comparison is misleading, since the factorisation space is indeed better explored by PMF, with 100 different sampling seeds, while only 4 seeds (chains) were used in BAMF-like models as usual procedure in Bayesian modelling for the sake of computational resources.”

It is a common practice in Bayesian modelling (*) to use around 1000 random samples and 4 chains (4 random seeds) for each modelling experiment in order to provide both intra- and inter- chain variability to assess both the drift and stability of the model (intra-chain variability) and the convergence of the model despite the solution space starting point (inter-chain variability). Due to the high complexity of BAMF+HS and datasets employed, the number of samples was enlarged to secure solution stability, which is also a common practice in the field. On top of that, the BAMF+HS model is not directly initialised by the random seeds, but by random-seed initialised maximum-a-priori estimations of the parameters, i.e. with values which are numerically pre-assessed to be viable.

This refined initialisation step helps stabilise the sampling and leads to more sensible solutions. Differently, in PMF, multiple runs with different random seeds are required to explore the factorisation space and mitigate the effect of local minima.

This intrinsic methodological difference complicates a rigorous comparison between the two approaches on the same terms. The BAMF+HS, due to the model complexity, requires the initialisation of the samples from plausible parameter values which are, in turn, calculated from the different seeds. Hence, BAMF+HS intrinsically promotes a more guided solution compared to PMF which performs a wider exploration of the space solution to then average the different explored runs. Consequently, while PMF may explore a broader range of potential factorisations, BAMF+HS tends to concentrate the posterior around plausible regions, providing stability and convergence assurances that are inherently different from the averaging strategy of PMF. This difference should be interpreted as a methodological distinction rather than a bias, highlighting that the comparison reflects the fundamental properties of the respective frameworks rather than a direct equivalence of sampling strategies.

(*) The common number of samples and chains has been mentioned in the revised version of the text as follows: *(Lines 239-241): "... The number of chains is consistent with standard practice in Bayesian modeling, and the number of samples was increased beyond commonly adopted values (e.g., 1000) in order to improve solution stability. ..."*

5. For the toy dataset it's not clear to me if it's intended to be ACSM-like. Please describe this further.

The toy dataset was constructed as a deliberately simplified test case intended to evaluate the basic characteristics of the models in a didactic manner, rather than to reproduce any realistic atmospheric scenario. Although it was based on ACSM-like time series and therefore reflects some of the temporal properties of such measurements, these three included sources do not represent combinations that would be expected in a real-world environment and were used solely for methodological testing purposes. In addition, the source profiles were intentionally designed to be highly simplified in order to facilitate an immediate visual assessment of the model fitting. For these reasons, the extracted components were not assigned environmental labels, but were instead referred to generically as Factor 1, Factor 2, and Factor 3.

To clarify the toy dataset concept for all readers, the text has been modified in the revised version of the dataset as follows:

(Lines 307-316): "... A simplistic synthetic toy dataset was designed as a deliberately simplified test case to perform basic control and performance tests, rather than to reproduce any realistic atmospheric scenario. It was devised by creating three very simple and sparse profiles and using three time series (HOA, SOAbio, BBOA) from modelled source time series of the city of Zurich (Rusanen et al. 2024, time resolution of 1h) in order to test how sparsity priors act on very uneven species contribution. Although it is based on ACSM-like time series and therefore reflects some of the temporal properties of such measurements, these three included sources do not represent combinations that would be expected in a real-world environment and were used solely for methodological testing purposes. In addition, the source profiles were intentionally designed to be highly simplified in order to facilitate an immediate visual assessment of the model fitting. For these reasons, the

extracted components were not assigned environmental labels, but were instead referred to generically as Factor 1, Factor 2, and Factor 3. ...)

6. **Please include somewhere in the manuscript the temporal resolution of any synthetic/real dataset and also what was the computation burden of each model. Would you suggest BAMF as a complementary to PMF method, or as a replacement to the traditional PMF? Maybe include a qualitative comparison including also the ease of applying both and how easily each model is to tune etc.**

The temporal resolution of each of each of the datasets is introduced now in the preamble of the 2.4 section as follows:

(Lines 306-309): The time resolution of modelled OA sources, used both in the chemically-sparse toy dataset and the chemically less sparse datasets is 1hour. The time resolution of offline datasets, used in the chemically sparse synthetic offline dataset and the chemically-sparse real-world offline dataset is 1 day. “

The computational time (with the computational specifications under which it is run) is shown at Table S1 with the HMC sampling parameters in the newer version of the manuscript:

Table S1. Hamiltonian MonteCarlo sampling parameters for all the conducted experiments

Experiment	Total number of samples	Number of warm-up samples	Number of chains	Time *
Toy dataset	4000	2000	4	~1h
European synthetic datasets	12000	6000	4	~7h
Filters synthetic dataset	6000	3000	4	~3h
Filters real-world dataset	6000	3000	4	~1.5h

* The time here shows the BAMF+HS running time plus the post-processing time (for sorting, averaging, etc.), the latter having a nearly neglectable contribution. Model runtimes were measured as wall-clock time on a high-performance computing cluster. Jobs were executed on compute nodes equipped with dual-socket AMD EPYC 7502 processors (64 physical cores, 2.5 GHz) and approximately 256 GB of RAM, running Linux. All BAMF+HS runs used four parallel chains using approximately 4 physical CPU cores (8 logical threads).

The BAMF+HS model aims to be an accurate, reliable alternative to PMF to conduct source apportionment on datasets whose sources are expected to be sparse and autocorrelated. For such cases, BAMF+HS has proved to better reproduce the truth and generate sparse profiles. However, its development status is still limited and it is not currently run in an user-friendly environment, indeed, it needs large computational resources and computational time exceeds substantially that of PMF. The use of BAMF+HS, currently, might be limited to experienced users and small datasets. Nevertheless, current efforts are being directed to decrease the amount of computational resources to use these models and a more easy-to-use platform could be developed in a short future.

7. **Line 348: How was the “truth performance” obtained? If derived through PMF, please discuss the appropriateness of comparing.**

The sentence is slightly confusing, Table 3 shows the factorisation performance and the comparison to truth scores, and not the “truth performance”. The authors acknowledge this readability issue and have rephrased this sentence in the revised version of the manuscript as follows:

Lines 426-427: “... Figure 2 shows the results of PMF, BAMF, and BAMF+HS models on the toy dataset and Table 3 shows their factorisation **performance** and comparison to truth **metrics**. ...”

8. **Lastly, Fig. 1 can be quite confusing for the reader. I would expand the explanation in the text and also in the caption. Describe the posterior density distributions and why the truth is always at zero in the y-axis.**

The confusion brought by Figure 1 was also remarked by reviewers 1 and 3. The several plots in Figure 1 represent the distributions of each of the species (rows) of three factors (columns). Hence, the y axis are unitless as m/z s are in source apportionment since they show normalized profiles. The truth is presented as a marker always in the zero of the y axis since it the y-axis represents the frequency of the modelled F components distributions across samples, and the truth is not sampled, hence cannot have such a distribution, but a fixed value. The authors find it very necessary to display the distributions of both CMB and CMB+HS models to highlight the sharpness of the distributions of the CMB+HS compared to those from CMB.

The revised version of the manuscript includes one more sentence to help with the understanding of the figure and some clarifications in the text.

Lines 411-418: “... Figure 1 shows the distribution of each **F** component for CMB with and without the horseshoe prior (CMB, CMB+HS, respectively, Table 1). **The distributions shown account for all the variability across samples of each F component for both models, and the truth is shown as a marker in the x-axis since it is a point value to be compared to the centers of the distributions. The presentation of the CMB and CMB+HS distributions aims to demonstrate the sparsity-inducing role of the horseshoe prior, which enforces shrinkage of the F component toward zero; this effect is more readily discernible when a strongly guided G matrix is used to isolate the evidence of sparsity.** Figure 1 showcases the horseshoe prior power to generate sparsity in **F** components, shrinking **more strongly** the lowest signals to zero than **CMB** and, as a consequence, enlarging the most prominent signals. ...”

Minor revisions

9. **Line 23 Improve the separation compared to what? Be more specific in such discussions (also in Line 514).**

The revised text has been modified to specify that the improvement has been compared to the BAMF solution through sparsity improvement.

Lines 23-26: “...This study introduces sparsity into the Bayesian Autocorrelated Matrix Factorisation (BAMF) model with the aim of removing non-essential species contribution in the unconstrained profiles, which is expected to improve the separation of factors **compared to BAMF...**”

Lines 619-620: “...The regularised horseshoe prior introduction in BAMF improved apportionment of offline synthetic and real-world datasets **with respect to BAMF**, promoting sparser profiles. ...”

10. Line 43 You mean "PM", not "PMF".

The revised version of the text has implemented this change:

Line 44: “... highlights the importance of developing mitigation plans grounded in detailed knowledge of **PM** sources composition and...”

11. Line 53 Add though; “even though it can lead...”

The revised version of the text has implemented this change:

Line 61 “...Unconstrained PMF, **although** it can lead to robust results, is usually insufficient...”

12. Line 55 Maybe end the sentence after the references.

The revised version of the text has separated that sentence into two for the sake of readability as follows:

Lines 62-65: “... In such cases, guiding the model by introducing a priori knowledge (common practice known as constraining the model) has been proven beneficial for the source disentanglement (Lingwall and Christensen et al. 2007, Belis et al. 2014, Dinh et al. 2025). However, it can still introduce substantial bias in the solution (Viana et al. 2022). ...”

13. Line 57-58 What do you mean “cover the whole range of prior knowledge required” ?

The authors tried to convey that receptor models can range from total to non-existing amounts of knowledge required about pollution sources prior to receptor modelling as shown in Figure 1 in Viana et al. (2009a). The revised version of the manuscript is rephrased to present it more clearly.

Lines 65-67: Globally, the RMs cover the **whole range of pollution sources knowledge required prior to receptor modelling** (Viana et al. 2008, Belis et al. 2013).

14. Line 61 Add "was introduced by Park..."t

The revised version of the text has implemented this change:

Lines 70-71: “... The first application of Bayesian models in atmospheric source apportionment **was introduced** in Park et al. (2001, 2002) ...”

15. Line 79 BMF; you have already introduced the acronym (also in Line 96 and 108).

The authors make a distinction between Bayesian Matrix Factorisation models, the collection of receptor models which use the Bayesian inference to resolve matrix factorisation (e.g. BAMF, BAMF+HS, BAMF-AR1, etc.), and the Bayesian Matrix Factorisation model (BMF) which represents the analogous model to BAMF but without the autocorrelation prior (as outlined in Table 1). In order to amplify this distinction and be coherent throughout the

text, the use of the acronym “BMF” will be used only for the particular model without the autocorrelation prior, whilst the ensemble of Bayesian models used for matrix factorisation will be referred without the acronym. The use (or lack of) this notation has been applied in the mentioned lines and elsewhere of the revised version of the manuscript:

Line 81: “... However, studies using the Bayesian Matrix Factorisation (~~BMF~~) framework are still scarce. ...“

16. **Line 96 By (a): do you mean Equation (1)?**

Yes, the revised version of the text has corrected this mistake:

Line 113: “... Bayesian Matrix factorisation models, like other RMs, are based on the chemical mass balance equation (Eq. 1). ...“

17. **Line 100 Similarly, replace (2) with Eq. (2) if you mean equation.**

The revised version of the text has implemented this change:

Lines 117-118: “... a standard deviation given by the positively-defined uncertainty matrix (Eq. 2). ...“

18. **Line 103 Delete “meaning”.**

The revised version of the text has implemented this change:

Lines 122-123: “... Whilst \mathbf{G} is not given any prior and is sampled then by default from a uniform distribution, ...”

19. **Line 107 Delete "hereinafter" or "here".**

The revised version of the text has implemented this change:

Lines 126-128: “...Usual notation for indices used hereinafter are i, j, k for elements in the range $(1, \dots, n)$, $(1, \dots, m)$ and $(1, \dots, p)$, for the timestamps, species, and factors, respectively. ...”

20. **Line 116 "Where" has different font. Also in Line 204.**

The revised version of the text has implemented these changes:

Line 141 “...where $i \in (2, \dots, n-1)$, C represents the...”

Lines 248-251: This coefficient compares the variance within chains and between chains of the Z matrix, hence if chains converge, $R \approx 1$, values of $R > 1$ imply chain divergence and values of $R < 1$ imply sampling divergence in chains...”

21. **Line 151 Maybe replace "unevenness" with "inequality".**

The revised version of the text has implemented this change:

Lines 181-812: “...Since it quantifies the inequality, it can be a proxy for sparsity;...”

22. **Line 162 Add () in γ .**

The revised version of the text has implemented this change:

Line 194: “... source-dependent slopes (α) and intercept (β), and width (γ). ...”

23. **Line 208 Change the Equation number to 14 and also the next ones.**

The revised version of the text has implemented these changes:

Line 254:
$$Q = \sum_{j=1}^m \sum_{i=1}^n \frac{e_{ij}^2}{\sigma_{ij}^2}$$

(14)

Line 265:
$$X^* = X / f_{norm} , \sigma^* = \sigma / f_{norm} \quad \text{where } f_{norm} = \sum_{i,j} X_{ij} / (n \cdot m)$$

(15)

24. **Line 210 What do you mean "sorted PMF runs"? Probably regular runs and you sort afterwards. Please rephrase.**

The use of that word represented a mistake and revised version of the text has implemented these changes:

Line 256: "... with 100 runs which are posteriorly sorted as in BAMF... .

25. **Line 216 Rephrase "Previously to model running,..."**.

The sentence has been rephrased to:

Line 247: "Before model running, ...".

26. **Line 222 This sentence is confusing, please rephrase.**

The sentence has been rephrased, since the Reviewer 1 also conveyed its bad readability:

Line 269: "...The factor ordering in the matrices is random in the model results, hence, the solution factors must be sorted. ...".

27. **Line 227, 486, 510, 542, 543 Use bold for F and/or G.**

The use of a regular type of font in line 227 (line 258 in the revised version of the manuscript) is delivered since it is referring to certain matrix components, F_k , G_k . However, it has been changed in the other cases in the revised version of the manuscript:

Line 587-588: "... The horseshoe prior adds more complexity to the **F** with ..."

Lines 614-615: "... the horseshoe prior in BAMF+HS effectively suppresses weak signals of **F** contributions as determined..."

Line 655: "... The BAMF-GS seemed to capture slightly better the **G** variability in comparison to ..."

Line 657: "... does not support enforcing sparsity in **F**, thereby..."

28. **Line 232 Use "was" instead of "is".**

The revised version of the text has implemented this change:

Line 280 "... The last step of the experimental process was to assess the model performance ..."

29. **Line 235 Use "was" instead of "will be".**

The revised version of the text has implemented this change:

Lines 283-284: "... The reconstruction performance was assessed by checking the cell-wise correlation..."

30. **Line 236 Use bold in Z, X.**

The revised version of the text has implemented this change:

Lines 285: "... value of relative deviations of **Z** and **X** with respect to..."

31. **Line 251 Please fix this sentence: Despite the truth's factorisation is unknown....**

The sentence has been rephrased to:

Line 303: "... **Although the truth** factorisation is unknown and the results cannot be directly validated, ...".

32. **Line 255 Please change 2.3.1 to 2.4.1 and apply the same to the following subsections in 2.4.**

The revised version of the text has implemented these changes.

Line 310: 2.4.1 "Chemically-sparse toy dataset"

Line 327: 2.4.2 "Chemically-sparse synthetic offline dataset"

Line 356: 2.4.3 "Chemically sparse real-world offline dataset"

Line 370: 2.4.4 "Chemically less sparse synthetic online ACSM datasets"

33. **Line 255 Maybe add "synthetic" here too, so it's easier for the reader.**

The revised version of the text has implemented this change.

Line 375: 3.1 "Chemically sparse **synthetic** toy dataset".

34. **Line 303 Please add a comma after "models".**

The revised version of the text has implemented this change.

Line 342: With the aim of recreating more complex real-world datasets to test the models, we generated 6 datasets ...".

35. **Line 304 Replace "is" with "was".**

The revised version of the text has implemented this change.

Line 343: "... The objective **was** to recreate OA matrices as given by..."

36. **Line 321 Maybe transfer "slightly" before "modifying".**

The revised version of the text has implemented this change.

Line 391: "... Lastly, a sensitivity analysis was carried out by **slightly** modifying the original **F**, **G** matrices upon which..."

37. **Lines 322-323 Why did you choose to perturbate only one factor? It would be helpful to see the sensitivity analysis on all factors of one dataset (e.g. Zurich).**

The aim of perturbing only one factor for this analysis is to understand how model outputs change upon minimal, isolated input perturbances. If the perturbances were to be applied to all factors, the differences across solutions would be more difficult to acknowledge. When one factor only is modified, the resulting solution differences can be directly attributed to the perturbation of that factor. If all factors were perturbed, the final **F**, **G** matrices would

be substantially distinct from the base case since the perturbations in the initial matrices might cause severe factors remixing even if the perturbations are small. Hence, the F, G metrics would drop suddenly even for slight perturbations, which would make it difficult to judge the worsening of the solution as a function of the magnitude of the perturbation. Hence, the authors believe the single-factor perturbation approach is more informative on the model processes.

38. **Line 328 Replace “with this framework” with “within this framework”.**

The revised version of the text has implemented this change.

Line 399: “... Consequently, **within** this framework,...”

39. **Line 329 Replace “tried” with “implemented”.**

The authors assume the reviewer was referring to line 366 in the previous version of the manuscript (line 414 in the revised version). This change was implemented in the revised version of the manuscript.

Line 451: “... Additionally, different autocorrelation formulations were **implemented** with and without the horseshoe prior ...”

40. **Line 370 Replace “to prove this further.” With “to further highlight this.”.**

The revised version of the text has implemented this change.

Lines 454-455: “... although these models are also tried on the other datasets to **further highlight this.** ...”

41. **Line 372 Replace “is intended” with “was used”.**

The revised version of the text has implemented this change.

Lines 457: “... This synthetic offline dataset **was used** to assess the performance. ...”

42. **Line 378 It is not very clear what the initialization failure means.**

The initialisation failure in Stan occurs when no set of initial parameter values satisfying the model constraints and result in a finite log posterior can be found prior to sampling. This can be solved by introducing more constrained priors into the model for it to launch more sensible parameter values. An additional sentence is added to the text to properly explain this issue:

Lines 464-467: “... **Model initialisation fails when no set of initial parameter values satisfying the model result in valid Bayesian solutions, and are usually solved by imposing more informative priors constraints on the model parameters.** ...”

43. **Line 404 You have used however three times in a paragraph. Maybe remove this one.**

The revised version of the text has implemented this change.

Lines 495-496: “... We found it valuable to present different model performances on different datasets, ...”

44. **Line 414 Repeated “offline”.**

The revised version of the text has implemented this change.

Line 505: “... in the real-world offline PM₁₀-PM_{2.5} dataset described in Section 2.4.3. ...”

45. **Line 426 Maybe replace “incapacitating“ with “preventing”.**

The revised version of the text has implemented this change.

Lines 507: “... BAMF and BAMF-AR1 models presented initialisation issues **preventing** them **from** properly **launching** the models. ...”

46. **Line 417 “we made them more robust”: this needs rephrasing.**

The sentence was rephrased as follows:

Lines 471-473: “... **To avoid this issue and make the model more robust, we implemented a prior in F so that its components are drawn from** Gaussian distributions centered at zero and with a standard deviation of 1 so that we restrict values to be bounded to 1. ...”

47. **Line 436 Replace with: “The next step was to test these models on more realistic synthetic datasets”.**

The revised version of the text has implemented this change.

Lines 530: “... The next step **was to test these models** on more realistic synthetic datasets. ...”

48. **Line 444 This sentence is a bit confusing.**

The sentences involved were rephrased as follows:

Lines 540-541: “... In this case, the **(not-squared)** Pearson correlation coefficient was used to compare the results of the ACSM-like datasets more easily to those presented in Rusanen et al. (2024), **which used this metric**. Figure 7 shows a good agreement between models and the truth, with **most solutions** with correlations with truth for **F** and **G** above 0.7, similarly to Rusanen et al. (2024). ...”

49. **Line 484-485 Repeating “solution”.**

The revised version of the text has implemented this change.

Lines 586-587: “... However, this plot reflects the deprecation of the solution with models when the horseshoe prior is applied ...”

50. **Line 526-528 This is a large sentence, please try to modify.**

The sentence was split in two for an easier readability.

Lines 634-636: “... The higher chain divergence found for the *horseshoed* models **causes a drop in solution** precision due to different landings on the solution space depending on the chain. **This issue could be** reduced by selecting chains a-posteriori upon user-defined criterion as is practiced in PMF. ...”

51. **Line 529 Maybe better to say: “suffers from in contrast to BAMF”.**

The sentence was rephrased as follows:

Lines 637-638: “... This is further confirmed by the insensitivity to **G** or **F** perturbations that **are visible for BAMF+HS but not for BAMF**. ...”

52. **Line 530 Remove “hence”.**

The revised version of the text has implemented this change.

Lines 638-639: “... Nonetheless, given that ACSM-like factor profiles exhibit low sparsity in the literature, the use of sparsity priors in these datasets is less justified. ...”

53. **Line 539 Replace “The other tried out models” with “The results from the other models tested”.**

The revised version of the text has implemented this change.

Lines 612: “... **The results of the other models tested** (BAMF-AR1, BAMF-GS) did not show ...”

54. **Line 547 What do you mean by essentialise?.**

The meaning of “essentialise” here refers to reducing the profiles contributions to only relevant contributions, and hence, keeping the essential information only. However, in the revised version of the text, it has been replaced by “condense”!, with similar meaning..

Lines 652: “... which intends to **condense** source apportionment profiles removing noisy signals.”

55. **Line 555 Maybe replace “essayed” with “tried” or “tested”.**

The revised version of the text has implemented this change.

Lines 670-671: “... In the Bayesian framework, we **tested** a different formulation of the autocorrelation term ”

References

Paatero, P., Hopke, P. K., Song, X. H., & Ramadan, Z. (2002). Understanding and controlling rotations in factor analytic models. *Chemometrics and intelligent laboratory systems*, 60(1-2), 253-264.

Rusanen, A., Björklund, A., Manousakas, M., Jiang, J., Kulmala, M. T., Puolamäki, K., & Daellenbach, K. R. (2023). A novel probabilistic source apportionment approach: Bayesian auto-correlated matrix factorization. *Atmospheric Measurement Techniques Discussions*, 2023, 1-28.

Viana, M., Kuhlbusch, T. A., Querol, X., Alastuey, A., Harrison, R. M., Hopke, P. K., ... & Hitzenberger, R. (2008). Source apportionment of particulate matter in Europe: a review of methods and results. *Journal of aerosol science*, 39(10), 827-849.

Reply to Reviewer 3

We thank Reviewer 3 for the thorough review of our manuscript and their thoughtful comments focusing on improving the scientific quality of the revised manuscript and its readability to reach a broader audience.

The reviewer addresses the review by separating their comments in general and specific comments. The reply will be structured in the same manner and replying point-by-point. We present the plain text in the interactive discussion and the formatted reply as an attachment. Firstly, the reviewer comments will be copied in **bold, blue color** and will be followed by the replies of the authors, for the sake of the answer legibility. All the manuscript editions from the Reviewer 3 suggestions are incorporated into the attached newer version of the manuscript (marked **in blue** in the text), providing in the revised document their exact location in terms of manuscript lines, and copied after the answer in grey color when possible.

General comments

1. **Although the paper is technically rigorous, the presentation is often dense, with extensive mathematical detail that can obscure the broader modelling logic. Some sections would benefit from clearer explanatory text and more consistent definition of notation to help guide the reader through the methodology. Adding brief intuitive descriptions alongside the formal equations would improve accessibility and make the work more approachable. In particular, a schematic workflow diagram for BAMF+HS would help clarify where the sparsity prior operates relative to temporal autocorrelation and the overall factorisation process.**

Similar concerns were also raised by reviewers 1 and 2 and some clarification sentences were added to make the technical details more accessible. The notation employed has also been more detailed. Some examples of these changes in the revised version of the manuscript are:

(Lines 57-59): "... where \mathbf{X} is the input matrix, a $n \cdot m$ matrix of n timepoints and m species, which is decomposed into \mathbf{G} and \mathbf{F} , matrices of dimensions $n \cdot p$ and $p \cdot m$, respectively, where p is the number of factors, and \mathbf{E} is the residuals matrix of dimensions $n \cdot m$"

(Lines 100-103): "... The accomplishment of sparse source fingerprints could represent "cleaner" emission sources with less mixing among resolved factor profiles, since substituting non-significant contributions in a factor by zeros might allow allocating more importance to the actually relevant contributions of species in factors. ..."

(Line 122): "... where N represents the normal distribution. ..."

(Lines 151-156): "... The introduction of sparsity in BAMF entails the addition of several hyperpriors in the \mathbf{F} prior to implement the shrinkage mechanism. In this study, we used the regularised horseshoe prior (Pironen and Vehtari, 2017), which is a global-local complex of hyperpriors, i.e. the shrinkage power is both regulated globally source-wise and \mathbf{F} -component-wise. The idea behind this prior is that species with very small contributions to a factor are shrunk toward zero through an automatic shrinkage mechanism, whereas species with substantial support from the data are largely unaffected. ..."

(Lines 222-229): “... In the Hamiltonian analogy, the evolution of these parameters across samples is computed as the trajectory of a fictitious particle. This particle moves through the parameter space driven by random momentum in all directions. **This approach avoids the random-walk behavior of simpler sampling methods and enables faster convergence.** The trajectory is hence simulated using a discretized approximation, and candidate positions are accepted or rejected according to the Metropolis criterion (Metropolis et al. 1953). **Accepted positions correspond to plausible parameter values given both the model assumptions and the data. ...**”

Plus, as also suggested by Reviewer 1, a workflow diagram was added in Figure S1(b) to also convey the process visually.

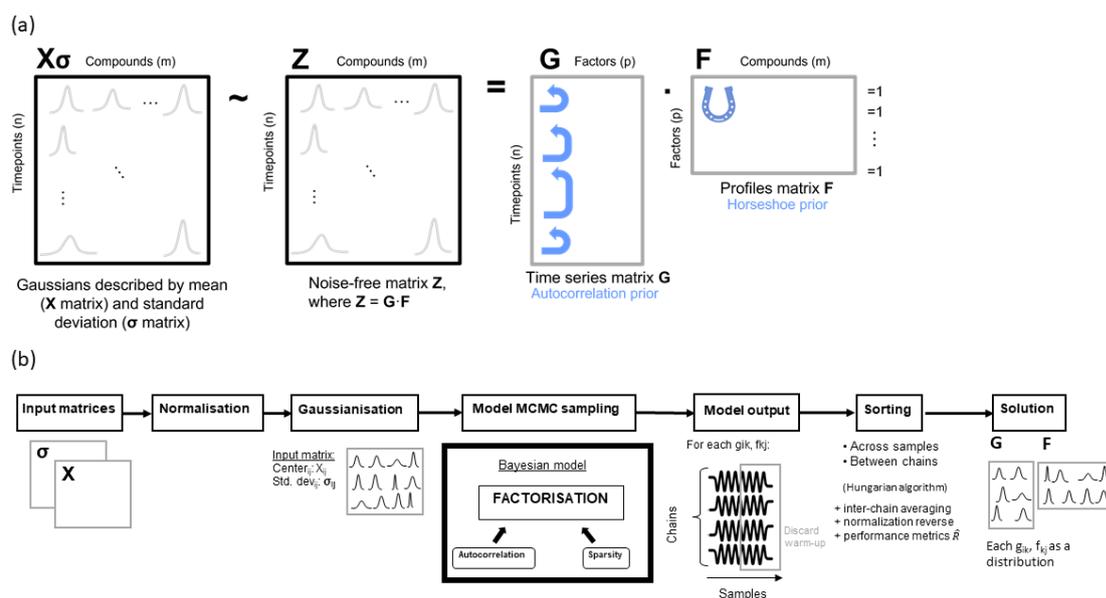


Figure S1. (a) Bayesian matrix factorisation sketch with autocorrelation and horseshoe priors. (b) Workflow diagram showing the aerosol source apportionment stages with the BAMF+HS model.

2. The potential benefits and limitations of BAMF+HS in real atmospheric data contexts (beyond the synthetic and offline examples) require deeper discussion. Readers would benefit from explicit guidance on dataset characteristics that favour the use of sparsity priors, when chemical profiles are inherently sparse versus when they are mixed and correlated.

The authors acknowledge the lack of guidance for end-users who might consider this methodology for the dataset. Specifically, sparsity priors are most beneficial in contexts where the underlying source profiles are expected to exhibit a limited number of dominant species with negligible contributions from others, such as in filter-based datasets, controlled laboratory experiments, or source types with well-defined chemical signatures. In contrast, datasets derived from ACSM observations often exhibit chemically mixed and correlated profiles due to instrumental noise, fragmentation patterns, and overlapping source contributions. In these cases, strict sparsification of the factor profiles may be neither achievable nor chemically meaningful.

We therefore emphasize that BAMF+HS should be regarded as a flexible framework that can promote sparsity when supported by the data, while naturally reverting to denser solutions when the chemical structure of the dataset does not support sparse representations. This perspective has been incorporated into the revised discussion and conclusion sections as shown below:

(Lines 625-626): “This result encourages the usage of the horseshoe prior for sparsity introduction in datasets whose solutions are expected to be strongly sparse, such as elemental datasets. ”

(Lines 639-641): “... Also, because usually ACSM profiles obtained in chamber or ambient experiments are not usually sparse, as seen in Ulbrich et al. (2009), the BAMF+HS is not as pertinent in these kinds of datasets as for filter-based datasets. ...”).

(Lines 6693-697): “... Using BAMF+HS in such datasets, the solutions reflect the sparsity of filter-based chemical profiles, hence, this newly introduced method is encouraged when source fingerprints are expected to be substantially sparse. ...”).

3. The introduction of the regularised HS prior increases model complexity and sampling demands. A brief discussion on computational performance, convergence diagnostics, and sample efficiency across dataset types would inform practitioners considering BAMF+HS for large-scale studies.

The computational time (as well as the computational resources) employed were added in Table S1. In addition, we now briefly discuss convergence diagnostics and running time across the different dataset types considered in this study as follows:

(Lines 241-246): “... Different settings were used according to the type of experiment, shown in Table S1. As seen in the table, the more complex the datasets are, the more samples they take to become stable and more time is spent on sampling. Since the BAMF+HS running times are high at this development stage, BAMF+HS in large datasets might currently be more adequate for exhaustive source apportionment refinement than real-time monitoring. The convergence of all runs has been assessed using standard Bayesian diagnostics, including visual inspection of trace plots, and the effective sample size and \hat{R} statistics, and all experiments shown in the manuscript fall within satisfactory stability ranges for these criteria. ...”).

While the introduction of the regularised HS prior increases model complexity and sampling demands relative to BAMF, we found that the additional computational cost remained manageable for the dataset sizes examined and did not compromise convergence severely.

Specific comments

4. While the authors provide quantitative metrics (e.g., correlation coefficients, ratios) to compare methods, additional visual comparisons of source fingerprints and residuals would enhance interpretability and illustrate practical differences between BAMF+HS, BAMF, and PMF solutions.

The chemical profiles of BAMF+HS, BAMF, PMF solutions are shown for the toy datasets (Figure 2), the synthetic offline dataset (Figure 3), and the real-world dataset (Figure 4). The only dataset for which the profiles are not shown are the ACSM-like datasets, which is a decision taken on the impossibility to show the results of 24 datasets (4 cities, 6 datasets per city). Plus, the differences in the solutions of receptor models might be difficult to appreciate since the differences are minimal and hard to depict in profiles who account for ~80-100 species. However, the authors present here and in the SI an example profile graph (Zurich, dates 01/01/2019-14/01/2019) with the results of BAMF, BAMF+HS, PMF in comparison to the truth in case the reader might need an example of such plots.

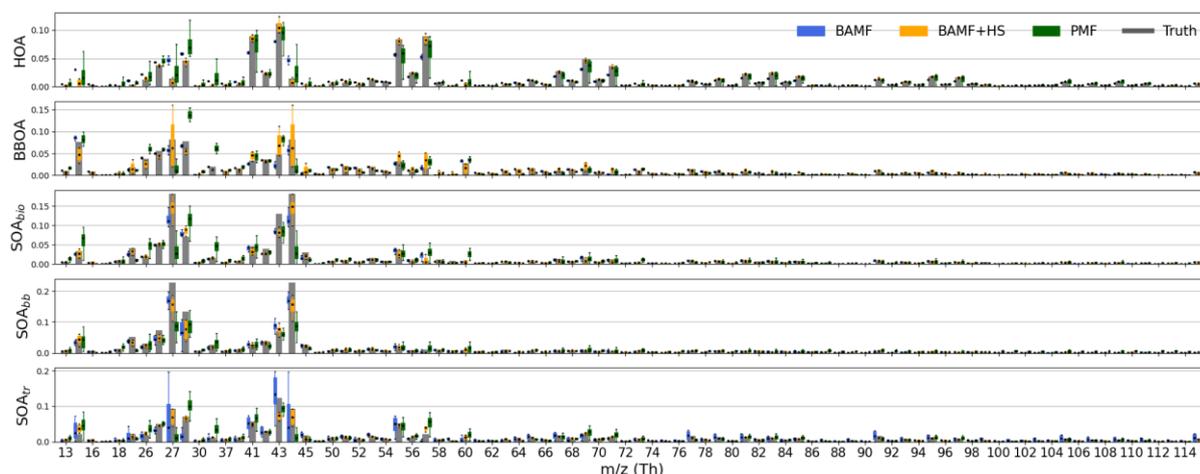


Figure S11. Example results obtained with BAMF, BAMF+HS, and PMF for one of the 24 datasets included in the ACSM-like experiment. The results correspond to Zurich, covering the period from 1 January 2019 to 14 January 2019 (dataset 0).

The text has been modified as follows in the revised version of the manuscript:

(Lines 528-530): “... All metrics over cities, datasets and sources are presented in Table S5, and an example for one site (Zurich) and one dataset (dataset 0, from 01/01/2019 to 14/01/2019) is shown in Figure S11 as an example of the results obtained by the three models in 1 out of the 24 datasets. ...”

Due to the addition of one more figure, the SI supplementary figure numbering has been adapted in the revised version of the text and the SI.

- 5. The limited achievement of sparsity in the ACSM-like datasets suggests that the HS prior may not always be appropriate. The paper should more explicitly address why the chemical structure of such data resists sparsification and whether alternative priors or hybrid techniques could overcome this challenge.**

The lack of sparsity obtained in the ACSM-like datasets is, as explained in the text, likely motivated by subtle chain divergence, which even if provides acceptable \hat{R} values, makes the values which might be close to zero substantially higher because of the chain averaging. In other words, if one or more chains present solutions in which a given factor species is different from zero, this will cause a non-zero contribution after averaging the chains, even if most of them were ~0.

The authors are not aware of any technique which can achieve to overcome this issue, driven for challenging datasets. However, the authors want to remark that most laboratory chamber ACSM profiles, as proxies for the “truth” emission profiles of environmental sources, are not sparse, probably due to instrumental noise. Hence, the sparsification of ACSM profiles is not even desirable since reference profiles are not sparse. With that, BAMF is more appropriate for these datasets, although the BAMF+HS does not cause severe damage to the source apportionment as mentioned in the manuscript. This idea has been restated in the conclusions section in the revised version of the dataset as follows:

(Lines 689-691): “... However, for ACSM-like datasets, the sparsity is not fully achieved due to converge issues, although the quality of the solution is not substantially deprecated with respect to BAMF. ...”

6. It would be beneficial and significantly increase the method’s practical usefulness to provide short recommendations on how users might diagnose when to apply BAMF+HS and to describe suggested diagnostic checks prior to deployment.

The authors acknowledge the need to enhance the explanations on suitability of the method to different air pollution datasets. However, the authors are aware that the results obtained for ACSM-like and elemental datasets found in the paper cannot be extrapolated to other types of datasets, however its use is recommended for solutions whose chemical profiles are expected to be sparse, since it does not worsen the solution with respect to BAMF. The authors have added a few sentences to the discussion and conclusions sections.

(Lines 625-626): “This result encourages the usage of the horseshoe prior for sparsity introduction in datasets whose solutions are expected to be strongly sparse, such as elemental datasets. ”

(Lines 639-641): “... Also, because usually ACSM profiles obtained in chamber or ambient experiments are not usually sparse, as seen in Ulbrich et al. (2009), the BAMF+HS is not as pertinent in these kinds of datasets as for filter-based datasets. ...”

(Lines 693-695): “... Using BAMF+HS in such datasets, the solutions reflect the sparsity of filter-based chemical profiles, hence, this newly introduced method is encouraged when source fingerprints are expected to be substantially sparse. ...”

7. Abstract: define “Toy”, “ACSM”

Reviewer 1 already rose these remarks and the text in the revised version of the manuscript has been changed as follows:

(Lines 28-30): “evaluated using three synthetic datasets designed to reflect increasing levels of data complexity (Toy, representing a highly simplified dataset; Offline; and Online), and a real-world multi-site filter dataset. ”

(Lines 32-33): “However, its application to higher-complexity Aerosol Chemical Speciation Monitor (ACSM) datasets ...”

8. Line 43-47: add other types of receptor model.

Later in the text other receptor models such as CMB and Bayesian models are introduced. Moreover, the reader is also pointed to Figure 1 in Viana et al. (2009a) for further exploration of receptor models.

9. Line 53-58: please introduce constrained/unconstrained PMF

As also pointed out in comment 12 from reviewer 2, the unconstrained/constrained distinction is currently further discussed in the revised version of the manuscript.

(Lines 62-64): "... In such cases, guiding the model by introducing a priori knowledge (common practice known as constraining the model) has been proven beneficial for the source deconvolution (Lingwall and Christensen et al. 2007, Belis et al. 2014, Dinh et al. 2025). ..."

10. Line 55: I suggest replacing "disentanglement" with deconvolution

The reviewer 1 also suggested this replacement (comment 7), which is implemented in the revised version of the manuscript as follows:

(Lines 63-64): "been proven beneficial for the source deconvolution (Lingwall and Christensen et al. 2007, Belis et al. 2014, Dinh et al. 2025)."

and also later in the text:

(Line 200): "This swap should allow \mathbf{F} to retain the \mathbf{X} matrix mass and could potentially help deconvoluting profiles ..."

11. Line 83-87: please rephrase the definition of sparsity to be clearer to the reader and add its usage together with PMF

The authors have rephrased the sentence to improve its clarity as follows:

(Lines 192): "steps towards \mathbf{F} refining should result in overall source apportionment method improvement. In this context, sparsity, defined as the property of a dataset, model or solution in which only a limited number of elements are substantial contributions while most are zero or close to zero, could be favourable for this problem. ..."

12. Line 116, 204: check the fonts

This remark was also raised by reviewer 2 in their comment #20. The revised version of the text has implemented these changes:

Line 141: "...where $i \in (2, \dots, n-1)$, C represents the...".

Lines 249-250: This coefficient compares the variance within chains and between chains of the Z matrix, hence if chains converge, $R \approx 1$, values of $R \gg 1$ imply chain divergence and values of $R \ll 1$ imply sampling divergence in chains..."

13. Line 119-120: add a proper reference to the statement

The authors added a standard reference applied to Bayesian models describing the heavy-tailed nature of the Cauchy distribution and its suitability for enabling larger jumps compared to Gaussian distributions.

Lines 144-145: "...The Cauchy distribution was chosen due to its heavier tails which enable more probable jumps between consecutive i 's than a Gaussian distribution (Gelman et al., 2013). ..."

14. Line 125: I suggest to put "involves", instead of "entails"

The revised version of the text has implemented this replacement:

Line 151: "... The introduction of sparsity in BAMF involves the addition of several hyperpriors in the \mathbf{F} prior...
."

15. Line 222: "hence" is repeated

This concern was also raised by reviewer 1 (comment #22) and reviewer 2 (comment #52) and the text has been modified as follows:

(Line 269): "The factor ordering in the matrices is random in the model results, hence, the solutions factors must be sorted. ..."

16. Line 235: define the parameters of "computational performance".

The computational performance has been evaluated through convergence metrics in STAN such as \hat{R} and only those runs which lay in acceptable ranges have been shown in the paper. The authors added one sentence in the revised version of the manuscript to describe that aspect.

(Lines 293-294): "Computational performance assessment will be based on the metrics of convergence metrics of the Hamiltonian Monte Carlo methods embedded in STAN software (e.g. \hat{R})."

17. Section 2.3.4: please mention if you tried to perturbate more factors simultaneously and if the model still catches the truth.

A similar question was raised by Reviewer 2 in comment #37, where this discussion is more extendedly addressed. The perturbation of more than one source simultaneously was not carried out and the authors believe this would prevent the isolation effect that the perturbation in a single factor enables, which would decrease the experiment interpretability. Hence, this all-factor sensitivity test is beyond the scope of this study and will be left out from the manuscript text.

18. Line 330: replace "grasp" with a more appropriate synonyms

The revised version of the text has implemented this replacement:

Lines 401-402: "... were compared to the original truth in order to comprehend the sensitivity of the models upon ...".

19. Line 347: mention the models

The models tested in this experiment are indicated in the following sentence:

Lines 426-427: "... Figure 2 shows the results of PMF, BAMF, and BAMF+HS models on the toy dataset and Table 3 shows their factorisation and comparison to truth performances. ...".

20. Line 378: please rephrase

The revised version of the text has implemented this replacement:

Lines 464-465: "... To avoid initialisation failure, BAMF was run by initialising \mathbf{F} as a normal distribution to ensure a more sturdy sampling. ...".

21. Line 380: explain why this specific factor, it has something particular?

There is nothing particular about this factor for it not to pass the t-tests, however, as pointed out there, in any case the correlation of this factor to the regular BAMF one is very strong.

22. Add in the Discussion section if the method can be applied to other types of datasets, which are the minimum requirements for these datasets

The authors are confident that BAMF+HS can be applied to datasets similar to those explored in this study, yielding results comparable to standard BAMF (e.g Xact). However, the extent to which these findings generalize to other datasets is likely to depend on their specific characteristics, and for a positive answer, more experiments should be performed. The use of BAMF+HS is particularly recommended in cases where the underlying chemical profiles are expected to be sparse, as the inclusion of the HS prior does not degrade model performance relative to BAMF when sparsity is not supported by the data. To clarify these points, we have added explanatory sentences to both the Discussion and Conclusions sections, as detailed in response to comment #2.

23. Please avoid abbreviation in the conclusion section.

The revised version of the text has applied these changes:

Lines 663-664: "... The BAMF+HS model is built in STAN, an open-source framework for statistical modelling with [Hamiltonian-Montecarlo Markov Chain](#). ...".

Lines 673-676: "... datasets and to not deprecate the solution for the more complex [datasets mimicking Aerosol Chemical Speciation Monitor \(ACSM\) data](#). ...".

Line 686: "... higher chain divergence during [Hamiltonian-Montecarlo Markov Chain](#) sampling as suggested by ...".

24. Figure 1 and 4 are difficult to follow, please improve the representation

Figures 1 and 4 are intentionally presented as distributions of the **F** components, as the aim of the experiment is to characterize the uncertainty of the factor profiles rather than single point estimates. This representation reflects the Bayesian nature of the BAMF+HS framework and differs from the more conventional point-solution visualizations commonly used in source apportionment studies. We acknowledge that this may represent a conceptual shift compared to standard PMF-based presentations. To improve clarity, we have revised the figure captions and accompanying text to more explicitly explain how these distributions should be interpreted and what information they convey (as shown in comment #25 or in #6 from Reviewer 1)

25. Figure 4 include more info in the caption (colour, elements, factors)

This caption was also modified in Reviewer 1 comment #6. The modification, applied to the revised version of the manuscript, is shown below:

Lines 881-883: "... Figure 4. Profile components distribution for PMF, BAMF, BAMF+HS ([solid colored lines](#)) in comparison to the truth ([markers](#)) on the real-world filters dataset. [Rows represent the species of the source apportionment and columns represent sources](#). ...".

26. Figure 7: use the colour code for the cities, explain the higher variability

We considered using colors instead of markers; however, the current marker-based representation was intentionally chosen to ensure readability in grayscale printing and accessibility for color-vision-deficient readers. Moreover, markers allow a clearer distinction between cities. For these reasons, we have retained the original representation, while improving the figure caption to clarify the visual encoding.

References

Gelman, A., & Shalizi, C. R. (2013). Philosophy and the practice of Bayesian statistics. *British Journal of Mathematical and Statistical Psychology*, 66(1), 8-38.

Ulbrich, I. M., Canagaratna, M. R., Zhang, Q., Worsnop, D. R., & Jimenez, J. L. (2009). Interpretation of organic components from Positive Matrix Factorization of aerosol mass spectrometric data. *Atmospheric Chemistry and Physics*, 9(9), 2891-2918.