Reply to the questions and comments of Reviewer #1 for the MS "Technical note: Mind the gap – benchmarking of various imputation approaches for precipitation stable isotope time series" by Kern and Hatvani.

We would sincerely like to thank the reviewers for the efforts and constructive comments, all of which have been of great assistance in preparing the improved version of the manuscript. The whole manuscript has been revised. The reviewers pointed out week points in the text and it is our hope that these have now been strengthened.

The questions of the reviewers are in bold.

The study seeks to compare different approaches for data imputation in monthly precipitation isotope time series records, such as those most commonly archived by the Global Network of Isotopes in Precipitation. The authors present several existing methods as well as a new method that they developed and test the efficacies of each using a study domain located largely in Austria with a few additional sites in Hungary and Slovenia. The methods are compared by omitting varying percentages of the data to evaluate the accuracy of imputation approaches against known values for missing data. Overall, I found the manuscript readable and informative, but in places the writing could be expanded for clarity, while in other cases condensed for brevity. I marked a few such examples in the line specific comments below. I think the topic is relevant for the journal and would be of interest to its readers. While I appreciate that this text was submitted as a technical note, I do also think that it might be worthwhile to expand the analyses to include other regions and to test for how the approaches perform when whole stations are removed, not just some data points from records at each station. My personal recommendation is that this should be published after at least some minor work to address the small comments, but also that it could be greatly improved with some additional analyses, and that the impact of the work would ultimately be higher if the authors did decide to take things a step further.

Thank you for the endorsement.

General comments:

The evaluation of the different methods was done in a single geographic region with relatively high station density. The effectiveness of different gap filling procedures was evaluated by randomly dropping varying numbers of observations, but retaining all stations (no whole datasets were dropped). This leaves me wondering how the results might differ if one were to take this approach to another part of the world with a different climate regime, different geography, such as more or fewer mountains or coastlines, and fewer stations overall. Since the code is all written, it actually wouldn't be a tremendous effort to pick, let's say, two more geographic regions in other parts of the world to repeat the analysis in and evaluate and compare the results.

We agree repeating the analysis in other regions would be valuable. However, this would require thoroughly quality-controlled input datasets, such as the one used in the current study. The present work relies on data that were previously screened and processed for possible discrepancies. For other regions, comparable cleaning would first have to be performed, otherwise outliers and database errors could bias the benchmarking results; this would likely require months of additional work.

Erdélyi, D., Hatvani, I. G., Derx, J., & Kern, Z. (2024). *Screening a precipitation stable isotope database for inconsistencies prior to hydrological applications – examples from the Austrian Network for Isotopes in Precipitation.* **Austrian Journal of Earth Sciences**, 117, 221–233. https://doi.org/10.17738/ajes.2024.0014

Fórizs I., Futó I., Kern Z., Hatvani I., Cserny T. (2025) *Stable isotope composition of event-based and monthly composite precipitation collected between 2003 and 2009 at two sites at the southern shore of Lake Balaton (Hungary).* **Central European Geology** 68: 18–30. https://doi.org/10.1556/24.2025.00146

Hatvani, I.G., Erdélyi, D., Vreča, P., Lojen, S., Žagar, K., Gačnik, J., Kern, Z. (2026). *Online screening tool for precipitation stable isotopes records: hybrid distance / density based outlier filtering approach via interactive web application*. **Journal of Hydrology** 672: 135401. https://doi.org/10.1016/j.jhydrol.2026.135401

We also note that the study area includes highly diverse settings, from the Alps to one of Europe's largest lowland areas, with different elevations, continentality gradients and Köppen climate types. Nevertheless, we agree that extending the approach to other climatic and geographic regions is a relevant future direction once suitable quality-controlled datasets are available.

While it would require a bit of new coding, it also wouldn't be extremely difficult to analyze the role that station density has on the results. In other words, how do the results look and compare if increasing numbers of complete time series are omitted, and not just individual observations within each time series?

We agree that station density is an important factor for spatially based approaches. However, the main aim of the present study was to benchmark methods for filling temporal gaps within existing precipitation isotope time series, not to evaluate spatial interpolation performance after removing entire stations. Such a work for a sub-domain of the studied region was conducted in a separate study (Hatvani et al. 2021). More importantly, if too many complete records were removed, parts of the analysis could no longer be performed reliably. For example, the distance-band evaluation shown in Fig. 4 requires sufficient neighboring observations at different distance classes; with sparse station coverage, especially at short distances, these comparisons would become poorly constrained or impossible.

Hatvani IG, Szatmári G, Kern Z, Erdélyi D, Vreča P, Kanduč T, Czuppon G, Lojen S, Kohán B. *Geostatistical evaluation of the design of the precipitation stable isotope monitoring network for Slovenia and Hungary*. **Environ Int.** 2021 146:106263. doi: 10.1016/j.envint.2020.106263.

Another suggestion is that it would be useful to compare the imputation approaches to model predictions of monthly values from model products that can predict monthly values, including piso.ai and isogsm, and even possibly OIPC and RCWIP. Since these models already exist and can predict monthly values, is it actually better to do imputation, or to use these existing models? Of course, OIPC and RCWIP don't make predictions for specific months, but even so, it might be useful to compare the imputation methods tested here to them as a benchmark.

Thank you for the suggestion. We accepted the idea however instead of isoGSM a recently released refined version was considered. The ICM is based on the framework originally used in isoGSM (Yoshimura et al. 2003) and incorporates isotopic variability associated with surface evaporation and sub-cloud raindrop evaporation. Meteorological forcing was taken from the JRA-3Q reanalysis dataset (Kosaka et al., 2024). The ICM-grid monthly values (Adachi and Yamanaka 2026) were extracted for the nearest grid cell to each station and evaluated against the same withheld station-month observations used for the imputation methods. This allows a direct comparison between station-based imputation and an external monthly model product. They show that ICM-grid is useful for comparison, but in our dataset the SPbI imputation approach still performs best overall.

As an illustration, please see the revised Fig. 2 draft.

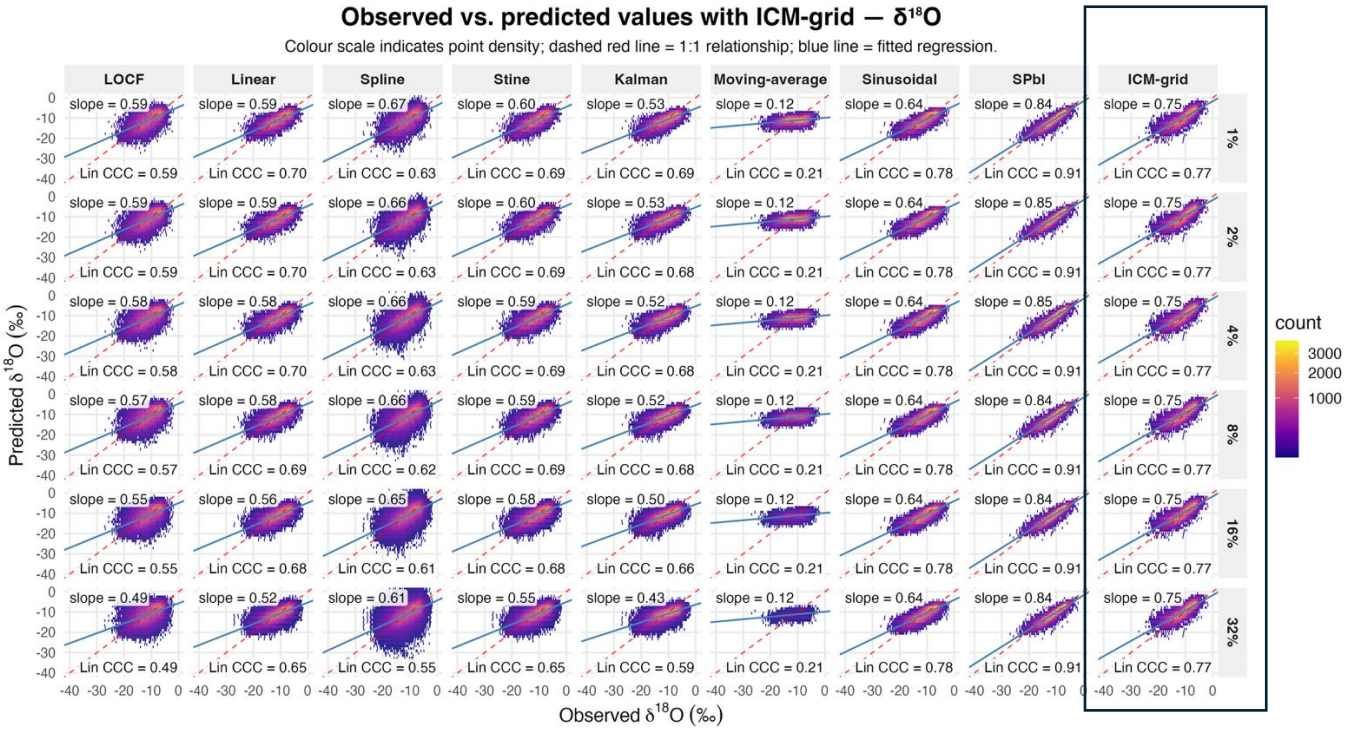

Fig. 2. (*draft*) Bland–Altman plots comparing predicted and observed monthly $\delta^{18}O$ values across all masking percentages (X %). Predictions include the tested station-based imputation methods and the external ICM-grid monthly isotope product (Adachi and Yamanaka, 2026). The y-axis shows the prediction error (predicted − observed), where positive values indicate overestimation and negative values indicate underestimation. Hex-binning is used to visualize point density. The horizontal blue line marks the mean bias within each panel. Limits of agreement are omitted for clarity. A tilted cloud indicates proportional bias, i.e., error dependence on the magnitude of the observed value. Values shown at the top of each panel are [median MAD; median RMSE]. Results were aggregated across stations and bootstrap

replicates for each rarefaction level X; ICM-grid was evaluated against the same withheld station-month observations.

Adachi, I., & Yamanaka, T. (2026). *Global dynamic precipitation isoscapes over three-quarters of a century*. **Earth System Science Data Discussions**. https://doi.org/10.5194/essd-2026-52. For the dataset itself see https://doi.org/10.5281/zenodo.17301045

Kosaka, Y., Kobayashi, S., Harada, Y., Kobayashi, C., Naoe, H., Yoshimoto, K., Harada, M. et al. 2024: *The jra-3q reanalysis*. **Journal of the Meteorological Society of Japan**. Ser. II 102-1. https://doi.org/10.2151/jmsj.2024-004

Yoshimura, K., Oki, T., Ohte, N., Kanae, S. 2003: *A quantitative analysis of short-term 18o variability with a rayleigh-type isotope circulation model*. **Journal of Geophysical Research: Atmospheres** 108-D20. https://doi.org/https://doi.org/10.1029/2003JD003477

Line specific comments:

Line 33: "Stable isotope" to "The stable isotope" Accepted and corrected.

Line 34: Could update reference to the Coplen (2011) paper. Thank you for the suggestion, we think the paper (Coplen 1994) better reflect what we meant to say.

Line 43: suggest changing "data sets" to "datasets" Accepted and corrected.

Line 53: clarify what a "decadal problem" is. Accepted and corrected, the sentence has been rewritten.

Line 54: "issue" to "issues" Accepted and corrected.

Line 72: "would outperforms" to "outperforms" Accepted and corrected.

Line 84: This is a general comment, but using this citation as a specific example, some of the references aren't formatted consistently. In this case, I wanted to look for the Hatvani (2026) reference, but I couldn't find it. I think it relates to this, but I couldn't find the paper.

The paper was in revision, now it is published and the reference has been updated.

Hatvani I.G., Erdélyi D., Vreča P., Lojen S., Žagar K., Gačnik, J., Kern Z. (2026) Online screening tool for precipitation stable isotopes records: Hybrid distance / density based outlier filtering approach via Interactive web application. *Journal of Hydrology* **672**: 135401 https://doi.org/10.1016/j.jhydrol.2026.135401

the webApp is also available with full functionality. The link is presented in the study and in the Author comment posted at 30 March 2026 already in the open discussion.

Lines 87-89: I don't follow this. So, you calculated the d-excess, and then "This metric was applied consistently to identify the same candidate period with the most continuous station data for both parameters." Please expand/clarify what you mean by this exactly.

I also don't understand the next sentence. What does it mean to compare the set of months with d-excess measurements against the full time span of a given record? I think ideas in these two sentences could be clarified by separating statements about what was done from statements about why it was done. It would make for a few more, but shorter sentences. It would also offer more space to clearly explain what exactly was done.

We thank the reviewer for pointing out that this description was unclear in the main text, however at the Github repository it is clear from the code provided what we did. The text has been revised to clarify that d-excess was not used at this stage as an interpretive isotope-hydrological metric, but as an *indicator of joint availability* of $\delta^{18}O$ and $\delta^{2}H$. Since d-excess can only be calculated when both primary isotope variables are available for the same month, months with d-excess values define the months with complete paired isotope observations. We then used this paired-availability series to identify the longest uninterrupted monthly segment at each station and retained stations with at least one continuous 84-month period. This ensured that the same months were used for $\delta^{18}O$, $\delta^{2}H$, and the subsequent d-excess evaluation.

https://github.com/istvan60/SPbI

Line 90: "required" for what? We still don't have an overview of what's going on.
Line 92: "became meaningless" for what? We are again missing some context.
Thank you for the two questions related to line 90-92. Both paragraphs in relation to the specific sentences were extensively revised, please find these below. These begin from the second paragraph of the original Sect. 2.1.

… This metric was then used as an indicator of joint data availability, to identify the same candidate period with the most continuous station data for both parameters. For each station, the full monthly time axis between the first and last available isotope observation was defined, then the months for which both $\delta^{18}O$ and $\delta^{2}H$ were present were flagged. On this binary availability series, we identified the longest uninterrupted sequence of monthly paired observations. Stations were retained as focus sites only if they contained at least one continuous 84-month period between January 1973 and December 2024 (Fig. 1).
Several test iterations showed when longer (e.g. > 10 years) record lengths were required, the number of stations meeting the selection criteria decreased substantially. Conversely, when the record length was set below 55 months, introducing a 1% artificial gap became meaningless given the monthly resolution of the data. An at least 84-month period was then used for each focus station (Fig. 1) for $\delta^{18}O$, $\delta^{2}H$, and d-excess in the imputation benchmark. Where multiple non-overlapping qualifying periods existed for a given station, all were included. This ensured that artificial masking was applied to identical months for both primary isotope variables, and that d-excess errors could subsequently be evaluated without changing the temporal subset. Most of the chosen stations — hereinafter called focus sites (n = 53) — came from Austria (n = 50), complemented by two from Slovenia and one from Hungary (Fig. 1). These continuous periods were used to test imputation methods while the remaining fragmented records were only used to calculate regional average δ_p values in the spatial imputation exercise (see sect. 2.2.1).

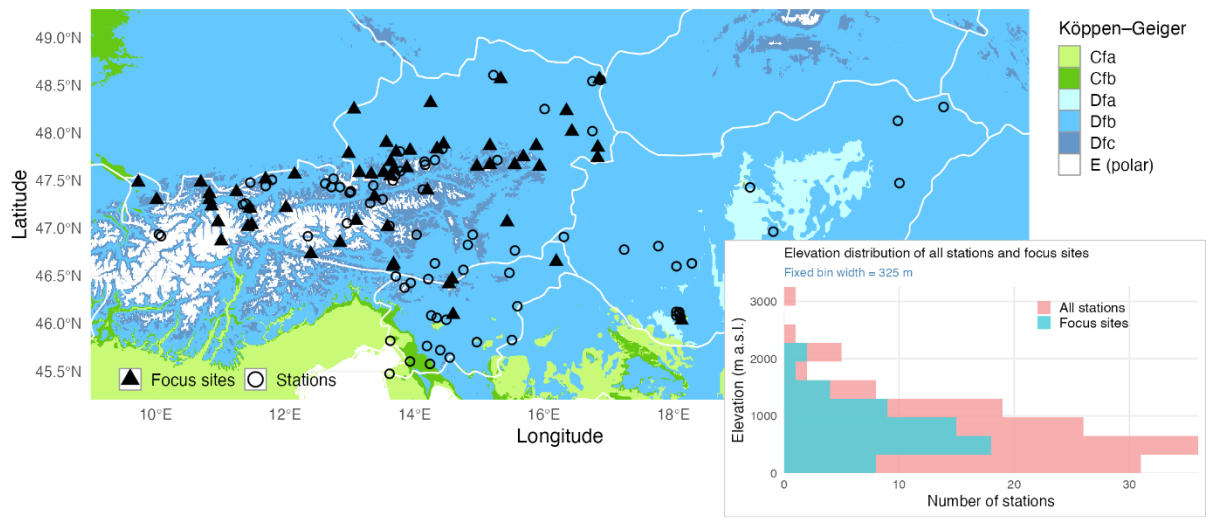

Fig. 1. Spatial distribution of the precipitation stable isotope monitoring sites across the study area. Circles denote the locations with δ_p data, while triangles mark the sites with at least 7 years uninterrupted monthly $\delta^{18}O$ and $\delta^{2}H$ observations between 1973 and 2022. Background colors indicate Köppen–Geiger climate classification zones (Beck et al., 2018). The inset histogram shows the hypsographic distribution of all stations (pink) and focus sites (teal) in 325 m elevation bands, demonstrating that the focus sites are representative of the full station network across the sampled elevation range.

Lines 146-156: This paragraph may not be necessary. I'm not sure that you need to give such a detailed explanation with equations for RMSE and MAD. In any case, the mean absolute difference uses the difference between each prediction and the mean of all predictions. It does not use the actual value of the predictand. So, in other words, it characterizes precision rather than accuracy. If you want to use the known value in an equation like this, then you want to calculate the mean absolute error (MAE). Accepted and corrected, the paragraph in question has been removed from the MS.

Lines 177-178: "In the meanwhile" is incorrect/uncommon phrasing. Accepted and corrected. A new paragraph has been started and the formulation removed.

Lines 183-184: It's not exactly fair to say this. The patten is certainly smaller than it is for other methods, but the point cloud is in all cases still not flat. Thank you for the comment, the text has been extend with: "Although the proportional bias pattern was less pronounced for SPbI than for most other methods, the point cloud was not completely horizontal, indicating that some magnitude-dependent error remained."

Yours sincerely,

Budapest, 06.09.2016. István Gábor Hatvani