

Dear Author,

("Technical note: Mind the gap – benchmarking of various imputation approaches for precipitation stable isotope time series" Zoltán Kern and István Gábor Hatvani. [Research Article MS#egusphere-2025-5493].)

General comments:

The manuscript “Technical note: Mind the gap – benchmarking of various imputation approaches for precipitation stable isotope time series” addresses an important problem in isotope hydrology: gap filling in monthly precipitation isotope records. The authors benchmark eight imputation approaches, including LOCF, Linear, Spline, Stine, Kalman, Moving-average, Sinusoidal fitting, and a newly proposed spatial proximity-based imputation method, SPbI.

The topic is relevant and timely. Missing values are common in GNIP and regional isotope monitoring networks, and complete precipitation isotope time series are often required for hydrological modelling, machine learning applications, and environmental tracing. The manuscript is generally well motivated, and the inclusion of $\delta^{18}\text{O}$, $\delta^2\text{H}$, and d-excess is valuable.

However, several aspects require substantial revision. In particular, the artificial missingness design, the reproducibility of SPbI, the justification of the 250 km threshold, the representativeness of the focus sites, and the interpretation of d-excess errors need to be strengthened.

Recommendation: Major revision.

Major comments

1. The artificial missingness mechanism is too idealized

Location: Lines 15–17, 73–75, 105–108

The authors randomly masked monthly data at 1%, 2%, 4%, 8%, 16%, and 32%. This is useful for controlled benchmarking, but real gaps in precipitation isotope records are often not completely random. Missing data may cluster in winter, during low-precipitation months, after extreme events, or during instrument failure periods.

The authors should discuss or test:

- whether the masking scheme corresponds to MCAR;
- how seasonal or consecutive missingness may affect the results;
- whether 2–3 month or 6-month gap scenarios could be evaluated;
- whether missing values near isotopic extremes influence method performance.

If no additional tests are added, the conclusions should be restricted to randomly distributed missingness.

2. The SPbI method needs more detailed description

Location: Lines 128–134

SPbI is the central methodological contribution, but the current description is not sufficiently detailed for reproducibility. Please clarify:

the minimum number of neighbouring stations required;

- whether imputation is performed when only one neighbour is available;
- whether the regional mean is unweighted;
- why distance weighting or correlation weighting was not considered;
- whether the isotope lapse rates are seasonally and spatially constant;
- whether the method is applicable across different moisture-source or climatic regimes.

I recommend adding a formula, pseudo-code, or workflow diagram for SPbI.

3. The 250 km threshold requires stronger quantitative support

Location: Lines 25–28, 157–158, 230–234, 251–252; Fig. 4

The manuscript recommends SPbI when neighbouring data are available within 250 km. This is an important practical conclusion, but the threshold appears to be based mainly on visual interpretation of Fig. 4.

Please provide:

- a clear criterion for defining the 250 km threshold;
- confidence intervals for MAD/RMSE in each distance band;
- statistical tests among distance bands;
- separate distance-dependent results for $\delta^{18}\text{O}$ and $\delta^2\text{H}$;
- a discussion of whether the threshold is region-specific.

The statement should be made more cautiously, as the 250 km threshold likely

depends on station density, topography, and regional isotope spatial coherence.

4. Representativeness of the focus sites should be clarified

Location: Lines 79–94; Fig. 1

Although 132 stations were used, the benchmarking relies on focus sites with at least seven years of uninterrupted records. The manuscript states that most focus sites are from Austria, with only two from Slovenia and one from Hungary. The exact number of focus sites is not clearly stated.

Please add:

- the total number of focus sites;
- the number of focus sites by country;
- their elevation range;
- their spatial and climatic representativeness;
- potential bias caused by the dominance of Austrian stations.

5. This is important for assessing the transferability of the conclusions.

The d-excess analysis should be deepened

Location: Lines 136–144, 192–216; Fig. 3

Including d-excess is a strength of the paper. However, the current interpretation is mostly descriptive. Please discuss:

- how $\delta^{18}\text{O}$ and $\delta^2\text{H}$ errors propagate into d-excess;
- whether errors in $\delta^{18}\text{O}$ and $\delta^2\text{H}$ are correlated;
- whether Sinusoidal imputation damps d-excess seasonality;
- whether SPbI preserves anomalous d-excess events;
- whether d-excess errors vary seasonally.

Additional metrics such as bias, correlation, or extreme-value preservation would strengthen the analysis.

6. The evaluation metrics are somewhat limited

Location: Lines 137–156

MAD, RMSE, and Bland–Altman plots are useful, but precipitation isotope time series also require evaluation of:

- isotopic extremes;

- $\delta^2\text{H}$ – $\delta^{18}\text{O}$ relationships;
- local meteoric water line properties;
- precipitation-weighted annual or seasonal means.

Since amplitude loss is discussed in Lines 180–183, a quantitative amplitude retention metric would be helpful.

7. The choice of benchmark methods should be better justified

Location: Lines 109–123, 128–134

The manuscript compares eight methods, but does not explain why other relevant methods were excluded, such as regression models using meteorological covariates, kriging, regression kriging, Bayesian models, random forest, gradient boosting, isoscape-based imputation, or bivariate methods preserving $\delta^2\text{H}$ – $\delta^{18}\text{O}$ covariance.

The authors do not need to implement all these methods, but they should clarify the scope and limitations of the benchmark.

Minor comments

1. Line 12: “in wide range” should be “in a wide range”.
2. Lines 23–24: “has shown a tendency of reducing extremes” should be “showed a tendency to reduce extremes”.
3. Line 71: “would outperforms” should be “would outperform” or “that outperforms”.
4. Lines 83–85: Please format the d-excess equation consistently.
5. Lines 89–92: State how many focus sites remain after applying the 84-month criterion.
6. Line 94: Discuss the imbalance among countries.
7. Lines 105–108: “~200,o00” appears to be a typo.
8. Line 121: “by sine curve” should be “by a sine curve”.
9. Line 123: “A 12-month sinusoidal was fit” should be “A 12-month sinusoidal curve was fitted”.
10. Line 132: Use “ km^{-1} ” instead of “km-1”.
11. Lines 146–151: RMSE and MAD equations appear incorrectly rendered.
12. Line 162: “In the case of both δp ” is unclear. Specify $\delta^{18}\text{O}$ and $\delta^2\text{H}$.

13. Lines 175–177: Consider reporting fallback rates in a table.
14. Fig. 2 and Fig. 3: Some labels are difficult to read.
15. Line 203: Use “Fig. 3” rather than “Fig 3.”
16. Lines 207–210: The sentence appears incomplete and should be checked.
17. Fig. 4: The x-axis label “Radius” is misleading; “Distance band” would be clearer.
18. Line 249: Remove the duplicated period in “isotopes..”.
19. Lines 255–259: Add punctuation between “sources” and “Umweltbundesamt”.
20. Lines 272–273: The AI-use statement should be checked against journal policy.

Final recommendation

The manuscript addresses an important and practical issue in isotope hydrology and proposes a promising spatial proximity-based imputation method. However, substantial revisions are required before publication, especially regarding missingness assumptions, SPbI reproducibility, the 250 km threshold, focus-site representativeness, and d-excess interpretation.

Recommendation: Major revision.