https://doi.org/10.5194/egusphere-2025-5493
Preprint. Discussion started: 17 December 2025
© Author(s) 2025. CC BY 4.0 License.

Technical note: Mind the gap – benchmarking of various imputation approaches for precipitation stable isotope time series

Zoltán Kern1,2, István Gábor Hatvani1,2,3*

[1]Institute for Geological and Geochemical Research, HUN-REN Research, Centre for Astronomy and Earth Sciences, Budapest, Budaörsi út 45, H-1112, Hungary
[2]HUN-REN CSFK, MTA Centre of Excellence, Budapest, Konkoly Thege Miklós út 15-17., H-1121, Hungary
[3]Center of Environmental Science, Eötvös Loránd University, Pázmány Péter stny. 1, H-1117 Budapest, Hungary

Correspondence to: István G. Hatvani (hatvaniig@gmail.com)

Abstract: Stable isotopes of hydrogen and oxygen in precipitation (δ_p) are important natural tracers in wide range of environmental applications (e.g., the exploration of the water cycle, ecology and food authenticity), yet observational records commonly contain gaps, although applications in hydrology and earth science frequently require complete cases. Eight imputation approaches were benchmarked using monthly δ_p time series from Austria, Slovenia, and Hungary. Uninterrupted periods were selected, and monthly data were masked site-wise with an increasing degree of missingness, removing 1 to 32% of the data using bootstrapping. The imputation performance of the following methods was assessed on the masked monthly data using the mean absolute difference and root mean square error between the observed and imputed values for primary and secondary isotopic parameters: Last Observation Carried Forward, Linear Interpolation, Spline Interpolation, Stineman Interpolation, Kalman Smoothing, Moving Average Imputation, Sinusoidal fit, and a spatial proximity-based imputation (SPbI) approach introduced in the present paper. SPbI estimates missing δ_p values using the mean of altitude-corrected δ_p data from within a predefined search radius. Across masking levels, SPbI was the most accurate and least prone to amplitude damping in δ_p records. Sinusoidal imputation remained robust under increasing missingness but has shown a tendency of reducing extremes, indicating amplitude loss in both δ_p and d-excess. Spline performed worst overall with the rest performing similarly up to ~16% masking beyond which their performance deteriorated. A sensitivity analysis using non-cumulative 50-km distance bands up to 400 km showed that SPbI errors increase with distance; beyond ~250 km, mean errors approach those of the sinusoidal method, making the sinusoidal—or even the simpler linear interpolation—a viable alternative when proximal observations are sparse. The benchmarking results recommend the use of SPbI where station data are available within 250 km distance, and the sinusoidal or linear approach otherwise.

https://doi.org/10.5194/egusphere-2025-5493
Preprint. Discussion started: 17 December 2025
© Author(s) 2025. CC BY 4.0 License.

1.Introduction

The ratio between the heavy and light stable isotopes in the water molecule ($^{18}O/^{16}O$;$^{2}H/^{1}H$) is an effective tool in solving practical problems in environmental isotope geochemistry (Coplen et al., 2000). Stable isotope composition of oxygen and hydrogen is conventionally expressed as δ values (δ^2H and δ^{18}O respectively) reported in per mill (‰) (Coplen, 1994). Systematic observations of the stable isotope composition of precipitation (δ_p) began in the mid-20th century (Baertschi, 1953; Dansgaard, 1953), and the global isotope monitoring network for hydrogen and oxygen isotopes in precipitation, called Global Network of Isotopes in Precipitation (GNIP), was launched in 1960 by the International Atomic Energy Agency (IAEA) and the World Meteorological Organization (WMO) (Araguas-Araguas et al., 1996). Besides the GNIP, individual stations (Vodila et al., 2011; Crawford et al., 2023; Wu et al., 2022) and regional precipitation monitoring networks (Vreča et al., 2022; Liu et al., 2014; Vachon et al., 2010; Garcia-Moya et al., 2024) were launched in the past decades providing essential isotope hydrological observation in assessing the state and understanding the change of the atmospheric branch of the hydrological cycle.

However, data sets collected during precipitation isotope monitoring efforts, like any environmental monitoring, suffer from data deficiencies. Reasons for incomplete data are manifold. Data gaps can emerge as early as the sample collection stage for numerous natural reasons, such as insufficient amount of precipitation for analysis (Von Freyberg et al., 2022), or damage to the sample collector during a severe storm (Friedman et al., 2002). Anthropogenic causes also occur, such as loss of the sample or vandalism of the monitoring site (Dores et al., 2020; Michelsen et al., 2019), as well as technical issues, such as malfunction of the sampling device (Friedman et al., 2002; Fischer et al., 2017), can also lead to missing values in an observation record. Data deficiency can arise also at a later stage of sample processing, for instance sample evaporation due to improper storage conditions (Nigro et al., 2024) may result in a lack of data assigned to a sampling date despite correct sample collection. Last but not least, inconsistent data or untraceable database errors can be identified during database screening, in which case certain data must be retrospectively excluded from the data set (IAEA, 1992; Erdélyi et al., 2024).

The comprehensive handling of missing data is a decadal problem (Rubin, 1976) and is one of the most frequently occurring data quality issue (Jäger et al., 2021; Little and Rubin, 2019; Rubin, 1976) in data science. Missing data in precipitation stable isotope time series can cause various challenges for some advanced time-series analysis methods (Yiou et al., 1996; Huang, 2014; Shumway and Stoffer, 2011). Its solutions include discarding records with incomplete cases — case-wise or variable-wise — restricting the data to the longest continuous segment, or filling the gaps with reasonable values (Yapiyev et al., 2023; Koeniger et al., 2025; Gačnik et al., 2026) to avoid losing potentially critical observations for evaluating environmental processes.

Gapless precipitation stable isotope time series may be required if one wishes to use δ_p time series as input data for isotope hydrological models (Delavau et al., 2017; Watson et al., 2024) or as predictors in machine learning models (Rácz and Gere, 2025; O' Sullivan et al., 2023; Azhar et al., 2025). Although different strategies can help maintain the operational stability of data-processing workflows, they frequently reduce the usable dataset size and — depending on the nature of the missingness

https://doi.org/10.5194/egusphere-2025-5493
Preprint. Discussion started: 17 December 2025
© Author(s) 2025. CC BY 4.0 License.

— may introduce bias into subsequent analyses, thereby further diminishing overall data quality (Little and Rubin, 2019; Schafer and Graham, 2002). Therefore, considering discipline and problem specific criteria is of the essence.

Imputation of precipitation stable isotopic time series is a challenge. In most cases it relies on the equation describing the regional linear covariance of δ^{18}O and δ^{2}H data (Cui et al., 2025; Nelson et al., 2021). This, however, can only be considered if at least one of the primary isotopic parameters is available. Another option is to exploit an empirical relationship with a highly correlating environmental variable (Dansgaard, 1964). Nevertheless, in this case the relationship of the δ_p time series should not be evaluated with the given environmental variable due to circular reasoning.

While a variety of studies exist on benchmarking imputation approaches, see Hassani et al. (2019); Jäger et al. (2021), (i) there is no imputation method which would outperforms all others in all occasions, and (ii) no benchmarking has been performed to evaluate the imputation-efficiency on water stable isotope time series. Therefore, the aim of the present work is to evaluate and compare various imputation methods for handling data deficiencies in precipitation stable isotopic time series with a focus on assessing their performance under different levels of artificially generated and controlled data sparsity taking isotope hydrological principles into account.

2.Materials and Methods

2.1. Used data and preprocessing

Monthly δ_p values were acquired from a total of 132 monitoring stations operating in three contiguous countries in Europe: Austria, Slovenia and Hungary **(Fig. 1).** The selected domain is particularly suitable for the planned study since the available monthly precipitation stable isotope records have undergone rigorous screening for potential database errors and inconsistencies with — in several cases — proposed solutions for the identified inconsistencies (Erdélyi et al., 2024; Hatvani et al., 2026; Fórizs et al., 2025). Deuterium excess (d-excess) (Dansgaard, 1964) was calculated following the conventional formula (Eq. 1) for all sites

$$d\text{-}excess = \delta^{2}\mathrm{H} - 8 \times \delta^{18}\mathrm{O} \qquad (1)$$

This metric was then applied consistently to identify the same candidate period with the most continuous station data for both parameters. Data deficiency was evaluated site-by-site by (i) comparing the set of months with d-excess measurements against the full-time span of a given record and temporal bounds of the longest uninterrupted run of data. 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 50 months, introducing a 1% artificial gap became meaningless given the monthly resolution of the data. Therefore, a duration of 84 months (7 years) of uninterrupted coverage between January 1973 to December 2022 was adopted as a practical compromise **(Fig. 1).** Most of the chosen stations —

https://doi.org/10.5194/egusphere-2025-5493
Preprint. Discussion started: 17 December 2025
© Author(s) 2025. CC BY 4.0 License.

hereinafter called focus sites — came from Austria, 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).

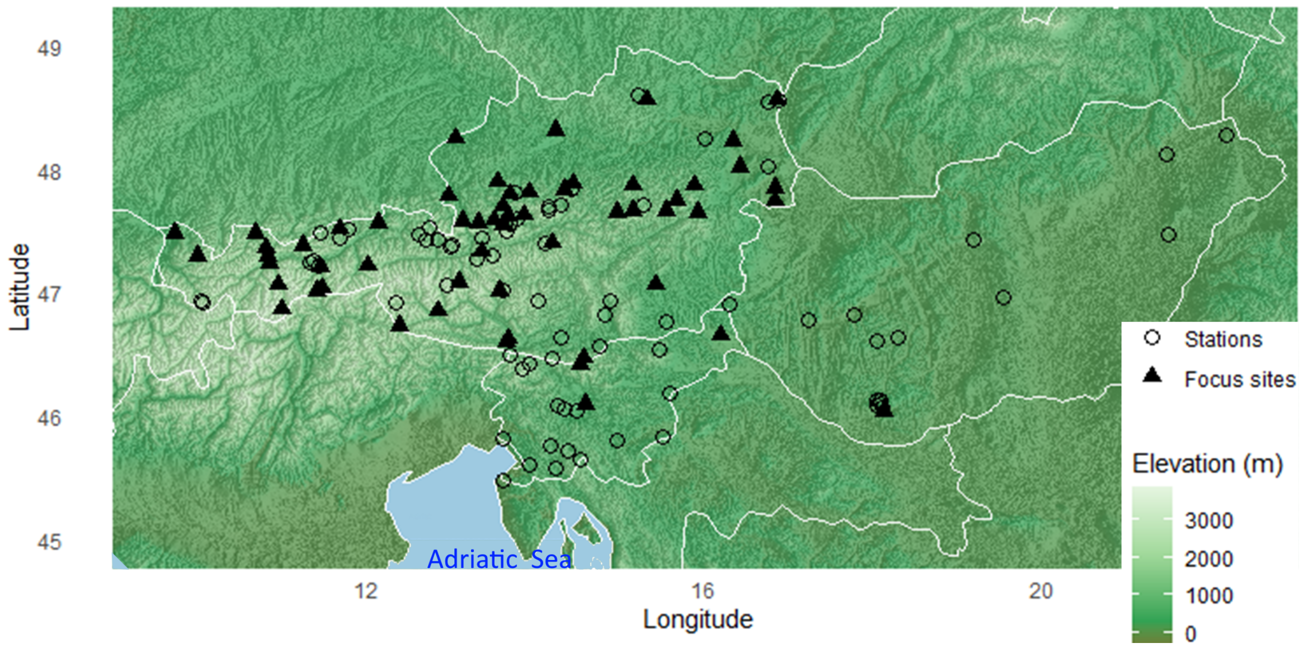

Fig. 1. Spatial distribution of the precipitation stable isotope monitoring sites across the study area. Circles denote the locations with δ_p data, while triangles mark the sites with at least 7 years uninterrupted monthly δ^{18}O and δ^{2}H observations between 1973 and 2022. Basemap digital elevation data acquired from ETOPO 2022 (NOAA, 2022).

2.2. Methodology

2.2.1. Imputation framework

The time series of the focus sites were rarefied by masking monthly data (i.e. set to *NA*) by $X \in \{1, 2, 4, 8, 16, 32\%\}$. For each masking level, the number of bootstrap replicates was set dynamically (ranging from ~950 to 35, as X increased) to obtain ~200,000 cases (imputed monthly values in the time series) across all sites. This yields comparable sample sizes at each X avoiding a replication bias among the bootstrap datasets (Davison and Hinkley, 1997) for the different masking levels.

Six frequently used imputing techniques were taken from the `imputeTS R` (Steffen and Thomas, 2017) package, namely:

LOCF: `na_locf`: last observation carried forward, replaces each missing value with the most recent present value prior to it;

https://doi.org/10.5194/egusphere-2025-5493
Preprint. Discussion started: 17 December 2025
© Author(s) 2025. CC BY 4.0 License.

Linear: `na_interpolation(linear)`;

Spline: `na_interpolation(option = "spline")`, performs cubic spline interpolation of given data points (De Boor, 1978);

Stine: (`na_interpolation(option = "stine")` performs an interpolation using a function that runs through a set of points in the xy-plane according to the algorithm of Stineman (1980);

Kalman: `na_kalman` It uses Kalman Smoothing on structural time series models (or on the state space representation of an ARIMA model) for imputation (Harvey, 1990);

Moving-average: `na_ma (k = 5)`: simple symmetric moving-average (window ± 2 months when available).

Besides these, two additional imputation approaches were involved considering specific isotope hydrological characteristics of precipitation. It was widely reported that isotope ratios in precipitation display distinct seasonal cycles that can be approximated by sine curve (Dutton et al., 2005; Feng et al., 2009; Allen et al., 2019), thus a **Sinusoidal** imputation was also applied (Eq. 2). A 12-month sinusoidal was fit per site × replicate on the rarefied series' observed values only:

$$\hat{\delta}_t = A + B \times \sin(2\pi t / 12 + \varphi), \qquad (2)$$

where t is the month index counted from the site's first observed month. Parameters (A, B, φ) are estimated by bounded nonlinear least squares using the `nlsLM()` function of the `minpack.lm` package (Elzhov et al., 2023) with starting values: offset $A = mean(\delta_p)$, amplitude $B = (max - min)/2$, $\varphi = 0$. Predictions are produced for all months, and any convergence failure or non-finite result yields *NA* at those months, which are logged as fallbacks.

Lastly a spatial proximity-based imputation approach (**SPbI**) was applied building on the foundations of spatial autocorrelation characterizing δ_p records (Terzer et al., 2013; Hatvani et al., 2020) inherited from the environmental parameters regulating their spatiotemporal variations. For each masked month the surrounding sites situated closer than 100 km with available data were identified and their corresponding δ_p values corrected to the focal site's elevation using the regional empirical isotopic lapse rates ($\delta^{18}O$: 1.2‰ km^{-1}; δ^2H: 7.9‰ km^{-1}; (Kern et al., 2020)). Masked monthly values were imputed with the regional mean of the available altitude-corrected neighbor values. If no neighbor value was available at a given month, the imputation was unsuccessful and logged as a fallback.

2.2.2. Performance evaluation

The performance of the eight imputation methods was evaluated by excluding the fallbacks and calculating the absolute differences between the observed and the imputed values for the primary isotopic parameters and d-excess by imputation method and masking percentage. For every masked month, the observed, and the imputed values were recorded, and the corresponding d-excess$_{orig}$ and d-excess$_{imp}$ were computed for further evaluation of not only the primary (δ_p), but the secondary isotopic parameters unique to isotope hydrological studies. In the procedure it was ensured that the same masked months are

https://doi.org/10.5194/egusphere-2025-5493
Preprint. Discussion started: 17 December 2025
© Author(s) 2025. CC BY 4.0 License.

used for δ^{18}O and δ^{2}H in a given replicate, ensuring consistent d-excess comparison. Then, root mean square error (RMSE) and mean absolute difference (MAD) between the observed and imputed values for δ_p and d-excess were summarized by imputation method and masking percentage.

RMSE is the quadratic mean of the differences between the observed values and predicted ones (Eq. 3).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\delta_i - \hat{\delta_i})^2} \qquad (3)$$

where n is the number of data points, $\hat{\delta}_i$ the estimated value returned by imputation method and δ_i the actual value for data point i. To avoid the amplification of the higher errors compared to the lower ones (Willmott and Matsuura, 2005) and get a more comprehensive picture MAD (Eq. 4) was also used besides root mean square error.

$$MAD = \frac{1}{n}\sum_{i=1}^{n}\left|\delta_i - \hat{\delta_i}\right| \qquad (4)$$

where n is the number of data points, $\hat{\delta}$i the value returned by the model and δi the actual value for data point i. For each (*method*, *X*), the fallback proportion was reported in percentage as 100 × (count of *NA* in $*_{imp}$)/(*count of masked rows*), which is also considered as an important factor when benchmarking the different imputation approaches. Simplified Bland–Altman plots (Bland and Altman, 1999) were also produced by each method and masking level to see the agreement between the observed and the imputed values by plotting their pairwise difference against the observations.

A sensitivity analysis was carried out to quantify how the performance of the SPbI for the primary isotopic values changes with distance from 50 to 400 km with 50 km increments in a non-cumulative way, band-wise. For each spatial band the same months' δ_p records were the imputation target as in the core run.

3. Results and discussion

3.1. Overall performance evaluation

In the case of both δ_p the pattern of the differences between the measured and the estimated values is quite similar (**Fig. 2, Fig. S1**). The average performance of all the methods is good, as the differences (imputed – observed) mostly scatter around zero. However, uncertainty (i.e. scatter) clearly increases as the degree of rarefaction (*X*%) rises (Hassani et al., 2019; Jäger et al., 2021), indicating more missing data, leads to greater uncertainty (Jäger et al., 2021) particularly with the Spline approach and less so with the Moving average (Fig. 2). The Spline approach performs poorly despite its capability to restrain the spectral characteristics of a given dataset (Hatvani et al., 2018), highlighting the importance of field – in our case isotope hydrological – exercises on imputation benchmarking. In contrast, the methods considering isotope hydrological principles to at least some

https://doi.org/10.5194/egusphere-2025-5493
Preprint. Discussion started: 17 December 2025
© Author(s) 2025. CC BY 4.0 License.

extent (Sinusoidal and SPbI) remain the most stable, with bias staying close to zero and only a slight increase in spread (**Fig. 2, Fig. S1**).

According to the overall picture there is a continuous increase in both MAD and RMSE for the Spline method, while Kalman, Linear, LOCF, Moving-average, and Stine methods perform quite similarly regardless of the rarefaction up to about 16% where a clear drop in performance is seen. As for the Sinusoidal and SPbI approaches, a steady performance is seen with SPbI performing about twice as good as the Sinusoidal with regard to MAD and RMSE (**Fig. 2**), with the SPbI approach performing the best overall. It should also be noted that fallback only occurred occasionally (< 0.15%) in the case of the SPbI approach, and only when there were no δ_p observations to calculate the regional mean in a given month for a given station. In the meanwhile, the bias (imputed minus observed) versus the observed values offers a more detailed view of the error distribution along the ordinate. For given approaches a quite striking difference is seen. It should also be noted, there is a tilt seen in the case of most of the tested imputing methods, most obviously for the smoothing methods (Kalman/Stine/Moving-average), confirming growing amplitude loss. Regarding the Sinusoidal approach, a negative slope is also observed indicating that it shrinks extremes toward the center— a classic case of smoothing or regularization, which underestimates highs and overestimates lows (**Fig. 2**). In the meanwhile SPbI is the only approach which does not show such a pattern on the Bland-Altman plots, implying it results in the least amplitude loss compared to the others (**Fig. 2**).

https://doi.org/10.5194/egusphere-2025-5493
Preprint. Discussion started: 17 December 2025
© Author(s) 2025. CC BY 4.0 License.

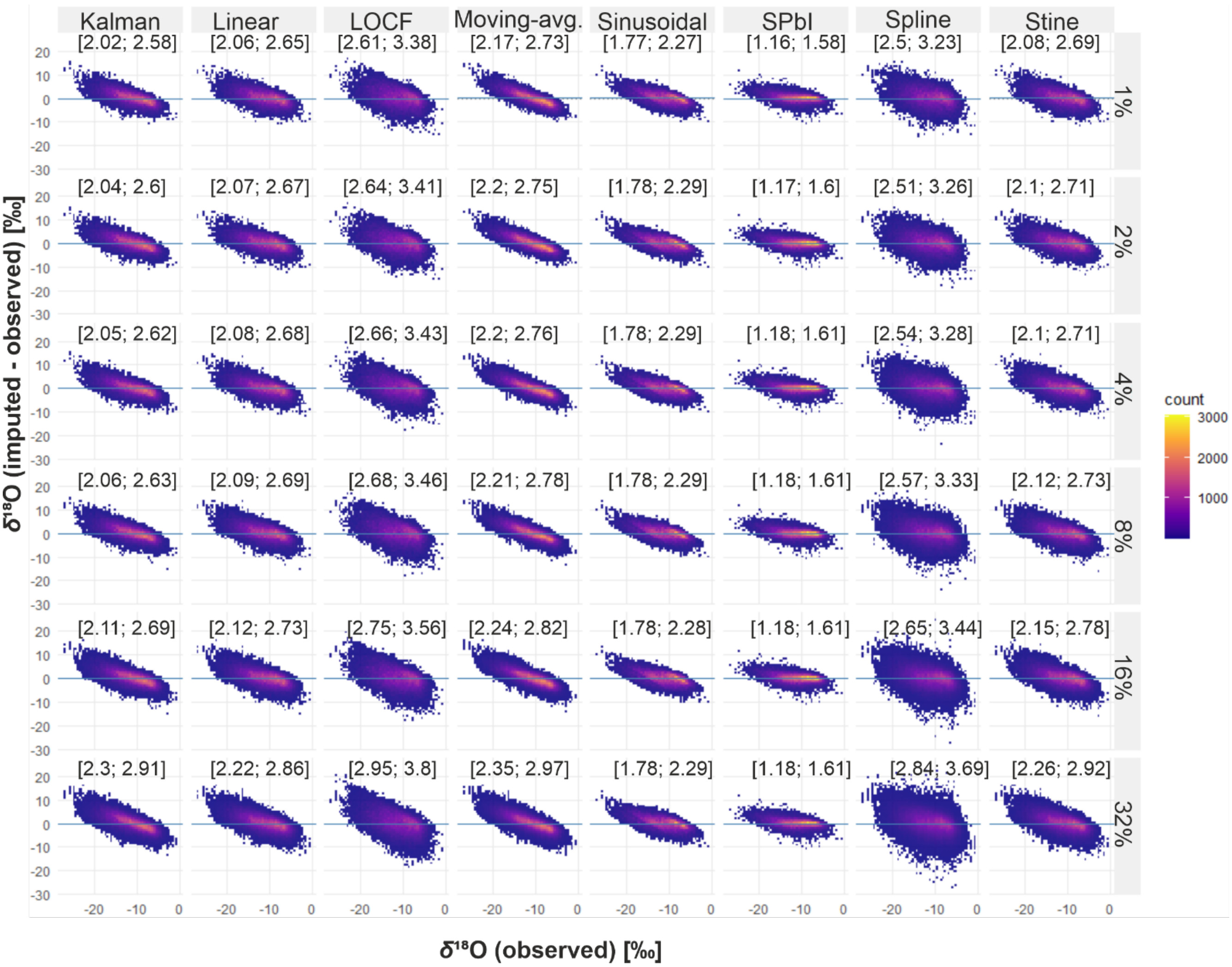

Fig. 2. Bland–Altman plots showing the differences between imputed and observed values (imputed − observed) versus the observed $\delta^{18}O$ value across all masking percentages (*X* %). Hex-binning is used to visualize the density of points. The Y-axis represents the bias for each point (solid horizontal blue line indicates no bias; positive values indicate overestimation, negative values underestimation). Limits of agreement are omitted for clarity. A tilted cloud indicates proportional bias (bias depends on magnitude). The numbers at the top of the panels are the [median MAD; median RMSE]. Results were aggregated across stations and bootstrap replicates for each rarefaction level *X*.

3.1.1. Effect of imputation on d-excess

There is a clear need for discipline-specific evaluations of imputation techniques (Gendre et al., 2024) as it remains difficult to determine which method performs best under realistic missingness conditions across diverse datasets (Jäger et al., 2021). Therefore, the performance of the imputation methods was evaluated from an isotope hydrological perspective using the

https://doi.org/10.5194/egusphere-2025-5493
Preprint. Discussion started: 17 December 2025
© Author(s) 2025. CC BY 4.0 License.

secondary isotopic variable d-excess. The errors in the primary isotopic variables (Gröning, 2011) are propagated into secondary isotopic variables (Terzer-Wassmuth et al., 2023), making d-excess a sensitive error-indicator to shed a light on any systematic bias. While the bias of the imputed values is generally close to zero, the scatter of the errors tends to increase as the degree of rarefaction rises (**Fig. 3**) unanimously with the observations on the primary variables (**Fig. 2, Fig. S1**).

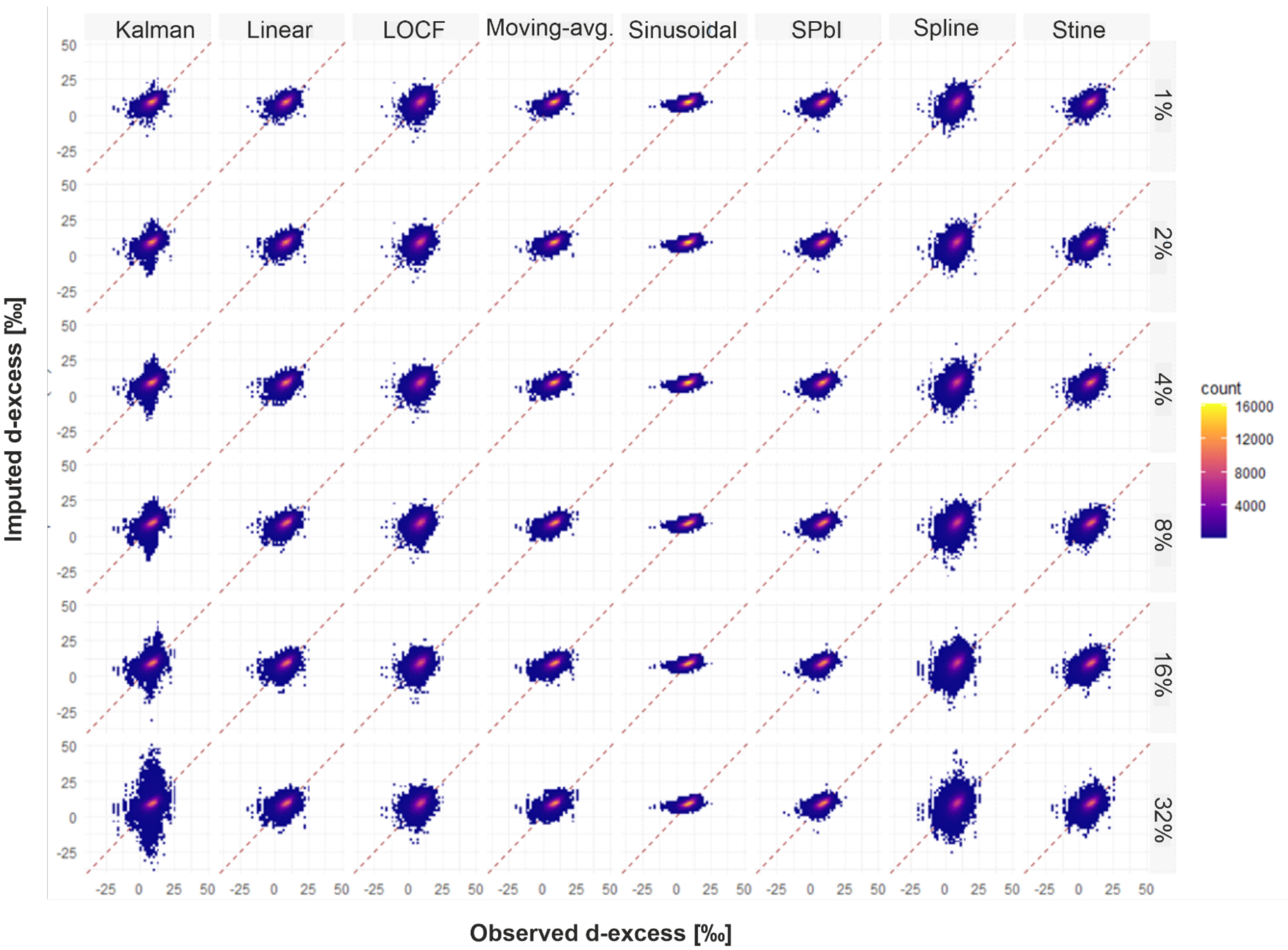

Fig 3. Comparison of observed and imputed *d-excess* values across all masking percentages (*X* %). Color scale = point density. Hex-binning is used to visualize the density of points. The dashed red line indicates the 1:1 relationship. Results were aggregated across stations and bootstrap replicates for each rarefaction level *X*.

Errors in d-excess highly increased with the fraction removed (*X*) for the Kalman and Spline methods, whereas the Stine, Linear, LOCF, and Moving-average showed modest degradation (**Fig. 3**). The Sinusoidal and SPbI approaches were generally insensitive to the increasing masking percentage. This pattern reflects the underlying design of the methods since the spatial estimator at a target month is the same altitude-corrected regional mean, independent of how many other months at the focal

https://doi.org/10.5194/egusphere-2025-5493
Preprint. Discussion started: 17 December 2025
© Author(s) 2025. CC BY 4.0 License.

site were withheld, and the Sinusoidal model imposes a single seasonal curve whose predictions at a given month change only marginally as additional points are removed. Consequently, increasing masking percentage primarily raises the number of imputed months drawn from an essentially unchanged error distribution for the Sinusoidal and SPbI imputations, rather than shifting that distribution itself.

Nevertheless, there is a crucial distinction between the Sinusoidal and SPbI approaches. The error-cloud is spread horizontally in the case of the former, while in the latter points are more tightly aligned along 1:1 line, suggesting a better match between the observed and imputed values, consequently the highest accuracy in the secondary isotopic parameter.

3.2. In-depth benchmarking of the best performing interpolation methods

It is well known that there are remarkable changes in δ_p already on a continental scale (Terzer-Wassmuth et al., 2021; Terzer et al., 2013), while on intercontinental level even seasonality of δ_p can reverse (Feng et al., 2009; Allen et al., 2019). Therefore, it is necessary to explore how the performance of the spatial proximity-based imputation method depends on the increasing distance to select data to derive the regional averages by.

To test this, SPbI is compared to the two other best performing approaches, the Sinusoidal and the Linear methods. The former relies on a general rule of isotopic seasonality of precipitation, whereas the latter represents a commonly applied simple and computationally inexpensive approach. The benchmarking shows that errors tend to increase as farther stations are used to estimate the regional mean for imputation (**Fig. 4, Fig. S2**). However, if relatively nearby data are available the SPbI clearly outperforms both Sinusoidal and Linear imputation methods. Not only the mean error, either MAD or RMSE (Fig. 4, S2), are lower but even upper quartile of the SPbI error range is below the corresponding mean error of the other methods across all tested masking percentages.

When the data used for calculating the imputed values originates from stations located more than 250 km apart at $X < 5\%$ the interquartile interval of the errors in SPbI reaches the baseline of the Sinusoidal approach (**Fig. 4, Fig. S2**). Due to the decay of spatial association between δ_p data (Terzer et al., 2013; Hatvani et al., 2020, 2021) it is reasonable to assume that beyond a certain search radius — where increasingly distant δ_p stations are incorporated into the regional mean — the Linear or Sinusoidal approaches may become more suitable than SPbI, the predictive power of which decays with spatial coherence.

https://doi.org/10.5194/egusphere-2025-5493
Preprint. Discussion started: 17 December 2025
© Author(s) 2025. CC BY 4.0 License.

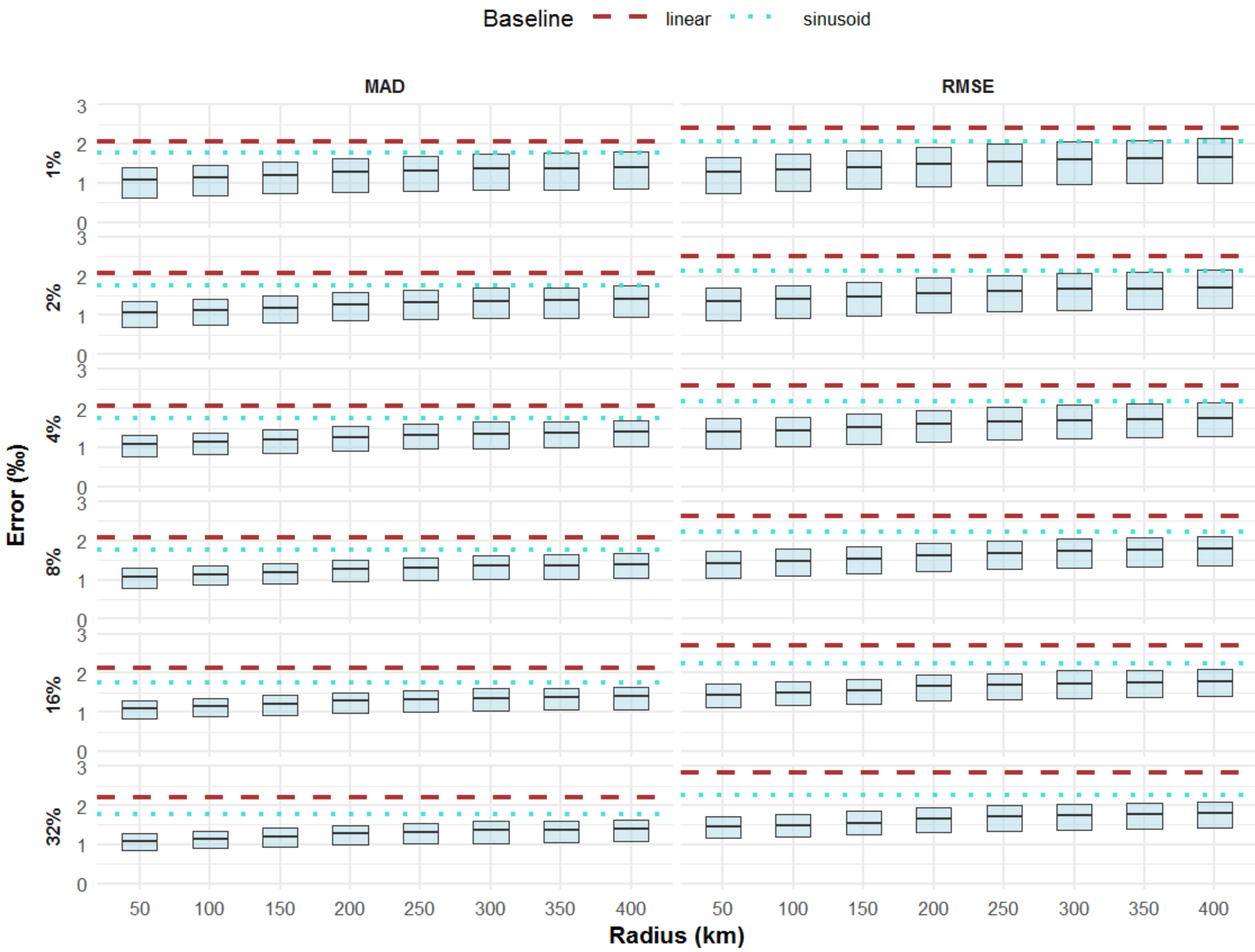

Fig. 4. Comparison of the performance of the SPbI approach using $\delta^{18}O$ values of the surrounding stations in non-overlapping 50 km spatial bands up to 400 km with 50 km increments in a non-cumulative way vs. the mean error of the Linear and Sinusoidal approaches. The Y-axis shows the absolute error of the imputed values, while the X-axis represents the spatial band used for selecting surrounding stations of SPbI. Each box shows the interquartile range (Q1–Q3) of the errors, with the mean represented by the central line. Dashed horizontal lines indicate the mean error of the Linear (red) and Sinusoidal (cyan) approaches.

4. Conclusions and outlook

The performance of eight imputation approaches was evaluated using monthly precipitation stable isotope time series from a merged dataset covering both high- and low-relief subdomains across three adjacent countries. In addition to six commonly applied imputation methods widely used across disciplines (Steffen and Thomas, 2017; Van Buuren, 2018), two approaches,

https://doi.org/10.5194/egusphere-2025-5493
Preprint. Discussion started: 17 December 2025
© Author(s) 2025. CC BY 4.0 License.

considering isotope hydrological features were tested. The Sinusoidal imputation exploits the general rule of isotopic seasonality of precipitation, whereas the approach of spatial proximity-based imputation (SPbI) – introduced in the present paper - leverages the spatial autocorrelation characterizing precipitation stable isotopes.. The results revealed that SPbI provides the most accurate imputation less prone to amplitude loss among the considered methods. However, when available data are located more than ~250 km from the target station whose gappy δ_p time series should be imputed, the Sinusoidal imputation—or, alternatively, the computationally simpler Linear interpolation—can be considered as viable options.

5. Code and data availability

The code to perform the analysis is available at: https://github.com/istvan60/SPbI/blob/main/v4 , while the used data was compiled from the following sourcesUmweltbundesamt, 2019; Erdélyi et al., 2024; Vreča et al., 2024; Vreča et al., 2014; Vreča et al., 2017; Gačnik et al., 2026; Vreča et al., 2022; Krklec et al., 2018; Domínguez-Villar et al., 2018; Rusjan et al., 2019; Mali, 2006; Doctor, 2002; Zavadlav, 2013; Kern et al., 2020; Fórizs et al., 2020; Fórizs et al., 2025; Vodila et al., 2011; Czuppon et al., 2021.

6. Author contribution

ZK came up with the original idea. ZK and IGH designed the experiments and IGH carried them out. IGH developed the model code and performed the simulations with conceptual input from ZK. ZK and IGH interpreted the results and prepared the manuscript together.

7. Competing interests

The authors declare that they have no conflict of interest.

8. Acknowledgments

On behalf of project 'Machine learning prediction of precipitation stable isotopes' we thank for the usage of HUN-REN Cloud (https://science-cloud.hu/) which helped us achieve the results published in this paper. This is contribution no. 97 of the 2ka Palæoclimatology Research Group.

https://doi.org/10.5194/egusphere-2025-5493
Preprint. Discussion started: 17 December 2025
© Author(s) 2025. CC BY 4.0 License.

During the preparation of this work the author(s) used ChatGPT in order to debug R scripts and polish the language. After using this tool/ service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

References

Allen, S. T., Jasechko, S., Berghuijs, W. R., Welker, J. M., Goldsmith, G. R., and Kirchner, J. W.: Global sinusoidal seasonality in precipitation isotopes, Hydrol. Earth Syst. Sci., 23, 3423-3436, 10.5194/hess-23-3423-2019, 2019.

Araguas-Araguas, L., Danesi, P., Froehlich, K., and Rozanski, K.: Global monitoring of the isotopic composition of precipitation, Journal of Radioanalytical and Nuclear Chemistry, 205, 189-200, 10.1007/BF02039404, 1996.

Azhar, A., Chakraborty, I., Visser, A., Liu, Y., Lerback, J. C., and Oerter, E.: Machine Learning Prediction of Tritium-Helium Groundwater Ages in the Central Valley, California, USA, Water Resources Research, 61, e2024WR038031, https://doi.org/10.1029/2024WR038031, 2025.

Baertschi, P.: Über die relativen Unterschiede im H218O-Gehalt natürlicher Wässer, Helvetica Chimica Acta, 36, 1352-1369, https://doi.org/10.1002/hlca.19530360619, 1953.

Bland, J. M. and Altman, D. G.: Measuring agreement in method comparison studies, Statistical Methods in Medical Research, 8, 135-160, 10.1177/096228029900800204, 1999.

Coplen, T. B.: Reporting of stable hydrogen, carbon and oxygen isotopic abundances, Pure App. Chem., 66, 273-276, 1994.

Coplen, T. B., Herczeg, A. L., and Barnes, C.: Isotope Engineering—Using Stable Isotopes of the Water Molecule to Solve Practical Problems, in: Environmental Tracers in Subsurface Hydrology, edited by: Cook, P. G., and Herczeg, A. L., Springer US, Boston, MA, 79-110, 10.1007/978-1-4615-4557-6_3, 2000.

Crawford, J., Hughes, C. E., and Chambers, S. D.: Controls on the stable isotope composition of daily precipitation in Sydney Australia: 9 years of daily data, including Radon-222, Journal of Hydrology, 617, 129123, https://doi.org/10.1016/j.jhydrol.2023.129123, 2023.

Cui, Y., Tian, L., Cai, Z., and Wang, S.: Spatially inhomogeneous response of precipitation δ18O in China to ENSO cycles, npj Climate and Atmospheric Science, 8, 164, 10.1038/s41612-025-01057-1, 2025.

Czuppon, G., Bottyán, E., Kristóf, E., Weidinger, T., Haszpra, L., and Kármán, K.: Stable isotope data of daily precipitation during the period of 2013-2017 from K-puszta (regional background monitoring station), Hungary, Data in Brief, 36, 10.1016/j.dib.2021.106962, 2021.

Dansgaard, W.: The Abundance of O18 in Atmospheric Water and Water Vapour, Tellus, 5, 461-469, https://doi.org/10.1111/j.2153-3490.1953.tb01076.x, 1953.

Dansgaard, W.: Stable isotopes in precipitation, Tellus, 16, 436-468, 1964.

Davison, A. C. and Hinkley, D. V.: Bootstrap Methods and Their Application, 1, Cambridge university press, UK, https://doi.org/10.1017/CBO9780511802843, 1997.

Delavau, C. J., Stadnyk, T., and Holmes, T.: Examining the impacts of precipitation isotope input (δ18Oppt) on distributed, tracer-aided hydrological modelling, Hydrol. Earth Syst. Sci., 21, 2595-2614, 10.5194/hess-21-2595-2017, 2017.

Doctor, D. H.: The hydrogeology of the classical karst (Kras) aquifer of southwestern Slovenia, University of Minnesota2002.

Domínguez-Villar, D., Lojen, S., Krklec, K., Kozdon, R., Edwards, R. L., and Cheng, H.: Ion microprobe δ18O analyses to calibrate slow growth rate speleothem records with regional δ18O records of precipitation, Earth and Planetary Science Letters, 482, 367-376, https://doi.org/10.1016/j.epsl.2017.11.012, 2018.

Dores, D., Glenn, C. R., Torri, G., Whittier, R. B., and Popp, B. N.: Implications for groundwater recharge from stable isotopic composition of precipitation in Hawai'i during the 2017–2018 La Niña, Hydrological Processes, 34, 4675-4696, https://doi.org/10.1002/hyp.13907, 2020.

Dutton, A., Wilkinson, B. H., Welker, J. M., Bowen, G. J., and Lohmann, K. C.: Spatial distribution and seasonal variation in 18O/16O of modern precipitation and river water across the conterminous USA, Hydrological Processes, 19, 4121-4146, https://doi.org/10.1002/hyp.5876, 2005.

Elzhov, T. V., Mullen, K. M., Spiess, A.-N., and Bolker, B.: minpack.lm: R Interface to the Levenberg-Marquardt Nonlinear Least-Squares Algorithm Found in MINPACK, Plus Support for Bounds (1.2-4) [code], 2023.

© Author(s) 2025. CC BY 4.0 License.

Erdélyi, D., Hatvani, I. G., Derx, J., and Kern, Z.: 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, 10.17738/ajes.2024.0014, 2024.

Feng, X., Faiia, A. M., and Posmentier, E. S.: Seasonality of isotopes in precipitation: A global perspective, Journal of Geophysical Research: Atmospheres, 114, https://doi.org/10.1029/2008JD011279, 2009.

Fischer, B. M. C., van Meerveld, H. J., and Seibert, J.: Spatial variability in the isotopic composition of rainfall in a small headwater catchment and its effect on hydrograph separation, Journal of Hydrology, 547, 755-769, https://doi.org/10.1016/j.jhydrol.2017.01.045, 2017.

Fórizs, I., Kern, Z., Csicsák, J., and Csurgó, G.: Monthly data of stable isotopic composition (δ18O, δ2H) and tritium activity in precipitation from 2004 to 2017 in the Mecsek Hills, Hungary, Data in Brief, 32, 106206, https://doi.org/10.1016/j.dib.2020.106206, 2020.

Fórizs, I., Futó, I., Kern, Z., Gábor, H. I., and Cserny, T.: Stable isotope composition of event-based and monthly composite precipitation collected between 2003 and 2009 in the south-western sector of Lake Balaton (Hungary), Central European Geology, in press, 2025.

Friedman, I., Smith, G. I., Johnson, C. A., and Moscati, R. J.: Stable isotope compositions of waters in the Great Basin, United States 2. Modern precipitation, Journal of Geophysical Research: Atmospheres, 107, ACL 15-11-ACL 15-22, https://doi.org/10.1029/2001JD000566, 2002.

Gačnik, J., Žagar, K., Hatvani, I. G., Kern, Z., and Vreča, P.: Climate change reflected in 40-year isotopic composition trends of precipitation in Slovenia, Environmental Research, 123286, https://doi.org/10.1016/j.envres.2025.123286, 2026.

Garcia-Moya, A., Alonso-Hernández, C. M., Sánchez-Murillo, R., Morera-Gómez, Y., Sánchez-Llull, M., Díaz Rizo, O., Cuesta Santos, O., López Lee, R., Brígido Flores, O., Ramos Viltre, E. O., and Ortega, L.: Spatiotemporal characterization of the isotopic composition of meteoric waters in Cuba, Scientific Data, 11, 1398, 10.1038/s41597-024-04178-z, 2024.

Gendre, M., Hauffe, T., Pimiento, C., and Silvestro, D.: Benchmarking imputation methods for categorical biological data, Methods in Ecology and Evolution, 15, 1624-1638, https://doi.org/10.1111/2041-210X.14339, 2024.

Gröning, M.: Improved water δ2H and δ18O calibration and calculation of measurement uncertainty using a simple software tool, Rapid Communications in Mass Spectrometry, 25, 2711-2720, https://doi.org/10.1002/rcm.5074, 2011.

Harvey, A. C.: Forecasting, structural time series models and the Kalman filter, 1990.

Hassani, H., Kalantari, M., and Ghodsi, Z.: Evaluating the Performance of Multiple Imputation Methods for Handling Missing Values in Time Series Data: A Study Focused on East Africa, Soil-Carbonate-Stable Isotope Data, Stats, 2, 457-467, 2019.

Hatvani, I. G., Erdélyi, D., Vreča, P., and Kern, Z.: Analysis of the Spatial Distribution of Stable Oxygen and Hydrogen Isotopes in Precipitation Across the Iberian Peninsula, Water, 12, 481, 10.3390/w12020481, 2020.

Hatvani, I. G., Kern, Z., Leél-Őssy, S., and Demény, A.: Speleothem stable isotope records for east-central Europe: resampling sedimentary proxy records to obtain evenly spaced time series with spectral guidance, Earth System Science Data, 10, 139-149, 2018.

Hatvani, I. G., Erdélyi, D., Vreča, P., Lojen, S., Žagar, K., Gačnik, J., and Kern, Z.: Online screening tool for precipitation stable isotopes records: hybrid distance / density based outlier filtering approach via interactive web application, Journal of Hydrology, HYDROL69130R69131, 2026.

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

Huang, N. E.: Hilbert-Huang transform and its applications, World scientific2014.

IAEA: Statistical treatment of data on environmental isotopes in precipitation, Technical Report Series 331, International Atomic Energy Agency, Vienna, 781 pp.1992.

Jäger, S., Allhorn, A., and Bießmann, F.: A Benchmark for Data Imputation Methods, Frontiers in Big Data, Volume 4 - 2021, 10.3389/fdata.2021.693674, 2021.

Kern, Z., Hatvani, I. G., Czuppon, G., Fórizs, I., Erdélyi, D., Kanduč, T., László, P., and Vreča, P.: Isotopic ‘Altitude’ and ‘Continental’ Effects in Modern Precipitation across the Adriatic–Pannonian Region, Water, 12, 1797, 10.3390/w12061797, 2020.

© Author(s) 2025. CC BY 4.0 License.

Koeniger, P., Neukum, C., Stadler, S., Noell, U., Marshall, J. D., Ahrends, B., Fleck, S., and Meesenburg, H.: Dynamics of stable isotopes in precipitation, soil water and groundwater at a Norway spruce and a European beech site at Solling, Germany, Isotopes in Environmental and Health Studies, 1-19, 10.1080/10256016.2025.2509756, 2025.

Krklec, K., Domínguez-Villar, D., and Lojen, S.: The impact of moisture sources on the oxygen isotope composition of precipitation at a continental site in central Europe, Journal of Hydrology, 561, 810-821, https://doi.org/10.1016/j.jhydrol.2018.04.045, 2018.

Little, R. J. and Rubin, D. B.: Statistical analysis with missing data, John Wiley & Sons2019.

Liu, J., Song, X., Yuan, G., Sun, X., and Yang, L.: Stable isotopic compositions of precipitation in China, Tellus B: Chemical and Physical Meteorology, 66, 22567, 10.3402/tellusb.v66.22567, 2014.

Mali, N.: Characterization of transport processes in the unsaturated zone of a gravel aquifer by natural and artificial tracers, Univerza v Novi Gorici, Fakulteta za podiplomski študij, 2006.

Michelsen, N., Laube, G., Friesen, J., Weise, S. M., Bait Said, A. B. A., and Müller, T.: Technical note: A microcontroller-based automatic rain sampler for stable isotope studies, Hydrol. Earth Syst. Sci., 23, 2637-2645, 10.5194/hess-23-2637-2019, 2019.

Nelson, D. B., Basler, D., and Kahmen, A.: Precipitation isotope time series predictions from machine learning applied in Europe, Proceedings of the National Academy of Sciences, 118, e2024107118, doi:10.1073/pnas.2024107118, 2021.

Nigro M, Žagar K, Vreča P.: A Simple Water Sample Storage Test for Water Isotope Analysis. Sustainability, 16(11):4740. https://doi.org/10.3390/su16114740, 2024

NOAA, N. C. f. E. I.: ETOPO 2022 15 Arc-Second Global Relief Model. [dataset], doi:10.25921/fd45-gt74, 2022.

O' Sullivan, R., Cama-Moncunill, R., Salter-Townshend, M., Schmidt, O., and Monahan, F. J.: Verifying origin claims on dairy products using stable isotope ratio analysis and random forest classification, Food Chem X, 19, 100858, 10.1016/j.fochx.2023.100858, 2023.

Rácz, A. and Gere, A.: Comparison of missing value imputation tools for machine learning models based on product development cases studies, LWT, 221, 117585, https://doi.org/10.1016/j.lwt.2025.117585, 2025.

Rubin, D. B.: Inference and missing data, Biometrika, 63, 581-592, 10.1093/biomet/63.3.581, 1976.

Rusjan, S., Sapač, K., Petrič, M., Lojen, S., and Bezak, N.: Identifying the hydrological behavior of a complex karst system using stable isotopes, Journal of Hydrology, 577, 123956, https://doi.org/10.1016/j.jhydrol.2019.123956, 2019.

Schafer, J. L. and Graham, J. W.: Missing data: our view of the state of the art, Psychological methods, 7, 147, 2002.

Shumway, R. and Stoffer, D.: Time series analysis and its application, 3, Springer - Verlag New York2011.

Steffen, M. and Thomas, B.-B.: imputeTS: Time Series Missing Value Imputation in R., The R Journal, 9, 207–218, doi:10.32614/RJ-2017-009, 2017.

Stineman, R. W.: A consistently well-behaved method of interpolation, Creative Computing, 1980.

Terzer-Wassmuth, S., Wassenaar, L. I., Araguás-Araguás, L. J., and Stumpp, C.: Balancing precision and throughput of δ17O and Δ'17O analysis of natural waters by Cavity Ringdown Spectroscopy, MethodsX, 10, 102150, https://doi.org/10.1016/j.mex.2023.102150, 2023.

Terzer-Wassmuth, S., Wassenaar, L. I., Welker, J. M., and Araguás-Araguás, L. J.: Improved high-resolution global and regionalized isoscapes of δ18O, δ2H and d-excess in precipitation, Hydrological Processes, 35, e14254, https://doi.org/10.1002/hyp.14254, 2021.

Terzer, S., Wassenaar, L. I., Araguás-Araguás, L. J., and Aggarwal, P. K.: Global isoscapes for δ18O and δ2H in precipitation: improved prediction using regionalized climatic regression models, Hydrol. Earth Syst. Sci., 17, 4713-4728, 10.5194/hess-17-4713-2013, 2013.

Umweltbundesamt, H. O. F.: [dataset], 2019.

Vachon, R. W., Welker, J. M., White, J. W. C., and Vaughn, B. H.: Monthly precipitation isoscapes (δ18O) of the United States: Connections with surface temperatures, moisture source conditions, and air mass trajectories, Journal of Geophysical Research: Atmospheres, 115, 10.1029/2010jd014105, 2010.

Van Buuren, S.: Flexible imputation of missing data, CRC press, Boca Raton, FL, USA, 342 pp.2018.

Vodila, G., Palcsu, L., Futó, I., and Szántó, Z.: A 9-year record of stable isotope ratios of precipitation in Eastern Hungary: Implications on isotope hydrology and regional palaeoclimatology, Journal of Hydrology, 400, 144-153, https://doi.org/10.1016/j.jhydrol.2011.01.030, 2011.

https://doi.org/10.5194/egusphere-2025-5493
Preprint. Discussion started: 17 December 2025
© Author(s) 2025. CC BY 4.0 License.

von Freyberg, J., Rücker, A., Zappa, M., Schlumpf, A., Studer, B., and Kirchner, J. W.: Four years of daily stable water isotope data in stream water and precipitation from three Swiss catchments, Scientific Data, 9, 46, 10.1038/s41597-022-01148-1, 2022.
Vreča, P., Pavšek, A., and Kocman, D.: SLONIP-A Slovenian Web-Based Interactive Research Platform on Water Isotopes in Precipitation, Water, 14, 2127, https://doi.org/10.3390/w14132127, 2022.
Vreča, P., Krajcar Bronić, I., Leis, A., and Demšar, M.: Isotopic composition of precipitation at the Station Ljubljana (Reaktor), Slovenia - period 2007-2010, Geologija, 57, 217 - 230, https://doi.org/10.5474/geologija.2014.019 2014.
Vreča, P., Štrok, M., Žigon, S., and Svetek, B.: Isotope Composition of Precipitation at Stations Ljubljana-Reaktor and Portorož airport: period 2011-2015, Jožef Stefan InstituteIJS-DP-12383, 32, 2017.
Vreča, P., Kanduč, T., Štrok, M., Žagar, K., Nigro, M., and Barsanti, M.: An Assessment of Six Years of Precipitation Stable Isotope and Tritium Activity Concentration Records at Station Sv. Urban, Eastern Slovenia, Water, 16, 469, 2024.
Watson, A., Birkel, C., Arciniega-Esparza, S., de Waal, J., Miller, J., Vystavna, Y., van Rooyen, J., Welham, A., Bong, H., Yoshimura, K., Helmschrot, J., Künne, A., and Kralisch, S.: Evaluating input data sources for isotope-enabled rainfall-runoff models, Hydrological Processes, 38, e15276, https://doi.org/10.1002/hyp.15276, 2024.
Willmott, C. J. and Matsuura, K.: Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance, Climate research, 30, 79-82, 2005.
Wu, H., Fu, C., Zhang, C., Zhang, J., Wei, Z., and Zhang, X.: Temporal Variations of Stable Isotopes in Precipitation from Yungui Plateau: Insights from Moisture Source and Rainout Effect, Journal of Hydrometeorology, 23, 39-51, https://doi.org/10.1175/JHM-D-21-0098.1, 2022.
Yapiyev, V., Rossi, P. M., Ala-Aho, P., and Marttila, H.: Stable Water Isotopes as an Indicator of Surface Water Intrusion in Shallow Aquifer Wells: A Cold Climate Perspective, Water Resources Research, 59, e2022WR033056, https://doi.org/10.1029/2022WR033056, 2023.
Yiou, P., Baert, E., and Loutre, M. F.: Spectral analysis of climate data, Surveys in Geophysics, 17, 619-663, 10.1007/BF01931784, 1996.
Zavadlav, S.: CO [sub] 2 Dynamics in a River System: Mass Balance, Hydrological, Geochemical and Biochemical Impacts: Doctoral Dissertation, S. Zavadlav, 2013.