



## Research papers

# Predicting spatial distribution of stable isotopes in precipitation by classical geostatistical- and machine learning methods

Dániel Erdélyi<sup>a,b</sup>, István Gábor Hatvani<sup>c,d,\*</sup>, Hyeonjeon Jeon<sup>e</sup>, Matthew Jones<sup>f</sup>, Jonathan Tyler<sup>g</sup>, Zoltán Kern<sup>c,d</sup>

<sup>a</sup> Department of Geology, Eötvös Loránd University, Pázmány Péter stry. 1, H-1117 Budapest, Hungary

<sup>b</sup> Doctoral School of Environmental Sciences, Eötvös Loránd University, Pázmány Péter stry 1, H-1117 Budapest, Hungary

<sup>c</sup> Institute for Geological and Geochemical Research, Research Centre for Astronomy and Earth Sciences, Budaörsi út 45, H-1112 Budapest, Hungary

<sup>d</sup> CSFK, MTA Centre of Excellence, Budapest, Konkoly Thege Miklós út 15-17, 1121 Budapest, Hungary

<sup>e</sup> Department of Statistics, Iowa State University, 2438 Osborn Dr, Ames, IA 50011, United States

<sup>f</sup> School of Geography, University of Nottingham, University Park, Nottingham NG7 2D, United Kingdom

<sup>g</sup> Department of Earth Sciences, University of Adelaide, Adelaide 5005 South Australia, Australia



## ARTICLE INFO

This manuscript was handled by Marco Borga, Editor-in-Chief, with the assistance of Huade Guan, Associate Editor

## Keywords:

Water stable isotopes  
Geostatistics  
Random forest  
Multiple regression  
Lasso regression  
Regression kriging

## ABSTRACT

Stable isotopes of precipitation are important natural tracers in hydrology, ecology, and forensics. The spatially explicit predictions of oxygen and hydrogen isotopes in precipitation are obtained through different interpolation techniques. In the present study we aim to examine the performance of various interpolation techniques when predicting the spatial distribution of precipitation stable isotopes. The efficiency of combined geostatistical tools (i.e. regression kriging; RK) and various machine learning methods (including regression enhanced random forest methods: MRRF, RERF) are compared in interpolating the spatial variability of precipitation stable oxygen isotope values from two different sampling networks in Europe. To assess the performance of the models, mean squared error (MSE), nonparametric Kling Gupta efficiency (KGE), absolute differences and relative mean absolute error metrics were employed. It was found that the combination of the different regression techniques with Random Forest can produce estimations with comparable accuracy in terms of descending order of overall average MSE, MRRF: 2.61, RK: 2.77, RERF: 2.99, RF: 3.08. The best performing combined random forest variant (MRRF) outperformed regression kriging in terms of a hybrid error metric (KGE) by 7.5%. Sequential random rarefying the station networks showed that machine-learning methods are more capable of maintaining high prediction accuracy even with fewer input data. This can be a great advantage when a suitable method is needed to predict the stable isotope composition of precipitation for large spatial domains where the spatial density of data stations shows large differences.

## 1. Introduction

The ratio of heavy and light stable isotopes of the water molecule ( $^{18}\text{O}/^{16}\text{O}$ ;  $^2\text{H}/^1\text{H}$ ) is a frequently used tool in environmental isotope geochemistry, specifically in hydrology, climatology and biogeochemistry (Coplen et al., 2000). Stable isotope composition is conventionally expressed as  $\delta$  values in per mill (Coplen, 1994). The isotopic composition of precipitation ( $\delta_p$ ) provides an insight into the origin of water vapor, and the conditions attained during condensation and

precipitation (Dansgaard, 1964). Using these variations, stable isotopes in meteoric water have become important natural tracers in the study of the water cycle (Bowen and Good, 2015; Fórizs, 2003). With the continuous advancement in effectiveness and availability of analytical tools, the spatiotemporal abundance of precipitation stable isotope measurements is steadily increasing (Yoshimura, 2015) providing sufficient ground for the development of spatially continuous datasets of isotope composition. These datasets can be utilized for advanced hydrological applications of precipitation stable isotopes where excess

**Abbreviations:** RK, regression kriging; MRRF, multiple regression random forest; RERF, regression enhanced random forest; MSE, mean squared error; KGE, Kling Gupta efficiency; RF, random forest.

\* Corresponding author at: Institute for Geological and Geochemical Research, Research Centre for Astronomy and Earth Sciences, Budaörsi út 45, H-1112 Budapest, Hungary.

E-mail address: [hatvaniig@gmail.com](mailto:hatvaniig@gmail.com) (I.G. Hatvani).

<https://doi.org/10.1016/j.jhydrol.2023.129129>

Received 8 August 2022; Received in revised form 1 December 2022; Accepted 12 January 2023

Available online 16 January 2023

0022-1694/© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

information can be gained from not only having point data, but spatially continuous information. Such applications can be found in: hydrogeology (Bowen and Good, 2015; Clark and Fritz, 1997); limnology (Birkel et al., 2018; Nan et al., 2019; Tyler et al., 2022); water resource management (Bowen and Good, 2015; Gibson and Edwards, 2002); exploration of changes in moisture source conditions (Amundson et al., 1996); animal migration studies (Hobson, 1999; Hobson and Wassenaar, 1996); food source traceability (Heaton et al., 2008); as well as forensic sciences studies (Ehleringer et al., 2008).

To interpolate isotope monitoring data across space, a pioneering approach combined (i) an initial empirical model of trend driven by independent (geographical) parameters, and (ii) kriging of the residuals (Bowen and Revenaugh, 2003; Bowen and Wilkinson, 2002). Two decades ago, this particular approach was tested against potential alternatives to interpolate  $\delta_p$  variability and has been proven to provide better estimates (Bowen and Revenaugh, 2003). This method has been widely applied and became the “gold standard” in mapping the spatial distribution of  $\delta_p$  globally (e.g. Bowen (2010); Terzer et al. (2013); van der Veer et al. (2009)), and regionally; e.g. Chan et al. (2012); Hatvani et al. (2020); Hatvani et al. (2017); Kaseke et al. (2016); Kern et al. (2014); Lykoudis and Argiriou (2007).

Machine learning approaches (random forests and similar techniques) are increasingly used to extract patterns and insights from the ever-increasing stream of geospatial data (Reichstein et al., 2019), becoming a frequently used tool in geosciences for interpolating environmental variables (Hengl et al., 2018). Examples of the application of machine learning methods in deriving isoscapes are scarce and are related to e.g. bioavailable strontium (Bataille et al., 2020; Funck et al., 2021); on sulfur isotopes in human remains (Bataille et al., 2021); nitrogen and carbon stable isotopes in particulate matter from the Northwest Atlantic Continental Shelf (Oczkowski et al., 2016); and continental studies on the isotopic composition of shallow groundwater (Stahl et al., 2020). An especially relevant work related to the present study is the precipitation isoscape derived by gradient boosting (Nelson et al., 2021).

There have been comparative studies on the performance of different interpolation techniques in relation to mapping geochemical and meteorological data, including stable isotope values. In these comparisons regression kriging (RK) has performed better than basic interpolation techniques such as triangulation or inverse distance weighting (Bowen and Revenaugh, 2003; Li et al., 2011). Moreover, machine learning algorithms seemed to perform similarly (Hengl et al., 2018), or even better than combined geostatistical tools (Bataille et al., 2018; Li et al., 2011; Zandi et al., 2022). Nevertheless, random forest interpolation has not yet been applied to map the spatial structure of precipitation stable isotopes.

With the vast number of possible methods to apply when interpolating environmental variables, benchmarks are needed to be able to decide which particular approach suits best the data at hand. Thus, the present study aims to compare the performance of combined geostatistical tools (i.e., regression kriging) and various machine learning (ML) methods (including the novel Regression Enhanced Random Forest methods) in interpolating the spatial variability of precipitation oxygen stable isotopes ( $\delta^{18}\text{O}$ ) from two different sampling networks in Europe.

## 2. Materials and methods

### 2.1. Used data and preprocessing

In the course of the research the efficiency of interpolation methods were tested on two datasets chosen to provide sufficient data for interpolation and to enable independent validation of results. Daily precipitation stable isotope observations from the United Kingdom (Dataset 1), and monthly data from Germany and its surroundings (Dataset 2) were used.

The daily precipitation stable isotopic data (Dataset 1) was acquired

from three consecutive days between 23 and 25 January 2012 by Tyler et al. (2016) conducted in the frame of the British Isotopes in Rainfall Project (BIRP). BIRP included 67, 68, and 67 stations providing data parallel for 23rd, 24th, and 25th of January, respectively, which is an exceptionally high spatial coverage, although for just a short amount of time. Out of the three days, 23 January was the most variable and 25 the least (Fig. S1).

In the case of Dataset 2, a subdomain of Central Europe was selected covering various physiographic conditions. The backbone of Dataset 2 is the German sampling network of stable water isotopes in precipitation (Stumpp et al., 2014) which had the highest number of active stations ( $n = 30$ ) in 1998. Hence monthly data from this year was chosen for the methodological comparison of the present study. This dataset was extended with additional monitoring stations from the neighboring countries within a 10 km band in order to stretch out the interpolated maps across the entirety of Germany (Fig. 1A). The only exception was Groeningen (NL), 30 km from the German border in the NW. Altogether 10 stations from the Austrian national network (Umweltbundesamt, 2019) and two stations from the Global Network for Isotopes in Precipitation (GNIP; (IAEA, 2019)) were included to improve the spatial coverage for Dataset 2 representing the Central European domain (Fig. 1). Out of the four months the values of January (representing winter) were the most variable and July (representing summer) the least. Median values were highest in the summer (Fig. S1).

To account for seasonal hydrometeorological differences and keep the number of cases assessed from Dataset 1 and 2 comparable, not all of the twelve months, but the central month of every season (January, April, July, October) was considered from Dataset 2 alongside the three days of Dataset 1.

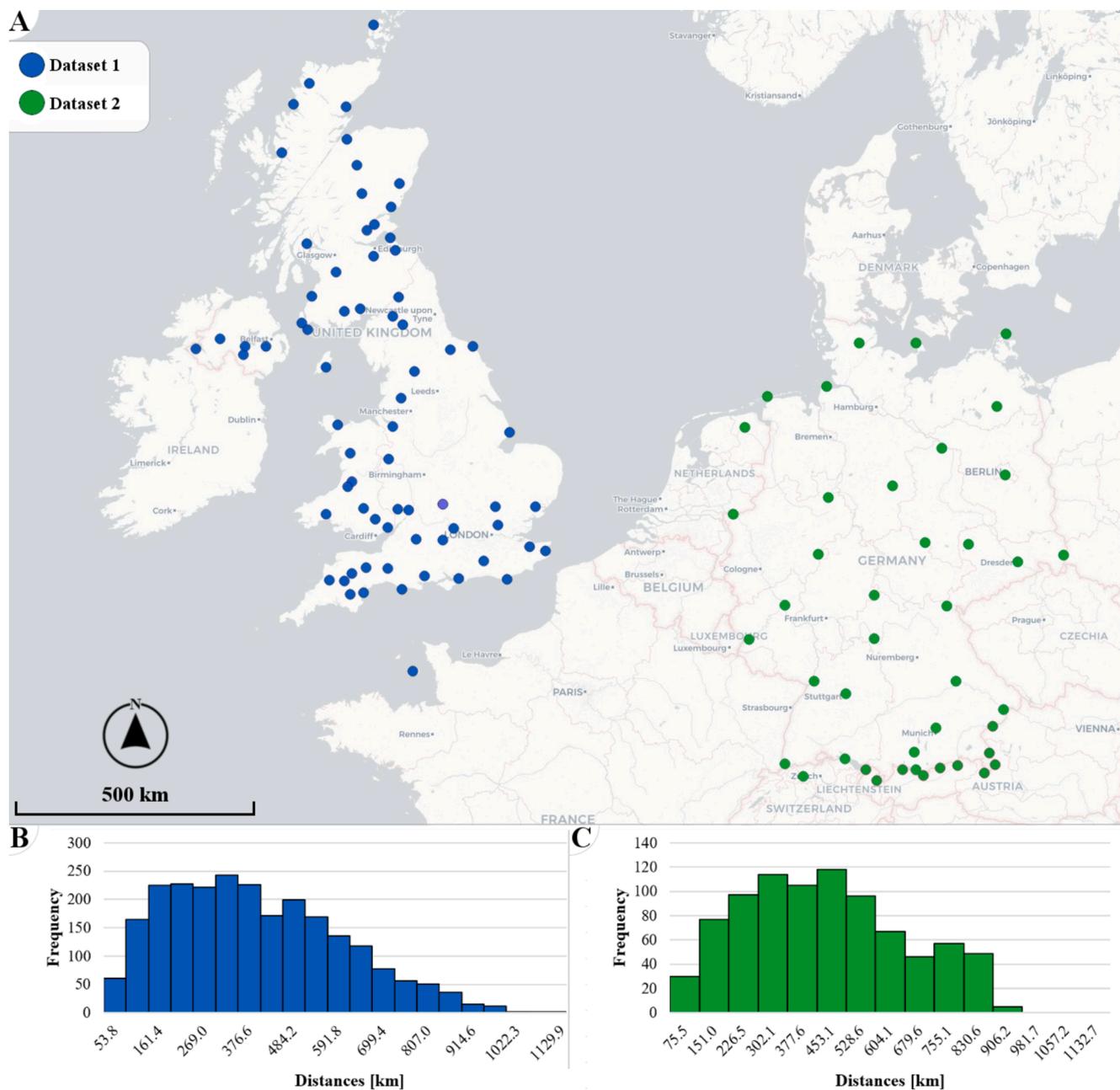
In Dataset 1 the distances between the stations are mostly (~50%) spread between 160 and 430 km (*median* = 339 km), while in Dataset 2 the range is considerably wider (~60% of the distances between 230 and 600 km; *median* = 383 km) reinforcing the visual impression (Fig. 1A) that Dataset 1 is much more densely populated (Fig. 1B) than Dataset 2 (Fig. 1C). Nevertheless, clustering can be observed in both cases to a different extent. In Dataset 1 the eastern and north-eastern parts were scarcely covered with no data on the coastlines, while the southern parts of England and Wales were most densely sampled. In Dataset 2, the high spatial abundance of stations in the south (primarily due to the inclusion of stations from the Austrian national network data) is noteworthy (Fig. 1).

The first step in data preprocessing was to numerically check the  $\delta$  values for potential database errors (e.g. typos, sign errors) (IAEA, 1992). This was done by local indicator of spatial association (LISA) which is used to identify “outliers” as in Anselin’s Moran scatterplot, (Anselin, 1995) and Supplement Sect. 1. It highlighted if nearby values were exceptionally dissimilar, although it is expected that the closer the sites are the more the values measured at them resemble each other (Tobler, 1970).

A second-order stable isotope variable is d-excess defined as  $d = \delta^2\text{H} - 8 \times \delta^{18}\text{O}$  (Dansgaard, 1964). In the present case it is used for detecting outlying values in the dataset. If a monthly  $\delta^{18}\text{O}$  value appeared to be an outlier according to the LISA statistics, the corresponding d-excess value was explored as well, similarly to (Hatvani et al., 2023). For example, very negative d-excess values, occasionally even lower than  $-10\text{‰}$  were interpreted as an indication of evaporative enrichment of the sample (Bowen et al., 2018) and it was discarded from the evaluation if no neighboring stations reported similarly extreme d-excess values for the given month. With this procedure the January value at Goerlitz (Fig. S2A) and the April value at Berlin (Fig. S2B) were discarded leaving January and April with 40, and July & October with 41 data points in the Dataset 2, while Dataset 1 remained unchanged.

### 2.2. Methods applied

After data preprocessing the efficiency of a classical geostatistical



**Fig. 1.** Spatial distribution of precipitation stable isotope monitoring sites used to test the performance of a classical geostatistical approach and machine learning methods (A). Dataset 1: Stations reporting 23rd, 24th, and 25th of January 2012 in the British Isotopes in Rainfall Project (Tyler et al., 2016) Dataset 2: selected GNIP stations (green) ANIP stations (red) active in 1998. Histogram of the distances between the precipitation monitoring stations in Dataset 1 (B) and Dataset 2 (C). The bin width was chosen following Scott (1979). Basemap: © OpenStreetMap contributors. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

method and three additional machine learning tools were explored in interpolating and replicating precipitation stable isotope variability.

Multiple regression kriging (RK; (Hengl et al., 2007)) is one of the most basic and widely used geostatistical interpolation techniques relying on both spatial autocorrelation and covariate information (Sekulić et al., 2020); for details see Sect. 2.2.1. Secondly, a machine learning method: random forest (Breiman, 2001) in regression exercises was applied, which was chosen for its previous application for interpolating isotope data (e.g. Bataille et al. (2018)) and other environmental variables (Hengl et al., 2018). Lastly, machine learning methods combined with regression analysis (i) multiple regression random forest (MRRF; Sect. 2.2.2; (Sekulić et al., 2020)) and (ii) regression enhanced random forest (RERF; Sect. 2.2.2; (Zhang et al., 2019)) were used. For a

summary see Table S1.

The investigations were conducted uniformly in Web Mercator projection (EPSG: 3857) since variography should be performed on a metric scale (Hatvani et al., 2021).

### 2.2.1. Regression kriging

Over large scale distances, trend-like predisposition of stable isotopes in precipitation were documented (Dansgaard, 1964; Rozanski et al., 1993) which could mask the local spatial autocorrelation patterns. Regression kriging (RK) is specifically a tool which accounts for these trend-like processes in its (multiple) regression component with geographic and climatic variables as the independent predictors, and the primary precipitation stable oxygen isotope values as the dependent

variables.

The spatial variability determined by multiple regression used to define an *initial grid*. Second, a *residual grid* is created using universal kriging (Cressie, 1990; Hengl et al., 2007) and its weight function - theoretical semivariograms (Mathéron, 1965) - fitted on to the residuals of the multivariate regression model; finally, the *initial* and *residual grids* are summarized to obtain the final predicted  $\delta^{18}\hat{O}$ .

In RK, the employed multiple regression and geographic variables used to explain stable isotope variance in the regions were:

- latitude, longitude (in Web Mercator EPSG: 3857);
- digital elevation model obtained from Amazon Terrain Tiles (AWS, 2021) downloaded via R using the `elevatr` package v0.3.4, `get_elev_raster` function (Hollister, 2021), setting the zoom (`z`) level to 5, resulting in the effective spatial resolution of  $2.446 \times 2.446$  km. Amazon Terrain Tiles provides open elevation data by aggregating, standardizing, and tiling multiple source datasets into common web mapping and desktop GIS formats and dimensions;
- minimum-, maximum-, and mean temperature; precipitation amount obtained from the ERA5 reanalysis (Copernicus Climate Change Service, 2017). ERA5 provides hourly estimates of a large number of atmospheric, land and oceanic climate variables at a 30 km spatial resolution. Daily and monthly aggregates of the hourly fields are available as well. In the study, the meteorological variables were extracted from the daily aggregates for Dataset 1, and monthly aggregates for Dataset 2;
- geodetic distance from the coast, calculated in R using the spatial coordinates with the `gdistance` function of the `rgeos` package version 0.5–9 (Bivand and Rundel, 2021). Being the dominant wind direction and moisture source pathway (Tyler et al., 2016), it could have been a viable approach to consider the distance from the Atlantic Ocean in the case of Dataset 1. However, this would have violated the general applicability of the present exercise and comparability of the results from Datasets 1 and 2.

The multiple regression model was constructed by backward elimination in a stepwise process (O'Brien, 2007) of the possible predictors considering significance,  $adj. R^2$ , and the variance inflation factor (VIF) applied in numerous similar studies, e.g. (Hatvani et al., 2020; Hatvani et al., 2017). In this study only the  $\delta^{18}\hat{O}$  values at the validation stations' locations are used and not the whole grid. The semivariograms were fitted to the data with least squares fitting in all cases resulting in the fit of a spherical model (Fig. S3).

Out of the possible predictors in Dataset 1, latitude, longitude and minimum temperature were found to be applicable. In Dataset 2 altitude was also a significant predictor in certain months, however it was discarded, considering (i) its strong covariance with temperature, and (ii) a uniform set of environmental variables for both datasets when constructing a multiple regression model assuming that the environmental variables driving precipitation stable isotope composition should be the same in neighboring regions. Yet their relative importance may vary over time/seasonally. In July an alternative version of regression model was derived since in this particular case precipitation was also found to be a meaningful predictor in minimizing the effect of geographical factors influencing the raw  $\delta_p$  records. All obtained models were significant ( $p < 0.093$ ) and VIF was negligible ( $VIF < 4.935$ ), while the average  $adj. R^2$  varied between 0.3 and 0.5 peaking at 0.37 in the case of Dataset 1 (Fig. S4A) and between 0.1 and 0.8 in a quite equally distributed way in the case of Dataset 2 (Fig. S4B).

### 2.2.2. Random forest

Random forest (RF) is a nonparametric method well-known for solving prediction problems. Its predictions in regression problems are based on the average results of the random decision trees which use bootstrap sampling (bagging) to eliminate the possibility of over-fitting.

Biau and Scornet (2016); Breiman (2001); Prasad et al. (2006). In RF, each tree is built using a subset of the original data set, which is used to derive the tree partition and to make the tree prediction (Biau and Scornet, 2016). Once the observations are selected, a recursive partitioning is performed of the covariates space. In each cell, a number `mtry` of variables are selected uniformly at random among all covariates (Fig. S5; Scornet (2017)). Then, the procedure selects the split minimizing the quadratic risk of the tree estimate at each step using the out-of-bag observations to evaluate the trained algorithm (Breiman, 1996; Prasad et al., 2006), being repeated until each cell contains less than a prespecified number, `nodesize`, of observations. After tree partition has been completed, the prediction at a new point is computed by averaging observations falling into the cell of the new point. Then each one of the trees in the forest gives a prediction, and the forest prediction is simply the average of the predicted values. Therefore, the key tuning parameters of the RF algorithm are `mtry` and `node.size` (Biau and Scornet, 2016; Breiman, 2001; Scornet, 2017). The specific parametrization of the random forest methods used here are shown in Fig. S6.

A weakness of the algorithm is that the prediction is in the form of a weighted average of responses in the training set, and the results can only be obtained within the range of the response variable, resulting in prediction bias (Zhang and Lu, 2012). Specifically, the bias can be quite large in the extrapolation problems we typically encounter in spatial analysis. To get around the problem of extrapolation, RF is combined with regression analysis (Zhang et al., 2019). In the present study LASSO regression combined with RF, called Regression Enhanced Random Forest (RERF) (Zhang et al., 2019), and common multivariate regression combined with RF, called multiple regression random forest (MRRF) are employed.

RERF is a semiparametric prediction approach. Predictions of the RERF are derived from the sum of the LASSO regression and the RF prediction for the LASSO regression residual. LASSO regression employs a coefficient estimate derived from a penalized least squares estimator as opposed to the linear regression model's standard least squares estimator. For a given tuning parameter, the L1-penalization of LASSO regression enables variable selection for a large number of covariates (Tibshirani, 1997), excluding those predictors that resulted in a coefficient equal to zero (Fig. S6: LASSO regression-right column), thereby improving prediction, or one can provide the set of predictors the LASSO regression can rank. However, LASSO regression assumes the linear regression model, which has a potentially limited structure, limiting its prediction ability to data with a complex structure. In other words, bias can be anticipated when applying only LASSO regression. Therefore, in the RERF approach, the random forest method is applied to the residual of the LASSO regression to correct the possible bias (Zhang et al., 2019). Each method of LASSO regression and RF method has distinct tuning parameters, which certainly affect both LASSO regression and RF performance. Therefore, selecting a tuning parameter combination that reduces the overall prediction error would be reasonable. In this instance, a tuning parameter was chosen using cross-validation, a reliable approach for estimating the prediction error (Zhang et al., 2019). Note that a grid search algorithm was applied when choosing the combination of tuning parameters. Using RERF higher prediction performance is expected (smaller errors) than with ordinary RF, especially in the case where extrapolation is needed. In MRRF the same multiple regression model was applied to complement the RF as with RK, the predictors were manually chosen.

### 2.3. Validation statistics

For validation purposes 15% of the available data were randomly retained while ensuring that these test stations are not closer to each other than 50 km to avoid clustering of the validation points. This random selection was repeated ten times for each day investigated from Dataset 1 and each month from Dataset 2 to provide ensembles to better depict the performance of the compared methods.

In the validation process ordinary and hybrid error metrics were used to explore the performance of the different methods in replicating the actual measured  $\delta^{18}\text{O}$  values. Two ordinary metrics, the mean standard error (MSE; Eq. (1)) and the relative mean absolute error (RMAE) were calculated, out of which the former is highly suggested to be included in such studies to make it comparable to others (Li and Heap, 2008).

$$MSE = \frac{1}{N} \sum_{i=1}^N (\hat{\delta}_i - \delta_i)^2 \quad (1)$$

where  $N$  is the number of data points,  $\hat{\delta}_i$  the value returned by the model and  $\delta_i$  the actual value for data point  $i$ . To complement this metric and make the results comparable with studies in which the primary variables are in a different measurement unit or resolution RMAE was also calculated (Supplement Sect. 2, Fig. S7) due to its scale independency (Li and Heap, 2008). Lastly the newly developed non-parametric version of the Kling-Gupta efficiency (KGE) was used (Pool et al., 2018) as a hybrid error metric, since ordinary KGE implicitly assumes data normality and the absence of outliers (Clark et al., 2021). The KGE can take values between  $-\infty$  and 1, with 1 indicating a perfect agreement between the simulations and observations and the lower the KGE gets, the less accurate the estimation is compared to the observations (Gupta et al., 2009; Knoben et al., 2019).

To test how sensitive the prediction performance under reduced data availability, the training data in both datasets were randomly rarefied one-by-one. In Dataset 1, 46 stations were excluded this way, corresponding to  $\sim 79\%$  of the training set, while in Dataset 2, 27 stations were left out at the end corresponding similarly to  $\sim 79\%$  of the training set. The models were recalculated using the reduced training sets and the results were compared to the previously used validation stations (see above). Due to high demand in computational resources in the case of RERF, the rarefaction was carried out at a lower resolution, i.e. pairs of stations were removed in one step. For even more specifics the reader is referred to the used R code Supplement section 4.

## 2.4. Software used

Variography was conducted using the `gstat` package, version 2.0–9 (Pebesma, 2004). The LASSO regression was calculated using the `glmnet` package (Friedman et al., 2010) and the RF, MRRF, RERF models are grown using 500 decision trees each and the specific parameter values for e.g. `node.size` and `mtry` were determined by the `tune` function of `randomForestSRC` package, version 3.1.1 (Ishwaran et al., 2021; Ishwaran et al., 2008) in R (R Core Team, 2019).

## 3. Results and discussion

### 3.1. Importance of predictors

The importance of the predictors was assessed for all models: RF, RK & MRRF, and RERF. It should be noted that no uniform metric yet exists, therefore, the comparison is done using the relative importance of the predictors for each approach for both datasets. The following metrics were applied: variable importance for RF (Fig S6: left column), adjusted  $R^2$  for RK & MRRF (Fig. S6: middle column), and normalized LASSO regression coefficient using the Agresti and Coull (1998) method for RERF (Fig. S6: right column).

The geographical position of the stations (X, Y coordinates) are generally considered to be important predictors (Bowen and Wilkinson, 2002; Dansgaard, 1964; Rozanski et al., 1993) and indeed in the case of Dataset 1 are among the strongest first two predictors in the case of the linear (Fig. S6B, E, H) and among the strongest four in the case of the non-linear approaches (Fig. S6: Dataset 1 left and right columns). On the contrary, precipitation seems unimportant in the case of the linear models (Fig. S6: middle column and sect. 2.2.1.). In the meanwhile precipitation leads the set of predictors regarding their importance in

the case of RF and RERF (Fig. S6A, C) on 23 Jan, and can also be considered meaningful on 25 January (Fig. S6G, I) suggesting a non-linear relationship with  $\delta^{18}\text{O}$  in precipitation.  $T_{\min}$  was the leading variable in the case of RK and MRRF (Fig. S6B) on 23 January and taken as important on 24 January (Fig. S6E). It was also among the most important predictors in the case of all approaches on 25 January (Fig. S6G-I). In the case of Dataset 2, we first focus on altitude which is traditionally an important predictor in the region (Kern et al., 2014; Siegenthaler and Oeschger, 1980) it takes the leading role in the non-linear methods. July is a special case (see Sect. 2.2.1) in which both the nonlinear approaches (e.g. Fig. S6P, S) and even multiple regression (Fig. S6R) considered precipitation as an important predictor.

### 3.2. Comparison of the interpolation schemes

The MSE values ranged from 2.43 to 6.36 and for the KGE values stood above 0.7 without any remarkable differences considering the RK, RF, MRRF, and RERF methods in case of the daily data (Dataset 1, Fig. 2A, C). In case of the Dataset 2, the MSE values ranged from 0.52 to 2.69 (Fig. 2B). Relatively larger errors occurred for January and October, but generally smaller errors were seen compared to Dataset 1. The KGE values for Dataset 2 (Fig. 2B, D) were usually also above 0.7, the exceptions are the traditional methods (RK) in July considering both variants.

The performance of the two ML approaches (MRRF and RERF) was comparable, with MRRF producing somewhat smaller MSE (Fig. 2A, B) and bigger nonparametric KGE values (Fig. 2C, D) closer to one indicating better predictions. In contrast, the standard RF method produced the highest MSE among all methods in Dataset 1 peaking on 23 January,  $MSE = 6.4$ . This was the date when the data range was greatest (Fig. S1) (Hashimoto et al., 2019).

The error metrics calculated between the measured and predicted sample values are usually utilized to evaluate the performance of interpolation methods (Burrough et al., 1998). MSE (employed here) fundamentally provides estimates of the average error but does not

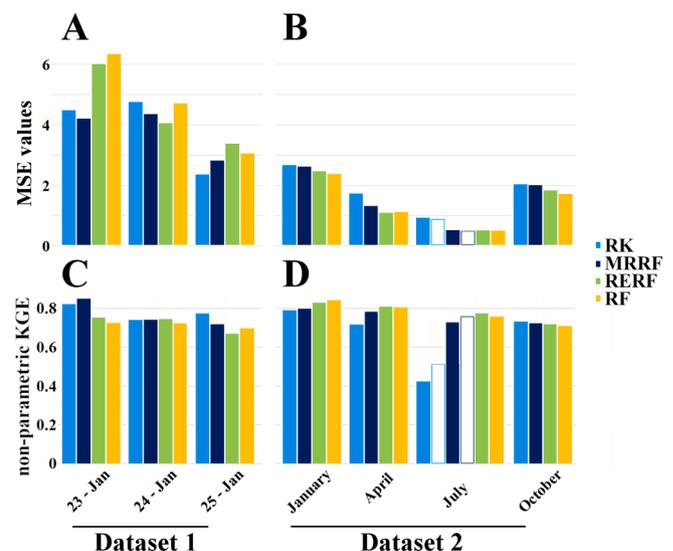


Fig. 2. Mean Squared Errors (MSE; A, B) and the non-parametric version of the Kling-Gupta efficiency (KGE; C, D) obtained during the validation of the interpolated products with the different methods for Dataset 1 (A, C) and Dataset 2 (B, D). The empty bars in Dataset 2 represent the RK (light blue) and MRRF (dark blue) model MSEs and non-parametric KGE's, where precipitation served as an independent variable during the multiple regression. Acronyms in the legend, mean RK: Regression Kriging, MRRF: multiple regression random forest, RERF: Regression Enhanced Random Forest and RF: Random Forest. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

provide information about the relative size of the average difference and the nature of differences (Willmott, 1982). Therefore, the absolute differences between the measured and predicted values of  $\delta^{18}\text{O}$  were calculated for each validation-ensemble member (Sect. 2.3) and compiled for the observed time points in Dataset 1 (Fig. 3A) and Dataset 2 (Fig. 3B) to get a detailed picture on the error distribution.

The distribution of the absolute differences (Fig. 3) follows a similar pattern as the MSE (Fig. 2A, B). The ML methods gave a smaller interquartile interval of absolute errors than RK in Dataset 2 (Fig. 3B) in every month, and two out of three days in Dataset 1 (Fig. 3A). The median of absolute differences obtained with the ML methods combined with regression analysis were smaller median  $\sim 1.6$  ‰ for MRRF  $\sim 1.9$  ‰ for RERF than the ones obtained with the common RF  $\sim 2.1$  ‰ for RF in the case of Dataset 1, especially MRRF on 23 January in Dataset 1. However, in the case of Dataset 2, the ML methods (especially RERF) performed better than the geostatistical interpolation in all cases considering the median values (Fig. 3B).

The inability of RF to extrapolate is a well-known problem (Hashimoto et al., 2019) There was a significant difference in the MSE values of the two datasets (Fig. 2A, B), with the range of Dataset 1 being significantly greater than that of Dataset 2 (Fig. S1). In Dataset 2 the smaller range of errors may be due to the better spatial distribution of the station data (Fig. 1). The advantage of RF application over RK is that it can manifest itself in quasi-automated interpolation of environmental isotopic parameters, especially where the point sampling is representative (extrapolation minimized) (Hengl et al., 2018).

In the future, the precision of the ML estimations could be further increased by including the buffer distances between the grid cells and the station data as a predictor - e.g. in Hengl et al. (2018) -, or by even adding a weight function to these predictors based on their distances (Sekulić et al., 2020). It should be noted that one of the main advantages of ML methods over classical geostatistical ones is that these require less knowledge about the spatial autocorrelation structure of the data which is a major disadvantage at the same time (Hengl et al., 2007). On the other hand, these ML tools such as the most recent one applied in predicting water  $\delta^{18}\text{O}$  are not able to assess the spatial autocorrelation

structure of the monitoring network (Nelson et al., 2021), where the data originates from, unlike the previously mentioned geostatistical tools; e.g. Hatvani et al. (2021).

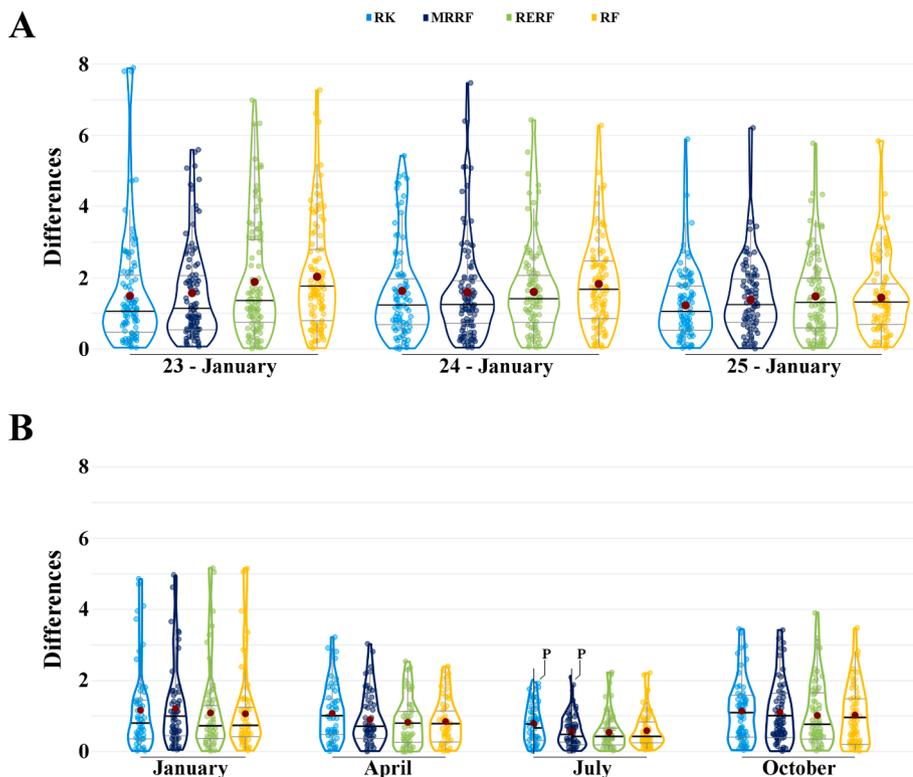
### 3.3. Model sensitivity to decreased station density

As a last step, the decrease in prediction potential was investigated with respect to artificial rarefaction of the monitoring stations (see Sect. 2.3 for details). In the case of Dataset 1, the RERF and RF methods gave the highest MSE values most of the time on 23 and 25 January, while on 24 January, the RERF predicted the best results up to a station density reduced by  $\sim 50\%$  (Fig. 4). The performance of MRRF on 24 and 25 January sustained its initial performance up to a substantial rarefaction of the station data; ca. 50–60%. The MSE in the case of RK method indicates a similar pattern as MRRF on 25 January, however, increases to a steeper extent on 23 and 24 January (Fig. 4A) resembling the pattern of RF, but with lower values.

For Dataset 2, RK performed the worst in July whether the precipitation was included or excluded (Fig. 4B). In all other cases, it maintains good performance up to a station density reduced by 50–60%. RF and MRRF performed similarly well tolerating a reduction in input data to about 50% without any drop in performance. Afterward the error attributed to the interpolation performance begins to increase with a much steeper gradient for RF, while MRRF indicates a much more moderate increase. Moreover, the initial good performance of MRRF is sustained in January, April, and October up to even  $\sim 80\%$  station density reduction (Fig. 4B). In the case of RERF, its prediction performance was sustained up to  $\sim 50\%$  and  $\sim 80\%$  of station reduction for January and October, respectively.

Regarding RK, it should be noted that the underlying theoretical semivariograms did not change to a great extent depending on the degree of rarefaction Fig. S3 regarding their main characteristics and types (Cressie, 1993). This highlights the robustness of the spatial autocorrelation structure against rarefaction, implying that the pattern seen prevails relatively independent of the chosen subset of sites.

Overall, MRRF seems to better reproduce the spatial variability of the



**Fig. 3.** Absolute differences of the measured and predicted values of  $\delta^{18}\text{O}$  using the different methods on Dataset 1 (A) and Dataset 2 (B). The outline of the violin plot covers the probability density of the differences (points inside), the grey horizontal lines indicate the interquartile range (IQR). The data points are not aligned on a vertical line in order to make the figure less crowded. Two upright lines represent the data within the 1.5 IQR. The data between 1.5 times the IQR are indicated with a circle (outliers) as in the case of a box-and-whiskers plot (Kovács et al., 2012). In July, the violin plots for RK and MRRF are not horizontally symmetric, the right half (marked with “P”) represents the model in which precipitation was used in the multiple regression as an additional predictor. RK: Regression Kriging, MRRF: multiple regression random forest, RERF: Regression Enhanced Random Forest and RF: Random Forest.

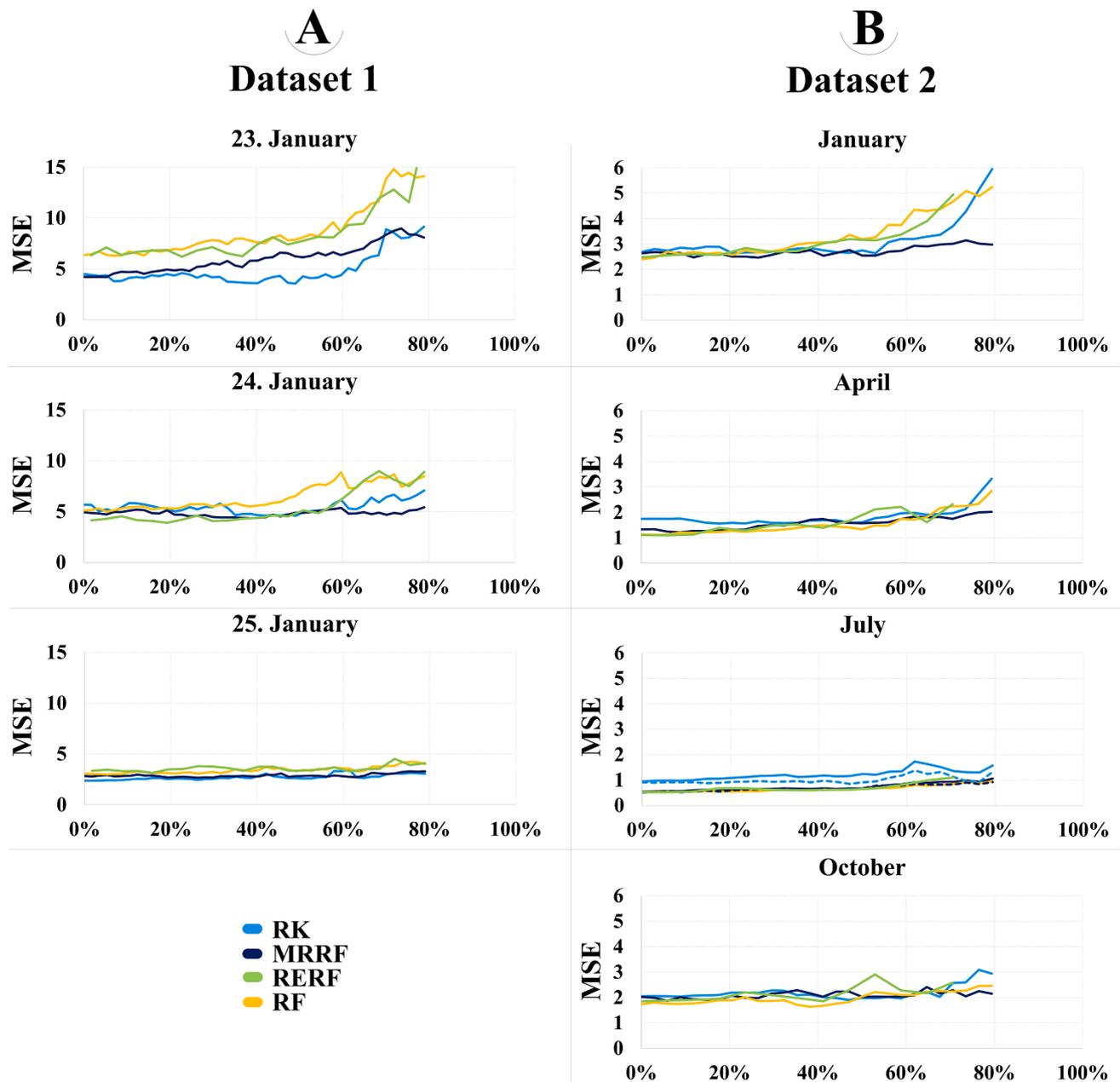


Fig. 4. The change in mean squared error (MSE) as a function of percentage of stations removed for Dataset 1 (A) and Dataset 2 (B) for the different interpolation methods; RK: Regression Kriging, MRRF: Multiple Regression Random Forest, RERF: Regression Enhanced Random Forest, RF: Random Forest. The dashed lines in Dataset 2 – July (RK -light blue; MRRF - dark blue) represent the interpolation model, where precipitation served as an independent variable during the multiple regression. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

two datasets even under circumstances with more than 50% of the sites not providing information. Similar conclusions were recently drawn regarding the robustness and less sensitivity of ML models towards stations density and their less dependency on expert knowledge, comparing their performances with linear models in high-resolution monthly precipitation interpolation (Zandi et al., 2022).

#### 4. Conclusions and outlook

In this study, the performance of four interpolation approaches was compared using two datasets of precipitation  $\delta^{18}\text{O}$  to understand how the interpolation schemes behave under various circumstances of spatial distribution/organization in order to determine which interpolation approach to utilize for producing isoscapes of different scales.

The results indicate that machine learning algorithms are more than capable of creating continuous datasets of precipitation stable isotopes of even superior quality than the ‘classical’ geostatistical tools, which were formerly considered as the gold standard in this discipline. In particular, machine learning methods tolerated the rarefaction of the dataset to a much greater extent especially above 50% of the data removed, indicating that they can be employed with the expectation of producing comparably accurate predictions if one or more stations become inoperable or in regions where monitoring stations are sparser or unevenly distributed. Moreover, machine learning tools require less “expert knowledge” in handling the predictors, are less sensitive to outliers and no stationarity assumption is needed to be followed, nor is there a need to deal with anisotropy or fitting variograms (Bataille et al., 2018; Hengl et al., 2018) as in the case of regression kriging. On the

other hand, unlike regression kriging, machine learning algorithms do not provide overall information about the spatial autocorrelation structure of precipitation  $\delta^{18}\text{O}$  in the studied regions, nor is it possible to estimate the representativity of the assessed monitoring network.

Based on the results, the application of a random forest algorithm is highly viable option for deriving isoscapes of subcontinental/continental extent, especially if the practical advantages of machine learning tools over their geostatistical counterparts are considered. Equipping these tools to take into account additional complex predictors for instance buffer distances (Hengl et al., 2018) or weight functions derived from  $\delta_p$  values (e.g. Sekulić et al. (2020)) is expected to further improve their performance.

## Funding

This research was supported by the ÚNKP-20-3 New National Excellence Program of the Ministry for Innovation and Technology from the source of the National Research, Development and Innovation Fund.

## Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Data availability

Publicly available datasets were analyzed in this study, which can be found at: <https://nucleus.iaea.org/wiser/index.aspx>, the Supplementary Material 2 of Tyler et al. (2016) and in the papers cited; see Sect. 2.1 for further details.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jhydrol.2023.129129>.

## References

- Agresti, A., Coull, B.A., 1998. Approximate is better than "Exact" for interval estimation of binomial proportions. *Am. Stat.* 52 (2), 119–126. <https://doi.org/10.2307/2685469>.
- Amundson, R., Chadwick, O., Kendall, C., Wang, Y., DeNiro, M., 1996. Isotopic evidence for shifts in atmospheric circulation patterns during the late Quaternary in mid-North America. *Geology* 24 (1), 23–26.
- Anselin, L., 1995. Local Indicators of Spatial Association—LISA. *Geogr. Anal.* 27 (2), 93–115. <https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>.
- AWS, A.W.S., 2021. Terrain Tiles. In: Amazon (Ed.). Amazon, Amazon.
- Bataille, C.P., et al., 2018. A bioavailable strontium isotope for Western Europe: A machine learning approach. *PLoS One* 13 (5), e0197386.
- Bataille, C.P., et al., 2021. Triple sulfur-oxygen-strontium isotopes probabilistic geographic assignment of archaeological remains using a novel sulfur isotope of western Europe. *PLoS One* 16 (5), e0250383.
- Bataille, C.P., Crowley, B.E., Wooller, M.J., Bowen, G.J., 2020. Advances in global bioavailable strontium isotopes. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 555, 109849 <https://doi.org/10.1016/j.palaeo.2020.109849>.
- Biau, G., Scornet, E., 2016. A random forest guided tour. *TEST* 25 (2), 197–227. <https://doi.org/10.1007/s11749-016-0481-7>.
- Birkel, C., et al., 2018. Characterization of surface water isotope spatial patterns of Scotland. *J. Geochem. Explor.* 194, 71–80. <https://doi.org/10.1016/j.gexplo.2018.07.011>.
- Bivand, R., Rundel, C., 2021. rgeos: Interface to Geometry Engine - Open Source ("GEOS"). R package version 0.5-9. DOI:10.5281/zenodo.5809645.
- Bowen, G.J., et al., 2018. Inferring the source of evaporated waters using stable H and O isotopes. *Oecologia* 187 (4), 1025–1039. <https://doi.org/10.1007/s00442-018-4192-5>.
- Bowen, G.J., Good, S.P., 2015. Incorporating water isoscapes in hydrological and water resource investigations. *WIREs Water* 2 (2), 107–119. <https://doi.org/10.1002/wat2.1069>.
- Bowen, G.J., Revenaugh, J., 2003. Interpolating the isotopic composition of modern meteoric precipitation. *Water Resour. Res.* 39 (10), 1299. <https://doi.org/10.1029/2003WR002086>.
- Bowen, G.J., Wilkinson, B., 2002. Spatial distribution of  $\delta^{18}\text{O}$  in meteoric precipitation. *Geology* 30 (4), 315–318. [https://doi.org/10.1130/0091-7613\(2002\)030<0315:sdooim>2.0.co;2](https://doi.org/10.1130/0091-7613(2002)030<0315:sdooim>2.0.co;2).
- Bowen, G.J., 2010. Isoscapes: Spatial Pattern in Isotopic Biogeochemistry. *Annual Review of Earth and Planetary Sciences*, 38(1): 161–187. DOI:doi:10.1146/annurev-earth-040809-152429.
- Breiman, L., 1996. Out-of-Bag Estimation. Department of Statistics, University of California.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45 (1), 5–32. <https://doi.org/10.1023/A:1010933404324>.
- Burrough, P.A., McDonnell, R.A., McDonnell, R., Lloyd, C.D., 1998. Principles of geographical information systems. Oxford University Press, Oxford, UK.
- Chan, W.-P., et al., 2012. Regional scale high resolution  $\delta^{18}\text{O}$  prediction in precipitation using MODIS EVI. *PLoS One* 7 (9), e45496.
- Clark, M.P. et al., 2021. The Abuse of Popular Performance Metrics in Hydrologic Modeling. *Water Resources Research*, 57(9): e2020WR029001. DOI:https://doi.org/10.1029/2020WR029001.
- Clark, I.D., Fritz, P., 1997. *Environmental Isotopes in Hydrogeology*. Taylor & Francis.
- Copernicus Climate Change Service (C3S), 2017. ERA5: Fifth generation of ECMWF atmospheric reanalyses of the global climate. In: (CDS), C.C.C.S.C.D.S. (Ed.), <https://cds.climate.copernicus.eu/cdsapp#!/home>.
- Coplen, T.B., 1994. Reporting of stable hydrogen, carbon and oxygen isotopic abundances. *Pure App. Chem.* 66, 273–276.
- Coplen, T.B., Herczeg, A.L., Barnes, C., 2000. Isotope Engineering—Using Stable Isotopes of the Water Molecule to Solve Practical Problems. In: Cook, P.G., Herczeg, A.L. (Eds.), *Environmental Tracers in Subsurface Hydrology*. Springer, US, Boston, MA, pp. 79–110. [https://doi.org/10.1007/978-1-4615-4557-6\\_3](https://doi.org/10.1007/978-1-4615-4557-6_3).
- Cressie, N., 1990. The origins of kriging. *Math Geol* 22 (3), 239–252. <https://doi.org/10.1007/BF00889887>.
- Cressie, N.A.C., 1993. *Statistics for Spatial Data*. John Wiley & Sons Inc, p. 900.
- Dansgaard, W., 1964. Stable isotopes in precipitation. *Tellus* 16, 436–468.
- Ehleringer, J.R., et al., 2008. Hydrogen and oxygen isotope ratios in human hair are related to geography. *Proc. Natl. Acad. Sci.* 105 (8), 2788–2793. <https://doi.org/10.1073/pnas.0712228105>.
- Fórizs, I., 2003. Isotopes as natural tracers in the watercycle: Examples from the Carpathian Basin. *Studia UBB Physica* 1 (48), 69–77.
- Friedman, J.H., Hastie, T., Tibshirani, R., 2010. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* 33 (1), 1–22. <https://doi.org/10.18637/jss.v033.i01>.
- Funck, J., Bataille, C., Rasic, J., Wooller, M., 2021. A bio-available strontium isotope for eastern Beringia: a tool for tracking landscape use of Pleistocene megafauna. *J. Quat. Sci.* 36 (1), 76–90. <https://doi.org/10.1002/jqs.3262>.
- Gibson, J.J., Edwards, T.W.D., 2002. Regional water balance trends and evaporation-transpiration partitioning from a stable isotope survey of lakes in northern Canada. *Global Biogeochemical Cycles*, 16(2): 10-1-10-14. DOI:https://doi.org/10.1029/2001GB001839.
- Gupta, H.V., Kling, H., Yilmaz, K.K., Martinez, G.F., 2009. Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *J. Hydrol.* 377 (1), 80–91. <https://doi.org/10.1016/j.jhydrol.2009.08.003>.
- Hashimoto, H., et al., 2019. High-resolution mapping of daily climate variables by aggregating multiple spatial data sets with the random forest algorithm over the conterminous United States. *Int. J. Climatol.* 39 (6), 2964–2983. <https://doi.org/10.1002/joc.5995>.
- Hatvani, I.G., et al., 2021. Geostatistical evaluation of the design of the precipitation stable isotope monitoring network for Slovenia and Hungary. *Environ. Int.* 146, 106263 <https://doi.org/10.1016/j.envint.2020.106263>.
- Hatvani, I.G., Leuenberger, M., Kohán, B., Kern, Z., 2017. Geostatistical analysis and isotope of ice core derived water stable isotope records in an Antarctic macro region. *Polar Sci.* 13, 23–32. <https://doi.org/10.1016/j.polar.2017.04.001>.
- Hatvani, I.G., Erdélyi, D., Vreča, P., Kern, Z., 2020. Analysis of the Spatial Distribution of Stable Oxygen and Hydrogen Isotopes in Precipitation Across the Iberian Peninsula. *Water* 12 (2), 481. <https://doi.org/10.3390/w12020481>.
- Hatvani, I.G., Smati, A.E., Erdélyi, D., Szatmári, G., Vreča, P., Kern, Z., 2023. Modeling the spatial distribution of the meteoric water line of modern precipitation across the broader Mediterranean region. *J. Hydrol.* 617 (Part B), 128925. <https://doi.org/10.1016/j.jhydrol.2022.128925>.
- Heaton, K., Kelly, S.D., Hoogewerff, J., Woolfe, M., 2008. Verifying the geographical origin of beef: The application of multi-element isotope and trace element analysis. *Food Chem.* 107 (1), 506–515. <https://doi.org/10.1016/j.foodchem.2007.08.010>.
- Hengl, T., Heuvelink, G.B.M., Rossiter, D.G., 2007. About regression-kriging: From equations to case studies. *Comput. Geosci.* 33 (10), 1301–1315. <https://doi.org/10.1016/j.cageo.2007.05.001>.
- Hengl, T., Nussbaum, M., Wright, M.N., Heuvelink, G.B., Gräler, B., 2018. Random forest as a generic framework for predictive modeling of spatial and spatio-temporal variables. *PeerJ* 6, e5518.
- Hobson, K.A., 1999. Tracing origins and migration of wildlife using stable isotopes: a review. *Oecologia* 120 (3), 314–326. <https://doi.org/10.1007/s004420050865>.
- Hobson, K.A., Wassenaar, L.I., 1996. Linking breeding and wintering grounds of neotropical migrant songbirds using stable hydrogen isotopic analysis of feathers. *Oecologia* 109 (1), 142–148. <https://doi.org/10.1007/s004420050068>.
- Ehleringer, J.R. et al., 2008. Hydrogen and oxygen isotope ratios in human hair are related to geography. *Proceedings of the National Academy of Sciences*, 105(8): 2788–2793. DOI:10.1073/pnas.0712228105.
- IAEA, 1992. *Statistical treatment of data on environmental isotopes in precipitation*. In: Technical Report Series International Atomic Energy Agency, Vienna, p. 781place.

- IAEA, 2019. Global Network of Isotopes in Precipitation. The GNIP Database.
- Ishwaran, H., Kogalur, U.B., Blackstone, E.H., Lauer, M.S., 2008. Random survival forests. *The Annals of Applied Statistics*, 2(3): 841–860, 20.
- Ishwaran, H., Kogalur, U., Kogalur, M., 2021. RandomForestSRC: Fast Unified Random Forests for Survival, Regression, and Classification (RF-SRC)[(accessed on 16 July 2020)].
- Kaseke, K.F., Wang, L., Wanke, H., Turewicz, V., Koeniger, P., 2016. An analysis of precipitation isotope distributions across Namibia using historical data. *PLoS One* 11 (5), e0154598.
- Kern, Z., Kohán, B., Leuenberger, M., 2014. Precipitation isoscape of high reliefs: interpolation scheme designed and tested for monthly resolved precipitation oxygen isotope records of an Alpine domain. *Atmos. Chem. Phys.* 14 (4), 1897–1907. <https://doi.org/10.5194/acp-14-1897-2014>.
- Knoben, W.J.M., Freer, J.E., Woods, R.A., 2019. Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. *Hydrol. Earth Syst. Sci.* 23 (10), 4323–4331. <https://doi.org/10.5194/hess-23-4323-2019>.
- Kovács, J., Tanos, P., Korponai, J., Kovácsné Székely, I., Gondár, K., Gondár-Sőregi, K., Hatvani, I.G., 2012. Analysis of water quality data for scientists. In: Voudouris, Kostas, Voutsas, D. (Eds.), *Water Quality Monitoring and Assessment*. InTech, pp. 65–94. <https://doi.org/10.5772/32173>.
- Li, J., Heap, A.D., 2008. A review of spatial interpolation methods for environmental scientists.
- Li, J., Heap, A.D., Potter, A., Daniell, J.J., 2011. Application of machine learning methods to spatial interpolation of environmental variables. *Environ. Model. Softw.* 26 (12), 1647–1659. <https://doi.org/10.1016/j.envsoft.2011.07.004>.
- Lykoudis, S.P., Argiriou, A.A., 2007. Gridded data set of the stable isotopic composition of precipitation over the eastern and central Mediterranean. *J. Geophys. Res. Atmos.* 112 (D18) <https://doi.org/10.1029/2007jd008472>.
- Matheron, G., 1965. *Les Variables régionalisées et leur estimation: une application de la théorie des fonctions aléatoires aux sciences de la nature*. Masson Et Cie Lusaing-Chartres, Impr. Durand, Paris.
- Nan, Y., Tian, F., Hu, H., Wang, L., Zhao, S., 2019. Stable Isotope Composition of River Waters across the World. *Water* 11 (9), 1760.
- Nelson, D.B., Basler, D., Kahmen, A., 2021. Precipitation isotope time series predictions from machine learning applied in Europe. e2024107118 *Proc. Natl. Acad. Sci.* 118 (26). <https://doi.org/10.1073/pnas.2024107118>.
- O'Brien, R.M., 2007. A caution regarding rules of thumb for variance inflation factors. *Qual. Quant.* 41 (5), 673–690. <https://doi.org/10.1007/s11335-006-9018-6>.
- Oczkowski, A., Kreakie, B., McKinney, R.A., Prezioso, J., 2016. Patterns in stable isotope values of nitrogen and carbon in particulate matter from the northwest Atlantic Continental Shelf, from the Gulf of Maine to Cape Hatteras. *Front. Marine Sci.* 3 (252) <https://doi.org/10.3389/fmars.2016.00252>.
- Pebesma, E.J., 2004. Multivariable geostatistics in S: the gstat package. *Comput. Geosci.* 30 (7), 683–691. <https://doi.org/10.1016/j.cageo.2004.03.012>.
- Pool, S., Vis, M., Seibert, J., 2018. Evaluating model performance: towards a non-parametric variant of the Kling-Gupta efficiency. *Hydrol. Sci. J.* 63 (13–14), 1941–1953. <https://doi.org/10.1080/02626667.2018.1552002>.
- Prasad, A.M., Iverson, L.R., Liaw, A., 2006. Newer classification and regression tree techniques: bagging and random forests for ecological prediction. *Ecosystems* 9 (2), 181–199. <https://doi.org/10.1007/s10021-005-0054-1>.
- R Core Team, 2019. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Reichstein, M., et al., 2019. Deep learning and process understanding for data-driven Earth system science. *Nature* 566 (7743), 195–204. <https://doi.org/10.1038/s41586-019-0912-1>.
- Rozanski, K., Araguás-Araguás, L., Gonfiantini, R., 1993. Isotopic patterns in modern global precipitation. In: Swart, P.K., Lohmann, K.C., Mckenzie, J., Savin, S. (Eds.), *Climate Change in Continental Isotopic Records*. American Geophysical Union, USA, pp. 1–36. <https://doi.org/10.1029/GM078p0001>.
- Scornet, E., 2017. Tuning parameters in random forests. *ESAIM: Procs* 60, 144–162.
- Scott, D.W., 1979. On optimal and data-based histograms. *Biometrika* 66 (3), 605–610. <https://doi.org/10.1093/biomet/66.3.605>.
- Sekulić, A., Kilibarda, M., Heuvelink, G.B.M., Nikolić, M., Bajat, B., 2020. Random forest spatial interpolation. *Remote Sensing* 12 (10), 1687.
- Siegenthaler, U., Oeschger, H., 1980. Correlation of 18O in precipitation with temperature and altitude. *Nature* 285 (5763), 314–317. <https://doi.org/10.1038/285314a0>.
- Stahl, M.O., Gehring, J., Jameel, Y., 2020. Isotopic variation in groundwater across the conterminous United States – Insight into hydrologic processes. *Hydrol. Process.* 34 (16), 3506–3523. <https://doi.org/10.1002/hyp.13832>.
- Stumpp, C., Klaus, J., Stichler, W., 2014. Analysis of long-term stable isotopic composition in German precipitation. *J. Hydrol.* 517, 351–361. <https://doi.org/10.1016/j.jhydrol.2014.05.034>.
- Terzer, S., Wassenaar, L.I., Araguás-Araguás, L.J., Aggarwal, P.K., 2013. Global isoscapes for  $\delta^{18}\text{O}$  and  $\delta^2\text{H}$  in precipitation: improved prediction using regionalized climatic regression models. *Hydrol. Earth Syst. Sci.* 17 (11), 4713–4728. <https://doi.org/10.5194/hess-17-4713-2013>.
- Tibshirani, R., 1997. The lasso method for variable selection in the cox model. *Stat. Med.* 16 (4), 385–395. [https://doi.org/10.1002/\(SICI\)1097-0258\(19970228\)16:4<385::AID-SIM380>3.0.CO;2-3](https://doi.org/10.1002/(SICI)1097-0258(19970228)16:4<385::AID-SIM380>3.0.CO;2-3).
- Tobler, W.R., 1970. A computer movie simulating urban growth in the detroit region. *Econ. Geogr.* 46 (sup1), 234–240. <https://doi.org/10.2307/143141>.
- Tyler, J.J., et al., 2022. Hydrological and Isotopic Variability of Perched Wetlands on North Stradbroke Island (Minjerribah), Australia: Implications for Understanding the Effects of Past and Future Climate Change. *Front. Environ. Sci.* 10 <https://doi.org/10.3389/fenvs.2022.868114>.
- Tyler, J.J., Jones, M., Arrowsmith, C., Allott, T., Leng, M.J., 2016. Spatial patterns in the oxygen isotope composition of daily rainfall in the British Isles. *Clim. Dyn.* 47 (5), 1971–1987. <https://doi.org/10.1007/s00382-015-2945-y>.
- Umweltbundesamt, H.O.F., 2019. In: Bundesministerium für Land-und Forstwirtschaft, U.u.W.B. (Ed.).
- van der Veer, G., Voerkelius, S., Lorentz, G., Heiss, G., Hoogewerff, J.A., 2009. Spatial interpolation of the deuterium and oxygen-18 composition of global precipitation using temperature as ancillary variable. *J. Geochem. Explor.* 101 (2), 175–184. <https://doi.org/10.1016/j.gexplo.2008.06.008>.
- Willmott, C.J., 1982. Some comments on the evaluation of model performance. *Bull. Am. Meteorol. Soc.* 63 (11), 1309–1313. [https://doi.org/10.1175/1520-0477\(1982\)063<1309:SCOTEO>2.0.CO;2](https://doi.org/10.1175/1520-0477(1982)063<1309:SCOTEO>2.0.CO;2).
- Yoshimura, K., 2015. Stable Water Isotopes in Climatology, Meteorology, and Hydrology: A Review. *Journal of the Meteorological Society of Japan. Ser. II*, 93(5), 513–533. <https://doi.org/10.2151/jmsj.2015-036>.
- Zandi, O., Zahraie, B., Nasser, M., Behrangi, A., 2022. Stacking machine learning models versus a locally weighted linear model to generate high-resolution monthly precipitation over a topographically complex area. *Atmos. Res.* 272, 106159 <https://doi.org/10.1016/j.atmosres.2022.106159>.
- Zhang, H., Nettleton, D., Zhu, Z., 2019. Regression-enhanced random forests. arXiv preprint arXiv:1904.10416.
- Zhang, G., Lu, Y., 2012. Bias-corrected random forests in regression. *J. Appl. Stat.* 39 (1), 151–160. <https://doi.org/10.1080/02664763.2011.578621>.