

How Much Information Is Needed for Reliable Groundwater Level Imputation? Evidence from Drought Conditions in Sweden

Master's thesis in Infrastructure and Environmental Engineering

Yang Zhang

DEPARTMENT OF Architecture and Civil Engineering

Division of Geology and Geotechnics

CHALMERS UNIVERSITY OF TECHNOLOGY

Gothenburg, Sweden 2026

www.chalmers.se

Master's thesis 2026

How Much Information Is Needed for Reliable Groundwater Level Imputation? Evidence from Drought Conditions in Sweden

Yang Zhang



CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Architecture and Civil Engineering
Division of Geology and Geotechnics
CHALMERS UNIVERSITY OF TECHNOLOGY
Göteborg, Sweden, 2026

How Much Information Is Needed for Reliable Groundwater Level Imputation? Evidence from Drought Conditions in Sweden
YANG ZHANG

© YANG ZHANG, 2026.

Supervisor: Ezra Haaf, Markus Giese (University of Gothenburg)

Examiner: Lars Rosen

Master’s thesis 2026

Department of Architecture and Civil Engineering
Division of Geology and Geotechnics
Chalmers University of Technology
SE-412 96 Göteborg
Sweden
Telephone + 46 79-323 3141

Printed by Chalmers

Göteborg, Sweden 2026

How Much Information Is Needed for Reliable Groundwater Level Imputation? Evidence from Drought Conditions in Sweden

YANG ZHANG
Department of Architecture and Civil Engineering
Chalmers University of Technology

Abstract

Reconstructing groundwater levels during drought is important for assessing water availability. This study investigated the information required for groundwater-level imputation under drought conditions in Sweden. Daily records from 297 monitoring wells were analysed, with drought periods identified using the Standardized Groundwater Index (SGI) and removed to create artificial gaps. Predictors included nearby groundwater-level records, rolling sums of net precipitation (NP), and time variables representing day of year, month and continuous time. Two methods, missForest and ARCHI, were evaluated using RMSE, MAE and Pearson correlation. Configuration comparisons used 147 eligible wells for missForest and 210 for ARCHI, with a separate comparison on 140 common wells.

Nearby groundwater-level records were the most important source of information for missForest. NP and time variables provided complementary benefits when reference-well information was limited. Within the eligible sample, approximately 3 to 5 reference wells were sufficient to achieve near-optimal performance, depending on the metric and tolerance, with limited further improvements beyond about 8. Near-optimal performance was defined using 5% and 10% tolerances relative to the best tested configuration. Pearson correlation metric satisfied these criteria with fewer reference wells. On the 140 common wells, missForest W5 (Top10 + NP1–3 + Time) achieved 35.9% lower median RMSE and 39.6% lower median MAE than ARCHI n8, together with higher median correlation.

Overall, combining reference-well observations with climatic and time information improved reconstruction performance. Minimum information requirements remained dependent on the evaluation sample, metric and tolerance definition. The descriptive SGI analysis characterised drought variability, while preservation of drought-event characteristics after imputation remains to be evaluated.

Keywords: groundwater level time series; groundwater level imputation; minimum information analysis; drought; SGI; missForest; ARCHI package

Acknowledgements

As my two years of master's studies gradually come to an end, I would like to take this opportunity to express my sincere gratitude to all the teachers, classmates, friends, and family members whose guidance, encouragement, and companionship made the completion of this thesis possible.

First and foremost, I would like to thank my supervisor, Dr. Ezra Haaf, for his guidance and advice throughout the process of selecting the thesis topic, designing the study, analysing the data, and revising the manuscript. In particular, his support in shaping the research direction of this thesis was invaluable. I would also like to express my sincere gratitude to my other supervisor, Dr. Markus Giese from the University of Gothenburg. I am very grateful for his support in providing the data, developing the research direction of the thesis, and offering valuable guidance and feedback during the revision process.

I would like to thank my examiner, Professor Lars Rosen, for his many valuable and constructive comments and suggestions. His feedback helped me identify important parts for improvement in the thesis and also deepened my understanding of academic research.

I am also grateful to my classmates, Yunlong Zhao and Haozhe Sun, who served as opponents during my thesis defence. Their thoughtful comments and questions helped me address several weaknesses in the thesis and further improve the final version.

I would especially like to thank my girlfriend, Leyao Zhang. Although you have been thousands of kilometres away for most of the past two years, you have always shared your life with me and made the distance between us feel much smaller. Thank you for travelling all the way from China to Sweden on your own in August and spending the lonely summer with me. I believe there will be many more happy days like that summer waiting for us in the future.

I would also like to thank my friends, Han Zhou, Wen Zhou, and Jiyang Fan. We became close friends from my very first day in Sweden, and I am grateful for your friendship and companionship throughout these years. I wish all of you every success in your future careers.

My deepest gratitude also goes to my family, especially my mother, grandfather and grandmother. Without your support, I would never have had the opportunity to study far away from home and experience a much broader world. Thank you for your understanding, encouragement, and unwavering support.

Finally, I would like to thank everyone who helped me during my time studying in Sweden. Your kindness made this sometimes difficult world a little brighter.

Contents

Abstract I
Acknowledgements II
1. Introduction 1
1.1 Background 1
1.2 Aim 3
1.3 Specification of research questions 3
2. Materials and methods 4
2.1 Study region 4
2.2Workflow 5
2.3 Data description 7
2.3.1 Data source 7
2.3.2 Data preprocessing 8
2.4 Designing gap patterns 8
2.4.1 Identification of drought period 8
2.4.2 Introducing artificial gaps 10
2.5 Variables construction 11
2.5.1 Time variables 11
2.5.2 Climatic variables 11
2.5.3 Groundwater levels from reference wells 12
2.6 Imputation methods and evaluation 13
2.6.1 The missForest algorithm 13
2.6.2 ARCHI package 13
2.6.3 Model performance metrics 14
2.6.4 NP variable importance analysis 15
3. Results 16
3.1 Groundwater drought characteristics 16
3.2 The missForest algorithm 16
3.2.1 Lerum local experiment 17
3.2.2 Global experiment 20
3.3 ARCHI package 27
3.4 Comparison of 2 models on common target wells 30
3.5 Minimum information analysis 31
3.5.1 The missForest algorithm 31
3.5.2 ARCHI package 33
4. Discussion 35

4.1 Imputation methods 35
4.2 Imputation variables 35
4.3 Spatial variation of imputation performance 37
4.4 Limitations and suggestions for future studies 40

5. Conclusions 41

References 42

List of figures

Fig. 1. (a) Köppen-Geiger climate classes information from Kottek et al. (2006), Rubel et al. (2017). (b) Aquifer lithology and bedrock information from Geological Survey of Sweden (2024a). (c) Soil type information from (Geological Survey of Sweden, 2024c). 4

Fig. 2. Workflow to impute groundwater time series and evaluate imputation performance 7

Fig 3. 307 groundwater monitoring wells in Sweden 8

Fig. 4. Groundwater drought event identification and merging 10

Fig. 5. Distribution of observed and artificially masked groundwater level data across the 297 wells 10

Fig. 6. Spatial and temporal patterns of groundwater drought across 297 monitoring wells during 1990–2023. (a) SGI heatmap. Red indicates dry conditions, blue indicates wet conditions, and grey indicates missing data. (b) Monthly percentage of wells experiencing drought, together with the number of available wells. 11

Fig. 7. Groundwater drought event characteristics across 297 monitoring wells during 1990–2023. Colours indicate (a) peak intensity, defined as the maximum monthly deficit below the SGI threshold of −1, and (b) cumulative severity, defined as the sum of monthly deficits below this threshold. 16

Fig. 8. Effects of different variables on missForest performance in Lerum. (a) Time variables; (b) NP accumulation windows; (c) groundwater level variables; (d) number of nearby wells. 18

Fig. 9. Median RMSE and MAE for different combinations of nearby wells and NP rolling-sum windows, with and without time variables. (a) Median RMSE without time variables; (b) median RMSE with time variables; (c) median MAE without time variables; and (d) median MAE with time variables. 19

Fig. 10. SHAP value of NP rolling sum windows in the NP-only model 20

Fig. 11. RMSE distributions for different configurations in the global missForest experiment: (a) baseline configurations; (b) main experiment; (c) fixed number of nearby wells; (d) fixed NP accumulation window; (e) combined configurations; and (f) configurations with NP12. 23

Fig. 12. MAE distributions for different configurations in the global missForest experiment: (a) baseline configurations; (b) main experiment; (c) fixed number of nearby wells; (d) fixed NP accumulation window; (e) combined configurations; and (f) configurations with NP12. 24

Fig. 13. Pearson r distributions for different configurations in the global missForest experiment: (a) baseline configurations; (b) main experiment; (c) fixed number of nearby wells; (d) fixed NP accumulation window; (e) combined configurations; and (f) configurations with NP12. 25

Fig. 14. Effects of reference-well number and NP accumulation windows on missForest performance across 147 target wells. (a) Median RMSE, MAE and Pearson r as the number of reference wells increases, with NP1–3 and time variables fixed. (b) Median RMSE and MAE for different NP input sets, with and without 5 reference wells. (c) Median Pearson r for different NP input sets, with and without 5 reference wells. 27

Fig. 15. RMSE and MAE distributions of ARCHI imputation results. 28

Fig. 16. Pearson r distributions of ARCHI imputation results. 29

Fig. 17. The number of reference wells used when “n_refwl_choice = 10”. 30

Fig. 18. Minimum number of nearby wells needed for near optimal RMSE. (a) Lowest median RMSE obtained for each number of nearby wells; (b) changes in the required number of nearby wells under different NP and time variables. 31

Fig. 19. Minimum number of nearby wells needed for near optimal MAE. (a) Lowest median MAE obtained for each number of nearby wells; (b) changes in the required number of nearby wells under different NP and time variables. 32

Fig. 20. Minimum number of nearby wells needed for near optimal Pearson r. (a) Highest median Pearson r obtained for each number of nearby wells; (b) changes in the required number of nearby wells under different NP and time variables. 33

Fig. 21. Comparison of missForest and ARCHI. (a) Median RMSE; (b) median MAE; (c) Pearson r. 34

Fig. 22. Spatial distribution of imputation errors across Sweden: (a) median RMSE and (b) median MAE 37

Fig. 23. Engineering and agricultural activities surrounding high error wells (a) Sigtuna_2 and (b) Motala_71. Location and satellite maps from OpenStreetMap and Google Earth. 39

Fig. 24. Historical deforestation activities surrounding Ronnerby_39 between 2009 and 2024. Location and satellite maps from OpenStreetMap and Google Earth. 39

List of tables

Tab. 1 Classification of groundwater drought 9

Tab. 2 31 global experiment configurations 21

Tab. 3 Number of eligible target wells by minimum required number of reference wells. 22

Tab. 4 Median RMSE, MAE and Pearson r for the 31 global missForest configurations. 26

Tab. 5 Median RMSE, MAE and Pearson r for ARCHI with different maximum numbers of reference wells 29

Tab. 6 Comparison of missForest W5 and ARCHI n8 on 140 common target wells. 31

1. Introduction

1.1 Background

Groundwater provides drinking water, irrigation supply, and many other essential uses for billions of people, making it the largest accessible freshwater resource on earth (Shiklomanov, 2000). In Sweden, approximately 80% of municipalities rely on groundwater as their local drinking water source and there are more than 500,000 private wells serving over 1.2 million local residents each year (Barthel et al., 2021; Eveborn et al., 2017). However, Swedish groundwater resources are confined to numerous small aquifers, so insufficient recharge can quickly threaten water supply, as was clearly demonstrated during the groundwater droughts in 2017 and 2018 (Bakke et al., 2020; Barthel et al., 2021; Stensen et al., 2019). Therefore, it is necessary and critical to have sustainable management of regional groundwater resources. Groundwater level data are a key basis for such management, supporting the long-term sustainability of water resources(Barthel et al., 2021). However, groundwater level records may contain missing observations, interruptions and uneven sampling intervals. These gaps significantly affect the accuracy of analyses of trends, extreme events, and the impacts of climate variability on groundwater systems. Consequently, the development and evaluation of imputation methods for reconstructing groundwater level time series have become an emerging and important research topic in hydrogeology. It is essential to use reliable imputation techniques when reconstructing incomplete groundwater-level time series historical records.

Currently, there is no standardized method for estimating missing groundwater level values (Dwivedi et al., 2022; Rajaee et al., 2019) , and a variety of imputation strategies have been suggested. The simplest approach is to remove time series containing missing values. However, this can substantially reduce the size of datasets and lead to a potential loss of valuable information (Haaf & Barthel, 2018; Nygren, Giese, et al., 2020; Retike et al., 2022; Wendt et al., 2020), making it unsuitable for realistic applications. Linear interpolation is commonly used for short-term gap filling and often performs well for short gaps in large datasets due to its simplicity and ease of implementation (Asgharinia & Petroselli, 2020; Bikše et al., 2023). However, its performance deteriorates when applied to longer periods of missing data (Meggiorin et al., 2023).

Advances in computational techniques have led to the increased application of artificial intelligence and machine learning methods in groundwater level modelling, interpolation and prediction in worldwide scenarios (Bikše et al., 2023; Dwivedi et al., 2022; Evans et al., 2020; Rajaee et al., 2019; Sharma et al., 2025; Solanki & Mishra, 2026; Taie Semiromi & Koch, 2019). To further improve imputation accuracy, additional auxiliary data are often integrated with groundwater level observation data, such as data from neighbouring monitoring wells (Dwivedi et al., 2022; Ramirez et al., 2023; Sejine et al., 2026; Taie Semiromi & Koch, 2019), Earth

observation data (Evans et al., 2020), climate data (Ramirez et al., 2022; Schoppa et al., 2020) and satellite remote sensing data (Chucuya et al., 2025; Haaf et al., 2023; Noori & Singh, 2021; Ramirez et al., 2022).

In recent years, groundwater level imputation methods have evolved toward more specialized, interpretable, and flexible integrated models. For example, Collenteur developed the open-source Python software Pastas, which is based on transfer function–noise (TFN) models. Several studies have demonstrated that Pastas performs well in reconstructing long-term gaps in groundwater level time series (Collenteur et al., 2019; Collenteur et al., 2025; Nair et al., 2025). Another approach is the ARCHI (Automated Regional Correlation Analysis for Hydrologic Record Imputation) package developed by Levy et al. (2025), which performs automatic imputation based on linear regression relationships between target wells and reference wells. ARCHI has demonstrated high computational efficiency for large-scale groundwater datasets and achieves reconstruction accuracy comparable to or even better than missForest in two regional datasets (Levy et al., 2025).

However, most existing studies mainly rely on introducing random artificial missing values and comparing the imputed results with the original observations (Arriagada et al., 2021; Chucuya et al., 2025; Dwivedi et al., 2022; Garciarena & Santana, 2017; Yadav & Roychoudhury, 2018; Zhou et al., 2023). While some studies have attempted to simulate more realistic missing-data patterns using cluster analysis (Bikše et al., 2023; Sejine et al., 2026; Solanki & Mishra, 2026), research on the reconstruction performance of groundwater level imputation methods under real extreme conditions, particularly drought scenarios, remains limited.

Meanwhile, previous studies have primarily focused on specific catchments or individual basins and have not sufficiently considered a wider range of climatic conditions and geomorphological controls (Arriagada et al., 2021; Sidibe et al., 2018; Zhou et al., 2023). Consequently, models validated in a single region may not necessarily generalize to basins with different climatic or hydrogeological conditions, and their cross-regional transferability remains to be systematically evaluated. Such transferability may be further challenged in areas strongly affected by human activities(Solanki & Mishra, 2026; Zhou et al., 2023).

Although previous studies have demonstrated that auxiliary variables can improve groundwater level imputation performance, little attention has been given to how much information is required to achieve reliable results. This question is particularly relevant for large scale monitoring networks, where the availability of reference wells and climatic observations may vary considerably between locations. Including more reference wells or climatic information may improve model performance, but it also increases data requirements and may introduce redundant or weakly relevant information. It therefore remains unclear at what point additional information provides only marginal improvements, and whether climatic and time variables can compensate for limited groundwater observations from nearby wells. To address this gap, this

study focuses on identifying the minimum information required for reliable groundwater level time series imputation during drought periods in Sweden.

1.2 Aim

The primary aim of this study is to determine the minimum information required for reliable groundwater level time series imputation during drought periods in Sweden. Specifically, the study quantifies the individual and combined contributions of three types of input information: groundwater level data from correlated nearby wells, net precipitation accumulated over different time windows, and time variables representing seasonality and long-term trends.

This study also aims to evaluate the performance of MissForest and ARCHI algorithms for groundwater level imputation on a national scale. Finally, the study aims to provide practical guidance on balancing imputation accuracy, data availability and computational efficiency in large scale groundwater modelling applications.

1.3 Specification of research questions

- How do MissForest and ARCHI perform in reconstructing groundwater level time series during drought-induced missing periods in Sweden?
- How do groundwater level data, climatic and time variables affect the imputation accuracy of missForest, and at what point do additional inputs provide only marginal improvements?
- What are the minimum variables required for missForest and ARCHI package to achieve performance within 5% and 10% of the best-performing tested configuration, and to what extent can net precipitation and time variables reduce this requirement?

2. Materials and methods

2.1 Study region

Sweden is located on the Scandinavian Peninsula in northern Europe, extending approximately from 55° to 69°N and from 10° to 24°E. According to the Köppen–Geiger climate classification(Kottek et al., 2006), northern Sweden has subarctic climate (Dfc) and arctic tundra climate (ET) , parts of central Sweden have a warm-summer humid continental climate (Dfb), and southern Sweden has a temperate oceanic climate (Cfb) (Fig. 1a). Across Sweden, the median annual mean temperature ranges from −4.4 to 9.1°C, while the median annual total precipitation ranges from 363 to 1188 mm/year (1980-2024, E-OBS data set: (Cornes et al., 2018)).

Sweden is a part of the Fennoscandian Shield, and its bedrock consists predominantly of ancient crystalline and metamorphic rocks. Precambrian crystalline rocks, including granite, granodiorite, and gneiss, underlie most of the country, whereas sedimentary rock formations occur locally in central and southern Sweden(Fig. 1b)(Fredén et al., 2009; Geological Survey of Sweden, 2024a).

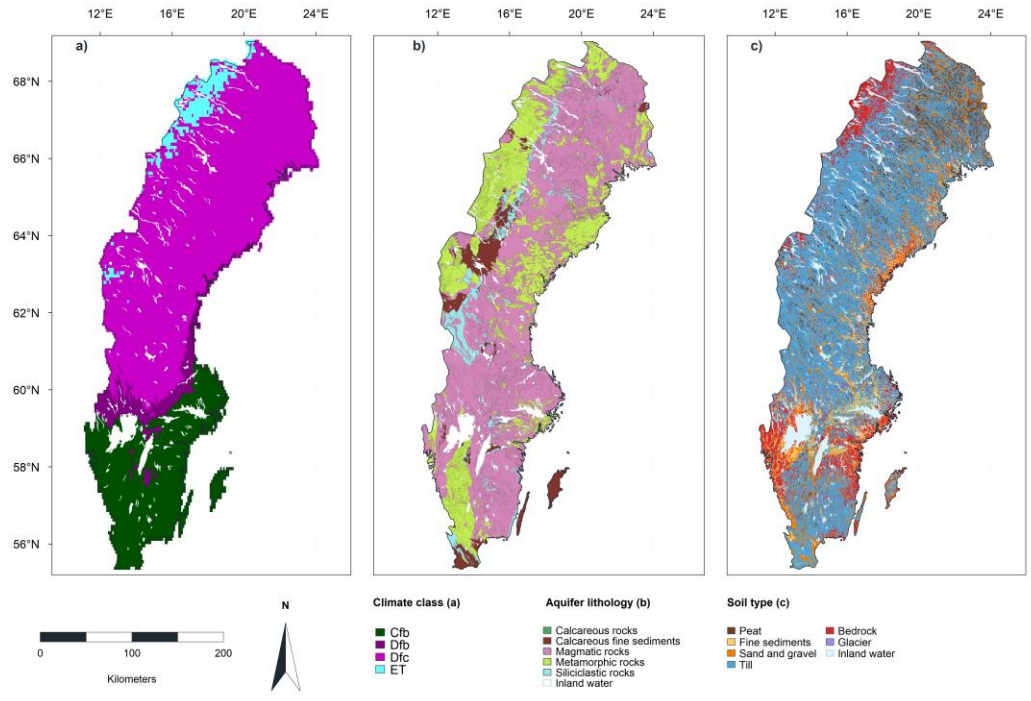


Fig. 1. (a) Köppen-Geiger climate classes information from Kottek et al. (2006), Rubel et al. (2017). (b) Aquifer lithology and bedrock information from Geological Survey of Sweden (2024a). (c) Soil type information from (Geological Survey of Sweden, 2024c).

Groundwater in Sweden occurs mainly in two types of aquifers. The first type of aquifer occurs within unconsolidated Quaternary deposits, such as sandy and gravelly glaciofluvial sediments.

Typical landforms include eskers and end moraines, which host some of Sweden's larger and more productive groundwater aquifers. The second type are fractured rock aquifers, where groundwater is primarily stored within fracture systems in bedrock such as granite, pegmatite, limestone, and sandstone (Englund et al., 1986). Prolonged glacial activity shaped Sweden's relatively smooth bedrock surface and removed much of the weathering layer (Karin et al., 2015; Karna et al., 2013; Kleman et al., 2008). In the underlying crystalline bedrock, groundwater flow is concentrated within fractures, while aquifer properties are strongly controlled by the density, length, connectivity, and hydraulic conductance of these fractures. Consequently, the hydraulic properties of bedrock aquifers in Sweden may vary considerably over short distances (Giese & Barthel, 2021; Richts et al., 2011).

Groundwater in Sweden is recharged mainly through the infiltration of rainfall and snowmelt. The recharge process is jointly controlled by precipitation, temperature, snow accumulation and melting, evapotranspiration, vegetation, soil freezing conditions, and the permeability of subsurface materials (Barthel et al., 2021; Nygren, Markus, et al., 2020). Groundwater recharge processes also show regional and seasonal variability. In southern Sweden, winter temperatures are relatively mild, and a greater proportion of precipitation falls as rain. Combined with low evapotranspiration, these conditions make autumn and winter the principal groundwater recharge periods. During the growing season in late spring and summer, increased plant transpiration and surface evaporation substantially reduce groundwater recharge, resulting in declining groundwater levels (Nygren, Markus, et al., 2020).

By contrast, winter precipitation in central and northern Sweden is stored as snow and groundwater recharge is limited during winter, and groundwater levels generally decline (Du et al., 2019). As temperatures rise in spring, soil thawing and snowmelt generate a concentrated recharge event, causing groundwater levels to increase rapidly (Barthel et al., 2021; Geological Survey of Sweden, 2024b; Nygren, Markus, et al., 2020). Although summer precipitation may be more abundant, strong evapotranspiration and vegetation interception considerably reduce the amount of water that supplies the groundwater. Coniferous forests in Sweden may intercept approximately 20–40% of summer precipitation (Grip & Rodhe, 2000). As a result, groundwater recharge during the growing season may be very low or even close to zero in some areas (Barthel et al., 2021). Groundwater drought is therefore more likely to occur during summer (Geological Survey of Sweden, 2024b).

2.2 Workflow

To reach the research objectives while considering both imputation performance and data completeness, the workflow of this study (Fig. 2) consisted of the following steps:

1. **Data collection and preprocessing.** Groundwater level time series data, Standardized Groundwater Index (SGI), and well coordinate data were obtained from the European

groundwater drought evaluation dataset (EDGE-1, (Nair et al.)), while E-OBS climate data were acquired from the Copernicus Climate Change Service. During preprocessing, wells with insufficient data or anomalous groundwater level fluctuations (Unreasonable, drastic fluctuations and an excessive disparity in long-term average water levels etc.) were excluded through visual inspection.

2. **Artificial gap design and variables construction.**

   2.1 Groundwater drought events were identified using SGI. Groundwater level observations corresponding to the identified drought periods were then removed to introduce artificial gaps.

   2.2 Meteorological variables were extracted at each well location based on its coordinates, and potential evapotranspiration (PET) and net precipitation (NP) were calculated. Groundwater level time series from nearby wells were also selected using Spearman cross-correlation analysis. These variables were subsequently used as variables for imputation.

3. **Groundwater-level imputation.** The artificially introduced gaps were imputed using the missForest algorithm and the ARCHI package.

4. **Performance assessment.** The imputed values were compared with the original groundwater level observations. The performance of the two imputation methods was evaluated using the root mean square error (RMSE), mean absolute error (MAE) and Pearson correlation coefficient.

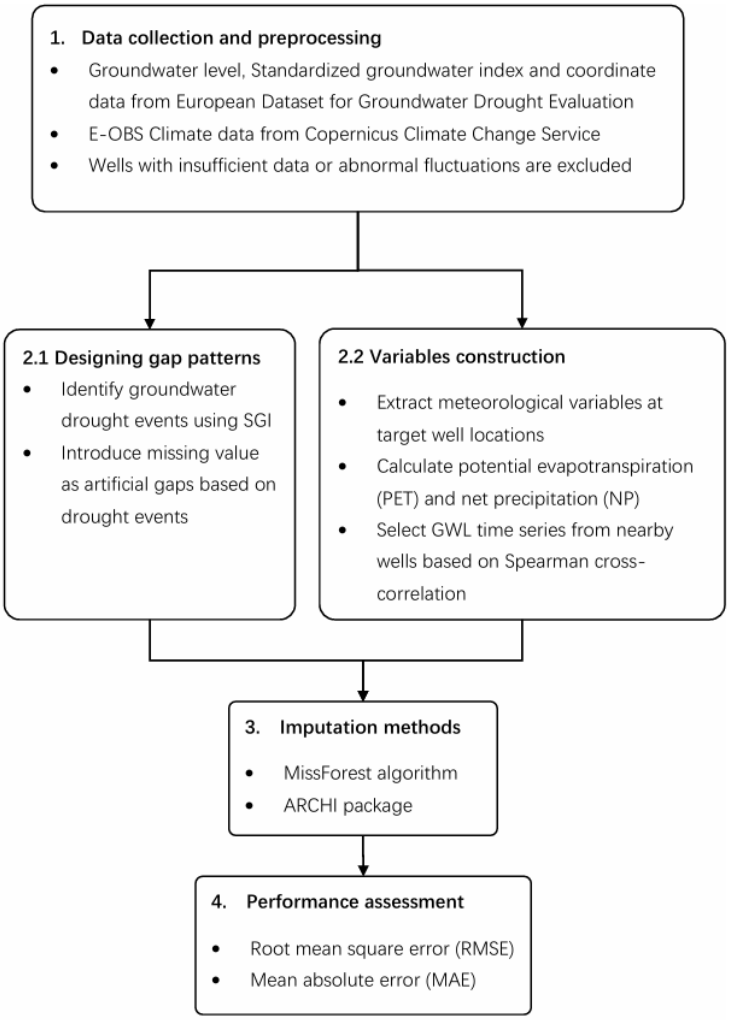


Fig. 2. Workflow to impute groundwater time series and evaluate imputation performance

2.3 Data description

2.3.1 Data source

This study used meteorological, groundwater level, well coordinate, and SGI data for imputation. The meteorological data was obtained from the E-OBS European daily gridded meteorological dataset provided through the Climate Data Store (CDS) of the Copernicus Climate Change Service (C3S). The dataset includes daily precipitation, mean air temperature, minimum and maximum air temperature, relative humidity, and wind speed (Cornes et al., 2018). Groundwater level, well coordinate, and SGI data were obtained from the European groundwater drought evaluation dataset (EDGE-1, (Nair et al.)) published on the Open Science Framework platform (Nair et al., 2026). The dataset has 313 monitoring wells with observations period from 1990 to 2023.

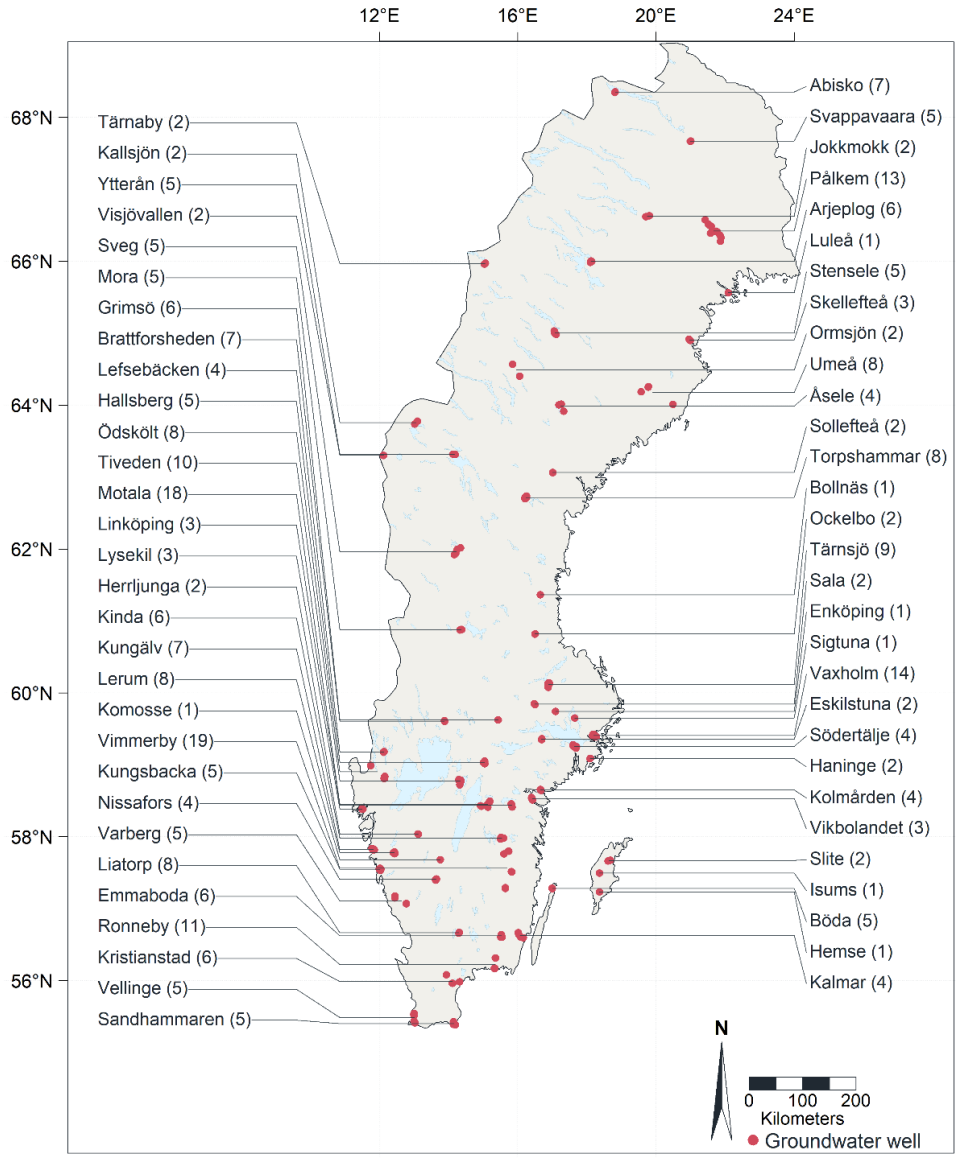


Fig 3. 307 groundwater monitoring wells in Sweden

2.3.2 Data preprocessing

Among the 313 wells, 16 were excluded through visual inspection. 6 wells because of insufficient observation data (Malgomaj_41, Malgomaj_44, Malgomaj_45, Sollefteå_2, Ströms_Vattudal_22, and Vaxholm_6) and 10 wells because of anomalous groundwater level fluctuations (Brattforsheden_3, Kolmården_3, Lysekil_4, Motala_31, Ormsjön_11, Ronneby_101, Torpshammar_24, Torpshammar_28, Vellinge_8, and Vimmerby_54). The excluded wells were used neither as target wells nor as reference wells in the imputation models. Consequently, the remaining 297 wells were retained for the groundwater level imputation.

2.4 Designing gap patterns

2.4.1 Identification of drought period

Various indices have been developed to assess meteorological drought. However, groundwater responses to climatic changes are often nonlinear and time lagging. Therefore, the Standardized Groundwater Index (SGI) (Bloomfield & Marchant, 2013) is more appropriate for assessing groundwater drought events. The SGI has been widely used to identify groundwater drought events, characterize their properties, and examine their relationships with other types of drought (Babre et al., 2022; Guo et al., 2021). The SGI classification is presented in Tab. 1.

Tab. 1 Classification of groundwater drought

Drought grades	SGI
No drought	>0
Mild drought	-1~0
Moderate drought	-1.5~ -1
Severe drought	-2~ -1.5
Extreme drought	⩽-2

Based on the SGI classification, drought periods were identified using the following procedure (Fig. 4):

1. **Drought definition.** The drought period was set when SGI was below −1.
2. **Preliminary drought event identification.** Each drought period during which SGI remained below −1 was initially identified as a drought event. In the example, six preliminary drought events, labelled a, b, c, d, e and f were identified.
3. **Merging adjacent drought events.** If two adjacent drought events, each lasting at least two months, were separated by only one or two months (e.g. events c and d), and the SGI remained below 1 throughout the intervening period, the two events were merged into a single drought event (e.g. event g). Otherwise, they were treated as independent drought events (e.g. events e and f).

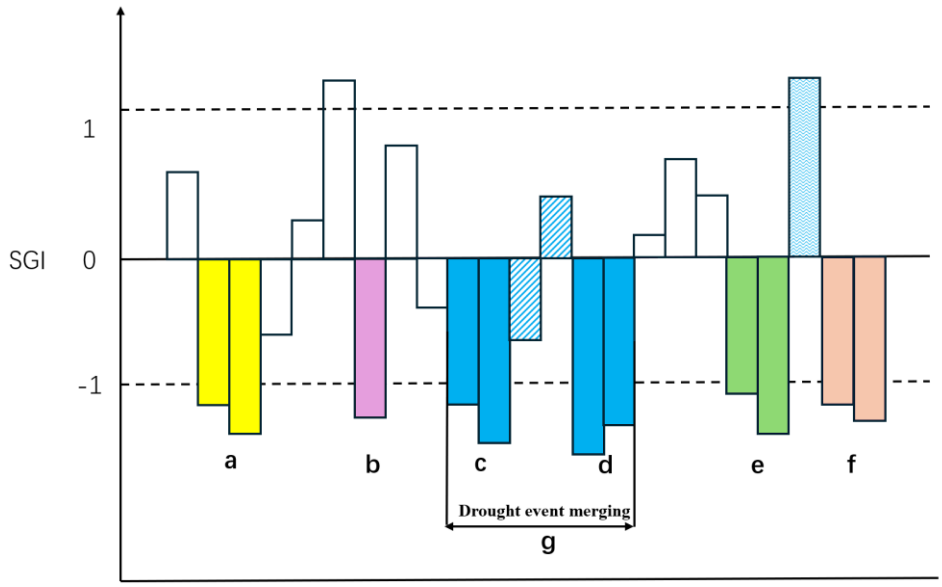


Fig. 4. Groundwater drought event identification and merging

2.4.2 Introducing artificial gaps

After the groundwater drought events were identified, groundwater level values within the drought periods were removed as missing values to create artificial gaps for imputation. The dataset after masking is illustrated on Fig. 5.

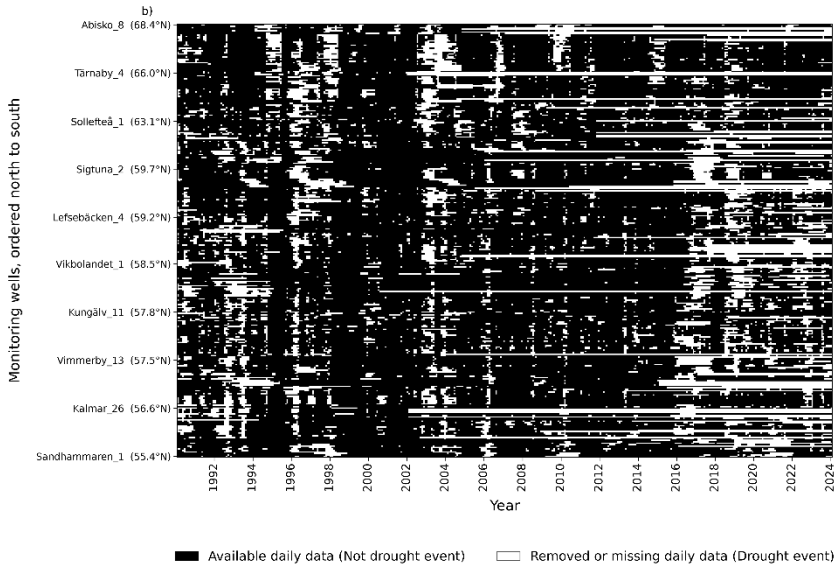


Fig. 5. Distribution of observed and artificially masked groundwater level data across the 297 wells.

The SGI heatmap (Fig. 6) shows marked temporal and spatial variability. Widespread drought signals were apparent around 1996 and 2003, while prolonged dry conditions affected some central and southern wells during 2016–2019. The proportion of monitored wells with SGI below −1 reached its maximum of approximately 62% in March 1996.

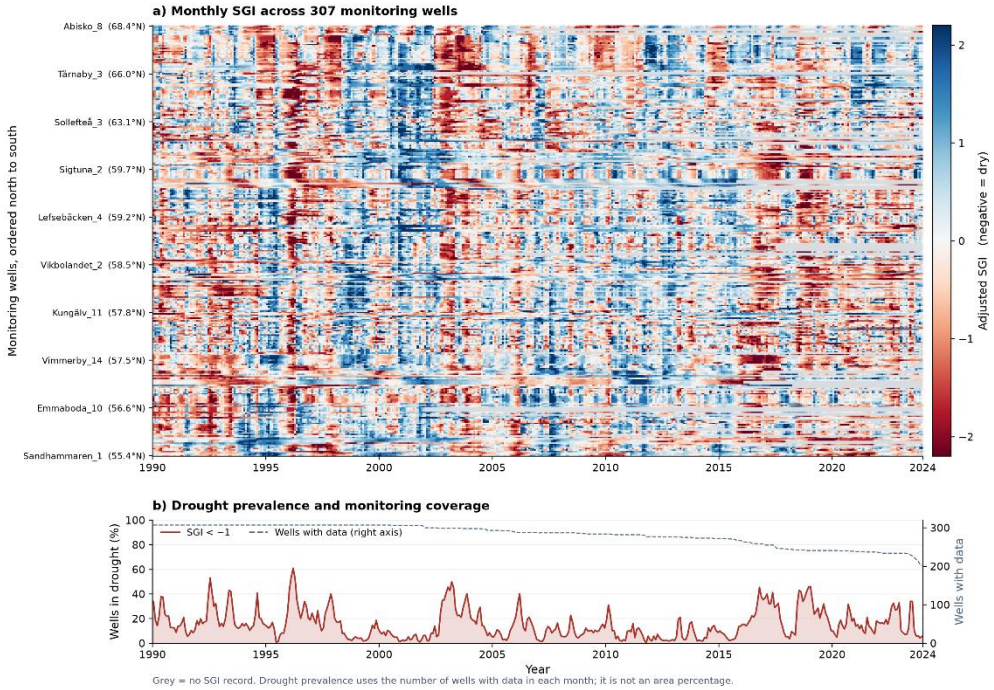


Fig. 6. Spatial and temporal patterns of groundwater drought across 297 monitoring wells during 1990–2023. (a) SGI heatmap. Red indicates dry conditions, blue indicates wet conditions, and grey indicates missing data. (b) Monthly percentage of wells experiencing drought, together with the number of available wells.

2.5 Variables construction

Three main variables were used in the missForest imputation process: time variables, climatic variables, and groundwater levels from nearby reference wells.

2.5.1 Time variables

Three time variables were used for groundwater-level imputation: day of year, month, and a continuous time variable (Unix time) (Bikše et al., 2023). Since missForest does not explicitly account for temporal order or periodicity, these variables were included to provide temporal information that can be learned by the random forest models. Day of year and month were used to represent seasonal variations in groundwater levels, while Unix time was defined as the cumulative number of days since 1990/1/1 and was used to represent time order and long-term trends.

2.5.2 Climatic variables

Understanding groundwater recharge conditions is important for groundwater modelling; however, groundwater recharge is difficult to measure directly(Barthel et al., 2021). Therefore, this study used a water budget method to approximate recharge conditions (Healy, 2010). Net precipitation (NP), defined as the difference between precipitation (P) input and potential

evapotranspiration (PET) loss, was used as a climatic variable for groundwater level imputation:

$$\mathrm{NP} = \mathrm{P} - \mathrm{PET}$$

where P is precipitation [mm] and PET is potential evapotranspiration [mm]. PET was estimated using Python package PyEt (Vremec et al., 2024) with the Hamon method. Net precipitation was subsequently accumulated over rolling periods of 1, 3, 6, 9, and 12 months, denoted as NP_{1m} and NP_{nm}, where $n \in \{3,6,9,12\}$, to represent water surpluses and deficits at different timescales and their delayed effects on groundwater level variations.

2.5.3 Groundwater levels from reference wells

The groundwater levels data were processed in three ways: original groundwater-level (GWL) values, seasonal anomalies (Anomaly), and z-score. For a given date, observations from the same reference well within 15 days before and after that date in previous years were used to calculate the historical seasonal mean and standard deviation (Boutt, 2017). The Anomaly value was calculated by subtracting the historical seasonal mean from the groundwater level on that date, and the z-score was calculated by dividing the seasonal anomaly by the historical seasonal standard deviation. The seasonal anomaly shows the absolute difference from normal seasonal conditions, while the z-score removes differences in the range of groundwater-level changes between wells, making anomalies easier to compare. For each date being calculated, the historical seasonal mean and standard deviation were calculated only using observations from at least two previous calendar years. Consequently, anomaly and z-score values were set as NaN during the first 2 years of each well record.

For missForest algorithm, spearman correlation coefficient was used to evaluate the relevance between the groundwater level time series of each target well and reference wells (Sejine et al., 2026). Spearman's correlation measures the strength of a relationship between two groundwater level time series. The coefficient value approaching -1 or 1 indicates a strong negative or positive association, respectively, whereas a coefficient close to 0 indicates a weak or negligible association. A strong correlation between two wells suggests that their groundwater levels exhibit similar or systematically related variations, allowing the reference well to provide useful spatial information for groundwater level imputation.

In this study, candidate reference wells were first identified within a 50 km radius of each target well, and Spearman's rank correlation coefficient was calculated between the target well and each candidate well. To balance the amount of spatial information included with data availability, model input size, and computational cost, the number of reference-well predictors was capped at 10. When fewer than 10 reference wells were available, all available wells were retained; otherwise, the 10 wells with the highest absolute Spearman correlation coefficients were selected. This upper limit was intended to retain the most relevant groundwater-level

information while avoiding unnecessary increases in model complexity and computational demand.

2.6 Imputation methods and evaluation

2.6.1 The missForest algorithm

MissForest is an iterative missing data imputation algorithm based on the Random forest algorithm (Stekhoven & Bühlmann, 2012). The algorithm can handle continuous and mixed type variables and can be applied to data series of different lengths. In hydrological studies, the missForest algorithm has been shown to be effective for imputing missing values in daily scale time series (Arriagada et al., 2021; Bikše et al., 2023).

The missForest algorithm imputation procedure consists of three main steps. First, missing values are initialized using the mean for continuous variables and the mode for categorical variables. Second, each variable containing missing values is treated in turn as the target variable, while the observed data from the other variables are used to train a random forest model. The trained model is then used to predict the missing values of the target variable, and the dataset is updated with these predictions. One iteration is completed after all variables containing missing values have been updated. The algorithm was allowed to terminate before reaching the maximum number of iterations when the stopping criterion was satisfied.

In this study, missForest was implemented in the RStudio environment. The versions of R, RStudio, and the missForest package used were 4.4.2, 2024.12.0.467, and 1.6.1, respectively. The number of trees in each random forest model (ntree) was set to 100, and the maximum number of imputation iterations (maxiter) was set to 10, following the standard missForest settings (Stekhoven & Bühlmann, 2012). These values were used to balance prediction stability and computational efficiency.

2.6.2 ARCHI package

ARCHI is a R package developed for imputing missing values in groundwater level time series exhibiting regional correlations (Levy et al., 2025). The package automatically identifies reference series correlated with a target series and imputes missing values using regression models. Unlike missForest, ARCHI uses only original groundwater time series data and does not require additional variables, such as time or climate variables.

The ARCHI imputation procedure consists of three main steps. First, reference series are selected based on the overlap of observations and the correlations between time series. Second, regression models are fitted using the selected reference series to predict missing values in the target series. Finally, the imputed series are added to the pool of potential reference series and used to fill gaps in other records. This process continues until no additional missing values

satisfy the imputation criteria. ARCHI supports ordinary linear regression, ridge regression, and MOVE.1 models. Correlation and error thresholds can also be changed to control the reliability of the imputed results. The package has been applied to impute hydrological records at different time scales, including groundwater levels, precipitation, and storm surge data (Glas et al., 2026; Levy et al., 2025).

ARCHI was implemented in the RStudio environment using version 1.0.1 of the ARCHI package.

2.6.3 Model performance metrics

In this study, the performance of the different imputation methods was evaluated by comparing the imputed groundwater levels with the original observation data. Imputation accuracy was assessed using the root mean square error (RMSE), mean absolute error (MAE) and Pearson's correlation coefficient (r). RMSE gives greater weight to large errors, whereas MAE measures the average absolute difference between reconstructed and reference values. Lower RMSE and MAE indicate smaller reconstruction errors. Pearson's (r) measures linear association, with values closer to +1 indicating stronger positive correspondence between the two series. The metrics were calculated as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|P_i - O_i|$$

$$r = \frac{\sum_{i=1}^{n}(P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^{n}(P_i - \bar{P})^2}\sqrt{\sum_{i=1}^{n}(O_i - \bar{O})^2}}$$

where O_i is the original observed groundwater level at the ith data point [m], P_i is the corresponding imputed groundwater level [m], $\bar{O}$ and $\bar{P}$ are their means over the evaluated pairs, n is the total number of artificially removed observations included in the evaluation, and i denotes the index of each evaluated data point.

To provide hydrological background information, groundwater droughts were further analyzed using the monthly SGI value. Following the definitions of drought characteristics used by Parry et al. (2024), duration was defined as the number of months in an event, severity as the cumulative deficit below the threshold, and intensity as the maximum monthly deficit below that threshold:

$$D_k = |E_k|$$

$$S_k = \sum_{t \in E_k} (u - SGI_t)$$

$$I_k = \max_{t \in E_k} (u - SGI_t)$$

$$u = -1$$

where E_k is the set of consecutive months forming event k. Duration (D_k) is expressed in months, severity (S_k) in SGI-months, and intensity (I_k) in SGI units. Here, intensity denotes the peak deficit rather than the mean deficit over an event.

2.6.4 NP variable importance analysis

SHapley Additive exPlanations (SHAP;(Lundberg & Lee, 2017)) were used to assess the relative importance of different NP accumulation windows in predicting groundwater levels. For each of the Lerum target wells, a separate random forest model was fitted using NP1–NP12 as predictors and available groundwater-level observations as the response. SHAP values were calculated using TreeExplainer in Python. SHAP decomposes each prediction into a baseline value and contributions from individual predictors. For each NP variable, absolute SHAP values were averaged over the explained records within each well and then averaged across the 8 wells. Larger mean absolute SHAP values indicated greater contributions to the model predictions. This analysis was used to identify the NP accumulation windows most strongly represented in the model.

3. Results

3.1 Groundwater drought characteristics

Groundwater drought events differed in timing, duration and intensity among the monitoring wells. Even during periods of widespread drought, events did not begin or end simultaneously. The analysis identified 4,611 well-level drought event segments. Fig. 7 highlights the distinction between peak intensity and cumulative severity. Short events with pronounced monthly SGI deficits could exhibit high peak intensity without accumulating particularly large severity. Conversely, prolonged events could accumulate substantial severity despite less extreme monthly deficits. Duration, peak intensity and cumulative severity therefore provide complementary descriptions of groundwater drought characteristics.

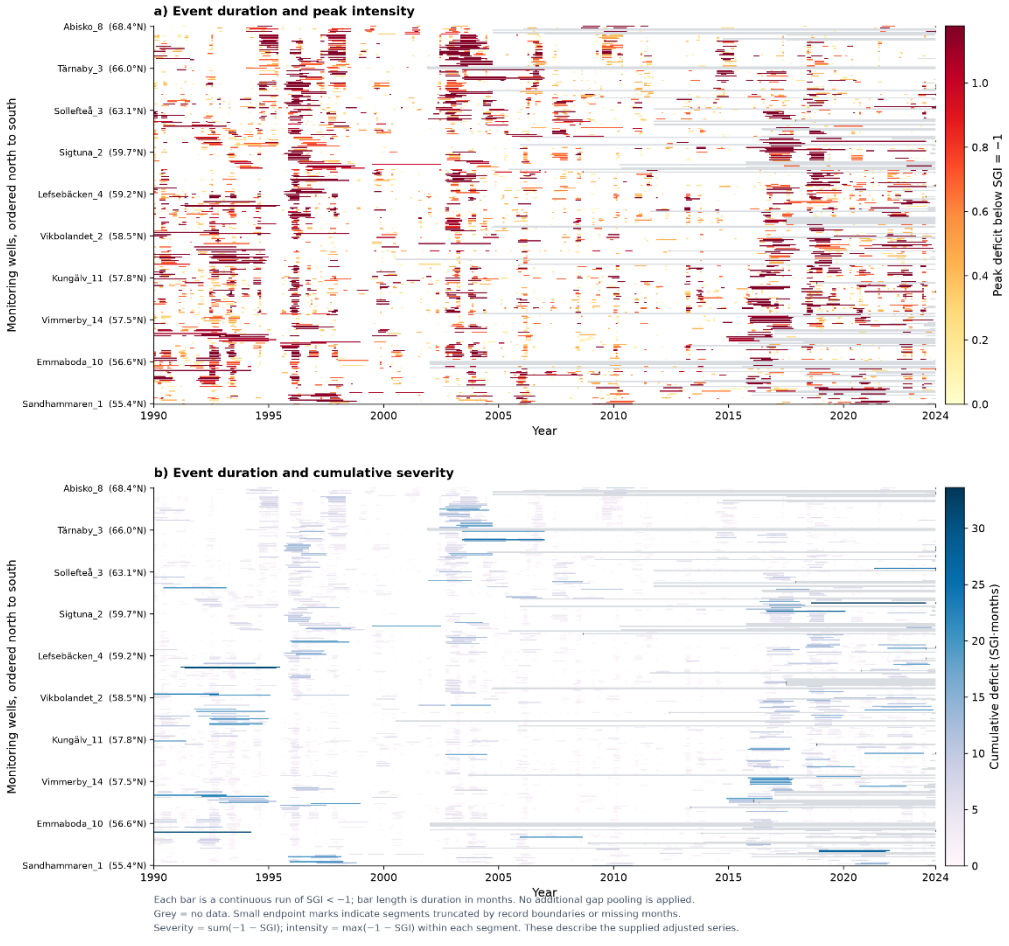


Fig. 7. Groundwater drought event characteristics across 297 monitoring wells during 1990–2023. Colours indicate (a) peak intensity, defined as the maximum monthly deficit below the SGI threshold of −1, and (b) cumulative severity, defined as the sum of monthly deficits below this threshold.

3.2 The missForest algorithm

3.2.1 Lerum local experiment

To select the variables for the global groundwater level imputation, local experiments were first conducted using 8 target wells in the Lerum area. The results of all experiments are presented in Appendix 1. Overall, the model performed poorly when only time or the NP variables were used. When the groundwater levels of nearby wells were combined with the time and NP variables, the model errors decreased (Fig. 8).

In the time variable experiment (Fig. 8a), Unix time had median RMSE and MAE values of 0.116 and 0.089, respectively. These values were lower than those for day of year (0.222 and 0.208), month (0.222 and 0.207), and the combination of all time variables (0.167 and 0.143). However, to maintain possible seasonal information, all time variables (Unix time, day of year, month) were used together as the time variable in the following experiments.

When only the net precipitation (NP) variable was used, the model error generally decreased as the NP accumulation period became longer (Fig. 8b). The median RMSE and MAE decreased from 0.204 and 0.167 for NP1 to 0.164 and 0.137 for NP1–12. However, the improvement became small after NP1–NP8. After the time variables were added, the model errors decreased further. The errors decreased clearly from NP1 to NP1-3, but there was little further improvement after NP1-3.

In the groundwater level variables experiment, when only the original groundwater levels were used, the median RMSE decreased from 0.154 to 0.131 as the number of neighboring wells increased from 1 to 8. This was a decrease of about 14.9%. When the number of wells increased to 10, the RMSE only decreased slightly to 0.129. Using only the anomaly or z-score values of nearby wells increased the model errors. Adding these 2 variables together with the original groundwater levels also performed worse than using only the original groundwater levels (Fig. 8c).

After the time variables were added to the nearby well groundwater levels, the model performance improved further. When 1, 5, and 8 nearby wells were used, the median RMSE values were 0.120, 0.101, and 0.093, respectively, while the median MAE values were 0.097, 0.081, and 0.075. When the number of nearby wells increased to 10, the median RMSE and MAE increased slightly to 0.095 and 0.078. Therefore, the “Top 8 GWL + Time” setting had the lowest median RMSE and MAE in this experiment (Fig. 8d).

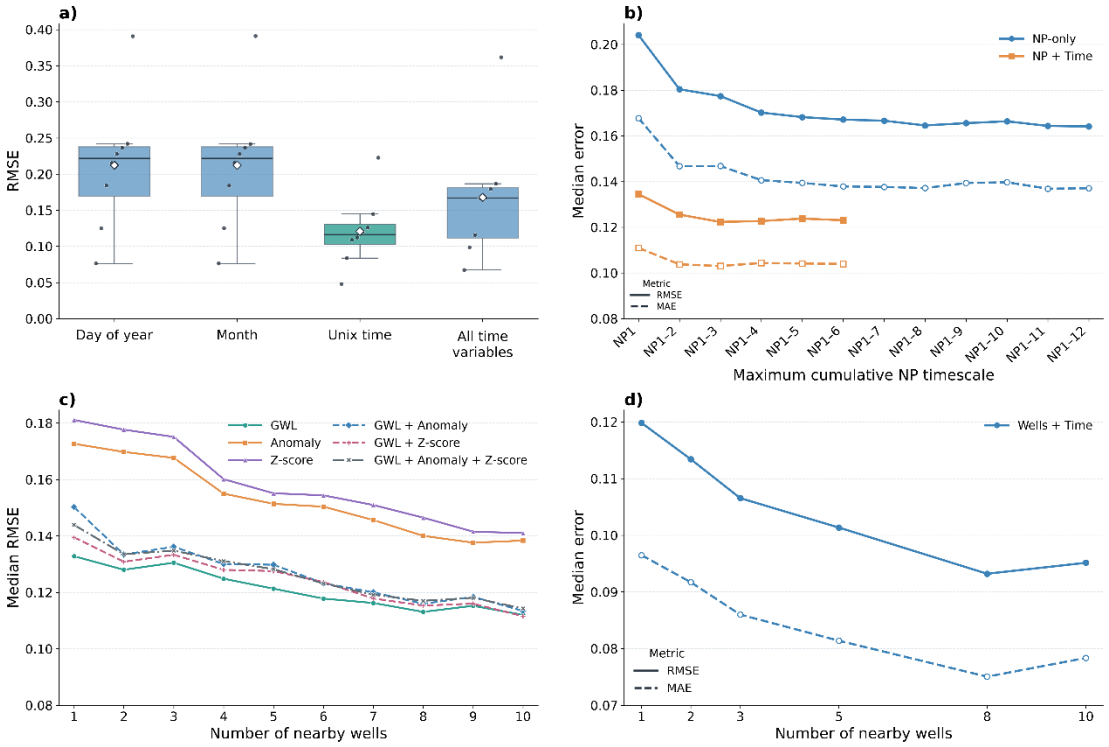


Fig. 8. Effects of different variables on missForest performance in Lerum. (a) Time variables; (b) NP accumulation windows; (c) groundwater level variables; (d) number of nearby wells.

The RMSE heatmap further shows the combined effects of the number of nearby wells, the NP accumulation window, and the time variables (Fig. 9). Without the time variables (Fig. 9a), the median RMSE ranged from 0.109 to 0.127. The lowest value 0.109, was found for “Top10 + NP1”. Simply increasing the NP accumulation window did not lead to a stable improvement. After adding the time variables (Fig. 9b), the overall median RMSE decreased to 0.090–0.117. Both “Top8 + NP1–4” and “Top8 + NP1–5” reached the lowest value of 0.090. Compared with the best result without the time variables (0.109), the lowest RMSE decreased by about 17.4%. The MAE heatmap shows a similar pattern (Fig. 9). Without the time variables (Fig. 9c), the median MAE ranged from 0.091 to 0.104, and the lowest value of 0.091 was found for “Top10 + NP1”. After adding the time variables (Fig. 9d), the median MAE decreased to 0.073–0.093. The lowest value of 0.073 was found for “Top8 + NP1–5”. This was about 19.8% lower than the best MAE without the time variables. These results show that adding the time variables improved model performance more than further increasing the NP window or the number of neighboring wells.

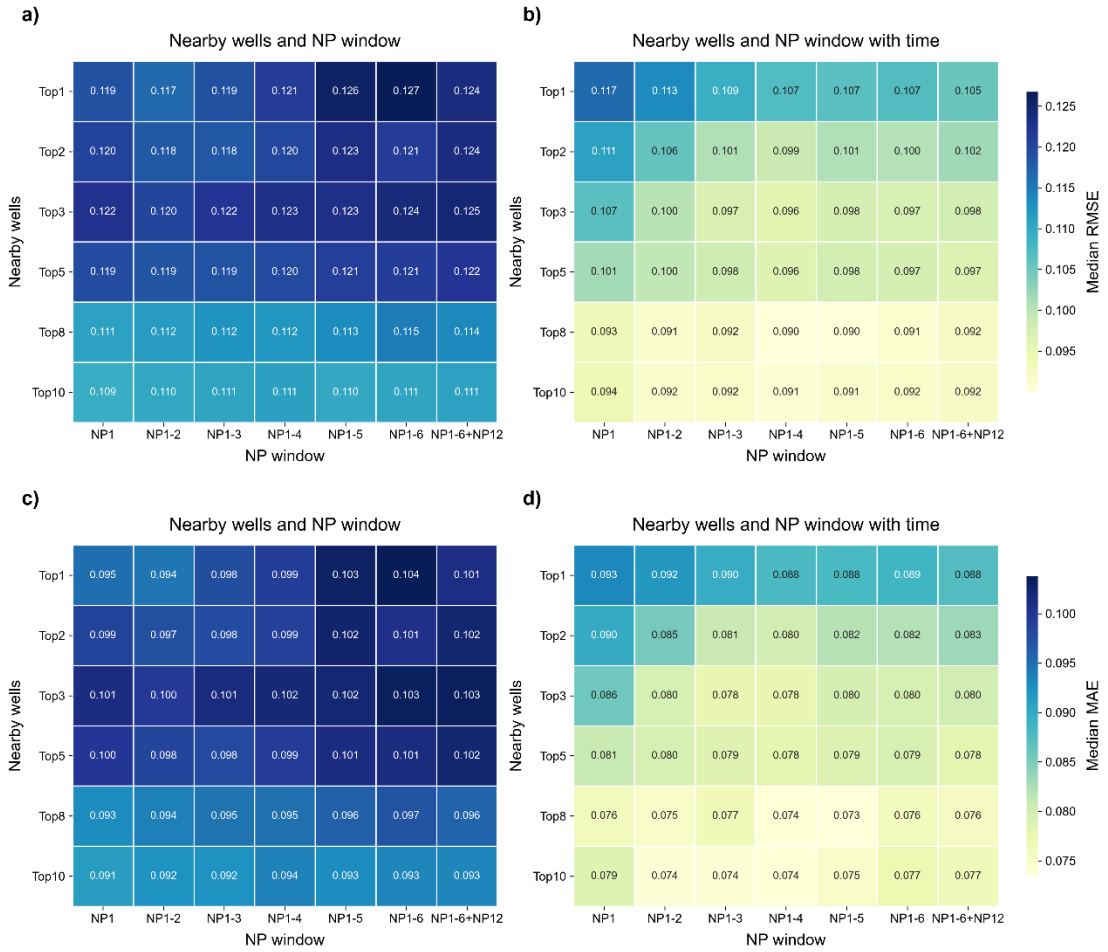


Fig. 9. Median RMSE and MAE for different combinations of nearby wells and NP rolling-sum windows, with and without time variables. (a) Median RMSE without time variables; (b) median RMSE with time variables; (c) median MAE without time variables; and (d) median MAE with time variables.

The SHAP analysis further shows how much each NP accumulation window contributed to the output of the NP only model. In the full NP1–12 model, NP5 had the highest mean absolute SHAP value (0.025), followed by NP4 (0.020) and NP1 (0.020). The mean absolute SHAP values of NP6, NP3, and NP12 were 0.016, 0.015, and 0.015, respectively. Together, NP1–NP6 and NP12 accounted for about 77.9% of the total SHAP value in the full NP model. In comparison, NP7–NP11 only made smaller contributions. This shows that the NP information used by the model mainly came from short to medium term accumulation periods. NP12 also provided some useful long-term information (Fig. 10).

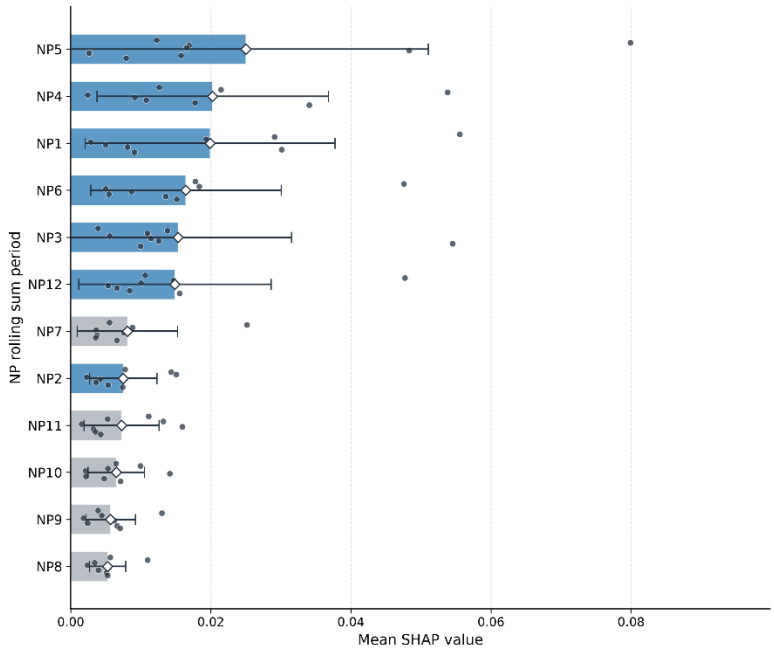


Fig. 10. SHAP value of NP rolling sum windows in the NP-only model

3.2.2 Global experiment

Based on the Lerum experiments, 31 configurations were evaluated using the global dataset of 297 groundwater wells (Tab. 2). The number of eligible target wells decreased as the required number of reference wells increased: 294 wells had at least one eligible reference well, whereas 147 had at least ten (Table 3). All configuration comparisons reported below therefore use the same 147 target wells. Median performance metrics are summarised in Tab. 4, while the distributions of RMSE, MAE and Pearson correlation are presented in Figs. 11–14, respectively. Detailed results for individual wells are provided in Appendix 2.

Tab. 2 31 global experiment configurations

Experiment group	Configuration	Reference wells (n)	NP variables	Time variables
Baseline	B1	0	None	Yes
	B2	0	NP1-NP3	
	B3	0	NP1-NP6	
	B4	0	NP1-NP6 and NP12	
	B5	1	None	No
	B6	3	None	
	B7	5	None	
	B8	8	None	
	B9	10	None	
Main route	M1	1	NP1	Yes
	M2	2	NP1-NP2	
	M3	3	NP1-NP3	
	M4	5	NP1-NP4	
	M5	8	NP1-NP5	
	M6	10	NP1-NP6	
Fixed wells	N1	5	NP1	Yes
	N2	5	NP1-NP2	
	N3	5	NP1-NP5	
	N4	5	NP1-NP6	
Fixed NP period	W1	1	NP1-NP3	Yes
	W2	2	NP1-NP3	
	W3	5	NP1-NP3	
	W4	8	NP1-NP3	
	W5	10	NP1-NP3	
Crossing	C1	1	NP1-NP6	Yes
	C2	3	NP1-NP6	
	C3	8	NP1-NP2	
	C4	10	NP1	
Additional NP12	L1	1	NP1-NP6 and NP12	Yes
	L2	5	NP1-NP6 and NP12	
	L3	10	NP1-NP6 and NP12	

Tab. 3 Number of eligible target wells by minimum required number of reference wells.

Required number of reference wells	Number of eligible target wells
1	294
3	273
5	220
8	159
10	147

The baseline configurations showed clear benefits from additional hydrological information. Adding NP1–3 to the time-only model reduced median RMSE by 13.7% and median MAE by 9.8%. Extending the NP input set to NP1–6 produced further reductions, although the additional benefit of NP12 was smaller (Fig. 11a and Fig. 12a). Within the GWL-only configurations, increasing the number of reference wells from 1 to 10 reduced median RMSE from 0.299 to 0.205 and median MAE from 0.239 to 0.161, corresponding to reductions of 31.3% and 32.8%, respectively. Median Pearson r also increased from 0.701 to 0.858 (Fig. 13a).

The effect of reference-well information was also evident when the climatic and time variables were fixed (Fig. 11d, Fig. 12d and Fig. 14a). Increasing the number of reference wells from 1 to 10 reduced median RMSE by 25.4% and median MAE by 30.1%. The largest improvement between successive tested settings occurred from 1 to 2 reference wells. By contrast, increasing the number from 8 to 10 reduced median RMSE by only 2.2%, while median MAE remained unchanged. The main route configurations showed a similar pattern: moving from M5 (Top8 + NP1-5 + Time) to the M6 (Top10 + NP1–6 + Time) produced little further reduction in RMSE and a slight increase in MAE (Fig. 11b and Fig. 12b). Thus, larger input sets did not consistently improve all performance metrics.

The contribution of NP depended on the availability of reference-well information (Fig. 14b). Without reference-well groundwater levels, extending NP1–3 to NP1–6 reduced median RMSE and MAE by a further 5.4% and 11.7%, respectively. However, when 5 reference wells and time variables were already included, extending the NP windows did not consistently improve performance.

Pearson correlation could also support these improvements, but the ranking of configurations differed slightly (Fig. 13). Median r ranged from 0.6806 for the time-only model to 0.8901 for C3 (TOP8 + NP1-2 + Time). Once 3 or more reference wells were included, differences in correlation between the configurations were small. The Pearson correlation results also illustrate how the contribution of longer NP accumulation windows depended on the availability of reference-well information (Fig. 14c). Without reference-well groundwater levels, extending the NP input set from NP1–3 to NP1–6 and subsequently adding NP12 increased median Pearson r from 0.779 to 0.789 and 0.799, respectively. In contrast, when 5

reference wells and time variables were included, median r remained within a range of 0.878–0.884.

Across all 31 configurations, W5 (Top10 + NP1–3 + Time) achieved the lowest median RMSE of 0.202, B9 (Top10 GWL-only) achieved the lowest median MAE of 0.161, C3 (TOP8 + NP1-2 + Time) achieved the highest median Pearson r. W5 (Top10 + NP1–3 + Time) was selected as the representative missForest configuration using median RMSE as the primary criterion; its median MAE and Pearson r were 0.161 and 0.887, respectively (Tab. 4). Overall, reference-well groundwater levels provided the clearest improvements in reconstruction accuracy, while NP and time variables offered complementary benefits that depended on the existing input information.

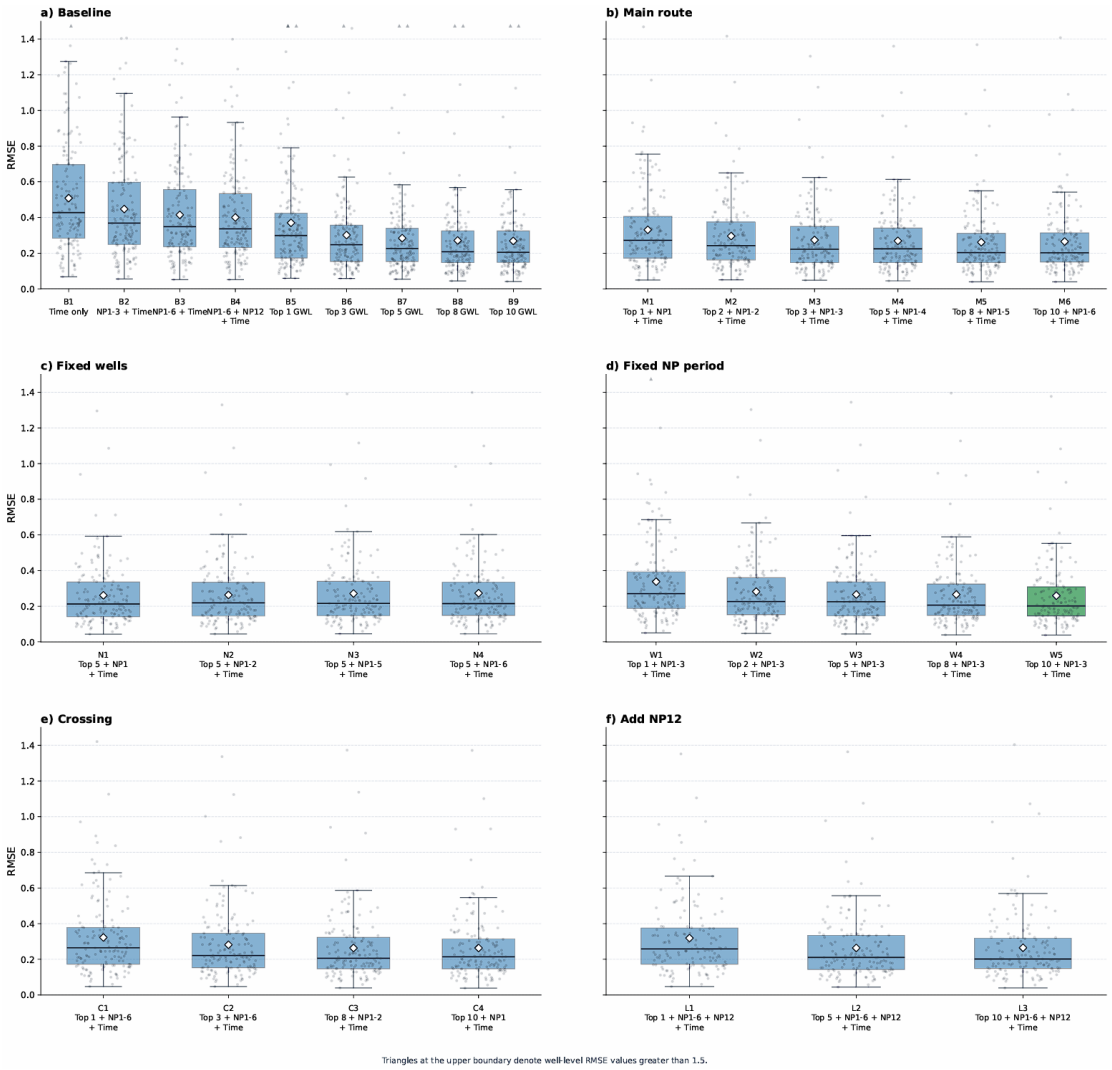


Fig. 11. RMSE distributions for different configurations in the global missForest experiment: (a) baseline configurations; (b) main experiment; (c) fixed number of nearby wells; (d) fixed NP accumulation window; (e) combined configurations; and (f) configurations with NP12.

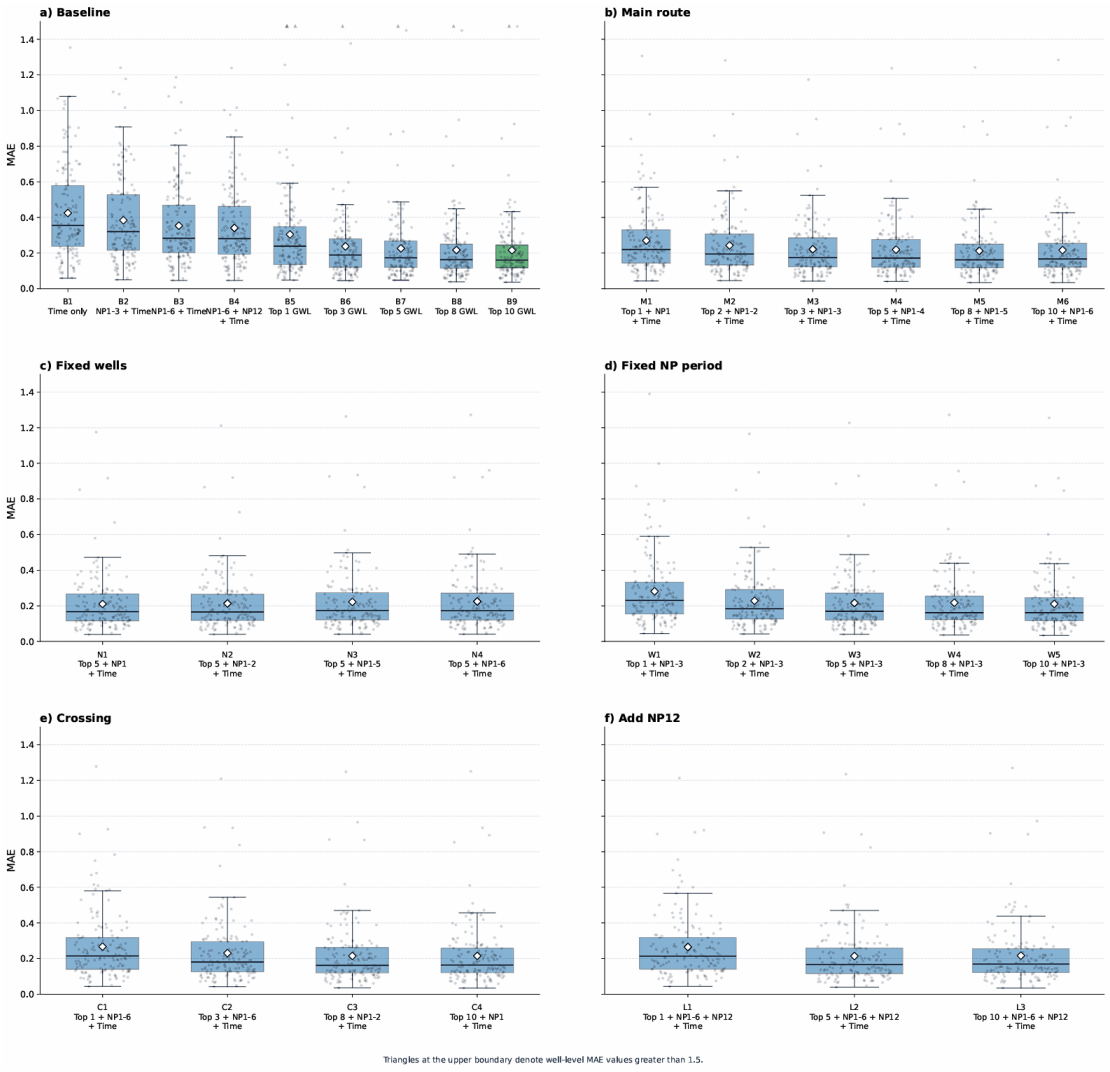


Fig. 12. MAE distributions for different configurations in the global missForest experiment: (a) baseline configurations; (b) main experiment; (c) fixed number of nearby wells; (d) fixed NP accumulation window; (e) combined configurations; and (f) configurations with NP12.

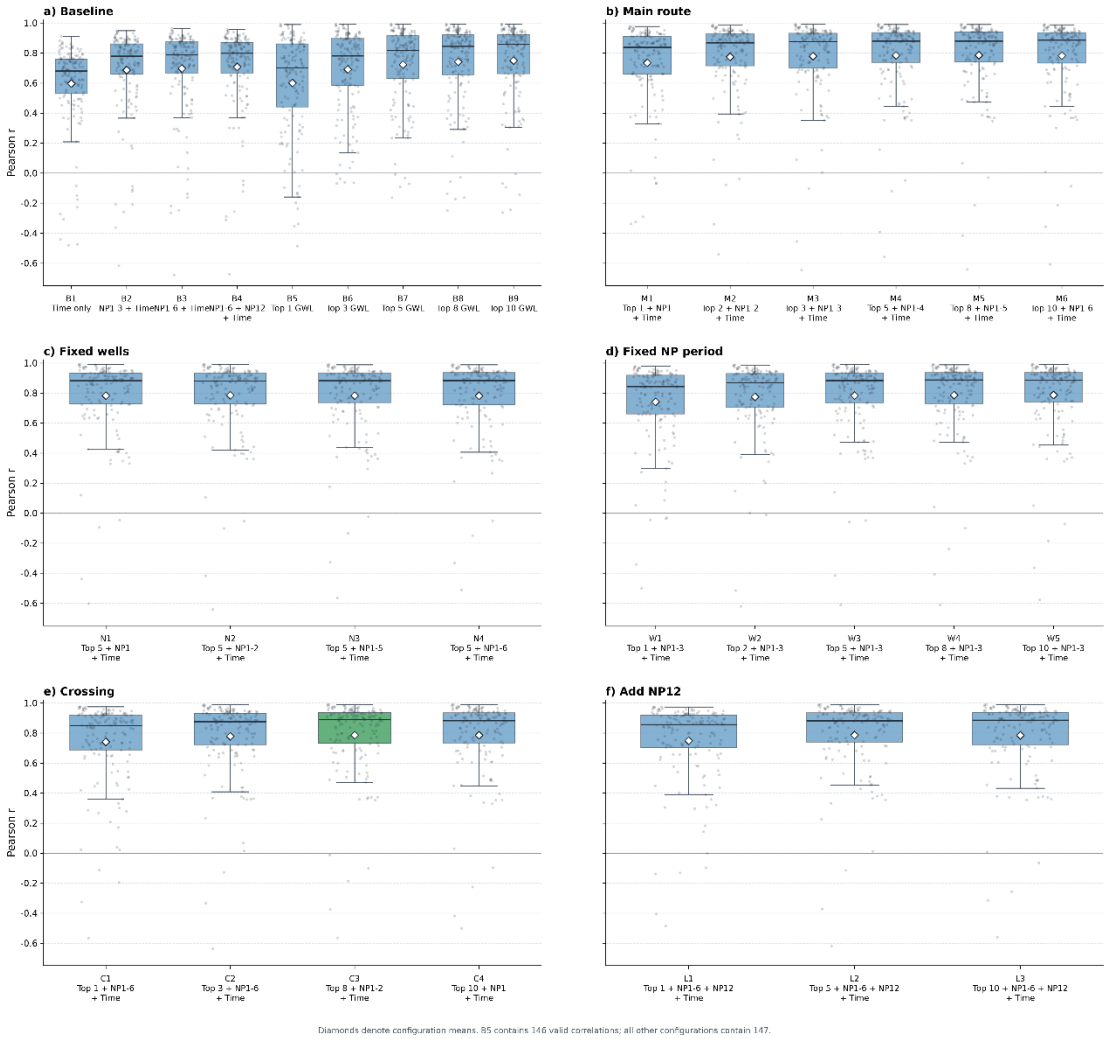


Fig. 13. Pearson r distributions for different configurations in the global missForest experiment: (a) baseline configurations; (b) main experiment; (c) fixed number of nearby wells; (d) fixed NP accumulation window; (e) combined configurations; and (f) configurations with NP12.

Tab. 4 Median RMSE, MAE and Pearson r for the 31 global missForest configurations.

Experiment group	Configuration	Median RMSE (m)	Median MAE (m)	Median Pearson r
Baseline	B1	0.4274	0.3551	0.6806
	B2	0.3688	0.3202	0.7786
	B3	0.3488	0.2829	0.7885
	B4	0.3367	0.2812	0.7994
	B5	0.2987	0.2389	0.7014
	B6	0.2468	0.1894	0.7815
	B7	0.2258	0.1734	0.8191
	B8	0.2081	0.1630	0.8456
	B9	0.2051	**0.1606**	0.8584
Main route	M1	0.2719	0.2201	0.8379
	M2	0.2416	0.1949	0.8692
	M3	0.2217	0.1745	0.8764
	M4	0.2245	0.1719	0.8799
	M5	0.2029	0.1621	0.8801
	M6	0.2016	0.1669	0.8863
Fixed wells	N1	0.2126	0.1671	0.8826
	N2	0.2193	0.1653	0.8805
	N3	0.2159	0.1738	0.8819
	N4	0.2145	0.1728	0.8840
Fixed NP period	W1	0.2701	0.2308	0.8436
	W2	0.2263	0.1838	0.8701
	W3	0.2248	0.1692	0.8830
	W4	0.2061	0.1615	0.8874
	W5	**0.2016**	0.1614	0.8865
Crossing	C1	0.2640	0.2145	0.8507
	C2	0.2216	0.1807	0.8764
	C3	0.2065	0.1618	**0.8901**
	C4	0.2138	0.1633	0.8836
Additional NP12	L1	0.2581	0.2136	0.8565
	L2	0.2107	0.1663	0.8822
	L3	0.2017	0.1692	0.8866

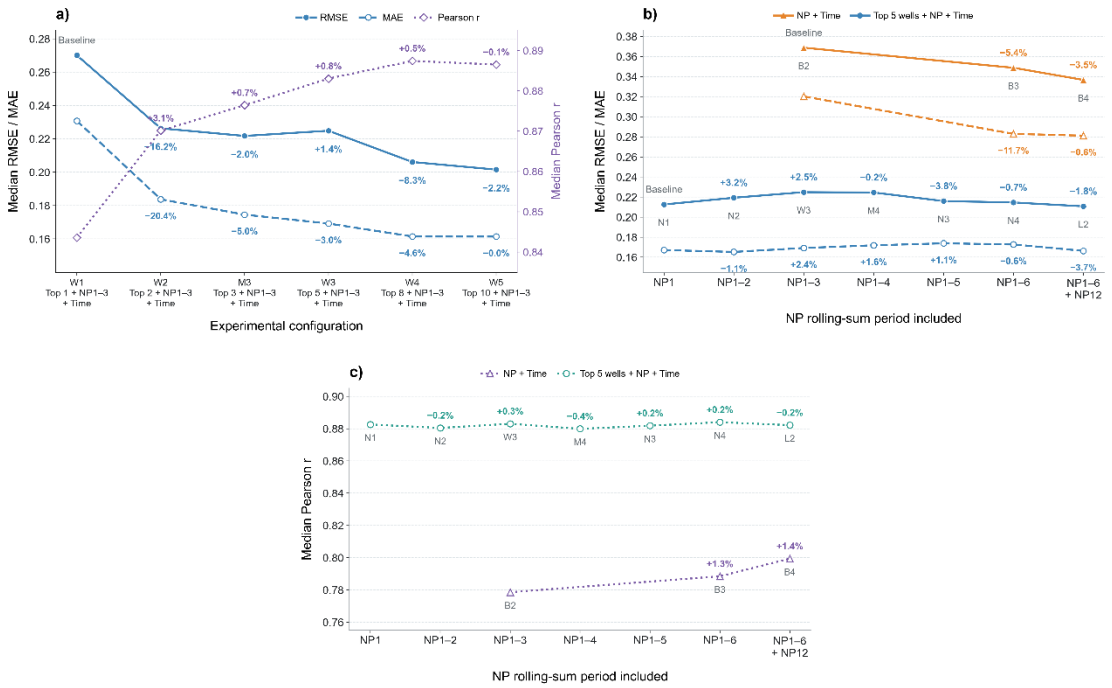


Fig. 14. Effects of reference-well number and NP accumulation windows on missForest performance across 147 target wells. (a) Median RMSE, MAE and Pearson r as the number of reference wells increases, with NP1–3 and time variables fixed. (b) Median RMSE and MAE for different NP input sets, with and without 5 reference wells. (c) Median Pearson r for different NP input sets, with and without 5 reference wells.

3.3 ARCHI package

ARCHI was applied to all 297 target wells, of which 245 runs were completed successfully. Among these successful runs, 35 target wells were imputed using only a regional mean groundwater-level series, without directly selecting any individual reference well due to the package's internal correlation-based screening procedure. This regional series was constructed by averaging standardised groundwater-level series from surrounding wells and therefore represented a composite reference rather than an individual monitoring well. Therefore, these 35 cases were excluded, leaving 210 wells for analysis. The RMSE and MAE distributions of these 210 target wells when "n_refwl_choice" (the maximum allowed number of reference wells) was set to 1, 2, 3, 5, 8, and 10 as shown in Fig. 15. Pearson correlation distributions in Fig. 16, and results in Tab. 5. Detailed results for individual wells are provided in Appendix 3.

Increasing the maximum allowed number of reference wells generally improved performance, although the overall changes were modest. Between n1 and n8, median RMSE decreased from 0.342 to 0.318 and median MAE from 0.291 to 0.267, corresponding to reductions of 7.0% and 8.1%, respectively (Tab. 5). Increasing the upper limit from 8 to 10 produced unchanged median errors (Fig. 15). Pearson correlation showed a similar overall tendency towards improvement (Fig. 16). Median r increased from 0.708 for n1 to 0.744 for n8. Nevertheless, the configurations

allowing 5, 8 or 10 reference wells all had a median r of 0.7439 at the reported precision (Tab. 5).

Overall, allowing more reference wells slightly improved the imputation performance of ARCHI, but the improvement became smaller as the upper limit increased. The median errors were the same when the upper limit was set to 8 or 10, while an upper limit of 8 required fewer candidate reference wells. Therefore, “n_refwl_choice = 8” was selected as the simpler optimal setting.

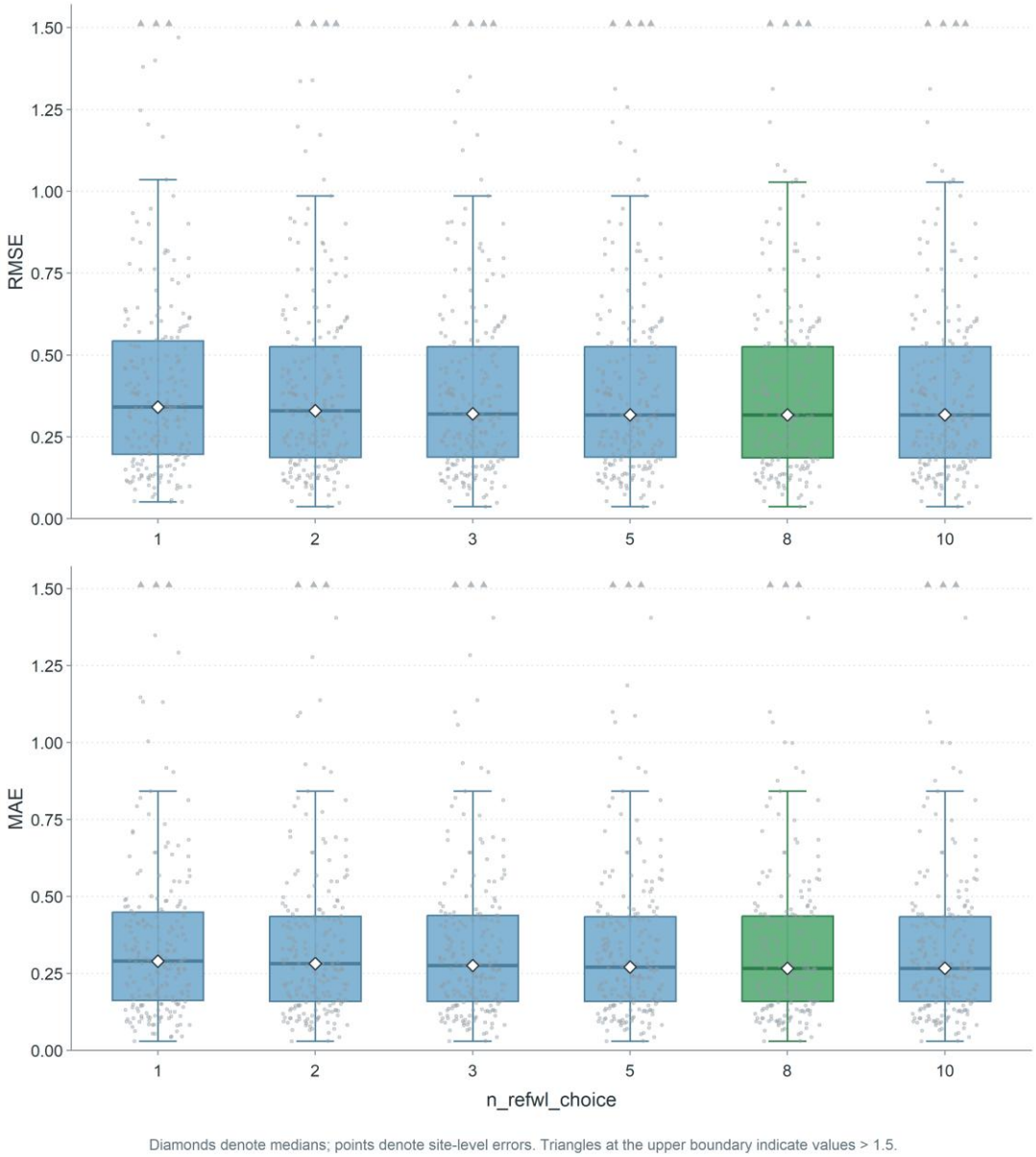


Fig. 15. RMSE and MAE distributions of ARCHI imputation results.

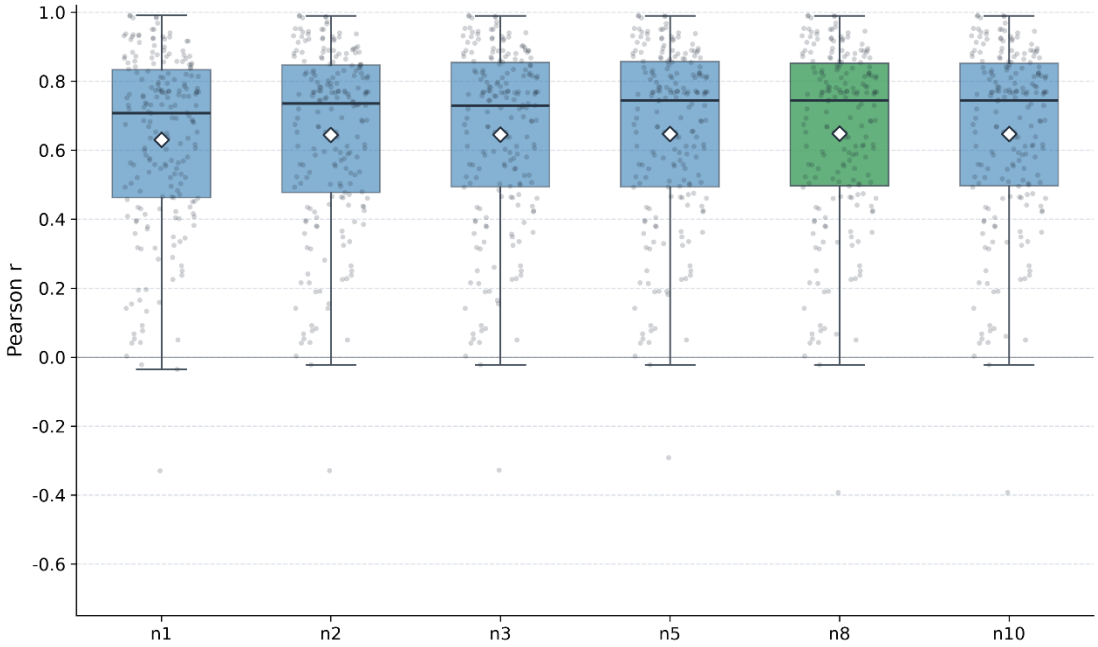


Fig. 16. Pearson r distributions of ARCHI imputation results.

Tab. 5 Median RMSE, MAE and Pearson r for ARCHI with different maximum numbers of reference wells

n_refwl_choice (the maximum allowed number of reference wells)	**Median RMSE (m)**	**Median MAE (m)**	**Median Pearson r**
1	0.3416	0.2905	0.7077
2	0.3303	0.2823	0.7354
3	0.3204	0.2763	0.7295
5	0.3176	0.2716	0.7439
8	**0.3176**	**0.2671**	**0.7439**
10	0.3176	0.2671	0.7439

Interestingly, the actual number of reference wells used was mostly limited (Fig. 17). When "n_refwl_choice = 10", 86 target wells used only one reference well (41.0%), 54 used two reference wells (25.7%), and 22 used three reference wells (10.5%). Together, these groups included 162 target wells, accounting for 77.1% of all target wells. The numbers of target wells that used 4, 5, 6, 7, 8, 9, and 10 reference wells were 7, 6, 14, 9, 6, 2, and 4, respectively. Only 35 target wells used 6 or more reference wells, accounting for 16.7%. This shows that even when the maximum number was increased to 10, ARCHI usually selected only a few reference wells for imputation. In other words, most wells were using only 1 to 3 reference wells.

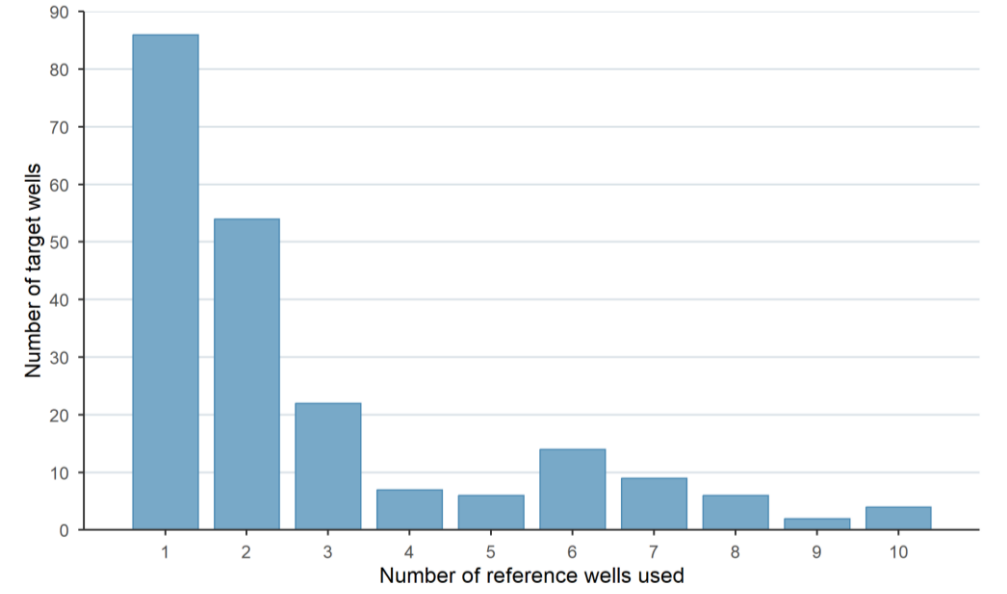


Fig. 17. The number of reference wells used when "n_refwl_choice = 10".

3.4 Comparison of 2 models on common target wells

The representative missForest configuration, W5 (Top10 GWL + NP1–3 + Time), was compared with ARCHI n8 using the 140 target wells common to both evaluation samples. The comparison is summarised in Tab. 6, with results for individual wells provided in Appendix 3.

missForest achieved lower median reconstruction errors than ARCHI. Median RMSE was 0.202 for missForest and 0.314 for ARCHI, while median MAE was 0.158 and 0.262, respectively. Relative to ARCHI, these corresponded to reductions of 35.9% in median RMSE and 39.6% in median MAE. The advantage was also apparent in the individual-well results: missForest produced lower RMSE and lower MAE in 124 wells (88.6%).

Pearson correlation supported the same overall conclusion. Median r was 0.888 for missForest and 0.768 for ARCHI, a difference of approximately 0.120. missForest achieved a higher correlation in 116 of the 140 common wells (82.9%).

Together, the 3 metrics show that W5 (Top10 GWL + NP1–3 + Time) generally combined smaller errors with stronger correlation than ARCHI n8 on the shared target-well sample. These results support the stronger overall performance of W5 (Top10 GWL + NP1–3 + Time) under the experiment conditions used in this study.

Tab. 6 Comparison of missForest W5 and ARCHI n8 on 140 common target wells.

Metric	missForest W5	ARCHI n8	Wells with better missForest performance
Median RMSE (m)	0.2015	0.3144	124/140 (88.6%)
Median MAE (m)	0.1581	0.2618	124/140 (88.6%)
Median pearson r	0.8881	0.7682	116/140 (82.9%)

3.5 Minimum information analysis

3.5.1 The missForest algorithm

The median RMSE of the best global configuration, W5 (Top10 GWL + NP1–3 + Time), was 0.202. This value was used as a reference. The 5% and 10% tolerance thresholds were therefore set to 0.212 and 0.222, respectively (Fig. 18a). When the best configuration was selected for each number of nearby wells, the median RMSE decreased from 0.337 without nearby wells. The result entered the 10% tolerance range when 3 nearby wells were used. With 5 nearby wells, the RMSE decreased to 0.211 and entered the 5% tolerance range. A comparison of different combinations gives more details (Fig. 18b). The GWL-only setting required 8 nearby wells to enter the 10% tolerance range and 10 wells to enter the 5% range. After NP1–3 and the time variables were added, these requirements decreased to 5 and 8 wells. After NP1–6 and the time variables were added, they decreased further to 3 and 5 wells. Therefore, the NP and time variables could partly make up for limited nearby well information.

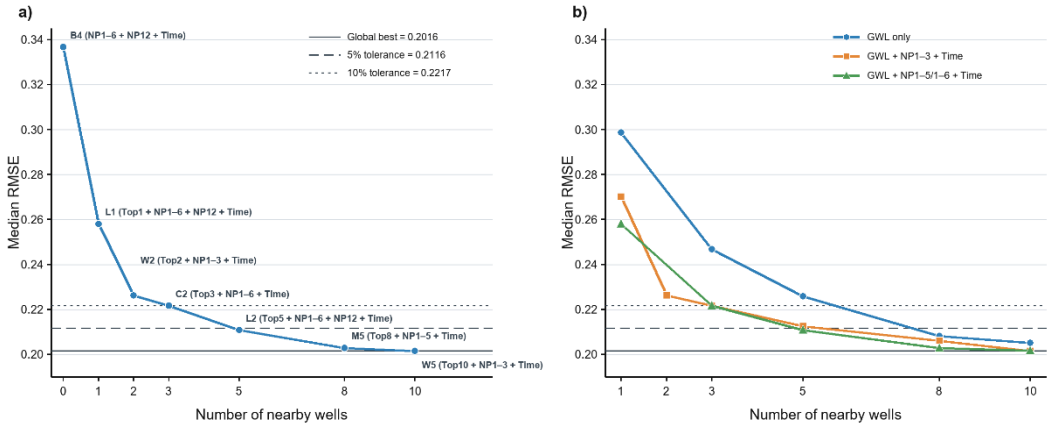


Fig. 18. Minimum number of nearby wells needed for near optimal RMSE. (a) Lowest median RMSE obtained for each number of nearby wells; (b) changes in the required number of nearby wells under different NP and time variables.

The MAE results showed a similar pattern (Fig. 19). The lowest median MAE among all configurations was 0.161, produced by B9 (Top10 GWL-only). The 5% and 10% tolerance thresholds were therefore set to 0.169 and 0.177, respectively. The best median MAE decreased from 0.281 without nearby wells. It entered the 10% tolerance range when three nearby wells

were used. With 5 nearby wells, it decreased further to 0.166 and entered the 5% tolerance range (Fig. 19a). The GWL-only configuration required 5 nearby wells to enter the 10% tolerance range and 8 wells to enter the 5% range. After NP1–3 and the time variables were added, these requirements decreased to 3 and 5 wells, respectively. However, when NP1–6 and the time variables were used, 5 nearby wells were needed to reach both tolerance ranges. In some cases, the error even increased slightly when the number of nearby wells increased from 8 to 10 (Fig. 19b).

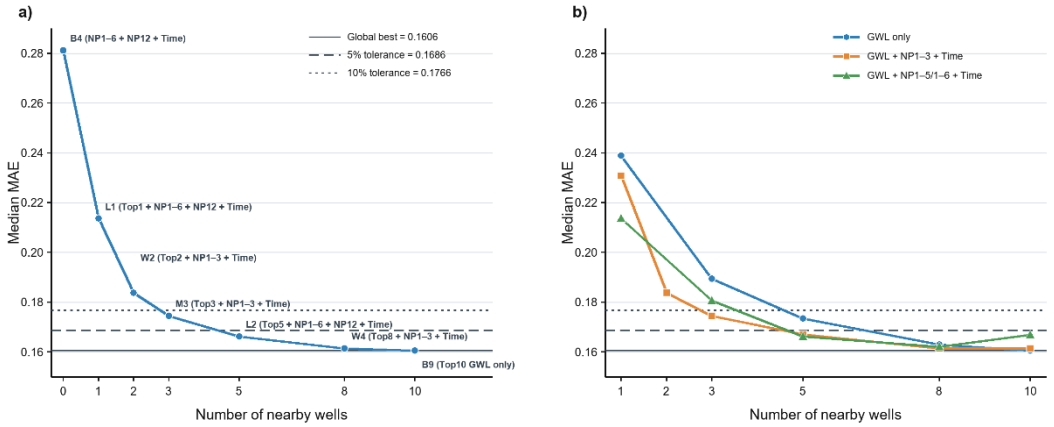


Fig. 19. Minimum number of nearby wells needed for near optimal MAE. (a) Lowest median MAE obtained for each number of nearby wells; (b) changes in the required number of nearby wells under different NP and time variables.

The highest median Pearson r among the global configurations was 0.890, achieved by C3 (Top8 GWL + NP1–2 + Time). The 5% and 10% tolerance thresholds were set to 0.846 and 0.801, respectively (Fig. 20a). When the best configuration was selected for each number of nearby wells, median r increased from 0.799 without nearby wells to 0.857 with 1 nearby well, bringing the result within both tolerance ranges. Correlation improved further as more nearby wells were included, reaching its maximum with 8 wells before decreasing slightly with 10. Climatic and time variables could reduce the reference-well requirement (Fig. 20b). The GWL-only configurations required 5 nearby wells to enter the 10% tolerance range and 8 to enter the 5% range. With short NP windows and time variables included, these requirements decreased to 1 and 2 nearby wells, respectively. With longer NP windows and time variables, 1 nearby well was sufficient to enter both tolerance ranges.

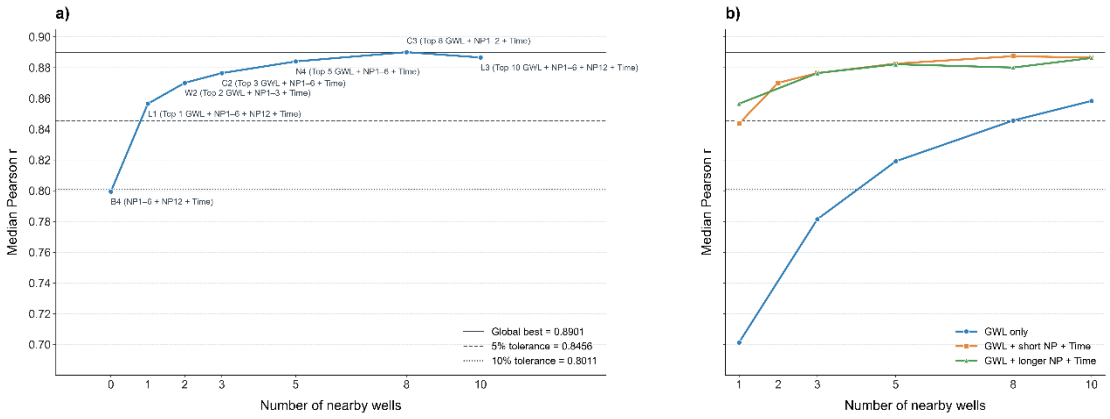


Fig. 20. Minimum number of nearby wells needed for near optimal Pearson r. (a) Highest median Pearson r obtained for each number of nearby wells; (b) changes in the required number of nearby wells under different NP and time variables.

Considering RMSE and MAE, when NP and time variables were included, at least 3 nearby wells were needed to achieve performance within 10% of the best result, while at least 5 were needed to achieve performance within 5%. Pearson correlation required fewer nearby wells: with longer NP windows and time variables, 1 nearby well was sufficient to achieve a median r within 5% of the highest value.

3.5.2 ARCHI package

Fig. 21 shows the median RMSE (Fig. 21a) and MAE (Fig. 21b) of ARCHI and missForest. The improvement in ARCHI was relatively small and began to level off when n_refwl_choice ≥ 5. Pearson correlation showed a similar pattern: median r increased from 0.708 with one permitted reference well to 0.744 with 5, and remained nearly unchanged when the upper limit increased to 8 or 10. It is important to note that the missForest were evaluated using 147 wells. In contrast, the ARCHI is 210 wells. In addition, ARCHI's imputation accuracy was affected by its internal correlation screening (Pearson value must be less than or equal 0.05) and the number of reference wells it selected. Therefore, it cannot be used to make a conclusion about which algorithm is better.

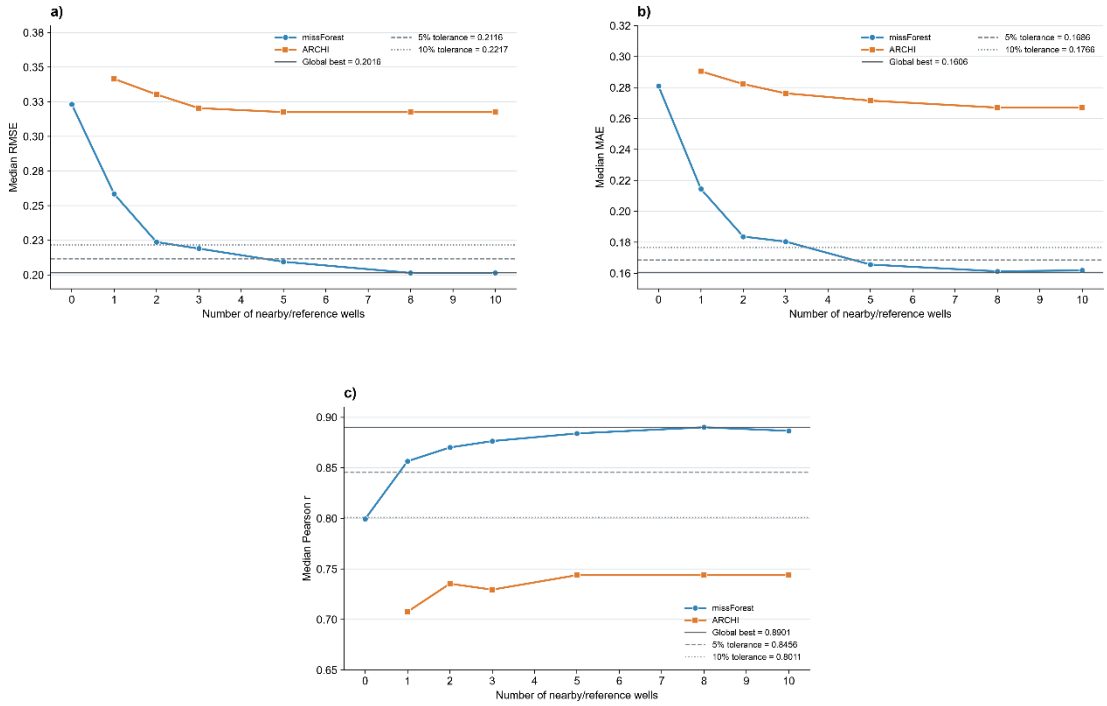


Fig. 21. Comparison of missForest and ARCHI. (a) Median RMSE; (b) median MAE; (c) Pearson r.

4. Discussion

4.1 Imputation methods

The comparison on 140 common target wells showed a better performance for the missForest than ARCHI under the conditions evaluated in this study. The missForest combined nearby groundwater-level records, NP, and time variables, for ARCHI only used groundwater-level records. These additional variables may contribute to the better performance of missForest.

Moreover, several differences between the tested configurations may further explain this performance gap. Firstly, missForest combined reference-well groundwater level data with climatic and time variables, whereas ARCHI relied on groundwater reference series. The additional climatic and time variables supplied information that was not included in the ARCHI modelling. Secondly, the number and selection of reference wells differed between the two methods. For ARCHI, 77.1% of the evaluated target wells were reconstructed using only one to three reference wells. However, a small number of reference wells does not necessarily imply insufficient information. Haaf et al. (2023) found that groundwater level reconstruction often reached optimal performance with only 1 to 3 donor wells, and adding further donors could result in little improvement or even reduced performance when the additional wells exhibited different hydrological responses from the target well. Therefore, the observed plateau in ARCHI performance may reflect not only the number of available reference wells, but also the suitability of the selected reference series. Thirdly, differences in models offer a further possible explanation. The missForest models can represent nonlinear relationships and interactions among inputs (Breiman, 2001), and ARCHI reconstructs through fitted regression relationships.

Although the comparison supports the stronger performance of missForest in this study, ARCHI still performed better at a minority of wells. To make the comparison fairer, further research can evaluate both methods using the same inputs. This would help determine whether the performance difference is caused by the imputation methods themselves or by differences in model inputs.

4.2 Imputation variables

Correlation screened groundwater level time series from nearby wells were the primary source of information for improving missForest imputation accuracy. When NP1–3 and the time variables were held constant, increasing the number of nearby wells from 1 to 10 reduced the median RMSE and MAE by 25.4% and 30.1%, respectively. The minimum information analysis further showed that using 3 and 5 nearby wells was sufficient to bring the errors within 10% and 5% of the global optimum, respectively. One possible explanation is that NP represents only external climatic inputs, whereas the time variables mainly describe seasonality and long-term trends. By contrast, nearby well GWL reflect the combined effects of recharge, changes

in groundwater storage, river–groundwater interactions, and groundwater abstraction (Giese et al., 2020). Therefore, when the target and reference wells have similar hydrogeological settings and external drivers, nearby well GWL can provide more direct information on groundwater level variation (Haaf & Barthel, 2018). Ramirez et al. (2022) also noted that the relationship between meteorological conditions and groundwater levels is very weak and site dependent, making it difficult to reliably reconstruct the periodicity and long-term variation of a well using meteorological variables alone.

Pearson correlation was also used as an evaluation metric to assess the linear relationship between observed and imputed groundwater levels. Another common metric is Spearman correlation, which converts values into ranks and checks whether the two series have the same ordering of high and low values (Schober et al., 2018). Therefore, Spearman correlation can remain high even when the imputed series does not accurately match the observed groundwater level fluctuations. In contrast, Pearson correlation uses the values to measure how the two series vary together in a linear relationship, making it more suitable than Spearman correlation for this study.

Interestingly, increasing the number of nearby wells did not continuously improve model performance. In this study, the error gradually reached a plateau after approximately 5 to 8 nearby wells, the additional improvement after this was small. This may be because the first few high correlation wells already contained most of the relevant information, while adding wells with lower correlations could introduce noise. Previous studies have found that although multi-well models generally better than single-well models, a small number of high quality reference wells are often sufficient to establish an effective model, whereas including irrelevant wells may reduce predictive accuracy (Ramirez et al., 2023; Sejine et al., 2026).

NP rolling sum periods and the time variables mainly played compensatory roles in this study. Because NP accounts for both precipitation inputs and evapotranspiration losses, it may represent the actual atmospheric water deficit more effectively than precipitation alone (Canedo Rosso et al., 2025). Therefore, NP rolling sum over different time windows may help represent water availability and delayed groundwater responses to recharge. However, NP should not be interpreted as a direct estimate of recharge, as the calculation does not explicitly represent soil-water storage, snow accumulation and melt, or water movement through the unsaturated zone. In southern Sweden, groundwater recharge occurs mainly during winter, when evapotranspiration is relatively low, whereas groundwater conditions in central and northern Sweden are strongly influenced by snow accumulation and snowmelt. During summer, increased vegetation evapotranspiration may reduce the proportion of precipitation that contributes to recharge (Barthel et al., 2021; Nygren, Giese, et al., 2020). The time variables can therefore help the model distinguish the different groundwater responses that may result when similar NP values occur in different seasons. When nearby well information was limited,

extending NP1–3 to NP1–6 reduced RMSE and MAE by 5.4% and 11.7%, respectively. After NP and the time variables were added, only 3 to 5 nearby wells were required to achieve near-optimal performance, indicating that climatic variables can help the model identify the timing and overall trend of groundwater level fluctuations when spatial information is insufficient. Nevertheless, NP represents only potential water availability and cannot directly describe soil and geological conditions, actual groundwater recharge, or groundwater abstraction. When the model already contains several high-quality reference wells, these integrated responses are partly represented by the reference wells, and extending the NP accumulation window therefore provides only limited additional improvement.

4.3 Spatial variation of imputation performance

Higher imputation errors were mainly observed in central-eastern and southeastern Sweden, whereas most wells in western Sweden exhibited relatively better (Fig. 22). One possible explanation is that the groundwater levels in some wells were affected by local human activities.

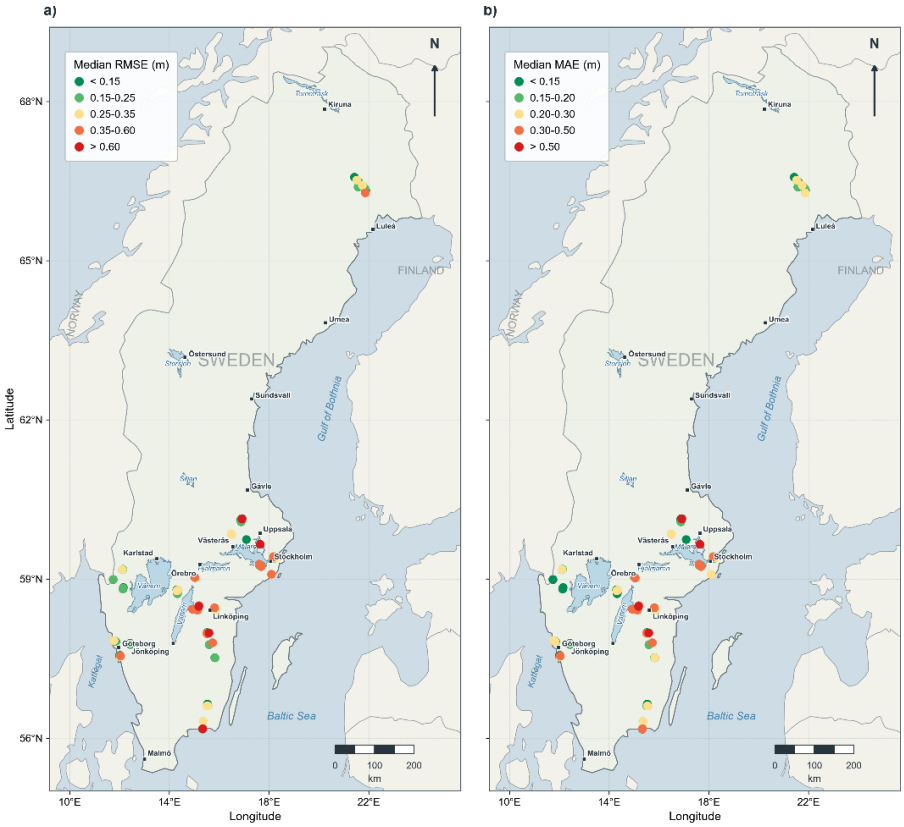


Fig. 22. Spatial distribution of imputation errors across Sweden: (a) median RMSE and (b) median MAE

For example, clear evidence of engineering activity was identified near Sigtuna_2 (Fig. 23a). Excavation, ground compaction, and drainage associated with construction activities may alter groundwater recharge and subsurface flow pathways (Mossmark et al., 2015). Motala_71 is located on a farm (Fig. 23b), and its groundwater level may be affected by crop

evapotranspiration, subsurface agricultural drainage, and possible groundwater abstraction for agricultural purposes. Agricultural drainage management can also directly regulate shallow groundwater levels (Wesström & Messing, 2007). As the Motala area is a typical agricultural region in Sweden, imputation in this region is more difficult than other regions due to anthropogenic influences. Moreover, historical satellite imagery shows that the area surrounding Ronnerby_39 and its nearby wells experienced at least four phases of forest harvesting over the approximately 15-year period from 2009 to 2024, occurring around 2013, 2018, 2022, and 2024 (Fig. 24). Deforestation may increase the proportion of precipitation and snowmelt entering the soil and groundwater system by reducing interception and evapotranspiration, potentially increasing groundwater levels, amplifying seasonal fluctuations, or altering groundwater response times (Schelker et al., 2013; Zannella et al., 2024).

However, these human activities can currently be considered only as possible explanations for the imputation errors and should not be regarded as confirmed direct causes. Small municipalities and rural communities in Sweden rely heavily on groundwater abstraction, but pumping records corresponding to these wells are often unavailable (Barthel et al., 2021; Kløve et al., 2017). It is therefore difficult to quantify the influence of groundwater abstraction on these wells. In addition to human activities, the small size and spatial isolation of aquifers may cause different wells to respond differently to precipitation and regional groundwater variations (Barthel et al., 2021). Consequently, the mechanisms responsible for the high error wells should be further examined by comparing the timing of abrupt changes in their groundwater level records with well use, aquifer type, and nearby groundwater abstraction or engineering records.

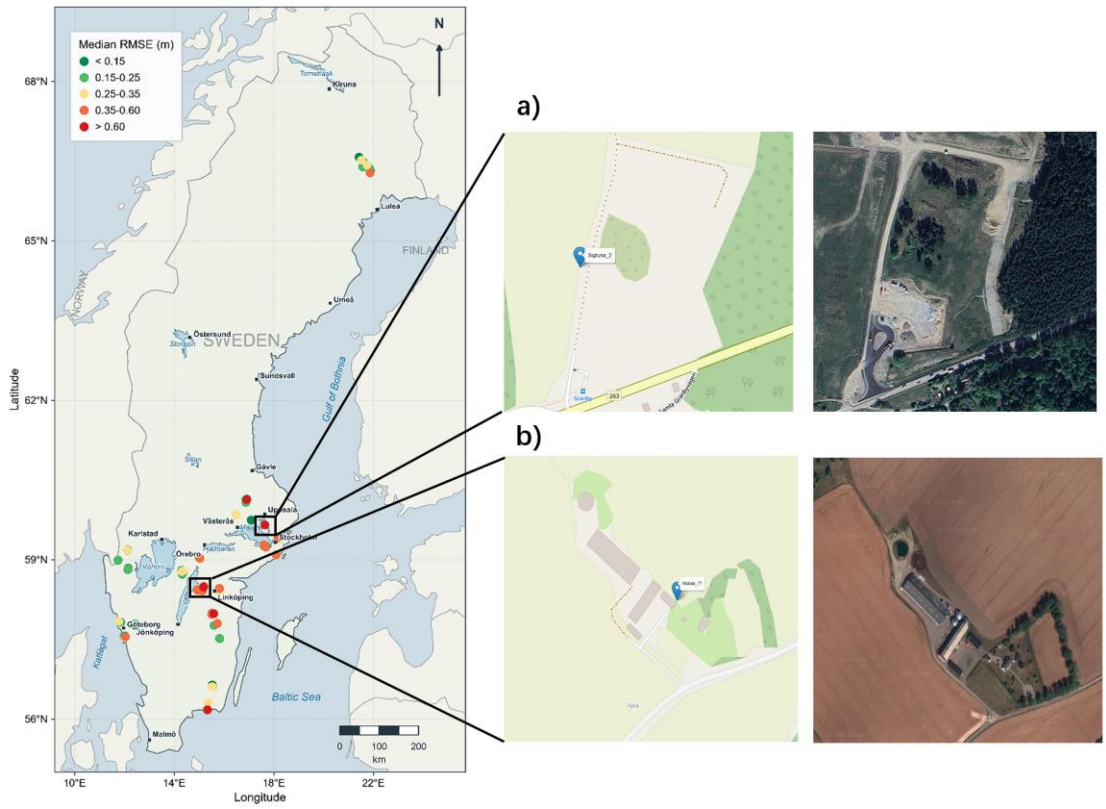


Fig. 23. Engineering and agricultural activities surrounding high error wells (a) Sigtuna_2 and (b) Motala_71. Location and satellite maps from OpenStreetMap and Google Earth.

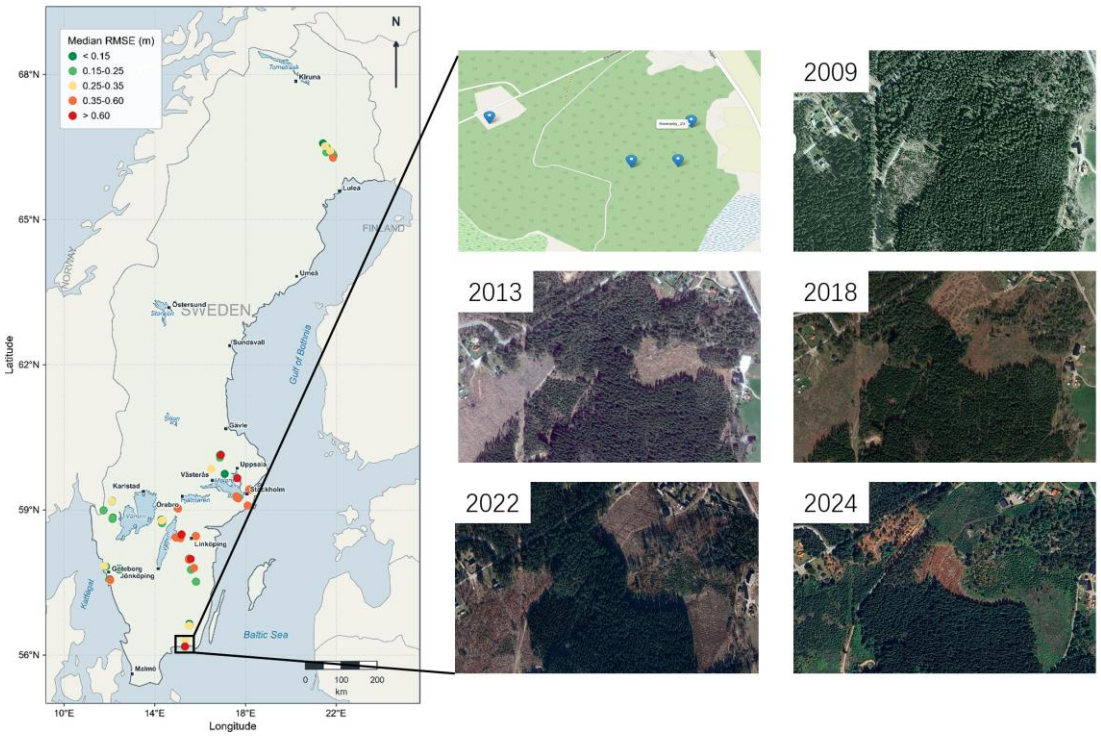


Fig. 24. Historical deforestation activities surrounding Ronnerby_39 between 2009 and 2024. Location and satellite maps from OpenStreetMap and Google Earth.

4.4 Limitations and suggestions for future studies

Although this study analyse the general effects of different input variables on a national scale, several limitations should be considered and discussed.

Firstly, the same NP accumulation windows were applied throughout Sweden without evaluating different climatic regions separately. In southern Sweden, recharge occurs mainly through autumn and winter rainfall, whereas central and northern regions are more strongly influenced by snow storage, soil freezing, and spring snowmelt (Nygren et al., 2022; Nygren et al., 2021; Nygren, Giese, et al., 2020). The NP variables did not distinguish rainfall from snowfall or represent the delayed release of snowmelt. Therefore, future studies could evaluate specific accumulation windows for different climate regions and include additional variables describing snow accumulation, snowmelt, and soil freezing conditions.

Secondly, this study evaluated NP accumulation windows on monthly scale but did not test shorter periods, such as 5, 10, or 20 days. Previous studies have found groundwater responses ranging from less than one day to several weeks, depending on groundwater depth, aquifer properties, and recharge pathways (Jiménez-Martínez et al., 2013; Lee et al., 2006; Lu et al., 2022). Monthly windows may be too coarse to represent rapid responses to individual precipitation events in shallow, highly permeable, or fractured aquifers. Future experiments could compare daily, weekly and monthly NP windows and explore their contribution to the model across different aquifer types and climate regions.

Moreover, the 5% and 10% tolerance thresholds used in this study require further investigation and validation. They were adopted to identify performance close to the best tested configuration, rather than to define universally acceptable reconstruction accuracy. The same relative tolerances are not equally suitable for all evaluation metrics. RMSE and MAE showed clearer separation between configurations, whereas Pearson correlation varied little among several combined-input configurations, allowing models with different errors to satisfy the same correlation threshold. Future work should therefore define tolerances according to each metric's mathematical properties, observed distribution and sensitivity to changes in model performance, together with the accuracy required for the intended application.

Finally, reference wells were selected within a fixed 50 km radius from target well. Although this is a consistent selection criterion, geographical distance does not always mean close by wells share a similarity of hydraulic connectivity or groundwater dynamics. Nearby wells may belong to different aquifers, while more distant wells may exhibit stronger temporal similarity (Ramirez et al., 2023). In addition, because the density of monitoring wells in northern Sweden is much lower than in the south, setting the search radius to 50km may result in no reference wells being found in northern Sweden. Future studies should combine hydrograph dynamics similarity or test different radius in different regions rather than setting one consistent distance.

5. Conclusions

This study shows that groundwater levels from nearby wells are the main source of information for missForest imputation. Precipitation and time variables can provide extra information, but they cannot fully replace groundwater level data. When precipitation and time variables were included, 3 nearby wells achieved RMSE and MAE values within 10% of their best tested values, while 5 nearby wells were needed to reach the 5% threshold. When the number of nearby wells increased to about 8 and adding more wells provided only a small improvement. Short- to medium-term accumulated NP was more useful than long-term information. NP1–NP3 and NP1–NP6 are the optimal choices. NP12 provided useful extra information only in some cases.

On the 140 common target wells, missForest achieved lower median RMSE and MAE and higher median Pearson correlation than ARCHI. The missForest combined nearby groundwater levels, NP, and time variables, but ARCHI only used groundwater-level records. These additional variables may have contributed to better performance.

Minimum-information requirements depended on the evaluation metric. Pearson correlation satisfied the tolerance criteria with fewer reference wells than RMSE and MAE, showing that the 5% and 10% thresholds should therefore be treated as thresholds awaiting determination through further research. Future work should refine these criteria and assess whether imputation preserves SGI-based drought duration, severity and intensity.

Finally, the spatial distribution of imputation errors was heterogeneous, with relatively large errors observed at some wells. This indicates that local factors, such as groundwater abstraction, forest harvesting, agricultural drainage, engineering activities, and differences in aquifer properties, may not be fully represented by the current variables. Future studies should therefore include additional information on local hydrogeological conditions and human activities. They should also prioritise a small number of high-quality reference wells with strong hydrological relationships to the target well, rather than simply using longer climatic accumulation windows or more complex models.

Use of Generative AI

Generative AI tools were used during the thesis writing as supporting tools for translating and polishing English, organizing section's structure, and obtaining suggestions for codes and analysis. All AI-generated suggestions were critically reviewed and modified by the author before being integrated into the thesis.

References

Arriagada, P., Karelovic, B., & Link, O. (2021). Automatic gap-filling of daily streamflow time series in data-scarce regions using a machine learning algorithm. *Journal of Hydrology*, *598*, 126454.

Asgharinia, S., & Petroselli, A. (2020). A comparison of statistical methods for evaluating missing data of monitoring wells in the Kazeroun Plain, Fars Province, Iran. *Groundwater for Sustainable Development*, *10*, 100294.

Babre, A., Kalvāns, A., Avotniece, Z., Retiķe, I., Bikše, J., Popovs, K., Jemeljanova, M., Zelenkevičs, A., & Dēliņa, A. (2022). The use of predefined drought indices for the assessment of groundwater drought episodes in the Baltic States over the period 1989–2018. *Journal of Hydrology: Regional Studies*, *40*, 101049.

Bakke, S. J., Ionita, M., & Tallaksen, L. M. (2020). The 2018 northern European hydrological drought and its drivers in a historical perspective. *Hydrology and Earth System Sciences*, *24*(11), 5621-5653. https://doi.org/10.5194/hess-24-5621-2020

Barthel, R., Stangefelt, M., Giese, M., Nygren, M., Seftigen, K., & Chen, D. (2021). Current understanding of groundwater recharge and groundwater drought in Sweden compared to countries with similar geology and climate. *Geografiska Annaler: Series A, Physical Geography*, *103*(4), 323-345. https://doi.org/10.1080/04353676.2021.1969130

Bikše, J., Retike, I., Haaf, E., & Kalvāns, A. (2023). Assessing automated gap imputation of regional scale groundwater level data sets with typical gap patterns. *Journal of Hydrology*, *620*. https://doi.org/10.1016/j.jhydrol.2023.129424

Bloomfield, J. P., & Marchant, B. P. (2013). Analysis of groundwater drought building on the standardised precipitation index approach. *Hydrology and Earth System Sciences*, *17*(12), 4769-4787. https://doi.org/10.5194/hess-17-4769-2013

Boutt, D. F. (2017). Assessing hydrogeologic controls on dynamic groundwater storage using long-term instrumental records of water table levels. *Hydrological Processes*, *31*(7), 1479-1497.

Breiman, L. (2001). Random forests. *Machine learning*, *45*(1), 5-32.

Canedo Rosso, C., Nyberg, L., & Pechlivanidis, I. (2025). Drought hazard assessment across Sweden's diverse hydro-climatic regimes. *Natural Hazards and Earth System Sciences*, *25*(11), 4577-4592.

Chucuya, S., Pacci, R., Bustincio, B., Taya-Acosta, E., Alfonso-Morales, W., Huayna, G., Pino-Vargas, E., Ingol-Blanco, E., Mora, A., Torres-Martínez, J. A., Narvaez-Montoya, C., & Mahlknecht, J. (2025). Reconstructing aquifer dynamics with machine learning: Linking irrigation expansion to groundwater decline in a data-scarce hyper-arid region. *Agricultural Water Management*, *322*. https://doi.org/10.1016/j.agwat.2025.110018

Collenteur, R. A., Bakker, M., Calje, R., Klop, S. A., & Schaars, F. (2019). Pastas: Open Source Software for the Analysis of Groundwater Time Series. *Ground Water*, *57*(6), 877-885. https://doi.org/10.1111/gwat.12925

Collenteur, R. A., Vonk, M. A., & Haaf, E. (2025). Quantification and Analysis of Hydrograph Behavior Using Groundwater Signatures. *Ground Water*, *63*(5), 779-789.

https://doi.org/10.1111/gwat.13486

Cornes, R. C., van der Schrier, G., van den Besselaar, E. J. M., & Jones, P. D. (2018). An Ensemble Version of the E-OBS Temperature and Precipitation Data Sets. *Journal of Geophysical Research: Atmospheres*, *123*(17), 9391-9409. https://doi.org/https://doi.org/10.1029/2017JD028200

Du, X., Fang, M., Lv, H., Cheng, T., Hong, P., & Liu, C. (2019). Effect of snowmelt infiltration on groundwater recharge in a seasonal soil frost area: a case study in Northeast China. *Environmental monitoring and assessment*, *191*(3), 151.

Dwivedi, D., Mital, U., Faybishenko, B., Dafflon, B., Varadharajan, C., Agarwal, D., Williams, K. H., Steefel, C. I., & Hubbard, S. S. (2022). Imputation of contiguous gaps and extremes of subhourly groundwater time series using random forests. *Journal of Machine Learning for Modeling and Computing*, *3*(2).

Englund, J., Knutsson, G., & Soveri, J. (1986). Studies on groundwater recharge in Finland, Norway and in Sweden. Proceedings of a Workshop, Mariehamn, Åland, Finland,

Evans, S., Williams, G. P., Jones, N. L., Ames, D. P., & Nelson, E. J. (2020). Exploiting Earth Observation Data to Impute Groundwater Level Measurements with an Extreme Learning Machine. *Remote Sensing*, *12*(12). https://doi.org/10.3390/rs12122044

Eveborn, D., Vikberg, E., Thunholm, B., Hjerne, C., & Gustafsson, M. (2017). Grundvattenbildning och grundvattentillgång i Sverige. *Uppsala: Sveriges Geologiska Undersöking.(SGUs diarie-nr: 21-2925/2016*.

Fredén, C., Wastenson, L., Arnberg, U., & Cramér, M. (2009). *Berg och Jord, Sveriges Nationalatlas (SNA)*.

Garciarena, U., & Santana, R. (2017). An extensive analysis of the interaction between missing data types, imputation methods, and supervised classifiers. *Expert Systems with Applications*, *89*, 52-65.

Geological Survey of Sweden. (2024a). *Bedrock 1:1 million* ([Bedrock geology vector data (GeoPackage/GeoJSON)]. Geological Survey of Sweden (SGU). https://www.sgu.se/en/products/geological-data/berggrund--geologisk-data/bedrock/

Geological Survey of Sweden. (2024b). *Grundvatten och förändrat klimat*. https://www.sgu.se/grundvatten/grundvatten-och-forandrat-klimat/

Geological Survey of Sweden. (2024c). *Jordarter 1:1 miljon* ([Soil type geology vector data (GeoPackage/GeoJSON)]. Geological Survey of Sweden (SGU). https://www.sgu.se/produkter-och-tjanster/geologiska-data/jordarter--geologiska-data/jordartsdata/

Giese, M., & Barthel, R. (2021). Review: Saltwater intrusion in fractured crystalline bedrock. *Hydrogeology Journal*, *29*(7), 2313-2328. https://doi.org/10.1007/s10040-021-02396-y

Giese, M., Haaf, E., Heudorfer, B., & Barthel, R. (2020). Comparative hydrogeology–reference analysis of groundwater dynamics from neighbouring observation wells. *Hydrological Sciences Journal*, *65*(10), 1685-1706.

Glas, R., Herdman, L., Cook, S., Howlader, A., & Masterson, K. (2026). Hazard potential of compound flooding from rainfall, storm surge, and groundwater in coastal New York and Connecticut. *Nat. Hazards Earth Syst. Sci.*, *26*(5), 2169-2188.

https://doi.org/10.5194/nhess-26-2169-2026

Grip, H., & Rodhe, A. (2000). *Vattnets väg från regn till bäck*. Hallgren och Fallgren, Uppsala.

Guo, M., Yue, W., Wang, T., Zheng, N., & Wu, L. (2021). Assessing the use of standardized groundwater index for quantifying groundwater drought over the conterminous US. *Journal of Hydrology*, *598*, 126227.

Haaf, E., & Barthel, R. (2018). An inter-comparison of similarity-based methods for organisation and classification of groundwater hydrographs. *Journal of Hydrology*, *559*, 222-237.

Haaf, E., Giese, M., Reimann, T., & Barthel, R. (2023). Data-Driven Estimation of Groundwater Level Time-Series at Unmonitored Sites Using Comparative Regional Analysis. *Water Resources Research*, *59*(7), e2022WR033470. https://doi.org/https://doi.org/10.1029/2022WR033470

Healy, R. W. (2010). *Estimating groundwater recharge*. Cambridge university press.

Jiménez-Martínez, J., Longuevergne, L., Le Borgne, T., Davy, P., Russian, A., & Bour, O. (2013). Temporal and spatial scaling of hydraulic response to recharge in fractured aquifers: Insights from a frequency domain analysis. *Water Resources Research*, *49*(5), 3007-3023. https://doi.org/10.1002/wrcr.20260

Karin, E., Adrian, M. H., Johan, K., & Jannike, A. (2015). Unequal ice-sheet erosional impacts across low-relief shield terrain in northern Fennoscandia. *Geomorphology*, *233*, 64-74. https://doi.org/https://doi.org/10.1016/j.geomorph.2014.09.024

Karna, L.-B., Johan, M. B., & Peter, J. (2013). Stratigraphic Landscape Analysis and geomorphological paradigms: Scandinavia as an example of Phanerozoic uplift and subsidence. *Global and Planetary Change*, *100*, 153-171. https://doi.org/https://doi.org/10.1016/j.gloplacha.2012.10.015

Kleman, J., Arjen, P. S., & Jan, L. (2008). Patterns of Quaternary ice sheet erosion and deposition in Fennoscandia and a theoretical framework for explanation. *Geomorphology*, *97*(1), 73-90. https://doi.org/https://doi.org/10.1016/j.geomorph.2007.02.049

Kløve, B., Kvitsand, H. M. L., Pitkänen, T., Gunnarsdottir, M. J., Gaut, S., Gardarsson, S. M., Rossi, P. M., & Miettinen, I. (2017). Overview of groundwater sources and water-supply systems, and associated microbial pollution, in Finland, Norway and Iceland. *Hydrogeology Journal*, *25*(4), 1033-1044.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World map of the Köppen-Geiger climate classification updated.

Lee, L. J. E., Lawrence, D. S. L., & Price, M. (2006). Analysis of water-level response to rainfall and implications for recharge pathways in the Chalk aquifer, SE England. *Journal of Hydrology*, *330*(3), 604-620. https://doi.org/https://doi.org/10.1016/j.jhydrol.2006.04.025

Levy, Z. F., Glas, R. L., Stagnitta, T. J., & Terry, N. (2025). ARCHI: A New R Package for Automated Imputation of Regionally Correlated Hydrologic Records. *Ground Water*, *63*(4), 595-610. https://doi.org/10.1111/gwat.13474

Lu, M., Rogiers, B., Beerten, K., Gedeon, M., & Huysmans, M. (2022). Exploring river–

aquifer interactions and hydrological system response using baseflow separation, impulse response modeling, and time series analysis in three temperate lowland catchments. *Hydrology and Earth System Sciences*, *26*(13), 3629-3649. https://doi.org/10.5194/hess-26-3629-2022

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

Meggiorin, M., Passadore, G., Bertoldo, S., Sottani, A., & Rinaldo, A. (2023). Comparison of Three Imputation Methods for Groundwater Level Timeseries. *Water*, *15*(4). https://doi.org/10.3390/w15040801

Mossmark, F., Ericsson, L. O., Norin, M., & Dahlström, L.-O. (2015). Hydrochemical changes caused by underground constructions—a case study of the Kattleberg rail tunnel. *Engineering Geology*, *191*, 86-98.

Nair, A. S., Tallaksen, L. M., Bjørbæk, T. A., & Collenteur, R. (2025). Gap-filling groundwater level time series of irregular temporal resolution using physical modeling (Pastas) and simple statistical scaling. EGU General Assembly Conference Abstracts,

Nair, A. S., Yebra, A. S., Cid-Escobar, D., Henao Casas, J. D., Giese, M., Huysmans, M., & Tallaksen, L. M. (2026). EDGE-1: The European Dataset for Groundwater Drought Evaluation (Version 1). *Scientific Data*. https://doi.org/10.17605/OSF.IO/SKXG5

Noori, A. R., & Singh, S. K. (2021). Spatial and temporal trend analysis of groundwater levels and regional groundwater drought assessment of Kabul, Afghanistan. *Environmental Earth Sciences*, *80*(20). https://doi.org/10.1007/s12665-021-10005-0

Nygren, M., Barthel, R., Allen, D. M., & Giese, M. (2022). Exploring groundwater drought responsiveness in lowland post-glacial environments. *Hydrogeology Journal*, *30*(7), 1937-1961. https://doi.org/10.1007/s10040-022-02521-5

Nygren, M., Giese, M., & Barthel, R. (2021). Recent trends in hydroclimate and groundwater levels in a region with seasonal frost cover. *Journal of Hydrology*, *602*. https://doi.org/10.1016/j.jhydrol.2021.126732

Nygren, M., Giese, M., Kløve, B., Haaf, E., Rossi, P. M., & Barthel, R. (2020). Changes in seasonality of groundwater level fluctuations in a temperate-cold climate transition zone. *Journal of Hydrology X*, *8*, 100062.

Nygren, M., Markus, G., Bjørn, K., Ezra, H., Pekka, M. R., & Roland, B. (2020). Changes in seasonality of groundwater level fluctuations in a temperate-cold climate transition zone. *Journal of Hydrology X*, *8*, 100062. https://doi.org/https://doi.org/10.1016/j.hydroa.2020.100062

Parry, S., Mackay, J. D., Chitson, T., Hannaford, J., Magee, E., Tanguy, M., Bell, V. A., Facer-Childs, K., Kay, A., Lane, R., Moore, R. J., Turner, S., & Wallbank, J. (2024). Divergent future drought projections in UK river flows and groundwater levels. *Hydrol. Earth Syst. Sci.*, *28*(3), 417-440. https://doi.org/10.5194/hess-28-417-2024

Rajaee, T., Ebrahimi, H., & Nourani, V. (2019). A review of the artificial intelligence methods in groundwater level modeling. *Journal of Hydrology*, *572*, 336-351. https://doi.org/10.1016/j.jhydrol.2018.12.037

Ramirez, S. G., Williams, G. P., & Jones, N. L. (2022). Groundwater Level Data Imputation Using Machine Learning and Remote Earth Observations Using Inductive Bias.

Remote Sensing, *14*(21). https://doi.org/10.3390/rs14215509

Ramirez, S. G., Williams, G. P., Jones, N. L., Ames, D. P., & Radebaugh, J. (2023). Improving Groundwater Imputation through Iterative Refinement Using Spatial and Temporal Correlations from In Situ Data with Machine Learning. *Water*, *15*(6). https://doi.org/10.3390/w15061236

Retike, I., Bikše, J., Kalvāns, A., Dēliņa, A., Avotniece, Z., Zaadnoordijk, W. J., Jemeljanova, M., Popovs, K., Babre, A., & Zelenkevičs, A. (2022). Rescue of groundwater level time series: How to visually identify and treat errors. *Journal of Hydrology*, *605*, 127294.

Richts, A., Struckmeier, W. F., & Zaepke, M. (2011). *WHYMAP and the Groundwater Resources Map of the World 1:25,000,000 , bookTitle= Sustaining Groundwater Resources: A Critical Element in the Global Water Crisis*. Springer Netherlands. https://doi.org/10.1007/978-90-481-3426-7_10

Rubel, F., Brugger, K., Haslinger, K., & Auer, I. (2017). The climate of the European Alps: Shift of very high resolution Köppen-Geiger climate zones 1800–2100. *Meteorologische Zeitschrift*, *26*(2), 115-125.

Schelker, J., Kuglerová, L., Eklöf, K., Bishop, K., & Laudon, H. (2013). Hydrological effects of clear-cutting in a boreal forest–Snowpack dynamics, snowmelt and streamflow responses. *Journal of Hydrology*, *484*, 105-114.

Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation Coefficients: Appropriate Use and Interpretation. *Anesth Analg*, *126*(5), 1763-1768. https://doi.org/10.1213/ANE.0000000000002864

Schoppa, L., Disse, M., & Bachmair, S. (2020). Evaluating the performance of random forest for large-scale flood discharge simulation. *Journal of Hydrology*, *590*. https://doi.org/10.1016/j.jhydrol.2020.125531

Sejine, H., Ouali, A., Pistre, S., & Anane, M. (2026). Bridging gaps in sparse groundwater data: classical, ensemble, and deep learning approaches for single- and multi-well imputation. *Frontiers in Water*, *7*. https://doi.org/10.3389/frwa.2025.1726853

Sharma, Y. K., Kim, S., Tayerani Charmchi, A. S., Kang, D., & Batelaan, O. (2025). Strategic imputation of groundwater data using machine learning: Insights from diverse aquifers in the Chao-Phraya River Basin. *Groundwater for Sustainable Development*, *28*. https://doi.org/10.1016/j.gsd.2024.101394

Shiklomanov, I. A. (2000). Appraisal and assessment of world water resources. *Water international*, *25*(1), 11-32.

Sidibe, M., Dieppois, B., Mahe, G., Paturel, J.-E., Amoussou, E., Anifowose, B., & Lawler, D. (2018). Trend and variability in a new, reconstructed streamflow dataset for West and Central Africa, and climatic interactions, 1950–2005. *Journal of Hydrology*, *561*, 478-493.

Solanki, H., & Mishra, V. (2026). Filling streamflow data gaps in Indian catchments using machine learning and K-means clustering. *Water Resources Research*, *62*(3), e2025WR042210.

Stekhoven, D. J., & Bühlmann, P. (2012). MissForest—non-parametric missing value imputation for mixed-type data. *Bioinformatics*, *28*(1), 112-118.

Stensen, K., Krunegård, A., Rasmusson, K., Matti, B., & Hjerdt, N. (2019). *Sveriges*

vattentillgång utifrån perspektivet vattenbrist och torka:–Delrapport 1 i regeringsuppdrag om åtgärder för att motverka vattenbrist i ytvattentäkter.

Taie Semiromi, M., & Koch, M. (2019). Reconstruction of groundwater levels to impute missing values using singular and multichannel spectrum analysis: application to the Ardabil Plain, Iran. *Hydrological Sciences Journal*, *64*(14), 1711-1726. https://doi.org/10.1080/02626667.2019.1669793

Vremec, M., Collenteur, R., & Birk, S. (2024). PyEt v1. 3.1: a Python package for the estimation of potential evapotranspiration. *Geoscientific Model Development Discussions*, *2024*, 1-26.

Wendt, D. E., Van Loon, A. F., Bloomfield, J. P., & Hannah, D. M. (2020). Asymmetric impact of groundwater use on groundwater droughts. *Hydrology and Earth System Sciences Discussions*, *2020*, 1-23.

Wesström, I., & Messing, I. (2007). Effects of controlled drainage on N and P losses and N dynamics in a loamy sand with spring crops. *Agricultural Water Management*, *87*(3), 229-240.

Yadav, M. L., & Roychoudhury, B. (2018). Handling missing values: A study of popular imputation packages in R. *Knowledge-Based Systems*, *160*, 104-118.

Zannella, A., Wallin, M. B., Sikström, U., Arvidsson, E., & Eklöf, K. (2024). Ditch cleaning in boreal catchments: Impacts on water chemistry and dissolved greenhouse gases in runoff. *Forest Ecology and Management*, *569*, 122146.

Zhou, Y., Tang, Q., & Zhao, G. (2023). Gap infilling of daily streamflow data using a machine learning algorithm (MissForest) for impact assessment of human activities. *Journal of Hydrology*, *627*. https://doi.org/10.1016/j.jhydrol.2023.130404