Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Groundwater Forecasting Under Missing Data Uncertainty: A Multi-Horizon Assessment Using Hyperbolic Deep Learning

Kübra Küllahcı

onerk@itu.edu.tr

Erzurum Teknik Üniversitesi: Erzurum Teknik Universitesi https://orcid.org/0000-0003-4699-5878

Kürşat Şekerci

Bingöl Üniversitesi: Bingol Universitesi

Research Article

Keywords: Hyperbolic deep learning, groundwater, multi-horizon prediction, data imputation

Posted Date: July 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10077915/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Groundwater Forecasting Under Missing Data Uncertainty: A Multi-Horizon Assessment Using Hyperbolic Deep Learning

Kübra KÜLLAHCI[1*], Kürşat ŞEKERCİ [2]

[1] Department of Civil Engineering Hydraulics Division /Erzurum Technical University / Erzurum, Türkiye

[2] Department of Civil Engineering Hydraulics Division /Bingöl University / Bingöl, Türkiye

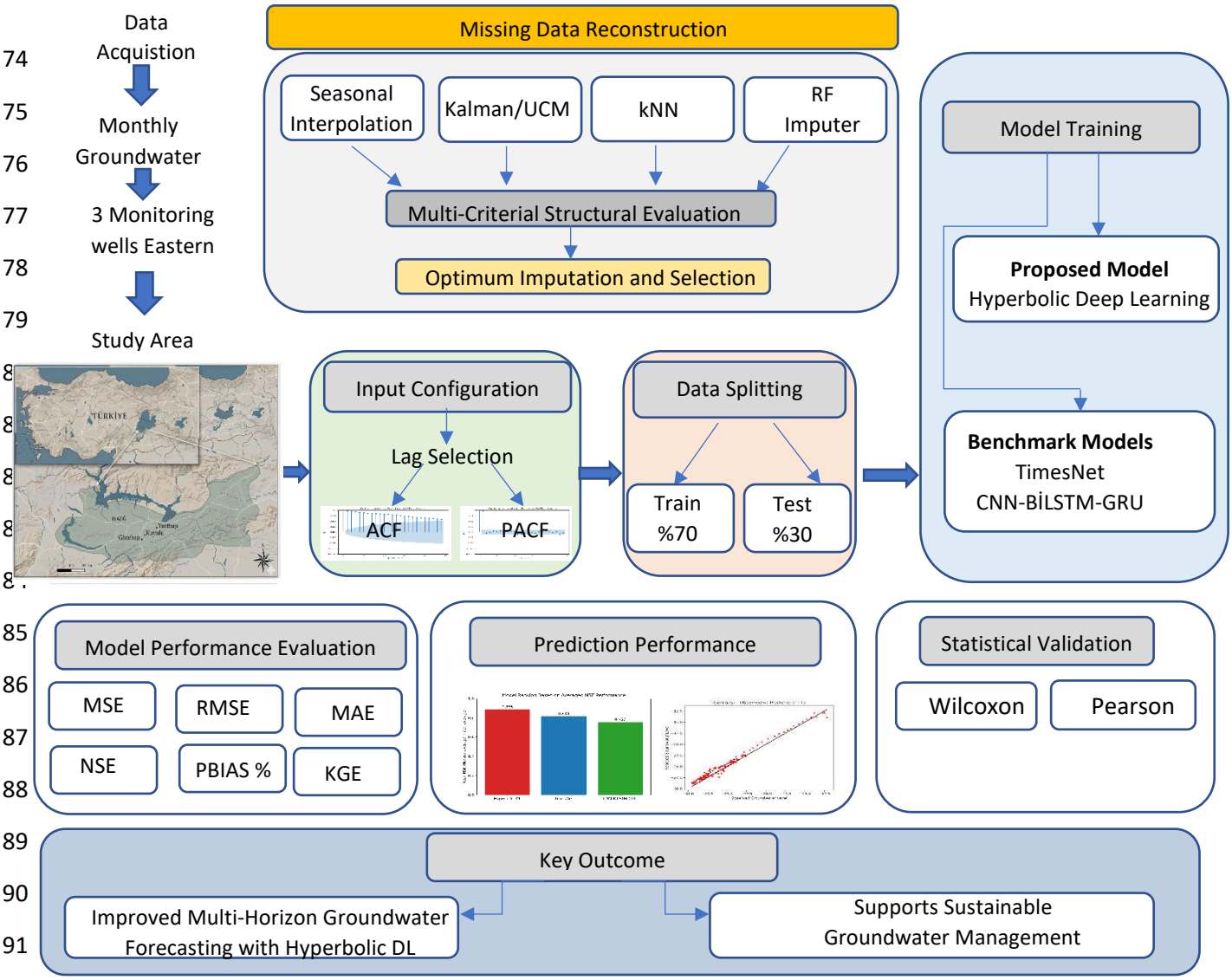
Corresponding Autor : Kübra KÜLLAHCI (onerk@itu.edu.tr)

Abstract: Groundwater resources are increasingly exposed to climatic variability, anthropogenic pressure, and unsustainable abstraction, creating an urgent need for reliable groundwater level (GWL) forecasting frameworks that can support adaptive aquifer management. Despite recent advances in deep learning, accurate multi-horizon groundwater forecasting remains hindered by two persistent limitations: incomplete monitoring records caused by observational gaps and the accumulation of prediction uncertainty at longer forecast horizons. Addressing these challenges requires forecasting frameworks capable of simultaneously handling missing information and preserving complex temporal structures in groundwater dynamics. This study proposes an integrated multi-horizon forecasting framework that combines optimal missing-data reconstruction with hyperbolic deep representation learning for robust groundwater prediction under incomplete observational conditions. Monthly in-situ GWL observations from three monitoring wells in eastern Türkiye covering the period 1978–2015 were used as the experimental dataset. Multiple imputation approaches were systematically evaluated using quantitative reconstruction metrics to identify the most reliable data completion strategy prior to forecasting. The reconstructed time series were subsequently used to generate predictions for five lead times ranging from t+1 to t+5 months. The predictive capability of the proposed hyperbolic framework was benchmarked against hybrid CNN-BiLSTM–GRU and TimesNet architectures under comprehensive statistical evaluation. Results demonstrated that the hyperbolic model consistently improved predictive stability across extended forecast horizons, particularly under long-range forecasting conditions where conventional architectures exhibited stronger error accumulation. The proposed framework produced lower RMSE values at longer lead times (particularly t+5), maintained stronger temporal consistency, and reduced forecast degradation across sequential prediction steps. Statistical validation using the Wilcoxon signed-rank test confirmed the significance and robustness of the forecasting results, while Pearson correlation analysis and residual diagnostics indicated strong agreement between observed and predicted groundwater levels across all prediction horizons. The findings suggest that hyperbolic representation learning provides a structurally efficient mechanism for capturing hierarchical temporal dependencies in groundwater systems and offers a promising pathway for improving multi-step forecasting reliability in datasets affected by missing observations and long-term monitoring uncertainties.

Key words: Hyperbolic deep learning, groundwater, multi-horizon prediction, data imputation

Highlights

- Proposes a hyperbolic deep learning framework for multi-horizon GWL prediction.
- Optimizes missing data reconstruction prior to neural forecasting.
- Reduces error propagation at extended t+5 lead times.
- Outperforms LSTM-GRU and TimesNet across all evaluation metrics.
- Statistically validated using Wilcoxon tests, Pearson correlation, and residual diagnostics.



Graphical Abstract Description: This visual summary serves as a pivotal entry point into the research, offering a concise overview of the study's core findings and methodologies. The study presents a comprehensive framework for multi-horizon groundwater level forecasting under incomplete observation conditions. Monthly groundwater level records collected from monitoring wells were first examined, where missing observations represented a major challenge for reliable groundwater assessment. To overcome this limitation, multiple reconstruction approaches, including Seasonal Interpolation, Kalman/UCM, kNN imputation, and Random Forest imputation, were comparatively evaluated to identify the most suitable strategy for recovering missing observations. Following data reconstruction, preprocessing and lag-based input generation procedures were applied to prepare the dataset for forecasting. Advanced forecasting architectures were then implemented to predict groundwater levels across multiple lead times ranging from t+1 to t+5. The forecasting stage incorporates Hyperbolic Deep Learning together with benchmark deep learning architectures, including TimesNet and CNN-BiLSTM-GRU, within a unified framework to evaluate forecasting robustness under increasing prediction horizons. To the best of our knowledge, this study represents one of the earliest applications of Hyperbolic Deep Learning for groundwater level forecasting. The final stage summarizes model performance using statistical metrics, residual analysis, and comparative evaluation procedures. Statistical significance and predictive reliability were further assessed using correlation analysis, residual diagnostics, and hypothesis testing procedures. The workflow highlights how

improved data reconstruction directly influences forecasting accuracy and robustness. Overall, the proposed framework provides a practical strategy for enhancing groundwater level prediction reliability and supporting sustainable groundwater resource management under data-limited conditions.

1. Introduction

Groundwater represents one of the most critical buffers against hydroclimatic variability, sustaining drinking water supply, agricultural irrigation, and industrial production across both developed and developing regions. (Konikow, 2013; Pointet, 2022)As climate change intensifies the frequency and magnitude of extreme events, international policy frameworks—including the United Nations Framework Convention on Climate Change (UNFCCC) and the Sustainable Development Goals (notably Goals 6 and 13)—increasingly emphasize systematic monitoring and anticipatory management of water resources. (Cadman, 2018; Nations, 2015) Within this context, reliable forecasting of groundwater level (GWL) dynamics is indispensable for sustainable abstraction planning, drought preparedness, and the resilient design and operation of water management infrastructure(Latuamury et al., 2025; Şekerci et al., 2023, 2024)

Despite its central importance, monthly-scale GWL prediction remains technically demanding due to the inherently nonlinear and multi-scale nature of subsurface hydrological processes. Groundwater dynamics emerge from complex interactions among precipitation recharge, delayed percolation and lag effects, the filtering behavior of the unsaturated zone, and anthropogenic pressures such as overextraction and land-use change. At monthly temporal resolution, short-term event-driven variability coexists with longer-term seasonal and interannual signals, requiring predictive models to simultaneously represent low-frequency periodic components and high-frequency fluctuations. (Liu et al., 2025)This multi-scale complexity places substantial demands on model expressiveness, stability, and generalization capacity.

Traditionally, groundwater level (GWL) forecasting has relied on classical time-series approaches, particularly models within the ARIMA family (e.g., ARIMA and SARIMA). (Gibrilla et al., 2018)These methods have proven effective in capturing linear temporal dependencies and seasonality in univariate settings. However, the inherently multi-scale, nonlinear, and multi-driver nature of groundwater systems shaped by precipitation recharge, evapotranspiration, abstraction, and land-use dynamics often exceeds the representational capacity of purely linear statistical models.(Alsumaiei, 2020; Raja & Edwin, 2025)

In parallel with advances in computational hydrology, machine learning (ML) and deep learning (DL) techniques have been increasingly adopted for GWL prediction.(Küllahcı & Altunkaynak, 2023, 2024a) A recent systematic review and meta-analysis covering the period 2000–2023 reports a marked rise in ML/DL applications for groundwater forecasting, with hybrid architectures gaining particular prominence due to their ability to integrate spatial and temporal information(Afrifa et al., 2022; Ahmadi et al., 2022; Khan et al., 2023; Küllahcı & Altunkaynak, 2023; Tao et al., 2022a; Uc-Castillo et al., 2023). Among DL approaches, recurrent neural networks (RNNs), especially Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models, have become standard benchmarks owing to their capacity to model sequential dependencies and memory effects. (Amiri et al., 2023; Bowes et al., 2019; Chang et al., 2025; Küllahcı & Altunkaynak, 2024a)Empirical evaluations have demonstrated

that model performance in groundwater applications depends not only on network architecture but also on data availability, predictor selection, and forecasting strategy (Altunkaynak et al., 2025; Wunsch et al., 2021a).

More recently, attention-based architectures such as Transformers have been introduced to hydrogeological forecasting, offering enhanced capability for capturing long-range dependencies in multivariate time series. For instance, Transformer-based models have been shown to outperform conventional CNN and LSTM configurations in selected groundwater prediction scenarios(Küllahcı & Altunkaynak, 2024b; Sun et al., 2024). In addition, architectures such as TimesNet, which model temporal variation through two-dimensional representations of time-series structure, have demonstrated strong performance in general time-series benchmarks and are increasingly considered in hydrological contexts (Wu et al., n.d.).

Several methodological gaps remain evident in the current hydrological modeling literature concerning groundwater level (GWL) forecasting.(Boo et al., 2024) First, although missing data is a recurrent issue in hydrogeological time series, the impact of different imputation strategies on downstream forecasting performance has been addressed only superficially in many studies. In practice, datasets are frequently completed using linear interpolation or other simple gap-filling techniques without systematically evaluating how these approaches alter variance structure, autocorrelation properties, or intrinsic temporal components of the series. As a result, the potential indirect influence of reconstructed data on multi-horizon forecasting performance remains insufficiently clarified.(Tao et al., 2022b)Second, relatively limited attention has been devoted to model behavior across extended prediction horizons. As the forecasting horizon increases, small initial discrepancies may accumulate in a nonlinear manner, leading to progressive error amplification and reduced predictive reliability. However, most existing studies evaluate model performance at a single forecast step, offering limited insight into the dynamic evolution of error propagation and model stability across sequential horizons. (Wunsch et al., 2021b)Third, the role of representational geometry in hydrological time-series modeling has not yet been comprehensively examined. While hyperbolic representation learning has demonstrated strong theoretical and empirical advantages in domains characterized by hierarchical and scale-free structures, its application to groundwater dynamics remains largely underexplored. In particular, the extent to which hyperbolic embeddings may better capture multi-scale responses and delayed hydrological processes in monthly GWL series has not been systematically investigated. Consequently, the relationship between model geometry (Euclidean versus hyperbolic) and the representation of long-range dependencies and hierarchical flow dynamics remains insufficiently established in the groundwater forecasting context.

Despite the growing body of literature on groundwater level forecasting, several critical gaps remain insufficiently addressed. First, the role of missing data reconstruction is often treated as a preliminary preprocessing step, with limited attention given to how different imputation strategies influence the statistical structure of the time series and, consequently, downstream forecasting performance. Second, most existing studies focus on single-step prediction, providing limited insight into model behavior and error propagation across extended forecasting

horizons. Third, the impact of representation geometry on modeling hydrological time series has not yet been systematically explored, particularly in the context of groundwater systems characterized by multi-scale dynamics and delayed responses. In this regard, hyperbolic representation learning offers a promising alternative due to its ability to capture hierarchical and scale-free structures; however, to the best of our knowledge, its application to groundwater level forecasting has not yet been systematically investigated.

To address these gaps, this study proposes a unified forecasting framework that integrates structurally consistent data reconstruction with geometry-aware deep learning. Unlike conventional approaches that rely solely on pointwise accuracy, the proposed imputation framework evaluates reconstruction quality using both statistical and structural criteria, ensuring preservation of intrinsic temporal properties. Furthermore, a Hyperbolic Deep Learning architecture is introduced to capture hierarchical and multi-scale dependencies in groundwater level dynamics, and its performance is systematically evaluated within a multi-horizon forecasting setting (t+1 to t+5).

The main contributions of this study can be summarized as follows:
(i) a multi-criteria imputation framework that explicitly accounts for structural consistency in groundwater time series,
(ii) the application of hyperbolic representation learning to groundwater level forecasting, providing a novel perspective on modeling complex temporal dependencies, and
(iii) a comprehensive multi-horizon evaluation combined with statistical validation, enabling a more rigorous assessment of model stability and predictive performance.

Through this integrated approach, the study aims to provide both methodological and practical insights for improving data-driven groundwater forecasting under incomplete observational conditions.

The remainder of this paper is organized as follows. Section 2 describes the study area and the groundwater level dataset used in the analysis. Section 3 presents the overall methodological framework of the proposed forecasting approach. Within this section, the missing data reconstruction strategy is first introduced, followed by the description of the hyperbolic deep learning architecture and the benchmark deep learning models used for comparison. The evaluation framework, including performance metrics and statistical validation procedures, is also detailed in this section. Section 4 presents the results of the data reconstruction and multi-horizon forecasting experiments, while Section 5 discusses the implications of the findings in the context of groundwater dynamics. Finally, Section 6 summarizes the main conclusions and outlines potential directions for future research.

2. Study Area and Data

2.1 Study Area

The study was conducted using groundwater monitoring data from Elazig Province, located in the eastern part of Türkiye. The geographical location of the study area and the spatial

distribution of the selected observation wells are presented in Figure 1. The region is characterized by a continental climate with pronounced seasonal variability, including cold winters and hot, dry summers. Precipitation is primarily concentrated in late autumn and spring, resulting in seasonal recharge patterns that directly influence groundwater dynamics.

Hydrogeologically, the area consists of heterogeneous alluvial and fractured formations, where groundwater systems exhibit delayed recharge responses and multi-scale variability driven by climatic inputs and local abstraction practices. Such characteristics make the region suitable for investigating nonlinear and lagged groundwater level (GWL) behavior at monthly temporal resolution.Three observation wells, representing distinct hydrogeological settings within the province, were selected for analysis. The wells were chosen based on data continuity and monitoring duration, ensuring sufficiently long records for multi-horizon forecasting assessment.

2.2 Data collection

Monthly groundwater level (GWL) observations from three monitoring wells in Elazig Province were analyzed for the period 1978–2015. After excluding missing records, the dataset consists of 415 observations for the Kuyulu well, 394 for Yurtbaşı, and 397 for Gözebaşı, providing long and sufficiently dense time series suitable for investigating seasonal behavior, long-term variability, and multi-horizon predictability. The temporal evolution of the monthly groundwater levels is illustrated in Figure 2. As shown in Figure 2, all three groundwater level series exhibit pronounced seasonal oscillations superimposed on longer-term fluctuations. Periods of gradual decline and partial recovery are clearly visible, particularly in the Kuyulu and Yurtbaşı wells, reflecting the combined influence of delayed recharge processes, storage effects, and cumulative abstraction. The coexistence of short-term variability and multi-year trends highlights the inherently multi-scale nature of the groundwater dynamics at monthly resolution.

To quantitatively characterize these behaviors, descriptive statistical properties of the series are summarized in Table 1. The mean groundwater levels vary substantially across wells, indicating site-specific hydrogeological conditions. Standard deviation values ranging between 5.27 m and 7.52 m reveal moderate to high variability, with the Kuyulu well exhibiting the largest dispersion. Skewness values indicate asymmetric distributions, particularly for Kuyulu (−1.38), suggesting that declining phases dominate over recovery periods. Kurtosis values deviate from zero for all wells, confirming departures from normality and the presence of non-Gaussian distributional characteristics.



Figure 1. Location of the study area and groundwater monitoring wells in Elazig Province, eastern Türkiye.

Well	**Kuyulu**	**Yurtbaşı**	**Gözebaşı**
Count	415	394	397
Mean	-12,9937	-33,9829	-20,8824
Median	-10,72	-33,415	-21,46
Mode	-10,5	-36,1	-28,5
Min	-37,5	-54,8	-31,65
Max	-3,85	-20,9	-5,6
Range	33,65	33,9	26,05
Variance	56,5633	37,7495	27,7831
Std. Dev.	7,5209	6,1441	5,271
CV	0,5788	0,1808	0,2524
Skewness	-1,3822	-0,8613	0,2129
Kurtosis	1,4079	0,7101	-1,0698
ACF(1)	0,9759	0,9466	0,9386
ACF(12)	0,8034	0,8055	0,6591
PermEntropy (m=5)	0,7241	0,836	0,8225
Hurst (H)	0,4773	0,3214	0,3382
ADF (stat)	0,2676	2,0909	-2,8375
ADF (p)	0,9758	0,9988	0,0531
KPSS (stat)	1,2432	1,3839	1,0471
KPSS (p)	0,01	0,01	0,01

Table 1. Descriptive statistics and time-series characteristics of monthly groundwater level observations for the monitoring wells used in this study.

The temporal dependence structure of the series is further examined using autocorrelation analysis. The autocorrelation functions (ACF), presented in Figure 3, reveal very strong short-term persistence, with lag-1 autocorrelation coefficients exceeding 0.93 for all wells. In addition, significant autocorrelation at the seasonal lag (12 months), with values ranging from 0.66 to 0.81, indicates robust annual memory consistent with climate-driven recharge cycles.

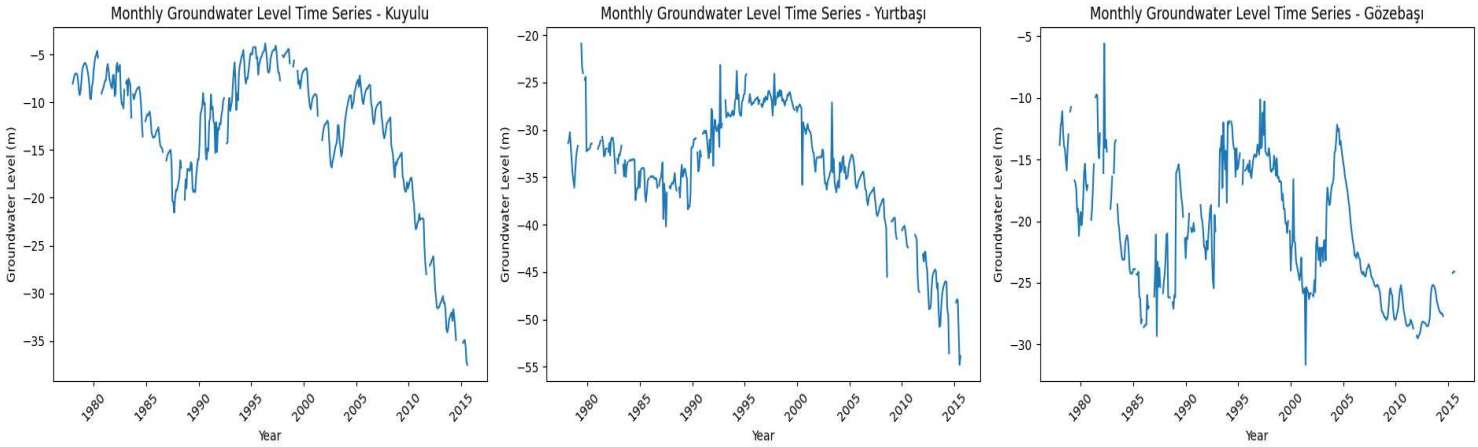
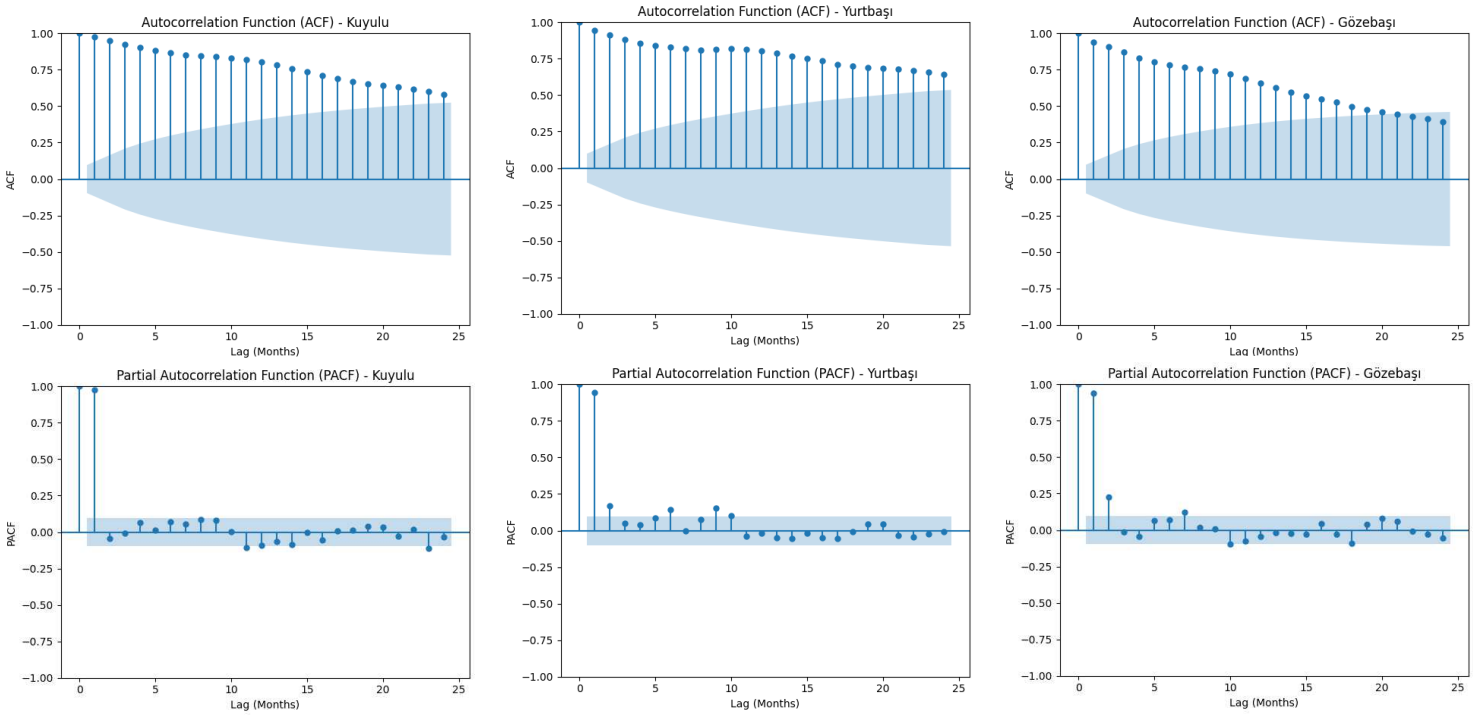


Figure 2. Monthly groundwater level time series for the Kuyulu, Yurtbaşı, and Gözebaşı monitoring wells (1978–2015).

The slow decay of the ACF curves suggests the presence of long-memory behavior rather than rapidly diminishing correlations. Complementary insights are provided by the partial autocorrelation functions (PACF) shown in Figure 3. Significant partial correlations at the first few lags imply the presence of autoregressive components in the groundwater level dynamics. However, the persistence of correlations beyond the initial lags, together with the gradual ACF decay, indicates that low-order linear autoregressive structures alone are insufficient to describe the observed temporal complexity fully.

To further assess the intrinsic complexity of the groundwater level series, normalized permutation entropy (embedding dimension $m = 5$, delay $\tau = 1$) was computed and is reported in Table 1. The resulting values range from 0.724 to 0.836, indicating moderate to high temporal complexity. The Yurtbaşı and Gözebaşı wells exhibit higher entropy values, suggesting more irregular and less predictable fluctuation patterns compared to the Kuyulu well.Long-range dependence was evaluated using the Hurst exponent, also reported in Table 1. Estimated H values range between 0.32 and 0.48. The Kuyulu well exhibits near-persistent behavior, while the other two wells show weaker persistence, reflecting differences in memory strength and recharge–storage dynamics across sites.



(a) (b) (c)

Figure 3. Autocorrelation (ACF) and partial autocorrelation (PACF) functions of the monthly groundwater level time series for the Kuyulu, Yurtbaşı, and Gözebaşı monitoring wells.

Finally, stationarity properties were examined using the Augmented Dickey–Fuller (ADF) and KPSS tests. The ADF test fails to reject the null hypothesis of a unit root for two of the three wells at conventional significance levels, while the KPSS test rejects level stationarity for all series ($p = 0.01$). These results indicate that the monthly groundwater level series are non-stationary, with embedded trends and seasonal components. Taken together, the combined descriptive statistics, autocorrelation structure, entropy measures, persistence indicators, and stationarity tests demonstrate that the groundwater level series are characterized by strong temporal dependence, seasonal memory, non-Gaussian distributions, structural complexity, and non-stationary behavior. These data characteristics underscore the need for forecasting frameworks capable of capturing nonlinear, multi-scale, and long-range dependencies inherent in groundwater systems.

3. Materials and Methods

3.1 Overall Methodological Framework

The methodological framework of this study was structured to systematically examine how data reconstruction quality and representational geometry influence multi-horizon groundwater level (GWL) forecasting performance. The proposed workflow comprises four sequential and interrelated stages: (i) missing data reconstruction, (ii) geometry-aware model training, (iii) multi-horizon forecasting assessment, and (iv) statistical performance validation.

In the first stage, incomplete groundwater level records were reconstructed using multiple imputation strategies, including Seasonal Interpolation, the Kalman Unobserved Components Model (KalmanUCM), k-Nearest Neighbors (kNN), and Random Forest–based imputation. Rather than relying solely on conventional pointwise error measures, each imputation technique was evaluated within a multi-criteria assessment framework. This framework jointly considered error magnitude (RMSE, MAE), variance preservation (VarRatio), autocorrelation distortion (ACF lag-1 and lag-12 differences), spectral consistency (SpecL2), and trend deviation (QDrift). The method exhibiting the most balanced structural and statistical performance was selected for downstream forecasting experiments.

In the second stage, the reconstructed GWL series were reformulated into a supervised learning structure using a sliding-window approach. The resulting input–output sequences were employed to train forecasting models under explicitly different geometric assumptions. The central modeling component of this study is a Hyperbolic Deep Learning architecture implemented within the Poincaré ball manifold. By operating in a negatively curved space, the proposed model is designed to more effectively represent hierarchical, scale-free, and long-range dependencies inherent in groundwater dynamics. In contrast to conventional neural architectures defined in Euclidean space, the hyperbolic framework leverages non-Euclidean geometry to enhance embedding efficiency and representational capacity for complex temporal structures.

To rigorously quantify the contribution of representational geometry, widely used Euclidean deep learning benchmarks — namely CNN-BiLSTM-GRU and TimesNet — were trained on the same reconstructed datasets under identical preprocessing, scaling, and data partitioning protocols. This controlled experimental design ensures that observed performance differences can be attributed to geometric representation rather than inconsistencies in data handling.

In the third stage, model performance was evaluated under a multi-horizon forecasting design spanning one- to five-month-ahead predictions (t+1 to t+5). This setup enables systematic examination of forecast degradation and nonlinear error propagation as the prediction horizon increases, thereby providing insight into model stability across extended lead times.

In the final stage, predictive performance was assessed using both conventional hydrological metrics (RMSE, MAE, NSE, R^2, KGE, and PBIAS) and formal statistical testing. Statistical validation procedures were applied to examine the reliability of the forecasting results. In particular, the Wilcoxon signed-rank test was used to assess the statistical agreement between observed and predicted groundwater levels, while Pearson correlation analysis quantified the strength of the linear relationship between forecasts and observations. Residual diagnostics were further conducted to evaluate potential bias and the dispersion characteristics of prediction errors. A schematic overview of the complete methodological workflow is presented in Figure 4.

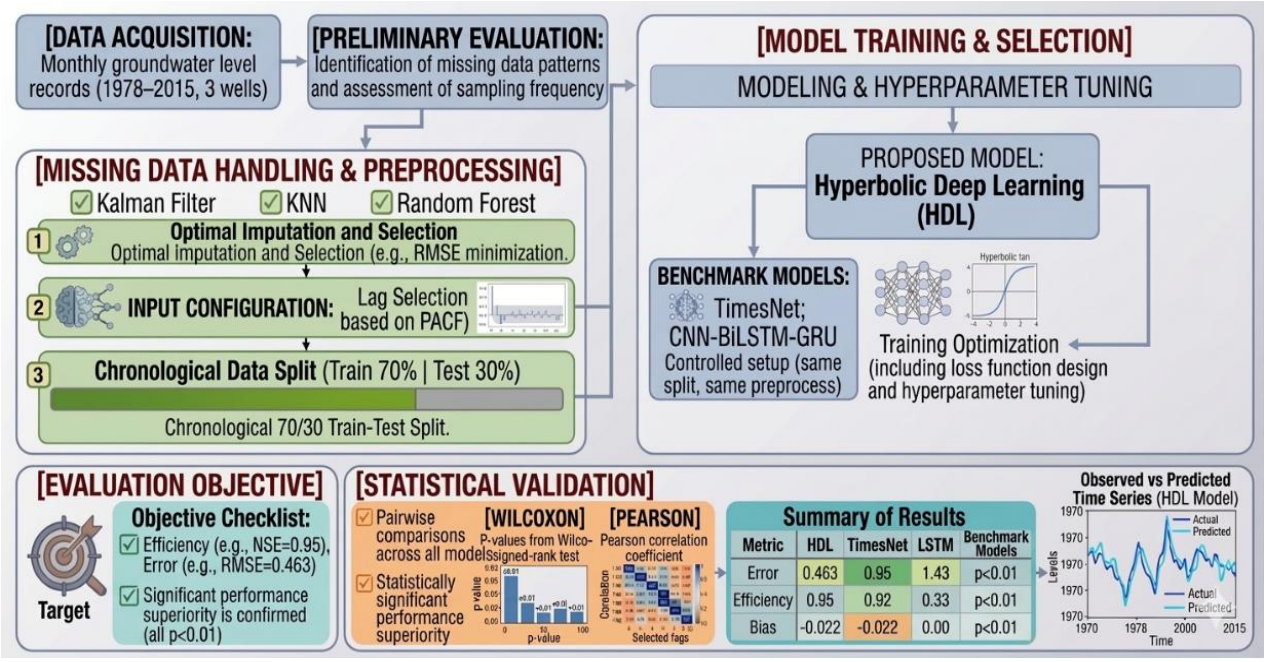


Figure 4. Overall methodological framework of the proposed groundwater level forecasting approach.

3.2 Missing Data Reconstruction

Long-term groundwater level (GWL) monitoring records are frequently affected by data gaps due to sensor malfunction, maintenance interruptions, or archival inconsistencies. In hydrogeological time-series analysis, improper handling of missing observations may alter the statistical and structural properties of the series, including variance, persistence, and seasonal amplitude, which in turn can bias downstream forecasting performance. Therefore, missing data reconstruction was treated as a dedicated methodological component rather than a routine preprocessing step. To reconstruct incomplete monthly GWL records, four imputation strategies representing different modeling assumptions were implemented: Seasonal Interpolation, Kalman Unobserved Components Model (KalmanUCM), k-Nearest Neighbors (kNN), and Random Forest–based imputation.

Seasonal Interpolation : Seasonal Interpolation was applied to exploit recurring periodic patterns in monthly time series. Seasonal methods estimate missing observations using information aligned across equivalent seasonal positions, thereby preserving systematic annual components (Hyndman & Athanasopoulos George, 2018; Shumway & Stoffer D. S., 2019). This approach assumes relatively stable seasonal structure and is computationally straightforward.

Kalman Unobserved Components Model (KalmanUCM) : The KalmanUCM framework formulates the observed series as a state-space process consisting of latent components such as trend and seasonality. Missing observations are inferred through recursive filtering and smoothing within the Kalman filtering framework (Durbin & Koopman, 2012). State-space

models explicitly accommodate unobserved states and allow statistically coherent estimation in the presence of missing data.

k-Nearest Neighbors (kNN) Imputation : The kNN method reconstructs missing values based on similarity to neighboring observations in feature space. As a non-parametric estimator, it does not impose distributional assumptions and infers missing values by averaging the nearest neighbors according to a distance metric (Troyanskaya et al., 2001). This method is widely used for missing value estimation in high-dimensional datasets.

Random Forest–Based Imputation : Random Forest imputation relies on ensemble tree-based regression to estimate missing entries by iteratively predicting incomplete variables using observed features (Stekhoven & Bühlmann, 2012). The method leverages bootstrap aggregation and nonlinear decision boundaries introduced in the original Random Forest algorithm (Breiman, 2001). It does not require parametric assumptions and can model complex nonlinear dependencies.

Rather than selecting an imputation method solely on the basis of pointwise reconstruction accuracy, a multi-criteria evaluation framework was adopted. Each method was assessed using both conventional error metrics (RMSE, MAE, R^2, Nash–Sutcliffe Efficiency) and structural diagnostics designed to evaluate preservation of intrinsic temporal properties. Structural criteria included variance preservation (Variance Ratio), autocorrelation distortion at lag-1 and lag-12, spectral consistency measured via L2 distance between power spectra, and long-term trend deviation quantified through drift metrics. This composite evaluation framework was designed to ensure that the selected reconstruction method maintains not only numerical fidelity but also hydrologically meaningful structural characteristics of the groundwater series. The imputation approach demonstrating the most balanced statistical and structural behavior was subsequently used as the input dataset for all deep learning experiments.

3.3 Hyperbolic Deep Learning Architecture

Recent advances in machine learning have shown that the geometric space in which data representations are learned can significantly influence model performance. Most conventional neural networks operate in Euclidean space, where distances grow linearly and the underlying geometry is flat. However, many real-world datasets exhibit hierarchical or multi-scale structures that cannot be efficiently represented in such flat spaces. In these cases, hyperbolic geometry, characterized by constant negative curvature, provides a more suitable representation space because distances expand exponentially with radius, allowing complex hierarchical relationships to be embedded with lower distortion (Ganea et al., 2018; Nickel & Kiela, 2017).

Hyperbolic deep learning extends conventional neural network architectures by performing representation learning directly on hyperbolic manifolds. Among different models of hyperbolic space, the Poincaré ball model is widely adopted due to its differentiable structure and closed-form expressions for core geometric operations (Ganea et al., 2018).

In this study, the proposed forecasting architecture is formulated within the Poincaré ball manifold, where neural transformations are carried out using hyperbolic operations such as exponential and logarithmic mappings and Möbius-based linear transformations.

3.3.1 Poincaré Ball Manifold

We denote the d-dimensional Poincaré ball with (negative) curvature -c by

$$\mathbb{B}_c^d = \{x \in \mathbb{R}^d : c\|x\|^2 < 1\}, \quad c > 0 \tag{1}$$

where

- d denotes the dimension of the embedding space**,**
- $x \in \mathbb{R}^d$ represents a point in the hyperbolic space,
- $\|\cdot\|$ denotes the Euclidean norm, and
- c is the curvature parameter controlling the negative curvature of the manifold**.**

The geodesic distance between two points x and y located in the Poincaré ball $\mathbb{B}_c^d$ is given by

$$d_c(x,y) = \frac{2}{\sqrt{c}} artanh(\sqrt{c}\|(-x) \oplus_c y\|) \tag{2}$$

where

- $d_c(x,y)$ denotes the hyperbolic distance between points x and y,
- $artanh(\cdot)$is the inverse hyperbolic tangent function, and
- $\oplus_c$ denotes the Möbius addition operation in hyperbolic space

3.3.2 Core Hyperbolic Operations

Möbius addition. For $x, y \in \mathbb{B}_c^d$, Möbius addition is defined as

$$x \oplus_c y = \frac{(1+2c\langle x,y\rangle + c\|y\|^2)x + (1+c\|x\|^2)y}{1+2c\langle x,y\rangle + c^2\|x\|^2\|y\|^2} \tag{3}$$

where

- $\langle x,y\rangle$ denotes the Euclidean inner product between vectors x and y.

Exponential and logarithmic maps at the origin**.** Hyperbolic layers typically operate by moving between the manifold and its tangent space. Let $\upsilon \in T_0\mathbb{B}_c^d \cong \mathbb{R}^d$ and x $\in \mathbb{B}_c^d$.The exponential and logarithmic maps at the origin are

$$exp_0^c(v) = \tanh\left(\sqrt{c}\|v\|\right)\frac{v}{\sqrt{c}\|v\|}, \qquad log_0^c(x) = artanh(\sqrt{c}\|x\|)\frac{x}{\sqrt{c}\|x\|}, \tag{4}$$

Möbius (hyperbolic) linear transformation**.** A Euclidean linear map is applied in the tangent space and then mapped back to the manifold:

$$HypLinear(x; W) = exp_0^c\left(W\ log_0^c(x)\right) \tag{5}$$

where W ∈ $\mathbb{R}^{mxd}$is the trainable weight matrix. This construction generalizes standard affine layers to hyperbolic space while respecting manifold geometry (Ganea et al., 2018).

3.3.3 Network Structure for Forecasting

Let the input vector at reference time t be $u_t \in \mathbb{R}^p$ (constructed from a fixed-length look-back window; see Section X). The model first embeds utu_tut into the hyperbolic manifold:

$$x_t = exp_0^c(Au_t) \tag{6}$$

followed by one or more hyperbolic linear layers *HypLinear*(·) and pointwise nonlinearities applied in a geometry-consistent manner (e.g., via tangent-space activations):

$$h_t = exp_0^c(\sigma(log_0^c\left(HypLinear(x_t; W_1)\right))) \tag{7}$$

and finally produces horizon-specific outputs through a final projection:

$$\hat{y}_t = g_h(log_0^c(h_t)) \qquad h=1,\ldots,H \tag{8}$$

where $g_h(\cdot)$ denotes an output head for horizon hhh (or a shared head if single-output training is adopted). This design permits learning hyperbolic representations while reporting predictions in the original Euclidean output space.

3.3.4 Optimization on Manifolds

Model parameters were optimized using gradient-based learning adapted to Riemannian manifolds. In practice, this involves computing gradients in the tangent space and updating parameters while ensuring they remain on the manifold via retraction or exponential mapping (Bonnabel, 2013). For adaptive optimization, Riemannian extensions of Adam have been proposed for hyperbolic learning (Bécigneul & Ganea, 2019). The training objective was formulated as a standard regression loss (e.g., mean squared error) computed in the output space, while representation learning was carried out on $\mathbb{B}_c^d$.

3.4 Benchmark Deep Learning Models

To evaluate the effectiveness of the proposed hyperbolic deep learning framework, its performance was compared with several established deep learning architectures widely used in time-series forecasting. In this study, a hybrid CNN–BiLSTM–GRU architecture and the recently proposed TimesNet model were selected as benchmark models.

Hybrid convolutional–recurrent architectures have been widely employed for time-series modeling because they can simultaneously capture local patterns and long-term temporal dependencies. Convolutional Neural Networks (CNN) are effective at extracting local features and short-term patterns from sequential data (Lecun et al., 1998), whereas recurrent neural networks such as Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) are designed to model temporal dependencies and mitigate the vanishing gradient problem commonly encountered in traditional recurrent networks (Cho et al., n.d.; Schmidhuber & Shmidhoobuh, 2003). Bidirectional LSTM (BiLSTM) further enhances sequence modeling by incorporating information from both past and future temporal contexts (Schuster & Paliwal, 1997). Such hybrid CNN–RNN architectures have been successfully applied in numerous time-series forecasting applications, including hydrological and environmental prediction problems (Qin et al., 2017; Zhang et al., 1998a).

In addition to hybrid recurrent architectures, the TimesNet model was included as a state-of-the-art benchmark for time-series forecasting. TimesNet introduces a novel representation mechanism by transforming one-dimensional time series into two-dimensional structures, enabling the model to capture complex temporal variations and multi-periodicity more effectively than traditional sequence models (Zhang et al., 1998b). This architecture has demonstrated strong performance across a wide range of time-series forecasting benchmarks and has recently been adopted in various forecasting applications due to its ability to model multi-scale temporal dynamics (Nie et al., 2023; Wen et al., 2023) By comparing the proposed hyperbolic deep learning architecture with these widely adopted deep learning models, the study aims to assess whether hyperbolic representation learning can provide additional predictive advantages for groundwater level forecasting.

3.5 Model Evaluation Metrics and Statistical Validation

The predictive performance of the forecasting models was evaluated using several widely adopted statistical metrics commonly employed in hydrological and time-series forecasting studies. These metrics quantify different aspects of model accuracy and reliability, allowing a comprehensive assessment of prediction quality. In this study, the evaluation metrics included Root Mean Square Error (RMSE), Mean Absolute Error (MAE), Nash–Sutcliffe Efficiency (NSE), Kling–Gupta Efficiency (KGE), and Percent Bias (PBIAS).

The Root Mean Square Error (RMSE) measures the average magnitude of prediction errors by giving greater weight to larger deviations between observed and predicted values:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{9}$$

where y_i represents the observed groundwater level, $\hat{y}_i$denotes the predicted value, and n is the number of observations.

The Mean Absolute Error (MAE) quantifies the average absolute difference between observations and predictions:

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad (10)$$

The Nash–Sutcliffe Efficiency (NSE) is widely used in hydrological model evaluation and measures how well the predicted values reproduce the variability of the observed data (Kartal & Altunkaynak, 2024; Nash & Sutcliffe, 1970)

$$NSE = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y}_i)^2} \quad (11)$$

where $\bar{y}$ represents the mean of observed values. An NSE value closer to 1 indicates better model performance.

To further assess model reliability, the Kling–Gupta Efficiency (KGE) was also computed. This metric simultaneously evaluates correlation, variability, and bias components of model performance(Gupta et al., 2009). The KGE index is defined as

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha - 1)^2 + (\beta - 1)^2} \quad (12)$$

Where *r* denotes the linear correlation coefficient between observed and predicted values,α represents the ratio of the standard deviations of simulated and observed values, and β denotes the ratio of the mean simulated values to the mean observed values.

In addition, Percent Bias (PBIAS**)** was used to quantify the tendency of a model to systematically overestimate or underestimate observed values (Moriasi et al., 2007). The PBIAS statistic is calculated as

$$PBIAS = 100x\frac{\sum_{i=1}^{n}(\hat{y}_i - y_i)}{\sum_{i=1}^{n} y_i} \quad (13)$$

where y_i and $\hat{y}$ denote the observed and predicted groundwater levels, respectively, and nnn is the number of observations. Negative PBIAS values indicate underestimation, whereas positive values indicate overestimation tendencies. Beyond numerical performance metrics, statistical validation was conducted to assess whether differences between competing forecasting models were statistically significant. Statistical validation was performed to assess the reliability of the forecasting results and to compare the performance of competing models. In addition to conventional performance metrics, non-parametric statistical tests were employed to evaluate whether differences between models were statistically significant. Specifically, the Wilcoxon signed-rank test was applied to the absolute prediction errors obtained from different models across each forecasting horizon. This test examines whether the median difference between paired error samples is significantly different from zero, thereby allowing a robust comparison without assuming normality of error distributions. In addition, Pearson correlation analysis was used to quantify the linear relationship between observed and predicted groundwater levels. Residual diagnostics, including mean error and standard deviation of residuals, were further examined to assess bias and dispersion characteristics of the model predictions. All evaluation procedures were conducted separately for each forecasting horizon (t+1 to t+5) to ensure a

comprehensive assessment of model stability and performance degradation over increasing lead times.

Metric	Range	Interpretation
MSE	0	Ideal prediction (no error)
	Increasing value	Indicates larger prediction errors
RMSE	0	Perfect model performance
	Lower values	Better predictive accuracy
MAE	0	Perfect prediction
	Lower values	Smaller average prediction errors
NSE	NSE = 1	Perfect agreement between observed and predicted values
	$0.85 < NSE \leq 1$	Very good model performance
	$0.75 < NSE \leq 0.85$	Good performance
	$0.50 < NSE \leq 0.75$	Acceptable performance
	$NSE \leq 0.50$	Unsatisfactory performance
KGE	KGE = 1	Perfect model performance
	$0.75 < KGE \leq 1$	Very good agreement
	$0.50 < KGE \leq 0.75$	Moderate performance
	$KGE \leq 0.50$	Poor performance
PBIAS	PBIAS = 0	Perfect model (no bias)
	$-10\% < PBIAS < 10\%$	Very good performance
	±10–15%	Good performance
	±15–25%	Satisfactory performance

Table 2. Interpretation ranges of the performance evaluation metrics used in this study.

4. Result

4.1 Performance of Data Reconstruction Methods

Since missing observations are a common issue in groundwater monitoring records, the first step of the analysis involved reconstructing incomplete groundwater level time series using several imputation techniques. The performance of the evaluated reconstruction methods was assessed using both traditional error metrics and structural consistency indicators to ensure that the reconstructed series preserved the intrinsic statistical characteristics of the original observations. Table 3 presents the comparative performance of the evaluated imputation methods, including Seasonal Interpolation, KalmanUCM, k-Nearest Neighbors (kNN), and Random Forest–based imputation. The evaluation framework considered not only prediction accuracy metrics such as RMSE and MAE but also structural indicators including variance preservation (VarRatio), autocorrelation deviations (ACF1 and ACF12 differences), spectral consistency (SpecL2), and trend deviation (QDrift). This multi-criteria evaluation approach allowed the selection of a reconstruction method that maintains both numerical accuracy and the underlying temporal structure of the groundwater level series.

Method	Well	RMSE	MAE	R2	NSE	N_eval	VarRatio	ACF1diff	ACF12diff	SpecL2	QDrift
SeasonalInterp	Kuyulu	1,242	0,983	0,958	0,958	41	0,997	0,002	0,001	0,006	0,017
	Yurtbaşı	1,974	1,120	0,879	0,879	39	0,981	0,008	0,003	0,017	0,016
	Gözebaşı	3,640	2,751	0,580	0,580	39	0,968	0,031	0,027	0,027	0,198
KalmanUCM	Kuyulu	0,875	0,694	0,979	0,979	41	0,999	0,000	0,001	0,004	0,075
	Yurtbaşı	1,978	1,061	0,878	0,878	39	1,008	0,006	0,001	0,011	0,083
	Gözebaşı	2,647	1,839	0,778	0,778	39	0,970	0,004	0,010	0,011	0,087
kNN	Kuyulu	7,143	5,317	-0,397	-0,397	41	0,993	0,091	0,085	0,068	0,063
	Yurtbaşı	6,331	4,636	-0,244	-0,244	39	0,944	0,112	0,103	0,056	0,209
	Gözebaşı	5,785	4,499	-0,060	-0,060	39	0,918	0,096	0,054	0,067	0,119
RF_Imputer	Kuyulu	3,632	2,420	0,639	0,639	41	1,007	0,021	0,017	0,028	0,099
	Yurtbaşı	3,272	2,523	0,668	0,668	39	0,986	0,019	0,021	0,031	0,129
	Gözebaşı	3,982	2,765	0,498	0,498	39	0,966	0,035	0,004	0,028	0,171

Table 3. Comparative performance of different imputation methods across the three groundwater monitoring wells.

Note: *Lower values of RMSE, MAE, ACF deviations, SpecL2, and QDrift indicate better reconstruction performance, whereas higher R^2 and NSE values indicate improved agreement between reconstructed and observed groundwater levels.*

From the perspective of classical error metrics, the KalmanUCM method consistently produced the lowest prediction errors for the majority of wells. For example, at the Kuyulu well, KalmanUCM achieved an RMSE of 0.874 and an MAE of 0.694, which are substantially lower than those obtained from Seasonal Interpolation (RMSE = 1.241) and Random Forest imputation (RMSE = 3.632). A similar trend was observed for the Gözebaşı well, where KalmanUCM again yielded the lowest RMSE (2.647) compared with Seasonal Interpolation (3.639) and Random Forest imputation (3.982). These results indicate that the KalmanUCM approach provides a more accurate reconstruction of missing groundwater level observations in terms of pointwise error measures.

The coefficient of determination (R^2) and Nash–Sutcliffe Efficiency (NSE) further confirm the superior predictive performance of KalmanUCM. For instance, the Kuyulu well exhibits extremely high values ($R^2 = 0.979$ and NSE = 0.979), indicating an excellent agreement between reconstructed and observed values. Seasonal Interpolation also produced relatively strong performance at some wells, particularly for Kuyulu (NSE ≈ 0.958), yet its performance deteriorated noticeably at Gözebaşı (NSE ≈ 0.581), suggesting that the method may be less robust across varying hydrogeological conditions.

In contrast, the k-Nearest Neighbors (kNN) method performed substantially worse than the other approaches. Negative R^2 and NSE values were obtained for all wells, indicating that the reconstructed series using kNN performs worse than simply using the mean of the observations as a predictor. Moreover, the error metrics associated with kNN were considerably higher (e.g., RMSE values exceeding 5–7 units), suggesting that this method is unsuitable for reconstructing the groundwater level series in the present dataset. The poor performance of kNN (negative NSE values) suggests that simple feature-space proximity is insufficient to capture the complex temporal dynamics of the studied aquifers, whereas the KalmanUCM's superior performance stems from its ability to decompose the series into stochastic trend and seasonal components.

Beyond traditional error metrics, structural consistency indicators provide further insight into the ability of each method to preserve the statistical characteristics of the original time series. The variance preservation ratio (VarRatio) remained very close to unity for KalmanUCM and Seasonal Interpolation,

indicating that these methods successfully maintained the overall variability of the groundwater level records. For example, VarRatio values ranged between 0.969 and 1.008, suggesting minimal distortion of the variance structure after reconstruction.(Figure 5)

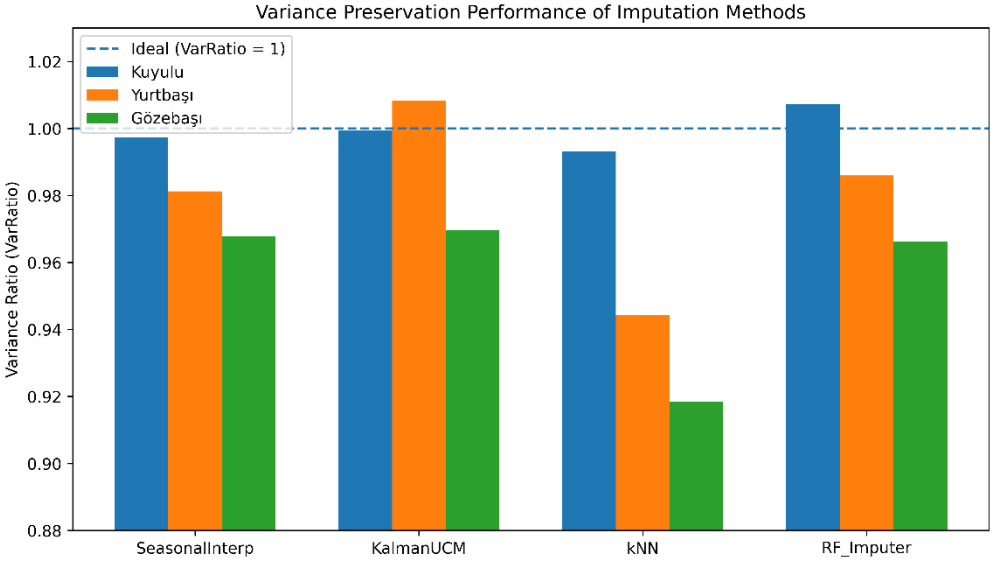


Figure 5. Variance preservation comparison of imputation methods for the three wells.

Similarly, the autocorrelation deviation metrics (ACF1diff and ACF12diff) indicate that KalmanUCM introduces the smallest distortions in the temporal dependency structure of the groundwater level series. For instance, the ACF1 difference for the Kuyulu well under KalmanUCM is only 0.00042, substantially lower than the values observed for Seasonal Interpolation and Random Forest imputation. This finding is particularly important because preserving autocorrelation structure is critical for reliable time-series forecasting.

Spectral consistency analysis further supports these findings. The SpecL2 metric, which evaluates differences in spectral energy distribution between reconstructed and observed series, remains lowest for KalmanUCM across most wells. This indicates that the method better preserves periodic and multi-scale components of groundwater fluctuations.

Figure 6 presents the comparison between the observed groundwater level series and the reconstructed series obtained using the evaluated imputation methods for the three monitoring wells. Overall, KalmanUCM and Seasonal Interpolation closely follow the temporal dynamics of the observed groundwater levels, successfully reproducing both seasonal variations and long-term trends. In contrast, the kNN method introduces noticeable irregular fluctuations and unrealistic spikes, indicating a poor representation of the underlying hydrological signal.

The Random Forest imputation generally captures the overall trend of the groundwater level variations but shows moderate deviations during periods with higher variability. Across all wells, KalmanUCM provides the most stable reconstruction, preserving the temporal structure and variability of the groundwater level series more effectively than the other methods. These visual findings are consistent with the quantitative evaluation results presented in Table 3, supporting the selection of KalmanUCM as the most suitable reconstruction method for the subsequent forecasting analysis.

Finally, the trend deviation indicator (QDrift) was examined to evaluate whether imputation methods introduce artificial long term trends into the reconstructed series. KalmanUCM generally produced relatively small QDrift values, suggesting minimal distortion of long-term groundwater level dynamics.

Considering the combined evidence from error metrics, variance preservation, autocorrelation consistency, spectral similarity, and trend stability, KalmanUCM demonstrates the most balanced performance across the evaluated criteria. Therefore, the groundwater level series reconstructed using KalmanUCM were selected for the subsequent forecasting experiments.

4.2 Multi-Horizon Groundwater Level Forecasting Results

To assess the predictive capability of the proposed framework, groundwater level forecasting experiments were conducted using the reconstructed datasets obtained from the optimal imputation method identified in the previous section. The objective of this stage was to evaluate how effectively deep learning models capture the temporal dynamics of groundwater level fluctuations and how their predictive accuracy evolves as the forecasting horizon increases. For model development, the reconstructed groundwater level series were partitioned into training and testing subsets, where 70% of the data were used for training and the remaining 30% for testing**.** Within the training dataset, 15% of the samples were further allocated for validation in order to monitor model convergence and mitigate potential overfitting during the training process. The forecasting framework was formulated using a sliding-window representation in which the input variables consisted of groundwater levels at times t and t−1, while the target variables corresponded to future groundwater levels at prediction horizons ranging from t+1 to t+5. The primary forecasting architecture adopted in this study is a Hyperbolic Deep Learning model, which represents temporal dependencies within a non-Euclidean embedding space. This representation enables the model to capture complex dependency structures that may arise in groundwater dynamics.

Model	Hyperparameter	Search Range
Hyperbolic DL	Input window size	{2, 3, 4}
	Hidden dimension	{16, 24, 32}
	Curvature parameter (c)	{0.3, 0.5, 0.7, 1.0}
	Learning rate	{0.01, 0.005, 0.001}
	Batch size	{16, 24, 32}
	Training epochs	{200, 500, 800}
CNN-BiLSTM-GRU	CNN filters	{16, 32, 64}
	Kernel size	{2, 3, 5}
	BiLSTM units	{32, 64, 128}
	GRU units	{16, 32, 64}
	Learning rate	{0.01, 0.005, 0.001}
TimesNet	Encoder layers	{1, 2, 3}
	Hidden dimension	{32, 64, 128}
	Top-k frequency components	{3, 5, 7}
	Learning rate	{0.01, 0.005, 0.001}

Table4. Hyperparameter search space and optimal configuration of the forecasting models

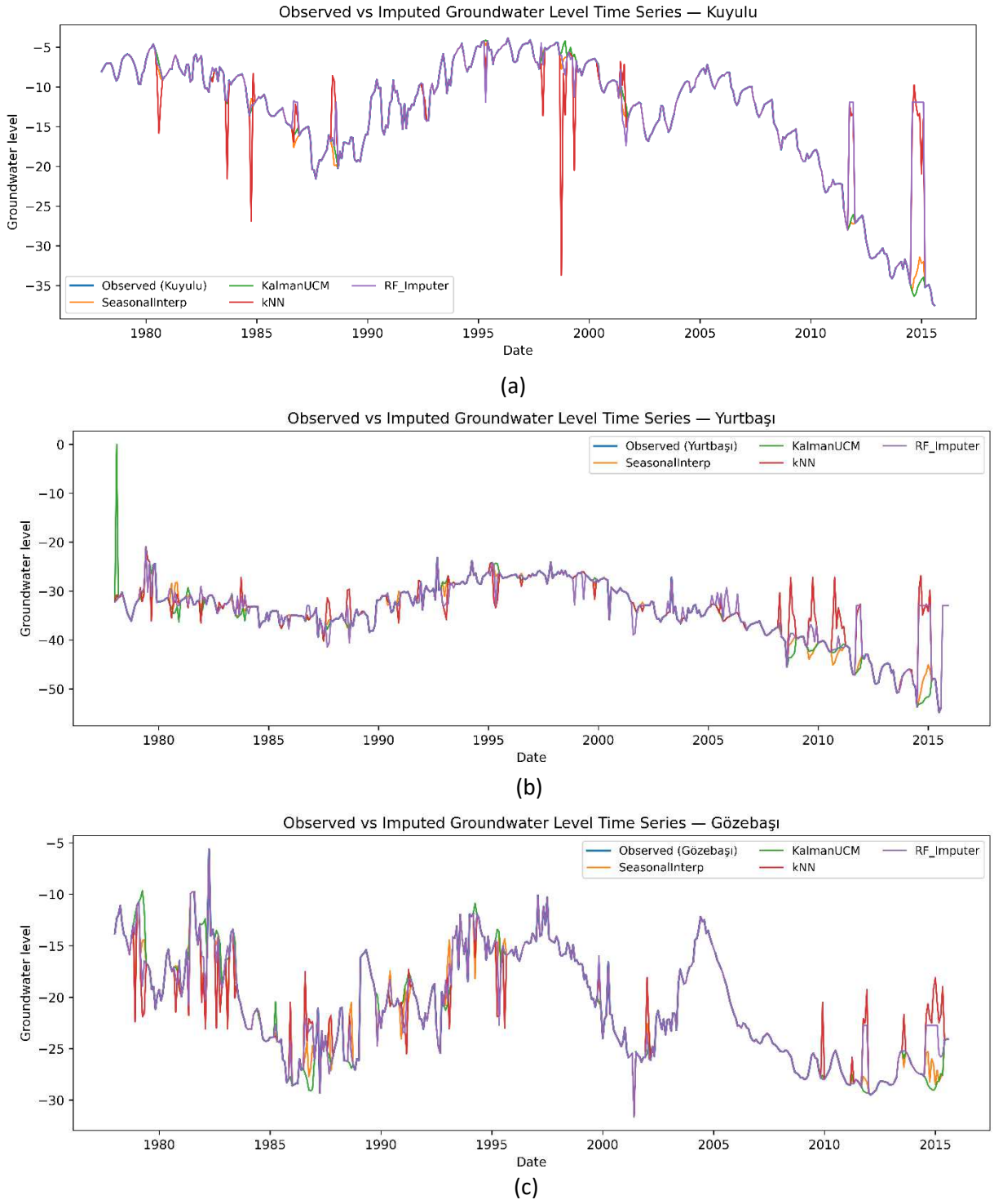


Figure 6. Observed and imputed groundwater level time series for the three monitoring well using different data reconstruction methods.(a) Kuyulu (b) Yurtbaşı (c)Gözebaşı

The hyperparameter settings of the forecasting models are summarized in Table 4. Given that the hyperbolic deep learning architecture constitutes the main methodological contribution of this study, its model parameters are reported in greater detail. All hyperparameters were determined through a k-fold cross-validation procedure, ensuring stable model training and improving the generalization capability of the forecasting models across different prediction horizons.(Hastie et al., 2009; James et al., 2013)

The forecasting performance of the evaluated models across different prediction horizons is summarized in Table 5(a-c). The results provide a comprehensive comparison of the predictive capability of the Hyperbolic Deep Learning model against the benchmark architectures, namely CNN-BiLSTM-GRU and TimesNet, across the three groundwater monitoring wells.

Overall, the Hyperbolic Deep Learning model consistently achieved the highest forecasting accuracy across nearly all prediction horizons and stations. In particular, the model demonstrated superior performance in terms of RMSE, MAE, NSE, and KGE, indicating its strong capability to capture the temporal dynamics of groundwater level variations. For one-step-ahead forecasting (t+1), the hyperbolic model produced exceptionally high NSE values reaching 0.989, 0.988, and 0.972 for the Kuyulu, Yurtbaşı, and Gözebaşı wells, respectively, while maintaining low error magnitudes. These results suggest that the hyperbolic representation effectively captures short-term temporal dependencies in groundwater level fluctuations.

Station	Horizon	Model	MSE	RMSE	MAE	NSE	PBIAS (%)	KGE
Kuyulu	**t+1**	Hyperbolic DL	1,026	1,013	0,752	0,989	2,678	0,950
		CNN-BiLSTM-GRU	4,621	2,15	1,721	0,949	7,351	0,849
		TimesNet	1,437	1,199	0,911	0,984	-3,603	0,933
	t+2	Hyperbolic DL	5,104	2,259	1,746	0,944	6,539	0,884
		CNN-BiLSTM-GRU	8,434	2,904	2,348	0,907	9,914	0,802
		TimesNet	4,725	2,174	1,724	0,948	-6,905	0,865
	t+3	Hyperbolic DL	7,307	2,703	2,159	0,919	8,662	0,836
		CNN-BiLSTM-GRU	13,137	3,625	2,950	0,855	12,469	0,753
		TimesNet	10,386	3,223	2,629	0,885	10,192	0,860
	t+4	Hyperbolic DL	11,668	3,416	2,741	0,871	11,571	0,855
		CNN-BiLSTM-GRU	26,359	5,134	3,810	0,708	16,590	0,594
		TimesNet	16,764	4,094	3,347	0,814	14,191	0,728
	t+5	Hyperbolic DL	14,637	3,826	3,141	0,838	13,018	0,758
		CNN-BiLSTM-GRU	29,295	5,412	4,392	0,677	19,163	0,609
		TimesNet	21,239	4,609	3,793	0,765	16,234	0,685

(a)

Station	Horizon	Model	MSE	RMSE	MAE	NSE	PBIAS (%)	KGE
Yurtbaşı	t+1	Hyperbolic DL	0,362	0,602	0,362	0,988	0,7859	0,968
		CNN-BiLSTM-GRU	3,197	1,788	1,375	0,893	3,215	0,854
		TimesNet	0,496	0,704	0,470	0,983	0,887	0,963
	t+2	Hyperbolic DL	2,574	1,604	1,091	0,913	1,672	0,919
		CNN-BiLSTM-GRU	6,940	2,634	1,870	0,767	3,540	0,853
		TimesNet	5,795	2,407	1,664	0,805	3,479	0,848
	t+3	Hyperbolic DL	3,186	1,785	1,278	0,893	2,501	0,886
		CNN-BiLSTM-GRU	8,814	2,969	2,206	0,704	4,972	0,804
		TimesNet	8,312	2,883	2,129	0,721	4,600	0,820
	t+4	Hyperbolic DL	5,431	2,330	1,678	0,817	3,280	0,875
		CNN-BiLSTM-GRU	10,624	3,260	2,500	0,643	6,283	0,781
		TimesNet	8,673	2,945	2,658	0,708	5,734	0,930
	t+5	Hyperbolic DL	8,958	2,993	2,169	0,699	4,428	0,824
		CNN-BiLSTM-GRU	19,824	4,452	3,663	0,333	7,864	0,821
		TimesNet	13,725	3,705	2,887	0,538	6,689	0,737

(b)

Station	Horizon	Model	MSE	RMSE	MAE	NSE	PBIAS (%)	KGE
Gözebaşı	t+1	Hyperbolic DL	0,417	0,646	0,547	0,972	1,816	0,974
		CNN-BiLSTM-GRU	1,126	1,061	0,908	0,925	3,439	0,922
		TimesNet	0,765	0,875	0,696	0,949	2,096	0,952
	t+2	Hyperbolic DL	0,923	0,961	0,843	0,933	2,477	0,899
		CNN-BiLSTM-GRU	1,949	1,396	1,215	0,859	4,178	0,938
		TimesNet	1,679	1,296	1,114	0,879	4,021	0,885
	t+3	Hyperbolic DL	1,308	1,144	0,962	0,906	2,722	0,919
		CNN-BiLSTM-GRU	2,870	1,694	1,475	0,808	5,309	0,876
		TimesNet	2,163	1,471	1,260	0,844	4,394	0,922
	t+4	Hyperbolic DL	2,016	1,420	1,234	0,865	-4,500	0,872
		CNN-BiLSTM-GRU	4,057	2,014	1,751	0,707	6,375	0,857
		TimesNet	3,274	1,809	1,563	0,764	5,537	0,898
	t+5	Hyperbolic DL	3,982	1,996	1,692	0,733	-4,354	0,895
		CNN-BiLSTM-GRU	5,350	2,313	2,021	0,614	7,357	0,825
		TimesNet	4,294	2,072	1,802	0,690	6,511	0,875

(c)

Table 5. Multi-horizon groundwater level forecasting performance of the evaluated models for three observation wells: (a) Kuyulu, (b) Yurtbaşı, and (c) Gözebaşı.

As the forecasting horizon increased, a gradual reduction in predictive performance was observed for all models, which is expected due to the increasing uncertainty associated with longer lead-time

predictions. Nevertheless, the Hyperbolic Deep Learning architecture maintained a relatively stable performance compared to the benchmark models. Even at longer horizons such as t+4 and t+5, the hyperbolic model preserved relatively high NSE values across the wells, whereas the competing models exhibited a more pronounced decline in forecasting skill. This behaviour indicates that the hyperbolic embedding space provides a more expressive representation of complex temporal relationships, enabling the model to maintain predictive stability over extended forecast horizons.

The TimesNet model demonstrated intermediate performance, generally outperforming the CNN-BiLSTM-GRU architecture while remaining slightly below the hyperbolic model. TimesNet maintained relatively stable results for short forecasting horizons; however, its predictive performance gradually decreased as the prediction horizon extended. These findings suggest that frequency-aware temporal modelling partially improves forecasting robustness but may still be limited in capturing hierarchical temporal dependencies.

Another notable observation is the consistent behaviour of the Hyperbolic Deep Learning model across the three monitoring wells. Despite the differences in groundwater dynamics among the Kuyulu, Yurtbaşı, and Gözebaşı stations, the hyperbolic architecture consistently achieved strong predictive performance. This robustness indicates that the proposed modelling framework is capable of generalizing across groundwater systems with different temporal variability patterns.

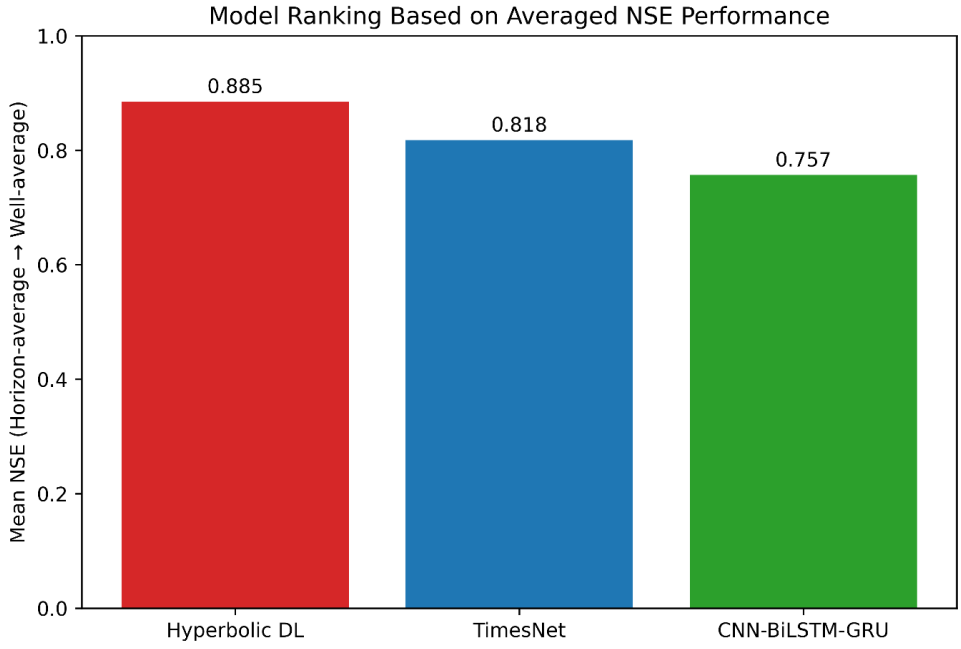


Figure 7. Overall ranking of the forecasting models based on the averaged NSE values across all prediction horizons and monitoring wells**.**

To provide an overall comparison of model performance, the forecasting accuracy of each model was aggregated across all wells and prediction horizons using the Nash–Sutcliffe Efficiency (NSE) metric. The resulting model ranking is presented in Figure 7.

The results clearly indicate that the Hyperbolic Deep Learning model achieved the highest average predictive performance, with a mean NSE value of 0.885, outperforming both benchmark deep learning architectures. In comparison, the TimesNet model yielded a slightly lower average NSE of 0.818, while

774 the CNN-BiLSTM-GRU model showed comparatively weaker performance with an average NSE of
775 0.757.

776 The superior performance of the Hyperbolic Deep Learning model suggests that incorporating non-
777 Euclidean representation geometry can improve the ability of neural networks to capture complex
778 temporal dependencies and hierarchical structures inherent in groundwater level dynamics. By
779 embedding temporal patterns in hyperbolic space, the model appears to better represent multi-scale
780 hydrological processes and delayed system responses. The model ranking results provide consistent
781 evidence that the proposed hyperbolic framework offers improved predictive capability compared with
782 conventional Euclidean deep learning architectures for multi-horizon groundwater level forecasting.

783 Scatter plots illustrating the agreement between observed and predicted groundwater levels produced by
784 the Hyperbolic Deep Learning model across different forecasting horizons are presented in Figure 8.
785 Overall, the predicted values show a strong agreement with the observations, as most data points are
786 concentrated around the 1:1 reference line. This pattern indicates that the model is capable of accurately
787 capturing the temporal dynamics of groundwater level variations across the monitoring wells. For the
788 one-step-ahead forecasts (t+1), the predicted values exhibit a very close alignment with the observations,
789 reflecting the high predictive skill of the model. As the forecasting horizon increases from t+1 to t+5, a
790 gradual increase in dispersion around the 1:1 line can be observed.

791 This behaviour is expected in multi-step forecasting tasks due to the accumulation of prediction
792 uncertainty at longer lead times. Nevertheless, even at longer horizons, the majority of the predictions
793 remain reasonably close to the reference line, indicating that the Hyperbolic Deep Learning model
794 maintains stable forecasting performance.A similar pattern is observed across the three monitoring
795 wells, although slight differences in dispersion can be noticed depending on the local groundwater
796 dynamics. Despite these variations, the overall structure of the scatter distributions confirms the
797 robustness and reliability of the proposed modelling framework for multi-horizon groundwater level
798 forecasting.

799 Taken together, the quantitative performance metrics, scatter plot analyses, and model ranking results
800 consistently indicate that the proposed hyperbolic deep learning framework provides superior predictive
801 capability for groundwater level forecasting. Across the three monitoring wells and all forecasting
802 horizons (t+1–t+5), the hyperbolic model achieved the highest overall accuracy and maintained a more
803 stable predictive performance compared with the benchmark Euclidean deep learning architectures. The
804 scatter plot distributions further confirm this finding, as the predicted values remain closely aligned with
805 the 1:1 reference line even at extended lead times.

806 These results suggest that incorporating hyperbolic representation geometry enhances the ability of deep
807 learning models to capture complex temporal dependencies and hierarchical dynamics present in
808 groundwater systems. Overall, the findings demonstrate that hyperbolic representation learning offers a
809 clear advantage for multi-horizon groundwater level forecasting and represents a promising modelling
810 paradigm for analysing complex hydrological time series.

811

812

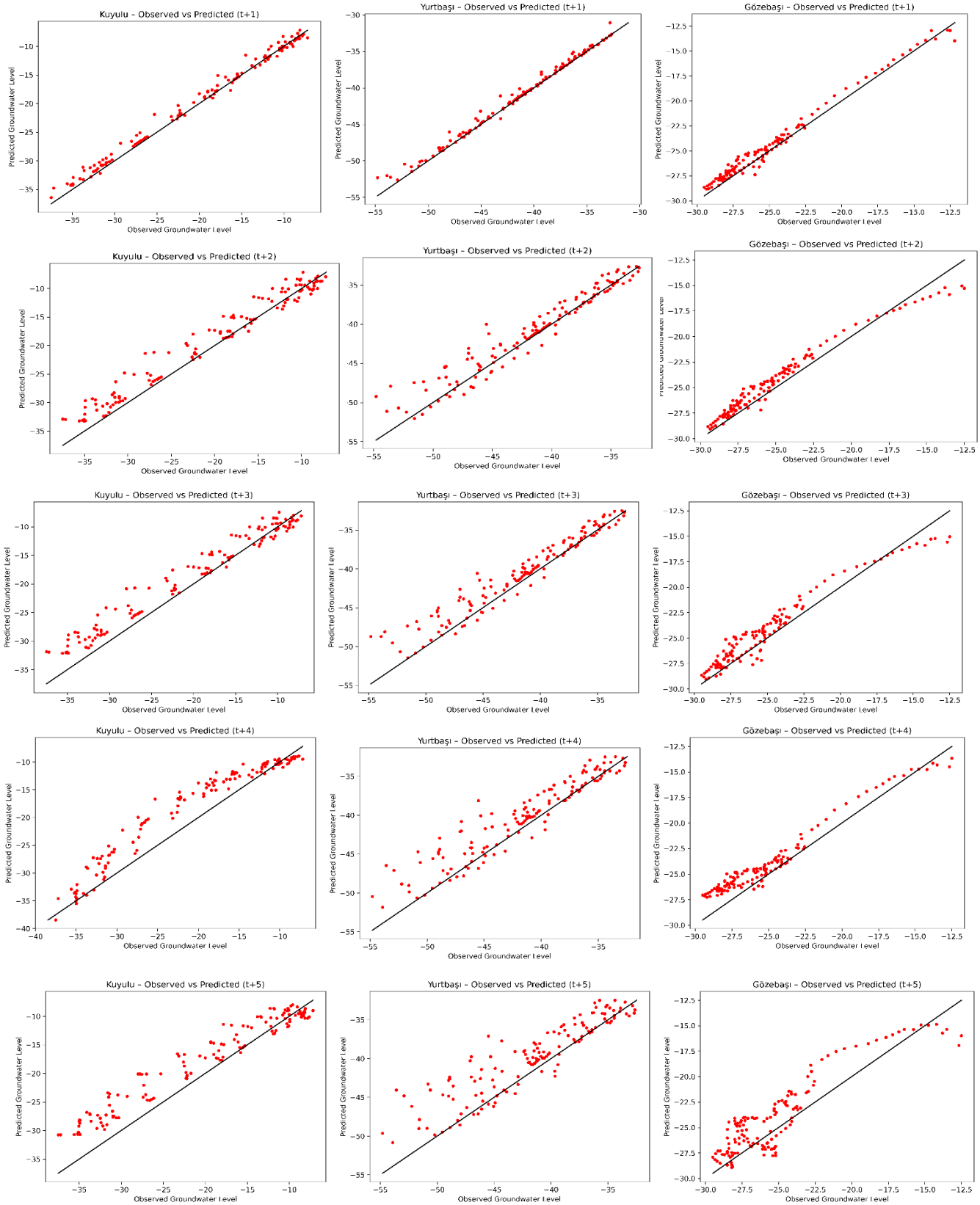


Figure 8. Scatter plots comparing observed and predicted groundwater levels generated by the Hyperbolic Deep Learning model for the monitoring wells across multiple forecasting horizons (t+1–t+5). The black diagonal line represents the 1:1 agreement line between observations and predictions

5. Discussion

The results of this study demonstrate that the proposed forecasting framework effectively captures the temporal dynamics of groundwater level variations across the investigated monitoring wells. In particular, the Hyperbolic Deep Learning (HDL) architecture consistently produced lower prediction errors and higher efficiency scores than the benchmark models across most forecasting horizons. This suggests that representing temporal dependencies within a non-Euclidean geometric space can provide a more expressive representation of complex groundwater dynamics.

Groundwater systems typically exhibit nonlinear responses, delayed recharge processes, and long-memory behavior resulting from interactions among precipitation, aquifer storage, and anthropogenic withdrawals. Such processes often generate hierarchical temporal structures that may not be fully captured by conventional Euclidean neural network architectures. Hyperbolic representation learning has been shown to provide efficient embeddings for hierarchical and complex relational structures(Chami et al., n.d.; Ganea et al., 2018; Nickel & Kiela, 2017), which may explain the improved forecasting performance observed in this study.

Forecasting accuracy gradually decreased as the prediction horizon increased from t+1 to t+5, a pattern commonly reported in hydrological forecasting due to the accumulation of prediction uncertainty with lead time (Kratzert et al., 2018; Shen, 2018). Nevertheless, the proposed hyperbolic model maintained relatively stable performance across longer horizons. Comparisons with CNN-BiLSTM-GRU and TimesNet further confirmed the effectiveness of the proposed framework. Although these models achieved reasonable predictive performance, the hyperbolic model generally yielded lower RMSE and MAE values and higher efficiency scores, indicating that geometry-aware representation learning can improve the modeling of groundwater dynamics. (Bai et al., 2018; Yang & Chen, 2019)

The reliability of the forecasting results was further examined through statistical validation analyses. The Wilcoxon signed-rank test was used to compare the error distributions of competing models across different forecasting horizons. The results indicated that the proposed Hyperbolic Deep Learning model achieved statistically significant improvements over the benchmark models in most cases, confirming the robustness of its predictive performance. Pearson correlation coefficients between observed and predicted groundwater levels were consistently high, generally exceeding 0.90, indicating a strong linear agreement between forecasts and observations. Residual diagnostics revealed relatively low bias and controlled dispersion of prediction errors, although a gradual increase in residual variability was observed for longer forecasting horizons. This behavior is consistent with the expected accumulation of uncertainty in multi-step forecasting systems. Overall, the combined use of performance metrics, statistical testing, and residual analysis provides strong evidence that the proposed model delivers both accurate and stable predictions across multiple lead times.

These findings suggest that, the findings highlight the potential of integrating advanced data reconstruction techniques with geometry-aware deep learning architectures for groundwater level forecasting. From a hydrogeological perspective, the results suggest that groundwater level dynamics may contain hierarchical temporal structures associated with recharge delays and aquifer storage processes. Although the present study focused on three monitoring wells and relied primarily on historical groundwater levels, future research could extend this framework by incorporating additional hydro-meteorological drivers and spatially distributed groundwater observations.

Kuyulu							
LeadTime	**Wilcoxon_stat**	**Wilcoxon_p**	**Pearson_r**	**Pearson_p**	**Shapiro_p (residual)**	**Residual_mean (pred-obs)**	**Residual_std**
1	1625	7,42E-11	0,99677	1,5E-147	0,002401	0,550854	0,853339
2	1348	1,79E-12	0,985145	7E-103	0,023963	1,34275	1,813498
3	981	3,69E-15	0,985022	1,2E-102	0,043129	1,778874	2,030192
4	813	1,75E-16	0,968165	2,85E-81	0,001432	2,380872	2,440773
5	767	7,41E-17	0,974936	4,96E-88	0,015242	2,675763	2,725029

(a)

Yurtbaşı							
LeadTime	**Wilcoxon_stat**	**Wilcoxon_p**	**Pearson_r**	**Pearson_p**	**Shapiro_p (residual)**	**Residual_mean (pred-obs)**	**Residual_std**
1	1409	2,82E-12	0,994501	3,2E-132	7,98E-09	0,373303	0,599745
2	891	7,33E-16	0,939926	1,8E-63	7,27E-09	1,455507	1,917197
3	639	6,44E-18	0,923655	7,77E-57	1,43E-06	1,915875	2,148759
4	18	1,47E-23	0,970985	6,85E-84	0,017237	2,629723	1,326356
5	170	4,2E-22	0,908013	1,02E-51	1,54E-06	2,781413	2,435599

(b)

Gözebaşı							
LeadTime	**Wilcoxon_stat**	**Wilcoxon_p**	**Pearson_r**	**Pearson_p**	**Shapiro_p (residual)**	**Residual_mean (pred-obs)**	**Residual_std**
1	884	3,97E-16	0,992506	2,6E-123	7,17E-06	0,4384	0,49775
2	1158	7,9E-14	0,983953	1,1E-100	4,04E-11	0,600587	0,775866
3	1405	4,41E-12	0,969962	6,52E-83	2,96E-05	0,657814	0,967203
4	398	5,2E-20	0,979654	6,08E-94	0,004634	1,127861	0,865636
5	1636	1,45E-10	0,905126	7,08E-51	0,008484	1,091299	1,677017

(c)

Table 6. Statistical validation of groundwater level forecasting results across prediction horizons for the three monitoring wells: (a) Kuyulu, (b) Yurtbaşı, and (c) Gözebaşı.

6. Conclusion

This study introduces a groundwater level forecasting framework that integrates multi-criteria data reconstruction with Hyperbolic Deep Learning (HDL) for multi-horizon prediction. The main findings and contributions can be summarized as follows:

•The proposed multi-criteria imputation framework preserved key statistical and temporal characteristics of groundwater level series, including variance, autocorrelation, and spectral properties. This ensured that the reconstructed data maintained its intrinsic structure, providing a reliable foundation for subsequent forecasting tasks.

• To the best of our knowledge, this study represents one of the first attempts to apply hyperbolic representation learning to groundwater level (GWL) forecasting. While hyperbolic deep learning has shown promising results in domains characterized by hierarchical and complex structures, its use in hydrogeological time-series modeling has remained largely unexplored. The findings of this study provide initial empirical evidence supporting its applicability in this context.

• The Hyperbolic Deep Learning architecture demonstrated consistently strong performance across all wells and forecasting horizons. The results suggest that hyperbolic embedding spaces can more effectively capture multi-scale temporal dependencies and delayed system responses commonly observed in groundwater systems, particularly when compared to conventional Euclidean-based models(CNN-BİLSTM-GRU, TimesNet).

• Although prediction accuracy decreased with increasing lead time, as expected in multi-step forecasting problems, the HDL model exhibited a more gradual performance degradation. This indicates a relatively stable predictive behavior, especially for extended forecasting horizons (t+4 and t+5).

• The performance of the proposed framework was assessed using multiple complementary metrics and statistical tools. The Wilcoxon signed-rank test, applied to model error distributions, indicated statistically significant differences between competing models, while Pearson correlation analysis and residual diagnostics confirmed strong agreement and controlled error dispersion.

• The combined use of structurally consistent data reconstruction and geometry-aware learning provides a flexible and data-driven framework that can support groundwater monitoring and management, particularly in systems characterized by limited or incomplete observations.

• While the results are encouraging, the analysis was conducted on univariate groundwater level series from a limited number of monitoring wells. It should also be noted that the strong predictive performance may partially be influenced by the high autocorrelation structure of groundwater systems, particularly for short-term predictions. Incorporating additional hydro-meteorological variables (e.g., precipitation and evapotranspiration) and evaluating the framework across different hydrogeological settings could further enhance model generalization and interpretability.

In summary, the integration of multi-criteria data reconstruction with hyperbolic representation learning offers a promising and conceptually innovative direction for groundwater level forecasting and related hydrological applications.

Declarations

Acknowledgments

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. The authors gratefully acknowledge the General Directorate of State Hydraulic Works (DSI) for providing the groundwater monitoring data utilized in this research.

Data availability

The data used in this study were provided by the General Directorate of State Hydraulic Works (DSI) and are not publicly available due to institutional data-sharing restrictions.

Funding

This study did not receive any financial support from any organization.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose

Author Contributions

Kübra Küllahcı were responsible for the conceptualization and proposing the research idea and developed the methodology, ran the models, carried out analysis and visualization, and wrote the original draft of the manuscript. Kürşat Şekerci assisted in hydrological and meteorological data collection and reviewing andediting the original draft.

References

"The United Nations World Water Development Report ," 2022.

"Groundwater Depletion in the United States (1900-2008)".

"United Nations Framework Convention On Climate Change," 2015.

"United Nations, Sustainable Development Goals," 2015.

K. Şekerci, M. C. Tuna, and M. S. Dogan, " Integrated water operations under climate change: Uluova Micro Basin example.," *Journal of Water and Climate Change*, vol. 15(8), pp. 3689–3701, 2024.

K. ŞEKERCİ, M. C. TUNA, and M. Ş. DOĞAN, "Seasonal Level Change and Overdraft of Water Wells," *Türk Doğa ve Fen Dergisi*, vol. 12, no. 1, pp. 1–5, Mar. 2023, doi: 10.46810/tdfd.1148326.

B. Latuamury, J. F. Sahusilawane, H. Marasabessy, W. N. Imlabla, G. Hartono, and M. H. Marasabessy, "Modeling Groundwater Flow Dynamics Using the Master Recession Curve: Insights from Small Island Watersheds in Ambon, Indonesia," *Forest and Nature*, vol. 1, no. 1, pp. 21–34, Jan. 2025, doi: 10.63357/fornature.v1i1.1.

P. Liu, Y. Peng, M. Shang, Y. Yuan, and Y. Wang, "Time-lag effects and periodic responses in precipitation-shallow groundwater system of Qiaocheng District, China: a combined diagnosis using wavelet and long-series statistical analysis," *Hydrology Research*, vol. 56, no. 12, pp. 1219–1240, Dec. 2025, doi: 10.2166/nh.2025.052.

K. Şekerci, M. C. Tuna, and M. Ş. Doğan, "Effects of reducing long-term groundwater overdraft in uluova micro-basin," *Appl. Water Sci.*, vol. 15, no. 8, Aug. 2025, doi: 10.1007/s13201-025-02581-4.

A. Gibrilla, G. Anornu, and D. Adomako, "Trend analysis and ARIMA modelling of recent groundwater levels in the White Volta River basin of Ghana," *Groundw. Sustain. Dev.*, vol. 6, pp. 150–163, Mar. 2018, doi: 10.1016/j.gsd.2017.12.006.

A. A. Alsumaiei, "A nonlinear autoregressive modeling approach for forecasting groundwater level fluctuation in urban aquifers," *Water (Switzerland)*, vol. 12, no. 3, Mar. 2020, doi: 10.3390/w12030820.

D. Raja and J. J. Edwin, "Comparative Analysis of Time Series Transformers on Multivariate Time Series Data," *Premier Journal of Science*, Dec. 2025, doi: 10.70389/pjs.100132.

K. Küllahcı and A. Altunkaynak, "Eigen time series modeling : a breakthrough approach to spatio-temporal rainfall forecasting in basins," vol. 1, 2024, doi: 10.1007/s00521-024-10864-1.

K. Küllahcı and A. Altunkaynak, "Enhanced rainfall prediction performance via hybrid empirical - singular - wavelet - fuzzy approaches," *Environmental Science and Pollution Research*, no. 0123456789, 2023, doi: 10.1007/s11356-023-26598-x.

J. L. Uc-Castillo, A. E. Marín-Celestino, D. A. Martínez-Cruz, J. Tuxpan-Vargas, and J. A. Ramos-Leal, "A systematic review and meta-analysis of groundwater level forecasting with machine learning

techniques: Current status and future directions," Oct. 01, 2023, *Elsevier Ltd*. doi: 10.1016/j.envsoft.2023.105788.

J. Khan, E. Lee, A. S. Balobaid, and K. Kim, "A Comprehensive Review of Conventional, Machine Leaning, and Deep Learning Models for Groundwater Level (GWL) Forecasting," Feb. 01, 2023, *MDPI*. doi: 10.3390/app13042743.

S. Afrifa, T. Zhang, P. Appiahene, and V. Varadarajan, "Mathematical and Machine Learning Models for Groundwater Level Changes: A Systematic Review and Bibliographic Analysis," Sep. 01, 2022, *MDPI*. doi: 10.3390/fi14090259.

A. Ahmadi *et al.*, "Groundwater Level Modeling with Machine Learning: A Systematic Review and Meta-Analysis," Mar. 01, 2022, *MDPI*. doi: 10.3390/w14060949.

H. Tao *et al.*, "Groundwater level prediction using machine learning models: A comprehensive review," Jun. 07, 2022, *Elsevier B.V.* doi: 10.1016/j.neucom.2022.03.014.

S. Amiri, A. Rajabi, S. Shabanlou, F. Yosefvand, and M. A. Izadbakhsh, "Prediction of groundwater level variations using deep learning methods and GMS numerical model," *Earth Sci. Inform.*, vol. 16, no. 4, pp. 3227–3241, Dec. 2023, doi: 10.1007/s12145-023-01052-1.

Y. W. Chang, W. Sun, P. Y. Kow, M. H. Lee, L. C. Chang, and F. J. Chang, "Advanced groundwater level forecasting with hybrid deep learning model: Tackling water challenges in Taiwan's largest alluvial fan," *J. Hydrol. (Amst).*, vol. 655, Jul. 2025, doi: 10.1016/j.jhydrol.2025.132887.

B. D. Bowes, J. M. Sadler, M. M. Morsy, M. Behl, and J. L. Goodall, "Forecasting groundwater table in a flood prone coastal city with long short-term memory and recurrent neural networks," *Water (Switzerland)*, vol. 11, no. 5, May 2019, doi: 10.3390/w11051098.

A. Altunkaynak, A. Çelik, and E. Kartal, "Eigen-ensemble-transformer model: An exceptionally accurate approach to spatiotemporal sea water level prediction," *Ocean Engineering*, vol. 335, Aug. 2025, doi: 10.1016/j.oceaneng.2025.121727.

A. Wunsch, T. Liesch, and S. Broda, "Groundwater level forecasting with artificial neural networks: A comparison of long short-term memory (LSTM), convolutional neural networks (CNNs), and non-linear autoregressive networks with exogenous input (NARX)," *Hydrol. Earth Syst. Sci.*, vol. 25, no. 3, pp. 1671–1687, Apr. 2021, doi: 10.5194/hess-25-1671-2021.

W. Sun, L. C. Chang, and F. J. Chang, "Deep dive into predictive excellence: Transformer's impact on groundwater level prediction," *J. Hydrol. (Amst).*, vol. 636, Jun. 2024, doi: 10.1016/j.jhydrol.2024.131250.

K. Küllahcı and A. Altunkaynak, "Maximizing daily rainfall prediction accuracy with maximum overlap discrete wavelet transform-based machine learning models," *International Journal of Climatology*, vol. 44, no. 10, pp. 3405–3426, 2024, doi: 10.1002/joc.8530.

H. Wu, T. Hu, Y. Liu, H. Zhou, J. Wang, and M. Long, "TIMESNET: TEMPORAL 2D-VARIATION MODELING FOR GENERAL TIME SERIES ANALYSIS." [Online]. Available: https://github.com/thuml/TimesNet.

K. B. W. Boo, A. El-Shafie, F. Othman, M. M. H. Khan, A. H. Birima, and A. N. Ahmed, "Groundwater level forecasting with machine learning models: A review," Mar. 15, 2024, *Elsevier Ltd*. doi: 10.1016/j.watres.2024.121249.

H. Tao *et al.*, "Groundwater level prediction using machine learning models: A comprehensive review," Jun. 07, 2022, *Elsevier B.V.* doi: 10.1016/j.neucom.2022.03.014.

A. Wunsch, T. Liesch, and S. Broda, “Groundwater level forecasting with artificial neural networks: A comparison of long short-term memory (LSTM), convolutional neural networks (CNNs), and non-linear autoregressive networks with exogenous input (NARX),” *Hydrol. Earth Syst. Sci.*, vol. 25, no. 3, pp. 1671–1687, Apr. 2021, doi: 10.5194/hess-25-1671-2021.

R. J. Hyndman and Athanasopoulos George, *Forecasting: principles and practice.* OTexts, 2018.

R. H. Shumway and Stoffer D. S., *Time series: a data analysis approach using R.* . Chapman and Hall/CRC, 2019.

J. Durbin and S. J. Koopman, *Time series analysis by state space methods.* . Oxford University Press (UK), 2012.

O. Troyanskaya *et al.*, “Missing value estimation methods for DNA microarrays,” 2001. [Online]. Available: http://smi-web.

D. J. Stekhoven and P. Bühlmann, “Missforest-Non-parametric missing value imputation for mixed-type data,” *Bioinformatics*, vol. 28, no. 1, pp. 112–118, Jan. 2012, doi: 10.1093/bioinformatics/btr597.

L. Breiman, “Random Forests,” 2001.

M. Nickel and D. Kiela, “ Poincaré embeddings for learning hierarchical representations.,” *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017.

O. Ganea, G. Bécigneul, and T. Hofmann, “ Hyperbolic neural networks.,” *Adv. Neural Inf. Process. Syst.*, vol. 31, 2018.

S. Bonnabel, “ Stochastic gradient descent on Riemannian manifolds. ,” *IEEE Trans. Automat. Contr.*, vol. 58(9), pp. 2217–2229, 2013.

G. Bécigneul and O.-E. Ganea, “Riemannian Adaptive Optimization Methods,” Feb. 2019, [Online]. Available: http://arxiv.org/abs/1810.00760

Y. Lecun, E. Bottou, Y. Bengio, and P. Haffner, “Gradient-Based Learning Applied to Document Recognition,” 1998.

J. Schmidhuber and Y. Shmidhoobuh, “Tutorial on LSTM Recurrent Nets,” 2003. [Online]. Available: www.idsia.ch

K. Cho *et al.*, “Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation.”

M. Schuster and K. K. Paliwal, “Bidirectional Recurrent Neural Networks,” 1997.

G. Zhang, B. E. Patuwo, and M. Y. Hu, “Forecasting with artificial neural networks: The state of the art,” 1998.

Y. Qin, D. Song, H. Chen, W. Cheng, G. Jiang, and G. Cottrell, “A Dual-Stage Attention-Based Recurrent Neural Network for Time Series Prediction,” Aug. 2017, [Online]. Available: http://arxiv.org/abs/1704.02971

G. Zhang, B. E. Patuwo, and M. Y. Hu, “Forecasting with artificial neural networks: The state of the art,” 1998.

Q. Wen *et al.*, “Transformers in Time Series: A Survey,” May 2023, [Online]. Available: http://arxiv.org/abs/2202.07125

Y. Nie, N. H. Nguyen, P. Sinthong, and J. Kalagnanam, "A Time Series is Worth 64 Words: Long-term Forecasting with Transformers," Mar. 2023, [Online]. Available: http://arxiv.org/abs/2211.14730

E. Kartal and A. Altunkaynak, "Empirical-singular-wavelet based machine learning models for sea level forecasting in the bosphorus strait: A performance analysis," *Ocean Model. (Oxf).*, vol. 188, no. December 2023, p. 102324, 2024, doi: 10.1016/j.ocemod.2024.102324.

J. E. Nash and J. V. Sutcliffe, "River flow forecasting through conceptual models part I — A discussion of principles," *J. Hydrol. (Amst).*, 1970, doi: 10.1016/0022-1694(70)90255-6.

H. V. Gupta, H. Kling, K. K. Yilmaz, and G. F. Martinez, "Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling," *J. Hydrol. (Amst).*, vol. 377, no. 1–2, pp. 80–91, Oct. 2009, doi: 10.1016/j.jhydrol.2009.08.003.

D. N. , , , R. D. , & Moriasi, J. G. Arnold, M. W. Van Liew, R. L. Bingner, R. D. Harmel, and T. L. Veith, " Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. ," *Trans. ASABE*, vol. 50(3), pp. 885–900, 2007.

G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning: with applications in R.* , vol. 103. New York: springer., 2013.

T. Hastie, R. Tibshirani, J. H. Friedman, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction.*, vol. 2. New York: springer., 2009.

I. Chami, A. Wolf, D.-C. Juan, F. Sala, S. Ravi, and C. Ré, "Low-Dimensional Hyperbolic Knowledge Graph Embeddings."

F. Kratzert, D. Klotz, C. Brenner, K. Schulz, and M. Herrnegger, "Rainfall-runoff modelling using Long Short-Term Memory (LSTM) networks," *Hydrol. Earth Syst. Sci.*, vol. 22, no. 11, pp. 6005–6022, Nov. 2018, doi: 10.5194/hess-22-6005-2018.

C. Shen, "A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists," Nov. 01, 2018, *Blackwell Publishing Ltd*. doi: 10.1029/2018WR022643.

H. F. Yang and Y. P. P. Chen, "Hybrid deep learning and empirical mode decomposition model for time series applications," *Expert Syst. Appl.*, vol. 120, pp. 128–138, Apr. 2019, doi: 10.1016/j.eswa.2018.11.019.

S. Bai, J. Z. Kolter, and V. Koltun, "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling," Apr. 2018, [Online]. Available: http://arxiv.org/abs/1803.01271