remote sensing

MDPI

Article

CT-TreeFlow: Probabilistic Groundwater-Potential Mapping Using Remote Sensing-Derived Environmental Predictors in Karst Aquifers

Saeid Pourmorad [1], Mostafa Kabolizade [2], Rui Ferreira [1], Samira Abbasi [3,*] and Luca Antonio Dimuccio [1]

[1] Centre of Studies in Geography and Spatial Planning (CEGOT), Department of Geography and Tourism, Faculty of Arts and Humanities, University of Coimbra, 3004-530 Coimbra, Portugal; uc2023140908@student.uc.pt (S.P.); luca@ci.uc.pt (L.A.D.)
[2] Department of Remote Sensing and GIS, Faculty of Earth Sciences, Shahid Chamran University of Ahvaz, Ahvaz 61357-83151, Khuzestan, Iran
[3] Department of Earth Sciences, Faculty of Science and Technology, University of Coimbra, 3030-790 Coimbra, Portugal
* Correspondence: s.abbasi.environment@gmail.com

Highlights

What are the main findings?

- CT-TreeFlow integrated probabilistic machine learning with Landsat-derived environmental predictors and Copernicus DEM morphometric variables to improve groundwater-potential assessment in structurally heterogeneous karst terrains.
- The framework achieved improved predictive accuracy, stronger probabilistic consistency, and reduced spatial prediction bias under geographically independent validation, while identifying fracture density, lithology, drainage organisation, and terrain-regulated recharge conditions as the dominant hydrogeological controls on groundwater occurrence.

What are the implications of the main findings?

- Integrating remote-sensing-derived environmental data with uncertainty-aware probabilistic modelling improves the physical interpretability and spatial reliability of regional groundwater assessments in complex carbonate aquifers.
- The proposed framework offers a transferable geospatial strategy for risk-informed groundwater exploration, hydrogeological decision support, and sustainable management of groundwater resources in environmentally heterogeneous karst systems.

Abstract

Groundwater-potential assessment in karst aquifers is complicated by pronounced spatial heterogeneity driven by structural permeability, lithological variability, recharge redistribution, and unresolved subsurface conduit connectivity. Although machine-learning approaches have improved regional groundwater mapping, most existing models provide only deterministic predictions and offer limited information on predictive uncertainty and hydrogeological reliability. To address this limitation, we propose CT-TreeFlow. This probabilistic groundwater assessment framework goes beyond conventional machine-learning models by explicitly learning the full conditional probability distribution of groundwater favourability rather than a single deterministic estimate. The framework integrates sparse probabilistic environmental routing, conditional density estimation, hydrogeologically constrained pseudo-absence generation, geographically structured spatial validation, and explainability-driven interpretation within a unified modelling architecture, enabling

Check for updates

Academic Editor: Teodosio Lacava

Received: 22 May 2026
Revised: 26 June 2026
Accepted: 30 June 2026
Published: 7 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

simultaneous groundwater prediction, uncertainty quantification, and hydrogeological interpretation. The framework was applied to the Zagros karst system in Khuzestan Province, Iran, using remote-sensing-derived environmental predictors, Copernicus DEM-based morphometric variables, geological–structural datasets, and hydroclimatic indicators. Performance was evaluated against LightGBM and XGBoost using GroupKFold spatial cross-validation. CT-TreeFlow achieved a mean RMSE of 2.737 and a mean R^2 of 0.852, while also providing spatially explicit uncertainty estimates and probabilistic prediction intervals. Explainability analyses identified fracture density, lithology, drainage organisation, and terrain-controlled recharge conditions as the dominant controls on groundwater favourability. Predicted high-favourability zones showed strong spatial correspondence with major carbonate formations and independent spring–cave inventories, supporting the hydrogeological plausibility of the mapped patterns. These results demonstrate that probabilistic modelling can provide more reliable and physically interpretable groundwater assessments than deterministic approaches in structurally complex karst environments. CT-TreeFlow offers a transferable framework for uncertainty-aware groundwater exploration and regional hydrogeological decision support in heterogeneous aquifer systems.

Keywords: groundwater-potential mapping; karst aquifers; probabilistic machine learning; remote sensing; predictive uncertainty; spatial cross-validation

1. Introduction

Groundwater stored in fractured and karstified carbonate aquifers represents one of the most important freshwater resources in many water-stressed regions. However, groundwater occurrence in these environments remains difficult to predict because flow and storage are controlled by complex interactions among lithology, structural deformation, topography, recharge dynamics, climatic forcing, and subsurface hydraulic connectivity [1,2]. Unlike porous-media aquifers, groundwater circulation in karst systems is dominated by fracture-enhanced permeability, dissolution-driven conduit development, and structurally controlled recharge, resulting in highly heterogeneous groundwater-flow patterns. These complexities are particularly evident in tectonically active carbonate provinces, where geological structures strongly influence aquifer geometry, groundwater connectivity, and recharge organisation. The Zagros fold–thrust belt represents one of the most structurally deformed hydrogeological regions in the Middle East, with folding, faulting, and fracture development exerting major controls on groundwater occurrence [3,4]. Carbonate formations, especially the Asmari Formation, form the principal regional aquifers and are characterised by extensive fracture systems and variable karstification [5,6]. Consequently, groundwater occurrence depends primarily on secondary porosity associated with fractures, joints, faults, and conduits rather than matrix permeability, making regional groundwater assessment particularly challenging.

Recent advances in remote sensing, geospatial analysis, and machine learning have substantially improved the investigation of groundwater-controlling factors [7–9]. Remote-sensing products and DEM-derived morphometric variables provide spatially continuous information on terrain organisation, recharge redistribution, infiltration potential, vegetation response, and hydrological processes [7,8], while tree-based machine-learning algorithms effectively model nonlinear environmental interactions in large geospatial datasets [10,11]. Accordingly, gradient-boosting frameworks such as XGBoost and LightGBM have become widely adopted for groundwater prediction [12–16], and hybrid and ensemble approaches have further improved groundwater-potential mapping in frac-

tured and karst aquifers [17–21]. Despite these advances, several important challenges remain. First, most groundwater-potential models remain deterministic, yielding a single prediction despite the inherently uncertain nature of groundwater occurrence in karst systems [22]. Second, many studies still rely on random validation strategies that ignore spatial autocorrelation and may therefore overestimate predictive performance and spatial transferability [23]. Third, although remote-sensing variables are increasingly incorporated into groundwater modelling, their influence is often interpreted statistically rather than hydrogeologically, limiting physical understanding of how structural, lithological, geomorphological, and hydrological factors jointly control groundwater occurrence. Consequently, an important gap persists between predictive performance, uncertainty quantification, and hydrogeological interpretability.

To address this gap, we propose CT-TreeFlow, an uncertainty-aware probabilistic framework for groundwater-potential assessment in structurally heterogeneous karst aquifers. Unlike conventional deterministic machine-learning models that predict a single groundwater-potential value, CT-TreeFlow learns the full conditional probability distribution of groundwater favourability, enabling simultaneous groundwater prediction, uncertainty quantification, and hydrogeological interpretation. The framework combines five complementary components within a unified workflow: (i) sparse probabilistic environmental routing, (ii) conditional density estimation, (iii) hydrogeologically informed pseudo-absence generation, (iv) geographically structured spatial validation, and (v) explainability-driven hydrogeological interpretation. By integrating probabilistic prediction with geographically independent validation and physically interpretable environmental controls, CT-TreeFlow addresses key limitations of existing groundwater-potential mapping frameworks. Accordingly, the objectives of this study are to: (i) develop CT-TreeFlow as an uncertainty-aware probabilistic framework for groundwater-potential mapping in structurally heterogeneous karst terrains; (ii) evaluate its predictive accuracy, uncertainty reliability, and spatial generalisation using geographically structured validation and probabilistic diagnostics; (iii) compare its performance with LightGBM and XGBoost benchmark models; and (iv) identify the dominant hydrogeological and environmental controls governing groundwater occurrence across the Zagros karst system. Beyond this case study, CT-TreeFlow provides a transferable framework for uncertainty-aware groundwater exploration, regional groundwater assessment, and hydrogeological decision support in fractured carbonate aquifers and other environmentally heterogeneous groundwater systems.

2. Geological and Geographical Setting of the Study Area

Khuzestan Province covers approximately 63,355 km^2 in southwestern Iran and occupies the transition between the Zagros Mountains and the Persian Gulf coastal plain (Figure 1). The study area is located within the foreland domain of the NW–SE-trending Zagros Fold-and-Thrust Belt (ZFTB), an active orogenic system formed by the long-term convergence of the Arabian and Eurasian plates from the Late Cretaceous to the Cenozoic [4,24]. Structurally, it lies within the Zagros Folded Belt between the High Zagros Fault (HZF) and the Mountain Front Fault (MFF), where folding, faulting, and fracture systems strongly control aquifer geometry, secondary permeability, groundwater-flow pathways, recharge, and discharge [25]. The spatial distribution of lithological units, structural features, and mapped karst springs is shown in Figure 1, while the regional stratigraphic framework is presented in Supplementary Figure S1. Groundwater occurrence is primarily controlled by the interaction of stratigraphy, tectonic deformation, and geomorphology. The exposed succession comprises the Pabdeh, Asmari, Gachsaran, Mishan, Aghajari, and Bakhtiari formations [26,27], forming a heterogeneous hydrostratigraphic system that includes pro-

ductive carbonate aquifers, semi-confining units, and regional aquitards. Among these, the fractured and karstified Asmari Formation constitutes the principal regional aquifer owing to its enhanced permeability and storage capacity [5], whereas the evaporitic Gachsaran Formation commonly restricts vertical groundwater movement [28]. The Mishan Formation locally behaves as a semi-confining unit, while the sandstone layers of the Aghajari Formation and the coarse clastic deposits of the Bakhtiari Formation locally promote infiltration and groundwater recharge [29].

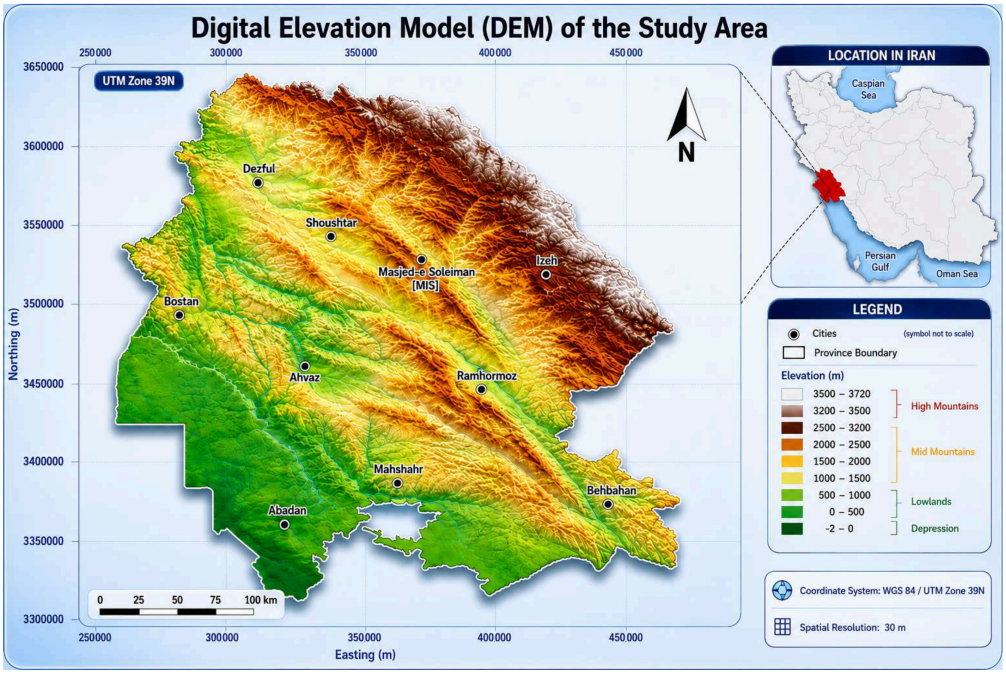


Figure 1. Location and topographic setting of the Khuzestan Province in southwestern Iran (study area). Elevation map derived from the digital elevation model (DEM) is classified into four altitudinal zones (0–3720 m), illustrating the transition from the Persian Gulf foreland plains to the Zagros Fold-and-Thrust Belt.

The lithological framework shown in Supplementary Figure S1 supported the hydrogeological interpretation of the lithology predictor used in the modelling workflow. Carbonate formations, particularly the Asmari Formation and structurally connected carbonate units, were considered highly favourable for karstification, fracture-controlled permeability, and groundwater storage, whereas evaporitic, marly, and shale-rich formations were regarded as less favourable or locally confining. Accordingly, the lithological susceptibility predictor was derived from the regional geological map based on the relative dissolution susceptibility, fracture development, and hydrogeological behaviour of the mapped lithological units, as summarised in Supplementary Table S1. Deeper carbonate formations, including the Sarvak Formation, may also contribute to regional groundwater storage where structural exposure or hydraulic connectivity exists through fracture systems. Karst features such as sinkholes, conduits, and underground drainage networks preferentially develop along anticlines, enhancing focused recharge into regional aquifers [3]. Geomorphology further influences groundwater occurrence, with steep dissected terrain promoting rapid runoff and preferential infiltration through fractures, whereas gentle slopes and topographic depressions favour diffuse recharge and moisture retention [30]. The combined effects of lithology, tectonic deformation, karst development, and terrain morphology create a highly heterogeneous hydrogeological environment that governs groundwater occurrence, recharge distribution, and flow pathways, making the Zagros karst system

an appropriate natural laboratory for evaluating uncertainty-aware groundwater-potential mapping frameworks.

3. Methodology

3.1. Hydrogeological Conceptual Model and Prediction Target

Groundwater-potential modelling was formulated as a hydrogeologically constrained probabilistic spatial prediction problem to identify areas with comparatively favourable groundwater-support conditions across southwestern Iran. Rather than treating groundwater prediction as a purely statistical pattern-recognition task, the framework incorporates hydrogeological processes related to structurally controlled recharge, lithological permeability contrasts, fracture-enhanced secondary permeability, terrain-controlled infiltration, and karst groundwater circulation within the Zagros Fold–Thrust Belt [2,4]. Groundwater occurrence is primarily controlled by tectonic deformation, lithological heterogeneity, fracture connectivity, karst conduit development, geomorphology, and structurally guided groundwater-flow pathways [4]. Fractured and karstified carbonate formations, particularly the Asmari Formation and hydraulically connected carbonate units, constitute the principal aquifers, whereas the Gurpi, Mishan, and Gachsaran formations generally act as aquitards or semi-confining units that restrict groundwater circulation [5,6]. This conceptual model is consistent with structurally heterogeneous karst systems, where conduit–fracture–matrix interactions and structural discontinuities generate strong spatial variability in groundwater occurrence and recharge connectivity [31]. Groundwater potential was interpreted as the combined effect of recharge concentration, lithological permeability, fracture development, climatic water balance, terrain-controlled infiltration, runoff redistribution, and karst conduit organisation. These nonlinear relationships are difficult to represent with conventional deterministic groundwater-potential mapping methods [7,8]. Although tree-based and ensemble learning algorithms effectively model complex environmental interactions [9,17,32], most existing approaches remain deterministic and provide limited information on predictive uncertainty and hydrogeological transition zones [30,33]. The proposed CT-TreeFlow framework (Figure 2) integrates hydrogeological predictor compilation, environmentally constrained spatial learning, sparse probabilistic routing, conditional density estimation, uncertainty-aware optimisation, and spatial uncertainty quantification within a unified modelling architecture. GroupKFold spatial validation was employed to minimise spatial leakage and autocorrelation between calibration and testing subsets [23].

Groundwater modelling was formulated as a probabilistic continuous-estimation problem. Let (Y_i) denote the continuous relative groundwater-support potential index (hereafter, groundwater-potential index) at location (i). Higher values indicate comparatively favourable groundwater-support conditions associated with enhanced recharge, fracture-enhanced permeability, structurally connected groundwater-flow pathways, and favourable carbonate aquifer settings, whereas lower values represent less favourable conditions. The response variable was derived from processed groundwater-depth observations from the regional monitoring-well inventory. The objective was not to predict groundwater-table depth, aquifer productivity, transmissivity, storage, sustainable yield, or groundwater reserves, but rather to identify the relative spatial distribution of groundwater-support potential for regional screening and prioritisation. Because groundwater depth is only an indirect proxy for groundwater-support conditions, particularly in karst aquifers, the groundwater-potential index was interpreted as a relative hydrogeological indicator rather than a direct measure of aquifer performance. Shallow groundwater conditions were considered comparatively more favourable, whereas deeper conditions indicated relatively less favourable settings within the limits of the available monitoring-well observations

and environmental predictors. Independent karst spring and cave inventories were excluded from model training and optimisation and were used solely for post-modelling hydrogeological plausibility assessment as indirect indicators of groundwater discharge concentration and conduit development rather than measures of sustainable groundwater availability or site-scale abstraction potential.

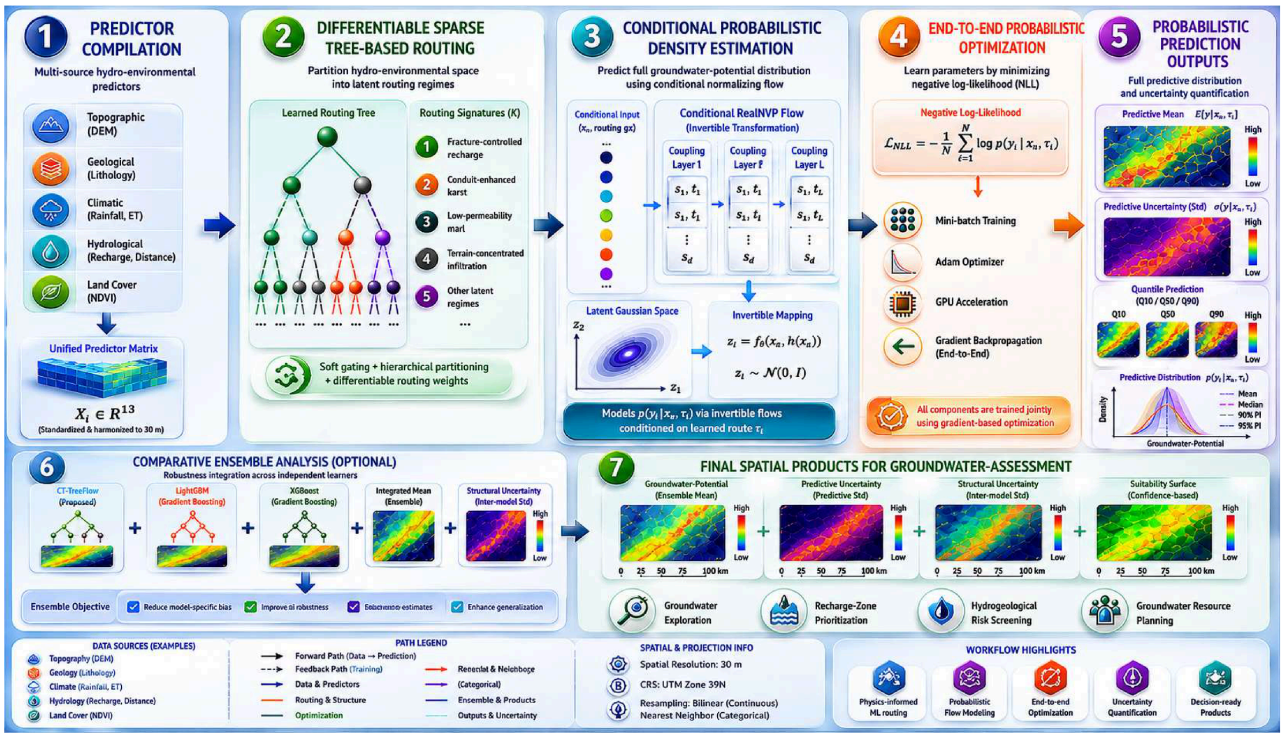


Figure 2. Conceptual workflow of the proposed CT-TreeFlow framework for probabilistic and uncertainty-aware groundwater-potential modelling under heterogeneous hydrogeological conditions.

Groundwater-Potential Index Transformation

Groundwater-depth observations were used to derive a relative groundwater-potential index for hydrogeological interpretation and regional prioritisation under structurally heterogeneous karst conditions [6,31]. The purpose of this transformation was to express groundwater-depth-derived information within a dimensionless groundwater-support framework, while avoiding direct interpretation of model outputs such as groundwater-table elevation, aquifer yield, transmissivity, groundwater storage, sustainable abstraction capacity, or groundwater reserve. The groundwater-potential index was derived using inverse min–max scaling:

$$GPI_i = 1 - \frac{D_i - D_{\min}}{D_{\max} - D_{\min}}$$

where GPI_i denotes the transformed groundwater-potential index at spatial location i, D_i is the observed groundwater depth, and D_{min} and D_{max} are the minimum and maximum groundwater-depth observations in the regional monitoring-well inventory, respectively [7,8]. Values approaching 1 indicate comparatively shallow groundwater-depth conditions and were interpreted as more favourable groundwater-support settings. In contrast, values approaching 0 indicate comparatively deep groundwater-depth conditions and were interpreted as less favourable groundwater-support settings within the regional hydrogeological continuum. This transformation was used only as a relative regional hydrogeological prioritisation framework. It does not imply that groundwater favourability varies linearly with groundwater depth under all hydrogeological conditions, particularly in karst aquifers where concealed conduits, perched conditions, deep

circulation, structural compartmentalisation, and local discharge controls may produce nonlinear groundwater-depth–favourability relationships [17,22]. Therefore, the transformed index was interpreted cautiously as an environmentally conditioned indicator of relative groundwater-support potential, not as a direct measure of aquifer productivity or sustainable groundwater availability.

To preserve hydrogeological meaning in model evaluation, predictive performance, residual diagnostics, and observed–predicted error statistics were reported on the original groundwater-depth-derived continuous response scale. The transformed index was used for relative interpretation and spatial prioritisation, whereas the original response scale was retained for assessing model error magnitude and predictive performance. This ensured that RMSE, MAE, residual behaviour, and cross-validation statistics remained physically interpretable while still allowing the final maps to be expressed as relative groundwater-potential patterns [6,31]. The supervised-learning unit in CT-TreeFlow consisted of geographically referenced point-level hydrogeological observations linked to extracted hydro-environmental predictor values. Environmentally constrained low-groundwater-support reference samples were incorporated as contrast information rather than as confirmed groundwater-absence locations [34,35]. This distinction is important because the absence of groundwater cannot be directly inferred in regional karst systems without site-specific hydraulic testing. Accordingly, CT-TreeFlow outputs should be interpreted as continuous probabilistic estimates of relative groundwater-support potential rather than binary groundwater-occurrence labels, direct groundwater-depth predictions, transmissivity estimates, aquifer-productivity metrics, or groundwater-reserve calculations [17,19]. The predictor vector x_i represents the hydro-environmental conditioning variables described in Section 3.2, including terrain morphology, lithological conditions, fracture-density characteristics, hydrological proximity indicators, climatic variables, and land-surface controls [3,31]. The modelling objective of CT-TreeFlow was therefore to estimate the conditional predictive distribution:

$$P(y_i \mid x_i)$$

where y_i denotes the groundwater-potential response at spatial location i, and x_i represents the corresponding hydro-environmental conditioning vector. This formulation enables uncertainty-aware regional groundwater assessment and spatial prioritisation of comparatively favourable groundwater-support conditions while preserving transparency regarding prediction uncertainty under environmentally heterogeneous karst settings [30,36]. CT-TreeFlow was therefore treated as a probabilistic regional groundwater-potential assessment framework rather than a deterministic groundwater-table prediction model. Springs and cave inventories were operationally separated from model fitting, predictor optimisation, pseudo-absence generation, probabilistic routing calibration, hyperparameter optimisation, and spatial validation procedures. They were reserved exclusively for post-modelling assessment of hydrogeological plausibility [36,37]. Consequently, spatial correspondence between predicted groundwater-potential patterns and mapped karst features was interpreted only as independent plausibility evidence of hydrogeological coherence, not as proof of sustainable groundwater availability or site-scale abstraction potential [36,38]. The complete dataset roles, inventory metadata, and pre-processing details are provided in Table 1 and Supplementary Table S2.

The environmental predictors incorporated into CT-TreeFlow were selected to represent hydrogeologically interpretable controls associated with recharge concentration, structural permeability development, climatic recharge availability, terrain-controlled infiltration, runoff redistribution, and karstification susceptibility. Rather than directly resolving subsurface conduit geometry or site-scale aquifer hydraulic properties, these variables provide spatially continuous environmental proxies that describe variability in regional ground-

water potential under heterogeneous karst hydrogeological conditions. Consequently, the adopted framework should be interpreted as a probabilistic regional groundwater assessment system designed for hydrogeological screening and support for groundwater exploration, rather than as a substitute for pumping tests, tracer experiments, hydrochemical investigations, or physically based groundwater-flow simulations [6,31]. The generated outputs therefore represent spatial patterns of groundwater-potential derived from environmentally conditioned hydrogeological relationships that influence groundwater occurrence and accessibility, rather than direct measurements of transmissivity, storativity, groundwater reserves, or well productivity (Figure 2).

Table 1. Operational role of hydrogeological and environmental datasets within the CT-TreeFlow framework.

Dataset Group	Hydrogeological Interpretation	Operational Role in CT-TreeFlow	Used for Model Fitting	Used for Independent Hydrogeological Plausibility Assessment
Groundwater monitoring wells	Groundwater-depth observations representing relative regional groundwater-support conditions	Calibration of the continuous groundwater-potential index used for probabilistic groundwater-potential modelling	✅ Yes	❌ No
Environmentally constrained pseudo-absence samples	Areas representing comparatively less favourable groundwater-support conditions within the regional hydrogeological continuum	Hydrogeological contrast learning using environmentally constrained low-groundwater-support reference conditions	✅ Yes	❌ No
Karst springs	Natural groundwater discharge zones and hydraulic convergence indicators	Independent post-modelling hydrogeological plausibility assessment	❌ No	✅ Yes
Cave entrances	Indicators of conduit-system development and structurally controlled subsurface groundwater connectivity	Independent karst-system plausibility assessment	❌ No	✅ Yes
Lithological formations	Aquifer–aquitard architecture, lithological permeability contrasts, and karstification susceptibility	Geological conditioning predictors representing lithological controls on groundwater-potential variability	✅ Yes	❌ No
Fault and fracture network	Secondary permeability development and structurally controlled groundwater connectivity	Structural predictor derivation describing fracture-controlled groundwater-flow pathways	✅ Yes	❌ No

Table 1. *Cont.*

Dataset Group	Hydrogeological Interpretation	Operational Role in CT-TreeFlow	Used for Model Fitting	Used for Independent Hydrogeological Plausibility Assessment
Terrain, drainage, climatic, and vegetation variables	Recharge redistribution, runoff–infiltration partitioning, climatic recharge controls, and land-surface environmental variability	Hydro-environmental predictor generation for probabilistic groundwater-potential modelling	✅ Yes	❌ No

Table note. Karst springs and cave entrances were completely excluded from pseudo-absence generation, predictor optimisation, probabilistic routing calibration, hyperparameter optimisation, model fitting, uncertainty optimisation, and geographically structured spatial cross-validation procedures. These datasets were reserved exclusively for independent post-modelling hydrogeological plausibility assessment after completion of model development and validation. Consequently, subsequent spatial correspondence between predicted groundwater-potential patterns and mapped karst hydrogeological features can be interpreted as independent evidence of hydrogeological plausibility, thereby supporting the environmental coherence of the CT-TreeFlow framework.

3.2. Geospatial Database Construction and Environmental Conditioning Factors

A harmonised geospatial database was constructed to represent the principal hydrogeological, geomorphological, climatic, structural, land-surface, and vegetation-related controls governing groundwater occurrence within the structurally heterogeneous Zagros karst domain. The database integrated topographic, geological, hydrological, climatic, hydrogeological, land-cover, Landsat-derived, and vegetation-index datasets obtained from national inventories, meteorological observations, remote sensing products, and groundwater monitoring archives. Complete metadata, spatial resolutions, temporal coverage, source institutions, pre-processing procedures, predictor derivation methods, operational modelling roles, and data sources are provided in Supplementary Table S1. The spatial distributions of the final conditioning layers are shown in Supplementary Figure S2.

All GIS pre-processing, raster harmonisation, environmental predictor derivation, spatial overlay, and map production procedures were performed within a reproducible GIS and programming workflow using ArcGIS Pro 3.3 (Esri, Redlands, CA, USA), QGIS 3.34 LTR (QGIS Development Team, Open-Source Geospatial Foundation, https://qgis.org) and Python 3.11 using NumPy 1.26, Pandas 2.2, Scikit-learn 1.5, PyTorch 2.3, SHAP 0.46, and Rasterio 1.3. All spatial layers were projected to UTM Zone 39N and harmonised to a common 30 m grid before predictor extraction and model implementation. Environmental predictors were selected based on their hydrogeological relevance to recharge concentration, runoff redistribution, structural permeability development, karstification susceptibility, climatic recharge availability, vegetation response, and terrain-controlled infiltration processes, rather than on purely statistical feature-selection criteria. This process-informed predictor-selection strategy is consistent with groundwater-assessment approaches developed for structurally complex and karst-dominated environments [7,17,39].

3.2.1. Geospatial Database and Hydrogeological Inventories

The hydrogeological database comprised four operational inventory groups: groundwater monitoring wells, environmentally constrained pseudo-absence samples, mapped karst springs, and documented cave entrances. The database was designed to ensure hydrogeological transparency, validation independence, and reproducible implementation within the CT-TreeFlow framework. Detailed metadata, spatial characteristics, temporal coverage, pre-processing procedures, and modelling roles are provided in Supplementary Table S2. Groundwater monitoring wells formed the primary quantitative hydrogeological dataset, comprising 7313 observations collected between 2018 and 2024 from the Ministry of Energy and Water Resources Management Company (MOE-WRMC). After coordinate verification,

duplicate removal, spatial consistency screening, and filtering of incomplete records, all observations were harmonised to a common 30 m spatial framework. Calibration, optimisation, and validation were performed using geographically referenced point observations rather than raster-derived groundwater surfaces to minimise information leakage, preserve observational independence, and avoid artificial inflation of sample size due to spatial autocorrelation. The response variable was derived from groundwater-depth observations that represent regional groundwater variability across the Zagros karst system. Comparatively shallow groundwater conditions indicate more favourable groundwater-support environments associated with enhanced recharge, secondary permeability, and hydraulically connected carbonate aquifers, whereas deeper groundwater reflects comparatively lower groundwater-support potential [6]. Consistent with karst hydrogeology, groundwater circulation is strongly controlled by conduit–fracture–matrix interactions and heterogeneous subsurface drainage organisation [31]. Unlike conventional groundwater-potential classification methods, CT-TreeFlow retained the response variable on its original continuous numerical scale throughout calibration, prediction, and validation [22,40].

Numerical standardisation was applied only to selected predictor pre-processing and optimisation steps to improve computational stability without altering the physical interpretation of the groundwater-response variable [41]. Accordingly, all residual analyses, uncertainty assessments, probabilistic calibration, and performance metrics were calculated using the original response scale. Environmentally constrained pseudo-absence samples were generated to support supervised probabilistic learning under hydrogeologically realistic conditions. Pseudo-absence allocation excluded areas surrounding groundwater wells, major structural permeability corridors, perennial drainage networks, highly karstifiable carbonate formations, and environmentally favourable fracture-density domains, with additional spatial filtering applied to reduce clustering and short-range spatial autocorrelation. The complete set of exclusion criteria is provided in Supplementary Table S3. Independent hydrogeological plausibility datasets comprised 250 mapped karst springs and 86 documented cave entrances, compiled from regional hydrogeological investigations, national groundwater databases, and speleological inventories. Springs represent groundwater-discharge concentration zones, whereas caves provide indirect evidence of conduit development and structurally controlled subsurface drainage within carbonate terrains [6,31]. These datasets were excluded from pseudo-absence generation, predictor derivation, model calibration, optimisation, uncertainty analysis, and spatial validation, and were used exclusively for independent post-modelling hydrogeological plausibility assessment. This separation minimises circularity bias in interpretation, strengthens hydrogeological interpretability, and provides an independent evaluation of the predicted groundwater-favourability patterns.

3.2.2. Environmental Predictor Derivation

A total of 13 hydro-environmental conditioning factors (distance from watercourses, terrain ruggedness index (TRI), evapotranspiration, topographic position index (TPI), relief energy, land cover, aspect, heat load index (HLI), temperature, normalised difference vegetation index (NDVI), drainage density, fracture density, and lithology) were retained for probabilistic groundwater-assessment modelling following hydrogeological screening, multicollinearity assessment, and spatial harmonisation (Supplementary Table S1; Supplementary Figure S2). The selected predictors represent the principal environmental controls on recharge redistribution, structural permeability, climatic water balance, terrain-controlled infiltration, drainage organisation, vegetation response, and susceptibility to karstification within the heterogeneous Zagros karst system [7,17]. They were designed to characterise regional groundwater-response variability rather than local aquifer hydraulic

properties or site-scale groundwater productivity [6,31]. Topographic predictors were derived from the 30 m Copernicus digital elevation model (DEM) and included TRI, TPI, relief energy, aspect, and heat load index (HLI), representing geomorphological heterogeneity, runoff redistribution, solar-radiation variability, and preferential infiltration processes that influence recharge in mountainous karst terrains [6,42]. In these structurally dissected carbonate environments, terrain morphology strongly controls recharge concentration, epikarst development, and groundwater-flow pathways [31].

Structural and geological predictors included fracture density and lithological susceptibility, derived from regional tectonic and geological datasets provided by the Geological Survey of Iran (GSI). Fracture density was calculated using kernel density analysis of mapped tectonic discontinuities, while lithological units were classified based on susceptibility to karstification, structural characteristics, and hydrogeological behaviour. Structural deformation within the Zagros Fold–Thrust Belt strongly influences secondary permeability, conduit development, and structurally controlled recharge, whereas fractured carbonate formations generally exhibit higher groundwater-storage potential and preferential subsurface flow [2,4,31,43]. Vegetation and land-surface conditions were represented using NDVI and land-cover data derived from Landsat-9 imagery and national land-cover inventories. The Landsat-9 scenes used for NDVI generation, including acquisition dates and WRS-2 Path/Row identifiers, are listed in Supplementary Table S4. NDVI was calculated from the red and near-infrared bands following the formulations of Rouse et al. [39] and Tucker [44]. Land-cover classes were reclassified into hydrologically meaningful categories representing vegetation density, infiltration capacity, runoff potential, and land-surface characteristics following established groundwater and hydrological classification approaches [17,30].

3.2.3. Spatial Harmonisation and Pre-Processing

All spatial datasets were harmonised within a unified geodetic and raster-processing framework before probabilistic modelling. Raster and vector layers were projected to UTM Zone 39N and resampled to a common 30 m grid to ensure geometric consistency during predictor extraction, spatial overlay, and groundwater-response modelling. Continuous variables were resampled using bilinear interpolation, whereas categorical variables, including lithology and land cover, were resampled using nearest-neighbour assignment to preserve class integrity. To improve computational stability during optimisation, continuous predictor variables were standardised to z-scores before model calibration [41]. This procedure was applied only to predictor variables and did not modify the original groundwater-response scale used for prediction, residual analysis, uncertainty assessment, probabilistic calibration, or model evaluation. Consequently, all predictive outputs and performance metrics were reported using the original groundwater-response scale. Missing raster cells, local geometric artefacts, and topological inconsistencies were corrected using neighbourhood-based interpolation and spatial consistency filtering. A 3 × 3 moving-window kernel (90 m × 90 m) was applied, with continuous variables interpolated from neighbouring cells and categorical datasets corrected using nearest-neighbour assignment to preserve class integrity. GIS pre-processing, raster harmonisation, and predictor derivation followed established groundwater-modelling protocols [20,32], while maintaining spatial consistency among terrain, geological, hydrological, climatic, and land-surface datasets. Because spatial autocorrelation can artificially inflate model performance under random validation, all model evaluation procedures employed spatial GroupKFold partitioning rather than random sampling [23]. This geographically structured validation framework minimises spatial information leakage between calibration and testing subsets and provides more reliable estimates of predictive generalisation under environmentally

heterogeneous hydrogeological conditions. Fold-wise sample distributions, spatial partition statistics, and validation geometries are provided in Supplementary Table S5.

3.2.4. Predictor Screening and Final Conditioning Stack

Before developing probabilistic models, an extended hydro-environmental database was compiled to represent geomorphological, climatic, structural, hydrological, vegetation-related, and hydrogeological controls on groundwater response behaviour within the Zagros karst system (Supplementary Table S1). Predictor screening was subsequently performed using a combined hydrogeological interpretability assessment and statistical redundancy analysis to construct a physically meaningful, environmentally stable, and reproducible modelling framework. Pairwise multicollinearity among continuous predictors was evaluated using Pearson correlation analysis following Rodgers and Nicewander [45,46], and the resulting correlation structure is presented in Figure 3. Predictors exhibiting strong correlation ($|r| > 0.80$), duplicated environmental meaning, unstable spatial behaviour, or limited hydrogeological interpretability were excluded before model implementation. Variable selection was guided not only by statistical redundancy but also by hydrogeological process overlap, conceptual redundancy, environmental stability, and physical interpretability in karst groundwater systems [6,31,47]. This process-informed screening strategy reduced multicollinearity, improved the stability of probabilistic learning, and enhanced the interpretability, reproducibility, and hydrogeological transparency of the final framework. Karst spring locations and documented cave inventories, although compiled during database development, were intentionally excluded from pseudo-absence generation, predictor optimisation, probabilistic routing calibration, model fitting, hyperparameter tuning, uncertainty optimisation, and geographically structured spatial validation [48]. These datasets were reserved exclusively for independent post-modelling hydrogeological plausibility assessment, thereby reducing circular interpretation bias and preserving methodological independence. Consequently, spatial agreement between predicted groundwater-response patterns and independently mapped karst features provides evidence of hydrogeological plausibility for the CT-TreeFlow framework.

The final CT-TreeFlow implementation retained 13 hydro-environmental conditioning factors representing the principal controls on recharge concentration, structural permeability, terrain-controlled infiltration, climatic recharge availability, runoff redistribution, drainage organisation, and karstification susceptibility within folded carbonate terrains. The retained predictors were selected to maximise hydrogeological interpretability while maintaining reproducible GIS derivation and environmentally stable probabilistic learning behaviour. Accordingly, the final database should be viewed as a process-informed hydrogeological predictor framework rather than an unconstrained collection of statistical features. All retained variables possess explicit hydrogeological meaning, are reproducible from documented datasets and workflows, and were evaluated independently of the karst spring and cave inventories reserved for external plausibility assessment. Complete model architectures, optimisation settings, routing constraints, validation procedures, uncertainty diagnostics, and implementation details for CT-TreeFlow, LightGBM, and XGBoost are provided in Supplementary Table S6.

3.3. Environmentally Constrained Pseudo-Absence Sampling and Spatial Validation Design

Because verified low-groundwater-support observations are rarely available in regional karst studies, pseudo-absence samples were generated using an environmentally constrained strategy rather than unrestricted random sampling to reduce hydrogeological contradiction, sampling bias, and spatial leakage during probabilistic model development [20,35]. This approach is consistent with previous groundwater studies that

recommend environmentally informed pseudo-absence allocation to improve consistency between hydrogeological processes and statistical sampling [17,30]. Candidate pseudo-absence locations were selected from environmentally plausible low-groundwater-support areas after applying hydrogeological exclusion masks based on groundwater wells, perennial drainage corridors, highly karstifiable carbonate formations, and favourable fracture-density domains [6,49]. A minimum inter-sample spacing criterion was also applied to reduce clustering and short-range spatial autocorrelation. These pseudo-absence samples represent environmentally constrained low-support reference conditions rather than confirmed groundwater-absence observations because concealed conduits and unresolved subsurface connectivity may occur outside mapped favourable zones. Mapped karst springs and documented cave entrances were excluded from pseudo-absence generation, predictor derivation, model calibration, hyperparameter optimisation, uncertainty calibration, and spatial validation, and were reserved exclusively for independent post-training hydrogeological plausibility assessment. Complete exclusion criteria and allocation settings are provided in Supplementary Table S3. The sensitivity of CT-TreeFlow to pseudo-absence allocation was evaluated by comparing environmentally constrained and unconstrained random sampling strategies using identical predictor sets, pre-processing procedures, optimisation settings, and spatial validation partitions. Comparative results showed that environmentally constrained sampling slightly improved predictive stability and uncertainty estimation while preserving the overall groundwater-response patterns, indicating that CT-TreeFlow learned robust environmental relationships rather than simply reproducing the imposed exclusion criteria.

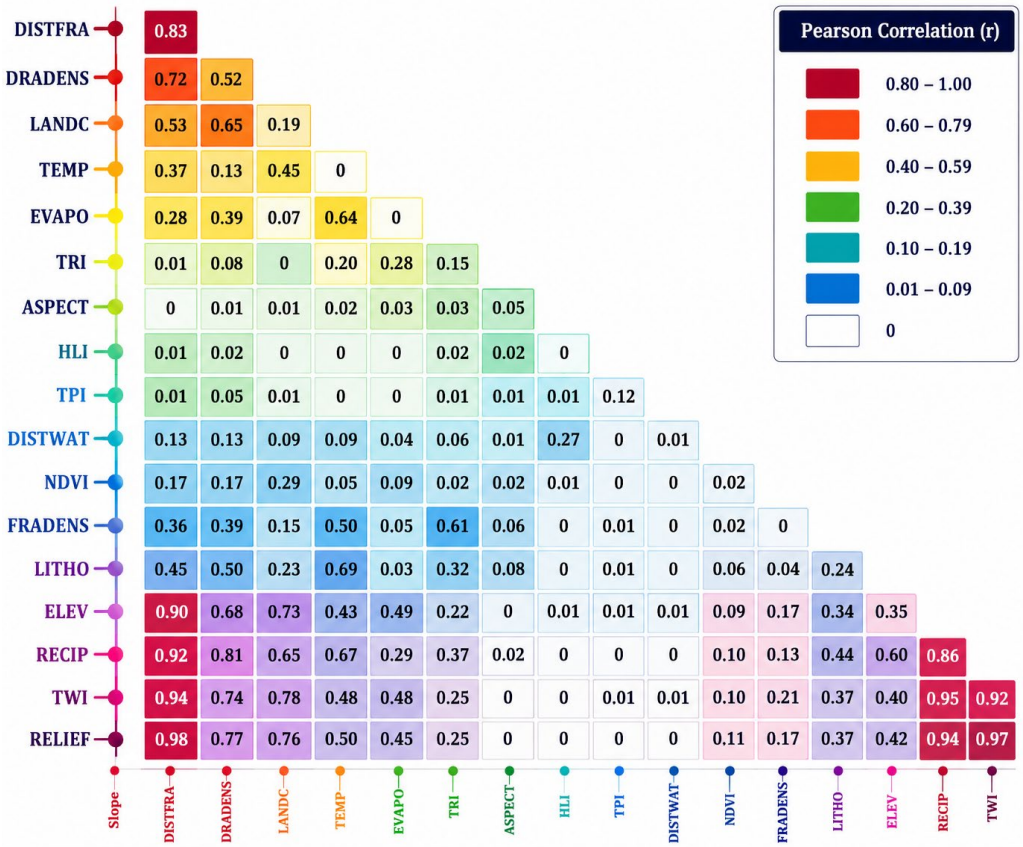


Figure 3. Pearson correlation matrix of environmental conditioning factors used for groundwater-potential modelling. Cell values represent Pearson correlation coefficients (r). Predictor screening was performed to identify multicollinearity and remove highly redundant variables ($|r| > 0.80$) before model development.

Spatial validation was performed using geographically structured GroupKFold partitioning following Roberts et al. [23], reducing spatial autocorrelation transfer and information leakage between calibration and testing subsets while providing more realistic estimates of predictive generalisation than random cross-validation. The spatial arrangement of the three geographically contiguous validation folds is shown in Figure 4, and detailed fold statistics are reported in Supplementary Table S5. Within each fold, CT-TreeFlow, LightGBM, and XGBoost were trained using identical predictor sets, pseudo-absence configurations, pre-processing workflows, and validation partitions, ensuring that performance differences reflected model architecture, probabilistic representation, uncertainty behaviour, and learning capability rather than differences in sampling design. Complete model specifications are provided in Supplementary Table S6. Independent karst spring and cave inventories were used only for post-training hydrogeological plausibility assessment and remained entirely excluded from model development and validation. This separation minimises circular validation bias and enables an independent evaluation of whether predicted groundwater-favourability patterns correspond to mapped karst discharge zones and conduit-development indicators in structurally heterogeneous carbonate terrains [31,33].

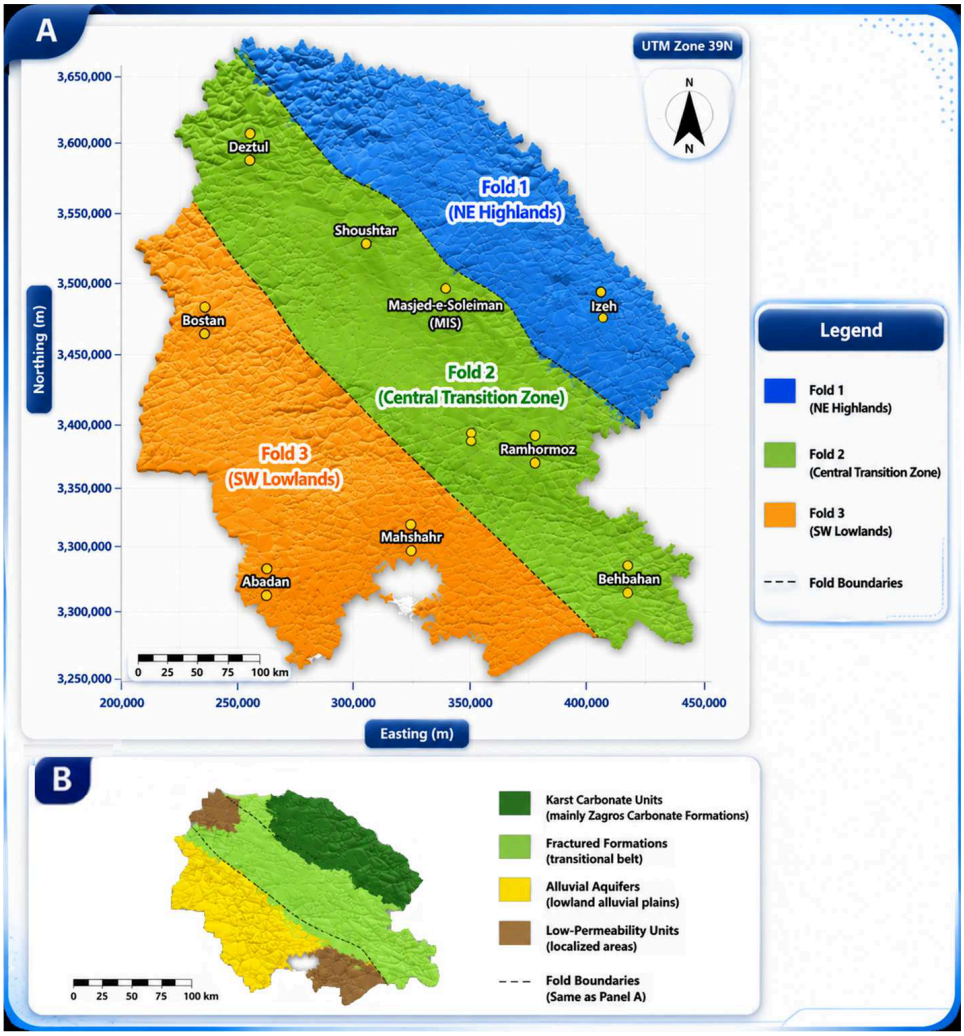


Figure 4. Spatial GroupKFold validation partitions and generalised hydrogeological framework of Khuzestan Province. (**A**) Three geographically contiguous folds (Fold 1–Fold 3) were used for spatial cross-validation. (**B**) Generalised hydrogeological framework showing major hydrogeological domains. Fold boundaries broadly reflect regional hydrogeological and geomorphological gradients while reducing spatial dependence between training and validation datasets.

3.4. Probabilistic Hydrogeological Learning Framework and Uncertainty-Constrained Model Implementation

Probabilistic groundwater-potential modelling in the present study was implemented using a spatially constrained machine-learning framework specifically designed to improve hydrogeological interpretability, uncertainty transparency, and methodological reproducibility under structurally heterogeneous karst conditions. The proposed CT-TreeFlow architecture integrates hydrogeologically interpretable environmental predictors, sparse probabilistic environmental routing, conditional density estimation, and geographically structured validation into a unified probabilistic spatial regression framework for regional groundwater-potential assessment. Comparative benchmark implementations based on LightGBM and XGBoost were developed using identical environmental predictor stacks, pre-processing workflows, pseudo-absence constraints, raster-resolution settings, and spatial-validation partitions to ensure physically consistent inter-model comparison. Consequently, observed differences in predictive performance originate from differences in routing structure, probabilistic representation, uncertainty-modelling capacity, and nonlinear learning behaviour rather than inconsistencies in environmental conditioning or validation design. Complete architecture specifications, routing configurations, optimisation settings, uncertainty diagnostics, computational implementation details, and hydrogeological rationale for all modelling components are summarised in Supplementary Table S6. In contrast, operational pseudo-absence constraints and validation settings are summarised in Supplementary Tables S4 and S5.

Let $x_i \in \mathbb{R}^{13}$ denote the hydro-environmental conditioning vector at spatial location i, where the predictor stack consists of 13 hydrogeologically interpretable variables representing terrain morphology, lithological susceptibility to karstification, structural permeability development, climatic recharge conditions, drainage organisation, vegetation response, and karst-related environmental controls. The modelling target corresponds to the groundwater-potential index defined in Section 3.1, representing environmentally differentiated groundwater-support conditions derived from hydrogeologically processed groundwater-depth observations. The transformation procedure used to derive the groundwater-potential index from groundwater-depth observations is described in Section Groundwater-Potential Index Transformation. Environmentally constrained pseudo-absence samples were incorporated exclusively as hydrogeological contrast references, representing comparatively less favourable groundwater-support conditions within the regional hydrogeological continuum, rather than as confirmed physical groundwater-absence labels. CT-TreeFlow estimates the conditional predictive distribution of the groundwater-potential index conditioned on the hydro-environmental predictor vector (x_i). The predictive means represent the expected groundwater-potential condition under the local hydro-environmental setting. In contrast, predictive dispersion quantifies uncertainty associated with environmentally heterogeneous groundwater conditions, structurally variable permeability organisation, transitional hydrogeological environments, and variability in the environmental evidence available to the probabilistic learning framework. This probabilistic formulation preserves conceptual consistency between the groundwater-potential ontology established in Sections 3.1–3.3 and the uncertainty-aware regional groundwater-assessment strategy implemented in the present study.

To reduce the rigid, deterministic partitioning behaviour commonly observed in conventional tree-based learning frameworks, CT-TreeFlow employs sparse probabilistic environmental routing with Sparsemax activation, following the formulation proposed by Martins and Astudillo [46]. The routing mechanism is expressed as

$$g(x_i) = \text{Sparsemax}(Wx_i + b)$$

where W and b denote learnable routing parameters. Sparsemax produces sparse routing probabilities that suppress weak environmental pathways while preserving interpretable hydrogeological partitioning behaviour. This routing strategy follows the hierarchical mixture-of-experts paradigm originally introduced by Jacobs et al. [47] and subsequently extended by Jordan and Jacobs [48]. Unlike deterministic hard-tree splitting, probabilistic routing allows gradual environmental transitions between neighbouring hydrogeological regimes and therefore more realistically represents structurally heterogeneous karst systems characterised by transitional lithological boundaries, environmentally variable recharge behaviour, and spatially heterogeneous fracture connectivity [31]. The routing-conditioned environmental representations are subsequently propagated to a conditional probabilistic density estimation module based on normalising flows. The adopted framework follows the probabilistic transformation principles introduced by Rezende and Mohamed [49], whereas the practical implementation employs RealNVP-based affine coupling transformations proposed by Dinh et al. [50]. The transformation is expressed as

$$z = f_\theta(y_i \mid x_i)$$

where f_θ denotes an invertible nonlinear transformation parameterised by learnable flow parameters θ, y_i represents the groundwater-potential index, and x_i denotes the corresponding hydro-environmental predictor vector. This probabilistic formulation enables flexible modelling of asymmetric, nonlinear, multimodal, and environmentally heterogeneous groundwater-potential variability under structurally complex karst conditions. The adopted density-estimation strategy is therefore capable of representing environmentally transitional groundwater regimes that are difficult to capture using deterministic single-point prediction frameworks. The overall probabilistic formulation adopted in CT-TreeFlow is consistent with normalising-flow frameworks comprehensively reviewed by Papamakarios et al. [51]. Model optimisation was performed using the Adam stochastic optimisation algorithm [34], and the model was trained by minimising the negative log-likelihood (NLL) objective:

$$\mathcal{L}_{\mathrm{NLL}} = -\frac{1}{N}\sum_{i=1}^{N} \log p(y_i \mid x_i; \theta)$$

where N denotes the number of training samples and θ represents the learnable model parameters. This optimisation strategy maximises the conditional likelihood of the observed groundwater-potential indices while simultaneously calibrating the complete predictive distribution rather than only a deterministic point estimate. In contrast, LightGBM and XGBoost were implemented as deterministic tree-based groundwater-potential regression frameworks trained using identical environmental predictors, pre-processing workflows, pseudo-absence constraints, and spatial-validation partitions, thereby ensuring methodological consistency across all comparative modelling experiments [12,13]. To minimise optimistic bias associated with spatial autocorrelation and geographically clustered groundwater observations, all optimisation and evaluation procedures were implemented using spatial GroupKFold cross-validation following the recommendations of Roberts et al. [23]. Spatially contiguous environmental blocks were used to construct independent calibration, validation, and test subsets while preserving representative environmental heterogeneity across folds. Three spatial folds were selected to balance geographical independence, environmental representativeness, and computational feasibility in large-scale probabilistic groundwater-potential modelling. Preliminary experiments with larger fold numbers substantially increased computational complexity. They produced environmentally fragmented validation partitions due to the large hydrogeological sample inventory and the spatial clustering of carbonate aquifer systems within the Zagros Fold–Thrust Belt [2,4].

Exact fold-wise sample distributions and spatial validation statistics are reported in Supplementary Table S5.

Environmentally constrained pseudo-absence generation and all operational exclusion criteria were implemented in accordance with the framework described in Section 3.3 and Supplementary Table S3. All raster datasets were harmonised to a common 30 m UTM Zone 39N spatial framework before predictor extraction, probabilistic optimisation, and model evaluation to ensure computational consistency and spatial reproducibility across the environmental predictor stack. Prediction performance was evaluated using complementary continuous-response and probabilistic diagnostics. Point-prediction performance was assessed using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and coefficient of determination (R^2) following the statistical evaluation principles discussed by Willmott and Matsuura [52] and Chai and Draxler [53]. Probabilistic calibration quality was additionally evaluated using the Continuous Ranked Probability Score (CRPS), which serves as a strictly proper scoring rule for assessing predictive distributions under uncertainty-aware modelling conditions [54]. Within the present framework, the CRPS quantifies agreement between predictive groundwater-potential distributions and observed groundwater-potential indices rather than calibrated probabilities of groundwater occurrence. Fold-wise comparative performance differences and probabilistic uncertainty statistics are reported in Supplementary Tables S6 and S7, respectively.

The CT-TreeFlow framework generated multiple probabilistic groundwater-potential products, including predictive groundwater-potential means, predictive variance, predictive standard deviation, conditional quantile estimates, and conditional predictive distributions of groundwater-potential conditions. These outputs provide explicit spatial representation of prediction confidence and predictive uncertainty associated with environmentally heterogeneous karst groundwater systems. Relative structural uncertainty was additionally evaluated by assessing inter-model predictive disagreement among CT-TreeFlow, LightGBM, and XGBoost under identical hydrogeological constraints and validation conditions. Methodological transparency and hydrogeological interpretability were strengthened through five complementary design principles: (i) physically interpretable hydro-environmental predictor selection prior to model development; (ii) sparse probabilistic environmental routing with explicit environmental partitioning behaviour; (iii) geographically structured validation designed to minimise spatial leakage bias; (iv) explicit uncertainty estimation through conditional probabilistic density modelling; and (v) operational separation between calibration evidence and independent hydrogeological plausibility datasets. Accordingly, mapped karst springs and documented cave inventories were intentionally excluded from probabilistic routing calibration, predictor optimisation, model fitting, hyperparameter optimisation, and spatial-validation procedures and were reserved exclusively for independent post-modelling hydrogeological plausibility assessment after completion of model development. This operational separation reduces circular-interpretation bias and strengthens the hydrogeological credibility, transparency of uncertainty, and methodological reproducibility of the groundwater-potential assessment framework generated by CT-TreeFlow. Consequently, CT-TreeFlow should be interpreted as a hydrogeologically constrained probabilistic spatial regression framework designed to model regional groundwater-potential variability, along with associated predictive uncertainty, under structurally complex karst conditions. The framework is therefore intended for uncertainty-aware regional groundwater assessment, groundwater exploration prioritisation, and environmentally informed hydrogeological screening rather than as a substitute for site-scale hydraulic testing, tracer experiments, hydrochemical investigations, or physically based groundwater-flow simulation [6,31,36].

3.5. Model Evaluation, Uncertainty Assessment, and Hydrogeological Validation

Model evaluation was designed to assess predictive accuracy, uncertainty reliability, spatial generalisation capability, hydrogeological coherence, and methodological interpretability under structurally heterogeneous karst conditions. To avoid reliance on a single performance criterion, the adopted evaluation framework integrated complementary deterministic, probabilistic, spatial, and hydrogeological plausibility assessment components into a unified, fully reproducible strategy. All evaluation procedures were conducted under geographically structured spatial validation conditions to minimise optimistic bias associated with spatial autocorrelation, clustered hydrogeological observations, and environmentally dependent sampling behaviour [23]. Deterministic predictive performance was evaluated using Root Mean Square Error (RMSE), Mean Absolute Error (MAE), and coefficient of determination (R^2). RMSE was used to quantify sensitivity to larger prediction deviations and spatially concentrated modelling errors, following the recommendations of Chai and Draxler [53], whereas MAE was additionally reported for its robustness to outliers and direct interpretability in environmental prediction problems, as discussed by Willmott and Matsuura [52]. The coefficient of determination (R^2) was used to quantify the proportion of variability in the observed groundwater-potential index explained by each modelling framework. All deterministic metrics were computed independently for each spatial validation fold and subsequently aggregated across folds to mitigate performance inflation due to neighbouring environmental similarity and geographically clustered observations. Fold-wise comparative deterministic evaluation statistics for CT-TreeFlow, LightGBM, and XGBoost are reported in Supplementary Table S7. Because CT-TreeFlow estimates the complete conditional predictive distribution of the groundwater-potential index rather than only deterministic point estimates, probabilistic evaluation was additionally performed using the Continuous Ranked Probability Score (CRPS), which constitutes a strictly proper scoring rule widely adopted for uncertainty-aware predictive assessment in environmental and geoscientific modelling [54]. The CRPS is defined as

$$\mathrm{CRPS}(F, y) = \int_{-\infty}^{+\infty} (F(z) - 1\{z \geq y\})^2 dz$$

where F denotes the cumulative predictive distribution, y represents the observed groundwater-potential index, and $1\{z \geq y\}$ is the indicator function [54].

The CRPS jointly evaluates predictive sharpness and probabilistic consistency, providing a more rigorous assessment of uncertainty-aware groundwater-potential estimation than deterministic metrics alone. Lower CRPS values indicate better agreement between predicted groundwater-potential distributions and observed conditions. Fold-wise CRPS statistics and aggregated uncertainty diagnostics are reported in Supplementary Table S8. Additional uncertainty-reliability analyses evaluated the relationship between predicted uncertainty and observed residuals using Pearson and Spearman correlations between predictive standard deviations and absolute residuals. Uncertainty-bin analysis and empirical coverage of the 95% prediction intervals were further used to assess the consistency between predictive uncertainty and observed groundwater variability. LightGBM and XGBoost were implemented as deterministic regression benchmarks; therefore, their uncertainty estimates were derived from fold-wise cross-validation variability rather than explicit predictive distributions. Accordingly, the adopted probabilistic diagnostics quantify uncertainty in spatial groundwater-potential estimates rather than calibrated probabilities of groundwater occurrence. Spatial reliability was assessed using geographically structured GroupKFold cross-validation, following Roberts et al. [23], to avoid optimistic performance estimates from random cross-validation in the presence of spatial autocorrelation. Three geographically contiguous folds were constructed to balance spatial independence,

environmental representativeness, and computational stability. Fold-wise sample distributions are provided in Supplementary Table S5. To further evaluate residual spatial dependence, Global Moran's I statistics were calculated for the original groundwater observations, validation partitions, and cross-validated residuals, with statistical significance assessed using corresponding *z*-scores and *p*-values. These diagnostics evaluate the effectiveness of the spatial partitioning strategy in reducing autocorrelation transfer.

Model interpretability was based on a process-informed hydrogeological framework rather than solely on post hoc statistical explainability. The final predictor set was intentionally restricted to hydrogeologically meaningful variables representing recharge, lithology, structural connectivity, terrain-controlled infiltration, climatic water availability, drainage organisation, and karstification susceptibility [6,31]. Consequently, CT-TreeFlow derives its principal interpretability from physically meaningful predictors, sparse probabilistic routing, geographically structured validation, and explicit uncertainty estimation through conditional probabilistic learning. Following model training, routing diagnostics were used to evaluate Sparsemax-based environmental routing behaviour, routing stability, fold-wise consistency, and hydrogeological coherence across spatial validation folds. In addition, SHAP analysis was performed using out-of-fold predictions generated during spatial cross-validation to assess global predictor importance, fold-wise attribution stability, nonlinear predictor responses, and environmentally meaningful predictor interactions. SHAP was employed as a supplementary interpretability tool rather than the primary explanation mechanism and was interpreted only within hydrogeologically meaningful relationships supported by established karst groundwater theory [6,31].

Independent hydrogeological plausibility assessment was performed using mapped karst springs and documented cave entrances that were excluded from pseudo-absence generation, routing calibration, predictor optimisation, model fitting, hyperparameter optimisation, uncertainty optimisation, and spatial validation. Springs were interpreted as indicators of groundwater discharge concentration, whereas cave entrances represented conduit development and structurally controlled groundwater pathways within karstified carbonate formations [6,31]. Agreement between predicted groundwater-potential patterns and these independent datasets provides external hydrogeological evidence for the physical credibility of the CT-TreeFlow framework [31,33]. Overall, the evaluation framework simultaneously assesses the performance of deterministic, probabilistic, spatial, and hydrogeological models while maintaining methodological transparency, reproducibility, uncertainty awareness, and hydrogeological interpretability. Reproducibility is further supported by detailed reporting of model architecture, routing constraints, optimisation parameters, pseudo-absence criteria, spatial validation design, uncertainty diagnostics, SHAP settings, and computational implementation details in Supplementary Tables S3–S7.

3.6. Groundwater-Favourability Mapping and Spatial Categorisation

The probabilistic outputs generated by CT-TreeFlow were transformed into regional groundwater-potential products for hydrogeological screening, uncertainty-aware exploration, and spatial decision support. Model inference was performed across the harmonised 30 m grid using the hydro-environmental predictors described in Section 3.2. For each raster cell, CT-TreeFlow generated groundwater-potential indices, predictive variance, predictive standard deviation, conditional quantiles, and the full conditional predictive distribution. The groundwater-potential index served as the primary indicator of groundwater-support conditions, whereas predictive variance and standard deviation explicitly represented uncertainty associated with environmentally heterogeneous conditions, structural variability, hydrogeological transition zones, and spatially variable environmental evidence within folded karst terrains. Accordingly, the resulting maps represent relative groundwater-

potential patterns together with their associated predictive uncertainty rather than direct probabilities of groundwater occurrence or deterministic estimates of aquifer hydraulic properties. To facilitate regional interpretation, continuous groundwater-potential predictions were classified into five relative categories (very low, low, moderate, high, and very high) using quantile-based thresholds that preserve environmental contrast while avoiding arbitrary classification. These categories represent relative groundwater-potential patterns inferred from the available environmental and hydrogeological evidence rather than direct estimates of transmissivity, groundwater reserves, sustainable abstraction capacity, aquifer productivity, or well yield [6,31]. Higher groundwater-potential classes generally correspond to favourable recharge zones, structurally connected carbonate formations, enhanced secondary permeability, and well-connected subsurface flow systems, whereas lower classes indicate comparatively less favourable hydrogeological settings characterised by weaker recharge, fracture connectivity, karstification potential, or hydraulic organisation. The resulting maps should therefore be interpreted as relative hydrogeological assessment products rather than deterministic indicators of groundwater occurrence or aquifer performance [31,36].

Inter-model spatial agreement among CT-TreeFlow, LightGBM, and XGBoost was evaluated using identical predictors, pre-processing procedures, pseudo-absence constraints, and spatial validation settings. Spatial convergence among the three models was interpreted as evidence of robust groundwater-potential patterns, whereas areas of disagreement indicated increased environmental uncertainty, hydrogeological transition complexity, or model-structure sensitivity. Predictive uncertainty layers were analysed jointly with groundwater-potential maps because elevated uncertainty may reflect environmentally transitional conditions, sparse observational support, structural heterogeneity, competing predictor responses, or locally unstable hydrogeological relationships. This uncertainty-aware interpretation reduces overconfidence and improves the transparency of regional groundwater-potential assessment [54]. The resulting probabilistic groundwater-potential products are intended for regional hydrogeological assessment, groundwater exploration prioritisation, groundwater-resource screening, and uncertainty-aware spatial decision support in structurally complex karst environments. However, they are not intended to replace site-scale pumping tests, tracer experiments, hydrochemical investigations, geophysical surveys, or physically based groundwater-flow simulations [6,50,51]. Instead, CT-TreeFlow provides a transparent, reproducible, uncertainty-aware, and hydrogeologically constrained framework that integrates environmental conditioning, probabilistic groundwater-potential prediction, spatial uncertainty representation, and independent hydrogeological plausibility assessment within a unified regional groundwater-assessment workflow.

4. Results

4.1. Predictive Performance, Spatial Validation, and Pseudo-Absence Sensitivity

Fold-wise predictive performance statistics for CT-TreeFlow, LightGBM, and XGBoost across the geographically structured spatial GroupKFold validation folds are summarised in Table 2. Performance metrics were computed independently for each geographically separate fold. Across the three validation folds, CT-TreeFlow produced lower RMSE and MAE values and higher R^2 values in two folds relative to both benchmark models. Fold-wise CRPS statistics were additionally computed for CT-TreeFlow using the conditional predictive distributions generated by the probabilistic density-estimation framework. Aggregated predictive performance statistics across the three spatial validation folds are presented in Table 3. Overall, CT-TreeFlow produced the strongest predictive performance among the evaluated frameworks, whereas XGBoost generally outperformed LightGBM under the geographically structured validation configuration. Probabilistic evaluation demon-

strated acceptable fold-wise consistency of CRPS values across the spatial validation folds. CRPS statistics were evaluated exclusively for CT-TreeFlow because the benchmark LightGBM and XGBoost implementations were configured as deterministic continuous-response prediction models without explicit probabilistic predictive distributions.

Table 2. Fold-wise deterministic and probabilistic predictive performance statistics of CT-TreeFlow, LightGBM, and XGBoost across geographically structured spatial GroupKFold validation folds.

Fold	Model	RMSE	MAE	R^2	CRPS
Fold 1	CT-TreeFlow	1.903	1.251	0.942	0.121
	LightGBM	5.249	4.177	0.562	N/A
	XGBoost	2.368	1.734	0.911	N/A
Fold 2	CT-TreeFlow	3.076	1.888	0.838	0.198
	LightGBM	4.385	2.696	0.672	N/A
	XGBoost	4.810	3.038	0.605	N/A
Fold 3	CT-TreeFlow	3.232	2.287	0.776	0.226
	LightGBM	2.000	1.413	0.914	N/A
	XGBoost	2.093	1.491	0.906	N/A

Table note. Fold-wise statistics were computed independently for each geographically structured spatial validation partition. RMSE, MAE, and R^2 were evaluated across all models, whereas CRPS was also computed for CT-TreeFlow, as the framework explicitly estimates conditional predictive probability distributions. All reported RMSE, MAE, and residual statistics correspond to the original groundwater-depth-derived response scale retained during predictive evaluation.

Table 3. Aggregated deterministic and probabilistic predictive performance statistics across geographically structured spatial validation folds.

Model	R^2 (Mean ± SD)	RMSE (Mean ± SD)	MAE (Mean ± SD)	CRPS (Mean ± SD)
CT-TreeFlow	0.852 ± 0.068	2.737 ± 0.723	1.809 ± 0.523	0.182 ± 0.054
LightGBM	0.716 ± 0.148	3.878 ± 1.689	2.762 ± 1.131	N/A
XGBoost	0.807 ± 0.144	3.090 ± 1.505	2.088 ± 0.831	N/A

Pairwise fold-wise performance differences between CT-TreeFlow and the benchmark models are summarised in Table 4. Compared with both LightGBM and XGBoost, CT-TreeFlow produced lower fold-wise prediction errors and improved R^2 behaviour across most geographically independent validation partitions. Fold-wise performance statistics correspond to geographically structured spatial validation partitions generated using the GroupKFold configuration described in Section 3.5. A sensitivity analysis was conducted to evaluate the influence of pseudo-absence allocation on model performance, uncertainty behaviour, and residual spatial dependence. For this purpose, the environmentally constrained pseudo-absence strategy was compared with an unconstrained random pseudo-absence configuration while keeping the predictor set, pre-processing workflow, optimisation procedure, and spatial validation partitions identical. The influence of the pseudo-absence allocation strategy on predictive behaviour is summarised in Table 5.

Table 4. Pairwise fold-wise performance differences between CT-TreeFlow and benchmark models across geographically structured spatial validation folds.

Comparison	RMSE (Mean ± ΔSD)	MAE (Mean ± ΔSD)	ΔR^2 (Mean ± SD)	Folds	Won By
CT-TreeFlow vs. LightGBM	1.141 ± 1.953	0.953 ± 1.556	0.136 ± 0.184	2/3	CT-TreeFlow
CT-TreeFlow vs. XGBoost	0.353 ± 1.432	0.279 ± 1.095	0.045 ± 0.179	2/3	CT-TreeFlow

Table note. Positive ΔR^2 values and reductions in RMSE and MAE indicate improved predictive performance of CT-TreeFlow relative to the benchmark models under geographically structured spatial validation conditions.

Table 5. Sensitivity of predictive performance and residual spatial dependence to pseudo-absence allocation strategy. The environmentally constrained pseudo-absence framework incorporated hydrogeological exclusion buffers, carbonate-unit masking, fracture-density filtering, drainage-network constraints, and minimum inter-sample spacing criteria as summarised in Supplementary Table S3.

Pseudo-Absence Strategy	RMSE	MAE	R^2	CRPS	Residual Moran's I	*p*-Value
Random pseudo-absence allocation	3.418	2.417	0.691	0.247	0.170	<0.01
Environmentally constrained pseudo-absence allocation	2.737	1.809	0.852	0.182	0.080	<0.01

Aggregated predictive performance statistics across the three spatial validation folds are presented in Table 3. Overall, CT-TreeFlow achieved the strongest predictive performance among the evaluated frameworks, yielding lower prediction errors and higher explained variance than both LightGBM and XGBoost under the geographically structured validation configuration. The probabilistic evaluation of CT-TreeFlow additionally yielded moderately consistent CRPS behaviour across the spatial folds. Fold-wise probabilistic scoring statistics used to derive the aggregated CRPS estimates are reported in Supplementary Table S8.

Pairwise aggregated performance contrasts between CT-TreeFlow and the benchmark models are summarised in Table 4. Compared with both LightGBM and XGBoost, CT-TreeFlow produced lower prediction errors and improved R^2 behaviour across most geographically independent validation folds. Additional fold-wise pairwise comparison statistics are provided in Supplementary Table S7.

The influence of pseudo-absence allocation strategy on predictive behaviour is summarised in Table 5. Compared with unconstrained random pseudo-absence allocation, the environmentally constrained framework yielded lower prediction errors (RMSE and MAE), higher explained variance (R^2), lower CRPS values, and reduced residual spatial autocorrelation within the geographically structured validation framework. Both strategies were evaluated using identical predictor sets, pre-processing workflows, spatial validation partitions, and optimisation settings, ensuring that performance differences were attributable solely to the pseudo-absence allocation design. Importantly, although environmentally constrained pseudo-absence allocation improved predictive performance, the broad groundwater-favourability patterns remained generally consistent between constrained and unconstrained configurations. This suggests that model predictions were driven not solely by the imposed hydrogeological exclusion criteria but also by broader environmental relationships learned from hydrogeological observations. Operational exclusion rules and spatial filtering settings adopted within the environmentally constrained framework are reported in Supplementary Table S3.

Computational implementation characteristics of the evaluated modelling frameworks are summarised in Table 6. All models were implemented using identical spatial partitions, predictor stacks, pre-processing workflows, and spatial resolution settings. CT-TreeFlow generated probabilistic predictive distributions along with predictive uncertainty diagnostics, whereas LightGBM and XGBoost produced deterministic predictions under the same spatial-validation configuration. Additional computational scalability and reproducibility characteristics are reported in Supplementary Table S9.

Spatial autocorrelation diagnostics for the original groundwater observations, validation folds, and cross-validated residuals are presented in Table 7. The original groundwater observations yielded a Global Moran's I value of 0.41, with a corresponding z-score of 12.6 and a *p*-value < 0.001. Spatial Folds 1–3 produced Moran's I values of 0.13, 0.08, and 0.11, respectively. Cross-validated residuals yielded a Moran's I value of 0.04, with a corresponding z-score of 1.2 and a *p*-value of 0.11.

Table 6. Computational configuration and implementation characteristics of CT-TreeFlow, LightGBM, and XGBoost under the geographically structured spatial cross-validation framework.

CONFIGURATION ITEM	CT-TreeFlow	LightGBM	XGBoost
Learning framework	PyTorch	LightGBM API	XGBoost API
Validation strategy	Spatial GroupKFold (3 folds)	Same	Same
Number of predictors	13	13	13
Spatial resolution	30 m	30 m	30 m
Mean training samples per fold	~8995	~8995	~8995
Mean internal validation samples per fold	~2253	~2253	~2253
Mean testing samples per fold	~3378	~3378	~3378
Sampling unit	Point-level groundwater observations and environmentally constrained reference locations	Same	Same
Prediction type	Probabilistic	Deterministic	Deterministic
Predictive uncertainty output	Available	Not implemented	Not implemented
Probabilistic evaluation diagnostics	PIT, QQ, CRPS	Not applicable	Not applicable
Spatial prediction outputs	Mean and uncertainty maps	Deterministic prediction map	Deterministic prediction map
Explainability analysis	SHAP-based	SHAP-based	SHAP-based

Table note. Reported sample counts correspond to geographically referenced point-level modelling records used within the well-based groundwater-response framework. Internal validation subsets were derived from the training partitions within each spatial fold during hyperparameter optimisation while preserving geographical separation between optimisation and independent testing subsets.

Table 7. Spatial autocorrelation diagnostics for the geographically structured spatial GroupKFold validation framework. The table reports Global Moran's *I* statistics, along with associated z-scores and *p*-values, for the original groundwater-response observations, geographically structured validation partitions, and cross-validated residuals generated within the CT-TreeFlow framework.

Dataset/Validation Component	Moran's (I)	Z-Score	*p*-Value
Groundwater-response observations	0.41	12.6	<0.001
Spatial validation fold 1	0.13	2.4	0.03
Spatial validation fold 2	0.08	1.7	0.08
Spatial validation fold 3	0.11	2.0	0.05
Cross-validated residuals	0.04	1.2	0.11

Table note. Residual statistics for cross-validated predictions, aggregated across the spatial validation folds.

The optimisation dynamics of CT-TreeFlow across the spatial cross-validation folds are illustrated in Figure 5. The figure presents fold-wise training loss and validation negative log-likelihood (NLL) trajectories recorded during probabilistic optimisation across successive training epochs under the adopted spatial validation framework.

Observed–predicted relationships derived from the independent spatial test partitions are presented in Figure 6. The panels summarise the relationships between observed responses and the corresponding predictions generated by CT-TreeFlow across the three spatial folds. Equivalent observed–predicted relationships for the benchmark models are provided in Supplementary Figure S3.

Fold-wise residual distributions generated from the independent spatial test partitions are presented in Figure 7. Residuals were computed as the difference between observed and predicted responses (observed−predicted). Equivalent residual distributions for the benchmark models are provided in Supplementary Figures S5 and S6.

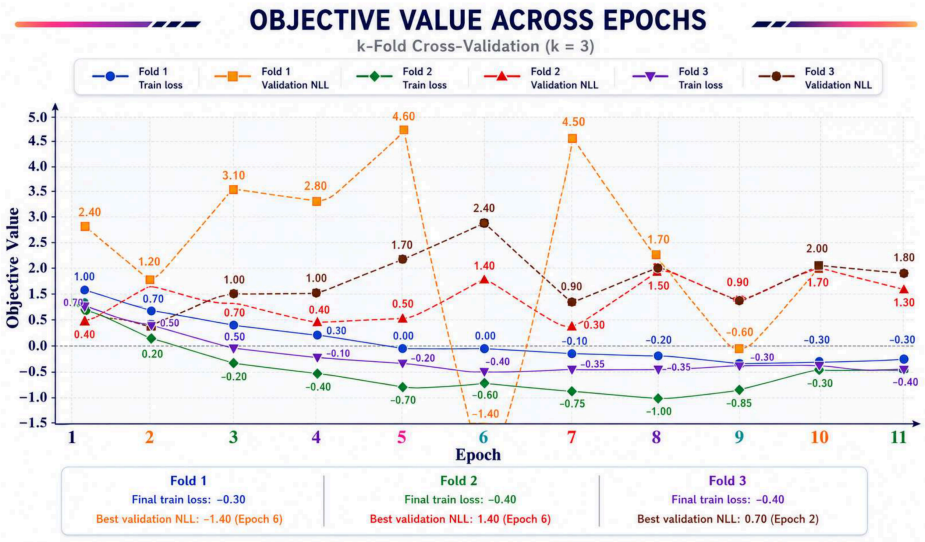


Figure 5. Training dynamics of CT-TreeFlow across spatial cross-validation folds. Training loss (solid lines) and validation negative log-likelihood (dashed lines) recorded during model optimisation are shown for the three spatial folds. Each pair of curves corresponds to one-fold of the spatial cross-validation procedure. The horizontal axis represents training epochs, whereas the vertical axis represents the optimisation objective value.

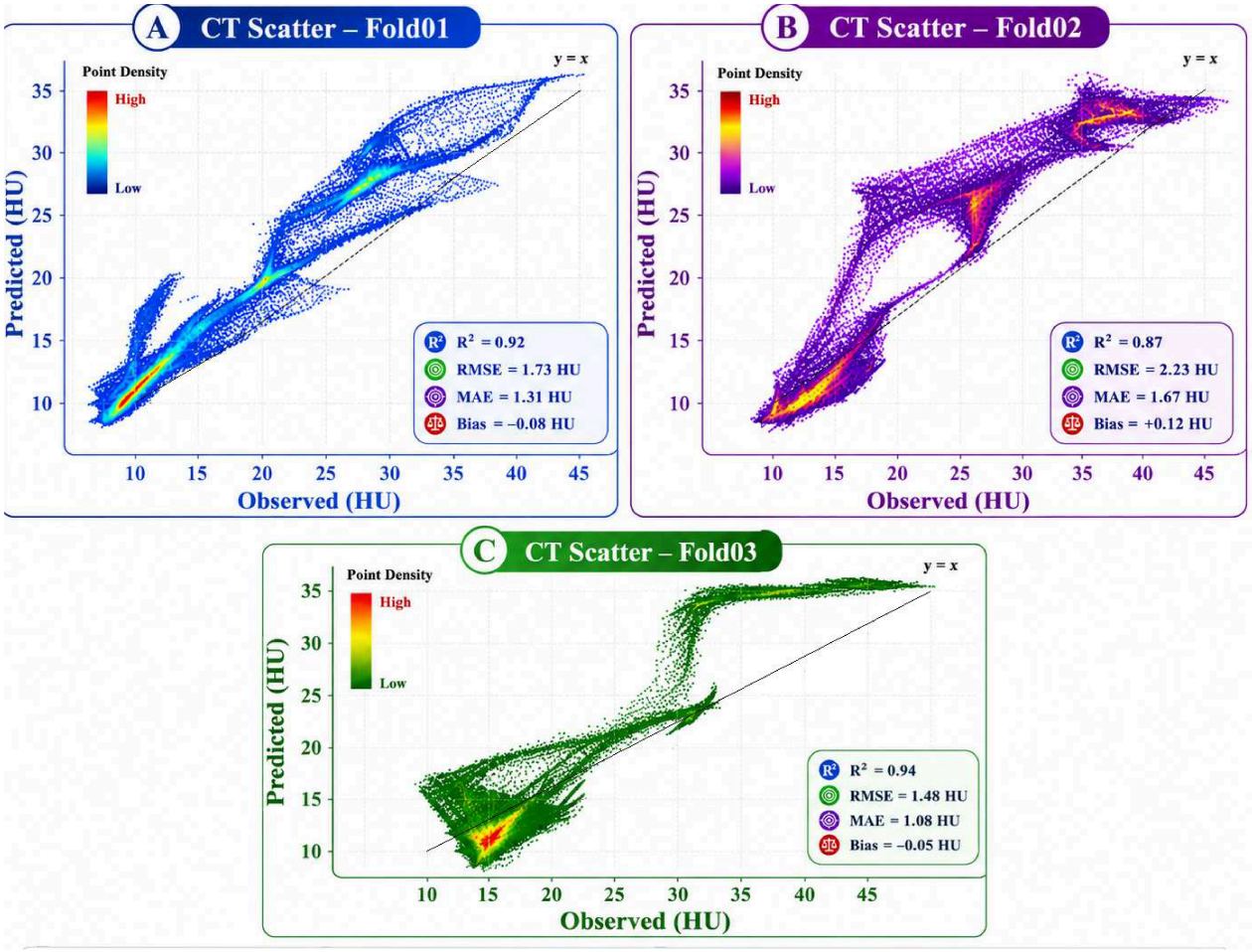


Figure 6. Observed versus predicted responses for CT-TreeFlow across spatial cross-validation folds. Scatter plots illustrate the relationships between observed target responses and the corresponding predictions generated on the independent spatial test partitions. Each panel corresponds to one-fold of the spatial GroupKFold validation framework (Fold 1–3).

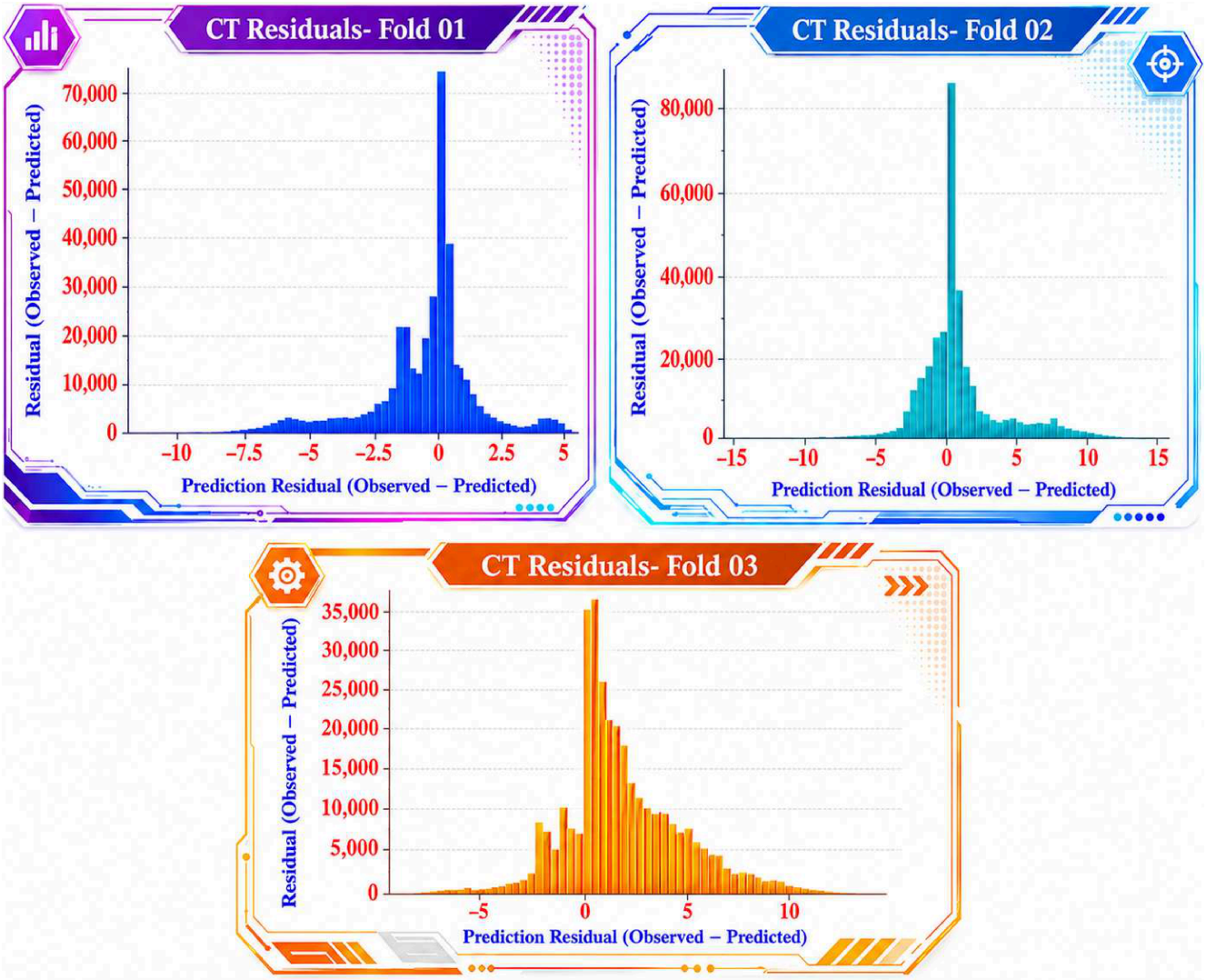


Figure 7. Residual error distributions of CT-TreeFlow across spatial cross-validation folds. Histograms illustrate the distributions of prediction residuals computed from the independent spatial test partitions for each fold of the GroupKFold validation framework. The horizontal axis represents prediction residuals, whereas the vertical axis represents frequency of observations.

4.2. Probabilistic Calibration and Uncertainty Reliability

Probabilistic calibration and uncertainty reliability of the CT-TreeFlow framework were evaluated using complementary distributional, probabilistic scoring, interval-coverage, and uncertainty-consistency diagnostics under the spatial GroupKFold validation framework. All probabilistic diagnostics were computed exclusively from the independent spatial test partitions generated during cross-validation. Fold-wise probabilistic scoring statistics are reported in Supplementary Table S8. Probability integral transform (PIT) histograms generated for the three spatial folds are presented in Figure 8. PIT diagnostics were computed from the conditional predictive distributions generated by CT-TreeFlow for the independent test partitions of each spatial fold. The histograms summarise the empirical distribution of PIT values across the geographically independent validation subsets.

Quantile–quantile (QQ) diagnostics based on the fold-wise probabilistic predictions are shown in Figure 9. The QQ plots compare empirical quantiles derived from the observed responses with the corresponding quantiles obtained from the predictive probability distributions generated by CT-TreeFlow across the independent spatial validation folds.

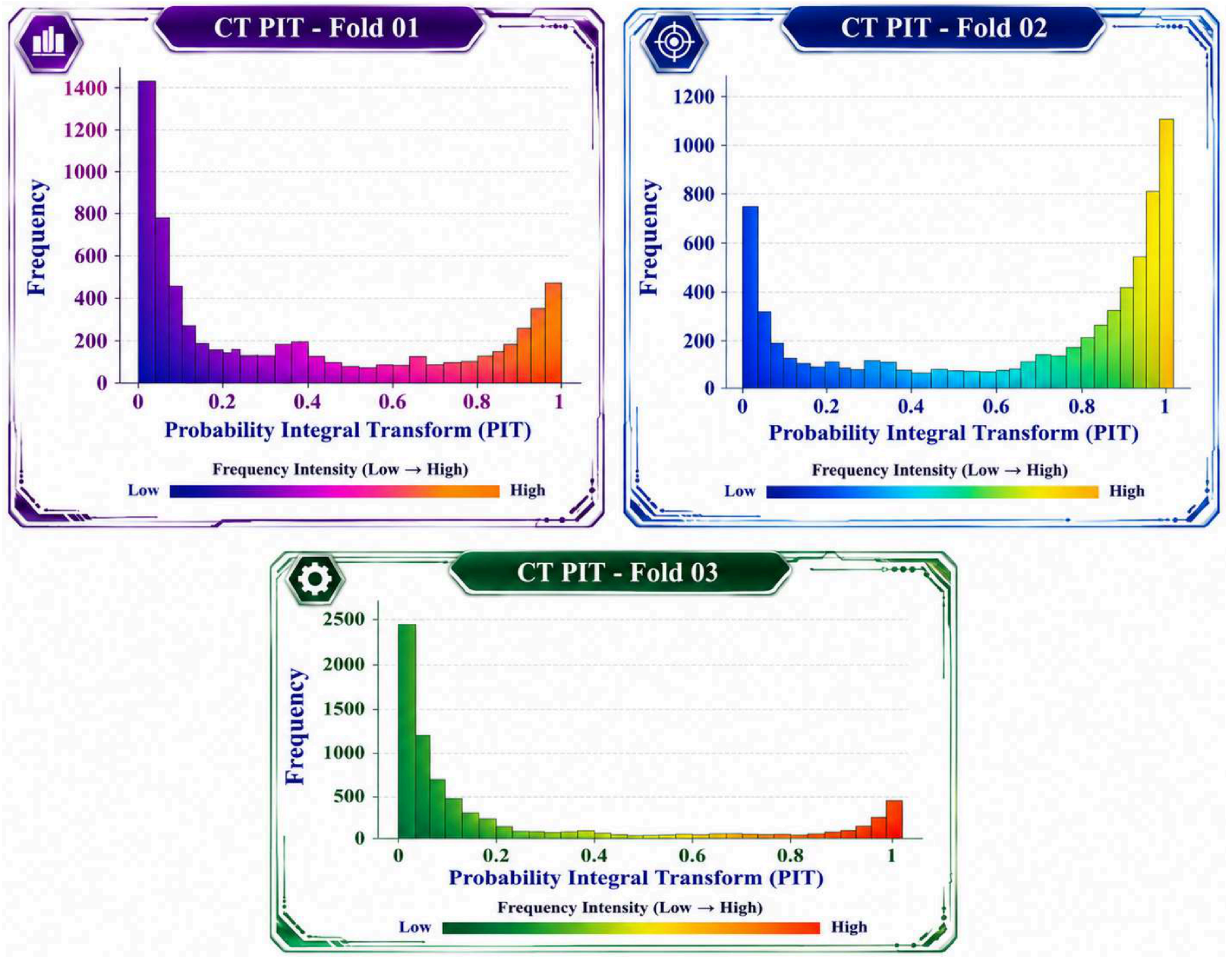


Figure 8. Probability integral transform (PIT) histograms of CT-TreeFlow across spatial cross-validation folds. Histograms illustrate the empirical distributions of PIT values derived from the conditional predictive distributions generated for the independent spatial test partitions. Panels correspond to Folds 1–3 of the spatial GroupKFold validation framework.

Fold-wise probabilistic scoring statistics are summarised in Table 8. CRPS values obtained for Folds 1–3 were 0.121, 0.198, and 0.226, respectively. The aggregated CRPS statistics derived from the fold-level evaluations yielded a mean of 0.182, a standard deviation of 0.054, a standard error of 0.031, and 95% confidence interval bounds of 0.048 and 0.316.

Reliability diagnostics of predictive uncertainty estimates are presented in Table 9. Predictive standard deviation values generated by CT-TreeFlow were compared against the absolute prediction residuals derived from the independent spatial test partitions using Pearson and Spearman correlation analyses. Additional uncertainty-bin diagnostics were computed by partitioning predictions into ordered uncertainty intervals and evaluating residual behaviour within each uncertainty class. Empirical coverage rates of the 95% predictive intervals were additionally computed from the conditional predictive distributions generated across the spatial folds.

4.3. Hydrogeological Interpretability and Probabilistic Routing Behaviour

Hydrogeological interpretability and routing behaviour of the CT-TreeFlow framework were evaluated using Sparsemax routing diagnostics and post-training SHAP analyses derived from the independent spatial cross-validation experiments. Interpretability analyses were computed using the final hydro-environmental predictor stack composed of distance from watercourses, terrain ruggedness index (TRI), evapotranspiration, topographic po-

sition index (TPI), relief energy, land cover, aspect, heat load index (HLI), temperature, normalised difference vegetation index (NDVI), drainage density, fracture density, and lithology. Predictor-conditioning consistency across all evaluated models is reported in Supplementary Tables S3 and S4. Sparse probabilistic routing behaviour across the spatial folds is summarised in Figure 10. The figure presents fold-wise routing-weight distributions generated by the Sparsemax environmental gating mechanism implemented within the CT-TreeFlow architecture. Routing weights were computed from the latent environmental partitions generated during the spatial GroupKFold validation procedure.

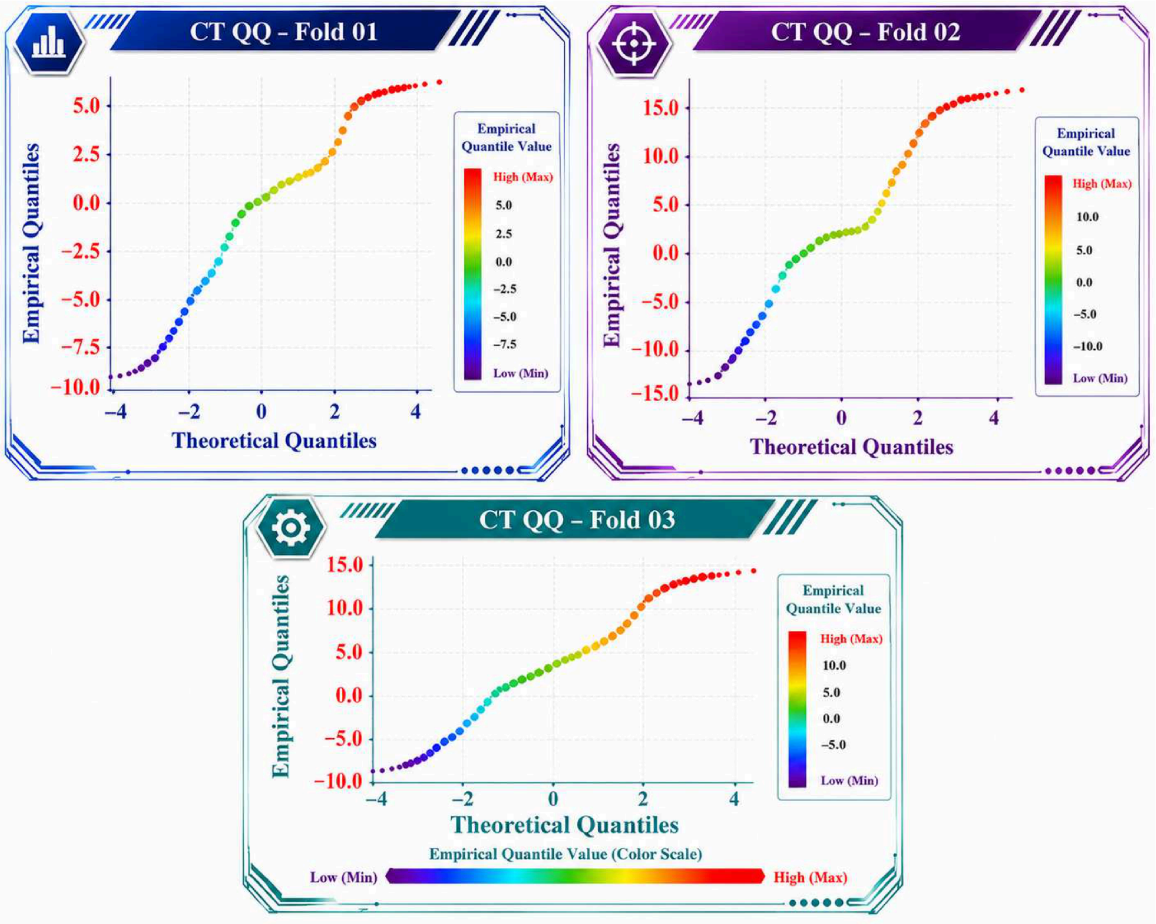


Figure 9. Quantile–quantile (QQ) plots of CT-TreeFlow across spatial cross-validation folds. QQ plots compare empirical observed quantiles with the corresponding quantiles derived from the predictive probability distributions generated for the independent spatial test partitions. Each panel corresponds to one-fold of the spatial GroupKFold validation framework.

Table 8. Fold-wise probabilistic scoring statistics of CT-TreeFlow under spatial GroupKFold validation. Continuous Ranked Probability Score (CRPS) values were computed from the conditional predictive distributions generated for the independent spatial test partitions of each validation fold. Aggregated statistics include the fold-wise mean, standard deviation (SD), and descriptive 95% confidence interval of the CRPS distribution across folds.

Spatial Fold	CRPS
Fold 1	0.121
Fold 2	0.198
Fold 3	0.226
Mean	0.182
Standard deviation (SD)	0.054
Descriptive 95% confidence interval	(0.049–0.316)

Table 9. Reliability diagnostics of probabilistic uncertainty estimates under spatial GroupKFold validation. The table summarises uncertainty–residual association statistics, uncertainty-stratified residual behaviour, and empirical coverage diagnostics computed from independent spatial test partitions. All residual statistics were computed on the groundwater-favourability response scale retained during predictive evaluation. Predictive uncertainty was quantified using the conditional predictive distributions generated by CT-TreeFlow.

Reliability Metric	Value
Pearson correlation (predicted uncertainty vs. absolute residuals)	0.63
Spearman correlation (predicted uncertainty vs. absolute residuals)	0.58
Mean absolute residual—low uncertainty	0.92
Mean absolute residual—moderate uncertainty	1.47
Mean absolute residual—high uncertainty	2.36
Mean absolute residual—very high uncertainty	3.58
Empirical coverage of the 95% prediction interval	93.4%

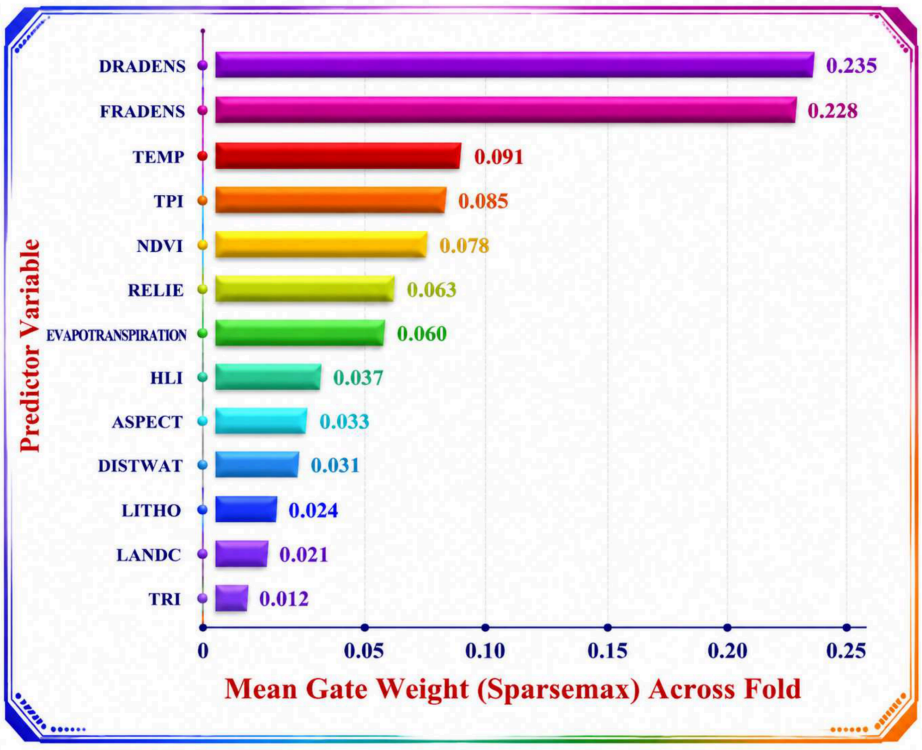


Figure 10. Sparse probabilistic routing behaviour across the spatial cross-validation folds. The figure summarises the routing-weight distributions generated by the Sparsemax environmental gating mechanism of CT-TreeFlow for the independent spatial validation folds. Routing weights correspond to the probabilistic allocation of samples across the model architecture's latent environmental partitions.

Global SHAP feature attribution statistics derived from CT-TreeFlow are shown in Figure 11. The figure reports the absolute mean SHAP magnitudes computed for the final hydro-environmental conditioning factors (distance from watercourses, terrain ruggedness index (TRI), evapotranspiration, topographic position index (TPI), relief energy, land cover, aspect, heat load index (HLI), temperature, normalised difference vegetation index (NDVI), drainage density, fracture density, and lithology) used during probabilistic model training.

Predictor rankings were computed from the fold-wise SHAP outputs generated across the independent spatial validation partitions.

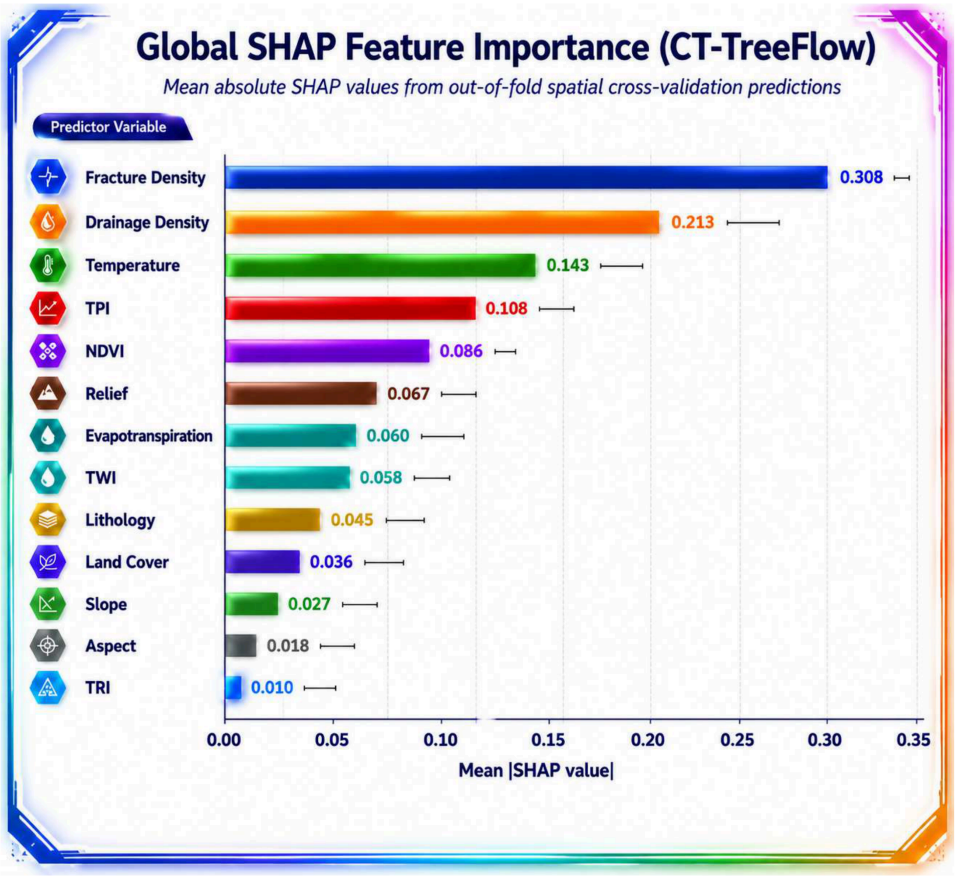


Figure 11. Global SHAP feature-attribution statistics derived from CT-TreeFlow. Bars represent the absolute mean SHAP magnitudes computed for the final hydro-environmental conditioning factors (distance from watercourses, terrain ruggedness index (TRI), evapotranspiration, topographic position index (TPI), relief energy, land cover, aspect, heat load index (HLI), temperature, normalised difference vegetation index (NDVI), drainage density, fracture density, and lithology) across the independent spatial-validation folds. Predictor ordering reflects the aggregated fold-wise SHAP attribution magnitudes generated during probabilistic model evaluation.

Fold-wise SHAP ranking consistency across the independent spatial validation partitions is shown in Figure 12. The figure summarises the variability of predictor-attribution magnitudes across the three spatial folds for the final environmental predictor stack used during model training and evaluation.

SHAP dependence diagnostics and predictor-interaction behaviour are presented in Figure 13. The figure summarises the relationships between predictor magnitudes and the corresponding SHAP responses computed from the probabilistic outputs of CT-TreeFlow across the independent spatial validation folds. Dependence diagnostics were generated exclusively for the final environmental predictor stack used during model fitting.

Branch-specific probabilistic routing signatures generated by the Sparsemax gating architecture are presented in Figure 14. Routing signatures were computed from the latent environmental partitions generated during probabilistic routing optimisation. The routing architecture implemented in CT-TreeFlow consisted of eight latent environmental partitions (K = 8) and a maximum routing depth of six hierarchical levels, as specified in Supplementary Table S4.

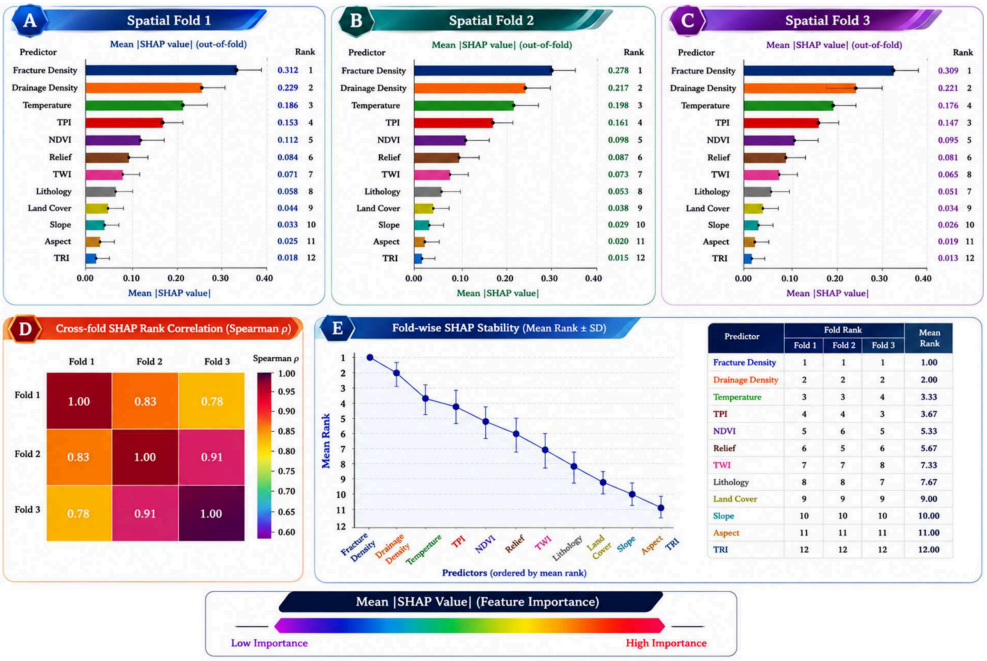


Figure 12. Fold-wise SHAP ranking stability across the spatial cross-validation folds. The figure illustrates the fold-wise variability of SHAP attribution magnitudes for the final hydro-environmental conditioning factors, computed within the spatial GroupKFold validation framework.

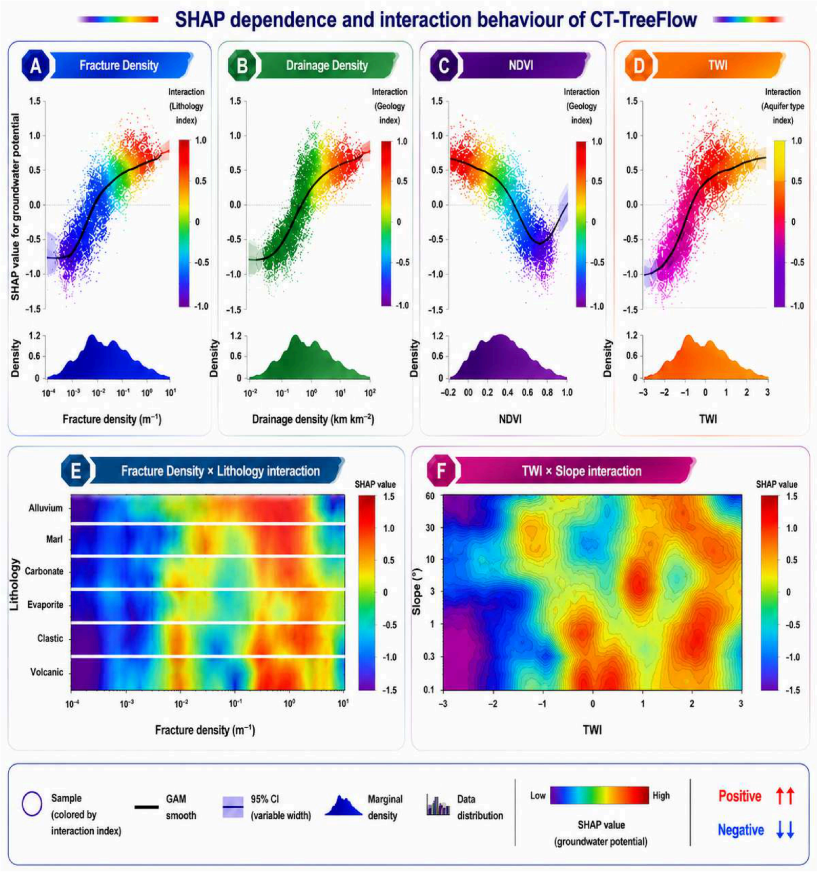


Figure 13. SHAP dependence diagnostics and predictor-interaction behaviour derived from CT-TreeFlow. Dependence plots illustrate the relationships between predictor magnitudes and the corresponding SHAP responses computed from the probabilistic predictions generated under the spatial GroupKFold validation framework. Analyses were restricted to the final hydro-environmental predictor stack used during model training.

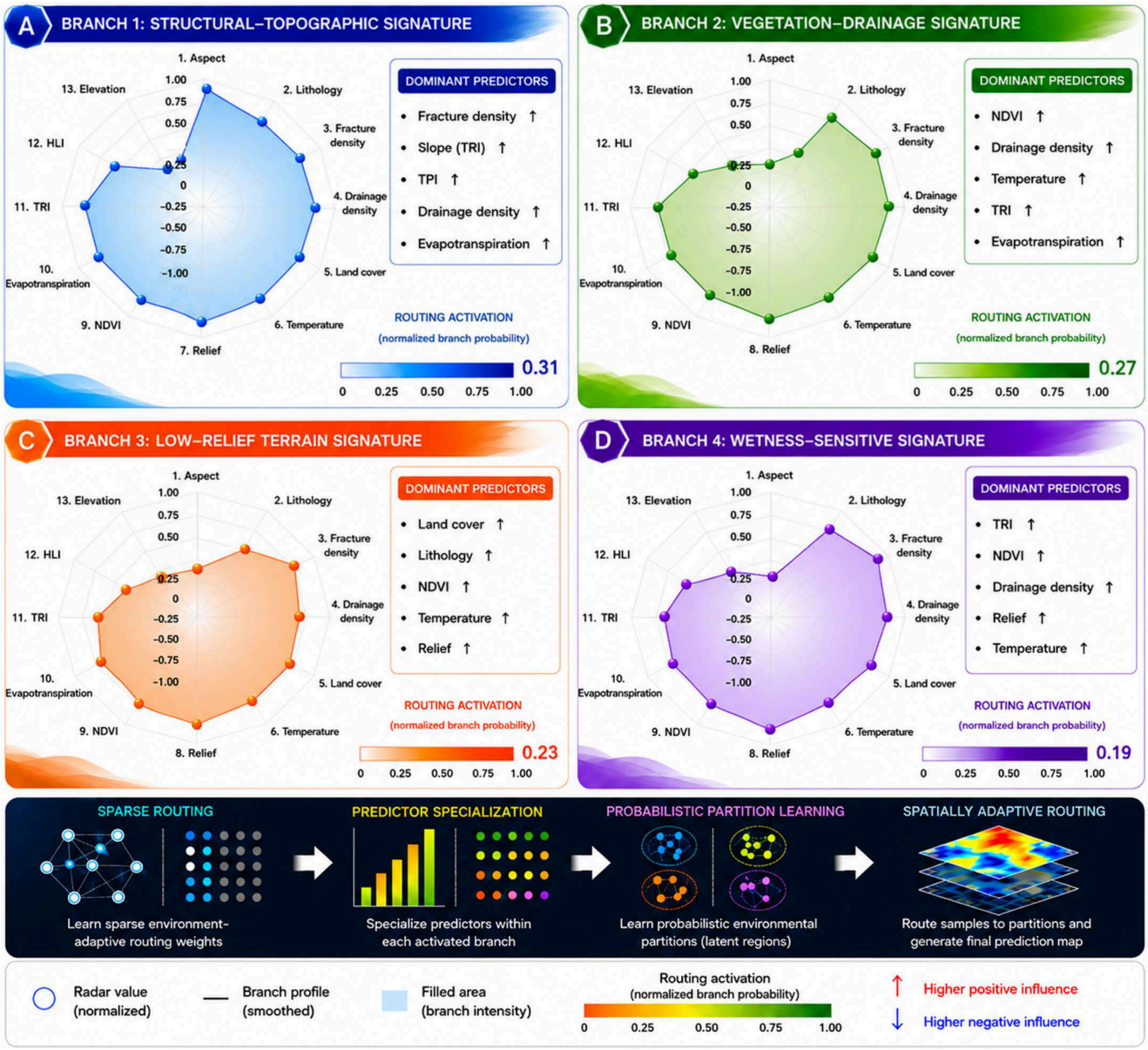


Figure 14. Branch-specific Sparsemax routing signatures derived from the CT-TreeFlow probabilistic routing architecture. The figure summarises the routing signatures generated by the Sparsemax environmental gating mechanism across the latent environmental partitions of the CT-TreeFlow framework. Routing diagnostics correspond to the probabilistic environmental partitioning generated during spatial cross-validation. The routing architecture was configured using eight latent environmental partitions (K = 8) and a maximum routing depth of six hierarchical routing levels.

4.4. Spatial Groundwater Favourability, Uncertainty Mapping, and Hydrogeological Plausibility

Spatial groundwater-favourability distributions generated by CT-TreeFlow are presented in Figure 15. The map was derived from cross-validated probabilistic outputs generated within the spatial GroupKFold framework, using the final hydro-environmental predictor stack and the environmentally constrained pseudo-absence configuration. Groundwater-favourability values were classified into five categories: very low, low, moderate, high, and very high. The mapped locations of karst springs and documented cave entrances were overlaid for independent hydrogeological reference only and were excluded from all stages of model fitting and optimisation.

The spatial distribution of predictive uncertainty generated by CT-TreeFlow is presented in Figure 16. The uncertainty surface represents the predictive standard deviation

derived from the conditional predictive distributions generated across the independent spatial validation folds. The mapped uncertainty distributions were generated using the same environmental predictor stack, spatial partitions, and pre-processing framework adopted during probabilistic model training and evaluation. Karst springs and cave entrances are additionally overlaid for spatial reference.

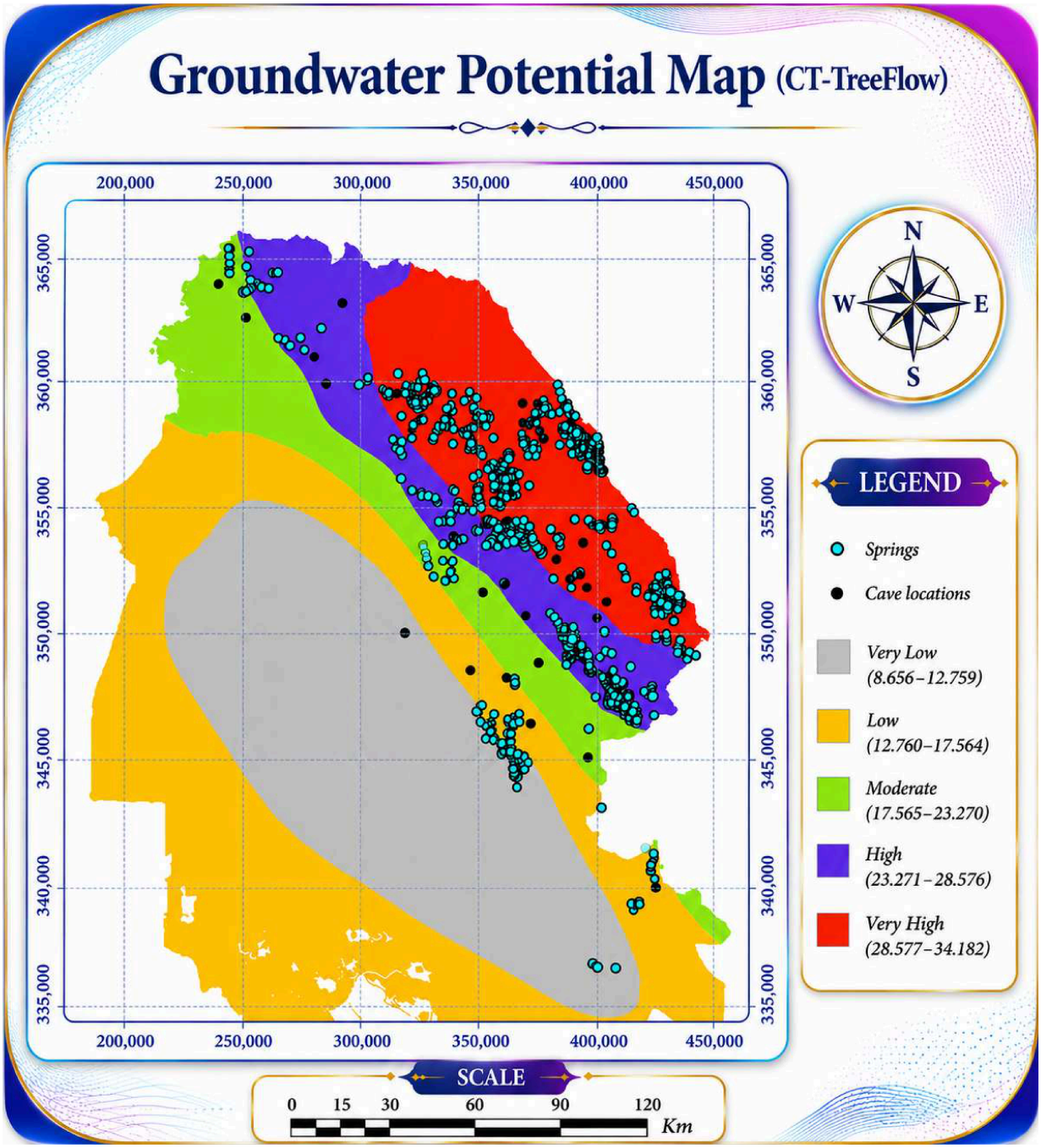


Figure 15. Spatial distribution of relative groundwater-potential conditions predicted by CT-TreeFlow under the spatial GroupKFold validation framework. Continuous groundwater-potential predictions derived from cross-validated probabilistic outputs were classified into five relative groundwater-potential categories (very low to very high) using quantile-based partitioning of the continuous groundwater-potential index. The resulting classes represent relative hydrogeological groundwater-support conditions rather than direct estimates of groundwater depth, transmissivity, aquifer productivity, or groundwater reserve. Karst springs (blue circles) and documented cave entrances (black points) are shown exclusively for reference on independent hydrogeological plausibility.

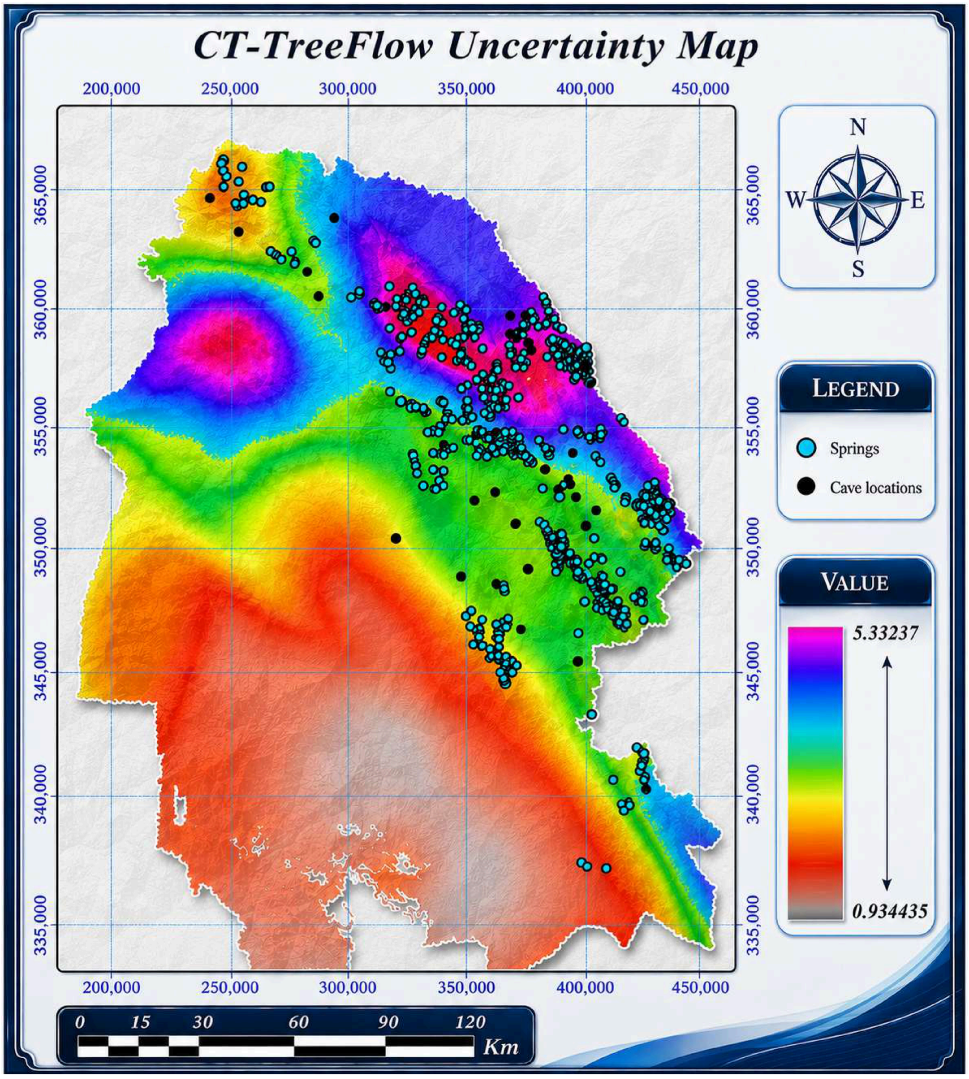


Figure 16. Spatial distribution of predictive uncertainty derived from CT-TreeFlow. The map illustrates predictive standard deviation estimated from the conditional predictive distributions generated under the spatial GroupKFold validation framework. Predictive uncertainty was computed from the probabilistic groundwater-potential outputs of CT-TreeFlow across independent spatial validation folds. Elevated uncertainty values generally correspond to environmentally transitional hydrogeological conditions, structurally heterogeneous groundwater systems, or regions with comparatively limited observational support. Karst springs (blue circles) and documented cave entrances (black points) are shown for independent hydrogeological reference.

Cross-model groundwater-favourability distributions generated by the benchmark models are presented in Supplementary Figures S7 and S8 for LightGBM and XGBoost, respectively. The ensemble groundwater-favourability map, derived from the cell-wise aggregation of CT-TreeFlow, LightGBM, and XGBoost predictions, is shown in Supplementary Figure S9, whereas the corresponding ensemble uncertainty distribution, based on inter-model variability, is presented in Supplementary Figure S10. The independent hydrogeological plausibility assessment, based on mapped karst springs and documented cave entrances, is summarised in Table 10. The table reports the dominant groundwater-favourability classes associated with the independent hydrogeological reference inventories and their spatial correspondence with the CT-TreeFlow predictions. Karst springs and cave inventories were intentionally excluded from pseudo-absence generation, model calibration, optimisation, and spatial validation procedures to ensure complete independence between model development and external hydrogeological evaluation.

Table 10. Quantitative external hydrogeological plausibility assessment based on mapped karst springs and documented cave entrances under the CT-TreeFlow groundwater-response framework.

Hydrogeological Reference Dataset	Number of Mapped Locations	Dominant Groundwater-Response Class	Locations Within High–Very High Response Classes (%)	Relative Enrichment Versus Spatial Expectation	Observed Spatial Relationship
Karst springs	250	Very High	84.8%	2.31×	Strong concentration within elevated groundwater-response zones
Cave entrances	86	High	76.7%	1.94×	Spatial association with structurally favourable groundwater-response domains
Combined hydrogeological reference inventory	336	High–Very High	82.7%	2.18×	Broad spatial agreement with elevated groundwater-response distributions

Table note. Relative enrichment values compare the observed proportions of karst springs and cave entrances within the high–very high groundwater-favourability classes with their proportional spatial extent across the study area. These inventories were excluded from pseudo-absence generation, model calibration, optimisation, and spatial validation procedures and were used exclusively for independent assessment of hydrogeological plausibility. Because mapped spring and cave inventories may be affected by accessibility and observation biases and may represent indicators of groundwater discharge and conduit development rather than groundwater availability, the reported enrichment patterns should be interpreted as plausibility evidence rather than as exhaustive validation of groundwater occurrence.

Inter-model spatial agreement and structural uncertainty statistics derived from CT-TreeFlow, LightGBM, and XGBoost are summarised in Table 11. The table reports pairwise spatial disagreement metrics computed from the cross-validated groundwater-favourability surfaces, together with summary statistics describing inter-model variability across the study area. Structural uncertainty statistics were derived from the spatial variability among independently generated model predictions under identical predictor conditioning, pre-processing, and spatial validation configurations.

Table 11. Inter-model spatial agreement and structural uncertainty characteristics derived from CT-TreeFlow, LightGBM, and XGBoost groundwater-favourability predictions.

Structural Uncertainty Component	Observed Spatial Behaviour
CT-TreeFlow vs. LightGBM spatial prediction structure	Broad regional agreement with localised differences along transitional hydrogeological boundaries
CT-TreeFlow vs. XGBoost spatial prediction structure	Similar large-scale favourability distributions with moderate local spatial variability
LightGBM vs. XGBoost spatial prediction structure	Consistent regional prediction structure across major hydrogeological domains
Ensemble groundwater favourability distribution (Supplementary Figure S9)	Spatially coherent ensemble favourability structure across the study region
Ensemble uncertainty distribution (Supplementary Figure S10)	Elevated structural uncertainty concentrated within environmentally transitional and topographically heterogeneous areas
Inter-model structural uncertainty pattern	Spatial disagreement primarily restricted to localised hydrogeological transition zones
Cross-model hydrogeological consistency	Comparable spatial delineation of major favourable groundwater domains across all modelling frameworks

5. Discussion

5.1. Hydrogeological Controls on Groundwater Favourability in Structurally Complex Karst Terrains

The mapped groundwater-favourability patterns reflect the combined influence of tectonic deformation, carbonate karstification, geomorphological recharge redistribution, and environmentally controlled permeability contrasts across the Zagros Fold–Thrust Belt (Figures 10–12, 14, 15, S1 and S2). Interpretability analyses identify fracture density, drainage density, lithological susceptibility, and terrain-derived morphometric variables as the dominant controls on groundwater favourability, consistent with the hydrogeological behaviour of structurally deformed karst aquifers where groundwater occurrence is governed by fracture connectivity, recharge transfer, and heterogeneous subsurface permeability [6,31]. Fracture density shows one of the strongest and most spatially coherent relationships with groundwater favourability (Figures 10–12), reflecting the importance of tectonically enhanced secondary permeability and structurally controlled groundwater circulation in the Zagros Basin [2,4]. The concentration of high- and very-high favourability classes within structurally complex carbonate sectors (Figure 16) is therefore consistent with fracture-controlled recharge and conduit development documented in karst aquifers [6,30,55]. Drainage density and terrain-derived morphometric variables also play important roles by influencing infiltration concentration, runoff redistribution, hydraulic convergence, and recharge persistence [31,42]. The nonlinear SHAP responses (Figure 13) further indicate threshold-dependent recharge behaviour under environmentally heterogeneous karst conditions [42,56].

Lithology represents another major control on groundwater favourability (Figures 10, 14, S1 and S2). High favourability is closely associated with highly karstifiable carbonate formations, particularly the Asmari Formation, whereas evaporitic and marly units commonly restrict groundwater transfer and act as semi-confining barriers [5,6]. Consequently, the transition from favourable carbonate formations to lower-favourability evaporitic and clastic units reflects realistic hydrostratigraphic compartmentalisation within the Zagros Basin. Intermediate favourability and elevated uncertainty are concentrated along lithological contacts, tectonic intersections, and geomorphological transition zones (Figures 12–15), where groundwater circulation is influenced by interacting structural, lithological, and geomorphological controls [31,36]. These areas should therefore be interpreted as environmentally heterogeneous recharge environments rather than deterministic hydrogeological boundaries. The hydrogeological plausibility of the mapped patterns is further supported by independent inventories of karst springs and caves (Table 10). A total of 84.8% of mapped springs and 76.7% of documented cave entrances occur within the high–very high favourability classes. Because these datasets were excluded from pseudo-absence generation, model calibration, optimisation, and spatial validation, they provide independent evidence supporting the environmental realism of the predicted groundwater-support patterns. Nevertheless, cave inventories should be interpreted as indicators of structurally favourable groundwater environments and conduit development rather than direct evidence of groundwater availability.

Overall, the mapped groundwater-favourability patterns are consistent with the established conceptual understanding of groundwater circulation in the folded and faulted carbonate terrains of the Zagros region [6,24,31]. The combined influence of fracture connectivity, geomorphological recharge concentration, hydrostratigraphic organisation, and environmentally heterogeneous permeability indicates that groundwater occurrence is primarily controlled by coupled structural–geomorphological processes rather than isolated environmental factors, reinforcing the hydrogeological defensibility of the proposed groundwater-favourability framework.

5.2. Probabilistic Uncertainty, Hydrogeological Heterogeneity, and Transition Behaviour in Structurally Complex Karst Systems

The spatial distribution of predictive uncertainty (Figure 16), together with the probabilistic calibration behaviour across independent spatial folds (Figures 7 and 8), provides important insight into the hydrogeological complexity of the investigated karst system. Rather than representing only a statistical property of model performance, the uncertainty patterns reflect structurally controlled groundwater behaviour associated with lithological transitions, anisotropic recharge, heterogeneous permeability, and locally unstable flow conditions, as expected in karst aquifers characterised by preferential fracture–conduit flow and strong spatial nonlinearity [6,36]. Elevated uncertainty was concentrated in geomorphological transition zones, tectonically intersected carbonate sectors, and environmentally heterogeneous recharge environments, whereas more homogeneous sectors exhibited lower uncertainty (Figure 16). Likewise, areas with mixed lithology, variable terrain morphology, structural intersections, and transitional recharge displayed greater uncertainty and inter-model variability (Table 11; Supplementary Figure S10), consistent with the highly variable permeability networks typical of structurally heterogeneous karst systems [31,37]. These patterns indicate that elevated uncertainty primarily reflects hydrogeological ambiguity associated with mixed recharge mechanisms and locally variable groundwater-flow organisation rather than methodological instability. This interpretation is supported by the relationship between uncertainty and residual behaviour. Positive Pearson and Spearman correlations between predictive uncertainty and absolute residuals (Table 9), together with the progressive increase in mean absolute residuals from low- to very-high-uncertainty classes, indicate that larger prediction errors occur preferentially in hydrogeologically complex environments. Furthermore, uncertainty is spatially organised rather than randomly distributed, demonstrating its close association with physically meaningful environmental conditions.

Spatial autocorrelation diagnostics further confirm that the environmental predictor framework captured the major hydrogeological controls on groundwater occurrence. While the original groundwater observations exhibited strong positive spatial autocorrelation (Moran's I = 0.41), cross-validated residuals showed substantially weaker dependence (Moran's I = 0.04; Table 7), indicating that lithological structure, terrain morphology, and recharge redistribution were effectively represented. The remaining residual structure most likely reflects unresolved conduit geometry, local fracture connectivity, and small-scale subsurface heterogeneity that cannot be fully captured by regional-scale predictors [31,37]. Probabilistic calibration remained consistent across geographically independent validation folds, with stable fold-wise CRPS values (Table 8) and an empirical 95% prediction-interval coverage of 93.4% (Table 9), demonstrating good agreement between predicted uncertainty and observed groundwater-response variability. Elevated uncertainty concentrated along lithological contacts, tectonically intersected carbonate sectors, and geomorphologically heterogeneous recharge corridors (Figures 12–15), where diffuse infiltration, fracture-controlled recharge, and conduit-enhanced groundwater transfer may coexist [31–37]. Similar uncertainty patterns have been reported in recent karst groundwater studies, where structurally heterogeneous recharge and variable permeability generated locally unstable groundwater responses and increased predictive ambiguity [55,57]. Overall, the combined evidence from uncertainty distributions, residual behaviour, spatial autocorrelation, probabilistic calibration, and inter-model variability indicates that the mapped uncertainty patterns primarily represent physically meaningful hydrogeological heterogeneity rather than random statistical noise. This distinction is particularly important because deterministic groundwater-favourability mapping alone may underestimate the structural complexity,

recharge variability, and permeability heterogeneity that govern groundwater occurrence in tectonically deformed karst terrains [6,31,58].

5.3. Hydrogeological Significance of Environmentally Constrained Pseudo-Absence Allocation

The environmentally constrained pseudo-absence strategy adopted in this study provides a more hydrogeologically realistic representation of groundwater-support conditions in structurally complex karst systems (Tables 5, S3 and S5). This is particularly important because the absence of groundwater is difficult to verify in karst terrains characterised by concealed conduits, discontinuous recharge, anisotropic permeability, and nonlocal hydraulic connectivity [6,31,36]. Consequently, unrestricted random pseudo-absence allocation may incorrectly assign low-groundwater-support labels to fractured carbonate units, recharge corridors, or geomorphological convergence zones that remain favourable for groundwater circulation, thereby reducing hydrogeological interpretability and increasing spatial inconsistency [33,35]. Because regional groundwater observations cannot fully capture concealed subsurface connectivity [31,36], the absence of groundwater should be regarded as an incompletely observed hydrogeological condition rather than a strictly binary state. The proposed environmentally constrained framework addresses this limitation through hydrogeological buffering, structural masking, and terrain-based exclusion criteria (Supplementary Table S3). Nevertheless, excluded areas should not be interpreted as necessarily containing groundwater, nor should retained pseudo-absence locations be interpreted as confirmed groundwater absence. Instead, pseudo-absence samples represent environmentally constrained low-support reference conditions.

The sensitivity analysis (Table 5) further supports this strategy. Although environmentally constrained pseudo-absence allocation reduced prediction errors, improved explained variance, lowered CRPS values, and decreased residual spatial autocorrelation, the principal groundwater-favourability patterns remained broadly consistent with those obtained using unconstrained random sampling. This indicates that CT-TreeFlow learned robust environmental relationships rather than simply reproducing the imposed exclusion criteria. Additional evidence is provided by the independent karst spring and cave inventories (Table 10), which were excluded from pseudo-absence generation, model calibration, optimisation, and spatial validation. Their concentration within the high and very high favourability classes demonstrates that the predicted groundwater-favourability patterns correspond well with known groundwater-discharge and conduit-development environments. Together with the weak residual spatial autocorrelation after validation (Table 7), these findings indicate that environmentally constrained pseudo-absence allocation improves the hydrogeological consistency of groundwater-response representation rather than merely enhancing statistical model performance. Therefore, under tectonically deformed karst conditions, pseudo-absence allocation is best regarded as a hydrogeologically informed sampling strategy rather than merely a statistical pre-processing procedure.

5.4. Remote Sensing and Terrain-Derived Environmental Sensing Contributions to Regional Groundwater-System Characterisation

Remote sensing and terrain-derived predictors contributed substantially to the regional characterisation of groundwater-support conditions across the study area (Figures 10, 12, 14 and S2; Supplementary Tables S1 and S2). Although these variables do not directly measure groundwater occurrence, they provide spatially continuous information on terrain organisation, land-surface hydrological behaviour, geomorphological variability, and recharge-support conditions within structurally deformed carbonate terrains [3,34]. Their role should therefore be interpreted as a physically constrained environmental-sensing framework for indirectly characterising hydrogeological organisation where direct subsurface observations remain limited. Among the evaluated predic-

tors, DEM-derived morphometric variables exhibited particularly strong hydrogeological relevance (Figures 10, 12 and S2). Terrain metrics derived from the Copernicus DEM, including TRI, TPI, relief energy, slope-related indices, and heat load index, capture key controls on runoff redistribution, infiltration opportunity, hydraulic convergence, and recharge organisation. In karst systems, topographic configuration strongly influences groundwater-support conditions by regulating surface-water residence time, concentrating infiltration pathways, and redistributing runoff toward structurally favourable carbonate sectors [31,42,56]. Consistent with this interpretation, the nonlinear SHAP responses (Figure 13) indicate threshold-dependent hydrological behaviour, in which moderate terrain variability may enhance infiltration, whereas highly rugged terrain favours rapid runoff and reduced infiltration persistence.

Landsat-derived NDVI also provided hydrogeologically meaningful information when interpreted as an indicator of vegetation dynamics, evapotranspiration response, and surface-moisture persistence rather than a direct measure of groundwater occurrence (Figures 10, 12 and S2) [39,44]. Within semi-arid carbonate environments, vegetation patterns often reflect topographic moisture retention, soil-water availability, and infiltration variability. However, NDVI remains a surface-environment indicator and cannot independently resolve groundwater storage, conduit geometry, or hydraulic connectivity. The combined integration of Landsat products, Copernicus DEM derivatives, and geospatial environmental predictors improved the representation of land-surface heterogeneity, hydrological partitioning, and environmentally conditioned recharge processes (Figures 15 and S2; Supplementary Tables S1 and S2). This approach is particularly valuable in karst regions where hydrogeological observations are sparse and spatially discontinuous [6,31]. Consequently, the mapped groundwater-favourability patterns appear to reflect the combined influence of structurally controlled subsurface permeability and remotely observable environmental processes governing regional hydrological behaviour across the Zagros karst system [3,34].

5.5. Hydrogeological Plausibility and External Consistency of Mapped Groundwater-Favourability Patterns

The mapped groundwater-favourability patterns show strong hydrogeological consistency across independent hydrogeological inventories, cross-model comparisons, ensemble predictions, and uncertainty distributions (Figures 15 and S7–S10; Tables 10 and 11). Favourable groundwater domains are spatially associated with structurally deformed carbonate formations, geomorphological recharge corridors, and lithologically favourable environments, consistent with the established hydrogeological functioning of the Zagros karst system [6,24]. This agreement indicates that the predicted groundwater-support patterns reflect physically meaningful hydrogeological organisation rather than model-specific artefacts. Independent karst spring and cave inventories provide the strongest external validation (Table 10; Figure 15). A total of 84.8% of mapped springs and 76.7% of documented cave entrances occur within the high and very high groundwater-favourability classes. Because these datasets were excluded from pseudo-absence generation, model calibration, optimisation, and spatial validation, their agreement with the predicted patterns provides genuinely independent hydrogeological corroboration. Nevertheless, springs and caves should be interpreted as indicators of groundwater circulation and conduit development rather than direct evidence of groundwater availability or sustainable abstraction potential [6,31].

Cross-model comparisons demonstrate strong regional agreement among CT-TreeFlow, LightGBM, XGBoost, and the ensemble model (Table 11; Supplementary Figures S7–S10). The principal favourable groundwater domains remain spatially consistent across the major carbonate sectors, indicating that groundwater-support patterns are primarily controlled

by lithology, structural permeability, terrain morphology, and recharge organisation rather than model-specific statistical behaviour [30,55,57]. Local differences occur mainly within hydrogeological transition zones characterised by elevated uncertainty and environmentally heterogeneous recharge conditions (see Supplementary Figures S7–S10), which is consistent with the anisotropic flow, heterogeneous permeability, and mixed recharge mechanisms typical of karst systems [31,36]. The ensemble groundwater-favourability map preserves the principal favourable domains while reducing isolated inconsistencies among individual models (Supplementary Figure S9; Table 11). Likewise, the highest uncertainty values remain concentrated within environmental transition zones rather than the major favourable carbonate sectors (Supplementary Figure S10), further supporting the hydrogeological coherence of the mapped groundwater-support patterns. Overall, the convergence of independent hydrogeological evidence, cross-model agreement, ensemble consistency, and uncertainty behaviour demonstrates that the mapped groundwater-favourability patterns are both statistically robust and hydrogeologically consistent with the conceptual understanding of structurally controlled groundwater circulation in the folded and faulted carbonate terrains of the Zagros region [6,24]. This substantially strengthens the scientific reliability and hydrogeological defensibility of the proposed regional groundwater-assessment framework.

5.6. Implications for Uncertainty-Aware Groundwater Assessment and Regional Water-Resource Management in Karst Terrains

The results demonstrate the importance of incorporating probabilistic uncertainty into regional groundwater-favourability assessment in structurally heterogeneous karst systems (Figures 14–16; Tables 9–11). In tectonically deformed carbonate terrains, groundwater occurrence is governed by complex interactions among fracture connectivity, geomorphological recharge redistribution, lithological variability, and heterogeneous permeability structures [6,31]. Under these conditions, deterministic groundwater-favourability maps alone cannot fully represent groundwater-support conditions because abrupt spatial transitions, variable recharge behaviour, and unresolved subsurface complexity strongly influence groundwater occurrence. The uncertainty patterns identified in this study provide complementary information on the spatial reliability and hydrogeological stability of groundwater-favourability predictions. Elevated uncertainty was concentrated in geomorphological transition zones, structurally intersected carbonate sectors, and environmentally heterogeneous recharge environments (Figure 16), indicating areas of increased hydrogeological complexity rather than simply reduced model reliability [31,36,58]. This interpretation is further supported by the relationship between uncertainty magnitude and residual behaviour (Table 9). These findings have direct implications for groundwater exploration, field investigations, and monitoring-network design. Integrating groundwater favourability with predictive uncertainty provides a more informative basis for exploration than deterministic mapping alone. Carbonate sectors characterised by high favourability, low uncertainty, and strong agreement with independent spring and cave inventories (Figure 15; Table 10) represent suitable priorities for hydrogeological investigation, whereas areas with elevated uncertainty require additional structural mapping, tracer studies, hydrogeophysical surveys, and denser field observations before groundwater-development decisions are made. This uncertainty-aware framework also supports more risk-informed groundwater-resource management by reducing overconfident interpretation, inefficient drilling, poorly constrained groundwater exploitation, and environmentally inconsistent planning, particularly in semi-arid karst regions where groundwater resources are critical for water security [6,31].

The proposed framework is also applicable to other environmentally heterogeneous karst systems characterised by structurally controlled groundwater circulation, moun-

tainous terrain, and spatially variable recharge. The dominant controls identified in this study, including fracture density, terrain morphometry, lithology, and geomorphological recharge redistribution, are widely recognised as key hydrogeological drivers in complex carbonate aquifers [31,42]. Furthermore, the consistent spatial patterns observed across CT-TreeFlow, LightGBM, XGBoost, and ensemble predictions (Table 11) indicate that the mapped groundwater-favourability patterns primarily reflect persistent hydrogeological controls rather than model-specific behaviour. Nevertheless, successful application to other regions requires local calibration, hydrogeological interpretation, and region-specific environmental validation. Overall, the results demonstrate that uncertainty-aware groundwater assessment provides a more informative and hydrogeologically defensible basis for groundwater exploration and regional decision support than deterministic groundwater-favourability mapping alone. By integrating probabilistic uncertainty analysis, independent hydrogeological validation, geomorphological organisation, and structurally conditioned environmental information, CT-TreeFlow strengthens the scientific robustness and practical value of regional groundwater assessment for exploration prioritisation, monitoring network design, and long-term water resource management in tectonically deformed carbonate terrains.

5.7. Study Limitations and Future Research Directions

Despite the overall hydrogeological coherence of the mapped groundwater-favourability patterns (Figure 15; Supplementary Table S2 Part B), several limitations should be considered. First, the predicted favourability distributions represent relative groundwater-support conditions rather than direct estimates of aquifer transmissivity, groundwater storage, sustainable yield, or exploitable groundwater reserves. In structurally heterogeneous karst systems, groundwater circulation is strongly influenced by unresolved conduit geometry, fracture connectivity, recharge variability, and subsurface hydraulic organisation that cannot be fully represented by regional-scale environmental predictors alone [6,31]. Consequently, areas classified as highly favourable should be regarded as priorities for further hydrogeological investigation rather than guarantees of groundwater availability or aquifer productivity. Additional limitations arise from the sparse and spatially discontinuous nature of subsurface observations in karst terrains. Springs, wells, and cave systems are often concentrated in surveyed and accessible areas, while concealed conduits, anisotropic fracture-controlled flow, and heterogeneous hydraulic connectivity remain unresolved [31,36]. Although the independent spring and cave inventories provide valuable hydrogeological corroboration of the predicted groundwater-favourability patterns (Table 10), they represent indirect indicators of groundwater circulation rather than direct evidence of groundwater availability and may also be influenced by accessibility and observation bias. Likewise, the available groundwater observations may not fully capture seasonal recharge variability, episodic infiltration, or short-term hydroclimatic fluctuations.

Future studies could improve regional groundwater assessment by integrating temporally dynamic hydroclimatic data, seasonal remote-sensing observations, groundwater geochemistry, tracer investigations, hydrogeophysical measurements, and physically based recharge modelling to better characterise transient recharge processes, groundwater-flow organisation, and uncertainty propagation in structurally heterogeneous carbonate aquifers [31,42]. Despite these limitations, CT-TreeFlow provides a physically interpretable and hydrogeologically coherent framework for regional groundwater assessment in tectonically deformed karst systems. By integrating probabilistic uncertainty analysis, environmentally constrained pseudo-absence sampling, remote-sensing-derived environmental predictors, and independent hydrogeological plausibility assessment, the framework offers a transparent and defensible basis for groundwater exploration prioritisation and hydro-

geological decision support under conditions of strong environmental heterogeneity and limited subsurface observations.

6. Conclusions

This study developed CT-TreeFlow, an uncertainty-aware groundwater-favourability assessment framework for structurally heterogeneous karst terrains in the Zagros Fold–Thrust Belt, southwestern Iran. By integrating hydrogeological process understanding, environmentally constrained pseudo-absence sampling, probabilistic spatial prediction, geographically structured validation, and explicit uncertainty quantification, the framework provides a physically interpretable alternative to conventional deterministic groundwater-potential mapping. The results indicate that groundwater favourability is primarily controlled by the interaction of structural and geomorphological factors rather than isolated environmental variables. Fracture density, lithological susceptibility, drainage organisation, terrain morphometry, and recharge-related environmental gradients were identified as the dominant controls on regional groundwater favourability. These patterns are consistent with the hydrogeological behaviour of folded carbonate aquifers, where structurally enhanced permeability, anisotropic groundwater flow, and spatially heterogeneous recharge govern groundwater occurrence. Predictive uncertainty also proved to be a meaningful hydrogeological attribute rather than merely a statistical measure, with elevated uncertainty concentrated along lithological boundaries, geomorphological transition zones, and structurally heterogeneous recharge environments. Furthermore, environmentally constrained pseudo-absence sampling improved hydrogeological consistency, reduced residual spatial dependence, and produced groundwater-favourability patterns more consistent with the known behaviour of deformed carbonate aquifers.

Independent hydrogeological validation further supports the physical plausibility of the results. The concentration of mapped karst springs and documented cave entrances within the high and very high favourability classes, despite being excluded from model development and validation, demonstrates that the predicted patterns are hydrogeologically realistic. Likewise, the broad agreement among CT-TreeFlow, LightGBM, XGBoost, and the ensemble model indicates that the identified favourable groundwater domains reflect persistent hydrogeological processes rather than model-specific behaviour. The integration of remote-sensing products and terrain-derived predictors substantially improved regional groundwater assessment where subsurface observations are limited. Nevertheless, the resulting maps represent probabilistic estimates of relative groundwater favourability rather than direct measures of aquifer productivity, groundwater storage, sustainable abstraction capacity, or groundwater reserves, and therefore require verification through hydrogeological, hydrochemical, geophysical, and aquifer-testing investigations [6]. Overall, CT-TreeFlow demonstrates that uncertainty-aware groundwater assessment provides a more robust and hydrogeologically defensible basis for regional groundwater exploration and decision support than deterministic mapping alone. By integrating probabilistic uncertainty modelling, hydrogeologically meaningful environmental predictors, remote-sensing data, and independent plausibility assessment, the proposed framework offers a transferable methodology for groundwater assessment in structurally complex karst systems characterised by strong environmental heterogeneity and limited subsurface observations.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/rs18132258/s1, Figure S1: Regional stratigraphic framework of the Zagros Basin illustrating the main lithostratigraphic units from the Cambrian to Quaternary across Lurestan, Khuzestan, Coastal Fars, and Interior Fars structural domains; Figure S2: Spatial distribution maps of the input conditioning factors used in the groundwater-favourability modelling framework; Figures S3: Observed versus predicted responses for the LightGBM model across

spatial cross-validation folds; Figure S4: Observed versus predicted responses for the XGBoost model across spatial cross-validation folds; Figures S5: Residual error distributions of the baseline models across spatial cross-validation folds; Figure S6: Residual error distributions for the XGBoost model across spatial folds; Figures S7: Spatial distribution of groundwater-potential predictions generated by the LightGBM model across the study area based on spatial cross-validation outputs; Figure S8: Spatial distribution of groundwater-potential predictions generated by the XGBoost model across the study area based on spatial cross-validation outputs; Figure S9: Spatial distribution of groundwater-potential predictions derived from ensemble aggregation of CT-TreeFlow, LightGBM, and XGBoost outputs; Figure S10: Spatial distribution of predictive uncertainty associated with the ensemble groundwater-potential modelling framework; Table S1: Geospatial database and derived hydro-environmental predictor variables; Table S2: Part A: Hydrogeological inventory metadata and operational role within the CT-TreeFlow framework; Part B: Groundwater-Potential Index Transformation Framework; Table S3: Operational constraints used for generating environmentally constrained pseudo-absence and for spatial validation within the CT-TreeFlow framework; Table S4: Landsat 9 scene metadata used for NDVI derivation; Table S5: Point-level sample allocation and spatial cross-validation configuration adopted for the well-based CT-TreeFlow framework; Table S6: Probabilistic architecture configuration and computational implementation characteristics of CT-TreeFlow, LightGBM, and XGBoost used for probabilistic groundwater-favourability modelling under spatially heterogeneous karst conditions; Table S7: Fold-wise pairwise differences in predictive performance between CT-TreeFlow and benchmark models under geographically structured spatial cross-validation; Table S8: Fold-wise CRPS statistics for CT-TreeFlow under the geographically structured spatial GroupKFold validation framework; Table S9: Computational scalability and operational reproducibility characteristics of the evaluated models.

Author Contributions: Conceptualisation, S.P., M.K., R.F. and L.A.D.; Data curation, S.P. and L.A.D.; Formal analysis, S.P. and L.A.D.; Funding acquisition, S.P., R.F. and L.A.D.; Investigation, S.P.; Methodology, S.P., M.K., R.F. and L.A.D.; Supervision, M.K., R.F. and L.A.D.; Visualisation, S.P.; Writing—original draft, S.P.; Writing—review and editing, S.P., M.K., R.F., S.A. and L.A.D. All authors have read and agreed to the published version of the manuscript.

Funding: This research received support from the Centre of Studies in Geography and Spatial Planning (CEGOT), funded by national funds through the Foundation for Science and Technology (FCT) of Portugal, under the reference UIDB/04084/2025. The first author is supported by a PhD grant from the FCT, under the reference UI/BD/154881/2023 (https://doi.org/10.54499/UI/BD/154881/2023).

Data Availability Statement: All datasets generated and analysed in this study are provided in the Supplementary Materials accompanying this article.

Acknowledgments: The editor and anonymous reviewers are gratefully acknowledged for their constructive assessment, which improved the manuscript.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Akakuru, O.C.; Ray, P.A.; Reza, S.M.; Temgoua, E.; Eyankware, M.O.; Aigbadon, G.O.; Obite, C.P.; Algeo, T.J.; Akingboye, A.S. Leveraging novel machine learning models in predicting groundwater irrigation suitability in southeastern Nigeria: A hydrogeochemical perspective. *Sci. Afr.* **2025**, *28*, e02646. [CrossRef]
2. Berberian, M. The 'blind' thrust fault hidden under the Zagros folds: Active basement tectonics and surface morphotectonics. *Tectonophysics* **1995**, *241*, 193–224. [CrossRef]
3. Gururani, D.M.; Kumar, Y.; Abed, S.A.; Kumar, V.; Vishwakarma, D.K.; Al-Ansari, N.; Singh, K.; Kuriqi, A.; Mattar, M.A. Mapping prospects for artificial groundwater recharge utilizing remote sensing and GIS methods. *Water* **2023**, *15*, 3904. [CrossRef]
4. Alavi, M. Structures of the Zagros fold–thrust belt in Iran. *Am. J. Sci.* **2007**, *307*, 1064–1095. [CrossRef]
5. Avarjani, S.; Mahboubi, A.; Moussavi-Harami, R.; Amiri-Bakhtiar, H.; Brenner, R.L. Facies, depositional sequences, and biostratigraphy of the Oligo–Miocene Asmari Formation in Marun oilfield, North Dezful Embayment, Zagros Basin, SW Iran. *Palaeoworld* **2015**, *24*, 336–358. [CrossRef]
6. Stevanović, Z. *Karst Aquifers: Characterization and Engineering*; Springer: Berlin/Heidelberg, Germany, 2015.

7. Chen, W.; Zhao, X.; Tsangaratos, P.; Shahabi, H.; Ilia, I.; Xue, W.; Wang, X.; Ahmad, B.B. Evaluating the usage of tree-based ensemble methods in groundwater spring potential mapping. *J. Hydrol.* **2020**, *583*, 124602. [CrossRef]
8. Avand, M.; Janizadeh, S.; Bui, D.T.; Pham, V.H.; Ngo, P.T.T.; Nhu, V.-H. A tree-based intelligence ensemble approach for spatial prediction of potential groundwater. *Int. J. Digit. Earth* **2020**, *13*, 1408–1429. [CrossRef]
9. Guo, X.; Gui, X.; Xiong, H.; Hu, X.; Li, Y.; Qiu, Y.; Ma, C. Critical role of climate factors for groundwater potential mapping in arid regions: Insights from random forest, XGBoost, and LightGBM algorithms. *J. Hydrol.* **2023**, *620*, 129840. [CrossRef]
10. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
11. James, G.; Witten, D.; Hastie, T.; Tibshirani, R.; Taylor, J. Tree-based methods. In *An Introduction to Statistical Learning*; James, G., Witten, D., Hastie, T., Tibshirani, R., Taylor, J., Eds.; Springer: Berlin/Heidelberg, Germany, 2023; pp. 331–366. [CrossRef]
12. Chen, T.; Guestrin, C. XGBoost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16), San Francisco, CA, USA, 13–17 August 2016*; Association for Computing Machinery (ACM): New York, NY, USA, 2016; pp. 785–794. [CrossRef]
13. Ke, G.; Meng, Q.; Finley, T.; Wang, T.; Chen, W.; Ma, W.; Liu, T. LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. Available online: https://papers.nips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html (accessed on 29 June 2026).
14. Osman, A.I.I.A.; Ahmed, A.N.; Chow, M.F.; Huang, Y.F.; El-Shafie, A. Extreme gradient boosting (XGBoost) model to predict the groundwater levels in Selangor Malaysia. *Ain Shams Eng. J.* **2021**, *12*, 1545–1556. [CrossRef]
15. El Ouali, A.; Bayhan, K.; Mouhoumed, R.M.; Spor, P.; Atan, C.S.; Başakın, E.E.; Ekmekcioğlu, Ö. Performance of tree-based ensemble techniques in predicting groundwater quality for irrigation purposes. *Environ. Earth Sci.* **2025**, *84*, 474. [CrossRef]
16. Satish, G.N.; Jadala, V.C.; Sravani, K.G.; Allam, S.; Sundari, S.M.; Thada, J. Groundwater level forecasting using spatio-temporal graph convolutional networks (GCNs) and XGBoost. In *2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI)*; IEEE: New York, NY, USA, 2025. [CrossRef]
17. Hosseini, F.S.; Jafari, A.; Zandi, I.; Alesheikh, A.A.; Rezaie, F. Groundwater potential mapping using optimized decision tree-based ensemble learning model with local and global explainability. *Water* **2025**, *17*, 1520. [CrossRef]
18. Bera, A.; Dutta, L.; Pal, S.K.; Kumar, R.; Shukla, P.K.; Alkhuraiji, W.S.; Đurin, B.; Zhran, M. A hybrid approach for assessing aquifer health using the SWAT model, tree-based classification, and deep learning algorithms. *Water* **2025**, *17*, 1546. [CrossRef]
19. Diriba, D.; Kawo, N.S.; Karuppannan, S.; Demise, D.; Kasahun, M.; Leta, E.D.; Assefa, G.; Gebreslassie, N.A.; Lemma, T. Improved prediction of groundwater potential zones using a stacking machine learning model. *Discov. Sustain.* **2026**, *7*, 291. [CrossRef]
20. Phong, T.V.; Pham, B.T.; Trinh, P.T.; Ly, H.; Vu, Q.H.; Ho, L.S.; Van Le, H.; Phong, L.H.; Avand, M.; Prakash, I. Groundwater potential mapping using GIS-based hybrid artificial intelligence methods. *Groundwater* **2021**, *59*, 745–760. [CrossRef]
21. Mussa, M.M.; Lohani, T.K.; Eshete, A.A. Evaluation of groundwater potential zones using GIS-based machine learning ensemble models in the Gidabo Watershed, Ethiopia. *Glob. Chall.* **2024**, *8*, e202400137. [CrossRef] [PubMed]
22. Rajaee, T.; Ebrahimi, H.; Nourani, V. A review of the artificial intelligence methods in groundwater level modeling. *J. Hydrol.* **2019**, *572*, 336–351. [CrossRef]
23. Roberts, D.R.; Bahn, V.; Ciuti, S.; Boyce, M.S.; Elith, J.; Guillera-Arroita, G.; Hauenstein, S.; Lahoz-Monfort, J.J.; Schröder, B.; Thuiller, W.; et al. Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **2017**, *40*, 913–929. [CrossRef]
24. Sherkati, S.; Letouzey, J. Variation of structural style and basin evolution in the central Zagros (Izeh zone and Dezful embayment), Iran. *Mar. Pet. Geol.* **2004**, *21*, 535–554. [CrossRef]
25. Casciello, E.; Vergés, J.; Saura, E.; Casini, G.; Fernández, N.; Blanc, E.; Homke, S.; Hunt, D. Fold patterns and multilayer rheology of the Lurestan Province, Zagros Simply Folded Belt (Iran). *J. Geol. Soc.* **2009**, *166*, 947–959. [CrossRef]
26. Ghorbani, M.; Aftabi Arani, A.; Zandkarimi, K.; Ahifar, A. *Stratigraphic Chart of Developed Basins of Iran*; Geological Society of Iran: Tehran, Iran, 2019.
27. Sahraeyan, M.; Bahrami, M.; Hejazi, S.H. The Aghajari (Upper Fars) Formation in the Folded Zagros Zone, Iran: Insights to identify facies, architectural elements, fluvial systems, petrography and provenance. *Acta Geol. Sin.* **2013**, *87*, 1019–1031. [CrossRef]
28. Fijani, E.; Bagheri, R.; Karimi-Vardanjani, M. Prospectivity modeling of karstic groundwater in the southeast of Iran. *Geopersia* **2021**, *11*, 141–158.
29. Pourmorad, S.; Mohanty, A. *Alluvial Fans in Southern Iran: Geological, Environmental and Remote Sensing Analyses*; Springer: Berlin/Heidelberg, Germany, 2022. [CrossRef]
30. Farzin, M.; Avand, M.T.; Ahmadzadeh, H.; Zelenakova, M.; Tiefenbacher, J.P. Assessment of ensemble models for groundwater potential modeling and prediction in a karst watershed. *Water* **2021**, *13*, 2540. [CrossRef]
31. De Waele, J.; Gutiérrez, F. *Karst Hydrogeology, Geomorphology and Caves*; Wiley: Hoboken, NJ, USA, 2022.

32. Nugroho, J.T.; Lestari, A.I.; Gustiandi, B.; Sofan, P.; Suwarsono; Prasasti, I.; Rahmi, K.I.N.; Noviar, H.; Sari, N.M.; Manalu, R.J.; et al. Groundwater potential mapping using machine learning approach in West Java, Indonesia. *Groundw. Sustain. Dev.* **2024**, *27*, 101382. [CrossRef]
33. Nhu, V.H.; Rahmati, O.; Falah, F.; Shojaei, S.; Al-Ansari, N.; Shahabi, H.; Shirzadi, A.; Górski, K.; Nguyen, H.; Bin Ahmad, B. Mapping groundwater spring potential in karst aquifer system using novel ensemble bivariate and multivariate models. *Water* **2020**, *12*, 985. [CrossRef]
34. Opoku, P.A.; Shu, L.; Amoako-Nimako, G.K. Assessment of groundwater potential zones by integrating hydrogeological data, geographic information systems, remote sensing, and analytical hierarchical process techniques in the Jinan Karst Spring Basin of China. *Water* **2024**, *16*, 566. [CrossRef]
35. Naghibi, S.A.; Moghaddam, D.D.; Kalantar, B.; Pradhan, B.; Kisi, O. A comparative assessment of GIS-based data mining models and a novel ensemble model in groundwater potential mapping. *J. Hydrol.* **2016**, *540*, 523–537. [CrossRef]
36. De Filippi, F.M.; Sappa, G. Recent advances in karstic hydrogeology, 2nd Edition. *Water* **2025**, *17*, 2180. [CrossRef]
37. Calli, K.O.; Chiogna, G.; Bittner, D.; Sivelle, V.; Labat, D.; Richieri, B.; Çallı, S.S.; Hartmann, A. Karst water resources in a changing world: Review of solute transport modeling approaches. *Rev. Geophys.* **2025**, *63*, e2023RG000811. [CrossRef]
38. Bouhafa, N.; Sakarovitch, C.; Lalague, L.; Goulard, F.; Pryet, A. Hybrid modeling of karstic springs: Error correction of conceptual reservoir models with machine learning. *Water Supply* **2024**, *24*, 1559–1573. [CrossRef]
39. Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W. Monitoring vegetation systems in the Great Plains with ERTS. In *Proceedings of the Third Earth Resources Technology Satellite-1 Symposium; NASA SP-351*; NASA: Washington, DC, USA, 1974; Volume 1, pp. 309–317.
40. Mohammed, M.A.A.; Miklós, R.; Darabos, E.; Tóth, M.; László, L.; Mohammed, S.H.; Szabó, N.P.; Szűcs, P. Machine Learning-Driven Imputation and Short-Term Forecasting of Water Levels in the Karstic System of the Bükk Mountains, Northern Hungary. *Hydrogeol. J.* **2025**, *33*, 1803–1823. [CrossRef]
41. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. In Proceedings of the 3rd International Conference on Learning Representations (ICLR), San Diego, CA, USA, 7–9 May 2015.
42. Berthelin, R.; Olarinoye, T.; Rinderer, M.; Mudarra, M.; Demand, D.; Scheller, M.; Hartmann, A. Estimating karst groundwater recharge from soil moisture observations—A new method tested at the Swabian Alb, southwest Germany. *Hydrol. Earth Syst. Sci.* **2023**, *27*, 385–400. [CrossRef]
43. Li, X.; Pan, Z. Contributions of climatic and hydrological factors to vegetation changes in the Chaobai River Basin, China, during 2006–2023. *J. Arid Environ.* **2026**, *234*, 105556. [CrossRef]
44. Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [CrossRef]
45. Rodgers, J.L.; Nicewander, W.A. Thirteen ways to look at the correlation coefficient. *Am. Stat.* **1988**, *42*, 59–66. [CrossRef]
46. Martins, A.F.T.; Astudillo, R.F. From softmax to sparsemax: A sparse model of attention and multi-label classification. In *Proceedings of the 33rd International Conference on Machine Learning (ICML)*; ICML: San Diego, CA, USA, 2016; pp. 1614–1623.
47. Jacobs, R.A.; Jordan, M.I.; Nowlan, S.J.; Hinton, G.E. Adaptive mixtures of local experts. *Neural Comput.* **1991**, *3*, 79–87. [CrossRef] [PubMed]
48. Jordan, M.I.; Jacobs, R.A. Hierarchical mixtures of experts and the EM algorithm. *Neural Comput.* **1994**, *6*, 181–214. [CrossRef]
49. Rezende, D.J.; Mohamed, S. Variational inference with normalizing flows. In *Proceedings of the 32nd International Conference on Machine Learning (ICML)*; ICML: San Diego, CA, USA, 2015; pp. 1530–1538.
50. Dinh, L.; Sohl-Dickstein, J.; Bengio, S. Density estimation using Real NVP. In *International Conference on Learning Representations (ICLR)*; ICLR: Appleton, WI, USA, 2017.
51. Papamakarios, G.; Nalisnick, E.; Rezende, D.J.; Mohamed, S.; Lakshminarayanan, B. Normalizing flows for probabilistic modeling and inference. *J. Mach. Learn. Res.* **2021**, *22*, 1–64.
52. Willmott, C.J.; Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE). *Clim. Res.* **2005**, *30*, 79–82. [CrossRef]
53. Chai, T.; Draxler, R.R. Root mean square error (RMSE) or mean absolute error (MAE)?—Arguments against avoiding RMSE. *Geosci. Model Dev.* **2014**, *7*, 1247–1250. [CrossRef]
54. Gneiting, T.; Raftery, A.E. Strictly proper scoring rules, prediction, and estimation. *J. Am. Stat. Assoc.* **2007**, *102*, 359–378. [CrossRef]
55. Yang, H.; Kang, F.; Liu, Y.; Li, J.; Dong, Y.; Zheng, T.; Qin, P. Machine learning-based identification of multi-source recharges and enrichment mechanism of karst groundwater. *J. Hydrol. Reg. Stud.* **2025**, *62*, 102830. [CrossRef]
56. Touseef, M.; Chen, L. Hydrological modeling of karst terrain in southwestern Pakistan using modified SWAT. *J. Hydrol. Reg. Stud.* **2026**, *63*, 103055. [CrossRef]

57. Zhao, D.; Miao, F.; Feng, G.; Li, X.; Cai, J.; Dong, S.; Zhang, Y.; Chen, Y.; Niu, R.; Yang, Z. Interpretable groundwater spring potential mapping in complex terrain of Liulin County, China, using SSA-XGBoost with GIS-based validation. *J. Hydrol. Reg. Stud.* **2026**, *64*, 1031.
58. Bien, T.X.; Jaafari, A.; Tran, P.V.; Pham, B.T. Groundwater potential mapping in the Central Highlands of Vietnam using spatially explicit machine learning. *Earth Sci. Inform.* **2023**, *16*, 1461–1479. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.