Journal of Hydrology: Regional Studies 65 (2026) 103372

Contents lists available at ScienceDirect

Journal of Hydrology: Regional Studies

journal homepage: www.elsevier.com/locate/ejrh

Sinkhole risk forecasting in the Lithuania–Latvia Karst region using artificial intelligence

Vytautas Samalavičius [a,*,1], Jānis Bikše [b,2], Ilya Zaslavsky [c,3], Ieva Lekstutytė [a,4], Jurga Arustienė [a,5], Gintaras Žaržojus [a,6], Assemzhan Kunsakova [a,7], Inga Retike [b,8], Sonata Gadeikienė [a,9], Saulius Gadeikis [a,10]

[a] Institute of Geosciences at Vilnius University, Čiurlionio str. 21/27, Vilnius LT-03101, Lithuania
[b] Faculty of Science and Technology at University of Latvia, Jelgavas str. 1, LV-1004, Riga, Latvia
[c] San Diego Supercomputer Center, University of California San Diego, La Jolla, CA, USA

ARTICLE INFO

Keywords:
Karst aquifers
Machine learning
Sinkholes
Remote sensing
Geohazard
Time-series

ABSTRACT

Study region: The Lithuania–Latvia transboundary gypsum karst region is highly prone to sinkhole formation, posing a significant geohazard to infrastructure, agriculture, and groundwater resources. Risk assessment is challenged by sparse groundwater monitoring networks and strongly heterogeneous karst hydrogeology.

Study focus: This study develops an end-to-end, remote-sensing–informed and data-driven workflow to reconstruct missing daily groundwater-level (GWL) records and to forecast monthly sinkhole formation risk. Daily GWL gaps were reconstructed using supervised machine-learning models driven by satellite-derived climate and water-storage variables. Reconstructed signals were aggregated to monthly resolution and translated into sinkhole risk classes using a Random Forest classifier. A defensible operational target was applied at each well using an empirical 90th-percentile threshold (≥4 newly formed sinkholes per month). Model training employed fold-scoped preprocessing and class-imbalance controls to ensure robust evaluation.

New hydrological insights for the region: Across seven wells (2003–2024), models combining groundwater level, seasonal encoding and hydroclimatic features outperformed single-domain baselines, achieving an accuracy of ~0.96, high-risk precision of ~0.98, and recall of ~0.85.

* Corresponding author.
E-mail addresses: vytautas.samalavicius@chgf.vu.lt (V. Samalavičius), janis.bikse@lu.lv (J. Bikše), izaslavsky@ucsd.edu (I. Zaslavsky), ieva.lekstutyte@chgf.vu.lt (I. Lekstutytė), jurga.arustiene@chgf.vu.lt (J. Arustienė), gintaras.zarzojus@gf.vu.lt (G. Žaržojus), assemzhan.kunsakova@chgf.stud.vu.lt (A. Kunsakova), inga.retike@lu.lv (I. Retike), sonata.gadeikyte@gf.vu.lt (S. Gadeikienė), saulius.gadeikis@gf.vu.lt (S. Gadeikis).
[1] ORCID: 0000-0001-5791-7483.
[2] ORCID: 0000-0003-2380-2690.
[3] ORCID: 0000-0003-4191-8275.
[4] ORCID: 0000-0003-0104-0605.
[5] ORCID: 0000-0003-0216-5063.
[6] ORCID: 0000-0003-2079-4680.
[7] ORCID: 0009-0003-4814-8864.
[8] ORCID: 0000-0002-3931-1143.
[9] ORCID: 0000-0001-6504-5769.
[10] ORCID: 0000-0001-5054-440X.

https://doi.org/10.1016/j.ejrh.2026.103372
Received 12 December 2025; Received in revised form 25 February 2026; Accepted 21 March 2026
Available online 30 March 2026

Explainable analyses highlight multi-week hydroclimatic preconditioning as the dominant driver, with sinkhole clusters occurring within ±30 days of groundwater-level peaks. By integrating forecasted groundwater and hydroclimatic features with remote-sensing inputs, the framework can be implemented as an operational decision-support tool or dashboard to deliver up-to-date sinkhole risk alerts, supporting coordinated cross-border infrastructure protection and groundwater management.

1. Introduction

Karst terrain is characterised by a unique set of geotechnical and environmental problems, which in turn have a significant impact on land use in these areas. A sinkhole is defined as a geologic feature that is characteristic of karst areas. It is a sudden collapse of the ground surface or a slow, imperceptible subsidence, which have the potential to cause damage to structures. Collapse sinkholes are the result of mechanical breakdowns, though chemical processes are important in the development of the cave in the underlaying rock and the removal of the collapsed materials. The formation of caves is precipitated by the dissolution process occurring within the rock. In the event of the rocks above the cave being rigid, they will only collapse when they are unable to support the weight above the cave. Karst terrains are characterized by highly heterogeneous subsurface structures formed by the dissolution of soluble rocks such as limestone, dolomite, and gypsum, leading to the development of conduits, caves, and sinkholes (Hartmann et al., 2014a). Sinkholes may form abruptly as collapse features or gradually through subsidence, posing significant risks to infrastructure, land use, and groundwater resources. Water plays a dominant role in these processes, as infiltration, groundwater-level fluctuations, and rapid recharge can trigger mechanical failure of overburden materials above dissolution voids (Hartmann et al., 2014a). Due to the non-linear and localized nature of karst processes, sinkhole formation remains difficult to predict using conventional deterministic approaches.

Karst aquifers develop in soluble rocks (limestone, dolomite, gypsum, halite) where dissolution enlarges fractures, joints, and bedding planes into conduits, caves, and sinkholes. Unlike porous-media aquifers that transmit water through intergranular pores, flow in karst is dominated by highly heterogeneous fracture–conduit networks, sometimes with turbulent conditions and limited matrix storage. This structure creates preferential pathways, rapid throughflow, and flashy responses of springs and water levels to recharge. The same properties also heighten contamination risk: pollutants can bypass natural filtration and move quickly with little attenuation, making vulnerability assessment and protection strategies essential.

Karst research is crucial for understanding groundwater resources, engineering geology, and environmental sustainability, as these systems provide drinking water for many communities worldwide. Climate change and increasing freshwater demand may intensify groundwater extraction from karst aquifers, potentially accelerating dissolution processes and sinkhole formation (Džeriņš, 2023; Linares et al., 2016). Water plays a central role in the development of cover-collapse, cave-collapse, and cover-subsidence sinkholes. Downward percolation of water through unconsolidated overburden can erode soil into solution-enlarged fractures, while fluctuations in groundwater levels alter effective stresses and hydraulic gradients within the subsurface. Numerous studies have demonstrated empirical relationships between sinkhole occurrence, precipitation, and groundwater-level variability, supported by both observational analyses and numerical modeling (Alrowaimi, 2016; Byle et al., 2024; Liu et al., 2024; Öztürk et al., 2025; Satkūnas et al., 2007).

In the transboundary karst region of Lithuania and Latvia, gypsum karst is actively developing within Late Devonian (Frasnian) evaporite formations (Paukštys and Narbutas, 1996). Although sulfate and carbonate karst occupy relatively small areas within the Baltic region, they have significant engineering and geo-ecological implications, such as elevated sulfate levels and other chemical alterations (Paukštys and Narbutas, 1996) pollution from nutrient leaching, which harms both surface and groundwater quality (Rudzianskaitė, 2017). Given that the hydrogeological system spans national boundaries, karst-related hazards and groundwater-quality issues necessitate coordinated cross-border monitoring and management strategies between Lithuania and Latvia. However, effective risk assessment is hindered by the limited density of groundwater observation wells operated by national monitoring agencies, constraining the ability to characterize groundwater fluctuations and detect early warning signals of instability. Sparse monitoring is particularly problematic in karst aquifers, where groundwater levels can respond rapidly to precipitation, pumping, and dissolution processes. Traditional hydrogeological modeling approaches often struggle under these conditions due to strong heterogeneity, non-linearity, and data scarcity (Jourde and Wang, 2023). As a result, there is growing interest in data-driven methods capable of integrating diverse data sources to improve groundwater-level estimation and hazard assessment. Filtration studies using electrical tomography show that high infiltration through cover rocks promotes active karst development in the Tatula gypsum-bearing deposits, mainly driven by intense rainfall infiltration. Where cover deposits are thin or poorly sealed, sinkholes form, and groundwater is highly vulnerable to surface pollution due to the high permeability of sandy soils and fractured Devonian carbonate rocks (Klizas and Šečkus, 2007).

The application of AI-driven time-series machine learning models has proven successful in estimating groundwater levels by integrating satellite-derived environmental data as predictive features (Collenteur et al., 2024; Rohde et al., 2021; Sun, 2013). Key inputs such as river runoff, temperature, soil moisture, and precipitation, combined with historical groundwater level monitoring data, enable the development of accurate predictive models that capture both short-term fluctuations and long-term trends in groundwater dynamics. ML modeling for groundwater level imputation and sinkhole risk prediction benefit from feature engineering (Bianchini et al., 2022; Bikše et al., 2023; Bilgilioğlu et al., 2025; Collenteur et al., 2024; Ramirez et al., 2022; Sharma et al., 2025).

AI models can effectively learn complex, non-linear relationships between environmental variables and groundwater fluctuations,

improving forecasting accuracy even in data-scarce regions like the Lithuanian-Latvian karst area, where monitoring well density is low. Recent work shows how complementary ML approaches advance sinkhole analysis and predicting the risks of their formation.In an Italian travertine karst plain, a full ML workflow produced susceptibility and risk maps—from inventory building and feature engineering through model training and planning-ready outputs (Bianchini et al., 2022). A U-Net to high-resolution imagery to delineate sinkholes and demonstrated that models trained on drone data can transfer effectively to satellite scales (Alrabayah et al., 2024). In the Konya Basin, Turkey explainable ensemble models (RF, XGBoost, LightGBM) combined with SHAP yielded accurate susceptibility maps while highlighting the most influential predictors, improving transparency for decision-makers (Bilgilioğlu et al., 2025). The effectiveness of tree-based ensemble methods in this study is further supported by who show that random forests robustly capture non-linear interactions between geomorphology and hydrological response in karst aquifers (Janža et al., 2025).

Despite these advances, applications in the Baltic region remain limited. Existing studies have focused primarily on groundwater isotope-geochemistry classification, gap imputation in groundwater time series, aquifer–spring associations, and impulse response modeling, with relatively few remotely associated with karst region (Bikše et al., 2023; Jemeļjanova et al., 2023; Koit et al., 2023; Retike et al., 2016; Samalavičius et al., 2024, 2025a).

Sinkhole occurrence is known to be strongly linked to groundwater dynamics, yet operational prediction is hindered by sparse groundwater monitoring and fragmented modeling approaches. Satellite-derived water-storage states can substitute for dense in situ monitoring and act as integrative predictors of karst instability. In this study the argument states that sinkhole formation is governed less by isolated meteorological events than by the cumulative hydraulic state and seasonal memory of the aquifer–epikarst system.The key novelty of this work lies in coupling groundwater-level gap reconstruction and sinkhole risk prediction into a unified, remote-sensing–driven workflow, rather than treating these challenges as separate modeling tasks. This study address two difficult problems simultaneously: 1) reconstructing daily gaps in sparse groundwater-level (GWL) records and; 2) producing operational predictors of sinkhole risk—all at monthly resolution and without relying on dense in situ networks. In this study point-scale GWL observations with satellite-derived climate forcings (precipitation, temperature), actual and potential evapotranspiration (AET, PET), and large-scale terrestrial groundwater storage proxies (TWS and GWS) are used to train four supervised ML models. Unlike typical single-purpose studies, our framework couples an ML regression for GWL imputation with a categorical classifier for sinkhole susceptibility, using process-guided feature engineering (e.g., multi-scale hydroclimatic anomalies, AET–PET imbalance, antecedent moisture/recharge indices, GLDAS-based storage anomalies and their rates of change, and recent GWL drawdown/recovery). Applied to separate wells across the heterogeneous Lithuania–Latvia karst belt, this approach demonstrates that physically meaningful, satellite-informed features can both stabilize groundwater time series and yield planning-ready risk signals in data-scarce transboundary settings.

By fusing limited in situ GWL data with satellite products, the approach both improves continuity of groundwater records and generates operational predictors for sinkhole risk assessment, offering a scalable, cost-effective complement to traditional monitoring in data-scarce karst settings.

2. Methods

2.1. Dataset

The study assembled a multi-source dataset integrating climate, water-storage, and groundwater observations for a transboundary karst region spanning Latvia and Lithuania. Daily near-surface air temperature (°C) and precipitation (mm) were obtained from the European gridded observational climate dataset (E-OBS) gridded dataset; available from 1950 onward (Cornes et al., 2018). AET and PET were taken from Global Land Evaporation Amsterdam Model (GLEAM) available for 1980–2023 (Miralles et al., 2025). GRACE-based TWS and GWS were sourced from Global Land Data Assimilation System (GLDAS); available from 2003-02-01 (Li et al., 2020, 2019).

Groundwater observations were compiled from national agencies. The Latvian observation dataset was provided by the Latvian Environment, Geology and Meteorology Centre. Lithuanian well records were supplied by the Lithuanian Geological Survey (LGT) PožVIS information system; these wells form part of the national groundwater monitoring network. Open-access data on sinkholes were also obtained from the LGT GEOLIS where sinkhole formation dates are comprised. Considering that it may take a few days to weeks for sinkhole to be noticed the real data accuracy is monthly.

Seven wells provide GWL measurement time frame suitable for temporal overlap with the climate and storage series; analyses that require the full set of variables are therefore constrained by the shortest record, namely the GLDAS products beginning on 2003-02-01. After merging the climatic and storage series with the groundwater records, the number of daily GWL entries available for each well was in range 3410–6186 entries with the single exception of LT_839, which has only 351 entries and was excluded from subsequent analyses to avoid bias arising from its substantially limited sample size.

Feature engineering encompassed seasonal encodings, lagged terms, rolling summaries, and monthly averages to represent temporal structure and data quality (Appendix 1).

These features capture the core hydrologic behaviors that govern groundwater and sinkhole dynamics while giving ML models clean, learnable signals. Lagged variables encode delayed aquifer responses to meteorological forcing—e.g., the time it takes for rainfall or snowmelt to percolate through the vadose zone and reach the water table—so models can learn cause–effect relationships that unfold over weeks to months. Rolling sums and means approximate short-term memory: sums of precipitation or evapotranspiration act like accumulated inputs/outputs to storage, whereas rolling means denoise high-frequency variability in temperature, GWS, or GWL, improving the signal-to-noise ratio. Cyclical encodings of day- and month-of-year let models learn seasonality without artificial discontinuities at year boundaries, and the seasonal phase provides a compact representation of timing shifts (e.g., an early or

late wet season). Climatological monthly baselines turn raw values into context-aware anomalies, helping models distinguish "normal for this month" from genuinely unusual conditions.

Together, these engineered variables improve modeling by providing redundant, physically meaningful predictors when observations are missing, and they boost predictive models by embedding domain structure that reduces overfitting and enhances generalization across wells and years.

2.2. Machine learning

The programming workflow was implemented in *Python* (version 3.11.13), employing a suite of open-source scientific computing libraries. Additional graphical and tabular outputs were produced using *AutoCAD* for cross-sectional drawing and *Microsoft Excel* for tabular summaries.

The modeling workflow addresses two linked objectives: (1) reconstruction of missing groundwater-level (GWL) observations and (2) prediction of sinkhole formation risk. Groundwater-level imputation provides continuous, gap-free time series that serve as inputs for subsequent sinkhole risk modeling, forming an integrated, two-stage framework (Fig. 1).

In the present study, both regression and classification approaches were implemented. For regression tasks—predicting continuous numerical outcomes—ensemble-based algorithms such as *ExtraTreesRegressor (ETR), RandomForestRegressor (RFR)* and *GradientBoostinRegresor (GBR)* were employed to capture complex, non-linear relationships between predictor variables and target responses. For classification tasks—predicting discrete categorical outcomes—models such as the *RandomForestClassifier (RFC), LightGBM (LGBM), XGBoost (XGB) and ExtraTreesClassifier (ETC)* were utilized, enabling the discrimination between classes based on learned patterns in the feature space.

In this context, sinkhole risk was formulated as a categorical classification problem, distinguishing between low-risk and high-risk conditions. The classification models aimed to identify the complex interplay between groundwater dynamics, climatic forcing, and subsurface conditions that predispose an area to sinkhole occurrence. This dual-stage strategy ensures that the final predictive models are trained on the most complete and representative datasets available, thereby enhancing both model performance and interpretability in the context of groundwater-related hazard assessment.

2.2.1. Groundwater Level Imputation

Imputation was performed to reconstruct continuous groundwater level time series. Predictors were derived from hydroclimatic and storage variables assembled into predefined feature sets. Core predictors included precipitation, air temperature, potential evapotranspiration PET), actual evapotranspiration (AET), and their rolling sums or means with lags of 30, 90, 270, and 360 days. The use of rolling aggregates of hydroclimatic variables follows the feature-construction framework proposed in Collenteur et al. (2024), although that study addressed groundwater-level forecasting rather than gap imputation. Two feature-set configurations were evaluated: a *no_seasonal* set comprising only hydroclimatic and storage predictors and a *with_seasonal* set that additionally included a seasonal phase encoding.

The target variable was groundwater level referenced to mean sea level (*gw_level_m_asl*). Records were grouped by monitoring point (*well_no*) and indexed by date. For each well, analyses were restricted to the period spanning the first to the last observed (non-null) groundwater level to avoid leading/trailing regions with no targets.

Predictors were drawn from climate and storage variables (and potentially their transformations) assembled into predefined feature sets.

To limit multicollinearity and stabilize model estimates, predictors were screened using the variance inflation factor (VIF). After median imputation and standardization applied within training folds only, an iterative backward elimination was performed per well and per candidate feature set until all retained predictors satisfied VIF < 5 (James et al., 2017). VIF diagnostics are provide in Appendix 2. Predictors exceeding this threshold, or rendered redundant by retained correlates, were removed. The resulting feature space comprised low-collinearity climate and storage drivers. Modelling was conducted independently for each well with observations indexed by date. Three machine learning algorithms were evaluated: Random Forest (RFC), LightGBM (LGBM), and Extra Trees (ETC). Model performance was assessed using the coefficient of determination (R^2) and mean absolute error (MAE) and the best performing model was selected to be used for imputation for all wells.

Following established methodology for groundwater imputation assessment (Bikše et al., 2023), realistic missingness was simulated by masking contiguous intervals that reflect empirically observed gap structures. Rather than randomly removing individual observations, continuous blocks of 7–60 days were withheld for testing. The full 2003–2024 record was evaluated using a time-aware

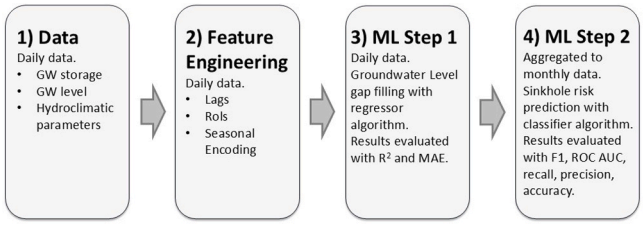

Fig. 1. Workflow of the data processing and machine learning pipeline.

cross-validation framework instead of a single static train-test split. Specifically, temporal cross-validation with block masking was applied (Fig. 2a), whereby continuous intervals of 7–60 days were withheld for testing and the remaining observations were used for training. This approach preserves temporal dependence in groundwater dynamics and accounts for potential non-stationarity in hydroclimatic controls. The masked indices were explicitly flagged and excluded from all model-selection procedures, ensuring that evaluation remained strictly out-of-sample and suitable for subsequent external validation or sensitivity analyses.

To achieve the best performing model per well ETR was wrapped in GridSearchCV (3-fold, scoring = 'r2', n_jobs = -1) to tune key hyperparameters: n_estimators ∈ {100, 200}, max_depth ∈ {None, 20}, min_samples_split = 2, min_samples_leaf ∈ {1, 2}. The hyperparameter search ranges were defined based on an initial screening analysis. Higher values of n_estimators were tested, but no further performance improvement was observed beyond 200 trees. Consequently, the search space for n_estimators was limited to {100, 200} to reduce computational cost without sacrificing model performance. The max_depth parameter showed sensitivity depending on the selected feature set, with shallow and moderately deep trees producing comparable results; therefore, the search was restricted to {None, 20} to capture this influence while avoiding overfitting. The winning (model, feature set) per well, were used to impute missing GW levels were imputed.

2.2.2. *Sinkhole formation risk assessment*

In the sinkhole risk assesment ML step risk indicator derived from *risk_level*, restricted to the *Low* and *High* categories was selected as a target. Observations were stratified by well and ordered by year and month to preserve temporal dependence. Class labels were encoded as binary_label ∈ {0,1}, mapping Low → 0 and High → 1.

Performance reporting included class-specific precision, recall, and F1 for both *Low risk* and *High risk* labels, alongside overall accuracy and area Under the Receiver Operating Characteristic Curve (ROC AUC). Precision for a class quantifies the proportion of samples predicted as that class that are truly members of it (e.g., *High risk* precision measures the reliability of positive alarms); recall quantifies the proportion of true members of that class that are correctly identified (e.g., *High risk* recall reflects sensitivity to minority events).

In this pipeline, ROC AUC is computed on the available predictions as implemented and reported alongside accuracy (Fig. 1b). To explicitly characterize prevalence, label counts were aggregated across all wells and feature sets, yielding a pronounced class imbalance with approximately 89% Low risk and 11% High risk. These totals both motivate the use of imbalance-aware training

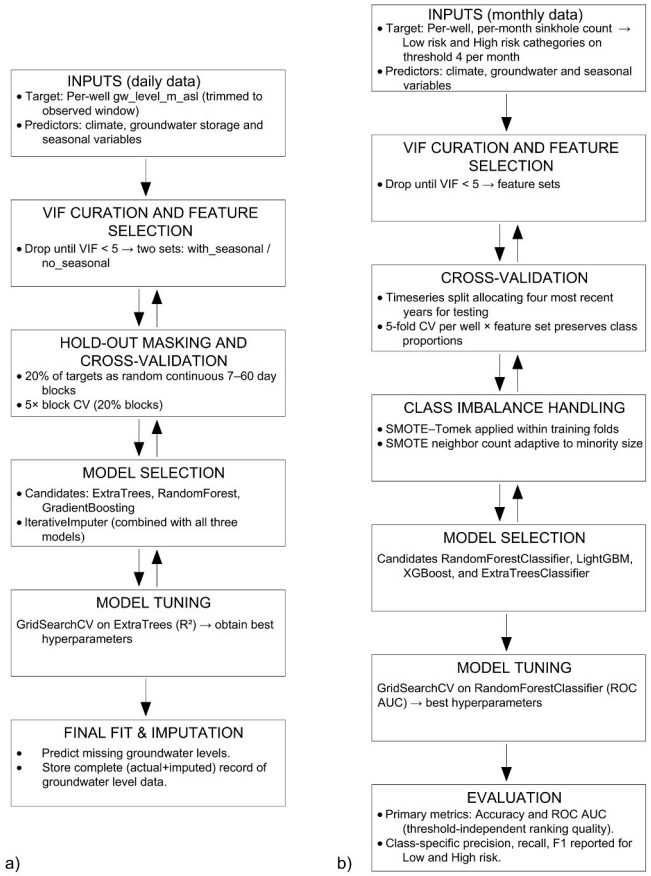

Fig. 2. Study design. a) Imputation of missing groundwater level data; b) Sinkhole formation risk assessment ML modeling.

(resampling and class-weight exploration) and contextualize the metrics: stable or high accuracy and strong *Low risk* scores can co-occur with deficient *High risk* recall, whereas improved *High risk* recall and F1 indicate better minority detection at the potential cost of some majority accuracy.

For each well, numeric columns were subset, features missing from a set were skipped, and VIF values were obtained on the standardized design matrix. Per-feature VIFs were then aggregated across wells to yield mean ± SD, together with counts of contributing wells (Appendix 3).

Predictor specification relied on predefined feature sets, with the sets motivated by the study's hydrologic conceptual model and exploratory analyses (Appendix 3): Climatic Features; Groundwater Features; Groundwater-storage (GWS) Features; GWS-Seasonal Features; Total Water-Storage (TWS) Features; and three combined specifications—CS (Climatic + Seasonal), GS Combined, and CGS Combined (Climatic + Groundwater/Storage + Seasonal). To control multicollinearity, each feature set was first diminished through variance inflation factor (VIF) screening For every well and VIF-curated set, models were trained and evaluated within the blocked time-series cross-validation pipeline: within each training fold, class imbalance was corrected using SMOTE–Tomek, categorical variables were encoded, and predictors were standardized using a scaler fit only on the training data; an inner grid search selected the classifier and its hyperparameters. The tuned model was then refit on the training fold and evaluated on the corresponding test split, and out-of-fold predictions were aggregated across repeats to compare feature sets, with all preprocessing and tuning strictly confined to the training data to prevent information leakage (Fig. 2b).

Model performance was estimated with a timeseries five-fold cross-validation design applied independently for each well and feature set. The test set was delegated with the values of four recent years 2020–2023, while the training set was left with the previous observation ensuring no data leakage could occur between them (Fig. 2b). Within each cross-validation fold, all data-dependent operations were fitted strictly on the training portion to also prevent information leakage.

Model selection adopted RFC as the base learner following a comparative benchmark in which LightGBM, XGBoost, and ExtraTreesClassifier were trained under identical preprocessing (SMOTE–Tomek, fold-scoped encoding), feature sets, and timeseries five-fold cross-validation; these alternatives exhibited approximately 10–15% lower accuracy on average across wells and feature sets, establishing RFC as the preferred model.

In order to achieve better performance, hyperparameters were tuned by an inner three-fold grid search that maximized mean area under the ROC curve. The objective optimized mean ROC AUC (Fig. 2b). The grid comprised *n_estimators* ∈ {200, 400}, *max_depth* ∈ {20, 30}, *min_samples_leaf* ∈ {2, 3, 5}, and *class_weight* ∈ {*balanced_subsample*, {0:1,1:4}, {0:1,1:6}}. The design targets class imbalance and temporal structure by conducting grid search on the native class distribution (avoiding resampling inside CV to prevent leakage and instability in small temporal folds), constraining tree depth and enforcing larger leaf sizes to reduce variance and suppress brittle minority splits, and comparing data-driven (*balanced_subsample*) versus fixed positive-class weighting to calibrate emphasis on the minority class. The configuration achieving the highest inner-fold ROC AUC was retained. The tuned model was refitted on the selected training features and evaluated on the held-out fold; both discrete class predictions and class probabilities were recorded.

For each well and feature set, out-of-fold predictions from the five splits were concatenated to form well-level assessments. These well-level results were subsequently pooled across wells to produce feature-set summaries. Primary metrics comprised overall accuracy and ROC AUC (computed on the discrete predictions as implemented), complemented by class-wise precision, recall, and F1 scores. Confusion matrices were aggregated across wells, normalized by row to emphasize per-class recall, and visualized in a tiled panel figure with one panel per feature set. To ensure proper data wrangling, an exception was set for wells containing only a single class after label filtering to be excluded from analysis for that feature set. This exception was not executed during the analysis and

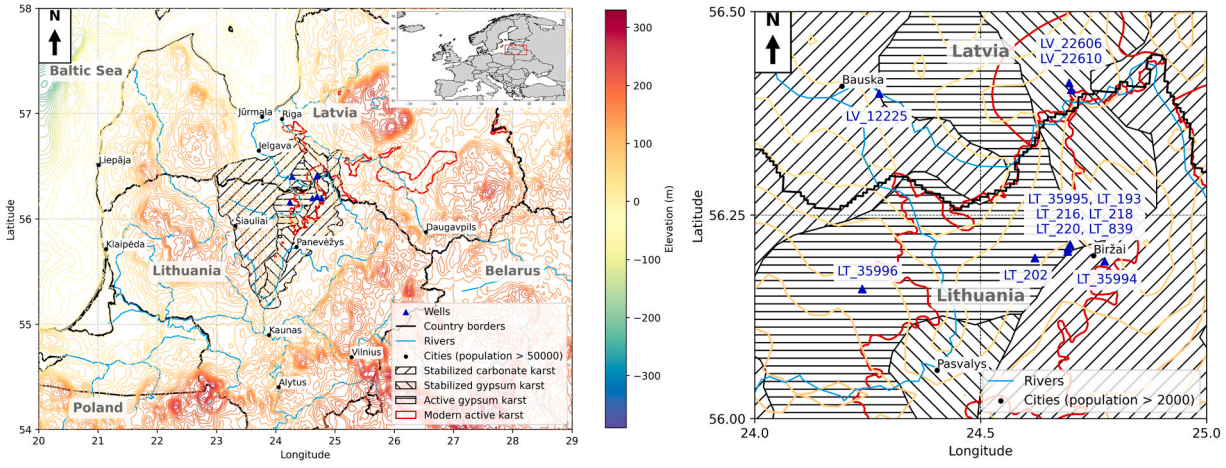

Fig. 3. Location of the transboundary karst region, rivers (blue), borders (black), cities ≥ 50k (black dots), and groundwater-monitoring wells location (blue triangles, labeled by national IDs). Karst domains: stabilized carbonate (hatched), stabilized gypsum (cross-hatched), active gypsum (horizontal lines) – after (Paukštys and Narbutas, 1996), original map authored by Paukstys, Vodzinskas, Heinsalu), and modern active karst (dark red) after (Levins and Buzajevs, 1999; Lithuanian Geological Survey under the Ministry of Environment, 2025).

modeling process.

3. Environmental setting

The transboundary karst region is a part of the Baltic Artesian Basin. The land surface altitude in the region varies from 25 to 150 m above sea level (Fig. 3). The Region's climate is classified as Dfb (snow, fully humid, warm summer) (Kottek et al., 2006).

The map differentiates four karst classes: stabilized carbonate karst, stabilized gypsum karst, active gypsum karst (shown with hatch fills, after (Paukštys and Narbutas, 1996), original map authored by Paukstys, Vodzinskas, Heinsalu), active gypsum karst (red linework), and modern active karst (dark-red linework) (Levins and Buzajevs, 1999; Lithuanian Geological Survey under the Ministry of Environment, 2025) (Fig. 3). These units delineate a NW–SE–trending belt from north-central Lithuania into southern Latvia, broadly coincident with the outcrop/subcrop of Devonian evaporites and carbonates. The hatched polygon that defines the principal

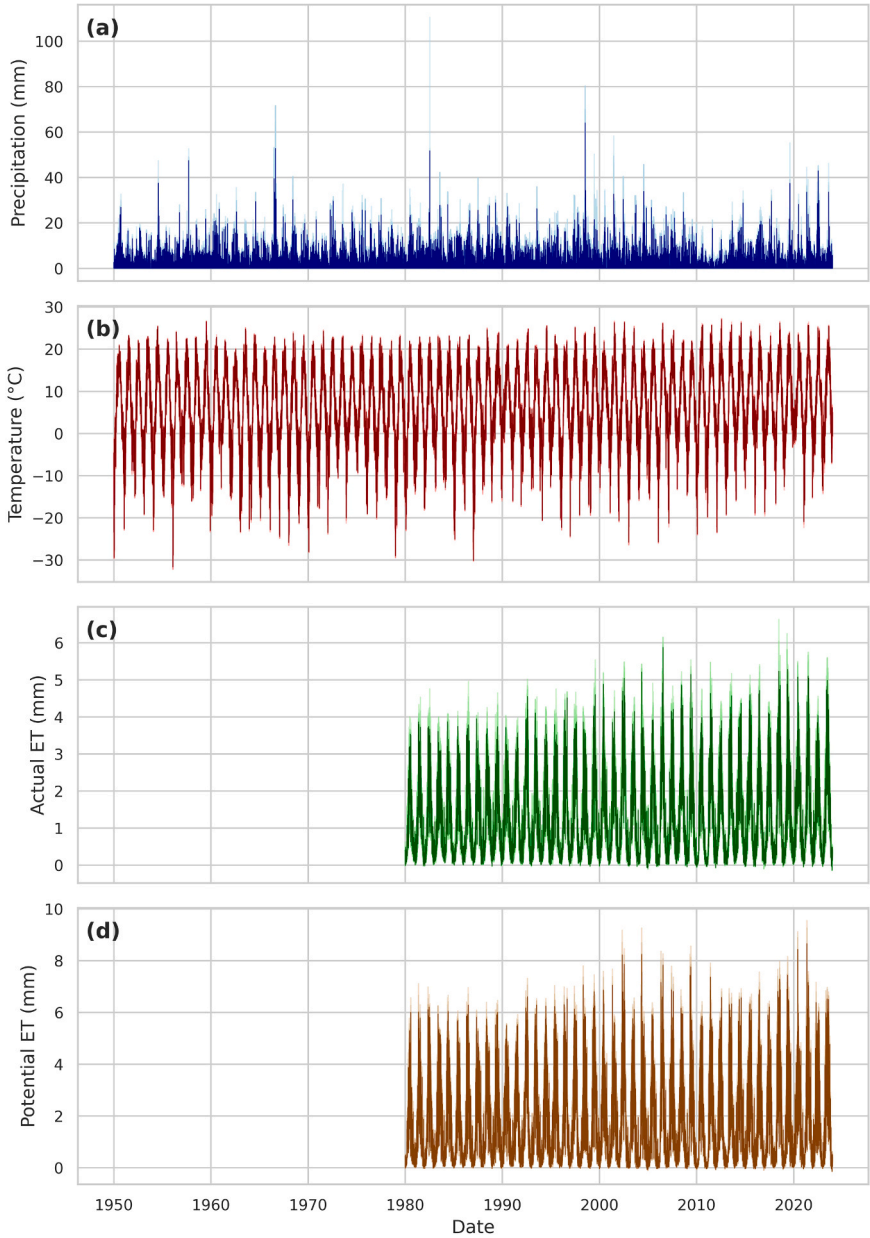

Fig. 4. Daily climate variables for the Lithuanian–Latvian transboundary area: (a) precipitation (mm/d, E-OBS), (b) air temperature (°C, E-OBS), (c) actual evapotranspiration (mm/d, GLEAM), and (d) potential evapotranspiration (mm/d, GLEAM). The dark tone line shows the per-date average across wells; faint lines show individual wells.

karst belt, its long axis is ~150–170 km, with a typical width of ~50–80 km (Fig. 3).

The "stabilized" zones represent relict or presently quiescent features where karstification in carbonate or gypsum strata is subdued by thicker Quaternary cover, lower hydraulic gradients, or infilled sinkholes. In contrast, the active gypsum karst traces narrow corridors and clusters, particularly northeast of Panevėžys and toward the Latvian border, where shallow Upper Devonian gypsum is intersected by river valleys. The modern active karst highlights the most dynamic sectors with ongoing subsidence and sinkhole formation; these sectors are topographically low, intersect fluvial corridors, and coincide with the highest density of mapped wells, underscoring groundwater sensitivity.

3.1. Climate

The Lithuanian–Latvian border region is shaped by a cool-temperate hydroclimate with pronounced annual cycles that are broadly coherent across monitoring locations (Fig. 4a–d). Precipitation and temperature span the full mid-twentieth-century to present record, whereas evapotranspiration (ET) commences later, yielding fewer observations for AET and PET (Fig. 4c–d; Table 1). The daily precipitation distribution is highly right-skewed—many dry or low-intensity days (median 0 mm/d; IQR 0–2.1 mm/day) punctuated by rare extremes up to 110.5 mm/d—while temperature exhibits large seasonal amplitude around a mean of 6.45 °C (range −32.2 to 27.3 °C) (Fig. 4a–b; Table 1). These features mirror regional analyses showing frequent light precipitation punctuated by heavy-rain episodes and an observed/expected increase in extremes through the 21st century (Rimkus et al., 2011). Evaporative demand generally exceeds supply (PET mean 1.77, AET 1.31 mm/d), with seasonal maxima aligned with summer warmth; small negative ET minima reflect processing artefacts and are negligible for water-balance purposes (Fig. 4c–d; Table 1). A similar warm-season evaporative deficit has been quantified for the Nemunas basin, where ET_0 increasingly outpaces precipitation from April–August (Stonevičius et al., 2018).

These joint characteristics translate directly into boundary conditions for karst development where carbonate or evaporite strata are present. The precipitation regime favors event-driven recharge: long sequences of low inputs followed by short, intense storms overcome soil moisture deficits and drive focused infiltration through karst and fractures (Fig. 4a; Table 1). Such hydraulic bursts efficiently widen preferential pathways and, in evaporitic settings, can accelerate dissolution and cover-collapse. Field evidence of sinkhole clustering and rapid hydraulic responses is well documented in the Biržai–Pasvalys/Skaistkalne gypsum-karst belt (Delina et al., 2012; Paukštys et al., 1999).

Thermal seasonality structures the chemical aggressiveness of infiltrating waters. Cool-season and snowmelt recharge, with higher CO_2 solubility and typically undersaturated chemistry, enhances carbonic-acid dissolution; warm-season soils, with elevated pCO_2 from respiration, can be equally effective when moisture is available. Consequently, transitional periods (spring, autumn) often combine chemically aggressive water with sufficient supply, producing disproportionate dissolution relative to their duration (Fig. 4b–d). The regional snow-cover regime (shorter seasons, frequent late-winter thaws) further concentrates effective recharge into shoulder seasons, sharpening these windows of karst aggressiveness (Rimkus et al., 2018)

The persistent evaporative deficit (PET > AET) implies hydrologic intermittency: extended dry periods interrupted by recharge pulses (Fig. 4c–d). Drying promotes desiccation cracking and soil piping, increasing connectivity to the subsurface; the subsequent first-flush after drought commonly carries high CO_2 and organic acids and can generate large hydraulic gradients, fostering both dissolution and suffosion. The same climate also heightens sensitivity to runoff focusing, which localizes recharge into karst-prone zones and amplifies subsidence hazard.

The regime summarized above—many dry days, strong thermal seasonality, and episodic high-intensity rainfall—suggests a suitable hydroclimatic set features for sinkhole risk modelling and groundwater level (GWL) reconstruction in the Lithuanian–Latvian karst transboundary area. The observed tendency toward heavier precipitation events and shorter, more variable snow seasons (Rimkus et al., 2018, 2011) suggests these contrasts—leading to intensified geohazards.

3.2. Geology and karst formation

Karstified gypsum and gypsiferous-dolomitic rocks in the Baltic States occur at two primary stratigraphic levels: the Narva Formation (Middle Devonian) and the Tatula and Salaspils formations (Dubnik regional stage) (Late Devonian, Frasnian age) (Paukštys and Narbutas, 1996). The gypsum deposits of the Narva Formation are buried at depths exceeding 100 m, preventing the development

Table 1
Summary statistics (count, mean, standard deviation, minimum, quartiles, maximum, and median) of daily precipitation and temperature (E-OBS) and daily actual and potential evapotranspiration (GLEAM) for the Lithuanian–Latvian transboundary area.

	Count	Mean	std	min	25%	50%	75%	max
Precipitation, mm E-OBS	324,344	1.58	3.24	0	0	0	2.1	110.5
Temperature, C° E-OBS	324,344	6.45	9.09	-32.2	0.24	6.61	14.12	27.31
Actual Evapotranspiration, mm GLEAM	192,860	1.31	1.10	-0.14	0.351	0.974	2.114	6.63
Potential Evapotranspiration, mm GLEAM	192,860	1.77	1.59	-0.15	0.376	1.268	2.929	9.553

of visible karst features at the surface (Fig. 5). However, evidence of ancient palaeokarst structures, including breccia-filled pipes and carbonate debris, indicates past dissolution and collapse processes (Narbutas, 1979; Paukštys and Narbutas, 1996). Variations in level and chemical composition of groundwater are determined by the rate of infiltration of precipitation (Satkūnas et al., 2007, 2006; Taminskas et al., 2020).

Dubnik regional stage formations in Lithuania and Latvia host well-developed gypsum successions. These gypsiferous strata overlie dolomites and marls of the Plavinas Formation, which rest upon a thin (2–9 m) interval of clays and marls. Beneath this, thick sandstone and argillaceous sandstone packages of the Sventoji and Upninkai formations constitute a major regional aquifer system. At the base of the gypsum-karst succession, clays of the Narva Formation act as a regional aquitard, restricting vertical groundwater flow (Narbutas, 1979; Paukštys and Narbutas, 1996).

In this study three-stage regional scheme is derived from the stratigraphy of D_3fr (Upper Devonian Frasnian) (Lukševičs et al., 2012). The Dubnik Stage comprises the Late Devonian gypsum units—Tatula Formation in Lithuania (Pasvalys (D_3tps), Kirdonys (D_3tkd), Nemunėlis (D_3tnm)) and Salaspils Formation in Latvia. The Daugava Stage corresponds to the Istras Formation in Lithuania and the Daugava Formation in Latvia. The Plavinas Stage is represented regionally by the dolomitic Pļaviņas/Pliavinias Formation in both countries (Fig. 5).

The active gypsum-karst belt shows dense dissolution features. In Lithuania, ~1100 sinkholes are mapped across 400 km^2, with local densities > 20 km^{-2} and a maximum of 200 km^{-2} in the Karajimiškis Geological Reserve (Paukštys and Narbutas, 1996; Samalavičius et al., 2025b). In Latvia, Skaistkalne averages 13 km^{-2} with peaks to 138 km^{-2}, while Baldone, Ādaži, and Saulkalne register lower densities (≈2–5 km^{-2}) (Paukštys and Narbutas, 1996).

Sinkholes in the Gypsum Karst Region are predominantly small to medium-sized dolines, 1–2 to 10–15 m in diameter and 1–6 m deep. Most remain dry, though some are intermittently water-filled. The deepest structures commonly intersect the phreatic surface of the unconfined aquifer and, locally, the potentiometric surface of confined (artesian) systems, thereby inducing spring discharge, particularly along river valleys (Paukštys and Narbutas, 1996).

Karst lakes produced by the coalescence of dolines are characteristic of the region. In Lithuania, Lake Ilgasis (Kirkilai) comprises ~30 hydraulically connected sinkholes that form an elongate basin ~1.1 km long and ~0.2 km wide (Kilkus, 1977; Paukštys and Narbutas, 1996). These water bodies receive mixed meteoric and groundwater recharge and commonly exhibit faint hydrogen-sulfide odors reflecting the ambient groundwater geochemistry (Paukštys and Narbutas, 1996).

The sinkholes often correspond with surface depressions on the bog, though some remain undetectable without geophysical methods (Džeriņš et al., 2023). Active karst processes drive peat subsidence, forming sinkholes with disrupted or concave peat layering, highlighting the need for further geohazard assessments in other peatland areas with potential karst activity (Džeriņš et al., 2023).

Monthly sinkhole counts in Lithuania (2003–2024) show a low background rate punctuated by sharp bursts, mostly in early spring, with peaks up to 40 events (Fig. 6a). Aggregated by month, occurrences are overwhelmingly concentrated in April, with smaller rises in March and September–October; February and July are minima (Fig. 6b). The pattern indicates hydrologic control: snowmelt and spring rains rapidly raise heads and infiltration, triggering collapses; weaker autumn peaks align with rainfall recharge, while frozen ground (winter) and low water tables plus high evapotranspiration (summer) suppress activity. No clear long-term trend is evident—hazard appears event-driven.

3.3. Hydrogeology

The Lithuania Latvia transboundary area karst region forms part of the Baltic Artesian Basin, within an active groundwater circulation zone, where freshwater (< 1 g/L) is a dominant. However, karst region is unique due to gypsum dissolution leading to elevated TDS (up to 3 g/L). The hydrogeological system consists of multiple interconnected aquifers spanning both Devonian and

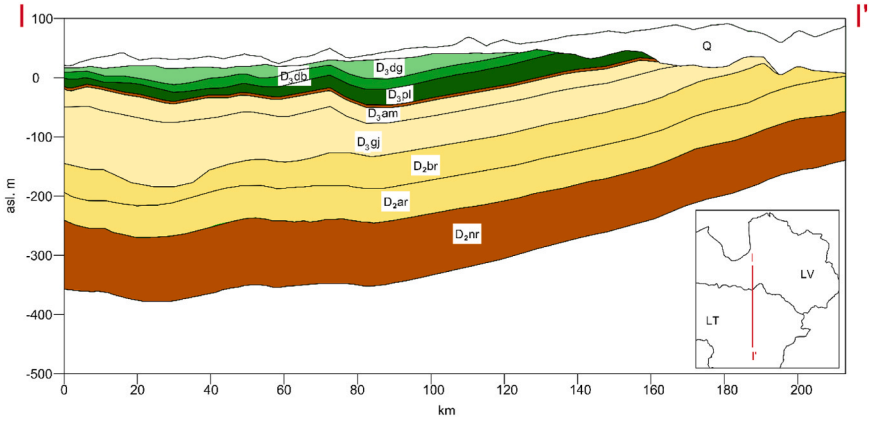

Fig. 5. Cross-section of karst region. In green shades – karst affected layers. Dark brown – aquitards, light brown aquifers. With regional stages (Spalvins et al., 2013; Virbulis et al., 2013).

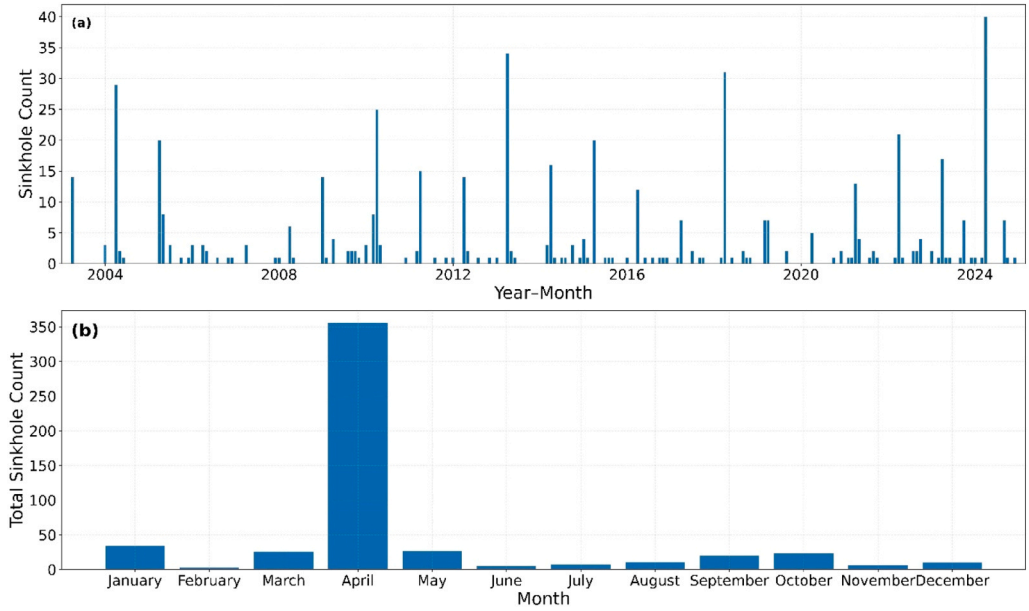

Fig. 6. Lithuania sinkhole dataset—(a) monthly counts, 2003–2024; (b) monthly totals.

Quaternary formations (Fig. 5), which are utilized for domestic and industrial water supply.

Time series (2003–2024) of TWS and GWS anomalies for the Lithuania–Latvia transboundary area display a pronounced, repeatable annual cycle (Fig. 7). TWS exhibits the larger seasonal amplitude and typically reaches its maximum in late winter–early spring; GWS covaries but peaks later, with a systematic lag of ~1–2 months. Both records show interannual variability—a relatively drier phase in the mid-2000s and higher peaks around 2018–2021—yet no clear monotonic long-term trend (Fig. 7). The seasonal pattern reflects cold-season accumulation and spring recharge from snowmelt and precipitation, followed by summer drawdown driven by evapotranspiration. The consistent TWS to GWS phase lag indicates delayed percolation through the vadose zone and aquifer response times. Year-to-year swings represent wet and dry years rather than a persistent increase or decline in storage, underscoring event- and climate-variability control on regional water availability (Fig. 7).

The clear phase lag (TWS leading GWS by ~1–2 months) and strong seasonality argue for usage of lagged and rolling-window features in ML models. Hydrologically, vadose percolation and aquifer response introduce memory, so past states carry predictive power for present groundwater level.

Quaternary aquifer system: Shallow glacial-sand aquifer with mainly fresh Ca–HCO_3 water (TDS 0.5–0.8 g/L); highly vulnerable to agricultural and urban pollution, with frequent elevations of nitrogen and organic contaminants (Paukštys and Narbutas, 1996).

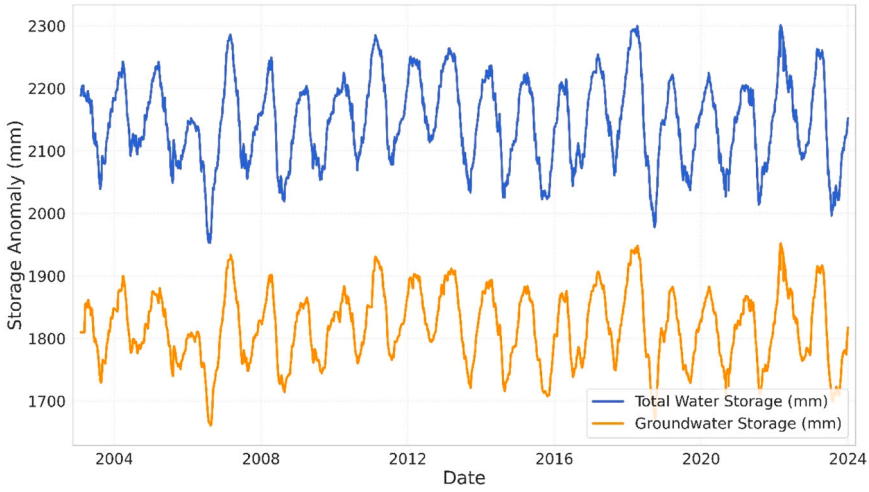

Fig. 7. Lithuania–Latvia transboundary karst area: 2003–2024 TWS and GWS show a pronounced annual cycle with GWS lagging TWS by ~1–2 months, strong interannual variability, and little long-term trend.

Daugava & Dubnik Stages: Gypsum formations yielding slightly mineralized Ca–SO_4 water (TDS 1.5–2.4 g/L); karst conduits make this aquifer extremely susceptible to contamination, and nitrogen compounds and organics often exceed drinking-water limits (Klimas and Paukštys, 1993). Plavinas Stage: Beneath the gypsum, chiefly Ca–Mg–HCO_3 water (TDS 0.5–0.8 g/L); local hydraulic links to the overlying karst introduce Ca–SO_4 water (TDS 1–2 g/L), and some wells show contamination (Paukštys and Narbutas, 1996).

4. Results and discussion

4.1. Imputation of missing groundwater level data

Before detailed interpretation, candidate algorithms were compared under identical conditions, including VIF-curated feature sets, block-masked 5-fold temporal cross-validation, and training-only preprocessing. Random Forest (RFR), Gradient Boosting (GBR), and pipelines incorporating IterativeImputer were evaluated against the Extra Trees model (ETR). Across wells and for both feature-set configurations, ETR consistently achieved higher predictive skill, with mean CV R^2 values approximately 5–10% greater and correspondingly lower MAE relative to competing approaches. IterativeImputer-based pipelines did not yield systematic improvements. On this basis, Extra Trees was selected as the primary model for subsequent imputation.

Imputation of missing groundwater-level records by ETR model was evaluated using block-based temporal cross-validation to assess both predictive performance and hydrogeological consistency. Across wells, reconstruction skill varied substantially (CV R^2 = 0.44–0.79; MAE = 0.12–0.40 m), indicating marked site-specific differences in the recoverability of groundwater dynamics from hydroclimatic and storage predictors. To examine the controlling factors behind this variability, feature attribution was analysed using SHAP diagnostics for the best-performing Extra Trees models (Fig. 8).

SHAP beeswarm plots reveal that same-day GLDAS groundwater storage (*gws_mm_tavg_gldas*) is the dominant predictor in both feature-set configurations, with higher storage values systematically associated with increased predicted groundwater levels. In the *no_seasonal* specification (Fig. 8a), importance is distributed across contemporaneous storage, 90–360-day storage lags, and multi-month precipitation aggregates, whereas temperature lags exhibit comparatively weak and inconsistent effects. When an explicit

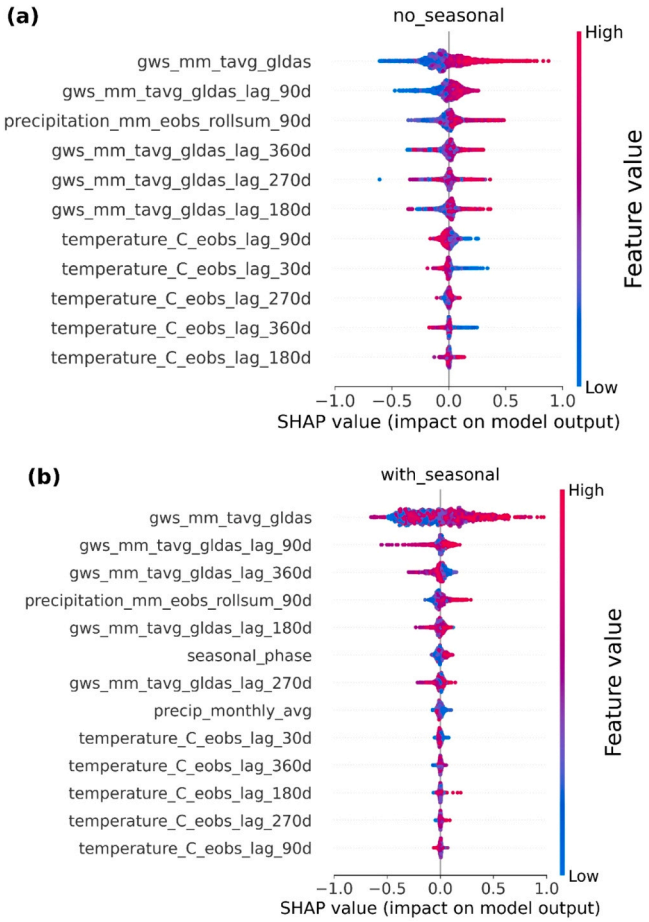

Fig. 8. SHAP feature attribution for groundwater-level imputation. Beeswarm plots show tuned ExtraTrees models (a) no_seasonal and (b) with_seasonal. Horizontal position = SHAP value (impact on predicted head); color = feature value (blue = low, pink = high).

seasonal term is included (Fig. 8b), attribution concentrates more strongly on present storage state and seasonal phase, and the contribution of long storage lags diminishes, indicating that the seasonal encoding absorbs part of the cyclic signal previously captured through lagged predictors.

Dominance of the GWS state, together with appreciable weights on 3–12 month lags, indicates strong state persistence in groundwater levels characteristic of karst systems with dual-porosity storage and delayed epikarst and matrix drainage. These findings are consistent with established representations of dual-porosity karst systems, where delayed epikarst and matrix drainage generate multi-month memory and attenuated recharge signals (Hartmann et al., 2012; Jódar et al., 2020; Martín-Rodríguez et al., 2023).

While attribution patterns indicate a broadly consistent hydroclimatic control across sites, imputation performance varies substantially between wells (Table 2). Cross-validated R^2 ranges from 0.44 to 0.79, suggesting that storage persistence alone does not fully determine reconstruction skill. Wells exhibiting smoother seasonal oscillations and gradual recessions tend to achieve higher predictive accuracy, whereas those characterized by abrupt drawdowns or short-timescale variability show reduced recoverability from hydroclimatic and storage predictors.

Across wells, optimal Extra Trees configurations favored deep trees (max_depth = None or 20) and small leaf sizes (min_samples_leaf = 1–2), indicating the importance of capturing nonlinear thresholds and sharp transitions typical of karst dynamics. Ensemble averaging maintained variance control despite model depth, consistent with the observed cross-validated stability.Across blocked CV, the best Extra-Trees setups favored very deep trees (*max_depth = None* or *20*), minimal node constraints (*min_samples_leaf = 2, min_samples_split = 2*), and moderate ensemble sizes (*n_estimators = 100–200*) (Table 2). In Extra-Trees, random split thresholds decorrelate trees and act as built-in regularization, so added depth reduces bias without a matching rise in variance (Geurts et al., 2006). Such flexibility is particularly relevant in karst systems, where recharge-discharge dynamics often exhibit sharp transitions associated with conduit activation and variable hydraulic connectivity (Hartmann et al., 2014b, 2012; Jódar et al., 2020). Ensemble sizes of 100–200 trees were sufficient to stabilize cross-validated performance, in line with established findings on diminishing marginal gains from larger forests (Probst et al., 2019), and the overall configuration yields a low-bias, variance-controlled fit consistent with the cross-validated gains in Table 2.

Hydrogeologically, the wide spread in metrics is consistent with the intrinsic heterogeneity of karst systems. Wells strongly connected to conduit networks typically exhibit rapid, event-driven responses and step-like drawdowns, while matrix-dominated settings show smoother, more diffusive behavior (Jódar et al., 2020; Martín-Rodríguez et al., 2023). Elevated errors and reduced CV R^2 are compatible with a noisy groundwater-level signal generated by high-frequency recharge pulses, variable boundary conditions (e.g., stream–aquifer exchange) or episodic pumping. The well LV_22610 results, in particular, align with disturbance by groundwater abstraction or other anthropogenic operations that introduce abrupt, nonseasonal fluctuations not captured by meteorological predictors alone.

Improved average performance under the with_seasonal specification indicates a pronounced seasonal component at several sites; however, limited gains or increased variance at others point to dominance of nonseasonal variability and nonstationarity (e.g., intermittent pumping, transient hydraulic connectivity). The elevated fold-to-fold variability at LT_35996 (CV R^2 SD = 0.19) further supports temporal instability in controlling processes (Table 2).

Short term GWL noise has previously been identified as a limiting factor in data-driven modeling approaches (Bo et al., 2025). The imputation results in this study (R^2 from 0.44 to 0.79) for masked windows of 7–60 days is therefore consistent with the challenges imposed by high-frequency dynamics. The groundwater-level series were used in their raw form, without smoothing or filtering, and no autoregressive groundwater lags were included and the reconstruction relied exclusively on hydroclimatic and remotely sensed storage predictors. A study within the Baltic countries focusing solely on GWL gap imputation likewise did not include any lagged or smoothed GWL values but instead used all available GWL observations from other sites as inputs, achieving median Nash–Sutcliffe efficiencies (NSE) of 0.58–0.76 for 24–63-day gaps and substantially higher accuracy for shorter gaps (median NSE = 0.92 for mean gap length of 7.8 days), achieving median Nash–Sutcliffe efficiencies (NSE) of 0.58–0.76 for 24–63-day gaps and substantially higher

Table 2
Classification performance for predicting monthly sinkhole risk using alternative feature sets.

Well No	CV R^2 mean	CV R^2 std	CV MAE mean	CV MAEstd	Feature set	Best Hyperparameters
LT_35994	0.79	0.08	0.16	0.01	with_seasonal	{'max_depth': None, 'min_samples_leaf': 2, 'min_samples_split': 2, 'n_estimators': 100}
LV_12225	0.67	0.08	0.12	0.01	no_seasonal	{'max_depth': 20, 'min_samples_leaf': 2, 'min_samples_split': 2, 'n_estimators': 200}
LT_35995	0.64	0.08	0.12	0.01	no_seasonal	{'max_depth': 20, 'min_samples_leaf': 2, 'min_samples_split': 2, 'n_estimators': 100}
LT_220	0.58	0.13	0.17	0.02	no_seasonal	{'max_depth': None, 'min_samples_leaf': 2, 'min_samples_split': 2, 'n_estimators': 200}
LV_22606	0.58	0.13	0.16	0.01	with_seasonal	{'max_depth': None, 'min_samples_leaf': 2, 'min_samples_split': 2, 'n_estimators': 200}
LV_22610	0.49	0.09	0.4	0.03	no_seasonal	{'max_depth': None, 'min_samples_leaf': 2, 'min_samples_split': 2, 'n_estimators': 100}
LT_35996	0.44	0.19	0.16	0.02	no_seasonal	{'max_depth': 20, 'min_samples_leaf': 2, 'min_samples_split': 2, 'n_estimators': 200}
AVERAGE	0.6	0.11	0.18	0.02		

accuracy for shorter gaps (median NSE = 0.92 for mean gap length of 7.8 days) (Bikše et al., 2023).

Overall, the large inter-well differences in CV R^2 and CV MAE reflect varying mixtures of diffuse and versus conduit flow, differing levels of anthropogenic influence, and site-specific noise characteristics that modulate the recoverability of groundwater dynamics from hydroclimatic and storage predictors.

To quantify the influence of short-timescale variability on imputation performance, groundwater-level noisiness was characterized using the standard deviation of first differences (Eqs. (1), (2))

$$\Delta h_t = \Delta h_t = \quad h_t \quad - \quad h_{t-1}, \quad t \quad = \quad 2, \ldots, N \tag{1}$$

The dispersion of these increments defines the noise metric:

$$\sigma_{\Delta h} = \quad sqrt\left(\frac{1}{N-1}\sum_{t=2}^{N}(\Delta h_t - \quad \overline{\Delta h})^2\right) \tag{2}$$

where: h_t – groundwater level at time index ttt (meters, relative to the chosen datum), Δh_t – first difference, between consecutive observations (meters per sampling interval), N – total number of observations in the groundwater-level series, $\overline{\Delta h}$ – sample mean of the first differences, $\sigma_{\Delta h}$ – sample standard deviation of the first differences, used here as a metric of high-frequency variability (meters per sampling interval).

Larger $\sigma_{\Delta h}$ reflects greater high-frequency variability and step-like disturbances, associated with recharge pulses, stream–aquifer exchanges, or episodic pumping.

A negative linear trend (Fig. 9) albeit with a broad confidence intervals due to the limited number of wells, indicates that increasing signal noise is associated with reduced model performance. Wells exhibiting larger standard deviation of first differences ($\sigma_{\Delta h}$) consistently return lower mean cross-validated R^2, implying diminished recoverability of GWL dynamics from the available predictors.

Several hydrogeological and operational mechanisms can generate the observed noise. Conduit-dominated flow in karst produces rapid, event-scale fluctuations; stream–aquifer exchanges and introduce short-term variability; and episodic pumping impose step changes not represented by meteorological predictors. Measurement artifacts—such as venting issues, sensor drift, or irregular sampling—can further elevate the first-difference variance.

The observed degradation of performance with increasing noise is consistent with the model's reliance on low- to mid-frequency signals (seasonality, multi-month recharge integration, and state persistence). When high-frequency disturbances dominate, predictor–response alignment weakens, and cross-validated R^2 declines despite moderate absolute errors.

Reconstruction accuracy is highest in wells with smooth, low-noise groundwater dynamics (LT_35994, LV_22606, LV_12225; Fig. 10). Performance degrades in wells dominated by short-timescale variability and abrupt drawdowns (LV_22610; Fig. 10 g), consistent with reduced predictor–response coherence under conduit-driven recharge, stream–aquifer interactions, or episodic pumping.

Several wells (LT_220, LT_35996, LT_35995) exhibit regime shifts and nonstationary baselines, visible as level offsets, amplitude changes, and variance breaks, with markedly noisier signals after 2014 (Fig. 10 a, c, d). These discontinuities likely reflect instrumentation changes, datum re-referencing, maintenance, or well rehabilitation. When cross-validation spans such regime boundaries, distribution shifts reduce model performance even if absolute errors remain moderate. This highlights the need for homogenization and metadata-informed preprocessing—such as offset correction, variance stabilization, regime segmentation, or intervention indicators—to suppress instrumentation artifacts and better isolate the hydroclimatic signal governing imputation.

Similar process-guided variable screening has proven effective in karst soil-moisture modeling (Zhu et al., 2026). The present study

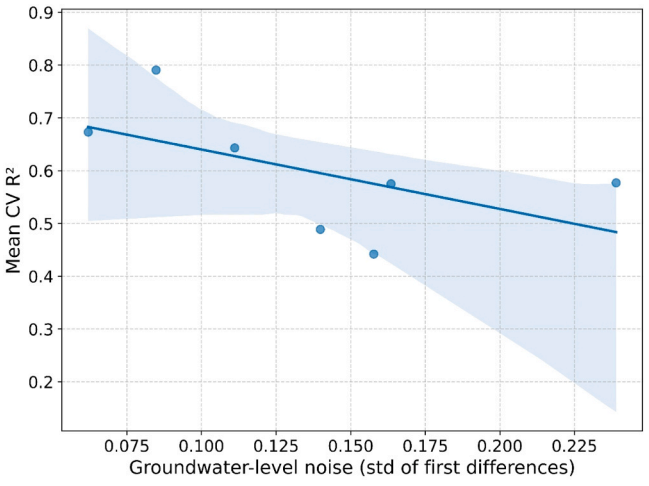

Fig. 9. Relationship between groundwater-level noise ($\sigma_{\Delta h}$, standard deviation of first differences) and mean cross-validated R^2 across wells.

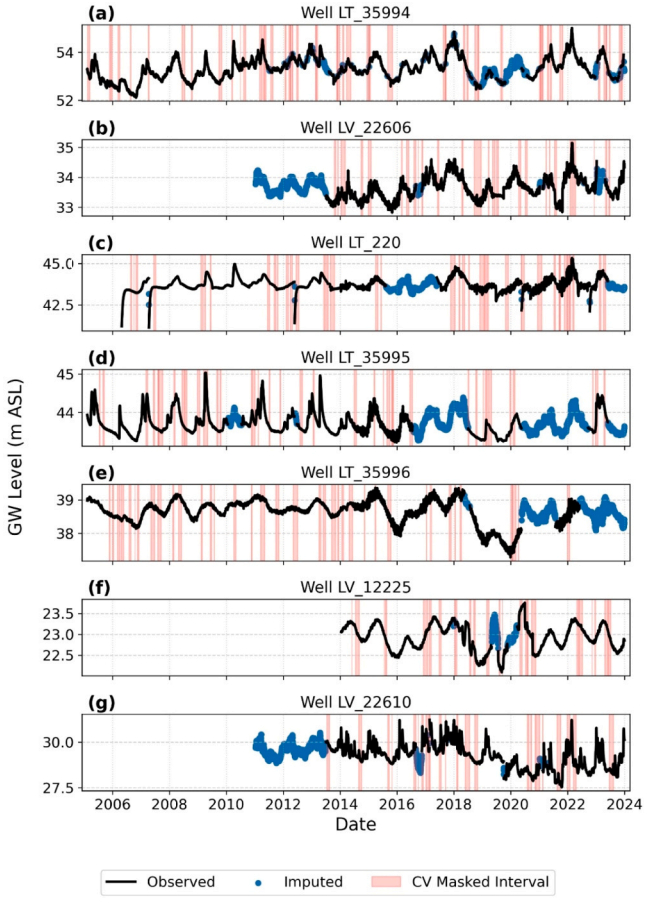

Fig. 10. Groundwater level time series (black) with imputed values (blue) and masked validation windows (pink) for a randomly selected fold of the cross-validation. Each panel (a–g) corresponds to one monitoring well (LT_35994, LV_22606, LT_220, LT_35995, LT_35996, LV_12225, LV_22610).

extends this framework to groundwater-level reconstruction by explicitly quantifying storage memory, seasonal structure, and high-frequency noise effects in saturated-zone dynamics.

4.2. Sinkhole formation risk assesment

4.2.1. Risk threshold definition and feature importance

Sinkholes in evaporate karst terrain are promoted by rapid GWL rise, intense rainfall or snowmelt, and human-induced drainage (Gutiérrez et al., 2008, 2007; Lucha et al., 2008). Because collapses may occur in clusters, a small monthly mean does not necessarily

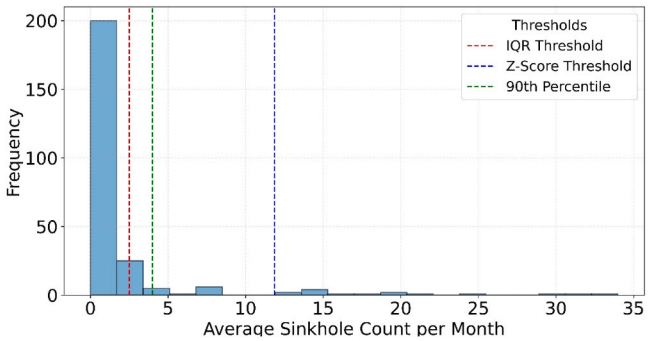

Fig. 11. Histogram of the average sinkhole count per month across all sites and years. Vertical dashed lines denote candidate anomaly thresholds: IQR-based (red), empirical 90th percentile (green), and z-score–based (blue), illustrating how different criteria select progressively rarer, more extreme months.

indicate low risk to infrastructure, agriculture, or settlements; consequently, a clear threshold is required to separate routine variability from months with conditions unusually favorable for multiple collapses. The ≥ 4 sinkholes month^{-1} threshold refers to the regional monthly total for the Lithuanian gypsum-karst area (national sinkhole inventory), rather than a per-well or local monitoring-area rate. The histogram of average monthly counts exhibits a zero-inflated, strongly right-skewed distribution: most months register 0–1 events, with a long tail extending beyond 30 (Fig. 11). Among candidate cutoffs, the interquartile-range rule (~2–3) is overly sensitive and prone to false alarms, whereas a z-score criterion (>≈11) is too permissive and would miss practically significant clusters. A percentile-based criterion is preferable: the 90th percentile occurs at approximately four sinkholes per month, indicating that 90% of months fall at or below this level (Fig. 18). The 90th-percentile threshold was selected based on the empirical distribution of monthly regional sinkhole counts, where it lies between more conservative IQR-based and more extreme z-score thresholds (Fig. 11). This choice balances sensitivity and robustness, capturing anomalously active months while remaining minimally influenced by rare

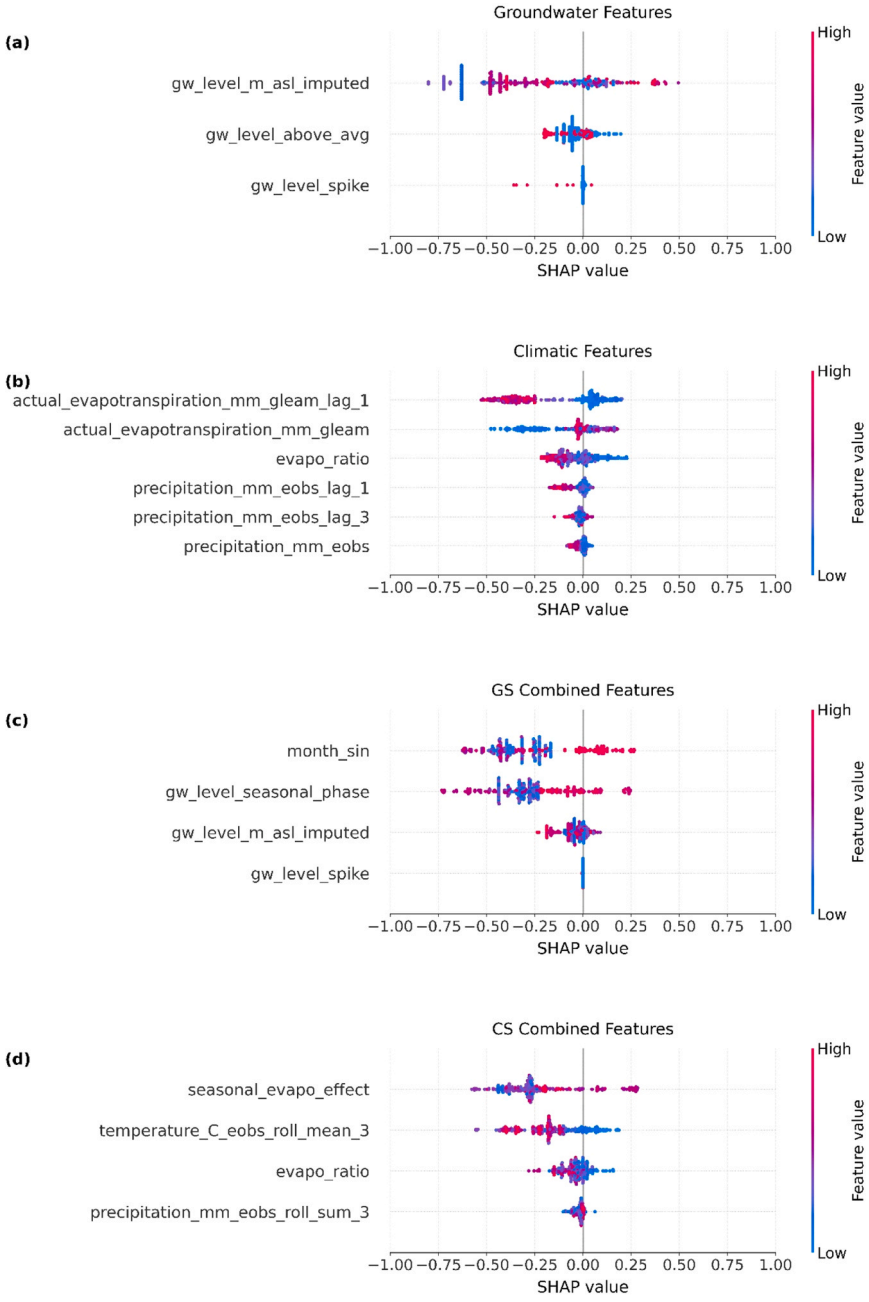

Fig. 12. SHAP summary plots for models using (a) groundwater, (b) climate, (c) groundwater + season (GS), and (d) climate + season (CS) features. Points represent individual predictions; color denotes feature value (blue = low, pink = high). Positive SHAP values increase predicted sinkhole risk.

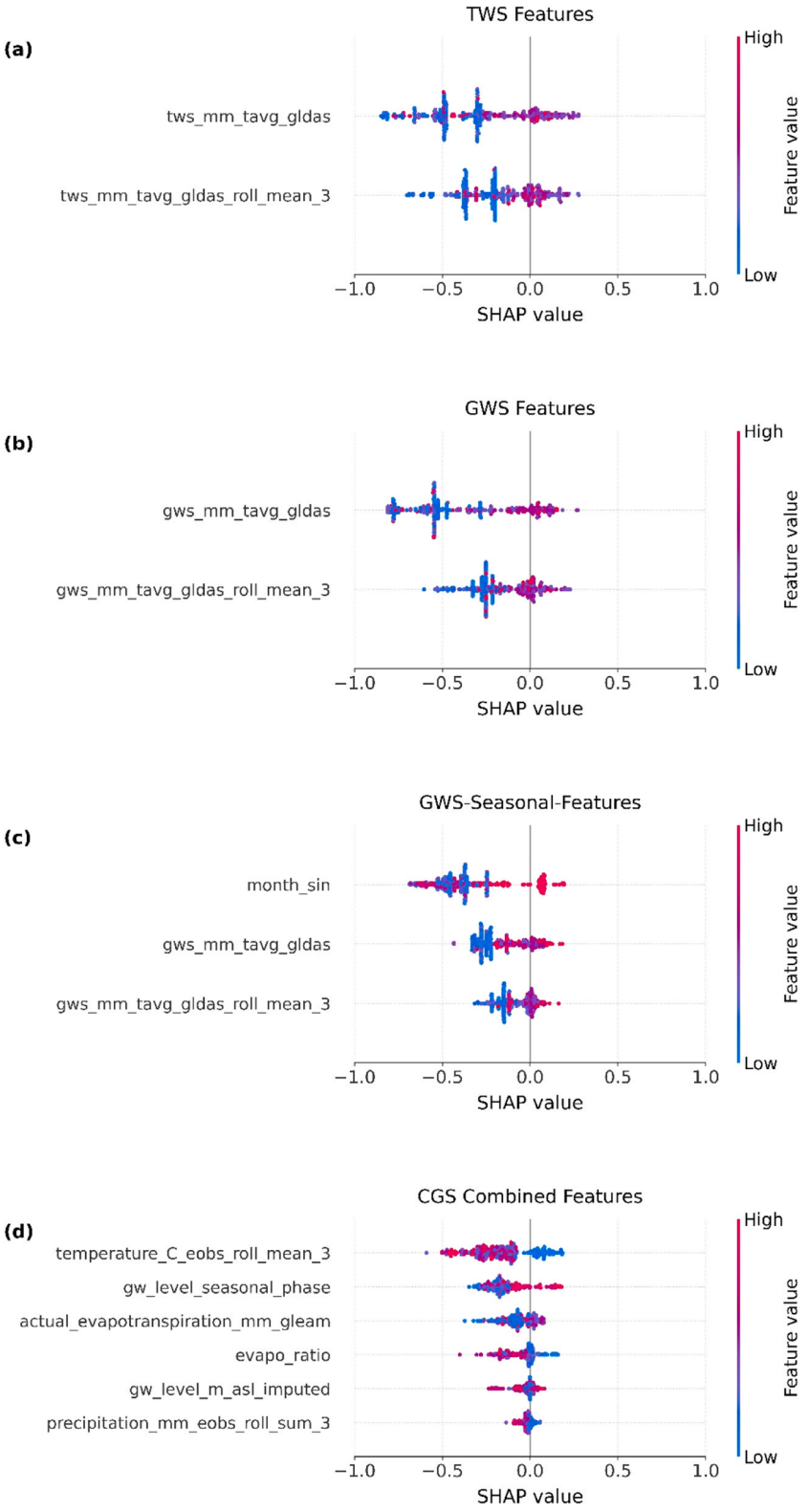

Fig. 13. SHAP summary plots for models using (a) terrestrial water storage (TWS), (b) groundwater storage (GWS), (c) GWS + seasonal terms, and (d) combined climate–groundwater–seasonal (CGS) features. Color indicates feature value (blue = low, pink = high); positive SHAP values raise predicted risk.

extremes in the heavy-tailed sinkhole distribution. Importantly, months with 0–4 sinkholes constitute a moderate occurrence that warrants routine monitoring and seasonal preparedness, with enhanced measures triggered once the monthly average exceeds four. This threshold provides consistent categorical labels for model training and evaluation and enables rigorous testing of the added value of GWL relative to climatic, TWS/GWS, and seasonal predictors.

SHAP results indicate that sinkhole hazard is governed primarily by the long-term hydraulic state of the groundwater system and its seasonal modulation. Absolute GW level and site-specific anomalies consistently dominate predictions, whereas short-lived transients have little influence (Fig. 12 a, c), consistent with earlier sinkhole susceptibility studies emphasizing hydrogeological and climatic controls over event-scale trigger, where modeling likewise emphasize geological–tectonic (fault proximity, lithology), hydrogeological (bicarbonate difference, groundwater-level change), and climatic (annual average precipitation) factors as the principal drivers of sinkhole occurrence (Bilgilioğlu et al., 2025). Elevated groundwater levels and positive anomalies increase risk, while spike indicators contribute weakly, underscoring the importance of persistence.

Climatic effects are led by actual evapotranspiration (current and lagged) and the evaporative ratio, with precipitation playing a secondary, indirect role through storage and groundwater recharge (Fig. 12 b, d). Storage metrics corroborate these patterns: higher terrestrial and groundwater storage and their multi-month means are associated with increased risk, highlighting system memory (Fig. 13 a, b). In the combined model, three-month mean temperature and groundwater seasonal phase dominate, confirming that sinkhole occurrence reflects longer-term aquifer and epikarst conditions rather than individual storm events (Fig. 13 d). This interpretation aligns closely with hydrological analyses of karst aquifer functioning with RF models link spring-discharge dynamics to geomorphological controls such as catchment area, cave density, and slope gradient (Janža et al., 2025).

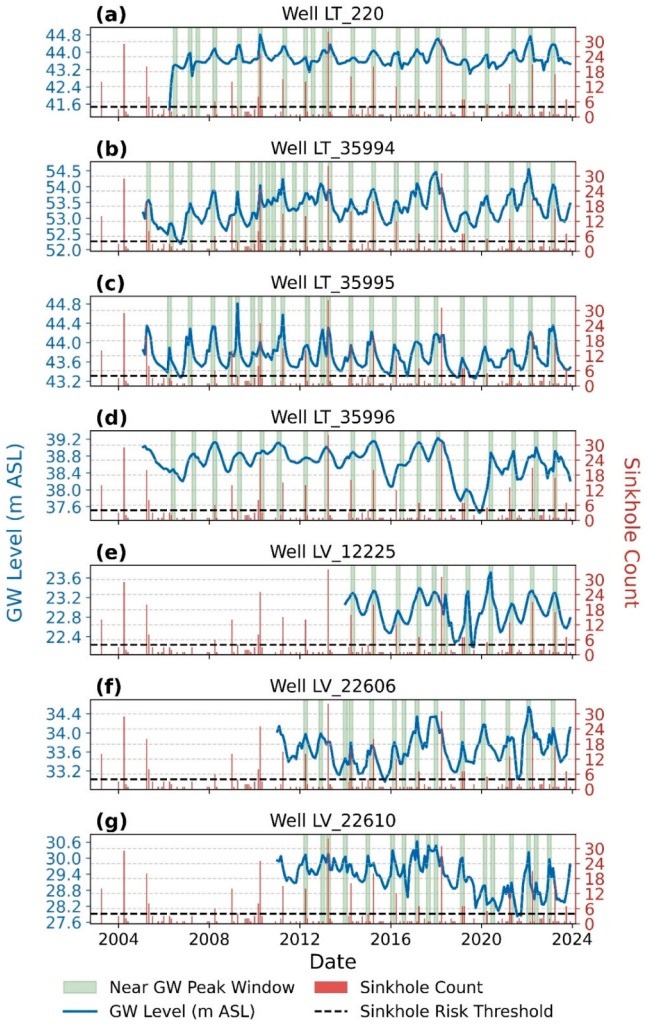

Fig. 14. Time series of groundwater level and sinkhole activity for seven monitoring wells: (a) LT_220, (b) LT_35994, (c) LT_35995, (d) LT_35996, (e) LV_12225, (f) LV_22606, and (g) LV_22610 (2003–2024). The blue line shows groundwater level (m ASL; left axis). Red bars are monthly sinkhole counts (right axis); the black dashed line is the sinkhole-risk threshold. Green bands mark "near–GW peak" windows defined as ±30 days around identified groundwater maxima.

4.2.2. Model performance assessment

Across all seven wells, the largest sinkhole counts (red bars, right axis) systematically cluster within the shaded "near-peak" windows (green), which mark ±30 days around each flagged center of the GWL peak (Fig. 14). Visual inspection shows that episodes of elevated sinkhole formation frequently coincide with—or follow within the ±30-day interval of—local maxima in GWL (blue curve, left axis). In many instances these event clusters also exceed the risk threshold (black dashed line), reinforcing a consistent temporal co-occurrence between peak GWL conditions and heightened sinkhole activity (Fig. 14).

An increase in the groundwater level from below the soil-rock interface to above will increase the degree of saturation of the soil, thereby decreasing its strength i.e. decrease in cohesion (c) and angles of internal friction (ϕ). It is important to note that the process of erosion may not be instantaneous, as the rising water can provide a temporary stabilising effect on the soil. However, a subsequent drop in that groundwater level is accompanied by the loss of buoyancy support, which may initiate the acceleration of the raveling process. Field-based evidence for drawdown-induced sinkhole formation is provided by who document widespread collapses in northeastern Thailand linked to rapid groundwater descent caused by quarry dewatering and intensive pumping (Arjwech and Everett, 2024). The presence of a large void may not be a necessary prerequisite for a cover-collapse sinkhole to form if rapid water level drawdown occurs (Alrowaimi, 2016; Byle et al., 2024; Liu et al., 2024; Öztürk et al., 2025).

This pattern suggests that groundwater peaks act as a short-term trigger for sinkhole collapse primarily through suffosion: the upward hydraulic gradient mobilizes and flushes fine overburden particles into pre-existing karst voids, undermining the cover and causing sudden surface failure. By contrast, rapid dissolution is generally too slow to explain the abrupt timing of collapse. Nevertheless, dissolution plays an essential preparatory role over longer timescales (Romanov et al., 2020): water percolation gradually dissolves carbonate or evaporite minerals, creating cavities and weakened rock frameworks. These dissolution-induced voids then provide the pathways and storage space into which soil can be suffused during groundwater rise. In other words, slow dissolution primes the subsurface, while suffosion governs the rapid collapse once groundwater gradients fluctuate. Lowering of groundwater levels may further remove buoyant support for loose fill and accelerate particle flushing, meaning that both rising and falling GWL can destabilize the system. Together, these coupled processes—long-term dissolution and short-term suffosion—explain why near-peak GWL conditions consistently align with heightened sinkhole activity. Evidence for multi-source recharge mixing and structurally guided connectivity in karst groundwater systems is shown by Yang et al. (2025), where combined ML (SOM–K-means clustering) where used to resolve triple mixing between distinct recharge contributions

The strength of the association varies among sites in ways that are hydrogeologically sensible and visible in the figure. Wells LT_35994, LT_35995, and LV_22606 exhibit sharp groundwater peaks and the tightest synchronization with elevated sinkhole counts, a pattern consistent with strong hydraulic connectivity to high-permeability conduit networks (Fig. 14 b, c, f). In contrast, LT_35996 and LV_12225 show more muted or irregular hydrographs with weaker temporal alignment to sinkhole activity, as would be expected in more diffuse or partially confined systems (Fig. 14 d, e). LT_220 and LV_22610 fall between these endmembers, showing moderate synchronization (Fig. 14 a, g). karst systems are spatially partitioned into recharge, transfer, and discharge/enrichment zones controlled by lithostratigraphy and structures (e.g., fault-guided conduit flow), which naturally produces site-to-site differences in signal coherence (Yang et al., 2025). In our case, wells more strongly linked to high-connectivity pathways show sharper peaks and tighter alignment with hazard windows.

On contrary, the case in Florida show that below-normal base water levels combined with cold temperatures increase the likelihood of widespread sinkhole formation (Aurit et al., 2013). Drop of GWL may initiate the collapse due to increased effective weight of the sediments (Gutiérrez et al., 2008; Lamoreaux and Newton, 1986). However, this seems not the case in Lithuania-Latvia karst region where elevated groundwater level clearly show the rapid increase in sinkhole formation.

The dataset is strongly imbalanced (301 low- versus 35 high-risk months), making ROC AUC and high-risk recall the most informative metrics (Table 3). Climate-only models provide the best class separability (Accuracy = 0.90; AUC = 0.91) and high precision (0.94) but moderate recall (0.52), missing many high-risk months. Groundwater-only models perform poorly (Accuracy = 0.67; AUC =

Table 3
Classification performance for predicting monthly sinkhole risk using alternative feature sets.

Feature Set	Accuracy	ROC AUC	Low Risk Count	High Risk Count		
Groundwater Features	0.67	0.61	301	35		
Climatic Features	0.90	0.91	301	35		
GS Combined Features	0.96	0.87	301	35		
CS Combined Features	0.91	0.84	301	35		
TWS Features	0.78	0.75	301	35		
GWS Features	0.78	0.77	301	35		
GWS-Seasonal-Features	0.86	0.82	301	35		
CGS Combined Features	0.93	0.86	301	35		
	Precision (Low)	**Recall (Low)**	**F1 (Low)**	**Precision (High)**	**Recall (High)**	**F1 (High)**
Groundwater Features	0.93	0.93	0.68	0.79	0.17	0.54
Climatic Features	0.97	0.97	0.92	0.94	0.52	0.71
GS Combined Features	0.98	0.98	0.98	0.98	0.85	0.80
CS Combined Features	0.98	0.98	0.92	0.95	0.54	0.80
TWS Features	0.97	0.96	0.79	0.87	0.28	0.71
GWS Features	0.96	0.96	0.79	0.86	0.28	0.71
GWS-Seasonal-Features	0.97	0.97	0.87	0.92	0.40	0.77
CGS Combined Features	0.97	0.97	0.96	0.96	0.66	0.71

0.61; recall = 0.17), indicating limited standalone predictive skill.

Adding seasonality to groundwater (GS) yields the strongest performance (Accuracy = 0.96; AUC = 0.86) and the highest high-risk recall (0.85) with minimal false alarms (precision = 0.98), indicating high early-warning potential. Climate–season models improve over climate alone but remain less sensitive. Storage-based models show intermediate skill (AUC ≈ 0.75–0.77), with recall increasing when seasonality is included. The full CGS model balances all predictors (Accuracy = 0.93; AUC = 0.86; recall = 0.66), confirming complementary information across domains.

Overall, results indicate that sinkhole occurrence is controlled less by instantaneous conditions than by seasonally modulated groundwater–climate coupling. Climate energy variables define broad susceptibility, while groundwater levels become hazardous when elevated within the seasonal window of maximum recharge. Storage metrics capture system memory but lack timing unless combined with seasonality. These findings support early-warning approaches that track groundwater anomalies in seasonal context and use climate metrics to anticipate multi-week to multi-month preconditioning.

Table 3 and the confusion matrices (Fig. 15) collectively show that models that encode seasonally modulated groundwater–climate

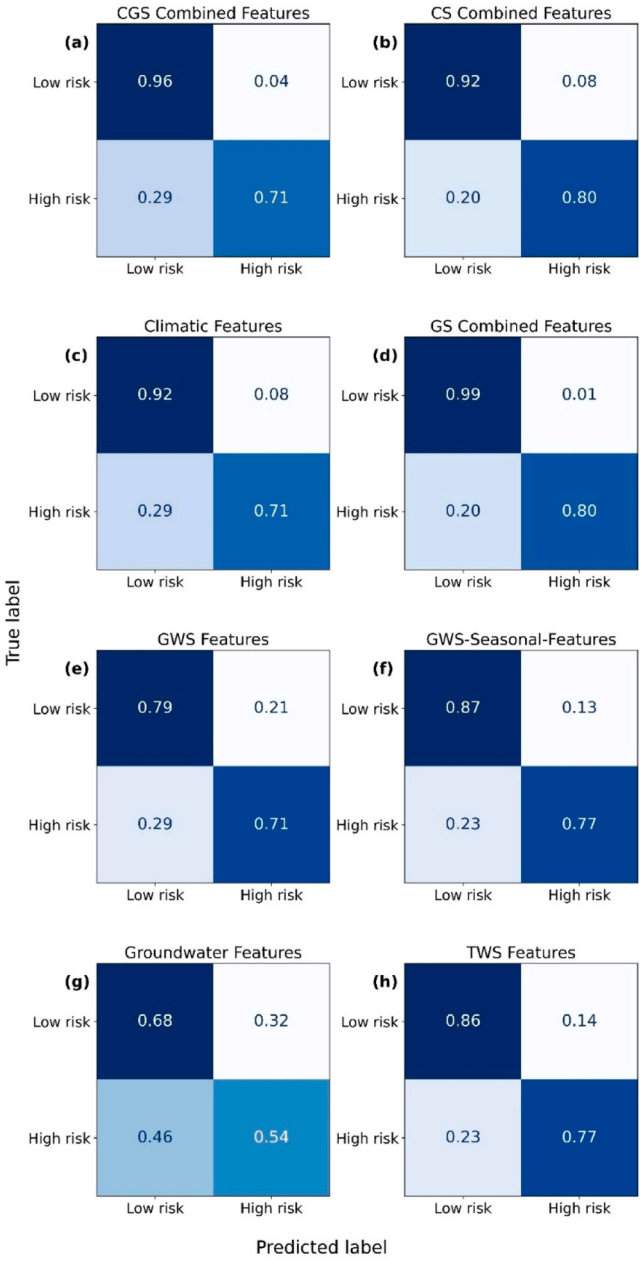

Fig. 15. Normalized confusion matrices for monthly sinkhole-risk classification using eight feature sets: (a) CGS Combined, (b) CS Combined, (c) Climatic, (d) GS Combined, (e) GWS, (f) GWS-Seasonal, (g) Groundwater, and (h) TWS. C – climatic, G – groundwater, S – seasonal features.

states outperform those based on a single data domain. The GS Combined model attains the highest overall accuracy and the best sensitivity to the high-risk class, while maintaining a very low false-alarm rate in the confusion matrix (≈1% of low-risk months misclassified; Fig. 15). The CS Combined and CGS models also perform strongly: CS retrieves most high-risk months but tolerates a slightly higher false-positive rate (Fig. 15 b), whereas CGS is more conservative, trading a small reduction in high-risk recall for fewer false positives (Fig. 15 a). By contrast, groundwater-only models show the weakest separability and the highest burden of false alarms (Fig. 15 g), and TWS/GWS-only models achieve intermediate skill, which improves when explicit seasonal terms are added (Fig. 15 e–f, h).

Well-specific model selection (Table 4) confirms spatial heterogeneity in the controlling signals. Four of seven wells are best predicted by climate-only features (ROC AUC = 0.88–0.97), two wells favor the fully combined CGS feature set (AUC = 0.91–0.93), and one well is best described by GS.

The well-level optimizations reveal a consistent pattern: high skill is obtained with deep ensembles and explicit imbalance handling and small leaf sizes. All selected models use *max_depth* = 20, effectively permitting near-unconstrained trees given the monthly sample sizes. This choice allows the forest to model strong non-linearities and higher-order interactions among hydroclimatic variables (e.g., temperature × evapotranspiration × groundwater phase) that are expected in karst systems. Model complexity is tempered by *min_samples_leaf* = 2 or 5. Wells requiring *min_samples_leaf* = 5 (e.g., LT_35995) benefited from additional smoothing, indicating that their signal-to-noise ratio is lower or that several predictors are collinear; leaves must then aggregate more observations to avoid spurious, well-specific partitions. Where *min_samples_leaf* = 2 was optimal, the data supported finer partitions without overfitting, consistent with clearer class separation (Table 4).

All best models employ large ensembles (*n_estimators* = 200–400). Increasing tree count primarily reduces variance; the jump from 200 to 400 trees was selected only for the most complex feature set (CGS at LT_35996), suggesting diminishing returns beyond ~200 trees for most wells but a measurable gain when interactions between climate, groundwater, and season are all present.

Given the pronounced class imbalance (≈1:9 high-risk to low-risk months overall), class weighting was essential. Most wells favored *class_weight* = *"balanced_subsample"*, which reweights classes inversely to their frequency within each bootstrap sample and thereby shifts split criteria to protect recall of the high-risk class while still averaging over resampled prevalence. One well (LT_35996) selected an explicit weight of {0:1, 1:6}, i.e., a milder up-weighting than the global imbalance (~8.6:1). This tuned weight trades a modest increase in false positives for a larger gain in true positives, aligning with the operational objective of not missing hazardous months. The need for explicit up-weighting at this well is consistent with site-specific attenuation of the signal (e.g., stronger pumping influence or weaker seasonal contrast), which requires a more aggressive prior toward the minority class (Table 4).

Overall, persistent hydroclimatic conditions—groundwater state and seasonal phase, basin storage, and temperature/evapotranspiration—are the primary determinants of sinkhole risk, while short-lived precipitation pulses add limited predictive value. Increase anthropogenic activity can also lead to intensified karst formation, as was demonstrated by Brinkmann et al. (2008).

4.2.3. *Advantages and limitations of the AI-based pipeline*

The proposed AI-based pipeline offers several advantages for sinkhole risk assessment in data-limited karst regions. Its primary strength lies in the integration of sparse in situ groundwater observations with satellite-derived hydroclimatic and water-storage variables, enabling both the reconstruction of near-continuous groundwater-level time series and the generation of planning-relevant sinkhole risk indicators without reliance on dense monitoring networks. The use of process-guided feature engineering and explainable ML methods allows the models to capture physically meaningful relationships, such as multi-week recharge preconditioning and groundwater-level peaks preceding sinkhole formation, thereby supporting both prediction and process understanding. The workflow is modular, transferable, and scalable, making it suitable for other karst regions with similar monitoring constraints.

Uncertainty in the proposed framework arises from three main sources: groundwater-level (GWL) reconstruction, sinkhole inventory completeness, and risk classification. In this study, uncertainty related to GWL reconstruction is limited because most

Table 4
Best-performing model by well for predicting monthly sinkhole risk. Selected by the best ROC AUC.

Well No.	Feature Set	Accuracy	ROC AUC	Best Hyperparameters
LT_220	Climatic Features	0.92	0.88	{"class_weight": "balanced_subsample", "max_depth": 20, "min_samples_leaf": 2, "n_estimators": 200}
LT_35994	Climatic Features	0.96	0.97	{"class_weight": "balanced_subsample", "max_depth": 20, "min_samples_leaf": 2, "n_estimators": 200}
LT_35995	CGS Combined Features	0.96	0.91	{"class_weight": "balanced_subsample", "max_depth": 20, "min_samples_leaf": 5, "n_estimators": 200}
LT_35996	CGS Combined Features	0.92	0.93	{"class_weight": {"0": 1, "1": 6}, "max_depth": 20, "min_samples_leaf": 2, "n_estimators": 400}
LV_12225	GS Combined Features	0.98	0.92	{"class_weight": "balanced_subsample", "max_depth": 20, "min_samples_leaf": 2, "n_estimators": 200}
LV_22606	Climatic Features	0.88	0.90	{"class_weight": "balanced_subsample", "max_depth": 20, "min_samples_leaf": 5, "n_estimators": 400}
LV_22610	Climatic Features	0.88	0.90	{"class_weight": "balanced_subsample", "max_depth": 20, "min_samples_leaf": 5, "n_estimators": 200}

monitoring wells contain long and relatively complete observation records, with gaps occurring intermittently rather than systematically. The imputation models are therefore constrained by abundant local observations and physically consistent hydroclimatic predictors. Uncertainty in the sinkhole inventory is also moderate: sinkhole formation in the Lithuanian gypsum-karst region is typically abrupt and geomorphologically distinct, occurring predominantly in agricultural land where detection likelihood is high. Some underreporting may occur in forested areas, and short delays between formation and recording may introduce temporal noise, but these effects are mitigated by monthly aggregation. Finally, uncertainty in risk classification reflects both data limitations and process complexity. Importantly, the machine-learning pipeline functions not only as a predictive tool but also as a process-exploration framework: its tuning and explainability analyses link sinkhole occurrence to hydroclimatic and groundwater dynamics, strengthening confidence in the inferred mechanisms despite inherent uncertainties.

Rather than investing in expensive, dense groundwater monitoring networks, authorities should maintain a limited number of strategically placed, high-quality observation wells and ensure their long-term continuity. When combined with satellite-derived hydroclimatic and water-storage products, even sparse in situ data can support reliable sinkhole risk assessment. Policy should therefore emphasize data continuity, sensor maintenance, and metadata documentation over spatial density.

Sinkhole management policies should move beyond static susceptibility maps toward dynamic, time-varying risk indicators that reflect seasonal groundwater and hydroclimatic conditions. Monthly or seasonal sinkhole risk bulletins—based on groundwater anomalies, recharge preconditioning, and storage trends—can support early-warning systems, agricultural planning, and infrastructure maintenance scheduling. Generative AI approaches should encompass real time data collection and model retraining when more data become available.

Models that integrate groundwater with seasonality (and, secondarily, climate) achieve the best balance of accuracy and sensitivity, whereas single-domain models perform inconsistently. Site-to-site variability is substantial, so the optimal feature mix and decision threshold should be tuned locally, with random-forest settings that allow deep, non-linear interactions and apply explicit class weighting. Operational monitoring should therefore emphasize seasonally contextualized groundwater anomalies, storage indicators, and energy-climate metrics to anticipate hazardous periods. Results are constrained by sparse wells and monthly sinkhole timing, storage product availability from 2003 onward, and instrumentation regime shifts that introduce distribution changes. Incorporating anthropogenic pressures (e.g., pumping, drainage works), richer recharge proxies, and targeted homogenization should further stabilize models and improve minority-class recall. Finally, operational deployment should use well-level tuning and uncertainty-aware alerting to match local hydrogeologic conditions.

5. Conclusions

This study highlights the importance of integrated ML pipelines for geohazard analysis in data-scarce karst regions. By combining remote-sensing products with limited in situ groundwater observations, the proposed workflow demonstrates how fragmented monitoring data can be converted into statistically meaningful indicators of sinkhole hazard.

The results show that satellite-derived hydroclimatic and water-storage information, when coupled with reconstructed groundwater states, contains sufficient signal to support categorical sinkhole risk prognosis. The pipeline structure—linking groundwater reconstruction and hazard classification—enhances consistency, interpretability, and robustness compared to isolated modeling approaches.

The main scientific value lies in demonstrating that process-informed ML pipelines can bridge the gap between sparse observations and operationally relevant geohazard assessment. This approach is transferable to other karst and data-limited regions and provides a foundation for scalable, statistically based early-warning systems when dense monitoring or deterministic modeling is not feasible.

CRediT authorship contribution statement

Gintaras Žaržojus: Writing – original draft, Visualization, Project administration, Methodology, Data curation. **Assemzhan Kunsakova:** Writing – original draft, Visualization, Data curation. **Ieva Lekstutytė:** Writing – original draft, Visualization, Data curation. **Jurga Arustienė:** Writing – original draft, Visualization, Formal analysis, Data curation. **Saulius Gadeikis:** Writing – original draft, Visualization, Formal analysis. **Inga Retiķe:** Writing – original draft, Funding acquisition, Data curation. **Sonata Gadeikienė:** Writing – original draft, Visualization, Investigation. **Jānis Bikše:** Writing – original draft, Investigation, Formal analysis, Data curation. **Ilya Zaslavsky:** Writing – original draft, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization. **Vytautas Samalavičius:** Writing – original draft, Visualization, Validation, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT (OpenAI) to help refine wording and improve clarity in short text segments (no more than five sentences per prompt), and Google Colab AI tools to assist with generating portions of the analysis code. After using these tools, the authors thoroughly reviewed, verified, and edited all generated material, and take full responsibility for the content and integrity of the published article.

Funding Information

Vilnius University has received funding from the Research Council of Lithuania (LMTLT), Agreement no. S-IMPRESSU-24-3. The GRANDE-U "Groundwater Resilience Assessment through Integrated Data Exploration for Ukraine" project (NSF Award 2409395) unites researchers from six countries - the U.S., Ukraine, Poland, Lithuania, Latvia, and Estonia. The University of Latvia receives funding from the Latvian Council of Science under the Contract no. 11-1.N-462.

Declaration of Competing Interest

The authors declare no conflict of interest.

Acknowledgements

We acknowledge the E-OBS dataset from the EU-FP6 project UERRA (http://www.uerra.eu) and the Copernicus Climate Change Service, and the data providers in the ECA&D project (https://www.ecad.eu). The GRANDE-U "Groundwater Resilience Assessment through Integrated Data Exploration for Ukraine" project (NSF Award 2409395) unites researchers from six countries - the U.S., Ukraine, Poland, Lithuania, Latvia, and Estonia. The University of Latvia receives funding from the Latvian Council of Science under the contract No. 11–1.N-462. Vilnius University has received funding from the Research Council of Lithuania (LMTLT), agreement No. S-IMPRESSU-24–3. We sincerely thank the three reviewers and the editors for their thoughtful comments and valuable guidance.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.ejrh.2026.103372.

Data availability

The data is available upon request from authors.

References

Alrabayah, O., Caus, D., Watson, R.A., Schulten, H.Z., Weigel, T., Rüpke, L., Al-Halbouni, D., 2024. Deep-learning-based automatic sinkhole recognition: application to the Eastern Dead Sea. Remote Sens. 16. https://doi.org/10.3390/rs16132264.

Alrowaimi, M., 2016. Experimental Study of Sinkhole Failure Related to Groundwater. University of Central Florida, Orlando.

Arjwech, R., Everett, M.E., 2024. Sinkhole formation induced by descending groundwater in a karst aquifer near a limestone quarry. Earth Surf. Process. Land. 49, 5027–5037. https://doi.org/10.1002/esp.6010.

Aurit, M.D., Peterson, R.O., Blanford, J.I., 2013. A GIS analysis of the relationship between sinkholes, dry-well complaints and groundwater pumping for frost-freeze protection of winter strawberry production in Florida. PLoS One 8. https://doi.org/10.1371/journal.pone.0053832.

Bianchini, S., Confuorto, P., Intrieri, E., Sbarra, P., Di Martire, D., Calcaterra, D., Fanti, R., 2022. Machine learning for sinkhole risk mapping in Guidonia-Bagni di Tivoli plain (Rome), Italy. Geocarto Int. 37, 16687–16715. https://doi.org/10.1080/10106049.2022.2113455.

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

Bilgilioğlu, S.S., Gezgin, C., Iban, M.C., Bilgilioğlu, H., Gündüz, H.I., Arslan, 2025. Explainable SInkhole Susceptibility Mapping Using Machine-learning-based SHAP: quantifying and comparing the effects of contributing factors in Konya, Türkiye. Appl. Sci. 15. https://doi.org/10.3390/app15063139.

Bo, Y., Zhang, C., Fang, X., Sun, Y., Li, C., An, M., Peng, Y., Lu, Y., 2025. Application of HP-LSTM models for groundwater level prediction in karst regions: a case study in Qingzhen City. Water 17. https://doi.org/10.3390/w17030362.

Brinkmann, R., Parise, M., Dye, D., 2008. Sinkhole distribution in a rapidly developing urban environment: Hillsborough County, Tampa Bay area, Florida. Eng. Geol. 99, 169–184. https://doi.org/10.1016/j.enggeo.2007.11.020.

Byle, M.J., Sasowsky, I.D., Rana, A.J., 2024. Cover collapse sinkhole formation is delayed in time and uncorrelated to distance from quarry in a long-term study of a karst basin. Eng. Geol. 338. https://doi.org/10.1016/j.enggeo.2024.107610.

Collenteur, R.A., Haaf, E., Bakker, M., Liesch, T., Wunsch, A., Soonthornrangsan, J., White, J., Martin, N., Hugman, R., De Sousa, E., Vanden Berghe, D., Fan, X., Peterson, T.J., Bikše, J., Di Ciacca, A., Wang, X., Zheng, Y., Nölscher, M., Koch, J., Schneider, R., Benavides Höglund, N., Krishna Reddy Chidepudi, S., Henriot, A., Massei, N., Jardani, A., Rudolph, M.G., Rouhani, A., Gómez-Hernández, J.J., Jomaa, S., Pölz, A., Franken, T., Behbooei, M., Lin, J., Meysami, R., 2024. Data-driven modelling of hydraulic-head time series: results and lessons learned from the 2022 Groundwater Time Series Modelling Challenge. Hydrol. Earth Syst. Sci. 28, 5193–5208. https://doi.org/10.5194/hess-28-5193-2024.

Cornes, R.C., van der Schrier, G., van den Besselaar, E.J.M., Jones, P.D., 2018. An ensemble version of the E-OBS temperature and precipitation data sets. J. Geophys. Res. Atmos. 123, 9391–9409. https://doi.org/10.1029/2017JD028200.

Delina, A., Babre, A., Popovs, K., Sennikovs, J., Grinberga, B., 2012. Effects of karst processes on surface water and groundwater hydrology at Skaistkalne Vicinity, Latvia. Hydrol. Res. 43, 445–459. https://doi.org/10.2166/nh.2012.123.

Džeriņš, P., 2023. Application of Electrical Resistivity Tomography for Investigation of Subsurface Karst Features (Doctoral Dissertation). University of Latvia, Riga.

Džeriņš, P., Karušs, J., Lamsters, K., Ješkins, J., Ķelpe, A., 2023. Investigation of buried karst sinkholes under a bog using ground penetrating radar (GPR) and electrical resistivity tomography (ERT). Earth Surf. Process. Land. 48, 1909–1925. https://doi.org/10.1002/esp.5593.

Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Mach. Learn. 63, 3–42. https://doi.org/10.1007/s10994-006-6226-1.

Gutiérrez, F., Cooper, A.H., Johnson, K.S., 2008. Identification, prediction, and mitigation of sinkhole hazards in evaporite karst areas. Environ. Geol. 1007–1022. https://doi.org/10.1007/s00254-007-0728-4.

Gutiérrez, F., Galve, J.P., Guerrero, J., Lucha, P., Cendrero, A., Remondo, J., Bonachea, J., Gutiérrez, M., Sánchez, J.A., 2007. The origin, typology, spatial distribution and detrimental effects of the sinkholes developed in the alluvial evaporite karst of the Ebro River valley downstream of Zaragoza city (NE Spain). Earth Surf. Process. Landforms 32, 912–928. https://doi.org/10.1002/esp.

Hartmann, A., Goldscheider, N., Wagener, T., Lange, J., Weiler, M., 2014a. Karst water resources in a changing world: review of hydrological modeling approaches. Rev. Geophys. 52, 218–242. https://doi.org/10.1002/2013RG000443.
Hartmann, A., Lange, J., Weiler, M., Arbel, Y., Greenbaum, N., 2012. A new approach to model the spatial and temporal variability of recharge to karst aquifers. Hydrol. Earth Syst. Sci. 16, 2219–2231. https://doi.org/10.5194/hess-16-2219-2012.
Hartmann, Andreas, Mudarra, M., Andreo, B., Marín, A., Wagener, T., Lange, J., 2014b. Modeling spatiotemporal impacts of hydroclimatic extremes on groundwater recharge at a mediterranean karst aquifer. Water Resour. Res. 50, 6507–6521. https://doi.org/10.1002/2014WR015685.
James, Gareth, Witten, Daniela, Hastie, Trevor, Tibshirani, Robert, 2017. An Introduction to Statistical Learning: with Applications in R. Springer: Springer Science+ Business Media.
Janža, M., Hudovernik, V., Serianz, L., Stroj, A., 2025. Modeling hydrological functioning of karst aquifer systems in Slovenia using geomorphological features and random forest algorithm. J. Hydrol. Reg. Stud. 62. https://doi.org/10.1016/j.ejrh.2025.102774.
Jemeļjanova, M., Collenteur, R.A., Kmoch, A., Bikše, J., Popovs, K., Kalvāns, A., 2023. Modeling hydraulic heads with impulse response functions in different environmental settings of the Baltic countries. J. Hydrol. Reg. Stud. 47, 101416. https://doi.org/10.1016/J.EJRH.2023.101416.
Jódar, J., González-Ramón, A., Martos-Rosillo, S., Heredia, J., Herrera, C., Urrutia, J., Caballero, Y., Zabaleta, A., Antigüedad, I., Custodio, E., Lambán, L.J., 2020. Snowmelt as a determinant factor in the hydrogeological behaviour of high mountain karst aquifers: the Garcés karst system, Central Pyrenees (Spain). Sci. Total Environ. 748. https://doi.org/10.1016/j.scitotenv.2020.141363.
Jourde, H., Wang, X., 2023. Advances, challenges and perspective in modelling the functioning of karst systems: a review. Environ. Earth Sci. 82. https://doi.org/10.1007/s12665-023-11034-7.
Kilkus, K., 1977. Karst of North Lithuania. Karst Lakes. Proc. Acad. Sci. Lith. 103, 125–133.
Klimas, A., Paukštys, B., 1993. Nitrate contamination of groundwater in the Republic of Lithuania. Geol. Surv. Nor. Bull. 75–85.
Klizas, P., Šečkus, R., 2007. Filtration and geoelectrical investigations in the karst region of North Lithuania. Geologija 59, 77–81.
Koit, O., Retiķe, I., Bikše, J., Terasmaa, J., Tarros, S., Abreldaal, P., Babre, A., Hunt, M., Pärn, J., Vainu, M., Marandi, A., Sisask, K., Lode, E., Männik, M., 2023. Hydrochemical signatures of springs for conceptual model development to support monitoring of transboundary aquifers. Groundw. Sustain. Dev. 21. https://doi.org/10.1016/j.gsd.2023.100927.
Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World map of the Köppen-Geiger climate classification updated. Meteorol. Z. 15, 259–263. https://doi.org/10.1127/0941-2948/2006/0130.
Lamoreaux, P.E., Newton, J.G., 1986. Catastrophic subsidence: an environmental hazard, shelby county, Alabama. Environ. Geol. Water Sci. 8, 25–40. https://doi.org/10.1007/BF02525556.
Levins, I., Buzajevs, V., 1999. Groundwater Vulnerability Map of Latvia, scale 1:500,000. State Geological Survey (Nr. 12074). State Geological Archive, Riga, Latvia.
Li, B., Beaudoing, H., Rodell, M., 2020. GLDAS catchment land surface model L4 daily 0.25° × 0.25°. GRACE-DA1 V2. https://doi.org/10.5067/TXBMLX370XX8.
Li, B., Rodell, M., Kumar, S., Beaudoing, H.K., Getirana, A., Zaitchik, B.F., de Goncalves, L.G., Cossetin, C., Bhanja, S., Mukherjee, A., Tian, S., Tangdamrongsub, N., Long, D., Nanteza, J., Lee, J., Policelli, F., Goni, I.B., Daira, D., Bila, M., de Lannoy, G., Mocko, D., Steele-Dunne, S.C., Save, H., Bettadpur, S., 2019. Global GRACE data assimilation for groundwater and drought monitoring: advances and challenges. Water Resour. Res. 55, 7564–7586. https://doi.org/10.1029/2018WR024618.
Linares, R., Roqué, C., Gutiérrez, F., Zarroca, M., Carbonel, D., Bach, J., Fabregat, I., 2016. The impact of droughts and climate change on sinkhole occurrence. A case study from the evaporite karst of the Fluvia Valley, NE Spain. Sci. Total Environ. https://doi.org/10.1016/j.scitotenv.2016.11.091.
Lithuanian Geological Survey under the Ministry of Environment, 2025. Boundaries of Lithuanian Karst Regions (WMS viewing service) [Data set].
Liu, X., Chen, H., Gong, B., Jiang, G., Wang, J., 2024. Formation process of cover collapse sinkholes related to groundwater level decline in karst areas. J. Mt. Sci. 21, 3832–3846. https://doi.org/10.1007/s11629-024-8944-x.
Lucha, P., Cardona, F., Gutiérrez, F., Guerrero, J., 2008. Natural and human-induced dissolution and subsidence processes in the salt outcrop of the Cardona Diapir (NE Spain). Environ. Geol. 1023–1035. https://doi.org/10.1007/s00254-007-0729-3.
Lukševičs, E., Stinkulis, Ģ., Mūrnieks, A., Popovs, K., 2012. Geological evolution of the Baltic Artesian Basin. In: Dēliņa, A., Kalvāns, A., Saks, T., Bethers, U., Vircavs, V. (Eds.), Highlights of Groundwater Research in the Baltic Artesian Basin. University of Latvia, pp. 7–52.
Martín-Rodríguez, J.F., Mudarra, M., De la Torre, B., Andreo, B., 2023. Towards a better understanding of time-lags in karst aquifers by combining hydrological analysis tools and dye tracer tests. Application to a binary karst aquifer in southern Spain. J. Hydrol. 621. https://doi.org/10.1016/j.jhydrol.2023.129643.
Miralles, D.G., Bonte, O., Koppa, A., Baez-Villanueva, O.M., Tronquo, E., Zhong, F., Beck, H.E., Hulsman, P., Dorigo, W., Verhoest, N.E.C., Haghdoost, S., 2025. GLEAM4: global land evaporation and soil moisture dataset at 0.1 resolution from 1980 to near present. Sci. Data 12. https://doi.org/10.1038/s41597-025-04610-y.
Narbutas, V., 1979. Geological pecularities of Lithuanian karst region and problems of its protection [in lithuanian]. Ann. Geogr. 17, 155–166.
Öztürk, M.Z., Poyraz, M., Duman, H., Taşoğlu, E., 2025. A geospatial approach to understanding sinkhole formation in Akgöl Wetland, Türkiye. Environ. Earth Sci. 84. https://doi.org/10.1007/s12665-025-12225-0.
Paukštys, B., Cooper, A.H., Arustienė, J., 1999. Planning for gypsum geohazards in Lithuania and England. Eng. Geol. 52, 93–103. https://doi.org/10.1016/S0013-7952(98)00061-1.
Paukštys, B., Narbutas, V., 1996. Gypsum karst of the Baltic Republics. Int. J. Speleol. 25, 279–284. https://doi.org/10.5038/1827-806x.25.3.21.
Probst, P., Wright, M.N., Boulesteix, A.L., 2019. Hyperparameters and tuning strategies for random forest. Wiley Interdiscip. Rev. Data Min. Knowl. Discov. https://doi.org/10.1002/widm.1301.
Ramirez, S.G., Williams, G.P., Jones, N.L., 2022. Groundwater level data imputation using machine learning and remote Earth observations using inductive bias. Remote Sens. 14. https://doi.org/10.3390/rs14215509.
Retike, I., Delina, A., Bikse, J., Kalvans, A., Popovs, K., Pipira, D., 2016. Quaternary groundwater vulnerability assessment in Latvia using multivariate statistical analysis. Res. Rural Dev. 1, 210–215.
Rimkus, E., Briede, A., Jaagus, J., Stonevicius, E., Kilpys, J., Viru, B., 2018. Snow-cover regime in Lithuania, Latvia and Estonia and its relationship to climatic and geographical factors in 1961-2015. Boreal Environ. Res. 23, 193–208.
Rimkus, E., Justas, K., Bukantis, A., Krotovas, A., 2011. Temporal variation of extreme precipitation events in Lithuania. Oceanologia 53, 259–277. https://doi.org/10.5697/oc.53-1-TI.259.
Rohde, M.M., Biswas, T., Housman, I.W., Campbell, L.S., Klausmeyer, K.R., Howard, J.K., 2021. A machine learning approach to predict groundwater levels in california reveals ecosystems at risk. Front. Earth Sci. 9. https://doi.org/10.3389/feart.2021.784499.
Romanov, D., Kaufmann, G., Al-Halbouni, D., 2020. Basic processes and factors determining the evolution of collapse sinkholes – a sensitivity study. Eng. Geol. 270. https://doi.org/10.1016/j.enggeo.2020.105589.
Rudzianskaitė, A., 2017. Investigation of water quality in the agricultural area of Lithuanian Karst region. Cave Investig. https://doi.org/10.5772/intechopen.68536.
Samalavičius, V., Gadeikienė, S., Žaržojus, G., Gadeikis, S., Lekstutytė, I., 2025a. Oxygen-18 prediction using machine learning in the Baltic Artesian Basin groundwater. Stoch. Environ. Res. Risk Assess. 39, 765–787. https://doi.org/10.1007/s00477-024-02896-9.
Samalavičius, V., Vanhala, E.K.-M., Lekstutytė, I., Gadeikienė, S., Gadeikis, S., Žaržojus, G., 2024. Hydraulic conductivity determination of Lithuanian soils using machine learning. Baltica 137–150. https://doi.org/10.5200/baltica.2024.2.5.
Samalavičius, V., Žaržojus, G., Kunsakova, A., Arustienė, J., 2025. Lithuanian Karst Region Hydrogeology: Available Data and Future Research Prospects. https://doi.org/10.5194/egusphere-egu25-11761.
Satkūnas, J., Marcinkevičius, V., Mikulėnas, V., Taminskas, J., 2007. Rapid development of karst landscape in North Lithuania – monitoring of denudation rate, site investigations and implications for management. GFF 129, 345–350. https://doi.org/10.1080/11035890701294345.
Satkūnas, J., Taminskas, J., Dilys, K., 2006. Geoindicators of changing landscapes-an example of karst development in North Lithuania. Geol. Q. 50, 457–464.
Sharma, Y.K., Kim, S., Tayerani Charmchi, A.S., Kang, D., Batelaan, O., 2025. Strategic imputation of groundwater data using machine learning: insights from diverse aquifers in the Chao-Phraya River Basin. Groundw. Sustain. Dev. 28. https://doi.org/10.1016/j.gsd.2024.101394.

Spalvins, A., Slangens, J., Lace, I., Aleksans, O., Krauklis, K., 2013. Efficient methods used to create hydrogeological model of Latvia. Int. Rev. Model. Simul. 1718–1726.

Stonevičius, E., Rimkus, E., Kažys, J., Bukantis, A., Kriaučiūniene, J., Akstinas, V., Jakimavičius, D., Povilaitis, A., Ložys, L., Kesminas, V., Virbickas, T., Pliūraite, V., 2018. Recent aridity trends and future projections in the nemunas river basin. Clim. Res. 75, 143–154. https://doi.org/10.3354/cr01514.

Sun, A.Y., 2013. Predicting groundwater level changes using GRACE data. Water Resour. Res. 49, 5900–5912. https://doi.org/10.1002/wrcr.20421.

Taminskas, J., Mikulėnas, V., Danielius, S., Satkūnas, J., Minkevičius, V., 2020. Karst denudation intensity and climate change – case of Lithuania. Vilnius Univ. Proc. 10, 38. https://doi.org/10.15388/klimatokaita.2020.30.

Virbulis, J., Bethers, U., Saks, T., Sennikovs, J., Timuhins, A., 2013. Hydrogeological model of the Baltic Artesian Basin. Hydrogeol. J. 21, 845–862. https://doi.org/10.1007/s10040-013-0970-7.

Yang, H., Kang, F., Liu, Y., Li, J., Dong, Y., Zheng, T., Qin, P., 2025. Machine learning-based identification of multi-source recharges and enrichment mechanism of karst groundwater. J. Hydrol. Reg. Stud. 62. https://doi.org/10.1016/j.ejrh.2025.102830.

Zhu, Z., Liu, D., Guo, X., Yang, Y., Yang, S., Wang, L., 2026. Multi-source remote sensing retrieval and spatiotemporal distribution characteristics of soil moisture content in typical karst farmlands of southwestern China. J. Hydrol. Reg. Stud. 63, 103052. https://doi.org/10.1016/j.ejrh.2025.103052.