

## Recognition of karst hydrographs using machine learning

### Context and objectives

Karst aquifers are characterized by highly heterogeneous medium, where groundwater circulation takes place through a network of well-developed conduits and storage is located in the fissures and matrix around conduits. This duality leads to complex and site-specific hydrodynamic responses, particularly visible in spring discharge hydrographs. Identifying whether a system is predominantly conduit-controlled or matrix-controlled, or hybrid, is essential for understanding recharge processes, vulnerability to contamination, and resource management.

Recent advances in machine learning provide a promising framework to automatically extract meaningful patterns from hydrological time series. However, most existing applications focus on prediction (e.g., flow forecasting) rather than process-based classification. There is therefore a need to develop methodologies capable of linking hydrograph dynamics to underlying flow regimes in karst systems.

The main objective of this project is to develop a machine learning approach to classify spring discharge hydrographs according to their dominant flow regime, distinguishing between conduit-dominated and matrix-dominated systems (at least to state between proof of conduits, expectation of conduits, expectation of the absence of conduits or proof of the absence of conduits). Specifically, the project aims to:

- Build a robust, labeled database of spring discharge time series with independently established hydrodynamic behavior;
- Identify and quantify hydrological signatures that discriminate between conduit and matrix flow regimes;
- Develop and evaluate machine learning models capable of classifying hydrographs and estimating the degree of conduit vs. matrix dominance;
- Assess the generalization capability of the approach on previously unseen systems;
- Provide an interpretable framework that links model predictions to physically meaningful hydrological descriptors.

Depending on the time and on the available data, this approach might also include measurements of water heads in drillholes for which similar questions arise (proof of conduits, etc.).

### Methodology

The proposed methodology is structured around four main components: data collection and labelling, feature extraction, model development, and validation.

#### 1. Data collection and labelling

A database of spring discharge time series is already compiled at SSKA from well-characterized karst systems. Only sites with independently established flow regimes (conduit-dominated, matrix-dominated, or mixed) will be included.

Time series will undergo quality control (gap handling, outlier removal, temporal harmonization) and

will be segmented into hydrologically relevant units, such as recharge events or fixed-length windows.

## 2. Extraction of hydrological signatures

Each segment will be described using a set of hydrological features capturing system behavior, including: response dynamics (e.g., rise time, peak discharge), recession characteristics, variability metrics, etc.

## 3. Model development

Machine learning models will be trained to classify hydrograph segments into conduit- or matrix-dominated behavior. Initial models will prioritize robustness and interpretability (e.g., Random Forest, etc.) producing either class labels or probabilistic scores. If sufficient data are available, deep learning approaches (like Convolutional Neural Networks) may be explored to learn directly from raw time series.

## 4. Validation and application

Model performance will be assessed using grouped cross-validation, ensuring evaluation on unseen systems. Results from pure hydraulic models might also be used as validation. Metrics will include accuracy, score, et. with particular attention to generalization and performance on mixed cases.

The final model will be applied to unlabeled time series to infer dominant flow regimes, with results interpreted in light of hydrological signatures and, where possible, compared to independent hydrogeological evidence.

## Supervision and collaboration

Dr. Arnauld Malard, ISSKA

Pierre-Yves Jeannin, ISSKA-CHYN

Contact : [arnauld.malard@isska.ch](mailto:arnauld.malard@isska.ch), [pierre-yves.jeannin@isska.ch](mailto:pierre-yves.jeannin@isska.ch)