water.ai

WATER MANAGEMENT

HOW WE CAN ANTICIPATE A VITAL YET THREATENED NEED

Water made life - and society - possible

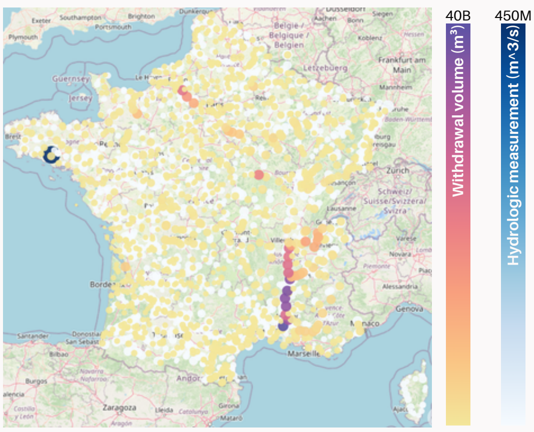
Whether it is used for agriculture, industry, or more simply to drink, water is a precious resource that we extract and use everywhere. Its availability is one of the most important development factors.

Groundwater represents in France...

of drinking water

of water used by agriculture

Spatial distribution of main measurements



Water management: a key challenge for local authorities

- **Resource Allocation**

  The water crisis is expected to cost more and more each year. In this context it is essential to take our current water allocation practices to the next level

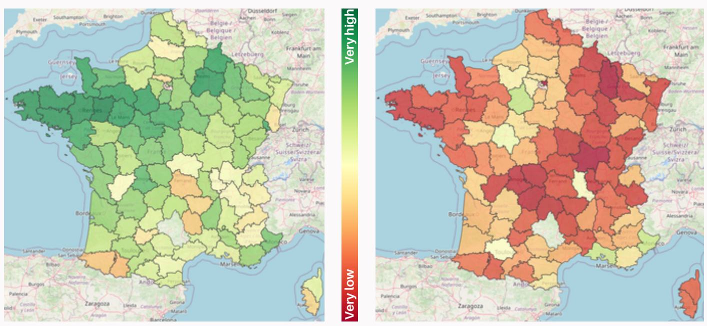
- **The data and AI revolution to the rescue**

  Using state-of-the-art models in machine learning, water.ai offers solutions to predict the groundwater levels like never before.

- **Working together with local authorities**

  At water.ai, we believe that the solution to climate change is in local actors' hands, knowing true, down-to-earth problems.

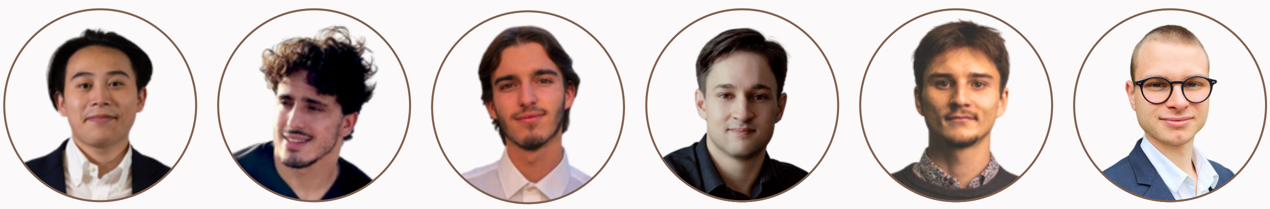
Groundwater levels in France



March 2020 September 2020

As climate change and population growth will put increasing pressure on water supplies, we need to use modern technology for optimized management and distribution, preventing the worst: shortages.

Meet our team



Jimmy BAO | Anas CHAOUI | Eliott HENRY | Louis NEL | Arnaud PELISSIER | Théo VIDAL

BUSINESS APPROACH

Our structure

We are a **young startup** funded by engineers and data scientists. We want to provide a clear vision of water usage and a groundwater level forecast to local authorities, especially towns and regions.

We plan to offer our services through a **subscription of 30,000€/year**, including everything to start planning on day one :

- **Data fetching and update**
- **Maintenance and features additions over time**
- **Custom 1-to-1 support**

This will cover our yearly 20,000€ costs, shared between servers, maintenance, data retrieval and licensing.

We'll first deploy in PACA region, by targetting 20% of "Communautés d'agglomérations": we should gain 240,000€/year.

In three years, we plan to reach clients all over the southern regions of France, where needs for water management are the highest. After 5 years, the service should reach every region. We plan to gain 3,3M€ yearly.

Main advantages of AI prediction

- **More security for the population**

  Water is ubiquitous in modern society. Thus securing its access is a significant challenge to protect populations

- **Better visualization of tensions**

  Our enhanced visualization tools will allow public authorities to more effectively target areas that are under tension.

- **Cost saving proactivity**

  Taking action as soon as possible will save precious resources in the long run .

Environmental approach

Our model is lightweight and only uses accessible computing resources: CPUs. It can be run on-device, removing the need for heavy server farms

Scenario #1

A municipality wants to apply preventive rules

Thanks to our approach, they can rather apply **predictive rules** as they have a complete vision on the future.

Scenario #2

A farmer needs to plan crops watering

They can enter their planned withdrawal volume and calculate the impact on the rest of the community

Scenario #3

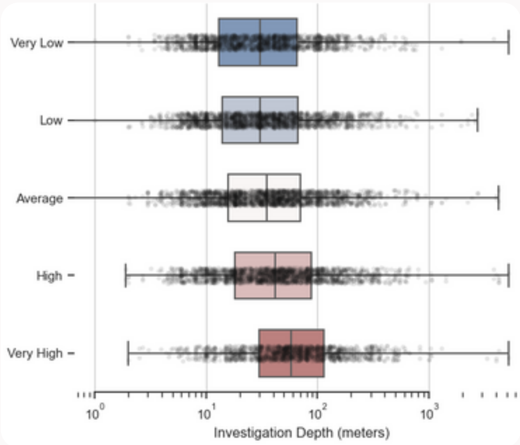
A city plans to build a new residential area

Using our models, they know which location will **minimize water tensions** in the long run.

water.ai

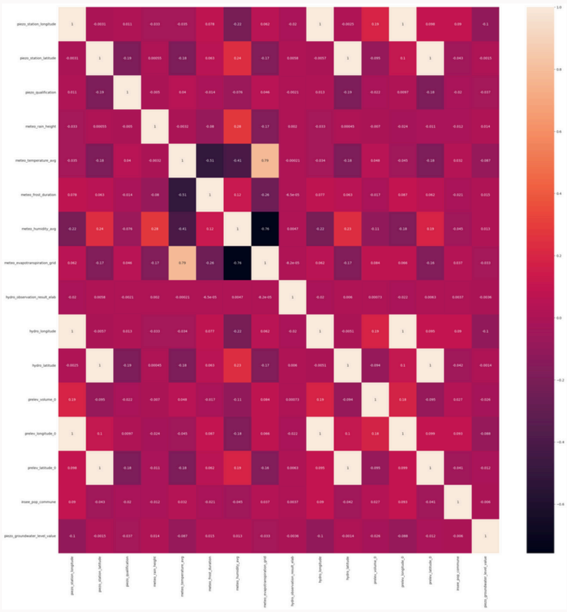
SCIENTIFIC APPROACH

Example of a significant feature



interpretation: groundwater level tends to increase with investigation depth

Correlation matrix of the most significant features



Understanding the data

To better understand the data, we made **several visualizations**. Since the data set includes multiple spatial measurements, we found it helpful to create maps showing how features vary across locations. One of the challenges that we found in using the data set is the presence of large numbers of missing values.

Building a model

Our approach to building a good predictive model is based on **gradient-boosted trees**. The data set is quite complex, consisting of more than a hundred features, few of which are obviously related to the predictor variable. Gradient-boosted trees are highly flexible models that are able to adapt to this complexity. In particular, we used the **XGBoost** library's implementations. This brings two main advantages: first, since it has been under development for so long, this library has been highly optimized to reduce training time. Second, the XGBoost library can seamlessly handle missing values in features, alleviating the need for extensive feature engineering.

Preprocessing data

- **INSEE Data cleaning**
- **Remove features with >10% NaN**
- **One-hot encoding quality and method categorical features**
- **Standardization of numerical features**

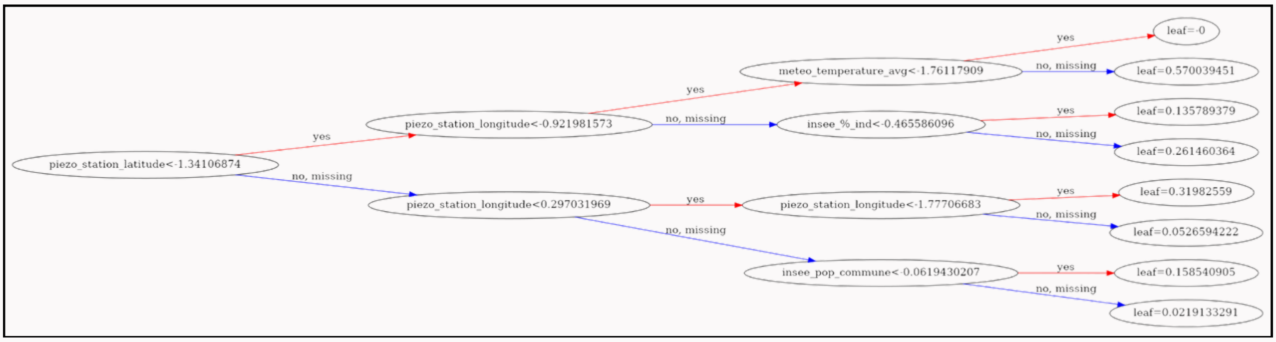
  Ensure every feature is in the same range

What we also tried

- **Day encoding, independent of the year**

  Create a cyclic numerical feature

Overview of a gradient-boosted tree

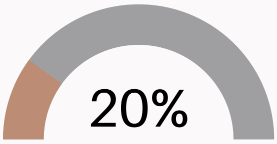


water.ai

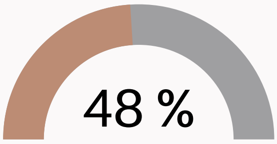
RESULTS AND FUTURE

Performance metrics

We obtained a weighted **F1 score of 48%** on a five-labels classification task. The F1 score takes into account predictions that are false positives or false negatives (unlike accuracy, which only looks at true positives/negatives) and is more robust to class imbalances. This is a noticeable improvement over random guessing, which would score approximately 20%.



Baseline F1-Score

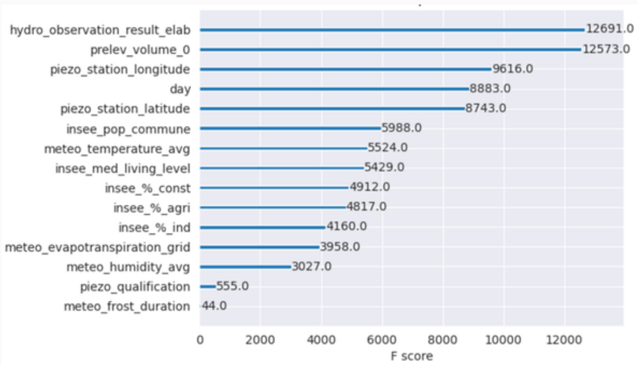


Our Model F1-Score

Future use

We plan to build a complete web dashboard, where local and national authorities can find all relevant metrics concerning water usage and shortages prediction. An AI will provide further assistance for clear insights: shown during councils, decision-making is faster and proactive.

Feature importance



Explanation

- **Hydrometric measurement**
  Measurement on a near hydrometric station.
- **Withdrawals' volume**
  Biggest withdrawal within range of a given zone.
- **Demographic data**
  May indicate nearby usage and needs for water
- **Temperature**
  Water evaporates if it's too high

To go further

- **Snow: freshwater during summer**
  A future model should take into account snow during winter, for its impact on the watershed.