EDHEC BUSINESS SCHOOL

Master Project:

Using Time Series to predict the level of French Groundwater till 2030.

Mathieu MARCHAND

MSc Data Analysis & Artificial Intelligence

Master Project Director: Jérémy Leymarie

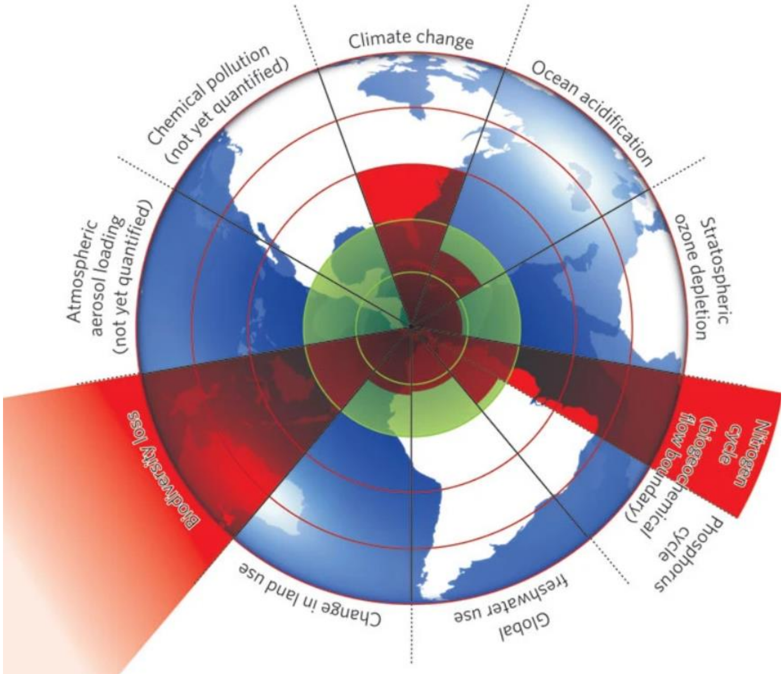
ABSTRACT

The purpose of this master project is to provide a prediction of the level of French groundwater in France. The database is provided by the government organization ADES and contains the observations of 4771 piezometers in Metropolitan France (including the Corse region). The time scope is from January 2000 to December 2021. The predictions cover the years 2022, 2025 and 2030. Various models have been used: ARIMA model and SARIMA model. The baseline model is composed of the mean of the observations on the train period. Statistical indicators are used to choose the best model for each prediction and the results are gathered by department and by basin to ease the visualization and the interpretation.

LIST OF CONTENTS

INTRODUCTION

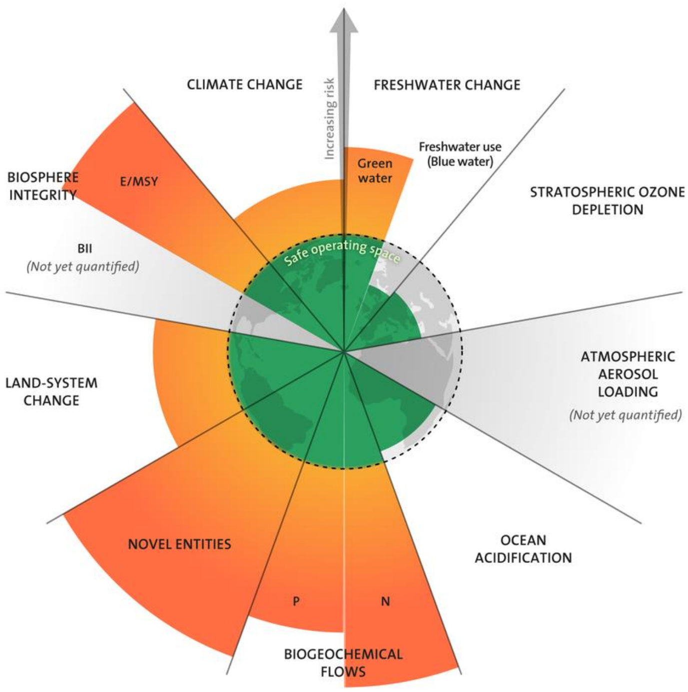
The 26th of April 2022, an article from the scientific review Nature entitled “A planetary boundary for green water” showed that the human activity has deeply disturb the water cycle (Wang-Erlandsson et al., 2022). It represents a milestone in the study on the consequences of global warning on the environment. The study of the nine planetary boundaries started in 2009 and gave thresholds on how to be sustainable for humanity on various topics (Rockström et al., 2009).



Picture 1: Beyond the boundary. The inner greed shading represents the proposed safe operating space for nice planetary systems. The red wedges represent an estimate of the current position for each variable.

From ‘A safe operating space for humanity’ Nature 2009

As we can see on the graph, three boundaries have already been crossed in 2009 and the situation is getting worse over the years.



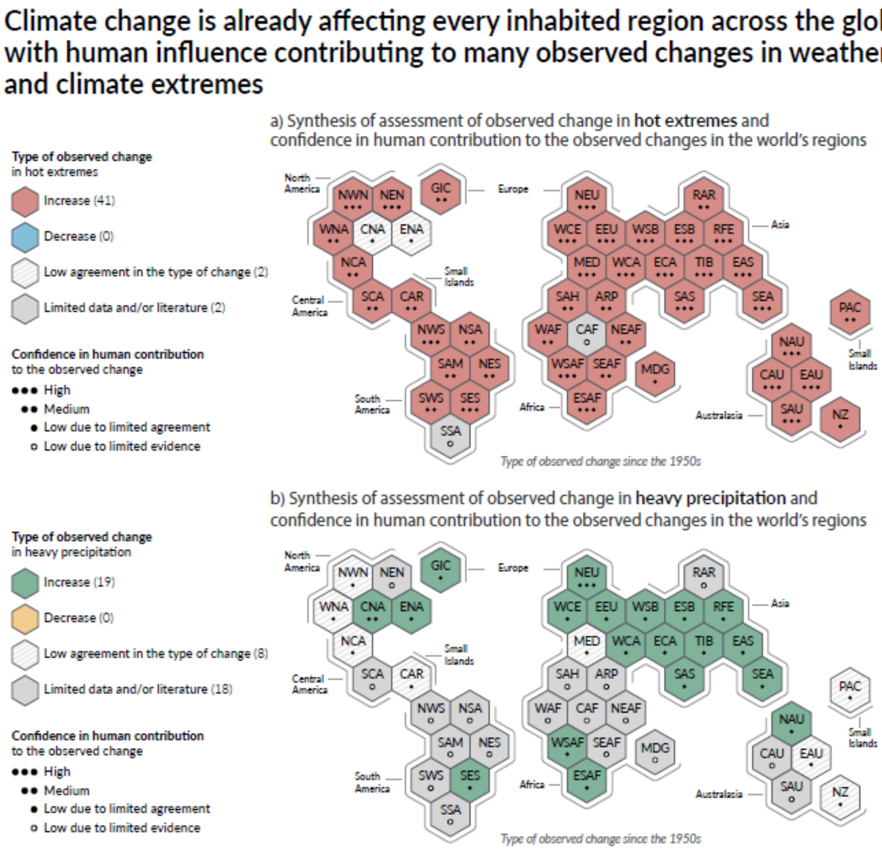
Picture 2: Updated planetary boundaries.

From 'A planetary boundary for green water' Nature 2022

But what is the distinction between green water and blue water? Green water is the water from precipitations, evaporation, and soil moisture, whereas blue water is the water in the rivers, the lakes, and the groundwater stores or aquifers. Indeed, the humidity in the ground, used by vegetables for instance, is part of the green water, while the groundwater is categorized as blue water. By introducing the notion of green water in the estimation of freshwater planet boundary, Lan Wang- Erlandsson and his team highlight a problem that threatens our entire interconnected environment (Wang-Erlandsson et al., 2022).

Another reference of the study of climate change is The Intergovernmental Panel on Climate Change (IPCC) which delivered a complete report on the Climate change in 2021 and 2022 through three working groups. The first Working group is dedicated to the physical science basis, it gives a summary of the academic knowledge on all the subjects linked to climate change. An entire chapter in that summary is dedicated to water (Douville et al., 2021).

In this chapter, we discovered the complexity of the issue. The chapter describes with high confidence a future with more precipitations, more evaporation, and more droughts and all this phenomenon is mainly caused by human activity.



Map 1: Synthesis of world climate extreme conditions

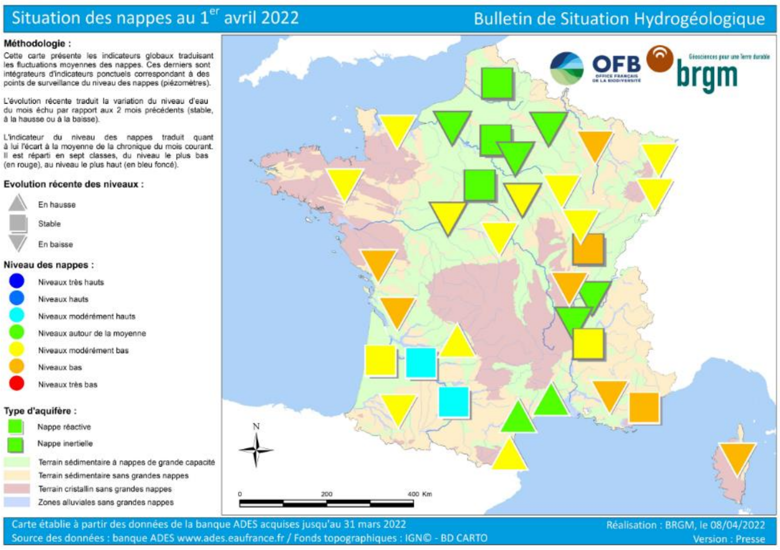
From IPCC, summary for policymakers, 2021 (Masson-Delmotte et al. 2021)

However, climate data are difficult to interpret for the final user of the water. There is a link between water availability and precipitations, between water stress and drought, but it is hard to measure. In this complex context, how to study the water resources and make estimations for our future?

Firstly, we decided to focus on one type of water resources: groundwater. According to the national geographic encyclopaedia, (National Geographic Society, 2022) “Groundwater is water that has infiltrated the ground to fill the spaces between sediments and cracks in rock.

Groundwater is fed by precipitation and can resurface to replenish streams, rivers, and lakes." So, it is the water you can extract from the ground, and that can be stocked in the aquifers. As its volume depends on the climate, it is an interesting approach to study the impact of the environmental crisis.

We also decided to narrow the geographic scale to France. It is an interesting country to analyse because its use of water relies a lot on its groundwater. According to J-C Maréchal and J Rouillard, (Maréchal, Rouillard, 2020) "groundwater resources represent about 66% of France's domestic water supply, 31% of industrial water supply and 37% of the total water use in agriculture". So, the level of groundwater is an important issue for the country, it is carefully tracked by the French government. All the data are available on the ADES (Accès aux Données sur les Eaux Souterraines) website managed by EauFrance. It gathers the data collected by more than 6000 piezometers with quality and quantity indicators. Each month, the BRGM (Bureau de Recherches Géologiques et Minières) publishes a map which resumes the current groundwater situation.



Map 2: On April 1rst 2022, Situation of French Aquifers

From BRGM, 2022 (BRGM, 2022)

Water level predictions are crucial for the government, who is used to implementing measures to limit the use of water in dry period. According to the datalab environment report of 2020, in 2019 more than 67% of the French Metropolitan territory has been concerned by water restrictions (SDSE, OFB, 2020).

Our objective will be to use this data to make predictions on the level of groundwater in the following years and deliver several French maps which show the future level of groundwater at different times.

Our hypothesis is that in most of the country, the quantity of groundwater is decreasing with the time.

PART 1: Literature review and hypothesis

1.1 The water issues

Today, we are certain that human activity is the main cause of the global warming which disturbs all of our ecosystems, including the global water cycle (Masson-Delmotte et al. 2021).

The consequences of global warming on the water cycle have been widely studied. In 1993, the book ‘Water in crisis’ is published (Shiklomaniv. 1993). The chapter 2, written by Igor A. Shiklomaniv shows figures on the water reserves on Earth: only 2.53% of the global reserves are fresh water. ‘Fresh water lakes and rivers, which are the main sources for water consumption, contain on average about 90,000 km^3 of water, or just 0.26% of total global freshwater resources.’ Indeed, even if water is considered as a renewable resource, we do not have an unlimited quantity of water, and its cycle must be respected. Already in 1993, the article warns about the problems related to water in regard to economic activity. In the 60’s - 90’s, the water withdrawal has increased in all continents and in the most developed regions, all the major rivers have been disturbed by human activity. The water exploitation and pollution will globally hinder our society’s development in the future.

In France, a milestone on water issue sensibilization was the interview of René Dumont, French engineer, environmental politician, and candidate to the presidential election of 1974. ‘We will soon run out of water. And that is why I drink, in front of you, a glass of precious water. Precious, because by the end of the century, if we continue such an overflow, it will be missing.’ (Delport. 2015) This famous intervention was a starting point for the French Environmental political party and showed the growing importance of water in scientists’ preoccupations globally. In April 1965, William Bowen published an article entitled ‘How real is the water shortage?’ and it contained all the issues we are facing now (Bowen. 1965). The article dealt with the water in the United States in the 60’s. Indeed, W Bowen referred to the terrible water predictions for the 80's. It demonstrated how the Southwest of the United States, which used a lot of water in their agriculture, would need to import water to feed the population. However, the main point of this 1965 article was that the real danger will come from water pollution. Therefore, how to measure pollution? What kind of controls should be implemented and how to enforce them?

In 1974, the Safe Drinking Water Act ruled the minimum level of quality required by public water supply in the United States. However, according to the scientists, this was not enough. Daniel Okun published in 1976 'Drinking Water for the future' which showed the limit of the Safe Drinking Water Act (Okun. 1976). Once again, the article emphases the importance of water quality rather than quantity, but the necessity to diligently manage these resources is not excluded.

Today, the world water issue is better understood and documented by scientists. The recent figures on water resources shows that less than 3% of fresh water is considered easily accessible to human society. And this small fraction of freshwater is largely contained in the groundwater (75% = 630 thousand km^3) (Douville et al., 2021). This source of water is actually overexploited 'Human water use is sustainable for some regions at some times, but for large portions of the globe, groundwater pumping exceeds recharge, river discharge is overallocated, and water pollution causes rampant human disease and ecosystem degradation' (Abbott et al. 2019). The article shows that we cannot keep an analysis of the water cycle without taking into consideration the human use of water.

To conclude this part, we can already feel the consequences of water shortage in many countries. The World Resources Institute published a list of countries based on their water stress situation, 17 countries, which represent a quarter of the world's population, face extremely high-water stress (Rutger et al. 2019).

EXTREMELY HIGH BASELINE WATER STRESS			
1. Qatar	6. Libya	10. United Arab Emirates	14. Pakistan
2. Israel	7. Kuwait	11. San Marino	15. Turkmenistan
3. Lebanon	8. Saudi Arabia	12. Bahrain	16. Oman
4. Iran	9. Eritrea	13. India	17. Botswana
5. Jordan			

HIGH BASELINE WATER STRESS			
18. Chile	25. Uzbekistan	32. Turkey	39. Niger
19. Cyprus	26. Greece	33. Albania	40. Nepal
20. Yemen	27. Afghanistan	34. Armenia	41. Portugal
21. Andorra	28. Spain	35. Burkina Faso	42. Iraq
22. Morocco	29. Algeria	36. Djibouti	43. Egypt
23. Belgium	30. Tunisia	37. Namibia	44. Italy
24. Mexico	31. Syria	38. Kyrgyzstan	

MEDIUM-HIGH BASELINE WATER STRESS			
45. Thailand	51. Tajikistan	57. Guatemala	63. Lesotho
46. Azerbaijan	52. Macedonia	58. Estonia	64. Denmark
47. Sudan	53. South Korea	59. France	65. Indonesia
48. South Africa	54. Bulgaria	60. Kazakhstan	66. Peru
49. Luxembourg	55. Mongolia	61. Mauritania	67. Venezuela
50. Australia	56. China	62. Germany	68. Cuba

Table 1: list of country based on their water stress, data from WRI's Aqueduct

From World Resources Institute

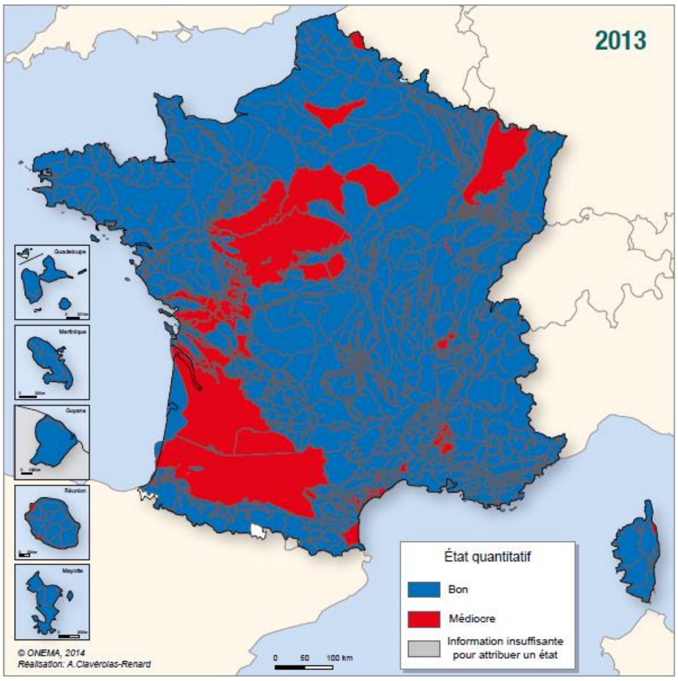
A country is considered in extremely high-water stress when the water withdraws for the country represent more than 80% of their available supply on average on a year. This is a very precarious situation that can degenerate in the event of droughts or any other water-related problem. It is what happen in this moment in India where the 2022 early droughts are deeply disturbing the country.

1.2 Global warming on groundwater

As we seen before, groundwater is the first freshwater reserve in the world in terms of quantity, and we have already started to exploit these reserves. According to the IPCC: '*There is high confidence that groundwater depletion has occurred since at least the start of the 21st century as a consequence of groundwater withdrawals for irrigation in agricultural areas in drylands (e.g., the United States southern High Plains, California Central Valley, North China Plain, and northwest India)*' (Douville et al., 2021). Indeed, a third of the world's annual freshwater is used for agriculture and also for industry and for domestic usage come from the groundwater. Our society relies on this resource. But, when the withdraw of groundwater are superior to the groundwater recharge, this exploitation increase the aridity of the area. As it is the case in the countries in the table 1.

In France, ‘groundwater resources represent about 66% of France’s domestic water supply, 31% of industrial water supply and 37% of total water use in agriculture’ (Maréchal, Rouillard, 2020)**.** The French territory contains different climates, from the dry South to the rainier North. On average, the country has good water characteristics with higher than average precipitation and large volumes of groundwater, especially along the rivers. But, despite of this good predispositions, the intensive used of extracted groundwater have damaged the French aquifers, in terms of quality and quantity.

On the quantity side, in 2013, 10% of the groundwater bodies was considered in a bad quantitative state, as we can see in red in the following map.



Map 3: Quantitative state of French groundwater in 2013.

From Onema (Petit, Michon, 2015).

A groundwater body is considered in a bad quantitative state when the withdraws are superior to the groundwater body recharge. The basin in bad quantitative state are Adour-Garonne, Loire-Bretagne, and Seine-Normandie.

1.3 Statistical case studies on groundwater

As we seen before, the groundwater is a crucial water resource for the countries. Especially in regions in water stress. The climate change is disturbing the water cycle and it is vital to make predictions on the level of groundwater. These studies use the climate predictions to produce statistical models which will estimate the future quantity and quality of groundwater. We will see their methodology and their results.

1.3.1 Explore 2070

Explore 2070 is a French project which answered the questions of the Ministry of Ecology on the French hydric adaptation face to climate change in 2050-2070 (Office Français de la Biodiversité. 2012). Around one hundred of searchers worked from June 2010 to October 2012 to elaborate various climate scenarios, their impact on the water management and some solutions. They start by making some hypotheses on the future climate conditions: sea level, greenhouse gases emissions, temperature, demography… And they used its conditions to build hydrological models. The reference period was 1961 to 1990 for making predictions on the 2050 to 2070 period. The models have first been tested in the period 1991 to 2010 and they have been selecting by comparing the variations between the result and the test period.

One part of Explore 2070 is dedicated to groundwaters (Office Français de la Biodiversité. 2012). The ADES database, which contains the data from piezometers in France, has been used to study the level of groundwater. After the data cleaning, the Pettitt test is applied to find rupture points in the dataset and add it to the other variables of the model.

The results of the Explore 2070 project show a general decrease of the level of water between (1960 – 1990) and (2045 - 2065). Even in the optimistic scenarios, the Loire' basin and the Southwest of France will lose between 30% and 50% of their recharge.

This project and its results are interesting because they used the same database and the same scope as our master project. We will see if we find similar results.

1.3.2 Statistical approach of the effect of climate change on groundwater, example in Morocco.

A statistical approach of the issue is developed in an article entitled: "Assessment by statistical approach of climate change impact on water resources: application to the Gharb perimeter (Morocco)" (Acharki et al. 2019). The study will take 1981-2016 as the reference period and 2021-2050 for the projections on various parameters as the temperatures, the precipitations, or the evapotranspiration.

The subject is not the same, but the methodology is interesting. As for our study, the team of S.Acharki has to deal with missing data, choosing a train and a test periods for his models, use some tests to detect ruptures, compare the performances of each models on different regions and deliver the results.

For the missing data, S.Acharki used the algorithm MICE and the method k-NN which allow to use the whole dataset to determine the values of the missing data.

As in the Explore 2070 project, the Pettitt test have been used to detect the rupture points in the dataset. Then, the searchers used statistical tests (Sen et Mann-Kendal) on the climatical data to find trends. The models have been compared with the Mean Absolute Error and the Root Mean Square Error which must have been minimized.

The result shows that evapotranspiration and the hydric deficit are increasing in the predictions for 2021-2050. Another conclusion is that the season of precipitations will change, with more precipitations in autumn and spring and less rain in winter.

The study shows other statistical tests and exogenous variables which can be used to study the water resources. We will use similar technics for comparing our models.

1.3.3 Groundwater vulnerability assessment in the Grand Est region

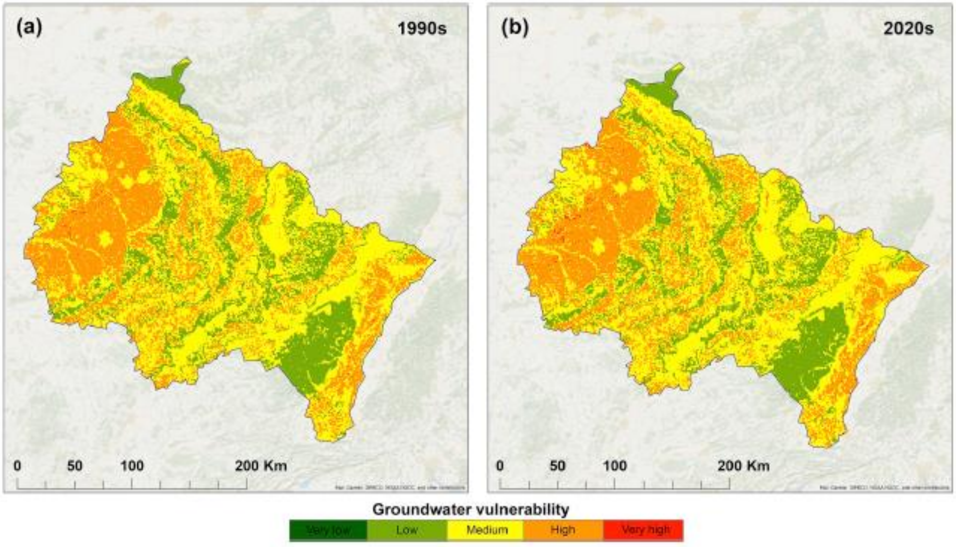
In this study, Ionel Haidu and Mărgărit-Mircea Nistor used a spatial approach to study the vulnerability of the groundwaters in the Grand Est Region (Haidu and Nistor. 2019). The major parameters to compute this vulnerability are the aquifers media, the potential infiltration map, the pollution load index, and the effective precipitations.

The searchers used several maps to measure each factor and compare data from the 90's and actual data. Finally, each parameters have a theorical weight which is used to determine the vulnerability of the place:

$$GW\ V = (1-EP)1.5 * 0.3 + AM * 0.2 + PIM * 0.2 + PLI1.5 * 0.3.$$

Where: GW V=Groundwater vulnerability, EP=Effective precipitation, AM=Aquifer media, PIM = Potential infiltration map and PLI = Pollution load index.

The index is between 0 and 1, above 0.6, the index shows a high groundwater vulnerability. The results for the Grand Est Region are a high vulnerability in the region. This spatial approach gives very precise results but need different types of data. We will see if we find different results for the Grand Est Region with our method.



Map 4: Groundwater vulnerability in Grand Est region in past and present period

From Groundwater vulnerability assessment in the Grand Est region, France

1.3.4 The VESPA Index

The VESPA (Validation of a Vulnerability Estimator for Spring Protection Areas) Index has been formalized by L.Galleani, B.Vigna, C.Banzato and S.Lo Russo (Galleani et al. 2011).This index is used to better understand and determine the groundwater vulnerability of an area thanks to one year of data from the spring. The VESPA index considers several parameters as the discharge of the aquifer, the temperature, and electrical conductivity derived from hydrographs.

Formula of VESPA Index: $V = c(\rho) * \beta * \gamma$. Where $c(\rho)$ is the correlation factor, β is the temperature variability, and γ is the discharge factor. Each parameter has also its formula and the multiplication of the factors gives the VESPA index.

1.4 Intuition and hypothesis

Groundwater is an important water resource which must be protected from global warming. The statistical models show a correlation between the change of climatical indicators as precipitations and temperature, and the drop of the level of groundwater, including in France.

Now, we want to know if we can make predictions on the level of groundwater by using other tools. The precedent models use a lot of exogenous variables to determine correlation between the parameters and find a formula to predict the level of groundwater.

In this master project, we will try to find similar results with a time series approach which only consider the level of groundwater, without using environmental variables.

We expect to find similar results, a drop of groundwater in France, especially in basin already in bad quantitative state as Adour-Garonne, Loire-Bretagne and Seine-Normandie.

PART 2: Methodology

2.1 Overview of the methodology

Making predictions on the level of French groundwater is a time series analysis conducted on each piezometer. So, we deal with a huge number of time series, and it is impossible to rely on visual interpretation. We cannot draw all the curves, so need to rely on statistical tests and parameters to gather the result and take decisions.

As every data analysis, our study will start on a data cleaning process. Thanks to the work of French government on the water management, our data are quite clean, but we still have some decisions to do on what can we use in our analysis.

Then, on each time series, we will test the stationary with the Augmented Dickey-Fuller Test to see if our time series tools are applicable on our time series. We will describe the test and the interpretations in a dedicated part.

The next step is the modelling step, we must find models, fit it on the train periods and compare the result with the test period with different indicators and a simple baseline model to evaluate the added value of our analysis.

The final step is the predictions, we apply our model on the future period and we show the results.

2.2 Extracting the data

What are the available data? On the ADES website, we have access to a large number a time series to analyse.

Picture 3: Screenshot from the ADES website

We have a total of 6161 piezometers. On each one, we extract a table which contains several dated data collected on this piezometer.

- The National BSS ID: A unique ID for each piezometer.
- The date of the measure: the date format is dd/mm/yyyy hh/mm/ss.
- The NGF value: our variable of interest. It is the General Levelling of France, the zero level (NGF value = 0) corresponds to sea level, measured at Marseille in France and Ajaccio in Corsa. One NGF value is one meter. For instance, if we measure a NGF value of 95 in 2000 and a NGF value of 94,8 in 2010, we can say that we lose 20 centimetres of water on the piezometer.
- Indicators on the method of observation: In our case, we will not make distinction between the observed values and the calculated values.
- Geographical parameters: latitude and longitude at the WGS84 format. We will not deal with geographical data so much. But we keep the department of the piezometer because it will be useful to product maps.

Actually, the NGF value is not really a good indicator to measure the quantity of water in the ground. And, it is difficult for us to say if we have a good or a bad level of water. But we can make conclusion on the evolution of this value.

We know that on the 21st of December 2002, the NGF value on the piezometer BSS001PZGD is 169.25. We are not specialist of the Ain department hydrology, so, it is impossible to correctly interpret the value. We do not know if it a lot or if it is a bad new, we do not know how much water there is below 169.25 meters. But we know that on the 22nd of December 2002, on the same piezometer, we have a NGF value of 169.27. So compared to yesterday, we gain 2 centimetres of water. And this is what we are looking for, determine if there a trend in the level of groundwater in France. And, even if we are not a specialist, we can assume that a constant reduction of groundwater is not good news for the environment.

So, for our analysis, we will keep the date of the observation, the department, the piezometer id and the NGF value.

	Identifiant national BSS	Date de la mesure	Côte NGF	Department
0	BSS001PZGD	2002-12-21 00:00:00	169.25	1
1	BSS001PZGD	2002-12-22 00:00:00	169.27	1
2	BSS001PZGD	2002-12-23 00:00:00	169.29	1
3	BSS001PZGD	2002-12-24 00:00:00	169.29	1
4	BSS001PZGD	2002-12-25 00:00:00	169.33	1
...	...	...	...	...
69112	BSS000NMUV	2019-06-23 17:00:00	45.71	94
69113	BSS000NMUV	2019-06-24 03:00:00	45.69	94
69114	BSS000NMUV	2019-06-25 05:00:00	45.68	94
69115	BSS000NMUV	2019-06-26 01:00:00	45.63	94
69116	BSS000NMUV	2019-06-27 00:00:00	45.60	94

17804755 rows × 4 columns

Table 2: Overview of our dataset

2.3 data cleaning

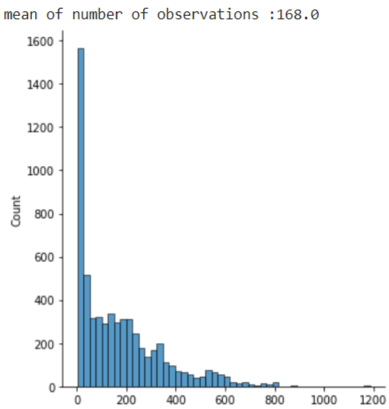
The data from ADES website are clean, on each piezometer, for each observation, we have the date, the NGP value and the id at the same format. But the data are collected on a daily basis, and it is not convenient because it increases the fluctuation. In our case, we would like to work with a monthly sampling frequency. There are many aggregations method to pass from daily data to monthly data. In our case, we choose the mean of the daily observations to have the month observation. This method is simple to understand and work with any length of sample.

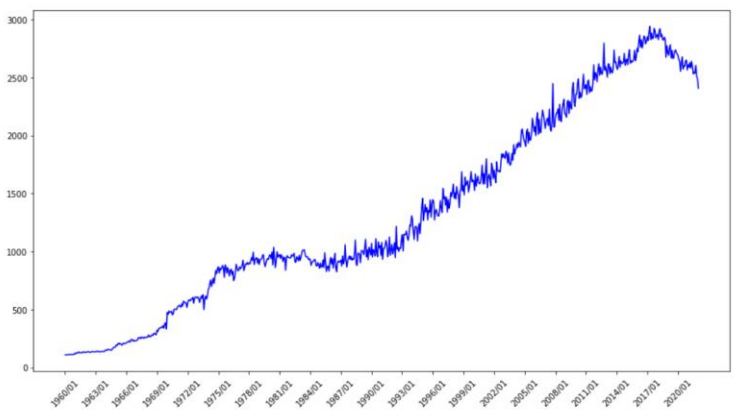
	Date de la mesure	Côte NGF	Department
		mean	mean
0	2007/03	44.975000	94.0
1	2007/04	44.775667	94.0
2	2007/05	44.659677	94.0
3	2007/06	44.844000	94.0
4	2007/07	44.803548	94.0
...	...	...	...
141	2019/02	45.773929	94.0
142	2019/03	45.822258	94.0
143	2019/04	45.611333	94.0
144	2019/05	45.590323	94.0
145	2019/06	45.669259	94.0

146 rows × 3 columns

Table 3: Overview of the data from one piezometer after the monthly sampling frequency.

The main goal of the data cleaning is to select the time series according to our objectives: making predictions till December 2030. To better understand our database, we plot the number of time series which have an observation on each date.

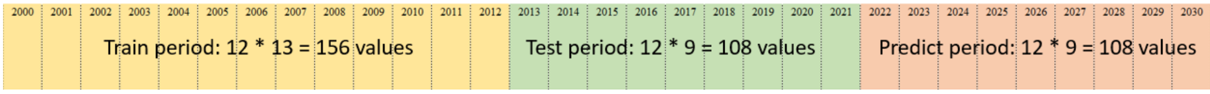




Graph 1: Number of observations by id / Graph 2: Number of observations by date

We can see on the graph 2 that a lot of our piezometers contains very few data. And, on the graph 3, we can observe that more recent is the date, more observations we have on this date.

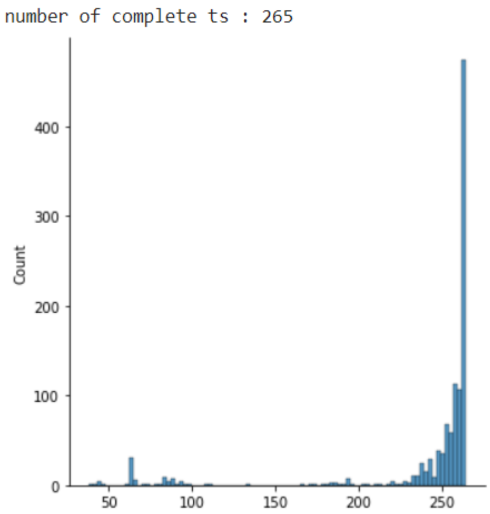
There is not really a point where it is more interesting to start. To make prediction till December 2030 with the data till December 2021, we need at least 9 years for our test period and 9 for our test period.



Picture 4: timeline of the different period with the needed number of values by time series.

The minimum timeline was to take 2004-2012 as train period and 2013-2021 as test period. But to have better results, we take all the data after 2000. It represents 4771 time series.

On this dataset, we apply a first filter, we keep only the time series which contains at least one observation by year between 2000 and 2021. It remains 1126 time series including 265 complete time series.

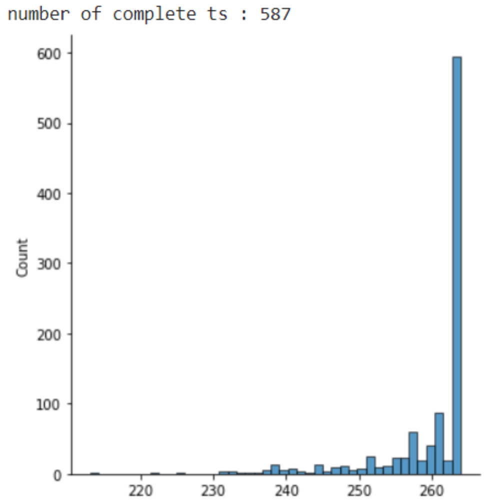


Graph 3: Number of observations by id after the first filter

We remove all the time series with less than 200 observations. It remains 1017 time series.

We really need to work on complete time series, so we decide to complete some missing data. In cases where we have one or two missing values in a row, we replace them with the more recent existing value.

Among 1017 time series, after the completion of missing data, we have 587 complete time series which are our dataset for the analysis.



Graph 4: Number of observations by id after filters and replacing some missing values.

With our database of 587 times series, we cover all the 6 French basins. At the department scale, 24 departments have more than 10 time series, 36 departments have between 1 and 8 time series and 35 departments have no time series at all. Our departmental maps will be uncompleted.

2.4 Advanced Dickey Fuller Test

Making predictions on time series is easier when the series is stationary. It means that the series does not contain a trend and fluctuates around the same value. The stationarity can be observed simply by observing the curve but, in our case, we do not want to do it manually because we have too much time series to analyse.

So, we use a statistical test to determine if our time series are stationary or not. The most popular statistical test for stationarity is the Advanced Dickey Fuller Test. It is a Unit Root test; it means that we will determine the value of the root of the time series and use it to determine if the series is stationary or not.

Unit Root of a series:

$$y_t = \beta_1 * y_{t-1} + \varepsilon_t$$

$With\ y_t = value\ at\ time\ t, \beta_1 = \ constant\ and\ \varepsilon_t = \ the\ error\ term\ at\ time\ t.$

$$y_t - \beta_1 y_{t-1} = \varepsilon_t$$

$$y_t(1 - \beta_1 L) = \varepsilon_t$$

$With\ L = \frac{y_{t-1}}{y_t}$

$NCe\ want\ to\ solve\ the\ equation\ (1 - \beta_1 L) = 0, if\ \beta_1 < 1, the\ series\ is\ stationnary.$

But why? We can have the intuition with some examples.

$$If\ \beta_1 = 0.5, L = 2, so\ \frac{y_{t-1}}{y_t} = 2 \equiv y_t = \frac{1}{2} * y_{t-1}$$

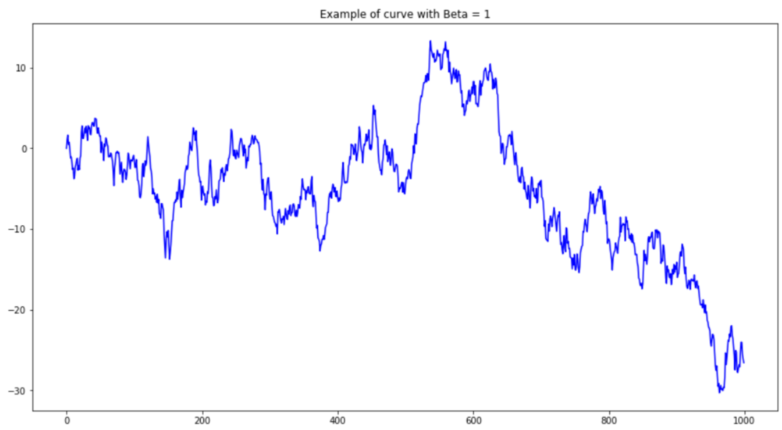
So, the value at time t is equal to the half of the precedent value. The series will decrease till 0 and become stationary.

$$If\ \beta_1 = 2, L = 0.5, so\ \frac{y_{t-1}}{y_t} = 0.5 \equiv y_t = 2 * y_{t-1}$$

The value at time t is equal to the double of the precedent value. The series will increase and never stabilize. The series is not stationary.

$$If\ \beta_1 = 1, y_t = y_{t-1} + \varepsilon_t$$

This series is not stationary as we can see in this graph, we set that the error term follows N(0,1):



Graph 5: example of no stationary curve with beta1 = 1

Dickey Fuller Test:

The Dickey Fuller Test starts with the same model and the same goal. Determine the value of beta 1 to check if the series is stationary or not.

$$y_t = \beta_1 * y_{t-1} + \varepsilon_t$$

$With\ y_t = value\ at\ time\ t, \beta_1 = constant\ and\ \varepsilon_t = the\ error\ term\ at\ time\ t.$

To perform the test, we confront the null hypothesis with the alternative hypothesis of Unit Root Test:

$$H_0: \beta_1 = 1\ versus\ H_A: \beta_1 < 1$$

But the Dickey-Fuller test solves another equation:

$$y_t - y_{t-1} = \beta_1 * y_{t-1} + \varepsilon_t - y_{t-1}$$

$$\Delta y_t = \rho * y_{t-1} + \varepsilon_t\ with\ H_0 : \rho = 0\ versus\ H_A: \rho < 0$$

The difference between two time series is more often stationary, we can now estimate the regression of the differenced series and obtain the statistic t-value associated with the ρ. And we compare this t-statistic with the Dickey Fuller critical value. If the t-statistic is below, the series is stationary.

The Augmented Dickey Fuller is a more complex version of the Dickey Fuller. It allows to deal with trend or more regressive component. This is the general regression formula:

$$\Delta y_t = c + \gamma_1 t + \rho y_{t-1} + \sum_{i=1}^{p} \beta_i \Delta y_{t-i} + \varepsilon_t$$

With c is a constant and $\gamma_1 t$ is a deterministic trend. But the test remains the same, we want to check if $\rho = 0$ (H_0: the series is not stationary) or if $\rho < 1$ (H_A: the series is stationary).

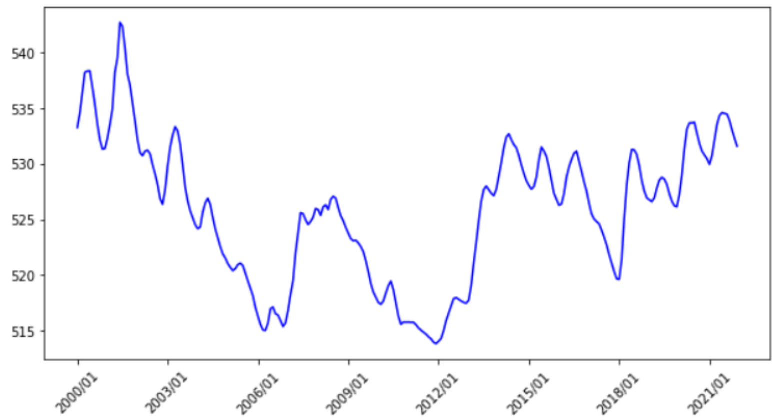
In Python, we can run the test on a time series thanks to the function adfuller(). The parameters of the test will determine the regression formula. On our dataset, we test with no constant, with a constant, with a constant and a trend and with a constant, a trend and a quadratic trend ($\gamma_2 t^2$).

We perform the test on our database, and we check the p-value.

Regression	Nb stationary time series	Proportion of the database
ctt	231	0.39
ct	291	0.50
c	371	0.63
n	9	0.02

Table 4: result of the ADF test on the series

We do not reach the 95% of the dataset stationary with one type of regression. Indeed, a lot of our time series are not stationary as the observation on the piezometer 'BSS001QCDZ' located in the city named Gex (01173).



Graph 6: Example of a series from our dataset which is not stationary

So, we try the same test on the differentiated series ($\Delta_t = y_t - y_{t-1}$). Study the difference between the value at time t and t+1 is a way to compute predictions.

Regression	Nb stationary time series	Proportion of the database
ctt	544	0.93
ct	566	0.96
c	580	0.99
n	584	0.99

Table 5: result of the ADF test on the differentiated series

The differentiated series are almost all stationary. We decide to use the regression with no constant and no trend for the models. We have slightly more stationary series with this formula, and it is the simplest form. But we keep in mind that we can also work with the regression with a constant.

Thus, we can formulate this stationary regression:

$$\Delta y_t = c + \sum_{i=1}^{p} \beta_i \Delta y_{t-i} + \varepsilon_t$$

With c a constant which can be equal to 0 (case with no constant).

2.5 Models

One example

In this part, we will take a unique time series as an example. We choose the time series with the id: ‘BSS001QSUX’. Its piezometer is located in the Loire department (42) and more precisely in the city of St-Galmier which is well-known for its production of water with gaz. The curve is quite clean, we clearly see the fall of the level of water years after years.

Graph 7: Level of water (NGF value) at St-Galmier

Indicator

We cannot compare all the curve manually, so we have to choose an indicator to compare our models. In the study, we will use the Root Mean Squared Error (RMSE). This is its formula:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{n}(y_{t,ref} - y_{t,pred})^2}{n}}$$

With $y_{t,ref}$ is the reference value of y at time t, $y_{t,pred}$ is the predicted value of y at time t and n is the number of observations or the length of the time period.

More the errors between the predictions and the reality are important, more the RMSE will be big. So, the aim of our models will be to minimize the RMSE to reach the best accuracy possible.

Dummies

As we saw in the first part, groundwater recharge is linked to weather. We add dummies to consider the month or the season of the prediction to improve the predictions. A dummy is a variable which can be equal only to 0 and 1.

Thanks to a linear model, we test if our monthly dummies are statistically significant to estimate the level of water. There are twelve months in a year, we take December as the month of reference. Meaning that if we are in December, all our dummies are equal to 0. If we are in March, all our dummies are equal to 0 except d_mar which is equal to 1. We test our dummies and look at the p-value for each of our time series.

dummy	p_value <= 0.1	p_value > 0.1
d_jan	437	150
d_feb	351	236
d_mar	269	318
d_apr	347	240
d_may	284	303
d_jun	452	135
d_jul	513	74
d_aug	455	132
d_sep	374	213
d_oct	364	223
d_nov	408	179

Table 6: p-values of our monthly dummies

We can see that some of our dummies are not significant as March or May. It shows that being in March or May is not helpful for more than 300 time series to determine the level of water. However, some of the months are really useful to estimate the level of water as July. We decide to gather our months in dummies for season, with spring (Mars, April, May) as the reference dummy. 'd_summer' is equal to 1 in June, July, and August, 'd_automn' is equal to 1 in September, October, and November, and 'd_winter' is equal to 1 in December, January, and February.

dummy	p_value <= 0.1	p_value > 0.1
d_summer	540	47
d_automn	454	133
d_winter	507	80

Table 7: p-values of our seasonal dummies

The seasonal dummies are more often statistically significant, we decide to integrate them to the following models.

2.5.1 The baseline model

A baseline model is a really basic model. We will not use it for predictions, only for compare the performances of the following models on the test period. In our case, we decide to take a constant equal to the mean of the values in the train period.

The formula for our baseline model is:

$$\forall t \in [T+1;\ N], y_t = \frac{1}{T+1} * \sum_{i=0}^{T} y_i$$

With t is time, T is the number of observations in the train period and N is the sum of the number of observations in the train and the test period. In other words, for each t in the test period, y_t is equal to the mean of the observation in the train period.

Example of St-Galmier

Graph 8: Predictions of the level of water (NGF value) at St-Galmier according to the baseline model (in red) compared to reality (in blue). RMSE = 1.25

In this example, the baseline model does not fit the real curve.

General performances

We compute the predictions for all of our time series, and we keep each RMSE. Computing the mean of this list of RMSE gives an idea of the global performances of the model.

We have a mean RMSE for our database of 1.44. The value is hard to interpret directly, we need to compare it with the following models.

2.5.2 The ARIMA model, without a constant

Our second model is an Auto-Regressive Integrated Moving-Average model (ARIMA model). It takes three parameters, the Auto-Regressive component (p), the Integrated component (d), and the Moving Average component (q).

Before looking at the formula, we focus on the Integrated component. It gives the number of times the series is differentiated before estimating the parameters. If d=1, the series is first-differentiated, it means we are working on $y_t - y_{t-1}$ instead of working on y_t (d=0). If d=2, we work on the first difference of the first difference: $(y_t - y_{t-1}) - (y_{t-1} - y_{t-2})$. In our case, we already first differentiated the series to reach stationarity in the Advanced Dickey Fuller part, so we set d=1 and we will express the formula of $y_t - y_{t-1} = \Delta_t$.

This is the formula of the ARIMA (p, 1, q) model without a constant.

$$\Delta_t = \beta_1 * \Delta_{t-1} + \ldots + \beta_p * \Delta_{t-p} + \theta_1 * \varepsilon_{t-1} + \ldots + \theta_{t-q} * \varepsilon_{t-q} + \varepsilon_t$$

With β_x is constant, θ_x is constant and ε_t is the error term at time t.

As we can see in the formula, the ARIMA model takes consider lags in the predictions. It means that it uses the precedent values of Δ_t to estimate the actual value. It is the Auto-Regressive component of the ARIMA, and the number of lags is p. Moreover, the ARIMA also considers the past error terms in the estimation. It is the Moving-Average component of the ARIMA, and the number of error terms is q.

Now, let's find the best order of our ARIMA model, which are the values of p and q for each time series of our database. So, we need an information criterion to compare the different order or the ARIMA.

In this project, we will use the Akaike Information Criterion (AIC) on ARIMA of order (p, 1, q) with p and q between 0 to 4. We will use all the available data from 2000 to 2021.

The AIC is value which consider the complexity of the model and the fit between the model and the data. Meaning how well the model reproduces the data. This is the AIC formula:

$$AIC = 2K - 2ln(L)$$

K is the number of independent variables in the model and L is the log-Likelihood of the model. We will not explain how is calculated Log-Likelihood, but it increases with the fitting of the model with the data. To sum up, AIC makes a compromise between complexity and accuracy to choose the order of the model.

id	p	q
BSS001QCDZ	4.0	4.0
BSS001RFRV	3.0	1.0
BSS001RGXM	4.0	4.0
BSS001SBPB	3.0	4.0
BSS001SEGB	2.0	3.0

Table 8: the optimized parameters of 5 series for the ARIMA model with no constant

Formula of our ARIMA model without constant.

$$\Delta_t = \beta_1 * \Delta_{t-1} + \ldots + \beta_p * \Delta_{t-p} + \theta_1 * \varepsilon_{t-1} + \ldots + \theta_{t-q} * \varepsilon_{t-q} + Ds_t + Da_t + Dw_t + \varepsilon_t$$

With β_x is constant, θ_x is constant, ε_t is the error term at time t, Ds_t is equal to 1 in summer and 0 otherwise, Da_t is equal to 1 in autumn and 0 otherwise and Dw_t is equal to 1 in winter and 0 otherwise.

Example of St-Galmier

The optimized parameters for St-Galmier are (1, 1, 1). Now, we apply the ARIMA(1,1,1) on the training period (2000-2012) to have predictions on the test period (2013-2021) and compare with the reality.

Graph 9: Predictions of the level of water (NGF value) at St-Galmier according to the ARIMA(1,1,1) model with no constant (in red) compared to reality (in blue). RMSE = 0.55

We can see on the graph that the predictions of the ARIMA are better than the baseline model on this series. Indeed, we have a smaller RMSE.

General performances

The general performances the ARIMA model without a constant are under the ones of the baseline model with a mean RMSE of 1.94.

2.5.3 The ARIMA model, with constant

The third model is very similar to the previous one. We only add a constant to the regression.

This is the formula of the ARIMA (p, 1, q) model with constant.

$$\Delta_t = c + \beta_1 * \Delta_{t-1} + \ldots + \beta_p * \Delta_{t-p} + \theta_1 * \varepsilon_{t-1} + \ldots + \theta_{t-q} * \varepsilon_{t-q} + \varepsilon_t$$

With β_x is constant, θ_x is constant, ε_t is the error term at time t and c is a constant.

Same method, we fit the model with the train and test period combined to estimate the order of the model based on the AIC.

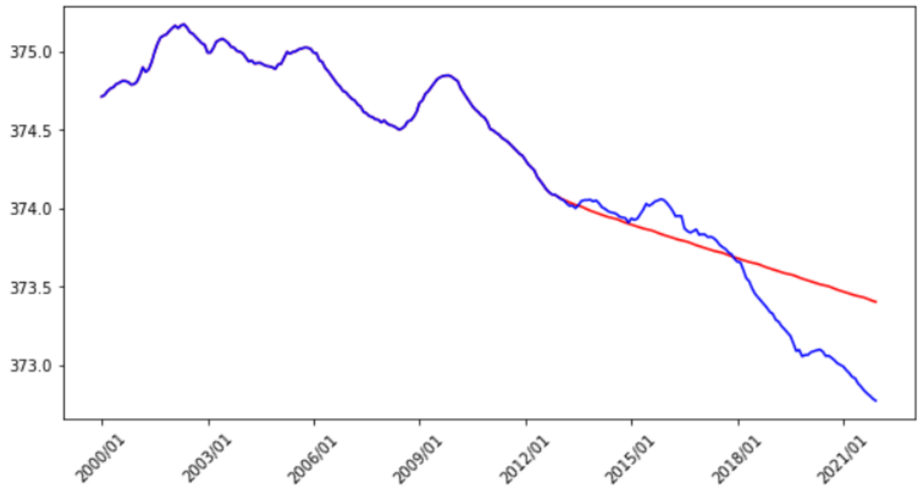
Formula of our ARIMA model with a constant.

$$\Delta_t = c + \beta_1 * \Delta_{t-1} + \dots + \beta_p * \Delta_{t-p} + \theta_1 * \varepsilon_{t-1} + \dots + \theta_{t-q} * \varepsilon_{t-q} + Ds_t + Da_t + Dw_t + \varepsilon_t$$

With c is a constant, β_x is constant, θ_x is constant, ε_t is the error term at time t, Ds_t is equal to 1 in summer and 0 otherwise, Da_t is equal to 1 in autumn and 0 otherwise and Dw_t is equal to 1 in winter and 0 otherwise.

Example of St-Galmier

As for the previous model, the optimized parameters for St-Galmier are (1, 1, 1). It is not every time the case. Now, as before, we apply the ARIMA(1,1,1) on the training period (2000-2012) to have predictions on the test period (2013-2021) and compare with the reality.



Graph 10: Predictions of the level of water (NGF value) at St-Galmier according to the ARIMA(1,1,1) model with constant (in red) compared to reality (in blue). RMSE = 0.29

We have a better fit. The RMSE is smaller than in the ARIMA model with no constant. The constant should be negative and helps the predictions to follow the decreasing trend of the curve.

General performances

On the whole dataset, the ARIMA model with a constant has worse performances than the ARIMA model without a constant. Indeed, the mean of the RMSE for all the time series is equal to 3.4 compared to the ARIMA model without a constant (RMSE = 1.94).

2.5.4 The SARIMA model, without a constant

The SARIMA model is an ARIMA model with a seasonal component (m), it is composed by the standard ARIMA(p,d,q) and the ARIMA at lag m: ARIMA(P,D,Q)m. Hence, a SARIMA model contains seven parameters: SARIMA(p,d,q)(P,D,Q)m.

As in the ARIMA model, we work with differentiated series, so we set d and D at 1. We have monthly values, so we set m at 12.

This is the formula of the SARIMA (p, d, q) (P, D, Q) m model without a constant.

$$AR(p) * I(d) * AR(P)_m * I(D)_m = MA(q) * MA(Q)_m$$

$$\begin{aligned}\left(1-\beta_1 L-\cdots-\beta_p L^p\right)(1-L)^d\left(1-\gamma_1 L^m-\cdots-\gamma_p L^{m*P}\right)(1-L^m)^D y_t \\ = \left(1+\theta_1 E+\cdots+\theta_q * E^q\right)\left(1+\alpha_1 E^m+\cdots+\alpha_Q E^{m*Q}\right)\varepsilon_t + Ds_t + Da_t \\ + Dw_t\end{aligned}$$

With βx is constant, θ_x is constant, γ_x is constant, α_x is constant, ε_t is the error term at time t, $L^x = y_t / y_{t-x}$, $E^x = \varepsilon_t / \varepsilon_{t-x}$, Ds_t is equal to 1 in summer and 0 otherwise, Da_t is equal to 1 in autumn and 0 otherwise and Dw_t is equal to 1 in winter and 0 otherwise.

Now, our objective is to find the best parameters p, d, P and Q to minimize the AIC. We test all the possibilities with p and q between 0 and 3 and P and Q between 0 and 1. We would like to test more models but with 587 time series, it is very time consuming.

We discover that P and Q are equal to 1 for each of the time series of our dataset.

Example of St-Galmier

For the piezometer of St-Galmier, the best SARIMA model with no constant is the SARIMA(3,1,0) $(1,1,1)_{12}$.

Graph 11: Predictions of the level of water (NGF value) at St-Galmier according to the SARIMA(3,1,0)(1,1,1)$_{12}$ model without constant (in red) compared to reality (in blue). RMSE = 0.60

We can see that the SARIMA does not fit well with this piezometer. The little oscillations at the beginning may fit more to other curves.

General performances

On the whole dataset, the SARIMA model without constant has worse performances than the baseline model (2.11 vs 1.44).

2.5.5 The SARIMA model, with a constant

Same method, we just add a constant to the formula of the SARIMA:

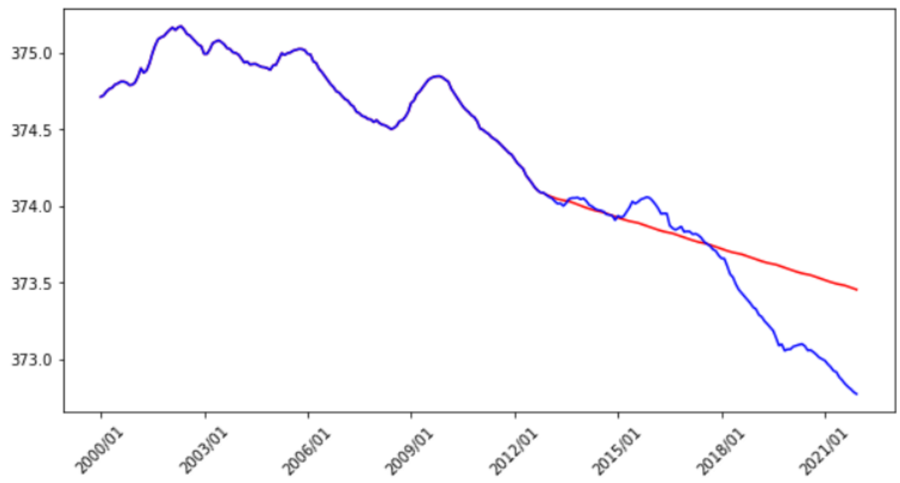
$$(1-\beta_1 L-\cdots-\beta_p L^p)(1-L)^d(1-\gamma_1 L^m-\cdots-\gamma_p L^{m*P})(1-L^m)^D y_t$$
$$= c+(1+\theta_1 E+\cdots+\theta_q * E^q)(1+\alpha_1 E^m+\cdots+\alpha_Q E^{m*Q})\varepsilon_t + Ds_t$$
$$+ Da_t + Dw_t$$

With c is constant, βx is constant, θ_x is constant, γ_x is constant, α_x is constant, ε_t is the error term at time t, $L^x = y_t / y_{t-x}$, $E^x = \varepsilon_t / \varepsilon_{t-x}$, Ds_t is equal to 1 in summer and 0 otherwise, Da_t is equal to 1 in autumn and 0 otherwise and Dw_t is equal to 1 in winter and 0 otherwise.

We choose the best parameters p, q, P, and Q with the same range as before (p and q between 0 and 3 and P and Q between 0 and 1). we observe a similar result with P and Q equal to 1 for each time series of the dataset.

Example of St-Galmier

For the piezometer of St-Galmier, the best SARIMA model with a constant is also the SARIMA(3,1,0) $(1,1,1)_{12}$.



Graph 12: Predictions of the level of water (NGF value) at St-Galmier according to the SARIMA(3,1,0)(1,1,1)$_{12}$ model without a constant (in red) compared to reality (in blue).

RMSE = 0.31

We have a better fit than with the SARIMA without constant but the ARIMA with a constant is better. For this piezometer, the model ARIMA with a constant is selected for the global predictions.

General performances

On the whole dataset, the SARIMA model with a constant has worse performances than the baseline model (1.95 vs 1.44). However, it is slightly better than the version without a constant (2.11) and almost equal to the ARIMA without a constant (1.94).

2.5.6 Selecting a model for each time series

For each of the 587 time series, we select the model which minimize the RMSE between the ARIMA without a constant, the ARIMA with a constant, the SARIMA with a constant and the SARIMA without a constant.

	Without constant	With a constant	Both
ARIMA	133	187	320
SARIMA	129	138	267

Table 9: Repartition of the best models according to the RMSE

This table shows that each of our model is appropriate for a certain part of the dataset. Indeed, we need various models to estimate the different kind of curves in the dataset.

Overall performances of taking the best model for each time series

By taking the model which minimize the RMSE for each time series, we obtain a RMSE of 1.47 which is slightly inferior to the RMSE of the baseline model (1.44). However, the mean is very sensitive to outlier, if one of our models has a huge RMSE, it spoils the global RMSE. So, we look at the number of time series where the predictions of the model are better than theses with the baseline model. In total, we have 400 time series with a RMSE inferior with an ARIMA or a SARIMA model than with the baseline model.

2.5.7 Selecting the time series with relevant model results

Diebold-Mariano test

To make predictions on the future period, we want to know if our models are pertinent to make predictions based on their RMSE and the RMSE of our baseline model. The Diebold-Mariano test is able to statistically compare the RMSEs from to different models. If the p-value is above 5%, we do not reject the null hypothesis, so the two models have similar predictive power on average.

How this test works?

The purpose of the test is to statistically show that the two RMSE are equal, that means that the difference between both tends to 0.

$$We\ note\ \hat{\epsilon}_{i,t}\ the\ forecast\ error\ of\ the\ model\ i\ at\ time\ t.$$

In our case, i = 1 corresponds to the baseline model and i = 2 is the model selected for the time series.

$$RMSE_1 - RMSE_2 = \frac{1}{N}\sum_{n=1}^{N}\hat{\epsilon}_{1,n}^2 - \frac{1}{N}\sum_{n=1}^{N}\hat{\epsilon}_{2,n}^2 = \frac{1}{N}\sum_{n=1}^{N} z_t$$

With N is the size of the sample.

The Diebold-Mariano test assumes that z_t is random variable and that we can apply the Central Limit Theorem which approximates a random variable with a normal distribution.

$$\frac{1}{N}\sum_{n=1}^{N} z_t \sim approx\ N(0, \frac{V(z_t)}{N})$$

We can compute the DM statistics, as the random variable has a normal distribution, we can assume, with 95% of accuracy, that:

$$DM\ stat = \frac{\frac{1}{N}\sum_{n=1}^{N} z_t}{\sqrt{\frac{V(z_t)}{N}}}$$

- If DM stat < -1.96, the baseline model gives on average smaller squared error.
- If -1.96 < DM stat < 1.96, the baseline model gives on average equivalent squared errors.
- If DM stat > 1.96, the baseline model on average larger squared errors.

In python, the Diebold-Mariano test gives a p-value which is above 5% if DM stat < 1.96 and below 5% if DM stat > 1.96. We use this p-value to detect the time series where our model has an added value compared to the baseline model.

Hence, there are three possible cases:

- RMSE baseline model < RMSE model, we remove the time series from the analysis.
- RMSE baseline model > RMSE model but the Diebold-Marin test p-value is above 5%. So, we remove the time series from the analysis.
- RMSE baseline model > RMSE model and the Diebold-Marin test p-value is below 5%. Our model has a significant superior predictive power compared to the baseline model. We keep the time series.

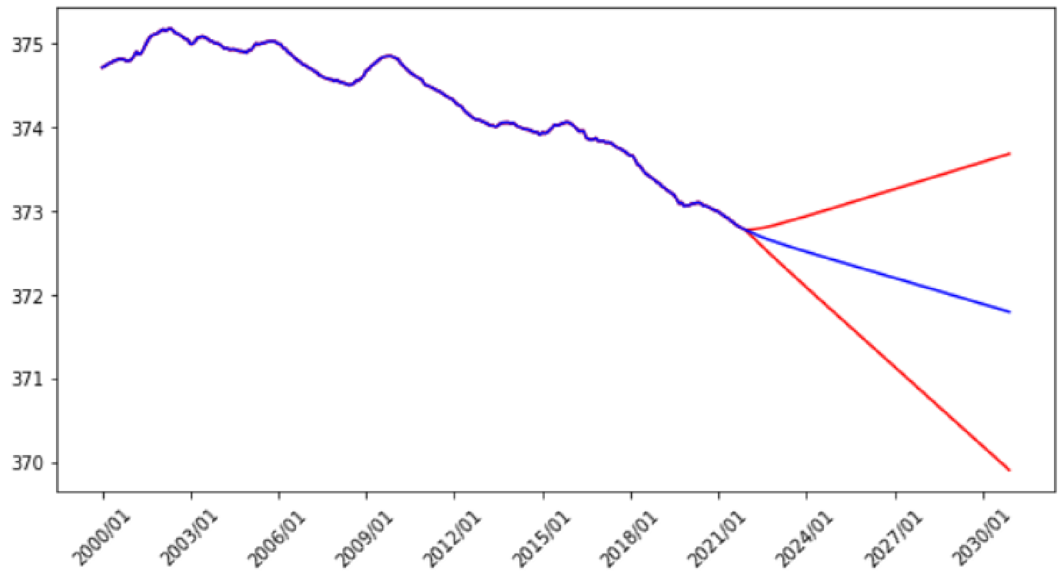
We test all of our series, and we finally keep 353 times series which are statistically superior to the baseline model.

2.6 Predictions

For each of our time series, we use the data from January 2000 to December 2021 to compute the predictions till 2030. We build a table which contains the season dummies to consider them in the predictions. For each id, for each month between January 2022 and December 2030, we have an estimation of the level of water in this piezometer and a confidence interval.

We decide to take a confidence interval of 60%, it means that our model estimates that there is 60% of probabilities that the real value is in the interval. The advantage of taking a confidence interval of 60% compared to a confidence interval of 95% or 90% is the range of the result. We have a smaller range with an interval of 60%.

Example of St-Galmier

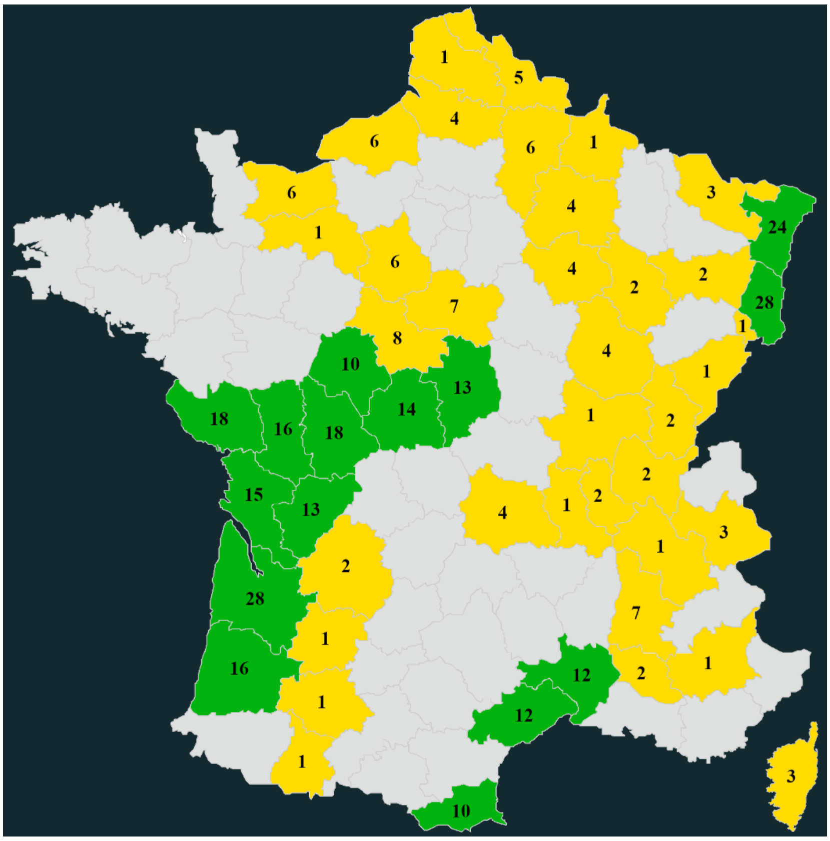


Graph 13: Predictions of the level of water (NGF value) at St-Galmier till December 2030 with a confidence interval of 60% (in red).

On the piezometer of St-Galmier, the estimation of the next values shows a decrease of the level of water in mean. The confidence interval is growing month after month because we work with an iterated approach. Meaning that the estimation for t+2 uses the estimation of t+1 which is already contained in a confidence interval.

Overview of the dataset

The map 3 shows a problem in our dataset, after data selection and focusing on models with decent results, we have a lot of departments without any data. We have to keep this in mind for the following maps which aggregate the data depending on basin or local climate.



Map 5: Number of observations by departments, in green, ten or more observations, in yellow less than 10 observations and in grey no observation. Map realised on the GIMP software.

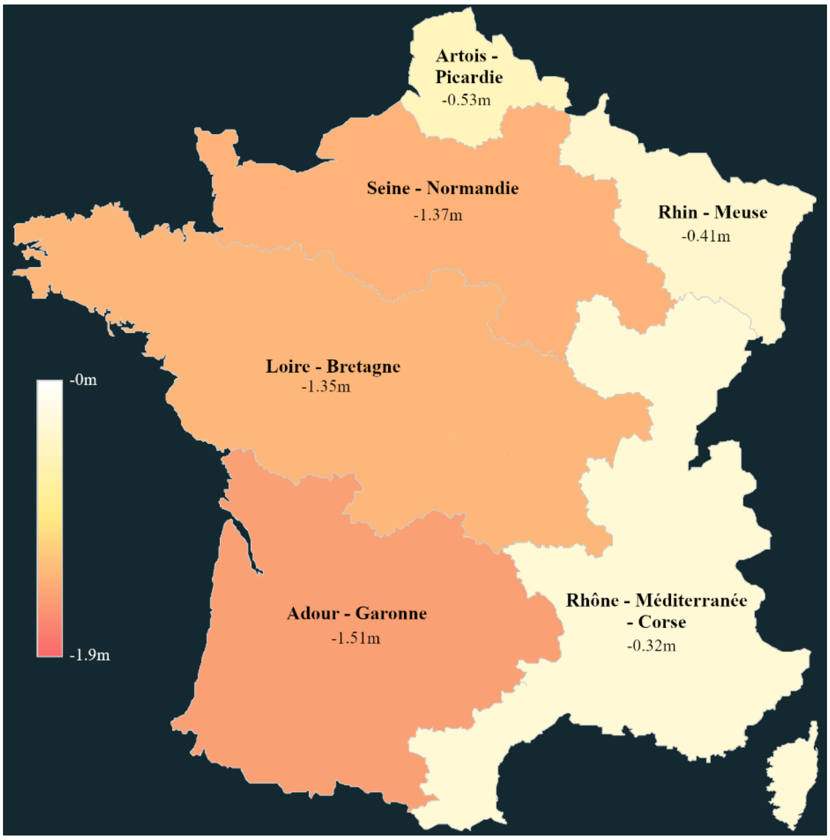
2.6.1 2022 maps

The following maps show the difference between the level of water predicted for 2022 compared to 2000.

To ease the interpretation, we gather the department in basin and in climate area. Basins are French administrative districts which manage the main French rivers. Each basin has its water agency and basin' committee. Some departments belong to various basins; hence, we assign the entire department to the basin which cover the more department territory.

We use the same method for the climate areas. Indeed, France has various meteorological conditions which influence precipitations, temperatures, and the groundwater recharges. We divide the territory in five climate area: oceanic, mountain, Mediterranean, semi continental and semi oceanic. Each department has been assigned to a climate depending on the one which cover the more department surface.

For each time series, we take the mean of the predictions between January 2022 and December 2022, and we compare with the mean of observations between January 2000 and December 2000.



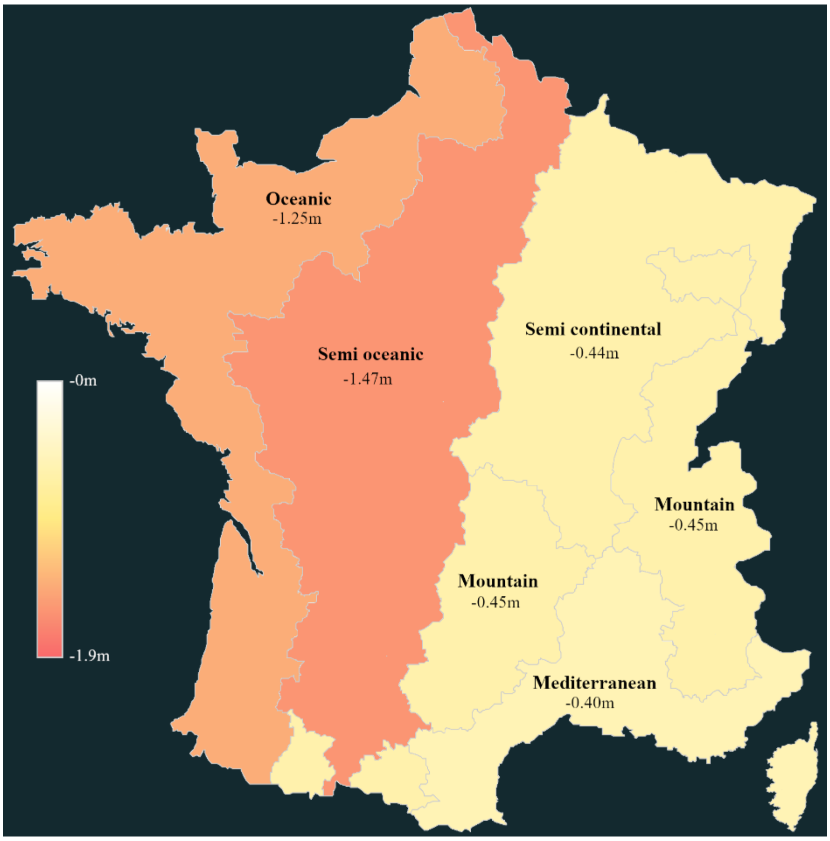
Map 6: Predictions about difference of groundwater level between 2022 and 2000 in France, by basin. Map realised on the GIMP software.

We can observe a general decrease of the groundwater level between 2022 and the reference year 2000. Nevertheless, the confidence intervals are large and if the mean is always negative, the intervals contain also positive values. So, we are not sure at 60% that the groundwater will drop in 2022 compared to the reference year (2000).

Bassin	Nb obs	Lower	Mean_diff	Upper	Int_range
Artois-Picardie	10.0	-2.84	-0.53	1.79	4.62
Seine-Normandie	35.0	-4.48	-1.37	1.73	6.21
Rhin-Meuse	58.0	-1.70	-0.41	0.88	2.58
Loire-Bretagne	110.0	-5.57	-1.35	2.87	8.45
Rhône-Méditerranée-Corse	63.0	-2.34	-0.32	1.70	4.04
Adour-Garonne	77.0	-5.81	-1.51	2.79	8.60

Table 10: Predictions about difference of groundwater level between 2022 and 2000 in France, by basin.

The more impacted basins are the Loire-Bretagne and the Seine-Normandie basins. It corresponds to the findings of the previous studies.

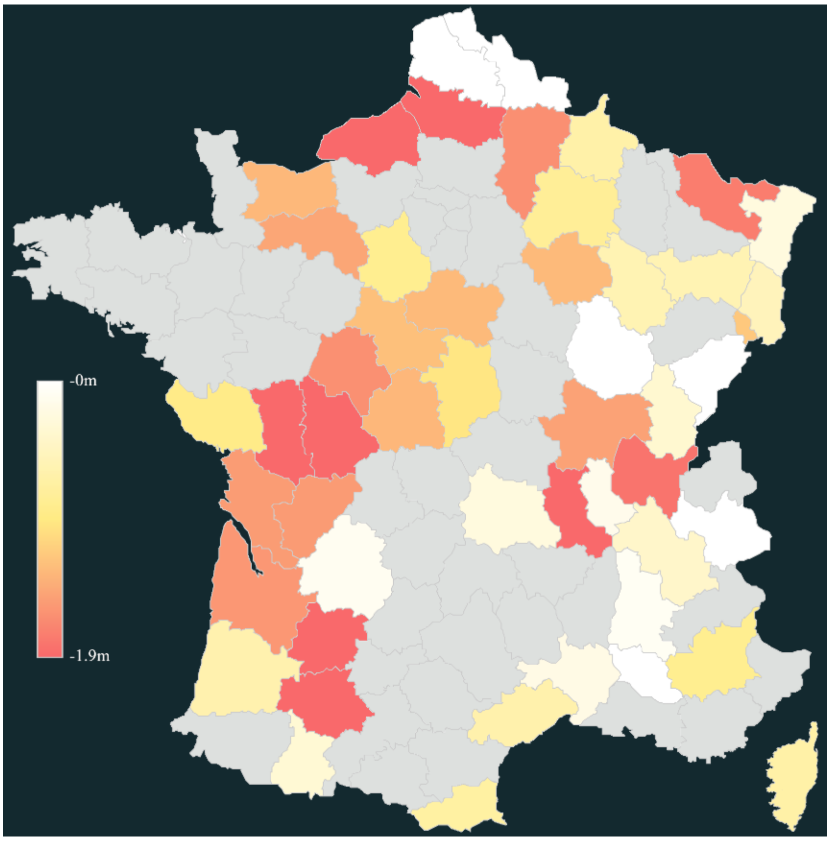


Map 7: Predictions about difference of groundwater level between 2022 and 2000 in France, by climate area. Map realised on the GIMP software.

Climate	Nb obs	Lower	Mean_diff	Upper	Int_range
Oceanic	95.0	-5.41	-1.25	2.91	8.33
Montain	18.0	-1.98	-0.45	1.09	3.06
Mediterranean	39.0	-2.35	-0.40	1.56	3.91
Semi oceanic	127.0	-5.38	-1.47	2.43	7.81
Semi continental	74.0	-2.26	-0.44	1.37	3.63

Table 11: Predictions about difference of groundwater level between 2022 and 2000 in France, by climate area.

Again, we observe a general decrease of the level of French groundwater. The oceanic and semi-oceanic climates are the most impacted ones and correspond to the Ouest of the country.



Map 8: Predictions about difference of groundwater level between 2022 and 2000 in France, by department. Map realised on the GIMP software.

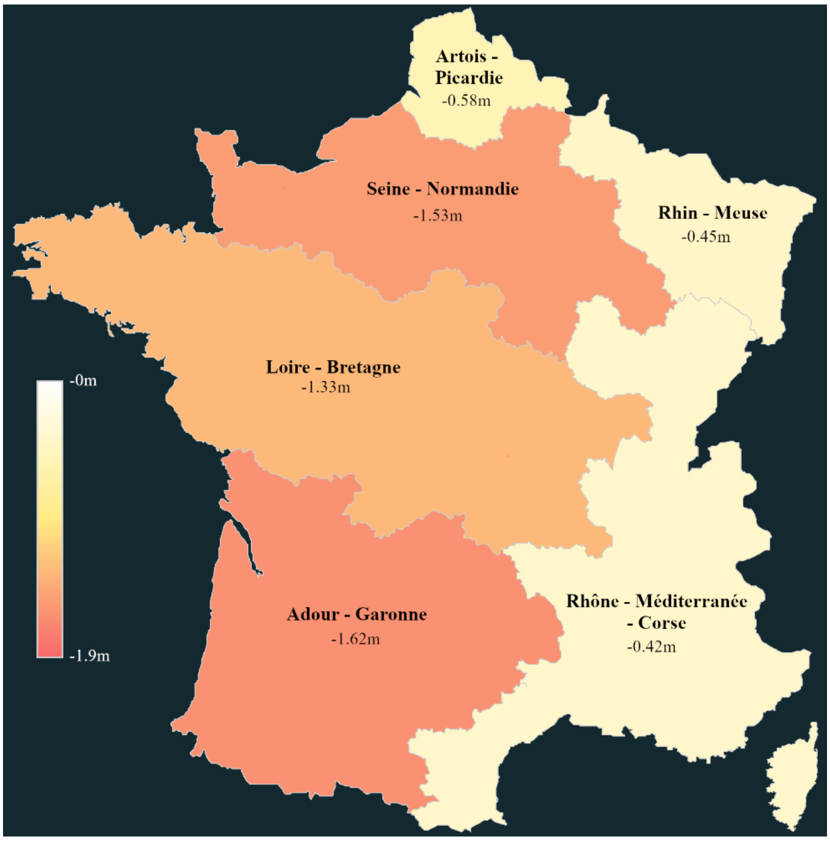
The values are the annexe. Depending on the department, the values are very different, we also observe a global decrease, but some departments have more groundwater than before as the

"Pas-de-Calais", "Nord" or "Côte d'Or" departments. However, even the departments with a positive mean of predictions have a negative lower confidence interval boundary.

Some departments are hardly impacted by the decrease of groundwater as the "Gers" department (-11.38m [-12.38; -10.38]) and the "Lot-et-Garonne" department (-13.64m [-14.11; -13.18]). 7 departments have confidence intervals which contain only negative values.

2.6.2 2025 maps

Same method with the 2025 predictions.



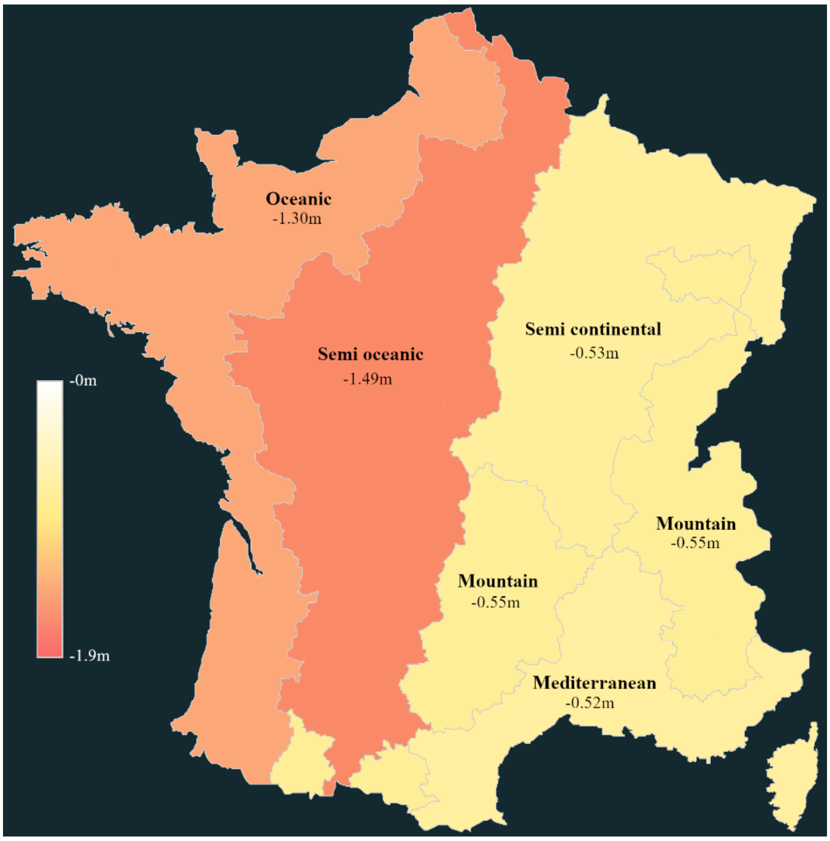
Map 9: Predictions about difference of groundwater level between 2025 and 2000 in France, by basin. Map realised on the GIMP software.

Compared to the map 4, the groundwater level continues to decrease except in the Loire Bretagne basin. It is an unexpected result; the basin should be one of the more impacted by the groundwater decrease.

Bassin	Nb obs	Lower	Mean_diff	Upper	Int_range
Artois-Picardie	10.0	-16.13	-0.58	14.98	31.11
Seine-Normandie	35.0	-22.37	-1.53	19.30	41.66
Rhin-Meuse	58.0	-9.10	-0.45	8.21	17.31
Loire-Bretagne	110.0	-29.85	-1.33	27.18	57.02
Rhône-Méditerranée-Corse	63.0	-14.18	-0.42	13.34	27.52
Adour-Garonne	77.0	-30.61	-1.62	27.38	57.99

Table 12: Predictions about difference of groundwater level between 2025 and 2000 in France, by basin.

We can see that the interval range is booming with the years. The results are less interpretable. However, the middle of the interval remains negative in any case.



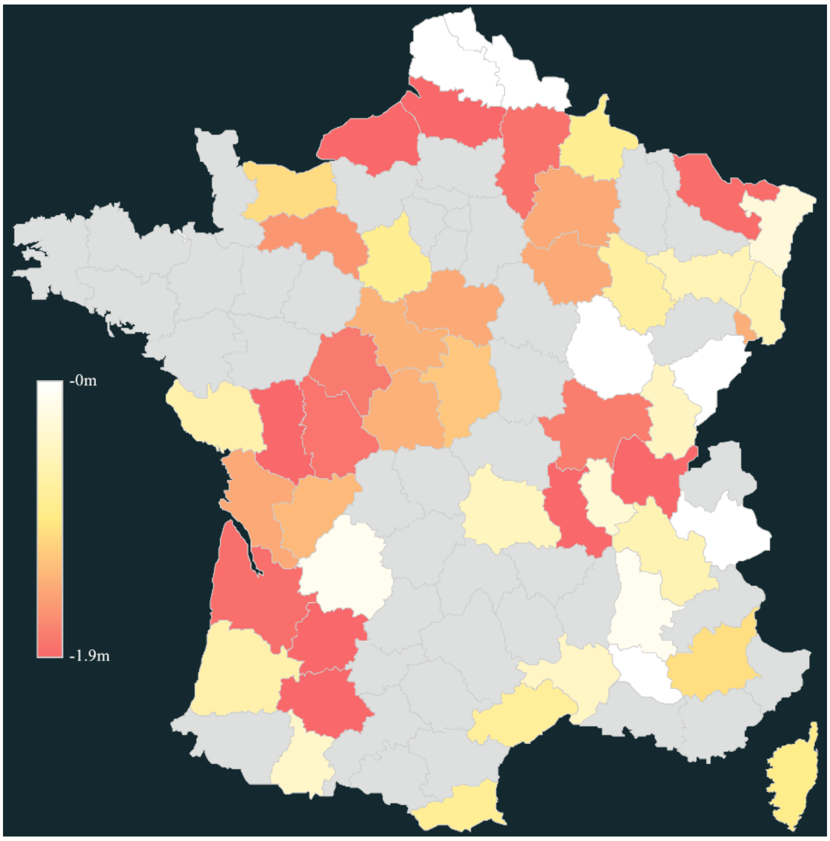
Map 10: Predictions about difference of groundwater level between 2025 and 2000 in France, by climate area. Map realised on the GIMP software.

The level of groundwater is decreasing in all climate area. It is faster in Mountain and Mediterranean which lost 0.10m and 0.08m between the previsions for 2022 and 2025.

Climate	Nb obs	Lower	Mean_diff	Upper	Int_range
Oceanic	95.0	-29.41	-1.30	26.80	56.21
Montain	18.0	-10.89	-0.55	9.79	20.68
Mediterranean	39.0	-13.84	-0.52	12.80	26.64
Semi oceanic	127.0	-27.80	-1.49	24.82	52.62
Semi continental	74.0	-12.76	-0.53	11.70	24.46

Table 13: Predictions about difference of groundwater level between 2025 and 2000 in France, by climate area.

Same observations, the confidence intervals are larger than in 2022 however, the mean difference remains negative.

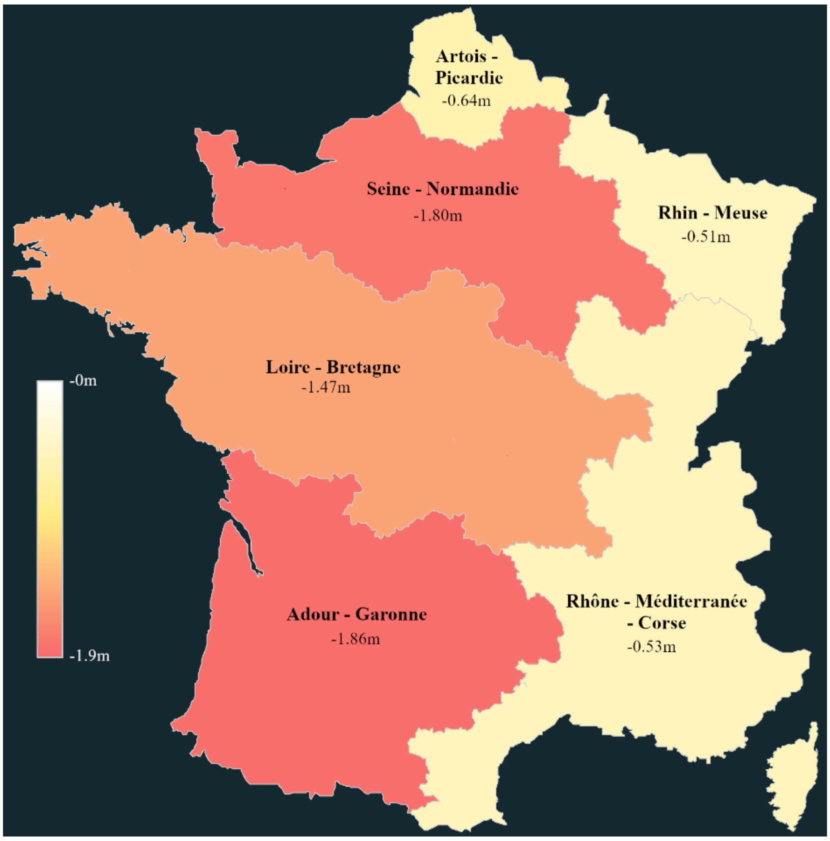


Map 11: Predictions about difference of groundwater level between 2025 and 2000 in France, by department. Map realised on the GIMP software.

We keep the same dynamic, with a majority of departments which are touch by a groundwater decrease. Among them, 4 departments have confidence intervals which contain only negative values.

2.6.3 2030 maps

The 2030 predictions have bigger confidence interval ranges.

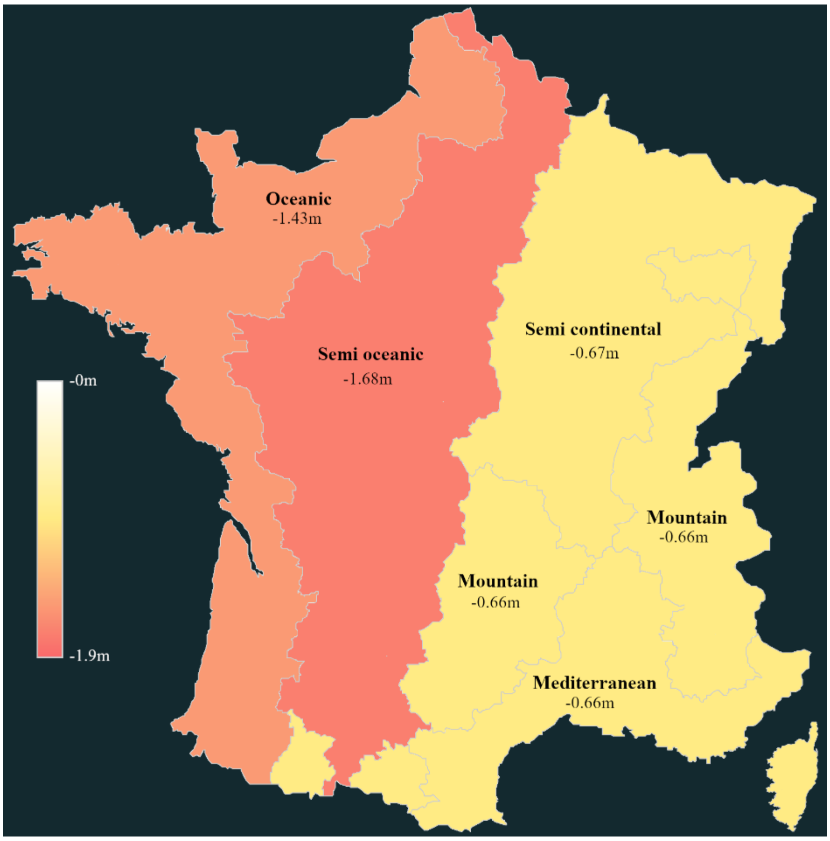


Map 12: Predictions about difference of groundwater level between 2030 and 2000 in France, by basin. Map realised on the GIMP software.

Finally, the predictions for 2030 shows a drop of the groundwater level between 2025 and 2030.

Bassin	Nb obs	Lower	Mean_diff	Upper	Int_range
Artois-Picardie	10.0	-38.61	-0.64	37.34	75.95
Seine-Normandie	35.0	-52.25	-1.80	48.65	100.90
Rhin-Meuse	58.0	-21.47	-0.51	20.45	41.93
Loire-Bretagne	110.0	-70.53	-1.47	67.59	138.12
Rhône-Méditerranée-Corse	63.0	-33.93	-0.53	32.87	66.80
Adour-Garonne	77.0	-72.08	-1.86	68.37	140.46

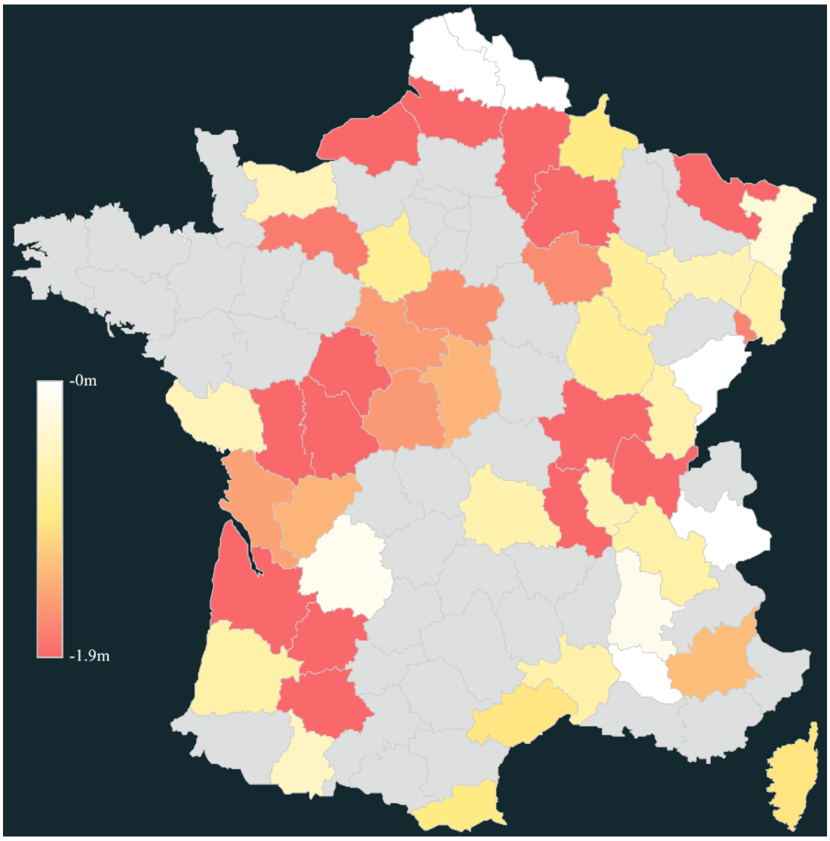
Table 14: Predictions about difference of groundwater level between 2030 and 2000 in France, by basin.



Map 13: Predictions about difference of groundwater level between 2030 and 2000 in France, by climate area. Map realised on the GIMP software.

Climate	Nb obs	Lower	Mean_diff	Upper	Int_range
Oceanic	95.0	-69.56	-1.43	66.70	136.26
Montain	18.0	-25.70	-0.66	24.37	50.07
Mediterranean	39.0	-33.00	-0.66	31.67	64.68
Semi oceanic	127.0	-65.40	-1.68	62.03	127.42
Semi continental	74.0	-30.32	-0.67	28.99	59.31

Table 15: Predictions about difference of groundwater level between 2030 and 2000 in France, by climate area.



Map 14: Predictions about difference of groundwater level between 2030 and 2000 in France, by department. Map realised on the GIMP software.

Same observations as in previous part, our results are hard to interpret with their huge confident interval. It shows a global decrease, but our models are not enough good to give a precise idea of the future level of groundwater.

PART 3: conclusions

This the results of our master project, we show that the iterated time series approach of the level of groundwater in France gives a good idea of the evolution of the groundwater recharge.

Indeed, our results confirm our hypothesis. We observe a general decrease of groundwater volume, and which is more important in the Adour-Garonne, Loire-Bretagne, and Seine-Normandie basins as in the studies described in the Literature Review.

However, we have a huge problem of accuracy with very large confident interval for our predictions. Maybe, a study of the rupture points in the series may improve the performances. In addition, we could try to test more parameters for the ARIMA and SARIMA models.

Another explanation is that climatical and environmental exogenous variables are necessary to give a precise prediction of the future of the groundwater recharge. It can explain in part the differences of predictions between the different climate area.

BIBLIOGRAPHY

Introduction

Wang-Erlandsson, L., Tobian, A., van der Ent, R.J. et al. (2022). Nature Reviews Earth Environment. A planetary boundary for green water. https://doi.org/10.1038/s43017-022-00287-8

Rockström, J., Steffen, W., Noone, K. et al. (2009). Nature. A safe operating space for humanity. 461, 472–475. https://doi.org/10.1038/461472a

Douville, H., K. Raghavan, J. Renwick, R. P. Allan, P. A. Arias, M. Barlow, R. Cerezo-Mota, A. Cherchi, T. Y. Gan, J. Gergis, D. Jiang, A. Khan, W. Pokam Mba, D. Rosenfeld, J. Tierney, O. Zolina. (2021). Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Chapter 8: Water Cycle Changes.

Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (2021). Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC, 2021: Summary for Policymaker. 10. https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_SPM.pdf

National Geographic Society (2022, May 20). Groundwater. National Geographic. https://education.nationalgeographic.org/resource/groundwater

Jean-Christophe Maréchal and Josselin Rouillard. (2020). Sustainable Groundwater Management: A Comparative Analysis of French and Australian Policies and Implications to Other Countries. Groundwater in France: resources, use and management issues. https://hal.archives-ouvertes.fr/hal-02532748

Bureau de Recherches Géologiques et Minières. (2022, April 12). Nappes d'eau souterraine au 1er avril 2022. https://www.brgm.fr/fr/actualite/communique-presse/nappes-eau-souterraine-au-1er-avril-2022

Service des données et études statistiques (SDES) and Office français de la biodiversité (December 2020). FR. Ministère de la Transition écologique. https://economiev2.eaufrance.fr/sites/default/files/2020-12/doc284-datalab_80_chiffres_cles_eau_edition_2020_decembre2020.pdf

Part 1

Masson-Delmotte, V., P. Zhai, A. Pirani, S.L. Connors, C. Péan, S. Berger, N. Caud, Y. Chen, L. Goldfarb, M.I. Gomis, M. Huang, K. Leitzell, E. Lonnoy, J.B.R. Matthews, T.K. Maycock, T. Waterfield, O. Yelekçi, R. Yu, and B. Zhou (2021). Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC, 2021: Summary for Policymaker. 4. https://www.ipcc.ch/report/ar6/wg1/downloads/report/IPCC_AR6_WGI_SPM.pdf

Shiklomaniv Igor A. (1993). Oxford University Press. Water in Crisis, A guide to the Wold's Fresh Water Ressources. Chapter 2.

Delporte Christian. (2015). Le Temps des Médias. René Dumont et la television. 148-161.

Bowen William. (1965). Fortune. How Real Is the Water Shortage?. 14-18

Okun A. Daniel. (1976). American Journal of Public Health. Drinking Water for the Future.

Douville, H., K. Raghavan, J. Renwick, R. P. Allan, P. A. Arias, M. Barlow, R. Cerezo-Mota, A. Cherchi, T. Y. Gan, J. Gergis, D. Jiang, A. Khan, W. Pokam Mba, D. Rosenfeld, J. Tierney, O. Zolina. (2021). Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Chapter 8: Water Cycle Changes. 8.1.1.1. 8.10

Abbott BW, Bishop K, Zarnetske JP,et al. (2019). Hydrological Processes. A water cycle for the Anthropocene. https://doi.org/10.1002/3052 ABBOTT ET AL.hyp.13544

Rutger Willem Hofste, Paul Reig and Leah Schleifer. (2019, October 31). 17 Countries, Home to One-Quarter of the World's Population, Face Extremely High-Water Stress. World Resources Institute. https://www.wri.org/insights/17-countries-home-one-quarter-worlds-population-face-extremely-high-water-stress

Douville, H., K. Raghavan, J. Renwick, R. P. Allan, P. A. Arias, M. Barlow, R. Cerezo-Mota, A. Cherchi, T. Y. Gan, J. Gergis, D. Jiang, A. Khan, W. Pokam Mba, D. Rosenfeld, J. Tierney, O. Zolina. (2021). Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Chapter 8: Water Cycle Changes. 8.1.1.1. 8.7

Jean-Christophe Maréchal and Josselin Rouillard. (2020). Sustainable Groundwater Management: A Comparative Analysis of French and Australian Policies and Implications to Other Countries. Groundwater in France: resources, use and management issues.

https://hal.archives-ouvertes.fr/hal-02532748

Petit Katell and Michon Janik. (2015). Office National de l'Eau et des Milieux Aquatiques. L'état des eaux de surfaces et des eaux souterraines.

Office français de la biodiversité (October 2012). FR. Direction de l'eau et Ministère de la Transition écologique. Synthèse du projet (4 pages) Explore 2070, hydrologie de surface. https://professionnels.ofb.fr/sites/default/files/pdf/explore2070_4pages_hydrologie_surface.pdf

Office français de la biodiversité (October 2012). FR. Direction de l'eau et Ministère de la Transition écologique. Synthèse du projet (synthèse complete) Explore 2070, hydrologie souterraine. 73

https://professionnels.ofb.fr/sites/default/files/pdf/RE_Explore2070_Eaux_Sout_Synthese.pdf

Acharki, S., Amharref, M., El Halimi, R. & Bernoussi, A.-S. (2019). Revue des Sciences de l'Eau / Journal of Water Science. Évaluation par approche statistique de l'impact des changements climatiques sur les ressources en eau: application au périmètre du Gharb (Maroc). 291–315. https://doi.org/10.7202/1067310ar

Ionel Haidu, Mărgărit-Mircea Nistor. (2019). Quaternary International. Groundwater vulnerability assessment in the Grand Est region, France

Lorenzo Galleani, Bartolomeo Vigna, Cinzia Banzato and Stefano LoRusso (2011). Journal of Hydrology. Validation of a Vulnerability Estimator for Spring Protection Areas: The VESPA index https://doi.org/10.1016/j.jhydrol.2010.11.012

ANNEXES

ANNEXE 1 – Previsions by French department for 2022

Year	Department	Nb obs	Lower	Mean_diff	Upper	Int_range
2022	1	2	-3,15	-1,80	-0,45	2,69
2022	2	6	-4,49	-1,57	1,35	5,84
2022	4	1	-1,70	-0,70	0,30	2,00
2022	8	1	-2,98	-0,55	1,88	4,86
2022	10	4	-8,03	-1,19	5,64	13,67
2022	14	6	-5,40	-1,21	2,97	8,37
2022	16	13	-8,11	-1,47	5,18	13,30
2022	17	15	-8,93	-1,46	6,02	14,95
2022	18	13	-5,75	-0,82	4,11	9,86
2022	20	3	-3,40	-0,55	2,30	5,70
2022	21	4	-5,40	0,49	6,38	11,79
2022	24	2	-1,31	-0,08	1,15	2,46
2022	25	1	-2,09	0,17	2,42	4,51
2022	26	7	-1,56	-0,07	1,41	2,97
2022	28	6	-1,82	-0,70	0,42	2,24
2022	30	12	-2,63	-0,16	2,30	4,93
2022	32	1	-12,38	-11,38	-10,38	2,01
2022	33	28	-3,59	-1,50	0,60	4,19
2022	34	12	-2,64	-0,52	1,59	4,22
2022	36	14	-4,65	-1,23	2,20	6,86
2022	37	10	-3,97	-1,56	0,84	4,81
2022	38	1	-3,31	-0,33	2,64	5,95
2022	39	2	-1,60	-0,30	1,01	2,60
2022	40	16	-4,85	-0,51	3,83	8,68
2022	41	8	-3,45	-1,15	1,15	4,59
2022	42	1	-2,18	-2,08	-1,98	0,21
2022	45	7	-3,37	-1,19	0,98	4,35
2022	47	1	-14,11	-13,64	-13,18	0,93
2022	51	4	-3,33	-0,66	2,02	5,36
2022	52	2	-1,47	-0,48	0,51	1,98
2022	54	3	-6,94	-1,73	3,49	10,43
2022	59	5	-1,69	0,37	2,43	4,12
2022	61	1	-2,32	-1,37	-0,43	1,89
2022	62	1	-1,20	0,78	2,76	3,96
2022	63	4	-1,43	-0,20	1,03	2,46
2022	65	1	-0,68	-0,27	0,15	0,84
2022	66	10	-1,64	-0,58	0,48	2,12
2022	67	24	-0,95	-0,21	0,53	1,49
2022	68	28	-1,55	-0,43	0,69	2,24
2022	69	2	-1,30	-0,12	1,06	2,36
2022	71	1	-2,27	-1,40	-0,54	1,73
2022	73	3	-0,99	0,04	1,08	2,07
2022	76	6	-5,97	-2,91	0,15	6,13
2022	79	16	-8,47	-2,11	4,26	12,73
2022	80	4	-4,69	-1,97	0,74	5,43
2022	84	2	-0,95	0,14	1,23	2,18
2022	85	18	-6,19	-0,75	4,69	10,89
2022	86	18	-6,96	-2,00	2,95	9,91
2022	88	2	-4,19	-0,48	3,24	7,43
2022	90	1	-1,25	-1,08	-0,90	0,35

ANNEXE 2 – Previsions by French department for 2025

Year	Department	Nb obs	Lower	Mean_diff	Upper	Int_range
2025	1	2	-11,03	-1,97	7,09	18,13
2025	2	6	-21,33	-1,82	17,68	39,01
2025	4	1	-7,50	-0,88	5,75	13,25
2025	8	1	-16,88	-0,69	15,49	32,36
2025	10	4	-47,26	-1,34	44,58	91,84
2025	14	6	-28,91	-0,90	27,11	56,02
2025	16	13	-45,84	-1,20	43,44	89,28
2025	17	15	-51,72	-1,34	49,03	100,75
2025	18	13	-34,45	-1,09	32,27	66,72
2025	20	3	-19,95	-0,73	18,48	38,43
2025	21	4	-40,08	0,55	41,18	81,26
2025	24	2	-8,33	-0,10	8,13	16,46
2025	25	1	-14,98	0,18	15,34	30,31
2025	26	7	-10,04	-0,09	9,86	19,90
2025	28	6	-8,21	-0,69	6,83	15,04
2025	30	12	-16,92	-0,36	16,19	33,11
2025	32	1	-19,60	-12,76	-5,93	13,67
2025	33	28	-16,01	-1,84	12,33	28,34
2025	34	12	-15,07	-0,63	13,80	28,87
2025	36	14	-24,32	-1,27	21,77	46,09
2025	37	10	-17,89	-1,75	14,39	32,28
2025	38	1	-20,72	-0,47	19,78	40,50
2025	39	2	-9,12	-0,39	8,34	17,46
2025	40	16	-29,95	-0,52	28,90	58,84
2025	41	8	-16,64	-1,26	14,11	30,74
2025	42	1	-3,15	-2,41	-1,68	1,46
2025	45	7	-15,95	-1,34	13,28	29,23
2025	47	1	-18,58	-15,52	-12,45	6,13
2025	51	4	-19,21	-1,34	16,53	35,74
2025	52	2	-7,22	-0,59	6,04	13,26
2025	54	3	-36,78	-1,86	33,06	69,85
2025	59	5	-13,21	0,39	14,00	27,21
2025	61	1	-7,71	-1,51	4,69	12,40
2025	62	1	-12,09	1,01	14,10	26,19
2025	63	4	-8,83	-0,41	8,01	16,84
2025	65	1	-3,11	-0,33	2,44	5,55
2025	66	10	-8,19	-0,66	6,87	15,07
2025	67	24	-5,21	-0,23	4,75	9,97
2025	68	28	-8,02	-0,47	7,08	15,10
2025	69	2	-8,15	-0,27	7,61	15,76
2025	71	1	-7,52	-1,72	4,08	11,59
2025	73	3	-6,90	0,04	6,98	13,88
2025	76	6	-24,01	-3,30	17,41	41,42
2025	79	16	-45,13	-1,90	41,34	86,47
2025	80	4	-20,80	-2,19	16,41	37,21
2025	84	2	-7,06	0,22	7,50	14,56
2025	85	18	-37,24	-0,52	36,20	73,44
2025	86	18	-35,15	-1,80	31,56	66,70
2025	88	2	-25,49	-0,46	24,56	50,05
2025	90	1	-2,48	-1,30	-0,11	2,37

ANNEXE 3 – Previsions by French department for 2030

Year	Department	Nb obs	Lower	Mean_diff	Upper	Int_range
2030	1	2	-24,17	-2,25	19,68	43,86
2030	2	6	-49,48	-2,31	44,86	94,34
2030	4	1	-17,17	-1,17	14,83	32,00
2030	8	1	-39,89	-0,78	38,32	78,21
2030	10	4	-112,94	-1,59	109,75	222,69
2030	14	6	-68,19	-0,45	67,29	135,48
2030	16	13	-109,29	-1,25	106,79	216,08
2030	17	15	-123,32	-1,40	120,52	243,84
2030	18	13	-82,01	-1,24	79,53	161,54
2030	20	3	-47,43	-0,84	45,76	93,19
2030	21	4	-98,13	0,73	99,58	197,70
2030	24	2	-20,00	-0,09	19,81	39,81
2030	25	1	-36,42	0,24	36,91	73,33
2030	26	7	-24,19	-0,12	23,95	48,15
2030	28	6	-19,01	-0,68	17,65	36,66
2030	30	12	-40,60	-0,54	39,53	80,13
2030	32	1	-31,77	-15,07	1,64	33,41
2030	33	28	-36,59	-2,24	32,11	68,70
2030	34	12	-35,94	-0,81	34,32	70,26
2030	36	14	-57,23	-1,47	54,28	111,51
2030	37	10	-41,10	-2,04	37,03	78,13
2030	38	1	-49,61	-0,56	48,49	98,11
2030	39	2	-21,64	-0,52	20,59	42,23
2030	40	16	-71,92	-0,55	70,83	142,75
2030	41	8	-38,64	-1,45	35,74	74,38
2030	42	1	-4,72	-2,93	-1,14	3,57
2030	45	7	-36,90	-1,55	33,81	70,71
2030	47	1	-26,03	-18,64	-11,25	14,78
2030	51	4	-45,71	-2,48	40,75	86,46
2030	52	2	-16,66	-0,63	15,39	32,05
2030	54	3	-86,84	-2,05	82,74	169,59
2030	59	5	-32,46	0,44	33,34	65,80
2030	61	1	-16,69	-1,73	13,23	29,92
2030	62	1	-30,25	1,38	33,01	63,26
2030	63	4	-20,95	-0,51	19,93	40,88
2030	65	1	-7,08	-0,38	6,33	13,42
2030	66	10	-19,16	-0,77	17,61	36,78
2030	67	24	-12,32	-0,26	11,80	24,11
2030	68	28	-18,84	-0,55	17,73	36,56
2030	69	2	-19,53	-0,47	18,59	38,12
2030	71	1	-16,11	-2,09	11,93	28,04
2030	73	3	-16,74	0,05	16,83	33,57
2030	76	6	-53,99	-3,82	46,35	100,34
2030	79	16	-106,87	-2,03	102,80	209,66
2030	80	4	-48,40	-2,49	43,41	91,81
2030	84	2	-17,36	0,25	17,86	35,22
2030	85	18	-89,44	-0,43	88,58	178,02
2030	86	18	-82,70	-1,95	78,80	161,49
2030	88	2	-61,06	-0,48	60,10	121,16
2030	90	1	-4,52	-1,65	1,22	5,74