Adrien Chabert

Table des matières

[Introduction 2](#_Toc11797678)

[Installation 3](#_Toc11797679)

[Humidité du sol 4](#_Toc11797680)

[Facteurs impactant l’humidité 4](#_Toc11797681)

[Taux d’humidité optimal 4](#_Toc11797682)

[Recherche de littérature 6](#_Toc11797683)

[ConservWater 6](#_Toc11797684)

[Autre études 7](#_Toc11797685)

[Machine Learning 8](#_Toc11797686)

[Linear Regression 8](#_Toc11797687)

[Méthodologie 8](#_Toc11797688)

[Analyse des résultats 9](#_Toc11797689)

[Elaboration d’un programme d’arrosage 9](#_Toc11797690)

[Point à améliorer 9](#_Toc11797691)

[Conclusion 9](#_Toc11797692)

## Introduction

L’eau, H20 pour sa composition atomique, est à première vue une ressource abondante sur notre terre. En effet, plus de 70% de la surface de la terre est recouvert d’eau. On la retrouve sous forme liquide (mers, lacs, rivières, nappes), sous forme de glace et sous forme de gaz (dans les nuages, dans l’atmosphère). Cependant sur l’ensemble de l’eau présente sur Terre, seulement 3% de cette eau est de l’eau douce, soit 35 millions de kilomètres cubes. A titre de comparaison, le lac Léman a un volume de 89km3.

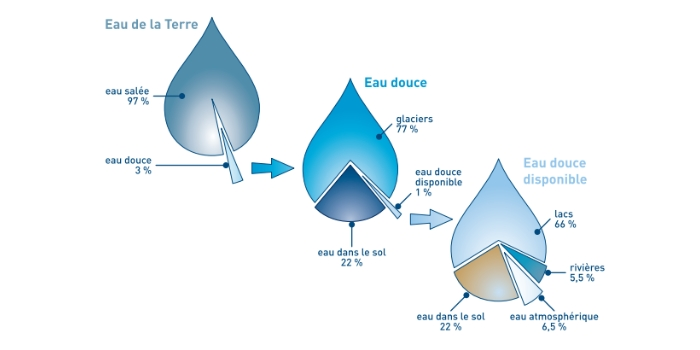


Figure 1 Eau sur Terre, http://www.septiemecontinent.com/pedagogie/lesson/eau-terre-origine-repartition/

Sur ces 2,5% d’eau douce 77% se trouve sous forme de glace et 22% sont profondément enfoui sous Terre. Ainsi sur l’ensemble de l’eau douce sur terre, seulement 1% de celle-ci est liquide et disponible que ce soit sous forme de lac, de rivière, dans les nappes phréatiques ou dans l’atmosphère. De plus, la Russie, l'Inde, le Canada, les États-Unis, l'Indonésie, le Congo et la Chine se partagent à eux neufs 60% du débit mondial d’eau. De ce fait, l’or bleu est inégalement réparti parmi les êtres humains sur Terre.

L’eau douce est essentielle à la vie sur Terre pour tous les êtres vivants. Pour l’Homme elle est essentielle pour l’hydratation, pour l’agriculture, pour l’usage ménager ou encore pour l’industrie. La consommation mondiale d’eau douche en 2000 est 4 km3 par an soit 1,3 millions de litres par an et par habitant. La demande en eau augmentera en vue des prévisions d’augmentation de la population mondiale. En considérant que 70% de notre consommation en eau est utilisé par l’agriculture et que de nombreux pays se trouve déjà en stresse hydrique une grande partie de l’année, rendre l’agriculture moins demandant en eau est un des enjeux majeurs de l’humanité pour les prochaines années.

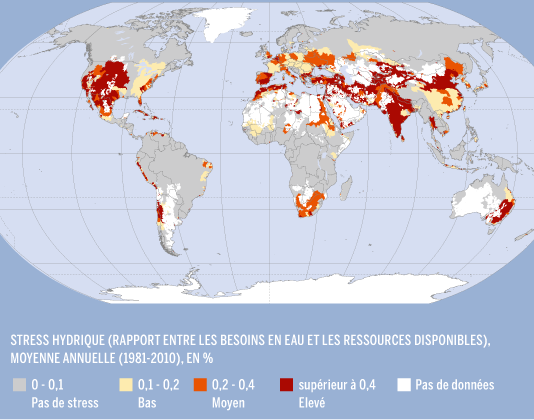


Figure 2: Carte du monde du stresse hydrique, <https://www.lemonde.fr/ressources-naturelles/article/2015/03/20/la-crise-de-l-eau-illustree-en-5-graphiques_4597592_1652731.html>

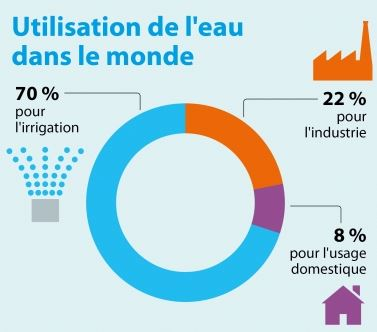


Figure 3Utilisation de l'eau, https://www.planetoscope.com/consommation-eau/239-consommation-d-eau-dans-le-monde.html

Différentes méthodes ont déjà été étudier comme cultiver des espèces moins qui nécessite une moins grande quantité d’eau ou encore améliorer les systèmes d’irrigation. C’est ce dernier point qui nous intéresse pour mon projet, et plus précisément d’arroser les plants de façon des plus efficient. Irriguer les champs avec uniquement la quantité d’eau qui est nécessaire au bien-être des plantations et de suivre un planning d’arrosage des plus efficients, permettent d’éviter un surplus d’arrosage et donc d’un gaspillage d’eau.

Ainsi le but de mon projet est de décider de la quantité d’eau qu’il faut arroser pour que la terre reste à une certaine humidité idéale à l’évolution de la plante. Ce taux d’humidité est décidé en fonction de la plante et du type de terre.

## Installation

Une importante phase du projet est la phase de récolte de donnée. Pour ce faire, deux bacs à plantes ont été mis à disposition avec pour chacun un arrosage automatique journalier programmable et des capteurs sensorielles. Les bacs se trouve dans des pièces fermées à côté de fenêtre. Un premier bac a été placé dans une salle bien-exposé, orienté sud-ouest. Ce bac sera nommé *Demeter*. Dans celui-ci est planté des oignons et du basilic. Le deuxième est placé dans une salle moins bien exposé, orienté nord-est et sera nommé *Ceres*. Dans ce bac, des épinards sont cultivés. Le choix de ce placement a été choisi car les épinards demandent une plus faible exposition que le basilic et les oignons. Dans les bacs, un arrosage goute à goute pour chaque pied de plante a été installé. Les capteurs sont placés à environ deux centimètres de la zone d’arrosage du goute à goute.

Image capteur et bac

Chaque bac dispose de deux capteurs enterrés et d’un capteur de luminosité indépendant. Ces capteurs fournissent de très nombreuses informations utiles au calcul de l’évolution de l’humidité dans le sol. Les deux capteurs enterrés permettent de récolter des informations sur l’humidité, la température et l’ensoleillement au-dessus du capteur. Les données sont récoltées toutes les 30 secondes.

## Humidité du sol

Le pourcentage d’humidité dans le sol est le facteur principal de ce projet. Le but de ce projet étant de prédire la quantité d’eau à arroser pour est à une certaine quantité d’humidité. Ce taux est calculé en pourcentage. Mettre le capteur dans de l’eau revient à obtenir une humidité de 100%.

### Facteurs impactant l’humidité

L’humidité dans le sol est influencée par différents facteurs. Selon les études, il s’agit de l’ensoleillement, la quantité d’eau arrosé, le vent, le type de terre, la qualité de la terre et des plants qui pompent l’eau dans la terre. Le vent qui est un facteur avec une forte influence à l’extérieur, le vent assèche les terres, n’a pas été pris en compte dans notre expérience car les bacs sont à l’intérieur de salle fermé. Le type de terre n’est également pas pris en compte, car la même terre a été utilisé dans les deux bacs. Il s’agit de terre universelle qu’on peut trouver communément dans le commerce. L’ensoleillement n’a pas été pris en compte également car les capteurs ont été vite protéger du soleil par les feuilles des plantes et que l’ensoleillement est fortement corrélé à la température.

L’évolution de la plante qui n’a pas été considéré car celui-ci est trop difficile à calculer et les résultats qui seront exposés par la suite montre le faible impacte de ce capteur sur l’humidité pendant l’expérience.

Ainsi la température, l’humidité et l’évolution de la terre ont été pris en compte pour l’analyse du taux d’humidité dans le sol.

### Taux d’humidité optimal

Le taux d’humidité optimal pour une plante dépend du type de sol. En effet, chaque type de sol a des caractéristiques différentes. La capacité de rétention et le point de flétrissement permanent sont les caractéristiques essentielles pour déterminer l’humidité d’un sol.

La capacité de rétention correspond à la quantité maximale d’eau qu’un sol peut contenir. Ce taux peut être calculer en quantifiant la quantité d’eau deux à trois jours après un arrosage saturant le sol. En dessus de ce taux, il est inutile d’arroser.

Le point de flétrissement permanent correspond à l’humidité du sol pour laquelle une plante à l’aide de ses racines n’est plus capable d’en extraire l’eau. Le sol est trop sec. Entre ces deux taux, se trouve la quantité d’eau disponible. Plus cette quantité d’eau est grande, plus un sol est propice à ……. Pour l’arrosage par goute à goute, il est conseillé d’arroser lorsqu’il ne reste plus que 80% d’eau disponible. Alors que pour l’irrigation par aspersion, il est conseillé d’irriguer à partir de 50% d’eau disponible.

Le sol utilisé pour notre expérience est une terre universelle qui peut facilement être acheter dans le commerce. Ceci correspond à un sol de type de loam limoneux. Ainsi le but de notre logiciel sera de rester au-dessus de 25% d’humidité dans notre sol.

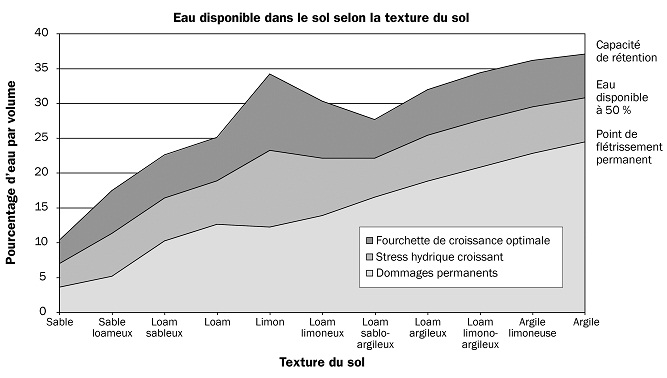


Figure 4 : Humidité optimal pour un sol. <http://www.omafra.gov.on.ca/french/engineer/facts/11-038.htm>

## Recherche de littérature

### ConservWater

Une large étude a été mené par l’institut de technologie de Californie à Pasedena et cette dernière a publié un approfondi article dans le IJOEAR (Journal international de recherche en environnement et en agriculture) nommé « *Global custom-tailored machine learning of soil water content for locale specific irrigation modeling with high accuracy »* écrit par Aadith Moorthy. Cet article fait une analyse de la situation actuelle, un comparatif des résultats de prédiction d’arrosage en utilisant différents algorithmes et propose un algorithme de Machine Learning dédié à la prédiction d’arrosage pour la culture de plants dans le monde entier. Ils ont créé un algorithme utilisant du machine Learning. Cet algorithme est nommé ConserWater™ algorithm.

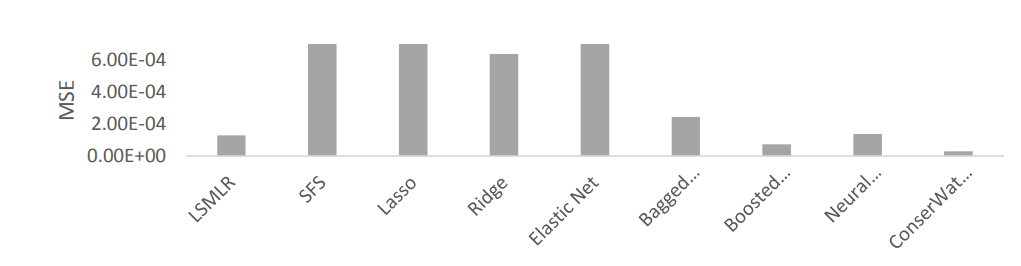
Selon leurs analyses, ils ont montré que leur algorithme se comporte mieux que des algorithmes statiques tel que Aqua Crop, proposé par Land and Water Division of the Food and Agricultural Organization (FAO). Leur algorithme obtient également de meilleurs résultats que des algorithmes de Machine Learning basé sur les itérations dans le temps tels que ARIMA et SARIMA. Ceci vient du fait que ARIMA et SARIMA aient besoin de données avec des propriétés statiques. Ceci n’est pas le cas étant que la météo ne suit aucune loi statiques (déviation standard ou moyenne). 

Figure 5: Erreur moyenne au carré (MSE) sur 30 jours pour différents algorithmes en ne donnant que l'humidité de la première journée, <https://ijoear.com/Paper-October-2016/IJOEAR-SEP-2016-34.pdf>

On peut voir que leur algorithme (ConserWater) obtient un bien meilleur résultat que d’autres algorithmes de machine Learning classique et fréquemment utilisé. Il est intéressant de noter que la régression linéaire multiple moindre carré (LSMLR) obtient de très bon résultat par rapport aux autres algorithmes de Machine Learning. Ceci est d’autant plus intéressant que c’est un algorithme avec une faible complexité d’implémentation.

A noter qu’il est important de séparer les données par rapport à leur région. L’humidité ne réagit pas de façon identique dans une région aride que dans une région tropicale.

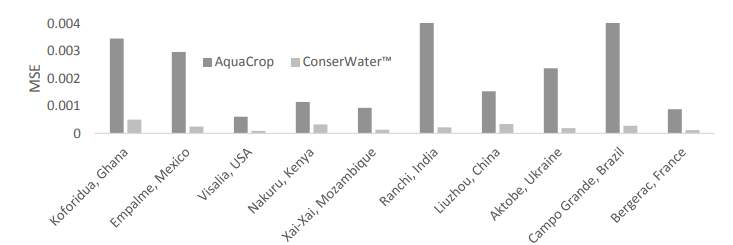


Figure 6: Erreur MSE entre Aqua Crop et ConserWater suivant les régions, <https://ijoear.com/Paper-October-2016/IJOEAR-SEP-2016-34.pdf>

L’article affirme que l’utilisation de machine Learning peut permettre une économie d’eau de 25% par rapport à l’usage habituel. Mais que l’utilisation de capteurs sensorielles installés par les maraichers peut revenir à un coût rédhibitoire pour les petites fermes. C’est pourquoi ils ont développé un algorithme utilisant les données récoltées par les nombreux capteurs météorologique déjà existant, notamment en utilisant le *World Weather online service*. Ils ont à disposition 30'000 stations de collecte de données. Les données ont été collecté sur un peu plus de 1 an.

Une telle étude n’est pas du tout comparable à l’étude qui a été pour ce projet. Le nombre de données collecté est amplement supérieur aux nôtres. De plus cette étude californienne a été mené en extérieur sur différents continents. Ainsi le nombre de facteur impactant l’humidité est beaucoup plus important que pour notre étude en intérieur. Il leur a fallu prendre en compte le vent, les averses, la condensation, la rosée, etc. Cependant il est intéressant de noter qu’une simple linéaire régression permet d’obtenir des résultats très satisfaisant et que l’économie possible réalisé est conséquente.

### Autre études

#### Estimating soil moisture using remote sensing data: A machine learning approach

C’est une étude réalisée en 2009 sur des données récoltées au Colorado pendant 5 ans. Il s'agit d'estimer l'humidité du sol avec les données du capteur en utilisant la technique appelée Support Vector Machine (SVM). "SVM est basé sur une théorie statistique d'apprentissage qui utilise un espace d'hypothèses de fonctions linéaires basé sur une approche par noyau." Ils ont obtenu de meilleurs résultats avec SVM qu'avec les modèles de réseau de neurones et qu’avec la régression linéaire à variables multiples.

#### Planning d’arrosage hebdomadaire avec Jojoba Israel

Jojoba est une société qui vend des capteurs d’humidité. L’entreprise a réalisé une étude sur des données récolté à l’aide de 22 capteurs sur 2 ans. Le but de leur expérience était de faire un planning hebdomadaire d’arrosage. Ils ont comparé différents algorithmes d’apprentissage et ont obtenu que les meilleures résultats était obtenu avec Gradient Boosted Regression Trees , avec un taux de précision de 93%. Ils ont jugé les résultats satisfaisants et qu’ils aidaient les ingénieur agronome à définir leur arrosage.

## Machine Learning

### Linear Regression

Ordinary Least Squares: linear model with coefficients w = (w1, …, wp) in purpose to minimize the residual sum of square between predicted approximation and observed dataset. When there is some dependencies between the columns, the least-squares become very sensitive to random errors. It’s problematic when data are collected without protocol and experimental design.

Ridge Regression: add a penalty L2 on the size of the coefficient. Higher is alpha, more robust to collinearity are the coefficients. Ordinary linear least squares regression can lead to overdetermined or underdetermined system of equations. With L2 regularization, it gives a preference to a particular solution with desirable properties.

Lasso: It use when you want to reduce the number of variables with which the given solution depends. Minimize the least-square with a L1 penalty. Some variable can have the value of 0 in contrast with ridge regression (rasoir d’Ocklam). If there is a strong correlation between component, Lasso will privilege one in the detriment of another.

Elastic net regularization: It combine the L1 and L2 penalties of the lasso and ridge methods. If there some component that are highly correlated, Lasso tend to select one, not Elastic net.

Bayesian Regression: Regression based on prior. Y is assumed to be Gaussian distributed. You can regularization parameters in the estimation procedure. Advantage: it adapts well to the data at hand.

ARIMA (Autoregressive integrated moving average): It’s fitted to time series data. It used when data show evidence of non-stationarity where a differencing step can be adding to eliminate it. Non-stationarity imply that static property as by example mean, variance are independent of the time.

Wikipedia: “The AR part of ARIMA indicates that the evolving variable of interest is regressed on its own lagged (i.e., prior) values. The MA part indicates that the regression error is actually a linear combination of error terms whose values occurred contemporaneously and at various times in the past. The I (for "integrated") indicates that the data values have been replaced with the difference between their values and the previous values (and this differencing process may have been performed more than once). The purpose of each of these features is to make the model fit the data as well as possible.”

SARIMA (Seasonal Autoregressive integrated moving average): It take in account season. When you implement it you have to add the number of period in each season.

## Méthodologie

Deux capteurs d’humidité et de température ce trouve dans chacun des deux bacs à plantes. Les données ont été récoltés sur trois mois, de mi-mars à mi-juin. Les deux bacs ne sont pas exposés pareil et ne contiennent pas les mêmes plantes. Les épinards, qui ont été planté dans le bac Ceres, sont arrivés à maturité deux mois après les avoir plantés, c’est-à-dire mi-mai. Alors que les basilics et oignons étaient toujours en bonne santé mi-juin. Ces différentes se manifestent sur les courbes d’humidité des capteurs de leur bac respectif. C’est pourquoi j’ai décidé de ne pas rassembler les données des deux bacs et d’appliquer des algorithmes d’apprentissage indépendamment entre les bacs. Les données sont récoltées toutes les 30 minutes afin de récupérer une moyenne sur 30 minutes du facteur. Ceci permet d’éviter l’apparition de bruit sur les valeurs obtenus. Chaque facteur suit ainsi une courbe homogène.

L’arrosage est journalier et toujours programmer à 11h pour l’heure d’hiver et midi pour l’heure d’été. Ainsi la représentation graphique de chaque journée comporte un pic à 11h ou 12h, puis une phase d’évaporation rapide de l’eau et enfin une troisième phase de stabilisation de l’humidité. La deuxième et troisième phases sont beaucoup plus marquées pour Demeter que Ceres.

Afin de trouver le meilleur algorithme pour prédire l’humidité sur une période donnée et ainsi créer un programme d’arrosage précis, j’ai effectué un comparatif entre plusieurs algorithmes d’apprentissage. Le but étant de faire la meilleure prédiction d’humidité sur une journée entière. Chacune des journées ont été séparées et ainsi chaque journée était considérée comme un tout et permettait d’améliorer l’algorithme de prédiction. Une seule régression était faite sur toute la journée.

Afin d’améliorer ma précision, chaque journée est séparée en 3 parties. Chaque partie représente une phase décrit ci-dessus dans ce même chapitre. Ceci a grandement permis d’améliorer la précision. La première partie est uniquement les 30 minutes qui suivent l’arrosage. La deuxième et troisième partie se partage le reste de la journée. Il a fallu trouver la meilleure séparation des deux parties qui minimise l’erreur de la prédiction.

## Analyse des résultats

## Elaboration d’un programme d’arrosage

## Point à améliorer

## Conclusion