



Prévision de la profondeur de la nappe phréatique d'un champ de canneberges à l'aide de deux approches de modélisation des arbres de décision

Mémoire

Jhemson Brédy

Maîtrise en génie agroalimentaire - avec mémoire
Maître ès sciences (M. Sc.)

Québec, Canada

© Jhemson Brédy, 2019

**Prévision de la profondeur de la nappe phréatique d'un champ
de canneberges à l'aide de deux approches de modélisation des
arbres de décision**

Mémoire

Jhemson Brédy

Sous la direction de :

Silvio José Gumiere, directeur de recherche

Jacques Gallichand, co-directeur de recherche

Résumé

La gestion intégrée de l'eau souterraine constitue un défi majeur pour les activités industrielles, agricoles et domestiques. Dans certains systèmes agricoles, une gestion optimisée de la nappe phréatique représente un facteur important pour améliorer les rendements des cultures et l'utilisation de l'eau. La prévision de la profondeur de la nappe phréatique (PNP) devient l'une des stratégies utiles pour planifier et gérer en temps réel l'eau souterraine. Cette étude propose une approche de modélisation basée sur les arbres de décision pour prédire la PNP en fonction des précipitations, des précédentes PNP et de l'évapotranspiration pour la gestion de l'eau souterraine des champs de canneberges. Premièrement, deux modèles : « Random Forest (RF) » et « Extreme Gradient Boosting (XGB) » ont été paramétrisés et comparés afin de prédire la PNP jusqu'à 48 heures. Deuxièmement, l'importance des variables prédictives a été déterminée pour analyser leur influence sur la simulation de PNP. Les mesures de PNP de trois puits d'observation dans un champ de canneberges, pour la période de croissance du 8 juillet au 30 août 2017, ont été utilisées pour entraîner et valider les modèles. Des statistiques tels que l'erreur quadratique moyenne, le coefficient de détermination et le coefficient d'efficacité de Nash-Sutcliffe sont utilisés pour mesurer la performance des modèles. Les résultats montrent que l'algorithme XGB est plus performant que le modèle RF pour prédire la PNP et est sélectionné comme le modèle optimal. Parmi les variables prédictives, les valeurs précédentes de PNP étaient les plus importantes pour la simulation de PNP, suivie par la précipitation. L'erreur de prédiction du modèle optimal pour la plage de PNP était de ± 5 cm pour les simulations de 1, 12, 24, 36 et 48 heures. Le modèle XGB fournit des informations utiles sur la dynamique de PNP et une simulation rigoureuse pour la gestion de l'irrigation des canneberges.

Mots-clés: Forêt aléatoire, arbres de décision renforcés, apprentissage automatique, niveau d'eau souterraine, évapotranspiration, précipitation

Abstract

Integrated groundwater management is a major challenge for industrial, agricultural and domestic activities. In some agricultural production systems, optimized water table management represents a significant factor to improve crop yields and water use. Therefore, predicting water table depth (WTD) becomes an important means to enable real-time planning and management of groundwater resources. This study proposes a decision-tree-based modelling approach for WTD forecasting as a function of precipitation, previous WTD values and evapotranspiration with applications in groundwater resources management for cranberry farming. Firstly, two models-based decision trees, namely Random Forest (RF) and Extrem Gradient Boosting (XGB), were parameterized and compared to predict the WTD up to 48-hours ahead for a cranberry farm located in Québec, Canada. Secondly, the importance of the predictor variables was analyzed to determine their influence on WTD simulation results. WTD measurements at three observation wells within a cranberry field, for the growing period from July 8, 2017 to August 30, 2017, were used for training and testing the models. Statistical parameters such as the mean squared error, coefficient of determination and Nash-Sutcliffe efficiency coefficient were used to measure models performance. The results show that the XGB algorithm outperformed the RF model for predictions of WTD and was selected as the optimal model. Among the predictor variables, the antecedent WTD was the most important for water table depth simulation, followed by the precipitation. Base on the most important variables and optimal model, the prediction error for entire WTD range was within ± 5 cm for 1-, 12-, 24-, 26- and 48- hour prediction. The XGB model can provide useful information on the WTD dynamics and a rigorous simulation for irrigation planning and management in cranberry fields.

Key words: Random Forest, Extreme Gradient Boosting, machine learning, groundwater level, evapotranspiration, precipitation

Table des matières

Résumé.....	III
Abstract	IV
Table des matières	V
Liste des tableaux	VII
Liste des figures	VIII
Liste des sigles	IX
Remerciement	XI
Avant-propos.....	XII
Introduction générale.....	1
1.1. Objectif général de l'étude	2
1.2. Objectifs spécifiques.....	2
1.3. Hypothèse	2
1.4. Portée de l'étude	2
Bibliographie.....	3
Chapitre 1 Revue de littérature.....	4
2.1. Situation de l'eau souterraine dans le monde	5
2.2. La canneberge et ses besoins en eau.....	5
2.3. Gestion de l'eau dans les champs de canneberges	6
2.4. Relation entre la profondeur de la nappe phréatique, la précipitation et l'évapotranspiration	7
2.5. Apprentissage automatique.....	9
2.6. Modélisation du niveau d'eau souterraine.....	10
Bibliographie.....	13
Chapitre 2 Water table depth forecasting in cranberry fields using two decision tree-modeling approaches.....	17
2.1. Résumé	18
2.2. Abstract.....	19
2.3. Introduction	20
2.4. Methodology.....	23
2.4.1. Study area	23
2.4.2. Data.....	23
2.4.2.1. Water table depth.....	23

2.4.2.2.	Precipitation.....	24
2.4.2.3.	Evapotranspiration.....	25
2.4.3.	Training with different decision tree-based models	26
2.4.3.1.	Random Forest.....	26
2.4.3.2.	Extreme Gradient Boosting	28
2.4.4.	Model development	29
2.4.4.1.	Input selection	29
2.4.4.2.	Data division.....	30
2.4.4.3.	Model calibration and validation.....	30
2.4.4.4.	Importance of predictor variables.....	31
2.4.5.	Performance criteria	32
2.5.	Results and discussion	34
2.5.1.	Groundwater movement	34
2.5.2.	Models parameterization	36
2.5.3.	Comparison of RF and XGB models.....	37
2.5.4.	Predictive variables relative importance.....	40
2.5.5.	Predicting water table depth	44
2.6.	Conclusions	48
	Acknowledgements	49
	Bibliographie.....	50
	Conclusion générale.....	54
	Bibliographie générale	56

Liste des tableaux

Table 1. Characteristics of the Extreme Gradient Boosting model employed for forecasting water table depth (WTD) 1, 12, 24, 36 and 48 hours ahead.	37
Table 2. RMSE (cm) values in training (Tr) and test (Ts) stages for both models (RF and XGB) in forecasting water table depth (WTD) at lead-time predictions of 1, 12, 24, 36 and 48-time steps for each observation well (P1, P2 and P3).	39
Table 3. NSE values in training (Tr) and test (Ts) stages for both models (RF and XGB) models in forecasting water table depth at lead-time predictions of 1, 12, 24, 36- and 48-time steps for each observation well (P1, P2 and P3).....	40
Table 4. XGB model testing accuracy obtained by the feature selection approach. Model I: all predictive variables (P, WTD and ET); Model II: the most important variables (P and WTD) for forecasting 1, 12, 24, 36 and 48 hours ahead the water table depth in the observation wells (P1, P2 and P3).....	44

Liste des figures

Fig. 1. Schéma du mémoire	XIII
Fig. 2. Approche typique de l'apprentissage automatique tirée de Liakos et al. (2018)	10
Fig. 3. Study area, Québec, Canada. Red landmarks represent the locations of water table depths observation wells.....	24
Fig. 4. The flowchart of Random Forest adapted from Guo et al. (2011).....	28
Fig. 5. Hourly measurements of (A) water table depths in the observation wells (P1, P2 and P3), precipitation and (B) evapotranspiration (ET) between July 8 to August 30, 2017.	35
Fig. 6. Correlation coefficient between 24-hour moving average water table depth (WTD) and climate variables as a function of time delay, with 24-hour moving average precipitation (dashed line) and evapotranspiration (solid line).	36
Fig. 7. Relative importance of the predictor variables of both models (RF and XGB) for forecasting 1, 12, 24, 36 and 48 hours ahead the water table depth in the observation wells (P1, P2 and P3).	43
Fig. 8. Observed and simulated water table depth Of XGB model for 24-hour forecast (A), 36- hour forecast (B) and 48-hour forecast (B) during the testing period for the observation well P1.	46
Fig. 9. The prediction error of XGB model for 24-hour forecast (A), 36-hour forecast (B) and 48-hour forecast (C) during the testing period for the observation well P1.....	47

Liste des sigles

AA	Apprentissage automatique
ACC	Autocorrelation coefficient
ANN	Artificial neurone network (réseau de neurone artificiel)
ATP	Adénosine triphosphate
CO ₂	Dioxyde de carbone
ET	Évapotranspiration
ET _{day}	Évapotranspiration journalière
ET _{hour}	Évapotranspiration horaire
K _{sat}	Conductivité hydraulique à saturation
NSE	Nash-Sutcliffe efficiency (Coefficient d'efficacité de Nash-Sutcliffe)
OOB	Out-of-bag
P	Précipitation
P1, P2 et P3	Puits d'observation de la profondeur de la nappe phréatique au sud, au centre et au nord du champ
PNP	Profondeur de la nappe phréatique
PVC	Polymère du chlorure de vinyle
R ²	Coefficient de détermination
R _a	Radiation extraterrestre
R _d	Radiation extraterrestre journalière
RF	Random Forest (Forêt aléatoire)
R _h	Radiation extraterrestre horaire
RMSE	Root mean square error (Erreur quadratique moyenne)
RT	Regression tree (arbre de régression)

\bar{T}	Température moyenne
T_r	Training period (Étape d'apprentissage des modèles d'apprentissage automatique)
T_s	Test period (Étape de vérification des modèles d'apprentissage automatique)
WTD	Water table depth
XGB	Extreme Gradient Boosting (Arbre décisionnel renforcé)

Remerciement

Mes premiers remerciements vont aux professeurs Jacques Gallichand et Silvio José Gumiere. Deux personnes qui ont grandement contribué à ce travail par leurs conseils, leurs accompagnements, leurs corrections et commentaires. Sachez que vos regards constructifs ont permis l'amélioration de mes écrits.

J'adresse un merci particulier à Jonathan Lafond et Guillaume Létourneau pour les données et les matériels qu'ils m'ont fournis lors de mon analyse. Ma reconnaissance va aussi à Cintia Racine pour ses conseils et ses réponses à mes interrogations tout au long de mes études.

À mes compagnons d'études, Lionel Petit-Homme, Mandela M. Jacques, Modeline Jean, Claude-Allah Joseph, Souleymane Dabo, un merci spécial pour vos conseils, vos corrections, nos rires fous et nos moments de détente. Votre passion pour les études m'a permis d'étudier et de rédiger ce travail avec beaucoup de plaisir. Grâce à vous, j'ai découvert la force de l'esprit d'équipe au sein d'une petite communauté.

À mon amie de cœur Alexandra Mervilus, merci pour ta présence et ton encouragement. Tu t'es toujours montrée attentive dans l'avancement de mon mémoire et toujours prête à m'écouter te parler de canneberges et d'irrigation sans jamais te plaindre.

Enfin, merci à ma mère et mes sœurs pour leur soutien tout au long de mes études. Vous m'encouragez continuellement à me dépasser et à aller au bout de ce que j'entreprends.

Voilà, je suis arrivé au bout de ces deux années de maîtrise grâce à l'apport de vous tous. Cela démontre bien l'importance d'être bien entouré dans la vie et de compter sur le soutien et l'amour de plus d'un.

Avant-propos

L'ensemble de ce mémoire est divisé en deux parties principalement : la revue de littérature et un article (Fig.1). L'introduction et la conclusion générales sont des chapitres complémentaires rédigés en français.

La **partie I**, rédigée en français, est une revue bibliographique des principaux concepts abordés dans le mémoire à savoir la situation mondiale de l'eau souterraine, les besoins en eau de la canneberge, les méthodes de gestion d'eau dans les champs de canneberges et une généralité de l'utilisation de l'apprentissage automatique dans l'hydrologie de l'eau souterraine.

La **partie II**, rédigée en anglais, présente sous forme d'article le développement et la comparaison de deux algorithmes d'apprentissage automatique pour prédire la profondeur de la nappe phréatique. L'article s'intitule « Water table depth forecasting in cranberry fields using two tree-modeling approaches ». La question principale est d'analyser si ces algorithmes de prédiction peuvent être utilisés pour la planification et la gestion de l'irrigation souterraine. Cet article sera soumis sous peu à la revue « Agricultural Water Management ». Il présente les résultats de deux modèles de prédiction sur un site de canneberges situé au Québec, Canada. En collaboration avec mon directeur de recherche Silvio José Gumiere et mon co-directeur Jacques Gallichand, j'ai traité et analysé les données puis j'ai rédigé l'article avec les commentaires de Paul Célécourt qui est un des co-auteurs. Les données ont été collectées par l'équipe des professionnels de recherche sous la direction de Jonathan Lafond et de Guillaume Létourneau.

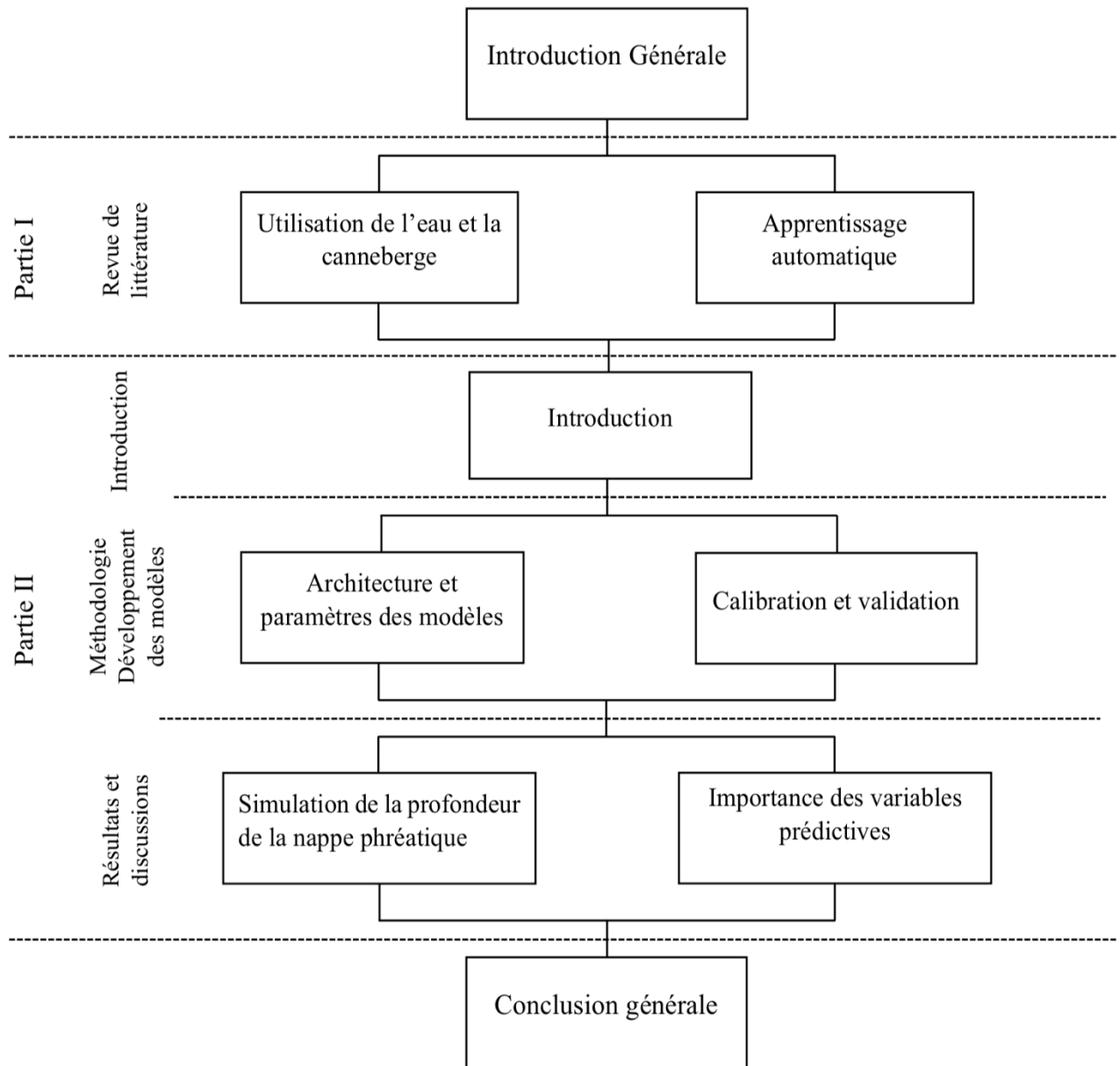


Fig. 1. Schéma du mémoire.

Introduction générale

Au cours des dernières années, au Québec (Canada), la productivité de la canneberge s'est grandement accrue avec un taux de croissance moyen de 17.45 % pour la période allant de 2005 à 2018 (Association des Producteurs de Canneberges du Québec, 2019). Cette croissance de rendement observée est en partie due aux travaux de mise en place des systèmes de drainage efficaces et à l'optimisation de l'irrigation (Bonin, 2009; Pelletier et al., 2016a, 2015a, 2015b). Ces travaux ont permis une gestion d'eau appropriée qui a facilité les échanges gazeux adéquats et les bonnes conditions de croissance et de développement pour la canneberge qui est une plante sensible à la fois aux conditions de stress hydrique et de stress hypoxique (Pelletier et al., 2015c).

Nombreuses stratégies ont été développées pour le contrôle de l'humidité du sol, car les conditions de sols plus sèches peuvent expliquer l'augmentation de rendements dans les récentes années dans la province de Québec (Canada) (Pelletier et al., 2016b). D'abord, des systèmes de drainage sont mis en place pour évacuer l'excès d'eau dans le profil du sol après une pluie ou une irrigation. En outre, la méthode de mesure de tension de l'eau du sol à l'aide des tensiomètres a été développée pour déclencher l'irrigation et aussi la stratégie de contrôle de la profondeur de la nappe phréatique dans le sol. Ces stratégies qui combinent les instruments de mesure et la technologie de communication sans fil ont permis de maintenir un système de précision de gestion d'eau en temps de réel. En dépit des travaux de gestion en eau, le problème de sol trop humide persiste encore dans les champs de canneberges à travers l'Amérique du Nord (Stafford, 2019). Ce problème de sol saturé en eau est une conséquence directe de la relation complexe et dynamique entre le niveau de la nappe d'eau dans le sol, la précipitation et l'évapotranspiration. Comprendre le mouvement des eaux dans un aquifère induit par le processus de recharge et de l'évapotranspiration est très important pour gérer un système d'eau souterraine. Une méthode de gestion fiable nécessite une prédiction du niveau d'eau souterraine (Coulibaly et al., 2001). Il existe des méthodes analytiques et des modèles empiriques d'écoulement d'eau (Chen et al., 2002; Rai et al., 2006; Carretero and Kruse, 2012) pour estimer la recharge, le stockage d'eau et l'extraction de l'eau sans le sol (drainage et évapotranspiration) mais dans la présente étude, l'apprentissage automatique est utilisé pour modéliser la relation entre le niveau d'eau souterraine et les variables météorologiques.

Le développement récent de l'apprentissage automatique dans l'hydrologie des eaux souterraines a permis d'analyser, de comprendre et de prédire le mouvement des eaux souterraines en fonction des variables météorologiques ou de tous facteurs externes qui conditionnent le niveau d'eau souterraine. Dans la présente étude, on propose un modèle de prédiction de la profondeur de la nappe phréatique dans les champs de canneberges.

1.1. Objectif général de l'étude

Développer et comparer deux modèles d'arbres décisionnels pour prédire la profondeur de la nappe phréatique dans les champs de canneberges en fonction des valeurs antécédentes de PNP et de deux variables météorologiques, la précipitation et l'évapotranspiration.

1.2. Objectifs spécifiques

Déterminer le meilleur modèle pour prédire la profondeur de la nappe phréatique dans un champ de canneberge jusqu'à 48 heures en comparant les modèles de forêt aléatoire et des arbres décisionnels renforcés.

Analyser l'importance des variables utilisées pour prédire la profondeur de la nappe phréatique.

1.3. Hypothèse

La prévision des profondeurs de la nappe phréatique en fonction des valeurs antécédentes de PNP et des conditions climatiques permet une gestion optimisée de l'irrigation des champs de canneberges et une identification des facteurs qui conditionnent la fluctuation de la profondeur de la nappe phréatique.

1.4. Portée de l'étude

L'étude est réalisée dans une perspective de gestion intégrée de l'eau à des fins de protection environnementale et de développement durable. Les modèles développés dans cette étude peuvent être utiles pour une gestion plus intelligente de l'eau afin de réduire les impacts associés aux activités agricoles, notamment en production de canneberges.

Bibliographie

- Association des Producteurs de Canneberges du Québec, 2019. Statistiques [WWW Document]. URL <http://www.notrecanneberge.com/Contenu/Page/Statistiques>
- Bonin, S., 2009. Département des sols et de génie agroalimentaire Faculté des Sciences de l'Agriculture et de l'Alimentation Université Laval Québec 114.
- Carretero, S.C., Kruse, E.E., 2012. Relationship between precipitation and water-table fluctuation in a coastal dune aquifer: northeastern coast of the Buenos Aires province, Argentina. *Hydrogeol. J.* 20, 1613–1621. <https://doi.org/10.1007/s10040-012-0890-y>
- Chen, Z., Grasby, S.E., Osadetz, K.G., 2002. Predicting average annual groundwater levels from climatic variables: an empirical model. *J. Hydrol.* 260, 102–117. [https://doi.org/10.1016/S0022-1694\(01\)00606-0](https://doi.org/10.1016/S0022-1694(01)00606-0)
- Coulibaly, P., Anctil, F., Aravena, R., Bobée, B., 2001. Artificial neural network modeling of water table depth fluctuations. *Water Resour. Res.* 37, 885–896. <https://doi.org/10.1029/2000WR900368>
- Pelletier, V., Gallichand, J., Caron, J., Jutras, S., Marchand, S., 2015a. Critical irrigation threshold and cranberry yield components. *Agric. Water Manag.* 148, 106–112. <https://doi.org/10.1016/j.agwat.2014.09.025>
- Pelletier, V., Gallichand, J., Gumiere, S., Caron, J., 2016a. Impact of drainage problems on cranberry yields: Two case studies. *Can. J. Soil Sci.* 97, 1–4.
- Pelletier, V., Gallichand, J., Gumiere, S., Pepin, S., Caron, J., 2015b. Water Table Control for Increasing Yield and Saving Water in Cranberry Production. *Sustainability* 7, 10602–10619. <https://doi.org/10.3390/su70810602>
- Pelletier, V., Gallichand, J., Gumiere, S., Pepin, S., Caron, J., 2015c. Water Table Control for Increasing Yield and Saving Water in Cranberry Production. *Sustainability* 7, 10602–10619. <https://doi.org/10.3390/su70810602>
- Pelletier, V., Pepin, S., Laurent, T., Gallichand, J., Caron, J., 2016b. Cranberry Gas Exchange under Short-term Hypoxic Soil Conditions. *HortScience* 51, 910–914. <https://doi.org/10.21273/HORTSCI.51.7.910>
- Rai, S.N., Manglik, A., Singh, V.S., 2006. Water table fluctuation owing to time-varying recharge, pumping and leakage. *J. Hydrol.* 324, 350–358. <https://doi.org/10.1016/j.jhydrol.2005.09.029>
- Stafford, J.V., 2019. Precision agriculture'19. Wageningen Academic Publishers.
- Wu, J., Zhang, R., Yang, J., 1996. Analysis of rainfall-recharge relationships. *J. Hydrol.* 177, 143–160. [https://doi.org/10.1016/0022-1694\(95\)02935-4](https://doi.org/10.1016/0022-1694(95)02935-4)

Chapitre 1

Revue de littérature

2.1. Situation de l'eau souterraine dans le monde

La préoccupation autour de la disponibilité de l'eau souterraine s'est amplifiée au cours des dernières années. La qualité et la quantité de l'eau souterraine sont de plus en plus menacées par la pollution et la surexploitation (Rodriguez-Galiano et al., 2014). En effet, la croissance économique et l'augmentation de la densité de la population humaine impactent considérablement la disponibilité de l'eau (Vörösmarty et al., 2000) en raison du secteur agricole et industriel qui prélèvent beaucoup d'eau pour répondre au besoin des opérations. À la longue, l'épuisement de l'eau souterraine peut conduire au rationnement local de l'eau, à la diminution des rendements des cultures, au changement des régimes d'écoulement, à l'assèchement des puits et à l'intrusion de l'eau de mer dans les régions côtières (Nayak et al., 2006). De plus, la pollution des eaux souterraines par le nitrate, les hydrocarbures aromatiques polycycliques, le rejet des eaux usées et les effluents agro-industriels augmentent le risque de détérioration de la qualité de l'eau souterraine en plus d'être un danger pour la santé humaine (Aghzar et al., 2002; Jarjoui et al., 2000). Combinés à ces problèmes de pollution et d'exploitation, le niveau des eaux souterraines diminue de plus en plus dû à la tendance décroissante globale de la précipitation engendrée par les changements climatiques (Taylor et al., 2013). Cette situation est de plus en plus inquiétante car elle n'est pas sans conséquence pour le futur des humains et de tous les êtres vivants. Au regard de cette situation alarmante, des mesures de gestion durable doivent être la priorité de tous les acteurs surtout les acteurs agricoles pour une utilisation minimale et plus efficace de l'eau. L'utilisation excessive de l'eau en agriculture est l'un des piliers économiques et environnementaux sur lequel il est primordial d'agir. Chaque année, l'agriculture utilise plus de 70 % des prélèvements d'eau douce dans le monde (Burton, 2010). Pour protéger les ressources d'eau globales, le raffinement de l'irrigation est un élément primordial (Samson et al., 2017). Aussi, une meilleure compréhension du processus hydrologique de chaque culture doit figurer parmi les moyens à privilégier pour parvenir à une application plus efficace de gestion de l'eau.

2.2. La canneberge et ses besoins en eau

La canneberge est une plante vivace et tempérée originaire de l'Amérique du Nord (Eck, 1990) où, au niveau mondial, 97.3% des canneberges sont produites (Food and Agriculture

Organization of the United Nations, 2017). La production de cette culture est fortement tributaire des ressources en eau souterraine et en eau de surface pour les inondations périodiques et les irrigations saisonnières (Jeranyama et al., 2014; Kennedy et al., 2017). Cette culture doit être irriguée notamment au printemps, durant la phase végétative, pour protéger ses organes reproducteurs contre le gel. Durant l'été, l'irrigation est essentiellement utilisée pour répondre aux besoins hydriques des plants, refroidir la culture et lutter contre les insectes et les mauvaises herbes. En automne et en hiver, les parcelles sont inondées respectivement pour la récolte et la protection contre les températures extrêmement basses. Les inondations d'hiver sont nécessaires pour former une couche de glace protectrice au-dessus des plants durant la dormance et pour éviter la levée pendant les redoux hivernaux (Asselin, 2005). Globalement, le volume total d'eau utilisé pour l'irrigation et la gestion de cette culture varie d'une année à une autre et est estimé en moyenne à 2.2 m an^{-1} (Kennedy et al., 2017).

2.3. Gestion de l'eau dans les champs de canneberges

L'eau dans les champs de canneberge est essentiellement utilisée pour maximiser le rendement et placer la plante dans des conditions optimales de développement et de croissance. La canneberge, bien que tolérante aux milieux humides, est très sensible aux conditions hypoxiques dans la rhizosphère (Laurent, 2015). En général, un sol mal drainé nuit à son développement et sa croissance (Baumann et al., 2005). Le manque d'oxygène dans la zone racinaire, causé par un sol saturé en eau, induit différentes réponses physiologiques qui réduisent l'assimilation du CO_2 et diminuent le rendement de cette culture (Pelletier et al., 2016). L'une des premières réactions physiologiques des plantes à un sol saturé en eau est la diminution de la capacité respiratoire des racines (Liao and Lin, 1995). Cette réduction de la capacité respiratoire des racines est accompagnée d'une inhibition des activités métaboliques et d'une réduction de la production d'ATP (Saglio et al., 1980), lesquelles limitent l'énergie pour la croissance racinaire et réduisent ainsi la croissance végétative (Liao and Lin 2001). Outre les dommages aux racines, un sol saturé en eau réduit significativement la capacité d'échange gazeux foliaire, limite le taux de transpiration et réduit la conductance stomatique (Liao and Lin, 2001). Dans le cas de la culture de canneberges, cette réduction de la photosynthèse dans des conditions de sol hypoxiques pourrait être due en partie à la limitation stomatique (Pelletier et al., 2016). De là, les effets

causés par une mauvaise aération plus ou moins prolongée du sol de canneberges restreignent leur développement racinaire, diminuent leur capacité de photosynthèse et leur rendement. En effet, des pertes de rendements de plus de 50 % ont été associées à une nappe phréatique peu profonde (25 cm de profondeur) dans le substrat des canneberges (Pelletier et al., 2015b). Pour éliminer les excès d'eau dans les champs, la solution passe premièrement par un drainage efficace et un contrôle continu de la profondeur de la nappe phréatique. Les problèmes de drainage tels que les défauts de construction et le colmatage des drains sont des facteurs importants limitant le rendement de canneberges (Samson et al., 2016). Dans l'un ou l'autre cas de problème de drainage, les pertes de rendements sont énormes et peuvent être estimées à 39 % dues au colmatage ou à 25 % dues au défaut de construction des drains. C'est important de bien drainer le sol afin de créer des conditions favorables pour la croissance et le développement de la canneberge. Cette condition optimale peut être maintenue d'une part par un système de drainage efficace pour évacuer l'excès d'eau après une période de pluie et d'autre part par une irrigation optimale. Pelletier et al. (2015a) proposent de déclencher l'irrigation quand la tension en eau du sol est à -7.5 kPa et de l'arrêter quand elle atteint - 3.5 kPa. Ils rapportent que l'utilisation de la tension en eau du sol pour déclencher l'irrigation aide les producteurs à devenir de meilleurs gestionnaires de l'eau.

Une autre stratégie de gestion d'eau dans les champs de canneberges consiste à maintenir la nappe phréatique à une profondeur de 60 cm au-dessous de la surface du sol pour favoriser l'alimentation hydrique par remontée capillaire (Pelletier et al., 2015b). Cette stratégie permet de maximiser le rendement de canneberges, d'économiser de l'eau et d'éviter les conditions hypoxiques (Pelletier et al., 2015b). Ces stratégies ont permis une amélioration dans la gestion de l'eau et ont contribué à maintenir des conditions de sol adéquates qui améliorent les conditions de développement des canneberges et de leur rendement.

2.4. Relation entre la profondeur de la nappe phréatique, la précipitation et l'évapotranspiration

Le niveau de la nappe phréatique dans les champs de canneberges varie spatialement et temporellement en réponse directe à l'évapotranspiration et à la précipitation (Vanderleest et al., 2016). D'une part, la réponse de la nappe phréatique est différente suivant les saisons pluvieuse et sèche qui conditionnent les processus d'évapotranspiration et de précipitation

(Carretero and Kruse, 2012). D'autre part, les propriétés hydrauliques du sol des canneberges varient spatialement (Gumiere et al., 2014), ce qui affecte le flux d'eau dans le sol qui alimente les eaux souterraines. Le processus de variation de la profondeur de la nappe phréatique dans un aquifère non confiné est un processus complexe, car il implique un écoulement d'eau à travers la région non saturée de l'aquifère (Viswanathan, 1983). Il est essentiel de comprendre le comportement dynamique de la nappe phréatique en fonction de la recharge et de l'extraction de l'eau dans le sol (évapotranspiration ou drainage) pour la mise en place d'une irrigation de précision. Dans le champ de canneberges, en raison du substrat sableux et d'une nappe phréatique peu profonde, l'aquifère est sensible à l'infiltration de la pluie et de l'irrigation. En effet, la nappe phréatique est peu profonde, car elle est généralement maintenue à 60 cm de profondeur en moyenne sous la surface du sol comme stratégie d'irrigation pour maximiser le rendement de canneberges (Pelletier et al., 2015b). Ainsi, pour un profil de sol avec une nappe phréatique peu profonde, la réponse de la recharge à la pluie est rapide et la correspondance entre elles est très élevée (Wu et al., 1996). Un événement de pluie peut occasionner une remontée rapide de la nappe phréatique près de la surface du sol, ce qui peut entraîner plusieurs problèmes tels que l'engorgement des sols, la salinité des sols (Rai et al., 2006) et des pertes de rendements. Néanmoins, en accord avec la dynamique du système des eaux souterraines, il y a un volume d'eau qui atteint la nappe phréatique, mais qui n'est pas suffisant pour déclencher une montée de la nappe phréatique parce qu'il est relié au flux sortant du système (Carretero and Kruse, 2012). En général, les changements dans le niveau de la nappe phréatique sont provoqués principalement par la précipitation qui est supérieure au flux sortant du système (Carretero and Kruse 2012). Outre la précipitation, l'évapotranspiration (ET) est un élément crucial pour la planification et la gestion de l'eau souterraine (Ndou et al., 2018). L'évapotranspiration représente le mouvement de l'eau du sol à la plante vers l'atmosphère. Ce paramètre climatique est contrôlé dynamiquement par la demande atmosphérique et la profondeur de la nappe phréatique (Vanderleest et al., 2016). Ses valeurs augmentent au fur et à mesure que la nappe phréatique se rapproche de la surface du sol (Ndou et al., 2018) et aussi quand la température de l'air augmente. En effet, l'évapotranspiration est affectée par la température de l'air qui joue un rôle significatif sur le bilan hydrique (Daliakopoulos et al., 2005). L'influence de l'évapotranspiration est plus élevée dans les mois avec les températures plus

élevées, ses effets se traduisent en une tendance à la baisse et en des oscillations quotidiennes du niveau de l'eau souterraine (Taormina et al., 2012). Dans le champ de canneberges, quand on compare les profondeurs de la nappe phréatique au taux d'évapotranspiration, on constate qu'une nappe phréatique se situant à une profondeur entre 50 cm et 60 cm peut compenser environ 30 % du volume d'eau perdu par évapotranspiration (Vanderleest et al., 2016). Quand elle est moins profonde que 50 cm, elle peut compenser presque la totalité d'évapotranspiration. Les valeurs d'ET varient autour de 0.8 à 6.2 mm j⁻¹ pour les canneberges (Vanderleest and Bland, 2016).

Cette relation entre la profondeur de la nappe phréatique, l'évapotranspiration et la précipitation dans le champ de canneberges peut être modélisée par l'apprentissage automatique, lequel est un outil très utilisé pour faire des prédictions et prendre des décisions. Dans les sections suivantes, on présente l'apprentissage automatique et ses utilisations dans l'hydrologie de l'eau souterraine.

2.5. Apprentissage automatique

L'apprentissage automatique (AA) est une discipline se situant à l'intersection des sciences de l'informatique et de la statistique, et au cœur de l'intelligence artificielle et de la science de données (Jordan and Mitchell, 2015). Dans de nombreux domaines scientifiques, il permet de démêler, de quantifier et de comprendre les processus à forte intensité de données (Liakos et al., 2018). Donc, c'est la taille de données qui rend indispensable le développement des procédures d'apprentissages évolutifs associant informatiques et statistiques (Jordan and Mitchell, 2015).

Dans sa méthodologie, l'AA implique un processus d'apprentissage avec l'objectif d'apprendre de l'expérience (données de formation) pour effectuer une tâche (Liakos et al., 2018) (*Fig. 2*). Le problème d'apprentissage classique est de construire un modèle capable de prédire correctement les classes de nouveaux objets par le biais d'une expérience de formation à partir d'anciens objets (Mitchell, 1997). Le développement de l'apprentissage automatique est essentiellement tributaire à l'obtention des informations, de la prise des décisions et de sa capacité à faire des prédictions à partir d'un ensemble de données observées.

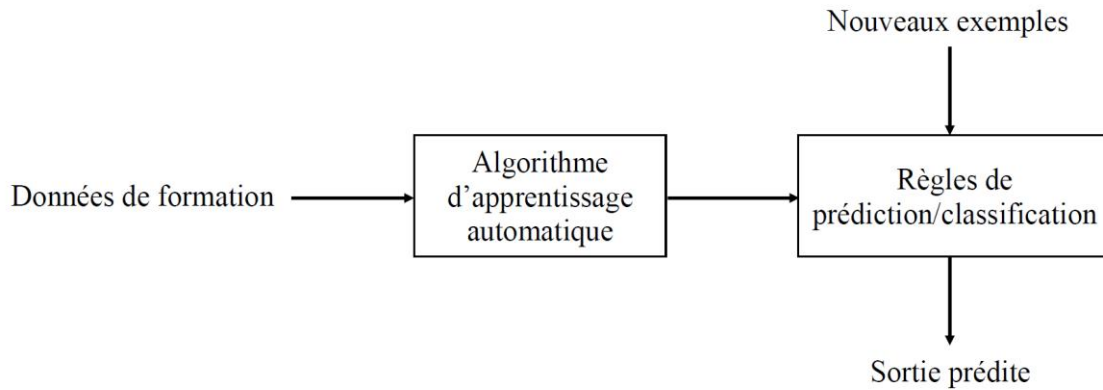


Fig. 2. Approche typique de l'apprentissage automatique tirée de Liakos et al. (2018)

De nos jours, de nombreuses techniques ont été développées pour apprendre automatiquement les relations et les structures dans les données d'observation afin de simplifier le processus d'acquisition de connaissance qui est souvent fastidieux et sujet à l'erreur (McQueen et al., 1995). L'ensemble de ces techniques d'apprentissage automatique sont regroupées en deux grandes approches principales en fonction des données d'apprentissage : l'apprentissage supervisé et l'apprentissage non supervisé. Le but de l'apprentissage supervisé est la construction des modèles concis de la distribution des étiquettes de classes en termes de variables prédictives (Kotsiantis et al., 2007). Cette approche supervisée contient plusieurs algorithmes très utilisés dans le monde des scientifiques et ingénieurs pour résoudre les problèmes de classification ou de régression. Les réseaux de neurones artificiels (Wu et al., 2014), les modèles d'arbres de décision (Breiman, 2001; Chen et al., 2015), les k-voisins plus proches et la régression logistique sont les principaux algorithmes d'apprentissage supervisé. Ces algorithmes apprennent à partir des données d'apprentissage étiquetées pour produire des règles qui aident à prédire des données non disponibles ou futures. Dans le cas où les données sont non étiquetées, l'apprentissage est appelé non-supervisé. Ils sont souvent utilisés à des fins exploratoires lorsque les structures dans les données sont recherchées sans connaître à l'avance ce qu'elles seront (McQueen et al., 1995).

2.6. Modélisation du niveau d'eau souterraine

La modélisation du niveau d'eau souterraine est généralement effectuée par des réseaux de neurones artificiels (Yoon et al. 2011; Adamowski and Chan 2011) et des modèles d'arbres

décisionnels (Li et al., 2016; Singh et al., 2014) qui sont des modèles avec des propriétés particulièrement adaptées pour modéliser les systèmes non linéaires dynamiques. À titre d'exemple, Taormina et al. (2012) ont entraîné des réseaux de neurones artificiels pour simuler les niveaux d'eaux souterraines horaires dans un système aquifère côtier de Venise en utilisant la pluie, l'évapotranspiration et les niveaux d'eaux souterraines comme variables prédictives. Aussi, Daliakopoulos et al. (2005) ont testé de nombreux types de réseaux de neurones artificiels pour prédire le niveau de l'eau souterraine en prenant comme variables prédictives la précipitation, la température, le niveau d'eau souterraine antécédent et le niveau de la rivière. Les modèles d'arbres de décision comme la « forêt aléatoire » et les « arbres décisionnels renforcés » sont aussi très utilisés pour prédire la profondeur de la nappe phréatique (Naghbi et al., 2016; Wang et al., 2018). Ces modèles sont capables de s'adapter aux changements de manière récurrente et de détecter les structures dans un système naturel complexe (Daliakopoulos et al., 2005). Pour prédire le niveau d'eau souterraine, ces modèles requièrent des données de séries temporelles du niveau d'eau souterraine et des variables d'entrée pertinentes (Yoon et al., 2011). Ces variables d'entrées sont pour la plupart du temps des données hydrométéorologiques : la précipitation, la température, l'évapotranspiration et les niveaux historiques d'eau souterraine.

Ces modèles de séries temporelles empiriques ont la capacité de dériver un ensemble de formulation explicite régissant les phénomènes et de décrire les relations entre les variables dépendantes et indépendantes à travers nombreux opérateurs (Kisi et al., 2012). De plus, ils n'exigent pas que la nature complexe des processus considérés soit décrite explicitement sous une forme mathématique (Nayak et al., 2006). En hydrologie, ils ajustent les relations complexes non linéaires entre les facteurs de conditionnant de l'eau souterraine et la potentialité de l'eau souterraine et analysent automatiquement les effets d'interaction entre les facteurs efficaces d'eau souterraine (Naghbi et al., 2016). Toutefois, ils présentent deux inconvénients. D'une part, ils sont limités à prédire la variation temporelle à un lieu fixe. D'autre part, ils ne sont pas adaptés pour la prévision lorsque le comportement dynamique du système hydrologique change avec le temps (Nayak et al., 2006).

Au-delà de ces modèles de série chronologique, il existe des modèles numériques physiques qui permettent aussi de caractériser et de prédire le niveau d'eau souterraine. Ces modèles exigent une définition des paramètres de l'aquifère afin de décrire la variabilité spatiale de

l'espace souterrain (Taormina et al., 2012). En général, la physique de l'écoulement du sous-sol est simplifiée par une équation établie par ces modèles et résolue avec des conditions initiales et limites appropriées à l'aide des méthodes numériques (Yoon et al., 2011).

Bibliographie

- Adamowski, J., Chan, H.F., 2011. A wavelet neural network conjunction model for groundwater level forecasting. *J. Hydrol.* 407, 28–40.
<https://doi.org/10.1016/j.jhydrol.2011.06.013>
- Aghzar, N., Berdai, H., Bellouti, A., Souidi, B., 2002. Pollution nitrique des eaux souterraines au Tadla (Maroc). *Rev. Sci. Eau* 15, 459. <https://doi.org/10.7202/705465ar>
- Asselin, R., 2005. Petit fruit deviendra grand... comme la canneberge. « Journées Agri-vision », Montérégie, Ministère de l'Agriculture, des Pêcheries et de l'Alimentation du Québec (MAPAQ).
- Baumann, D.L., Workmaster, B.A., Kosola, K.R., 2005. 'Ben Lear' and 'Stevens' Cranberry Root and Shoot Growth Response to Soil Water Potential 40, 4.
- Breiman, L., 2001. Random Forests. *Mach. Learn.* 45, 5–32.
<https://doi.org/10.1023/A:1010933404324>
- Burton, M., 2010. Irrigation management: principles and practices. CABI North American Office, Cambridge, MA.
- Carretero, S.C., Kruse, E.E., 2012. Relationship between precipitation and water-table fluctuation in a coastal dune aquifer: northeastern coast of the Buenos Aires province, Argentina. *Hydrogeol. J.* 20, 1613–1621. <https://doi.org/10.1007/s10040-012-0890-y>
- C.-C. Yang, S. O. Prasher, R. Lacroix, 1996. Applications of Artificial Neural Networks to Land Drainage Engineering. *Trans. ASAE* 39, 525–533.
<https://doi.org/10.13031/2013.27531>
- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., 2015. Xgboost: extreme gradient boosting. R Package Version 04-2 1–4.
- Daliakopoulos, I.N., Coulibaly, P., Tsanis, I.K., 2005. Groundwater level forecasting using artificial neural networks. *J. Hydrol.* 309, 229–240.
<https://doi.org/10.1016/j.jhydrol.2004.12.001>
- Eck, P., 1990. The American cranberry Rutgers Univ. Press N. B. NJ.
- Food and Agriculture Organization of the United Nations, 2017. Production share of cranberries by region [WWW Document].
<http://www.fao.org/faostat/en/#data/QC/visualize>
- Gumiere, S.J., Lafond, J.A., Hallema, D.W., Périard, Y., Caron, J., Gallichand, J., 2014. Mapping soil hydraulic conductivity and matric potential for water management of cranberry: Characterisation and spatial interpolation methods. *Biosyst. Eng.* 128, 29–40. <https://doi.org/10.1016/j.biosystemseng.2014.09.002>
- Jarjoui, M., Geahchan, A., Boutros, E., Abou-Kaïs, A., 2000. Pollution des eaux souterraines par les hydrocarbures aromatiques polycycliques et évaluation du risque. *Houille Blanche* 89–92. <https://doi.org/10.1051/lhb/2000080>
- Jeranyama, P., Demoranville, C., Waddell, J., 2014. Using canopy temperature, soil tension and moisture measurements as tools in cranberry irrigation. *Acta Hort.* 1017, 487–492. <https://doi.org/10.17660/ActaHortic.2014.1017.60>

- Jordan, M.I., Mitchell, T.M., 2015. Machine learning: Trends, perspectives, and prospects. *Science* 349, 255. <https://doi.org/10.1126/science.aaa8415>
- Kennedy, C., Jeranyama, P., Alverson, N., 2017. Agricultural water requirements for commercial production of cranberries. *Can. J. SOIL Sci.* 97, 38–45. <https://doi.org/10.1139/cjss-2015-0095>
- Kisi, O., Shiri, J., Nikoofar, B., 2012. Forecasting daily lake levels using artificial intelligence approaches. *Comput. Geosci.* 41, 169–180. <https://doi.org/10.1016/j.cageo.2011.08.027>
- Kotsiantis, S.B., Zaharakis, I., Pintelas, P., 2007. Supervised machine learning: A review of classification techniques. *Emerg. Artif. Intell. Appl. Comput. Eng.* 160, 3–24.
- Laurent, T., 2015. Réponse de la canneberge (*Vaccinium macrocarpon* Ait.) à l'aération du sol 101.
- Létourneau, G., 2017. Approche multicritère d'optimisation de l'irrigation goutte-à-goutte du fraisier 169.
- Li, B., Yang, G., Wan, R., Dai, X., Zhang, Y., 2016. Comparison of random forests and other statistical methods for the prediction of lake water level: a case study of the Poyang Lake in China. *Hydrol. Res.* 47, 69–83. <https://doi.org/10.2166/nh.2016.264>
- Liakos, K., Busato, P., Moshou, D., Pearson, S., Bochtis, D., 2018. Machine learning in agriculture: A review. *Sensors* 18, 2674.
- Liao, C.-T., Lin, C.-H., 2001. Physiological Adaptation of Crop Plants to Flooding Stress 10.
- Liao, C.T., Lin, C.H., 1995. Effect of flood stress on morphology and anaerobic metabolism of *Momordica charantia*. *Environ. Exp. Bot.* 35, 105–113. [https://doi.org/10.1016/0098-8472\(94\)00048-A](https://doi.org/10.1016/0098-8472(94)00048-A)
- McQueen, R.J., Garner, S.R., Nevill-Manning, C.G., Witten, I.H., 1995. Applying machine learning to agricultural data. *Comput. Electron. Agric.* 12, 275–293. [https://doi.org/10.1016/0168-1699\(95\)98601-9](https://doi.org/10.1016/0168-1699(95)98601-9)
- Mitchell, T.M., 1997. Machine learning.
- Naghibi, S.A., Pourghasemi, H.R., Dixon, B., 2016. GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. *Environ. Monit. Assess.* 188, 44. <https://doi.org/10.1007/s10661-015-5049-6>
- Nayak, P.C., Rao, Y.R.S., Sudheer, K.P., 2006. Groundwater Level Forecasting in a Shallow Aquifer Using Artificial Neural Network Approach. *Water Resour. Manag.* 20, 77–90. <https://doi.org/10.1007/s11269-006-4007-z>
- Ndou, N.N., Palamuleni, L.G., Ramoelo, A., 2018. Modelling depth to groundwater level using SEBAL-based dry season potential evapotranspiration in the upper Molopo River Catchment, South Africa. *Egypt. J. Remote Sens. Space Sci.* 21, 237–248. <https://doi.org/10.1016/j.ejrs.2017.08.003>

- Pelletier, V., Gallichand, J., Caron, J., Jutras, S., Marchand, S., 2015a. Critical irrigation threshold and cranberry yield components. *Agric. Water Manag.* 148, 106–112. <https://doi.org/10.1016/j.agwat.2014.09.025>
- Pelletier, V., Gallichand, J., Gumiere, S., Pepin, S., Caron, J., 2015b. Water Table Control for Increasing Yield and Saving Water in Cranberry Production. *Sustainability* 7, 10602–10619. <https://doi.org/10.3390/su70810602>
- Pelletier, V., Pepin, S., Laurent, T., Gallichand, J., Caron, J., 2016. Cranberry Gas Exchange under Short-term Hypoxic Soil Conditions. *HortScience* 51, 910–914. <https://doi.org/10.21273/HORTSCI.51.7.910>
- Rai, S.N., Manglik, A., Singh, V.S., 2006. Water table fluctuation owing to time-varying recharge, pumping and leakage. *J. Hydrol.* 324, 350–358. <https://doi.org/10.1016/j.jhydrol.2005.09.029>
- Rodriguez-Galiano, V., Mendes, M.P., Garcia-Soldado, M.J., Chica-Olmo, M., Ribeiro, L., 2014. Predictive modeling of groundwater nitrate pollution using Random Forest and multisource variables related to intrinsic and specific vulnerability: A case study in an agricultural setting (Southern Spain). *Sci. Total Environ.* 476–477, 189–206. <https://doi.org/10.1016/j.scitotenv.2014.01.001>
- Saglio, P.H., Raymond, P., Pradet, A., 1980. Metabolic Activity and Energy Charge of Excised Maize Root Tips under Anoxia: CONTROL BY SOLUBLE SUGARS. *Plant Physiol.* 66, 1053–1057. <https://doi.org/10.1104/pp.66.6.1053>
- Samson, M., Fortin, J., Pepin, S., Caron, J., 2016. Impact of potassium sulfate salinity on growth and development of cranberry plants subjected to overhead and subirrigation. *Can. J. Soil Sci.* 97, 20–30.
- Singh, K.P., Gupta, S., Mohan, D., 2014. Evaluating influences of seasonal variations and anthropogenic activities on alluvial groundwater hydrochemistry using ensemble learning approaches. *J. Hydrol.* 511, 254–266. <https://doi.org/10.1016/j.jhydrol.2014.01.004>
- Taormina, R., Chau, K., Sethi, R., 2012. Artificial neural network simulation of hourly groundwater levels in a coastal aquifer system of the Venice lagoon. *Eng. Appl. Artif. Intell.* 25, 1670–1676. <https://doi.org/10.1016/j.engappai.2012.02.009>
- Taylor, R.G., Scanlon, B., Döll, P., Rodell, M., van Beek, R., Wada, Y., Longuevergne, L., Leblanc, M., Famiglietti, J.S., Edmunds, M., Konikow, L., Green, T.R., Chen, J., Taniguchi, M., Bierkens, M.F.P., MacDonald, A., Fan, Y., Maxwell, R.M., Yechieli, Y., Gurdak, J.J., Allen, D.M., Shamsudduha, M., Hiscock, K., Yeh, P.J.-F., Holman, I., Treidel, H., 2013. Ground water and climate change. *Nat. Clim. Change* 3, 322–329. <https://doi.org/10.1038/nclimate1744>
- Vanderleest, C.P., Bland, W.L., 2016. Evapotranspiration from cranberry compared with the equilibrium rate. *Can. J. Soil Sci.* 97, 5–10.
- Vanderleest, C.P.L., Caron, J., Bland, W.L., 2016. Water table level management as an irrigation strategy for cranberry (*Vaccinium macrocarpon* Aiton) 1. *Can. J. Soil Sci.* 1–9. <https://doi.org/10.1139/cjss-2016-0001>

- Viswanathan, M.N., 1983. The Rainfall/Water-Table Level Relationship of an Unconfined Aquifer. *Ground Water* 21, 49. <https://doi.org/10.1111/j.1745-6584.1983.tb00704.x>
- Vörösmarty, C.J., Green, P., Salisbury, J., Lammers, R.B., 2000. Global water resources: vulnerability from climate change and population growth. *Science* 289, 284–288.
- Wang, X., Liu, T., Zheng, X., Peng, H., Xin, J., Zhang, B., 2018. Short-term prediction of groundwater level using improved random forest regression with a combination of random features. *Appl. Water Sci.* 8, 125. <https://doi.org/10.1007/s13201-018-0742-6>
- Wu, J., Zhang, R., Yang, J., 1996. Analysis of rainfall-recharge relationships. *J. Hydrol.* 177, 143–160. [https://doi.org/10.1016/0022-1694\(95\)02935-4](https://doi.org/10.1016/0022-1694(95)02935-4)
- Wu, W., Dandy, G.C., Maier, H.R., 2014. Protocol for developing ANN models and its application to the assessment of the quality of the ANN model development process in drinking water quality modelling. *Environ. Model. Softw.* 54, 108–127. <https://doi.org/10.1016/j.envsoft.2013.12.016>
- Yoon, H., Jun, S.-C., Hyun, Y., Bae, G.-O., Lee, K.-K., 2011. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *J. Hydrol.* 396, 128–138. <https://doi.org/10.1016/j.jhydrol.2010.11.002>

Chapitre 2

Water table depth forecasting in cranberry fields using two decision tree-modeling approaches

Jhemson Brédy ^{a, *}, Jacques Gallichand ^a, Paul Celicourt ^a, Silvio José Gumiere ^a

^a Faculté des Sciences de l'agriculture et de l'alimentation, Université Laval, 2425 rue de l'Agriculture,
Québec, G1V 0A6, Canada

* Corresponding author.

E-mail address: bredyhemson@gmail.com (J. Brédy)

2.1. Résumé

La gestion intégrée de l'eau souterraine constitue un défi majeur pour les activités industrielles, agricoles et domestiques. Dans certains systèmes agricoles, une gestion optimisée de la nappe phréatique représente un facteur important pour améliorer le rendement des cultures et l'utilisation de l'eau. La prévision de la profondeur de la nappe phréatique (PNP) devient l'une des stratégies utiles pour planifier et gérer en temps réel l'eau souterraine. Cette étude propose une approche de modélisation basée sur les arbres de décision pour prédire la PNP en fonction des précipitations, des précédentes PNP et de l'évapotranspiration pour la gestion de l'eau souterraine des champs de canneberges. Premièrement, deux modèles : « Random Forest (RF) » et « Extreme Gradient Boosting (XGB) » ont été paramétrisés et comparés afin de prédire la PNP jusqu'à 48 heures. Deuxièmement, l'importance des variables prédictives a été déterminée pour analyser leur influence sur la simulation de PNP. Les mesures de PNP de trois puits d'observation dans un champ de canneberges, pour la période de croissance du 8 juillet au 30 août 2017, ont été utilisées pour entraîner et valider les modèles. Des statistiques tels que l'erreur quadratique moyenne, le coefficient de détermination et le coefficient d'efficacité de Nash-Sutcliffe sont utilisés pour mesurer la performance des modèles. Les résultats montrent que l'algorithme XGB est plus performant que le modèle RF pour prédire la PNP et est sélectionné comme le modèle optimal. Parmi les variables prédictives, les valeurs précédentes de PNP étaient les plus importantes pour la simulation de PNP, suivie par la précipitation. L'erreur de prédiction du modèle optimal pour la plage de PNP était de ± 5 cm pour les simulations de 1, 12, 24, 36 et 48 heures. Le modèle XGB fournit des informations utiles sur la dynamique de PNP et une simulation rigoureuse pour la gestion de l'irrigation des canneberges.

Mots-clés: Forêt aléatoire, arbres de décision renforcés, apprentissage automatique, niveau d'eau souterraine, évapotranspiration, précipitation

2.2. Abstract

Integrated groundwater management is a major challenge for industrial, agricultural and domestic activities. In some agricultural production systems, optimized water table management represents a significant factor to improve crop yields and water use. Therefore, predicting water table depth (WTD) becomes an important means to enable real-time planning and management of groundwater resources. This study proposes a decision-tree-based modelling approach for WTD forecasting as a function of precipitation, previous WTD values and evapotranspiration with applications in groundwater resources management for cranberry farming. Firstly, two models-based decision trees, namely Random Forest (RF) and Extrem Gradient Boosting (XGB), were developed and compared to predict the WTD up to 48-hours ahead for a cranberry farm located in Québec, Canada. Secondly, the importance of the predictor variables was analyzed to determine their influence on WTD simulation results. WTD measurements at three observation wells within a cranberry field, for the growing period from July 8, 2017 to August 30, 2017, were used for training and testing the models. Statistical parameters such as the mean squared error, coefficient of determination and Nash-Sutcliffe efficiency coefficient were used to measure models performance. The results show that the XGB algorithm outperformed the RF model for predictions of WTD and was selected as the optimal model. Among the predictor variables, the antecedent WTD was the most important for water table depth simulation, followed by the precipitation. Base on the most important variables and optimal model, the prediction error for entire WTD range was within ± 5 cm for 1-, 12-, 24-, 26- and 48- hour prediction. The XGB model can provide information useful on the WTD dynamics and a rigorous simulation for irrigation planning and management in cranberry fields.

Key words: Random Forest, Extreme Gradient Boosting, machine learning, groundwater level, evapotranspiration, precipitation

2.3. Introduction

Integrated groundwater management remains a crucial challenge for industrial, agricultural and domestic activities. One fundamental concern is the major impact of their rapid expansions on groundwater quality and quantity (Rodriguez-Galiano et al., 2014). Furthermore, groundwater resources are growingly being threatened by climate change and variability (Ghose et al., 2010), which may lead to an overall decrease in the level of groundwater as a result of decreasing precipitation and increasing atmospheric temperature (Lee et al., 2014; Taylor et al., 2013).

In Québec, groundwater is the source of economically exploitable water due to its quality and availability (Ministère de l'Environnement et de la Lutte contre les Changements Climatiques, 2019). However, its vulnerability is increasing because the agricultural sector is extensively withdrawing water to meet the current demands for food (Godfray et al., 2010). For cranberry production particularly, large volumes of water are extracted and used for crop irrigation and other production practices (Samson et al., 2016).

Water requirement for cranberry production is on average of 2233 mm yr⁻¹ (Kennedy et al., 2017). Such an extensive use of water is due to several factors, including crop production (Vanderleest et al., 2016), soil moisture management, harvest and winter flooding and frost protection (Kennedy et al., 2017). Notwithstanding its essential water requirements, cranberry is highly sensitive to variations in soil water tension during the fruit growing season (Caron et al., 2016). As an example, Pelletier et al. (2015b) reported that a reduction of up to 50% in yield is due to hypoxic stress when the water table is close to the soil surface (25 cm). Furthermore, they reported that cranberry yield decreased by 8% for each WTD step of 10 cm for a range between 60 to 120 cm below the soil surface. It is, therefore, important to establish water management methods to reduce the excessive use of groundwater in cranberries fields.

Many studies have been conducted for a better understanding of hydrologic processes and to reduce water consumption in cranberry fields. These studies sought to develop a comprehensive knowledge of water management strategies necessary to optimize water consumption of cranberry crops. For instance, Pelletier et al. (2015a) and Caron et al. (2016) showed that, to optimize cranberry yield without excess of water, irrigation must be applied when the soil matric potential is between -7.5 and -4 kPa. The control of the water table depth

has also been used as an irrigation strategy to optimize cranberry yield. This strategy has resulted in a reduction of 77% of sprinkler water use, and a maximum yield for a water table depth of 60 cm (Pelletier et al., 2015b). However, despite the improvement in water resources management in cranberry fields, no study has been conducted to predict water table depths as a function of predicted meteorological conditions.

Predicting water table depth at different time intervals in the future using time series is an essential tool in water resources planning (Kisi et al., 2012). Nowadays, numerous machine learning methods have been developed to predict the water table depth. These models have been grouped in physically-based models and data-based time-series models (Nayak et al., 2006). The physical models require an adequate definition of aquifer parameters to describe the soil subsurface spatial variability (Taormina et al., 2012). The implementation of the physically-based modelling approach requires an enormous amount of data, which is difficult to obtain because of cost and time constraints (Nayak et al., 2006). A potential solution is a data-driven empirical model susceptible to provide useful results without costly calibration time (Daliakopoulos et al., 2005). Among the data-driven empirical models, Artificial Neural Networks (ANN) have been used to model the dynamics of a nonlinear system for which the water table depth represents the dependent variable (Coulibaly et al., 2001; Nayak et al., 2006; Yoon et al., 2011). Despite their performance in modelling groundwater hydrology, ANN present the possible inconveniences of being stuck in a local minimum and of over-adjustment during the training step. Besides, their inability to work with missing and correlated data leads to complex and less precise models with long calibration time (Maier et al., 2010). Conversely, decision-tree models are not affected by missing and correlated data and they have been widely used in modeling natural phenomena. Some authors have introduced boosted decision trees as an alternative to ANN to identify particles in subatomic physics (Roe et al., 2005). Li et al. (2016) have compared the Random Forest (RF) model and ANN in predicting a lake water level. The results demonstrate that the RF model has superior predictive capabilities with fewer parameters and training time. The decision tree-based models have been successfully applied in modeling groundwater hydrology (Singh et al., 2014a; Wang et al., 2018).

Based on the predictive power of decision tree models, the objective of this paper is to develop and compare decision-trees-based models to forecast the water table depth in

cranberry fields as a function of meteorological conditions (precipitation and evapotranspiration).

2.4. Methodology

2.4.1. Study area

The study area is located on a cranberry farm, near Québec City, Québec, Canada (46°14' N, 72°02' W). In summer, the study area is characterized by a warm and humid climate. The normal total rainfall in June is 48.0 mm, in July 75.0 mm, in August 86.4 mm and 70.6 mm in September. The average temperature is 17.1 °C in June, 19.6 °C in July, 18.4 in August and 13.3 °C in September (Station Arthabaska) (Environment Canada, 2019). The location of the study area is shown in *Fig. 3*. This study site covers approximately 18 ha (40 m-wide x 450 m-long) of “the Pilgrim” cranberries variety. The top 30 cm of soil of the field was excavated and replaced with sand to create favorable conditions for cranberry growth and production. This cranberry field is equipped with a subsurface drainage system with four drainage pipes (spacing = 10 m approximately) at an average depth of 65 cm below the soil surface. These subsurface drains flow out into a control chamber south of the site, which maintains a water table at a depth of 60 cm for keeping a matric potential between -7.5 and -4 kPa, ensuring that the cranberries receive an adequate water supply. Three PVC pipes were inserted vertically at the mid-spacing between the subsurface drains, serving as observation wells for water table depth measurements. A rain gauge had also been installed to measure precipitation.

2.4.2. Data

2.4.2.1. Water table depth

Water table depths (WTD) were measured from a continuously deployed sensing system consisting in a submersible pressure transducer (TDH80, Transducersdirect Inc., Ellington Court Cincinnati, Ohio, USA) and a data logger (ST-4, Hortau Inc., Lévis, Québec, Canada) equipped with a Global Positioning System (GPS) chip. The system automatically reads and corrects the water pressures measurements for the influence of atmospheric pressure. These instruments were inserted into the three observation wells (P1, P2 and P3) shown in *Fig. 3*. We used the ground surface level as the reference to the wells and pressure sensor depth measurements. Data were collected every 15 minutes and sent wirelessly to Hortau’s Irrolis (Irrolis 3 v.3.5.1, Hortau Inc., Lévis, Québec, Canada) website. The water pressure measurements in the observation wells were converted to water elevation and then to

groundwater depth (in cm) in reference to ground surface level. The ground surface level is the reference due to the fact that cranberry field have constructed with zero slope. Missing depth values of water table were interpolated from existing measurements with the help of “cubic spline”, a time series interpolation technique (Moritz and Bartz-Beielstein, 2017). Hourly water table depth values used in this study were the average of the measurements every 15 minutes.

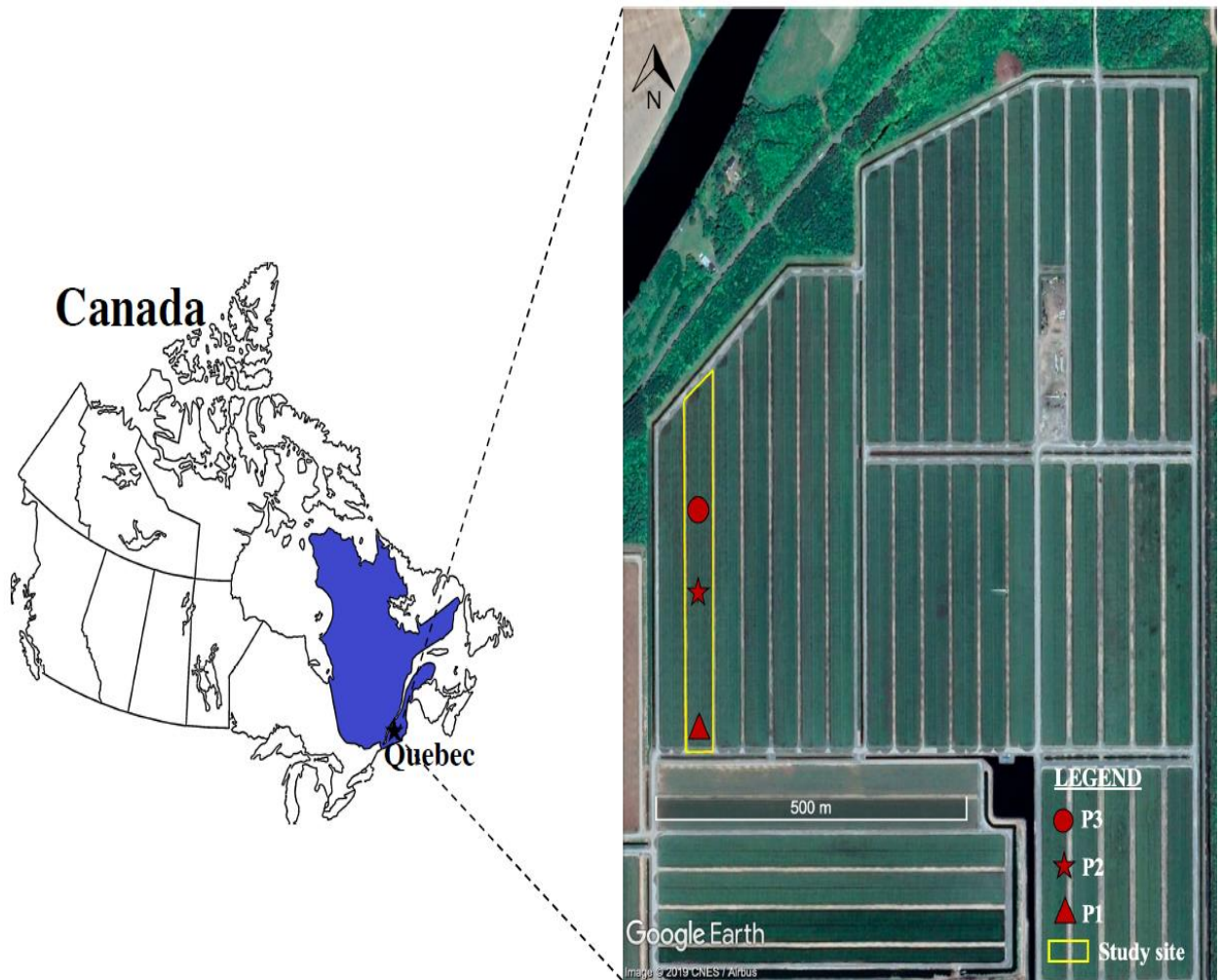


Fig. 3. Study area, Québec, Canada. Red landmarks represent the locations of water table depths observation wells.

2.4.2.2. Precipitation

Precipitations were measured every 15 minutes using a rain gauge (WatchDog 1120 Rain Gauge, Spectrum Technology, Inc., IL US) combined with a ST-4 Hortau data logger (Lévis, Québec). In summer, sprinkler irrigation is used for cooling crops in this study area. This rain

gauge also measured the amount of irrigation water. The total precipitation used in this study was a combination of irrigation water and rainfall. For modeling WTD, hourly precipitation values used were the sum of rain gauge measurements every 15 minutes.

2.4.2.3. Evapotranspiration

Evapotranspiration (ET) was estimated using the formula of Baier and Robertson (1965), which is :

$$ET_{day} = -2.40 + 0.065T_{max} + 0.083(T_{max} - T_{min}) + 0.0044R_a \quad (1)$$

Where R_a is the extraterrestrial radiation ($\text{cal cm}^{-2} \text{d}^{-1}$), ET_{day} is the evapotranspiration daily (mm d^{-1}) and T is the temperature ($^{\circ}\text{C}$).

This formula uses the maximum (T_{max}) and minimum (T_{min}) temperatures for the days that were measured by Hortau's weather station, which includes the WatchDog 2900ET (Spectrum Technology, Inc., IL USA). To obtain hourly ET, the results from Baier and Robertson's formula were multiplied by a ratio of extraterrestrials radiations for hourly (R_h) and daily (R_d) periods:

$$ET_{hour} = ET_{day} * \frac{R_h}{R_d} \quad (2)$$

The extraterrestrial radiation for the daily and hourly period were estimated using the equations of Allen et al. (1990), which for each day of the year and for different latitudes can be estimated from the solar constant, the solar declination and the time of the year by:

$$R_d = \frac{24}{\pi} G_{sc} d_r [\omega_s \sin \sin (\varphi) \sin \sin (\delta) + \cos \cos (\varphi) \cos \cos (\delta) \sin \sin (\omega_s)] \quad (3)$$

$$R_h = \frac{12}{\pi} G_{sc} d_r [(\omega_2 - \omega_1) \sin \sin (\varphi) \sin \sin (\delta) + \cos \cos (\varphi) \cos \cos (\delta) (\sin \sin (\omega_2) - \sin \sin (\omega_1))] \quad (4)$$

Where G_{sc} is the solar constant = $4.92 \text{ MJm}^{-2}\text{d}^{-1}$, d_r is the inverse relative distance Earth-Sun, ω_s sunset hour angle, δ solar declination, φ latitude and ω_1 , ω_2 are the solar time angle at beginning and end period respectively.

2.4.3. Training with different decision tree-based models

Decision-tree-based learning algorithms are one of the most powerful techniques used to build predictive models (Analytics Vidhya, 2016). Those machine learning algorithms can model natural non-linear phenomena and do not need prior statistical assumptions, elimination of outliers or data transformation. In this paper, Random Forest and Extreme Gradient Boosting models were chosen for learning the complex relationship between groundwater conditioning factors and water table depth. Furthermore, predictors importance can be estimated, which is essential in this study, for identifying the significant influencing factors for water table depth variations. We compare those methods for identifying the best one to predict water table depth in cranberry field. Those machine learning methods are described below.

2.4.3.1. Random Forest

Random Forest (RF) is a supervised machine learning method with multiple building blocks named decision trees that outputs an ensemble of predictive models of the same phenomenon (Rodriguez-Galiano et al., 2014). RF is one among many supervised machine learning algorithms suitable for either classification or regression depending on the nature of the targeted variable. The objective of this study being the water table depth modeling, therefore, we focus on regression trees (RT). A regression tree is an algorithm that imposes a set of hierarchically structured restrictions and sequentially applied from a root node up to a terminal node of the decision tree (Breiman et al., 1984). To prevent correlation of the different RTs, RF develop each tree independently from different bootstrap samples extracted from the learning dataset and a subset of the randomly selected variables out of the predictive variables (Breiman, 2001). Specifically, RF builds each tree by making them grow from two thirds ($\frac{2}{3}$) of each bootstrap sample (*inbag*). About one third ($\frac{1}{3}$) is excluded from the bootstrap sample (*out-of-bag*) for a non-biased estimation of the regression error when the trees are added to the forest. A prediction of the out-of-bag (*OOB*) data is generated for each tree. These predictions are subsequently averaged to obtain an estimation of the *OOB* error rate. The Random Forest generalization error depends, therefore, on the weight of the individual trees and the correlation between them. Fig. 4 represents the Random Forest architecture extracted from Guo et al. (2011). The building of a RT follows a recursive binary

partition approach starting at the root node and divides the non-correlated variables into two new branches. This recursive division is performed on the dependent variable as a function of the most significant independent variable that creates the best ensembles of homogeneous populations. Each node is then divided using the most homogeneous among the subsets of predictive variables selected randomly at the node. This division process continues up to when the predefined stop condition is reached. This stop condition is the number of observations at the terminal node (*nodesize*). The result obtained using the RF is, in the case of a regression, an average of the predictions from all decision trees.

In RF modeling, the three following parameters must be defined by the user:

- The number of trees to generate the forest, *ntree*. This parameter is a determinant factor for the predictions performed with RF. The higher the value of *ntree*, the smaller the root mean square error (Scornet, 2017).
- The number of the randomly selected variables at each node, *mtry*, among the set of predictive variables.
- The minimum number of observations at the terminal nodes of the trees, *nodesize*. The default value that always provides good results in regression is *nodesize* = 1 (Breiman, 2002).

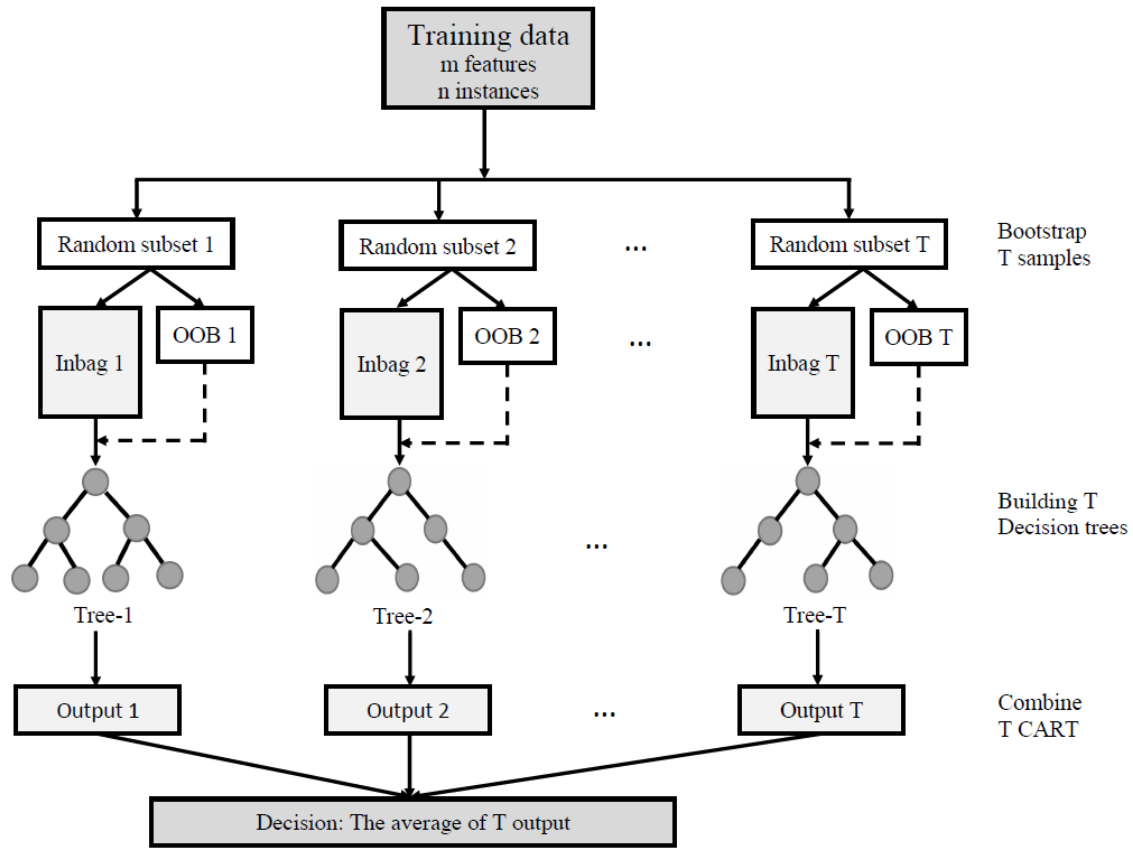


Fig. 4. The flowchart of Random Forest adapted from Guo et al. (2011)

2.4.3.2. Extreme Gradient Boosting

XGB is a supervised machine learning method for regression and classification problems that combines decision trees, which is an algorithm that predicts a targeted value based on several input using the recursive partitioning procedure, and a boosting algorithm. Boosting is a technique implemented to improve the predictive accuracy of the decision trees (Naghbi et al., 2016). XGB is one of several ensemble techniques that aim to improve the prediction accuracy of a model by fitting many models and combining them for prediction to a sequential stagewise procedure (Elith et al., 2008). At each step, XGB adds a new decision tree to improve the predictive accuracy of the current decision tree (Hu et al., 2017). This optimizes a loss function, which is a measure of the difference between observed and fitted values.

Initially, a first decision tree is fitted to the data with an equal weight for each observation. Then, the residuals, with a different and higher weight applied to the unclassified

observations, are introduced to the second tree that attempts to reduce the root mean square error (Singh et al., 2014b). This process is repeated from a tree to another successively until a higher precision is reached. The final prediction of XGB is the sum of the weighted contribution of all decision trees used.

To maximize the predictive accuracy of XGB, the following parameters must be defined by the user (full details, see Chen et al. 2015):

- The *learning rate*, rate at which the model learns pattern in training data. This parameter controls the model overfitting by retarding its convergence. It determines the contribution of each tree in the ensemble of models (Godinho et al., 2018).
- The *tree complexity (max_depth)*, which is the number of splits which must be performed in each tree. It determines the size of the tree.
- The *nrounds* parameter, which controls the maximum number of iterations.
- The subsamples ratio of features (*colsample_bytree*) controls the ratio of the number of predictive variables supplied to a tree (similar to *mtry* in RF).
- The minimal loss reduction (*gamma*) is required to make additional splits on a tree terminal node. This is a tuning parameter to prevent overfitting.
- The minimum sum of weights (*min_child_weight*) necessary at a terminal node. In regression tree, it refers to the minimum number of observations required at a terminal node (similar to *nodesize* in RF).
- The fraction of randomly selected observations (*subsample*) from the training data to build each tree.

2.4.4. Model development

2.4.4.1. Input selection

The Random Forest and Extreme Gradient Boosting models for water table depth forecasting were programmed in R (version 3.4.4) (R Core Team, 2018). The RF algorithm is from the `randomForest` package (Breiman and Cutler, 2018; Liaw and Wiener, 2002) and the XGB model is from the `XGBoost` package (Chen et al., 2015). To model the water table depth, we considered precipitation (P), evapotranspiration (ET) and WTD of the observation wells (P1, P2 and P3) as input variables. The models were built with a combination of input variables for each observation wells separately. Lag times of input variables were generated by means

of the sliding window method, which allowed to restructure these time series dataset (P, ET and WTD) as a supervised learning problem by using a size of the lag equal to 24. This size of lag was chosen to consider the influence of the water table depth over the previous 24-hour time period on the predictions. This means that we used also the values of 24 prior hour of variables (P, ET, WTD) and their values at time t as input. Multi-step forecast of 1, 12, 24, 36 and 48-time steps were used for forecasting water table depth in each observation wells (P1, P2 and P3). In addition, according to the number of time steps ahead to be forecasted ($t + 1, t + 12, t + 24, t + 36$ and $t + 48$), we used the values of next time steps of precipitation and evapotranspiration as input. Thus, the number of input features was between 77 and 171 and one output value to predict for each training pattern.

2.4.4.2. Data division

The experimental dataset consisted of 1280 hourly observations for each variable (P, ET and WTD) ranging from the 8 July 2017 to 30 August 2017, which is the cranberry growth period. This dataset has been split by the holdout method, using the 'CreateDataPartition' caret package function (Kuhn, 2008), in two subsets. A training subset, which contained 70% of data randomly selected for each dataset and the remaining 30% were used as test subset. For each step-ahead prediction, this data division was repeated 30 times using a uniform distribution. This allows models to be trained and validated with a total of 30 different training and test subsets for each training pattern and to reduce the variability of their results. Thus, the predictive performance statistics of models reported was the average for all prediction models with the training and test subsets.

2.4.4.3. Model calibration and validation

The calibration procedure aims to optimize a set of model parameter values that enable the model to map the relationship between the inputs and outputs of a given dataset (Wu et al., 2014). Therefore, two parameters (*n_{tree}* and *m_{try}*) of the RF model have been settled. For regression problems, the default value for *m_{try}* is one-third of the total number of variables (Liaw and Wiener, 2002). The number of trees (*n_{tree}*) was set to 1000. This is based on the study of Rodriguez-Galiano et al. (2014) which showed that *n_{tree}* value of 1000 results in low and stable generalization error. The XGB model parameters have been optimized by a step-by-step procedure. First, we used the default parameters *max_depth* ($md = 6$), *subsample*

($ss = 1$), *min_child_weight* ($mw = 1$), *gamma* ($\gamma = 0$) and *colsample_bytree* was set to 0.8 for introducing a random selection of features to supply to a tree. The learning-rate (*lr*) value was set to 0.009 because low values of *lr* ($lr < 0.1$) improve the model predictive accuracy (Elith et al., 2008). The starting value of the number of iterations (*nrounds*) was set to 3500 because low value of *lr* require a high number of iterations (Freeman et al., 2016). Elith et al. (2008) recommend models with at least 1000 iterations to obtain XGB models with balanced learning rate and number of iterations (*nrounds*). After, we performed a k-fold cross-validation with $k = 5$ at each boosting iteration to identify the optimal number of iterations. Subsequently, the tree complexity was optimized using different values for *max_depth* ($md = 3, 4, 5, 6$). At each training pattern, the *gamma* parameter has been activated or not to prevent overfitting. The *gamma* values were 5 or 10 based on the difference between training error and test error. The model was adjusted and trained using the training data to determine the optimal parameter values that minimize the training and testing error. The test data were used to assess the generalization ability of the model trained. The validation was essential to evaluate the model quality and accuracy with unused data for training.

2.4.4.4. Importance of predictor variables

RF and XGB provide estimates of features importance from the trained model, which allows for the comparison of input variables predictive power. To evaluate the importance of a variable, RF uses two parameters: “mean decrease in accuracy” and “mean decrease in node impurity”. The first parameter is based on how the accuracy decreases when the variable is excluded. The mean decrease in node impurity measures the reduction of heterogeneity defined from the Gini criteria (Breiman, 2001). The Gini importance is a measure of the contribution of each variable in the node splitting. The impurity importance for each variable is computed by the sum of all impurity decrease measures of all nodes in the forest at which a split on these variables has been conducted, normalized by the number of trees (Nembrini et al., 2018).

XGB also outputs the contribution score of the input variables to evaluate their significance in the final prediction. The relative importance of a variable is determined by a measure of its relevance (squared), based on the number of times it was selected in the splitting process, weighted by the squared improvement to the model of each split, and averaged over all trees

(Friedman and Meulman, 2003). In this study, to determine the features importance in RF model employed, the "mean decrease in node impurity" parameter was used. This parameter enables RF model to estimate the features importance in the same way as XGB model.

RF provides the absolute importance of the predictors in the construction of the decision trees within the model. However, the XGB model estimates the relative importance of each predictor variable. The relative influence of each variable of XGB model is scaled so that the sum adds to 100, with higher numbers indicating a stronger influence on the response (Elith et al., 2008). To compare importance scores of the two models, we scaled the predictor variables importance of RF model to have a sum of 100.

2.4.5. Performance criteria

Three statistical evaluation criteria were used to evaluate the models (RF and XGB) predictive power and efficiency:

The root mean square error (RMSE):

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{N}} \quad (5)$$

The Nash-Sutcliffe efficiency (NSE):

$$NSE = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (6)$$

And the coefficient of determination:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n y_i^2 - \frac{\sum_{i=1}^n \hat{y}_i^2}{n}} \quad (7)$$

Where y_i are the observed data, \hat{y}_i are the simulated data, \bar{y} is the mean of observed values and N the number of observations. For a model to yield a good fit between simulated and observed data, the RMSE should be close to zero and NSE and R^2 close to one. In this study, RMSE and NSE statistics are used to measure the model performance for forecasting water table depth whereas R^2 is used to analyze the linear regression goodness of fit between observed and simulated data. A model which systematically over- or under-predicts all the

time will still result in good R^2 values close to one even if all predictions were wrong (Krause et al., 2005). Thus, the intercept and the slope of the regression are provided as an additional information for which R^2 is calculated. For a model goodness of fit, the intercept and gradient should be close to zero and one respectively.

2.5. Results and discussion

2.5.1. Groundwater movement

The observed water table depths in the observation wells (P1, P2 and P3), the precipitation (P) and the evapotranspiration (ET) in the study area during the period from July 8 to August 30, 2017 are shown in *Fig. 5*. The mean groundwater depths are 66.5, 55.3 and 67.3 cm below the soil surface with a coefficient of variation of 10.31, 16.84 and 13.25% for the observation wells (P1, P2 and P3), respectively. The mean precipitation (\pm standard deviation) was 0.17 (\pm 0.70) mm h⁻¹ and the mean evapotranspiration was 0.16 (\pm 0.17) mm h⁻¹. In this site, the difference in water table depth responses to precipitation from well to well was due to the difference in soil hydraulic properties and natural drainage of underlying soil layers.

Fig. 5A shows a close correspondence between individual rainfall and water table depth variation, the water table levels increased after a significant rainfall or irrigation events. Most of the recharge due to precipitation appears to take place within few hours of rainfall or irrigation. We found that the average WTD at time t is positively correlated with average precipitation (*Fig. 6*), with a certain lag time. The maximum correlation coefficient is 0.62 for a 24-hour delay. This can be attributed to the high permeability ($K_{sat} = 19.41 \text{ cm h}^{-1}$) of the sandy soil and the shallow, unconfined aquifer of this study area. On the contrary, during periods of zero precipitation (rainfall and irrigation), the water table decreases due to drainage effect of the site and evapotranspiration. The high evapotranspiration due to high temperature ($\bar{T} = 18.4 (\pm 4.6) \text{ }^\circ\text{C}$) in the summer causes soil moisture deficits. The maximum correlation coefficient between average WTD and mean evapotranspiration is 0.16 for a 27-hour delay (*Fig. 6*). The physical significance of this result is that the lowered groundwater levels (deep percolation) were weakly related to the evapotranspiration effect. The deeper water table recorded for hours without precipitation (*Fig. 5A*) means that the soil profile is under drainage conditions, as found by Bigah et al. (2019). Therefore, in cranberries fields, the observed groundwater level variation remains a dynamic equilibrium between recharge due to significant precipitation and the discharge (drainage and evapotranspiration), resulting in changes in groundwater storage (Chen et al., 2002).

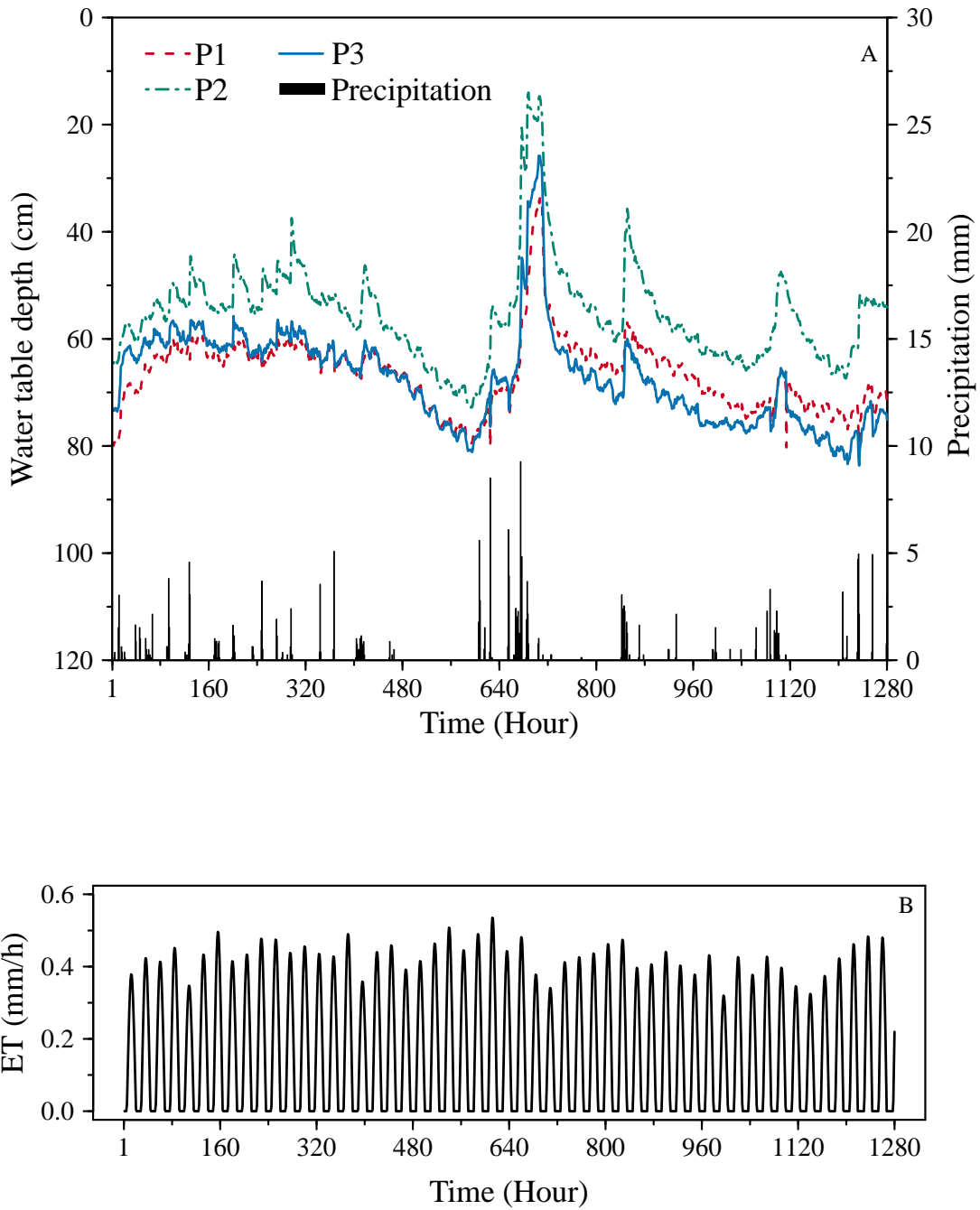


Fig. 5. Hourly measurements of (A) water table depths in the observation wells (P1, P2 and P3), precipitation and (B) evapotranspiration (ET) between July 8 to August 30, 2017.

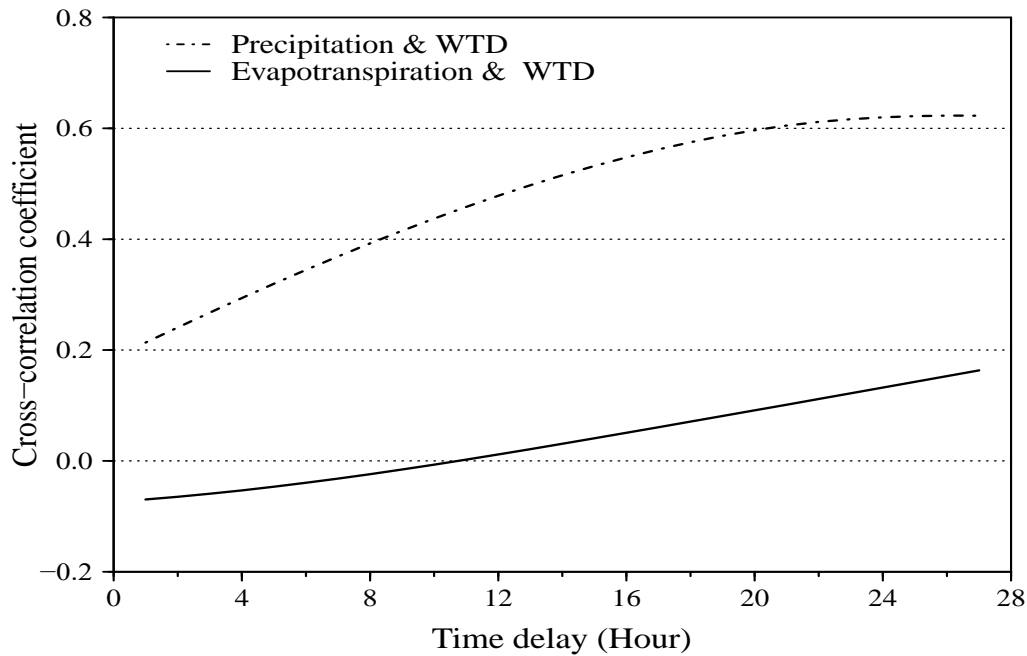


Fig. 6. Correlation coefficient between 24-hour moving average water table depth (WTD) and climate variables as a function of time delay, with 24-hour moving average precipitation (dashed line) and evapotranspiration (solid line).

2.5.2. Models parameterization

During the calibration stage, the RF and XGB parameters were adjusted to minimize the overtraining risk of decision trees. For RF, the values of $mtry$ was $1/3 M$, $n tree$ was 1000 and $nodesize$ was one, where M is the total number input variables. The parameters that showed the best XGB model are given in Table 1. These parameters resulted in the best performance of XGB model during training and test period.

Table 1. Characteristics of the Extreme Gradient Boosting model employed for forecasting water table depth (WTD) 1, 12, 24, 36 and 48 hours ahead.

Parameters	<i>WTD</i> ($t + 1$)	WTD ($t + 12$)	WTD ($t + 24$)	WTD ($t + 36$)	WTD ($t + 48$)
Gamma	10	5	5	10	10
Max depth	3	4	3	3	3
Nrounds	3500	3500	3500	3500	3500
Min weight child	1	1	1	1	1
Subsample	1	1	1	1	1
Colsample by tree	0.8	0.8	0.8	0.8	0.8
Learning rate	0.009	0.009	0.009	0.009	0.009

2.5.3. Comparison of RF and XGB models

The predictive performance statistics of RF and XGB models for water table depth forecasting from 1 to 48 hours ahead are shown in **Table 2** and **Table 3**. The RMSE statistic, which is a measure of residual variances between observed and simulated data, was good with low values during training (< 2 cm) and testing (< 5 cm). Also, the NSE statistic, which is a measure of the hydrological models predictive power, indicated that models accuracy was high ($NSE > 0.7$) in forecasting water table depth during calibration and validation period. According to **Table 2** and **Table 3**, the RF model yielded better RMSE and NSE scores than XGB in 1-hour prediction. However, for 12-, 24-, 26- and 48- hour prediction, the predictive performance of XGB model was superior to that of RF in terms of RMSE and NSE. The lower RMSE value and higher NSE value in both the training and testing stages show that XGB model had better calibration capability than RF model for the data used. The XGB model predictive accuracy is higher than that of RF for two reasons. Firstly, the XGB calibration is more rigorous and has more parameters that improve the generalization error against the simpler RF model. Secondly, the predictive accuracy of decision trees is adaptively optimized by the boosting algorithm (Elith et al., 2008). In the boosting, stage

decision trees are fitted iteratively to the training data, using appropriate methods to gradually increase emphasis on observations modelled poorly by the existing collection of trees. This operation had no equivalence in RF model.

In addition, the performance of both models in 1-hour prediction was better than prediction from 12 to 48 hours. The models predictive accuracy gradually decreased as the prediction time increased. These results are consistent with the findings of Nayak et al. (2006) which showed that the performance of any model was better for shorter times ahead but becomes worse as the prediction time increases. In general, all models displayed good performances in terms of RMSE and NSE during calibration and validation steps, which indicates that both methods were fully trained with the two-month data of cranberry growth to provide satisfactory forecasting. However, XGB was the optimal model chosen to establish the hourly water depth prediction at prediction times of 1 to 48 hours for the three wells.

Table 2. RMSE (cm) values in training (Tr) and test (Ts) stages for both models (RF and XGB) in forecasting water table depth (WTD) at lead-time predictions of 1, 12, 24, 36 and 48-time steps for each observation well (P1, P2 and P3).

		P1		P2		P3	
		Tr	Ts	Tr	Ts	Tr	Ts
WTD(t + 1)	RF	0.345	0.936	0.401	1.131	0.400	1.059
	XGB	0.650	0.992	0.521	0.946	0.638	1.070
WTD(t + 12)	RF	0.662	1.784	1.028	2.686	0.861	2.370
	XGB	0.752	1.728	0.755	2.372	0.785	2.346
WTD(t + 24)	RF	1.026	2.802	1.528	4.182	1.333	3.476
	XGB	0.884	2.655	0.950	3.427	0.954	3.322
WTD(t + 36)	RF	1.310	3.746	1.728	4.942	1.673	4.224
	XGB	1.088	3.727	1.191	4.231	1.187	4.003
WTD(t + 48)	RF	1.385	3.840	1.645	4.541	1.775	4.779
	XGB	1.093	3.618	1.184	4.023	1.183	4.414

Table 3. NSE values in training (Tr) and test (Ts) stages for both models (RF and XGB) models in forecasting water table depth at lead-time predictions of 1, 12, 24, 36- and 48-time steps for each observation well (P1, P2 and P3).

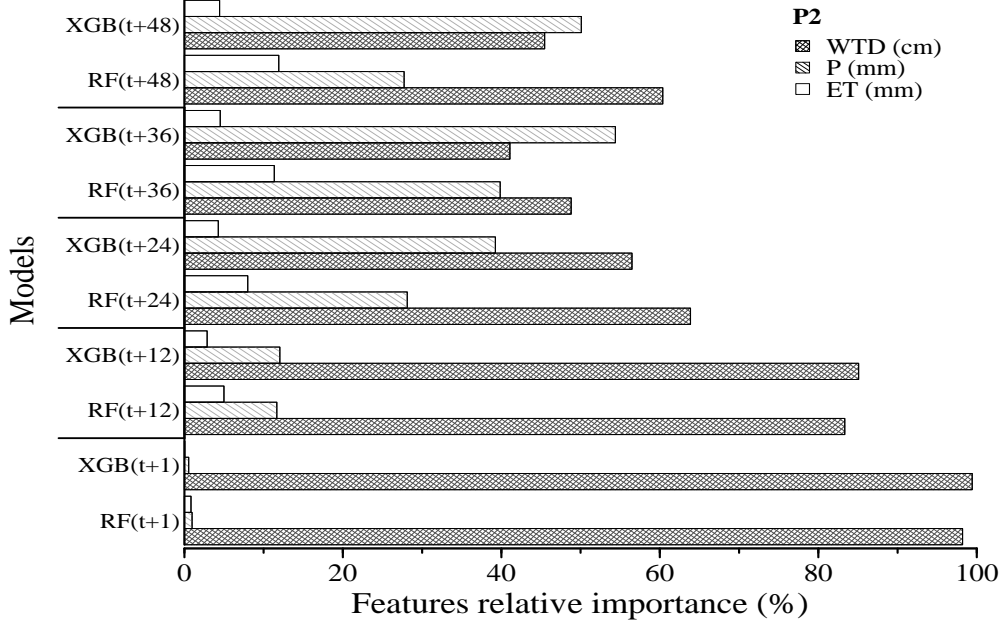
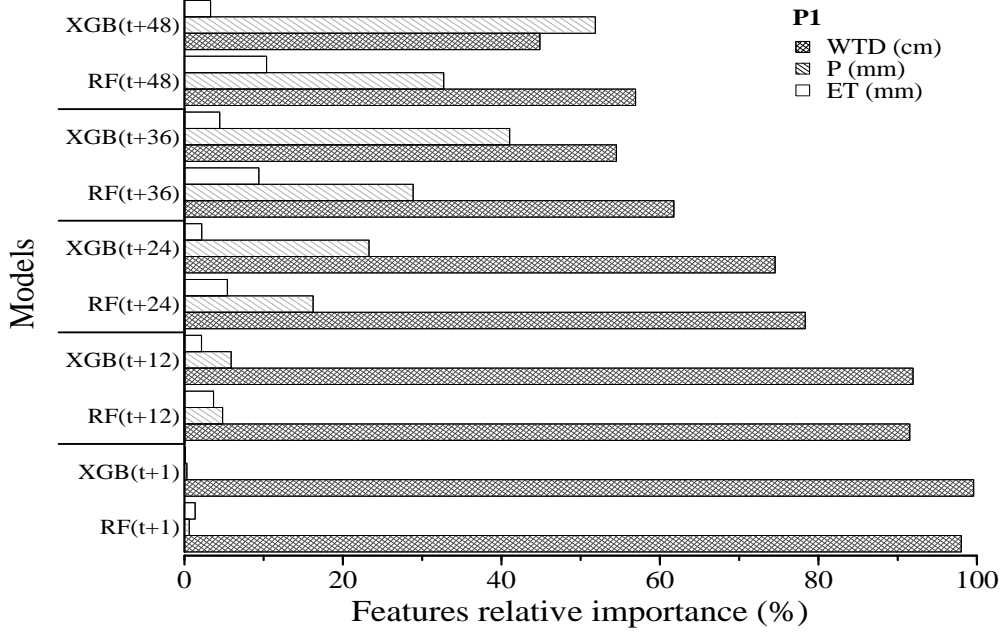
		P1		P2		P3	
		Tr	Ts	Tr	Ts	Tr	Ts
WTD(t + 1)	RF	1	0.981	1	0.985	1	0.986
	XGB	0.990	0.979	1	0.990	0.995	0.986
WTD(t + 12)	RF	0.990	0.931	0.990	0.917	0.990	0.928
	XGB	0.990	0.936	0.990	0.936	0.990	0.929
WTD(t + 24)	RF	0.980	0.833	0.971	0.805	0.979	0.847
	XGB	0.980	0.849	0.990	0.870	0.990	0.860
WTD(t + 36)	RF	0.961	0.711	0.968	0.734	0.968	0.774
	XGB	0.974	0.715	0.982	0.806	0.980	0.797
WTD(t + 48)	RF	0.960	0.689	0.970	0.769	0.961	0.720
	XGB	0.973	0.722	0.984	0.818	0.980	0.760

2.5.4. Predictive variables relative importance

Fig. 7 shows the relative importance of predictor variables for RF and XGB models used to predict the water table depths in the observation wells (P1, P2 and P3). As Fig. 7 shows, the previous water table depths were the most important predictor in the RF model, with mean relative importance between 98% and 56% for forecasting 1, 12, 24, 36 and 48 hours ahead water table depth. The effect of precipitation was also considerable with a mean relative importance between 0.72% and 34%. In the RF model, the WTD importance scores evolve in a manner that is inversely proportional with precipitation as the prediction time increases. On the other hand, the relative evapotranspiration importance was small, with mean relative importance between 0.98% and 10.18% for all hours forecast, probably because ET does not

vary much from hour to hour, compared to precipitation. The same results were obtained using XGB model with different ranks of importance. In the RF and XGB models, the water table depth fluctuation is mainly due to the potential recharge (precipitation) and historical underground water level. The WTD was the most frequently selected predictor variable because it was strongly autocorrelated. The autocorrelation coefficient (acc) of WTD was high (acc = 0.99 for $t - 1$ and acc = 0.48 for $t - 48$), which shows that the previous WTD values are useful and should be considered for forecasting the target variable. Then, the WTD fluctuation in this site is affected quickly by the precipitation because of the unconfined and shallow water table. For the effects of evapotranspiration on the water table depth fluctuation, Fig. 7 shows that hourly oscillations in groundwater levels due to evapotranspiration is less evident in the XGB and RF model, suggesting that the WTD and autoregressive precipitation data already include the information on evapotranspiration. The recharge rate of water table is one of the significant factors affecting the groundwater level, and it can be approximated as a function of precipitation and evapotranspiration (Freeze and Cherry, 1979).

The XGB model was applied to data with the most important predictor variables (model I) and compared to the one using all the variables (model II). For simplicity, only model performance during the test period was displayed (**Table 4**). There was no significant difference between the two forecasting models ($P = 0.980$, $P = 0.997$ and $P = 0.972$ respectively for observation wells P1, P2 and P3). Therefore, the XGB model was trained without evapotranspiration in order to simplify and accelerate its fitting and prediction.



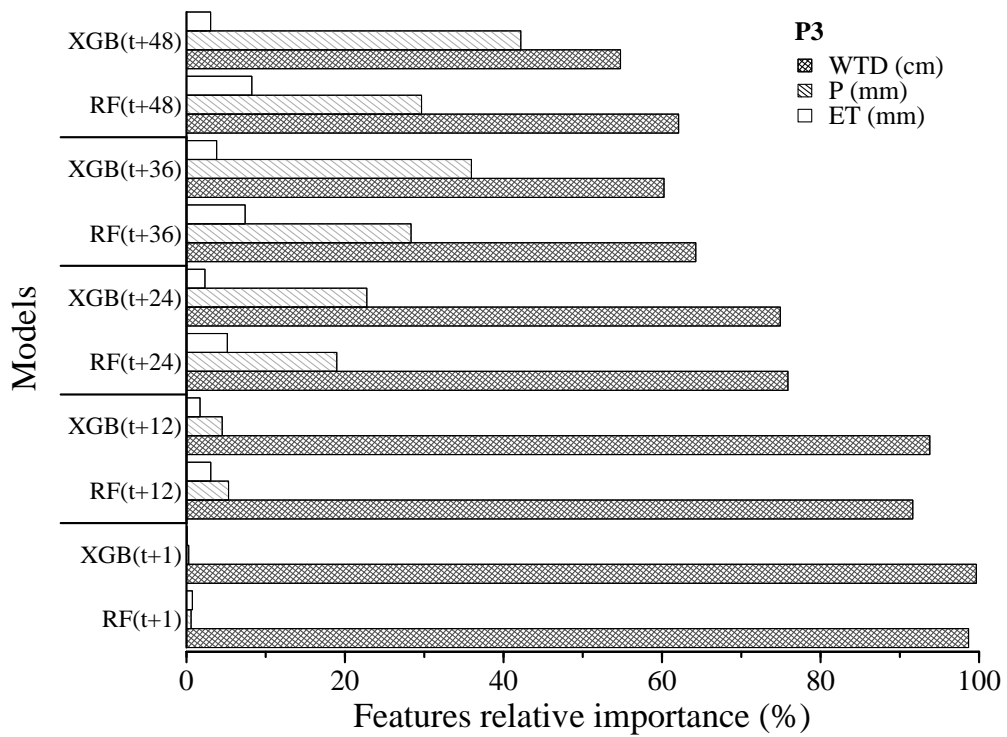


Fig. 7. Relative importance of the predictor variables of both models (RF and XGB) for forecasting 1, 12, 24, 36 and 48 hours ahead the water table depth in the observation wells (P1, P2 and P3).

Table 4. XGB model testing accuracy obtained by the feature selection approach. Model I: all predictive variables (P, WTD and ET); Model II: the most important variables (P and WTD) for forecasting 1, 12, 24, 36 and 48 hours ahead the water table depth in the observation wells (P1, P2 and P3).

	Model	P1		P2		P3	
		RMSE (cm)	NSE	RMSE (cm)	NSE	RMSE (cm)	NSE
$wtd(t + 1)$	I	0.99	0.98	0.95	0.99	1.07	0.99
	II	0.99	0.98	0.94	0.99	1.08	0.99
$wtd(t + 12)$	I	1.73	0.94	2.40	0.94	2.35	0.93
	II	1.85	0.93	2.50	0.93	2.48	0.92
$wtd(t + 24)$	I	2.66	0.85	3.43	0.87	3.32	0.86
	II	2.66	0.85	3.36	0.87	3.33	0.86
$wtd(t + 36)$	I	3.73	0.72	4.23	0.81	4.00	0.80
	II	3.73	0.71	4.20	0.81	4.06	0.79
$wtd(t + 48)$	I	3.62	0.72	4.02	0.82	4.41	0.76
	II	3.59	0.73	4.01	0.82	4.37	0.78

2.5.5. Predicting water table depth

Once the most sensitive variables were determined, the optimal model (XGB) was used for water table depth forecasting at time prediction of 24, 36 and 48-hour. We consider only these time steps because they are more helpful for water management of cranberry fields. Fig. 8 shows the scatter plot of XGB model used only for well P1 during the testing period. According to Fig. 8, the water table depth forecasting for 24-hour forecast is less dispersed than other predictions. The 36 and 48-hour predictions of water table depth were increasingly dispersed because the model tended to overestimate the observed dataset as the prediction time increased. However, the model accuracy increases slightly between 36 and 48-hour

predictions in terms of R^2 . The 48-hour prediction values were overall closer to the observations because XGB model probably fitted better some water table depth and rainfall events for much longer prediction horizons. In general, the XGB model showed good performances in terms of R^2 for all times ahead.

It was found that there is a wider dispersion around the regression line when the water table is close to the soil surface (<50 cm). When the water table was relatively deep (> 50 cm), the XGB model provides satisfactory forecasting, resulting in less deviation from the regression line. To analyze the magnitude of prediction errors, the differences between the observed and simulated water table depths were calculated. These differences (Fig. 9) allow a real appreciation of model errors for each water table depth observation. From Fig. 9, it was found that the model underestimated the high depths of the water table (positive values) and overestimated the shallow water table depths (negative values) in forecasting extreme events, as found by Wang et al. (2018). The prediction error for the entire water table depth range is within ± 5 cm for 24, 36 and 48-hour forecast. In addition, the prediction errors are very high in extreme occurrences present in the data. The prediction errors were found to deteriorate during 36 and 48-hour lead forecast for the low values of WTD (prediction errors between -25 and -30 cm). In this case, the XGB model was more susceptible to overestimate the shallow water table when precipitation relative importance exceeded 40 %. In cranberry fields, the groundwater level was kept to 60 cm below the ground surface level for production purposes, the water table depths between 15 and 30 cm are rather rare. Because this is a rare event caused by heavy rainfall, the XGB model could probably not learn a meaningful prediction for this range. Therefore, WTD forecasting should be trained from similar weather events (rainy, dry) to take into account a rapid drop or rise in groundwater (Wang et al., 2018).

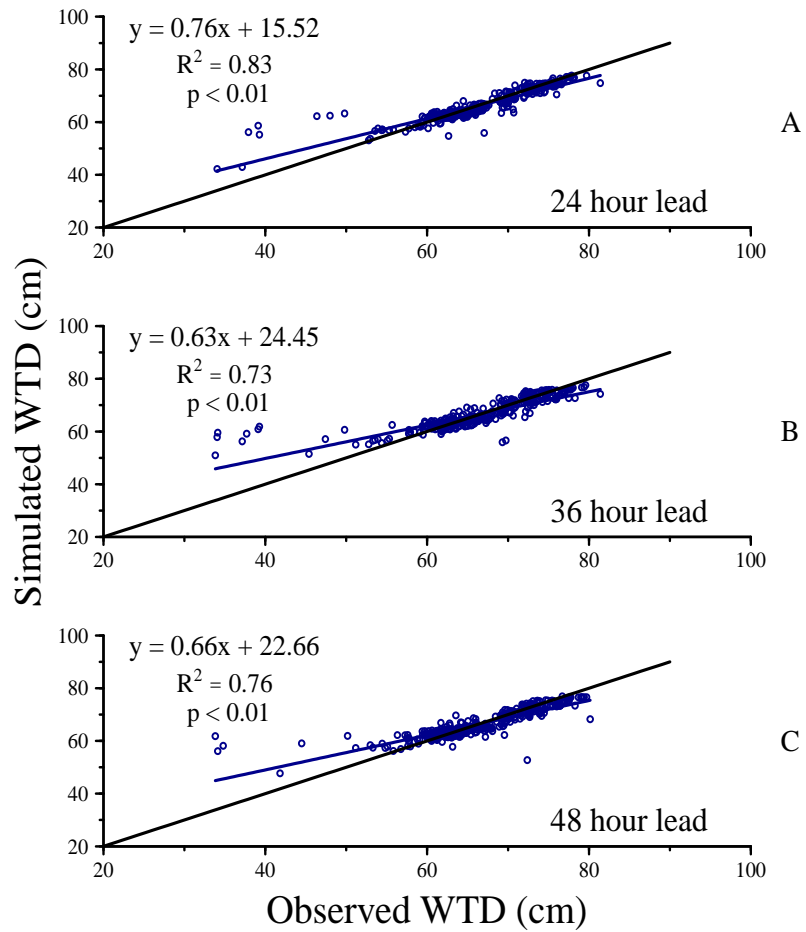


Fig. 8. Observed and simulated water table depth Of XGB model for 24-hour forecast (A), 36- hour forecast (B) and 48-hour forecast (B) during the testing period for the observation well P1.

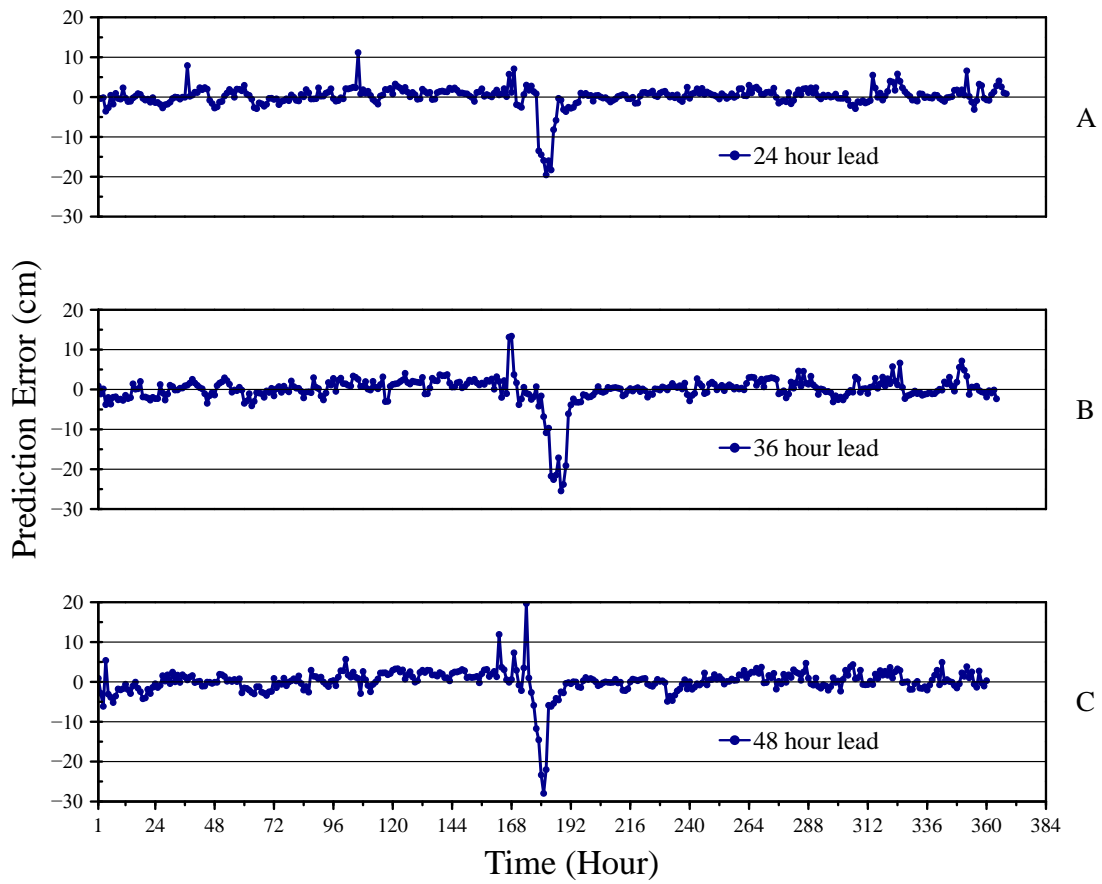


Fig. 9. The prediction error of XGB model for 24-hour forecast (A), 36-hour forecast (B) and 48-hour forecast (C) during the testing period for the observation well P1.

2.6. Conclusions

Two decision-tree models were applied to forecast the water table depth in cranberries fields. The RF and XGB models were trained and compared to perform 1, 12, 24, 36 and 48 hours ahead predictions using the historical observations of WTD, precipitation and evapotranspiration as predictor variables. It was found that the XGB model had superior predictive ability to produce accurate water table depth simulations for the longer periods than RF in terms of RMSE and NSE. Based on the optimal model, the prediction error for each water table depth observation was calculated for each observation well (P1, P2 and P3), and results were moderate and within an acceptable limit (± 5 cm). However, the prediction error for the shallow water table depths were found to deteriorate during 36 and 48-hour forecast. Overall, the decision tree models proposed in this study show satisfactory results for water table depths forecasting and offer a powerful and useful tool for water resource management of cranberries fields.

Furthermore, the predictor variables importance was evaluated to analyze the relationship between the target variable (water table depth) and the set of predictors in each well. The results indicated that the previous water depth was the most appropriate predictor to forecast water table depth. Moreover, the precipitation had also important effect on water table depth fluctuation. In contrast, evapotranspiration was the predictor that had no significant effect on the water table dynamics.

The developed models are a practical technique to represent the water flow pattern in cranberry fields. The definition of others hydrological parameters such as the aquifer hydraulic conductivity, storage coefficients variation and drainage rate can be essential to produce accurate water table depth simulations. In general, the XGB model offers cranberry growers a powerful tool that accurately simulates the water table depth. It could be used as an efficient decision tool in precision irrigation of cranberry fields, which provide information useful on the WTD variation dynamic for water planning and management purposes.

Acknowledgements

We want to acknowledge Jonathan Lafond and Guillaume Letourneau for providing the field data. Thanks to Cintia Racine, Jean Pascal Matteau and Thiago Gumiere for their useful comments that helped improve this manuscript.

Bibliographie

- Allen, R.G., Pereira, L.S., Raes, D., Smith, M., FAO, 1990. Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56.
- Analytics Vidhya, 2016. A Complete Tutorial on Tree Based Modeling from Scratch (in R & Python) [WWW Document]. Complete Tutor. Tree Based Model. Scratch R Python. URL <https://www.analyticsvidhya.com/blog/2016/04/complete-tutorial-tree-based-modeling-scratch-in-python/#> (accessed 5.29.19).
- Baier, W., Robertson, Geo.W., 1965. Estimation of latent evaporation from simple weather observations. *Can. J. Plant Sci.* 45, 276–284. <https://doi.org/10.4141/cjps65-051>
- Bigah, Y., Rousseau, A.N., Gumiere, S.J., 2019. Development of a steady-state model to predict daily water table depth and root zone soil matric potential of a cranberry field with a subirrigation system. *Agric. Water Manag.* 213, 1016–1027. <https://doi.org/10.1016/j.agwat.2018.12.024>
- Breiman, L., 2002. Manual on setting up, using, and understanding random forests v3. 1. Stat. Dep. Univ. Calif. Berkeley CA USA 1.
- Breiman, L., 2001. Random Forests. *Mach. Learn.* 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Breiman, L., Cutler, A., 2018. Breiman and Cutler’s Random Forests for Classification and Regression.
- Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification and regression trees. Chapman & Hal. Wadsworth. Inc N. Y. 368.
- Caron, J., Bonin, S., Pepin, S., Kummer, L., Vanderleest, C., Bland, W.L., 2016. Determination of irrigation set points for cranberries from soil-and plant-based measurements. *Can. J. Soil Sci.* 96, 37–50.
- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., 2015. Xgboost: extreme gradient boosting. R Package Version 04-2 1–4.
- Chen, Z., Grasby, S.E., Osadetz, K.G., 2002. Predicting average annual groundwater levels from climatic variables: an empirical model. *J. Hydrol.* 260, 102–117. [https://doi.org/10.1016/S0022-1694\(01\)00606-0](https://doi.org/10.1016/S0022-1694(01)00606-0)
- Coulibaly, P., Anctil, F., Aravena, R., Bobée, B., 2001. Artificial neural network modeling of water table depth fluctuations. *Water Resour. Res.* 37, 885–896. <https://doi.org/10.1029/2000WR900368>
- Daliakopoulos, I.N., Coulibaly, P., Tsanis, I.K., 2005. Groundwater level forecasting using artificial neural networks. *J. Hydrol.* 309, 229–240. <https://doi.org/10.1016/j.jhydrol.2004.12.001>
- Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. *J. Anim. Ecol.* 77, 802–813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>

- Environment Canada, 2019. Archives sur le climat [WWW Document]. URL http://climat.meteo.gc.ca/climate_normals/results_f.html?stnID=5310&dispBack=0&month1=0&month2=12 (accessed 8.16.19).
- Freeze, R.A., Cherry, J.A., 1979. Groundwater: Englewood Cliffs. N. J.
- Friedman, J.H., Meulman, J.J., 2003. Multiple additive regression trees with application in epidemiology. *Stat. Med.* 22, 1365–1381. <https://doi.org/10.1002/sim.1501>
- Ghose, D.K., Panda, S.S., Swain, P.C., 2010. Prediction of water table depth in western region, Orissa using BPNN and RBFN neural networks. *J. Hydrol.* 394, 296–304. <https://doi.org/10.1016/j.jhydrol.2010.09.003>
- Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M., Toulmin, C., 2010. Food security: the challenge of feeding 9 billion people. *science* 327, 812–818.
- Godinho, S., Guiomar, N., Gil, A., 2018. Estimating tree canopy cover percentage in a mediterranean silvopastoral systems using Sentinel-2A imagery and the stochastic gradient boosting algorithm. *Int. J. Remote Sens.* 39, 4640–4662. <https://doi.org/10.1080/01431161.2017.1399480>
- Guo, L., Chehata, N., Mallet, C., Boukir, S., 2011. Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests. *ISPRS J. Photogramm. Remote Sens.* 66, 56–66. <https://doi.org/10.1016/j.isprsjprs.2010.08.007>
- Hu, Y., Quinn, C.J., Cai, X., Garfinkle, N.W., 2017. Combining human and machine intelligence to derive agents' behavioral rules for groundwater irrigation. *Adv. Water Resour.* 109, 29–40. <https://doi.org/10.1016/j.advwatres.2017.08.009>
- Kennedy, C., Jeranyama, P., Alverson, N., 2017. Agricultural water requirements for commercial production of cranberries. *Can. J. SOIL Sci.* 97, 38–45. <https://doi.org/10.1139/cjss-2015-0095>
- Kisi, O., Shiri, J., Nikoofar, B., 2012. Forecasting daily lake levels using artificial intelligence approaches. *Comput. Geosci.* 41, 169–180. <https://doi.org/10.1016/j.cageo.2011.08.027>
- Krause, P., Boyle, D.P., Bäse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. *Adv. Geosci.* 5, 89–97. <https://doi.org/10.5194/adgeo-5-89-2005>
- Kuhn, M., 2008. Building Predictive Models in R Using the caret Package. *J. Stat. Softw.* 28. <https://doi.org/10.18637/jss.v028.i05>
- Lee, B., Hamm, S.-Y., Jang, S., Cheong, J.-Y., Kim, G.-B., 2014. Relationship between groundwater and climate change in South Korea. *Geosci. J.* 18, 209–218. <https://doi.org/10.1007/s12303-013-0062-7>
- Li, B., Yang, G., Wan, R., Dai, X., Zhang, Y., 2016. Comparison of random forests and other statistical methods for the prediction of lake water level: a case study of the Poyang Lake in China. *Hydrol. Res.* 47, 69–83. <https://doi.org/10.2166/nh.2016.264>

- Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. *R News* 2, 18–22.
- Maier, H.R., Jain, A., Dandy, G.C., Sudheer, K.P., 2010. Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. *Environ. Model. Softw.* 25, 891–909. <https://doi.org/10.1016/j.envsoft.2010.02.003>
- Ministère de l'Environnement et de la Lutte contre les Changements Climatiques, 2019. *Eaux souterraines* [WWW Document]. URL <http://www.environnement.gouv.qc.ca/eau/souterraines/index.htm> (accessed 6.3.19).
- Moritz, S., Bartz-Beielstein, T., 2017. imputeTS: time series missing value imputation in R. *R J.* 9, 207–218.
- Naghibi, S.A., Pourghasemi, H.R., Dixon, B., 2016. GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. *Environ. Monit. Assess.* 188, 44. <https://doi.org/10.1007/s10661-015-5049-6>
- Nayak, P.C., Rao, Y.R.S., Sudheer, K.P., 2006. Groundwater Level Forecasting in a Shallow Aquifer Using Artificial Neural Network Approach. *Water Resour. Manag.* 20, 77–90. <https://doi.org/10.1007/s11269-006-4007-z>
- Nembrini, S., König, I.R., Wright, M.N., 2018. The revival of the Gini importance? *Bioinformatics* 34, 3711–3718. <https://doi.org/10.1093/bioinformatics/bty373>
- Pelletier, V., Gallichand, J., Caron, J., Jutras, S., Marchand, S., 2015a. Critical irrigation threshold and cranberry yield components. *Agric. Water Manag.* 148, 106–112. <https://doi.org/10.1016/j.agwat.2014.09.025>
- Pelletier, V., Gallichand, J., Gumiere, S., Pepin, S., Caron, J., 2015b. Water Table Control for Increasing Yield and Saving Water in Cranberry Production. *Sustainability* 7, 10602–10619. <https://doi.org/10.3390/su70810602>
- R Core Team, 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2012. URL [Httpwww R-Proj. Org](http://www.R-Project.org).
- Rodriguez-Galiano, V., Mendes, M.P., Garcia-Soldado, M.J., Chica-Olmo, M., Ribeiro, L., 2014. Predictive modeling of groundwater nitrate pollution using Random Forest and multisource variables related to intrinsic and specific vulnerability: A case study in an agricultural setting (Southern Spain). *Sci. Total Environ.* 476–477, 189–206. <https://doi.org/10.1016/j.scitotenv.2014.01.001>
- Roe, B.P., Yang, H.-J., Zhu, J., Liu, Y., Stancu, I., McGregor, G., 2005. Boosted decision trees as an alternative to artificial neural networks for particle identification. *Nucl. Instrum. Methods Phys. Res. Sect. Accel. Spectrometers Detect. Assoc. Equip.* 543, 577–584. <https://doi.org/10.1016/j.nima.2004.12.018>
- Samson, M., Fortin, J., Pepin, S., Caron, J., 2016. Impact of potassium sulfate salinity on growth and development of cranberry plants subjected to overhead and subirrigation. *Can. J. Soil Sci.* 97, 20–30.

- Scornet, E., 2017. Tuning parameters in random forests. *ESAIM Proc. Surv.* 60, 144–162. <https://doi.org/10.1051/proc/201760144>
- Singh, K.P., Gupta, S., Mohan, D., 2014a. Evaluating influences of seasonal variations and anthropogenic activities on alluvial groundwater hydrochemistry using ensemble learning approaches. *J. Hydrol.* 511, 254–266. <https://doi.org/10.1016/j.jhydrol.2014.01.004>
- Singh, K.P., Gupta, S., Mohan, D., 2014b. Evaluating influences of seasonal variations and anthropogenic activities on alluvial groundwater hydrochemistry using ensemble learning approaches. *J. Hydrol.* 511, 254–266. <https://doi.org/10.1016/j.jhydrol.2014.01.004>
- Taormina, R., Chau, K., Sethi, R., 2012. Artificial neural network simulation of hourly groundwater levels in a coastal aquifer system of the Venice lagoon. *Eng. Appl. Artif. Intell.* 25, 1670–1676. <https://doi.org/10.1016/j.engappai.2012.02.009>
- Taylor, R.G., Scanlon, B., Döll, P., Rodell, M., van Beek, R., Wada, Y., Longuevergne, L., Leblanc, M., Famiglietti, J.S., Edmunds, M., Konikow, L., Green, T.R., Chen, J., Taniguchi, M., Bierkens, M.F.P., MacDonald, A., Fan, Y., Maxwell, R.M., Yecheili, Y., Gurdak, J.J., Allen, D.M., Shamsudduha, M., Hiscock, K., Yeh, P.J.-F., Holman, I., Treidel, H., 2013. Ground water and climate change. *Nat. Clim. Change* 3, 322–329. <https://doi.org/10.1038/nclimate1744>
- Vanderleest, C.P.L., Caron, J., Bland, W.L., 2016. Water table level management as an irrigation strategy for cranberry (*Vaccinium macrocarpon* Aiton) 1. *Can. J. Soil Sci.* 1–9. <https://doi.org/10.1139/cjss-2016-0001>
- Wang, X., Liu, T., Zheng, X., Peng, H., Xin, J., Zhang, B., 2018. Short-term prediction of groundwater level using improved random forest regression with a combination of random features. *Appl. Water Sci.* 8, 125. <https://doi.org/10.1007/s13201-018-0742-6>
- Wu, W., Dandy, G.C., Maier, H.R., 2014. Protocol for developing ANN models and its application to the assessment of the quality of the ANN model development process in drinking water quality modelling. *Environ. Model. Softw.* 54, 108–127. <https://doi.org/10.1016/j.envsoft.2013.12.016>
- Yoon, H., Jun, S.-C., Hyun, Y., Bae, G.-O., Lee, K.-K., 2011. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *J. Hydrol.* 396, 128–138. <https://doi.org/10.1016/j.jhydrol.2010.11.002>

Conclusion générale

Le but de cette étude a été de développer un modèle d'apprentissage automatique capable de prédire la profondeur de la nappe phréatique dans les champs de canneberges qui sont en condition de drainage et d'irrigation. Deux modèles d'arbres de décision ont été utilisés pour modéliser le comportement dynamique et non linéaire de la nappe phréatique. Le modèle « Arbres décisionnels renforcés », plus complexe, a été comparé à un modèle plus simple, « Forêt aléatoire », pour déterminer lequel a été le plus efficace à construire un modèle de prédiction de la profondeur de la nappe phréatique dans un champ de canneberges. Ces modèles ont été développés en utilisant les profondeurs antérieures de la nappe phréatique, la précipitation, l'irrigation et l'évapotranspiration comme variables prédictives. Ces mesures provenant des puits d'observation de la nappe phréatique et de la station météorologique ont été très utiles pour la procédure de calibration et de validation des modèles.

La calibration et la validation ont permis d'optimiser les modèles et de comparer les résultats simulés et observés. Dans la procédure de calibration, un ensemble de paramètres ont été testés pour permettre aux modèles de déterminer la relation entre les variables prédictives et la variable réponse. Après cette étape, les modèles ont été validés sur un ensemble de données non utilisé pour leur entraînement. Cela a permis de vérifier la capacité des modèles à bien représenter la tendance et le comportement dynamique du système réel.

Des deux modèles, les « Arbres décisionnels renforcés » ont été les plus performants dans la prédiction jusqu'à 48 heures de la profondeur de la nappe phréatique dans un champ de canneberges et ont été considérés comme le modèle optimal. Les erreurs de simulation de ce modèle optimal par rapport aux mesures d'observation de la profondeur de la nappe phréatique ont été faibles et dans la limite acceptable (environ de ± 5 cm). Cependant, les faibles profondeurs de nappes phréatiques (entre 15 à 30 cm) entraînent une surestimation des résultats du modèle et occasionnent des erreurs de prédiction plus élevées. Cela constitue une limitation du modèle qui doit être entraîné sur des plages de valeurs similaires pour avoir une précision plus élevée. Outre la prédiction de la profondeur de la nappe phréatique, « les arbres décisionnels renforcés » ont permis d'identifier les variables prédictives qui conditionnent la fluctuation de la profondeur de la nappe phréatique dans le champ de canneberges. Parmi les facteurs prédictifs utilisés, les profondeurs historiques de la nappe

phréatique et la précipitation ont été les variables les plus importantes dans la simulation de la profondeur de la nappe phréatique. L'effet de l'évapotranspiration estimée sur la dynamique de la nappe phréatique a été faible, car l'évapotranspiration varie beaucoup moins par rapport aux profondeurs de la nappe et la précipitation. Ce résultat reste vrai, car l'analyse du mouvement d'eau souterraine apporte la démonstration que, dans les champs de canneberges en situation de drainage et d'irrigation pour maintenir un niveau approprié de l'eau souterraine, l'effet de l'évapotranspiration sur la nappe reste faible.

Les modèles d'apprentissage automatique offrent un outil puissant pour prédire la profondeur de la nappe phréatique de manière précise afin de maintenir et de gérer une irrigation de précision. Cependant, leurs performances dans l'hydrologie des eaux souterraines des champs de canneberges restent à être améliorées en introduisant d'autres variables telles que les débits de drainage et la variation du niveau d'eau de la boîte de contrôle.

Cette étude a été réalisée avec l'approche des arbres de décision pour prédire la variation de la nappe phréatique. Des études futures pourraient inclure d'autres approches comme des méthodes analytiques ou d'autres catégories de l'apprentissage automatique afin d'élargir les horizons de prédiction de l'eau souterraine des champs de canneberges. Néanmoins, dans cette étude, le modèle d'apprentissage automatique développé ouvre la voie aux producteurs de canneberges à la possibilité de faire : 1) la prévision des profondeurs de nappe phréatique pour maintenir une irrigation de précision, 2) la reconstruction de périodes d'observation manquantes du niveau d'eau souterraine et 3) l'analyse des facteurs qui conditionnent la fluctuation de la profondeur de la nappe phréatique.

Bibliographie générale

- Adamowski, J., Chan, H.F., 2011. A wavelet neural network conjunction model for groundwater level forecasting. *J. Hydrol.* 407, 28–40.
<https://doi.org/10.1016/j.jhydrol.2011.06.013>
- Aghzar, N., Berdai, H., Bellouti, A., Soudi, B., 2002. Pollution nitrique des eaux souterraines au Tadla (Maroc). *Rev. Sci. Eau* 15, 459. <https://doi.org/10.7202/705465ar>
- Allen, R.G., Pereira, L.S., Raes, D., Smith, M., FAO, 1990. Crop evapotranspiration - Guidelines for computing crop water requirements - FAO Irrigation and drainage paper 56.
- Analytics Vidhya, 2016. A Complete Tutorial on Tree Based Modeling from Scratch (in R & Python) [WWW Document]. Complete Tutor. Tree Based Model. Scratch R Python. URL <https://www.analyticsvidhya.com/blog/2016/04/complete-tutorial-tree-based-modeling-scratch-in-python/#> (accessed 5.29.19).
- Asselin, R., 2005. Petit fruit deviendra grand... comme la canneberge. « Journées Agri-vision », Montérégie, Ministère de l’Agriculture, des Pêcheries et de l’Alimentation du Québec (MAPAQ).
- Association des Producteurs de Canneberges du Québec, 2019. Statistiques [WWW Document]. URL <http://www.notrecanneberge.com/Contenu/Page/Statistiques>
- Baier, W., Robertson, Geo.W., 1965. Estimation of latent evaporation from simple weather observations. *Can. J. Plant Sci.* 45, 276–284. <https://doi.org/10.4141/cjps65-051>
- Baumann, D.L., Workmaster, B.A., Kosola, K.R., 2005. ‘Ben Lear’ and ‘Stevens’ Cranberry Root and Shoot Growth Response to Soil Water Potential 40, 4.
- Bigah, Y., Rousseau, A.N., Gumiere, S.J., 2019. Development of a steady-state model to predict daily water table depth and root zone soil matric potential of a cranberry field with a subirrigation system. *Agric. Water Manag.* 213, 1016–1027.
<https://doi.org/10.1016/j.agwat.2018.12.024>
- Bonin, S., 2009. Département des sols et de génie agroalimentaire Faculté des Sciences de l’Agriculture et de l’Alimentation Université Laval Québec 114.
- Breiman, L., 2001. Random Forests. *Mach. Learn.* 45, 5–32.
<https://doi.org/10.1023/A:1010933404324>
- Breiman, L., 2001. Random Forests. *Mach. Learn.* 45, 5–32.
<https://doi.org/10.1023/A:1010933404324>
- Breiman, L., 2002. Manual on setting up, using, and understanding random forests v3. 1. Stat. Dep. Univ. Calif. Berkeley CA USA 1.
- Breiman, L., Cutler, A., 2018. randomForest: Breiman and Cutler’s Random Forests for Classification and Regression, Version 4.6 [WWW Document]. URL <https://www.stat.berkeley.edu/~breiman/RandomForests/> (accessed 10.23.19).
- Breiman, L., Friedman, J., Olshen, R., Stone, C., 1984. Classification and regression trees. Chapman & Hal. Wadsworth. Inc N. Y. 368.

- Burton, M., 2010. Irrigation management: principles and practices. CABI North American Office, Cambridge, MA.
- C.-C. Yang, S. O. Prasher, R. Lacroix, 1996. Applications of Artificial Neural Networks to Land Drainage Engineering. *Trans. ASAE* 39, 525–533.
<https://doi.org/10.13031/2013.27531>
- Caron, J., Bonin, S., Pepin, S., Kummer, L., Vanderleest, C., Bland, W.L., 2016. Determination of irrigation set points for cranberries from soil-and plant-based measurements. *Can. J. Soil Sci.* 96, 37–50.
- Carretero, S.C., Kruse, E.E., 2012. Relationship between precipitation and water-table fluctuation in a coastal dune aquifer: northeastern coast of the Buenos Aires province, Argentina. *Hydrogeol. J.* 20, 1613–1621. <https://doi.org/10.1007/s10040-012-0890-y>
- Carretero, S.C., Kruse, E.E., 2012. Relationship between precipitation and water-table fluctuation in a coastal dune aquifer: northeastern coast of the Buenos Aires province, Argentina. *Hydrogeol. J.* 20, 1613–1621. <https://doi.org/10.1007/s10040-012-0890-y>
- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., 2015. Xgboost: extreme gradient boosting. R Package Version 04-2 1–4.
- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., 2015. Xgboost: extreme gradient boosting. R Package Version 04-2 1–4.
- Chen, Z., Grasby, S.E., Osadetz, K.G., 2002. Predicting average annual groundwater levels from climatic variables: an empirical model. *J. Hydrol.* 260, 102–117.
[https://doi.org/10.1016/S0022-1694\(01\)00606-0](https://doi.org/10.1016/S0022-1694(01)00606-0).
- Chen, Z., Grasby, S.E., Osadetz, K.G., 2002. Predicting average annual groundwater levels from climatic variables: an empirical model. *J. Hydrol.* 260, 102–117.
[https://doi.org/10.1016/S0022-1694\(01\)00606-0](https://doi.org/10.1016/S0022-1694(01)00606-0)
- Coulibaly, P., Anctil, F., Aravena, R., Bobée, B., 2001. Artificial neural network modeling of water table depth fluctuations. *Water Resour. Res.* 37, 885–896.
<https://doi.org/10.1029/2000WR900368>
- Coulibaly, P., Anctil, F., Aravena, R., Bobée, B., 2001. Artificial neural network modeling of water table depth fluctuations. *Water Resour. Res.* 37, 885–896.
<https://doi.org/10.1029/2000WR900368>
- Daliakopoulos, I.N., Coulibaly, P., Tsanis, I.K., 2005. Groundwater level forecasting using artificial neural networks. *J. Hydrol.* 309, 229–240.
<https://doi.org/10.1016/j.jhydrol.2004.12.001>
- Daliakopoulos, I.N., Coulibaly, P., Tsanis, I.K., 2005. Groundwater level forecasting using artificial neural networks. *J. Hydrol.* 309, 229–240.
<https://doi.org/10.1016/j.jhydrol.2004.12.001>
- Eck, P., 1990. The American cranberry Rutgers Univ. Press N. B. NJ.
- Elith, J., Leathwick, J.R., Hastie, T., 2008. A working guide to boosted regression trees. *J. Anim. Ecol.* 77, 802–813. <https://doi.org/10.1111/j.1365-2656.2008.01390.x>

- Environment Canada, 2019. Archives sur le climat [WWW Document]. URL http://climat.meteo.gc.ca/climate_normals/results_f.html?stnID=5310&dispBack=0&month1=0&month2=12 (accessed 8.16.19).
- Food and Agriculture Organization of the United Nations, 2017. Production share of cranberries by region [WWW Document]. URL <http://www.fao.org/faostat/en/#data/QC/visualize>
- Freeman, E.A., Moisen, G.G., Coulston, J.W., Wilson, B.T., 2016. Random forests and stochastic gradient boosting for predicting tree canopy cover: comparing tuning processes and model performance 1. *Canadian Journal of Forest Research* 46, 323–339. <https://doi.org/10.1139/cjfr-2014-0562>
- Freeze, R.A., Cherry, J.A., 1979. *Groundwater*: Englewood Cliffs. N. J.
- Friedman, J.H., Meulman, J.J., 2003. Multiple additive regression trees with application in epidemiology. *Stat. Med.* 22, 1365–1381. <https://doi.org/10.1002/sim.1501>
- Ghose, D.K., Panda, S.S., Swain, P.C., 2010. Prediction of water table depth in western region, Orissa using BPNN and RBFN neural networks. *J. Hydrol.* 394, 296–304. <https://doi.org/10.1016/j.jhydrol.2010.09.003>
- Godfray, H.C.J., Beddington, J.R., Crute, I.R., Haddad, L., Lawrence, D., Muir, J.F., Pretty, J., Robinson, S., Thomas, S.M., Toulmin, C., 2010. Food security: the challenge of feeding 9 billion people. *science* 327, 812–818.
- Godinho, S., Guiomar, N., Gil, A., 2018. Estimating tree canopy cover percentage in a mediterranean silvopastoral systems using Sentinel-2A imagery and the stochastic gradient boosting algorithm. *Int. J. Remote Sens.* 39, 4640–4662. <https://doi.org/10.1080/01431161.2017.1399480>
- Gumiere, S.J., Lafond, J.A., Hallema, D.W., Périard, Y., Caron, J., Gallichand, J., 2014. Mapping soil hydraulic conductivity and matric potential for water management of cranberry: Characterisation and spatial interpolation methods. *Biosyst. Eng.* 128, 29–40. <https://doi.org/10.1016/j.biosystemseng.2014.09.002>
- Guo, L., Chehata, N., Mallet, C., Boukir, S., 2011. Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests. *ISPRS J. Photogramm. Remote Sens.* 66, 56–66. <https://doi.org/10.1016/j.isprsjprs.2010.08.007>
- Hu, Y., Quinn, C.J., Cai, X., Garfinkle, N.W., 2017. Combining human and machine intelligence to derive agents' behavioral rules for groundwater irrigation. *Adv. Water Resour.* 109, 29–40. <https://doi.org/10.1016/j.advwatres.2017.08.009>
- Jarjoui, M., Geahchan, A., Boutros, E., Abou-Kaïs, A., 2000. Pollution des eaux souterraines par les hydrocarbures aromatiques polycycliques et évaluation du risque. *Houille Blanche* 89–92. <https://doi.org/10.1051/lhb/2000080>
- Jeranyama, P., Demoranville, C., Waddell, J., 2014. Using canopy temperature, soil tension and moisture measurements as tools in cranberry irrigation. *acta hort.* 1017, 487–492. <https://doi.org/10.17660/ActaHortic.2014.1017.60>

- Jordan, M.I., Mitchell, T.M., 2015. Machine learning: Trends, perspectives, and prospects. *Science* 349, 255. <https://doi.org/10.1126/science.aaa8415>
- Kennedy, C., Jeranyama, P., Alverson, N., 2017. Agricultural water requirements for commercial production of cranberries. *Can. J. SOIL Sci.* 97, 38–45. <https://doi.org/10.1139/cjss-2015-0095>
- Kisi, O., Shiri, J., Nikoofar, B., 2012. Forecasting daily lake levels using artificial intelligence approaches. *Comput. Geosci.* 41, 169–180. <https://doi.org/10.1016/j.cageo.2011.08.027>
- Kotsiantis, S.B., Zaharakis, I., Pintelas, P., 2007. Supervised machine learning: A review of classification techniques. *Emerg. Artif. Intell. Appl. Comput. Eng.* 160, 3–24.
- Krause, P., Boyle, D.P., Bäse, F., 2005. Comparison of different efficiency criteria for hydrological model assessment. *Adv. Geosci.* 5, 89–97. <https://doi.org/10.5194/adgeo-5-89-2005>
- Kuhn, M., 2008. Building Predictive Models in R Using the caret Package. *J. Stat. Softw.* 28. <https://doi.org/10.18637/jss.v028.i05>
- Laurent, T., 2015. Réponse de la canneberge (*Vaccinium macrocarpon* Ait.) à l'aération du sol 101.
- Lee, B., Hamm, S.-Y., Jang, S., Cheong, J.-Y., Kim, G.-B., 2014. Relationship between groundwater and climate change in South Korea. *Geosci. J.* 18, 209–218. <https://doi.org/10.1007/s12303-013-0062-7>
- Létourneau, G., 2017. Approche multicritère d'optimisation de l'irrigation goutte-à-goutte du fraisier 169. <http://hdl.handle.net/20.500.11794/33286>
- Li, B., Yang, G., Wan, R., Dai, X., Zhang, Y., 2016. Comparison of random forests and other statistical methods for the prediction of lake water level: a case study of the Poyang Lake in China. *Hydrol. Res.* 47, 69–83. <https://doi.org/10.2166/nh.2016.264>
- Liakos, K., Busato, P., Moshou, D., Pearson, S., Bochtis, D., 2018. Machine learning in agriculture: A review. *Sensors* 18, 2674.
- Liao, C.T., Lin, C.H., 1995. Effect of flood stress on morphology and anaerobic metabolism of *Momordica charantia*. *Environ. Exp. Bot.* 35, 105–113. [https://doi.org/10.1016/0098-8472\(94\)00048-A](https://doi.org/10.1016/0098-8472(94)00048-A)
- Liao, C.-T., Lin, C.-H., 2001. Physiological Adaptation of Crop Plants to Flooding Stress 10.
- Liaw, A., Wiener, M., 2002. Classification and regression by randomForest. *R News* 2, 18–22.
- Maier, H.R., Jain, A., Dandy, G.C., Sudheer, K.P., 2010. Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. *Environ. Model. Softw.* 25, 891–909. <https://doi.org/10.1016/j.envsoft.2010.02.003>
- McQueen, R.J., Garner, S.R., Nevill-Manning, C.G., Witten, I.H., 1995. Applying machine learning to agricultural data. *Comput. Electron. Agric.* 12, 275–293. [https://doi.org/10.1016/0168-1699\(95\)98601-9](https://doi.org/10.1016/0168-1699(95)98601-9)

- Ministère de l'Environnement et de la Lutte contre les Changements Climatiques, 2019. Eaux souterraines [WWW Document]. URL <http://www.environnement.gouv.qc.ca/eau/souterraines/index.htm> (accessed 6.3.19).
- Mitchell, T.M., 1997. Machine learning.
- Moritz, S., Bartz-Beielstein, T., 2017. imputeTS: time series missing value imputation in R. *R J.* 9, 207–218.
- Naghibi, S.A., Pourghasemi, H.R., Dixon, B., 2016. GIS-based groundwater potential mapping using boosted regression tree, classification and regression tree, and random forest machine learning models in Iran. *Environ. Monit. Assess.* 188, 44. <https://doi.org/10.1007/s10661-015-5049-6>
- Nayak, P.C., Rao, Y.R.S., Sudheer, K.P., 2006. Groundwater Level Forecasting in a Shallow Aquifer Using Artificial Neural Network Approach. *Water Resour. Manag.* 20, 77–90. <https://doi.org/10.1007/s11269-006-4007-z>
- Ndou, N.N., Palamuleni, L.G., Ramoelo, A., 2018. Modelling depth to groundwater level using SEBAL-based dry season potential evapotranspiration in the upper Molopo River Catchment, South Africa. *Egypt. J. Remote Sens. Space Sci.* 21, 237–248. <https://doi.org/10.1016/j.ejrs.2017.08.003>
- Nembrini, S., König, I.R., Wright, M.N., 2018. The revival of the Gini importance? *Bioinformatics* 34, 3711–3718. <https://doi.org/10.1093/bioinformatics/bty373>
- Pelletier, V., Gallichand, J., Caron, J., Jutras, S., Marchand, S., 2015a. Critical irrigation threshold and cranberry yield components. *Agric. Water Manag.* 148, 106–112. <https://doi.org/10.1016/j.agwat.2014.09.025>
- Pelletier, V., Gallichand, J., Gumiere, S., Caron, J., 2016a. Impact of drainage problems on cranberry yields: Two case studies. *Can. J. Soil Sci.* 97, 1–4.
- Pelletier, V., Gallichand, J., Gumiere, S., Pepin, S., Caron, J., 2015b. Water Table Control for Increasing Yield and Saving Water in Cranberry Production. *Sustainability* 7, 10602–10619. <https://doi.org/10.3390/su70810602>
- Pelletier, V., Pepin, S., Laurent, T., Gallichand, J., Caron, J., 2016b. Cranberry Gas Exchange under Short-term Hypoxic Soil Conditions. *HortScience* 51, 910–914. <https://doi.org/10.21273/HORTSCI.51.7.910>
- R Core Team, 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2012. URL [Httpwww R-Project. Org](http://www.R-Project.org).
- Rai, S.N., Manglik, A., Singh, V.S., 2006. Water table fluctuation owing to time-varying recharge, pumping and leakage. *J. Hydrol.* 324, 350–358. <https://doi.org/10.1016/j.jhydrol.2005.09.029>
- Rodriguez-Galiano, V., Mendes, M.P., Garcia-Soldado, M.J., Chica-Olmo, M., Ribeiro, L., 2014. Predictive modeling of groundwater nitrate pollution using Random Forest and multisource variables related to intrinsic and specific vulnerability: A case study in an agricultural setting (Southern Spain). *Sci. Total Environ.* 476–477, 189–206. <https://doi.org/10.1016/j.scitotenv.2014.01.001>

- Roe, B.P., Yang, H.-J., Zhu, J., Liu, Y., Stancu, I., McGregor, G., 2005. Boosted decision trees as an alternative to artificial neural networks for particle identification. *Nucl. Instrum. Methods Phys. Res. Sect. Accel. Spectrometers Detect. Assoc. Equip.* 543, 577–584. <https://doi.org/10.1016/j.nima.2004.12.018>
- Saglio, P.H., Raymond, P., Pradet, A., 1980. Metabolic Activity and Energy Charge of Excised Maize Root Tips under Anoxia: CONTROL BY SOLUBLE SUGARS. *Plant Physiol.* 66, 1053–1057. <https://doi.org/10.1104/pp.66.6.1053>
- Samson, M., Fortin, J., Pepin, S., Caron, J., 2016. Impact of potassium sulfate salinity on growth and development of cranberry plants subjected to overhead and subirrigation. *Can. J. Soil Sci.* 97, 20–30.
- Scornet, E., 2017. Tuning parameters in random forests. *ESAIM Proc. Surv.* 60, 144–162. <https://doi.org/10.1051/proc/201760144>
- Singh, K.P., Gupta, S., Mohan, D., 2014. Evaluating influences of seasonal variations and anthropogenic activities on alluvial groundwater hydrochemistry using ensemble learning approaches. *J. Hydrol.* 511, 254–266. <https://doi.org/10.1016/j.jhydrol.2014.01.004>
- Stafford, J.V., 2019. Precision agriculture'19. Wageningen Academic Publishers.
- Taormina, R., Chau, K., Sethi, R., 2012. Artificial neural network simulation of hourly groundwater levels in a coastal aquifer system of the Venice lagoon. *Eng. Appl. Artif. Intell.* 25, 1670–1676. <https://doi.org/10.1016/j.engappai.2012.02.009>
- Taylor, R.G., Scanlon, B., Döll, P., Rodell, M., van Beek, R., Wada, Y., Longuevergne, L., Leblanc, M., Famiglietti, J.S., Edmunds, M., Konikow, L., Green, T.R., Chen, J., Taniguchi, M., Bierkens, M.F.P., MacDonald, A., Fan, Y., Maxwell, R.M., Yeichieli, Y., Gurdak, J.J., Allen, D.M., Shamsudduha, M., Hiscock, K., Yeh, P.J.-F., Holman, I., Treidel, H., 2013. Ground water and climate change. *Nat. Clim. Change* 3, 322–329. <https://doi.org/10.1038/nclimate1744>
- Vanderleest, C.P., Bland, W.L., 2016. Evapotranspiration from cranberry compared with the equilibrium rate. *Can. J. Soil Sci.* 97, 5–10.
- Vanderleest, C.P.L., Caron, J., Bland, W.L., 2016. Water table level management as an irrigation strategy for cranberry (*Vaccinium macrocarpon* Aiton) 1. *Can. J. Soil Sci.* 1–9. <https://doi.org/10.1139/cjss-2016-0001>
- Viswanathan, M.N., 1983. The Rainfall/Water-Table Level Relationship of an Unconfined Aquifer. *Ground Water* 21, 49. <https://doi.org/10.1111/j.1745-6584.1983.tb00704.x>
- Vörösmarty, C.J., Green, P., Salisbury, J., Lammers, R.B., 2000. Global water resources: vulnerability from climate change and population growth. *science* 289, 284–288.
- Wang, X., Liu, T., Zheng, X., Peng, H., Xin, J., Zhang, B., 2018. Short-term prediction of groundwater level using improved random forest regression with a combination of random features. *Appl. Water Sci.* 8, 125. <https://doi.org/10.1007/s13201-018-0742-6>
- Wu, J., Zhang, R., Yang, J., 1996. Analysis of rainfall-recharge relationships. *J. Hydrol.* 177, 143–160. [https://doi.org/10.1016/0022-1694\(95\)02935-4](https://doi.org/10.1016/0022-1694(95)02935-4)

- Wu, W., Dandy, G.C., Maier, H.R., 2014. Protocol for developing ANN models and its application to the assessment of the quality of the ANN model development process in drinking water quality modelling. *Environ. Model. Softw.* 54, 108–127. <https://doi.org/10.1016/j.envsoft.2013.12.016>
- Yoon, H., Jun, S.-C., Hyun, Y., Bae, G.-O., Lee, K.-K., 2011. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *J. Hydrol.* 396, 128–138. <https://doi.org/10.1016/j.jhydrol.2010.11.002>