

Centre Eau Terre Environnement

ESTIMATION DE L'HUMIDITÉ DU SOL À PARTIR DES FLUX THERMIQUES EN UTILISANT DES MODÈLES D'APPRENTISSAGE PROFOND

Par

Mohamed Youssef Efrit

Mémoire présenté pour l'obtention du grade de
Maître ès Sciences (M.Sc.)
en sciences de l'eau

Jury d'évaluation

Président du jury et
Examineur interne

Karem Chokmani

Examineur externe

Wassim Bouachir
Université TÉLUQ

Examineur interne
Président du jury

Saeid Homayouni

Directeur de recherche

Karem Chokmani

REMERCIEMENTS

Je tiens à exprimer ma profonde gratitude au Dr. Karem Chokmani, qui a joué un rôle central dans mon parcours académique. En 2023, il m'a offert l'opportunité d'effectuer un stage Mitacs, expérience déterminante qui m'a donné envie de poursuivre mes recherches dans le cadre de ce mémoire. Son accompagnement constant, ses conseils avisés et sa disponibilité m'ont permis de progresser non seulement sur le plan scientifique, mais aussi sur le plan personnel et professionnel. J'aimerais également remercier le Dr. Saied Homayouni, dont l'aide précieuse m'a guidé dans l'élaboration du devis de recherche, et Haythem Zidi, pour son soutien et ses compétences techniques qui ont grandement contribué à l'analyse des données. Ma reconnaissance va aussi à Marie-Christine Marmette, pour ses retours constructifs et ses conseils qui ont enrichi la qualité de ce travail. Je remercie chaleureusement l'INRS-ETE, qui a mis à ma disposition les ressources nécessaires pour mener à bien ce projet dans un environnement propice à l'apprentissage et à l'innovation.

Enfin, je souhaite exprimer toute ma gratitude à mes parents, Faten Khamassi et Mohamed Lotfi Efrit, dont le soutien indéfectible a été essentiel pour la réussite de mes études au Canada. Leur encouragement constant, leurs sacrifices et leur confiance m'ont donné la force d'avancer malgré les défis rencontrés. Je n'oublie pas non plus mes amis et ma famille élargie, qui ont su m'apporter réconfort et motivation tout au long de ce parcours exigeant. À toutes ces personnes et institutions, je suis profondément reconnaissant, car sans elles, ce travail n'aurait jamais pu aboutir.

RÉSUMÉ

L'humidité du sol est un paramètre clé pour le système de production agricole au même niveau d'importance que la gestion des ressources en eau et les conditions climatiques. Bien que la mesure directe de ce paramètre clé soit précise, elle demeure coûteuse, localisée et invasive. Les méthodes indirectes offrent une alternative, mais elles présentent des limites de calibration et de généralisation selon les types de sols. Dans ce contexte, ce projet visait à développer et comparer des modèles d'apprentissage profond capables d'estimer l'humidité du sol à partir du flux thermique, en exploitant les propriétés thermiques et les dynamiques temporelles de la température du sol.

Les données expérimentales ont été collectées sur deux sites au Québec : le Parc technologique de l'INRS durant les étés 2024 et 2025, et le champ 41 sur la Route du Pont Rouge à Sainte-Catherine pour les années 2020 et 2021. Des capteurs Decagon 5TM, reliés à un enregistreur de données Em50G, ont été installés entre 0 et 60 cm de profondeur, couvrant trois horizons contrastés : organique (0–20 cm), sableux (20–40 cm) et argileux (40–55 cm). Un profileur de température CS320 SDI-12 a également été installé et mesurait les températures tous les 5 cm de profondeur. Les variables utilisées comprenaient la température de surface (T_{sup}), la température en profondeur (T_{inf}), le gradient de température (T_{flux}) et l'humidité du sol mesurée in situ pour la calibration.

La base de données a été structurée en séquences temporelles de 24 à 96 heures avec une fenêtre glissante, puis augmentée en variant les distances de mesure (5 à 30 cm), ce qui a porté le jeu de données d'environ 9 000 à 24 500 séquences. Les données ont été nettoyées, normalisées et divisées en ensembles d'entraînement, de validation et de test.

Deux principales architectures ont été évaluées : un réseau de neurones convolutif (CNN) pour capturer les relations spatiales au sein des séries de flux thermiques, et un réseau de neurones à mémoire à long et à court terme (LSTM) pour exploiter les dépendances temporelles. Les hyperparamètres ont été optimisés par validation croisée à l'aide d'outils tels que Keras Tuner. Les résultats montrent des performances élevées : le meilleur CNN (96 heures) a atteint un coefficient de détermination $R^2 = 0,96$ avec un faible biais et une erreur quadratique moyenne réduite, tandis que le meilleur LSTM (48 heures) a obtenu $R^2 = 0,95$. Bien que les deux modèles soient efficaces, les CNN se distinguent par une convergence plus rapide et une précision légèrement supérieure. Ces résultats confirment le potentiel de l'apprentissage profond pour

estimer l'humidité du sol dans des milieux hétérogènes à partir de données thermiques, réduisant la dépendance aux mesures invasives.

En conclusion, cette étude démontre qu'il est possible de prédire avec une grande précision l'humidité du sol en appliquant des modèles CNN et LSTM à des données thermiques issues de capteurs à faible coût. Il demeure nécessaire d'étendre cette approche à une plus large gamme de textures de sols et de contextes environnementaux afin d'assurer la robustesse et la transférabilité des modèles.

Mots-clés : Humidité du sol, Flux thermique, Apprentissage profond, Réseau de neurones convolutif (CNN), Mémoire à long court terme (LSTM), Séries temporelles, Calibration.

ABSTRACT

Soil moisture is a key parameter for agricultural production systems, on the same level of importance as water resource management and climatic conditions. Although direct measurement of this parameter is accurate, it remains costly, localized, and often destructive. Indirect methods provide an alternative, but they face calibration and generalization limitations depending on soil types. In this context, this project aimed to develop and compare deep learning models capable of estimating soil moisture from thermal flux, leveraging the thermal properties and temporal dynamics of soil temperature.

Experimental data were collected at two sites in Quebec: the INRS Technology Park during the summers 2024 and 2025, and Field 41 on Route du Pont Rouge in Sainte-Catherine for 2020 and 2021. Decagon 5TM sensors connected to an Em50G data logger were installed between 0 and 60 cm depth, covering three contrasting horizons: organic (0–20 cm), sandy (20–40 cm), and clayey (40–55/60 cm). A CS320 SDI-12 temperature profiler was also installed and measured temperatures every 5 cm of depth. The variables used included surface temperature (T_{sup}), deep temperature (T_{inf}), temperature gradient (T_{flux}), and in-situ measured soil moisture for calibration.

The database was structured into time sequences of 24 to 96 hours with a sliding window, then augmented by varying measurement distances (5 to 30 cm), increasing the dataset from ~9 000 to ~24 500 sequences. The data were cleaned, normalized, and split into training, validation, and test sets.

Two main architectures were evaluated: a Convolutional Neural Network to capture spatial relationships within thermal flux series, and a Long Short-Term Memory network to exploit temporal dependencies. Hyperparameters were optimized using cross-validation and tools such as Keras Tuner. The results show high performance: the best CNN (96 hours) achieved a coefficient of determination $R^2 = 0.96$ with low bias and mean squared error, while the best LSTM (48 hours) reached $R^2 = 0.95$. Although both models are effective, CNNs stood out for faster convergence and slightly higher accuracy. These results confirm the potential of deep learning to estimate soil moisture in heterogeneous soils using thermal data, thereby reducing reliance on invasive measurements.

In conclusion, this study demonstrates that soil moisture can be predicted with high precision using CNN and LSTM models applied to thermal data from low-cost sensors. Nevertheless, extending the approach to a wider range of soil textures and environmental contexts remains necessary to ensure model robustness and transferability.

Keywords: Soil moisture, Thermal flux, Deep learning, Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), Time series, Calibration.

TABLE DES MATIÈRES

REMERCIEMENTS	III
RÉSUMÉ.....	V
ABSTRACT	VII
TABLE DES MATIÈRES	IX
LISTE DES FIGURES	XI
LISTE DES TABLEAUX.....	XIII
LISTE DES ABRÉVIATIONS	XV
1 INTRODUCTION.....	1
1.1 MISE EN CONTEXTE	1
1.2 PROBLEMATIQUES	2
1.3 OBJECTIFS ET HYPOTHESES	3
2 ETAT DE L'ART	5
2.1 PROPRIETES HYDRIQUES DU SOL.....	5
2.1.1 Teneur en eau	5
2.1.2 Capacité au champ.....	5
2.1.3 Point de flétrissement permanent.....	6
2.1.4 Tension de l'eau dans le sol.....	6
2.1.5 Conductivité hydraulique.....	6
2.1.6 Porosité du sol.....	7
2.2 PROPRIETES THERMIQUES DU SOL.....	7
2.2.1 Conductivité thermique	7
2.2.2 Diffusivité thermique	8
2.2.3 Capacité thermique volumique.....	8
2.2.4 Flux de chaleur.....	9
2.2.5 Variation de la température dans le sol.....	10
2.2.6 Hypothèses de travail	11
2.3 METHODES DIRECTES DE MESURE DE LA TENEUR EN EAU	12
2.4 METHODES INDIRECTES DE MESURE DE LA TENEUR EN EAU	13
2.4.1 Modèles physiques.....	15
2.4.2 Modèles stochastiques	16
2.4.3 Intelligence artificielle.....	16
2.4.4 Apprentissage profond.....	16
3 METHODOLOGIE.....	21
3.1 ZONES D'ETUDES.....	21

3.2	OUTILS ET MATERIELS	22
3.3	CALIBRATION DES CAPTEURS FDR POUR LA MESURE DE LA TENEUR EN EAU	25
3.4	DISPOSITIF EXPERIMENTAL	26
3.5	CREATION DU MODELE ET ANALYSE DE DONNEES	30
3.5.1	<i>Structure des données</i>	30
3.5.2	<i>Métriques d'évaluation</i>	32
3.5.3	<i>Modèles d'apprentissage profond</i>	33
3.5.4	<i>Comparaison avec des modèles de base</i>	39
4	RESULTATS ET DISCUSSION	44
4.1	ANALYSE DES DONNEES	44
4.2	RESULTATS DES MODELES LSTM.....	45
4.3	RESULTATS DES MODELES CNN	46
4.4	COMPARAISON DES MODELES CNN ET LSTM.....	46
4.5	COURBES D'APPRENTISSAGE	47
4.6	ÉVALUATION DES PERFORMANCES DES MODELES	49
4.7	COMPARAISON AVEC LES MODELES DE BASE.....	51
4.8	DISCUSSION.....	51
5	CONCLUSION ET RECOMMANDATIONS	55
6	BIBLIOGRAPHIE.....	57
	ANNEXE A.....	62
	ANNEXE B.....	65

LISTE DES FIGURES

FIGURE 1: L'ETAT DE L'EAU DANS LE SOL (BEAUCHAMP, 2006).....	5
FIGURE 2: L'EAU DANS LE SOL SELON LA TEXTURE (BEAUCHAMP, 2006)	6
FIGURE 3: LA RELATION ENTRE LA CONDUCTIVITE THERMIQUE ET LA TENEUR VOLUMIQUE EN EAU PAR TYPE DE SOL (SEPASKHAH & MAZAHERI-TEHRANI, 2024).....	7
FIGURE 4: LA TEMPERATURE DU SOL EN FONCTION DE LA PROFONDEUR ET DE LA SAISON (CUNY, 2017).....	10
FIGURE 5: STRUCTURE D'UN CNN (BENCHOHRA & MAHMOUDI, 2020)	18
FIGURE 6: STRUCTURE D'UN LSTM (BOURDOIS, 2019)	20
FIGURE 7: ZONE D'ETUDES PARC TECHNOLOGIQUE INRS.....	21
FIGURE 8: ZONE D'ETUDES CHAMP 41	21
FIGURE 9: EMBLACEMENT DES DEUX ZONES D'ETUDES AU QUEBEC	22
FIGURE 10: CAPTEURS 5TM DECAGON.....	22
FIGURE 11: CS230 SDI-12 PROFILEUR DE TEMPERATURE	23
FIGURE 12: DATA LOGGER EM50G	24
FIGURE 13: INTERFACE DU LOGICIEL ECH2O	24
FIGURE 14: METHODE GRAVIMETRIQUE POUR LA CALIBRATION DES CAPTEURS	25
FIGURE 15: COURBE DE CALIBRATION	26
FIGURE 16: PHOTO REELLE DE L'EMPLACEMENT DES CAPTEURS DANS LE SITE CHAMP 41	27
FIGURE 17: PHOTO REELLE DES HORIZONS ET L'EMPLACEMENT DES CAPTEURS DANS LE SITE DU PARC TECHNOLOGIQUE	27
FIGURE 18: DISPOSITIF EXPERIMENTAL FINAL DU SITE PARC TECHNOLOGIQUE	28
FIGURE 19: DISPOSITIF EXPERIMENTAL FINAL DU SITE CHAMP 41	28
FIGURE 20: DIAGRAMME DE L'EMPLACEMENT DES CAPTEURS ET LES HORIZONS.....	29
FIGURE 21: TYPES DES SOLS	30
FIGURE 22: EXEMPLE DE DONNEES.....	31
FIGURE 23: ARCHITECTURE CNN UTILISEE.....	34
FIGURE 24: ARCHITECTURE LSTM UTILISEE	34
FIGURE 25: TECHNIQUE D'AUGMENTATION DE DONNEES DANS LE SITE 2	35

FIGURE 26: TECHNIQUE D'AUGMENTATION DE DONNEES DANS LE SITE 1	36
FIGURE 27: APPROCHE UTILISEE POUR ESTIMER L'HUMIDITE EN UTILISANT DES MODELES AP ET DES MODELES PEU PROFONDS.....	39
FIGURE 28: ARCHITECTURE DE XGBOOST (CHEN & GUESTRIN, 2016)	40
FIGURE 29: EXEMPLE DE REGRESSION LINEAIRE SIMPLE	41
FIGURE 30: EXEMPLE DE REGRESSION LINEAIRE MULTIVARIEE	42
FIGURE 31: EXEMPLE D'UN MODELE DE PERSISTANCE.....	43
FIGURE 32: ANALYSE DE L'HORIZON 0-20CM	44
FIGURE 33: ANALYSE DE L'HORIZON 20-40CM	44
FIGURE 34: ANALYSE DE L'HORIZON 40-55CM	45
FIGURE 35: COURBE DE VALIDATION CNN96.....	47
FIGURE 36: COURBE DE VALIDATION LSTM48	48
FIGURE 37: COMPARAISON DES COEFFICIENTS DE DETERMINATION DES DIFFERENTES VERSIONS DES MODELES D'AP ETUDIES	49
FIGURE 38: COMPARAISON DES RMSE DES DIFFERENTES VERSIONS DES MODELES D'AP ETUDIES	49
FIGURE 39: COMPARAISON DES BIAIS DES DIFFERENTES VERSIONS DES MODELES D'AP ETUDIES	50
FIGURE 40: COMPARAISON DES COEFFICIENTS DE DETERMINATION DES MEILLEURES VERSIONS DES MODELES D'AP ETUDIES ET LES MODELES DE BASES	51

LISTE DES TABLEAUX

TABLEAU 1: TABLEAU COMPARATIF DES METHODES DE MESURES DIRECTES DE LA TENEUR EN EAU DU SOL	13
TABLEAU 2: TABLEAU COMPARATIF DES METHODES DE MESURES DIRECTES DE LA TENEUR EN EAU DU SOL	14
TABLEAU 3: LES LIBRAIRIES DU ML ET LEURS UTILITES	18
TABLEAU 4: HYPERPARAMETRES OPTIMISES DU MODELE LSTM.....	38
TABLEAU 5: HYPERPARAMETRES OPTIMISES DU MODELE CNN	38
TABLEAU 6: TABLEAU COMPARATIF DES RESULTATS DES MODELES LSTM	45
TABLEAU 7: TABLEAU COMPARATIF DES RESULTATS DES MODELES CNN	46

LISTE DES ABRÉVIATIONS

CNN	Convolutional Neural Network
RNN	Recurrent Neural Network
FCN	Fully Connected Network
DNN	Deep Neural Network
AI	Artificial Intelligence
DL	Deep Learning
GAN	Generative Adversarial Network
GRU	Gated Recurrent Unit
LSTM	Long Short-Term Memory
R ²	Coefficient de détermination
RMSE	Root Mean Squared Error
MSE	Mean Squared Error
MAE	Mean Absolute Error
BERT	Bidirectional Encoder Representations from Transformers
θ_v	Teneur en eau
TDR	Time Domain Reflectometry
FDR	Frequency Domain Reflectometry
κ	Diffusivité thermique
λ	Conductivité thermique
ρ	Masse Volumique
c_p	Capacité thermique Massique
q	Flux de chaleur
ET	évapotranspiration

1 INTRODUCTION

1.1 Mise en contexte

L'humidité du sol est un paramètre crucial dans de nombreux domaines, allant de l'activité agricole au domaine de la gestion des ressources en eau, en passant par le domaine de la météorologie et la prévention des catastrophes naturelles. La mesure précise de l'humidité du sol permet de mieux comprendre et gérer les cycles de l'eau, la productivité des cultures, et la santé des écosystèmes. Dans le secteur agricole la connaissance de l'humidité du sol est essentielle pour optimiser l'irrigation. Une irrigation bien gérée réduit le gaspillage d'eau, améliore le rendement des cultures et diminue les coûts énergétiques liés au pompage de l'eau. Elle influence donc directement les décisions liées à l'irrigation, la gestion des rendements agricoles, la conservation des sols, la planification agricole et la détection des tendances climatiques (Senanayake, Yeo, Willgoose, & Hancock, 2021). En particulier, la mesure de l'humidité du sol est essentielle pour l'optimisation de l'irrigation, permettant d'économiser l'eau tout en maximisant les rendements des cultures (Paolini, et al., 2023). Dans la gestion des ressources en eau, l'humidité du sol joue un rôle clé en influençant l'infiltration et le ruissellement de l'eau de pluie. Une estimation précise permet de mieux prévoir les réserves d'eau disponibles et de prendre des décisions éclairées pour l'allocation de cette ressource (Petit, 2003). En météorologie, les données sur l'humidité du sol contribuent à améliorer les prévisions climatiques et les modèles hydrologiques, car elles affectent l'évapotranspiration et le flux de chaleur à la surface terrestre.

Les types de sol varient considérablement en termes de composition, structure et capacité de rétention d'eau. Les sols sableux ont une capacité de rétention d'eau faible et un drainage rapide, tandis que les sols argileux retiennent l'eau plus efficacement mais peuvent souffrir de problèmes de drainage. Les sols limoneux combinent les avantages des sols sableux et argileux, offrant une bonne rétention d'eau et un drainage adéquat.

L'estimation de l'humidité du sol, quel que soit le type de sol, est donc essentielle pour plusieurs raisons. Utiliser des techniques avancées comme l'apprentissage profond pour estimer l'humidité du sol à partir des flux thermiques peut considérablement augmenter la précision et diminuer le temps nécessaire pour estimer l'humidité pour chaque type de sol spécifique.

1.2 Problématiques

La problématique de ce projet de recherche est complexe, impliquant plusieurs défis techniques et méthodologiques. Tout d'abord, il est crucial de déterminer la méthode de mesure de l'humidité du sol la plus adaptée pour notre contexte spécifique. Parmi les nombreuses techniques disponibles, chacune présente ses avantages et ses limitations. (Sharma, Srivastava, & Kumar, 2018)

La physique du sol est un domaine qui se concentre sur les propriétés physiques et les processus au sein des sols. L'eau présente dans le sol est influencée par de nombreux facteurs tels que le drainage, le ruissellement, le rayonnement et l'évapotranspiration. Le drainage est le processus par lequel l'excès d'eau est évacué du sol, ce qui est essentiel pour prévenir l'engorgement et maintenir un équilibre optimal entre l'eau et l'air dans le sol. Le ruissellement est la portion d'eau provenant des précipitations ou de l'irrigation qui s'écoule à la surface du sol. Le rayonnement affecte la température du sol et l'évaporation de l'humidité, tandis que l'évapotranspiration englobe à la fois l'évaporation de l'eau du sol et la transpiration des plantes. Il est donc essentiel de mesurer avec précision la teneur en eau du sol. Il existe deux principales catégories de méthodes pour mesurer la teneur en eau du sol : les méthodes directes et les méthodes indirectes.

. Les méthodes directes, comme la méthode gravimétrique, consistent à déterminer la masse d'eau contenue dans un échantillon de sol. Bien que précises, ces techniques sont généralement laborieuses, destructives, non continues et exigent des ressources matérielles et humaines importantes.

Les méthodes indirectes reposent sur l'exploitation de propriétés physiques ou empiriques pour déduire la teneur en eau. Les approches physiques déterministes, comme la diffusion de neutrons, l'atténuation gamma ou la réflectométrie (TDR, FDR), reposent sur des principes physiques bien définis. Ces techniques ont l'avantage d'être continues, non destructives et automatisées. Cependant, elles requièrent des calibrations spécifiques et peuvent être sensibles aux conditions environnementales. Les approches empiriques non déterministes, telles que les tensiomètres ou les capteurs capacitifs, s'appuient sur des relations issues de données expérimentales. Bien qu'elles soient parfois plus simples à mettre en œuvre, leur précision dépend de la qualité des calibrations.

Ces méthodes physiques et empiriques, bien qu'efficaces, présentent des limites dans les études complexes sur le sol, notamment en ce qui concerne l'intégration de multiples variables

environnementales et les variations spatiales et temporelles. Les modèles basés sur l'apprentissage machine constituent une alternative prometteuse. Les modèles plus classiques, comme les régressions linéaires ou les modèles non linéaires basés sur des algorithmes d'apprentissage automatique, permettent d'exploiter des jeux de données importants tout en évitant certains écueils des calibrations physiques complexes. Cependant, leur capacité à capturer des relations hautement non linéaires reste limitée.

Pour surmonter ces limites, les modèles d'apprentissage profond, tels que les réseaux de neurones convolutifs (CNN) et les réseaux de neurones à mémoire à long terme (LSTM), sont de plus en plus utilisés. Conçus pour analyser des données complexes comme les flux thermiques, ces modèles permettent de capturer les interactions non linéaires et les relations complexes entre les variables que d'autres approches ne peuvent pas appréhender. Des travaux récents ont mis en évidence la performance remarquable des réseaux de neurones profonds pour la modélisation des paramètres thermiques et hydriques du sol. Une étude en particulier a montré que les réseaux convolutionnels unidimensionnels surpassent les approches classiques, comme les perceptrons multicouches, dans la prédiction de la température du sol en conditions météorologiques variées (Farhangmehr, et al., 2023). Cette supériorité est attribuée à la capacité des CNN à extraire efficacement des motifs temporels locaux à partir de signaux continus tels que les variations de température. D'autres travaux ont également démontré l'intérêt des LSTM pour la prédiction de l'humidité du sol à partir de données de température, humidité relative et précipitations (Li, et al., 2022). Cependant, le développement de ces modèles profonds pose des défis importants, comme l'exigence en volumes de données, la généralisation à de nouveaux ensembles de données et l'adaptation à la diversité des caractéristiques physiques des sols. Cette variabilité des sols influence les mesures d'humidité et la réponse des capteurs, ce qui complique la tâche de généraliser le modèle pour qu'il puisse être utilisé pour tous les types de sols. Il est essentiel de choisir et de former une architecture de réseau appropriée, capable de s'ajuster à cette variabilité et de fournir des estimations fiables et précises dans une variété étendue de conditions de sol.

1.3 Objectifs et hypothèses

Le présent projet de de recherche a pour objectif de développer un modèle d'apprentissage profond capable d'estimer à haute précision l'humidité du sol à partir du flux thermique du sol. Ce projet repose sur les hypothèses suivantes :

- Sur un horizon de quelques mois, les phases solides du sol (matières minérale et organique) sont invariantes. Par conséquent, la seule composante du sol variante est la phase liquide (eau) ;
- Sous condition d'un flux thermique constant, la variation des propriétés thermiques du sol ne dépend que de la variation de l'humidité du sol.

2 ETAT DE L'ART

2.1 Propriétés hydriques du sol

2.1.1 Teneur en eau

La teneur en eau est la quantité d'eau présente dans le sol, généralement exprimée en pourcentage du volume ou de la masse de sol. L'équation de la teneur en eau volumique est la suivante :

$$\theta_v = \frac{V_l}{V_t} = \frac{\text{Volume de liquide (eau)}}{\text{Volume de sol}} \quad (1)$$

2.1.2 Capacité au champ

C'est la quantité maximale d'eau que le sol peut retenir contre la force de gravité après un drainage libre et complet. C'est une mesure importante pour déterminer la disponibilité en eau pour les plantes.

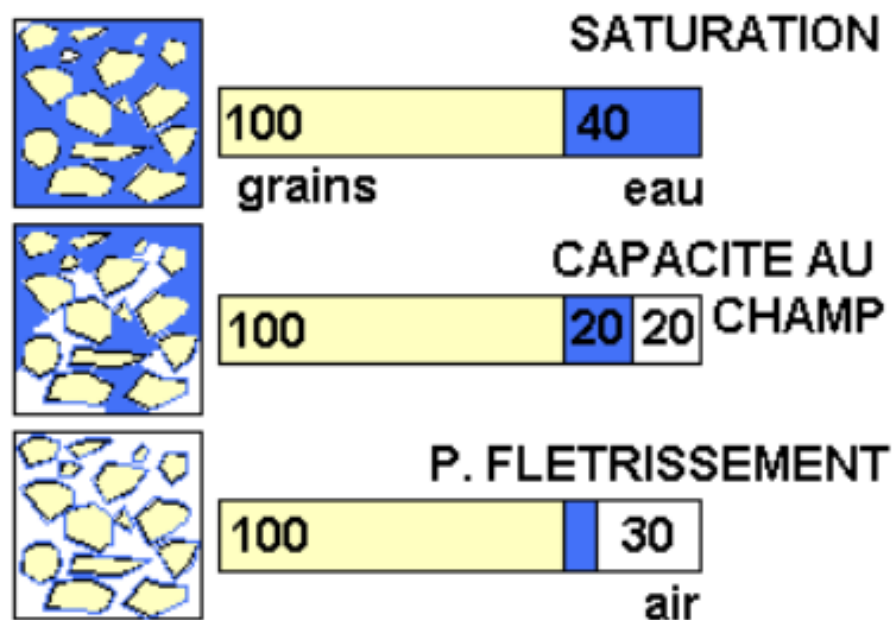


Figure 1: L'Etat de l'eau dans le sol (Beauchamp, 2006)

2.1.3 Point de flétrissement permanent

C'est la teneur en eau à laquelle les plantes ne peuvent plus extraire l'eau du sol, conduisant à un flétrissement irréversible.

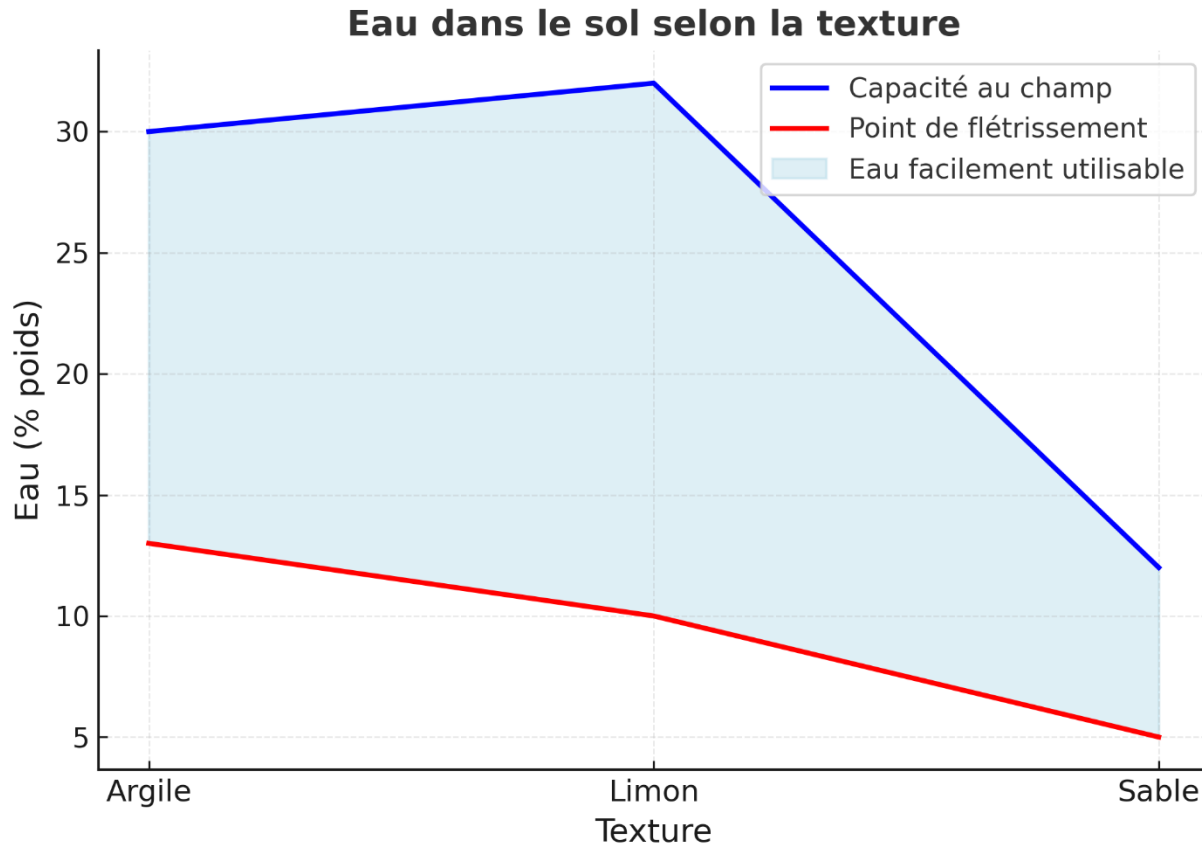


Figure 2: l'eau dans le sol selon la texture (Beauchamp, 2006)

On peut observer dans la figure 2 la capacité de champ et le point de flétrissement de différents types de sol.

2.1.4 Tension de l'eau dans le sol

La tension de l'eau dans le sol décrit la force avec laquelle l'eau est retenue dans le sol. Elle est mesurée en kilopascals (kPa) ou en bars. Des instruments comme les tensiomètres peuvent être utilisés pour mesurer cette tension.

2.1.5 Conductivité hydraulique

La conductivité hydraulique est la capacité du sol à transmettre l'eau. Elle dépend de la texture, de la structure et de la teneur en eau du sol. Elle est généralement exprimée en cm/s ou en mm/h.

2.1.6 Porosité du sol

La porosité du sol est le volume des pores par rapport au volume total du sol. Elle influence la quantité d'eau que le sol peut retenir et la vitesse à laquelle l'eau peut se déplacer à travers le sol. La distribution des tailles des pores dans le sol, affecte aussi la rétention et le mouvement de l'eau. Les pores plus grands drainent rapidement et retiennent moins d'eau, tandis que les pores plus petits retiennent plus d'eau mais drainent plus lentement (Brown, Biswas, Caron, Dyck, & Si, 2021).

$$f = \frac{V_f}{V_t} = \frac{V_l + V_g}{V_t} = \frac{\text{Volume de vide (liquide + gaz)}}{\text{Volume de sol}} = \frac{V_t - V_s}{V_t} \quad (2)$$

2.2 Propriétés thermiques du sol

2.2.1 Conductivité thermique

La conductivité thermique du sol mesure la capacité du sol à conduire la chaleur. Elle dépend de la composition du sol, l'humidité, de la densité et de la texture. Les sols avec une teneur élevée en eau ont généralement une conductivité thermique plus élevée que les sols secs ou sableux. (Boukhattem, 2008). La figure 3, montre l'évolution de la conductivité thermique en fonction de la teneur en eau volumique pour différents types de sols. On observe que la conductivité thermique augmente avec l'humidité pour tous les sols. Les sols sableux et sable-limoneux ont les valeurs de λ les plus élevées et les sols limoneux et argileux ont les valeurs plus faibles.

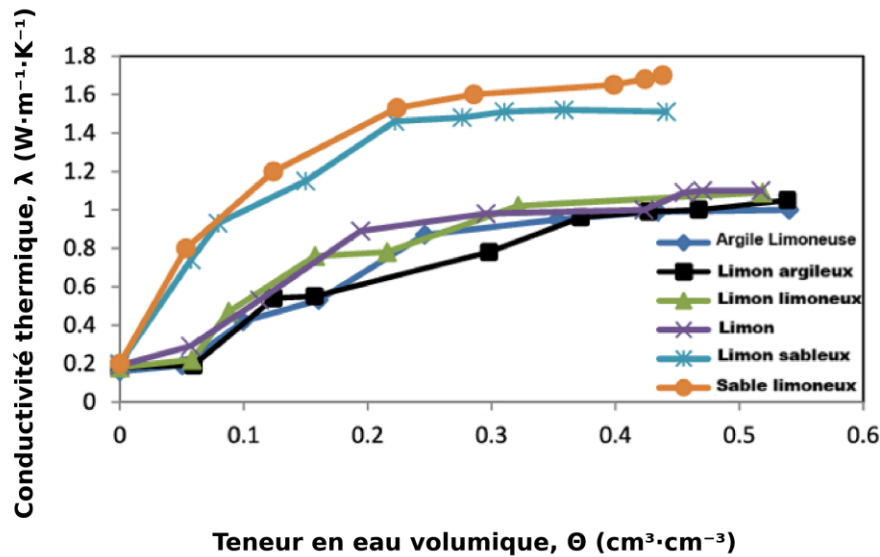


Figure 3: La relation entre la conductivité thermique et la teneur volumique en eau par type de sol (Sepaskhah & Mazaheri-Tehrani, 2024)

2.2.2 Diffusivité thermique

La diffusivité thermique est la capacité du sol à diffuser la chaleur à travers sa masse. Elle est définie comme le rapport entre la conductivité thermique et la capacité thermique volumique du sol. Les sols avec une haute diffusivité thermique réagissent rapidement aux variations de température de surface.

$$\kappa = \frac{\lambda}{\rho \cdot c_p} \quad (3)$$

λ : Conductivité thermique ($\text{W m}^{-1} \text{K}^{-1}$)

ρ : Masse volumique (kg/m^3)

c_p : Capacité thermique massique ($\text{J kg}^{-1} \text{K}^{-1}$)

2.2.3 Capacité thermique volumique

La capacité thermique volumique du sol est la quantité de chaleur nécessaire pour augmenter la température d'une unité de volume de sol d'un degré Celsius. Elle est influencée par la teneur en eau, la densité et la composition minérale du sol. La capacité thermique volumique augmente avec l'humidité. L'eau possède une capacité thermique massique élevée à $\sim 4,18 \text{ kJ/kg}\cdot\text{K}$ par rapport aux solides minéraux à $\sim 0,8\text{--}1,2 \text{ kJ/kg}\cdot\text{K}$ et à l'air. Lorsque l'humidité augmente, la proportion d'eau dans le volume du sol s'accroît, augmentant la quantité de chaleur nécessaire pour élever la température du sol :

Sol sec : La capacité thermique volumique (C_v) est dominée par la phase minérale.

Sol humide : C_v devient nettement plus élevée grâce à l'apport de la phase aqueuse.

Selon Tarara et al. (1997), un capteur de chaleur à double sonde permet de mesurer la capacité thermique volumique d'un sol, en utilisant une impulsion de chaleur appliquée à une distance fixe. Cela est aussi applicable sans impulsion de chaleur puisque la terre émet de la chaleur et en reçoit aussi du soleil. La capacité thermique volumique est calculée par la formule :

$$c_{pc} = \frac{q}{\pi \cdot e \cdot r^2 \cdot \Delta T_m} \quad (4)$$

q : flux de chaleur

r : la distance entre deux capteurs

ΔT_m : la température maximale observée

e : la base des logarithmes naturels

La relation entre c_{pc} et les fractions volumiques minérale X_m , organique X_o et en eau du sol θ est donnée par :

$$c_{pc} = 1,92X_m + 2,50X_o + 4,18\theta \quad (5)$$

2.2.4 Flux de chaleur

Le flux de chaleur est une mesure du transfert d'énergie thermique à travers une surface en raison d'une différence de température. Ce transfert conduit à l'équilibre thermique entre les substances concernées. Le flux de chaleur est exprimé en watts (W) ou en joules par seconde (J/s). Il existe trois modes de transfert de chaleur :

- Par conduction : Ce mode de transfert se produit à l'intérieur des solides ou entre des solides en contact direct. Les molécules plus chaudes transfèrent leur énergie aux molécules plus froides par collision. La loi de Fourier décrit ce phénomène.

$$q = -k \frac{dT}{dx} \quad (6)$$

q : Flux de chaleur (W/m²)

$\frac{dT}{dx}$: Gradient de température en une dimension (K/m)

k : conductivité thermique du matériau (W/m·K)

- Par convection : La convection est le transfert de chaleur entre une surface solide et un fluide en mouvement. La loi de Newton de refroidissement décrit le flux thermique par convection.

$$q = h(T_s - T_f) \quad (7)$$

q : Flux de chaleur (W/m²)

h : Coefficient de transfert de chaleur par convection (W/m²·K)

T_s : Température de la surface (K)

T_f : Température du fluide (K)

- Par radiation : Ce mode de transfert se produit lors du transfert de chaleur par émissions d'ondes électromagnétiques. La loi de Stefan-Boltzmann décrit le flux thermique par rayonnement.

$$q = \sigma T^4 \quad (8)$$

q : Flux de chaleur (W/m²)

σ : Constante de Stefan-Boltzmann (5.67×10^{-8} W/m²·K⁴)

T : Température absolue de la surface (K)

Le flux thermique dans le sol est important pour de nombreux processus physiques. Il régule la température du sol et influence la croissance des plantes, l'activité microbienne et les cycles biogéochimiques (Brown, Biswas, Caron, Dyck, & Si, 2021). Le flux joue également un rôle clé dans la production d'énergie géothermique et la détermination du microclimat local (Wang & Pang, 2021). La compréhension et la gestion du flux thermique dans le sol sont essentielles pour l'agriculture, l'écologie, la géologie et même certains sports, soulignant son impact sur notre environnement et nos activités (Castellvi & Gonzalez Dugo, 2021).

2.2.5 Variation de la température dans le sol

La figure 4 représente la température du sol en fonction de la profondeur et du mois, on note une variation tout au long de l'année seulement dans les 15 premiers mètres de la surface. La température après 15m demeure constante a 10 C.

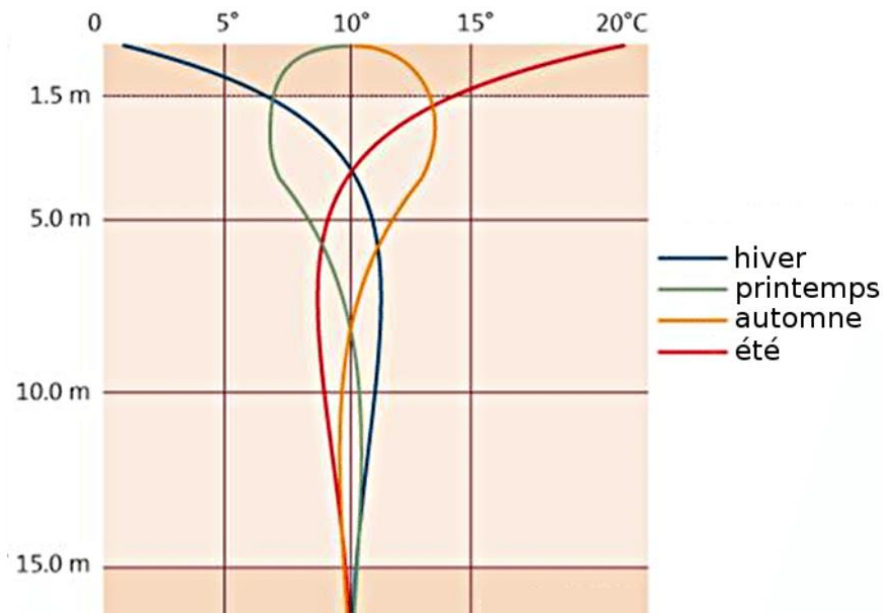


Figure 4: La température du sol en fonction de la profondeur et de la saison (Cuny, 2017)

Cette variation est causée par plusieurs facteurs, les conditions météorologiques particulièrement jouent un rôle direct sur la variation de la température du sol (Oliver, Oliver, Wallace, & Roberts,

1987). La température de l'air à la surface influence aussi directement la température du sol. Pendant la journée, le sol absorbe la chaleur de l'air chaud, augmentant sa température. La nuit, le sol perd de la chaleur vers l'air plus frais, ce qui entraîne une baisse de sa température. Le rayonnement solaire est une source majeure de chaleur pour le sol. L'ensoleillement direct chauffe la surface du sol, et cette chaleur est progressivement transférée vers les couches plus profondes par conduction. Les précipitations influencent la température du sol principalement par leur effet sur l'humidité du sol et l'évapotranspiration. Les sols humides, résultant de fortes précipitations, se réchauffent plus lentement en raison de l'évaporation de l'eau qui consomme de la chaleur (Patil, Laegdsmand, Olesen, & Porter, 2010). Cependant, en période de sécheresse, le sol sec se réchauffe plus rapidement car il y a moins d'humidité pour modérer la température. Les vents peuvent également affecter la température du sol en facilitant le transfert de chaleur entre l'air et le sol. Les vents forts augmentent l'échange de chaleur, ce qui peut refroidir ou réchauffer le sol en fonction de la température du vent par rapport à celle du sol. De plus, les vents peuvent augmenter le taux d'évaporation de l'eau de surface, influençant l'humidité du sol et sa température (Sauer & Horton, 2005). La présence de végétation modifie le microclimat du sol. Les plantes offrent une ombre qui réduit l'ensoleillement direct sur le sol, ce qui diminue le réchauffement diurne. Par ailleurs, l'évapotranspiration augmentée par la végétation a un effet de refroidissement, comme mentionné précédemment (Ni, Cheng, Wang, Ng, & Garg, 2019). Les propriétés physiques et thermiques du sol, comme la texture, la composition minérale, la densité et la capacité thermique, influencent également la manière dont le sol absorbe et stocke la chaleur. Par exemple, les sols sableux se réchauffent et se refroidissent plus rapidement que les sols argileux en raison de leurs différentes capacités de rétention d'eau et conductivités thermiques.

2.2.6 Hypothèses de travail

On peut donc obtenir les deux équations suivantes :

$$\theta = \frac{q}{\pi e r^2 \Delta T_m} - \frac{(1,92 X_m + 2,50 X_o)}{4,18} \quad (9)$$

$$\Delta T_m = \left(\frac{\pi \cdot e \cdot r^2}{q \cdot (1,92 X_m + 2,50 X_o + 4,18 \theta)} \right)^{-1} \quad (10)$$

Selon nos hypothèses, on obtient une nouvelle équation :

$$\Delta T_m = \frac{a}{\theta + b} \quad (11)$$

$$a = \frac{\pi \cdot e \cdot r^2}{4,18q}$$

$$b = \frac{1,92X_m + 2,50X_o}{4,18}$$

Par conséquent, la différence de température entre les deux faces d'une couche de sol assez homogène ne varie dans le temps qu'en fonction de la teneur en eau. Au cours de cette période temps, les propriétés physiques du sol restent constantes.

Maintenant qu'on a démontré que la différence en température dépend de la teneur en eau, on peut étudier cette relation avec des données et un modèle d'apprentissage profond, dans le but de connaître la quantité d'eau dans le sol en mm.

$$\text{Quantite d'eau (mm)} = \theta \times \text{Profondeur du sol (mm)} \quad (12)$$

2.3 Méthodes directes de mesure de la teneur en eau

Les méthodes directes reposent sur des mesures physiques réalisées à partir d'échantillons de sol pour déterminer la quantité d'eau dans le sol. Selon (Rasheed, et al., 2022), la méthode gravimétrique constitue la technique de référence. Cette méthode est très précise mais destructive et limitée à des mesures ponctuelles, ce qui restreint son utilisation pour le suivi continu de l'humidité. Les autres approches directes reposent sur des principes similaires de mesure physique de l'eau du sol et nécessitent aussi une instrumentation et un temps d'analyse importants (Rasheed, et al., 2022).

Tableau 1: Tableau comparatif des méthodes de mesures directes de la teneur en eau du sol

Méthode	Avantages	Inconvénients	Temps de Réponse	Automatique	Références
Méthode gravimétrique	Très précise, méthode de référence, simple à mettre en œuvre	Laborieuse, destructive, non continue	Long	Non	(Topp & Ferre, 2002) ; (Hillel, 2004)
Méthode volumétrique (anneau de prélèvement)	Donne directement la teneur volumique, simple	Erreurs de prélèvement possibles, destructive	Moyen	Non	(Hillel, 2004); (Evelt & Tolk, 2009)
Extraction sous vide / membrane de Richards	Permet de déterminer la courbe de rétention d'eau du sol	Longue, instrumentation lourde	Très long	Non	(Richards, 1948); (Dane & Hopmans, 2002)
Méthode chimique (Karl Fischer)	Très précise, applicable aux sols très fins	Coûteuse, nécessite solvants et réactifs	Moyen	Partiellement	(Al-Mahmoud, 2019)
Méthode thermique	Permet l'analyse énergétique et la calibration de capteurs	Complexe, non portable	Long	Non	(Campbell & Norman, 2019); (Oschner, Horton, & Ren, 2013)

2.4 Méthodes indirectes de mesure de la teneur en eau

Les méthodes indirectes mesurent des propriétés physiques du sol corrélées à la teneur en eau plutôt que la teneur en eau elle-même. Selon (Hunduma & Kebede Bunare, 2020), ces techniques comprennent les approches électromagnétiques, électriques, nucléaires et optiques. Les méthodes à réflectométrie temporelle (TDR) et fréquentielle (FDR) estiment la teneur en eau à partir de la constante diélectrique du sol, tandis que les sondes neutroniques ou gamma mesurent la diffusion de particules ou de rayonnements sensibles à la présence d'hydrogène. Ces approches présentent l'avantage d'être non destructives, rapides et automatisables, permettant ainsi un suivi spatial et temporel continu de l'humidité du sol. Toutefois, leur précision dépend

d'une calibration préalable propre à chaque type de sol, et elles peuvent être influencées par la texture, la salinité ou la température du milieu (Hunduma & Kebede Bunare, 2020).

Tableau 2: Tableau comparatif des méthodes de mesures directes de la teneur en eau du sol

Méthode	Propriétés Mesurées	Avantages	Inconvénients	Temps de Réponse	Auto	Références
Diffusion de neutrons	La diffusion de neutrons rapides	Très précise, facile à mettre en œuvre	Sources radioactives, faible sensibilité en surface	2 minutes	Oui	(Greacen, 1981)
Atténuation gamma	Contenu volumétrique en eau	Non destructive	Profondeur limitée (2,5 cm), faible précision	Instantané	Oui	Udagani, 2013
TDR	Contenu volumétrique en eau	Non destructive, mesures continues	Influencée par la salinité et calibration spécifique au sol	Instantané	Oui	Benson, 1999
FDR	Contenu volumétrique en eau	Non destructive, précise	Sensible aux changements de température et d'humidité	Instantané	Oui	Munoz-Carperna, et al., 2004
Tensiomètres	Potentiel matriciel	Non destructif, grande précision	Limité dans les sols très secs/humides, réponse lente	3 heures	Non	Barrett, et al., 2012
Technique hygrométrique	Potentiel hydrique	Mesures automatisées	Précision qui diminue avec le temps	3 minutes	Oui	Petropoulos, et al., 2015
Méthode du flux de chaleur	Conductivité thermique	Potentiel pour une estimation indirecte de l'eau du sol	Calibration précise requise, non automatisable	N/A	Non	(Tarara & Ham, 1997)

2.4.1 Modèles physiques

Les modèles physiques créés pour évaluer l'humidité du sol ont pour objectif de représenter les mécanismes naturels qui régulent le flux d'eau à travers le sol. Un modèle bien connu repose sur l'équation de Richards, qui associe la loi de Darcy, décrivant le mouvement de l'eau dans des sols poreux, et le principe de conservation de la masse énoncé par Richards en 1931. Cette équation facilite la simulation des fluctuations d'humidité tant à travers le temps qu'à différentes profondeurs.

COMSOL Multiphysics est un logiciel qui intègre cette équation pour modéliser la répartition de l'humidité dans le sol en prenant en compte des éléments comme les précipitations, l'évaporation et les propriétés du sol. (Multiphysics®, 2023) Bien que ces modèles affichent une très bonne précision, leur efficacité dépend de la disponibilité des données et du bon calibrage des paramètres. (Šimůnek & Van Genuchten, 2016)

D'autre part, les modèles de bilan hydrique utilisent aussi la loi de Darcy mais l'appliquent plus généralement. Ils estiment la teneur en eau moyenne du sol en comparant les apports comme la pluie et l'irrigation avec les pertes tels que l'évaporation, le ruissellement et le drainage. (Thornthwaite & Mather, 1955) Ces modèles sont très simples à utiliser et appliquer et permettent d'estimer la teneur en eau du sol sur une grande échelle. Le modèle SNOSWAB, par exemple, est un modèle physique déterministe conçu pour simuler la dynamique de la teneur en eau du sol et le bilan hydrique. Les précipitations et l'eau de fonte sont réparties entre l'infiltration, le ruissellement, l'évapotranspiration et la percolation en fonction des propriétés du sol et des variables météorologiques telles que la température de l'air, les précipitations et l'évapotranspiration. Cette structure permet d'estimer la teneur en eau du sol en suivant le mouvement vertical et le stockage de l'eau dans des couches de sol homogènes (Danielscu, 2025). Ce modèle est très simple à utiliser et possède une grande capacité d'adaptation à différents contextes climatiques et types de sols, mais il repose sur l'hypothèse d'homogénéité verticale et horizontale, ce qui limite sa précision dans les sols hétérogènes ou stratifiés. De plus, la performance du modèle dépend fortement de la qualité et de la précision des données d'entrée et de calibration. Ces modèles ont une grande sensibilité aux erreurs dans l'estimation des apports et des pertes, et ne prennent pas en compte les différentes textures du sol. Une erreur de mesure d'un élément dans le bilan hydrique suffit pour se propager et donner de mauvais résultats.

2.4.2 Modèles stochastiques

Les modèles stochastiques ont plusieurs avantages comme leur simplicité, la facilité d'interprétation et leurs faibles besoins en calcul. Ils reposent sur des relations mathématiques ou physiques entre les variables, ce qui les rend plus transparents et plus faciles à valider expérimentalement (Hillel, 2004). Ces approches sont largement utilisées pour décrire les processus de transfert d'eau et de chaleur dans le sol à travers des formulations analytiques dérivées de l'équation de Richards ou du bilan hydrique (Vereecken et al., 2016). Ces modèles utilisent des hypothèses simplificatrices comme l'homogénéité du sol, le flux unidimensionnel, les conditions stables aux limites, qui ne reflètent pas la complexité des milieux naturels (Pan, 2015). Leur performance diminue face à la variabilité spatiale, temporelle et non linéaire des paramètres hydriques. De plus, leur capacité à exploiter de grandes bases de données multidimensionnelles reste limitée comparativement aux approches d'apprentissage automatique et profond, qui peuvent mieux capter les interactions complexes entre variables (Moazen-zadeh et al., 2022). Dans le cadre de ce projet, où les relations entre flux thermiques et humidité du sol sont fortement non linéaires et partiellement inconnues, l'utilisation de modèles analytiques sera certainement inutile, d'où l'intérêt d'approches basées sur l'apprentissage machine.

2.4.3 Intelligence artificielle

L'objectif principal de l'IA dans le cadre de cette recherche est d'intégrer des scripts et des modules sur des outils et des machines qui leur permettraient de réfléchir et de résoudre des problèmes en prenant des décisions (Liew, Krishnaswamy, Cheng, & Tan, 2019). L'apprentissage automatique est un sous-ensemble de l'IA, dans lequel l'utilisation de données est nécessaire pour entraîner le modèle à résoudre des problèmes. Dans l'ensemble, l'IA est considérée comme le domaine primordial, tandis que le ML et le DL sont des outils permettant d'exploiter correctement l'utilisation de l'intelligence artificielle (Pedersen, Verspoor, Jenkinson, & Law, 2020).

2.4.4 Apprentissage profond

L'apprentissage profond est un sous-domaine de l'apprentissage automatique dans lequel le modèle a la capacité d'apprendre et de représenter automatiquement des modèles complexes à partir de grandes quantités de données. Fondamentalement, le DL utilise des réseaux de neurones artificiels inspirés des neurones interconnectés du cerveau humain. Les caractéristiques uniques propres au DL sont la présence de plusieurs couches au sein de ces

réseaux, leur permettant d'extraire automatiquement des fonctionnalités des ensembles de données. Les modèles DL excellent également dans l'apprentissage des fonctionnalités et réduisent la nécessité d'un développement manuel des fonctionnalités (Jakhan & Kaur, 2020). Ils prospèrent sur de grands ensembles de données et utilisent des techniques telles que la rétropropagation pour tirer les leçons des erreurs rencontrées entre les valeurs prédites et réelles (Goodfellow, Bengio, & Courville, 2016). Les fonctions d'activation, présentes dans chaque couche du réseau, aident le modèle à devenir non linéaire, permettant de modéliser des relations complexes (Nwankpa, Ijomah, Gachagan, & Marshall, 2018).

Les techniques d'apprentissage profond englobent plusieurs approches et architectures de réseaux de neurones artificiels conçues pour modéliser des données complexes et multidimensionnelles. Parmi ces techniques, on trouve les réseaux de neurones convolutifs (CNN), largement utilisés pour le traitement des images et la reconnaissance d'objets en raison de leur capacité à capturer les caractéristiques spatiales et les motifs locaux (Albawi, Mohamed, & Al-Zawi, 2017). Les réseaux de neurones récurrents (RNN), incluant les réseaux de mémoire à long court terme (LSTM) et les réseaux neuronaux à portes (GRU), sont spécialisés dans le traitement des séquences temporelles et des données textuelles, en capturant les dépendances temporelles (Yu, Si, & Hu, 2019).

Les réseaux génératifs adversariaux (GAN) sont utilisés pour générer des données synthétiques en mettant en concurrence deux réseaux, un générateur et un discriminateur, améliorant la qualité des données générées (Karras, et al., 2019). Les auto-encodeurs sont employés pour la réduction de dimensionnalité et la détection d'anomalies en apprenant une représentation compacte des données (Bank, Koenigstein, & Giryas, 2023). Enfin, les transformeurs, qui ont révolutionné le traitement du langage naturel avec des modèles comme BERT et GPT, utilisent des mécanismes d'attention pour gérer efficacement de longues séquences et capturer des relations complexes entre les éléments (Abada, Abubakar, & Bilal, 2022).

2.4.4.1 CNN

Une compréhension approfondie des méthodes de Deep Learning est essentielle, notamment lors de l'utilisation des CNN. La création d'un modèle fiable et précis nécessite une expertise en architecture de modèle pour les CNN, y compris des techniques de prétraitement telles que la normalisation et l'augmentation des données. Les couches convolutives, de pooling et entièrement connectées, que les fonctions d'activation, sont essentielles à la compréhension de

l'architecture afin de créer un réseau efficace qui extrait les informations pertinentes et reconnaît les modèles des données de flux thermique.

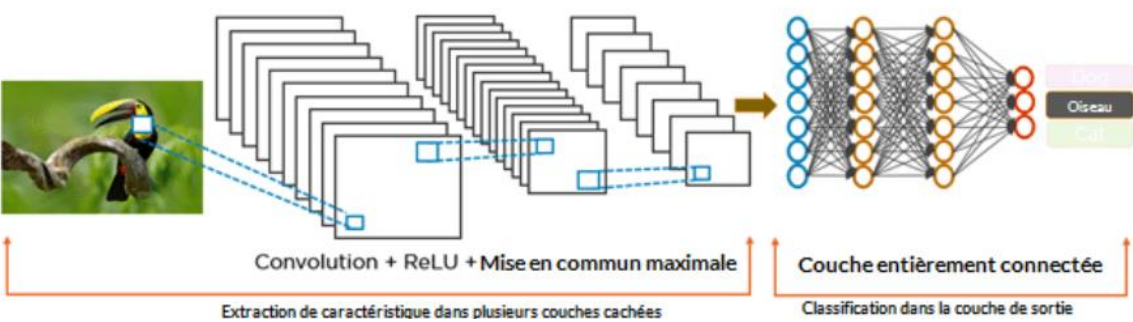


Figure 5: Structure d'un CNN (Benchohra & Mahmoudi, 2020)

Dans les CNNs, l'image passe à travers une série de couches convolutives, où des filtres sont utilisés pour numériser l'image dans le but de détecter des informations importantes, telles que les formes et les bords. (Krizhevsky, Sutskever, & Hinton, 2012) Vient ensuite la mise en commun des couches, dont l'objectif principal est de réduire la complexité et de se concentrer sur les fonctionnalités extraites. Ensuite, l'image passe par le FCN où sont appliquées les fonctions d'activation spécifiques à la tâche à accomplir. Enfin, la couche de sortie fournit le résultat final. (Prosper, 2025). La maîtrise du réglage des hyperparamètres et de la sélection des bibliothèques python est essentielle pour obtenir de bonnes performances.

Tableau 3: Les librairies du ML et leurs utilités

Librairie	Utilité
Tensorflow	Application des techniques d'apprentissage profond
Numpy	Operations mathématiques sur les tableaux
Pandas	Analyse des données
Matplotlib	Visualisation des données
Sklearn	Modélisation

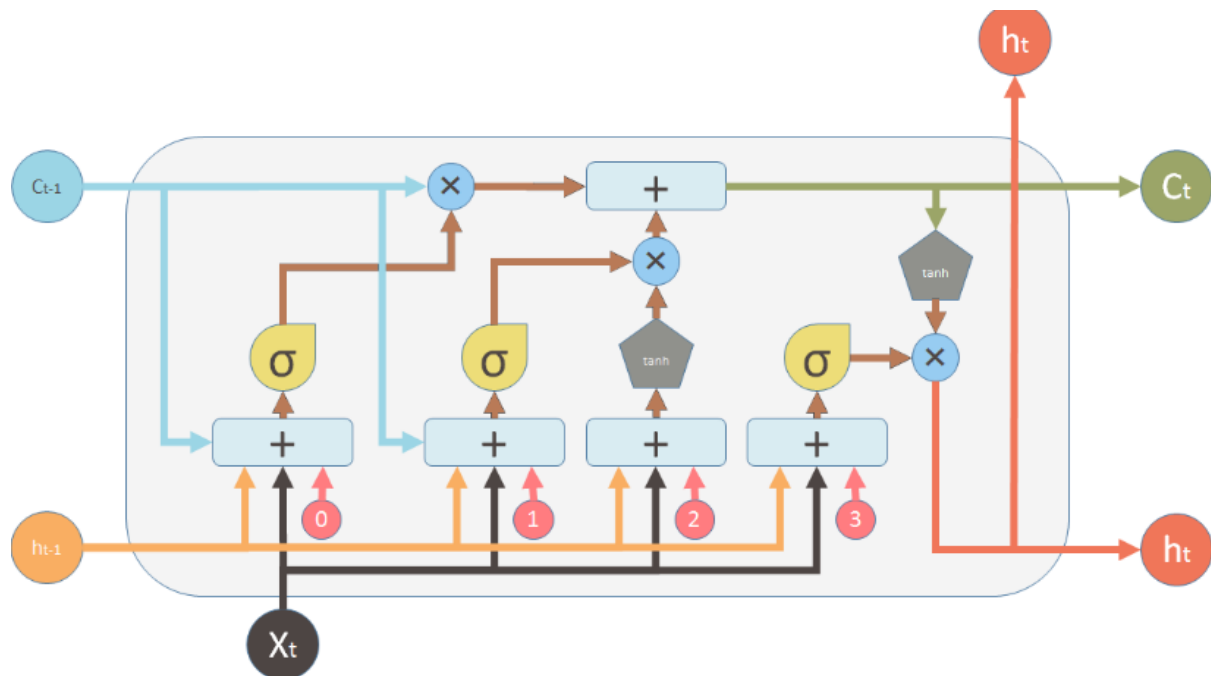
2.4.4.2 LSTM

Un LSTM est un type de réseau de neurones récurrents (RNN) conçu pour surmonter les problèmes de dépendances à long terme dans les séquences de données. Grâce à sa structure unique de cellules, portes d'entrée, de sortie et d'oubli, un LSTM peut conserver et accéder à des informations sur de longues périodes, ce qui le rend particulièrement adapté pour modéliser et prédire les dynamiques de systèmes non linéaires et variables dans le temps (Lindemann, Müller, Vietz, Jazdi, & Weyrich, 2021). Cette capacité à gérer les dépendances temporelles complexes permet aux LSTM d'exceller dans des tâches comme la prévision de séries temporelles, la reconnaissance vocale et la traduction automatique.

L'architecture d'un LSTM (Figure 6) contient une cellule de mémoire qui stocke l'information utile au fil du temps, et trois portes qui contrôlent cette information :

- La porte d'entrée décide quelle nouvelle information est ajoutée à la mémoire.
- La porte d'oubli décide quelles informations anciennes doivent être supprimées.
- La porte de sortie décide quelle partie de la mémoire sera utilisée pour produire la sortie à l'instant courant.

Ces mécanismes de portes permettent au réseau d'apprendre quelles informations garder ou ignorer, évitant ainsi la perte de contexte dans les longues séquences. En pratique, cela rend les LSTM capables de capturer les relations temporelles complexes et d'améliorer la précision des prédictions sur des données séquentielles comme les séries temporelles utilisées dans cette étude.



Entrées :	Sorties :	Fonctions d'activation :	Opérations vectorielles :
x_t Vecteur d'entrée	C_t Mémoire du bloc actuel	σ Sigmoidé	$\otimes$ Multiplication par éléments
C_{t-1} Mémoire du précédent bloc	h_t Sortie du bloc actuel	$\tanh$ Tangente hyperbolique	$+$ Sommation / concaténation par éléments
h_{t-1} Sortie du précédent bloc		Biais : 0	

Valeurs des sorties intermédiaires que nous ne détaillerons pas dans cet article : $\rightarrow$

Figure 6: Structure d'un LSTM (Bourdois, 2019)

3 METHODOLOGIE

3.1 Zones d'études

La zone choisie pour la récupération des données de température et d'humidité pour 2024 et 2025 est au parc technologique de l'INRS. Les coordonnées géographiques dans cette zone sont $46^{\circ} 47' 44.9''$ N et $71^{\circ} 18' 4.98''$ O.



Figure 7: Zone d'études parc technologique INRS

On a utilisé aussi des données mesurées en 2020 et 2021 de température et d'humidité durant la même période de l'année dans le champ 41 : Route du pont rouge-Ste-Catherine. Les coordonnées géographiques dans cette zone sont $46^{\circ} 50' 11.9''$ N, $71^{\circ} 39' 1.6''$ O. Les deux zones n'ont pas le même type de sol.



Figure 8: Zone d'études champ 41

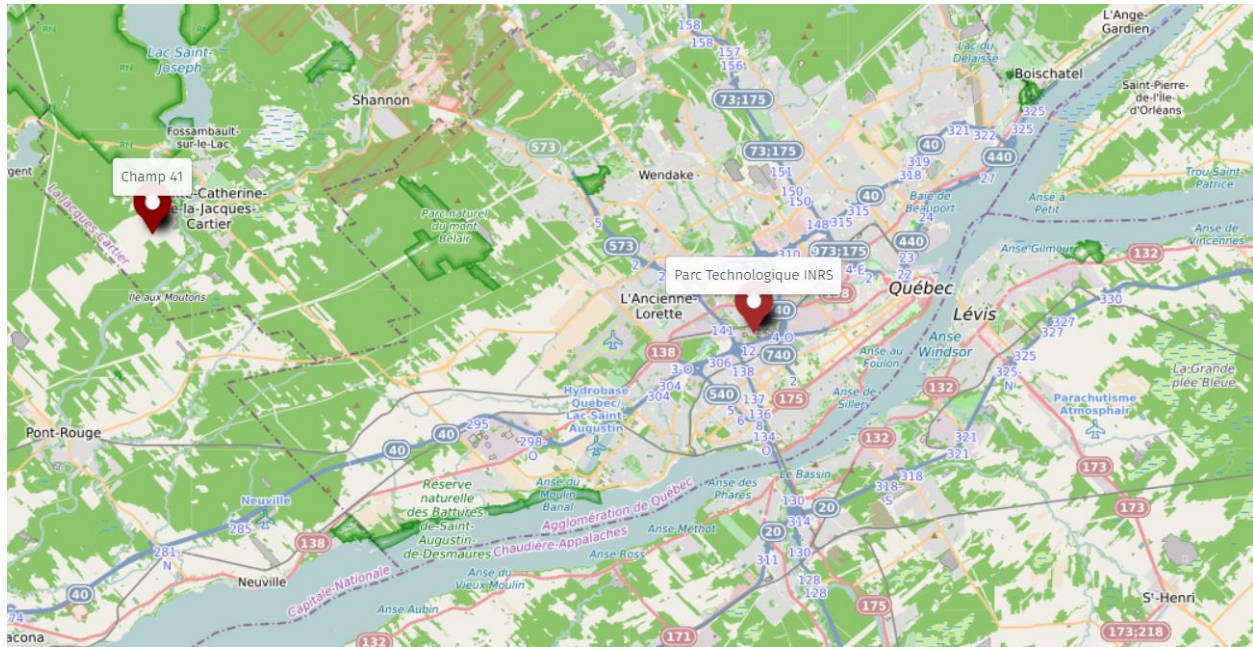


Figure 9: Emplacement des deux zones d'études au Québec

3.2 Outils et matériels

La mesure de ces données a été effectuée avec des capteurs FDR. Spécifiquement, on a utilisé les capteurs 5TM manufacturés par Decagon



Figure 10: Capteurs 5TM Decagon

Ces capteurs mesurent à la fois la température et la teneur en eau du sol. Ils ont une précision de la permittivité diélectrique apparente de ± 1 Ka pour une plage de 1 à 40 et de ± 15 % pour des valeurs comprises entre 40 et 80. La teneur en eau volumique est calculée selon l'équation de Topp et a une précision de $\pm 0,03 \text{ m}^3/\text{m}^3$ (± 3 %) dans les sols minéraux dont la conductivité électrique est inférieure à 10 dS/m, et de $\pm 0,02 \text{ m}^3/\text{m}^3$ (± 2 %) lorsqu'une calibration spécifique au milieu est appliquée. La mesure de température offre une précision de ± 1 °C. Enfin, la résolution du capteur est de $0,0008 \text{ m}^3/\text{m}^3$ pour la teneur en eau et de $0,1$ °C pour la température. (CampbellScientific, s.d.)

L'utilisation d'un profileur de température CS230 SDI-12 a aussi été effectuée pour mesurer la température du sol. Ce profileur est équipé de 32 capteurs de température espacés de 5 cm qui mesurent en continu les valeurs de température dans le sol créant un profil. Il peut fonctionner dans une large plage de températures allant de $-55\text{ }^{\circ}\text{C}$ à $+85\text{ }^{\circ}\text{C}$ et présente une précision typique de $\pm 0,2\text{ }^{\circ}\text{C}$ entre $-40\text{ }^{\circ}\text{C}$ et $+85\text{ }^{\circ}\text{C}$. Dans les conditions les plus défavorables, la précision atteint $\pm 0,4\text{ }^{\circ}\text{C}$ pour la plage de $-40\text{ }^{\circ}\text{C}$ à $+85\text{ }^{\circ}\text{C}$ et $\pm 0,5\text{ }^{\circ}\text{C}$ entre $-55\text{ }^{\circ}\text{C}$ et $-40\text{ }^{\circ}\text{C}$. Ce capteur offre ainsi une mesure fiable et stable pour les applications environnementales exigeant une grande robustesse. (CampbellScientific, CS230-L Temperature Probe [Fiche produit], s.d.)



Figure 11: CS230 SDI-12 profileur de température

Pour cette étude, nous avons utilisé l'enregistreur de données Em50, reliés avec les capteurs 5TM Decagon pour enregistrer les données. Il contient 6 ports différents, il peut donc contenir jusqu'à 5 capteurs et un port dédié à la communication qui sert à se connecter à un ordinateur et à transférer les données. La capacité de stockage des données de l'enregistreur est de (CaTeC, s.d.)



Figure 12: Data logger Em50G

Une fois connecté à un ordinateur le logiciel est capable de modifier le nom et les propriétés de l'enregistreur de données et de télécharger toutes les données collectées sous forme de fichier Excel. (DecagonDevices, 2007)

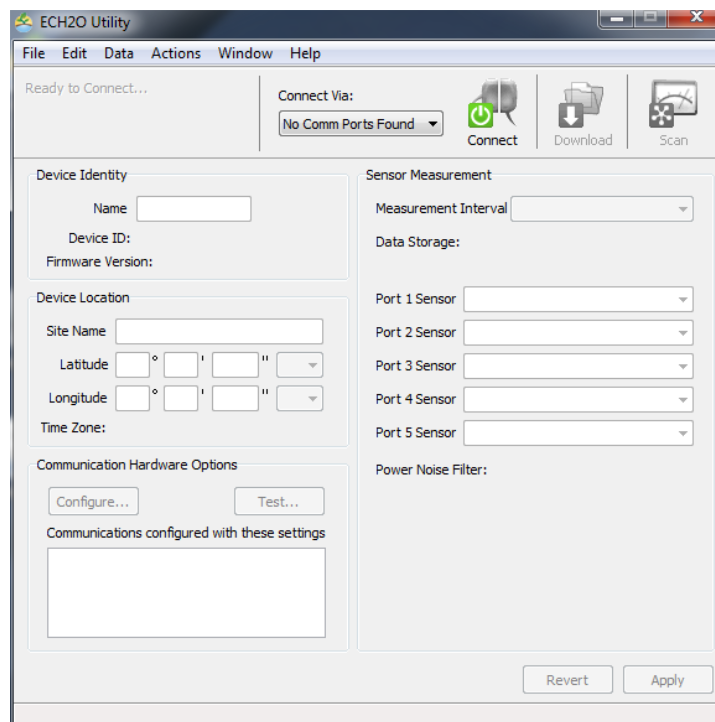


Figure 13: Interface du logiciel ECH2O

3.3 Calibration des capteurs FDR pour la mesure de la teneur en eau

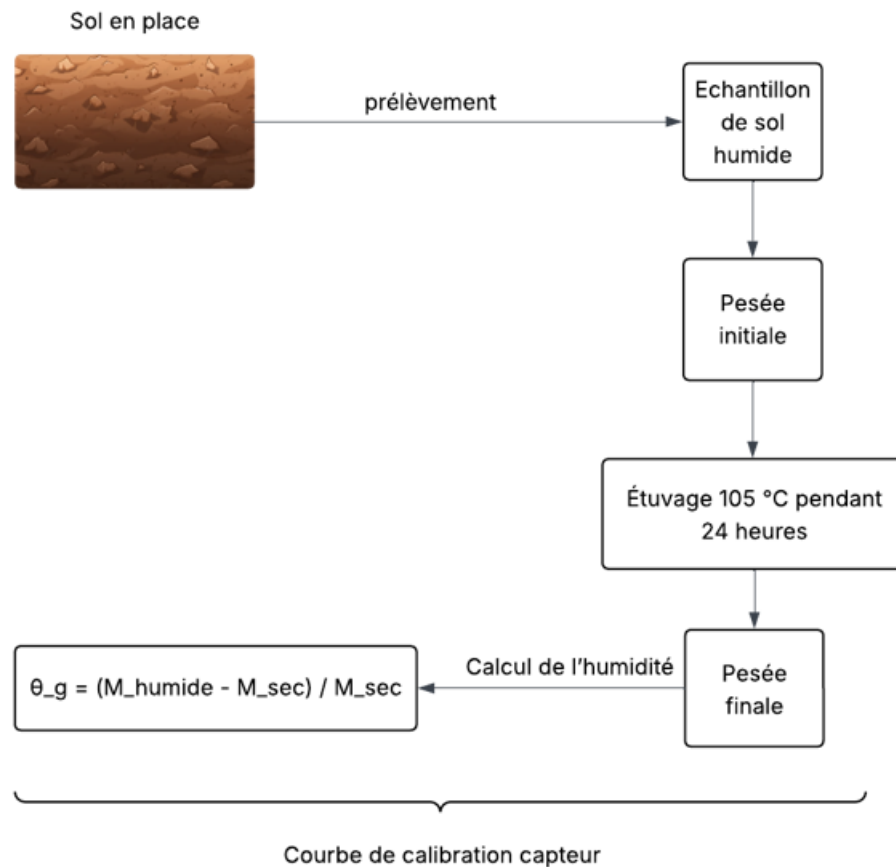


Figure 14: Méthode gravimétrique pour la calibration des capteurs

Cette figure décrit le protocole standard de la méthode gravimétrique, considérée comme la technique de référence pour mesurer l'humidité du sol et établir les courbes de calibration des capteurs. Le processus débute par le prélèvement d'un échantillon de sol humide directement in situ. Celui-ci est soumis à une pesée initiale afin de déterminer sa masse humide. Ensuite, l'échantillon est placé en étuve à 105 °C pendant 24 heures, ce qui permet d'éliminer totalement l'eau contenue dans les pores du sol, tout en conservant la masse des phases solides. Une fois séché, l'échantillon est de nouveau pesé pour obtenir sa masse sèche. Cette valeur exprime la proportion d'eau par rapport à la masse sèche du sol. Elle constitue une mesure fiable et reproductible, qui sert de référence pour la calibration des capteurs. En reliant les données gravimétriques aux signaux fournis par les capteurs, il est possible d'établir une courbe de calibration spécifique au sol étudié, garantissant une meilleure précision dans l'estimation de l'humidité volumique. (Agustirandi, Inayah, Siti Aminah, & Budiman, 2024)

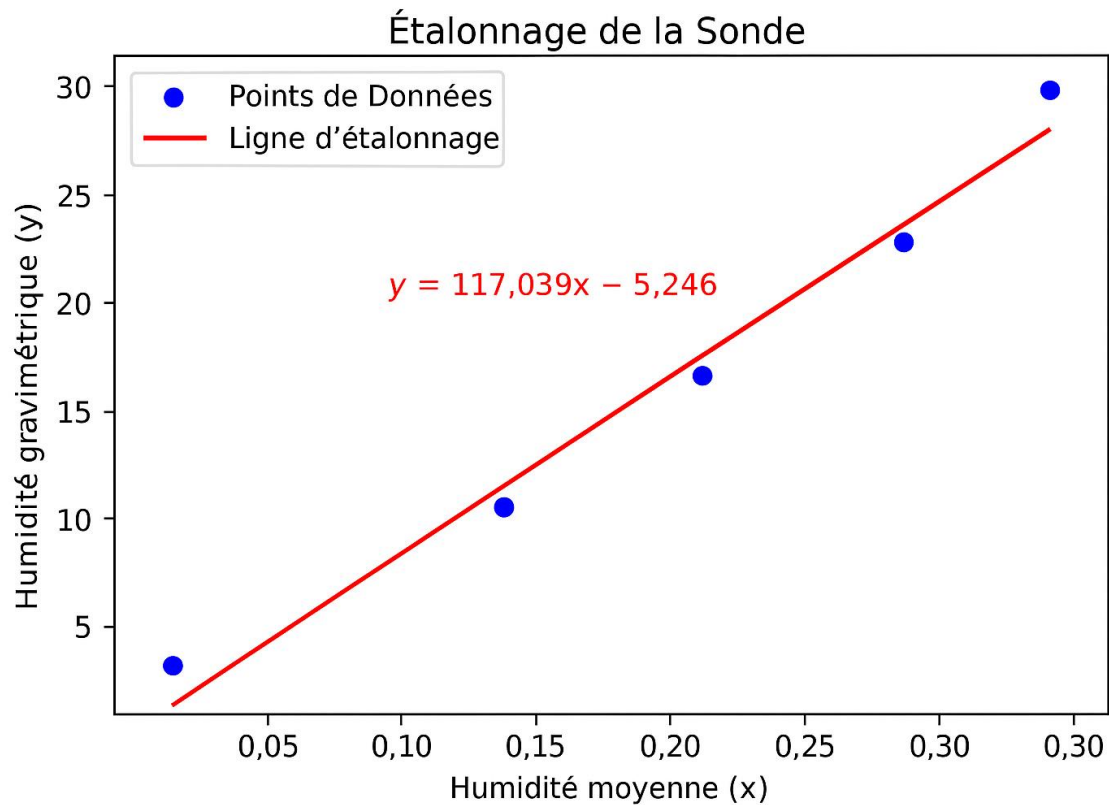


Figure 15: Courbe de calibration

On obtient ensuite une courbe de calibration et une équation pour calibrer les valeurs mesurées en des valeurs réelles.

3.4 Dispositif expérimental

Durant l'été 2020, des capteurs 5TM ont été installés dans la zone Champ 41 dans le cadre d'un projet mené par une chercheuse de l'INRS (*Araujo Martins Sarah*). Ces capteurs ont permis la mesure continue de l'humidité et de la température du sol sur deux saisons estivales consécutives, 2020 et 2021. La profondeur d'installation était de 45 cm, correspondant à un profil de sol homogène dominé par un seul horizon textural. Cette zone est caractérisée par une structure relativement sableuse, favorisant un drainage rapide et une faible rétention d'eau.



Figure 16: Photo réelle de l'emplacement des capteurs dans le site champ 41

En 2023, un second dispositif de mesure a été mis en place dans la zone du Parc technologique de Québec, également sous la supervision de l'équipe de l'INRS. Contrairement à Champ 41, cette zone présente une plus grande variabilité verticale, comprenant trois horizons distincts (0–20 cm, 20–40 cm et 40–55 cm), observables sur la photo de terrain (figure 18) et décrits en détail dans le triangle textural (figure 20). Les données collectées sur le site du Parc technologique couvrent les périodes 2023, 2024 et 2025, et ont été utilisées pour l'entraînement et la validation complémentaire des modèles d'apprentissage développés dans le cadre de ce projet.

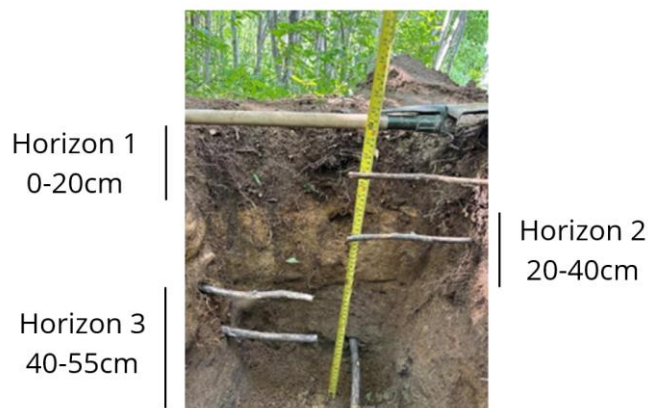


Figure 17: Photo réelle des horizons et l'emplacement des capteurs dans le site du parc technologique

La figure 18 illustre la variabilité texturale des horizons étudiés, des couches superficielles sableuses et des horizons plus argileux en profondeur. Cette diversité est un facteur important qui influence directement la relation entre flux thermique et humidité, et justifie l'utilisation de modèles d'apprentissage profond capables de généraliser sur des textures différentes.



Figure 18: Dispositif expérimental final du site parc technologique

Dans la figure 18 et 19 nous pouvons observer le dispositif expérimental final dans les deux sites. Les sondes ont été placées, l'enregistreur de données peut être vu au-dessus du sol, le profileur de température, et un ruban adhésif a été utilisé pour isoler la zone. De plus, l'ajout de terre au-dessus du niveau du sol pour empêcher l'eau de déborder et d'affecter les mesures a été effectué.



Figure 19: Dispositif expérimental final du site champ 41

Pour avoir une meilleure idée de l'emplacement des capteurs dans le sol, la figure suivante représente un diagramme contenant la position exacte des capteurs et les différents horizons du sol du site parc technologique.

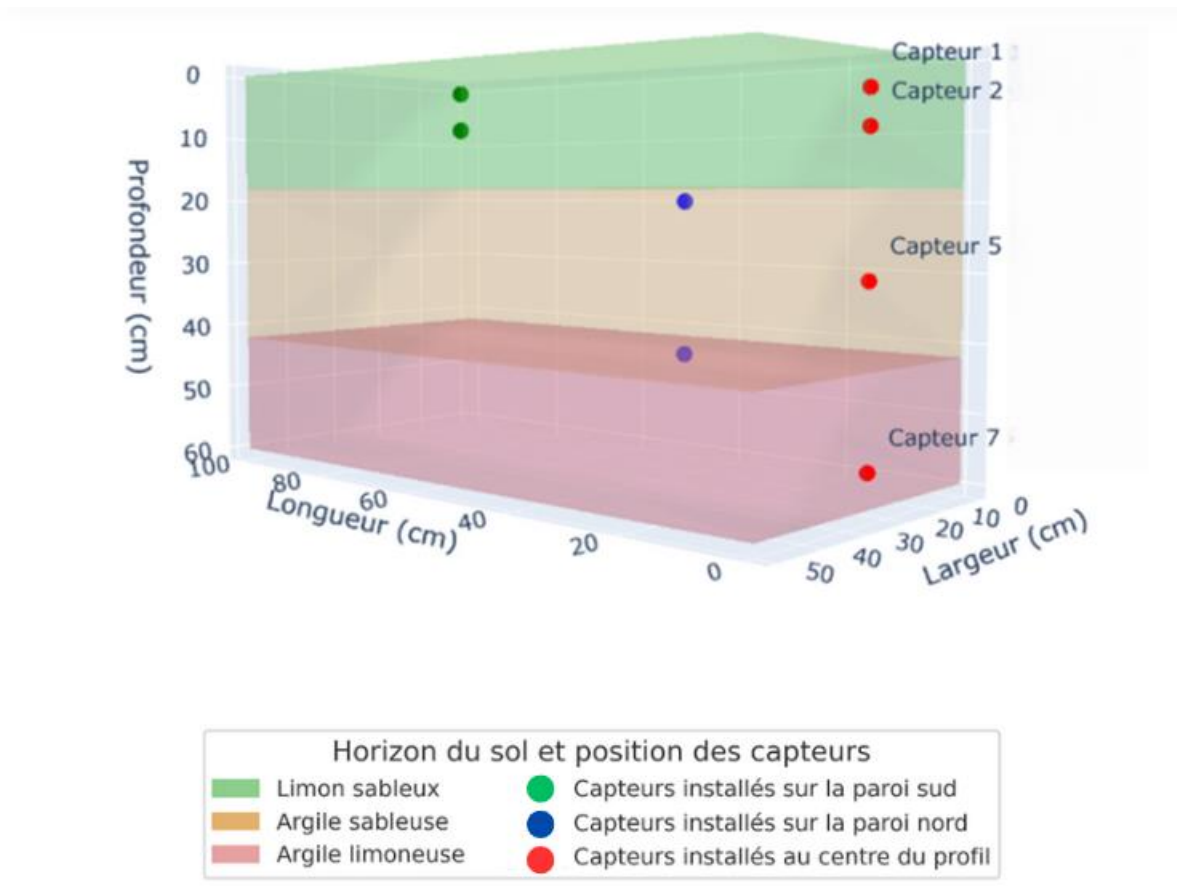


Figure 20: Diagramme de l'emplacement des capteurs et les horizons

La figure 21 présente le triangle textural des sols permettant de classer les différents horizons étudiés en fonction de leurs proportions relatives en sable, limon et argile. Les zones cerclées et annotées correspondent aux échantillons collectés dans le cadre du projet.

- Site 1 : Horizon 1 (0–20 cm) est identifié comme un limon sableux, ce qui traduit une forte proportion de sable combinée à une fraction notable de limon. Ce type de sol favorise un drainage rapide mais retient une humidité modérée.
- Site 1 : Horizon 2 (20–40 cm) se situe dans la classe argile sableuse, caractérisée par une teneur plus importante en argile qui améliore la rétention d'eau, mais peut limiter la perméabilité.

- Site 1 : Horizon 3 (40–55 cm) est classé comme argile limoneuse, avec une proportion élevée d'argile et une composante limoneuse significative. Cette texture entraîne une forte capacité de rétention d'eau mais une aération réduite.
- Site 2 : Route du Pont Rouge – Ste-Catherine correspond à un sol sableux, avec une dominance claire de sable. Ce type de sol est très drainant, mais présente une faible capacité de stockage en eau, le rendant plus sensible à la sécheresse.

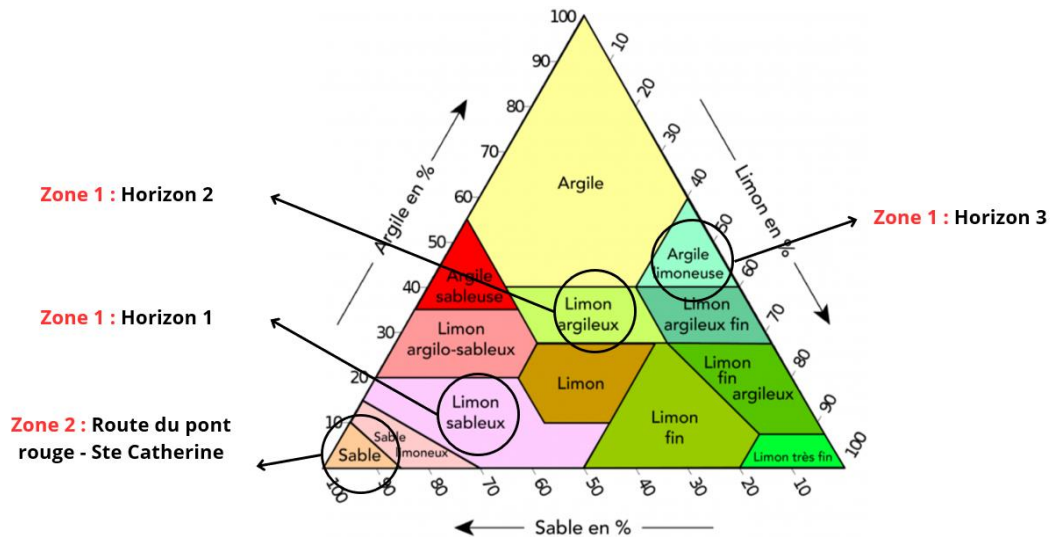


Figure 21: Types des sols

3.5 Création du modèle et analyse de données

3.5.1 Structure des données

Les données utilisées dans cette étude proviennent de deux périodes de mesure distinctes : les années 2020-2021, et les années 2023-2024-2025. Ces deux ensembles de données présentent des caractéristiques différentes en termes de types de capteurs, de positionnement et de paramètres mesurés, nécessitant une structuration spécifique afin de les rendre comparables et exploitables dans le cadre de la modélisation.

- Données de 2020 et 2021

La température a été mesurée à l'aide d'une sonde thermique équipée de capteurs espacés tous les 5 cm en profondeur (CS320 SDI-12 ; Figure 13). Les mesures de température ont été prises

toutes les 5 minutes, offrant une résolution temporelle fine. L'humidité du sol a été mesurée grâce à des capteurs FDR, installés dans trois profondeurs : 0–15 cm, 15–30 cm et 30–45 cm.

- Données de 2023, 2024 et 2025

Des capteurs FDR ont été installés aux profondeurs suivantes : 5 cm, 10 cm, 20 cm, 30 cm, 40 cm et 55 cm. Ces capteurs mesurent à la fois la température et l'humidité à chaque niveau de profondeur. Les mesures sont également prises toutes les 5 minutes. L'utilisation du profileur de température CS320-SDI12 (Figure 13) a aussi été effectuée.

Afin de faciliter l'analyse et l'entraînement des modèles, toutes les données ont été agrégées sur une base horaire. Pour cela, chaque heure (60 minutes) a été divisée en 12 intervalles de 5 minutes. Pour chaque variable (température ou humidité), la moyenne des 12 valeurs a été calculée, produisant un jeu de données horaire plus stable et plus adapté à la modélisation.

Date et heure	Humidite	Tsup	Tinf	Tflux
7/18/2020 12:00	0.15309	24.39667	20.79667	0.24
7/18/2020 13:00	0.152039	25.735	21.36333	0.291444
7/18/2020 14:00	0.15067	26.82583	22.035	0.319389
7/18/2020 15:00	0.14972	27.3875	23.07333	0.287611
7/18/2020 16:00	0.148697	28.36417	24.10333	0.284056
7/18/2020 17:00	0.147672	30.17667	24.81167	0.357667

Figure 22: Exemple de données

La structuration a permis de créer trois variables importantes :

- Tsup : Température mesurée à la partie haute d'un bloc de sol. Elle reflète les conditions thermiques proches de la surface du bloc.
- Tinf : Température mesurée à la partie basse du même bloc. Elle indique la quantité de chaleur qui pénètre plus en profondeur.
- Tflux : Différence entre Tsup et Tinf, divisée par l'épaisseur du bloc. Elle représente une approximation du flux de chaleur vertical.

Les fenêtres temporelles de 24 h, 48 h, 72 h et 96 h ont été choisies et correspondent à des échelles de temps associées aux principaux processus thermiques et hydriques du sol. La fenêtre de 24 h correspond au cycle thermique journalier imposé par le rayonnement solaire et le refroidissement nocturne, qui constitue la principale source de variation de la température du sol

près de la surface (Hillel, 2004). Ce cycle diurne affecte directement le gradient thermique et le flux de chaleur dans les couches superficielles du sol.

Les fenêtres de 48 h, 72 h et 96 ont été utilisées dans le cadre de cette étude pour prendre en compte les processus plus lents, comme la diffusion de la chaleur vers les couches plus profondes et la redistribution progressive de l'humidité après une précipitation. Ces processus se produisent entre 24 heures et quelques jours, en fonction de la diffusivité thermique du sol et de la profondeur considérée (Oschner, Horton, & Ren, 2013). De même, plusieurs études ont montré que les variations de l'humidité du sol peuvent persister sur plusieurs jours avant d'atteindre un nouvel état d'équilibre, en particulier dans les horizons intermédiaires et profonds.

L'utilisation de plusieurs fenêtres temporelles permet d'explorer différentes échelles physiques des processus thermiques et hydriques du sol et d'évaluer la capacité des modèles à exploiter des dépendances temporelles de durées croissantes.

3.5.2 Métriques d'évaluation

Les principales métriques d'évaluation utilisées pour les tâches de régression sont le R^2 , le RMSE, le MSE, le MAE et le biais. En plus de ces métriques, une analyse des courbes de validation au fil des époques est nécessaire pour valider le modèle (Botchkarev, 2018).

- **R^2 :**

Le R^2 est le coefficient de détermination, qui donne un aperçu de la qualité de l'ajustement du modèle aux données. Un R carré faible signifie que le modèle fonctionne mal, alors qu'un R carré plus élevé signifie que le modèle fonctionne bien. L'équation suivante représente sa formule.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (13)$$

- **RMSE :**

C'est la racine carrée de l'erreur quadratique moyenne et représente la différence entre les valeurs prédites et les valeurs observées. Plus le RMSE est bas, meilleur est le pouvoir de prédiction du mode. L'équation suivante représente sa formule.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N \frac{(x_i - \hat{x}_i)^2}{N}}{N}} \quad (14)$$

N est le nombre de données non manquantes, tandis que x_i est la série chronologique des observations réelles et $\hat{x}_i$ est la série chronologique estimée.

- **MAE :**

MAE est l'erreur absolue moyenne, représentant la différence absolue moyenne entre les valeurs prédites et observées. Comme pour le RMSE, plus le MAE d'un modèle est bas, mieux c'est.

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \quad (15)$$

- **Le biais :**

Le biais mesure la moyenne des écarts entre les valeurs prédites et les valeurs observées. Il indique si le modèle a tendance à surestimer (biais positif) ou sous-estimer (biais négatif) les résultats. Il permet de détecter un décalage entre les prédictions du modèle et la réalité.

$$Bias = \left(\frac{1}{n}\right) \sum_{i=1}^n (\hat{y}_i - y_i) \quad (16)$$

3.5.3 Modèles d'apprentissage profond

Deux types de modèles d'apprentissage profond ont été sélectionnés pour l'estimation de l'humidité du sol à partir des flux thermiques en raison de leurs capacités respectives :

- 4 modèles CNN ont été créés afin de capturer les relations spatiales entre les variables sur différentes séquences temporelles (24h, 48h, 72h, 96h).

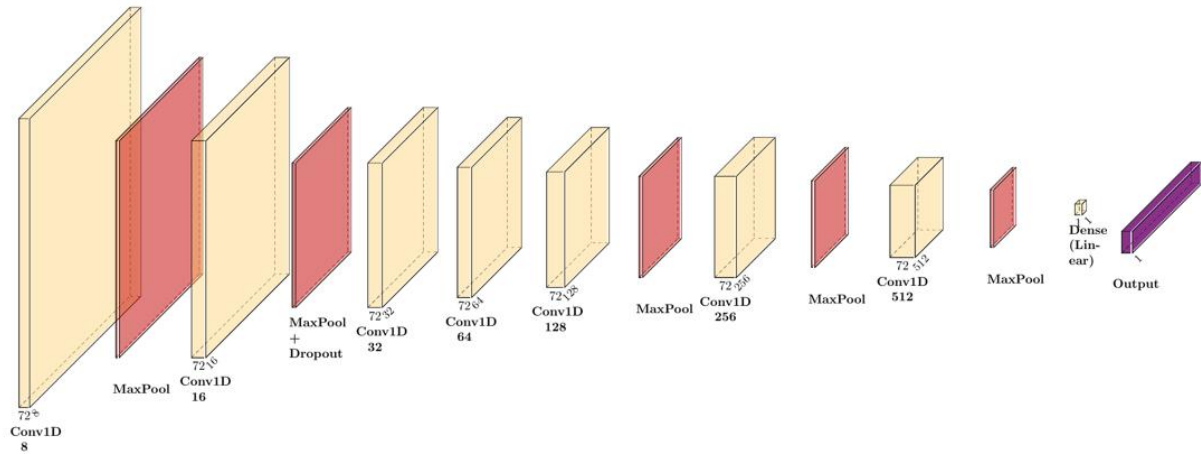


Figure 23: Architecture CNN utilisée

- 4 modèles LSTM ont été mis en œuvre pour exploiter les dépendances temporelles sur différentes séquences temporelles (24h, 48h, 72h, 96h).

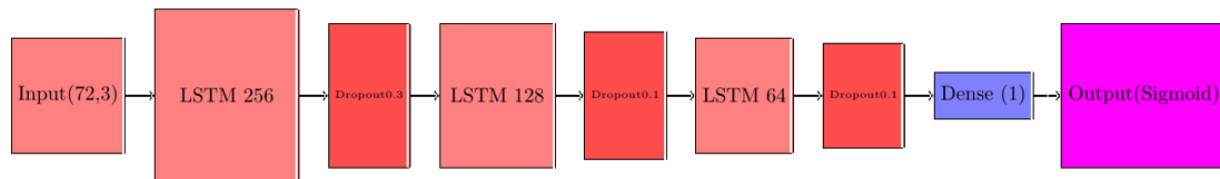


Figure 24: Architecture LSTM utilisée

On observe dans cette figure une architecture LSTM simple avec une succession de couches LSTM et dropout, suivie par une couche dense et un output en fonction sigmoïde.

Les performances de ces modèles avaient été comparées sur l'ensemble de test à l'aide de métriques telles que la RMSE, le MAE et le coefficient de corrélation R^2 . En complément, des modèles de base avaient également été utilisés pour établir un point de référence dans l'évaluation des performances. La comparaison a été effectuée avec 3 modèles de bases :

- XGBoost
- Régression linéaire
- Modèle de persistance

Cette comparaison avait été approfondie à l'aide d'analyses graphiques, notamment des nuages de points, des courbes de perte entraînement/validation, et le suivi de l'évolution du R^2 au fil des époques. Ces analyses avaient permis d'identifier les forces et faiblesses de chaque modèle en

termes de précision des prédictions et de capacité à capturer les relations complexes présentes dans les données.

3.5.3.1 Augmentation des données

Puisque les modèles d'apprentissage profond dépendent considérablement de la quantité de données, une augmentation de données peut être nécessaire. Cela peut être accompli en utilisant les caractéristiques de Tflux.

$$T_{flux} = \frac{T_{sup} - T_{inf}}{d} \quad (17)$$

d : Distance entre les deux mesures de température.

On peut changer la distance d et choisir des valeurs de T_{sup} et T_{inf} qui correspondent à cette distance. En suivant cette approche, on peut considérablement élargir notre base de données. La figure suivante éclaire cette méthode.

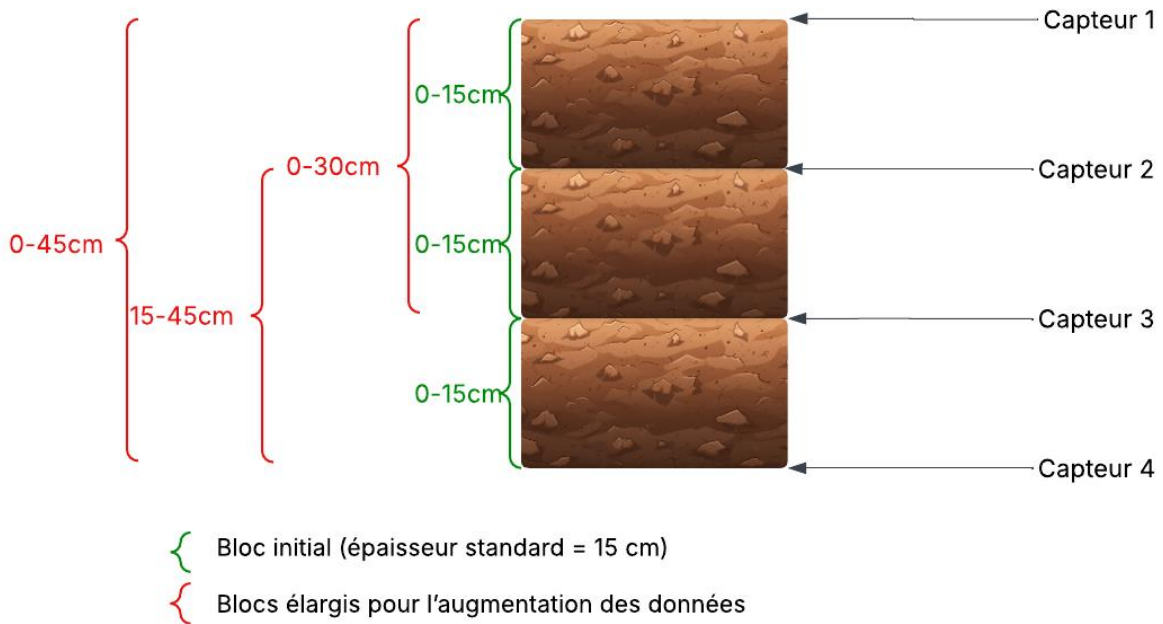


Figure 25: Technique d'augmentation de données dans le site 2

La séparation verte représente la séparation initiale de la structuration des données, correspondant à 15 cm dans le site 2 et 5 cm dans le site 1. La séparation rouge représente les séparations supplémentaires qui ont été effectuées. En augmentant la distance entre les deux mesures de températures supérieures et inférieures, on peut créer plus de données à utiliser. Il

est possible de diviser les données de manière encore plus fine en effectuant des séparations de 5 cm ou 10 cm, pour augmenter davantage la taille du jeu de données et c'est ce qui a été effectué dans le site 1. (Figure 26)

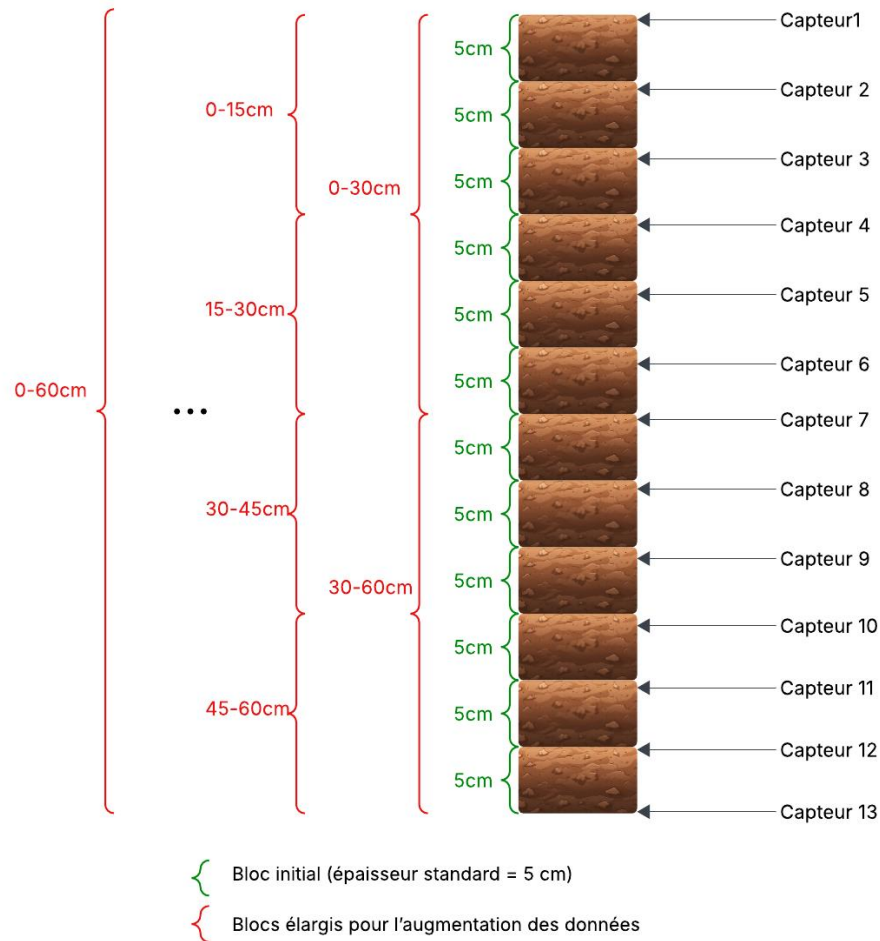


Figure 26: Technique d'augmentation de données dans le site 1

L'augmentation des données par la variation des distances entre les mesures de température a pour but d'exploiter la continuité verticale du profil thermique afin d'augmenter la taille du jeu de données utilisé pour l'entraînement des modèles. Cette approche repose sur l'hypothèse que, sur des distances verticales limitées, les propriétés thermiques et hydriques du sol restent comparables.

Cependant, cette stratégie peut introduire du biais. La génération de plusieurs séquences à partir d'un même profil vertical peut augmenter la redondance des données utilisées et renforcer la dépendance entre eux. De plus, lorsque la distance entre les points de mesure augmente, la

relation entre gradient thermique et humidité peut être influencée par des variations locales de texture ou de structure du sol, ce qui peut introduire des artefacts dans les données augmentées.

Dans le cadre de cette étude, seuls des modèles entraînés à partir de données augmentées ont été considérés, à cause de manque de données. Aucune comparaison directe avec des modèles entraînés sur les données originales sans augmentation n'a été réalisée. L'effet de la stratégie d'augmentation sur la performance des modèles n'a pas été évalué de manière indépendante. Cette limitation constitue un point à approfondir dans des travaux futurs, notamment par une analyse comparative explicite avec et sans augmentation des données.

3.5.3.2 Optimisation des hyperparamètres

Dans cette étude, Keras Tuner a été utilisé pour optimiser automatiquement les hyperparamètres des deux modèles. Keras Tuner est une bibliothèque open-source conçue pour automatiser le processus d'optimisation des hyperparamètres dans les modèles construits avec Keras et très utilisés dans les modèles complexes (Rogachev & Melikhova, 2020). Elle permet de définir un espace de recherche pour divers hyperparamètres tels que :

- Le nombre d'unités dans les couches LSTM ou CNN
- Le taux de dropout
- Le taux d'apprentissage
- Le choix de l'optimiseur
- La taille des mini-lots (batch size)

Une fois l'espace de recherche défini, Keras Tuner explore différentes combinaisons de ces hyperparamètres. Ce processus permet d'identifier automatiquement la configuration optimale qui minimise la fonction de perte sur les données de validation, améliorant la performance et la robustesse du modèle, tout en gagnant du temps. Cette bibliothèque est excellente pour trouver rapidement la meilleure combinaison des modèles (Pon & Krishna, 2021).

Tableau 4: Hyperparamètres optimisés du modèle LSTM

Paramètre	Valeur
Nombre de couches LSTM	3
Nombre de neurones par couche	256, 128, 64
Taux d'apprentissage	0.001
Optimiseur	Adam
Nombre d'époques	200 (avec early stopping)
Fonction d'activation	Sigmoïde
Fonction de perte	Huber Loss
Couches de régularisation et taux de dropout	3 couches : 30 %, 10 %, 20 %
Taille de la séquence (entrée)	24 / 48 / 72 / 96 heures
Batch size	128

Tableau 5: Hyperparamètres optimisés du modèle CNN

Paramètre	Valeur
Nombre de couches Conv1D	7
Nombre de filtres par couche	8, 16, 32, 64, 128, 256, 512
Taux d'apprentissage	0.0005
Optimiseur	Adam
Nombre d'époques	500
Fonction d'activation	ReLU
Fonction de perte	Huber Loss
Couches de régularisation et taux de dropout	2 couches, 10 %

3.5.3.3 Calibration et validation des modèles

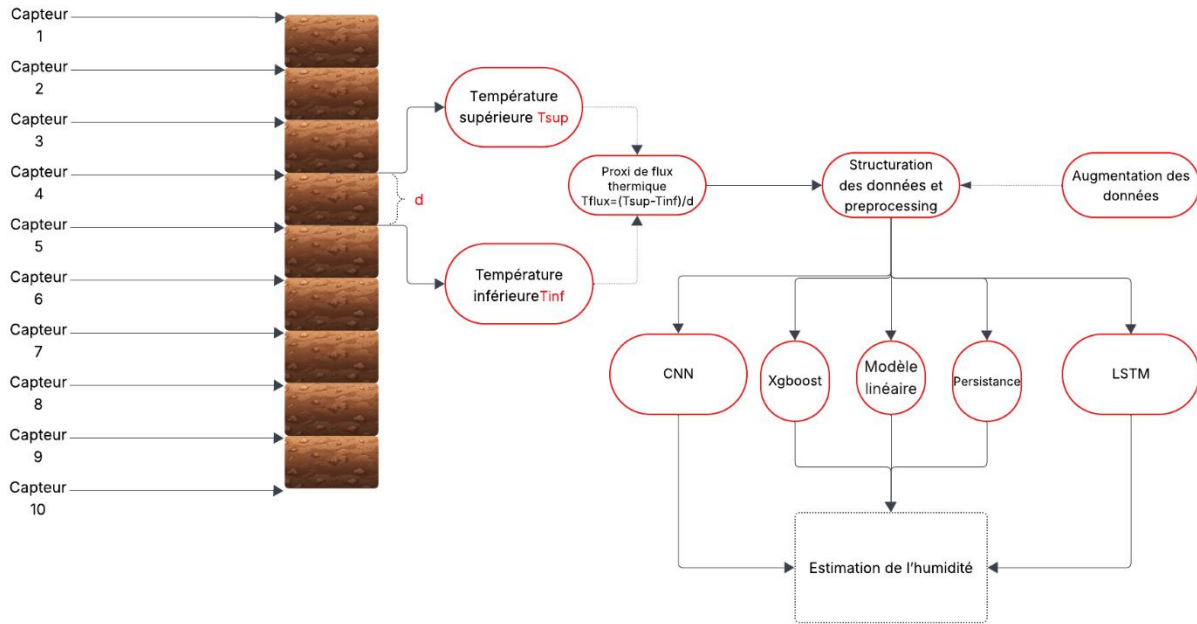


Figure 27: Approche utilisée pour estimer l'humidité en utilisant des modèles AP et des modèles peu profonds

Cette figure illustre le schéma méthodologique utilisé pour l'estimation de l'humidité du sol à partir de capteurs thermiques. Les données de température sont relevées à différentes profondeurs à l'aide de capteurs. L'écart entre ces températures permet de calculer une approximation du flux thermique (Equation 17). T_{flux} ne correspond pas à un flux thermique mesuré directement mais plutôt à une approximation basée sur l'écart des températures et la profondeur. Ces variables sont ensuite soumises à un processus de structuration et prétraitement des données, incluant une augmentation artificielle du jeu de données. Finalement, deux modèles d'apprentissage profond, CNN et LSTM, sont entraînés sur ces données ainsi que 3 modèles de bases pour fournir une estimation précise de l'humidité du sol.

3.5.4 Comparaison avec des modèles de base

3.5.4.1 XGboost

XGboost est un algorithme très performant sur des données tabulaires complexes développée par Chen Guestrin en 2016. C'est une version plus améliorée de l'algorithme Gradient Boosting (Friedman, 2001). Le schéma dans la figure 28 explique le fonctionnement de XGBoost. Le processus débute avec les données d'entraînement, qui sont utilisées pour construire un premier

arbre de décision à l'aide de la méthode CART. Après cette première prédiction, XGBoost identifie les erreurs commises et augmente la pondération des observations mal prédites, pour que le deuxième arbre se concentre sur ces cas difficiles. Ce mécanisme se répète pour chaque arbre de la séquence en cherchant à chaque fois à corriger les erreurs résiduelles des précédents arbres. Une fois tous les arbres construits, leurs prédictions partielles sont agrégées par moyenne pondérée. Grâce à cette architecture séquentielle et à la régularisation intégrée (L1/L2), XGBoost combine la puissance des arbres de décision avec un contrôle du surapprentissage. Cet algorithme est rapide, performant et robuste, particulièrement efficace pour la régression et la classification sur de grands ensembles de données, mais il reste complexe à paramétrer et peut nécessiter un temps d'entraînement important. (Chen & Guestrin, 2016)

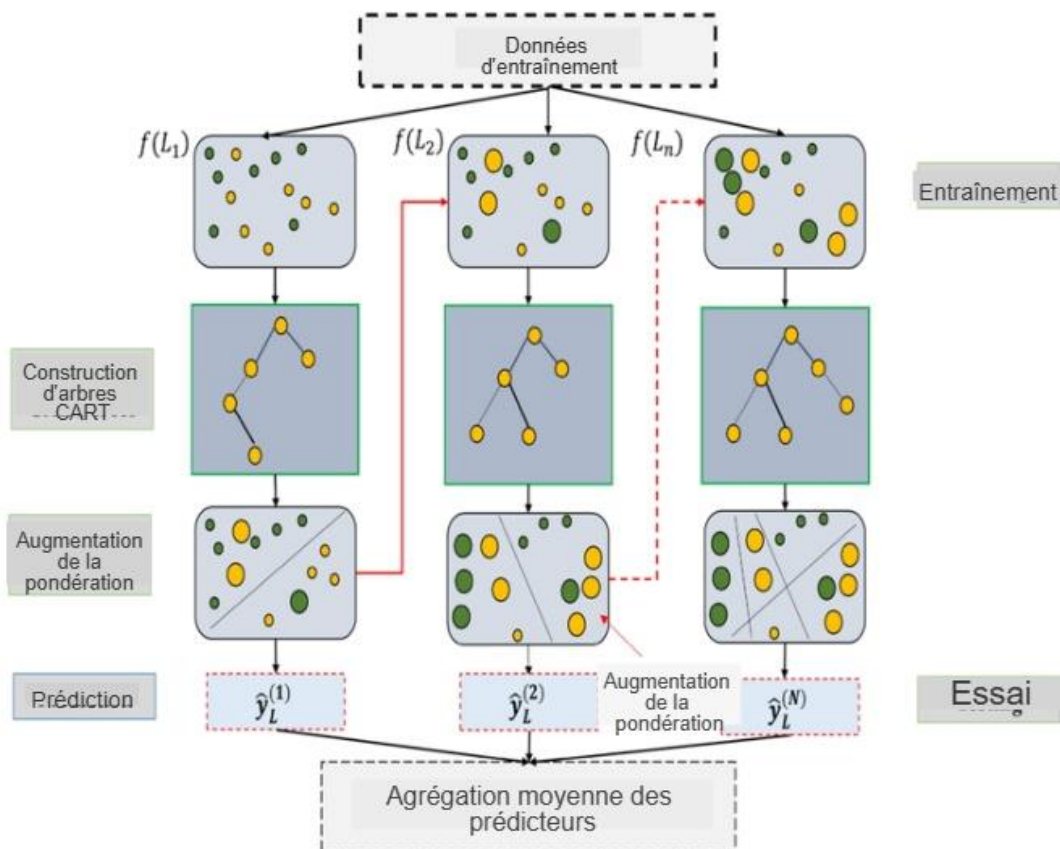


Figure 28: Architecture de XGboost (Chen & Guestrin, 2016)

3.5.4.2 Régression linéaire

La régression linéaire est un modèle statistique classique permettant d'estimer la relation entre une variable dépendante y et une ou plusieurs variables explicatives x_i (Kutner et al., 2005).

- Régression linéaire simple :

$$y = \beta_0 + \beta_1 x_i + \varepsilon_i \quad (18)$$

β_0 : l'ordonnée à l'origine

β_1 : le coefficient de régression

ε_i : terme d'erreur

La régression linéaire simple permet d'estimer la relation entre une seule variable explicative et la variable cible. Elle cherche à ajuster une droite qui minimise la somme des carrés des résidus. (Figure 29)

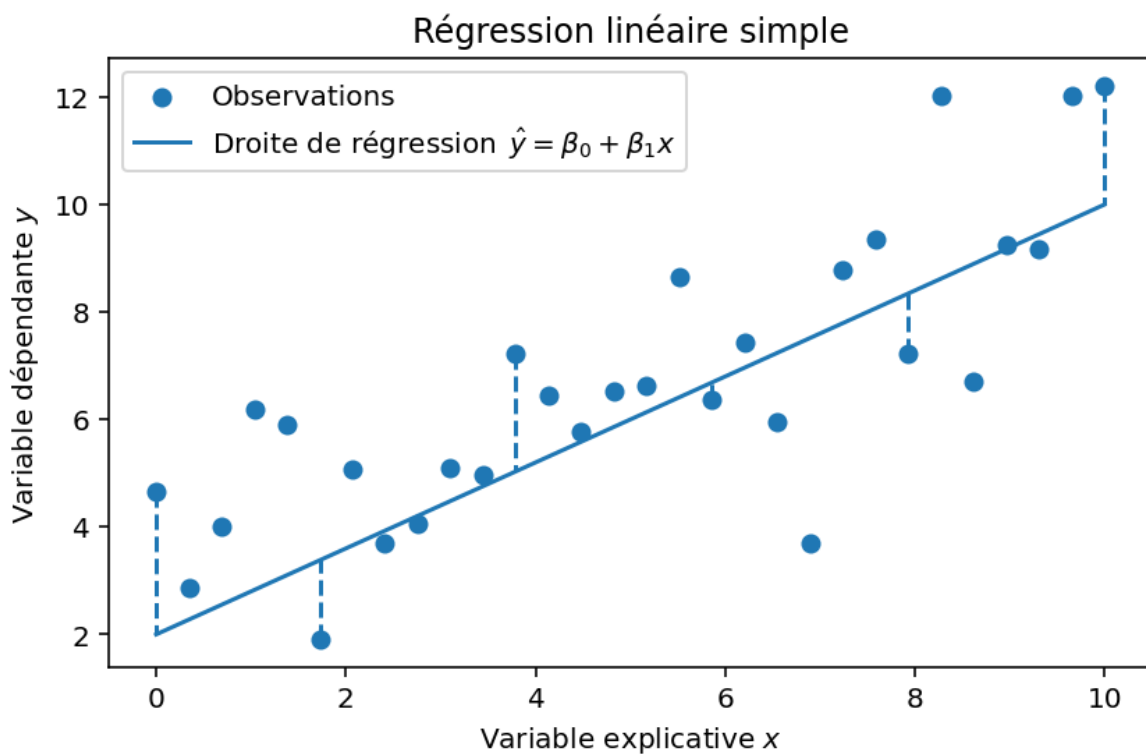


Figure 29: Exemple de régression linéaire simple

- Régression linéaire multivariée :

$$y = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_n x_{ni} + \varepsilon_i \quad (19)$$

β_0 : l'ordonnée à l'origine

β_1 : le coefficient de régression

ε_i : terme d'erreur

n : nombre de variables

La régression linéaire multiple généralise cette approche à n variables prédictives, ce qui permet de modéliser simultanément l'effet combiné de plusieurs facteurs (Figure 29).

Régression linéaire multiple

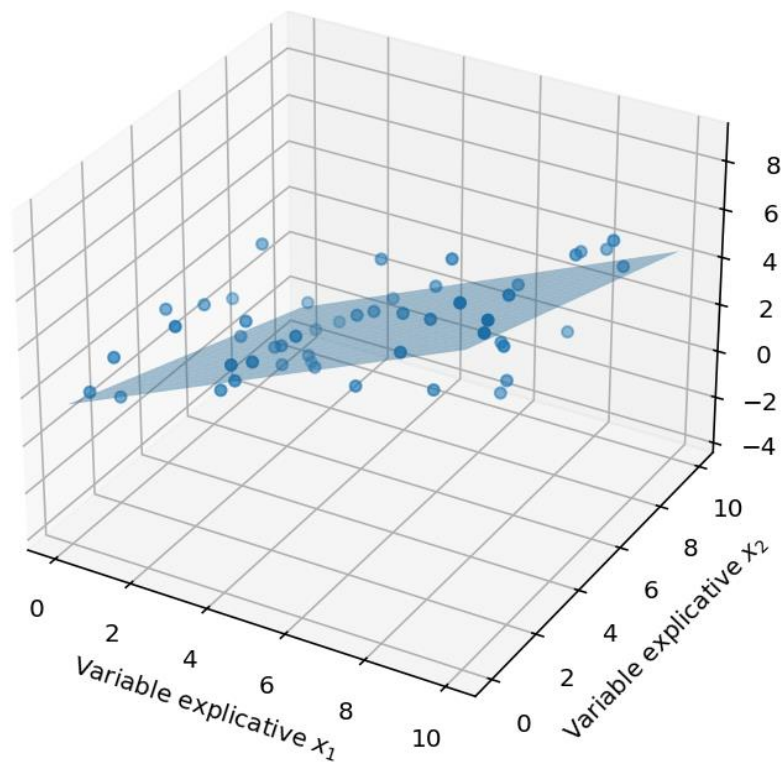


Figure 30: Exemple de régression linéaire multivariée

Dans le cadre de cette étude, elle est utilisée comme modèle de base pour vérifier si la relation entre les variables thermiques et l'humidité du sol peut être décrite de manière linéaire. En cas de déviation importante du modèle linéaire, cela met en évidence la présence de relations non linéaires nécessitant des approches d'apprentissage plus complexes comme l'apprentissage profond.

3.5.4.3 Persistance

Ce modèle prédit que la prochaine valeur d'humidité est identique à la dernière valeur observée. (Hyndman & Athanasopoulos, 2018). Ce modèle ne nécessite aucun entraînement et sert principalement de modèle de base pour évaluer la performance de méthodes prédictives plus avancées. Ce modèle est très simple mais peut fournir de bons résultats lorsque les variations des données sont faibles. (Figure 31)

$$y^{t+1} = y_t \quad (20)$$

y : Valeur observée

t : temps

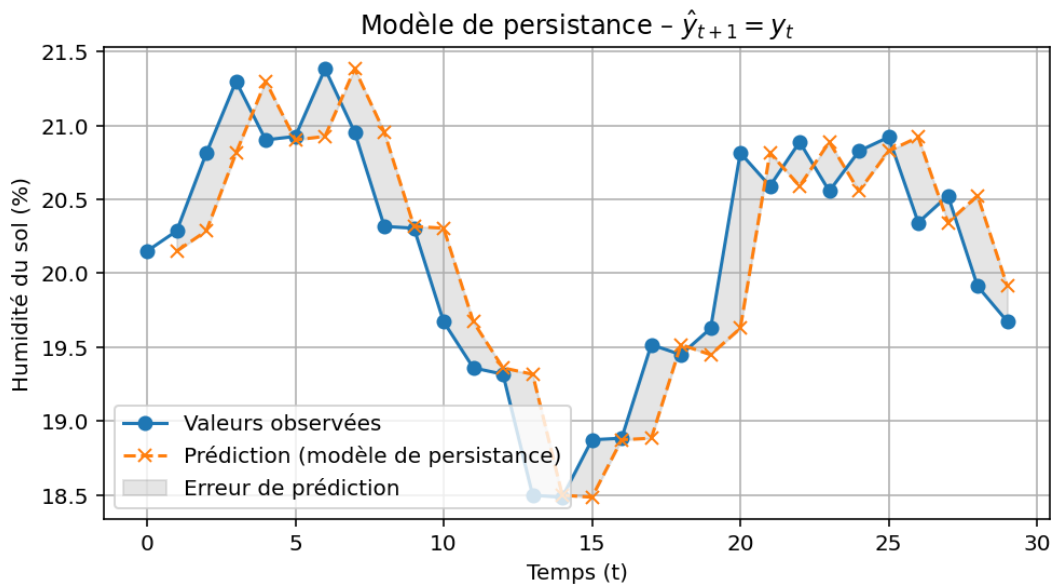


Figure 31: Exemple d'un modèle de persistance

4 RESULTATS ET DISCUSSION

4.1 Analyse des données

Une analyse détaillée de chaque horizon (0–20 cm, 20–40 cm et 40–55 cm) a été conduite.

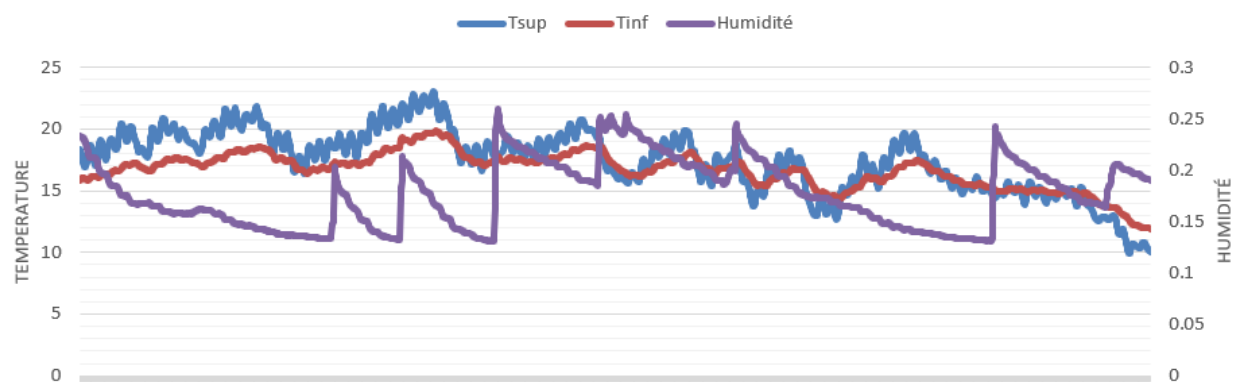


Figure 32: Analyse de l'horizon 0-20cm

On observe une variabilité importante des températures, avec Tsup toujours plus élevée que Tinf. Cela montre l'influence directe des conditions atmosphériques comme le rayonnement solaire, les précipitations et le refroidissement nocturne. Les écarts entre Tsup et Tinf indiquent un flux thermique plus important en surface. L'humidité présente de fortes fluctuations avec plusieurs pics soudains. Ces pics correspondent à des événements de pluie ou d'irrigation, qui saturent rapidement la couche. Après chaque pic, on observe une diminution rapide cause par le drainage et l'évapotranspiration.

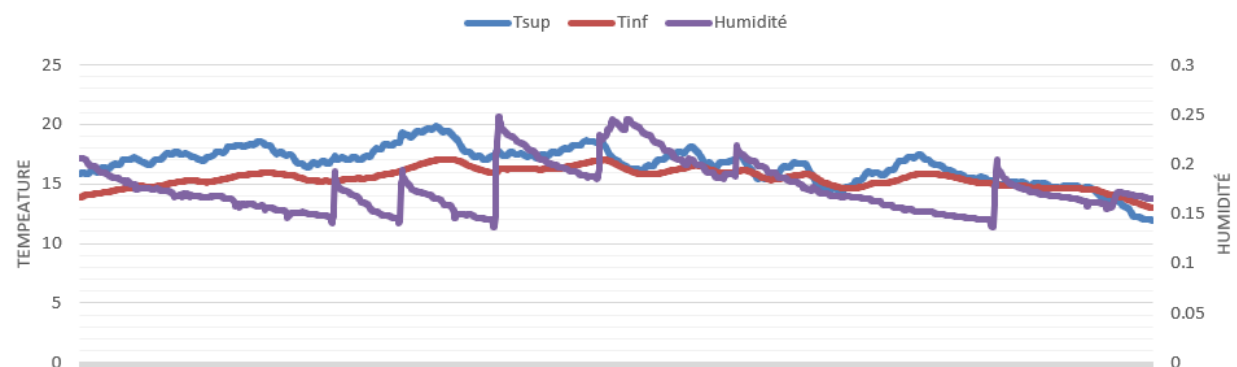


Figure 33: Analyse de l'horizon 20-40cm

Les valeurs de température sont plus stables que dans la couche superficielle et l'écart Tsup–Tinf reste modéré, montrant une atténuation des variations atmosphériques en profondeur. Les

fluctuations suivent la tendance générale de l'horizon supérieur mais avec un délai et une amplitude réduite. On note des pics d'humidité moins prononcés qu'en surface. Ces hausses correspondent à l'infiltration de l'eau depuis la couche 0–20 cm. La décroissance est plus progressive que dans la couche superficielle ce qui confirme les propriétés sableuses

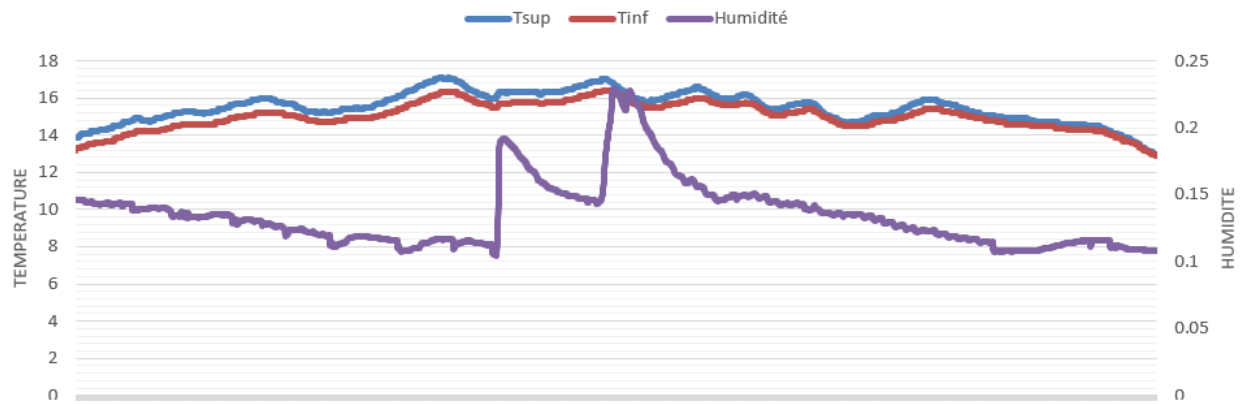


Figure 34: Analyse de l'horizon 40-55cm

Les variations de température sont beaucoup plus amorties, avec des courbes très proches. Cela traduit un effet tampon thermique dû à la profondeur et à la texture argileuse qui conserve mieux la chaleur et amortit les variations externes. L'humidité est plus élevée au départ mais diminue de façon régulière. Les pics observés sont plus atténués : l'eau met plus de temps à atteindre cette profondeur et à se redistribuer. Une fois infiltrée, l'eau est retenue plus longtemps par les particules argileuses, ce qui explique la lenteur du drainage et une décroissance moins brutale.

4.2 Résultats des modèles LSTM

Tableau 6: Tableau comparatif des résultats des modèles LSTM

Modèle	R ²	RMSE (m^3/m^3)	Biais
LSTM24	0.8873	0.00918	-0.0018
LSTM48	0.954	0.00464	-0.0018
LSTM72	0.9234	0.00687	-0.00059
LSTM96	0.9077	0.00972	0.00028

Les performances comparatives des modèles LSTM montrent que le modèle LSTM48 offre les meilleurs résultats globaux, avec un R² de 0,954 et un RMSE faible de 0,00464 m^3/m^3 . Ce modèle

présente également un biais très faible de -0,0018. Ces résultats montrent que ce modèle a la meilleure performance sur l'ensemble de test. Les autres modèles affichent des performances inférieures. LSTM72 conserve une bonne précision $R^2 = 0.9234$, et les LSTM24, LSTM96 montrent des erreurs légèrement plus élevées. Dans l'ensemble, les quatre modèles démontrent une bonne capacité de généralisation, mais la fenêtre temporelle de 48 heures semble la mieux adaptée pour capturer les tendances du jeu de données.

Les nuages de points représentant les résultats de chaque modèle sont présentés en annexe A.

4.3 Résultats des modèles CNN

Tableau 7: Tableau comparatif des résultats des modèles CNN

Modèle	R^2	RMSE (m^3/m^3)	Biais
CNN24	0.8857	0.012	-0.0032
CNN48	0.908	0.007	-0.00138
CNN72	0.954	0.0039	0.0008
CNN96	0.9602	0.00384	0.00064

Les performances comparatives entre les modèles CNN sur différentes fenêtres temporelles montrent que le modèle CNN96 a les meilleurs résultats avec un R^2 de 0,9602 et un RMSE de 0,00384 m^3/m^3 . Ce modèle affiche aussi un biais très faible de 0,00064, indiquant une excellente stabilité et une bonne précision de prédiction. Le modèle CNN72 obtient des résultats très proches, avec un R^2 de 0,954 et un RMSE de 0,0039 m^3/m^3 . Les modèles CNN24 et CNN48 montrent des performances inférieures, avec des R^2 plus faibles et des erreurs plus élevées.

Les nuages de points représentant les résultats de chaque modèle CNN sont présentés dans l'annexe B.

4.4 Comparaison des modèles CNN et LSTM

La comparaison entre les modèles a été réalisée à partir de métriques de performance déterministes (R^2 , RMSE, MAE et biais) calculées sur un ensemble de test. L'objectif principal était d'évaluer et de comparer le comportement prédictif des modèles dans un cadre opérationnel, plutôt que de réaliser une inférence statistique formelle sur leurs performances. Dans le cadre de

cette étude, aucun test de significativité statistique ni intervalle de confiance n'a été calculé. Les différences de performance présentées doivent donc être interprétées comme des écarts observés sur le jeu de données étudié, et non comme des conclusions statistiques. Cependant, la cohérence des résultats entre les différentes métriques et les différentes tailles de fenêtres temporelles permet de dégager des tendances robustes quant au comportement relatif des modèles. Une analyse statistique plus approfondie serait utile dans des travaux futurs.

4.5 Courbes d'apprentissage

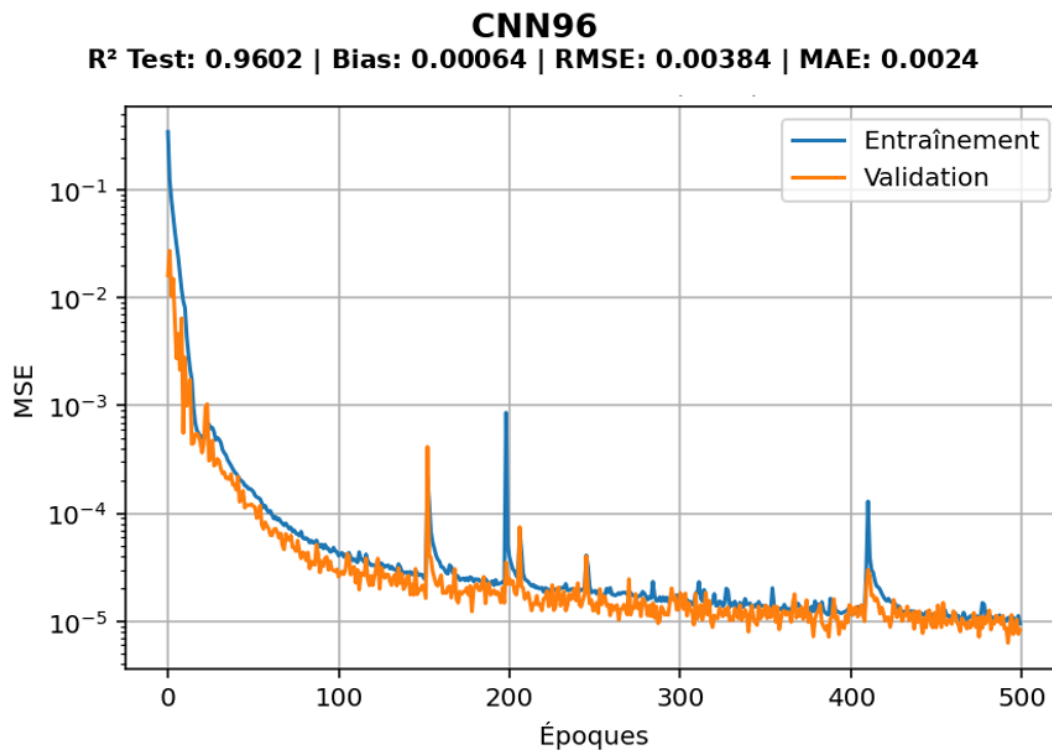


Figure 35: Courbe de validation CNN96

La figure montre l'évolution de la MSE pour l'entraînement et la validation sur 500 époques. Les deux courbes suivent une tendance décroissante rapide lors des premières itérations. Dès environ 100 époques, les courbes atteignent une zone de stabilité avec des valeurs de MSE proches de 10^{-4} puis descendent vers 10^{-5} , ce qui indique un ajustement très précis. Quelques pics apparaissent à cause de l'utilisation de données réelles, mais la courbe de validation continue de suivre celle de l'entraînement. Le modèle atteint une très haute précision avec un R^2 de 0,9602, un biais de 0,00064 et des erreurs très faibles, RMSE de 0,00384 et MAE de 0,0024.

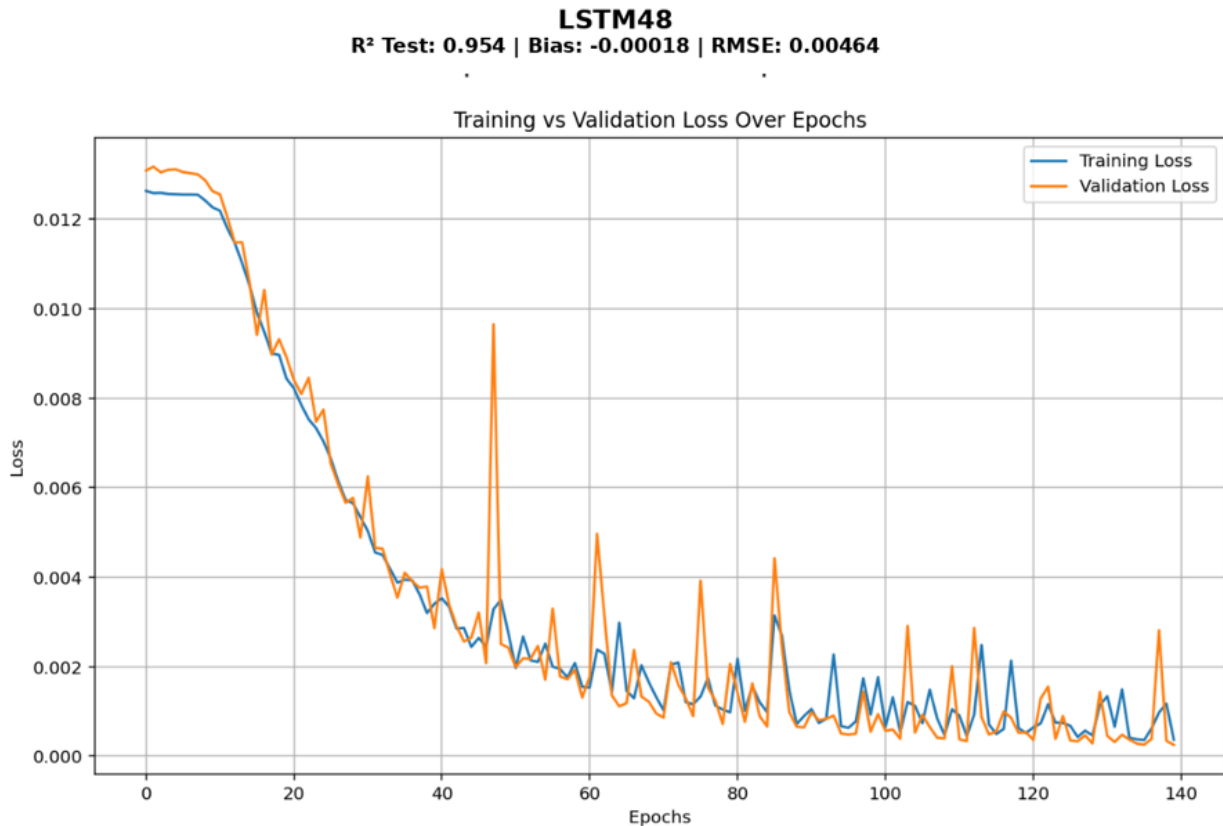


Figure 36: Courbe de validation LSTM48

Cette figure présente la perte pour l'entraînement et la validation sur environ 140 époques. Contrairement au CNN, le LSTM met plus de temps à converger. On peut observer plusieurs piques, notamment à cause des données réelles mentionnées, mais la courbe de validation suit celle de l'entraînement, ce qui témoigne d'un bon équilibre et d'une absence d'overfitting, et le modèle finit par converger. L'instabilité de la courbe peut aussi être liée à la nature séquentielle des données et à la sensibilité du modèle LSTM aux gradients. Malgré cela, le modèle atteint des performances remarquables, avec un R^2 de 0,954, un biais de -0,00018 et un RMSE de 0,00464.

Les deux modèles affichent une excellente capacité de prédiction de l'humidité du sol. Le CNN96 se distingue par une convergence plus rapide et plus stable sur un nombre élevé d'époques, tandis que le LSTM48 est plus sensible et montre des oscillations dans la validation. Le CNN96 est le plus performant en termes de précision et de stabilité, mais le LSTM48 se situe juste en dessous mais reste pertinent pour des séquences plus courtes.

4.6 Évaluation des performances des modèles

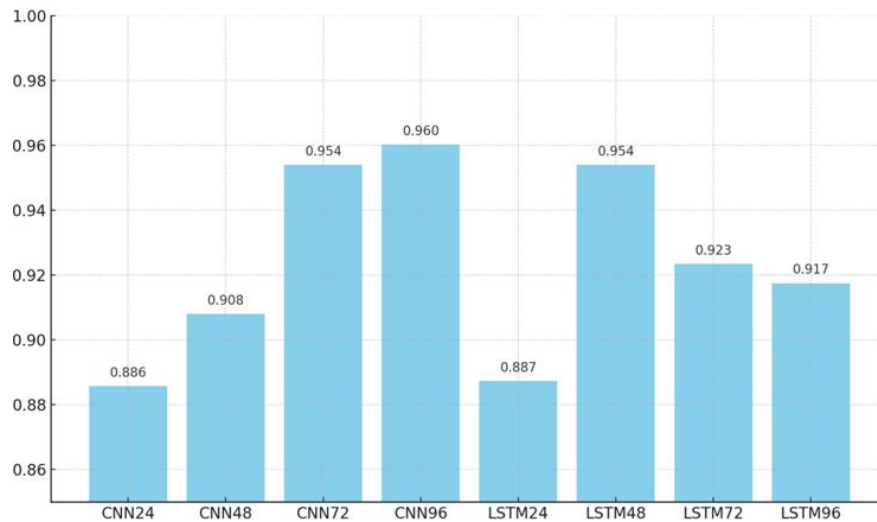


Figure 37: Comparaison des coefficients de détermination des différentes versions des modèles d'AP étudiés

Cette figure compare les valeurs de R^2 obtenues par les modèles CNN et LSTM pour différentes fenêtres temporelles. Les résultats montrent une progression de la performance des CNN avec l'augmentation de la séquence temporelle, atteignant un maximum de 0,960 pour CNN96, tandis que les LSTM atteignent leur meilleur niveau avec LSTM48 à 0,954 avant de perdre en précision pour les séquences plus longues.

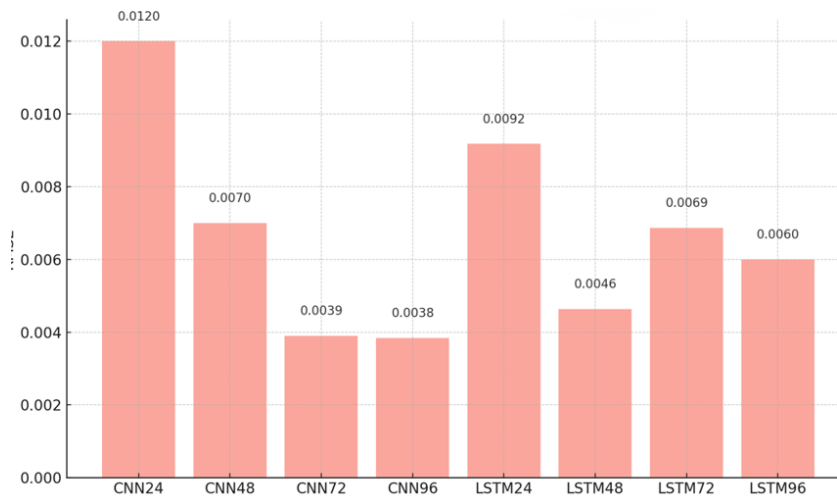


Figure 38: Comparaison des RMSE des différentes versions des modèles d'AP étudiés

Cette figure montre les valeurs de RMSE obtenues pour chaque modèle. Les CNN affichent une amélioration avec l'allongement de la fenêtre temporelle, passant de 0,012 pour CNN24 à seulement 0,0038 pour CNN96. Les LSTM présentent une tendance différente, avec une

performance optimale pour LSTM48 à 0,0046, puis une dégradation pour LSTM72 et LSTM96 qui atteignent respectivement 0,0069 et 0,0060.

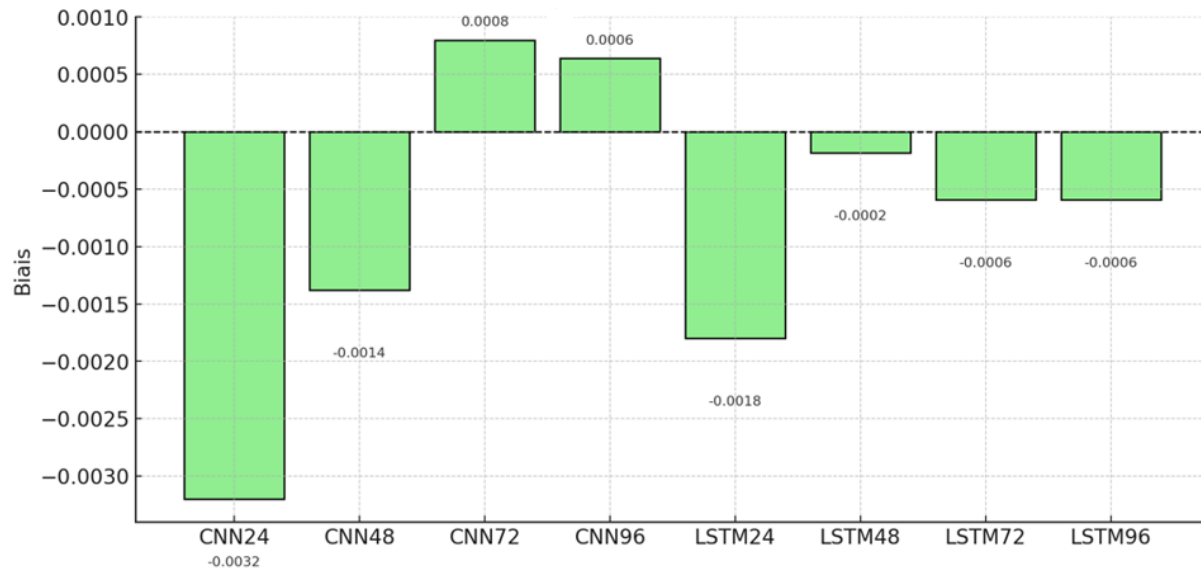


Figure 39: Comparaison des biais des différentes versions des modèles d'AP étudiés

Cette figure présente les biais obtenus pour les différents modèles. On observe que CNN24 et LSTM24 présentent les valeurs négatives les plus marquées, respectivement -0,0032 et -0,0018. CNN48 et les modèles LSTM72 et LSTM96 affichent également un biais négatif mais de faible amplitude qui indique une faible sous-estimation. Les CNN72 et CNN96 présentent un biais positif très faible, respectivement 0,0008 et 0,0006, indiquant une légère surestimation. LSTM48 se distingue par un biais à -0,0002, ce qui montre un excellent équilibre.

4.7 Comparaison avec les modèles de base

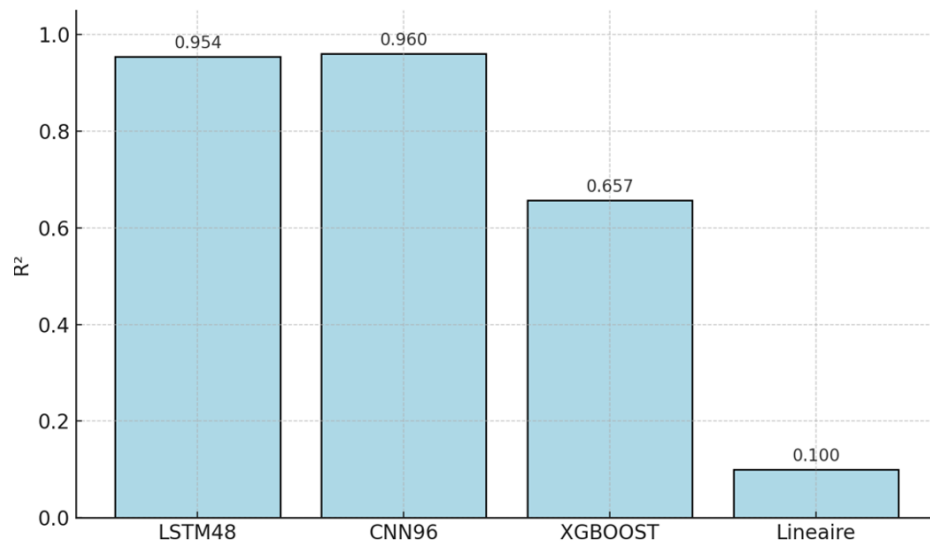


Figure 40: Comparaison des coefficients de détermination des meilleures versions des modèles d'AP étudiés et les modèles de bases

Ce graphique compare les valeurs de R^2 obtenues par les modèles qu'on a créés et des modèles de base. Les approches d'apprentissage profond dominent avec un R^2 de 0,96 pour CNN96 et 0,954 pour LSTM48. Le modèle XGBoost atteint 0,657, ce qui reste correct. La régression linéaire ne dépasse pas 0,100, révélant son incapacité à capturer la complexité des relations entre flux thermiques et humidité du sol. Cette comparaison met en évidence la supériorité des modèles profonds pour cette tâche.

4.8 Discussion

L'objectif principal de cette étude était d'estimer la teneur en eau du sol à partir des flux thermiques, en exploitant les variations naturelles de température comme indicateur indirect de l'humidité. L'analyse des profils thermiques et hydriques du sol montre des comportements cohérents avec les propriétés physiques attendues. Les couches superficielles réagissent rapidement aux variations atmosphériques alors que les horizons plus profonds présentent une réponse thermique plus lente et une humidité plus stable. Ces observations confirment la diminution de l'influence climatique avec la profondeur et le rôle des textures : la couche superficielle se vide d'eau et se réchauffe rapidement, les horizons sableux favorisent l'infiltration et les horizons argileux assurent un effet tampon thermique et hydrique. Ces résultats valident la fiabilité des mesures et la capacité des capteurs thermiques à suivre les dynamiques du sol.

D'autre part, les modèles d'apprentissage profond ont permis de reconnaître cette relation non-linéaire liant ces mesures thermiques à la teneur en eau volumique avec une bonne précision. Le modèle LSTM48 montre que la fenêtre temporelle de 48 h offre un équilibre optimal entre la capture des dépendances temporelles et la stabilité de l'apprentissage. Les modèles CNN améliorent leurs performances lorsque la fenêtre temporelle s'allonge comme le CNN96 qui atteint les meilleures métriques globales, confirmant sa capacité à extraire des relations complexes et à généraliser les tendances temporelles sur plusieurs jours. La comparaison entre CNN96 et LSTM48 révèle deux stratégies complémentaires : le LSTM excelle dans la modélisation séquentielle à court terme, tandis que le CNN capte mieux les régularités spatiales et temporelles à long terme. Les courbes d'apprentissage montrent également deux comportements de convergence différents. Les CNN convergent plus rapidement alors que les LSTM convergent progressivement mais plus lentement. Bien que le jeu de données contienne environ 24 500 séquences après augmentation, cette taille reste modeste au regard des standards usuels de l'apprentissage profond. Dans ce contexte, le choix d'architectures CNN et LSTM de complexité limitée constitue un compromis visant à exploiter les capacités de modélisation non linéaire tout en réduisant le risque de sur-apprentissage. Le nombre de couches a été volontairement limité afin de capter des dépendances temporelles de complexité croissante tout en conservant une architecture suffisamment simple pour assurer une convergence stable avec le volume de données disponible. Des architectures plus profondes ou plus complexes nécessiteraient un volume de données plus important ou des stratégies de régularisation plus poussées. Les performances obtenues indiquent que les modèles retenus offrent un niveau de complexité adapté au jeu de données disponible.

Bien que les métriques globales indiquent de très bonnes performances, certaines erreurs peuvent se concentrer dans des conditions particulières où la relation entre les variables thermiques et l'humidité devient plus difficile à modéliser. Les écarts les plus importants correspondent chronologiquement à des événements de précipitation, car l'humidité du sol augmente très rapidement alors que la réponse thermique présente un décalage temporel lié à l'infiltration. Dans ces situations, le modèle sous-estime les résultats. D'autre part, des erreurs peuvent être plus fréquentes lors de changements rapides de température car ces conditions modifient le gradient thermique et peuvent introduire des régimes transitoires où le signal thermique ne suit pas immédiatement l'évolution de l'humidité. Une analyse plus systématique pourrait être ajoutée dans des travaux futurs, en examinant la distribution des résidus aux

événements de pluie et à la variation horaire de température pour identifier précisément les régimes où les modèles sont moins performants.

Les résultats de cette étude sont cohérents avec ceux obtenus dans des travaux antérieurs sur la relation entre les propriétés thermiques et l'humidité du sol. Plusieurs études ont montré que la conductivité thermique augmente proportionnellement avec la teneur en eau : plus le sol est humide, plus il conduit efficacement la chaleur (Campbell & Norman, 2019). Les gradients thermiques mesurés ici confirment cette relation, les variations de température suivant directement les fluctuations hydriques dans les horizons observés. Les performances obtenues avec les architectures CNN et LSTM s'accordent avec les études récentes démontrant leur supériorité sur les modèles statistiques classiques. (Farhangmehr, et al., 2023) ont montré la pertinence des CNN pour la prévision horaire de la température du sol et (Li, et al., 2022) ont prouvé l'efficacité des LSTM pour la prédiction de l'humidité à partir de plusieurs variables. Les résultats confirment aussi les observations expérimentales de (Tarara & Ham, 1997), qui ont établi une relation entre la conductivité thermique et la teneur en eau du sol à l'aide de capteurs à double sonde. Cette étude s'est basée sur le concept de pic de chaleur introduit dans leur recherche. Leur approche exigeait des conditions contrôlées et des pics de chaleur constants, alors que l'approche proposée ici se distingue en mesurant les températures pour ensuite calculer une approximation de flux de chaleur T_{flux} dépendante de la profondeur entre les températures mesurées. Cette mesure est introduite comme variables d'entrée pour les modèles profonds, sans chauffage artificiel. Cette méthode permet de suivre la dynamique temporelle réelle du sol sur le long terme et démontre encore une fois que la température est un indicateur actif de l'humidité. Les mesures de température à différentes profondeurs permettent d'obtenir une estimation fiable de la teneur en eau et de relier la chaleur circulant dans le sol à la quantité d'eau présente. Cette étude se distingue aussi en comparant deux architectures d'apprentissage profond (CNN et LSTM) selon plusieurs fenêtres temporelles (24, 48, 72 et 96). Un autre point original réside dans la base de données, issue de plusieurs horizons et de quatre types de sols différents. Les résultats suggèrent que la méthode semble être agnostique au type de sol, bien que cette hypothèse doive être confirmée dans de futures recherches. Cette étude se focalise principalement sur des sols situés le long de l'axe sableux du triangle textural (Figure 21), ce qui limite la représentativité de l'ensemble des textures et conditions pédologiques possibles. Dans des contextes différents, notamment pour des sols limoneux ou argileux, les propriétés thermiques et hydriques peuvent suivre des régimes distincts, conduisant à des combinaisons de températures et d'humidité qui ne sont pas couvertes par le jeu de données utilisé. Les

performances des modèles pourraient ainsi être dégradées lorsqu'ils sont appliqués à ces textures non représentées. L'élargissement de la base de données à d'autres régions et types de sols, en particulier le long des axes limoneux et argileux du triangle textural, constitue donc une étape nécessaire pour tester la capacité de généralisation des modèles et valider la méthodologie à une échelle plus large. De plus, les données analysées concernent uniquement la saison estivale sans prendre en compte les autres saisons. On pourra aussi avoir recours à des architectures hybrides CNN-LSTM combinant les forces des deux modèles. Une application de cette idée pourrait être la création d'un système automatisé de suivi en temps réel de l'humidité du sol, utilisable dans les domaines agricoles et environnementaux.

Les modèles physiques classiques, tels que l'équation de Richards et les modèles de bilan hydrique comme SNOSWAB, ont été présentés dans ce mémoire afin d'établir le cadre physique reliant les propriétés thermiques du sol à sa dynamique hydrique. Dans cette étude, la comparaison des performances a été réalisée entre les modèles d'apprentissage profond et des modèles de base statistiques utilisant les mêmes données d'entrée, dans le but d'évaluer le gain apporté par l'apprentissage profond à partir de mesures thermiques locales. Les modèles physiques reposent sur des hypothèses, des paramètres et des forçages différents, et ne sont donc pas utilisés comme modèles comparatifs directs dans cette étude.

Cette étude se limite volontairement à la saison estivale. Ce choix repose sur l'une des hypothèses fondamentales du projet, selon laquelle, sur un horizon de quelques mois, les phases solides du sol être considérées comme invariantes, la seule composante variable étant la phase liquide. Dans ce cadre, la période estivale permet de se concentrer sur des variations hydriques actives et directement liées aux flux thermiques, tout en évitant les changements structuraux ou physiques propres aux saisons froides. Dans d'autres saisons, et en particulier en période hivernale, des phénomènes supplémentaires comme le gel-dégel et la présence de neige peuvent altérer la relation entre température et humidité. Ces conditions pourraient générer des régimes thermiques et hydriques différents de ceux observés en été, et conduire à des combinaisons de variables que les modèles n'ont pas rencontrées durant l'entraînement. L'intégration de données couvrant l'ensemble des saisons et des événements extrêmes constitue ainsi une perspective importante pour évaluer la robustesse des modèles et améliorer leur capacité de généralisation.

5 CONCLUSION ET RECOMMANDATIONS

L'objectif principal de cette étude était de démontrer la faisabilité et la précision de l'estimation de l'humidité du sol à partir des flux thermiques mesurés par des capteurs, en utilisant des modèles d'apprentissage profond. Les résultats obtenus montrent de manière claire et robuste que l'intégration de réseaux de neurones, notamment les architectures CNN et LSTM, permet d'atteindre des niveaux de performance nettement supérieurs à ceux des méthodes classiques ou des modèles de base comme la régression linéaire, la persistance ou encore XGBoost.

Les expérimentations ont été menées sur deux sites distincts au Québec : le parc technologique de l'INRS (campagnes 2023–2024–2025) et le champ 41 de la Route du Pont Rouge à Sainte-Catherine (campagnes 2020–2021). Ces sites, caractérisés par des horizons de sol variés, ont permis d'évaluer la transférabilité des modèles et leur capacité à capter la complexité des dynamiques hydriques et thermiques. Les mesures ont été effectuées grâce à des capteurs 5TM reliés à un enregistreur Em50G, avec une calibration préalable par méthode gravimétrique afin de garantir la fiabilité des mesures de référence. Un profileur de température a aussi été installé pour mesure en continue la température du sol chaque 5 cm.

La structuration des données, incluant la moyenne horaire et l'augmentation artificielle par variation de profondeur, a permis de porter la base de données à environ 24 500 séquences temporelles, ce qui a renforcé la robustesse des modèles. Quatre horizons temporels (24h, 48h, 72h et 96h) ont été explorés pour les deux architectures principales : CNN et LSTM. Les résultats montrent que les CNN surpassent légèrement les LSTM en termes de précision et de stabilité, particulièrement pour des séquences prolongées. Le modèle CNN96 a atteint un coefficient de détermination $R^2 = 0,9602$, un biais négligeable et un RMSE de $0,00384 \text{ m}^3/\text{m}^3$, ce qui témoigne d'une prédiction extrêmement fiable. Du côté des LSTM, le modèle LSTM48 a obtenu des performances proches, avec un $R^2 = 0,954$ et un RMSE de $0,00464$, confirmant sa pertinence mais révélant une convergence plus instable et une sensibilité accrue aux oscillations dans la validation.

Ces résultats mettent en évidence plusieurs points forts. Les flux thermiques se révèlent être un indicateur pertinent et exploitable pour la prédiction de l'humidité du sol. L'apprentissage profond permet de capturer les non-linéarités complexes inhérentes aux interactions sol-eau-température, interactions difficilement modélisables par des approches déterministes ou empiriques.

En définitive, la combinaison de méthodes de prétraitement et de techniques modernes d'optimisation a permis de faire la meilleure partition des données disponibles.

Il faut aussi noter que l'étude s'est restreinte à deux sites et à quatre types de sols. Si les résultats sont prometteurs, la capacité de généralisation à d'autres textures ou à des contextes climatiques différents reste à démontrer.

Plusieurs recommandations peuvent être formulées pour orienter les travaux futurs et maximiser l'impact des modèles développés. Premièrement, la diversification des types de sols. Les modèles actuels ont été entraînés et validés sur un nombre limité de textures. L'intégration de différents types de sols dans la base de données est indispensable pour renforcer la robustesse des modèles et réduire les biais liés aux spécificités du sol étudié.

D'autre part, une collecte de données sur une période plus longue et sur des sites plus variés est aussi recommandée. L'ajout de campagnes hivernales ou printanière, permettrait de tester le comportement des modèles dans des conditions de gel/dégel.

Bien que les deux approches testées soient performantes, les CNN se distinguent par une rapidité d'entraînement, une bonne stabilité et une meilleure précision. Leur utilisation est donc recommandée en priorité pour des applications similaires. Les LSTM restent utiles mais leur utilisation doit être optimisée afin de réduire leur instabilité. On peut aussi essayer une hybridation des modèles. En combinant les avantages des CNN et des LSTM dans des architectures hybrides.

Au terme de cette étude, les résultats obtenus ouvrent la voie à de multiples applications pratiques. En agriculture, ils pourraient servir à l'optimisation de l'irrigation de précision, réduisant la consommation d'eau tout en maintenant les rendements. Dans le domaine de la gestion hydrologique, ces modèles peuvent contribuer à la prévision du ruissellement et à l'amélioration des modèles climatiques.

En résumé, cette recherche démontre le potentiel considérable des approches d'apprentissage profond pour l'estimation de l'humidité du sol à partir de mesures thermiques. Le CNN96 et le LSTM48 sont les meilleures configurations testées avec un léger avantage pour les CNN. Pour que ces approches deviennent des outils universellement applicables, il sera essentiel de suivre les recommandations faites dans ce mémoire.

6 BIBLIOGRAPHIE

- Abada, M., Abubakar, M., & Bilal, M. (2022). Deep Learning Models and Their Applications: A Comprehensive Review. *Journal of Computer Science and Technology*.
- Abada, R., Abubakar, A., & Bilal, M. (2022). An Overview on Deep Learning Application of Big Data. (31-35).
- Agustirandi, B., Inayah, I., Siti Aminah, N., & Budiman, M. (2024). Analysis comparison, calibration, and application of low-cost soil moisture in smart agriculture based on internet of things. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*.
- Albawi, S., Mohamed, T., & Al-Zawi, S. (2017). Understanding of Convolutional Neural Network (CNN). *International Conference on Engineering and Technology*.
- Al-Mahmoud, M. (2019). Determination of water in soil by Karl Fischer titration. *Journal of Analytical Chemistry*.
- Bank, D., Koenigstein, N., & Giryas, R. (2023). Autoencoders. *arXiv preprint*.
- Barrett, B., & Petropoulos, G. (2012). Satellite Remote Sensing of Surface Soil Moisture. Dans B. Barrett, *Remote Sensing of Energy Fluxes and Soil Moisture Content*. Taylor & Francis.
- Beauchamp, J. (2006). *L'EAU ET LE SOL*. Université Picardie.
- Benchohra, M., & Mahmoudi, S. (2020). *Conception et mise en œuvre d'un système de contrôle d'accès pour les bâtiments intelligents basé sur la reconnaissance faciale et vocale conjointe*.
- Benson, C. (1999). Time-domain reflectometry (TDR) in geotechnics: A review.
- Botchkarev, A. (2018). Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology.
- Boukhattem, L. (2008). *Thermophysical characterization of construction materials and modeling of coupled heat transfer through a building*.
- Bourdois, L. (2019). *Les RNN, les LSTM, les GRU et ELMO*. Récupéré sur <https://lbourdois.github.io/blog/nlp/RNN-LSTM-GRU-ELMO/>
- Brown, S., Biswas, A., Caron, J., Dyck, M., & Si, B. (2021). La physique du sol. Dans *Introduction à la science du sol : de la théorie à la pratique en sols canadiens*.
- Campbell, G., & Norman, J. (2019). Determination of water in soil by Karl An Introduction to Environmental Biophysicstirration. *Springer*.
- CampbellScientific. (s.d.). *5TM – Soil Moisture and Temperature Probe (brochure)*. Récupéré sur campbellsci: https://s.campbellsci.com/documents/ca/product-brochures/5tm_br.pdf
- CampbellScientific. (s.d.). *CS230-L Temperature Probe [Fiche produit]*. Récupéré sur

- <https://www.campbellsci.com/cs230-l>
- Castellvi, F., & Gonzalez Dugo, M. (2021). A one – source model to estimate sensible heat flux in agricultural landscapes.
- CaTeC. (s.d.). *CaTeC – Climate measuring instrumentation*. Récupéré sur https://www.catec.nl/uploads/pdf/DEC-EM50-folder_756.pdf
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System.
- Cuny, M. (2017). *Etude de l'impact des conditions géologiques et climatiques sur l'efficacité énergétique des systèmes géothermiques de surface*.
- Dane, J., & Hopmans, J. (2002). Water retention and storage. *Methods of Soil Analysis*.
- Danielscu, S. (2025). Development and Application of the Snow, Soil Water and Water Balance Model (SNOSWAB), an Online Model for Daily Estimation of Snowpack Processes, Soil Water Content and Soil Water Balance. *Water (Switzerland)*.
- DecagonDevices. (2007). *EM50/EM50R/EM50G Operator's Manual [PDF]*. Récupéré sur <https://d3pcsg2wj9izr.cloudfront.net/files/24799/download/175891/29.Em50-Manual-R7.pdf>
- Dewi, C., Chen, R.-C., Yu, H., & Liu, Y.-T. (2021). Various Generative Adversarial Networks Model for Synthetic Prohibitory Sign Image Generation. *11*.
- Dobriyal, P., Qureshi, A., Badola, R., & Hussain, S. (2012). A review of the methods available for estimating soil moisture and its implications for water resource management. *458-459*.
- Evelt, S., & Tolk, J. (2009). Soil water sensing and calibration. *Agronomy Journal*.
- Farhangmehr, V., Cobo, J., Mohammadian, A., Payeur, P., Shirkhani, H., & Imanian, H. (2023). A Convolutional Neural Network Model for Soil Temperature Prediction under Ordinary and Hot Weather Conditions: Comparison with a Multilayer Perceptron Model. *Sustainability*.
- Friedman, J. (2001). Greedy Function Approximation: A Gradient Boosting Machine. *Annals of Statistics*, 1189–1232.
- Geng, Q., Yan, S., Li, Q., & Zhang, C. (2024). Enhancing data-driven soil moisture modeling with physically-guided LSTM networks. *Frontiers in Forests and Global Change*.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Greacen, E. (1981). *Soil water assessment by the neutron method*. Australia: East Melbourne, Vic., Australia : CSIRO Australia.

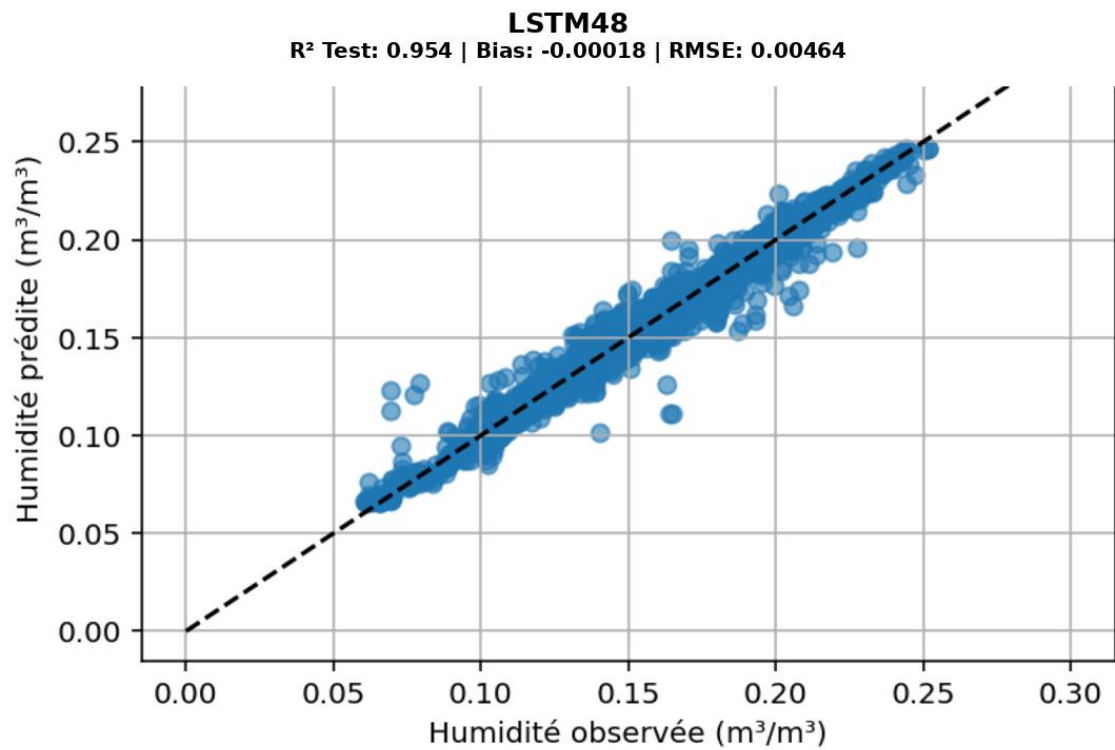
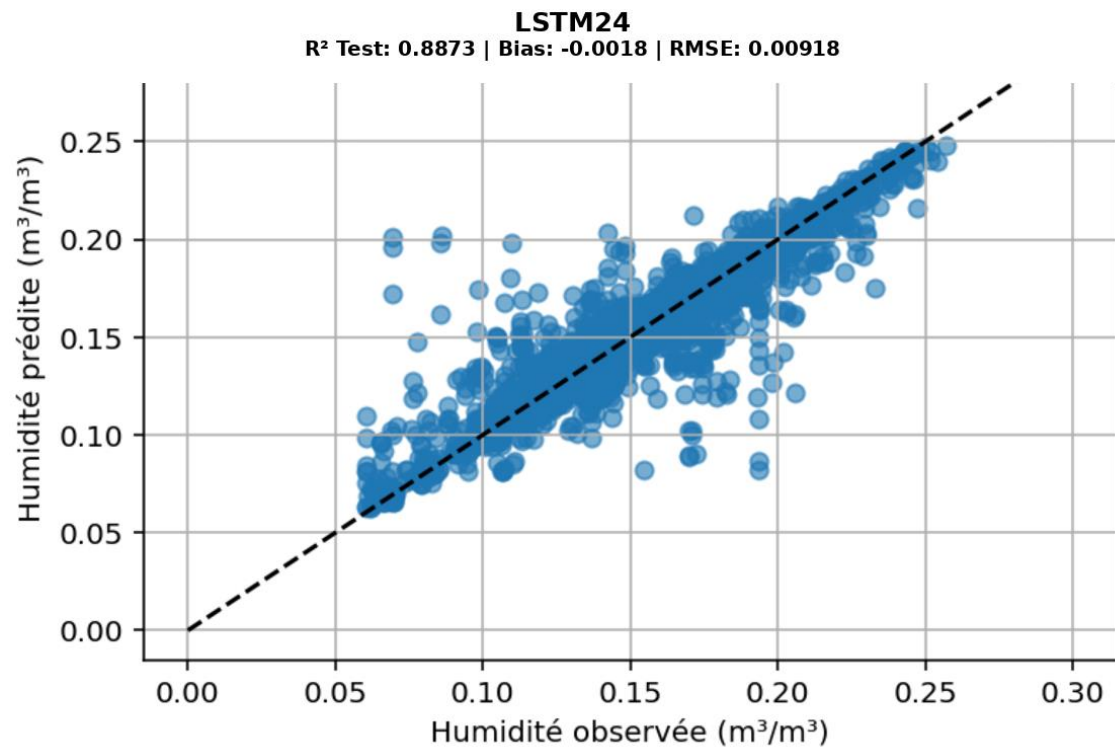
- Gu, Y., & Chen, Q. (2018). GAN-based Model for Residential Load Generation Considering Typical Consumption Patterns. *The 10th Conference on Innovative Smart Grid Technologies (ISGT 2019)*. Washington D.C.
- Gu, Y., & Chen, Q. (2018). GAN-based Model for Residential Load Generation Considering Typical Consumption Patterns. *The 10th Conference on Innovative Smart Grid Technologies (ISGT 2019)*. Washington D.C.
- Hillel, D. (2004). *Introduction to Environmental Soil Physics*. Academic Press.
- Hunduma, S., & Kebede Bunare, G. (2020). Indirect Methods of Measuring Soil Moisture Content Using Different Sensors. 37-55.
- Hyndman, R., & Athanasopoulos, G. (2018). *Forecasting: principles and practice*.
- Jakhan, D., & Kaur, I. (2020). Artificial Intelligence, machine learning and deep learning: definitions and differences. 45.
- Jakhan, J., & Kaur, T. (2020). *Deep Learning: A Comprehensive Overview on Techniques, Taxonomy, Applications and Research Directions*. Computer Science Review.
- Karras, T., Laine, S., Aittala, M., Hellsten, J., Lehtinen, J., & Aila, T. (2019). Analyzing and Improving the Image Quality of StyleGAN.
- Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *NeurIPS*.
- Le, X., Lee, G., Jung, S., & Ho, H. (2019). Application of Long Short-Term Memory (LSTM) Neural Network for Flood Forecasting. 11.
- Letafati, M., Khalaj, B., Jorswieck, E., & Behroozi, H. (2023). Learning-Based Secret Key Generation in Relay Channels Under Adversarial Attacks.
- Li, L., Dai, Y., Shangguan, W., Wei, Z., Wei, N., & Li, Q. (2022). Causality-Structured Deep Learning for Soil Moisture Predictions. *Journal of Hydrometeorology*.
- Li, X., Zhu, Y., Li, Q., Zhao, H., Zhy, J., & Zhang, C. (2023). Interpretable spatio-temporal modeling for soil temperature prediction. *Frontiers in Forests and Global Change*.
- Liew, C., Krishnaswamy, P., Cheng, L., & Tan, C. (2019). Artificial Intelligence and Radiology in Singapore: Championing a New Age of Augmented Imaging for Unsurpassed Patient CareArtificial Intelligence and Radiology in Singapore: Championing a New Age of Augmented Imaging for Unsurpassed Patient Care. 48.
- Lindemann, B., Müller, T., Vietz, H., Jazdi, N., & Weyrich, M. (2021). A survey on long short-term memory networks for time series prediction. 99.
- Multiphysics®, C. (2023). *COMSOL Multiphysics® User's Guide*. Récupéré sur COMSOL AB.
- Munoz-Carperna, R., Shukla, S., & Shukla, K. (2004). *Field Devices for Monitoring Soil Water*

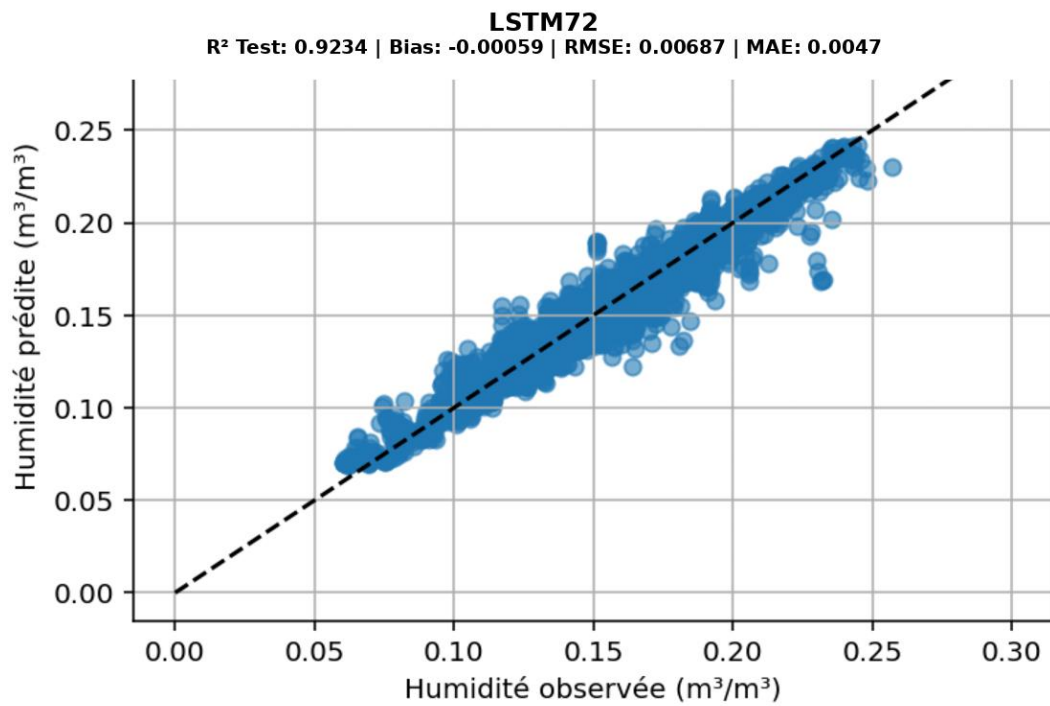
- Content.* Department of Agricultural and Biological Engineering, University of Florida.
- Ni, J., Cheng, Y., Wang, Q., Ng, C., & Garg, A. (2019). Effects of vegetation on soil temperature and water content: Field monitoring and numerical modelling. 571.
- Nwankpa, C., Ijomah, W., Gachagan, A., & Marshall, S. (2018). Activation Functions: Comparison of Trends in Practice and Research for Deep Learning.
- Oliver, S., Oliver, H., Wallace, J., & Roberts, A. (1987). Soil heat flux and temperature variation with vegetation, soil type and climate. 39(2-3).
- Osaragi, T. (2017). Classification Method by Information Loss Minimization for Visualizing Spatial Data. *International Conference on Computational Science and Its Applications*.
- Oschner, T., Horton, R., & Ren, T. (2013). Use of soil heat flux sensors in soil science. *Soil Science Society of America Journal*.
- Paolini, G., Escorihuela, M., Merlin, O., Lallet, P., Bellvert, J., & Pellarin, T. (2023). Estimating multi-scale irrigation amounts using multi-resolution soil moisture data: A data-driven approach using PrISM. *Agricultural Water Management*.
- Patil, R., Laegdsmand, M., Olesen, J., & Porter, J. (2010). Effect of soil warming and rainfall patterns on soil N cycling in Northern Europe. 139(1-2).
- Pedersen, M., Verspoor, K., Jenkinson, M., & Law, M. (2020). Artificial intelligence for clinical decision support in neurology.
- Petit, M. (2003). Développement durable à l'échelle de la planète et gestion des ressources en eau et en sols. *Comptes Rendus Geoscience*.
- Petropoulos, G., Ireland, G., & Barrett, B. (2015). Surface Soil Moisture Retrievals from Remote Sensing: Current Status, Products and Future Trends. 83-84.
- Pon, M., & Krishna, K. (2021). Hyperparameter Tuning of Deep learning Models in Keras. *Sparklinglight Transactions on Artificial Intelligence and Quantum Computing*, 36-40.
- Prosper, D. (2025). Comparative Study of CNN and Traditional Machine Learning Models in Network Security.
- Rasheed, M., Tang, J., Sarwar, A., Shah, S., Saddique, N., & Khan, M. (2022). Soil Moisture Measuring Techniques and Factors Affecting the Moisture Dynamics: A Comprehensive Review. *Sustainability*.
- Reynolds, S. (1970). The Gravimetric Method of Soil Moisture Determination.
- Richards, L. (1931). Capillary conduction of liquids through porous mediums. *Physics*.
- Richards, L. (1948). Porous plate apparatus for measuring capillary potential. *Soil Science*.
- Rogachev, A., & Melikhova, E. (2020). Automation of the process of selecting hyperparameters for artificial neural networks for processing retrospective text information. *IOP Conference*

Series: Earth and Environmental Science, 577.

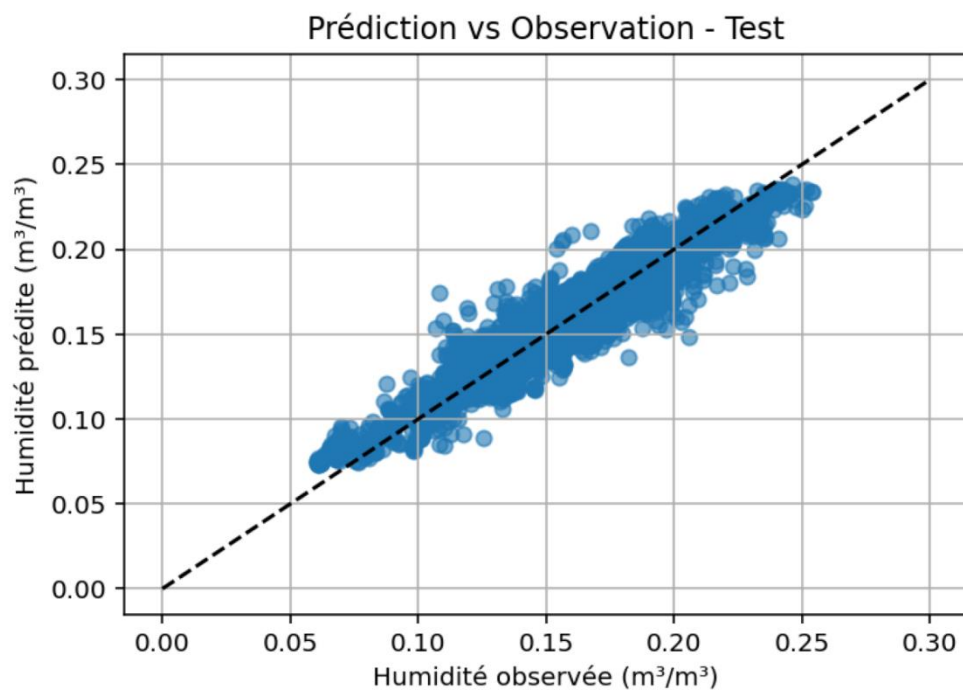
- Rohrer, P., Prieto, P., & Delais-Roussarie, E. (2022). Le rythme prosodique guide le rythme gestuel. 588-596.
- Sauer, T., & Horton, R. (2005). Soil Heat Flux.
- Senanayake, I., Yeo, I., Willgoose, G., & Hancock, G. (2021). Disaggregating satellite soil moisture products based on soil thermal inertia: A comparison of a downscaling model built at two spatial scales. *Journal of Hydrology*.
- Sepaskhah, A., & Mazaheri-Tehrani, M. (2024). A sigmoidal model for predicting soil thermal conductivity-water content function in room temperature. *Scientific Reports*.
- Sharma, P., Srivastava, H., & Kumar, D. (2018). Assessment of Different Methods for Soil Moisture Estimation: A Review. *Journal of Remote Sensing & GIS*.
- Šimůnek, J., & Van Genuchten, M. (2016). Recent developments and applications of the HYDRUS computer software packages. *Vadose Zone Journal*.
- Tarara, J., & Ham, J. (1997). Measuring Soil Water Content in the Laboratory and Field with Dual-Probe Heat-Capacity Sensors. 89(4).
- Thornthwaite, C., & Mather, J. (1955). The water balance. *Publications in Climatology*.
- Topp, G., & Ferre, P. (2002). *Soil water content*. Dane & Topp, Methods of Soil Analysis (SSSA).
- Udagani, C. (2013). Gamma ray attenuation Study with Varying Moisture Content of. 2(7).
- Verstraeten, W., Veroustraete, F., & Feyen, J. (2008). Assessment of Evapotranspiration and Soil Moisture Content Across Different Scales of Observation. 8.
- Wang, Y., & Pang, Z. (2021). Heat flux measurements and thermal potential of the Garze geothermal area in the eastern Himalayan Syntaxis. 420.
- Yu, R., Si, X., & Hu, C. (2019). A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. *Neural Computation*.
- Zhou, J., Ouyang, Y., Ye, J., & Tong, M. (2022). On Building Real Time Intelligent Agricultural Commodity Trading Models. *The 8th IEEE International Conference on Big Data Computing Service and Machine Learning Applications*. California.

ANNEXE A





LSTM96
R² Test: 0.9077 | Bias: 0.00028 | RMSE: 0.00972



ANNEXE B

