

WP4 : Principaux travaux réalisés et résultats obtenus

Développement d'une plateforme Web-GIS d'aide à la décision pour le pilotage de l'irrigation en temps réel

1- Conception et modélisation de la base de données :

Pour la conception de la base de données qui sera utilisée par la plateforme de pilotage d'irrigation, nous avons commencé par la réalisation du modèle relationnel qui représente le choix fondamental pour la conception sous-jacente. Dans cette approche, les données sont organisées en tables, ce qui permet d'abstraire la façon dont les informations sont physiquement stockées sur la machine. Chaque ensemble de données à traiter est ainsi modélisé par un ensemble de tables interconnectées, reflétant les relations entre les données.

Nous avons par la suite réalisé le diagramme de classes qui joue un rôle fondamental dans l'expression de la structure statique d'un système, en détaillant les composantes essentielles sous forme de classes et en établissant des connexions significatives entre celles-ci. Chaque classe, en substance, agit comme un modèle pour un ensemble d'objets partageant des caractéristiques et des comportements communs. Parallèlement, les relations entre les classes offrent un aperçu des interactions et des interdépendances entre les objets, renforçant ainsi la compréhension globale de la manière dont ces composants travaillent ensemble pour atteindre les objectifs du système.

Dans le contexte spécifique de la plateforme de GeanTech, cette approche se traduit par la décomposition des éléments du système en sept tables distinctes : Filed, Crop, Soil, Farmer, Remote Sensing DataSource, Season et Management Practice. Chacune de ces tables correspond à une classe et encapsule un ensemble spécifique d'attributs et de comportements pertinents. Par exemple, la classe "Crop" représente l'idée générale d'une culture agricole, tandis que la classe "Season" pourrait encapsuler les caractéristiques d'une saison agricole.

Le diagramme de classes fourni dans ce contexte est représenté sur la figure suivante :

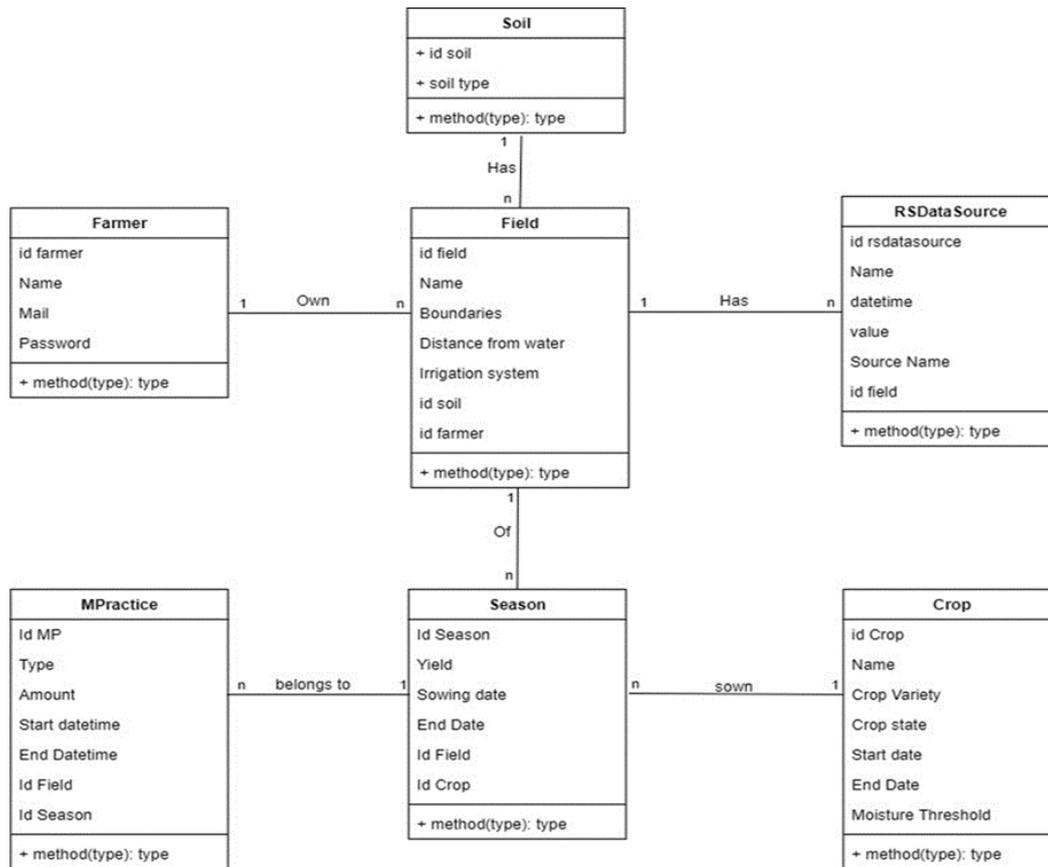


Figure 1 : Diagramme de classes de la plateforme de pilotage d'irrigation

2- Méthodologie

Notre plateforme est basée sur une approche systématique soigneusement élaborée pour renforcer la gestion efficace de l'irrigation. Cette méthodologie est constituée d'une série d'étapes cohérentes, chacune s'imbriquant harmonieusement pour former un processus global efficace, garantissant une gestion optimale de l'irrigation. Nous présentons ci-dessous ces principales étapes :

2.1. Collecte de données : Dans notre plateforme, la plupart des données collectées sont des séries temporelles (données météorologiques, données de réanalyse, données d'images satellitaires, etc.). Ce type de données est principalement caractérisé par la variété (multi-sources), le volume et la vitesse.

- a) **Stations Météorologiques In Situ :** Les stations météorologiques déployées sur le terrain sont le pilier de cette collecte de données. Ces installations physiques capturent directement les conditions météorologiques au niveau local, mesurant des paramètres cruciaux tels que la température, l'humidité, la vitesse du vent, les précipitations et le rayonnement solaire. En fournissant des données en temps réel et spécifiques à chaque emplacement, elles jouent un rôle important dans l'évaluation des modèles météorologiques actuels et la projection des conditions futures. Cette perspective permet de prendre des décisions éclairées en matière d'irrigation, garantissant une utilisation judicieuse et efficace des ressources hydriques.
- b) **Images satellitaires :** notre plateforme tire profit des données satellitaires pour obtenir une vue d'ensemble plus large et intégrée. Nous utilisons principalement les images provenant de Sentinel-2, qui offre des images détaillées de la surface terrestre et de l'atmosphère. Ces images sont exploitées pour calculer l'indice de végétation par différence normalisée (NDVI), une mesure

cruciale pour évaluer la santé et la vigueur de la végétation dans diverses régions. Ainsi, l'utilisation de données satellitaires renforce la précision de l'analyse et permet des recommandations d'irrigation plus précises et ciblées.

- c) **Images de drones** : En complément, la plateforme intégrera l'utilisation d'images capturées par des drones. Ces images aériennes à haute résolution offrent une perspective plus détaillée et localisée des champs agricoles. Les drones peuvent collecter des informations sur des paramètres tels que la santé des cultures, l'humidité du sol et la répartition des ressources en eau.
- d) **Données de réanalyse** : La plateforme intégrera des ensembles de données de réanalyse qui agrègent des données historiques de sources multiples. Des jeux de données tels que ERA5-Land et MERRA-2 sont assimilés, fournissant des informations essentielles sur les paramètres atmosphériques, la surface terrestre et les caractéristiques hydrologiques. Ces ensembles de données facilitent la compréhension des tendances climatiques, offrent une perspective historique et permettent des comparaisons avec les conditions actuelles.
- e) **Données des agriculteurs** : les utilisateurs finaux jouent un rôle clé en fournissant des données spécifiques sur leurs parcelles et leurs pratiques de gestion. Ces informations enrichissent la plateforme en lui fournissant des détails contextuels essentiels.

2.2. Gestion de l'information : Après la collecte des données, les informations sur les cultures, les sols, les parcelles, les conditions météorologiques, les pratiques agricoles et les ressources en eau sont soumises à un processus de transformation pour servir de fondement solide à l'analyse et à la prise de décision, incluant la correction de la résolution temporelle pour garantir une cohérence, ainsi que le traitement des valeurs manquantes qui pourraient obscurcir la vision globale. Une fois ces processus de révision et de correction achevés, les données sont archivées dans une base de données. Pour une meilleure gestion de ces données notre plateforme utilise une base de données orientée big data pour stocker et récupérer des données météorologiques dans un premier temps.

2.3 Traitement et de modélisation :

Une fois que les données ont été soumises à des transformations rigoureuses et ont été intégrées harmonieusement Notre plateforme intégrera un ensemble de modèles parmi lesquels on a les modèles d'estimation de l'évapotranspiration de référence et l'évapotranspiration réelle basés sur la télédétection et le machine learning et les modèles de complétion des séries temporelles climatiques, et les modèles de prévisions météorologiques à court terme. Nous présentons en détail dans la section suivante l'ensemble de ces modèles que nous avons implémenté et testé.

3- Modélisation

3.1 Estimation de l'évapotranspiration de référence par machine learning

Pour estimer l'évapotranspiration de référence en utilisant un nombre limité de paramètres météorologiques, nous avons évalué divers modèles d'apprentissage automatique pour identifier les caractéristiques les plus contributives à l'estimation de la valeur finale de l'ETO (Figure 2). L'étude a révélé que deux variables climatiques (rayonnement solaire global et température de l'air) sont les plus significatives et suffisantes pour prédire l'ETO dans notre zone d'étude, avec une importance de plus de 87% pour les deux équations de référence : FAO-56 Penman-Monteith et Hargreaves ;

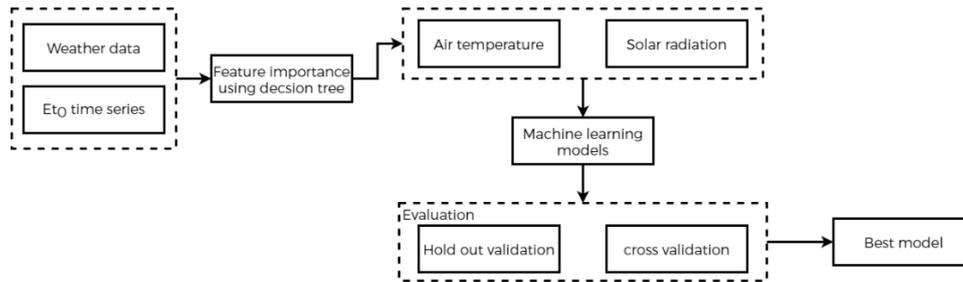


Figure 2 : Le schéma de flux pour l'estimation de l'ET0 utilisant le machine learning.

Pour évaluer la méthode, nous avons divisé notre ensemble de données en cinq parties mélangées aléatoirement. Nous avons utilisé une portion pour l'évaluation du modèle et les quatre restantes pour l'entraîner. Enfin, nous avons évalué les performances du modèle en calculant les mesures de régression sur les cinq parties de l'ensemble de données (Erreur Quadratique Moyenne RMSE et le coefficient de détermination R^2). Selon les résultats du Tableau 1, le point principal n'est pas d'obtenir des résultats parfaits en utilisant tous les paramètres comme estimateurs, mais d'utiliser uniquement la température de l'air et le rayonnement solaire global (RMSE = 0,27 et $R^2 = 0,93$), ce qui a donné des résultats prometteurs. Cela représente une approche pratique basée sur les données pour estimer avec précision l'ET0 en combinant deux critères : la précision de la méthode FAO Penman-Monteith et l'utilisation de peu de variables météorologiques et simples à obtenir (température de l'air avec le rayonnement solaire global ou seulement la température de l'air).

Tableau 1. Résultats de la validation croisée pour les trois scénarios.

Portion	Toutes les variables		Ta et Rg		Seulement Ta	
	R^2	RMSE	R^2	RMSE	R^2	RMSE
1	0.976094	0.080174	0.922928	0.258481	0.727668	0.913333
2	0.978442	0.092729	0.922629	0.332801	0.759759	1.033369
3	0.981453	0.066742	0.942489	0.206956	0.760640	0.861347
4	0.978672	0.085282	0.938638	0.245365	0.770725	0.916799
5	0.979803	0.081603	0.930292	0.281652	0.757509	0.979770

3.2. Complétion des séries temporelles climatiques (Gap-filling)

Des séries temporelles météorologiques complètes sont cruciales pour la modélisation précise des processus dans le continuum sol-plante-atmosphère. Cependant, les données météorologiques réelles mesurées sur le terrain présentent souvent des lacunes en raison de diverses raisons, notamment des dysfonctionnements d'équipement, des pannes de courant ou des erreurs de transmission de données, etc. Par conséquent, il est nécessaire de contrôler la qualité des données météorologiques avant de les utiliser pour des analyses ultérieures. Pour relever ce défi, nous avons développé un Framework en Python qui détecte automatiquement les valeurs manquantes et utilise les données de réanalyse ERA5-Land et MERRA2 ainsi que des algorithmes d'apprentissage automatique pour combler les lacunes (Figure 3). Le Framework utilise également diverses techniques axées sur les données pour identifier et corriger les valeurs aberrantes dans les données, et appliquer des contraintes sur les données.

Le Framework présente plusieurs avantages par rapport aux méthodes traditionnelles de complétion des données manquantes (Gap-filling). Tout d'abord, en exploitant seulement les données de réanalyse ERA5-Land et MERRA2, le Framework est capable de fournir des estimations précises des valeurs manquantes vue la fiabilité de ces sources. Deuxièmement, les algorithmes d'apprentissage automatique permettent de capturer des relations complexes entre les variables météorologiques et de compléter les données manquantes avec une plus grande précision. Enfin, nos méthodes de

détection et de correction des valeurs aberrantes contribuent à garantir que les données résultantes sont fiables et appropriées pour des analyses ultérieures (Figure 4).

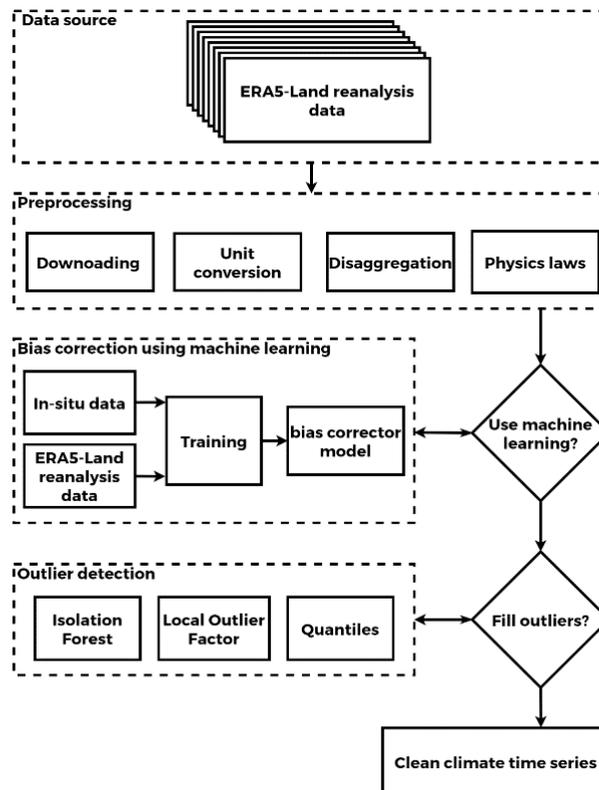


Figure 3 : Le schéma de flux de ClimateFiller.

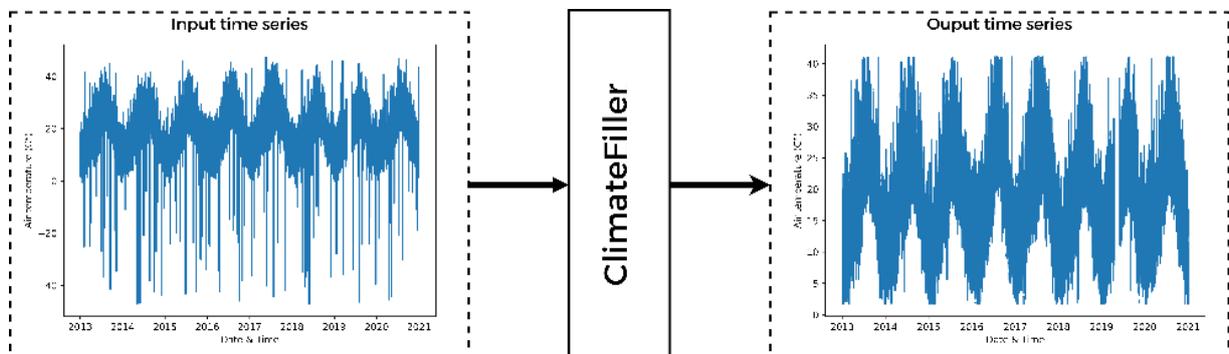


Figure 4 : Un exemple d'application de la détection des valeurs aberrantes et de la vérification des contraintes physiques sur la série temporelle de la température de l'air de la station météorologique automatique de Sidi Rahal.

Le Tableau 2 présente des statistiques sur les valeurs manquantes en nombre d'observations et en pourcentage pour les trois variables météorologiques : température de l'air (T_a), rayonnement solaire global (R_g) et humidité relative de l'air (H_r) pour les stations météorologiques de test. Le tableau met en évidence le problème des données manquantes dans les trois variables météorologiques (T_a , R_g et H_r) pour toutes les stations couvertes par notre étude. Le pourcentage de données manquantes variait de 1,86 % à 25,96 %.

Tableau 2 : Nombre et pourcentage de données manquantes de T_a , R_g et H_r pour chaque station météorologique.

Station	T_a		R_g		R_h	
	Missing data	Percentage (%)	Missing data	Percentage (%)	Missing data	Percentage (%)
Agdal	16371	5.76	13571	4.77	18891	6.65
Chichaoua	7892	2.62	7887	2.62	7886	2.62
Graoua	78270	25.96	78269	25.96	76887	25.5
Sidi Rahal	2631	1.86	5169	3.66	2626	1.86
Armed	7836	2.48	7886	2.5	8368	2.65
Oukaimden	11795	10.55	11466	10.26	11754	10.52

Pour évaluer la méthode de complétion des lacunes, les deux stations météorologiques Graoua et Sidi Rahal ont été utilisées. La première représente la station la plus touchée par les lacunes de données (25,5 %-25,96 %), tandis que la deuxième présente moins de lacunes ;

Lors de cette évaluation, nous avons d'abord généré des intervalles artificiels continus de valeurs manquantes (d'une journée, d'une semaine et d'un mois). Ensuite, les données d'origine ont été stockées, la méthode de complétion a été appliquée, puis nous avons comparé les données d'origine avec les nouvelles données complétées. Le Tableaux 3 présente les résultats de cette approche d'évaluation en termes de coefficient de corrélation (R) et d'erreur quadratique moyenne (RMSE) pour chaque variable pour les stations de Graoua et de Sidi Rahal, respectivement.

Tableau 3 : Évaluation de la méthode de complétion en utilisant les coefficients de corrélation et les valeurs RMSE pour T_a , R_g et H_r sur différents intervalles de temps pour les stations Graoua et Sidi Rahal.

	Station Graoua						Station Sidi Rahal					
	1 jour		1 semaine		1 mois		1 jour		1 semaine		1 mois	
	R	RMSE	R	RMSE	R	RMSE	R	RMSE	R	RMSE	R	RMSE
T_a	0.99	0.73 (°C)	0.96	1.54 (°C)	0.94	1.91 (°C)	0.98	0.70 (°C)	0.98	1.39 (°C)	0.97	1.58 (°C)
R_g	0.99	26.07 (W.m ⁻²)	0.96	60.10 (W.m ⁻²)	0.96	74.91 (W.m ⁻²)	0.99	46.71 (W.m ⁻²)	0.96	65.99 (W.m ⁻²)	0.97	88.60 (W.m ⁻²)
H_r	0.84	11.57 (%)	0.96	7.58 (%)	0.91	10.01 (%)	0.92	9.12 (%)	0.92	9.88 (%)	0.92	12.10 (%)

3.3. Prévisions météorologiques à court terme

Pour fournir des prévisions météorologiques qui peuvent être utilisées par divers modèles (modèle de culture, modèles hydrologiques, etc.), nous avons évalué différentes méthodes statistiques et d'apprentissage en profondeur pour cette tâche.

- **3.3.1. Modèles statistiques :** Trois principales méthodes de prévision des séries temporelles météorologiques ont été évaluées. La première méthode est ARIMA, basée sur l'autorégression et utilisant la corrélation naturelle entre les observations. La deuxième méthode est 4theta, une version modifiée de la méthode Thêta originale basée sur le lissage exponentiel. La dernière méthode est la Transformée de Fourier Rapide (FFT). Elle est dérivée des formules de la Transformée de Fourier Discrète (TFD).

- **3.3.2. Modèles d'apprentissage en profondeur :** Nous avons évalué également trois architectures de réseau neuronal de pointe dans le domaine de la prévision des séries temporelles. La première est une Mémoire à Long Terme avec Réseau de Neurones Récurrents (LSTM), une variante de réseau neuronal récurrent dont l'objectif principal est de surmonter les problèmes de gradient qui disparaît et de gradient explosif. Le deuxième modèle est l'Analyse d'Expansion de Base Neuronale pour la prévision interprétable des séries temporelles, abrégé en N-BEATS. Il utilise uniquement des blocs résiduels et des couches entièrement connectées pour réaliser des prévisions précises. Enfin, le Réseau de Neurones Convolutifs Temporels (TCNN) est une architecture basée sur le principe encodeur-décodeur, où l'encodeur est composé de couches de convolution causale empilées qui capturent les dépendances spatio-temporelles dans les données, tandis que la composante décodeur contient à son tour une variante de Réseaux de Neurones Résiduels (ResNet).

Nous avons adopté l'approche présentée sur la figure 5. Tout d'abord, les données de réanalyse ERA5-Land ont été utilisées pour imputer les valeurs manquantes dans les trois séries temporelles. Ensuite vient la modélisation, où les données propres sont fournies aux six modèles pour l'entraînement et les prévisions (1 jour, 3 jours et 1 semaine). Les résultats sont présentés dans le tableau 4.

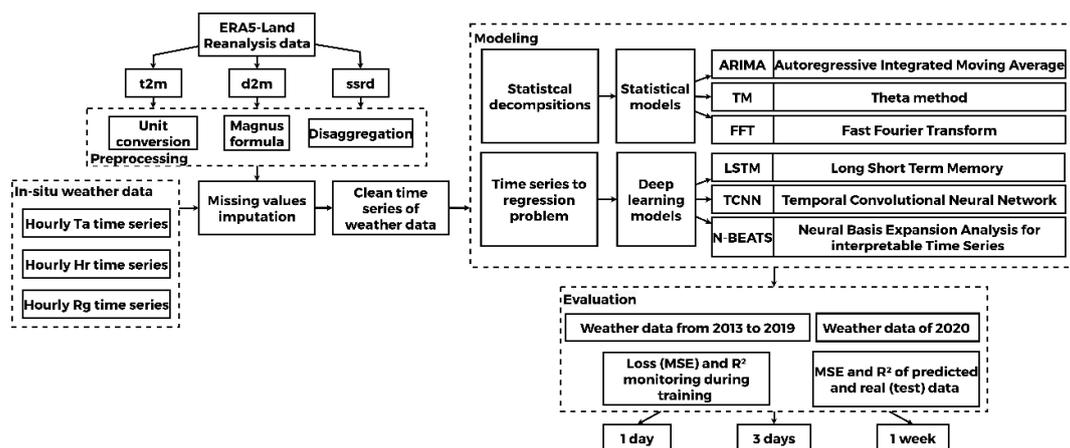


Figure 5 : Le schéma de flux pour la prévision météorologique à l'aide de méthodes statistiques et d'apprentissage en profondeur.

Tableau 4. RMSE des modèles comparés.

	Modèles de Deep learning			Modèles statistiques			
		LSTM	TCNN	N-BEATS	ARIMA	Fast Fourier Transform	Thêta Method
1 jour	Ta	1.30	0.93	1.64	5.08	1.13	6.69
	Hr	6.69	4.05	4.74	12.31	13.61	17.16
	Rg	322.14	30.38	57.27	107.05	128.93	128.76
3 jours	Ta	2.24	2.01	1.14	6.46	1.37	6.60
	Hr	9.25	7.12	6.69	16.60	11.67	15.65
	Rg	348.16	26.60	38.05	139.33	108.42	160.45
1 semaine	Ta	4.94	2.88	2.42	6.64	1.54	5.95
	Hr	14.96	13.73	15.02	23.17	14.30	16.88
	Rg	356.49	26.08	23.82	141.47	104.64	162.72

Concernant le compromis entre la complexité de la capture des séries temporelles et la précision par rapport à la période de prévision, le modèle TCNN surpasse tous les autres. Pour le meilleur modèle (TCNN), la figure 6 montre les résultats visuels des valeurs réelles et estimées de la température de l'air pour une période de prévision d'un jour (24 heures). Il y a un accord clair entre les prévisions du modèle TCNN et la réalité.

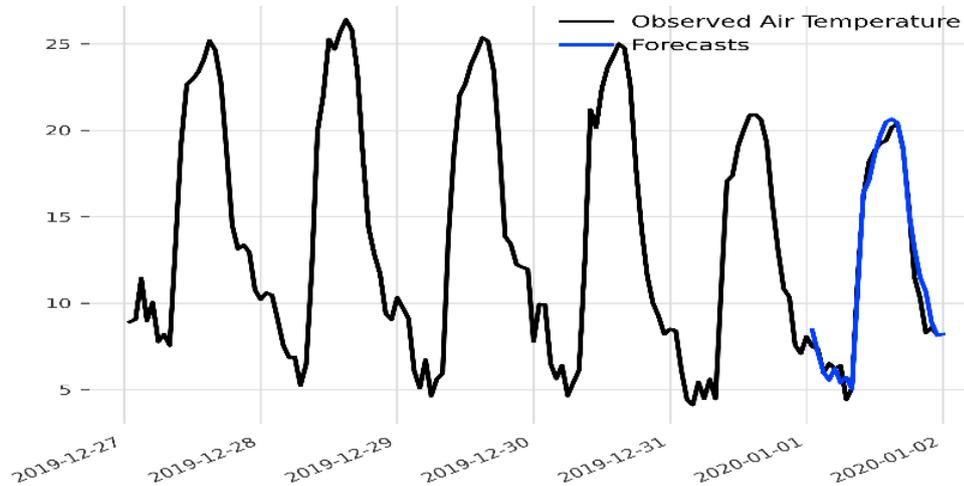


Figure 6 : Résultats visuels d'une prévision de la température de l'air sur 1 jour.

3.4. Estimation de l'évapotranspiration réel utilisant la télédétection et le machine learning

Estimer avec précision le flux de chaleur latente (LE) est crucial pour optimiser l'efficacité de l'irrigation. LE représente l'énergie utilisée par les cultures et la surface pour convertir l'eau de l'état liquide à l'état de vapeur par transpiration et évaporation. Ce flux est une composante fondamentale pour déterminer l'évapotranspiration réelle (ET_a), qui quantifie à son tour la quantité d'eau perdue et qui doit être compensée adéquatement par l'irrigation. Dans ce sens, plusieurs modèles empiriques ou physiques sont utilisés pour ce fait. Cependant, ces modèles présentent encore des faiblesses, notamment leur dépendance à beaucoup de données d'entrée, la simplification des hypothèses concernant les processus complexes et la nécessité de paramétrage et de calibration en cas de changement de site d'étude. En revanche, l'émergence de techniques basées sur les données combinées à la télédétection a montré des résultats prometteurs pour l'estimation du LE avec des données d'entrée minimales. Pour cette fin, nous avons évalué deux approches basées sur l'apprentissage automatique pour estimer le LE (Figure 7). La première utilise des données climatiques, l'indice de végétation par différence normalisée (NDVI) et la température de surface terrestre (LST), tandis que la deuxième utilise uniquement des données climatiques combinées à toutes les bandes satellitaires brutes disponibles. Les données in-situ proviennent d'une station de flux installée dans la région de Sidi Rahal (centre du Maroc), une parcelle de blé pluvial, choisi pour évaluer l'approche avant généralisation. Les données comprennent la température de l'air, le rayonnement solaire global et le LE mesuré pour la période 2015-2018. L'étude utilise Landsat 8 comme source de données de télédétection.

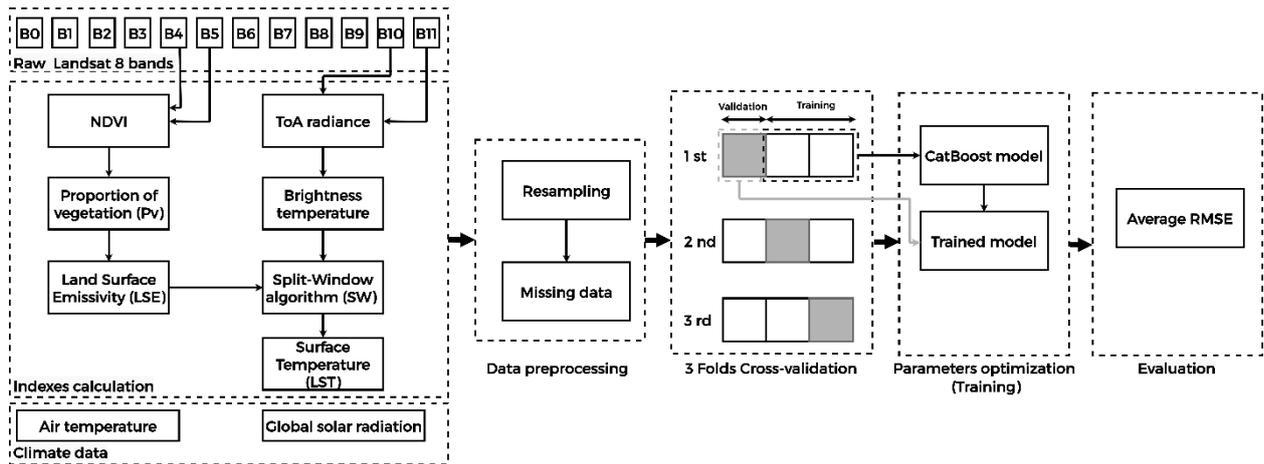


Figure 7 : Le schéma de flux pour l'estimation de l'évapotranspiration réel.

Lors de la modélisation durant l'apprentissage automatique, le modèle CatBoost est alimenté, entraîné et évalué en utilisant les deux approches de combinaison de données. La validation croisée à 3 portions a donné une RMSE moyenne de 27,54 W.m⁻² en utilisant la première approche et de 27,05 W.m⁻² en utilisant la deuxième approche (Tableau 5 et 6).

Tableau 5. Résultats en RMSE et R² du modèle utilisant les données Landsat 8 brutes.

Portion	RMSE
1	25.789654
2	28.713678
3	26.632358
Moyenne	27.045230

Tableau 6. Résultats en RMSE et R² du modèle utilisant les indices et LST.

Portion	RMSE
1	32.004610
2	26.063326
3	24.541034
Moyenne	27.536323

Comme perspectives, l'objectif est de généraliser cette approche en l'appliquant à des zones avec des conditions environnementales différentes et à la tester pour d'autres applications telles que l'estimation de l'humidité du sol et la prédiction des rendements.