

# Apprentissage Ensembliste Multicouche pour la prévision à court terme de variables agro-climatiques

**Résumé.** L'agriculture est l'un des domaines dont les activités dépendent fortement des prévisions météorologiques. En effet, dans le but d'optimiser leur production, les agriculteurs doivent pouvoir anticiper les conditions climatiques favorables ou non à leurs activités en déployant les plans d'action adéquats. Pour cela, ils consultent chaque jour les données de divers fournisseurs de prévisions météorologiques qu'ils soient payants ou gratuits. Cependant, la fiabilité des prévisions de chaque fournisseur est variable en fonction de la période, du climat ou encore de la zone géographique. Les agriculteurs doivent donc quotidiennement arbitrer entre les fournisseurs de façon subjective. Dans le but de faciliter la prise de décision des agriculteurs, cet article propose une nouvelle architecture d'apprentissage ensembliste permettant d'améliorer significativement la précision des prévisions météorologiques à court terme (échelle de 1 à 12 heures).

## 1 Introduction

La fiabilité des prévisions météorologiques est nécessaire dans de nombreux domaines dont les activités dépendent des conditions climatiques. L'agriculture est l'un de ses domaines pouvant être particulièrement impacté par les événements météorologiques tels que les températures extrêmes, le vent, la grêle, la pluie, etc. Ces événements peuvent causer des dommages conséquents et durables sur les récoltes en entraînant la perte totale ou partielle de la production. Outre les dégâts, certaines conditions climatiques peuvent aussi affecter les opérations culturales comme par exemple la possibilité de traitement limitée en cas de vent, mais aussi l'impraticabilité du sol avec les engins agricoles en cas de pluie. L'enjeu pour les agriculteurs est donc de prendre connaissance suffisamment tôt des risques climatiques à venir afin de pouvoir mettre en place des plans d'action pour minimiser les dégâts potentiels. Pour cela, ils consultent généralement plusieurs fois par jour les prévisions météorologiques. Conscients que ces dernières ont une fiabilité limitée, les agriculteurs consultent 3, 4, parfois davantage, sources de prévisions météorologiques et arbitrent entre ces sources de façon subjective. En effet, ces sources ne performant pas forcément toujours de la même manière. Certains fournisseurs de prévisions météorologiques vont parfois "sur performer" en fournissant des prévisions proches des valeurs réellement observées et d'autres "sous performer" avec des prévisions plus éloignées des valeurs réellement observées. La précision des prévisions météorologiques de chaque fournisseur peut varier selon la période, le type de climat ou encore la zone géographique. Face à la variabilité de la fiabilité des


prévisions météorologiques et aux multiples fournisseurs de prévisions dont disposent les agriculteurs, leur prise de décision n'est pas facilitée et reste de ce fait souvent imprécise.

Pour améliorer la prise de décision des agriculteurs, l'une des possibilités serait de mettre à leur disposition une unique source de prévisions plus performante que celles dont ils disposent. Cette dernière devrait permettre aux agriculteurs de ne consulter plus qu'une seule source de prévisions et ainsi de ne plus avoir à arbitrer entre les divers fournisseurs de prévisions habituels. La portée idéale des prévisions permettant aux agriculteurs de se préparer à certains événements climatiques est de 1 à 12 heures.

Dans cet article, l'objectif est de fournir des prévisions de la température et de l'humidité plus fiables que les fournisseurs de prévisions météorologiques pour les 1 à 12 heures à venir pour 2 sites agricoles distants de quelques kilomètres. Premièrement, nous établissons dans la section 2 un état de l'art des méthodes de prévision à court terme. Deuxièmement, nous présenterons dans la section 3 une nouvelle méthode d'apprentissage ensembliste composée de 4 couches distinctes permettant de réaliser des prévisions. Troisièmement, nous détaillerons le contexte et les caractéristiques du jeu de données. Et enfin, nous présenterons dans la section 5 les résultats et nous montrons que la méthode est efficace pour réaliser des prévisions fiables à court terme.

## 2 État de l'art

La météorologie, science interdisciplinaire, repose principalement sur les domaines de la physique, de la chimie et des mathématiques. L'évolution au cours du temps des variables climatiques telles que la température, le vent, la pression, etc. est principalement pronostiquée via la résolution numérique des équations de Navier-Stokes (qui décrivent le mouvement des gaz et des liquides) et de continuité de masse, mais aussi avec les lois de la thermodynamique notamment celle des gaz parfaits. La Prévision Numérique du Temps (PNT) (Bauer et al. (2015)) est une méthode pouvant prédire les probabilités des variables météorologiques en prenant en compte les équations et lois énoncées précédemment. Ces dernières sont intégrées dans divers modèles représentant le plus fidèlement possible la dynamique atmosphérique, chacun étant basé sur des conditions initiales différentes ou en utilisant des configurations de modèle et/ou des paramètres différents. Ces modèles sont ensuite alimentés par les observations météorologiques actuelles et fournissent des prévisions. Cet ensemble de prévision permet par la suite d'établir des probabilités de prévisions pour les heures et mois à venir. La PNT permet d'obtenir des prévisions fiables tout en estimant leur erreur. Cette méthode requiert cependant une très grande puissance de calcul et celle-ci ne cesse de croître avec les améliorations apportées aux systèmes de PNT. Néanmoins, d'autres méthodes de prévision moins exigeantes en terme de puissance peuvent permettre d'obtenir des prévisions tout aussi fiables. Le modèle hybride profond pour les prévisions météorologiques (Grover et al. (2015)) utilise les méthodes d'apprentissage automatique pour la prévision météorologique de 6 à 24 heures en tenant également en compte des fonctions thermodynamiques et de l'influence dynamique des lois atmosphériques sur les phénomènes météorologiques. Cette méthode propose une architecture comprenant 3 principaux composants, le premier utilise les arbres de décisions optimisés afin de construire



un ensemble de prédicteurs individuels pour chaque variable météorologique. Le second composant, dénommé interpolation spatiale, affine les inférences produites par les prédicteurs distincts en contraignant la sortie à être lisse dans l'espace et alignée avec les contraintes imposées par les lois physiques. Et enfin le troisième composant consiste en un réseau de croyances profondes (constitué de plusieurs couches de Machine de Boltzmann Restreinte Empilées) qui conduit à une préférence pour les solutions qui respectent les statistiques conjointes attendues des variables météorologiques.

Il est également possible de procéder à des prévisions météorologiques avec uniquement des techniques d'apprentissage machine. Une étude concernant la prévision des températures de 1 à 24 heures à Nashville dans l'Etat du Tennessee aux Etats-Unis (Jakaria et al. (2018)) compare les résultats de diverses méthodes d'apprentissage machine. L'historique des données météorologiques de nombreuses zones entourant la ville de Nashville et les conditions météorologiques actuelles de la ville en question sont utilisées en entrées des modèles de Régression à Vecteur de Support, Perceptron Multicouche, Régression Extra-arbre, Forêt Aléatoire, et enfin Régression Ridge. Ces modèles prouvent une certaine efficacité de prévision en présentant entre 1°C et 4°C de différence entre les températures prédites et les températures observées.

D'autres techniques d'apprentissage machine comme les réseaux de neurones récurrents (Zaytar et Amrani (2016)) et les réseaux bayésiens (Cofiño et al. (2002)) sont utilisées pour les prévisions à court terme de variables climatiques. Il est aussi possible d'utiliser des techniques de clustering comme par exemple celle des K plus proche voisin améliorée (Sharif et Burn (2007)). Cependant, 1 degré de différence entre les températures prédites et les températures observées peut être déterminant pour la prise de décision. De ce fait, notre méthode, via une architecture innovante, propose de minimiser l'écart entre les températures prédites et les températures observées.

## 3 Présentation du modèle

### 3.1 Approche Globale

L'approche présentée dans cet article se base sur des méthodes d'apprentissage machine pour la prévision à court terme de variables climatiques épargnant ainsi la modélisation complexe et exigeante des lois de la thermodynamique. Dans cette partie, l'architecture et le fonctionnement global de la méthode Ensembliste Multicouche seront présentés puis chaque composant de l'architecture et leurs paramètres seront détaillés. Dans cette partie, l'architecture et le fonctionnement global de la méthode Ensembliste Multicouche seront présentés puis chaque composant de l'architecture et leurs paramètres seront détaillés. Ces techniques permettant de répondre à des problématiques de régression et de classification, mais dans cet article nous nous focaliserons sur la problématique de régression qui permet la réalisation de prévisions.

#### 3.1.1 Introduction à l'Apprentissage Ensembliste

L'apprentissage Ensembliste consiste au fait d'entraîner un ensemble de modèles puis à les combiner dans le but d'obtenir des prévisions plus fiables que celles obtenues

avec un unique modèle (Džeroski et Ženko (2004)). De ce fait, chaque modèle de l'ensemble est nommé apprenant de base ou apprenant faible, car la méthode ensembliste permet d'améliorer ces derniers en un modèle dit d'apprentissage fort. Concernant la composition d'un ensemble, l'apprentissage Ensembliste se décompose en deux sous-parties principales (Zhou (2012)) :

**Les apprenants faibles homogènes.** consistent à utiliser un même algorithme d'apprentissage pour chaque apprenant faible de l'ensemble. Cependant, pour que ces derniers soient suffisamment différentiables, chaque algorithme de la méthode d'apprentissage homogène prend en entrée des données différentes que plusieurs techniques permettent de déterminer :

La technique de Bagging (Breiman (1996)) consiste à utiliser plusieurs échantillons des données d'entraînement en utilisant le bootstrap, c'est-à-dire un échantillonnage avec remplacement. Chacun de ces échantillons est utilisé par un seul apprenant faible. Les résultats des modèles (indépendants les uns des autres et exécutés en parallèle) sont ensuite combinés par moyenne pondérée pour créer un seul résultat. Cette technique est utilisée dans le but de réduire l'instabilité des modèles face aux modifications dans l'ensemble d'apprentissage et permet donc de réduire la variance de l'estimateur.

D'autre part, le boosting (Shapire et Freund (2013)), proposant lui aussi des prévisions en effectuant la moyenne pondérée des résultats de plusieurs modèles, diffère cependant du Bagging. En effet, chaque modèle de l'ensemble est exécuté de façon séquentielle, car chaque modèle est une version adaptative du précédent en donnant plus de poids, lors de l'estimation suivante, aux observations mal prédites. Le boosting, ayant pour objectif de réduire l'erreur de prévision, se concentre sur les observations les plus difficiles à prédire en essayant de mieux les modéliser au fil des itérations. Le boosting permet ainsi de réduire le biais.

**Les apprenants faibles hétérogènes.** consistent quant à eux, à utiliser divers algorithmes d'apprentissage se basant sur les mêmes données d'entraînement. La technique de Stacking regroupant les modèles hétérogènes se présente de la manière suivante :

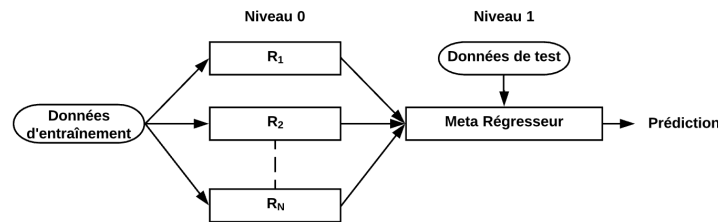


FIG. 1 : Représentation modèle d'apprentissage ensembliste empilé (Stacking).

Le stacking vise à combiner plusieurs régresseurs en utilisant un ensemble de  $N$  différents algorithmes d'apprentissage notés  $L_k$  où  $k = 1, \dots, N$ , sur un même jeu de données d'entraînement  $S$  composé des sous-ensembles,  $S_i = \{X_i, Y_i\}$ , avec les paires

de variables exogènes  $X = (x_1, \dots, x_W)$  où  $W$  représente le nombre de valeurs, et de variables endogènes  $Y = (x_{W+1}, \dots, x_{W+h})$  où  $h$  est le nombre de valeurs à prédire. Comme il est possible de voir sur la figure 1, le stacking est composé de deux principaux niveaux. Au niveau 0, un ensemble  $R_k$  de  $N$  régresseurs avec  $k = 1, \dots, N$ , est généré, où  $R_i = L_i(S)$ . Au niveau 1, un algorithme d'apprentissage nommé méta-régresseur, est chargé de prendre en entrée un nouvel ensemble de données formé par les résultats des  $N$  régresseurs du niveau 0 puis de les combiner. En d'autres termes, les régresseurs  $N$  de l'ensemble  $R_k$  génèrent les prévisions pour  $S_i$  :

$$R_k^i(x_i) = \hat{y}_i^k \quad (1)$$

Ces dernières sont ensuite combinées par la fonction  $f_j$  induite par un algorithme d'apprentissage qualifié de méta-régresseur. Le régresseur  $R_{k,j}$  où  $j = 1, \dots, h$  prend en entrée  $X$  pour prédire  $\hat{y}_{W+h}$  qui est obtenu par la fonction :

$$\hat{y}_{W+h} = f_j(R_{1,j}, \dots, R_{N,j}) \quad (2)$$

En pratique, la fonction  $f_j$  du méta-régresseur est un modèle induit par un algorithme d'apprentissage. Concernant le jeux de données, un sous-ensemble de données  $S_i$  où  $S_i \in S$  est conservé comme données de test du niveau 1, donc les données d'entraînement en entrée des  $N$  régresseurs du niveau 0,  $\forall i = 1, \dots, n$  et  $k = 1, \dots, N$ , correspondent à :

$$R_{k,i} = L_k(S - S_i) \quad (3)$$

### 3.1.2 Apprentissage Ensembliste Multicouche

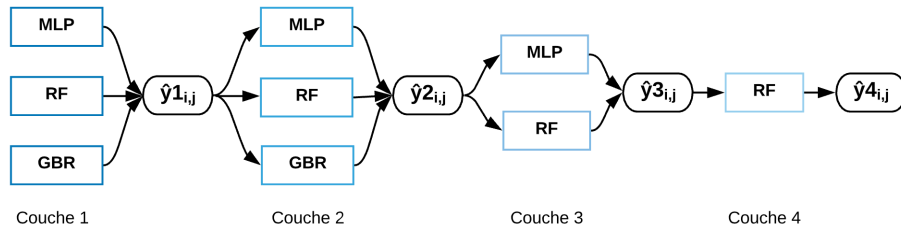


FIG. 2 : Architecture du modèle Ensembliste Multichouches.

Notre modèle d'Apprentissage Ensembliste Multichouche, représenté par la figure 2, s'inspire principalement de la technique de stacking. En effet, chaque couche du modèle est composée d'un ensemble d'apprenants faibles hétérogènes  $R_k$ , où  $k = 1, \dots, 3$ . On retrouve ainsi dans chaque couche un ensemble  $L_k$  de différents algorithmes d'apprentissage supervisés tels que le Perceptron Multicouche (MLP), Forêt Aléatoire (RF) et l'Amélioration du Gradient (GBR). Ces différents algorithmes permettent l'utilisation de régressions dont le but est d'établir des prévisions comme l'indique l'équation (1).

**Définition du format des données d'entrée.** Pour un ensemble de données  $S$ , un sous-ensemble  $S_i$  est conservé comme données de test, de ce fait chaque algorithme de chaque couche prend en entrée le complémentaire<sup>1</sup>  $S_i^c$  représentant les données d'entraînement. L'ensemble de données  $S$  est organisé sous format d'une matrice  $A = (a_{i,j})_{1 \leq i \leq m, 1 \leq j \leq n}$  de taille  $(n \times m)$  où les ensembles  $I$  et  $J$  sont finis et sont respectivement les ensembles de nombres entiers  $\{1, \dots, m\}$  et  $\{1, \dots, n\}$ . La matrice  $A$  est composée de variables exogènes  $X_{1 \leq i \leq m, 1 \leq j \leq n-1}$  et d'une variable endogène  $Y_{1 \leq i \leq m, j=n}$ .

**Couche 1.** Les trois<sup>2</sup> algorithmes d'apprentissage supervisé de la couche 1 prennent en entrée les mêmes données d'entraînement  $S_i^c$ , et apprennent chacun en parallèle une fonction de régression basée sur les variables exogènes  $X_{1 \leq i \leq m, 1 \leq j \leq n-1}$  pour prédire correctement la variable endogène  $Y_{1 \leq i \leq m, j=n}$ . Afin de tester l'efficacité de prévision des fonctions de régression apprises, les algorithmes vont prendre en entrée les variables exogènes des données de test  $S_i$  et y appliquer les fonctions de régressions apprises. Chacun de ces trois régresseurs va produire des prévisions  $\hat{y}_i$  et celles-ci vont être comparées aux variables endogènes des données de test  $S_i$  permettant ainsi de rendre compte des performances de chaque algorithme. Comme indiqué dans l'équation (1), les prévisions issues de l'ensemble des  $R_k$  régresseurs sont notées  $\hat{y}_i^k$  où dans notre cas  $k = 1, \dots, 3$  et forme la matrice  $\hat{y}_{1,i,j}$  où  $j = k$ .

**Couches 2 à 4.** Contrairement à la couche 1, les algorithmes des couches suivantes ne sont plus entraînés sur les données d'entraînements  $S_i^c$  du jeu de données initial  $S_i$ . Les algorithmes de la couche 2, par exemple, sont entraînés sur la matrice  $\hat{y}_{1,i,j}$  qui représente désormais les variables exogènes prises en entrée de chaque algorithme de la couche 2. Cependant, les variables endogènes issues des données de test  $S_i$  de la couche 1 sont conservées pour la phase d'entraînement des algorithmes de la couche 2, mais aussi pour la phase de test. Ainsi, à partir de la couche 2, chaque algorithme de chaque couche prend en entrée la matrice de prévision  $\hat{y}_{k,i,j}$  issue de la couche précédente. Ces algorithmes conservent également la variable endogène  $Y_{1 \leq i \leq m, j=n}$  des données de test, utilisée dans la couche précédente, pour leurs phases d'entraînement et de test.

**Définition du format des données de sortie.** Sachant que la quatrième couche comporte un seul algorithme  $k = 1$  et que  $j = k$  alors la prévision finale  $\hat{y}_{4,i,j}$  issue de la quatrième couche est un vecteur colonne.

Notre méthode a été conçue dans le but d'obtenir des prévisions les plus fiables possibles à court terme, c'est pourquoi elle s'inspire du fonctionnement du stacking reconnu pour améliorer les prévisions. Le principal intérêt de l'Apprentissage Ensembliste Multicouche consiste en l'amélioration progressive de la précision des prévisions au fur et à mesure de l'avancement dans les couches. En effet, à l'issue de chaque couche, l'écart

1. Un ensemble de référence  $S$  étant donné, le complémentaire du sous-ensemble  $S_i$  de  $S$  (sous-entendu relativement à  $S$ ) est l'ensemble des éléments de  $S$  qui n'appartiennent pas à  $S_i$ . Il est noté  $\mathbb{C}S_i : S_i^c = \{x \in S \mid x \notin S_i\}$ .

2. Le nombre d'algorithmes de chaque couche et le nombre de couche lui-même sont relatifs à nos données d'application (voir la section 3.2) et peuvent donc varier pour un sujet d'application différent

entre les prévisions et les valeurs réellement observées tend à diminuer. Pour cela, nous avons sélectionné 3 méthodes présentant différents avantages pour l'amélioration des prévisions.

### 3.2 Détails des méthodes

Les apprenants faibles utilisés dans les méthodes ensemblistes pour des prévisions sont variés (Ma et al. (2018)), on retrouve par exemple les algorithmes de Régression à Vecteurs de Support, Forêts Aléatoires et Réseaux de Neurones. Notre méthode ensembliste multicouche utilise trois apprenants faibles différents : les Perceptrons Multichouche (MLP), les Forêts Aléatoires (RF) et enfin l'Amélioration du Gradient en régression (GBR). Ces 3 algorithmes possèdent l'avantage de ne pas être sensibles aux changements d'échelle (dispense de l'étape de normalisation et donc de la perte du sens initial de l'information) et ne requièrent pas d'hypothèses statistiques particulières.

**Perceptron Multichouche (MLP)** fait parti de la famille des réseaux de neurones, cet algorithme est capable de fournir des prévisions précises dans la mesure où les hyperparamètres suivants sont correctement ajustés en fonction des données. Différentes valeurs pour les hyperparamètres ont été testées, celles permettant au réseaux de neurones de fournir des prévisions optimales ont été retenues :

- 3 couches de respectivement 40,15,1 neurones ;
- Régularisation des poids et du biais (0.03) ;
- Itérations (7000) ;
- Mini-Batch (10) ;
- Fonctions d'activation linéaire ;
- Algorithme d'optimisation (Adagrad).

**Forêt Aléatoire (RF)** cet algorithme est un cas particulier de bagging, en effet, il entraîne chaque modèle indépendamment puis fait la moyenne des prévisions de tous les modèles. Il a l'avantage de réduire la variance et donc de réduire la sensibilité aux petites fluctuations de l'échantillon d'apprentissage. Dans notre cas, nous utilisons l'hyperparamètre par défaut (mais qui est aussi optimal) où le nombre d'arbres est égal à 10.

**Gradient Boosting Regressor (GBR)** quant à lui s'apparente au fonctionnement du Boosting. Il construit des modèles un à un, chaque nouveau modèle aidant à corriger les erreurs commises par les modèles précédemment formés. Ce principe a pour avantage de réduire le biais et donc l'erreur de prévision. Nous utilisons l'algorithme de recherche automatique GridSearchCV<sup>3</sup> de la librairie Scikit-Learn, pour sélectionner les hyperparamètres tels que la pseudo-régularisation de la fonction de perte et la profondeur maximale des arbres.

---

3. [https://scikit-learn.org/stable/modules/generated/sklearn.model\\_selection.GridSearchCV.html](https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html)

## 4 Expérimentation

### 4.1 Présentation du jeu de données

Notre jeu de données s'étend sur la période du premier mars 2019 jusqu'à fin août 2019 et se concentre sur une zone réduite située dans le département du Gard, où l'on dispose de 2 stations météorologiques distantes de plusieurs kilomètres l'une de l'autre. Ces 2 stations collectent 1 à 4 fois par heure des données réellement mesurées où on retrouve les niveaux d'humidité, de précipitation, mais aussi la température. Nous possédons également des données issues de plusieurs fournisseurs de prévisions météorologiques qui comportent des prévisions, pour chaque station, ayant un pas variable compris entre 1 et 3 heures selon les fournisseurs. Les prévisions concernent la couverture nuageuse, l'humidité, la pression, la température et enfin l'intensité du vent et sa direction. L'ensemble de ce jeu de données présente quelques contraintes, premièrement, les stations météorologiques peuvent présenter des défaillances, aussi le jeu de données qu'elles fournissent est plus ou moins complet selon la station considérée. Deuxièmement, il est également possible que certains fournisseurs ne possèdent pas une ou plusieurs variables de prévision contrairement aux autres. Troisièmement, la période couverte par les données représente seulement 3 saisons de l'année en cours, le volume du jeu de données est donc relativement faible. Après nettoyage du jeu de données, le cadre expérimental a été défini de la manière suivante.

Matrice  $X$  composée des variables exogènes suivantes :

- couverture nuageuse (%);
- direction du vent (degrés);
- humidité (%);
- pression (hPa);
- température (°C);
- vent (m/s).

Vecteur colonne  $Y$  composée de la variable endogène suivante :

- humidité (%) ou température (°C).

Pour chacune des 2 stations (Station\_1 et Station\_2) 3 fournisseurs météorologiques sont sélectionnés (Api-Agro, Weatherbit, et Dark Sky) ayant tous un pas de prévision égal à 1 heure sont retenus. La matrice  $X$  est uniquement composée des variables de prévisions des fournisseurs et le vecteur colonne  $Y$  est constitué des observations concernant la température ou de l'humidité selon ce que nous cherchons à prédire. Concernant le volume des données, si nous prenons l'exemple de la Station\_1, la matrice  $X$  en entrée de la première couche est de la taille  $(2131 \times 15)$  et de  $(2131 \times 1)$  pour  $Y$ . À chaque ligne de  $X$  et  $Y$ , est associée une date en tant qu'index et ayant une granularité heure par heure. Le pourcentage de sélection des données d'entraînement en entrée des couches 1 à 3 du modèle est de 75% et les 25% restant représentent les données de test. Quant à la dernière couche, le pourcentage de sélection des données d'entraînement est de 80% et de 20% pour les données de test.



## 4.2 Objectifs et métriques d'évaluation

Dans cette étude, nous nous concentrons principalement sur l'amélioration des prévisions des températures et de l'humidité sur chacune des 2 stations. L'objet de l'étude sera donc de fournir des prévisions plus précises que celles des fournisseurs météorologiques. Pour déterminer la précision de prévision des fournisseurs pour chaque station, il suffit de comparer la valeur de leurs prévisions aux valeurs des observations de la station étudiée. Comme notre jeu de données dispose d'environ 4 mois d'observations passées nous pouvons exploiter ces observations pour évaluer la performance d'un fournisseur à travers les 3 dernières saisons. Pour calculer l'écart entre les prévisions et les observations nous utilisons la mesure de l'écart quadratique moyenne (Root Mean Square Error) :

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \tag{4}$$

L'équation (4) permet ainsi d'obtenir les différences entre les  $n$  prévisions  $\hat{y}_i$  et les  $n$  observations  $y_i$  ayant la même unité que les valeurs de prévisions et d'observation. Les 3 fournisseurs sélectionnés dans notre cadre expérimental sont ceux présentant l'écart quadratique le plus faible et donc ceux ayant des prévisions plus proches des valeurs observées. Chacun de ces fournisseurs utilise un modèle numérique qui simule l'évolution de l'atmosphère, ils procèdent à un découpage en une grille de trois dimensions, comprenant des mailles de quelques kilomètres de côté. C'est donc pour chacune de ces mailles que les prévisions sont établies. Cependant, il est possible qu'entre deux points distants d'une centaine de mètres seulement les données climatiques telles que la température et l'humidité par exemple soient significativement différentes. Ces variations locales sont des phénomènes dit de sous-maille, pour lesquels les fournisseurs visent, via l'utilisation d'algorithmes complexes, de prédire au mieux les phénomènes locaux. Les données d'observations réelles pour chaque station dont nous disposons permettent de pouvoir prendre en compte les phénomènes locaux pour chacune d'entre elles. Notre modèle va donc apprendre à prédire la température et l'humidité à l'endroit exact de chaque station dans l'objectif de d'éliminer le problème de généralisation des prévisions sur plusieurs kilomètres et donc de fournir des prévisions plus fiables à court terme.

## 5 Résultats et Discussion

	Api-Agro	Dark Sky	Weatherbit	Prévisions
Température Station_1	0.80	0.73	1.32	<b>0.50</b>
Température Station_2	1.19	1.80	2.13	<b>0.79</b>
Humidité Station_1	9.63	4.39	5.28	<b>4.02</b>
Humidité Station_2	4.98	4.74	7.12	<b>4.12</b>

TAB. 1 : Comparaison RMSE des prévisions des fournisseurs et celles du modèle d'Apprentissage Ensembliste Multicouche.

Le tableau 1 présente les écarts quadratiques moyens des prévisions (des températures et de l'humidité) du 28 août 12 heures jusqu'au 28 août 18 heures provenant des 3 fournisseurs (Api-Agro, Dark Sky et Weatherbit) et des prévisions réalisées avec notre modèle pour les Station\_1 et Station\_2. Les valeurs RMSE pour les températures s'interprètent en degrés puis en points de pourcentage pour l'humidité. Pour une meilleure représentativité des résultats, nous avons exécuté 5 fois l'ensemble des algorithmes de notre méthode et effectué la moyenne des prévisions pour chaque heure de prévision. Les RMSE permettent ici de juger de la différence moyenne des valeurs prédites et des valeurs observées sur la plage horaire 12 à 18 heures. Par exemple, pour la Station\_1, les températures prédites par notre algorithme (Prévisions), le 28 août entre 12 et 18 heures diffèrent en moyenne de 0,50 degrés des températures réellement observées. Concernant l'humidité, pour la Station\_2, toujours d'après le tableau 1, les prévisions issues de notre modèle diffèrent en moyenne de 4,02 points de l'humidité réellement observée.

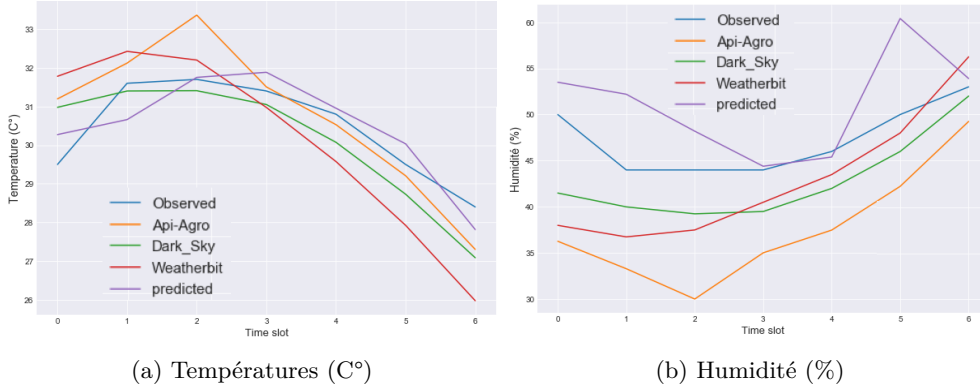


FIG. 3 : Comparaison des 7 heures de prévisions des températures et de l'humidité pour la Station\_1.

La figure 3a permet de comparer les prévisions de nos 3 fournisseurs et les prévisions de notre modèle (predicted) aux valeurs réellement mesurées (Observed) pour la Station\_1. Sur la figure de gauche, sont représentés les 7 heures de prévisions (toujours du 28 août 12 heures jusqu'à 18 heures) des températures en fonction de chaque heure. Cette figure permet de comparer les prévisions de notre modèle à celles des fournisseurs heure par heure. Ainsi, il est possible de voir que notre algorithme effectue les meilleures prévisions pour 5 cas sur 7. Il est cependant possible de voir sur les coordonnées 1 et 3 de l'axe des abscisses (correspondant respectivement aux 28 août 13 heures et 15 heures) que les prévisions des fournisseurs Dark Sky et Api-Agro sont les plus proches des valeurs observées. Quant à la figure 3b, celle-ci représente le pourcentage d'humidité en fonction de chaque heure. On constate ici que les prévisions de l'humidité de notre modèle sont plus précises que celles des fournisseurs pour 5 cas sur 7 où aux coordonnées 1 et 5 les fournisseurs Dark Sky et Weatherbit proposent les prévisions les plus précises. Pour les prévisions d'une portée de 1 à 7 heures notre méthode permet de fournir des prévisions plus précises que celles des fournisseurs dans 71% des cas.

**Impact de l'amélioration des prévisions : cas de la prévision de maladie Mildiou.** Si nous prenons par exemple le cas de la prévision de la maladie Mildiou, cette dernière peut apparaître si certains seuils (voir figure 2) de température et d'humidité sont franchis. Pour déterminer si la maladie est avérée, il suffit d'additionner les symboles des seuils de température et d'humidité, plus une cellule comporte de symboles (+) plus le risque est avéré. Si cependant la somme comporte un seul (0) alors la probabilité est très faible.

Température (°C)	11[	[11-15]	]15-20]	]20-25]	]25-30]	]30
	0	+	++	+++	++	0

Humidité (%)	[0-95]	]95-97]	]97
	0	+	+++

TAB. 2 : Seuil d'intensité de contamination pour les températures et l'humidité.

Nous pouvons voir ici que le passage d'une catégorie d'intensité à une autre pour les températures est l'humidité ne dépend que d'une seule unité. Par exemple, si nous considérons l'humidité égale à 97% et la température réelle à 11C°, une prévision présentant un seul degré de moins suffirait pour conclure que le risque de maladie est quasi nul alors qu'en réalité le risque est présent. Il en est de même pour la problématique de prévision de gel qui relève du même besoin de précision des prévisions, ces exemples reflètent donc l'importance de l'amélioration des prévisions.

## 6 Conclusion et Travaux futurs

La méthode d'Apprentissage Ensembliste Multicouche présentée dans ce papier est capable de fournir des prévisions des températures et de l'humidité heure par heure, majoritairement plus précises que celles de nos meilleurs fournisseurs météorologiques. Et en moyenne, sur une plage totale de 7 heures de prévision, notre méthode présente une précision de prévision plus fine que celle des 3 fournisseurs pour les deux sites étudiés. L'amélioration des prévisions peut ainsi directement simplifier la prise de décision des agriculteurs. Les prévisions issues de notre modèle seront utilisées pour prévoir l'apparition de diverses maladies, mais aussi de prévoir l'apparition de phénomènes extrêmes tels que le gel ou des pluies abondantes. Une amélioration possible de nos travaux peut consister à rendre notre méthode adaptable à tous jeux de données de façon automatique. En effet, le choix de l'architecture de notre modèle et plus précisément du nombre de couches, des algorithmes et enfin de leurs hyperparamètres dépend intrinsèquement de nos données. Le nombre de couches dépend par exemple du nombre de lignes de notre jeu de données, mais aussi du pourcentage de données d'entraînement et de test choisi au sein de chaque couche. Il en est de même pour la sélection d'algorithmes présents dans chaque couche, ceux fournissant les prévisions les plus proches de celles réellement observées ont été conservés. L'objectif serait de mettre en œuvre une méthode permettant d'optimiser les prévisions en sélectionnant

automatiquement les paramètres optimaux tels que les pourcentages d'entraînement et de test, le nombre de couches, les algorithmes et leurs hyperparamètres.

## Références

- Bauer, P., A. Thorpe, et G. Brunet (2015). The quiet revolution of numerical weather prediction. *Nature* 525, 47–55.
- Breiman, L. (1996). Bagging predictors. *Machine Learning* 24, 123–140.
- Cofiño, A., R. Cano, C. Sordo, et J. Gutiérrez (2002). Bayesian networks for probabilistic weather prediction. *Proceedings of the 15th European Conference on Artificial Intelligence*, 695–699.
- Džeroski, S. et B. Ženko (2004). Is combining classifiers with stacking better than selecting the best one? *Machine Learning* 54, 255–273.
- Grover, A., A. Kapoor, et E. Horvitz (2015). A deep hybrid model for weather forecasting. pp. 379–386.
- Jakaria, A., M. Hossain, et M. Rahman (2018). Smart weather forecasting using machine learning: A case study in tennessee.
- Ma, Z., P. Wang, Z. Gao, R. Wang, et K. Khalighi (2018). Ensemble of machine learning algorithms using the stacked generalization approach to estimate the warfarin dose. *Plos One* 13, e0205872.
- Shapire, R. et Y. Freund (2013). Boosting: Foundations and algorithm. *Kybernetes* 42, 164–166.
- Sharif, M. et D. Burn (2007). Improved k -nearest neighbor weather generating model. *Journal of Hydrologic Engineering* 12.
- Zaytar, M. et C. E. Amrani (2016). Sequence to sequence weather forecasting with long short-term memory recurrent neural networks. *International Journal of Computer Applications* 143, 7–11.
- Zhou, Z. (2012). *Ensemble Methods: Foundations and Algorithms*. Abingdon-on-Thames: Taylor & Francis.

## Summary

Agriculture is one of the areas whose activities depend heavily on weather forecasts. Indeed, in order to optimize their production, farmers must be able to anticipate climate conditions favorable or not to their activities by deploying the appropriate action plans. For this purpose, they consult the data daily from various suppliers of weather forecasts. However, the reliability of the forecasts of each supplier is variable according to the period, the climate or the geographical area. Farmers, therefore, have to arbitrate between suppliers daily. To facilitate farmers' decision-making, this paper proposes a new set of learning architecture that significantly improves the accuracy of weather short-term forecasts for the next 1-12h.