

# Chapitre III

## Approche d'ensembles multi-modèles pour la prédiction de l'irradiation solaire

### III.1 Introduction

La prédiction de l'irradiation solaire est devenue une étape clé pour prévoir correctement les performances des systèmes de conversion de l'énergie solaire et pour garantir la stabilité et la sécurité des réseaux électriques. Les compagnies d'électricité doivent contrôler les systèmes de production en se basant sur la connaissance préalable de la charge et de la capacité de production attendues au cours des prochaines minutes, heures et jours, car certains groupes électrogènes peuvent être démarrés rapidement, tandis que d'autres ont besoin de beaucoup de temps avant d'être prêts. De plus, le prix de la production d'électricité se négocie sur un marché en évolution constante et rapide. Cela explique pourquoi des prédictions précises de la production et de la consommation d'électricité sont nécessaires pour réduire les coûts de l'électricité [100].

La procédure normale de prédiction de la puissance électrique produite par un système de conversion solaire est basée sur le rayonnement solaire disponible, ainsi que sur plusieurs aspects concernant la performance du système et les problèmes environnementaux connexes au moment de l'intérêt. En utilisant des générateurs basés sur les technologies de conversion photovoltaïque (PV), l'électricité produite est, en principe au moins, linéairement liée à l'irradiation solaire incidente. Ainsi, à l'aide d'un outil de simulation d'énergie photovoltaïque approprié, la prévision de la production d'électricité est directement liée à celle de l'éclairement énergétique incident. Cette dernière est cependant très variable dans le temps pour des raisons météorologiques, ce qui rend difficile sa prévision exacte. D'un point de vue plus général, la prédiction de l'irradiation solaire avec une bonne précision devrait augmenter l'efficacité et la valeur de la production d'énergie de tout type de système d'énergie solaire.

La littérature que nous venons d'examiner dans le chapitre trois montre que de nombreuses méthodes stochastiques et de l'IA sont possibles et peuvent être combinées

efficacement. Bien que les méthodes individuelles soient largement utilisées, leur «force intrinsèque» n'est pas forcément la meilleure option possible. La nouvelle méthodologie présentée dans ce chapitre, tente d'améliorer les prédictions en sélectionnant de manière dynamique le prédicteur le plus approprié pour chaque série temporelle de GHI en utilisant plusieurs approches de combinaison et de classification [101].

On peut considérer qu'une méthode fiable serait une méthode qui diminue le risque de modèle (lorsque il est impliqué dans l'ensemble) et cherche à améliorer la précision des prédictions en tirant profit des différentes forces de divers modèles potentiellement efficaces, tout en compensant leurs limites. Le résultat de la nouvelle approche proposée dans cette étude est la prédiction produite par le modèle d'apprentissage distingué par la performance maximale dans la région de *l'espace des variables*. Ce dernier est également appelé «espace d'observations» car chaque valeur historique de la série GHI (dimension / coordonnée) correspond à une observation du GHI historique enregistré. Dans ce chapitre, divers types d'ensembles sont évalués à l'aide de plusieurs modèles statistiques et / ou modèles à base des techniques de l'IA [102].

### III.2 Méthodologie proposée

Considérons un ensemble de  $N$  modèles de prédiction  $f_i(X)$  ( $i = 1, 2, \dots, N$ ) formés indépendamment sur les données d'apprentissage disponibles. Le problème schématisé sur la figure III.1, consiste à définir une approche de combinaison  $\theta\{\}$  de telle sorte que les prédictions obtenues après la combinaison des différents modèles individuels pour un point de données particulier (c'est-à-dire non inclus dans l'ensemble d'apprentissage) puissent être décrites par (III.1):

$$\hat{F}(X) = \theta\{f_1(X), f_2(X), \dots, f_N(X)\} \quad (III.1)$$

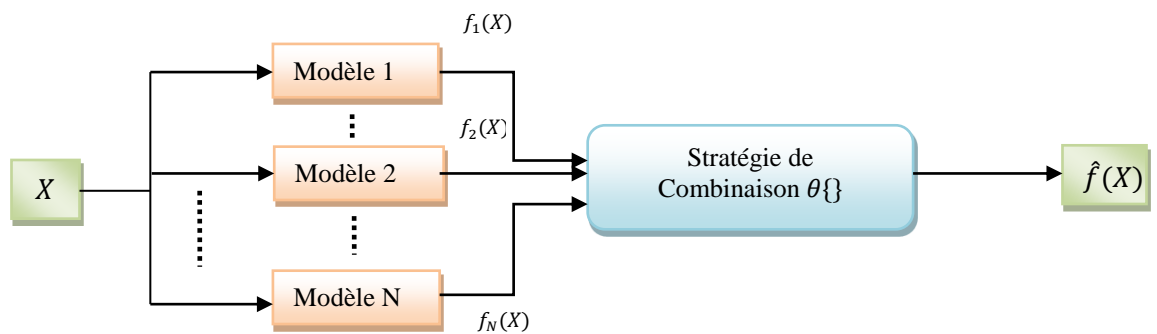


FIGURE III.1 Schéma fonctionnel du système de prédiction proposé.

### III.2.1 Modélisation d'ensemble

Les méthodes d'ensemble sont des algorithmes d'apprentissage qui construisent un ensemble de modèles, puis classifient / prédisent les nouveaux points de données en prenant une combinaison pondérée de leurs prédictions [103,104].

La construction d'un ensemble de différents modèles de prédiction peut être faite à l'aide du même modèle de prédiction avec différents paramètres de réglage (tuning) ou selon différentes méthodes d'apprentissage.

Dans une autre perspective, les méthodes d'ensemble sont des méta-algorithmes qui combinent de nombreuses méthodes d'apprentissage automatique en un seul modèle dans le but de réduire la variance (bagging [89] ou le biais (boosting) [105] ou d'améliorer les prédictions (stacking). Les méthodes existantes peuvent être divisées en deux groupes:

- ***Méthodes d'ensemble séquentielles***

Dans lequel les apprenants de base sont créés consécutivement (par exemple, AdaBoost) [106]. L'objectif ici est d'exploiter la dépendance entre les différents apprenants de base. Les performances générales peuvent être améliorées en analysant les mauvais exemples (mauvaises sorties) avec des poids de grandes valeurs, comme dans les méthodes de renforcement classiques.

- ***Méthodes d'ensemble parallèles***

Les apprenants de base sont créés en parallèle (par exemple, Random Forest) [107]. Le but ici est d'exploiter l'indépendance potentielle entre tous les apprenants de base, car l'erreur peut être fortement réduite en faisant la moyenne.

Habituellement, un seul algorithme d'apprentissage de base peut être utilisé pour produire des modèles de base homogènes à partir de modèles du même type, conduisant à des ensembles homogènes.

Alternativement, des modèles hétérogènes, c'est-à-dire des modèles différents qui conduisent à des ensembles hétérogènes, dont l'objectif est d'avoir un modèle global plus précis que chacun de ses membres individuels, les modèles individuels de l'ensemble doivent être aussi précis que possible et aussi diversifiés que possible [104]. Une question centrale dans la théorie des ensembles est que, pour améliorer le processus de combinaison, les modèles individuels doivent être différents. Sinon, les performances ne seraient pas améliorées si les modèles simples dupliqués étaient combinés.

Dans cette thèse, nous ne considérons que la seconde approche, dans laquelle la modélisation *hétérogène* est utilisée. Cette considération est confortée par l'intérêt porté à

l'examen de l'efficacité de l'approche pour la prédiction de GHI à court terme (de 1 minute à 1 heure à l'avance). A des taux de rafraîchissement aussi élevés, les variables exogènes qui pourraient être utilisées pour prédire le GHI avec un certain degré de fiabilité (par exemple, fraction nuageuse, précipitation ou profondeur optique d'aérosol) ne sont généralement pas disponibles à partir d'observations locales ou de données modélisées et ré-analysées (dont la résolution spatiale est quand même trop large).

Comme mentionné ci-dessus, dans ce qui suit, plusieurs approches de combinaison existent et sont conçues selon deux stratégies différentes. Dans la première, appelée «combinaison», la prédiction finale est obtenue après la combinaison des prédictions obtenues par les différents modèles uniques. L'autre stratégie, dite «sélection», consiste en une nouvelle méthode (détaillée ci-après), qui obtient la prédiction finale en sélectionnant le résultat (prédiction) du «meilleur modèle individuel» à l'aide d'une méthodologie qui définit une partition appropriée de l'espace des variables. Dans la présente étude, trois familles de modèles sont utilisées pour construire l'ensemble optimal, à savoir : (i) les modèles autorégressifs ; ii) modèles à base de Réseaux de Neurones Artificiels (RNA); et (iii) les modèles à noyaux.

### III.2.2 Méthodes individuelles

#### III.2.2.1 Modèles autorégressifs

Les modèles ARIMA [108] constituent l'origine de la plupart des techniques de prédiction statistiques. Un modèle ARIMA (p,d,q), généralement appelé modèle de Box-Jenkins, suppose que toute valeur future d'une série temporelle peut être exprimée sous forme de fonction linéaire de certaines observations passées, en plus de certains termes de bruit blanc. Il utilise une procédure de différenciation courante d'ordre  $d$  pour rendre la série temporelle stationnaire. Il se décompose en différents modèles, comme suit:

- ***Processus autorégressif AR***

Un processus stochastique  $Y_t$  avec des valeurs réelles est considéré comme un processus AR(p) d'ordre  $p$  si:

$$y_t = a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p} + \varepsilon_t. \quad (\text{III.2})$$

La valeur AR (p) à l'instant  $t$  est ensuite estimée à l'aide des  $p$  valeurs passées et d'un terme aléatoire  $\varepsilon_t$ .

- **Moyen mobile MA**

Un processus stochastique  $Y_t$  est dit processus MA d'ordre  $q$ , noté MA ( $q$ ), s'il existe  $q$  coefficients  $b_1, b_2, \dots, b_q$ , et un bruit blanc  $\varepsilon_t$  tel que :

$$y_t = b_0 \varepsilon_t + b_1 \varepsilon_{t-1} + \dots + b_q \varepsilon_{t-q}. \quad (\text{III.3})$$

La valeur du processus MA ( $q$ ) au temps  $t$  est donc régressée sur ses propres erreurs passées.

- **Autorégressif à moyenne mobile, ARMA**

Les processus AR ( $p$ ) sont des cas particuliers d'ARMA. Un processus stochastique  $Y_t$  est considéré comme un processus ARMA d'ordre  $p, q$  si:

$$y_t = \varepsilon_t + (a_1 y_{t-1} + a_2 y_{t-2} + \dots + a_p y_{t-p}) + (b_0 \varepsilon_t + b_1 \varepsilon_{t-1} + \dots + b_q \varepsilon_{t-q}) \quad (\text{III.4})$$

Ceci peut aussi être formulé comme  $\phi(z)Y_t = \theta(z)\varepsilon_t$ , ou  $\phi(z) = 1 + a_1 z + a_2 z^2 + \dots + a_p z^p$  et  $\theta(z) = 1 + b_1 z + b_2 z^2 + \dots + b_q z^q$  sont les termes polynomiaux des composantes AR et MA du processus ARMA( $p, q$ ).  $y_t$ , et  $z$  est l'opérateur de rétrogradation.

- **Autorégressif à moyenne mobile intégrée, ARIMA**

ARIMA est une extension de la classe de modèles ARMA dans laquelle une opération de différenciation est ajoutée. Un processus  $y_t$  est considéré comme un ARIMA( $p, d, q$ ) si  $(1 - z)^d Y_t$  est une ARMA ( $p, q$ ) causal. Le modèle ARIMA peut être formalisé comme suit:

$$\phi(z)(1 - z)^d Y_t = \theta(z)\varepsilon_t. \quad (\text{II.5})$$

- **Autorégressif à moyenne mobile saisonnière intégrée, SARIMA**

Le modèle SARIMA peut être décrit comme suit:

$$\phi_p(z)\Phi_P(z^S)(1 - z)^d(1 - z^S)^D Y_t = \theta_q(z)\Theta_Q(z^S)\varepsilon_t \quad (\text{III.6})$$

Où  $\phi_p(z), \Phi_P(z^S), \theta_q(z)$  et  $\Theta_Q(z^S)$  sont les termes polynomiaux d'ordres  $p, P, q$  et  $Q$ , respectivement,  $d$  et  $D$  sont les ordres de la différenciation non saisonnière et saisonnière, respectivement.

### III.2.2.2 Modèles à base de réseaux de neurones artificiels (RNA)

Les RNA ont été créés pour surmonter les limitations rencontrées avec des méthodes plus traditionnelles lorsqu'on essaie de résoudre des problèmes complexes. Les RNA apprennent de certains exemples en construisant une relation d'entrées-sorties permettant d'effectuer des prédictions. Les RNA contiennent une interconnexion de plusieurs neurones.

Dans la littérature, plusieurs architectures de réseaux avec différents algorithmes d'apprentissage ont été proposées [109]. Un algorithme d'apprentissage est essentiellement une procédure dont la fonction est de réguler les poids d'un réseau donné. Pour un ensemble de données d'apprentissage donné, cette opération peut être réalisée en minimisant une fonction objective entre la fonction cible souhaitée et estimée. La présente étude se concentre sur quatre types de RNA, à savoir: MLP (Perceptron Multicouches), RBF (Réseaux à Fonction de Base Radial), ELM (Extreme Learning Machines) et Réseau de neurones Elman (ENN).

- **Réseaux de neurones de perceptron multicouches PMC**

Un PMC est composé de plusieurs unités interconnectées (neurones) structurées en couches et fonctionnant en parallèle avec un flux d'informations en retour, comme illustré dans la figure III.2. Les signaux circulent successivement dans les différentes couches, de la couche d'entrée à la couche de sortie. Les couches entre les couches d'entrée et de sortie sont appelées couches cachées. Dans chaque couche, chaque unité calcule un produit scalaire entre le vecteur des poids (pondération) et le vecteur de sortie issu de la couche précédente. Ensuite, une fonction de transfert est appliquée au résultat précédent pour créer une entrée pour la couche suivante [111]. Généralement, la fonction de transfert pour les couches cachées est la sigmoïde:

$$f(x) = \frac{1}{1+e^{-x}} \quad (\text{III.6})$$

- **Réseaux de neurones à fonctions radiales, RBF**

Un réseau neuronal RBF est composé de trois couches (une couche entrée, une couche cachée et une couche de sortie). Comme illustré sur la figure III.3, les neurones d'entrée (c'est-à-dire les variables d'entrée) transmettent uniquement les variables d'entrée  $X_j$ . Chaque neurone de la couche cachée est caractérisé par une fonction noyau  $\varphi_j$  (généralement une fonction gaussienne) décrite par une largeur  $\sigma_j$  et un centre  $c_j$ , tels que :

$$\varphi_j(\|y - c_j\|) = \exp\left(-\frac{1}{2}\left(\frac{\|y - c_j\|}{\sigma_j}\right)^2\right). \quad (\text{III.7})$$

La couche de sortie contient un seul neurone fournissant la valeur prédite. La fonction de sortie peut être écrite comme suit:

$$f(y) = \sum_{j=1}^m \lambda_j \varphi_j(\|y - c_j\|) \quad (\text{III.8})$$

où  $\lambda_j$  et  $m$  sont respectivement le poids et le nombre de fonctions RBF [104,108].

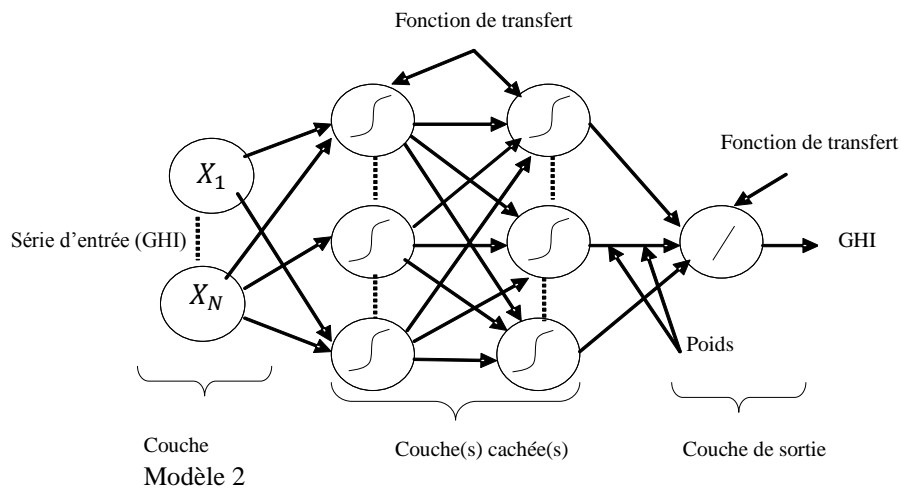


FIGURE III.2 Architecture d'un réseau de neurones de type perceptron multicouches.

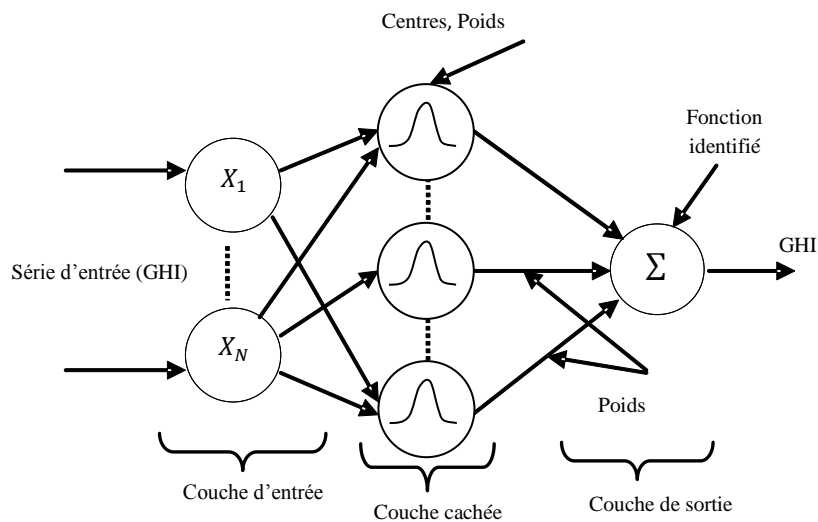


FIGURE III.3 Architecture d'un réseau de neurones à fonction radiale

- **Réseaux de Neurones Elman ENN**

Un ENN est un type de réseau de neurones dynamique, basé essentiellement sur la structure de réseaux de neurones artificiels classiques. Il stocke un état interne pour élaborer la fonction de ses caractéristiques dynamiques et fournit au système la possibilité de s'adapter à toute caractéristique variable dans le temps. Un ENN peut donc refléter directement les caractéristiques dynamiques d'un système, ce qui est une caractéristique importante d'un outil de prédiction avancé. La structure ENN est illustrée sur la figure III.4.

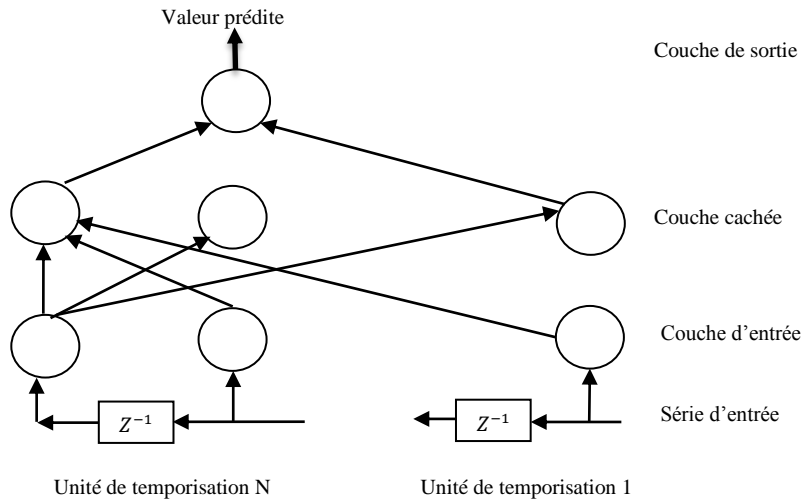


FIGURE III.4 Architecture du réseau Elman.

Généralement, les ENN sont divisés en quatre couches: la couche d'entrée, la couche cachée, couche de réalisation (*undertake*) et la couche de sortie. Les nœuds de la couche d'entrée agissent simplement comme des unités de transmission de signal. La couche de sortie réagit comme une fonction de sommation, tandis que la couche de réalisation (ou couche d'état) est utilisée pour mémoriser les valeurs de sortie précédentes des unités de la couche cachée qui devront être renvoyées à l'entrée du réseau.

La capacité de stockage de l'ENN le rend plus sensible aux données historiques, ce qui permet au retour d'information interne d'améliorer l'aptitude du réseau à gérer les variations dynamiques des données. De plus, un ENN peut approximer n'importe quelle fonction non linéaire avec un degré de précision très acceptable et peut éviter l'impact du bruit externe. Les équations mathématiques dans l'espace d'état non linéaire d'un ENN peuvent s'écrire comme suit :

$$y(x) = g(\omega^3 x_c(x)) \quad (\text{III.9})$$

$$x(k) = f(\omega^1 x_c(k) + \omega^2 (u(k-1))) \quad (\text{III.10})$$

$$x_c(k) = x(k-1) \quad (\text{III.11})$$

où

$y$  est le vecteur de sortie de dimension égale à  $M$ ;

$x$  est le vecteur de nœud de la couche cachée avec  $M$  dimensions;

$u$  est le vecteur d'entrée avec  $r$  dimensions;

$x$  est le vecteur d'état de retour avec  $N$  dimensions;

$\omega^3$  sont les poids reliant la couche cachée à la couche de sortie;

$g$  est la fonction de transfert du noeud de sortie;

$f$  est la fonction de transfert des nœuds des couche cachées.

Généralement, les ENNs utilisent également l'algorithme de rétro-propagation pour déterminer les poids du réseau.

- **Extreme Learning Machines, ELM**

Le modèle ELM a été proposé par Huang et al. [110]. Ce type de réseau neuronal à une seule couche cachée peut être utilisé pour les problèmes de régression et de classification. Supposons un ELM avec  $Q$  neurones cachés et une fonction d'activation  $g(x)$  qui est conçu pour apprendre  $N$  différents échantillons  $(X_i, t_i)$ , ou  $X_i = [x_{i1}, x_{i2}, \dots, x_{in}]^T$  et  $t_i = [t_{i1}, t_{i2}, \dots, t_{im}]^T$ . Les poids d'entrée et les biais cachés sont choisis aléatoirement. Par conséquent, le système non linéaire peut être converti en un système linéaire comme suit:

$$H\beta = T \quad (\text{III.12})$$

Où  $H = \{h_{ij}\}$  ( $i = 1, 2$ ) est la matrice de sortie de la couche cachée unique,  $h_{ij} = g(W_j * Y_i + b_j)$  désigne la sortie du  $j_{th}$  neurone caché par rapport à  $X_i$ ;  $W_j = [w_{j1}, w_{j2}, \dots, w_{jn}]^T$ . et  $b_j$  est le biais du neurone caché  $j_{th}$ ;  $W_j * Y_i$  est le produit intérieur de  $W_j$  et  $X_i$ ;  $\beta = [\beta_1, \beta_2, \dots, \beta_K]^T$  est la matrice des poids de sortie et  $\beta_j = [\beta_{j1}, \beta_{j2}, \dots, \beta_{jm}]^T$  ( $j=1, \dots, K$ ) indique le vecteur de poids reliant le neurone caché  $j_{th}$  au neurone de sortie. Enfin,  $T = [t_1, t_2, \dots, t_N]^T$  est la matrice de sortie (cibles) (c'est-à-dire la sortie souhaitée de GHI). Par conséquent, les poids de sorties sont déterminés en recherchant la solution des moindres carrés du système linéaire décrit par l'équation (III.13), tel que:

$$\hat{\beta} = H^+ T \quad (\text{III.13})$$

où  $H^+$  est l'inverse généralisé de Moore-Penrose de la matrice  $H$ . Habituellement, les ELM ont tendance à donner de bonnes performances de généralisation avec une vitesse d'apprentissage extrêmement rapide d'où vient leur nom [111,112].

### III.2.2.3 Les modèle à noyau

- **Machine à vecteurs de support, SVM**

Les SVM sont des techniques non paramétriques supervisées qui ne nécessitent donc aucune hypothèse sur la distribution des données. Le but d'une approche de régression SVM est de définir une fonction *lisse* ayant (au maximum)  $\varepsilon$  déviation par rapport aux échantillons d'apprentissage, qui sont dans le cas présent issus d'une série temporelle de rayonnement

solaire [113,114]. Ceci peut être réalisé en projetant la série originale d'entrée de dimension  $d$  dans un espace variable transformé, c'est-à-dire  $\Gamma(x) \in \mathbb{R}^{d'}$  ( $d' > d$ ) et en fixant une estimation linéaire comme suit:

$$f(x) = \omega^* \cdot \Gamma(x) + b^* \quad (\text{III.14})$$

La fonction linéaire dans l'espace variable de dimension supérieure est obtenue en optimisant une fonction de coût, dont l'objectif est de minimiser deux critères (norme euclidienne et minimisation d'erreur), elle est exprimée comme suit:

$$\Psi(\omega, \xi) = \frac{1}{2} \|\omega\|^2 + C \sum_{i=1}^N (\xi_i + \xi_i^*). \quad (\text{III.15})$$

Où  $C$  est un paramètre de régularisation. Cette fonction est soumise aux contraintes ci-dessous:

$$\begin{cases} y_i - (\omega \cdot \Gamma(x_i) + b) \leq \varepsilon + \xi_i \\ (\omega \cdot \Gamma(x_i) + b) - y_i \leq \varepsilon + \xi_i^* \end{cases}, \quad \xi_i, \xi_i^* \geq 0 \text{ et } i \in [1, N] \quad (\text{III.16})$$

où  $\xi_i$  et  $\xi_i^*$  sont des variables *slack* liées à des échantillons qui ne correspondent pas au tube de déviation  $\varepsilon$ . La formulation de la fonction d'erreur est conçue pour traiter une fonction de perte  $\varepsilon$ -insensible  $|\gamma|_\varepsilon$  définie comme suit :

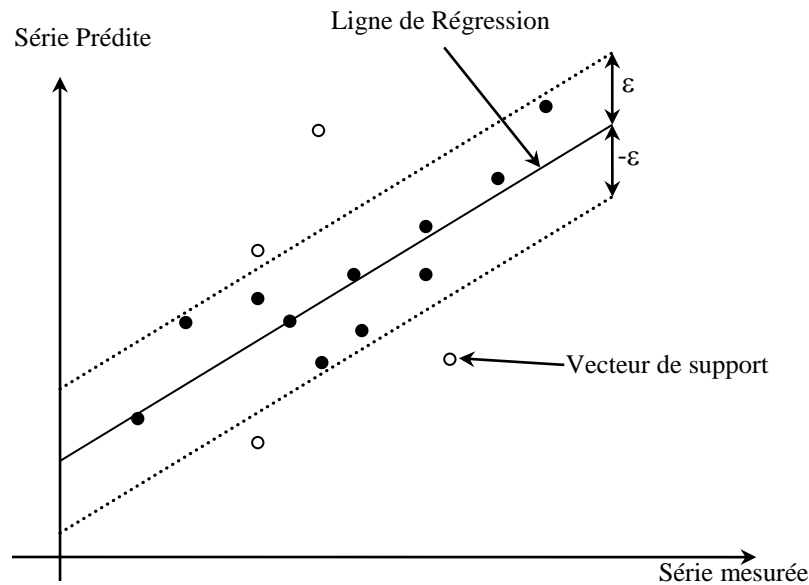
$$\begin{cases} 0, \text{ si } |\tau| \leq \varepsilon \\ |\tau| - \varepsilon, \text{ autrement} \end{cases} \quad (\text{III.17})$$

où  $\tau$  est représenté l'écart par rapport à la sortie souhaitée. Cela revient à dire que les différences entre les valeurs mesurées et les valeurs prédites sont acceptées à l'intérieur du tube (erreur inférieure à  $\varepsilon$ ), alors qu'une pénalité linéaire est imposée aux points situés à l'extérieur du tube insensible. Un exemple de SVM linéaire est présenté dans la figure III.5.

Le résultat du problème d'optimisation est une fonction représentée dans l'espace variable (original) d'entrée comme suit :

$$f(x) = \sum_{i \in S} (\alpha_i - \alpha_i^*) K(x_i, x) + b^* \quad (\text{III.18})$$

Où  $K(.)$  est une fonction noyau  $S$  est le sous-ensemble d'indices  $i = (1, 2, \dots, N)$  correspondant au multiplicateurs de Lagrange non nuls  $\alpha_i$  ou  $\alpha_i^*$ . En fonction de leur importance dans la détermination d'une solution, les multiplicateurs de Lagrange évaluent chaque point d'apprentissage. Les points d'apprentissage avec des poids non nuls sont appelés des vecteurs de support.

FIGURE III.5 Régression SVM linéaire avec  $\varepsilon$ -tube.

### III.2.3 Méthodes de combinaison

Une stratégie de combinaison tente d'améliorer les performances de prédiction globales en combinant différentes approches. Elle tente d'augmenter la précision des prédictions par le biais d'une solution pratique, mais aux dépens d'une complexité de calcul supplémentaire. Par conséquent, il convient de bien comprendre les propriétés de tous les modèles d'ensemble et de faire preuve de prudence dans l'utilisation des informations accessibles dans le but de développer des méthodes de combinaison efficaces [115]. Après avoir généré un ensemble d'apprenants de base, au lieu d'essayer de déterminer le meilleur modèle, les méthodes d'ensemble sélectionnent une combinaison pour obtenir la meilleure capacité de généralisation, ou la méthode de combinaison joue un rôle important.

Dietterich [116] a attribué le bénéfice obtenu avec le processus de combinaison aux trois facteurs fondamentaux suivants:

- **Statistique:** habituellement, l'espace des hypothèses est trop grand pour permettre la découverte de petits ensembles de données de formation, où d'autres hypothèses différentes peuvent donner la même précision sur les données de formation. Si le modèle d'apprentissage sélectionne un espace parmi ces hypothèses, il existe un risque qu'une hypothèse mal choisie ne prédit pas correctement les données futures. Par conséquent, en combinant des hypothèses, le risque de choisir une hypothèse incorrecte peut être réduit.

- **Coût de calcul:** plusieurs modèles d'apprentissage permettent de créer différents types de recherche locale pouvant être bloqués dans les optima locaux. Bien que les données de formation soient suffisantes, il reste très difficile de déterminer la meilleure hypothèse. La combinaison de recherches locales lancées à partir de plusieurs points de départ distincts peut fournir une meilleure estimation de la véritable hypothèse inconnue. Dans ce cas, la combinaison d'hypothèses réduit le risque de choisir un mauvais minimum local.

- **Représentationnel:** dans plusieurs problèmes d'apprentissage automatique, la vraie hypothèse inconnue ne peut pas être caractérisée. Dans ce cas, l'utilisation d'une combinaison d'hypothèses permet une expansion de l'espace représentable des fonctions. Par conséquent, le modèle d'apprentissage peut constituer une estimation plus précise de la véritable hypothèse inconnue.

Les trois classes de facteurs que nous venons de décrire, expliquent généralement pourquoi les techniques d'apprentissage classiques ont tendance à échouer [104]. Dans ce qui suit, on fournit un bref résumé des techniques utilisées dans cette étude pour la combinaison linéaire de plusieurs prédictions.

### III. 2.3.1 Moyen simple, SA

La méthode de combinaison la plus simple est la moyenne simple, où toutes les prédictions individuelles ont le même poids ( $w_j = 1/n$ ). Plusieurs études ont révélé que curieusement, la moyenne simple améliore souvent la précision globale des prédictions, même lorsque certaines prédictions individuelles apparaissent comme des valeurs aberrantes (outliers) ou des valeurs extrêmes [117, 118].

### III. 2.3.2 Méthode basée sur l'erreur, EB

Dans cette méthode, la série chronologique est divisée en deux sous-ensembles complémentaires, à savoir l'ensemble d'apprentissage et celui de validation. Les modèles individuels sont formés sur les données de formation et leurs erreurs de prédiction résultantes (en utilisant la donnée de validation) sont enregistrées. Le poids de la combinaison pour chaque prédiction individuelle est ensuite calculé comme étant inversement proportionnel à l'erreur de prédiction du modèle correspondant, c'est-à-dire

$$W_j = \frac{e_j^{-1}}{\sum_{j=1}^n e_j^{-1}} \quad (\text{III.19})$$

### III.2.3.3 Régression par moindres carrés, LSR

Dans cette méthode, les poids d'une combinaison linéaire sont attribués en minimisant la somme des erreurs carrées (SSE) entre les données prédites et les données mesurées. La formulation classique de la combinaison linéaire peut être exprimée comme suit:

$$\hat{Y} = UW \quad (\text{III.20})$$

Où

$$U = \begin{bmatrix} \hat{y}_1^{(1)} \hat{y}_1^{(2)} \dots \hat{y}_1^{(n)}; \hat{y}_2^{(1)} \hat{y}_2^{(2)} \dots \hat{y}_2^{(n)}; \dots; \hat{y}_M^{(1)} \hat{y}_M^{(2)} \dots \hat{y}_M^{(n)} \end{bmatrix}^T$$

$$W = [w_1, w_2, \dots, w_n]^T \quad (\text{III.21})$$

Et où  $\hat{Y}$  et  $Y$  sont les séries prédites et mesurées, respectivement, et  $W$  est lié avec les poids des modèles individuels. Les SSE de prédictions peuvent alors être exprimées comme suit:

$$\begin{aligned} \text{SSE} &= \sum_{t=1}^M (y_t - \hat{y}_t)^2 \\ &= (Y - UW)^T (Y - UW) \\ &= Y^T Y - 2W^T U^T Y + W^T U^T U W. \end{aligned} \quad (\text{III.22})$$

En minimisant les SSE par rapport à  $W$ , le vecteur de pondération est calculé comme suit:  
 $W = (U^T U)^{-1} U^T Y$ .

### III.2.3.4 Méthode de pondération différentielle, DW

Similairement à la méthode LSR, les poids d'une combinaison linéaire de prédictions sont obtenus en minimisant la variance de l'erreur de prédiction combinée [103]. En fait, cette approche nécessite la connaissance de la matrice de covariance des erreurs de prédiction, inconnue dans la pratique. Les poids sont calculés avec :

$$W_i = \frac{\sum_{s=t-v}^{t-1} (e_s^{(i)})^2}{\sum_{j=1}^n \sum_{s=t-v}^{t-1} (e_s^{(j)})^2} \quad (\text{III.23})$$

Où  $n$  est le nombre de modèles individuels,  $t$  est la période de prédiction,  $W_i$  est la pondération attribuée à la deuxième méthode en fonction de la période précédente des données  $(t-1)$ ,  $v$  est une constante et  $e_t^{(i)}$  est le pourcentage d'erreur de prédiction à l'instant  $t$ , défini comme suit:

$$e_t^{(i)} = \frac{y_t - \hat{y}_t^{(i)}}{y_t} \quad \forall i = 1, 2, \dots, n. \quad (\text{III.24})$$

### III.2.3.5 la méthode Outperformance, OP

Proposé par Lemke et Gabrys [119], OP adopte un cadre bayésien de probabilités subjectives d'attribution des poids appropriés à l'ensemble des prévisions individuelles. Le poids d'un modèle de prévision individuel est déterminé à partir du nombre de fois où il a dépassé les autres modèles. Pour simplifier, la combinaison de deux modèles individuels F1 et F2 est réalisée comme suit

$$\hat{y}_k = w\hat{y}_k^{(1)} + (1 - w)\hat{y}_k^{(2)}, \forall k = 1, 2, \dots, N. \quad (\text{III.25})$$

Dans le cas présent, on suppose que les performances des deux modèles sont évaluées à l'aide d'une erreur absolue et enregistrées M fois. Avec ce scénario, il est raisonnable d'accepter que le modèle avec l'erreur absolue la plus petite surpasse les autres modèles. Si le nombre de fois  $F_1$  bat  $F_2$  est une fraction  $k$  de  $M$ , on peut supposer que  $k$  suit une distribution bêta  $B(k|a_1, a_2)$  ( $0 \leq k \leq 1; a_1, a_2 > 0$ ) [119]. Pour chaque réalisation de prédiction  $i$ , la variable de Bernoulli  $\delta_i$  est exprimée:

$$\delta_i = \begin{cases} 1, & \text{si } F_1 \text{ bat } F_2 \\ 0, & \text{sinon} \end{cases} \quad (\text{III.26})$$

La structure générale peut être étendue aux modèles, en supposant que les fractions de surperformance  $k_i$  ( $i = 1, 2, \dots, n$ ) des modèles individuels suivent la distribution de Dirichlet à  $n$  paramètres, ce qui est équivalent à la distribution bêta.

### III.2.4 Stratégie de classification

Cette stratégie est motivée par l'idée que divers prédicteurs à hautes performances peuvent être connectés à différentes zones de l'espace des variables en entrée de la série temporelle. Le meilleur modèle individuel est celui qui minimise l'erreur dans différentes parties de l'espace des variables. L'approche consiste ici, à créer une partition de l'espace variable, où chaque point de l'espace variable (série temporelle unique) est lié au prédicteur de l'ensemble qui obtient l'erreur de prédiction la plus petite. Comme illustré sur la figure III.6, l'étape d'apprentissage comprend l'identification (parmi l'ensemble des modèles de prédiction individuels existants) du meilleur prédicteur individuel  $\hat{P}(X)$  pour chaque série  $S$  de l'espace variable. La division optimale de l'espace variable en un ensemble de régions est acquise en analysant les données d'apprentissage  $S_i$  ( $i = 1, 2, \dots, M$ ), chacune étant attribuée à un seul prédicteur. Par conséquent, l'objectif de l'élément de sélection est de rechercher le meilleur moyen possible d'obtenir la partition la plus appropriée de l'espace variable, de telle sorte que :

$$\hat{P}(X) = \operatorname{argmin}_{j=1,2,\dots,N} \{|f_j(S_i) - y_i|\} \quad (\text{III.27})$$

Néanmoins, cela ne modifierait pas la partition obtenue, car le résultat de l'opérateur «argmin» n'est pas influencé par des fonctions monotones dans la phase de prévision. Chaque série inconnue  $S \in \mathbb{R}^d$  est introduite en tant qu'entrée dans le modèle de prévision. Sa sortie est alors la prédiction  $\hat{P}(X) \in \{1, 2, \dots, N\}$  du prédicteur le plus précis pour la série considérée. La valeur prédite  $\hat{F}(X)$  fournie par divers prédicteurs peut être exprimée comme suit:

$$F(X) = f_{\hat{P}(X)}(X). \quad (\text{III.28})$$

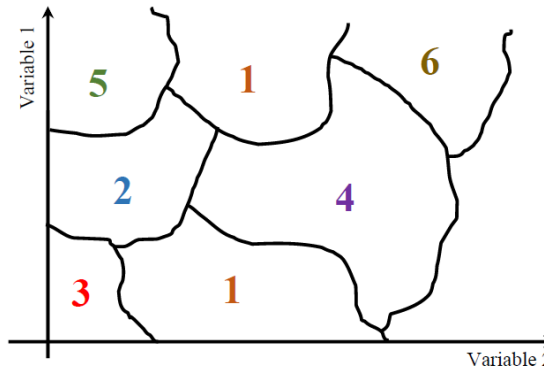


FIGURE III.6 Partition schématique d'un espace de variables bidimensionnel pour un exemple composé de six modèles. Chaque région désigne l'algorithme de prévision individuel qui fournit la précision de prédiction optimale dans la zone correspondante de l'espace variable de la série temporelle.

D'un point de vue opérationnel, l'approche de classification peut être utilisée en fonction de l'utilisation des informations contenues dans les enregistrements GHI accessibles. Cela implique deux parties: l'apprentissage et le test. Comme l'illustre la figure III.7, la phase d'apprentissage consiste à déterminer le modèle le plus probable pour chaque point de l'espace des variables. La meilleure partition de l'espace variable (constitué d'un ensemble de zones) est obtenue par l'analyse de la série de formations, chacune étant attribuée à un seul modèle de prédiction. Le but de l'approche de sélection est de construire la partition optimale de la meilleure façon possible. Ici, l'optimalité peut être formulée en termes d'erreur de prévision. Dans le cas présent, le critère de l'erreur minimale absolue (MAE) est sélectionné pour cette tâche. Dans la phase de test, chaque série (pour laquelle la valeur future de GHI est inconnue) est fournie en entrée au classifieur, qui fournit la valeur prévue du modèle le plus précis possible pour cette série [120 ,121].

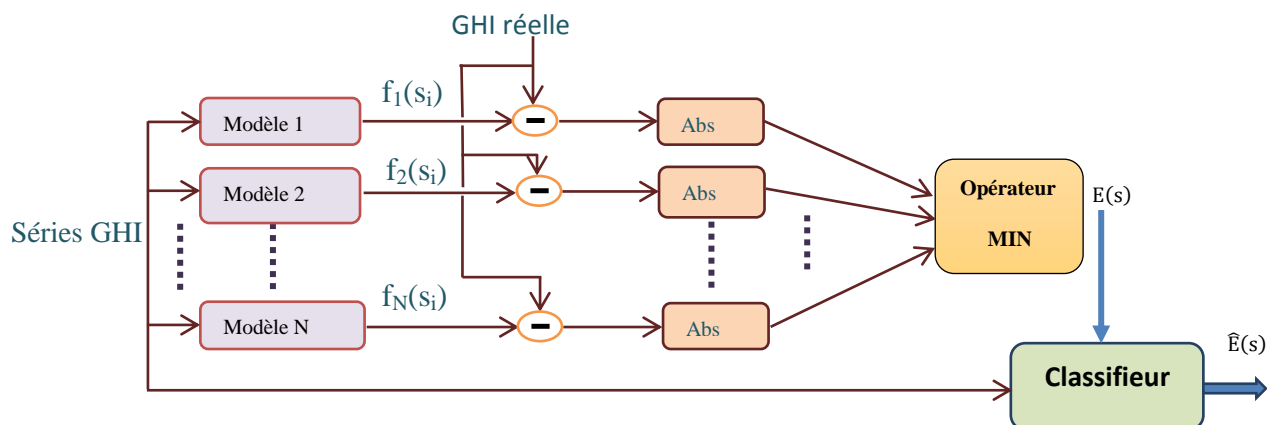


FIGURE III.7 Schéma fonctionnel du processus d'apprentissage utilisant l'approche basée sur la classification.

### III.3 Sources de données

Une base de données de haute qualité de mesures d'irradiation solaire en 1 minute provenant de quatre stations radiométriques qui observent les composants d'irradiation (y compris GHI) avec des radiomètres à thermopile, a été examinée [122]. Cette dernière référence fournit des informations détaillées sur ces stations et sur le processus de *contrôle de qualité* de leurs données. Par conséquent, le tableau III.2 donne une brève description. Les quatre stations illustrent respectivement des types de climat radiatif distincts: Aride, Tempéré, Tropical et Albédo élevé qui favorisent différents régimes de nuages et différents schémas saisonniers, comme le montre la figure III.8. Pour ce premier exemple d'application selon cette méthode, les données initiales sur une minute ont été compilées en données horaires pour prédire la prochaine période horaire, ce qui est un horizon commun à court terme dans la pratique. Les développements futurs examineront d'autres applications, telles que l'horizon de 15 minutes, qui est également très important dans la répartition de la distribution d'électricité et les négociations des prix.

TABLEAU III.1. Informations sur les quatre stations de test utilisées dans cette étude. Les valeurs GHI moyennes et maximales mesurées sont données en  $\text{W/m}^2$ .

Site	Source	Altitude [m]	Longitude [°]	Latitude [°]	Max_GHI [ $\text{W/m}^2$ ]	Moy_GHI [ $\text{W/m}^2$ ]	Période	Climat
Alice Springs	BSRN	547	133.88	-23.79	1260.7	543.2	2009–2010	Arid
Brasilia	BSRN	1023	-47.71	-15.60	1335.1	475.6	2006–2007	Tropical
Golden	NREL	1829	-05.18	39.74	1209.3	459.4	2007–2008	Tempéré
Ny-Ålesund	BSRN	11	11.93	78.92	899.0	184.3	2010–2011	Albédo élevé

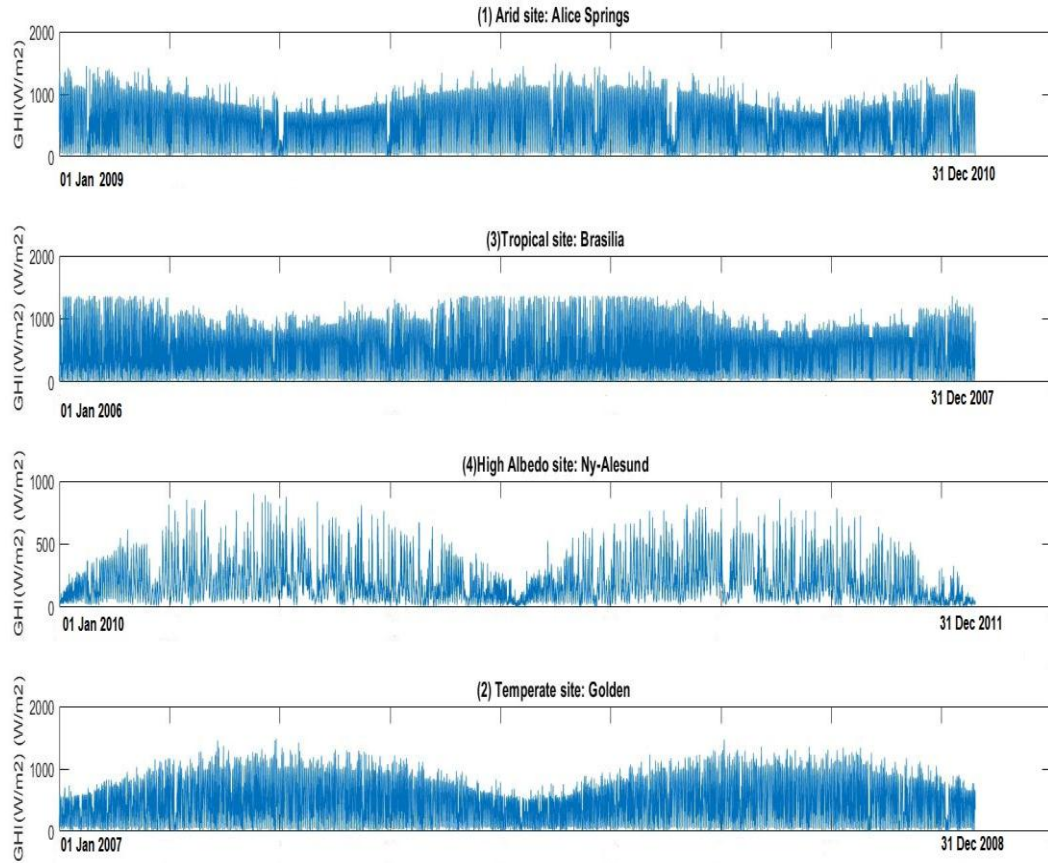


FIGURE III.8 Série chronologique de données GHI sur deux années successives, des stations (Alice Springs, Brasilia, Ny-Ålesund, and Golden, du haut jusqu'en bas).

### III.4 Estimation de l'erreur de prédiction

Différentes métriques ont été utilisées dans la littérature pour évaluer la précision des modèles de rayonnement solaire et des prévisions. Dans la présente étude, les indicateurs statistiques suivants sont utilisés.

- Coefficient de détermination ( $R^2$ ):

$$R^2 = 1 - \frac{\text{var}(y_t - \hat{y}_t)}{\text{var}(y_t)} \quad (\text{III.29})$$

- Erreur moyenne absolue en pourcentage (MAPE):

$$\text{MAPE} = \frac{1}{M} \sum_{t=1}^M \frac{|y_t - \hat{y}_t|}{|y_t|} \times 100 \quad (\text{III.30})$$

• **Erreur quadratique moyenne normalisée (NMSE):**

$$NMSE = \frac{\frac{1}{M} \sum_{t=1}^M (y_t - \hat{y}_t)^2}{\text{var}(y)} \quad (\text{III.31})$$

où  $\text{var}(y)$  est la variance des valeurs mesurées.

• **Racine de l'erreur quadratique moyenne (RMSE):**

$$RMSE = \sqrt{\frac{1}{M} \sum_{t=1}^M (y_t - \hat{y}_t)^2} \quad (\text{III.32})$$

où  $y_t$  et  $\hat{y}$  sont respectivement les valeurs mesurées et prédites et  $M$  est le nombre de séries prédites.

• **Compétence de prévision (FS):**

FS peut être exprimé comme le rapport de la performance du modèle à celle du modèle de persistance, en termes de RMSE:

$$FS = 1 - \frac{RMSE_m}{RMSE_p} \quad (\text{III.33})$$

### III.5 Résultats et discussion

Pour les expériences rapportées ici, la série chronologique GHI comprend deux années de données sur chacune des quatre stations de test. La série GHI, définie dans un espace des variables à 50 dimensions (c.-à-d. Les derniers points de l'heure à tout moment, où les périodes de nuit sont exclues), est subdivisée en trois ensembles : deux ensembles d'apprentissage (formation et validation, environ 25% de l'ensemble des données pour chacun d'eux) et un ensemble de test (environ la moitié de l'ensemble de données). L'ensemble d'apprentissage est utilisé pour apprendre les différents modèles individuels de l'ensemble, alors que l'ensemble de validation est nécessaire pour former les différents schémas de combinaison / classification supervisés proposés dans cette étude (SA, LSR, EB, DWS, OP et Class.).

#### III.5.1. Résultats obtenus avec des prédictors autorégressifs

Les prédictors autorégressifs sont la classe de modèles la plus générale pour la prédiction de séries chronologiques. Dans cette thèse, cinq modèles sont utilisés (AR, MA, ARMA, ARIMA et SARIMA). On constate que le meilleur modèle est ARIMA sur tous les

sites, probablement parce qu'il tire profit de deux modèles de base (AR et MA) avec la procédure mathématique AR. appelée «opération de différenciation» appliquée pour rendre la série immobile. En ce qui concerne les méthodes de combinaison, la méthode de classification donne les meilleurs résultats sur tous les sites (voir tableau III.2 et figure III. 9).

TABLEAU III.2. Résultats statistiques obtenus à toutes les stations de test par différents modèles autorégressifs et leurs méthodes de combinaison. Dans chaque groupe de colonnes, les meilleurs résultats apparaissent en gras.

Site/Méthode	Modèles Individuels				Méthodes de Combinaison					
	ARMA	ARIMA	SARIMA	Persist	SA	LSR	EB	DWS	OP	Class
<b>Alice Springs</b>										
FS	0.09	<b>0.11</b>	-0.07	0.04	0.09	0.22	0.11	0.03	0.11	<b>0.38</b>
R <sup>2</sup>	0.79	<b>0.80</b>	0.71	0.77	0.79	0.82	0.72	0.76	0.80	<b>0.90</b>
MABE	128.3	<b>115.8</b>	134.8	139.7	120.0	99.4	111.3	121.7	110.8	<b>71.9</b>
MAPE	25.0	<b>22.6</b>	26.3	27.4	23.4	19.7	21.8	23.7	21.6	<b>14.0</b>
RMSE	159.1	<b>155.4</b>	187.8	168.0	158.8	134.2	155.1	168.9	155.2	<b>108.1</b>
<b>Brasillia</b>										
FS	0.05	0.06	0.05	<b>0.14</b>	0.19	0.24	0.19	0.20	0.19	<b>0.43</b>
R <sup>2</sup>	0.72	0.73	0.72	<b>0.77</b>	0.80	0.83	0.79	0.80	0.80	<b>0.90</b>
MABE	128.9	<b>113.6</b>	123.0	121.0	97.6	95.3	97.6	102.8	97.4	<b>62.2</b>
MAPE	25.5	<b>22.4</b>	24.3	24.0	19.3	18.8	19.3	20.3	19.2	<b>12.3</b>
RMSE	162.6	160.5	163.5	<b>147.2</b>	138.5	129.3	137.9	136.6	138.5	<b>97.1</b>
<b>Ny-Ålesund</b>										
FS	0.06	0.04	-0.06	<b>0.08</b>	-0.03	0.08	-0.01	-0.17	0.006	<b>0.27</b>
R <sup>2</sup>	0.60	0.57	0.47	<b>0.61</b>	0.52	0.63	0.47	0.37	0.55	<b>0.76</b>
MABE	67.6	<b>60.5</b>	71.5	66.3	71.9	65.2	70.7	82.2	69.2	<b>44.7</b>
MAPE	32.7	<b>28.7</b>	34.0	32.0	34.8	31.5	34.2	39.8	33.4	<b>21.6</b>
RMSE	87.5	<b>83.1</b>	92.6	85.5	96.1	85.2	94.2	109.0	92.4	<b>67.7</b>
<b>Golden</b>										
FS	0.10	0.13	-0.02	<b>0.20</b>	0.11	0.19	0.12	-0.01	0.11	<b>0.40</b>
R <sup>2</sup>	0.76	0.77	0.69	<b>0.81</b>	0.76	0.81	0.76	0.69	0.76	<b>0.89</b>
MABE	102.1	<b>93.9</b>	106.6	94.6	92.2	89.7	92.2	103.8	92.1	<b>57.5</b>
MAPE	23.9	<b>22.0</b>	24.9	22.1	21.6	21.01	21.6	24.3	21.6	<b>13.4</b>
RMSE	132.7	128.4	150.8	<b>117.7</b>	130.8	119.1	129.6	149.2	130.5	<b>87.6</b>

### III.5.2. Résultats obtenus avec les prédicteurs neuronaux

Les réseaux de neurone sont devenus une technologie du monde réel, avec des applications efficaces dans différents domaines. La majorité de ces applications traitent des problèmes de reconnaissance de formes, principalement de classification, de régression et de regroupement. Ils exploitent des architectures de réseaux telles que le perceptron multicouche (MLP) ou le réseau à fonction de base radiale [123]. Dans la présente étude, plusieurs architectures sont étudiées, à savoir: MLP, RBF, ELM et ENN. Ici, ENN semble avoir les meilleurs résultats pour tous les indicateurs statistiques (voir tableau III.3 et figure III.10), ce qui peut s'expliquer par le fait qu'un ENN révèle le comportement dynamique temporel d'une séquence temporelle. Contrairement aux réseaux de neurones à anticipation directe

(feedforward), les ENN peuvent utiliser leur état interne (mémoire) pour traiter diverses séquences de mesures de GHI. Cela les rend plus appropriés pour ce genre de tâches.

TABLEAU III.3. Résultats des prédictions pour les quatre stations de test obtenues avec quatre modèles RNA et leurs méthodes de combinaison. Dans chaque groupe de colonnes, les meilleurs résultats apparaissent en gras.

Site/ Méthode	Modèles Individuels					Méthodes de Combinaison					
	MLP	RBF	ELM	ENN	Persist	SA	LSR	EB	DWS	OP	Class
<b>Alice Springs</b>											
FS	0.32	0.28	0.30	<b>0.34</b>	0.13	0.34	0.35	0.34	0.34	0.35	<b>0.49</b>
R <sup>2</sup>	0.89	0.87	0.87	<b>0.89</b>	0.81	0.89	0.89	0.89	0.89	0.89	<b>0.93</b>
MABE	80.6	86.9	85.0	<b>75.4</b>	129.7	76.7	74.8	76.5	76.8	76.1	<b>50.7</b>
MAPE	15.7	16.9	16.5	<b>14.7</b>	25.3	14.9	14.6	14.9	14.9	14.8	<b>9.8</b>
RMSE	118.5	125.6	122.5	<b>114.6</b>	151.9	114.3	113.0	114.2	114.2	113.9	<b>88.1</b>
<b>Brasilia</b>											
FS	0.36	0.32	0.34	<b>0.38</b>	0.12	0.38	0.42	0.39	0.35	0.39	<b>0.54</b>
R <sup>2</sup>	0.88	0.87	0.87	<b>0.88</b>	0.76	0.89	0.89	0.89	0.87	0.89	<b>0.93</b>
MABE	80.4	89.8	86.9	<b>80.2</b>	128.8	79.9	65.5	79.6	85.4	79.6	<b>50.6</b>
MAPE	15.9	17.7	17.2	<b>15.9</b>	25.5	15.8	12.9	15.7	16.9	15.7	<b>10.0</b>
RMSE	108.4	115.4	112.7	<b>106.1</b>	150.6	104.9	99.1	104.8	110.9	104.8	<b>78.8</b>
<b>Ny-Ålesund</b>											
FS	0.14	0.12	0.13	<b>0.15</b>	0.12	0.16	0.20	0.16	0.12	0.31	<b>0.37</b>
R <sup>2</sup>	0.69	0.66	0.67	<b>0.69</b>	0.62	0.70	0.71	0.69	0.66	0.79	<b>0.83</b>
MABE	61.4	60.8	59.6	<b>57.7</b>	59.2	57.2	55.2	57.2	60.7	41.6	<b>48.4</b>
MAPE	29.7	29.4	28.8	<b>27.9</b>	29.6	27.7	26.7	27.7	29.3	20.1	<b>23.4</b>
RMSE	79.5	81.5	80.0	<b>78.2</b>	80.2	77.4	73.7	77.3	81.5	63.6	<b>58.1</b>
<b>Golden</b>											
FS	0.34	0.29	0.29	<b>0.34</b>	0.20	0.35	0.38	0.35	0.35	0.35	<b>0.51</b>
R <sup>2</sup>	<b>0.877</b>	0.85	0.85	0.876	0.81	0.87	0.88	0.87	0.88	0.88	<b>0.93</b>
MABE	73.3	84.2	81.1	<b>71.5</b>	101.5	74.3	63.6	73.8	72.7	72.4	<b>47.1</b>
MAPE	17.1	19.7	19.0	<b>16.7</b>	23.7	17.4	14.9	17.3	17.0	16.9	<b>11.0</b>
RMSE	97.3	104.3	104.9	<b>96.7</b>	117.9	95.9	90.8	95.8	95.9	94.8	<b>71.5</b>

### III.5.3. Résultats obtenus avec les prédicteurs SVM

La sélection de ce type de prédicteur vient de l'intérêt porté à une évaluation globale et large de l'efficacité de la technique SVM lorsqu'elle est utilisée dans les problèmes de prédiction de GHI. Plus précisément, trois différents types de SVM sont considérés : un SVM linéaire (SVM-Linear; égal à un SVM sans transformation avec noyau), un SVM avec un noyau polynomial (SVM-Polynomial) et un SVM avec un noyau gaussien (SVM-RBF).

L'ajout de ce dernier type de SVM permet d'évaluer l'effet de la transformation noyau sur la méthode de prévision SVM. De plus, des indications précieuses sont ensuite obtenues sur les modèles de prédiction à choisir pour mettre en œuvre les diverses stratégies de combinaison et évaluer l'efficacité de l'ensemble. Comme indiqué dans le tableau III.4, les performances du SVM-RBF sont optimales sur tous les sites de test, comme l'indiquent toutes les mesures de performance. Cela peut être lié à la flexibilité du noyau RBF pour modéliser

les phénomènes non linéaires (voir figure III.11). En ce qui concerne les méthodes de combinaison, la stratégie basée sur la classification surpasse largement les autres.

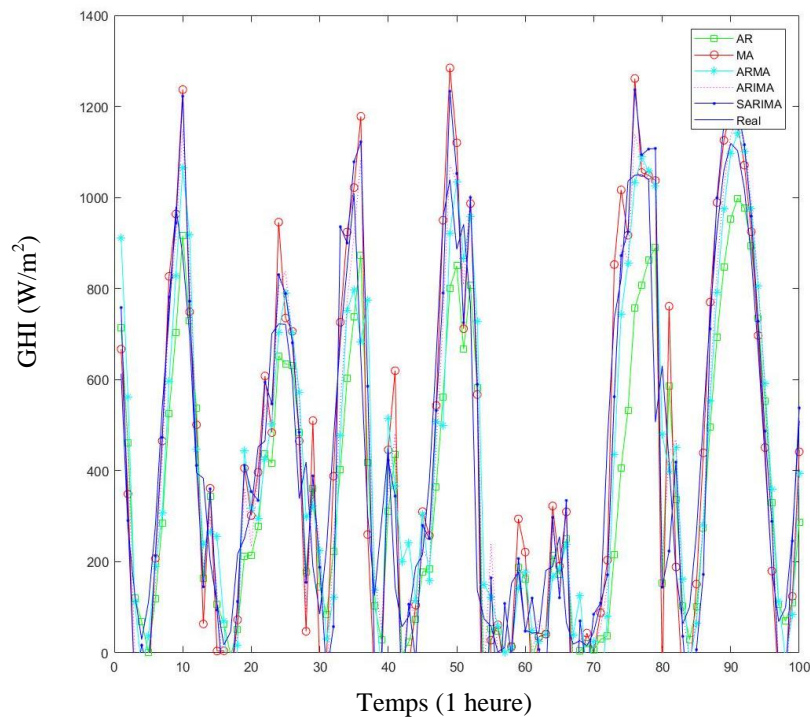


FIGURE III.9 : Une petite fenêtre sur les séries temporelles mesurées et prédites par différents modèles autorégressifs sur le site d'Alice Springs.

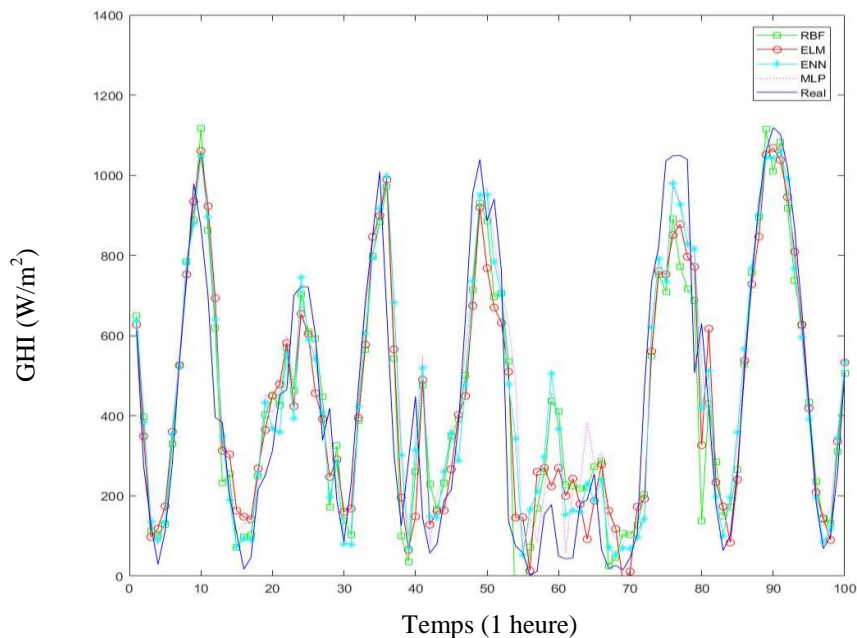


FIGURE III.10 Une petite fenêtre sur les séries temporelles mesurées et prédites par différents modèles d'RNA sur le site de test d'Alice Springs.

TABLEAU III.4. Résultats de prédiction obtenus avec trois modèles à noyau SVM et leurs méthodes de combinaison aux quatre stations de test. Dans chaque colonne, les meilleurs résultats apparaissent en gras.

Site/ Méthode	Modèles Individuels				Méthodes de Combinaison					
	SVM- RBF	SVM- Poly	SVM- Lin.	Persist.	SA	LSR	EB	DWS	OP	Class
<b>Alice Springs</b>										
FS	<b>0.35</b>	0.27	0.27	0.14	0.33	0.35	0.34	0.35	0.34	<b>0.48</b>
R <sup>2</sup>	<b>0.89</b>	0.86	0.86	0.81	0.89	0.89	0.89	0.89	0.89	<b>0.93</b>
MABE	<b>75.8</b>	86.2	89.5	127.3	78.6	76.05	78.3	75.8	77.9	<b>73.2</b>
MAPE	<b>14.7</b>	16.8	17.4	24.8	15.3	14.8	15.2	14.7	15.1	<b>10.7</b>
RMSE	<b>113.6</b>	126.6	126.9	149.2	115.8	113.5	115.67	113.6	115.3	<b>91.0</b>
<b>Brasilia</b>										
FS	<b>0.39</b>	0.38	0.37	0.16	0.39	0.42	0.39	0.39	0.39	<b>0.48</b>
R <sup>2</sup>	<b>0.89</b>	0.88	0.88	0.78	0.89	0.89	0.89	0.89	0.89	<b>0.92</b>
MABE	<b>80.5</b>	81.4	84.1	123.5	81.1	66.7	81.1	80.5	81.3	<b>63.4</b>
MAPE	<b>15.9</b>	16.1	16.6	24.5	16.0	13.2	16.0	15.9	16.1	<b>12.5</b>
RMSE	<b>105.0</b>	106.3	108.3	144.3	104.5	99.0	104.5	104.8	104.5	<b>88.6</b>
<b>Ny-Ålesund</b>										
FS	<b>0.15</b>	0.13	0.07	0.11	0.14	0.18	0.14	0.15	0.15	<b>0.26</b>
R <sup>2</sup>	<b>0.68</b>	0.66	0.63	0.40	0.68	0.70	0.68	0.68	0.68	<b>0.76</b>
MABE	<b>57.7</b>	59.6	65.2	78.2	58.4	56.1	58.3	57.5	58.2	<b>45.1</b>
MAPE	<b>27.9</b>	28.8	31.5	33.1	28.2	27.1	28.2	27.8	28.1	<b>21.8</b>
RMSE	<b>78.2</b>	80.5	85.8	99.5	79.3	75.7	79.	78.2	79.0	<b>68.1</b>
<b>Golden</b>										
FS	<b>0.34</b>	0.33	0.31	0.28	0.34	0.35	0.34	0.34	0.34	<b>0.45</b>
R <sup>2</sup>	<b>0.87</b>	0.86	0.86	0.84	0.87	0.87	0.87	0.87	0.87	<b>0.91</b>
MABE	<b>72.3</b>	75.4	78.4	92.6	73.2	70.1	73.1	72.3	73.2	<b>56.6</b>
MAPE	<b>16.9</b>	17.6	18.3	21.6	17.1	16.4	17.1	16.9	17.1	<b>13.2</b>
RMSE	<b>97.2</b>	98.7	101.1	106.3	96.3	95.4	96.3	97.2	96.3	<b>80.0</b>

#### III.5.4. Résultats obtenus avec un ensemble de prédictions hybride

Dans ce cas, pour chaque famille de prédictions, le plus performant est sélectionné. Le meilleur modèle individuel de l'ensemble semble être SVM-RBF. On pouvait s'y attendre dans une certaine mesure, en raison de la meilleure capacité de généralisation des SVM aux problèmes d'apprentissage. En ce qui concerne les stratégies de combinaison, comme pour les combinaisons précédentes, la stratégie basée sur la classification surpasse largement les autres (voir le tableau III.5 et la figure III.12).

TABLEAU III.5. Résultats de prédictions obtenus avec le meilleur modèle de chaque famille de prédiction (SVM-RBF, ENN et ARIMA) et leurs méthodes de combinaison aux quatre stations de test. Dans chaque groupe de colonnes, les meilleurs résultats apparaissent en gras. Les modèles individuels incluent le modèle ciel clair (CS) [124] et le Persistence (Persist).

Site/Method	Modèles Individuels				Méthodes de Combinaison						
	ARIMA	ENN	SVM-RBF	CS	Persist.	SA	LSR	EB	DWS	OP	Class
<b>Alice Springs</b>											
FS	0.11	0.34	<b>0.35</b>	0.15	0.17	0.35	0.36	0.36	0.35	0.36	<b>0.51</b>
R <sup>2</sup>	0.80	0.89	<b>0.89</b>	0.58	0.82	0.89	0.90	0.89	0.89	0.89	<b>0.94</b>
MABE	115.8	<b>75.4</b>	75.8	140.1	121.0	75.9	71.4	73.9	75.7	73.4	<b>49.3</b>
MAPE	22.6	<b>14.7</b>	14.8	27.3	23.6	14.8	13.9	14.4	14.7	14.3	<b>9.6</b>
RMSE	155.4	114.6	<b>113.6</b>	260.0	145.4	113.6	110.8	112.1	113.5	111.4	<b>84.5</b>
<b>Brasillia</b>											
FS	0.06	0.38	<b>0.39</b>	0.26	0.11	0.40	0.42	0.41	0.40	0.39	<b>0.59</b>
R <sup>2</sup>	0.73	0.88	<b>0.89</b>	0.71	0.75	0.89	0.89	0.88	0.89	0.88	<b>0.94</b>
MABE	113.6	<b>80.2</b>	80.5	91.5	131.4	73.2	65.6	73.1	77.0	73.4	<b>46.0</b>
MAPE	22.4	<b>15.8</b>	15.9	18.1	26.0	14.5	12.9	14.4	15.2	14.5	<b>9.1</b>
RMSE	160.5	106.1	<b>105.0</b>	186.3	152.5	102.8	99.2	100.4	101.8	104.1	<b>70.3</b>
<b>Ny-Ålesund</b>											
FS	0.04	0.15	<b>0.15</b>	-0.01	0.10	0.20	0.20	0.14	0.17	0.13	<b>0.38</b>
R <sup>2</sup>	0.57	<b>0.69</b>	0.68	-1.29	0.70	0.50	0.74	0.41	0.70	0.66	<b>0.83</b>
MABE	60.5	57.8	<b>57.7</b>	155.6	58.8	78.3	26.4	61.0	56.6	62.8	<b>39.3</b>
MAPE	28.7	27.9	<b>27.9</b>	117.4	28.4	37.8	54.7	29.5	27.3	30.4	<b>19.0</b>
RMSE	83.1	78.2	<b>78.2</b>	187.9	75.2	98.1	71.1	79.6	76.8	80.8	<b>56.8</b>
<b>Golden</b>											
FS	0.13	<b>0.34</b>	0.34	0.20	0.17	0.36	0.37	0.36	0.35	0.36	<b>0.53</b>
R <sup>2</sup>	0.77	<b>0.87</b>	0.87	0.60	0.79	0.88	0.88	0.87	0.87	0.88	<b>0.93</b>
MABE	93.9	71.5	<b>72.3</b>	106.0	102.5	67.0	65.0	66.8	68.8	66.9	<b>43.5</b>
MAPE	22.0	<b>16.7</b>	16.9	24.8	23.9	15.7	15.2	15.6	16.1	15.6	<b>10.1</b>
RMSE	128.4	<b>96.7</b>	97.2	197.0	121.5	94.2	92.6	93.3	95.5	93.8	<b>68.2</b>

### III.5.5. Résultats obtenus avec la sélection des variables

Une autre façon d'étudier les performances de la méthodologie de prédiction proposée consiste à utiliser une étape de traitement additif avant d'introduire la série temporelle dans les différents prédicteurs. Ici, la méthode de sélection de caractéristiques décrite dans [112] est utilisée et le meilleur ensemble de variables historiques sélectionnées est obtenu à l'aide du principe de sélection de modèle [125]. Pour chaque cas, le meilleur ensemble est obtenu en fonction des performances de la technique de prédiction dans l'ensemble de validation. Le tableau III.6 montre une légère amélioration des performances par rapport aux résultats du tableau III.5 (sans sélection de caractéristiques). Une exception importante est celle du modèle autorégressif (ARIMA), où les performances de prédiction dégradent.

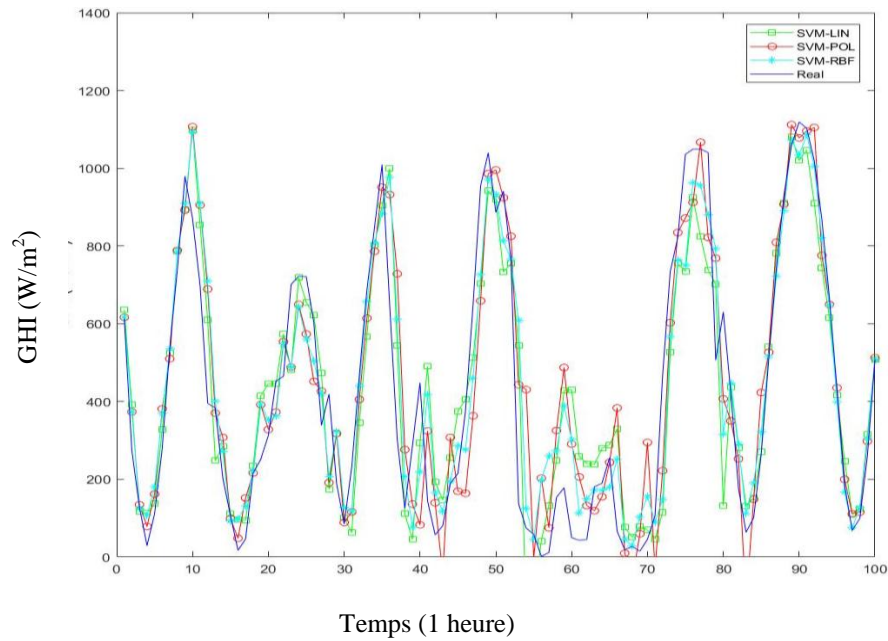


FIGURE III.11. Une petite fenêtre sur les séries temporelles mesurées et prédites par différents modèles de SVM pour le site de test d'Alice Springs.

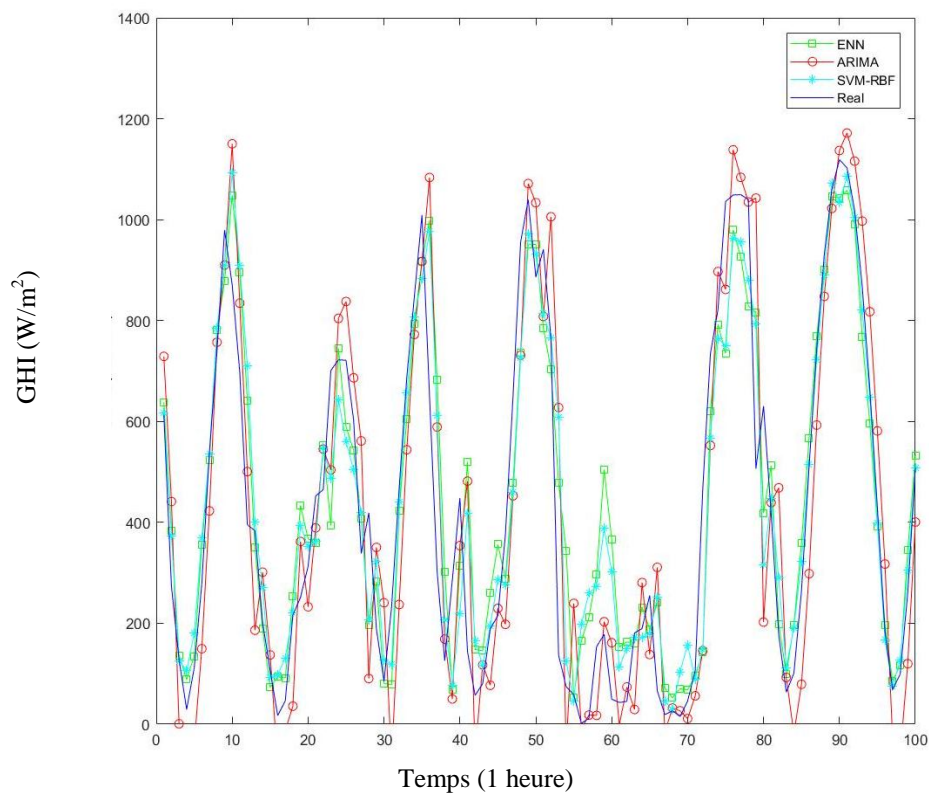


FIGURE III.12. Comparaison entre les modèles les plus performants pour chaque famille de prédiction sur le site de test d'Alice Springs.

Cela peut s'expliquer par le fait que les modèles ARIMA sont basés sur la modélisation polynomiale et nécessitent davantage de variables historiques pour comprendre le comportement de la série chronologique du signal de rayonnement solaire.

Néanmoins, lorsqu'on utilise différentes techniques de combinaison, le système conserve le même niveau de performance malgré les mauvais résultats de la technique ARIMA. Cela confirme l'utilité de la stratégie de combinaison lorsque les techniques d'entrée sont suffisamment diversifiées (différentes familles de prédiction).

Un autre moyen d'analyse pour comparer les performances des différents modèles consiste à utiliser le diagramme de Taylor [126], qui est une illustration graphique de la connexion statistique entre deux variables aléatoires. Il caractérise trois statistiques reliant deux variables en même temps, à savoir : la corrélation, l'écart type et la différence quadratique moyenne. Ces graphiques sont particulièrement adaptés pour évaluer plusieurs aspects des modèles complexes ou pour juger la compétence relative des différents modèles

Dans l'espace bidimensionnel d'un diagramme de Taylor, chaque point peut décrire la combinaison de trois statistiques énoncées ci-dessus, elles sont liées entre elles par :

$$E'^2 = \sigma_t^2 + \sigma_m^2 - 2\sigma_t\sigma_m R \quad (\text{III.34})$$

où  $E'$  est la différence quadratique moyenne centrée entre la valeur mesurée et celle prédite,  $R$  est le coefficient de corrélation entre la série mesurée et la série prédite et  $\sigma_m$  et  $\sigma_f$  sont les variances des champs de test et de modèle, respectivement. La différence RMS centrée entre les séries temporelles GHI prédites et mesurées est relative à la distance entre la prévision et le point de l'axe des  $x$  identifié comme "Mesuré". De même, l'écart type de la série prédite est proportionnel à la distance radiale entre la prédiction et l'origine.

En utilisant les mesures réelles et les modèles de prédiction, les diagrammes de Taylor pour les différents sites sont montrés aux figures III.13-16. Ces diagrammes indiquent que toutes les valeurs de corrélation entre les séries chronologiques mesurées et prévues se situent dans l'intervalle [0,95–0,99] pour Alice Springs, Ny-Ålesund et Brasilia et dans l'intervalle [0,90–0,95] pour le site de Golden. Ces diagrammes indiquent également que la stratégie de classification surpasse tous les autres modèles de combinaison. Il est à noter que le diagramme de Taylor ne fournit pas d'informations

sur les biais généraux, mais décrit simplement l'erreur du modèle centrée.

TABLEAU III.6. Résultats des prédictions obtenus à l'aide du système à deux blocs (technique de sélection des caractéristiques et de régression). Dans chaque groupe de colonnes, les meilleurs résultats apparaissent en gras. SV: variables historiques sélectionnées.

Site/Méthode	Modèles Individuels			Méthodes de combinaison					
	ARIMA	ENN	SVM_ RBF	SA	LSR	EB	DWS	OP	Class
<b>Alice Springs</b>	<b>SV=15</b>	<b>SV=12</b>	<b>SV=19</b>						
FS	-1.62	0.35	0.35	-0.09	0.37	0.28	0.35	0.31	<b>0.48</b>
R <sup>2</sup>	-0.72	0.89	0.89	0.70	0.90	0.88	0.89	0.88	<b>0.93</b>
MABE	386.5	77.5	75.5	159.2	14.4	91.0	75.6	89.4	<b>57.1</b>
MAPE	75.3	15.1	14.7	31.0	74.1	17.7	14.7	17.4	<b>11.1</b>
RMSE	460.4	113.0	113.2	191.6	109.3	124.9	112.9	120.1	<b>91.1</b>
<b>Brasilia</b>	<b>SV=12</b>	<b>SV=11</b>	<b>SV=16</b>						
FS	-1.31	0.38	0.38	-0.001	0.43	0.31	-0.02	0.30	<b>0.50</b>
R <sup>2</sup>	-0.65	0.895	0.89	0.69	0.90	0.85	0.68	0.85	<b>0.92</b>
MABE	334.1	81.4	81.3	144.8	65.5	94.5	146.8	97.3	<b>62.1</b>
MAPE	66.1	16.1	16.0	28.6	12.9	18.7	29.0	19.2	<b>12.2</b>
RMSE	397.4	105.5	105.3	172.3	96.7	118.3	175.7	119.6	<b>85.7</b>
<b>Ny-Ålesund</b>	<b>SV=17</b>	<b>SV=14</b>	<b>SV=11</b>						
FS	-1.83	-0.20	-0.15	-0.17	0.12	-0.04	-0.15	-0.04	<b>0.23</b>
R <sup>2</sup>	-0.66	0.78	0.78	0.75	0.84	0.75	0.80	0.82	<b>0.89</b>
MABE	113.5	43.1	39.9	47.3	29.6	40.0	41.7	40.3	<b>26.9</b>
MAPE	85.7	32.5	30.1	35.6	22.4	30.2	31.4	30.4	<b>20.3</b>
RMSE	145.1	61.5	58.9	60.3	44.6	53.6	59.1	53.5	<b>38.9</b>
<b>Golden</b>	<b>SV=17</b>	<b>SV=17</b>	<b>SV=13</b>						
FS	-0.60	0.36	0.33	0.17	0.39	0.30	-0.39	0.31	<b>0.53</b>
R <sup>2</sup>	0.23	0.88	0.87	0.80	0.89	0.85	0.41	0.86	<b>0.93</b>
MABE	193.2	70.8	78.2	99.7	14.8	81.9	169.7	81.6	<b>49.6</b>
MAPE	45.2	16.6	18.3	23.3	63.3	19.1	39.7	19.1	<b>11.6</b>
RMSE	236.6	94.0	98.9	121.3	89.6	102.4	206.7	101.5	<b>68.4</b>

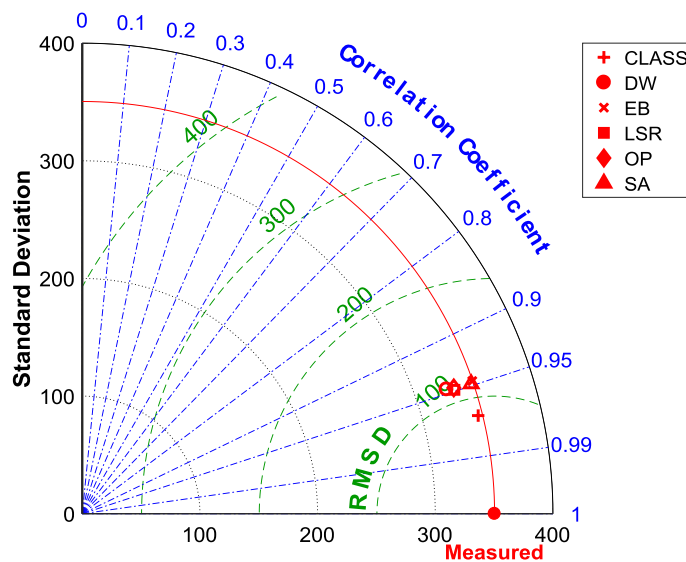


FIGURE III. 13. Diagramme de Taylor des valeurs de GHI mesurées et prédites par différents modèles de combinaison pour le site de Alice Springs.

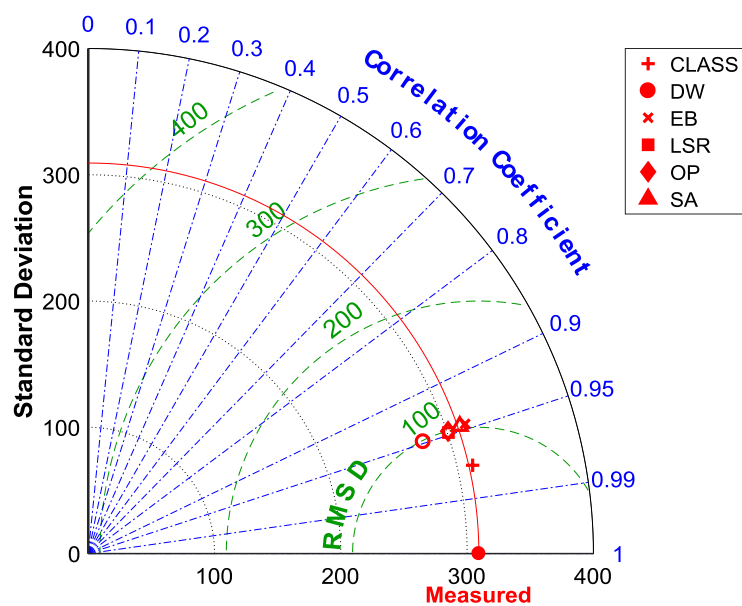


FIGURE III. 14. Diagramme de Taylor des valeurs GHI mesurées et prédites par différents modèles de combinaison pour le site de Brasilia.

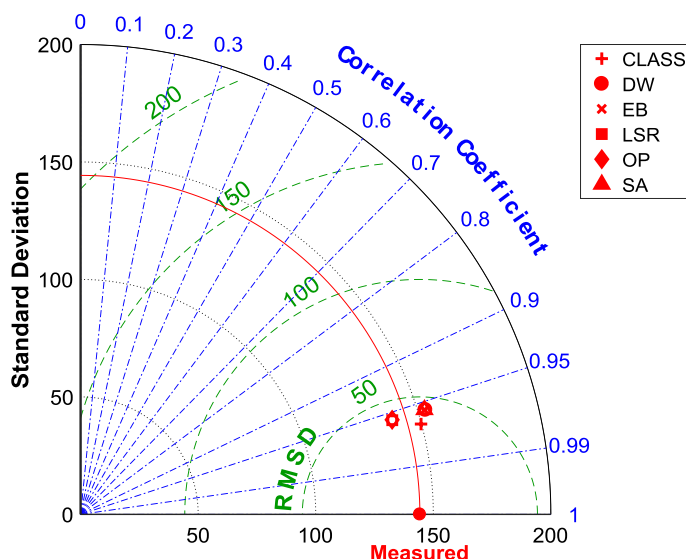


FIGURE III. 15. Diagramme de Taylor des valeurs GHI mesurées et prédites par différents modèles de combinaison pour le site de Ny-Ålesund.

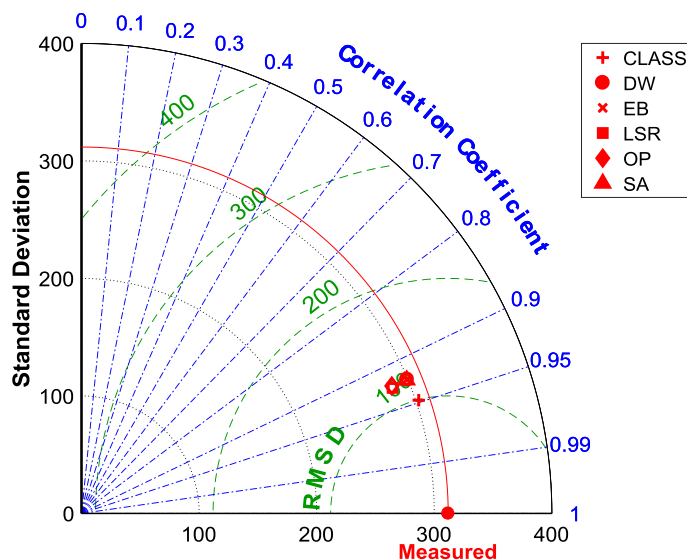


FIGURE III. 16. Diagramme de Taylor des valeurs GHI mesurées et prédites par différents modèles de combinaison pour le site de Golden.

### II.5.6 Test d'hypothèse

Le test d'hypothèse statistique utilisé dans cette étude est principalement basé sur le test  $t$  apparié et le test de Fisher signé. Ils sont utiles pour déterminer si une approche de combinaison particulière des meilleurs modèles individuels, à savoir (ARIMA) du modèle autorégressif, (MLP) des réseaux neuronaux artificiels et (SVM-RBF) des machines à vecteurs de support, est plus performante que la meilleure stratégie de combinaison. Cette dernière est la stratégie de classification [127,128]. Pour effectuer les tests statistiques, l'ensemble de données complet est subdivisé en  $M$  sous-ensembles de sorte que, chaque

modèle soit formé en sous-division  $i$  et testé en sous-division  $i + 1$ . Pour obtenir un groupe de 51 sous-ensembles différents, dans la présente étude,  $K$  est fixé à 50. Le test de *Kolmogorov–Smirnov* ( $K - S$ ) peut ensuite être exécuté pour confirmer la normalité des données. Si en effet la distribution des données est normale, un test  $t$  est exécuté pour comparer les performances des diverses stratégies combinées et découvrir les meilleurs. Sinon Le test non paramétrique de Fisher signé est effectué pour obtenir la même comparaison.

Le tableau III.6 compare les différentes stratégies de fusion et la méthode de classification proposée sur la base de test d'Alice Springs. Le niveau de signification,  $\alpha$ , est fixé à 0,05, ce qui correspond à un niveau de confiance de 95%. Les résultats du test  $K-S$  et ceux du  $t$ -test ou du test de Fisher sont également présentés. De plus, pour chaque cas, les valeurs *Win-Lose-Tie* ( $W-L-T$ ) sont fournies. Ici, la valeur probabiliste  $p$  est la probabilité que, lorsque l'hypothèse nulle est vraie, la performance statistique de la méthode de classification proposée est identique ou supérieure à celle d'autres scénarios. En d'autres termes, une valeur  $p$  inférieure indique une amélioration des performances du modèle proposé. D'après le tableau III.7, il est évident que l'approche de classification surpasse statistiquement toutes les autres stratégies de combinaison, comme en témoignent les valeurs de  $W-L-T$ .

TABLEAU III.7 Résultats du test d'hypothèse comparant la stratégie de classification à d'autres stratégies de combinaison pour le site d'Alice Springs.

Class. vs. autres méthodes de combinaison	W-L-T	Test de Fisher $p$	Test K-S $p$	Test $t$ $p$
LSR	27-22-1	—	0.840	0.900
OP	26-22-2	—	0.150	0.910
EB	44-6-0	—	0.004	0.020
DWS	36-14-0	—	0.030	0.650
SA	47-0-3	—	0.190	0.710

### III.6. Conclusion

L'analyse et la prédiction des séries temporelles constituent un domaine de recherche dynamique, qui revêt une importance capitale dans de nombreux domaines pratiques. L'amélioration de l'exactitude des prédictions de séries chronologiques est une tâche qui retient de plus en plus l'attention des chercheurs au cours des dernières décennies. Dans ce chapitre, une nouvelle méthodologie a été proposée pour prévoir l'irradiation solaire globale horizontale (GHI) de l'heure qui suit, sur la base d'un ensemble multi modèles. La méthodologie vise à exploiter de la diversité entre un ensemble de modèles de prédiction

variés, dans le but d'améliorer la précision du processus de prédiction. Plusieurs stratégies ont été décrites et mises en œuvre. Les différentes stratégies varient les unes des autres par rapport à : (i) la procédure de combinaison ; et (ii) la technique utilisée pour exploiter les connaissances antérieures accessibles de la série chronologique horaire de GHI.

Les divers ensembles multi modèles utilisés dans les expériences actuelles reposent sur différentes familles de prédictions, principalement : (1) modèles autorégressifs ; (2) modèles neuronaux ; et (3) modèles de machines à vecteurs de support. Les méthodologies proposées ici peuvent améliorer les performances de la prédiction de GHI, en combinant différentes techniques, simplement avec la condition de diversité, c'est-à-dire en utilisant différentes familles de prédicteurs.

Afin de réaliser une évaluation cohérente et complète de l'efficacité de la méthodologie proposée, différentes expériences (avec différentes architectures) ont été considérées. Les expériences ont simulé plusieurs architectures, dans lesquelles les membres de l'ensemble étaient soit basés sur la même famille de prédiction, soit sur le meilleur modèle de chaque famille de prédiction inclus dans l'ensemble.

L'évaluation expérimentale réalisée pour évaluer l'efficacité du système multi modèle a clairement montré qu'il pouvait fournir une précision et des compétences en prédiction meilleures que le meilleur modèle unique introduit dans l'ensemble, avec des performances plus élevées lors de l'utilisation de la stratégie de classification. Ce résultat est particulièrement encourageant car les compétences améliorées de la méthode d'ensemble proposée sont maintenues sur les quatre sites de test utilisés ici, lesquels sont représentatifs de divers climats et régimes de nuages. Les développements ultérieurs de la méthode permettront d'évaluer ses compétences en matière de prédiction à plus court terme, par exemple pour l'horizon 15 min.