

# THESE

En vue de l'obtention du : **DOCTORAT**

Structure de Recherche : Laboratoire de physique des Hautes Énergies  
Modélisation et  
Simulation (LPHE-MS)

Discipline : **Physique**

Spécialité : **Modélisation et simulation**

Présentée et soutenue le : 09/01/2021 par :

**Soukaina BARHMI**

## Nouvelles approches de machine learning pour la modélisation de la vitesse du vent et la prédiction énergétique

### JURY

|                            |     |                                                                                      |                      |
|----------------------------|-----|--------------------------------------------------------------------------------------|----------------------|
| <b>Souad EL HAJJAJI</b>    | PES | Faculté des Sciences, Université Mohammed V, Rabat                                   | Présidente           |
| <b>Omkaltoume EL FATNI</b> | PH  | Faculté des Sciences, Université Mohammed V, Rabat                                   | Directrice de thèse  |
| <b>El Hassan SAIDI</b>     | PES | Faculté des Sciences, Université Mohammed V, Rabat                                   | Rapporteur/Examineur |
| <b>Mohamed KISSI</b>       | PES | Faculté des Sciences et Techniques, Mohammedia FSTM, Université Hassan II Casablanca | Rapporteur/Examineur |
| <b>Rachid AHLAAMARA</b>    | PH  | Faculté des Sciences, Université Mohammed V, Rabat                                   | Rapporteur/Examineur |
| <b>Abdellah IDRISI</b>     | PH  | Faculté des Sciences, Université Mohammed V, Rabat                                   | Examineur            |

Année Universitaire 2020- 2021



## Dédicaces

Je dédie ce travail

À la mémoire de mon père

À ma très chère mère

À mon mari chakib El amrani

À mon cher bébé Mahmoud

À ma sœur, son mari sif Eddine et son fils mouad

À mes frères Fayçal, Jalil, Talal et Youness

À ma chère amie Ikram

À tous ceux qui m'ont aidé de près ou de loin dans la  
la réalisation de ce travail

### Remerciements

Ce travail de thèse a été réalisé au sein du Laboratoire de Physique des Hautes Énergies- Modélisation et Simulation (LPHE-MS), du département de physique de la faculté des sciences de Rabat, sous la direction de Madame Omkaltoume EL FATNI , Professeur Habilité à la faculté des sciences de rabat.

Je voudrais tout d'abord exprimer mes sincères remerciements à ma directrice de thèse, madame El FATNI Omkaltoume, Professeur Habilité à la faculté des sciences Rabat, pour m'avoir encadré et proposer ce sujet de thèse, pour sa disponibilité, son encouragement, sa confiance. Je la remercie également pour ses conseils, sa patience ainsi que sa grande qualité humaine qui m'a donné un grand soutien durant toutes ces années.

Je remercie Madame Souad EL HAJJAJI, Professeur de l'enseignement supérieur à la faculté des sciences Rabat, pour avoir accepté de me faire l'honneur de présider le jury de ma soutenance.

Mes vifs remerciements vont à Monsieur EL Hassan SAIDI, Professeur de l'Enseignement Supérieur à la Faculté des Sciences Rabat pour le grand honneur qu'il m'a donné en acceptant d'évaluer ce travail et d'avoir accepté de participer en tant que rapporteur dans le jury de ma soutenance.

J'adresse aussi mes remerciements à Mohamed KISSI, Professeur de l'Enseignement Supérieur à la Faculté des Sciences Techniques, Mohammadia, Casablanca, d'avoir accepté de rapporter et examiner ce travail de thèse. Je le remercie de l'intérêt qu'il a manifesté à l'égard de cette recherche en s'engageant à être rapporteur, ainsi que pour ses jugements très pertinents sur mon manuscrit.

Je suis particulièrement reconnaissant à Monsieur Rachid AHL LAAMARA, Professeur Habilité à la Faculté des Sciences de Rabat, d'avoir accepté de rapporter et examiner ce travail et de m'avoir fait profiter de ses expériences durant mes études en cycle de Master. Veuillez recevoir, Monsieur, l'expression de mes respectueux sentiments.

Je tiens également à remercier Monsieur, Abdellah IDRISSE, Professeur Habilité à la Faculté des Sciences Rabat, pour avoir accepté d'examiner ce travail et de faire partie du jury.

Je remercie ma mère et ma famille pour leurs aides et sacrifices.

Je remercie ma petite famille (mon mari et mon bébé adorable Mahmoud) pour leurs patiences et sacrifices durant toutes ces années.

Un grand merci à ma chère ikram qui m'a beaucoup encouragé et aidé durant toute la durée de cette thèse.

## *Résumé*

Le développement et l'exploitation des énergies renouvelables ont connu une forte augmentation ces dernières années dans le monde et plus particulièrement au Maroc vu sa situation géographique favorable qui montre un potentiel solaire et éolien très important.

Le travail de notre thèse met l'accent sur la prévision de la vitesse du vent qui est la composante principale du système éolien. Pour cette raison, nous avons développé différents modèles en utilisant les techniques de modélisation et d'intelligence artificielle. Une comparaison de la performance de ces méthodes est faite en se basant sur des indicateurs de performance qui seront exposés ultérieurement dans ce manuscrit.

Les sites étudiés sont Tanger au Nord et Tarfaya au Sud du Maroc en raison de leurs positions géographiques où la vitesse du vent dépasse 10m/s. La base de données utilisée contient trois années de données satellitaires météorologiques horaires.

Une étude statistique des données montre une colinéarité entre les paramètres d'entrée qui influencent la sortie du système et qui peut avoir une explication sur son comportement non linéaire. Les résultats obtenus montrent que les méthodes non linéaires ont une précision bien meilleurs que les méthodes linéaires.

**Mots clés :** prévision, vitesse du vent, modélisation, intelligence artificielle.

### *Abstract*

The development and exploitation of renewable energies have experienced a strong increase in recent years in the world and more particularly in Morocco given its favorable geographical location which shows a very important solar and wind potential.

The work of our thesis focuses on the prediction of the wind speed which is the main component of the wind system. For this reason, we have developed different models using modeling and artificial intelligence techniques. A comparison of the performance of these methods is made based on performance indicators that will be discussed later in this manuscript.

The regions studied are Tangier in the north and Tarfaya in the south of Morocco because of its geographical positions where the wind speed exceeds 10m/ s. The database used contains three years of hourly meteorological satellite data.

A statistical study of the data shows a collinearity between the input parameters that influence the system output and which may have an explanation for its non-linear behavior. The results obtained show a better performance of the non-linear methods to the linear methods.

**Keywords:** Forecast, wind speed, modeling, artificial intelligence.



# Table des matières

Remerciements

Introduction générale..... 19

**Chapitre 1. Notions et état de l'art sur l'énergie éolienne.....21**

1.1.Introduction.....21

1.2. Notions sur l'énergie éolienne..... 22

1.2.1. Historique de l'éolien ..... 22

1.2.2. Premier pas vers les éoliennes..... 22

1.2.3. Retour à l'éolien : une source de l'énergie de demain..... 23

1.2.4. Enjeux de l'éolienne..... 24

1.2.4.1. Sur le plan environnemental..... 24

1.2.4.2. Sur le plan économique..... 24

1.2.5. Futurs challenges et perspectives mondiales ..... 24

1.2.6. L'énergie éolienne dans le contexte international ..... 24

1.2.7. L'énergie éolienne dans le contexte national..... 27

1.2.7.1. Historique ..... 27

1.2.7.2. Répartition régionale de la vitesse du vent ..... 29

1.2.8. Avantages et inconvénients d'énergie éolienne ..... 31

1.3. Descriptif d'une éolienne..... 32

1.3.1. Eolienne à axe horizontal ..... 32

1.3.2. Eolienne à axe vertical ..... 33

1.3.3. Principaux constituants d'une éolienne à axe horizontal ..... 33

1.3.4. Un potentiel énorme encore peu exploité : l'éolien offshore.... 35

1.4. La source d'énergie éolienne : vitesse du vent..... 36

1.4.1. Origine du vent..... 36

1.4.2. Appareils de mesure ..... 37

□ Anémomètre :..... 37

|                                                                                    |                                                                                       |           |
|------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|-----------|
| □                                                                                  | Girouette :.....                                                                      | 37        |
| 1.5.                                                                               | Puissance récupérable par une turbine.....                                            | 37        |
| 1.6.                                                                               | Transformation de l'énergie mécanique du vent en énergie électrique par éolienne..... | 39        |
| 1.7                                                                                | Revue des techniques de prévision des vitesses de vent .....                          | 39        |
| 1.8                                                                                | Méthodes de prévision.....                                                            | 40        |
| 1.9                                                                                | Etat de l'art.....                                                                    | 41        |
| 1.10                                                                               | Conclusion.....                                                                       | 41        |
| <b>Chapitre 2. Généralités sur les techniques de machine learning utilisées ..</b> |                                                                                       | <b>43</b> |
| 2.1                                                                                | Introduction.....                                                                     | 43        |
| 2.2.                                                                               | Distinction entre la prédiction et la prévision.....                                  | 43        |
| 2.3.                                                                               | Les horizons de prévision.....                                                        | 43        |
| 2.4                                                                                | Différentes méthodes de prévision proposées .....                                     | 44        |
| 2.4.1                                                                              | Modélisation statistique par régression linéaire.....                                 | 44        |
| 2.4.1.1                                                                            | Régression simple .....                                                               | 45        |
| 2.4.1.2                                                                            | Régression polynomiale .....                                                          | 46        |
| 2.4.1.3                                                                            | Régression Multiple .....                                                             | 47        |
| 2.4.1.4                                                                            | Analyse de la variance (ANOVA) .....                                                  | 47        |
| 2.4.1.5                                                                            | Examiner les résidus.....                                                             | 49        |
| 2.4.2.                                                                             | Modélisation par réseaux de neurones artificiels .....                                | 50        |
| 2.4.2.1.                                                                           | Qu'est-ce qu'un neurone biologique ? .....                                            | 50        |
| 2.4.2.2.                                                                           | Qu'est-ce qu'un neurone formel ? .....                                                | 51        |
| 2.4.2.3.                                                                           | Qu'est-ce qu'un réseau de neurones ? .....                                            | 51        |
| 2.4.2.4.                                                                           | Architecture d'un réseau de neurones artificiels .....                                | 52        |
| 2.4.2.5.                                                                           | Modélisation d'un réseau de neurones artificiels .....                                | 54        |
| 2.4.2.5.1.                                                                         | Réseau neuronal monocouche .....                                                      | 54        |
| 2.4.2.5.2                                                                          | Réseau neuronal multicouche.....                                                      | 54        |
| 2.4.2.5.3                                                                          | Les différentes fonctions de transfert .....                                          | 54        |
| 2.4.2.6.                                                                           | Apprentissage des réseaux de neurones.....                                            | 55        |
| 2.4.2.6.1                                                                          | Principe .....                                                                        | 56        |

|                                                                |                                                    |           |
|----------------------------------------------------------------|----------------------------------------------------|-----------|
| 2.4.2.6.2                                                      | Procédures d'apprentissage .....                   | 56        |
| •                                                              | Apprentissage supervisé.....                       | 56        |
| •                                                              | Apprentissage non supervisé.....                   | 56        |
| •                                                              | Apprentissage semi supervisé .....                 | 56        |
| 2.4.2.6.3                                                      | Apprentissage du réseau neuronal monocouche .....  | 57        |
| •                                                              | a- Règle de Hebb .....                             | 57        |
| •                                                              | b- Règle de widrow-Hoff .....                      | 57        |
| 2.4.2.6.4                                                      | Apprentissage du réseau neuronal multicouche ..... | 58        |
| •                                                              | a- Algorithme de rétropropagation du gradient..... | 58        |
| •                                                              | b- Algorithme de levenberg- Marquardt .....        | 59        |
| 2.4.3                                                          | Machine à vecteurs support .....                   | 59        |
| 2.5.                                                           | Algorithmes métaheuristiques .....                 | 61        |
| 2.5.1                                                          | Algorithme génétique .....                         | 61        |
| 2.5.1.1                                                        | Principe.....                                      | 61        |
| 2.5.1.2                                                        | Opérateurs génétiques .....                        | 62        |
| □                                                              | La sélection : .....                               | 62        |
| □                                                              | Le croisement : .....                              | 63        |
| □                                                              | La mutation : .....                                | 63        |
| 2.5.2                                                          | Optimisation par essaims des particules.....       | 63        |
| 2.5.2.1                                                        | Définition .....                                   | 63        |
| 2.5.2.2                                                        | Les éléments de la PSO.....                        | 63        |
| 2.5.2.3                                                        | Notion de voisinage.....                           | 64        |
| 2.5.2.4                                                        | Principe fondamental.....                          | 66        |
| 2.6.                                                           | Les indicateurs de performance.....                | 67        |
| 2.7.                                                           | Conclusion.....                                    | 68        |
| <b>Chapitre 3: Prévion de la vitesse du vent au maroc.....</b> |                                                    | <b>70</b> |
| 3.1.                                                           | Introduction.....                                  | 70        |
| 3.2.                                                           | Données expérimentales et régions d'étude.....     | 70        |
| 3.2.1.                                                         | Source des données .....                           | 70        |

|                                                                                                                         |           |
|-------------------------------------------------------------------------------------------------------------------------|-----------|
| 3.2.2. les régions d'étude.....                                                                                         | 71        |
| 3.2.3. Prétraitement des données.....                                                                                   | 72        |
| 3.3. Détermination des paramètres d'entrée.....                                                                         | 72        |
| 3.3.1. Etude statistique des données.....                                                                               | 73        |
| 3.3.2. L'influence des paramètres choisis sur la sortie.....                                                            | 75        |
| 3.4. <b>Approche 1:</b> Modélisation de la vitesse du vent par Régression linéaire multiple.....                        | 76        |
| 3.4.1. Analyse des résultats.....                                                                                       | 76        |
| 3.4.2. Performance du modèle.....                                                                                       | 78        |
| 3.5. <b>Approche 2:</b> Prévion de la vitesse du vent par Réseaux de neurones artificiels.....                          | 79        |
| 3.5.1. Analyse des résultats.....                                                                                       | 79        |
| 3.5.2. Performance du modèle.....                                                                                       | 81        |
| 3.5.3. Hybridation RNA-GA et RNA-PSO.....                                                                               | 82        |
| 3.5.3.1. Méthodologie.....                                                                                              | 82        |
| 3.5.3.2. Analyse des résultats.....                                                                                     | 83        |
| 3.5.3.3. Performance du modèle.....                                                                                     | 84        |
| 3.6. <b>Approche 3 :</b> Prévion de la vitesse du vent par Machine à vecteurs supports .....                            | 85        |
| 3.6.1. Méthodologie.....                                                                                                | 85        |
| 3.6.2. Analyse des résultats.....                                                                                       | 86        |
| 3.6.3. Performance du modèle.....                                                                                       | 87        |
| 3.7. Choix du meilleur modèle.....                                                                                      | 88        |
| 3.8. Conclusion .....                                                                                                   | 88        |
| <b>Conclusion générale et perspectives.....</b>                                                                         | <b>89</b> |
| <b>Bibliographie.....</b>                                                                                               | <b>90</b> |
| <b>Annexes.....</b>                                                                                                     | <b>99</b> |
| <b>Contribution I:</b> Forecasting of wind speed using a multiple linear regression and artificial neural networks..... | 100       |

|                                                                                                                           |     |
|---------------------------------------------------------------------------------------------------------------------------|-----|
| Résumé.....                                                                                                               | 100 |
| Contenu.....                                                                                                              | 101 |
| <b>Contribution II:</b> Hourly wind speed forecasting based on support vector machine and artificial neural networks..... | 113 |
| Résumé.....                                                                                                               | 113 |
| Contenu.....                                                                                                              | 114 |
| <b>Contribution III:</b> A direct normal irradiation forecasting model based on artificial neural networks .....          | 120 |
| Résumé.....                                                                                                               | 120 |
| Contenu.....                                                                                                              | 121 |
| <b>Contribution IV :</b> Two developed methods for predicting hourly direct normal irradiation (DNI).....                 | 129 |
| Résumé.....                                                                                                               | 129 |
| Contenu.....                                                                                                              | 130 |

## Table des figures

|                                                                                      |    |
|--------------------------------------------------------------------------------------|----|
| Figure 1.9 : Première aérogénératrice de Poul La Cour en 1890 à Askov, Danemark..... | 22 |
| Figure 1.10 : Capacité éolienne mondiale par pays en 2018.....                       | 24 |
| Figure 1.11 : Régions ventées au Maroc.....                                          | 29 |
| Figure 1.12 : Parcs éoliens au Maroc (MASSEN).....                                   | 30 |
| Figure 1.13 : Éoliennes à axe horizontal.....                                        | 32 |
| Figure 1.14 : Éoliennes à axe vertical.....                                          | 32 |
| Figure 1.15 : Éléments constituant une éolienne.....                                 | 33 |
| Figure 1.16 : La première ferme éolienne offshore (1991 Vindeby).....                | 35 |
| Figure 1.17 : Concept d'offshore flottant.....                                       | 35 |
| Figure 1.18 : Anémomètre.....                                                        | 36 |
| Figure 1.19 : Girouette.....                                                         | 37 |
| Figure 2.1 : Boîte de dialogue des données d'entrée pour la régression simple.....   | 44 |
| Figure 2.2: Fenêtre d'analyse de la régression simple.....                           | 45 |
| Figure 2.3: Modèle ajusté du premier, deuxième, troisième et quatrième ordre.....    | 46 |
| Figure 2.4: Graphe du résidu studentisé.....                                         | 49 |
| Figure 2.5: Neurone biologique.....                                                  | 50 |
| Figure 2.6: Schéma d'un neurone formel.....                                          | 50 |
| Figure 2.7: Réseau neuronal non bouclé.....                                          | 51 |
| Figure 2.8: Réseaux neuronaux bouclés.....                                           | 52 |
| Figure 2.9: Réseau neuronal complètement connecté.....                               | 52 |
| Figure 2.10: Fonction sigmoïde.....                                                  | 54 |
| Figure 2.11: Fonction tangente hyperbolique.....                                     | 54 |
| Figure 2.12: Schéma des étapes de l'algorithme de rétropropagation.....              | 58 |
| Figure 2.13: Principe général des algorithmes génétiques.....                        | 61 |
| Figure 2.14: Un voisinage géographique.....                                          | 64 |
| Figure 2.15: Un voisinage en cercle.....                                             | 65 |
| Figure 2.16: Compromis entre deux déplacements possibles.....                        | 66 |
| Figure 2.17: Organigramme de la méthode d'optimisation des essaims particuliers..... | 66 |
| Figure 3.1: extrait de la base de données dans statgraphics.....                     | 72 |
| Figure 3.2: graphe de régression de la vitesse du vent pour Tanger.....              | 76 |
| Figure 3.3: graphe de régression de la vitesse du vent pour Tarfaya.....             | 77 |
| Figure 3.4: Réduction du MSE pour la structure 4-5-1 à Tanger.....                   | 79 |
| Figure 3.5: Réduction du MSE pour la structure 4-1-1 à Tarfaya.....                  | 80 |

|                                                        |    |
|--------------------------------------------------------|----|
| Figure 3.6: Valeur actuelle/ prédite par RNA-PSO ..... | 83 |
| Figure 3.7: Valeur actuelle/ prédite par RNA-GA.....   | 83 |
| Figure 3.8: Valeur actuelle/ prédite par SVM-PSO.....  | 86 |
| Figure 3.9: Valeur actuelle/ prédite par SVM-GA.....   | 86 |

## Liste des tableaux

|                                                                                                                  |    |
|------------------------------------------------------------------------------------------------------------------|----|
| 1.1: Classification des turbines éoliennes.....                                                                  | 34 |
| 3.1: Présentation d'une fiche de données du site de Tanger.....                                                  | 70 |
| 3.2: Situation géographique des sites.....                                                                       | 70 |
| 3.3: Présentation des coefficients de corrélation intervariables du Tanger .....                                 | 73 |
| 3.4: Présentation des coefficients de corrélation intervariables du Tarfaya .....                                | 73 |
| 3.5: Analyse statistique de la RLM à Tanger .....                                                                | 74 |
| 3.6: Analyse statistique de la RLM à Tarfaya .....                                                               | 74 |
| 3.7: Coefficient de corrélation entre les variables indépendantes et la variable dépendante pour<br>Tanger.....  | 75 |
| 3.8: Coefficient de corrélation entre les variables indépendantes et la variable dépendante pour<br>Tarfaya..... | 75 |
| 3.9: Performance de la régression linéaire multiple à Tanger et Tarfaya.....                                     | 76 |
| 3.10: Indicateurs statistiques du RNA pour les deux régions.....                                                 | 78 |
| 3.11: Performance des modèles hybrides RNA-GA et RNA-PSO.....                                                    | 82 |
| 3.12: les paramètres optimaux de SVM-GA.....                                                                     | 85 |
| 3.13: les paramètres optimaux de SVM-PSO.....                                                                    | 85 |
| 3.14: Performance des modèles hybrides SVM-GA et SVM-PSO.....                                                    | 85 |
| 3.15: Performance des modèles hybrides de prévision.....                                                         | 87 |

## Glossaire

|       |                                                    |
|-------|----------------------------------------------------|
| WWEA  | World Wind Energy Association                      |
| CDER  | Centre de Développement des Energies Renouvelables |
| ONE   | Office National d'Electricité                      |
| GES   | Gaz à Effet de Serre                               |
| GA    | Genetic Algorithm                                  |
| PSO   | Particle Swarm Optimization                        |
| SVM   | Support Vector Machine                             |
| AR    | Autorégressif                                      |
| ARMA  | Auto Regressive Moving Average                     |
| ARIMA | Auto Regressive Integrated Moving Average          |
| RNA   | Réseaux de Neurones Artificiels                    |
| NWP   | Numerical Weather Prediction                       |
| OLS   | Ordinary Least Square                              |
| ANOVA | Analyse de la variance                             |
| RLM   | Régression Linéaire Multiple                       |



## **Introduction générale**

L'augmentation croissante de la demande énergétique représente un enjeu majeur pour les pays du monde. Cela est dû au fait que l'activité industrielle dans les pays développés s'étend rapidement et par conséquent les entreprises en question investissent plus dans les pays à faible coût de production. En réalité, cette augmentation se reflète par une hausse du prix des carburants représentant une source importante d'énergie. Dans ce sens, il faut noter la diminution continue de la réserve pétrolière ce qui causera une pénurie de cette ressource et risquera ensuite de ne pas satisfaire la demande. Le recours à d'autres types de ressources d'énergie dites renouvelables est donc inévitable et devient plus qu'une nécessité pour le progrès de l'humanité.

Durant la première décennie de ce vingt et unième siècle, le Maroc a connu un intense élan d'activité tant aux niveaux industriel et économique qu'au niveau des structures sociales ; cet essor se traduit par une croissance soutenue des besoins en énergie. Pour réduire sa dépendance énergétique et vu sa situation géographique favorable, le Maroc s'est orienté vers le développement des sources d'énergies renouvelables en particulier le vent étant une source d'énergie libre, propre et inépuisable. En effet, la prévision précise de vent à court terme joue un rôle important dans l'optimisation des opérations des parcs éoliens, la maximisation de leur rendement et l'amélioration de leur développement. Aussi, la prévision de la vitesse du vent permet le contrôle et la gestion du réseau électrique en connaissant la quantité d'énergie produite, la protection du système contre les grandes vitesses et l'étude de faisabilité portant sur la construction des centrales éoliennes pour un site dont on connaît déjà le potentiel énergétique.

L'objectif de cette thèse est donc la réalisation d'une étude du parc éolien installé à la ville marocaine Tanger située au Nord et d'un autre localisé à Tarfaya située au Sud, ceci afin d'estimer un modèle de prévision à partir des données météorologiques acquises grâce à l'Institut de Recherche d'Energie Solaire et Energie Nouvelle (IRESEN). Tout d'abord, les données sont traitées par la méthode des moindres carrés ordinaires dans le but d'éliminer la multi colinéarité et la corrélation entre les variables pour choisir celles qui influencent plus la vitesse du vent dans les régions nord et sud du Maroc. Ensuite, les données traitées sont exploitées dans plusieurs approches de machine learning qu'on va développer pour chercher un modèle de prévision avec une meilleure précision. Nous nous sommes particulièrement penché sur les réseaux de

neurones artificiels qui sont très populaires et se montrent d'une grande efficacité dans le domaine de la prévision météorologique et les machines à vecteurs de support qui sont montrés ses efficacités de traitement des problèmes de discrimination non-linéaire, afin d'élaborer les logiciels suivants : Matlab et statgraphics.

## Aperçu de thèse

Cette thèse est organisée comme suit :

**Le chapitre 1** est divisé en quatre sections principales. La première section présente des notions importantes d'énergie éolienne. La deuxième section concerne les différentes éoliennes existantes et les principaux constituants d'une éolienne. La troisième section concerne les instrumentations de mesure utilisées dans le domaine éolienne ainsi que la puissance récupérable par une turbine afin de finaliser la section par une explication de la transformation d'énergie cinétique du vent en énergie électrique et la dernière section présente une revue des méthodes de prévision ainsi qu'une bibliographique des techniques utilisées pour la prévision de la vitesse du vent.

**Le chapitre 2** définit les principes de prévision et prédiction ainsi que les différents horizons de prévision afin de finaliser ce chapitre par une description détaillée aux méthodes mises en place pour développer un modèle de prévision de la vitesse du vent à court terme.

**Le chapitre 3** décrit les travaux réalisés tout au long de cette période de recherche. La première contribution traite les données météorologiques provenant de deux sites différents, couvrant une période de trois ans par la méthode des Moindres Carrés Ordinaires notée en anglais (OLS) dans le but de déterminer les paramètres qui influencent la vitesse du vent afin de les considérer comme entrées pour développer des modèle linéaire et non-linéaire de prédiction de la vitesse du vent. Dans la deuxième contribution, nous développerons quatre méthodes hybrides notamment RNA-GA, RNA-PSO, SVM-GA et SVM-PSO. Ensuite, nous conclurons cette étude en reprenant tous les résultats issus des travaux effectués et en essayant de dégager une synthèse de meilleur modèle de prévision pour les deux régions du Nord et du Sud au Maroc.

Enfin, une **conclusion générale** résumera les résultats de nos travaux et donnera des perspectives et futures pistes de recherches.

# Chapitre 1. Notions et état de l'art sur l'énergie éolienne

---

## 1.1 Introduction

Les pays dans le monde entier ont connu une croissance économique, politique et sociale au cours des deux dernières décennies. Ce développement se traduit par une croissance soutenue des besoins en énergie électrique.

Les sources d'énergie électrique traditionnelles telles que les combustibles fossiles comme le pétrole, le charbon, le gaz naturel et l'énergie de fission nucléaire ont des inconvénients sur l'humanité et la nature, par exemple la production électrique à partir de combustibles fossiles est à l'origine de 40 % des émissions mondiales de CO<sub>2</sub>.

Pour relever ce grand défi et garantir un développement responsable, un nouveau plan d'équipement a été élaboré sur la base d'une vision globale et intégrée. Il prévoit la mise en œuvre d'un bouquet électrique diversifié, dans lequel les énergies renouvelables occupent une place de choix. Parmi ces énergies, on trouve l'énergie éolienne qui est une source d'énergie propre et inépuisable et qui peut produire de l'électricité pratiquement sans l'émission des gaz polluants en utilisant la technologie des aérogénérateurs.

Dans ce présent chapitre, la première partie sera consacrée à l'historique d'un éolien ainsi que la capacité éolienne installée dans le monde, et plus particulièrement la capacité marocaine et aussi les projets en construction et ceux en perspectives. La deuxième partie présente les technologies d'éoliennes et les différents composants constituant l'aérogénérateur (couramment appelé éolienne). Après, on va expliquer la conversion d'énergie cinétique du vent en énergie électrique ainsi que la puissance récupérable par une turbine.

Dans la troisième partie, nous allons présenter une revue de littérature de différentes méthodes statistiques et physiques de prévision ainsi que l'état de l'art des techniques de prévision de la vitesse du vent.

## **1.2 Notions sur l'énergie éolienne**

Les énergies éoliennes sont notamment caractérisées par des déplacements de masse d'air à l'intérieur de l'atmosphère. C'est l'énergie mécanique de ces déplacements de masse d'air qui est à la base de l'énergie éolienne. L'énergie éolienne consiste ainsi à utiliser cette énergie mécanique par l'intermédiaire d'un rotor et de pales. Cette source est non polluante. On peut installer des éoliennes en mer et sur terre [1]. On va détailler ce type d'énergie dans les parties suivantes de ce chapitre.

### **1.2.1. Historique de l'éolien**

Parmi toutes les énergies renouvelables, à part l'énergie du bois, c'est l'énergie du vent qui a été exploitée en premier par l'homme. Ce n'est qu'à la fin du XIX<sup>ème</sup> siècle [2,3], lorsque l'électricité prit son essor, que l'aérogénérateur électrique fit ses premiers pas. Au Danemark, Poul La Cour fut, sans doute, le pionnier, il a associé une turbine éolienne à une dynamo en 1891.

En France, contrairement à ce que la situation actuelle pourrait laisser penser, les recherches allèrent bon train dans les années 1920 (éolienne bipale de 20 m de diamètre, Compagnie Electro-Mécanique (CEM) puis dans les années 1950-60 (tripale de 30 m est génératrice synchrone de 800 kW à Nogent-le-Roi, bipale de 35 m est génératrice asynchrone de 1 MW à Saint Rémy des Landes).

En 1941, une éolienne bipale de 1250 kW (machine synchrone) était expérimentée aux États-Unis dans le Vermont.

En 1978, les Danois, aujourd'hui leaders dans le domaine, réalisèrent une machine tripale de 54 m pour une puissance de 2 MW. Les nombreuses réalisations expérimentales (notamment en Californie) et leurs avatars permirent de perfectionner les turbines, les systèmes de freinage d'urgence, les pylônes et les divers dispositifs de contrôles et d'obtenir des aérogénérateurs viables et compétitifs à la fin du XX<sup>ème</sup> siècle. Diverses solutions de turbines furent explorées, notamment celles à axe vertical connu sous le nom de Darrieus (suite à leur inventeur), mais ne subsistent aujourd'hui pratiquement plus que des machines à axe horizontal à 2 ou 3 pales.

### **1.2.2. Premier pas vers les éoliennes**

C'est le physicien et mathématicien *Lord KELVIN* qui a eu le premier la brillante idée de coupler une génératrice électrique à un moulin à vent. Ce premier pas vers 1<sup>ère</sup> des aérogénérateurs modernes remonte à 1802. Mais les premières applications concluantes durent attendre un demi-siècle de plus, lorsque *Zénobe Gramme* inventa la dynamo. C'est en 1888 que la première éolienne reliée à un système de stockage du courant vit le jour. La première éolienne « industrielle » génératrice d'électricité est développée par le danois

*Poul La Cour* en 1890(voir la figure 1.9), pour fabriquer de l'hydrogène par électrolyse. Dans les années suivantes, il crée l'éolienne *Lykkegard*, dont il aura vendu 72 exemplaires en 1908.



Figure 1.1: Première aérogénératrice de Poul La Cour en 1890 à Askov, Danemark

### **1.2.3. Retour à l'éolien : une source de l'énergie de demain**

Avec l'avènement des énergies fossiles, les éoliennes furent reléguées au second plan jusqu'à la crise pétrolière de 1973[4]. Suite à la panique engendrée par les deux chocs pétroliers de 1973 et 1979, qui déboucha sur la recherche d'énergies alternatives, un regain d'intérêt sans cesse grandissant se fit sentir pour l'éolien. Les principales motivations actuelles sont la hausse des prix des énergies fossiles et l'alternative au nucléaire. Étant donné que les ressources naturelles s'épuisent inexorablement et que les opinions publiques sont de plus en plus sensibilisées aux énergies renouvelables, il ne fait nul doute que l'éolien s'impose comme une énergie du futur. Nous cherchons surtout à les utiliser pour produire de l'énergie électrique suivant le principe exploité dans toutes les centrales électriques : un fluide (ici : le vent) entraîne une turbine reliée mécaniquement à une machine électrique, appelée génératrice, qui va produire du courant.

#### **1.2.4. Enjeux de l'éolienne**

##### *1.2.4.1. Sur le plan environnemental*

Les énergies renouvelables sont des énergies propres, elles n'émettent pas de gaz à effet de serre et elles ne produisent aucun déchet radioactif. Le recours aux énergies renouvelables permet de conserver certaines ressources naturelles telles que le pétrole qui sert notamment à la fabrication de certains médicaments [5,6].

##### *1.2.4.2. Sur le plan économique*

Les différents accidents (de cause naturelle ou humaine) comme ceux survenus à Tchernobyl ou plus récemment à Fukushima se traduisent par des coûts importants et des problèmes écologiques souvent dramatiques pour les populations locales. Les déchets radioactifs créés par le nucléaire sont dangereux et le coût pour les contenir est très élevé. De plus, différentes études montrent que les énergies renouvelables sont plus économiques à long terme.

#### **1.2.5. Futurs challenges et perspectives mondiales**

Sur la base de l'expérience et des taux de croissance des années passées, l'association internationale d'énergie éolienne dans le monde World Wind Energy Association (WWEA) prévoit que l'énergie éolienne continuera son développement dynamique dans les années à venir. Malgré les impacts à court terme de la crise financière actuelle qui rendent difficiles les prévisions à court terme, on peut attendre, qu'à moyen terme, l'énergie éolienne attire les investisseurs grâce à son faible risque et au besoin d'énergie propre et fiable [7]. De plus en plus de gouvernements comprennent les bénéfices multiples de l'énergie éolienne et mettent en place des politiques favorables, incluant la promotion des investissements décentralisés par des producteurs indépendants, petits, moyens ou basés sur des communautés, tous ouvrant la voie vers un système de production d'énergies plus renouvelables pour le futur. En calculant rigoureusement et en prenant en compte des facteurs d'insécurité, on estime que l'énergie éolienne sera capable de contribuer en 2020 au moins 12% de la consommation électrique mondiale.

#### **1.2.6. L'énergie éolienne dans le contexte international**

Les systèmes éoliens connaissent un essor considérable depuis une quinzaine d'années, l'intégration de ces systèmes est devenue un élément clé pour des nouvelles politiques énergétiques dans nombreux pays.

La production d'énergie éolienne connaît toujours un taux de croissance très important comme il est montré sur la figure 1.10, on peut ainsi constater que la capacité en énergie éolienne de la Chine s'établissait à environ 184 gigawatts en

2018, plus que dans tout autre pays du monde à ce temps-là. Les États-Unis obtenaient la deuxième place du classement, disposant d'une capacité en énergie éolienne de près de 100 gigawatts et l'Allemagne s'occupe la troisième place avec une puissance de 60 gigawatts.

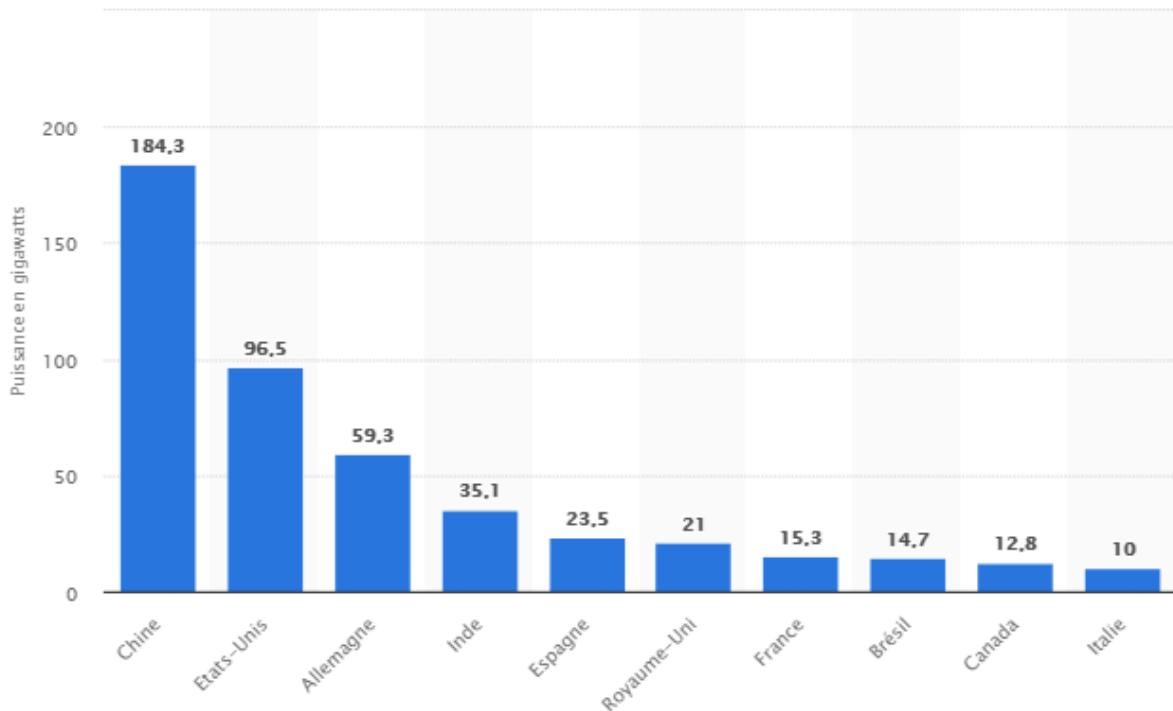


Figure 1.2: Capacité éolienne mondiale par pays en 2018

On remarque aussi d'après cette figure qu'il y a cinq pays européens qui jouent un rôle majeur pour augmenter la capacité éolienne dans le monde ce qui signifie l'importance d'énergie éolienne dans l'union européenne qui assure 14,5 % de la production d'électricité de l'Union européenne en 2019. Les pays européens qui sont très célèbres dans le domaine des éoliennes sont:

L'Allemagne est l'un des principaux producteurs d'électricité éolienne avec 22 247 MW de puissance installée à la fin de l'année 2007. L'Allemagne installe une moyenne de 2GW de puissance supplémentaire par an. L'éolien a couvert 20,4 % de la consommation d'électricité du pays en 2017-18. Pour la puissance installée par habitant, l'Allemagne se situait en 2017 au quatrième rang européen derrière le Danemark, l'Irlande et la Suède. En 2018, l'Allemagne se place au premier rang européen pour la production d'électricité éolienne : 111,6 TWh, soit 29,4 % du total de l'Union européenne, et pour la puissance installée : 32,9 %. Au niveau mondial, l'Allemagne se place au troisième rang avec 8,6 % de la production mondiale et 10,4 % de la puissance totale, derrière la Chine et les États-Unis, et au deuxième rang pour les éoliennes offshore (27,6 % du total mondial) derrière le Royaume-Uni. En 2019, la part de l'éolien dans la production d'électricité du pays atteint 20,5 %.

En France, le potentiel éolien est très important, 20 GW terrestres pour une production de 50 TWh par an, et 40 GW offshore pour une production de 150 TWh par an, soit un potentiel éolien théorique de 200 TWh par an. S'il était disponible en 2040, il représenterait alors 31 % de la consommation française prévisible d'électricité. Cette production de 200 TWh/an se répartirait ainsi : 8000 éoliennes offshore de 5 MW sur 40 grandes centrales installées entre 15 et 40 km de la côte et 8000 éoliennes terrestres de 3 MW installés en France (qui mesurent 50 à 55 m de haut - et jusqu'à 100 m dans les zones vallonnées, contre 80 à 100 m pour les mâts des grandes éoliennes).

Le Danemark est le plus important fabricant mondial d'éoliennes, il a joué un rôle pionnier dans le développement de cette technique. C'est le pays européen le plus densément équipé en éoliennes: 960,3 W/habitant en 2017. Les éoliennes installées au pays produisaient 48 % de l'électricité du pays en 2019 et couvraient 40,5 % de la consommation électrique danoise sur la période de mi-2017 à mi-2018, taux qui le situe de loin au premier rang européen. Le Danemark était en 2018 au huitième rang européen pour la production d'électricité éolienne et au septième rang par sa puissance installée ; cette haute proportion d'énergie intermittente est rendue compatible avec la demande grâce aux capacités de régulation apportées par les barrages hydroélectriques de la Norvège et de la Suède, et aux nombreux câbles sous-marins d'interconnexion qui relient le Danemark à ces deux pays. Sur le segment offshore, il tient le troisième rang européen, loin derrière le Royaume-Uni et l'Allemagne. Les éoliennes offshore ont produit 38,4 % de la production éolienne totale en 2019 avec 27,8 % de la puissance installée.

En Espagne, l'énergie éolienne est une source d'énergie très importante et depuis l'année 2011, l'énergie éolienne a été la principale source d'électricité espagnole pour la première fois de l'histoire avec une production de 4738 GWh. Au niveau mondial, l'Espagne se classait au 6<sup>ème</sup> rang pour la production d'électricité d'origine éolienne en 2017 avec 4,4 % de la production éolienne mondiale, loin derrière la Chine : 26,2 %, les États-Unis : 22,8 % et l'Allemagne : 9,4 %. L'Espagne est en 2018 le troisième producteur d'électricité éolienne en Europe avec 13,4 % du total européen contre 29,4 % pour l'Allemagne, qui lui a ravi le premier rang en 2014 et 14,7 % pour le Royaume-Uni, qui lui a ravi le deuxième rang en 2018. Sa puissance installée représente 13,4 % du total européen fin 2019 contre 31,9 % pour l'Allemagne. En 2019, année où l'Espagne s'est classée au deuxième rang européen en termes de mises en service ; l'éolien assurait 21 % de la production électrique espagnole en 2019 et couvrait 18,8 % de la consommation en 2017-18, taux qui la situait

au cinquième rang européen derrière le Danemark, l'Irlande, le Portugal et l'Allemagne.

Le Royaume-Uni dispose d'un potentiel éolien parmi les meilleurs au monde, il était en 2018 au deuxième rang européen pour la production d'électricité éolienne et pour la puissance installée au sixième rang mondial avec 3,5% du total mondial. Sur le segment offshore, il tient le premier rang européen et mondial, les parcs éoliens en mer du Royaume-Uni ont produit 47% de la production éolienne du pays et 44,5% du total de l'éolien en mer européen. En 2019, l'éolien fournissait 22 % de la production électrique britannique et sa progression est rapide : la puissance installée de ses éoliennes s'est accrue de 29 % en 2017, 7 % en 2018 et 11 % en 2019. Le rythme d'installation s'accélère pour les éoliennes en mer alors qu'il chute pour les éoliennes terrestres.

En Belgique, les éoliennes ont le vent en poupe. La puissance installée est passée de 170 MW en 2005 à 911 MW en 2010. La production maximale disponible dans le pays est toutefois estimée à environ 2000 MW. La Belgique était en 2018 le quatorzième producteur d'électricité éolienne d'Europe et le douzième pour la puissance installée en 2019 ; elle se situait en 2017 au dixième rang européen pour la puissance installée par habitant. L'éolien produisait près de 10 % de l'électricité du pays en 2019 et couvrait 8,1 % de la consommation d'électricité belge en 2017-18. Le développement de l'éolien se fait en grande partie en mer : 40 % de la puissance installée fin 2019 et 51,6 % de la production en 2018 ; la Belgique se classe au quatrième rang européen et cinquième rang mondial.

### **1.2.7. L'énergie éolienne dans le contexte national**

#### **1.2.7.1. Historique**

Pour réduire sa dépendance énergétique et vu que sa situation géographique favorable, le Maroc s'est orienté entre autres vers le développement des sources d'énergies renouvelables notamment l'énergie éolienne dans plusieurs régions. C'est ainsi que, dès 1982, avec la création du Centre de Développement des Énergies Renouvelables (CDER), établissement public sous la tutelle du Ministère de l'Énergie et des Mines, le Département de l'Énergie a intégré les énergies renouvelables dans la politique énergétique nationale. Après des phases de prospection, de maturation de l'offre technique et de développement de projets pilotes importants, les énergies renouvelables ont été adoptées par des programmes nationaux de développement, notamment d'électrification et d'approvisionnement en eau potable des zones rurales pour le photovoltaïque et

par la mise en œuvre de projets de parcs éoliens de puissance et de centrales thermo-solaires [8].

Concernant l'énergie éolienne, il faut remarquer qu'un des principaux atouts pour le Maroc est tout d'abord son gisement. En effet, il bénéficie d'un gisement éolien important avec des régions dépassant dix mètres par seconde (10 m/s) de vitesse annuelle moyenne du vent. La mise en évidence de ce gisement a été réalisée par le Centre de Développement des Énergies Renouvelables (CDER) avec l'aide de la coopération technique allemande (GTZ : Gesellschaft für technische Zusammenarbeit) qui a permis l'installation de plusieurs mâts de mesure sur différents sites et le suivi des données sur des périodes d'au moins une année. Une première carte éolienne du pays a montré que la zone nord (Tanger à Tétouan) et la bande côtière allant de Tarfaya jusqu'à Lagouira présentent des sites exceptionnels avec des vents réguliers et des vitesses moyennes suffisantes pour développer des projets rentables [9]. Donc le premier parc installé est le parc Abdelkhalek Torres à Tétouan d'une puissance de 50 MW, ce parc a été lancé par l'Office National d'Électricité (ONE) en 1994 et mis en service en août 2000. Située sur le même site qu'Abdelkhalek Torres, l'installation d'un parc de 3.5 MW a été faite. Après, Lafarge a réalisé un parc éolien de 10.2 MW. La puissance de 10 MW n'est due qu'à la limite autorisée par la loi concernant l'autoproduction. Cette expérience a été très concluante pour un parc composé de 12 éoliennes de 850 KW et un productible dépassant 38 GWh. Ensuite, la réalisation du parc éolien sur le site Cap Sim au sud de la ville d'Essaouira a été lancée par l'ONE en 2004, et la mise en service en 2007. En 2006, le lancement du projet de Tanger avec une puissance prévue de 140 MW qui sera composé de 165 aérogénérateurs, les travaux dureront dix-huit mois et la centrale devrait être opérationnelle au cours du premier trimestre 2009. l'ONE a préparé un autre projet de 100 MW dans la région de Taza. Jusqu'à nos jours, autres parcs éoliens ont été réalisés dans différentes régions ventées au Maroc par exemple : parc El Koudia El Baida, Amogdoul à Essaouira, Dhar Saadan à Tanger, parc éolien à Laayoune, etc..À cause de ce contexte du développement d'énergie renouvelable, le Maroc arrive ainsi à la troisième place du classement des plus gros marchés éoliens africains et moyen-orientaux, derrière l'Afrique du Sud (1.053 MW installé) et l'Égypte (810 MW).

Dans le cadre de sa stratégie énergétique, le Maroc s'engage dans un vaste programme éolien, pour accompagner le développement des énergies renouvelables et de l'efficacité énergétique dans le pays dans le but d'augmenter la capacité électrique totale à 14% à l'horizon 2020 et économiser en combustibles 1,5 million de tonnes équivalent pétrole annuellement, soit 750

millions de dollars US et aussi éviter l'émission de 5,6 millions de tonnes de CO<sub>2</sub> par an[10,11].

Le Projet marocain intégré de l'Énergie éolienne, s'étalant sur une période de 10 ans pour un investissement total estimé à 31,5 milliards de dirhams, permettra au pays de porter la puissance électrique installée, d'origine éolienne, de 280 MW en 2010 à 2000 MW à l'horizon 2020, c'est pour cela le développement de 1720 MW de nouveaux parcs éoliens à l'horizon 2020 est prévu dans le cadre du projet éolien [12,13]:

- 720 MW en cours de développement à Tarfaya (300 MW), Akhfenir (200 MW), Bab El Oued (50 MW), Haouma (50 MW) et Jbel Khalladi (120 MW)
- 1 000 MW prévus sur 5 nouveaux sites choisis pour leur grand potentiel: Tanger 2 (150 MW), Koudia El Baida à Tétouan (300 MW), Taza (150 MW), Tiskrad à Laayoune (300 MW) et Boujdour (100 MW).

#### ***1.2.7.2. Répartition régionale de la vitesse du vent***

Pour installer une éolienne, il faut choisir un site où la vitesse du vent fournirait l'énergie nécessaire; des mesures déjà faites de la vitesse du vent dans plusieurs sites. La carte des vents au Maroc (voir la figure 1.11) estimée à 10 m du sol présente les vitesses moyennes annuelles obtenues varient de 3 à plus 8 m/s.

D'après la figure (1.11) nous remarquons que les régions les plus ventées au Maroc se trouvent à l'extrême nord, du côté de détroit de Gibraltar dans la région de Tanger, Tétouan et région d'Essaouira. Pour la zone du sud, on trouve la zone du Tarfaya à Lagouira, et finalement le couloir de Taza entre les chaines montagneuses de l'Atlas et du Rif [14,15].

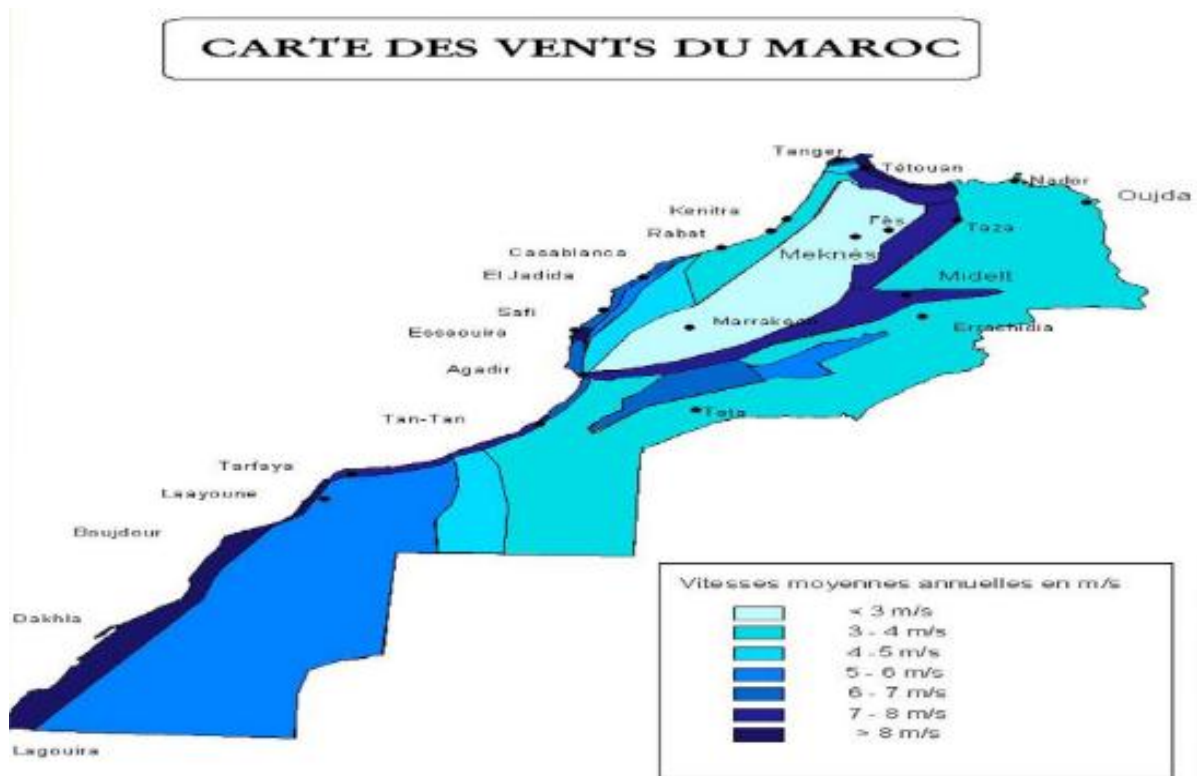


Figure 1.3: Régions ventées au Maroc

Le gisement éolien se caractérise par des vitesses moyennes de vent supérieures à 8 m/s pour les régions les plus ventées. Et d'après les mesures de la vitesse du vent pour de différentes régions au Maroc, la figure 1.12 indique les différents parcs éoliens situés au Maroc. Le premier parc avait été installé dès l'an 2000 dans la région de Tétouan (Abdelkhalek Torrès), pour une puissance installée de 50 mégawatts. 7 ans après, c'est le parc d'Amogdoul d'une puissance de 60 mégawatts qui avait été implanté dans la région d'Essaouira. S'en sont suivis les parcs de Tanger en 2010 totalisant 140 mégawatts, et de Laayoune en 2013 ajoutant 60 mégawatts à la production éolienne nationale.

Le mégaprojet éolien de Tarfaya a vu le jour en 2014, pour une puissance installée de 300 mégawatts. Doté de 131 turbines, il est le plus grand parc d'Afrique et a été réalisé par le groupe français Engie, en partenariat avec Nareva.

Autre projet installé est celui de Ksar Sghir, situé entre Tanger et Tétouan, inauguré fin 2016 avec une capacité de production de 120 mégawatts.

Il y a aussi d'autres parcs comme le parc de Taza et au nord de Midelt avec une capacité de 150 mégawatts, et celui de boujdour, au sud du pays avec une capacité de 100 mégawatts.



Figure 1.4: Parcs éoliens au Maroc (MASÉN)

### 1.2.8. Avantages et inconvénients d'énergie éolienne

Les avantages de l'énergie éolienne sont nombreux. Non émettrice de Gaz à Effet de Serre (GES), son impact sur l'environnement est quasi nul, contrairement aux énergies traditionnelles telles que gaz, pétrole et charbon (GES) ou au nucléaire (déchets radioactifs). Rappelons que c'est une énergie renouvelable dont la ressource ne sera pas épuisée dans l'avenir. De plus, elle n'est pas dangereuse comme peut l'être le nucléaire. Aussi, son coût a tendance à baisser (il a été divisé par 2 au cours des années 90) et est désormais inférieur à celui des énergies fossiles qui augmente (notons que c'est aussi la moins chère des énergies renouvelables). En outre, son coût peut être prédit avec très peu d'incertitude à la différence des fluctuations de prix qui touchent le charbon, le gaz ou le pétrole. Enfin, l'énergie éolienne permet d'alimenter des sites isolés qui ne peuvent être connectés au réseau électrique. L'éolien a également quelques inconvénients qu'il ne faut pas omettre. Son impact visuel tout d'abord qui n'est pas encore accepté du grand public. Le bruit engendré par les éoliennes a, quant à lui, nettement diminué. L'éolien serait de plus impliqué dans la mortalité des oiseaux ; sa responsabilité relative reste tout de même 19000 fois

inférieure à celle des bâtiments/fenêtres. Néanmoins, des efforts sont faits pour éviter leur implantation, notamment sur les parcours migratoires. Le principal désavantage de cette énergie est en réalité le caractère intermittent de sa ressource, le vent. La puissance électrique d'origine éolienne est fortement dépendante de la vitesse du vent puisqu'elle est proportionnelle au cube de celle-ci et est donc elle-même hautement instable, ce qui la rend difficilement prévisible. Or, l'un des intérêts des exploitants du réseau électrique est justement de contrôler la production à venir et de la répartir sur le réseau de manière à répondre au mieux à la demande. Ils se doivent donc de maîtriser les ressources et notamment la ressource éolienne pour une intégration efficace de celle-ci dans le réseau, efficacité qui passe par le choix des sites d'implantation d'une part et par la prévision de la production énergétique future d'autre part. Une exploitation optimale de l'énergie éolienne repose donc sur des outils de prévision de la ressource performants. Quelques notions d'énergie éolienne et la vitesse du vent, introduites dans la partie suivante avant d'aborder les problématiques et les techniques de prédiction du vent.

### **1.3. Descriptif d'une éolienne**

Une éolienne est une machine qui transforme l'énergie cinétique du vent (déplacement d'une masse d'air) en énergie mécanique ou électrique. Dans la technologie concernant les dispositifs de conversion d'énergie, parmi toutes les énergies renouvelables contribuant à la production d'électricité, l'énergie éolienne tient actuellement le rôle de vedette. Dans cette partie, nous allons nous intéresser aux dispositifs de conversion permettant de transformer l'énergie éolienne en énergie électrique. Dans un premier temps, nous présenterons les caractéristiques et les différents types d'aérogénérateurs existants, puis nous analyserons les chaînes de conversion électrique associées à ces turbines [16]. Il existe deux grandes catégories d'éoliennes selon la disposition géométrique de l'arbre sur lequel est montée l'hélice.

#### **1.3.1. Éolienne à axe horizontal**

Les éoliennes à axe horizontal captent le vent en hauteur et loin du sol, le vent étant beaucoup moins ralenti ou influencé par le relief. L'axe de rotation de ces éoliennes est placé horizontalement par rapport au sol (voir la figure 1.13). Elles sont montées sur pylône et possèdent un système d'orientation dans le vent. Elles ont également deux types de rotation, la première Lente (multipales) qui ne sert quasiment qu'au pompage de l'eau et le deuxième rapide (aérogénérateur) qui sont les plus utilisés pour produire de l'électricité.



Figure 1.5: Éoliennes à axe horizontal

### 1.3.2. Éolienne à axe vertical

L'éolienne à axe vertical ne dispose pas d'un système d'orientation par rapport au vent. Elle possède l'énorme avantage de pouvoir capter des vents très faibles, ce qui en fait un excellent moteur pour le pompage de l'eau, la conception verticale offre l'avantage de mettre le multiplicateur, et la génératrice au sol directement, mais cela impose que l'éolienne fonctionne avec le vent proche de sol, en conséquence, moins fort qu'en hauteur, en ce qui concerne la production électrique, elle est moins rapide que les modèles à axe horizontal. Les deux modèles les plus connus sont le type "Savonius" et le type "Darrieus", du nom de leur inventeur respectif (voir la figure 1.14).



Figure 1.6: Éoliennes à axe vertical

### 1.3.3. Principaux constituants d'une éolienne à axe horizontal

L'éolienne comporte généralement des éléments mécaniques et électriques qui sont illustrés sur la figure (1.15) [17].

**Le mat ou la tour:** c'est un tube d'acier, il doit être le plus haut possible pour bénéficier du maximum de l'énergie cinétique du vent, et d'éviter les perturbations près du sol. Au sommet du mât se trouve la nacelle.

**La nacelle:** regroupe les éléments mécaniques permettant de coupler la génératrice électrique à l'arbre de l'éolienne.

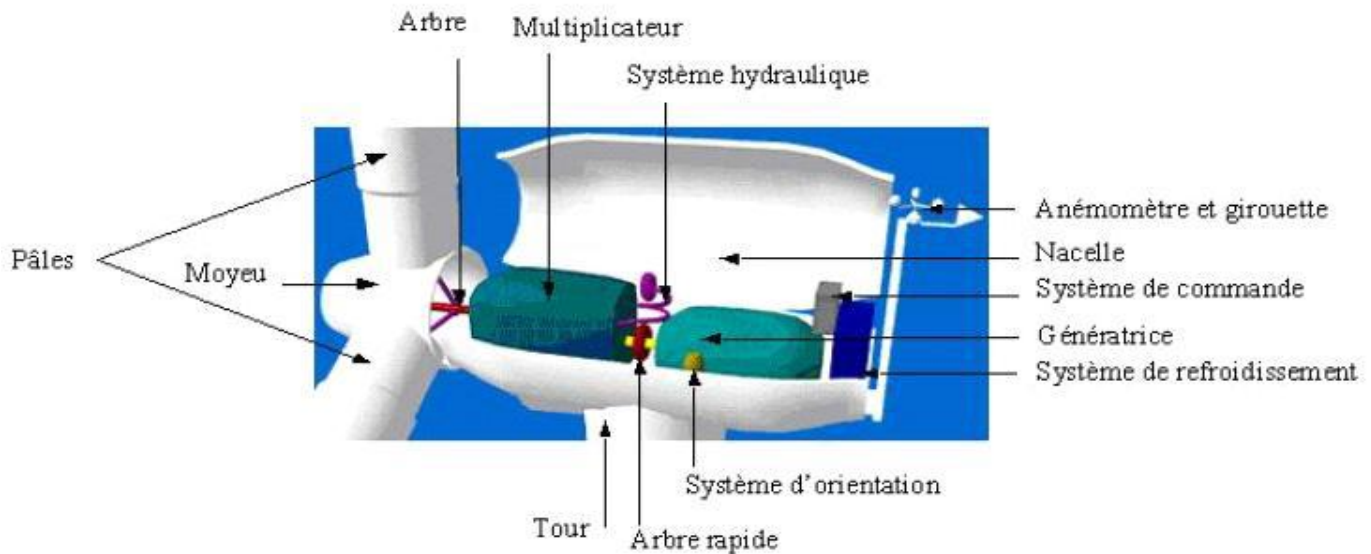


Figure 1.7: Éléments constituant une éolienne

**Le multiplicateur:** Sert à adapter la vitesse de la turbine éolienne à celle de la génératrice électrique.

**Le système de refroidissement:** se compose généralement d'un ventilateur électrique utilisé pour refroidir la génératrice et d'un refroidisseur à l'huile pour le multiplicateur.

**La génératrice électrique:** c'est l'élément principal de la conversion mécano électrique qui est généralement une machine synchrone, asynchrone à cage ou à rotor bobiné. La puissance électrique de cette génératrice peut varier entre quelque kW à 10 MW.

**Le système de commande:** qui contrôle en permanence le bon fonctionnement de l'éolienne et qui intervient automatiquement, en cas de défaillance pour l'arrêter.

**L'arbre:** qui relie le moyeu au multiplicateur, il contient un système hydraulique permettant le freinage aérodynamique en cas de besoin.

**Le système d'orientation des pâles:** qui sert à la régulation de la puissance (réglage aérodynamique). En plus de ces éléments, la turbine est munie des pâles fixes ou orientables et qui tournent à des vitesses nominales inférieures à 40 tr/min. Les éoliennes sont divisées en trois catégories selon leur puissance nominale, il y a :

- Éoliennes de petite puissance : inférieure à 40 KW.
- Éoliennes de moyenne puissance : de 40 à quelques centaines de KW.

- Éoliennes de forte puissance : supérieure à 1 MW.

À titre de comparaison, le tableau ci-dessous propose une classification de ces turbines selon la puissance qu'elles délivrent et le diamètre de leur hélice. On constate que lorsque le diamètre augmente la puissance du l'éolien augmente aussi ce qui signifie qu'il y a une relation de proportionnalité entre le diamètre et la puissance.

Tableau 1.1: Classification des turbines éoliennes

| Échelle | Puissance         | Diamètre de l'hélice |
|---------|-------------------|----------------------|
| Petite  | Moins de 40 KW    | Moins de 12 m        |
| Moyenne | De 40 KW à 999 KW | 12 m à 45 m          |
| Grande  | 1 MW et plus      | 46 et plus           |

#### 1.4.4 Un potentiel énorme encore peu exploité : l'éolien offshore

Une éolienne offshore, c'est-à-dire installée en mer, permet de convertir la force du vent en électricité. Le terme anglais « *offshore* » signifie littéralement « hors côtes », par opposition aux éoliennes terrestres ou « *onshore* ». Les éoliennes offshore fonctionnent selon le même principe que les modèles terrestres traditionnels : elles utilisent l'énergie cinétique du vent pour la transformer en électricité [18].

Le premier parc éolien offshore est installé au Danemark en 1990 sur le site de Vindeby à 2,5 km de la côte en eaux très peu profondes (moins de 5 mètres). Les 11 éoliennes installées d'une puissance unitaire de 450 kW ne sont alors que des éoliennes terrestres légèrement modifiées (voir figure 1.16).

Au milieu des années 2000 : le Royaume-Uni et l'Allemagne se lancent à leur tour dans l'éolien offshore à grande échelle. Avec le parc de Middelgrunden, le Danemark inaugure en 2001 le plus grand parc éolien offshore de l'époque et une nouvelle génération d'éoliennes : 20 éoliennes de 2 MW, distantes de 180 m est disposées en un arc de cercle de 3,4 km de long.

En 2017 : plus de 3,1 GW de nouvelles capacités éoliennes offshore sont installés. Et pour l'année 2019, Plus de 5 000 éoliennes offshore sont connectées.



Figure 1.8: La première ferme éolienne offshore (1991 Vindeby)

Pour la réalisation de ces fermes éoliennes, en proche offshore, les mâts sont montés sur des fondations ancrées dans le fond, mais lorsque l'on s'éloigne des côtes, il devient difficilement concevable de conserver ces moyens. C'est ainsi que s'est développé le concept d'offshore flottant, issu de la technologie des plates-formes pétrolières. Les architectures de conversion d'énergie sont souvent spécialement adaptées pour les fermes éoliennes offshore.

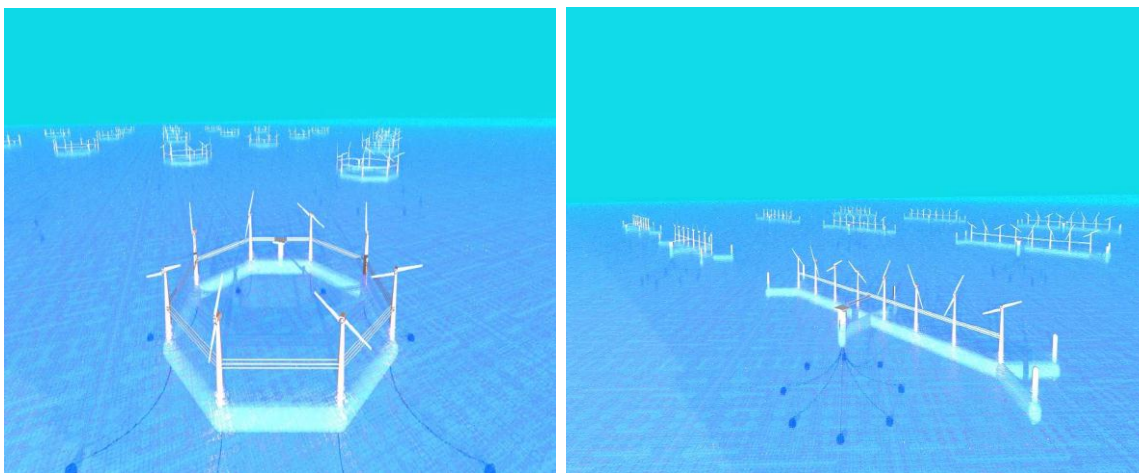


Figure 1.9: Concept d'offshore flottant

## 1.4. La source d'énergie éolienne : vitesse du vent

### 1.4.1. Origine du vent

L'atmosphère autour de la terre ne reçoit pas partout la même quantité d'énergie du soleil ; les écarts de température qui en résultent provoquent des différences de densité des masses d'air qui se traduisent par des variations de pression atmosphérique. Cette différence d'énergie est poursuivie d'une différence de pression qui provoque des déplacements d'air, car l'air chaud est moins dense

(plus léger) que l'air froid. La différence de température dans l'air est l'une des clés de tout mouvement dans l'atmosphère. Toutefois, le mouvement vertical de l'air est relativement faible par rapport au mouvement horizontal et c'est ce dernier type de mouvement qui nous intéresse ici. Ainsi le vent trouve son origine du rayonnement solaire qui rayonne la terre d'une façon non uniforme [19].

#### **1.4.2. Appareils de mesure**

- Anémomètre :

Un anémomètre est un appareil qui permet de mesurer la vitesse du vent (du grec anémos=vent). Il existe un certain nombre d'appareils pour la mesure de la vitesse du vent, plus au moins sophistiqués, mais l'appareil le plus répandu est encore l'anémomètre à coupelles. L'anémomètre à coupelles se compose de trois demi-sphères fixées sur trois bras horizontaux et tournants par rapport à un axe vertical. La vitesse de rotation de l'anémomètre est proportionnelle à la vitesse du vent.



Figure 1.10: Anémomètre

- Girouette :

La direction du vent est toujours donnée par la direction d'origine. On parle par exemple d'un vent du nord lorsque le vent souffle du nord vers le sud. L'instrument qui sert à déterminer la direction du vent est nommé : La girouette



Figure 1.11: Girouette

### 1.5. Puissance récupérable par une turbine

L'énergie éolienne provient de l'énergie cinétique du vent. En effet, si nous considérons une masse d'air,  $m$ , qui se déplace avec la vitesse  $v$ , l'énergie cinétique de cette masse est :

$$E_c = \frac{1}{2} m v^2 \quad (1.1)$$

Si, pendant l'unité de temps, cette énergie pouvait être complètement récupérée à l'aide d'une hélice qui balaie une surface  $A$ , située perpendiculairement à la direction de la vitesse du vent, la puissance instantanée fournie serait [20], alors:

$$P_v = \frac{1}{2} \rho A v^3 \quad (1.2)$$

Où  $\rho$  est la masse volumique de l'air.

Cependant, le dispositif de conversion extrait une puissance  $P_e$  inférieure à la puissance incidente  $P_v$ , à cause de la vitesse non nulle des masses d'air derrière l'aéromoteur. On définit alors, le coefficient de puissance de l'aéromoteur par la relation :

$$C_p = \frac{P_e}{P_v} ; \quad C_p < 1 \quad (1.3)$$

Ce coefficient caractérise l'aptitude de l'aérogénérateur à capter de l'énergie éolienne. La puissance correspondante est donc donnée par :

$$P_e = \frac{1}{2} \rho \pi R^2 v^3 C_p \quad (1.4)$$

Où  $R$  est le rayon de l'hélice.

On peut estimer la valeur maximale de ce coefficient, donc la puissance maximale qui peut être récupérée avec une turbine éolienne, en s'appuyant sur la théorie Rankine – Froude de l'hélice dans un fluide incompressible. C'est la valeur maximale de ce coefficient qui correspond à la limite dénommée la limite de Betz (théorie de Betz années 1920) du coefficient de puissance :

$$C_{p_{\max}} = \frac{16}{27} = 0.593 \quad (1.5)$$

La valeur du coefficient de puissance  $C_p$  dépend de la vitesse de rotation de la turbine et peut s'exprimer en fonction de la vitesse spécifique  $\lambda$ :

$$C_p = C_p(\lambda) \quad (1.6)$$

Avec :

$$\lambda = \frac{R\Omega}{v} \quad (1.7)$$

Où  $R\Omega$  est la vitesse linéaire périphérique en bout de pale de l'hélice.

La courbe  $C_p(\lambda)$  passe un maximum  $\lambda_{\max}$ , compris entre 5 et 16 environ, selon le nombre de pales de la turbine et c'est seulement pour cette valeur que  $C_p$  approche la valeur de 16/27.

On peut définir la notion de rendement aérodynamique de l'éolienne par le rapport :

$$\eta_a = \frac{C_p}{C_p^{\max}} = \frac{C_p}{0.593}$$

Qui est représentatif de la « qualité aérodynamique » de l'éolienne.

## **1.6. Transformation de l'énergie mécanique du vent en énergie électrique par éolienne**

Les pales tournent à une vitesse relativement lente, de l'ordre de 5 à 15 tours par minute, d'autant plus lente que l'éolienne est grande. La plupart des générateurs ont besoin de tourner à très grande vitesse (de 1000 à 2000 tours par minute) pour produire de l'électricité. C'est pourquoi le mouvement lent du rotor est accéléré par un multiplicateur. L'énergie mécanique transmise par le multiplicateur est transformée en énergie électrique par le générateur. Le rotor du générateur tourne à grande vitesse et produit de l'électricité. Cette électricité ne peut pas être utilisée directement ; elle est traitée grâce à un convertisseur, et acheminée à travers un câble enterré jusqu'à un poste de transformation, pour être injecté sur le réseau électrique, puis distribuée aux consommateurs [21].

### **2.1. Revue des techniques de prévision des vitesses du vent**

Comme nous l'avons déjà évoqué, un des principaux inconvénients de l'énergie éolienne est sa forte intermittence, directement issue de la volatilité du vent. Or, une bonne connaissance du vent et de la production d'énergie éolienne à venir est primordiale, car elle peut aider à la décision de l'intégration de l'énergie éolienne dans le réseau électrique ou au déploiement d'autres stations ou centrales. Des outils de prévision de la vitesse du vent et de la puissance en sortie d'une turbine ou d'une ferme éolienne sont donc plus que nécessaires.

Si l'on s'intéresse à l'évolution dans le temps d'une variable (notée  $V$  dans le cas de la vitesse du vent), une prévision consiste à estimer sa valeur future, connaissant la situation actuelle et passée. Depuis une trentaine d'années et avec l'intérêt grandissant porté à cette forme d'énergie renouvelable, de nombreuses techniques de prévision ont été développées pour améliorer les performances de ces systèmes dans la recherche actuelle [22]. Il existe deux sortes d'approches à la prédiction de la vitesse du vent. Les modèles dits "physiques" s'appuient sur des considérations atmosphériques. Ils ont démontré leur efficacité pour des

horizons allant de plus de six heures à quelques jours. Pour des horizons plus courts, de l'ordre de quelques minutes à quelques heures, on préférera les approches "statistiques" qui consistent en des modèles stochastiques et en l'analyse de séries temporelles.

## **2.2. Méthodes de prévision**

Les méthodes de prévision de vent peuvent être divisées en deux principales catégories : les méthodes à base physique et les méthodes statistiques.

### ➤ **Méthodes statistiques**

La première approche statistique est celle du modèle autorégressif (AR) qui s'appuie sur les propriétés d'autocorrélation des vitesses de vent. Un AR d'ordre  $n$  est un modèle selon lequel la valeur future est obtenue à un bruit près, comme une combinaison linéaire des  $n$  dernières valeurs mesurées. De même, leur généralisation autorégressive et moyenne mobile notée en anglais ARMA ou ARIMA (Auto-Regressive- Integrated- Moving Average) [23, 24] a été utilisée pour modéliser différentes séries météorologiques [25] comme les précipitations mensuelles [26], les débits annuels [27], l'indice mensuel de sécheresse [28] et bien d'autres applications [29]. Ainsi, malgré la simplicité initiale des processus ARMA, le modèle final reste relativement difficile à estimer et il est loin d'être parcimonieux. Parmi les approches basées sur les séries temporelles appliquées aux données de vitesses de vent, citons également les méthodes de type filtre de Kalman [30], des outils de type chaînes de Markov [31] et des ondelettes [32]. D'autres techniques plus récentes, basées sur l'intelligence artificielle, ont été prises en considération dans le cadre de la prévision de la ressource éolienne. Notons qu'une base de données importante est nécessaire à la phase d'apprentissage de ces modèles qui définissent une relation entre leurs entrées et leur sortie, c.-à-d. la variable que l'on cherche à prédire à un horizon  $h$ . Les modèles de type boîtes noires permettraient en outre de modéliser des processus non linéaires contrairement aux méthodes linéaires classiques.

Cette famille comprend les Réseaux de Neurones Artificiels (RNA) [33, 34, 35, 36] (voir section 2.5 pour plus de détails) et les techniques de logique floue [37].

### ➤ **Méthodes physiques**

Les modèles dits "physiques" s'appuient sur des considérations de la physique de l'atmosphère et conduisent à un système numérique de prévision du temps (NWP, Numerical Weather Prediction). Les modèles NWP (voir [38] pour une introduction aux modèles NWP) fournissent de meilleures prévisions que les méthodes purement statistiques quand on considère des horizons allant de plusieurs heures à plusieurs jours. Ils sont basés sur des équations et des paramètres physiques de l'atmosphère comme le relief, la rugosité du terrain, la pression et la température, ainsi que l'interaction de plusieurs phénomènes physiques entre eux. Lorsqu'on connaît l'état de l'atmosphère, on peut à l'aide

de simulations informatiques, simuler l'évolution du système actuel en se basant sur une multitude de lois physiques [39]. Le développement d'un tel modèle est très complexe, et les ressources informatiques nécessaires à son fonctionnement sont substantielles. La prévision météorologique étant stratégique pour un pays, les modèles de NWP sont par conséquent généralement réalisés par des agences gouvernementales. Par exemple, elle est produite aux États-Unis par l'Agence américaine d'observation océanique et atmosphérique notée en anglais NOAA, en France par Météo-France et au Canada par Environnement Canada. Bien que très répandus et largement utilisés partout dans le monde, les modèles physiques présentent certaines faiblesses. D'abord, comme ces modèles se basent sur les conditions initiales des paramètres de l'atmosphère, la moindre erreur dans les variables de départ peut se propager et s'amplifier avec l'évolution dans le temps. En général, des méthodes statistiques de sortie de modèle, Model Output Statistics (MOS), sont utilisées pour réduire les erreurs reliées à la NWP [40]. Par ailleurs, comme le temps de calcul est très élevé, les prévisions numériques sont exécutées généralement entre 2 et 4 fois par jour seulement, ce qui contraint leur utilisation au moyen terme et au long terme [41].

### **2.3. État de l'art**

Dans la littérature éolienne, certains auteurs ont dédié une grande partie de leur recherche à la prédiction de la vitesse du vent ou de la puissance éolienne. C'est notamment le cas de Debnath et coll. [42] ont passé en revue 483 modèle de planification énergétique notée en anglais (EPM) et ont révélé l'utilisation de cinquante méthodes différentes pour la période entre 1985 et 2017. Parmi ces 50 méthodes identifiées, nous avons trouvé l'intelligence informatique, la programmation mathématique (MP) et les méthodes statistiques. Pour les méthodes d'intelligence informatique, il y avait des réseaux de neurones artificiels (RNA) utilisés dans 194 EPM, suivis des machines à vecteur support (SVM) appliquées dans 58 modèles, de la logique floue dans 40 modèles, de l'algorithme génétique dans 39 modèles, de l'optimisation des essais de particules dans 34 modèles et de la prédiction des gris dans 29 modèles. Dans le cas des méthodes statistiques, la moyenne mobile intégrée autorégressive (ARIMA), la régression linéaire et la moyenne mobile autorégressive (ARMA) sont utilisées pour la prévision. Même ces trois méthodes sont appliquées à la charge électrique et aux énergies renouvelables, la précision des prévisions des méthodes d'intelligence informatique était meilleure que celle des méthodes statistiques et programmations mathématiques.

Concernant les prédictions horaires de la vitesse du vent Mohandes et al [43] ont exploité dans leurs travaux les réseaux de neurones artificiels avec trois couches de type perceptron multicouche avec un algorithme d'apprentissage de

rétropropagation du gradient, les résultats trouvés pour l'erreur de prédiction (MAPE) et le coefficient de corrélation étaient 8.9% et 0.9380 respectivement. De la même manière, Velo et al. [44] ont utilisé aussi RNA avec trois couches et la rétropropagation du gradient comme algorithme d'apprentissage et les résultats ont été 0.5 pour R et 6% pour MAPE. Chandel et al. [45] ont employé les RNA pour prédire la vitesse du vent en utilisant comme données la température, la pression, rayonnement solaire et altitude. Les résultats obtenus ont été 6.489% pour MAPE et 0.99 pour R. Dans ces travaux, Palomares-Salas et al [46] ont été exploités les RNA et comparés leur performance avec ARIMA et le modèle de persistance. Le meilleur modèle a été le RNA avec rétropropagation de gradient comme algorithme d'apprentissage et une seule couche cachée de neuf neurones ( 0.18 m/s pour MAE et 0.05 m<sup>2</sup>/s<sup>2</sup> pour MSE).

### **1.7. Conclusion**

Le thème abordé dans ce chapitre permet d'exploiter la conversion de l'énergie éolienne, non polluante, en énergie électrique. L'étude consiste à travailler sur la vitesse du vent qui est le cœur et la source principale de l'énergie éolienne. Premièrement, nous avons vu dans ce chapitre les informations générales et les différents types des aérogénérateurs. D'autre part, nous avons présenté les caractéristiques de la vitesse du vent ainsi que les différents appareils de mesure. Finalement, une étude approfondie du fonctionnement d'une éolienne pour savoir le passage d'une énergie cinétique du vent à une énergie électrique afin d'injecter dans un réseau. Finalement, nous avons fait une comparaison bibliographique des méthodes qui sont employées dans le domaine de la prévision de la vitesse du vent.

### 2.1. Introduction

Le présent chapitre est consacré à une catégorie de méthodes de prévision pour la modélisation linéaire et non linéaire de la vitesse du vent. Nous allons commencer par une distinction entre la prédiction et la prévision ainsi que les horizons de prévision de la vitesse du vent. Ensuite, nous présenterons les différents types de la modélisation linéaire ainsi que les équations de régression et les tests adoptés. Ensuite, Nous allons décrire deux méthodes très connues de prévision, la première est les réseaux de neurones artificiels qui sont très populaires et se montrent d'une grande efficacité dans le domaine de la prévision. La deuxième méthode est les machines à vecteurs supports qui sont montrés ses efficacités de traitement des problèmes non linéaire. Après, nous allons définir deux algorithmes métaheuristiques qui sont très connus à savoir l'Algorithme Génétique (GA) et Optimisation par Essaim des Particules (appelé en anglais PSO) pour développer des modèles hybrides de prévision de la vitesse du vent à court terme. A la fin, nous allons présenter les indicateurs de performance utilisés dans ce travail.

### 2.2. Distinction entre la prédiction et la prévision

En général, la prédiction ou la prévision veulent dire une extrapolation dans le temps en utilisant un historique d'observations du passé. Cependant, il existe une nuance entre ces deux termes importants et il y a plusieurs explication pour faire la distinction entre ces deux termes scientifiques.

D'après l'explication du Professeure Galit Shmueli de Tsing Hua University, "Prédire" ou "Prévoir" signifient la même chose. Dans le monde des statistiques, nous avons tendance à utiliser le mot « prédire » dans le contexte « transversal » où nous avons des données sur de nombreuses observations à un moment donné. Nous réservons le terme « prévision » pour un contexte de séries temporelles où nous mesurons une chose au fil du temps. C'est similaire à la distinction entre une "vidéo fixe" et une "photo". Et d'après l'explication du Dr. Cyril Voyant, le terme prévision est employé pour définir une estimation analytique d'une grandeur sur un temps continu et la prédiction pour connaître cette grandeur à un temps donnée [47].

Dans notre travail, nous opterons pour la première explication car nous travaillons avec des séries chronologiques.

### 2.3. Les horizons de prévision

La prévision de n'importe quelle grandeur physique nécessite de fixer l'horizon de prévision. Les prévisions éoliennes à différents horizons sont nécessaires pour accroître la part de l'énergie éolienne dans les réseaux électriques. En effet,

le caractère intermittent de la vitesse du vent peut entraîner des déséquilibres entre l'offre et la demande d'électricité [48] ce qui impose des mesures de gestion et de planification pour différents horizons spatio-temporels. Ces horizons de prévision diffèrent de par la granularité des données utilisées, le besoin exprimé et les méthodes de prévision utilisées. Chaque horizon de prévision répond à un besoin précis et nécessite l'utilisation de méthodes différentes. On distingue trois horizons principaux de prévision :

- La prévision à court terme : de quelques secondes à quelques jours.
- La prévision à moyen terme : de quelques jours à une semaine.
- La prévision à long terme : d'une semaine à des mois.

## **2.4. Différentes méthodes de prévision proposées**

Dans cette section, nous mettrons l'accent sur les méthodes de prédiction qu'on va utiliser le long de ce travail. Suite à une revue de littérature approfondie des techniques de prédiction de la vitesse du vent, nous nous sommes particulièrement penchés sur deux méthodes de prédiction très populaires. La première est le réseau de neurones artificiel qui montre une grande efficacité dans le domaine de la prévision météorologique. La deuxième est les machines à vecteurs de support qui sont montrés ses efficacités de traitement des problèmes de discrimination non linéaire.

### **2.4.1. Modélisation statistique par régression linéaire:**

La modélisation statistique est une méthode ou technique qui permet de collecter les données, le traitement et l'interprétation des données collectées. La modélisation statistique se fait à l'aide des logiciels et dans notre thèse, on a choisi STATGRAPHICS comme logiciel de modélisation [49,50].

Une des parties les plus utilisées de STATGRAPHICS est celle relative à la modélisation statistique par méthodes de régression. Dans un modèle de régression, une variable de réponse Y est exprimée comme une fonction d'une ou plusieurs variables prédictives X plus un bruit. Dans beaucoup de cas, mais pas tous, la forme de la fonction est linéaire par rapport aux coefficients inconnus, si bien que le modèle peut être exprimé sous la forme :

$$y_i = \sum_{i=1}^n \beta_i x_i + \varepsilon \quad (2.1)$$

Où l'indice i représente la i-ème observation dans l'échantillon des données, les  $\beta_i$  sont les Coefficients inconnus du modèle et  $\varepsilon$  est un terme résiduel aléatoire.

À partir d'un jeu de données contenant une variable de réponse Y et une ou plusieurs Variables prédictives, le but de l'analyse de régression est de construire un modèle qui :

1) Décrit la relation entre les variables de façon à permettre de bien prévoir Y à partir de valeurs connues des X.

2) Ne contient pas plus de variables X que nécessaire pour obtenir une bonne prévision

Il y a plusieurs types de modèles de régression.

#### 2.4.1.1. Régression simple

Dans le cas de la régression simple, dans laquelle se retrouve une seule variable indépendante [51], la relation peut s'écrire de la façon suivante :

$$Y = \beta_0 + \beta_1 X + \varepsilon \quad (2.2)$$

Y : variable dépendante

X : variable indépendante

$\beta_0$  : ordonnée à l'origine (constante)

$\beta_1$  : coefficient de régression (paramètre estimé, la pente)

$\varepsilon$  : erreur d'estimation

La boîte de dialogue d'entrée des données doit être renseignée comme montrée ci-dessous :

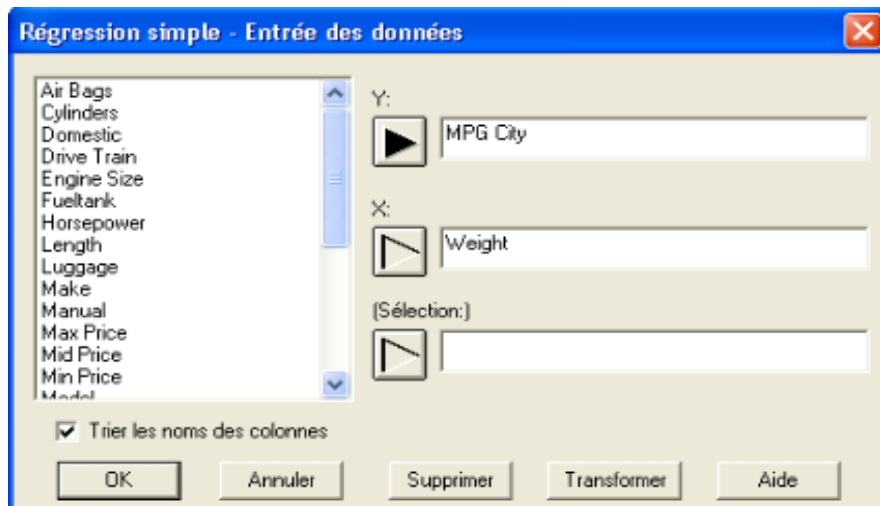


Figure 2.1: Boîte de dialogue des données d'entrée pour la régression simple

La fenêtre d'analyse initiale comporte des sous-fenêtres donnant des informations sur le modèle ajusté. Le résumé de l'analyse, affiché dans la fenêtre en haut à gauche, résume l'ajustement : parmi les nombreuses statistiques dans ce tableau, les statistiques les plus importantes sont : les coefficients estimés du modèle, le R-carré et la valeur de la probabilité. On va donner dans la partie suivante la définition de chaque statistique. Le graphique en haut à droite de la fenêtre d'analyse affiche le modèle ajusté.

Ce graphique affiche la droite de régression des moindres carrés et deux jeux de limites. Les limites internes donnent des intervalles de confiance pour la valeur

moyenne de Y. Les limites externes donnent des intervalles de prévision pour de nouvelles observations.

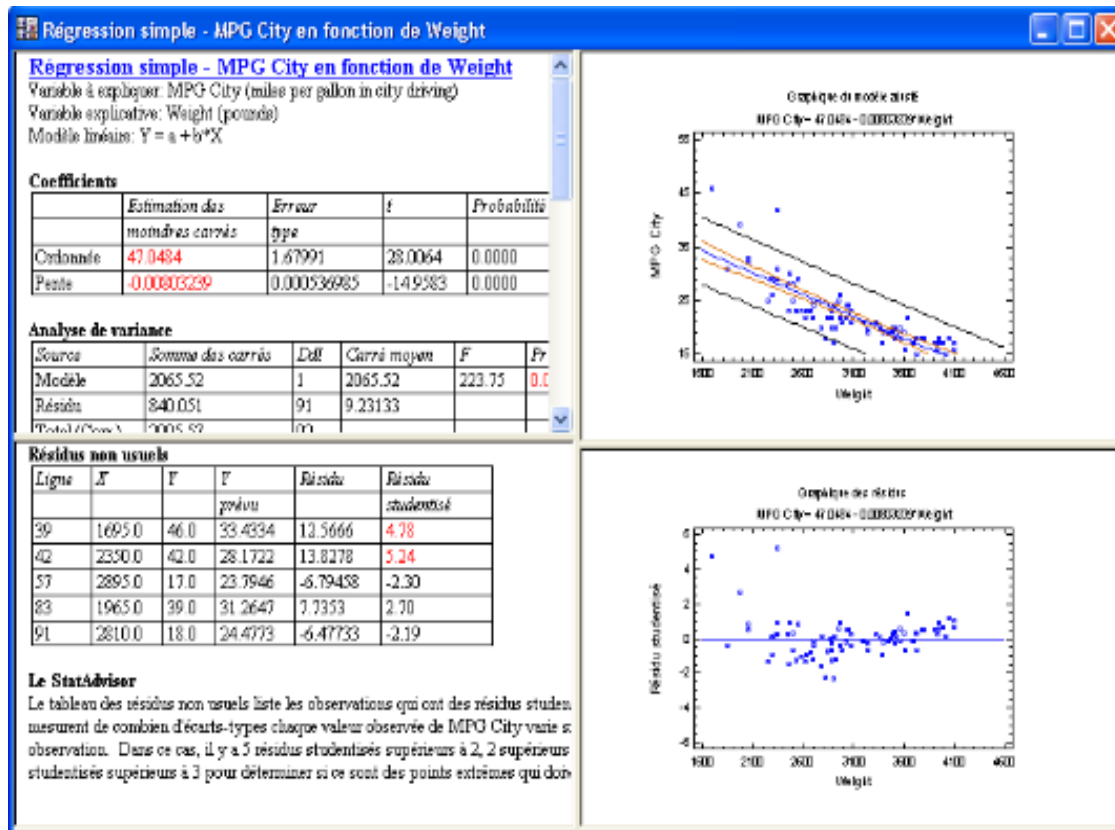


Figure 2.2: Fenêtre d'analyse de la régression simple

#### 2.4.1.2. Régression polynomiale

Cette méthode consiste à ajouter à la variable explicative  $x$  de nouvelles variables construites en mettant  $x$  au carré, au cube, etc. [52]. L'équation devient donc, pour un polynôme du nième ordre et une variable explicative  $x$  :

$$Y = P_n(x) + \varepsilon \quad (2.3)$$

$$= \beta_0 + \beta_1 x + \beta_2 X^2 + \beta_3 X^3 + \beta_n X^n + \varepsilon$$

$Y$  : variable à expliquer,  $X$  : variable explicative,  $P_n$  : polynôme de degré  $n$ , et  $\varepsilon$  : perturbation

Une façon simple de se représenter l'effet de l'ajout d'un ordre est la suivante : Chaque nouvel ordre permet d'ajouter un pli à la courbe. Une équation du premier ordre est celle d'une droite, du deuxième ordre une parabole, le troisième ordre donne un S couché, etc. : Cette figure montre comment l'ajout

d'éléments de souplesse permet de mieux ajuster le modèle de régression aux données. Cependant, comme toujours, il faut savoir trouver un compromis entre un modèle trop rudimentaire et mal ajusté et un modèle ajustant bien les données.

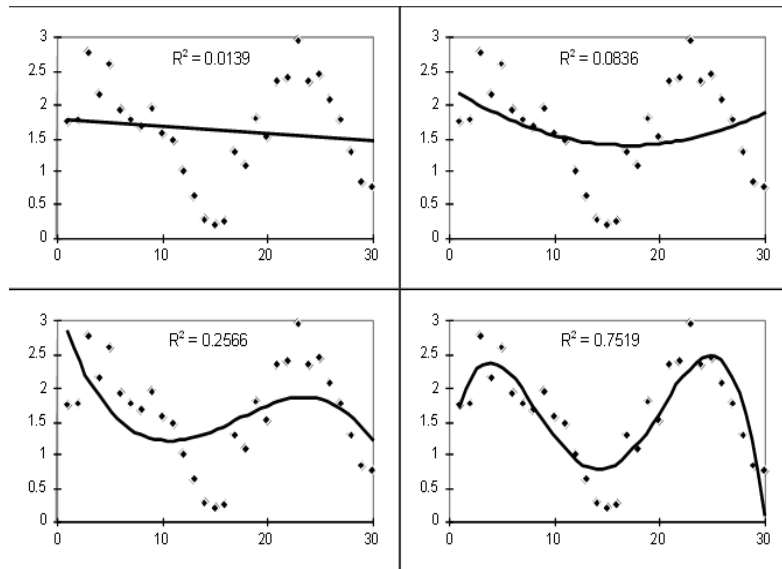


Figure 2.3: Modèle ajusté du premier, deuxième, troisième et quatrième ordre

#### 2.4.1.3. Régression Multiple

Pour améliorer le modèle, d'autres variables prédictrices doivent être ajoutées. Cela peut être fait aisément par l'analyse de régression multiple [53]. Voici la boîte de dialogue d'entrée des données de cette analyse : La procédure de régression multiple ajuste un modèle reliant une variable de réponse  $Y$  à de multiples variables prédictrices  $X_1, X_2, \dots, X_p$

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \dots + \varepsilon \quad (2.4)$$

#### 2.4.1.4. Analyse de la variance (ANOVA)

Les résultats de l'ANOVA sont contenus dans le tableau de l'ANOVA affiché dans la sous-fenêtre en haut à gauche de la fenêtre d'analyse [54,55]. Pour généraliser les différentes définitions que nous allons donner ici, nous nous plaçons dans un modèle de régression multiple qui fait l'objet de l'étude du chapitre suivant. La première table d'analyse de la variance donne les résultats du test des hypothèses :

$$H_0 = \{Y_i = \beta_0 + \varepsilon_i\} \text{ contre } H_1 = \{Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \beta_p X_{pi} + \varepsilon_i\}$$

Ainsi que d'autres informations que nous allons préciser. En régression linéaire simple, ce test s'écrit.

$$H_0 = \{Y_i = \beta_0 + \varepsilon_i\} \text{ contre } H_1 = \{Y_i = \beta_0 + \beta_1 x + \varepsilon_i\}$$

Et qui s'exprime littéralement par  $H_0 = \{\text{il n'existe pas de liaison entre } Y \text{ et } x\}$  et  $H_1 = \{\text{il existe une liaison entre ces deux variables qui est linéaires}\}$ . Dans le cadre de la régression linéaire multiple, les quantités de la table d'analyse de la variance se définissent comme suit dans le tableau ci-dessous.

Tableau 2.1 : Analyse de la variance

| Source | Degrés de liberté | Somme de carrés | Carré moyen | Statistique de test             | Probabilité critique     |
|--------|-------------------|-----------------|-------------|---------------------------------|--------------------------|
| Modèle | p                 | SCM             | SCM /p      | $F = \frac{SCM/p}{SCR/(n-p-1)}$ | $P = (f_{p, n-p-1} > F)$ |
| Résidu | n-p-1             | SCR             | SCR/(n-p-1) |                                 |                          |
| Total  | n-1               | SCT             | SCT/(n-1)   |                                 |                          |

Où les différentes sommes des carrés sont définies par :

$$- SCT = \sum_{i=1}^n (Y_i - \bar{Y})^2 = \text{Somme des carrés totale} \quad (2.5)$$

$$- SCM = \sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2 = \text{Somme des carrés due au modèle} \quad (2.6)$$

$$- SCR = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \text{Somme des carrés des erreurs} \quad (2.7)$$

Avec  $\bar{Y}$  est la moyenne générale des observations

$$\bar{Y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (2.8)$$

$\hat{Y}_i$  : Les valeurs prédites

$F_{p, n-p-1}$  : Désigne la loi de Fisher à p et (n-p-1) degrés de liberté

Il est facile de voir que SCT se décompose en la somme des deux autres termes ( $SCT = SCM + SCR$ ). Cela signifie que la variabilité totale de Y est décomposée en variabilité due au modèle et en variabilité résiduelle. La statistique de test F suit sous l'hypothèse  $H_0$  une loi de Fisher à p et (n-p-1) degrés de liberté  $f_{p, n-p-1}$ . On utilise cette propriété pour calculer la probabilité critique (Si la probabilité critique est inférieure à 0,05, on rejette l'hypothèse  $H_0$ ) [56,57,58].

- L'erreur type du coefficient  $\beta_i$  de la régression (Standard Error). C'est une mesure de dispersion de données, elle est équivalente à l'écart type pour la moyenne d'une variable.
- T-statistics : C'est un rapport du paramètre estimé et son erreur standard. Il est utilisé dans des tests d'hypothèses.
- Le  $R^2$  ou R-squared s'appelle le coefficient de détermination. Il représente la proportion de la variation totale de Y expliquée par l'ensemble des variables explicatives.
- Le " $R^2$ " ajusté ou R-squared (adjusted for df) : est l'ajustement du  $R^2$  au nombre p de variables explicatives. En effet, il est clair que plus on ajoute de variables explicatives, meilleur sera l'ajustement du modèle aux données. Mais en considérant trop de variables, on risque d'obtenir un modèle qui est sur-ajusté et ce n'est pas ce que l'on cherche.
- L'erreur type d'estimation (Standard Error of Estimate) est l'indicateur de la performance prédictive du modèle de régression.
- Durbin watson statistics : Durbin-Watson est une statistique de test utilisé pour détecter la présence d'autocorrélation dans les résidus (erreurs de prédiction) à partir d'une analyse de régression. Il est nommé d'après James Durbin et Watson Geoffrey.

#### **2.4.1.5. Examiner les résidus**

Une fois qu'un modèle acceptable a été choisi, les résidus du modèle ajusté doivent être examinés [59]. En général, un résidu peut être vu comme la différence entre la valeur observée de

Y et la valeur prévue par le modèle :

$$\text{Résidu} = Y_{\text{observé}} - Y_{\text{prévu}} \quad (2.9)$$

En utilisant les options pour la fenêtre, il est possible de choisir un graphique des résidus ou un graphique des résidus studentisés. Les résidus studentisés sont calculés comme les résidus ordinaires définis ci-dessus divisés par leurs erreurs types estimées. Un résidu studentisé indique donc de combien d'erreurs types un point est éloigné du modèle comme montre le graphe ci-dessous.

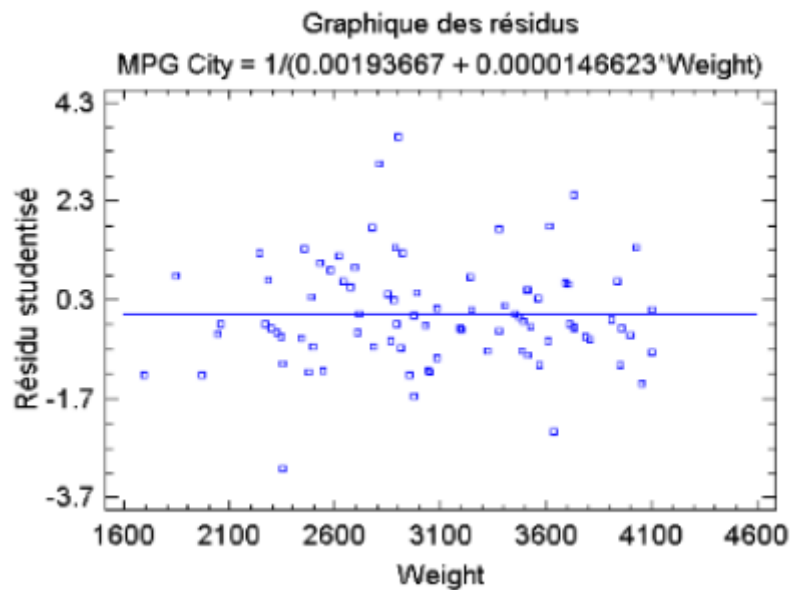


Figure 2.4 : Graphe du résidu studentisé

## 2.4.2. Modélisation par réseaux de neurones artificiels

### 2.4.2.1 *Qu'est-ce qu'un neurone biologique ?*

Un neurone est une cellule nerveuse constituant le système nerveux central qui en possède environ cent milliards. Il réalise de nombreuses applications telles que la communication et le traitement d'information. Un neurone comprend trois parties [60] :

- Le corps cellulaire : Ce sont des cellules somatiques, il fait la somme des informations qui lui parviennent. Ils traitent ensuite ces informations et renvoient les résultats sous forme des signaux électriques.
- Les dendrites : ou entrées des neurones, ce sont les récepteurs principaux du neurone pour capter les signaux qui lui parviennent.
- L'axone : est la sortie du neurone vers d'autres neurones; les axones reliant les neurones entre eux jouent donc un rôle important dans le comportement logique de l'ensemble.
- Les synapses : Ils reçoivent les informations des autres neurones via l'axone et permettent donc aux neurones de communiquer entre eux.

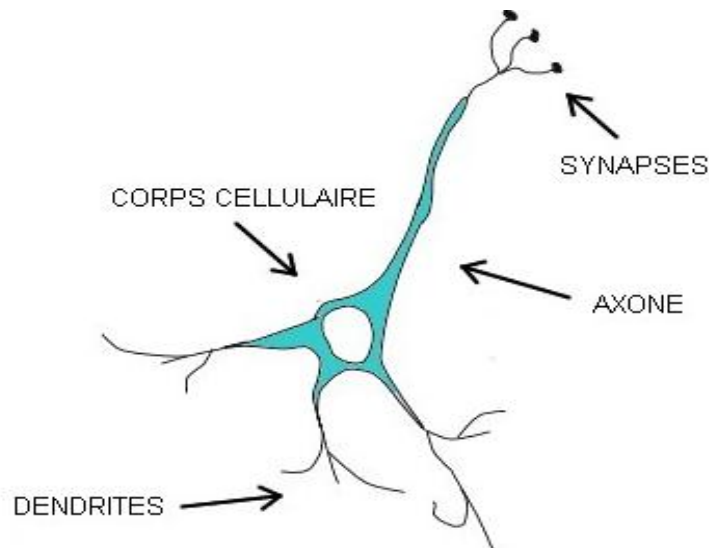


Figure 2.5: neurone biologique

#### 2.4.2.2. *Qu'est-ce qu'un neurone formel ?*

Les neurones formels sont dotés de caractéristiques inspirées de celles des neurones biologiques que nous avons passées en revue dans la section précédente. Par analogie avec le neurone biologique, le neurone artificiel est un système dont les "dendrites" représentent ses entrées et l'axone sa sortie. Il est l'unité élémentaire des réseaux de neurones artificiels [61,62,63].

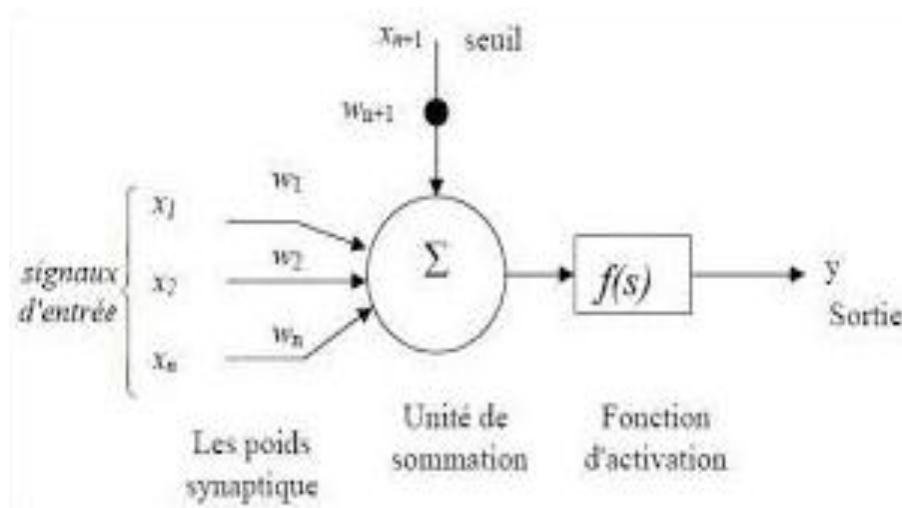


Figure 2.6: Schéma d'un neurone formel

#### 2.4.2.3. *Qu'est-ce qu'un réseau de neurones ?*

Le neurone artificiel étant défini, il suffit d'en associer plusieurs pour former un réseau. Il existe plusieurs types de réseaux suivant la façon dont sont reliés les neurones.

#### 2.4.2.4. Architecture d'un réseau de neurones artificiels

Dans les réseaux de neurones artificiels, la première couche constitue les neurones d'entrée (c'est la d'entrée, ou couche sensorielle) et la dernière couche est la couche de sortie. S'il n'y a que la couche d'entrée et celle de sortie, on parle de **réseau monocouche**, sinon de **réseau multicouche** et les couches intermédiaires sont dites **couches cachées**. Il existe trois architectures de réseaux de neurones artificiels [64,65]:

##### ➤ Réseaux neuronaux non bouclés

Un réseau neuronal non bouclé (appelé aussi feed-forward) est un réseau neuronal sans rebouclage, c'est-à-dire sans retour d'information dans les neurones précédemment parcourus. L'information chemine sans retour en arrière et donc en un temps restreint [66]. Dans ce type de réseau, chaque neurone peut avoir des connexions à des neurones de la couche inférieure et de la couche supérieure, mais pas aux autres neurones (pas même ceux de la même couche).

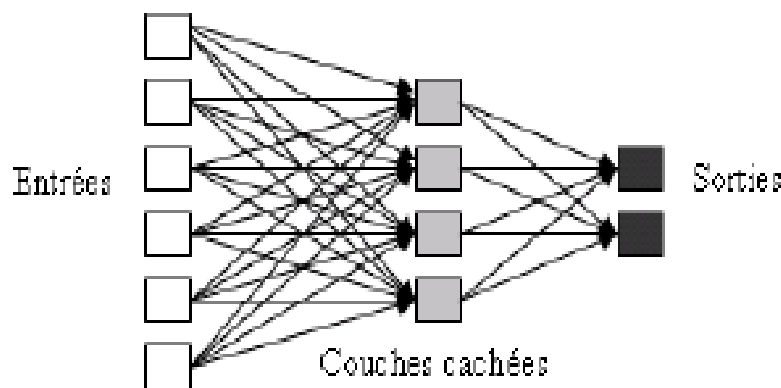


Figure 2.7: Réseau neuronal non bouclé

##### ➤ Réseaux neuronaux bouclés

Les réseaux neuronaux bouclés ou récurrents contiennent des boucles et retours en arrière. La sortie des neurones peut être rebouclée vers des neurones précédemment parcourus, ainsi le fonctionnement asynchrone de ce réseau, bien que plus lent à cause de la réentrance, permet, à nombre égal de neurones, des capacités énormément plus vastes [67].

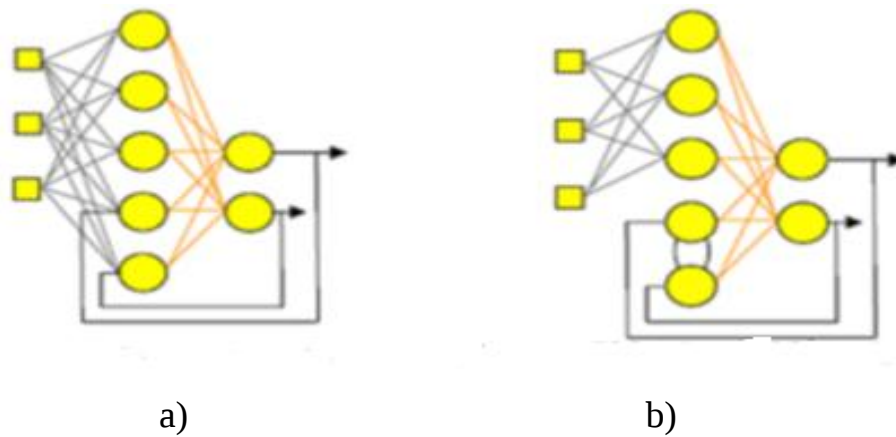


Figure 2.8: Réseaux neuronaux bouclés : a) réseau de Jordan b) réseau récurrent partiel avec doubles connexions

➤ Réseaux neuronaux complètement connectés

Les réseaux neuronaux complètement connectés sont des réseaux ayant des connexions entre toutes les cellules et chacune possède des retours sur elles-mêmes[68,69]. Les neurones sont numérotés de telle sorte que pour chaque neurone:

- Les entrées sont toutes les entrées du réseau ainsi que les sorties des neurones de numéro inférieur.
- La sortie est connectée aux entrées de tous les neurones de numéro supérieur.

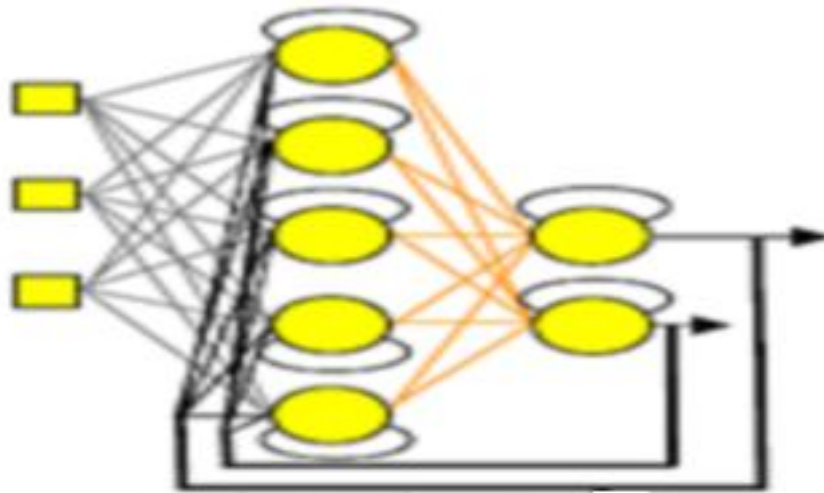


Figure 2.9 : Réseau neuronal complètement connecté

#### **2.4.2.5. Modélisation d'un réseau de neurones artificiels**

##### **2.4.2.5.1 Réseau neuronal monocouche**

Un réseau de neurones monocouche, aussi appelé perceptron, est caractérisé de la manière suivante. Il possède  $n$  informations en entrée ; il est composé de  $p$  neurones, que l'on représente généralement alignés verticalement.

##### **2.4.2.5.2 Réseau neuronal multicouche**

Pour le réseau neuronal multicouche, on considère une couche d'entrée qui correspond aux variables d'entrée, une couche de sorties, et un certain nombre de couches intermédiaires. Les liens n'existent qu'entre les cellules d'une couche avec les cellules de la couche suivante. Nous allons restreindre notre étude aux réseaux neuronaux à une seule couche cachée. La couche cachée permet plus d'interaction et instinctivement, on est conscient que notre réseau de neurones pourra apprendre des "fonctions" plus complexes en rajoutant une couche. Le fonctionnement n'est pas pour autant complexe. On a toujours le choix de la fonction d'activation, et l'évaluation de la sortie d'un neurone se déroule de la même manière. La seule différence est que la sortie d'un neurone de la couche cachée sera l'une des informations d'entrées des neurones de la couche de sortie [70].

##### **2.4.2.5.3. Les différentes fonctions de transfert**

La fonction d'activation, ou fonction de transfert est une fonction qui doit renvoyer un réel proche de 1 quand les "bonnes" informations d'entrée sont données et un réel proche de 0 quand elles sont "mauvaises". On utilise généralement des fonctions à valeurs dans l'intervalle réelles  $[0,1]$ . Quand le réel est proche de 1, on dit que l'unité (le neurone) est active alors que quand le réel est proche de 0, on dit que l'unité est inactive. Le réel en question est appelé la sortie du neurone et sera noté  $a$ . Si la fonction d'activation est linéaire, le réseau de neurones se réduirait à une simple fonction linéaire.

En effet, si les fonctions d'activations sont linéaires, alors le réseau est l'équivalent d'une régression multilinéaire (méthode utilisée en statistiques) . L'utilisation du réseau de neurones est toutefois bien plus intéressante lorsque l'on utilise des fonctions d'activations non linéaires. Il existe de nombreuses formes possibles pour la fonction de transfert. Les plus courantes sont présentées sur la figure 2.10 et 2.11.

**La fonction sigmoïde** présente l'avantage d'être dérivable (ce qui va être utile par la suite) ainsi que de donner des valeurs intermédiaires (des réels compris entre 0 et 1), définie par :

$$f(x) = \frac{1}{1 + e^{-x}}$$

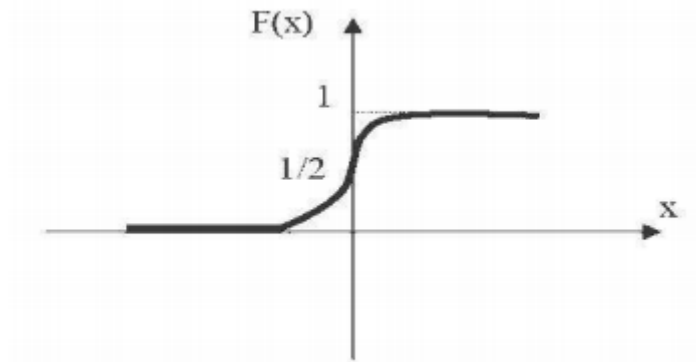


Figure 2.10 : Fonction sigmoïde

**La fonction tangente hyperbolique** est une fonction non linéaire définie sur l'intervalle  $[-1,1]$  par :

$$f(x) = \tanh(x)$$

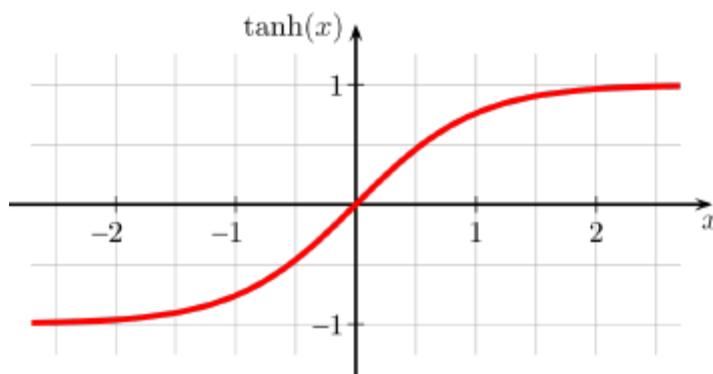


Figure 2.11 : Fonction tangente hyperbolique

#### 2.4.2.6 Apprentissage des réseaux de neurones

L'apprentissage (sous-entendu : artificiel, automatique, Machine Learning) est une notion qui englobe toute méthode permettant de construire un modèle de la réalité à partir de données, soit en améliorant un modèle partiel ou moins général, soit en créant complètement le modèle. Il existe deux tendances principales en apprentissage, celle issue de l'intelligence artificielle et qualifiée de symbolique, et celle issue des statistiques et qualifiée de numérique.

#### **2.4.2.6.1. Principe**

L'apprentissage est la phase durant laquelle le comportement du réseau est modifié jusqu'à l'obtention du comportement désiré. Cette phase agit sur les poids, cependant après le calcul d'erreur des sorties désirées par rapport aux sorties trouvées, les poids sont modifiés de sorte que l'erreur soit optimale pour le prochain balayage du réseau neuronal tout en utilisant des algorithmes appelés algorithmes d'apprentissage ou règles d'apprentissage [71].

#### **2.4.2.6.2. Procédures d'apprentissage**

On présente au réseau une base des entrées  $x_i$ , et on lui demande de modifier sa pondération  $w_i$  (les poids synaptiques) de manière à s'approcher au mieux de la sortie  $y_i$ , correspondante, en se basant sur les algorithmes d'apprentissage jusqu'à ce que l'erreur commise entre la sortie calculée et celle connue soit minimisée.

Plusieurs modes d'apprentissage peuvent être distingués, ceux-ci se différencient principalement par le type des exemples disponibles [72].

- *Apprentissage supervisé*

L'apprentissage supervisé est une tâche d'apprentissage automatique consistant à apprendre une fonction de prédiction à partir d'exemples annotés. Il consiste en des variables d'entrée ( $X$ ) et une variable ( $Y$ ). Il faut utiliser un algorithme pour apprendre la fonction de mappage de l'entrée à la sortie ( $y=f(x)$ ), si le mappage est bien déterminé, vous pouvez prévoir les variables de sortie  $Y$  pour de nouvelles données d'entrée ( $x$ ). La majorité des experts machine learning utilisent un apprentissage supervisé.

- *Apprentissage non supervisé*

L'apprentissage non supervisé consiste à ne disposer que des données d'entrée ( $X$ ) et pas de variables de sortie correspondantes. L'objectif de l'apprentissage non supervisé est de modéliser la structure ou la distribution sous-jacente dans les données afin d'en apprendre davantage sur les données. Celles-ci sont appelées apprentissage non supervisé, car, contrairement à l'apprentissage supervisé ci-dessus, il n'y a pas de réponse correcte ni d'enseignant. Les algorithmes sont laissés à leurs propres mécanismes pour découvrir et présenter la structure intéressante des données.

- *Apprentissage semi-supervisé*

L'apprentissage supervisé est la technique permettant d'accomplir une tâche en fournissant aux systèmes une formation, des modèles d'entrée et de sortie, tandis que l'apprentissage non supervisé est une technique d'autoapprentissage dans laquelle le système doit découvrir les caractéristiques de la population d'entrée

par lui-même. Alors l'apprentissage semi-supervisé se situe entre les deux types d'apprentissages supervisés et celui non supervisé en fournissant des entrées et des indicateurs sur les sorties désirées (vrai/ faux).

#### 2.4.2.6.3. Apprentissage du réseau neuronal monocouche

L'apprentissage du réseau monocouche est un apprentissage qui concerne un modèle perceptron qui est considéré comme le modèle le plus basic d'un réseau neuronal, car il ne contient qu'une seule couche cachée. En ce qui suit, nous allons donner les règles d'apprentissage concernant ce type de réseaux [73].

##### a) Règle de Hebb

La première règle d'apprentissage a été formulée en 1949 par Hebb. Son principe considère que si deux neurones connectés entre eux sont activés au même moment, la connexion qui les relie doit être renforcée et par conséquent le poids de cette connexion doit être augmenté pour lui permettre de jouer un rôle déterminant lors du mécanisme d'apprentissage. La règle de Hebb s'écrit sous la forme suivante :

$$W_{ij}(t + \partial t) = W_{ij}(t) + \mu a_i a_j \quad (2.10)$$

$W_{ij}(t)$  est le poids de la connexion entre le neurone  $j$  et le neurone  $i$  à l'instant  $t$

$\mu(\mu > 0)$  est un paramètre de l'intensité d'apprentissage.

##### b) Règle de widrow-Hoff

Cette règle est appelée règle delta, ou règle Adaline, ou encore règle de widrow-Hoff d'après le nom de ses inventeurs. Cet algorithme est une variante très utilisée de l'algorithme précédent. Au lieu de calculer les variations des poids en sommant sur tous les exemples, l'idée est de modifier les poids à chaque présentation des exemples dans le but de trouver une frontière ou un hyperplan qui soit le plus médian possible par rapport à ces deux classes.

Pour la règle de widrow-hoff, l'apprentissage est fini lorsque la frontière entre deux classes d'exemples est équidistance. Cette règle consiste à minimiser l'expression quadratique :  $\sum_{r=1}^R (d_{(r)m} - \sum_{n=1}^{N+1} W_{mn} x_{(r)n})^2$ , afin que l'activité  $\sum_{n=1}^{N+1} W_{mn} X_{(r)n}$  Passe au milieu du nuage d'exemples de chaque classe, tout en ajoutant au poids la correction  $\Delta W$ .

$$\text{Avec : } \Delta W_{(r)m} = \eta (d_{(r)m} - \sum_{n=1}^{N+1} W_{mn} x_{(r)m} x_{(r)n}) \quad (2.11)$$

Où  $\eta$  est une valeur réelle positive.

$W_{mn}$  : Les poids de connexion

$d_{(r)m}$ : Valeurs attendues en sortie

L'apprentissage est considéré comme terminé lorsque cette somme de dérivées est nulle :

$$\begin{aligned}\sum_{r=1}^R \Delta W_{(r)mn} &= \frac{\eta}{2} \frac{\partial (\sum_{r=1}^R (d_{(r)m} - \sum_{n=1}^N W_{mn} x_{(r)n})^2)}{\partial W_{mn}} \\ &= \sum_{r=1}^R \eta \cdot (d_{(r)m} - \sum_{n=1}^N W_{mn} x_{(r)n}) \cdot x_{(r)n}\end{aligned}\quad (2.12)$$

#### 2.4.2.6.4 Apprentissage du réseau neuronal multicouche

De la même manière que le perceptron monocouche, le perceptron multicouche est lui aussi capable d'apprentissage. En effet, il existe également des algorithmes permettant de corriger les poids vis-à-vis d'un ensemble d'exemples donnés [74]. Parmi ces algorithmes il y a des algorithmes de premier ordre comme celui de rétropropagation du gradient de l'erreur, ou bien de deuxième ordre comme de levenberg-marquard.

##### ***a- Algorithme de rétropropagation du gradient (backpropagation)***

L'algorithme de rétropropagation du gradient consiste à calculer l'erreur du réseau de neurone artificiel en propageant cette dernière sur tout le réseau afin de converger vers la sortie désirée tout en agissant sur les poids. Cela signifie que dans cet algorithme de même qu'on est capable de propager un signal provenant des cellules d'entrées vers la couche de sortie, on peut, en suivant le chemin inverse, rétropropager l'erreur commise en sortie vers les couches internes. Le schéma suivant résume les étapes de cette méthode :

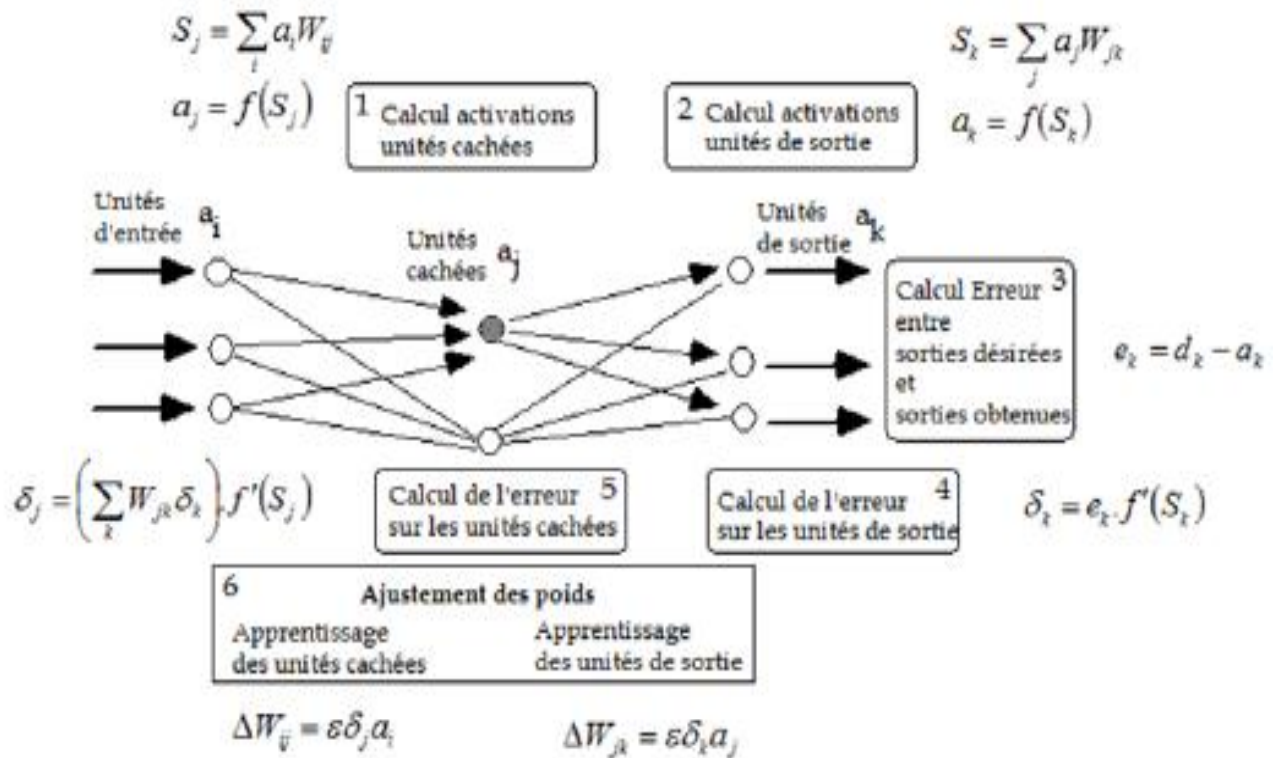


Figure 2.12 : Schéma des étapes de l'algorithme de rétropropagation

### b- Algorithme de levenberg- Marquardt

Cet algorithme consiste à minimiser l'erreur quadratique tout en agissant sur les poids et l'erreur en même temps, en considérant l'erreur comme fonction des poids et en changeant ces derniers avec la formule :

$$\Delta W = -[\nabla^2 E(W) + \mu I]^{-1} \nabla E(W) \quad (2.13)$$

$\mu$  : taux d'apprentissage

E : l'erreur

$\Delta W$  : correction des poids

Il est multiplié par un facteur  $\beta$  lorsque  $E(W)$  augmente et est divisé par ce même facteur lorsque  $E(W)$  diminue.

Les valeurs initiales de ces facteurs sont :  $\beta = 10$  et  $\mu = 0.1$  .

### 2.4.3. Machine à vecteurs support

La méthode des machines à vecteur support noté en anglais (SVM) ou séparateurs à vaste marge, développée par Vapnik (1999), permet de classifier un nombre élevé de variables de façon non linéaire par apprentissage supervisé. L'apprentissage, c'est de collecter des informations sur des couples (entrées, sorties) pour pouvoir formuler des hypothèses sur le comportement du couple au

futur. La classification est réalisée avec une marge maximale à l'aide de fonctions de similarité (ou kernels) qui a le but de transformer les données et choisir les vecteurs de support [75,76]. L'algorithme doit être entraîné sur un premier ensemble de patients pour déterminer ces vecteurs, qui seront ensuite utilisés pour d'autres patients dont le devenir est inconnu. On peut décrire brièvement l'algorithme du SVM comme suit :

Soit  $D=\{x_i, y_i\}_{i=1}^M$  où  $x_i \in \mathbb{N}^M$  représente la matrice d'entrée et  $y_i \in \{0, 1\}$  représente le vecteur de sortie. Dans le cas où les données sont linéairement séparables, le séparateur optimal prend la forme affine suivante  $\omega \cdot x + b = 0$ . Par ailleurs, dans le cas non linéaire, on change l'espace où les données sont distribuées et on passe d'un espace à deux dimensions à un autre de plus grande dimension appelée espace de redescription. On fait, plus on augmente la dimension du problème à résoudre, plus on accroît la probabilité de trouver un séparateur linéaire entre les classes. On appelle cette transformation non linéaire,  $\Phi(x)$  [77,78]. La contrainte d'optimisation s'écrit comme suit :

$$\text{Min } \frac{1}{2} \omega^T \omega + C \sum_{i=1}^M \xi_i$$

$$\text{Sous les contraintes } \begin{cases} y_i [\omega^T \phi(x_i) + b] \geq 1 - \xi_i & (i = 1, 2, \dots, M) \\ \xi_i \geq 0 & (i = 1, 2, \dots, M) \end{cases} \quad (2.14)$$

Où  $\xi_i$  Sont les termes d'erreurs, C le paramètre de contrôle de pénalisation.

Pour résoudre cette contrainte, on applique le lagrangien à (2.14), on obtient la forme duale suivante :

$$\text{Min } \frac{1}{2} \alpha^T Q \alpha - e^T \alpha$$

Sous les contraintes:

$$\begin{cases} 0 \leq \alpha_i \leq C & (i = 1, 2, \dots, M) \\ y^T \alpha = 0 \end{cases} \quad (2.15)$$

Où  $e^T$  est les vecteurs de tous les 1, Q représente la matrice semi-définie positive et  $\alpha_i$  est le multiplicateur de Lagrange. Les éléments  $Q_{ij} = y_i y_j K(x_i, x_j)$  sont des éléments de Q tels que :  $K(x_i, x_j) = \phi(x_i)^T \phi(x_j)$  est la fonction noyau (kernel function). Ils existent plusieurs autres fonctions noyau, linéaires, polynomiale et le noyau gaussien de type radial (radial basis function [RBF]).

D'après l'analyse précédente, la fonction de classification du SVM est donnée par :

$$F(x) = \text{sign}(\sum_{i=1}^M \alpha_i y_i K(x_i, x_j) + b) \quad (2.16)$$

## 2.5. Algorithmes métaheuristiques

### 2.5.1. Algorithme génétique

Les algorithmes génétiques ont été inventés par Jonh Holland dans les années 60. Repris notamment par Golberg dans les années 70, elles s'inspirent des mécanismes biologiques de mutation et de sélection naturelle des espèces [79,80,81]. Elles imitent et exploitent l'évolution naturelle issue des espèces vivantes, afin de trouver des solutions optimales à des problèmes complexes de combinatoire. Cette technique connaît aujourd'hui un franc succès. On l'utilise dans la résolution de problèmes complexes, nécessitant des temps de calcul élevés comme :

- Prédiction de systèmes non linéaires : bio/ socio/ météo/ écono, etc.
- Conception, design et mise en œuvre des réseaux neuronaux.
- Optimisation des trajectoires des robots.
- Traitement d'images, par exemple la resynthèse des textures naturelles.
- Conception et organisation des circuits électroniques.

#### 2.5.1.1. Principe

Le principe général du fonctionnement d'un algorithme génétique est représenté comme suit : l'initialisation de ce type d'optimisation se fait à partir d'une population qui peut être créée d'une façon aléatoire. Pour passer d'une génération  $k$  à la génération  $k+1$ , les trois opérations suivantes sont répétées pour tous les éléments de la population  $k$ . Des couples de parents  $P_1$  et  $P_2$  sont sélectionnés en fonction de leurs adaptations. L'opérateur de croisement est appliqué avec une probabilité  $P_c$  (généralement autour de 0.6) et génère des couples d'enfants  $C_1$  et  $C_2$ . D'autres éléments  $P$  sont sélectionnés en fonction de leur adaptation. L'opérateur de mutation est appliqué avec la probabilité  $P_m$  ( $P_m$  est généralement très inférieur à  $P_c$ ) et génère des individus mutés  $P_0$ . Le niveau d'adaptation des enfants ( $C_1$ ,  $C_2$ ) et des individus mutés  $P_0$  est ensuite évalué avant insertion dans la nouvelle population [82,83,84]. Différents critères d'arrêt de l'algorithme peuvent être choisis (voir figure 2.13)

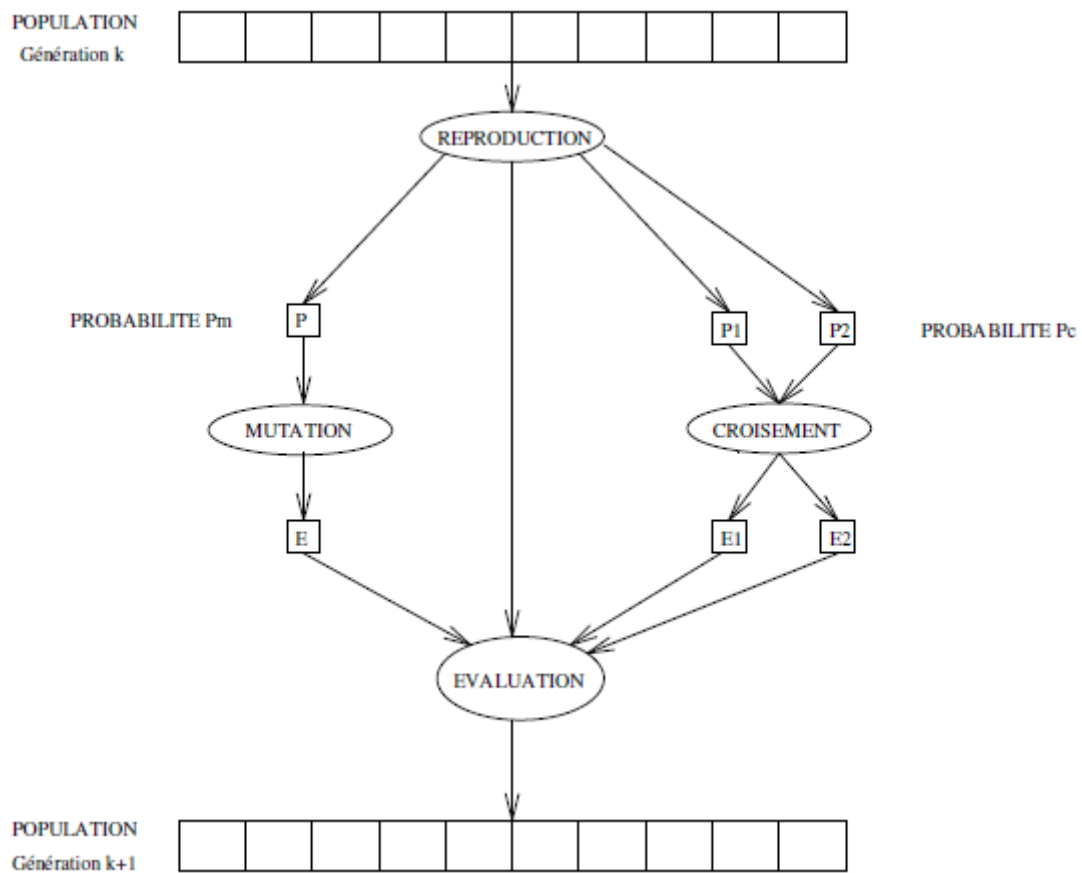


Figure 2.13 : Principe général des algorithmes génétiques

Un algorithme génétique recherche le ou les extrêmes d'une fonction définie sur un espace de données et pour l'utiliser, on doit disposer des cinq éléments suivants :

- Un principe de codage de l'élément de population.
- Un mécanisme de génération de la population initiale.
- Une fonction à optimiser.
- Des opérateurs permettant de diversifier la population au cours des générations et d'explorer l'espace d'état.
- Des paramètres de dimensionnement.

#### 2.5.1.2. Opérateurs génétiques

Les algorithmes génétiques sont des algorithmes d'optimisation s'appuyant sur des techniques dérivées de la génétique et de l'évolution naturelle telles que [85,86, 87]:

##### ➤ La sélection :

Elle a pour but de choisir des individus dans la population courante à partir de leur fitness il existe plusieurs méthodes de sélection :

- La méthode de roulette Wheel.
- La méthode élitiste.
- La sélection par tournois.
- La sélection universelle stochastique.

➤ *Le croisement :*

Il permet à partir de deux individus parents d'obtenir deux enfants, en y propageant certaines des caractéristiques génétiques des parents

➤ *La mutation :*

Elle engendre de nouveaux individus en changeant certains gènes de façon aléatoire ; le but est de parcourir l'espace de recherche.

## 2.5.2.Optimisation par essaims des particules

### 2.5.2.1. Définition

Les algorithmes d'optimisation par essaim des particules (appelé en anglais PSO) ont été introduits en 1995 par Kennedy et Eberhart comme méthode d'optimisation stochastique, pour des fonctions non linéaires, basée sur la reproduction d'un comportement social. Elle est utilisée pour explorer l'espace de recherche d'un problème quelconque pour trouver l'ensemble des paramètres qui maximise/ minimise un objectif particulier [88,89, 90]. Cet algorithme est inspiré des essaims d'insectes (ou des bancs de poissons ou des nuées d'oiseaux) et de leurs mouvements coordonnés. En effet, tout comme ces animaux se déplacent en groupe pour trouver de la nourriture ou éviter les prédateurs, les algorithmes à essaim de particules recherchent des solutions pour un problème d'optimisation.

L'optimisation par essaim des particules est une technique évolutionnaire qui utilise « une population » de solutions candidates pour développer une solution optimale au problème d'optimisation. Le degré d'optimalité est mesuré par une fonction fitness définie par l'utilisateur. Il s'agit d'une méthode fondée sur la notion de coopérations des agents, appelés ici « particules »[91, 92].

### 2.5.2.2. Les éléments de la PSO

Pour appliquer la PSO il faut définir un espace de recherche constitué de particules et une fonction objective à optimiser. Le principe de l'algorithme est de déplacer ces particules afin qu'elles trouvent l'optimum.

Chacune de ces particules est dotée :

- D'une position, c'est-à-dire ses coordonnées dans l'ensemble de définition.

- D'une vitesse qui permet à la particule de se déplacer. De cette façon, au cours des itérations, chaque particule change de position. Elle évolue en fonction de son meilleur voisin, de sa meilleure position, et de sa position précédente. C'est cette évolution qui permet de tomber sur une particule optimale.
- D'un voisinage, c'est-à-dire un ensemble de particules qui interagissent directement sur la particule, en particulier celle qui a le meilleur critère.

À tout instant, chaque particule connaît :

- Sa meilleure position visitée. On retient essentiellement la valeur du critère calculée ainsi que ses coordonnées.
- La position du meilleur voisin de l'essaim qui correspond à l'ordonnancement optimal.
- La valeur qu'elle donne à la fonction objectif, car à chaque itération il faut une comparaison entre la valeur du critère donnée par la particule courante et la valeur optimale.

### **2.5.2.3. Notion de voisinage**

Le voisinage d'une particule est le sous-ensemble de particules de l'essaim avec lequel il a une communication directe. Ce réseau de rapports entre toutes les particules est connu comme la sociométrie, ou la topologie de l'essaim [93,94]. Il existe deux principaux types de voisinage :

- Les voisinages géographiques : les voisins sont considérés comme les particules les plus proches. Cependant, à chaque itération, les nouveaux voisins doivent être recalculés à partir d'une distance prédéfinie dans l'espace de recherche. C'est donc un voisinage dynamique.

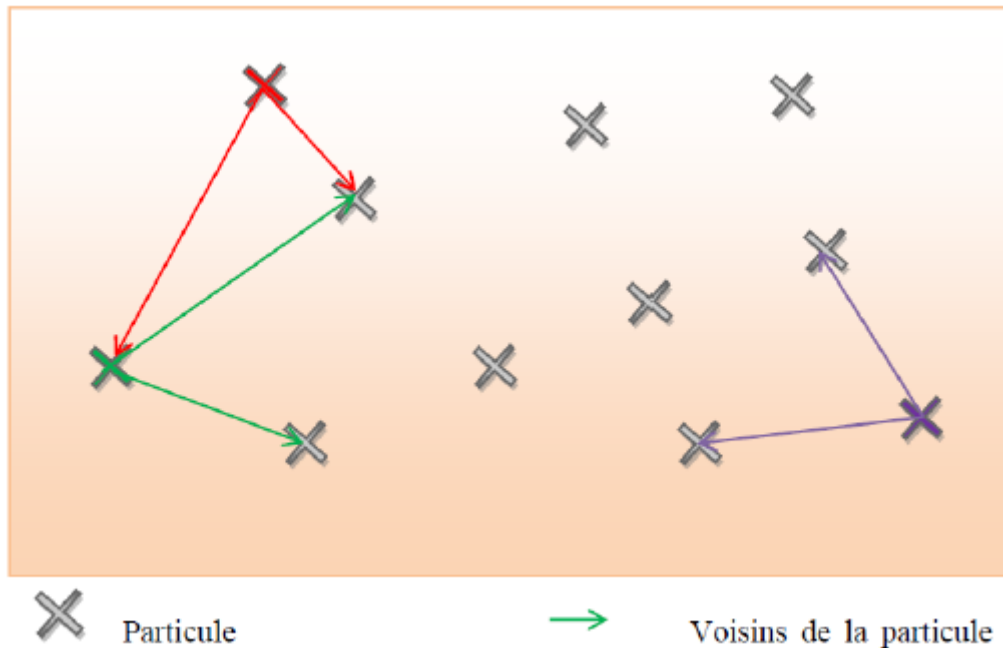


Figure 2.14 : Un voisinage géographique

Dans cet exemple, le voisinage de la particule est composé des deux particules les plus proches.

- Les voisinages sociaux : les voisins sont définis à l'initialisation et ne sont pas modifiés ensuite. C'est le voisinage le plus utilisé, pour plusieurs raisons :
  - Il est plus simple à programmer.
  - Il est moins coûteux en temps de calcul.

En cas de convergence, un voisinage social tend à devenir un voisinage géographique. Pour ce faire, on dispose (virtuellement) les particules en cercle (voir figure 2.15). Puis, pour la particule étudiée, on inclut progressivement dans ses informatrices, d'abord elle-même, puis les plus proches à sa droite et à sa gauche, jusqu'à atteindre la taille voulue. On peut aussi choisir les informatrices au hasard.

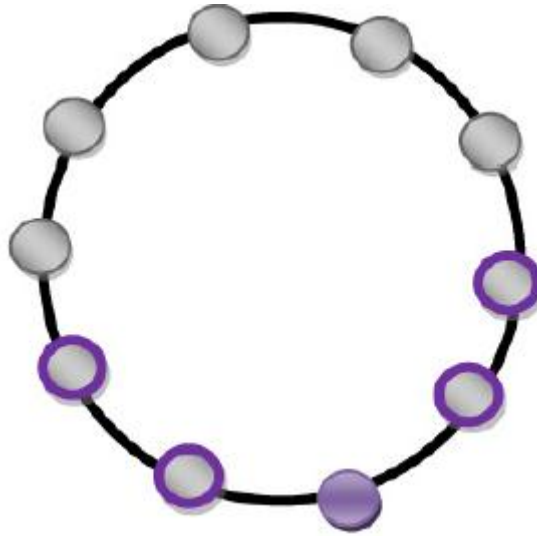


Figure 2.15: Un voisinage en cercle

Dans cet exemple, la particule principale est en bas et ses informatrices correspondent aux deux particules directement à sa droite et à sa gauche.

#### 2.5.2.4. Principe fondamental

L'algorithme de base de la PSO travaille sur une population appelée essaim de solutions possibles, elle-même appelée particules. Ces particules sont placées aléatoirement dans l'espace de recherche de la fonction objectif. À chaque itération, les particules se déplacent en prenant en compte leur meilleure position (déplacement égoïste), mais aussi la meilleure position de son voisinage (déplacement panurgien) [95-96]. Dans les faits, on calcule la nouvelle vitesse à partir de la formule suivante :

$$V_{k+1} = c_1 * V_k + c_2 * (\text{best}_{\text{particule}} - \text{position}_{\text{particule}}) + c_3 * (\text{best}_{\text{voisin}} - \text{position}_{\text{particule}}) \quad (2.17)$$

Où :  $V_{k+1}$  et  $V_k$  sont les vitesses de la particule aux itérations  $k$  et  $k+1$ .

$\text{Best}_{\text{particule}}$  est la meilleure position de la particule

$\text{Best}_{\text{voisin}}$  est la meilleure position de son voisinage à l'itération  $k$

$\text{Position}_{\text{particule}}$  est la position de la particule à l'itération  $k$

$c_1$ ,  $c_2$ ,  $c_3$  sont des coefficients fixés,  $c_2$  est généré aléatoirement à chaque itération et, en général,  $c_3 = c_2$

On peut ensuite déterminer la position suivante de la particule grâce à la vitesse que l'on vient de calculer :

$$X_{k+1} = X_k + V_{k+1} \quad (2.18)$$

Où :  $X_k$  est la position de la particule à l'itération  $k$

On génère  $X_0$  et  $V_0$  au début de notre algorithme.

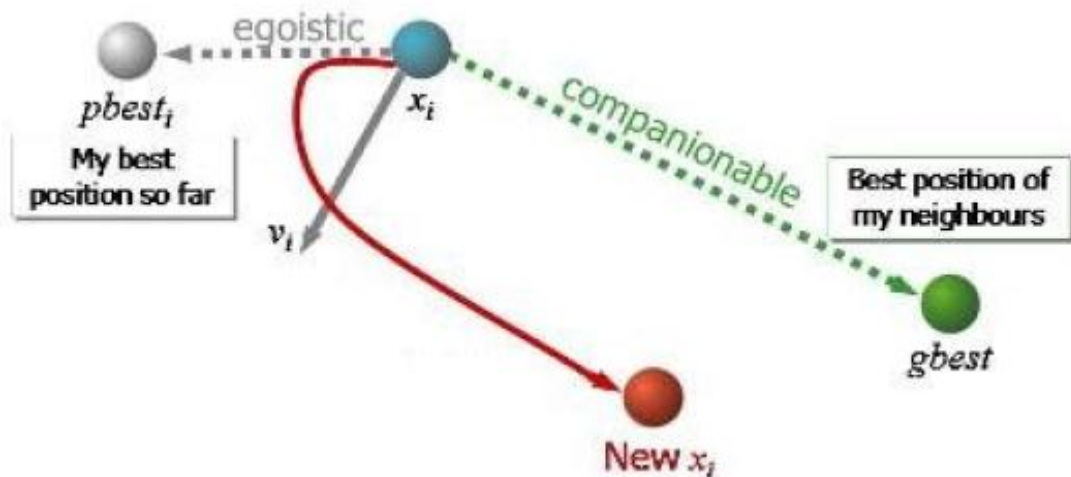


Figure 2.16 : Compromis entre deux déplacements possibles

Le principe de l'algorithme peut être plus facilement visualisé grâce à la figure suivante :

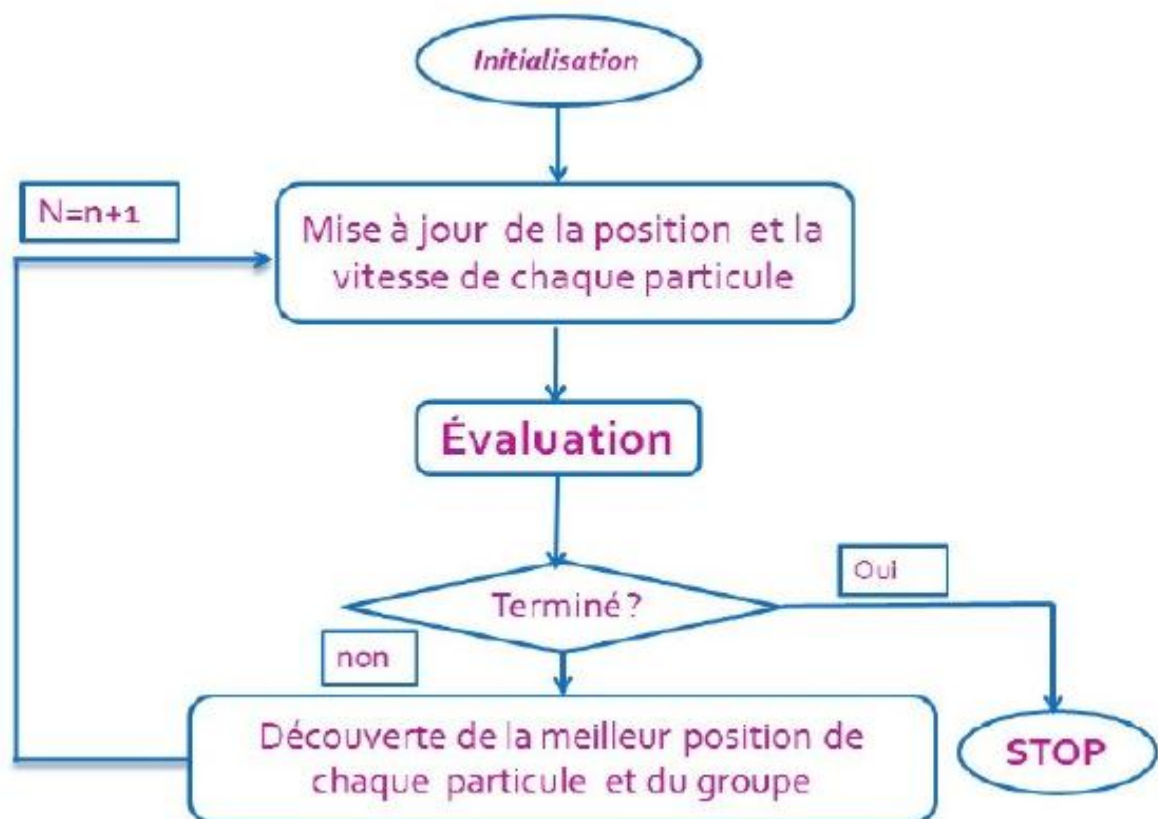


Figure 2.17 : Organigramme de la méthode des essaims particuliers

## 2.6. Les indicateurs de performance

Dans le cadre de la prévision horaire de la vitesse du vent, il existe de nombreux paramètres d'erreur permettant de comparer les approches de prévision [97, 98, 99,100]. Les outils que nous avons utilisés durant cette thèse sont décrits ci-après :

- Le coefficient de détermination

Le coefficient de détermination ( $R^2$ ) est un indicateur qui permet de juger la qualité d'une régression linéaire, simple ou multiple. D'une valeur comprise entre 0 et 1, il mesure l'adéquation entre le modèle et les données observées. Dans le cadre d'une régression linéaire simple, c'est le carré du coefficient de corrélation. Le  $R^2$  se définit comme la part de variance expliquée par rapport à la variance totale, autrement dit par  $[1 - (\text{somme des carrés des résidus} / \text{variance totale})]$ . Ce coefficient s'applique aussi bien à une régression simple qu'à une régression multiple.

$$R^2 = 1 - \frac{\sum_{i=1}^N (T_i - Y_i)^2}{\sum_{i=1}^N (T_i - \bar{T})^2} \quad (2.19)$$

avec  $T_i$  la sortie ,  $\bar{T}$  la moyenne,  $Y_i$  la sortie calculée

- L'erreur quadratique moyenne

L'erreur quadratique moyenne ( en anglais Mean Square Error ; MSE) est très utile pour comparer plusieurs estimateurs, il mesure la moyenne des carrés des erreurs c'est-à-dire la différence quadratique moyenne entre les valeur calculées et les valeurs réelles. On peut effectivement exprimer l'erreur quadratique moyenne par:

$$MSE = \frac{1}{N} \sum_{i=1}^N (T_i - Y_i)^2 \quad (2.20)$$

L'erreur quadratique moyenne est une mesure de la qualité d'un estimateur et les valeurs plus proches de zéro sont meilleures. Pour manipuler une erreur basée sur la même échelle, on peut utiliser le *RMSE* (Root Mean Square Error) défini par:

$$RMSE = \sqrt{MSE} \quad (2.21)$$

- L'erreur moyenne absolue

L'erreur moyenne absolue (Mean Absolute Error ; *MAE*), est une quantité souvent utilisée pour mesurer l'écart des prédictions et des mesures. Sa formule est donnée par :

$$MAE = \frac{1}{N} \sum_{i=1}^N |T_i - Y_i| \quad (2.22)$$

Tout comme dans le cas du *RMSE*, il est possible d'utiliser une forme relative de cette erreur, en divisant par la valeur moyenne . Le paramètre *MAPE* (Mean

Absolute Percentage Error) que l'on a choisi d'utiliser dans cette étude est décrit par :

$$MAPE = \left( \frac{1}{N} \sum_{i=1}^N \left| \frac{T_i - Y_i}{T_i} \right| \right) * 100\% \quad (2.23)$$

## 2.7 Conclusion

Comme nous l'avons vu, la vitesse du vent peut être perçue comme une succession de mesures qui changent aléatoirement au cours du temps pour lesquelles plusieurs outils d'analyse sont potentiellement dédiés. Tout d'abord, nous avons présenté une distinction entre la prédiction et la prévision ainsi que les horizons de prévision de la vitesse du vent. Ensuite, nous avons détaillé les différentes méthodes utilisées pour développer un modèle de prévision. Ce qui a été présenté ici, n'est bien entendu, pas une liste exhaustive de tout ce qui peut exister, ou peut être étudié comme méthodes de prévision mais nous nous sommes attachés à ne détailler que ce qui a été concrètement utilisé lors de nos différentes simulations. Finalement, nous avons défini les différents indicateurs de performance utilisés dans ce travail de thèse.

## Chapitre 3 : Pr vision horaire de la vitesse du vent au Maroc

---

### 3.1 Introduction :

Dans ce chapitre, nous pr sentons les travaux r alis s tout au long de cette th se. Ceux-ci incluent la pr vision horaire de la vitesse du vent en utilisant plusieurs m thodes statistiques lin aires et non-lin aires. Les analyses, ainsi que l'impl mentation des mod les, ont  t  r alis es   partir des logiciels suivants : STATGRAPHIS et MATLAB.

### 3.2. Donn es exp rimentales et r gions d' tude

Il est clair que pour d terminer les potentialit s  oliennes d'un site, il faut passer par le traitement statistique des mesures de la vitesse du vent sachant que le vent est l'un des ph nom nes m t orologiques, les plus difficiles   pr dire. Nous pr senterons tout d'abord les deux sites  tudi s. Ensuite, nous nous int resserons aux param tres qui influencent la vitesse du vent. Enfin, nous faisons une normalisation pour l'utilisation comme entr e pour les mod les de pr vision.

#### 3.2.1. Source des donn es

Les donn es m t orologiques utilis es dans cette  tude sont des donn es recueillies par IRESEN (Institut de Recherche d' nergie Solaire et d' nergie Nouvelle). Ces donn es sont enregistr es chaque jour par intervalle d'une heure   une hauteur de 10 m au-dessus du sol et sont obtenues   travers la base de donn es m t orologique d'IRESEN, comporte pour chaque enregistrement : le nom de site, la date et l'heure d'enregistrement, la temp rature ambiante en  C, la pression atmosph rique en hPa, la direction du vent en degr , l'humidit  relative en % et la vitesse du vent en m/s. Ainsi les donn es recueillies suivant le format du tableau 3.1 doivent subir un pr traitement afin d' tre le plus commodes et le plus fiables   l'exploitation.

Tableau 3.1 : Présentation d'une fiche de données du site de Tanger

| A1            f_x      Site latitude (positive means North) |                |            |     |      |              |              |             |             |                 |   |
|-------------------------------------------------------------|----------------|------------|-----|------|--------------|--------------|-------------|-------------|-----------------|---|
|                                                             | A              | B          | C   | D    | E            | F            | G           | H           | I               | J |
| 1                                                           | Site latitude  | 35.800     |     |      |              |              |             |             |                 |   |
| 2                                                           | Site longitude | -5.840     |     |      |              |              |             |             |                 |   |
| 3                                                           | Site altitude  | 8          |     |      |              |              |             |             |                 |   |
| 4                                                           | Date beginning | 01/01/2011 |     |      |              |              |             |             |                 |   |
| 5                                                           | Date end       | 31/01/2014 |     |      |              |              |             |             |                 |   |
| 6                                                           |                |            |     |      |              |              |             |             |                 |   |
| 7                                                           |                |            |     |      |              |              |             |             |                 |   |
| 8                                                           |                |            |     |      |              |              |             |             |                 |   |
| 9                                                           |                |            |     |      |              |              |             |             |                 |   |
| 10                                                          |                |            |     |      |              |              |             |             |                 |   |
| 11                                                          |                |            |     |      |              |              |             |             |                 |   |
| 12                                                          |                |            |     |      |              |              |             |             |                 |   |
| 13                                                          |                |            |     |      |              |              |             |             |                 |   |
| 14                                                          | Year           | Month      | Day | Hour | Temperature  | Relative Hum | Pressure    | Wind speed  | Wind direction  |   |
| 15                                                          | 2011           |            | 1   | 1    | 1 287.654236 | 76.38947070  | 1008.031953 | 2.906730796 | 251.58987839981 |   |
| 16                                                          | 2011           |            | 1   | 1    | 2 287.650116 | 76.42615692  | 1007.861875 | 2.872621678 | 252.10365295223 |   |
| 17                                                          | 2011           |            | 1   | 1    | 3 287.636414 | 76.46167903  | 1007.616875 | 2.851333620 | 254.38306123655 |   |
| 18                                                          | 2011           |            | 1   | 1    | 4 287.617096 | 76.52877171  | 1007.6575   | 2.834714092 | 258.01694786608 |   |
| 19                                                          | 2011           |            | 1   | 1    | 5 287.614287 | 76.58884173  | 1007.617345 | 2.785047338 | 259.78188348333 |   |

### 3.2.2. Les régions d'étude

- *Région du Nord : site Tanger*

Tanger est une ville du nord du Maroc, située à l'extrémité du nord-ouest du pays sur le détroit de Gibraltar, la ville se trouve à 14 kilomètres de la côte espagnole et a l'avantage unique de s'ouvrir à la fois sur l'atlantique et la méditerranée. Le climat de Tanger est Méditerranéen martiné d'influence océanique et de souffle chergui. Les saisons sont bien différenciées. À noter qu'il peut se produire des intempéries quelquefois violentes avec des vents puissants et des pluies déferlantes.

- *Région du Sud: site Tarfaya*

Tarfaya est une ville côtière marocaine. Elle est située au bord de l'océan atlantique, au niveau du cap Juby, à environ 890 km au Sud-Ouest de la capitale rabat, 100 km au nord-est de Laàyoune et 100 km au Sud-Est de l'île de Lanzarote (îles Canaries). Le tableau 3.2 présente les coordonnées géographiques des deux sites.

Tableau 3.2 : Situation géographique des sites

| Région  | Latitude | Longitude | Altitude |
|---------|----------|-----------|----------|
| Tanger  | 35.800   | -5.840    | 8        |
| Tarfaya | 27.900   | -12.900   | 16       |

### 3.2.3. Prétraitement des données

Nous avons utilisé le logiciel STATGRAPHICS pour effectuer le prétraitement des données météorologiques. La modélisation a été réalisée par l'outil Régression Linéaire Multiple (RLM) que nous avons utilisée dans l'environnement STATGRAPHICS (nous rappelons que logiciel STATGRAPHICS est un logiciel de modélisation qui fait le traitement et l'analyse des données collectées). Pour être conformes aux spécificités de STATGRAPHICS, les données ont été prétraitées dans EXCEL. Ce prétraitement est d'autant plus nécessaire que, d'une part, certaines données se présentent dans le logiciel sous forme de chaîne de caractères alors qu'il nous faut des données numériques, et d'autre part, il existe des informations se présentant sous des formats non utilisables par STATGRAPHICS. Les données prétraitées sont la température, la pression, l'humidité, la direction et la vitesse du vent. Afin d'adapter nos données à la réalisation des modèles de prévision, nous avons procédé à la normalisation de leurs valeurs dans un intervalle [0 ,1] par la relation (3.1).

$$X_n = \frac{(X_{max}-X)}{(X_{max}-X_{min})} \quad (3.1)$$

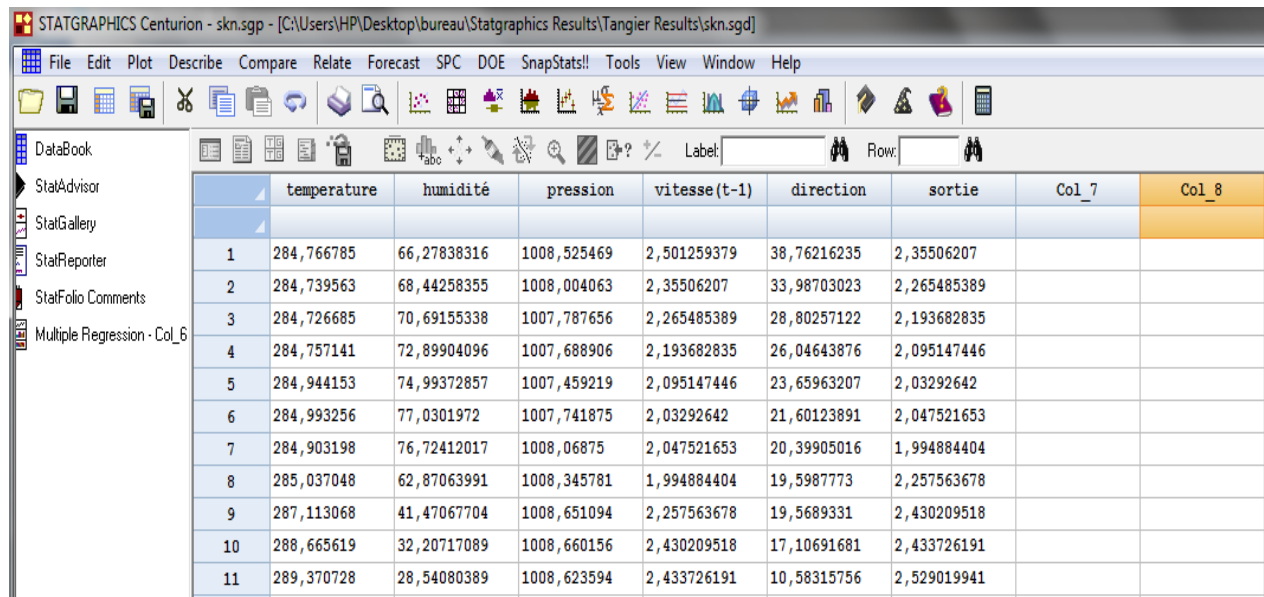
Où :  $X_{max}$  et  $X_{min}$  représentent respectivement les valeurs maximale et minimale des données représentées sous forme d'un vecteur X. Les données normalisées sont utilisées pour réaliser des modèles de prévision suivant des approches que nous allons présenter dans la suite de ce travail.

### 3.3. Détermination des paramètres d'entrée

La première tâche cruciale pour la conception d'un modèle de prévision de la vitesse du vent est de déterminer les entrées du système. Pour ce faire, nous allons voir la pertinence de quelques éléments dont nous jugeons d'influence réelle sur la valeur de la sortie en analysant avec le logiciel de modélisation « statgraphics ». Un élément très important aussi dans cette analyse reste la compacité des entrées avec la conservation de la même quantité d'informations ainsi que l'indépendance de ces éléments d'entrées. En effet, un mauvais choix de certains paramètres peut influencer nos résultats concernant le temps de simulation qui doit être minimal et la fiabilité du modèle sachant que le modèle sera plus performant et fiable s'il contient moins de variable d'entrée.

Dans ce travail, on dispose d'une base de données de trois ans de granularité horaire allant de 2011 à 2013 pour deux sites différents Tanger et Tarfaya au Nord et Sud du Maroc respectivement. Les paramètres à analyser sont : la

température, l'humidité, la pression, la direction et la vitesse(t-1). Un extrait des données classées dans le logiciel STATGRAPHICS est présenté dans la figure ci-dessous.



|    | temperature | humidité    | pression    | vitesse(t-1) | direction   | sortie      | Col_7 | Col_8 |
|----|-------------|-------------|-------------|--------------|-------------|-------------|-------|-------|
| 1  | 284,766785  | 66,27838316 | 1008,525469 | 2,501259379  | 38,76216235 | 2,35506207  |       |       |
| 2  | 284,739563  | 68,44258355 | 1008,004063 | 2,35506207   | 33,98703023 | 2,265485389 |       |       |
| 3  | 284,726685  | 70,69155338 | 1007,787656 | 2,265485389  | 28,80257122 | 2,193682835 |       |       |
| 4  | 284,757141  | 72,89904096 | 1007,688906 | 2,193682835  | 26,04643876 | 2,095147446 |       |       |
| 5  | 284,944153  | 74,99372857 | 1007,459219 | 2,095147446  | 23,65963207 | 2,03292642  |       |       |
| 6  | 284,993256  | 77,0301972  | 1007,741875 | 2,03292642   | 21,60123891 | 2,047521653 |       |       |
| 7  | 284,903198  | 76,72412017 | 1008,06875  | 2,047521653  | 20,39905016 | 1,994884404 |       |       |
| 8  | 285,037048  | 62,87063991 | 1008,345781 | 1,994884404  | 19,5987773  | 2,257563678 |       |       |
| 9  | 287,113068  | 41,47067704 | 1008,651094 | 2,257563678  | 19,5689331  | 2,430209518 |       |       |
| 10 | 288,665619  | 32,20717089 | 1008,660156 | 2,430209518  | 17,10691681 | 2,433726191 |       |       |
| 11 | 289,370728  | 28,54080389 | 1008,623594 | 2,433726191  | 10,58315756 | 2,529019941 |       |       |

Figure 3.1 : extrait de la base de données dans statgraphics

### 3.3.1. Étude statistique des données

Nous allons utiliser la méthode des moindres carrées ordinaires notée en anglais (OLS) et l'analyse de la variance (ANOVA) pour enlever la colinéarité entre les variables et les procédures de sélection de variables que nous présentons dans ce travail répondent aux questions suivantes : peut-on simplement éliminer les variables qui ne sont pas significatives ? Doit-on les éliminer une à une, comment doit-on organiser la suppression? Souvent, certaines variables indépendantes sont redondantes, elles emmènent le même type d'information : c'est le problème de la colinéarité, elles se gênent mutuellement dans la régression.

Dans un premier temps, pour mieux comprendre la relation entre ces paramètres, il est nécessaire de déterminer l'interaction entre eux pour supprimer la colinéarité entre les paramètres en calculant le coefficient de corrélation entre toutes les variables.

D'après les tableaux 3.3 et 3.4 qui représentent les coefficients de corrélation entre toutes les variables citées précédemment, nous constatons une forte corrélation significative entre les variables suivantes : température et pression

avec un coefficient de corrélation de -0,9968 pour Tanger, la température et la direction avec un coefficient de corrélation de -0,9974 pour Tarfaya. Une autre corrélation moins forte que la première apparaissent aussi entre les variables comme la corrélation entre la pression et l'humidité avec un coefficient de corrélation de - 0,6160 pour Tanger et entre la pression et la direction avec un coefficient de corrélation de 0,7109 pour Tarfaya. Ainsi deux groupes de variables corrélées se dégagent : {température, pression, direction} pour le site de Tanger et {pression, humidité, direction} pour le site de Tarfaya.

**Remarque :** un coefficient de corrélation négatif, mais proche de 1 en valeur absolue ne signifie pas qu'il n'existe pas de liaison entre les deux variables. Cela indique au contraire qu'elles sont liées et qu'elles ont tendance à évoluer en sens contraire.

Tableau 3.3 - Présentation des coefficients de corrélation intervariables du Tanger

|                       | Température | Humidité | Pression        | Vitesse du vent (t-1) | Direction |
|-----------------------|-------------|----------|-----------------|-----------------------|-----------|
| Température           | 1.0000      | 0.4699   | - <b>0.9968</b> | 0.0340                | 0.0646    |
| Humidité              |             | 1.0000   | - <b>0.6160</b> | - 0.0452              | 0.0705    |
| Pression              |             |          | 1.0000          | - 0.0709              | - 0.1042  |
| Vitesse du vent (t-1) |             |          |                 | 1.0000                | 0.1480    |
| Direction             |             |          |                 |                       | 1.0000    |

Tableau3.4- Présentation des coefficients de corrélation intervariables du Tarfaya

|                       | Température | Humidité | Pression | Vitesse du vent (t-1) | Direction       |
|-----------------------|-------------|----------|----------|-----------------------|-----------------|
| Température           | 1.0000      | 0.4274   | -0.0408  | 0.0035                | - <b>0.9974</b> |
| Humidité              |             | 1.0000   | - 0.4734 | - 0.0920              | 0.0793          |
| Pression              |             |          | 1.0000   | - 0.0447              | <b>0.7109</b>   |
| Vitesse du vent (t-1) |             |          |          | 1.0000                | 0.2910          |
| Direction             |             |          |          |                       | 1.0000          |

Dans la deuxième étape, nous éliminons la multicollinéarité entre les paramètres en utilisant la méthode des moindres carrés ordinaires. Pour cela, si la valeur de P est inférieure à 0,05, les paramètres doivent être acceptés tandis que dans le cas contraire (c'est-à-dire que la valeur de P est supérieure à 0,05), les

paramètres doivent être rejetés. D'après la valeur de P indiquée dans les tableaux 3.5 et 3.6, nous remarquons que la plus grande valeur de P sur les variables indépendantes est de 0,3999, appartenant à la pression pour le site de Tanger et 0.8620 appartenant à la direction pour le site de Tarfaya. Puisque la valeur de P est supérieure à 0,05, ce terme est non statistiquement significatif au niveau de confiance de 95,0%. Par conséquent, nous devons envisager de supprimer la pression du modèle de Tanger et la direction du modèle de Tarfaya.

Tableau 3.5 : Analyse statistique de la RLM à Tanger

| Parameter       | code           | Coefficient value | Standard error | t-value  | P-value       |
|-----------------|----------------|-------------------|----------------|----------|---------------|
| Direction       | X <sub>1</sub> | -0.0000987504     | 0.0000144857   | -6.8171  | 0.0000        |
| Humidity        | X <sub>2</sub> | -0.00127188       | 0.0000830857   | -15.308  | 0.0000        |
| Temperature     | X <sub>3</sub> | 0.000407384       | 0.0000214706   | 18.9741  | 0.0000        |
| Past wind speed | X <sub>4</sub> | 0.990865          | 0.00101643     | 974.848  | 0.0000        |
| Pressure        | X <sub>5</sub> | -0.000060839      | 0.0000722721   | -0.84181 | <b>0.3999</b> |

Tableau 3.6 : Analyse statistique de la RLM à Tarfaya

| Parameter       | Code           | Coefficient value | Standard error | t-value  | P-value       |
|-----------------|----------------|-------------------|----------------|----------|---------------|
| Humidity        | X <sub>1</sub> | -0.00534814       | 0.0001524      | -35.0928 | 0.000         |
| Past wind speed | X <sub>2</sub> | 0.995477          | 0.0010076      | 987.964  | 0.000         |
| Pressure        | X <sub>3</sub> | -0.00240208       | 0.000137537    | -17.465  | 0.000         |
| Temperature     | X <sub>4</sub> | 0.00940481        | 0.000458722    | 20.5022  | 0.000         |
| Direction       | X <sub>5</sub> | 0.00000280115     | 0.0000161195   | 0.173775 | <b>0.8620</b> |

### 3.3.2. L'influence des paramètres choisis sur la sortie

En comparant les coefficients de corrélation entre les quatre variables indépendantes et la variable dépendante montrée dans les tableaux 3.7 et 3.8, on remarque que le coefficient de corrélation a des valeurs positives et négatives. Un coefficient de corrélation est positif signifie que si l'un des paramètres augmente, le taux de la vitesse du vent augmente aussi. De plus, si la corrélation est négative cela montre une relation inverse entre eux.

La direction et la vitesse du vent à l'instant (t-1) ont la moindre et la plus grande corrélation avec la vitesse du vent à Tanger tandis que la température et la vitesse du vent à (t-1) ont la moindre et la plus grande corrélation à Tarfaya, ce

qui signifie que les paramètres n'ont pas la même influence sur la vitesse du vent.

Tableau 3.7: Coefficient de corrélation entre les variables indépendantes et la variable dépendante pour Tanger

| City    | Dependent variable | Direction | Humidity | Temperature | Past wind speed |
|---------|--------------------|-----------|----------|-------------|-----------------|
| Tangier | Wind speed         | -0.1954   | -0.0201  | -0.0608     | 0.9906          |

Tableau 3.8: Coefficient de corrélation entre les variables indépendantes et la variable dépendante pour Tarfaya

| City    | Dependent variable | Pressure | Humidity | Temperature | Past wind speed |
|---------|--------------------|----------|----------|-------------|-----------------|
| Tarfaya | Wind speed         | 0.1246   | 0.1020   | -0.1852     | 0.9874          |

#### Conclusion :

Après une étude faite sur les données météorologiques, nous concluons que les paramètres qui influencent la vitesse du vent pour le site du nord sont : Direction, humidité, température et vitesse du vent à (t-1) ainsi que les paramètres qui influencent notre sortie pour le site du sud sont : Humidité, pression, température et la vitesse du vent à (t-1). Après avoir généré ces variables indépendantes sans aucune colinéarité, un modèle de régression linéaire multiple sera développé dans le travail suivant.

### **3.4. Approche 1 : Modélisation de la vitesse du vent par Régression linéaire multiple**

#### **3.4.1. Analyse des résultats**

On cherche à expliquer la vitesse du vent à partir de quatre variables explicatives décrites précédemment pour chaque région étudiée sachant que pour utiliser avec confiance un modèle, il est important de vérifier sa validité en se basant sur les critères suivants :

- Qualité de l'ajustement (erreur minimale)
- Coefficient de détermination  $R^2$  élevé
- Analyse des résidus

Une fois les paramètres d'entrée sont choisis, on a développé le modèle de régression linéaire pour la prévision de la vitesse du vent. D'une part, on remarque que ce modèle a une valeur élevée du coefficient de détermination  $R^2$  qui est égale à 0,9964 et 0,9981 à Tanger et Tarfaya respectivement (voir tableau 3.9). Les valeurs obtenues indiquent qu'il y a un bon accord entre les valeurs expérimentales et prédites obtenues à partir du modèle MLR, comme illustré dans les figures 3.2 et 3.3, d'autre part, l'erreur du modèle est 0.0388 et 0.1009 à Tanger et Tarfaya respectivement. C'est une erreur d'un modèle linéaire qui peut être s'améliorer en utilisant un modèle non linéaire.

Tableau 3.9 : Performance de la régression linéaire multiple à Tanger et Tarfaya

| Paramètres statistiques | Tanger | Tarfaya |
|-------------------------|--------|---------|
| $R^2$                   | 0.9816 | 0.9726  |
| MSE                     | 0.0388 | 0.1009  |
| RMSE                    | 0.1969 | 0.3177  |
| MAE                     | 0.1509 | 0.2410  |
| MAPE                    | 7.3020 | 5.9648  |

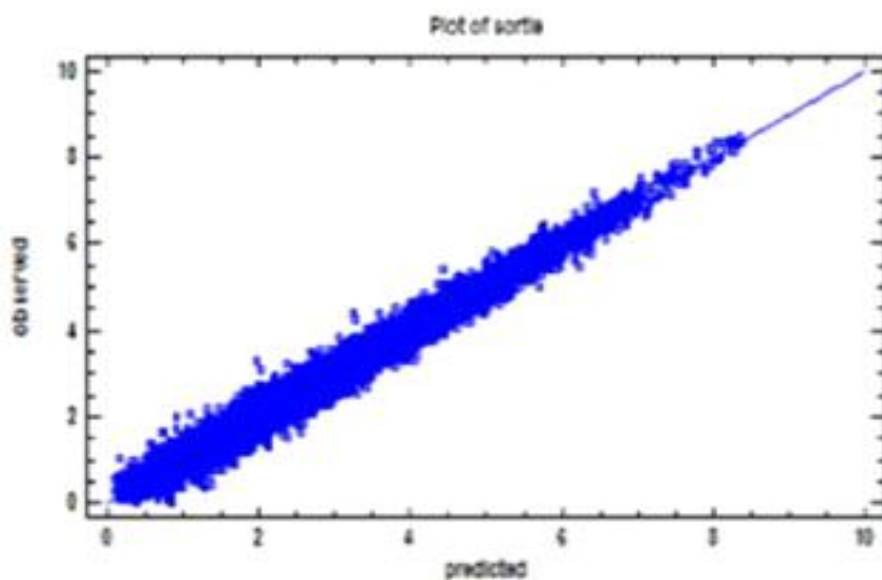


Figure 3.2 : graphe de régression de la vitesse de vent pour Tanger

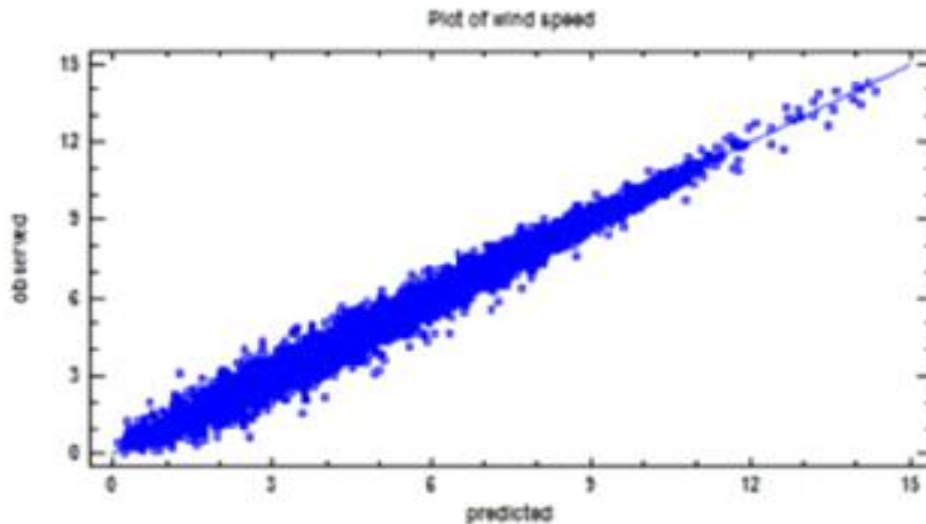


Figure 3.3 : graphe de régression de la vitesse du vent pour Tarfaya

### 3.4.2. Performance du modèle

Finalement, les modèles linéaires que nous avons développé le plus performants et qui représentent convenablement notre système sont cités dans le tableau ci-dessous :

| Site    | Modèle                                                                            |
|---------|-----------------------------------------------------------------------------------|
| Tanger  | $y = -0.0000987504 * X_1 - 0.00127188 * X_2 + 0.000407384 * X_3 + 0.990865 * X_4$ |
| Tarfaya | $y = -0.00534814 * X_1 + 0.995477 * X_2 - 0.00240208 * X_3 + 0.00940481 * X_4$    |

Sachant que :

Pour le site de Tanger  $\left\{ \begin{array}{l} X_1: \text{direction} \\ X_2: \text{humidité} \\ X_3: \text{température} \\ X_4: \text{vitesse du vent}(t-1) \\ Y: \text{sortie} \end{array} \right.$

Pour le site de Tarfaya  $\left\{ \begin{array}{l} X_1: \text{humidité} \\ X_2: \text{vitesse du vent}(t-1) \\ X_3: \text{pression} \\ X_4: \text{température} \\ Y: \text{sortie} \end{array} \right.$

Après avoir estimé les modèles de régression linéaire multiple pour les deux sites étudiés, nous concluons que la performance de ces modèles n'est pas de grande précision car il faut toujours rappeler qu'une très bonne corrélation ne

veut pas dire que le modèle de bons résultats. La combinaison du coefficient de détermination et les métriques de performance donne une idée globale et plus approfondie du modèle prédictif. Par conséquent, le travail suivant sera consacré sur le développement d'un modèle de prévision non linéaire en se basant sur les réseaux de neurones artificiels pour améliorer la précision.

### 3.5. Approche 2 : Prévion de la vitesse du vent par Réseaux de neurones artificiels

#### 3.5.1. Analyse des résultats

Dans ce travail, nous proposons la technique des réseaux de neurones artificiels comme méthode statistique non linéaire car elle est très populaire et se montre une grande efficacité dans le domaine de la prévision et aussi son utilité dans l'élaboration de notre objectif. Il existe un grand nombre de types de réseaux de neurones, dont chacun a des avantages et des inconvénients. Le réseau choisi dans notre cas est un perceptron multicouche avec un algorithme d'apprentissage de rétropropagation. Ce choix est fait pour la facilité et la rapidité de sa construction et encore par le fait que notre problème présente un nombre limite de variables d'entrées.

La base de données est divisée en trois parties, l'une pour l'apprentissage comprend 70% de la base, 15% pour le test et le 15% qui reste pour la validation. A noter que les données utilisées dans ce travail sont les données traitées dans le travail précédent.

Afin d'avoir un modèle parcimonieux, nous avons testé plusieurs structures du RNA en changeant le nombre des couches et le nombre de neurones de chaque couche cachée (voir le tableau 3.10) . Nous avons limité à une couche cachée en utilisant une fonction sigmoïde comme fonction d'activation et dans le but de déterminer le nombre de neurones dans la couche cachée, nous avons varié le nombre de neurones de cette couche plusieurs fois. À partir des résultats obtenus, nous avons constaté qu'une couche cachée avec cinq neurones pour le site de Tanger et une couche cachée avec un seul neurone pour le site de Tarfaya présente de bonnes performances en termes de  $R^2$  élevées et une valeur réduite du MSE.

Tableau 3.10 : indicateurs statistiques du RNA pour les deux régions

| Site   | structure    | $R^2$         | MSE           | RMSE   | MAE    | MAPE          |
|--------|--------------|---------------|---------------|--------|--------|---------------|
| Tanger | 4-1-1        | 0.9981        | 0.0040        | 0.0630 | 0.0215 | 1.9715        |
|        | 4-2-1        | 0.9981        | 0.0039        | 0.0628 | 0.0214 | 1.9805        |
|        | 4-4-1        | 0.9982        | 0.0039        | 0.0625 | 0.0212 | 1.9705        |
|        | <b>4-5-1</b> | <b>0.9982</b> | <b>0.0039</b> | 0.0628 | 0.0213 | <b>1.9644</b> |

|         |              |               |               |        |        |               |
|---------|--------------|---------------|---------------|--------|--------|---------------|
| Tarfaya | <b>4-1-1</b> | <b>0.9957</b> | <b>0.0157</b> | 0.1254 | 0.0522 | <b>1.9027</b> |
|         | 4-2-1        | 0.9956        | 0.0163        | 0.1278 | 0.0571 | 2.0647        |
|         | 4-3-1        | 0.9956        | 0.0163        | 0.1275 | 0.0573 | 2.0692        |
|         | 4-5-1        | 0.9955        | 0.0167        | 0.1294 | 0.0558 | 2.0212        |

Les figures 3.4 et 3.5 illustrent les graphiques des erreurs de formation, on remarque que les résultats sont raisonnables parce que:

- L'erreur tend vers une valeur minimale pour les trois phases (apprentissage, test et validation).
- Aucun surajustement significatif n'est survenu pour la formation du réseau de prévision de la vitesse du vent.
- L'erreur de la phase du test et la phase de la validation ont des caractéristiques similaires.

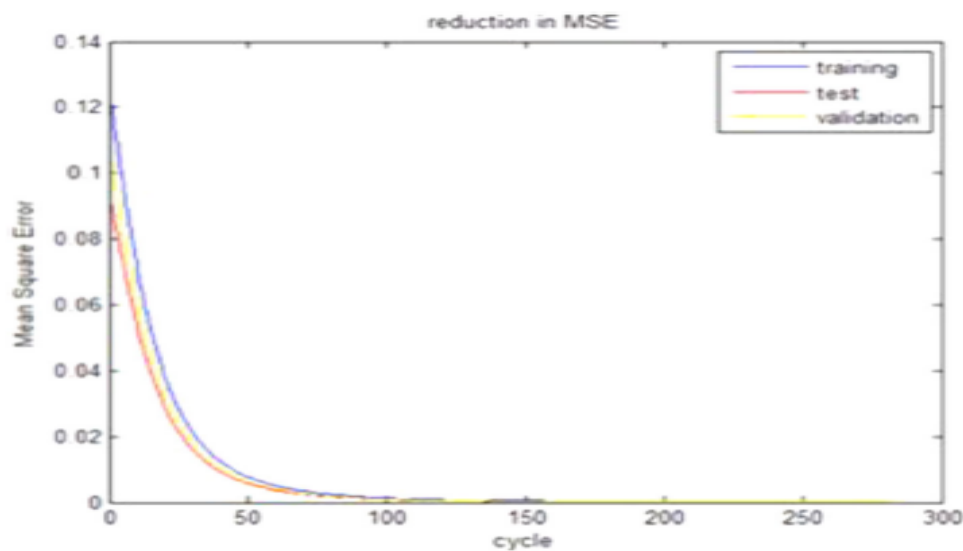


Figure 3.4: Réduction du MSE pour la structure 4-5-1 à Tanger

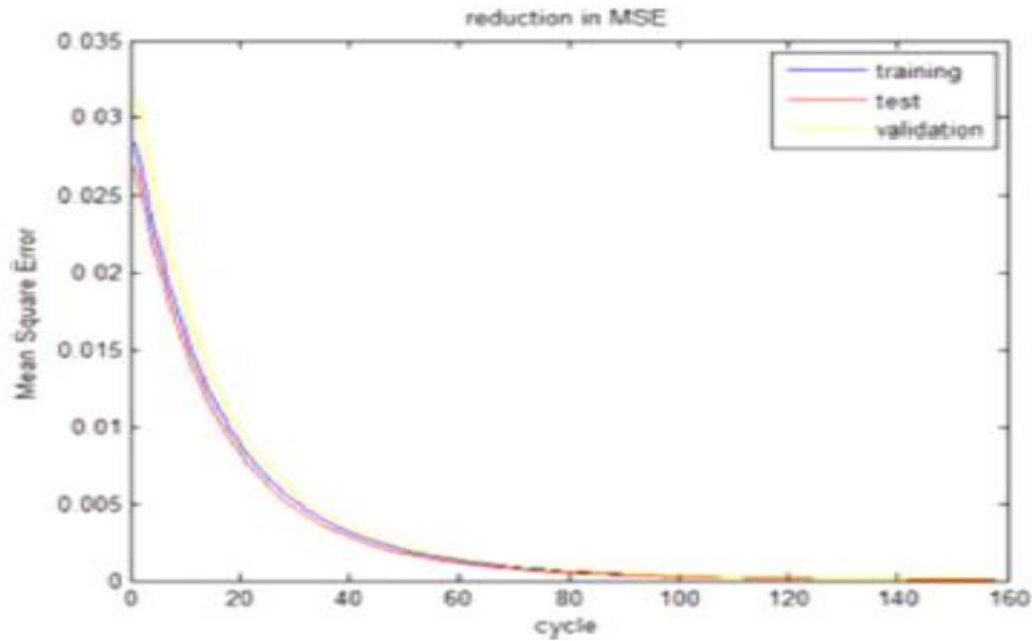


Figure 3.5 : Réduction du MSE pour la structure 4-1-1 à Tarfaya

### 3.5.2. Performance du modèle

La qualité de l'ajustement du modèle de régression est évaluée en tenant compte le coefficient de détermination  $R^2$  qui est proche de 1 et MSE est faible (c'est-à-dire proche de 0). Les résultats de la prévision de la vitesse du vent à l'aide du réseau neuronal et la régression linéaire multiple montrent que RNA est capable de prédire la vitesse du vent de deux villes marocaines dans cette étude avec  $R^2$  et Valeurs MSE égales respectivement à 0,9982 et 0,0039 pour le site du Tanger, 0,9957 et 0,0157 pour le site de Tarfaya. Cependant dans le modèle du RLM, ces paramètres statistiques sont de 0,9816 et 0,0388 pour Tanger, 0,9726 et 0,1009 pour Tarfaya. Les valeurs du coefficient de détermination sont élevées pour les RLM et les RNA alors que les erreurs statistiques MSE, RMSE, MAE et MAPE sont plus faibles pour les RNA par rapport au RLM. Cela démontre que les performances des modèles non linéaires sont meilleures que celle du modèle RLM linéaire. La méthode RLM est choisie uniquement comme représentant du modèle statistique linéaire de cette étude afin de déterminer les variables qui influencent le plus la vitesse du vent. La raison pour laquelle le modèle non linéaire fonctionne mieux que le modèle linéaire. Cela signifie que le modèle peut être attribué une relation non linéaire complexe entre la vitesse du vent et les variables météorologiques. Ce qui implique la non-linéarité du phénomène. Vu que notre but est de minimiser l'erreur et d'augmenter la performance du modèle, on va attaquer dans le travail suivant les algorithmes métaheuristiques.

### 3.5.3. Hybridation RNA-GA et RNA-PSO

L'objectif général dans ce travail est de faire une étude du parc éolien installé à Tanger dans la partie nord du Maroc pour estimer un modèle de prévision plus performant à partir des données météorologiques traitées dans le premier travail. Premièrement, les données sont exploitées pour estimer deux modèles hybrides de prévision du vent en se basant sur les réseaux de neurones artificiels: le premier modèle est Réseau de neurones artificiels avec Génétique Algorithmique (RNA-GA) noté aussi NN-GA, le deuxième modèle est le réseau de neurones artificiels avec l'algorithme d'optimisation par essaim particulaire (RNA-PSO) noté aussi NN-PSO. En faisant une comparaison entre ces deux modèles hybrides pour choisir le modèle qui est capable de sélectionner les meilleurs indicateurs, d'une part. Ainsi il sera apte à fournir des résultats interprétables avec une meilleure performance prédictive, d'autre part.

#### 3.5.3.1. Méthodologie :

La technique d'hybridation adoptée dans notre travail est celle de la méthode des réseaux de neurones artificiels avec deux algorithmes différents GA et PSO en insérant RNA dans GA pour former RNA-GA et PSO dans RNA pour former RNA-PSO. Nous allons discuter la structure de nos méthodes RNA-GA et RNA-PSO comme il est indiqué en dessous.

Premièrement, l'algorithme génétique comme nous avons décrit précédemment est formé par trois parties principales : la sélection, le croisement, et la mutation. Deuxièmement, dans la phase d'apprentissage pour les réseaux de neurones artificiels la génération du poids se fait aléatoirement. Cependant nous avons proposé d'hybrider ces deux algorithmes avec le RNA dans la phase d'apprentissage pour bien choisir les poids synaptiques. Nous avons limité à une couche cachée en utilisant une fonction sigmoïde comme fonction d'activation et dans le but de déterminer le nombre de neurones dans la couche cachée, nous avons varié le nombre de neurones de cette couche plusieurs fois.

#### Algorithme d'hybridation RNA-GA

1. Normaliser les données entre [0,1]
2. Initialiser la population pour GA.
3. Evaluer la valeur fitness de chaque individu (i.e., calculer la sortie par le RNA et ensuite l'erreur quadratique moyenne entre cette sortie et la valeur désirée).
4. Appliquer les opérations de sélection, croisement, mutation et remplacement sur les chromosomes.

5. Si la condition d'arrêt est satisfaite alors on obtient les poids synaptiques optimaux sinon on reprend l'algorithme à partir du point 3.

### **Algorithme d'hybridation RNA-PSO**

1. Normaliser les données entre [0,1]
2. Initialiser l'essaim de particules pour PSO
3. Evaluer la valeur fitness de chaque particule (i.e., calculer la sortie par le RNA et ensuite l'erreur quadratique moyenne entre cette sortie et la valeur désirée)
4. Comparer la valeur fitness calculée avec les deux valeurs pbest et gbest
5. Mettre à jour la vitesse et la position pour chaque particule
6. Si la valeur cible ou le nombre d'itérations est atteint alors on obtient les hyper-paramètres optimaux sinon on reprend l'algorithme à partir du point 3

#### **3.5.3.2. Analyse des résultats :**

Afin d'avoir un modèle performant et fiable, nous avons testé plusieurs structures du RNA-GA et RNA-PSO en changeant le nombre des couches et le nombre de neurones de chaque couche cachée (voir le tableau 3.11) . Nous avons limité à une couche cachée en utilisant une fonction sigmoïde comme fonction d'activation et pour la phase d'apprentissage, on a choisi les deux algorithmes métaheuristiques comme algorithme génétique et optimisation par essaim des particules. Dans le but de déterminer le nombre de neurones dans la couche cachée, nous avons varié le nombre de neurones de cette couche plusieurs fois. À partir des résultats obtenus, nous avons constaté qu'une couche cachée avec trois neurones pour l'algorithme GA et une couche cachée avec deux neurones pour l'algorithme PSO présente de bonnes performances en termes de  $R^2$  élevées et une valeur réduite du MSE.

Tableau 3.11 : Performance des modèles hybrides RNA-GA et RNA-PSO

| Méthode | structure    | MSE                                    | $R^2$         |
|---------|--------------|----------------------------------------|---------------|
| RNA-GA  | 4-1-1        | $4.97 \cdot 10^{-4}$                   | 0.9833        |
|         | 4-2-1        | $4.88 \cdot 10^{-4}$                   | 0.9834        |
|         | <b>4-3-1</b> | <b><math>4.71 \cdot 10^{-4}</math></b> | <b>0.9834</b> |
|         | 4-5-1        | $4.99 \cdot 10^{-4}$                   | 0.9833        |
| RNA-PSO | 4-1-1        | $5.67 \cdot 10^{-3}$                   | 0.9877        |
|         | <b>4-2-1</b> | <b><math>5.27 \cdot 10^{-4}</math></b> | <b>0.9881</b> |
|         | 4-4-1        | $6.01 \cdot 10^{-4}$                   | 0.9787        |
|         | 4-5-1        | $5.99 \cdot 10^{-4}$                   | 0.9880        |

### 3.5.3.3. Performance du modèle :

L'évaluation du pouvoir prédictif des modèles hybrides RNA-GA et RNA-PSO s'effectue par l'étude de l'aire entre les deux courbes pour chaque modèle. En effet, l'étude de l'aire entre les courbes montre que les deux modèles hybrides ayant des valeurs proches par rapport aux valeurs actuelles. Ceci est clairement illustré par les Figures 3.5 et 3.6 où nous observons la courbe du modèle hybride RNA-PSO et RNA-GA, qui s'affiche en rouge, par rapport à la courbe relative à valeurs actuelles, qui est au-dessous.

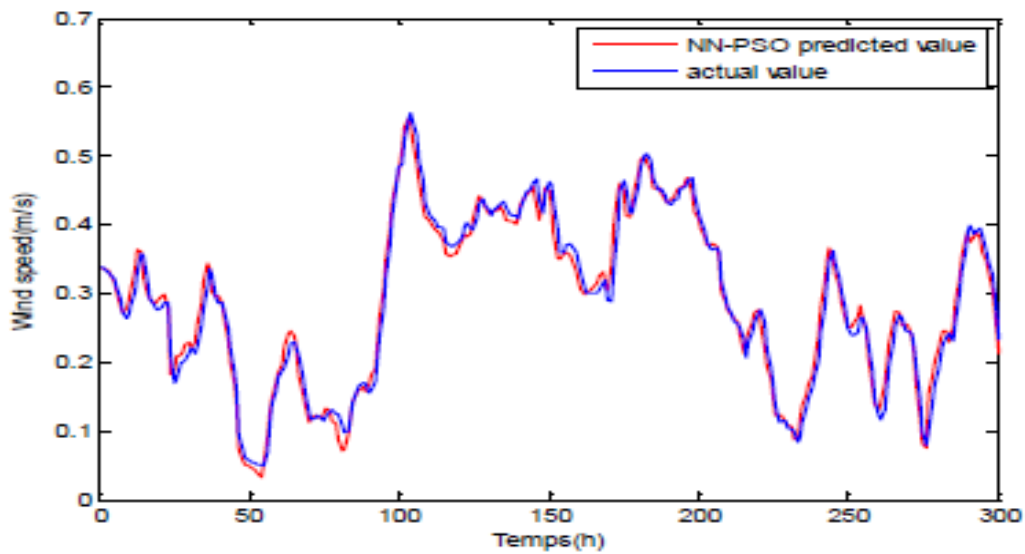


Figure 3.6 : Valeur actuelle/ prédite par RNA-PSO

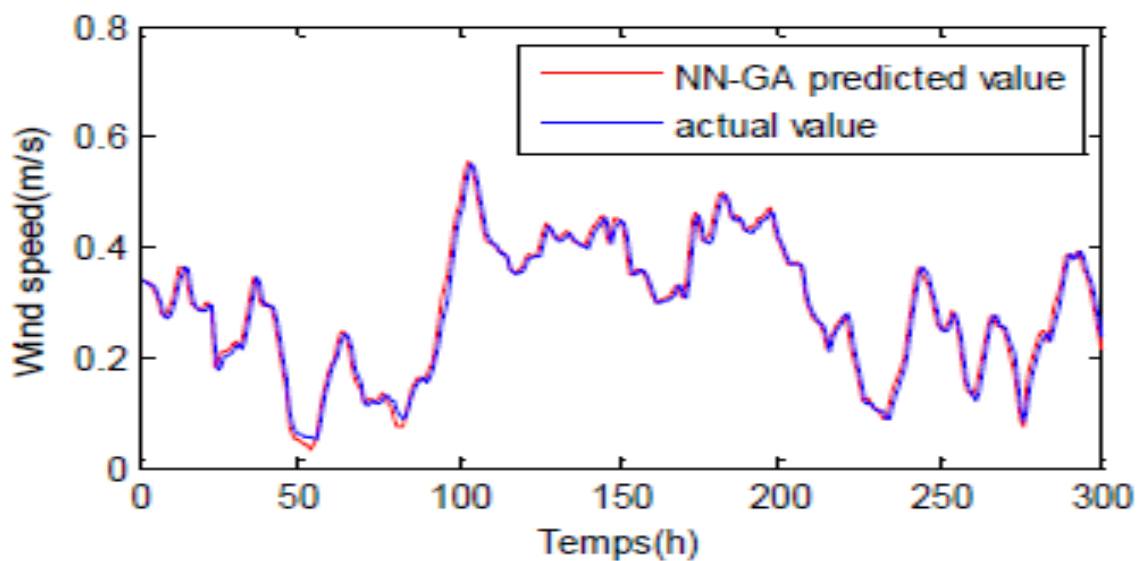


Figure 3.7 : Valeur actuelle/prédite par RNA-GA

### **3.6. Approche 3 : Pr vision de la vitesse du vent par Machine   vecteurs supports**

#### **3.6.1. M thodologie**

Dans ce travail, d'autres mod les hybrides sont d velopp s en se basant sur la m thode de machine   vecteur support qui a montr  son efficacit  de traitement des probl mes de discrimination non lin aire et qu'on va hybrider premi rement avec l'algorithme g n tique not e en anglais (SVM-GA) puis avec l'algorithme d'optimisation par essaim particulaire not e en anglais (SVM-PSO).

Les m mes variables explicatives ind pendantes du site Tanger qui influencent le plus la vitesse du vent sont utilis es pour d velopper ces deux mod les hybrides. L'algorithme g n tique (GA) et l'optimisation des essaims de particules (PSO) sont utilis s pour optimiser les trois param tres du SVM tels que le co t (C), gamma (g), epsilon (e) comme indiqu  ci-dessous :

#### **Algorithme d'hybridation SVM-GA**

1. Normaliser les donn es entre [0,1]
2. Initialiser la population pour GA.
3. Evaluer la valeur fitness de chaque individu (i.e., calculer la sortie par le SVM et ensuite l'erreur quadratique moyenne entre cette sortie et la valeur d sir e).
4. Appliquer les op rations de s lection, croisement, mutation et remplacement sur les chromosomes.
5. Si la condition d'arr t est satisfaite alors on obtient les hyper-param tres optimaux sinon on reprend l'algorithme   partir du point 3.

#### **Algorithme d'hybridation SVM-PSO**

7. Normaliser les donn es entre [0,1]
8. Initialiser l'essaim de particules pour PSO
9. Evaluer la valeur fitness de chaque particule (i.e., calculer la sortie par le SVM et ensuite l'erreur quadratique moyenne entre cette sortie et la valeur d sir e)
10. Comparer la valeur fitness calcul e avec les deux valeurs pbest et gbest
11. Mettre   jour la vitesse et la position pour chaque particule

12. Si la valeur cible ou le nombre d'itérations est atteint alors on obtient les hyper-paramètres optimaux sinon on reprend l'algorithme à partir du point 3

### 3.6.2. Analyse des résultats

Les paramètres du SVM tels que le coût (C), gamma (g), epsilon (e) ont un impact significatif sur les performances de prévision, leurs valeurs sont listées dans les tableaux 3.12 et 3.13. L'algorithme génétique (GA) et l'optimisation des essaims de particules (PSO) sont utilisés pour optimiser les trois paramètres ci-dessous en minimisant la valeur MSE. Pour évaluer les deux approches hybrides proposées et déterminer quantitativement le meilleur modèle, le coefficient de détermination ( $R^2$ ) et trois indices statistiques sont utilisés pour mesurer la précision des prévisions. Ces indices sont l'erreur moyenne (ME), l'erreur quadratique moyenne (MSE) et l'erreur absolue moyenne (MAE) entre les valeurs prévues et réelles de la vitesse du vent.

Tableau 3.12 : les paramètres optimaux de SVM-GA

| Site   | Valeurs des paramètres |           |              |
|--------|------------------------|-----------|--------------|
| Tanger | Cost (C)               | Gamma (g) | Epsilon (e)  |
|        | 15.2749                | 3.144599  | 0.0006444887 |

Tableau 3.13 : les paramètres optimaux de SVM-PSO

| Site   | Valeurs des paramètres |           |              |
|--------|------------------------|-----------|--------------|
| Tanger | Cost (C)               | Gamma (g) | Epsilon (e)  |
|        | 10.8001                | 0.167599  | 0.0003052174 |

Par ailleurs et d'après les résultats affichés par le tableau 3.14, on remarque que les deux modèles SVM-GA et SVM-PSO ayant une valeur très faible du  $R^2$  entre données observées et données prédites ce qui signifie que la qualité d'ajustement du modèle est faible.

Tableau 3.14 : Performance des modèles hybrides SVM-GA et SVM-PSO

| Méthode | MSE                  | ME                   | MAE    | R-squared |
|---------|----------------------|----------------------|--------|-----------|
| SVM-GA  | $2.42 \cdot 10^{-2}$ | $6 \cdot 10^{-3}$    | 0.1238 | 0.1947    |
| SVM-PSO | $2.60 \cdot 10^{-2}$ | $1.31 \cdot 10^{-2}$ | 0.1281 | 0.1355    |

### 3.6.4. Performance du modèle

L'étude de l'aire entre les courbes montre que les modèles SVM-PSO et SVM-GA ayant des valeurs très loin par rapport aux valeurs actuelles. Ceci est clairement illustré par les Figures 3.8 et 3.9 où nous observons la courbe du modèle SVM-PSO et SVM-GA, qui s'affiche en rouge, par rapport à la courbe relative à valeurs actuelles, qui est au-dessous. Cela signifie que la méthode de machine à vecteurs supports ne montre pas une grande capacité prédictive dans le domaine de prévision de la vitesse du vent comme peut le montrer dans le domaine de classification.

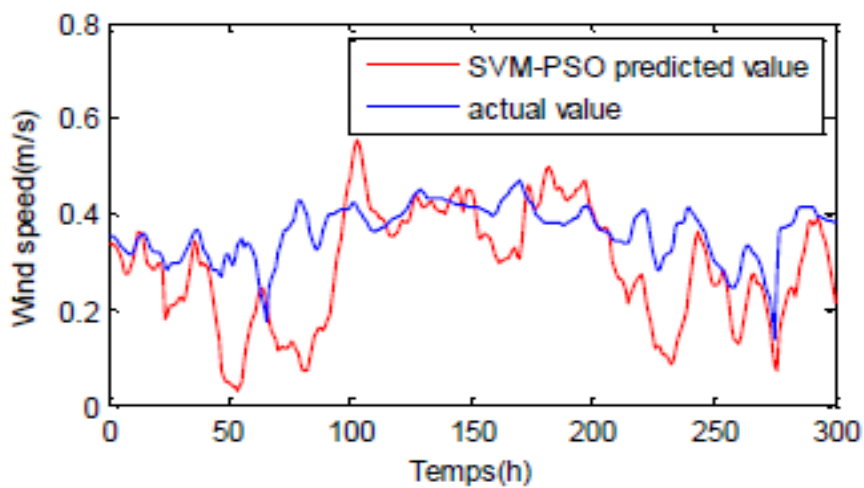


Figure 3.8 : Valeur actuelle/ prédite par SVM-PSO

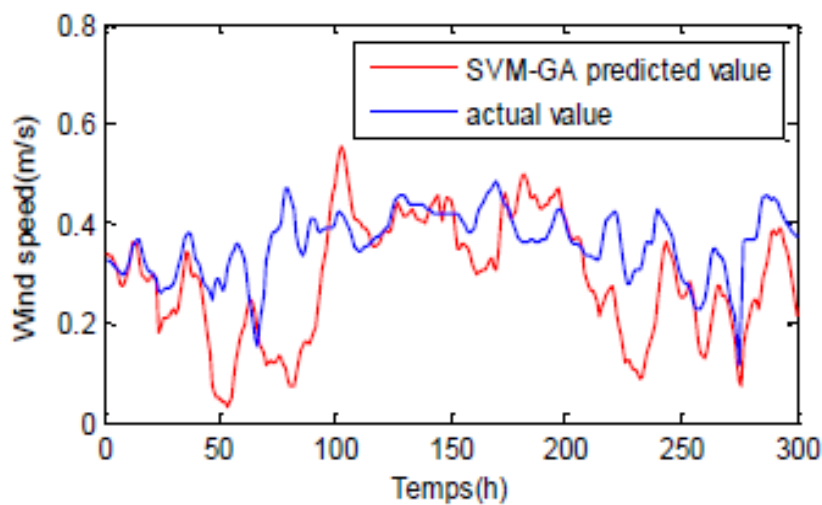


Figure 3.9 : Valeur actuelle/prédite par SVM-GA

### 3.7. Choix du meilleur modèle : Performance des modèles

Dans cette section, nous avons comparé nos approches utilisées pour développer un modèle de prévision de la vitesse du vent. Le tableau 3.14 présente les erreurs des approches proposées. Le coefficient de détermination ( $R^2$ ), également appelé qualité d'ajustement, il montre une très bonne qualité de prévision pour les modèles RLM, RNA-BP, RNA-GA et RNA-PSO puisque les valeurs sont très proches de 1. De plus, les MSE, ME et MAE montrent une erreur très faible dans les modèles qui peut s'expliquer par le fait que les modèles décrivent la plupart des données, mais le RNA-GA est le mieux adapté pour la prévision avec  $MSE = 4,71 \cdot 10^{-4} \text{ m}^2/\text{s}^2$  avec trois neurones dans la couche cachée. En revanche, le  $R^2$  des modèles SVM-GA et SVM-PSO est très proche de 0. Cependant, les MSE, ME et MAE montrent une erreur relativement faible dans les modèles qui peut s'expliquer par le fait que les modèles décrivent la plupart des données, mais pas avec une grande précision.

Tableau 3.15 : Performances des modèles hybrides de prévision

| Méthode | MSE ( $\text{m}^2/\text{s}^2$ ) | MAE (m/s)              | R-squared |
|---------|---------------------------------|------------------------|-----------|
| RLM     | 0.0388                          | 0.1509                 | 0.9816    |
| RNA-BP  | 0.0039                          | 0.0213                 | 0.9982    |
| SVM-GA  | $2.42 \cdot 10^{-2}$            | 0.1238                 | 0.1947    |
| SVM-PSO | $2.60 \cdot 10^{-2}$            | 0.1281                 | 0.1355    |
| RNA-PSO | $5.27 \cdot 10^{-4}$            | 0.0124                 | 0.9881    |
| RNA-GA  | $4.71 \cdot 10^{-4}$            | $9.0892 \cdot 10^{-4}$ | 0.9834    |

### 3.8. Conclusion :

Dans ce chapitre, nous avons tenté d'établir une synthèse globale sur les travaux qui ont été réalisés dans cette thèse. On a vu et discuté les résultats obtenus pour la prévision de la vitesse du vent, où on a proposé une technique basée sur la modélisation statistique pour déterminer les paramètres d'entrée, un type très connu du réseau de neurones artificiel « MLP » et aussi quatre modèles hybrides.

En général, l'application de ces techniques nous a permis d'améliorer les critères de performance et la qualité de prévision. Finalement, après une comparaison des résultats obtenus nous avons pu conclure que notre technique proposée RNA-GA conduit à la prévision de la vitesse du vent avec une meilleure qualité et aussi avec grande précision, ce qui démontre l'utilité d'une telle méthode et permet d'envisager sa mise en service opérationnel.

## Conclusion générale et perspectives

Dans ce travail de thèse, nous avons pu réaliser des prévisions à court terme en se basant sur des approches de machine learning. Tout d'abord un historique d'énergie éolienne dans le cadre national et international ainsi que les caractéristiques du vent comme phénomène variable, suivi d'une meilleure compréhension du fonctionnement des turbines éoliennes, nous avons décrit de façon explicite les différents éléments constitutifs de l'éolienne. Afin d'arriver à un état de l'art qui a permis de se positionner par rapport aux méthodes développées dans ce travail de thèse pour estimer l'horizon de prévision et les marges d'erreur acceptables pour chaque méthode. Un chapitre a été consacré aux différentes techniques de prévision de la vitesse du vent utilisées dans ce travail de thèse ainsi que les indicateurs de performance adoptés.

Finalement, nous avons atteint nos objectifs d'analyser notre base de données collectée des deux sites différents qui sont choisis vu la disponibilité des données et l'importance de ces deux régions dans le plan éolien Marocain où la vitesse du vent dans ces régions dépasse 10 m/s. Après, nous avons pu déterminer les paramètres qui influencent la vitesse du vent pour chacun des deux sites étudiés.

Les méthodes développées dans cette thèse ont donné des résultats remarquables et pertinents, elles ont montré de bonnes concordances avec les résultats de la littérature dans le domaine, mais la recherche continue dans le but d'améliorer la performance parce que comme tout travail de recherche, ce travail connaît des limitations dans les méthodes utilisées et l'horizon de prévision.

Comme perspectives, nous proposons d'hybrider dans le futur d'autres méthodes avec les algorithmes métaheuristiques et aussi un nouveau cadre de recherche basé sur les méthodes physiques qui prendront en considération des paramètres physiques. Nous comptons aussi par la suite de travailler sur des problèmes multiobjectifs plus complexes pour bien améliorer les limites de nos approches.

## Bibliographie

- 1- A. ARGAND, « L'énergie électrique d'origine éolienne », Revue général de l'électricité 82(1982) : 43-50.
- 2- E. HAU, « Wind-Turbines: Fundamentals, Technologies, Application, Economics », livre, 2<sup>ème</sup> édition springer, (2005): 1-801.
- 3- T. Lubin, « Modélisation et commande de la machine synchrone à réluctance variable Prise en compte de la saturation magnétique », thèse de doctorat en génie électrique soutenue à l'université Henri Poincaré, Nancy-I, avril 2003.
- 4- A.E.M.E, « Guide du porteur de projet éolien », livre, 1999.
- 5- A.E.M.E, « Guide de l'étude d'impact sur l'environnement de parcs éoliens », Manuel préliminaire 2000.
- 6- J.-L. Bal, B. Chabot, « Les énergies renouvelables. État de l'art et perspectives de développement », C. R. Geoscience 333 (2001)827–834.
- 7- H. NFAOUI , « Caractéristiques du gisement éolien marocain et optimisation d'un système aérogénérateur/Groupe Electrogène pour l'Electrification des villages isolés », thèse de Doctorat, université mohammed V-AGDAL.
- 8- S. DELENCLOS, « l'énergie éolienne », cours de génie thermique et énergie, université du Littoral côte d'opale,2011-2012.
- 9- M.Ghirez et N.Coquelet, « les éoliennes », cours, faculté des sciences,Université Libre de Bruxelles, 2010.
- 10- M.Mustapha ENZILI, « L'énergie éolienne au Maroc », Session de formation des acteurs de l'Électrification rurale des pays africains francophones",CDER du 19 au 28 novembre 2007.
- 11- O.N.E.E, « Programme intégré de production électrique éolienne »,Tanger le 28 juin 2010.
- 12- A. LARAKI, « Énergies solaire et éolienne au Royaume du Maroc », session de formation de Amisole.
- 13- M. ENZILI, « L'énergie éolienne au Maroc(Ressources,Potentiel et Projets) »,6<sup>ième</sup> conférence Maroc-Allemande sur l'énergie éolienne-Casablanca-Le 15 novembre 2011.

- 14-** C.D.E.R, « L'Énergie éolienne au Maroc: Gisement et Dimensionnement », session de formation, édition 2002.
- 15-** H. Mabchour, « Étude, Modélisation et Expérimentation des Composantes d'un Système Hybride Couplant les Énergies Solaire et éolienne – Performances et Méthodologie de Dimensionnement », Thèse de Doctorat, Université Hassan II Mohammedia, 1999.
- 16-** British Wind Energy Agency, « Offshore Wind Energy : Building a new industry for Britain », report June 1998.
- 17-** C.D.E.R, « Les Ressources Éoliennes du Maroc », session de formation, édition 2007.
- 18-** M. Lopez, « Contribution à l'optimisation d'un système de conversion éolien pour une unité de production isolée », thèse de doctorat. Université PARIS-SUD11.
- 19-** C. NICHITA, « chaîne de conversion électromécanique des éoliennes de petite puissance, Fonctionnement d'un système éolien de petite puissance », cours, université du Havre.
- 20-** STATGRAPHICS Centurion XVI-Manuel de l'utilisateur, imprimé aux États-Unis d'Amérique, Traduction par Christian R.Charles.
- 21-** Ph. BESSE, « Pratique de la modélisation statistique », cours, Université Paul Sabatier.
- 22-** S. V. Thor and P. Weis-Taylor, "Long-term research and development needs for wind energy for the time frame 2000-2020". Wind Energy, 5(1)(2002) :73–75.
- 23-** G. E. P. Box and G. M. Jenkins, "Times series analysis : forecasting and control", livre, 1976.
- 24-** R. B. Corotis, A. B. Sigl, and P. M. Cohen, "Variance analysis of wind characteristics for energy conversion". J. of Applied Meteorology, 16 (1977) :1149–1157.
- 25-** R.W. Katz and R. H. Skaggs, "On the use of autoregressive moving-average processes to model meteorological time series", Monthly Weather Review, 109 (1981) :479–484.

- 26- J. W. Delleur and M. Levent Kavvas,” Stochastic models for monthly rainfall forecasting and synthetic generation”. *Journal of Applied Meteorology*, 17 (1978) :1528–1536.
- 27- R. F. Carlson, A. J. A. MacCormick, and D. G. Watts, “Application of linear random models to four annual stream-flow series”. *Water Resources Research*, 6 (1970) :1070–1078.
- 28- J. M. Davis and P. N. Rappaport, “The use of time series analysis techniques in forecasting meteorological drought” . *Monthly Weather Review*, 102 (1974) :176–180.
- 29- S. Makridakis, S. C. Wheelwright et V. E. McGee, “Forecasting : methods and applications”, livre 2nd ed. JohnWiley & Sons, 1983.
- 30- E. A. Bossanyi, “Short-term wind prediction using kalman filters”. *Wind engineering*, 9(1985) :1–8.
- 31- M. C. Torre, P. Poggi et A. Louche, “Markovian model for studying wind speed time series in corsica”. *International Journal of Renewable Energy Engineering*, 3 (2001) :311–319.
- 32- T. Kitagawa and T. Nomura, “A wavelet-based method to generate artificial wind fluctuation data”. *Journal of Wind Engineering and Industrial Aerodynamics*, 91(2003) :943–964.
- 33- H. G. Beyer, T. Degner, J. Haussmann, M. Hoffman, and P. Rujan. “Short term prediction of wind speed and power output of a wind turbine with neural networks”. In *Proceedings of the Second European Congress on Intelligent Techniques and Soft Computing*, Aachen, Germany, 1994.
- 34- M. Alexiadis, P. Dokopoulos, H. Sahsamanoglou et I. Manousaridis, “Short term forecasting of wind speed and related electrical power”. *Solar Energy*, 3(1998) :61–68.
- 35- A. Sfetsos, “Time series forecasting of wind speed and solar radiation for renewable energy sources”, Thèse de doctorat, Imperial College, London, 1999.
- 36- A. Sfetsos,”A comparison of various forecasting techniques applied to mean hourly wind speed-time series”. *Renewable Energy*, 21(1)(2000) :23–35.
- 37- G. Wu et Z. Dou,”Nonlinear wind prediction using a fuzzy modular temporal neural network”. In *Windpower '95. American Wind Energy Association Conference*, (1995): 285–294.

- 38-** E. Kalnay, S. J. Lord et R. D. McPherson. “Maturity of operational numerical weather prediction: Medium range”, Bulletin of American Meteorological Society, 79(1998): 2753-2770.
- 39-** A. Woods, “weather prediction : medium range”, Bulletin of American Meteorological Society, 79(12) (1998) :2753–2769.
- 40-** [https ://meteo. gc.ca/model\\_forecast/about\\_these\\_products\\_f.html](https://meteo.gc.ca/model_forecast/about_these_products_f.html), Avr. 2013.
- 41-** G. Giebel, “The State-Of-The-Art in Short-Term Prediction of Wind Power”. Energy, Environment and Sustainable Development, 1(2009): 1-36.
- 42-** S. S. Soman, H. Zareipour, O. Malik, et P. Mandal, “A review of wind power and wind speed forecasting methods with different time horizons”, North American Power Symposium (NAPS), 3(2010):1–8.
- 43-** L. Ferrara, « Introduction aux modèles statistiques », cours, Université Paris-ouest, Février 2012.
- 44-** D. Borcard, « Régression Polynomiale », cours ,Département de sciences biologiques, Université de Montréal.
- 45-** Sh. Amiri, S.Mottahedi, S.Asadi, “Using multiple regression analysis to develop energy consumption indicators for commercial buildings in the U.S”. Energy Build. 109 (2015), 209–216.
- 46-** C. Voyant, « Prédiction de séries temporelles de rayonnement solaire global et de production d'énergie photovoltaïque à partir de réseaux de neurones artificiels ». Thèse de Doctorat (2011). Energie électrique, Université Pascal Paoli, France.
- 47-** E. Ghysels, H.Qian,”Estimating MIDAS regressions via OLS with a polynomial parameter profiling”. Econom. Stat. 9(2018): 1–6.
- 48-** S.S.Chong, A.A.Aziz, S.W. Harun, H. Arof, S. Shamshirband, “Application of multiple linear regression, central composite design, and ANFIS models in dye concentration measurement and prediction using plastic optical fiber sensor”. Measurement 74(2015): 78–86.
- 49-** D.G. Kleinbaum, L.L. Kupper, K.E. Muller, “Applied Regression Analysis and Other Multivariable Methods”, livre. PWS Publishing Co., Boston (1988)

- 50-** J.C.M. Pires, F.G. Martins, S.I.V. Sousa, M.C.M. Alvim-Ferraz, M.C. Pereira: Selection and validation of parameters in multiple linear and principal component regressions. *Environ. Model. Softw.* **23**(1)(2008), 50–55.
- 51-** E.LEBARBIER et S.Robin, Exemples d'application du modèle linéaire, Agros Paris Tech.
- 52-** S. Azadi, A. Karimi-Jashni, "Verifying the performance of artificial neural network and multiple linear regression in predicting the mean seasonal municipal solid waste generation rate: A case study of fars province. Iran". *Waste Manag.* **48**(2016): 14–23.
- 53.** D. Graupe, *Principles of Biological Neural Networks*, World Scientific 2007.
- 54-** M.C. Alexiadis, P.S. Dokopoulos, H.S. Sahsamanoglou, I.M. Manousaridis, "Short-term forecasting of wind speed and related electrical power". *Solar Energy* **63**(1998): 61–68.
- 55-** B.A. Yang, X. Li, H. Ji, J. Xu, "An early warning system for loan risk assessment using artificial neural networks". *Knowl. Based Syst.* **14**(2001): 303–306.
- 56-** S. Azadi, A.R. Sepaskhah, "Annual precipitation forecast for west, southwest, and south provinces of Iran using artificial neural networks". *Theor. Appl. Climatol.* **109**(2011): 175–189.
- 57-** Y. Wang, J. Li, J. Gu, Z. Zhou, Z. Wang, "Artificial neural networks for infectious diarrhea prediction using meteorological factors in Shanghai (China)". *Appl. Soft Comput.* **35**(2015) :280–290.
- 58-** O. EL FATNI, "Modélisation Physicochimique et neuronale du procédé de préparation des eaux acidifiées d'extraction du sucre à partir de la betterave : Application au contrôle du PH de ces eaux", thèse de Doctorat en Génie des Procédés, Ecole Mohammadia des ingénieurs, 28 Avril 2006.
- 59-** F. Gaofeng, W. Weisheng, L. Chun, D. Huizhu, "Wind power prediction based on artificial neural network". *Renew. Energy* **28**(34)(2008): 118–123.
- 60-** A. Sozen, E. Arcaklioglu, M. Ozalp, E.G. Kanit, "Use of artificial neural networks for mapping of solar potential in Turkey". *Appl. Energy*, **77**(2004): 273–286.

- 61-** B. Curry, D.E. Rumelhart, “MSnet: a neural network which classifies mass spectra”. Tetra. Comput. Methodol. 3 (1990): 213–237.
- 62-** R. Pasti, L.N. Castro,”Bio-inspired and gradient-based algorithms to train MLPs: the influence of diversity”. Inf. Sci. 179(10)(2009): 1441–1453.
- 63-** D.E. Rumelhart, J.L. McClelland, “An interactive activation model of context effects in letter perception: part I. An account of basic findings”. Computer science, 37(1988): 580–596.
- 64-** K. HORNIK, M. STINCHCOMBE, H. WHITE, "Multilayer Feedforward Networks are Universal Approximators", Neural Networks, 2(1989): 359-366.
- 65-** Anil K. Jain, Jianchang Mao, “Artificial Neural Networks “. A Tutorial, March 1996, IEEE.
- 66-** D. Graupe, “Principles of Biological Neural Networks”, World Scientific 7(2013): 1-364.
- 67-** G. Hinton, "Neural Networks for Machine Learning", Coursera.
- 68-** V. N .Vapnik, “Statistical learning theory”. Livre, John Wiley & Sons ISBN: 0-4761-03003-1(1999).
- 69-** S. Chen, S. Zhou, F.F. Yin, L.B. Marks, S.K. Das, “Investigation of the support vector machine algorithm to predict lung radiation-induced pneumonitis”. Med Phys, 34(2007):3808–14.
- 70-** Z. Zhou, M. Folkert, N. Cannon, P. Iyengar, K. Westover, Y. Zhang, et al, “Predicting distant failure in early stage NSCLC treated with SBRT using clinical parameters”. Radiother Oncol, 119(2016) :501–4.
- 71-** Y. Hayashida, K. Honda, Y. Osaka, T. Hara, T. Umaki, A. Tsuchida, et al, “Possible prediction of chemoradio sensitivity of esophageal cancer by serum protein profiling”. Clin Cancer Res, 11(2005):8042–7.
- 72-** D.E. Goldberg , “Genetic Algorithms in Search, Optimization and Machine Learning”, livre, Addison-Wesley, Reading, MA, 1989.
- 73-** J.M. Alliot et T. Schiex, « Intelligence artificielle et informatique théorique », livre, Cépadués Editions, 1993.
- 74-** J. Arifovic, “Genetic algorithm learning and the cobwell model”. Journal of Economic Dynamic and Control, 18(1994): 2-28.

- 75-** D. Goldberg, "Genetic Algorithms", livre, Addison Wesley editor, 1989.
- 76-** D.E Goldberg, "Real-coded genetic algorithms, virtual alphabets and blocking". Complex Systems, 5(1991):139–167.
- 77-** J. Holland, "Outline for a logical theory of adaptive systems". Journal of the Association of Computing Machinery, 3(1962): 297-314.
- 78-** A.H Wright, "Genetic algorithms for real parameter optimization". In Proceeding of the Foundation Of Genetic Algorithms, 1(1991): 205-218.
- 79-** P.C. Chu et J.E. Beasley, "Constraint Handling in genetic algorithms: The Set Partitioning Problem". Journal of Heuristics, 11(1998): 323-357.
- 80-** C.L Bridges et D.E Goldberg. "An analysis of multipoint crossover". In Proceedings of the Foundation of Genetic Algorithms. FOGA, 1(1991): 301-315.
- 81-** L.J Fogel, A.J Owens et M.J Walsh, "Artificial Intelligence Through Simulated Evolution". Wiley and sons. NY, 14(1966): 170.
- 82-** M. Clerc, "Discrete Particle Swarm Optimization illustrated by the Traveling SalesmanProbleme". Studies in Fuzziness and Soft Computing, 141(2003): 219-239.
- 83-** M. Clerc et Siarry. Une nouvelle métaheuristique pour l'optimisation difficile : la méthode des essais particuliers. Journales sur l'enseignement des sciences et technologies de l'information et des systèmes, 3(2004) : 1-13.
- 84-** Eudes et Rioland, « Optimisation par essaim particulier pour un problème d'ordonnancement et d'affectation de ressources », Projet de Modélisation et Calcul Scientifique, 2007.
- 85-** Gourgand et Kemmoe. "Particle Swarm Optimization : A study of particle displacement for solving continuous and combinatorial optimization problems", International Journal of Production Economics, 121(2009): 57-67.
- 86-** M. Clerc et J.Kennedy, The Particle Swarm: Explosion, Stability and Convergence in a Multi- Dimensional complex Space, IEEE Transactions on Evolutionary Computation, 6(2001): 58-73.

- 87-** J. Kennedy et R. Eberhart. “Particle Swarm Optimization”, conférence internationale de l'IEEE sur les réseaux de neurones , Perth, Australia, 4(1995) : 1-7.
- 88-** J. Kennedy et R.C. Eberhart, “Swarm Intelligence”, livre. Morgan Kaufman Publishers, Academic Press, 2001.
- 89-** M.G.H.Omran, “Particle Swarm Optimization Methods for Pattern Recognition and Image Processing”, Thèse du doctorat, Université de Pretoria, Novembre 2004.
- 90-** F. Van den Bergh, “An Analysis of Particle Swarm Optimizers”, Thèse du doctorat, Département d'informatique, Université de Pretoria, Afrique du Sud, 2002.
- 91-** M.N. Vrahatis, K.E. Parsopoulos, “Parameter selection and adaptation in Unified Particle Swarm Optimization”, Mathematical and Computer Modelling, 46(2007): 198-213.
- 92-** D. Ahlburg, “Error measures and the choice of a forecast method. International Journal of Forecasting n°6”, 8(1992): 99-100.
- 93-** S. Kalogirou, “Artificial neural networks in renewable energy systems applications: a review”. Renewable and Sustainable Energy Reviews, n°12, 5(2001): 373-401.
- 94-** K.B. Debnath, M. Mourshed, “Forecasting methods in energy planning models”. Renew. Sustain. Energy Rev. 88(2018): 297–325.
- 95-** M.A. Mohandes, S. Rehman, T.O. Halawani, “A neural networks approach for wind speed prediction”. Renew. Energy 13(1998): 345–354.
- 96-** R. Velo, P. Lopez, F. Maseda, “Wind speed estimation using multilayer perceptron”. Energy Convers. Manag. 81(2014): 1–9.
- 97-** S.S. Chandel, P. Ramasamy, A.K. Yadav, “Wind speed prediction in the mountainous region of India using an artificial neural network model”. Renew. Energy 80(2015): 338–347.
- 98-** A. Mellit, S.A. Kalogirou, “Artificial intelligence techniques for photovoltaic applications: A review”. Progress in Energy and Combustion Science , 34(2008): 574-632.

**99-** C. Voyant, « Prédiction de séries temporelles de rayonnement solaire global et de production d'énergie photovoltaïque à partir de réseaux de neurones artificiels », thèse de Doctorat en Energétique, Ecole doctorale environnement et société, Université de Corse-Pascal Paoli, 16 novembre 2011.

**100-** J.C. Lomares-Salas, A. Aguera-Pérez, J.J.G. de la Rosa, A. Moreno-Munoz, "A novel neural network method for wind speed forecasting using exogenous measurements from agriculture stations". Measurement 55(2014): 295-304.

## **Annexes**

---

## Contribution I

Forecasting of wind speed using a multiple linear regression and artificial neural networks

---

### *Résumé*

Dans cet article, deux méthodes sont développées pour la prédiction de la vitesse du vent, à savoir la Régression linéaire multiple (RLM) et les réseaux de neurones artificiels (RNA) pour deux sites éoliens différents au nord et au sud du Maroc pendant trois ans (2011–2012–2013). La première méthode consiste à déterminer les paramètres qui influencent la vitesse du vent afin de construire un modèle de régression entre les prédicteurs et la variable dépendante. La deuxième approche proposée est les RNA où le réseau neuronal choisi est le perceptron multicouche qui utilise le back-propagation (BP) comme algorithme d'apprentissage. Les résultats montrent que les modèles RLM et RNA ont un coefficient de détermination acceptable entre la valeur actuelle et prédite de la vitesse du vent. Cependant, les RNA sont plus performants en termes d'erreurs statistiques notamment en termes d'erreur absolue moyenne, d'erreur absolue moyenne en pourcentage et d'erreur quadratique moyenne.



# Forecasting of wind speed using multiple linear regression and artificial neural networks

Soukaina Barhmi<sup>1</sup> · Omkaltoume Elfatni<sup>1</sup> · Ismail Belhaj<sup>1</sup>

Received: 8 February 2018 / Accepted: 11 April 2019  
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

## Abstract

In this paper, two methods are developed for the prediction of wind speed, namely, the Multiple Linear Regression (MLR) and Artificial Neural Networks (ANNs) in north and south regions of Morocco for three years (i.e., 2011–2012–2013). The first method consists of determining the parameters which most significantly influence the wind speed in order to build a regression model between the predictors and the dependent variable. The second proposed approach is ANNs where the neural network chosen is the multilayer perceptron that uses the back-propagation (BP) as a supervised learning technique for training. The results show that both MLR and ANNs models predict the wind speed at an acceptable correlation coefficient between the actual and predicted wind speed. However, the ANNs perform better in terms of statistical errors notably in terms of mean absolute error, mean absolute percentage error and mean square error.

**Keywords** Wind energy · Wind speed · Prediction · Artificial neural networks · Multiple linear regression

## 1 Introduction

Wind energy has become increasingly popular due to the limitations of the global fossil and nuclear fuel resources. Furthermore, wind energy is considered as a clean resource energy which doesn't emit dangerous emissions, like those produced by fossil fuel power stations, that cause several human health problems. The wind power is greatly affected by the wind speed, and there exists a strong positive correlation

---

✉ Ismail Belhaj  
ibelhaj13@gmail.com  
Soukaina Barhmi  
soukainabarhmi90@gmail.com  
Omkaltoume Elfatni  
omfatni@gmail.com

<sup>1</sup> Faculty of Sciences, Laboratory of High Energy Physics-Modeling and Simulations, Rabat, Morocco

between them. In this sense, the accurate prediction of wind speed is indispensable in order to protect the security of the running systems, decrease the possibilities of wind power break down and protect the security of the wind power integration.

Morocco is a Northern African country located between latitude of  $21^{\circ}$  and  $36^{\circ}$  in north and longitude of  $1^{\circ}$  and  $17^{\circ}$  in the west. It is bordered by the Atlantic Ocean to the west, the Mediterranean Sea to the north, Algeria to the east and Mauritania to the southeast, with an area of  $750.810 \text{ km}^2$  ( $289.90 \text{ mi}^2$ ) [11]. Along with its key geographical location, Morocco benefits from a great solar and wind energy potential. Two major renewable energy source initiatives (i.e., the Moroccan wind and solar projects) have been launched in order to reach the national target of increasing the share of renewable energy sources in the energy mix up to 42% by 2020 [12].

In recent years, several scientists have focused their research on the wind speed prediction [24]. Debnath et al. [7] reviewed 483 energy planning models (EPMs) and revealed the use of fifty different methods for the period between 1985 and 2017. Among those 50 identified methods, we found computational intelligence (CI), Mathematical Programming (MP) and statistical methods. For CI methods, there were Artificial Neural Networks (ANNs) used in 194 EPMs, followed by Support Vector Machines (SVM) applied in 58 models, Fuzzy Logic in 40 models, Genetic Algorithm in 39 models, Particle Swarm Optimization in 34 models and Grey prediction in 29 models. In the case of statistical methods, Autoregressive Integrated Moving Average (ARIMA), Linear Regression and Autoregressive Moving Average (ARMA) are used for forecasting. Even these three methods are applied to electric load and renewable energy, the forecasting accuracy of CI methods was better than that of statistical and MP methods.

Mohandes et al. exploited in their work [15] the use of ANNs with three layered, feed-forward, back-propagation and the results obtained were 8.9% for the Mean Absolute Percentage Error (MAPE) and 0.9380 for the correlation coefficient (R). In the same way, Velo et al. use ANNs with three layers trained by the Back-Propagation algorithm and the results were 0.5 for R and 6% for MAPE [25]. Chandel et al. employed the ANNs model to predict the daily mean wind speed using measured data of temperature, air pressure, solar radiation and altitude. The results obtained were 6.489% for MAPE and 0.99 for R [4]. In their study, Palomares-Salas et al. exploited the use of ANNs and compared their performance to ARIMA and persistence model. The best model found was a backpropagation network with one hidden layer and nine neurons ( $0.18 \text{ m/s}$  for MAE and  $0.05 \text{ m}^2/\text{s}^2$  for MSE) [16].

On the other hand, Liu et al. [14] proposed a hybrid method for wind speed prediction based on the adaptive boosting (Adaboost) algorithm and the multilayer perceptron (MLP) neural networks. They adopted four different algorithms in their approach for the training and modeling of the MLP neural networks, including the gradient descent with adaptive learning rate back-propagation algorithm, gradient descent with momentum and adaptive learning rate back-propagation algorithm, conjugate gradient back-propagation with Fletcher-Reeves algorithm and Broyden-Fletcher-Goldfarb-Shanno algorithm. According to simulation results, they concluded that this hybrid method was effective for the wind speed prediction. In their study, Shukur et al. introduced another new hybrid model KF-ANN based on the ANN and Kalman filter (KF) and compared it with ARIMA, ANN and hybrid ARIMA-KF models. The results

showed that all these models were effective but the MAPE values indicated that the hybrid KF-ANN model was the most effective [22].

The selection of any model input parameters is of crucial importance to enhance the output forecasting performance. Unfortunately, most of the research works in literature proposed just methods to predict the wind speed based on meteorological parameters without taking into consideration the multicollinearity between them. In this sense, the main contribution of our study consists on carrying out a rigorous analysis of the meteorological parameters by using the ordinary least squares (OLS) analysis in order to eliminate the multicollinearity between these variables and to determine the parameters which most influence the wind speed in northern and southern regions of Morocco. Based on the evaluation of meteorological data, we find that the selection of parameters influencing the wind speed is different for each region due to the climate changes between northern and southern Moroccan regions. Moreover, we perform the data processing before applying two suggested models, namely, the multiple linear regression (MLR) and artificial neural networks (ANNs) to predict the wind speed.

This paper is divided into five sections: Sect. 2 provides a description of input data and some information about the regions. Section 3 discusses in details the MLR and ANNs methods. Section 4 analyses the input data and discusses the results. Section 5 closes the paper.

## 2 Data

The data handled are the hourly meteorological satellite data from 2011 to 2013 (i.e., 3 years), namely, temperature, pressure, past wind speed, humidity and direction provided by the Research Institute in Solar Energy and New Energies (IRESEN). The regions studied are Tangier in the south Mediterranean ocean (i.e., located in the latitude of 35,800, longitude of  $-5840$  and altitude of 8 in north Morocco) and Tarfaya next to the Pacific Ocean (i.e., located in the latitude of 27,900 and longitude of  $-12,900$  and altitude of 16 in southern Morocco).

## 3 Methods

### 3.1 Multiple linear regression (MLR)

Here, we develop MLR to predict the hourly wind speed. The MLR is one of the classical statistical tools to describe relationships between a dependent and independent variables [2,5]. The developed regression model for prediction is represented by the following equation:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \varepsilon \quad (1)$$

where  $y$  is the dependent variable,  $x_i (i = 1, 2, 3, \dots, k)$  are the explanatory independent variables,  $\beta_i (i = 1, 2, 3, \dots, k)$  are the regression coefficients, and  $\varepsilon$  is the residual error. In this study, different independent variables, namely, direction, temperature, humidity, past wind speed and pressure are considered. The wind speed is

considered as a dependent variable. The database is divided into two groups: 80% of the data are selected randomly and used to train the regression model while 20% are used to validate the developed regression model [21].

To develop the MLR model, it is necessary to remove the multicollinearity of independent variables and to estimate the parameters in a linear regression model using the ordinary least squares (OLS) analysis. OLS chooses the parameters of a linear function of a set of independent variables by minimizing the sum of the squared differences between the observed dependent variable (i.e., values of the variable being predicted) in the given dataset and those predicted by the linear function [10]. The selected variables for each method are validated with t-test using a significance level of 0.05 [13,18].

### 3.2 Artificial neural networks (ANNs)

In this subsection, we introduce the method used in the prediction of wind speed. The ANNs is composed of an input layer, one or more hidden layers and an output layer [1,27]. These inputs are normalized in the range [0, 1] by the following equation:

$$\text{Normalized input, } X'_i = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad (2)$$

where  $X_i$  are the actual data,  $X_{min}$  and  $X_{max}$  are the minimum and maximum values in the input data respectively.

A particular structure of ANNs is feed-forward neural network which is popular and widely applied in many applications; including forecasting [3,26]. Any layer that is formed between the input layer and the output layer is called hidden layer. These layers are interconnected with each other by weights [9,23].

The Tangent Sigmoid (TanSig) activation function is applied over the inputs to calculate the output. The output of the neural network is given by the following equation:

$$Y = \frac{1}{1 + \exp(-z)} \quad \left( z = \sum_{i=1}^N X_i * W_i + b_i \right) \quad (3)$$

where  $X_i$  are the inputs,  $W_i$  are the weights and  $b_i$  are the biases.

The computation of the network can be divided into two steps. The first step is to train the ANNs in order to find the optimal weights that minimize the mean square error (MSE) using Back-Propagation (BP) algorithm [6,17,20]. The second step is to validate the already trained model.

## 4 Results and discussion

### 4.1 MLR results

In this study, the selection of input parameters for the MLR and ANNs models is of crucial importance to enhance the accuracy of forecasting. Here, five different param-

eters, namely, temperature, pressure, humidity, past wind speed, and wind direction are available.

In the first step, to better understand the relation between these parameters, it is necessary to determine the presence of interaction between them to remove the multicollinearity. From Tables 1 and 2, it can be seen that the pressure has a collinearity with temperature and humidity in Tangier and the direction has a collinearity with temperature and pressure in Tarfaya.

In the second step, to remove multicollinearity, the OLS is used for the elimination of P-value  $> 0.05$ . If the P-value is less than 0.05, the parameters have to be accepted while in the other case (i.e., the P-value is greater than 0.05) the parameters have to be rejected. According to the P-value shown in Tables 3 and 4, we notice that the highest P-value on the independent variables is 0.3999, belonging to pressure in Tangier and 0.8620 to direction in Tarfaya. Since the P-value is greater than 0.05, that term is not statistically significant at the 95.0% or higher confidence level. Consequently, we should consider removing pressure from the model of Tangier and direction from the model of Tarfaya.

Finally, four explanatory independent variables that most influence the wind speed are direction, humidity, temperature and past wind speed (in Tangier) and humidity, temperature, pressure and past wind speed (in Tarfaya). All were used to develop the MLR model.

By comparing the correlation coefficients between independent variables and dependent variable shown in Tables 5 and 6, we notice that the correlation coefficient has positive and negative values. This means that if any of the parameters increases in its amount, the wind speed rate increases as well. Furthermore, the negative correlation coefficient shows the inverse relation between them. The direction and past wind speed

**Table 1** Correlation matrix of the explanatory variables in Tangier

|                 | Temperature | Humidity | Pressure        | Past wind speed | Direction |
|-----------------|-------------|----------|-----------------|-----------------|-----------|
| Temperature     | 1.0000      | 0.4699   | − <b>0.9968</b> | 0.0340          | 0.0646    |
| Humidity        |             | 1.0000   | − <b>0.6160</b> | − 0.0452        | 0.0705    |
| Pressure        |             |          | 1.0000          | − 0.0709        | − 0.1042  |
| Past wind speed |             |          |                 | 1.0000          | 0.1480    |
| Direction       |             |          |                 |                 | 1.0000    |

**Table 2** Correlation matrix of the explanatory variables in Tarfaya

|                 | Temperature | Humidity | Pressure | Past wind speed | Direction       |
|-----------------|-------------|----------|----------|-----------------|-----------------|
| Temperature     | 1.0000      | 0.4274   | − 0.0408 | 0.0035          | − <b>0.9974</b> |
| Humidity        |             | 1.0000   | − 0.4734 | − 0.0920        | 0.0793          |
| Pressure        |             |          | 1.0000   | − 0.0447        | <b>0.7109</b>   |
| Past wind speed |             |          |          | 1.0000          | 0.2910          |
| Direction       |             |          |          |                 | 1.0000          |

**Table 3** Statistical analysis obtained from the MLR at Tangier station

| Parameter       | Code  | Coefficient value | Standard error | t-value  | P-value       |
|-----------------|-------|-------------------|----------------|----------|---------------|
| Direction       | $X_1$ | -0.0000987504     | 0.0000144857   | -6.8171  | 0.0000        |
| Humidity        | $X_2$ | -0.00127188       | 0.0000830857   | -15.308  | 0.0000        |
| Temperature     | $X_3$ | 0.000407384       | 0.0000214706   | 18.9741  | 0.0000        |
| Past wind speed | $X_4$ | 0.990865          | 0.00101643     | 974.848  | 0.0000        |
| Pressure        | $X_5$ | -0.000060839      | 0.0000722721   | -0.84181 | <b>0.3999</b> |

**Table 4** Statistical analysis obtained from the MLR at Tarfaya station

| Parameter       | Code  | Coefficient value | Standard error | t-value  | P-value       |
|-----------------|-------|-------------------|----------------|----------|---------------|
| Humidity        | $X_1$ | -0.00534814       | 0.0001524      | -35.0928 | 0.000         |
| Past wind speed | $X_2$ | 0.995477          | 0.0010076      | 987.964  | 0.000         |
| Pressure        | $X_3$ | -0.00240208       | 0.000137537    | -17.465  | 0.000         |
| Temperature     | $X_4$ | 0.00940481        | 0.000458722    | 20.5022  | 0.000         |
| Direction       | $X_5$ | 0.00000280115     | 0.0000161195   | 0.173775 | <b>0.8620</b> |

**Table 5** Correlation coefficients between independent and dependent variables in Tangier

| City    | Dependent variable | Direction | Humidity | Temperature | Past wind speed |
|---------|--------------------|-----------|----------|-------------|-----------------|
| Tangier | Wind speed         | -0.1954   | -0.0201  | -0.0608     | 0.9906          |

**Table 6** Correlation coefficients between independent and dependent variables in Tarfaya

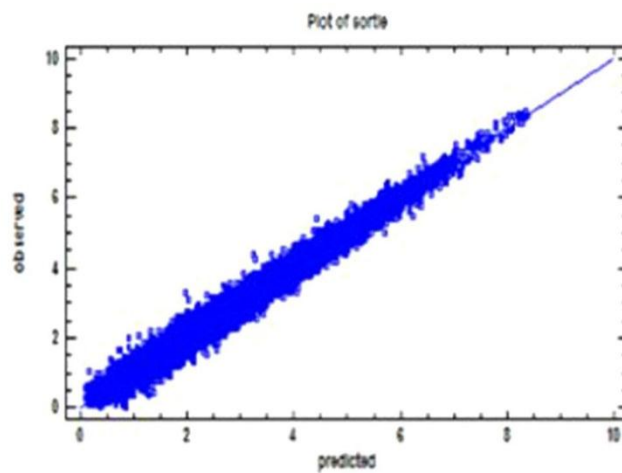
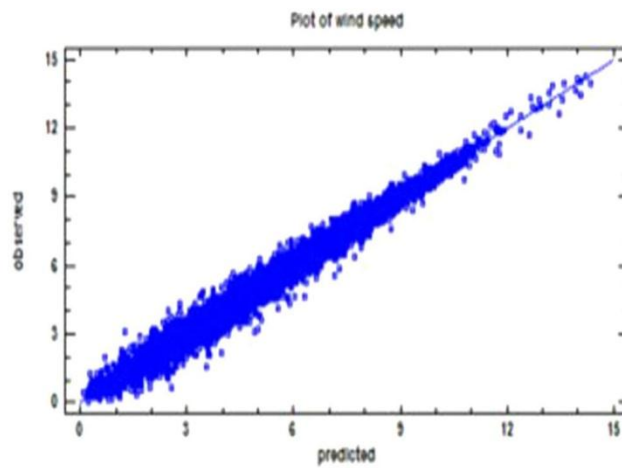
| City    | Dependent variable | Pressure | Humidity | Temperature | Past wind speed |
|---------|--------------------|----------|----------|-------------|-----------------|
| Tarfaya | Wind speed         | 0.1246   | 0.1020   | -0.1852     | 0.9874          |

have the least and most correlation with wind speed in Tangier while temperature and past wind speed have the least and most correlation in Tarfaya, which means that the parameters don't have the same influence on the wind speed.

After generating four independent variables without any multicollinearity, a cross validation OLS-MLR model is developed and t-test (significance level of 0.05) is used to calculate the statistically valid coefficients. Finally, the equation for MLR model shown in Table 7 is used to predict the wind speed. This regression model has a high value of square correlation coefficient  $R^2$  equal to 0.9964, 0.9981 in Tangier and Tarfaya respectively. The obtained values indicate that there is a good agreement between experimental and predicted values attained from the MLR model, as illustrated in Figures 1 and 2.

**Table 7** Regression model from MLR method

| Region  | Equation                                                                          | $R^2$  |
|---------|-----------------------------------------------------------------------------------|--------|
| Tangier | $Y = -0,0000987504 * X_1 - 0,00127188 * X_2 + 0,000407384 * X_3 + 0,990865 * X_4$ | 0.9964 |
| Tarfaya | $Y = -0,00534814 * X_1 + 0,995477 * X_2 - 0,00240208 * X_3 + 0,00940481 * X_4$    | 0.9981 |

**Fig. 1** Regression plot in Tangier**Fig. 2** Regression plot in Tarfaya

## 4.2 ANNs results

To choose the best ANN model for predicting the wind speed, different structures of ANNs with three layers (i.e., an input layer, one hidden layer and an output layer) and different numbers of neurons in the hidden layer were analyzed. The wind speed prediction accuracy is determined by the coefficient of determination ( $R^2$ ) [18,19], Mean Square Error (MSE), Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) between predicted and measured values of wind speed, their formulas are given below:

$$MSE = \frac{1}{N} \sum_{i=1}^N (T_i - Y_i)^2 \quad (4)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |T_i - Y_i| \quad (5)$$

$$RMSE = \sqrt{\sum_{i=1}^N \frac{(T_i - Y_i)^2}{N}} \quad (6)$$

$$MAPE = \left( \frac{1}{N} \sum_{i=1}^N \left| \frac{T_i - Y_i}{T_i} \right| \right) * 100\% \quad (7)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (T_i - Y_i)^2}{\sum_{i=1}^N (T_i - \bar{T})^2} \quad (8)$$

where  $T_i$  are the targets,  $\bar{T}$  is the mean of all target values,  $Y_i$  are the neural network outputs and  $N$  is the number of samples.

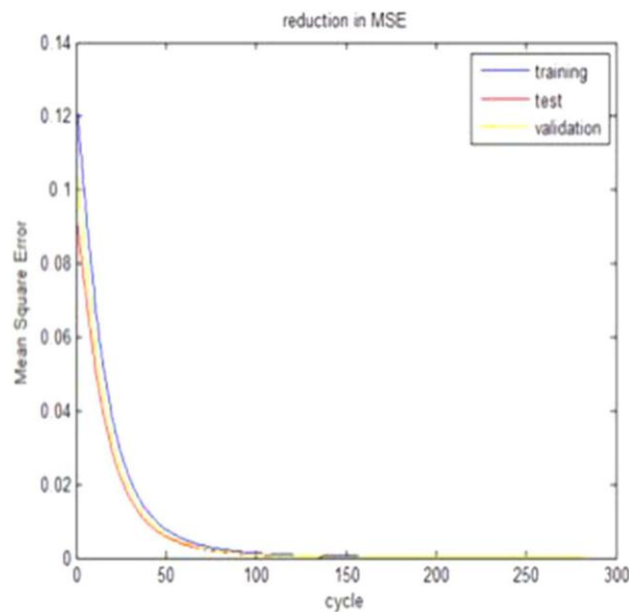
The networks with one hidden layer having five neurons (4-5-1) for Tangier and one neuron (4-1-1) for Tarfaya trained with the back-propagation algorithm have the lowest values of MAPE of 1.9644%, 1.9027% in Tangier and Tarfaya respectively as presented in Table 8. These ranges of error provide high prediction accuracy in wind speed data prediction [4]. Moreover, the obtained results are better than those reported by Fadare. [8] (i.e., MAPE of 8.9%) which used similar ANNs models for the prediction of wind speed. The optimum network, which has the coefficient of determination of 0.9982, 0.9957 in Tangier and Tarfaya, respectively, are obtained based on the testing dataset.

Figures 3 and 4 illustrate the plots of training errors, validation errors and testing errors of wind speed prediction. In these cases, the results are reasonable because of the following considerations:

- The final mean-square errors are small.
- No significant overfitting occurred for wind speed prediction network training.
- The test set error and the validation set error have similar characteristics.

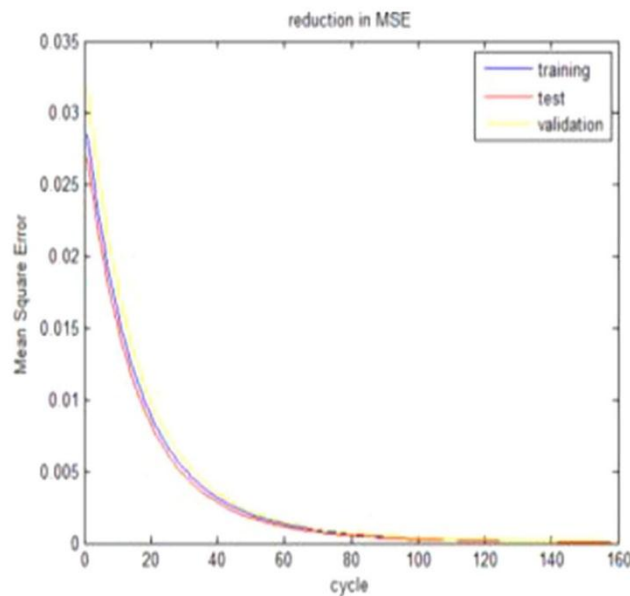
**Table 8** Statistical indicators of the ANN model in Tangier and Tarfaya

| City    | Structure    | R-squared | MSE    | RMSE   | MAE    | MAPE   |
|---------|--------------|-----------|--------|--------|--------|--------|
| Tangier | 4-1-1        | 0.9981    | 0.0040 | 0.0630 | 0.0215 | 1.9715 |
|         | 4-2-1        | 0.9981    | 0.0039 | 0.0628 | 0.0214 | 1.9805 |
|         | 4-4-1        | 0.9982    | 0.0039 | 0.0625 | 0.0212 | 1.9705 |
|         | <b>4-5-1</b> | 0.9982    | 0.0039 | 0.0628 | 0.0213 | 1.9644 |
| Tarfaya | <b>4-1-1</b> | 0.9957    | 0.0157 | 0.1254 | 0.0522 | 1.9027 |
|         | 4-2-1        | 0.9956    | 0.0163 | 0.1278 | 0.0571 | 2.0647 |
|         | 4-3-1        | 0.9956    | 0.0163 | 0.1275 | 0.0573 | 2.0692 |
|         | 4-5-1        | 0.9955    | 0.0167 | 0.1294 | 0.0558 | 2.0212 |

**Fig. 3** Reduction in MSE for the network with 4-5-1 configuration in Tangier

### 4.3 Discussion

The goodness-of-fit of regression model is evaluated giving the fact that the coefficient of determination  $R^2$  is close to 1 and MSE is low (i.e., close to 0). The results of wind speed prediction using neural network (4-5-1, TanSig, BP), (4-1-1, TanSig, BP) and MLR model are shown in Tables 9 and 10. According to this table, the optimum ANNs is able to predict the wind speed of two Moroccan cities in this study with  $R^2$  and MSE values equal to 0.9982 and 0.0039 respectively in Tangier, 0.9957 and 0.0157 in Tarfaya. However in the MLR model, these statistical parameters are 0.9816 and 0.0388 in Tangier, 0.9726 and 0.1009 in Tarfaya.



**Fig. 4** Reduction in MSE for the network with 4-1-1 configuration in Tarfaya

**Table 9** Performance of MLR and ANNs methods in Tangier

| Statistical parameter | MLR    | ANN    |
|-----------------------|--------|--------|
| R-squared             | 0.9816 | 0.9982 |
| MSE                   | 0.0388 | 0.0039 |
| RMSE                  | 0.1969 | 0.0628 |
| MAE                   | 0.1509 | 0.0213 |
| MAPE                  | 7.3020 | 1.9644 |

**Table 10** Performance of MLR and ANNs methods in Tarfaya

| Statistical parameter | MLR    | ANNs   |
|-----------------------|--------|--------|
| R-squared             | 0.9726 | 0.9957 |
| MSE                   | 0.1009 | 0.0157 |
| RMSE                  | 0.3177 | 0.1254 |
| MAE                   | 0.2410 | 0.0522 |
| MAPE                  | 5.9648 | 1.9027 |

The coefficient of determination values are high for both MLR and ANNs while the statistical errors MSE, RMSE, MAE and MAPE are lower for ANNs when compared to the MLR. This demonstrates that the performances of nonlinear models are better than that of linear MLR model. The MLR method is chosen only as a representative statistical model in this study in order to determine the variables that most influence the wind speed. The reason why the nonlinear model performs better than the MLR model may be attributed to the complex nonlinear relationship between the wind speed

and the meteorological variables. Which implies the nonlinearity of the investigated phenomenon.

## 5 Conclusion

This paper uses MLR and ANNs to predict the wind speed based on historical data. The goal of this work is to determine the parameters which most influence the wind speed and to create the MLR and ANNs models. The simulation results show that both models predict the wind speed at a high value of the determination coefficient between actual and predicted wind speed. Based on these results, ANNs has an acceptable accuracy with MSE of 0.0039, 0.0157 and MAPE of 1.9705%, 1.9027% in Tangier and Tarfaya respectively. Hence, it could be concluded that the ANNs is an effective tool for predicting wind speed.

## References

1. Alexiadis, M.C., Dokopoulos, P.S., Sahsamanoglou, H.S., Manousaridis, I.M.: Short-term forecasting of wind speed and related electrical power. *Solar Energy* **63**, 61–68 (1998)
2. Azadi, S., Karimi-Jashni, A.: Verifying the performance of artificial neural network and multiple linear regression in predicting the mean seasonal municipal solid waste generation rate: A case study of fars province. Iran. *Waste Manag.* **48**, 14–23 (2016)
3. Azadi, S., Sepaskhah, A.R.: Annual precipitation forecast for west, southwest, and south provinces of Iran using artificial neural networks. *Theor. Appl. Climatol.* (2011)
4. Chandel, S.S., Ramasamy, P., Yadav, A.K.: Wind speed prediction in the mountainous region of India using an artificial neural network model. *Renew. Energy* **80**, 338–347 (2015)
5. Chong, S.S., Aziz, A.A., Harun, S.W., Arof, H., Shamshirband, S.: Application of multiple linear regression, central composite design, and ANFIS models in dye concentration measurement and prediction using plastic optical fiber sensor. *Measurement* **74**, 78–86 (2015)
6. Curry, B., Rumelhart, D.E.: An MSnet: a neural network which classifies mass spectra. *Tetra. Comput. Methodol.* **3**, 213–237 (1990)
7. Debnath, K.B., Mourshed, M.: Forecasting methods in energy planning models. *Renew. Sustain. Energy Rev.* **88**, 297–325 (2018)
8. Fadare, D.A.: The application of artificial neural networks to mapping of wind speed profile for energy application in Nigeria. *Appl. Energy* **87**, 934–942 (2010)
9. Gaofeng, F., Weisheng, W., Chun, L., Huizhu, D.: Wind power prediction based on artificial neural network. *Renew. Energy* **28**(34), 118–123 (2008)
10. Ghysels, E., Qian, H.: Estimating MIDAS regressions via OLS with a polynomial parameter profiling. *Econom. Stat.* **9**, 1–6 (2018)
11. Kingdom of Morocco: Moroccan Regions 2010. [http://www.hcp.ma/downloads/Maroc-en-chiffres\\_t13053.html/](http://www.hcp.ma/downloads/Maroc-en-chiffres_t13053.html/). Last checked on Dec 29, 2017 (2009)
12. Kingdom of Morocco: Moroccan Investment Development Agency. <http://invest.gov.ma/>. Last checked on Dec 29, 2017 (2011)
13. Kleinbaum, D.G., Kupper, L.L., Muller, K.E.: *Applied Regression Analysis and Other Multivariable Methods*. PWS Publishing Co., Boston (1988)
14. Liu, H., Tian, H.Q., Li, Y.F.: Comparison of four Adaboost algorithm based artificial neural network in wind speed predictions. *Energy Convers. Manag.* **92**, 67–81 (2015)
15. Mohandes, M.A., Rehman, S., Halawani, T.O.: A neural networks approach for wind speed prediction. *Renew. Energy* **13**(3), 345–354 (1998)
16. Palomares-Salas, J.C., Agüera-Pérez, A., de la Rosa, J.J.G., Moreno-Munoz, A.: A novel neural network method for wind speed forecasting using exogenous measurements from agriculture stations. *Measurement* (2014)

17. Pasti, R., de Castro, L.N.: Bio-inspired and gradient-based algorithms to train MLPs: the influence of diversity. *Inf. Sci.* **179**(10), 1441–1453 (2009)
18. Pires, J.C.M., Martins, F.G., Sousa, S.I.V., Alvim-Ferraz, M.C.M., Pereira, M.C.: Selection and validation of parameters in multiple linear and principal component regressions. *Environ. Model. Softw.* **23**(1), 50–55 (2008)
19. Rajendra, M., Jena, P.C., Raheman, H.: Prediction of optimized pretreatment process parameters for biodiesel production using ANN and GA. *Fuel* **88**(5), 868–875 (2009)
20. Rumelhart, D.E., McClelland, J.L.: An interactive activation model of context effects in letter perception: part I. An account of basic findings. *Readings in cognitive science: a perspective from psychology and artificial intelligence*, pp 580–596 (1988)
21. Shams Amiri, S., Mottahedi, S., Asadi, S.: Using multiple regression analysis to develop energy consumption indicators for commercial buildings in the U.S. *Energy Build.* **109**, 209–216 (2015)
22. Shukur, O.B., Lee, M.H.: Daily wind speed forecasting through hybrid KF-ANN model based on ARIMA. *Renew. Energy* **76**, 637–647 (2015)
23. Sozen, A., Arcaklioglu, E., Ozalp, M., Kanit, E.G.: Use of artificial neural networks for mapping of solar potential in Turkey. *Appl. Energy* **77**, 273–286 (2004)
24. Tasckaraoglu, A., Uzunoglu, M.: A review of combined approaches for prediction of short-term wind speed and power. *Renew. Sustain. Energy Rev.* **34**, 243–254 (2014)
25. Velo, R., Lopez, P., Maseda, F.: Wind speed estimation using multilayer perceptron. *Energy Convers. Manag.* **81**, 1–9 (2014)
26. Wang, Y., Li, J., Gu, J., Zhou, Z., Wang, Z.: Artificial neural networks for infectious diarrhea prediction using meteorological factors in Shanghai (China). *Appl. Soft Comput.* **35**, 280–290 (2015)
27. Yang, B.A., Li, X., Ji, H., Xu, J.: An early warning system for loan risk assessment using artificial neural networks. *Knowl. Based Syst.* **14**, 303–306 (2001)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Contribution II

### Hourly wind speed forecasting based on support vector machine and artificial neural networks

---

#### *Résumé*

Dans cet article, nous proposons quatre modèles hybrides basés sur la machine à vecteurs supports notée en anglais (SVM) et les réseaux de neurones artificiels (RNA) pour développer un modèle de prévision de la vitesse du vent. Tout d'abord, une analyse par la méthode des moindres carrés ordinaires notée en anglais (OLS) est faite pour sélectionner les paramètres qui influencent la vitesse du vent. Ensuite, les quatre modèles hybrides proposés sont : SVM-GA , SVM-PSO, NN-GA et NN-PSO. La performance de ces modèles est évaluée à l'aide de trois indicateurs statistiques: l'erreur quadratique moyenne (MSE), l'erreur moyenne (ME) et l'erreur absolue moyenne (MAE). Les résultats montrent une meilleure performance du modèle neuronale par rapport à la machine à vecteurs supports.

## Hourly wind speed forecasting based on support vector machine and artificial neural networks

Soukaina Barhmi, Omkalthoume El Fatni

Laboratory of High Energy Physics-Modeling and Simulations, Rabat, Morocco

### Article Info

#### Article history:

Received May 17, 2019

Revised Jul 27, 2019

Accepted Aug 23, 2019

#### Keywords:

Genetic algorithm

Neural Network

Particle swarm optimization

Support vector machine

Wind speed forecasting

### ABSTRACT

Wind speed is the main component of wind power. Therefore, wind speed forecasting is of big importance due to its uses. It permits to plan the dispatch, determine the hours of storage needed, the amount of energy stored that should be used and avoid the big fluctuations in the electrical grid caused by the nature of the renewable energy resources. In this paper, we propose four hybrid models based on Support Vector Machine (SVM) and Artificial Neural Networks (ANNs) or just Neural Networks (NN) for wind speed forecasting. Using the Ordinary Least Squares (OLS) analysis for selecting the parameters more influencing wind speed. Then, a Support Vector Machine and Artificial Neural Networks models are tuned by Genetic Algorithm (GA) and Particle Swarm Optimization (PSO). The performance of these models is evaluated using three statistical indicators: the Mean Square Error (MSE), Mean Error (ME) and Mean Absolute Error (MAE). The results show a better performance of the neural model compared to the support vector machine.

Copyright © 2019 Institute of Advanced Engineering and Science.

All rights reserved.

### Corresponding Author:

Soukaina Barhmi,  
Laboratory of High Energy Physics-Modeling and Simulations,  
Rabat, Morocco.  
Email: soukainabarhmi90@gmailcom

## 1. INTRODUCTION

Wind power is nowadays one of the predominant alternative energy sources since the energy produced by wind is clean, also it helps in reducing global warming and environmental pollution because it doesn't emit dangerous emissions (e.g., those produced by fossil fuel power stations that cause several human health issues). In this sense, the accurate forecast of wind speed is important for the safety of renewable energy utilization. At the present time, Morocco is the largest producer of wind power in Africa since it benefits from great solar and wind energy potential [1] because of its key geographic location [2]. In recent years, many methods have been developed to forecast wind speed. Classical and hybrid models [3] were subject to many works. In [4] presented a hybrid model that consisted of the Ensemble Empirical Mode Decomposition (EEMD) and the Genetic Algorithm- Backpropagation (GA-BP) neural network and compared their performance to Empirical Mode Decomposition (EMD) and the Genetic Algorithm-Backpropagation (GA-BP), traditional Genetic Algorithm-Backpropagation (GA-BP) and Wavelet Neural Network method (WNN) and The best model found was the hybrid EEMD and GA-BP neural network method (6.82% for MAPE and 0.59 for RMSE). On the other hand, in [5] designed a new hybrid model KF-ANN based on the artificial neural network (ANN) and Kalman filter (KF) and compared to Autoregressive Integrated Moving Average (ARIMA), Artificial Neural Network (ANN), hybrid ARIMA-KF models, the results showed that all these models were effective but the MAPE values indicated that the hybrid KF-ANN model was the most effective. In [6] exploited ANN and compared with persistence and ARIMA, the best model was a backpropagation network with one hidden layer and nine neurons (0.18 m/s for MAE and 0.05m<sup>2</sup>/s<sup>2</sup> for MSE). In [7] applied the evolutionary-SVM algorithm for short-term wind speed prediction in

Journal homepage: <http://iaescore.com/online/index.php/IJAI>

wind turbines of Spanish wind farm. In [8] presented a combination method PSO-BP neural network and the results indicate that the proposed method is effective for the wind speed prediction more than the basic back propagation neural network and ARIMA model.

Artificial neural networks based methods [9-10] prove to be the best for nonlinear problems due to its capacity of generalization and fitting any kind of functions which makes it suitable for any random variable based process in particular forecasting the renewable energy resources. Due to their nonlinear nature, artificial neural networks are used as part of the many forecasting hybrid model and we choose the SVM because it is shown to be efficient in addressing nonlinear input-output mapping [11]. On the other hand, the simplicity and flexibility of GA and PSO justify its use to optimize the NN connection weights and SVM parameters during the training phase. In this study, we find based on the Ordinary Least Squares (OLS) analysis four inputs influencing the wind speed, namely, direction, humidity, temperature and past wind speed. The data handled are the hourly meteorological satellite data from 2011 to 2013 (3 years) provided by the Research Institute of Solar Energy and New Energy (IRESEN). The region studied is Tangier located in the south Mediterranean ocean (located in the latitude 35.800, longitude -5.840 and altitude 8 in north Morocco). Afterwards, we apply the proposed hybrid models SVM-GA, SVM-PSO, NN-GA and NN-PSO to capture the relationship between the aforementioned inputs and the wind speed and to verify the performance of these models for the prediction of hourly wind speed in Tangier. The remainder of this article is organized as follows: Section 2 discusses in details the NN, SVM, GA and PSO. In section 3 experimental results analysis are given in detail, and the last section concludes this article.

## 2. METHODOLOGY

In this section, the methodologies and algorithms used for prediction are presented. The first subsection will be dedicated to technique used for analyzing the inputs. The remainder subsections will handle the Support Vector Machines, Artificial Neural Network and defining each algorithms used, namely, Genetic Algorithm and Particle Swarm Optimization.

### 2.1. The ordinary least squares (OLS)

The ordinary least squares (OLS) is a statistical method that fits a model using all of the independent variables. It estimates the linear regression coefficients by minimizing the vertical squared distance between the regression line and the data points [12-13]. In our models, no constant will be used because it affects the results since it has no meaningful physical interpretation.

### 2.2. Support vector machines

Support Vector Machines (SVMs) [14-19] are supervised learning models commonly used for classification problems. Given a set of data, the SVM will attempt to separate the two classes using a hyperplane, which is a subspace one dimension less than the ambient space. Various mathematical techniques such as dual quadratic programming and non-linear kernel transformations can be used to maximize the margin between the classes of data. A hard-margin SVM will enforce the principle that the data must be separable. However, it may not be possible to achieve perfect separation of the data in some cases. Therefore, a soft-margin SVM classifier can be used to maximize the hyperplane margin while allowing some misclassifications. In our approach, we adopt  $\varepsilon$ -Support Vector Regression ( $\varepsilon$ -SVR), the basic principle of SVM for regression is to map the data into a high dimensional feature space via nonlinear mapping, after which a linear regression is performed in this feature space. The regression formula can be expressed as:

$$f(x) = \sum_{i=1}^N w_i \Phi_i(x) + b \quad (1)$$

where  $\{\Phi_i(x)\}_{i=1}^n$  are named features and  $b$  is the bias term. The coefficients  $\{w_i\}_{i=1}^n$  can be obtained from the data by optimizing the following quadratic programming problem:

$$\begin{aligned} & \text{Minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ & \text{Subject to } \begin{cases} y_i - \langle w, \Phi(x_i) \rangle - b \leq \varepsilon + \xi_i \\ \langle w, \Phi(x_i) \rangle - y_i + b \leq \varepsilon + \xi_i^* \\ \xi_i \geq 0, \xi_i^* \geq 0, i = 1, \dots, n \end{cases} \end{aligned} \quad (2)$$

where  $\xi_i^*$  is a slack variable and  $C$  is a positive constant determining the trade off between the flatness of  $f$  and the amount up to which deviations larger than  $\varepsilon$  are tolerated. Here, by solving the optimization problem through the introduction of Lagrange multipliers and optimality constraints:

---

*Hourly wind speed forecasting based on Support Vector Machine and... (Soukaina Barhmi)*

$$f(x, \alpha, \alpha^*) = \sum_{i=1}^D (\alpha_i - \alpha_i^*) k(x_i, x) + b \quad (3)$$

where  $\alpha_i$  and  $\alpha_i^*$  are called Lagrange multipliers. They are obtained by maximizing the dual function of (2) as follows:

$$\begin{aligned} & \text{Maximize } \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(x_i, x_j) \\ & \text{Subject to } \begin{cases} \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \\ 0 \leq \alpha_i, \alpha_i^* \leq C \end{cases} \end{aligned} \quad (4)$$

For  $k(x_i, x_j)$ , it denotes the kernel function equal to the inner product of two vectors  $x_i$  and  $x_j$  in the feature space  $\Phi(x_i)$  and  $\Phi(x_j)$ , given by  $k(x_i, x_j) = \Phi(x_i) \cdot \Phi(x_j)$ . Any function that satisfies Mercer's condition [20] can be used as the kernel function and hence it depends on the problem at hand. In our approach, we opted Radial Basis Function (RBF) as the kernel function of SVR model with good analyticity and arbitrary rank derivatives existence, good performance under general smoothness assumptions. Further details concerning the  $\varepsilon$ -SVR are presented in [21] and the expression of radial basis function is.

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (5)$$

### 2.3. Neural network

Neural network is the modeling and prediction tool [22]. The network forms by connecting the output of certain neurons to the input of other neurons through synaptic weights used to store the knowledge. Any layer that is formed between the input layer and the output layer is called hidden layer. The output of the neural network is given by the following equation:

$$Y = \frac{1}{1 + \exp(-z)} \quad (z = \sum_{i=1}^N X_i * W_i + b_i) \quad (6)$$

where  $X_i$  are the inputs,  $W_i$  the weights and  $b_i$  are the biases. It should be noted that the number of neurons in the input/output layer is related to the variables that are dealt with.

The input/target datasets were divided randomly into two subsets. The first set is to train the neural network using the learning algorithm in order to find the optimal weights that minimize the Mean Square Error (MSE). The second one is to validate the already trained model [23].

### 2.4. Genetic algorithm

A genetic algorithm is a stochastic general search method [24-27]. It was proposed by John Holland et al and widely applied in various forecasting and optimization fields. The genetic algorithm selects the individuals of random from the current population and these individuals use to produce the children for the next generation and repeatedly modifies a population of individual solutions. At each step, over successive generations, the population "evolves" toward an optimal solution.

### 2.5. Particle swarm optimization

The PSO is a new evolutionary computation method [28-32]. The system is initialized with a population of random solutions and searches for optima by updating generations. In PSO, the particles fly through the problem space by following the current optimum particles. The basic mathematical expressions of PSO are as follows:

$$\begin{aligned} v_{sd}(t+1) &= v_{sd}(t) + c_1(t)r_1(t)(p_{sd}(t) - x_{sd}(t)) + c_2(t)r_2(t)(p_{gd}(t) - x_{sd}(t)) \\ x_{sd}(t+1) &= x_{sd}(t) + v_{sd}(t+1) \end{aligned} \quad (7)$$

where  $t$  is iteration number; random variable of  $r_1$  and  $r_2$  obey uniform distribution of the interval (0, 1);  $c_1(t)$  and  $c_2(t)$  are acceleration constants;  $x_{sd}(t)$  is the position of the particle  $S$  in the  $t$  iterations;  $p_{sd}(t)$  is the optimal location that all particles in the species search for in the  $t$  iterations.

### 3. RESULTS AND DISCUSSION

In this study, we tried to fit 5 inputs to wind speed, namely, Direction, past wind speed, Humidity, pressure and temperature, selecting the predictors is a necessary phase. In order to do that, some assumptions must be verified. The first one is to check the correlation between the dependent variable and the predictors in order to determine the degree of dependence of the predicted values to the predictors. A p-value of 0.05 or more means that it would not be rejected at the 5% level [33] and their value are given in Table 1. Consequently, you should consider removing pressure from the model of Tangier.

The second one is to remove the multicollinearity between independent variables. Finally, four explanatory independent variable most influencing wind speed are direction, humidity, temperature and past wind speed to develop the proposed hybrid models [34]. The SVM model is used to perform wind speed forecasting. The parameters, cost (C), gamma (g), epsilon (e) has a significant impact on the forecasting performance, their values are listed in Tables 2 and 3. The Genetic Algorithm (GA) and Particle Swarm Optimization (PSO) are used to optimize the above three parameters by minimizing the MSE value. To evaluate the proposed hybrid approach and to determine quantitatively the best model, the coefficient of determination ( $R^2$ ) and three statistical indices are utilized to measure the forecasting accuracy. These indices are the Mean Error (ME), Mean Square Error (MSE) and Mean Absolute Error (MAE) between the predicted and the actual values of wind speed, then the errors are defined as:

Table 1. Statistical analysis obtained from the OLS at Tangier station

| Parameter       | Coefficient value | Standard error | t-value  | P-value |
|-----------------|-------------------|----------------|----------|---------|
| Direction       | -0.0000987504     | 0.0000144857   | -6.8171  | 0.0000  |
| Humidity        | -0.00127188       | 0.0000830857   | -15.308  | 0.0000  |
| Temperature     | 0.000407384       | 0.0000214706   | 18.9741  | 0.0000  |
| Past wind speed | 0.990865          | 0.00101643     | 974.848  | 0.0000  |
| Pressure        | -0.000060839      | 0.0000722721   | -0.84181 | 0.0039  |

Table 2. Optimal parameters of SVM-GA

| City    | Values of parameters |          |              |
|---------|----------------------|----------|--------------|
| Tangier | Cost(C)              | Gamma(g) | Epsilon(e)   |
|         | 15.2749              | 3.144599 | 0.0006444887 |

Table 3. Optimal parameters of SVM-PSO

| City    | Values of parameters |          |              |
|---------|----------------------|----------|--------------|
| Tangier | Cost(C)              | Gamma(g) | Epsilon(e)   |
|         | 10.8001              | 0.167599 | 0.0003052174 |

$$ME = \frac{1}{N} \sum_{i=1}^N (T_i - Y_i) \quad (8)$$

$$MSE = \frac{1}{N} \sum_{i=1}^N (T_i - Y_i)^2 \quad (9)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |T_i - Y_i| \quad (10)$$

$$R^2 = 1 - \frac{\sum_{i=1}^N (T_i - Y_i)^2}{\sum_{i=1}^N (T_i - T)^2} \quad (11)$$

where  $T_i$  are the targets,  $T$  is the mean of all target values,  $Y_i$  are the neural network outputs and  $N$  is the number of samples.

Table 4 shows the errors of the four hybrid models. The coefficient of determination ( $R$ -squared), also called goodness of fit, shows a very good quality of prediction for the NN-GA and NN-PSO models since the values are very close to 1. Additionally, the MSE, ME and MAE show a relatively low error in the models which can be explained by the fact that the models describe most of the data, but the NN-GA is the best suited for the prediction with  $MSE = 4.71 \cdot 10^{-4} m^2/s^2$  with 3 neurons in the hidden layer. On the other hand, the  $R$ -squared of SVM-GA and SVM-PSO models is very close to 0. However, the MSE, ME and MAE show a relatively low error in the models which can be explained by the fact that the models describe most of the data but not with a high accuracy. Furthermore, Figure 1 and Figure 2 show a large gap between actual and predicted SVM-GA and SVM-PSO values, but in Figure 3 and Figure 4 the two curves are almost confused for NN-GA and NN-PSO. According to these results, we conclude that the NN-GA is able to predict the wind speed with a higher accuracy than the other models.

Table 4. Performance evaluation of different models for forecasts in Tangier

| Methods | MSE(m <sup>2</sup> /s <sup>2</sup> ) | ME(m/s)                | MAE(m/s)               | R-squared |
|---------|--------------------------------------|------------------------|------------------------|-----------|
| SVMGA   | $2.42 \cdot 10^{-2}$                 | $6 \cdot 10^{-3}$      | 0.1238                 | 0.1947    |
| SVMPSO  | $2.60 \cdot 10^{-2}$                 | $1.31 \cdot 10^{-2}$   | 0.1281                 | 0.1355    |
| NNPSO   | $5.27 \cdot 10^{-4}$                 | $4.0765 \cdot 10^{-5}$ | 0.0124                 | 0.9881    |
| NNGA    | $4.71 \cdot 10^{-4}$                 | 0.0169                 | $9.0892 \cdot 10^{-4}$ | 0.9834    |

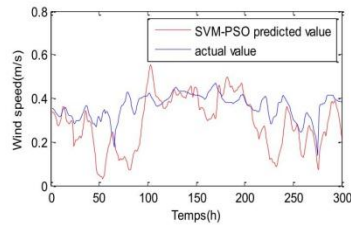


Figure 1. The actual/forecast wind speed by SVM-PSO

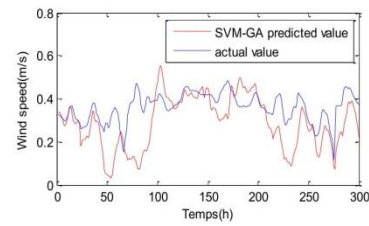


Figure 2. The actual/ forecast wind speed by SVM-GA

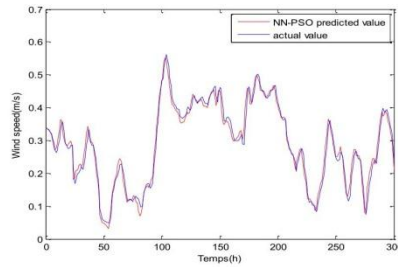


Figure 3. The actual/forecast wind speed by NN-PSO

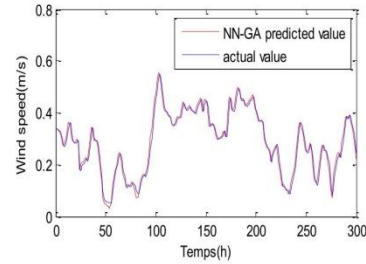


Figure 4. The actual/forecast wind speed by NN-GA

#### 4. CONCLUSION

The objective of this study was the prediction of wind speed using four hybrid models. In the first, we find based on the OLS analysis four inputs influencing the wind speed, namely, direction, humidity, temperature and past wind speed. Afterwards, we apply the proposed hybrid models NN-GA, SVM-GA, SVM-PSO and NN-PSO to capture the relationship between the aforementioned inputs and the wind speed. The performance of these models is evaluated using R-squared and three statistical indicators: the Mean Square Error (MSE), Mean Error (ME) and Mean Absolute Error (MAE). The conclusion made is that the SVM-GA and SVM-PSO are not suitable for estimating the behavior of wind speed in the region of Tangier using the selected inputs. On the other hand, the nonlinear method of neural networks tuned by genetic algorithm using the same inputs gives excellent results which indicate its high accuracy in the prediction of wind speed.

#### REFERENCES

- [1] Kingdom of Morocco (2009). Moroccan Regions 2010. [http://www.hcp.ma/downloads/Maroc-en-chiffres\\_t13053.html](http://www.hcp.ma/downloads/Maroc-en-chiffres_t13053.html). Last checked on Dec 29, 2017.
- [2] Kingdom of Morocco (2011). Moroccan Investment Development Agency. <http://invest.gov.ma/>. Last checked on Dec 29, 2017.
- [3] J. S. Jung, R.P. Broadwater, "Current status and future advances for wind speed and power forecasting," *Renew Sustain Energy Rev*, vol. 31, pp. 762-777, 2014.
- [4] S. Wang, N. Zhang, L. Wu, and Y. Wang, "Wind speed forecasting based on the hybrid ensemble empirical mode decomposition and GA-BP neural network method," *Renewable Energy*, vol. 94, pp. 629-636, 2016.

- [5] O. B. Shukur, and M. H. Lee, "Daily wind speed forecasting through hybrid KF-ANN model based on ARIMA," *Renewable Energy*, vol. 76, pp. 637-647, 2015.
- [6] J. C. Palomares-Salas, A. Agüera-Pérez, J. J. G. de la Rosa, and A. Moreno-Munoz, "A novel neural network method for wind speed forecasting using exogenous measurements from agriculture stations," *Measurement*, vol. 55, pp. 295-304, September 2014.
- [7] Salcedo-Sanz Sancho, Ortiz-García Emilio G, Pérez-Bellido Ángel M, Portilla-Figueras Antonio, Prieto Luis. Short term wind speed prediction based on evolutionary support vector regression algorithms. *Expert Syst Appl* 2011; 38(4):4052-7.
- [8] Ren Chao, An Ning, Wang Jianzhou, Li Lian, Hu Bin, Shang Duo. Optimal parameters selection for BP neural network based on particle swarm optimization: a case study of wind speed forecasting. *Knowl – Based Syst* 2014; 56:226-39.
- [9] H.N. Wang, Y.M. Cui, R. Li, L.Y. Zhang, H. Han, "Solar flare forecasting model supported with artificial neural network techniques," *Advances in space research*, vol. 42, pp. 1464-1468, 3 November 2008.
- [10] A. Koca, H. F. Oztup, Y. Varol, G. O. Koca, "Estimation of solar radiation using artificial neural networks with different input parameters for Mediterranean region of Anatolia in Turkey," *Expert systems with applications*, vol. 38, pp. 8756-8762, 2011.
- [11] E. Ghysels, and H. Qian, "Estimating MIDAS Regressions via OLS with a Polynomial Parameter Profiling," *Econometrics and Statistics.Cortes*, vol. 9, pp. 1-16, 2018.
- [12] D. G. Kleinbaum, L. L. Kupper, and K. E. Muller, "Applied Regression Analysis and Other Multi-variable Methods," PWS Publishing Co., Boston, MA, USA, 1988.
- [13] C. Vapnik V N, "Support vector networks. Machine Learning," vol. 20, pp. 273-297, 1995.
- [14] H.M. Pei, Y. Y Chen, Y.K. Wu, " Laplacian total margin support vector machine based on within-class scatter," *Knowledge-Based Systems*, vol. 119, pp. 152-165, 2017.
- [15] Ch. Sudheer, R. Maheswaran,B. K. Panigrahi, "A hybrid SVM-PSO model for forecasting monthly streamflow," *Neural Computing and Applications*, vol. 24(6), pp. 1381-1389, 2014.
- [16] SC. Huang, B. Chen, " Automatic moving object extraction through a real-world variable-bandwidth network for traffic monitoring system," *IEEE Transactions on Industrial Electronics*, vol. 61(4), pp. 2099-2112, 2014.
- [17] JX. Wu, "Efficient HIK SVM learning for Image Classification," *IEEE Transactions on Image Processing*, vol. 21(10), pp. 4442-4453, 2012.
- [18] F. Zhu, N. Ye, W. Yu, "Boundary detection and sample reduction for one-class support vector machine," *Neurocomputing*, vol. 123, pp. 174-184, 2014.
- [19] J.X. Du, D.S.Huang, X.F. Wang, X. Gu, "Shape recognition based on neural networks trained by differential evolutionary algorithm," *Neurocomputing*, vol. 70(4-6), pp. 896-903, 2007.
- [20] D. Basak, S. Pal, "Patranabis DC Support vector regression," *Neural Inf Process Lett Rev*, vol. 11, pp. 203-224, 2007.
- [21] AJ. Smola, B. Schölkopf, "A tutorial on support vector regression," *Stat Comput*, vol. 14, pp. 199-222, 2004.
- [22] SA. Kalogirou, "Artificial neural networks in renewable energy systems applications: a review" *Renew Sust Energy Rev*, vol. 5(4), pp. 373-401, 2001.
- [23] B. Curry, and D. E. Rumelhart, "An MSnet: A Neural Network which Classifies Mass Spectra," *Tetrahedron Computer Methodology*, vol.3, pp. 213-237, 1990.
- [24] D. Liu, D. Niu, M. Xing, Q. Nie, " Day-ahead price forecast with genetic-algorithm-optimized support vector machines based on garch error calibration," *Automation of Electric Power Systems*, pp. 31-40, 2007.
- [25] S. Marsili-Libelli, P. Alba. Adaptive mutation in genetic algorithms. *Soft Comput* 2000; 4:76-80.
- [26] JE. Baker. Reducing bias and inefficiency in the selection algorithm, genetic algorithms and their applications. *Proc of the Second Int Conf on Gen Alg* 1987:14-21.
- [27] JR. Fernández, JA. López-Campos, A. Segade, JA. Vilán. A genetic algorithm for the characterization of hyperelastic materials. *Appl Math Comput* 2018; 329:239-50.
- [28] R. Jursa, K. Rohrig, "Short-term wind power forecasting using evolutionary algorithms for the automated specification of artificial intelligence models," *International Journal of Forecasting*, vol.24, pp. 694-709, 2008.
- [29] J. Kennedy, R. Eberhart, Particle swarm optimization, *Neural Networks*, 1995.In: Proceedings. IEEE International Conference on, November/December 1995, 4, Perth, WA, vol. 1944, pp. 1942-1948.
- [30] R. Eberhart, J. Kennedy, A new optimizer using particle swarm theory, *MicroMachine and Human Science*, 1995. MHS '95, in: Proceedings of the Sixth International Symposium on, 4-6 October 1995, Nagoya, Japan, pp. 39-43.
- [31] R.C. Eberhart,Y.H. Shi, Particle swarm optimization: Developments,applications and resources, in: Proceedings of the 2001 Congress on Evolutionary Computation, Vols 1 and 2, 27-30 May 2001, Seoul, South Korea, pp. 81-86.
- [32] R.C. Eberhart, Y.H. ve Shi, Special issue on Particle Swarm Optimization, in: *IEEE Transactions on Evolutionary Computation*, June 2004. pp. 201-203
- [33] C. Dougherty, Introduction to Econometrics, Oxford University Press, USA, 2007.
- [34] S. Barhmi, O. El fatni, I. Belhaj, " Forecasting of wind speed using Multiple Linear Regression and Artificial Neural Networks", *Energy Syst* (2019). <https://doi.org/10.1007/s12667-019-00338-y>.

### **Contribution III**

A direct normal irradiation forecasting model based on artificial  
neural networks

---

#### *Résumé*

Dans ce travail, on a étudié la prévision de l'irradiation horaire normale directe notée en anglais (DNI) en utilisant les réseaux de neurones artificiels (RNA). Les données utilisées sont des données satellitaires horaires pour la région d'Ouarzazate située dans la région Sud-Ouest du bassin méditerranéen. La base de données est un total de 19 ans. Le type de réseau utilisé est un perceptron multicouche feed-forward avec dix paramètres d'entrée. Les résultats montrent une très bonne précision pour les quatre saisons des années prévues.

## A direct normal irradiation forecasting model based on artificial neural networks

I. Belhaj, O. El Fatni, S. Barhmi and E.H. Saidi

Laboratory of High Energy Physics, Modelling and Simulations  
Faculty of Sciences Rabat, Mohammed V University,  
4 Avenue Ibn Battouta B.P. 1014 RP, Rabat, Morocco

(reçu le 10 Novembre 2015 – accepté le 29 Mars 2016)

**Abstract** - We investigate the forecasting of the hourly Direct Normal Irradiation (DNI) using Artificial Neural Networks (ANN). The data used are hourly satellite data for the region of Ouarzazate in the South West Mediterranean basin region. The database is a total of 19 years. A feed-forward multilayer Perceptron is used with ten input parameters. The results show a very good accuracy for the four seasons of the years forecasted.

**Résumé** - Nous étudions la prédiction de l'irradiation horaire normale directe (DNI) en utilisant les réseaux de neurones artificiels (RNA). Les données utilisées sont les données satellitaires horaires pour la région de Ouarzazate dans la région Sud-Ouest du bassin méditerranéen. La base de données est un total de 19 ans. Un Perceptron multicouche feed-forward est utilisé avec dix paramètres d'entrée. Les résultats montrent une très bonne précision pour les quatre saisons des années prévues.

**Keywords:** Forecasting - Artificial neural networks - Direct normal irradiation (DNI) – Backpropagation – Perceptron.

### 1. INTRODUCTION

Forecasting solar irradiation is of big importance due to applications in electricity generation, planning, dispatch and water heating energy. Variability of electricity produced from unstable renewable energy resources leads to fluctuations that affect the power grid and the limited storage hour's capacity; therefore forecasting of solar irradiation proves to be a very important solution.

Direct Normal Irradiation (DNI) is the main component of Concentrated Solar Power (CSP), Solar Thermal (ST) systems and Concentrated Photovoltaic's (CPV) and present a challenging task to foresee, since its high fluctuations caused by many parameters such as temperature, humidity and cloudiness that can affect it in a very effective manner; therefore, its forecasting is crucial.

Various energy demand forecasting models exist [1], such as time series [2-4], artificial neural networks [5-8] and Markal and Leap models. In general, forecasting methods can be divided into: statistical approaches, techniques based on cloud and satellite images [9], NWP models and hybrid models [10-12]. For example, an image processing methodology was elaborated to generate deterministic DNI forecasts for horizons ranging from 3 to 15 min ahead called the Sector-ladder method [13].

A hybrid model of a time delay neural network (TDNN) and Arma was used for predicting the hourly solar radiation which was then compared to ANN and Arma [14]. A univariate Dynamic Harmonic Regression (DHR) model was proposed for short-term (1-24h) solar irradiation forecasting using GHI and DNI from ground-based weather stations located in Spain [15].

A hybrid Arma/ANN model was tested to predict hourly global radiation using Numerical Weather Prediction (NWP) model database [16]. An exponential smoothing

state space (ESSS) model was used to forecast high-resolution solar irradiance time series [17].

The solar forecasting methods are linked to the horizon of forecasting and its purpose [18]. For short-term and medium-term forecasting, Artificial Neural networks techniques are known to be the best method to give accurate results in comparison to conventional methods [18, 19].

In this paper, we study forecasting hourly Direct Normal Irradiation (DNI) using Artificial Neural Network (ANN). The data used are Ouarzazate's region in south-west Mediterranean basin, (south-east Morocco). The importance of Ouarzazate's platform for Africa is similar to the one of Almeria's platform for Europe since it has the same purpose for the Mediterranean region which is to exploit the high CSP potential with high solar incidence and proximity to Europe to export energy to region countries.

Ouarzazate is also part of the Moroccan solar plan of electrical energy production from renewable energy. The data are calculated from Meteosat MSG and MFG satellite data (EUMETSAT) and from atmospheric data (ECMWF and NOAA) by Solar GIS method. Since the forecasting accuracy of ANN models is dependent on input parameter combinations, training algorithm and architecture configurations [20], ten meteorological parameters are used as inputs based on a statistical test selection, namely: sun altitude angle, Sun azimuth angle, atmospheric pressure, relative humidity, air temperature at 2 m, wind speed at 10 m, wind direction, dew point temperature, wet bulb temperature and Direct Normal Irradiation (DNI) at a time  $t$ .

The forecasting model used is a Feed-forward Multilayer Perceptron (FFMLP) with a Back-Propagation Learning (BPL) algorithm. Ten ANN forecasting models are compared.

## 2. DATA

The Direct Normal Irradiation (DNI) and the meteorological data used are the hourly data provided by MASEN (Moroccan Agency for Solar ENergy) calculated from Meteosat MSG and MFG satellite data (Eumetsat) and from atmospheric data (ECMWF and NOAA) by Solar GIS method for the region of Ouarzazate. The spacial resolution for solar radiation data is  $250 \text{ m} \times 250 \text{ m}$  and for meteorological data is  $1000 \text{ m} \times 1000 \text{ m}$ , which reveals the quality of the data used. The region of Ouarzazate located in the latitude  $31^{\circ}00'83''$  and longitude  $-06^{\circ}86'27''$  in South Morocco is a strategic region for the Moroccan Solar Plan (MSP) that aims to produce 2000 MW by 2020, making of Morocco one of the first countries in the world in terms of producing electricity from renewable energy resources. The entire hourly database used represents 19 years of data starting from 1<sup>st</sup> January 1994 to 31<sup>st</sup> December 2012.

## 3. METHODS

In this section, we present the methodology used for forecasting the hourly Direct Normal Irradiation (DNI). The data are pre-processed by excluding the 29<sup>th</sup> February day from each year to avoid the leap years and homogenize the data. The period between 07:30 AM to 17:30 is considered as it represents the daytime solar duration. The method used here is the artificial neural networks ANN; this is a method of artificial intelligence (AI) very used in finance, process analysis and solar forecasting.

It consists of three layers: the input layer, the hidden layer and the output layer. Each layer has a number of nodes called 'neurons'. Every layer is connected to the following one by its nodes. Each connection is represented with the multiplication of the node by

a random weight all summed with a bias in a transfer function given by the following equation,

$$y = f \left( \sum_{i=1}^N W_i x_i + b_i \right) \quad (1)$$

where  $f$  is the transfer function,  $W_i$  the random weights assigned to each input,  $N$  the number of the neurons and  $b_i$  the bias. First, the database is divided randomly into 3 parts: training data, test data and validation data as shown in **Table 1**.

**Table 1:** Partition of data

|       | Training | Test | Validation |
|-------|----------|------|------------|
| Years | 1994     | 1995 | 1996       |
|       | 1997     | 1998 | 1999       |
|       | 2000     | 2001 | 2002       |
|       | 2003     | 2004 | 2005       |
|       | 2006     | 2007 | 2008       |
|       | 2009     | 2010 | 2011       |
|       | 2012     |      |            |

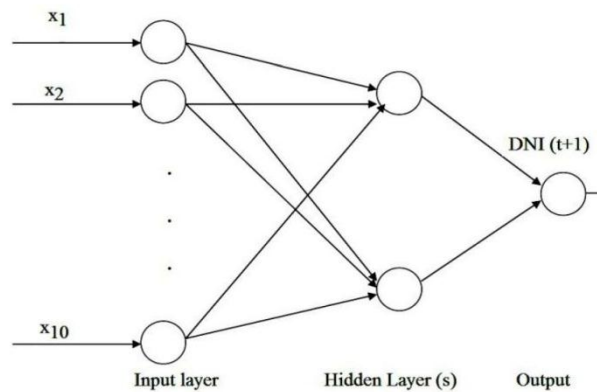
Ten meteorological data are used as inputs, namely: sun altitude angle, Sun azimuth angle, atmospheric pressure, relative humidity, air temperature at 2 m, wind speed at 10 m, wind direction, dew point temperature, wet bulb temperature and Direct Normal Irradiation (DNI) at a time  $t$ .

These inputs are then normalized in the range  $[0, 1]$  using the equation

$$x_{\text{normalized}} = \frac{x_{\text{actual}} - x_{\text{min}}}{x_{\text{max}} - x_{\text{min}}} \quad (2)$$

to reduce the discrepancies between the data since it's the domain of definition of the sigmoid function used as a transfer function in this work. The sigmoid function is given by the following equation

$$f(x) = \frac{1}{1 + e^{-x}} \quad (3)$$



**Fig. 1:** ANN Architecture

Figure 1 shows the architecture of the feed-forward multilayer perceptron (MLP) used, which consists of one hidden layer with one neuron and a bias. The Back-propagation algorithm is used for training the forecasting model and the cross validation method is used for validation. DNI forecasts are then evaluated by computing the Root Mean Squared Error (RMSE) and Mean Squared Error (MSE) given by

$$\text{RMSE} = \left( \frac{1}{n} \sum_{i=1}^n (t_i - y_i)^2 \right)^{\frac{1}{2}} \quad (4)$$

And

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (t_i - y_i)^2 \quad (5)$$

where  $n$  is the number of samples,  $t_i$  the targets and  $y_i$  the neural network outputs.

#### 4. RESULTS

In this study, a feed-forward MLP is used with ten inputs to forecast the hourly normal irradiation (DNI). The database consists of 19 years of data where the period from 07:30 to 17:30 is considered to omit the period of darkness. As mentioned earlier, many studies have been conducted to forecast solar irradiation.

It was shown that ANN gave the best forecasting skill in one hour horizon for different regions [18]. In our work, the optimal ANN architecture selected in this study is 10-1-1, which consists of the least number of neurons and hidden layers possible.

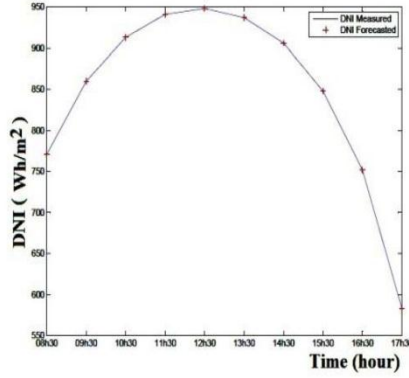


Fig. 2a- 15 April 2002

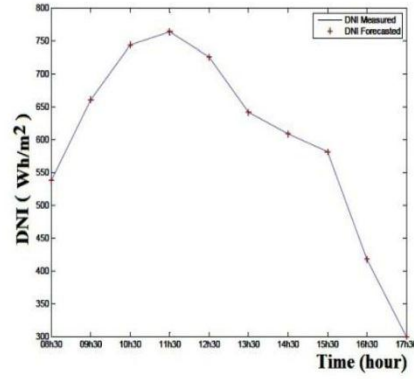


Fig. 2b- 15 July 2005

**Table 2** shows the metric results of the validation set for the forecasting models tested. As can be seen from the table, the least error found is  $\text{RMSE} = 0.3317$  while we can notice that exceeding two hidden layers yields to a filter behaviour. The value of the error found shows the good accuracy of the model since the DNI in the validation set ranges from 0 to 1049  $\text{Wh/m}^2$ . The determination coefficient  $R^2$  is then computed by the following equation

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - t_i)^2}{\sum_{i=1}^n (t_i - \bar{t})^2} \quad (6)$$

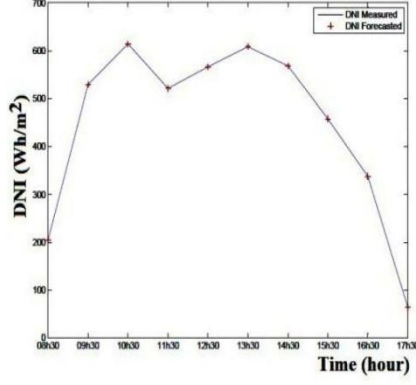


Fig. 2c- 15 October 2008

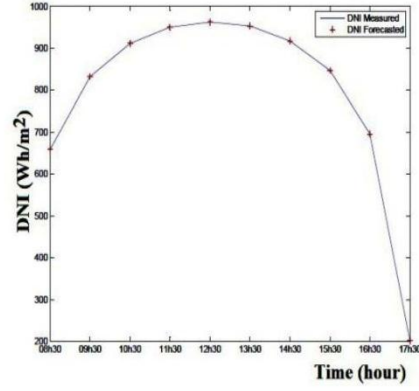


Fig. 2d- 15 January 2011

Fig. 2: Comparison between forecasted DNI and actual DNI for four random days

where  $y_i$  the forecasts from ANN,  $t_i$  measured values and  $\bar{t}$  the average of DNI. The value shows a perfect description of the data with  $R^2 = 1:000$ . Figure 2 displays the results of four random days: 15<sup>th</sup> April 2002 Fig. 2a-, 15<sup>th</sup> July 2005 Fig. 2b-, 15<sup>th</sup> October 2008 Fig. 2c- and 15<sup>th</sup> January 2011 Fig. 2d-.

The nRMSE, normalized RMSE, is used to measure the error in each season to view the accuracy of the forecasting model used with the following equation

$$nRMSE = \frac{\frac{1}{N} \left( \sum_{i=1}^N (t_i - y_i)^2 \right)^{1/2}}{\frac{1}{N} \sum_{i=1}^N y_i}$$

As listed in **Table 3**, the normalized root mean square errors for the four seasons don't exceed 1%. This could be explained by the fact that the hours of sunshine and sunset induce global errors in the database as a whole but when the error is normalized, their influence is negligible as predicted since most of these errors appear for values are less than 5 Wh/m<sup>2</sup>.

Figure 3 illustrates how the forecasts are in very good agreement with the actual DNI values for the four seasons: summer (Fig. 3a-), Winter (Fig. 3b-), Spring (Fig. 3c-) and Autumn (Fig. 3d-).

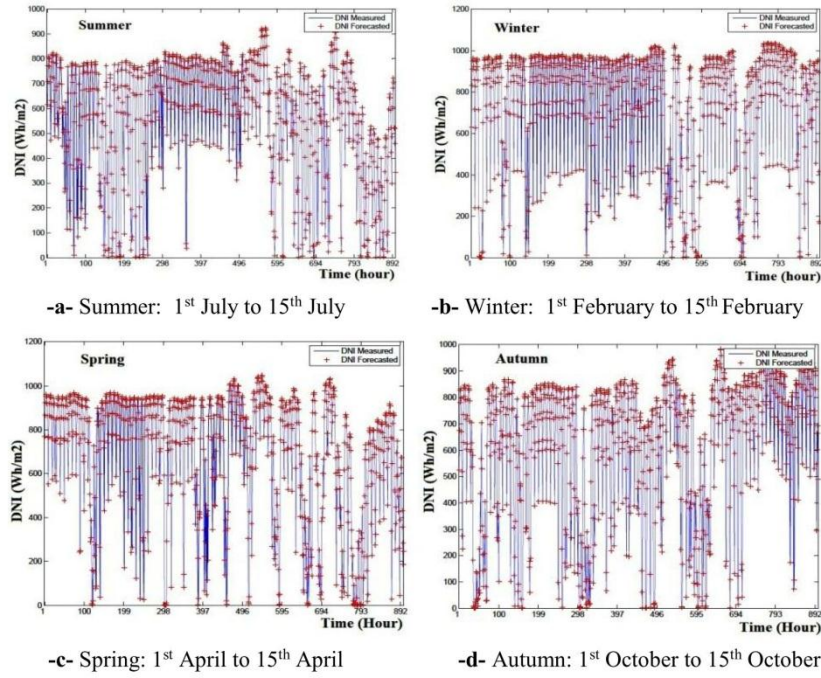


Fig. 3: Comparison between forecasted DNI and actual DNI for four seasons

**Table 2:** Statistical metrics for the models tested  
 Bold values indicate the model selected using the validation dataset

| Network<br>N° | N° of hidden<br>layers | N° of<br>neurons | MSE     | RMSE<br>(Wh/m <sup>2</sup> ) |
|---------------|------------------------|------------------|---------|------------------------------|
| 1             | 1                      | 1                | 0.11000 | 0.3317                       |
| 2             | 1                      | 2                | 0.11000 | 0.3317                       |
| 3             | 1                      | 3                | 0.11000 | 0.3317                       |
| 4             | 2                      | 2                | 0.11000 | 0.3317                       |
| 5             | 2                      | 3                | 0.11000 | 0.3317                       |
| 6             | 2                      | 4                | 0.11000 | 0.3317                       |
| 7             | 3                      | 3                | Filter  | Filter                       |
| 8             | 3                      | 4                | Filter  | Filter                       |
| 9             | 3                      | 5                | Filter  | Filter                       |
| 10            | 3                      | 6                | Filter  | Filter                       |

**Table 3:** Statistical metrics for seasons from the forecasting model chosen

| Season | RMSE (Wh/m <sup>2</sup> ) | MSE (Wh/m <sup>2</sup> ) | nRMSE (%)  |
|--------|---------------------------|--------------------------|------------|
| Winter | 0.3009                    | 0.0906                   | 4.3058e-04 |
| Spring | 0.3364                    | 0.1131                   | 5.0385e-04 |
| Summer | 0.3156                    | 0.0996                   | 5.6977e-04 |
| Autumn | 0.3490                    | 0.1218                   | 5.8206e-04 |

## 5. CONCLUSION

Forecasting solar irradiation has known many improvements in terms of experimental methodologies and results accuracy. From time series, to artificial techniques to hybrid models, satellite imagery and cloud tracking, depending on the time-scale needed and the forecasting purposes, results differ.

However, the results of the basic methods can be further improved by exploring different new sets of independent variables that affect the dependent variable.

In this study, DNI forecasting using the artificial neural networks (ANN) is explored with a new different set of independent variables. Ten exogenous parameters were used based on a statistical P-value selection.

A feed-forward MLP is used with a back propagation algorithm. The results show that ANN gives very accurate results using these inputs demonstrated by the very low mean squared error. The model reveals also a good forecasting accuracy in each of the four seasons which can be shown by an MSE equal to 0.0906 Wh/m<sup>2</sup> in Winter, 0.1131 Wh/m<sup>2</sup> in Spring, 0.0996 Wh/m<sup>2</sup> in Summer and 0.1218 Wh/m<sup>2</sup> in Autumn.

Artificial neural networks prove to be a very good tool for forecasting although, the choice of independent variables and the huge database (69 350 data in our case) play an important role in the enhancement and achievement of good results which is still an ill-defined subject of matter in the artificial intelligence field.

Further improvement shall come up in next works by using optimizing algorithms, multiple regression and hybrid models with a comparison between these results using the same inputs.

## ACKNOWLEDGEMENTS

The authors thank the Moroccan agency for solar energy for putting a disposal data.

## REFERENCES

- [1] L. Suganthi, Anand A. Samuel., '*Energy Models for Demand Forecasting- A Review*', Renewable & Sustainable Energy Reviews, Vol. 16, N°2, pp. 1223 - 1240, 2012.
- [2] C. Voyant, P. Haurant, M. Muselli, C. Paoli, M.L. Nivet, '*Time Series Modelling and Large Scale Global Solar Radiation Forecasting From Geostationary Satellites Data*', Solar Energy, Vol. 102, pp. 131 - 142, 2014.
- [3] G. Reikard, '*Predicting Solar Radiation at High Resolutions: A Comparison of Time Series Forecasts*', Solar Energy, Vol. 83, N°3, pp. 342 - 349, 2009.
- [4] J.G. De Gooijer and R.J. Hyndman, '*25 Years of Time Series Forecasting*', International Journal of Forecasting, Vol. 22, pp. 443 - 473, 2006.
- [5] C. Paoli, C. Voyant, M. Muselli and M.L. Nivet, '*Forecasting of Preprocessed Daily Solar Radiation Time Series Using Neural Networks*', Solar Energy, Vol. 84, N°12, pp. 2146 - 2160, 2010.
- [6] Y. Kashyap, A. Bansal and A.K. Sao, '*Solar Radiation Forecasting with Multiple Parameters Neural Networks*', Renewable and Sustainable Energy Reviews, Vol. 49, pp. 825 - 835, 2015.

- [7] H.N. Wang, Y.M. Cui, R. Li, L.Y. Zhang and H. Han, 'Solar Flare Forecasting Model Supported with Artificial Neural Networks Techniques', *Advances in Space Research*, Vol. 42, N°9, pp. 1464 – 1468, 2008.
- [8] S.X. Chen, H.B. Gooi and M.Q. Wang, 2013. 'Solar Radiation Forecast Based on Fuzzy Logic and Neural Networks', *Renewable Energy*, Vol. 60, N°C, pp. 195 - 201, 2013.
- [9] C. Rigollier, M. Lefèvre and L. Wald, 'The Method Heliosat-2 for Deriving Shortwave Solar Radiation from Satellite Images', *Solar Energy*, Vol. 77, N°2, pp. 159 - 169, 2004.
- [10] J.Z. Wang, H. Jiang, Y.J. Wu and Y. Dong, 'Forecasting Solar Radiation using an Optimized Hybrid Model by Cuckoo Search Algorithm', *Energy*, Vol. 81, N°C, pp. 627 – 644, 2015.
- [11] M. Diagne, M. David, Ph. Lauret, J. Boland and Nicolas Schmutz, 'Review of Solar Irradiance Forecasting Methods and a Proposition for Small-Scale Insular Grids', *Renewable & Sustainable Energy Reviews*, Vol. 27, pp. 65 – 76, 2013.
- [12] Y. Gala, A. Ferandez, J. Diaz and J.R. Dorronsoro, In press, 'Hybrid Machine Learning Forecasting of Solar Radiation Values', *Neurocomputing*, Vol. 176, pp. 48 – 59, 2016.
- [13] S. Quesada-Ruiz, Y. Chu, J. Tovar-Pescador, H.T.C. Pedro and C.F.M Coimbra, 'Cloud Tracking Methodology for Intra-Hour DNI Forecasting', *Solar Energy*, Vol. 102, pp. 267 – 275, 2014.
- [14] Ji Wu and Chee Keong Chan, 'Prediction of Hourly Solar Radiation Using a Novel Hybrid Model of ARMA and TDNN', *Solar Energy*, Vol. 85, N°5, pp. 808 - 817, 2011.
- [15] J.R. Trapero, N. Kourentzes and A. Martin, 'Short-Term Solar Irradiation Forecasting Based on Dynamic Harmonic Regression', *Energy*, Vol. 84, N°2, pp. 289 – 295, 2015.
- [16] C. Voyant, M. Muselli, Ch. Paoli and M.L. Nivet, 'Numerical Weather Prediction (NWP) and Hybrid ARMA/ANN Model to Predict Global Radiation', *Energy*, Vol. 39, N°1, pp. 341 - 355, 2012.
- [17] Z. Dong, D. Yang, T. Reindl and W.M. Walsh, 'Short-Term Solar Irradiance Forecasting using Exponential Smoothing State Space Model', *Energy*, Vol. 55, Issue C, pp. 1104 – 1113, 2013.
- [18] R.H. Inman, H.T.C. Pedro and Carlos F.M. Coimbra, 'Solar Forecasting Methods for Renewable Energy Integration-Review', *Progress in Energy and Combustion Science*, Vol. 39, N°6, pp. 535 - 576, 2013.
- [19] R. Marquez and Carlos F.M. Coimbra, 'Forecasting of Global and Direct Solar Irradiance using Stochastic Learning Methods, Ground Experiments and the NWS Database', *Solar Energy*, Vol. 85, N°5, pp. 746 – 756, 2011.
- [20] A.K. Yadav and S.S. Chandel, 'Solar Radiation Prediction using Artificial Neural Network Techniques: A Review', *Renewable and Sustainable Energy Reviews*, Vol. 33, pp. 772 – 781, 2014.

## Contribution IV

### Two developed methods for predicting hourly direct normal irradiation (DNI)

---

#### *Résumé*

L'irradiation normale directe (DNI) est le principal composant de l'énergie solaire concentrée (CSP), les systèmes solaires thermiques (ST) et le photovoltaïque concentré (CPV) qui sont très utilisés dans de nombreux parcs solaires dans les zones très ensoleillées, mais ce composant présente un gros défi en termes de fortes fluctuations causées par de nombreux paramètres d'où l'importance de la prévision. Cette étude est consacrée au développement de deux méthodes de prévision d'irradiation normale directe (DNI) et la comparaison de leurs performances. La première méthode de prévision est la régression multiple qui permet l'étude d'un système avec un groupe de prédicteurs. La deuxième méthode est celle des réseaux de neurones artificiels, qui est connue comme l'une des meilleures méthodes de prévision du rayonnement solaire à court terme. La région d'étude est Ouarzazate située au sud-ouest du bassin méditerranéen. Les données utilisées sont des données satellitaires de 19 ans d'un historique horaire allant de 07h30 à 17h30. Les entrées des modèles trouvés sont sélectionnées à l'aide d'une étude d'analyse de variance notée ANOVA. L'architecture du modèle neuronal est un Perceptron multicouche de cinq entrées. Les résultats montrent une meilleure performance du modèle neuronal par rapport à la régression multiple.



## Two developed methods for predicting hourly direct normal irradiation (DNI)

I. Belhaj<sup>1\*</sup>, S. Barhmi<sup>1</sup>, O. Elfatni<sup>1\*</sup>

<sup>1</sup>Laboratory of high energy physics-Modeling and Simulations, Faculty of Sciences, Rabat, Morocco

Received 22 Feb 2017,  
Revised 28 Jan 2018,  
Accepted 08 Feb 2018

### Keywords

- ✓ Prediction,
- ✓ Artificial neural network,
- ✓ Multiple regression,
- ✓ Solar radiation,
- ✓ DNI,
- ✓ ANOVA.

[ibelhaj13@gmail.com](mailto:ibelhaj13@gmail.com) ;  
[omfatni@gmail.com](mailto:omfatni@gmail.com) ;

Phone: +212671694767;  
+212657131082

### Abstract

Direct Normal Irradiation (DNI) is the main component of Concentrated Solar Power (CSP), Solar Thermal (ST) systems and Concentrated Photovoltaic (CPV) which are highly utilized in many solar parks in very sunny areas, however this component presents a big challenge in terms of high fluctuations caused by many parameters such as temperature, humidity, cloudiness and turbidity that can affect electricity generation, planning, dispatch and water heating energy which in turns affect the power grid and the limited storage hours capacity hence the importance of prediction. This study is devoted to the development of two direct normal irradiation (DNI) prediction methods and the comparison of their performance. The first prediction method is the Multiple Regression which enables the study of a system with a group of predictors and the understanding of how they affect each other by the means of a mathematical equation. This method happens to be more used by economists and social scientists, but not in the field of solar prediction. The second method is that of artificial neural networks, which is known as one of the best methods for short term solar radiation forecasting and prediction. It is a nonlinear method that is suitable to describe in a good manner the chaotic nature of the atmosphere. The case study is the region of Ouarzazate in the south west of the Mediterranean basin, which houses the large CSP Park Noor. The data used are 19 years of hourly historical satellite weather data from 07h30 am to 17h30 pm. The inputs of the models found are selected based on an analysis of variance study (ANOVA). The architecture of the neural model is a multilayer Perceptron of five entries. The same predictors are used for the Multiple Linear Regression and the neural network model. The results show a better performance of the neural model compared to multiple regression.

### 1. Introduction

Solar prediction is of big importance due to its uses. It permits to plan the dispatch, determine the hours of storage needed, the amount of energy stored that should be used and avoid the big fluctuations in the electrical grid caused by the nature of the renewable energy resources. Many methods have been developed to forecast solar radiation. Classical and hybrid models [1-6] were subject to many works. Artificial neural networks based methods [7-21] prove to be the best for nonlinear problems due to its capacity of generalization and fitting any kind of functions which makes it suitable for any random variable based process in particular forecasting the renewable energy resources. Due to their nonlinear nature, artificial neural networks are used as part of many forecasting hybrid models.

Contrary to time series models [22-24], multiple regression focuses on the dynamics of data as physical systems without considering time as a parameter which implies stationary of the process. Generally known for its application in economics, which knows fluctuations in the stock market prices, Multiple Regression is also used in forecasting wind energy which happens to be a renewable energy resource too.

This work focuses on the development of two prediction methods based on multiple regression and artificial neural networks with a comparison between the two. An initiative of such comparison has been recently done for the prediction of solid wastes [25] which motivated to do this study for solar irradiation prediction. In this study, the solar component forecasted is the direct normal irradiation (DNI), which is the main component of CSP. The range of forecasting is one hour ahead using hourly data from 1994 to 2012. Ouarzazate, a region in the south west Mediterranean basin region and southern east Morocco, is chosen to be

the case of our study. The reason of our choice is the fact that Ouarzazate is part of the Moroccan solar plan which aims to produce 2000MW of electricity from solar energy by the horizon of 2020.

In this paper, the first section will be an introduction; the next one will be about data used for this work. The methods part will introduce the process used for pre-processing and post-processing data and the prediction methods and model selection criteria.

In the results section, two regression models are selected based on statistical selection criteria and an analysis of variance ANOVA. The neural networks models tested used different training algorithms with optimization procedures. The results show a better performance of the neural networks method compared to the multiple regression. Finally, a conclusion part will discuss the study done and give insights about future contributions.

## 2. Data

The Direct Normal Irradiation (DNI) and meteorological data used are the hourly data provided by MASEN (Moroccan Agency for Solar Energy) calculated from METEOSAT MSG and MFG satellite data (EUMETSAT) and from atmospheric data (ECMWF and NOAA) by SolarGIS method for the region of Ouarzazate. The spatial resolution for solar radiation is 250m × 250m and for meteorological data is 1000m × 1000m, which reveals the quality of the data used. The region of Ouarzazate located in the latitude 31° 00' 83" and longitude -06° 86' 27" in south Morocco is a strategic region for the Moroccan Solar Plan (MSP) that aims to produce 2000MW by 2020, making of Morocco one of the first countries in the world in terms of producing electricity from renewable energy resources. The entire hourly database used represents 19 years of data starting from 1st January 1994 to 31st December 2012. The true solar value (TSV) for the whole 19 years was determined to be from 07h30 am to 17h30 pm which is the time slot used in our study.

## 3. Methods

In this section the methodologies used for prediction are presented. The first subsection will be dedicated to the multiple regression method including the technique used for building the regression model and the input selection. The second subsection will handle the artificial neural networks techniques starting from the pre-processing and post-processing of data and defining each algorithm used.

### 3.1. Multiple Linear Regression (MLR)

The multiple regression method is a linear method that consists of modelling data based on finding a regression line that describes their sparseness and fit them by the mathematical equation [27]:

$$y = \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + b \quad (1)$$

This equation relates the predictors (or multivariate independent variables)  $X_1 \dots X_n$  to the predicted variable (or the dependent variable)  $y$ . There exist many criteria for fitting a model to the data. The method used in this work is the ordinary least squares (OLS) method for several reasons that will be discussed later. MLR is developed based on the assumption that the explanatory variables must be independent, and the dependent variable must be normally distributed with zero mean and constant variance.

#### 3.1.1. The Ordinary Least Squares (OLS)

The Ordinary Least Squares (OLS) is a statistical method that fits a model using all of the independent variables. It estimates the linear regression coefficients by minimizing the vertical squared distance between the regression line and the data points. In our models, no constant will be used because it affects the results since it has no meaningful physical interpretation.

#### 3.1.2. Input selection

While building a regression model, selecting the predictors is a necessary phase. In order to do that, some assumptions must be verified. The first one is to check the correlation between the dependent variable and the predictors in order to determine the degree of dependence of the predicted values to the predictors. The correlation has to be greater than 0.05. P-value is also used as a criterion for selecting predictors. A p-value less than 0.01 means that the probability is less than 1%, which in turn means that the null hypothesis would be rejected at the 1% level. A p-value between 0.01 and 0.05 means, that the null hypothesis would be rejected at the 5% but not the 1% level. A p-value of 0.05 or more means, that it would not be rejected at the 5% level [26].

The second one is to remove the multicollinearity between independent variables. This can be done by using the Principal Component Analysis (PCA), however, in this study, we will proceed by removing predictors that have the absolute value of the correlation of the coefficient estimate greater than 0.5 because it is more convenient in our case to use the actual satellite data for an easier implementation of the prediction tool. In this study, we tried to fit 10 inputs to the direct normal irradiation DNI namely: anterior value of Direct normal Irradiation (DNI) at time t-1 in Wh/m<sup>2</sup>, Sun altitude angle (SE) in degree, Sun azimuth angle (SA) in degree, air temperature (TEMP) at 2m in °C, atmospheric pressure (AP) in hPa, relative humidity (RH) in %, Wind speed (WS) at 10m in m/s, Wind direction (WD) in degree, Dew point temperature (TD) in °C, Wet bulb temperature (WBT) in °C

**Table1:** Correlation matrix between the predictors and the dependent variable.

|        | SE(t-1) | SA(t-1) | TEMP(t-1) | AP(t-1) | RH(t-1) | WS(t-1) | WD(t-1) | TD(t-1) | WBT(t-1) | DNI(t-1) |
|--------|---------|---------|-----------|---------|---------|---------|---------|---------|----------|----------|
| DNI(t) | 0.1665  | -0.3367 | -0.1413   | -0.0344 | -0.0747 | -0.1221 | -0.1406 | -0.2577 | -0.2188  | 0.8126   |

Table1 demonstrates how the strong relationship between the anterior value of DNI and the actual value hence the purpose of including it in the independent variables. In our analysis, the input correlation matrix for coefficient estimates of the fitting model detected the presence of 7 serious multicollinearity, i.e., correlation amongst the predictor variables with absolute values greater than 0.5. Two sets of predictors are identified for the potential regression model:

SET1: SA(t-1), TEMP(t-1), WS(t-1), WD(t-1), DNI(t-1)

SET2: SA(t-1), RH(t-1), WS(t-1), WD(t-1), DNI(t-1)

### 3.2. Artificial Neural Networks

Artificial neural networks are a nonlinear method of artificial intelligence very used for fitting problems such as prediction. It consists of: an input layer, a hidden layer and an output layer each layer with a number of nodes interconnecting between two layers and inside each layer.

$$y_j = f\left(\sum_{i=1}^m W_{ij} X_i + b_j\right) \quad (2)$$

Where:  $w_{ij}$  is the matrix of the weights,  $b_j$  the bias,  $X_i$  the inputs and  $f$  the transfer function also called the activation function. In this study, four different algorithms are used to train the neural network using a tangent sigmoid function (Tansig).

#### 3.2.1. Pre-processing and post-processing data

Before the use of any method, data have to be pre-processed in order to avoid errors in the results, to reduce the distance between the data and to make sure that they reflect as close as possible the reality. Since the inputs used for multiple regression and artificial neural networks have to be linearly independent, the correlation coefficient value between these independent variables is determined to eliminate the unnecessary variables. Next, the wide range between data used motivates to normalize them. Since there are some inputs that take negative values it is best suited to normalize data between -1 and 1. The equation used is:

$$p_{normalized} = \frac{2(p - \min_p)}{(\max_p - \min_p)} - 1 \quad (3)$$

The data are after that denormalized using:

$$p = \frac{(p_{normalized} + 1)(\max_p - \min_p)}{2} + \min_p \quad (4)$$

The database is divided randomly into three data sets using the following distribution: 70% of data for training, 15% for test and 15% for validation.

#### A. Variable Learning Rate Back-propagation (VLRB)

The variable learning rate back-propagation (VLRB) is a heuristic technique based on the elaboration of a rule for the change of the learning rate value of the gradient descent algorithm. It states that the determination of the learning rate cannot be determined before training since a too high value may cause the algorithm to oscillate

and become unstable and a too small rate can take the algorithm too long to converge hence the practical option of allowing the learning rate to change during the training process to find its optimal value improving by that the performance. The application of this algorithm suggests that if the error exceeds the old error by more than a predefined ratio the new weights and biases are discarded.

#### B. Polack-Ribière Update (PR)

The Polack-Ribière update (PR) is a version of the conjugate gradient algorithm proposed by Polack and Ribière. The search direction at each iteration is determined by:

$$P_k = -g_k + \beta_k P_{k-1} \quad (5)$$

Where the constant  $\beta_k$  is computed by

$$\beta_k = \frac{\Delta g_k^T g_k}{g_{k-1}^T g_{k-1}} \quad (6)$$

The inner product of the previous change in the gradient with the current gradient divided by the norm squared of the previous gradient.

#### C. One Step Secant (OSS) algorithm

The one step secant (OSS) algorithm is a numerical optimization technique that establishes a compromise between full Quasi-Newton algorithms and conjugate gradient algorithms. It is a secant approximation with smaller storage and computation requirements compared to the traditional training algorithms.

#### D. Levenberg-Marquardt (LM)

The Levenberg-Marquardt algorithm is a training method introduced by Kenneth Levenberg and Donald Marquardt designed to approach Newton's method, which is a second-order training method, as quickly as possible by approximating the Hessian matrix as:

$$H = J^T J \quad (7)$$

And the gradient as

$$g = J^T e \quad (8)$$

where J is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases, and e is a vector of network errors.

The weights are adjusted using the following Newton like update:

$$W_{k+1} = W_k - [J^T J + \mu I]^{-1} + J^T e \quad (9)$$

The scalar  $\mu$  is a learning rate.

## 4. Results and discussion

In this section we present the findings of our study. Several neural network models are tested to assess their predictive performance and compare it to the prediction using multiple linear regression models. Two regression models are constructed after the elimination of multicollinearity between inputs and selection of independent variables. The equations of each model based on its set of independent variables are given below:

The equation of the Model 1 based on set 1:

$$DNI(t) = 0.911822 * DNI(t-1) + 0.0910817 * WD(t-1) + 3.86395 * WS(t-1) - 0.238613 * TEMP(t-1) - 1.7183 * SA(t-1) \quad (10)$$

The equation of the Model 1 based on set 2:

$$DNI(t) = 0.896447 * DNI(t-1) + 0.032933 * WD(t-1) + 1.17254 * WS(t-1) + 1.06031 * RH(t-1) - 1.59746 * SA(t-1) \quad (11)$$

The quality of the performance of each model is reflected by the Mean Squared error (MSE) and Mean Absolute Error:

$$MSE = \frac{1}{N} \sum_{i=1}^N (target_i - output_i)^2 \quad (12)$$

$$MAE = \frac{1}{N} \sum_{i=1}^N |target_i - output_i| \quad (13)$$

N is the number of samples.

**Table 2:** Comparison of the results of the two regression models

| Models  | R-squared | R-squared adjusted for degrees of freedom | Std error of est. | MAE     | MSE        |
|---------|-----------|-------------------------------------------|-------------------|---------|------------|
| Model 1 | 0.957657  | 0.957654                                  | 143.161           | 97.2556 | 2.0494e+04 |
| Model 2 | 0.958602  | 0.958599                                  | 141.554           | 95.5198 | 2.0036e+04 |

Table 2 shows the errors of the two regression models. The coefficient of determination (R-squared), also called goodness of fit, shows a very good quality of prediction for the models since the value is very close to 1. However, the MSE and MAE show a relatively high error in the models which can be explained by the fact that the models describe most of the data but not with a high accuracy.

To select a regression model between the two, a test of Fisher is made based on the analysis of variance ANOVA to select the best model between the two and also define the best set of predictors that will be inserted for the neural network to make the comparison.

**Table 3:** Analysis of variance for model 1

| Source   | Sum of squares | Df    | Mean Square | F-Ratio   | P-value |
|----------|----------------|-------|-------------|-----------|---------|
| Model 1  | 3.21433E10     | 5     | 6.42866E9   | 313666.97 | 0.000   |
| Residual | 1.42124E9      | 69345 | 20495.2     |           |         |
| Total    | 3.35645E10     | 69350 |             |           |         |

**Table 4:** Analysis of variance for model 2

| Source   | Sum of squares | Df    | Mean Square | F-Ratio   | P-value |
|----------|----------------|-------|-------------|-----------|---------|
| Model 2  | 3.2175E10      | 5     | 6.435E9     | 321146.05 | 0.000   |
| Residual | 1.38951E9      | 69345 | 20037.6     |           |         |
| Total    | 3.35645E10     | 69350 |             |           |         |

Tables 3 and 4 display the analysis of variance for the two regression models. From the two tables, the variance of the first model is 20495.2 and the one for the second model is 20037.6. According to the table of Fisher, Model 1 is selected since:

The set 1 of the Model 1 (i.e. SA(t-1), TEMP(t-1), WS(t-1), WD(t-1), DNI(t-1)) is used as a set of inputs for the neural networks models. Table 5 gives the results of the artificial neural networks. The results show that the model using the Levenberg- Marquardt training algorithm with 30 neurons in the hidden layer is the best suited for the prediction with MSE=0.04135 and a relatively low number of epochs (74 iterations).

**Table 5:** The best results found different architectures of the artificial neural networks based on set 1

| Transfer function | Training algorithm | Structure     | MSE            | N° of iterations |
|-------------------|--------------------|---------------|----------------|------------------|
| Tansig            | VLRB               | 5-15-1        | 0.075295       | 91               |
|                   | PR                 | 5-30-1        | 0.043579       | 209              |
|                   | OSS                | 5-20-1        | 0.044867       | 271              |
|                   | <b>LM</b>          | <b>5-30-1</b> | <b>0.04135</b> | <b>74</b>        |

Figure1 illustrates the performance of the neural network model in terms of MSE versus the epochs. The graph indicates reasonable results and very good similarities between the test set and the validation which indicates that no over-fitting has occurred. Figure 2 plots the fitting regression line for the three phases. The R-value shows a strong correlation between the targets and the outputs of the neural model.

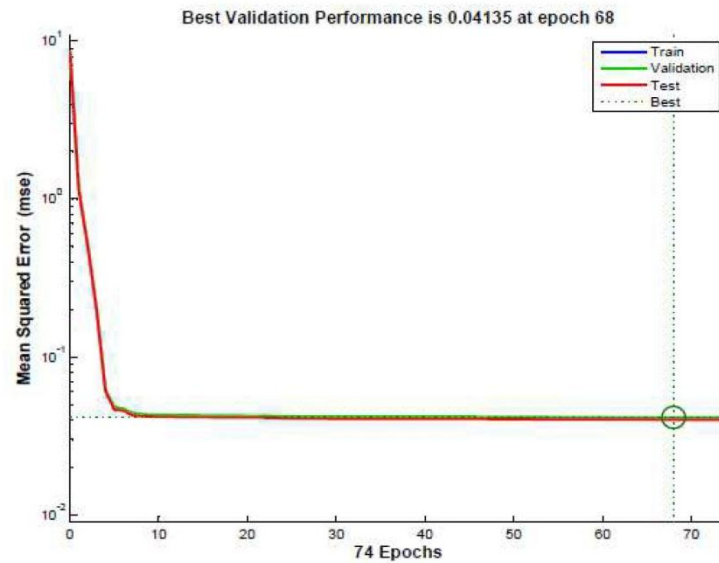


Figure 1: Plot of the performance of the neural network 5-30-1

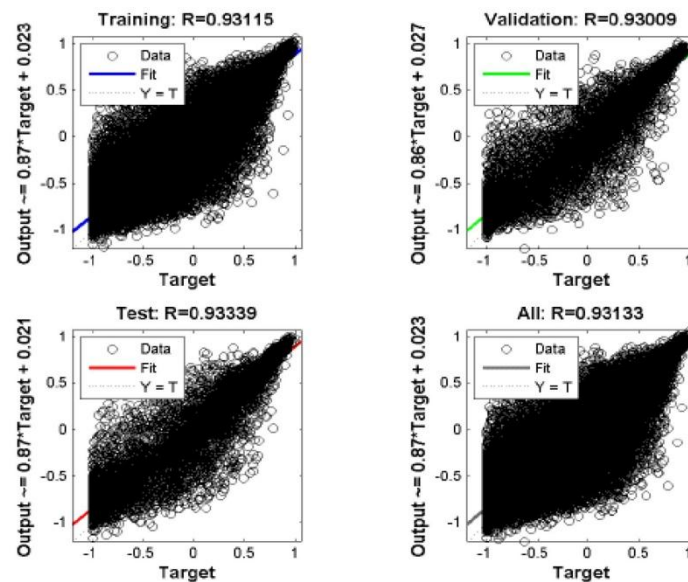


Figure 2: Plot of the regression line between the target and output of the ANN model 5-30-1.

## Conclusion

The objective of this study was the prediction of direct normal irradiation DNI using two different methods. The first method being the multiple linear regression method and the second is the artificial neural networks. In the first method, we used the ordinary least squares (OLS) method for building the regression model. In the second method, four algorithms were tested with different neural architectures. The results show the better performance of the neural based method than the regression one using sun azimuth angle, temperature, wind speed, wind direction and the anterior value of DNI as inputs with an MSE = 0.04135 and a high correlation between the targets and model outputs ( $R=0.93133$ ) compared to  $MSE=2.0494E4$  for the regression model.

The conclusion made is that the Multiple Linear Regression model is not suitable for estimating the behaviour of direct normal solar radiation for the region of Ouarzazate using the selected inputs. On the other hand, the nonlinear method of neural networks using the same inputs gives excellent results which indicate its high accuracy in the prediction of DNI in this area.

## References

1. C. Voyant, M. Muselli, *Energy*. 39 (2012) 341-355.
2. K. Benmouiza, A. Cheknane, *Enconman*. 75 (2013) 561-569.
3. C. Voyant, M. Muselli, C. Paoli, M.L. Nivet, *Renene*. 53 (2013) 1-11.
4. J. Wu, C.K. Chan, *Solener*. 85 (2011) 808-817.
5. M. Chaabene, M. Ben Ammar, *Renene*. 33 (2008) 1435-1443.
6. J. Wang, H. Jiang, Y. Wu, Y. Dong, *Energy*. 81 (2015) 627-644.
7. H.N. Wang, Y.M. Cui, R. Li, L.Y. Zhang, H. Han, *Asr*. 42 (2008) 1464-1468.
8. A. Koca, H.F. Oztup, Y. Varol, G.O. Koca, *Eswa*. 38 (2011) 8756-8762.
9. J.C. Cao, S.H. Cao, *Amer. Energy*. 31 (2006) 3435-3445.
10. C. Voyant, M. Muselli, C. Paoli, M.L. Nivet, *Energy*. 36 (2011) 348-359.
11. G. Notton, C. Paoli, S. Vasileva, M.L. Nivet, J.L. Canaletti, C. Cristofaria, *Energy*. 39 (2012) 166-179.
12. A. Kialashaki, J.R. Reisel, *Energy*. 76 (2014) 749-760.
13. J. Cao, X. Lin, *Enconman*. 49 (2008) 1396-1406.
14. A.N. Celik, T. Muneer, *Enconman*. 67 (2013) 117-124.
15. G. Zhang, B.E. Patuwo, M.Y. Hu, *Iiforecast*. 14 (1998) 35-62.
16. A. Ouammi, D. Zejli, H. Dagdougui, R. Benchrifa, *Rser*. 16 (2012) 4876-4889.
17. A. K. Yadav, S.S. Chandel, *Rser*. 33 (2014) 772-781.
18. S. X. Chen, H.B. Gooi, M.Q. Wang, *Renene*. 60 (2013) 195-201.
19. Y. Kashyap, A. Bansal, A.K. Sao, *Rser*. 49 (2015) 825-835.
20. A. Mellit, M. Benghanem, S.A. Kalogirou, *Apenergy*. 83 (2006) 705-722.
21. B. Amrouche, X. Le Pivert, *Apenergy*. 130 (2014) 333-341.
22. L. Suganthia, A.S. Anand, *Rser*. 16 (2012) 1223-1240.
23. M. Diagne, M. David, P. Lauret, J. Boland, N. Schmutz, *Rser*. 27 (2013) 65-76.
24. J.J. Flores, M. Graff, H. Rodriguez, *Renene*. 44 (2012) 225-230.
25. S. Azadi, A.K. Jashni, *Wasman*. 48 (2016) 14-23.
26. C. Dougherty, *Introduction to Econometrics*, Oxford University Press USA. (2007).
27. J. David Olive, *Linear Regression*, Springer. (2017).

(2018) ; <http://www.jmaterenvirosci.com>

## Résumé

Le développement et l'exploitation des énergies renouvelables ont connu une forte augmentation ces dernières années dans le monde et plus particulièrement au Maroc vu sa situation géographique favorable qui montre un potentiel solaire et éolien très important.

Le travail de notre thèse met l'accent sur la prévision de la vitesse du vent qui est la composante principale du système éolien. Pour cette raison, nous avons développé différents modèles en utilisant les techniques de modélisation et d'intelligence artificielle. Une comparaison de la performance de ces techniques est faite en se basant sur des indicateurs de performance qui seront exposés ultérieurement dans ce manuscrit. Les sites étudiés sont Tanger au Nord et Tarfaya au Sud du Maroc en raison de leurs positions géographiques où la vitesse du vent dépasse 10m/s. La base de données utilisée contient trois années de données satellitaires météorologiques horaires.

Une étude statistique des données montre une colinéarité entre les paramètres d'entrée qui influencent la sortie du système et qui peut avoir une explication sur son comportement non linéaire. Les résultats obtenus montrent que les méthodes non linéaires ont une précision bien meilleurs que les méthodes linéaires.

**Mots clés :** prévision, vitesse du vent, modélisation, intelligence artificielle.

## Abstract

The development and exploitation of renewable energies have experienced a strong increase in recent years in the world and more particularly in Morocco given its favorable geographical location which shows a very important solar and wind potential. The work of our thesis focuses on the prediction of the wind speed which is the main component of the wind system. For this reason, we have developed different models using modeling and artificial intelligence techniques. A comparison of the performance of these methods is made based on performance indicators that will be discussed later in this manuscript. The regions studied are Tangier in the north and Tarfaya in the south of Morocco because of its geographical positions where the wind speed exceeds 10m/ s. The database used contains three years of hourly meteorological satellite data. A statistical study of the data shows a collinearity between the input parameters that influence the system output and which may have an explanation for its non-linear behavior. The results obtained show a better performance of the non-linear methods to the linear methods.

**Keywords:** Forecast, wind speed, modeling, artificial intelligence.

Année Universitaire : 2020/2021

Faculté des Sciences, avenue Ibn Battouta, BP. 1014 RP, Rabat-Maroc.

Tél : 00212(0)5 37 77 18 76, Fax : 00212(0)5 37 77 42 61, <http://www.fsr.ac.ma>