

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère de l'enseignement supérieur et de la recherche scientifique

UNIVERSITE AHMED DRAIA
-ADRAR-



جامعة أحمد درايا-أدرار

Année/2022

Faculté des Sciences et de la Technologie
Département des Sciences de la Matière

Mémoire de fin d'étude, en vue de l'obtention du diplôme de Master en

Physique

Option : Physique Energétique et Energies Renouvelables

Thème

**Etude théorique des modèles spatio-temporels
probabilistes pour la prédiction à court terme de
la production éolienne et photovoltaïque**

Présenté Par :

Mr. SENOUCI REGUIG

et

Mr. OMAR DAHMOUNE

Devant le jury composé de:

Mr. BOUSSOUKAIA Tahar

Président

MCA

Université Ahmed Draia-Adrar

Mr. SAYHI Mostpha

Examinateur

MCB

Université Ahmed Draia-Adrar

Mr DALILE Ahmed

Promoteur

MAA

Université Ahmed Draia-Adrar

Année Universitaire 2021/2022



شهادة الترخيص بالإيداع

انا الأستاذ(ة): دليل أحمد
المشرف مذكرة الماستر الموسومة بـ: Étude Théorique des modèles spatio-temporels probabilistes pour la prédiction à court terme de la production éolienne et photovoltaïque
من إنجاز الطالب(ة): دحوى عمر
و الطالب(ة): هسي رقيف
كلية: العلوم والتكنولوجيا
القسم: علوم المادة
التخصص: فيزياء طاقوية و طاقات متجددة
تاريخ تقييم / مناقشة: 09 - 06 - 2022

أشهد ان الطلبة قد قاموا بالتعديلات والتصحيحات المطلوبة من طرف لجنة التقييم / المناقشة، وان المطابقة بين النسخة الورقية والإلكترونية استوفت جميع شروطها.
وإمكانهم إيداع النسخ الورقية (02) واليكترونية (PDF).

- امضاء المشرف:

مساعد رئيس القسم:



ادرار في: 09 - 06 - 2022

ملخص

تتميز الطاقات المتجددة بتذبذب الإنتاجية بسبب اعتمادها على الظروف المناخية. يتطلب التحكم في هذا النوع من الطاقات استخدام أدوات للتنبؤ المستقبلي بإنتاج محطات الطاقة. تتوفر العديد من طرق التنبؤ بالإنتاج في المصادر المختلفة، ويمكن تصنيفها إلى ثلاث مجموعات رئيسية: الأساليب الإحصائية، والطرق الفيزيائية، والطرق الهجينة. الهدف من هذا العمل هو استكشاف بعض المناهج الاحتمالية المكانية-الزمانية التي تم تطويرها مؤخراً للمساعدة في تحسين القدرة على التنبؤ على المدى القصير في إنتاج طاقة الرياح والطاقة الكهروضوئية.

Abstract

Renewable energies are characterized by high volatility due to their dependence on climatic conditions. Controlling the diversity of production requires the implementation of tools to predict the future production of power plants. In the literature, there are many production forecasting methods, but they can be classified into three main groups: statistical methods, physical methods and hybrid methods. The objective of this work is to explore some of the spatio-temporal probabilistic approaches recently developed in order to help improve the short-term predictability of wind and photovoltaic production.

Résumé

Les énergies renouvelables se caractérisent par une forte volatilité due à leur dépendance aux conditions climatiques. La maîtrise de la diversité de la production nécessite la mise en place d'outils permettant de prédire la production future des centrales. Dans la littérature, il existe de nombreuses méthodes de prévision de production, mais elles peuvent être classées en trois grands groupes : les méthodes statistiques, les méthodes physiques et les méthodes hybrides. L'objectif de ce travail est d'explorer certaines des approches probabilistes spatio-temporelles récemment développées afin de contribuer à améliorer la prévisibilité à court terme de la production éolienne et photovoltaïque.

Remerciements

Louanges et remerciements à Dieu Tout-Puissant qui nous a donné la foi, le courage, la volonté, la santé et la patience pour faire ce travail.

Tout d'abord, nous tenons à exprimer nos sincères remerciements à Mr Ahmed Dalile qui a accepté de nous encadrer durant ce travail, et pour son aide, ses conseils, ses encouragements et sa présence durant la réalisation de ce travail.

Nous remercions tous les professeurs et docteurs de l'Université Ahmed Draia Adrar pour leurs aides, leurs conseils, leurs encouragements, leurs gentillesse et leurs suivis tout au long de ces deux années d'enseignement.

Nous adressons nos sincères remerciements aux membres du jury ainsi qu'aux professeurs, docteurs et travailleurs du Département des sciences de la matière.

Nous remercions sincèrement et chaleureusement nos familles pour leurs compréhensions, leurs soutiens et leurs encouragements.

Enfin, nos sincères remerciements vont à tous nos amis, à ceux qui nous ont aidés d'un simple mot, d'un sourire ou d'un geste, de près ou de loin, à achever et à mener à bien ce travail.

Dédicace

Je dédie ce mémoire avec plaisir:

- À mon père et ma mère reconnaissants pour leur soutien tout le temps mes études et leurs innombrables sacrifices, je leur souhaite longue vie.
- À ma femme, mes enfants Abdelghani et Ilham , mes sœurs et toute ma famille je leur souhaite à tous joie.
- À tous mes amis qui ont toujours fait preuve d'esprit de coopération et d'aide.

Senouci R.

Dédicace

Je dédie ce mémoire:

- À mes parents pour leur amour inestimable, leur confiance, leur soutien, leurs sacrifices, et pour toutes les valeurs qu'ils ont su m'inculquer. Qu'ils trouvent en moi la source de leur fierté.
- À qui je dois tout, ma sœur mon âme ainsi qu'à mes deux frères pour leur tendresse, leur complicité et leur présence dans ma vie. Je leur souhaite un avenir radieux plain de réussit.
- À toute ma famille ainsi qu'à mes amis.

Omar D.

Table des matières

Introduction Générale	1
Chapitre 1: Energies renouvelables et état de l'art de la prévision	7
1.1 Introduction	8
1.2 Etat de l'art de la prévision photovoltaïque	8
1.2.1 Les modèles météorologiques	9
1.2.2 Les modèles de ciel-clair	12
1.2.3 Les modèles déterministes	13
1.3 Etat de l'art de la prévision éolienne	16
1.4 Présentation de la région d'étude	17
1.5 Les corrélations spatio-temporelles	18
1.5.1 Indice de Moran	19
1.5.2 Corrélogramme modifié	20
1.5.3 Impact de la saisonnalité	20
1.5.4 Variabilité de la production	21
1.6 Conclusion	22
Chapitre 2: Modèles spatio-temporels déterministes et probabilistes	23
2.1 Introduction	24
2.2 Les modèles spatio-temporels déterministes	25
2.2.1 Les modèles spatio-temporels déterministes de la production PV ..	25
2.2.2 Les modèles spatio-temporels déterministes de la production éolien	30
2.3 Modèles spatio-temporels probabilistes	34
2.3.1 Modèles probabilistes pour la prévision de la production PV	34
2.3.2 Modèles probabilistes pour la prévision de la production éolienne	39
2.4 Conclusion	45
Chapitre 3: Analyse comparative et perspectives	46
3.1 Introduction	47
3.2 Analyse comparative	47
3.2.1 Critères d'évaluation des prévisions probabilistes	47
3.2.2 Évaluation des méthodes de sélection des variables	49
3.2.3 Evaluation de la production de l'énergie éolienne	51
3.3 Perspectives	53
Conclusion Générale	55
Bibliographie	59

Liste des tableaux

2.1	Les données du modèle GFS	32
2.2	Les données du modèle ECMWF	33
3.1	Énergie produite annuellement par chaque éolienne	52
3.2	Résultats de la production électrique nette des deux sites sélectionnés	53

Table des figures

1.1	Principe de fonctionnement de base des modèles NWP	10
1.2	Description des systèmes de coordonnées et des procédés physiques dans un modèle NWP	11
1.3	Illustration des méthodes de prévision déterministe à court-terme de la production photovoltaïque	14
1.4	Les centrales de production d'énergie PV et éolienne en Algérie ..	18
2.1	Comparaison des valeurs de RMSE normalisées des modèles AR et persistance	28
2.2	Approche directe de la prévision de la production éolienne	31
2.3	Approche indirecte de la prévision de la production éolienne	31
2.4	Schéma récapitulatif de l'acheminement des données	33
2.5	La distribution de Weibull, l'histogramme et la rose des vents à 10 m de hauteur	40
2.6	La carte des vitesses des vents et la localisations des sites	42
2.7	Modélisation de la variation temporelle	42
2.8	Modélisation de la variation spatiale	43
2.9	Modèle de prévision indirect de la production	45
3.1	Les critères de KS et de MAEP	48
3.2	Fiabilité des quantiles estimés pour 3 heures d'horizon	50
3.3	Sharpness des densités prévues pour des horizons 3h	51

Introduction Générale

La production d'électricité à partir d'énergies renouvelables est une solution pour réduire le bilan global des émissions de gaz à effet de serre et accroître l'autonomie énergétique d'un territoire.

En fin 2016, 24.5% de l'électricité mondialement produite était d'origine renouvelable, dont 16,6% issues de barrages hydrauliques, 4% d'énergie éolienne, 2% de combustion de biomasse, et 1.5% d'énergie photovoltaïque. La part des énergies marines (houle, marée), du solaire thermodynamique et de la géothermie était de 0.4% [01].

Le solaire photovoltaïque et l'éolien sont les filières connaissant les plus fortes progressions mondiales d'installation de capacité, avec respectivement 28 et 12.6% de croissance entre fin 2015 et fin 2016 [02]. Ces deux sources d'énergie ont aussi toutes deux la particularité d'avoir une production qui varie étroitement avec les conditions météorologiques. L'énergie éolienne dépend de la vitesse du vent au niveau du hub (qui supporte les pales) et dans une moindre mesure de sa direction au niveau de la turbine. La production photovoltaïque dépend essentiellement du rayonnement solaire incident au sol (ci-après appelée éclairement) et dans une moindre mesure de la température du module photovoltaïque. Si les énergies hydrauliques et marines présentent une faible variabilité de production à l'échelle de quelques heures, voire plusieurs jours, les énergies solaires et éoliennes peuvent connaître des variations de très fortes amplitudes en quelques secondes. Ces énergies dites *variables*, souvent appelées à mauvais escient *intermittentes*, ont une production à variabilité stochastique: leur prévisibilité est limitée. Plus précisément, l'erreur de prévision de production électrique peut devenir significative au-delà de quelques minutes.

L'électricité se stockant difficilement, il est impératif que la quantité produite puisse, à tout instant et à l'échelle d'un pays, être égale à celle de la consommation. Les réseaux électriques actuels et leur gestion ont été pensés pour accueillir des énergies à production garantie (hydro-

électricité, combustion de charbon, fission nucléaire, etc.) et délivrer de l'électricité selon une répartition préexistante des bassins de population. La coexistence d'un service efficace de distribution d'électricité et d'une forte pénétration des énergies solaires et éoliennes nécessite une gestion du réseau active qui puisse bénéficier de capacités d'inter-connexion renforcées, de moyens de stockage au meilleur rendement possible, d'un contrôle de la demande (organisation d'effacement, politique de sobriété énergétique), mais aussi d'anticipation de la production. La prévision-météorologique tient donc un rôle majeur dans la transition énergétique en cours vers plus d'énergies renouvelables.

La prévision de production s'utilise sur plusieurs points distincts des filières éoliennes et solaires. Le développement d'un projet de ferme solaire ou éolienne a recours à des données climatologiques. Des prévisions à très court terme allant de quelques minutes à plusieurs jours sont généralement utilisées pour l'optimisation des transactions sur les marchés de l'électricité, le dimensionnement et la gestion des réseaux électriques, notamment des réserves, à différentes échelles spatiales et temporelles.

Avant de construire une ferme, le développeur de projet doit connaître le gisement d'un territoire afin d'optimiser le choix du terrain, la capacité de la centrale et sa rentabilité. La prévision de cette production à très long terme (souvent 20 ans) suppose que le gisement des 10 à 30 dernières années est représentatif du potentiel énergétique des décennies à venir. Plus concrètement, des séries temporelles de données solaires (éclairage énergétique solaire au sol sur un plan horizontal ou incliné, appelé aussi *irradiance*) ou éoliennes (vitesse et direction du vent) à échelle horaires, issues d'observations et de modèles, sont utilisées pour obtenir de manière statistique une année représentative du gisement, appelée *année météorologique typique* (TMY en anglais). La TMY contient, heure par heure, les moyennes et les quantiles du gisement qui permettent d'en déduire la production la plus probable et de

quantifier les risques que celle-ci dépasse ou n'atteigne pas des seuils critiques de rentabilité. Si des recherches sont initiées pour estimer l'impact des changements climatiques sur la ressource future, les incertitudes des résultats restent difficilement intégrables dans les modèles de décision actuels des développeurs de projets et de leurs investisseurs [03, 04].

Les limitations des méthodes de prévision basées sur les données de modèles numériques de prévision ont conduit au développement de méthodes qui exploitent les informations spatiales dans la prévision temporelle de la production photovoltaïque. Les premières méthodes proposées pour cette modélisation spatiale et temporelle ont été introduites dans le cadre de la prévision de production éolienne notamment pour la prévision des vitesses et directions de vent, l'étude de la propagation spatiale et temporelle des erreurs de prévision, l'identification de région à fort potentiel de production, et la prise en compte de la variabilité spatiale [05]. Le principe de base de ces modèles est qu'ils exploitent les corrélations entre les conditions atmosphériques observées sur plusieurs sites au même moment, mais aussi avec des décalages temporels. L'utilisation de ces corrélations permet d'améliorer la qualité des prévisions.

La plupart des références sur la prévision spatio-temporelle de production photovoltaïque traitent de la prévision de l'irradiation solaire. La spatialité des modèles provient souvent de l'utilisation d'images de caméras ou de satellites. La résolution spatio-temporelle de l'irradiation globale au sol peut être améliorée par combinaison des mesures d'irradiation et des observations fournies par les images satellites. Les prévisions d'irradiation peuvent être aussi améliorées par le calcul de la vitesse de déplacement des nuages au-dessus des centrales photovoltaïque. Ces informations sur le déplacement des nuages sont consignées dans un vecteur appelé *cloud motion vector* (CMV) [06]. L'impact de la vitesse de déplacement des nuages sur la variabilité des mesures d'irradiation peut être calculé en utilisant des transformations en ondelettes ou des algo-

rythmes de traitement d'images. Les *cloud motion vector* et les prévisions météorologiques sont combinées pour fournir de meilleures prévisions à moyen-terme de l'irradiation. La plupart des modèles auto-régressifs dans le temps ou l'espace peuvent aussi être utilisés pour la prévision spatio-temporelle de l'irradiation de même que les modèles géostatistiques. Les méthodes de classification et la prévision semi-paramétrique sont aussi utilisées dans le cadre de la prévision spatio-temporelle de l'irradiation. Les prévisions par des modèles spatio-temporels de l'irradiation peuvent certes être transformées en prévision de production photovoltaïque par des modèles physiques qui prennent en compte les caractéristiques des systèmes photovoltaïque mais cela correspond à l'introduction d'un niveau d'incertitude (ou d'erreur) supplémentaire.

Le développement des modèles de prévision *probabiliste* a été porté par la volonté d'avoir plus d'informations sur la distribution future de la production photovoltaïque ou éolienne et l'incertitude liée aux prévisions. Les méthodes de prévision probabiliste de la production photovoltaïque ou éolienne peuvent se présenter sous deux approches: la prévision des niveaux de confiance associés à des prévisions déterministes et l'approche directe qui fournit une représentation de la densité de la production. La majorité des méthodes de prévision probabiliste de la production photovoltaïque ou éolienne repose sur la régression quantile. Les quantiles sont ainsi considérés comme une bonne représentation de la densité future à prévoir. La régression quantile peut être utilisée sous différentes formes allant de simples modèles linaires à des modèles plus complexes. Elle peut être aussi couplée à des méthodes d'estimation non paramétrique, à des méthodes de *machine learning* ou à des méthodes de classification pour prévoir la production photovoltaïque ou éolienne. Ces méthodes de classification peuvent aussi être utilisées indépendamment de la régression quantile pour obtenir des prévisions probabilistes de la production [07].

Objectifs et démarche de ce mémoire

L'objectif principal de ce mémoire est l'étude théorique des modèles spatio-temporels pour la prédiction à court terme de la production éolienne et photovoltaïque. Les horizons temporels qui nous intéressent varient de quelques heures à quelques jours. Contrairement aux méthodes de prévision qui exploitent les données de production d'un site ou d'une agrégation de sites mais rarement l'ensemble des informations des sites voisins, ce travail considère des méthodes de prévision qui exploitent les relations spatiales et temporelles entre les productions de différents sites géographiquement distribués. Les ensembles de centrales voisines sont utilisés comme des réseaux de capteurs qui apportent de l'information pour améliorer la qualité des prévisions pour chaque centrale.

Dans le **premier chapitre**, nous présentons l'état de l'art de la prévision de production énergétique photovoltaïque et éolienne. Nous présentons ensuite les différentes étapes de stationnarisation et on discute les corrélations spatio-temporelles à travers les modèles spatio-temporels déterministes de la production.

Dans le **deuxième chapitre**, nous présentons les modèles de prévision spatio-temporels probabilistes pour la prévision de la production éolienne et photovoltaïque. Le formalisme théorique de chaque modèle et la nature probabiliste des résultats issus sont développés.

Le **troisième chapitre** est consacré à une analyse comparative des différents modèles probabilistes considérés. Les évaluations de ces modèles sont faites avec les critères spécifiques à la prévision probabiliste. Nous donnons aussi dans ce chapitre des perspectives sur l'amélioration et le développement théorique des modèles probabilistes pour la prévision de production éolienne et photovoltaïque.

Chapitre 1

Energies renouvelables et état de l'art de la prévision

1.1 Introduction

Les prévisions spatiales et temporelles sont d'une grande importance pour la production d'énergie photovoltaïque et d'énergie éolienne, car avant de placer une ferme pour produire de l'énergie dans un lieu, il est nécessaire de connaître les caractéristiques du terrain et les conditions météorologiques de cette zone. La prévision de production s'utilise sur plusieurs points distincts des filières éolienne et solaire. Le développement d'un projet de ferme solaire ou éolienne a recours à des données climatologiques. Des prévisions à très court terme allant de quelques minutes à plusieurs jours seront utilisées pour l'optimisation des transactions sur les marchés de l'électricité, le dimensionnement et la gestion des réseaux électriques, notamment des réserves, à différentes échelles spatiales et temporelles.

1.2 Etat de l'art de la prévision photovoltaïque

Dans la littérature, il existe de nombreuses méthodes de prévision de la production photovoltaïque. Ces méthodes peuvent être regroupées en trois grandes familles: les méthodes statistiques, les méthodes physiques et les méthodes hybrides [08].

Ces méthodes fournissent soit des prévisions de l'irradiation soit des prévisions directes de production. Le choix d'une méthode de prévision peut être guidé par plusieurs paramètres à savoir le besoin auquel doit répondre la prévision (planification, management de la réserve, participation au marché, etc.), l'horizon de prévision envisagé et aussi le type de données disponibles. Ces paramètres sont étroitement liés. En effet, le besoin à l'origine de la prévision permet de déterminer quels sont les horizons intéressants et donc quelles sont les données à utiliser. Il existe diverses sources de données utilisables dans le cadre de la prévi-

sion de production photovoltaïque à savoir les mesures de production et des variables météorologiques comme l'irradiation solaire, les prévisions météorologiques, les images de caméras ou de satellites.

Une approche intéressante est de regrouper les modèles de prévisions par horizons croissants de quelques minutes à plusieurs jours. On retrouve alors la classification en fonction des horizons de prévision suivante [09]:

- des prévisions intra-horaires (de 15 min à 2h avec un pas temps de 1min);
- des prévisions à très court-terme pour des horizons de quelques heures ($\leq 1\text{h} - 6\text{h}$);
- des prévisions à court terme pour des horizons de quelques jours (1 jour - 3 jours);
- des prévisions à moyen-terme (1 semaine - 3/4 mois);
- des prévisions à long-terme (≥ 1 an).

Les prévisions intra-horaires et très court-terme qui couvrent des horizons allant de moins de quelques minutes à quelques heures sont essentielles aux activités de traitement de la variabilité, de suivi de la production, d'ajustement de la charge et de gestion du stockage. La prévision à moyen terme est utilisée dans le cadre du management et du trading d'énergie. La prévision à long terme quant à elle permet une meilleure planification et optimisation des ressources. On retrouve dans la littérature des comparaisons de méthodes de prévision pour des horizons court et très court terme et des analyses détaillées de ces méthodes suivant le type de données d'entrées [10].

1.2.1 Les modèles météorologiques

Les principaux modèles utilisés dans la littérature pour la prévision de l'irradiation solaire sont des modèles météorologiques [11]. Ils sont dénommés modèles NWP (Numerical Weather Prediction) à cause de

leur caractère numérique et sont des modèles physiques complexes (résolution de système non linéaire d'équations différentielles) qui permettent de prévoir différentes variables météorologiques. Le principe général de ces méthodes repose sur une bonne connaissance au temps initial de l'état de l'atmosphère et des lois physiques qui régissent les changements d'état de l'atmosphère.

Le fonctionnement des modèles NWP peut être décrit par la figure 1.1. On commence par choisir une zone (un domaine), qu'on discrétise ensuite spatialement suivant une résolution choisie. Enfin les modèles NWP prévoient les informations désirées en résolvant des équations de thermodynamique. Les modèles NWP ont beaucoup évolué et leur amélioration est très dépendante des capacités de calcul.

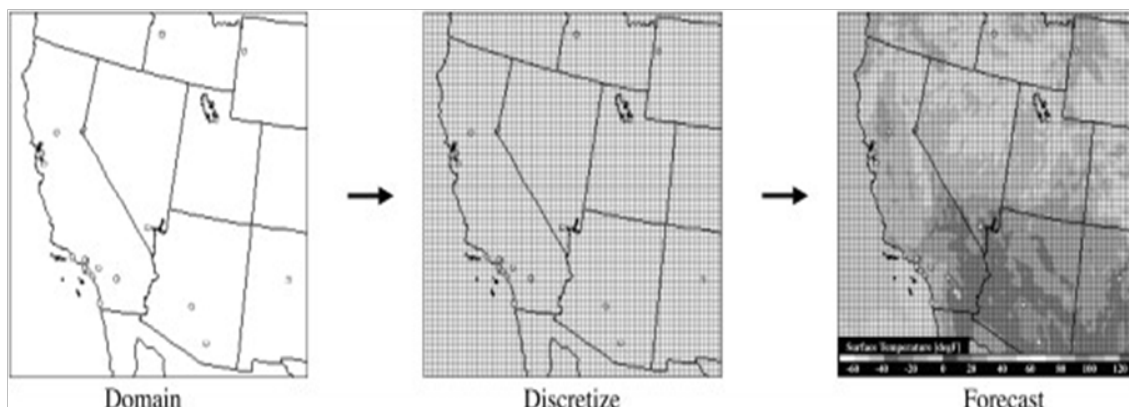


Figure 1.1 Principe de fonctionnement de base des modèles NWP.

La variable prédite ici est la température à la surface et l'horizon de prévision est 18 h [43].

Les modèles NWP se divisent en modèles globaux ou régionaux. Parmi les modèles globaux, on peut citer le modèle ECMWF (European Centre for Medium-Range Weather Forecasts) et le modèle GFS (Global Forecast System) [12, 13, 14]. Le modèle GFS a été développé aux Etats-Unis (National Centers for Environmental Prediction) et propose un accès gratuit aux données. Le modèle ECMWF développé par l'organisation intergouvernementale Européenne propose un accès payant aux données opérationnelles à travers les organismes météorologiques natio-

naux. Parmi les modèles régionaux on peut citer les modèles NAM (North American Mesoscale), RAP ou RAR (Rapid Refresh) qui offrent un accès gratuit aux données. La figure 2 montre le système de coordonnées utilisées pour le découpage.

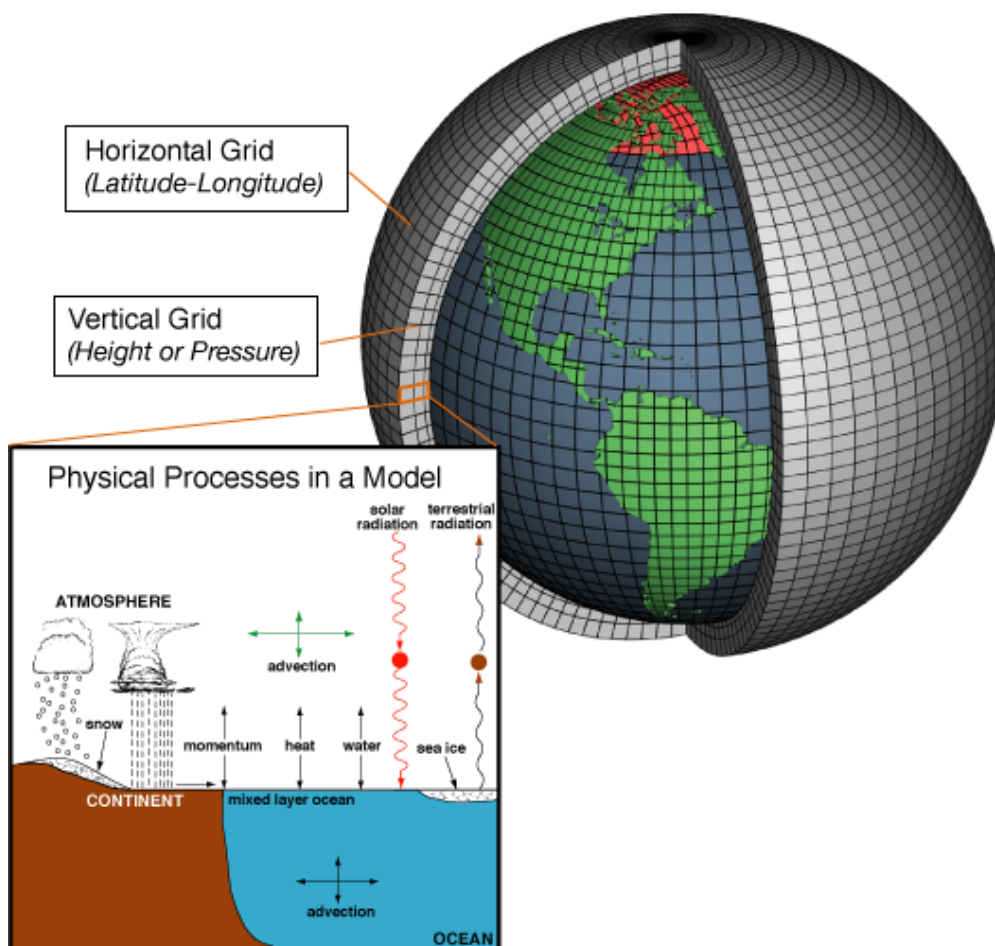


Figure 1.2 Description des systèmes de coordonnées et des procédés physiques dans un modèle NWP [43].

Les sorties des modèles NWP sont très rarement utilisées brutes. Elles peuvent être couplées à des observations d'images satellites ou de caméras hémisphériques qui permettent de décrire la couverture nuageuse au niveau des centrales par l'étude des mouvements des nuages. Ces sorties peuvent aussi être injectées dans des modèles statistiques afin de prévoir l'irradiation ou la production photovoltaïque.

Ces modèles présentent plusieurs limitations. La première concerne les temps de calcul (et de mise à jour) importants. En effet malgré les

avancées dans le domaine pour construire des modèles plus rapides en réduisant le nombre de paramètres ré-estimés par exemple, les temps de calcul et de mise à jour restent des freins importants pour l’exploitation des produits de ces modèles dans le cadre de la prévision à court terme. Une autre limitation concerne la résolution spatiale de ces modèles qui rend impossible la descente à des niveaux microscopiques de la physique qui sont associés à la formation des nuages. Les séries d’irradiation fournies en sortie des modèles NWP sont souvent traitées par un second modèle “filtre” qui permet de les épurer de la variabilité due au cycle solaire. Cette variabilité étant connue avec une bonne précision ces modèles filtres sont intéressants en ce sens où ils permettent de se concentrer sur les autres sources de variabilité. Ces modèles “filtres” sont dits modèles de ciel-clair.

1.2.2 Les modèles de ciel-clair

Le rayonnement solaire au sol est variable dans le temps (et dans l’espace). Cette variabilité est due à la position du soleil, à l’état de l’atmosphère (vapeur d’eau, ozone, aérosol) et enfin à la couverture nuageuse. Cette variabilité de la ressource solaire entraîne celle de la production d’énergie grâce aux systèmes photovoltaïque. Le rayonnement solaire au sol peut être modélisé par le produit du rayonnement ciel-clair par les effets d’atténuation dus aux nuages y compris l’albédo. Le rayonnement ciel-clair est observé lorsqu’il n’y a aucun nuage. Il dépend donc essentiellement de la distance Terre/Soleil, de l’altitude et aussi de paramètres décrivant les gaz et particules de l’atmosphère même sans nuages. Il existe dans la littérature plusieurs modèles pour décrire le rayonnement ciel-clair . Dans le modèle ESRA (European Solar Radiation Atlas), la variation de composition de l’atmosphère est prise en compte par le trouble de Linke (Linke Turbidity Factor) calculé à partir d’une base de données en moyenne mensuelle [16].

Il existe aussi dans la littérature des indices ciel-clair purement statistiques utilisés pour normaliser non pas les séries d’irradiation mais

celles de production photovoltaïque. Ces modèles ne nécessitent aucune connaissance physique préalable des interactions dans l’atmosphère mais seulement les informations tirées des observations .

Dans les modèles ciel-clair statistiques, la régression par noyau et la régression quantile sont utilisés pour estimer les fonctions de lien adéquates entre la production et les variables temporelles. Cette approche d’apprentissage statistique permet de définir *un indice ciel-clair* pour la production photovoltaïque. Cet indice sert non seulement à normaliser les séries de production mais aussi de réduire leur non stationnarité. Les séries normalisées sont plus proches des propriétés de stationnarité que les séries originelles ce qui facilite l’application de méthodes classiques de traitement de séries temporelles.

La modélisation statistique des modèles ciel-clair rencontre plusieurs problèmes. Le premier est l’absence dans la littérature de définition claire du processus d’estimation des fonctions de lien entre la production ciel-clair et le moment de la journée. De plus, dans la littérature, les références proposant une méthode de ciel-clair statistique pour normaliser les séries de production ne présentent pas d’analyse post-stationnarisation des propriétés des nouvelles séries que ce soit au sens des indicateurs de stationnarité ou de la performance de ces séries pour la prévision [17, 18].

1.2.3 Les modèles déterministes

Il existe plusieurs modèles dans la littérature qui permettent de fournir des prévisions de la production photovoltaïque faisant recours à des méthodes dites déterministes dans le ce sens qu’elles fournissent pour chaque instant de l’horizon de prévision une seule valeur qui est la production moyenne attendue. La production d’un système photovoltaïque peut être prévue par couplage des prévisions météorologiques de type NWP avec les caractéristiques des centrales (position, orientation, puissance crête, etc.) Les prévisions de production peuvent aussi

être obtenues grâce à des modèles de traitement de séries temporelles qui peuvent combiner les données de mesures avec des variables exogènes.

Dans la littérature, on retrouve un large éventail de méthodes utilisables pour prévoir la production photovoltaïque à un instant donné. La figure 1.3 présente une schématisation de la prévision déterministe de la production photovoltaïque.

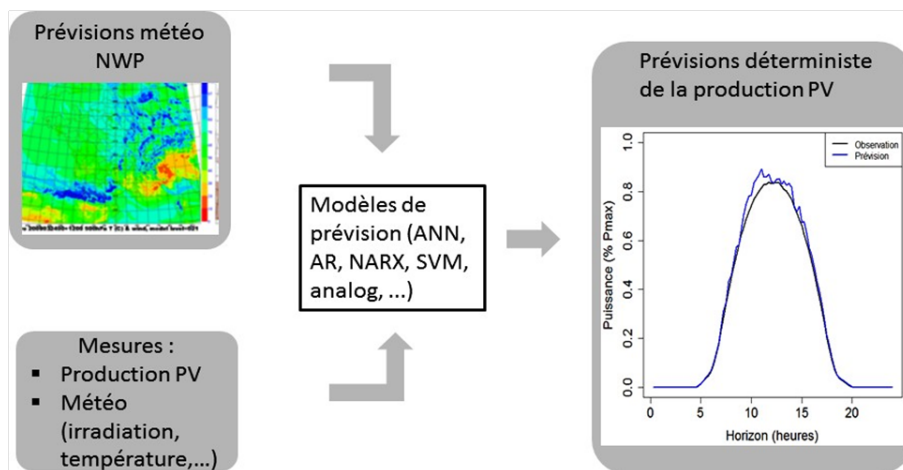


Figure 1.3 Illustration des méthodes de prévision déterministe à court-terme de la production photovoltaïque [43].

Les modèles autorégressifs et de type Box et Jenkins par exemple sont très couramment utilisés pour prévoir la production photovoltaïque [19, 20, 21]. Ces modèles exploitent les historiques de mesure sur site et de prévisions météorologiques en recherchant les similarités avec les observations passées. Dans la même optique, certains modèles fournissent des prévisions de production en construisant des classes de *situations* de production ou météorologiques semblables à celle observées au moment auquel la prévision est faite. Ces modèles sont appelés modèles *analogues*.

Les méthodes d'intelligence artificielle sont aussi très utilisées pour prévoir la production photovoltaïque, les réseaux de neurones étant les plus fréquemment utilisés [22]. Il existe différents types de réseaux de neurones comme les Feed Forward Neural Network (FFNN), les Radial

Basis Function Neural Network (RBFNN) et les Recurrent Neural Network (RNN). Ces méthodes peuvent être implémentées et comparées dans le but de choisir la meilleure fonction d'activation et les caractéristiques optimales du réseau pour la prévision de la production photovoltaïque.

Les réseaux de neurones peuvent être combinés à diverses autres techniques afin d'obtenir des méthodes de prévision de la production plus performantes. Ils peuvent être associés à un modèle linéaire autorégressif avec variable exogène (NARX) ou à des modèles physiques de rayonnement solaire [23]. Les réseaux de neurones sont aussi utilisés dans des processus de modélisation à deux étapes où la première étape consiste à prévoir les variables météorologiques en utilisant les modèles NWP. Les données de production photovoltaïque peuvent aussi être classifiées selon les niveaux de couverture nuageuse (ciel dégagé, partiellement couvert, totalement couvert) et des réseaux de neurones spécifiques à chacune de ces classes peuvent être construits. Des scénarios de la production photovoltaïque peuvent aussi être construits et intégrés aux réseaux de neurones pour la prévision.

Les Support Vector Machine (SVM) fait référence à des méthodes utilisés pour la prévision de la production photovoltaïque [24]. Il existe plusieurs formulations de l'algorithme des SVM mais ils peuvent être interprétés comme une méthode d'estimation où le critère à estimer est la norme carrée de la fonction d'estimation dans un espace bien défini (Hilbertien reproductible). Les SVM peuvent être utilisés avec des prévisions météorologiques de type NWP pour prévoir la production photovoltaïque. L'utilisation des SVM pour prévoir la production peut aussi intervenir dans un processus à deux étapes où la première est une classification des situations météorologiques passées. Les SVM ont été couplées à un modèle SARIMA ou à des chaînes de Markov pour améliorer les performances de prévision de la production photovoltaïque [25, 26]. La méthode de Gradient Boosting ou les algorithmes génétiques

sont aussi des méthodes utilisées pour la prévision de la production photovoltaïque.

La principale limitation de ces méthodes déterministes est qu'elles ne permettent pas de quantifier les incertitudes associées aux prévisions fournies [27, 28].

1.3 Etat de l'art de la prévision éolienne

La prévision éolienne peut être aussi classifiée selon qu'elle intègre les modèles numériques de prévision du temps (Numerical Weather Prediction model, NWP) ou pas [29]. L'intégration des données de modèle NWP dépend de l'horizon de prévision. Selon l'horizon de prévision, on distingue les trois catégories de prévision suivantes :

- Prévision immédiate à court terme (de l'instantané à environ 8 heures),
- Prévision à court terme (de 1 jour à environ 3 jours),
- Prévision à long terme.

Pour un horizon de prévision inférieur à 3 à 6 heures environ, les séries chronologiques utilisant juste les données du SCADA (Supervisory Control and Data Acquisition) donnent des résultats satisfaisants. Après environ 3 à 6 heures, les modèles intégrant les données NWP surpassent les approches des séries chronologiques [29]. Par conséquent, tous les modèles utilisés par les services publics utilisent l'approche avec le NWP.

Il existe deux écoles de pensée différentes en matière de prévision à court terme avec NWP:

- *L'approche physique*: les modèles physiques tentent d'utiliser les considérations physiques aussi longtemps que possible pour parvenir à la meilleure estimation possible de la vitesse du vent local avant d'utiliser les Model Output Statistics (MOS) ou différentes techniques statistiques relativement simples pour réduire l'erreur

restante [29].

- *L’approche statistique*: les modèles statistiques dans leur forme pure tentent de trouver les relations entre une multitude de variables explicatives, y compris les résultats des modèles NWP, et les données de puissance mesurées, en utilisant généralement des techniques récursives. Souvent, des modèles de *Machine Learning* sont utilisés [30].

1.4 Présentation de la région d’étude

En Algérie, comme dans le reste du monde, il existe plusieurs recherches menées annuellement afin de développer la capacité de production en énergies renouvelables à l’aide de prévisions spatiales et temporelles.

Dans la production d’énergie photovoltaïque, dix (10) stations de mesures ont été acquises et installées le premier semestre 2015. Elles sont actuellement en exploitation au niveau des sites retenus conformément aux résultats des études satellitaires réalisées (Identification des sites éligibles). Il s’agit de: Khenchela, El Bayadh, Naama, Tiaret, El Goléa, Timimoun, M’sila, Labreg, Batna et In Salah [31].

Dans la production éolien, on trouve le Projet de réalisation d’une ferme éolienne de 10 MW à Kabertene (Adrar) et Projet de réalisation d’une centrale photovoltaïque de 1,1 MW à Ghardaïa. Les deux projets réalisés dans le cadre de la phase expérimentale, sont entrés en exploitation depuis juillet 2014 [31].

Pour ce qui est du technopôle de Ghardaia, une équipe (SKTM/ CREDEG/ CDER) est sur site pour analyser les restitutions et le comportement des différents sous-champs (différentes technologies de panneaux).

On peut résumer la capacité de production des énergies renouvelables à l’état actuel dans les points suivants [31]:

- Nombre des centrales installée: 21 Centrales (21 Cle PV et 01 Cle Eolien).

- Puissance Installée: 354,3 MW (344,1 MWc PV et 10,2 MW Eolien).
- Energie Produite depuis la MES: 930 GWh (865 GWh PV et 65 GWh Eolien).

La figure 1.4 montre les centrales de production d'énergie photovoltaïque et éolienne en Algérie.

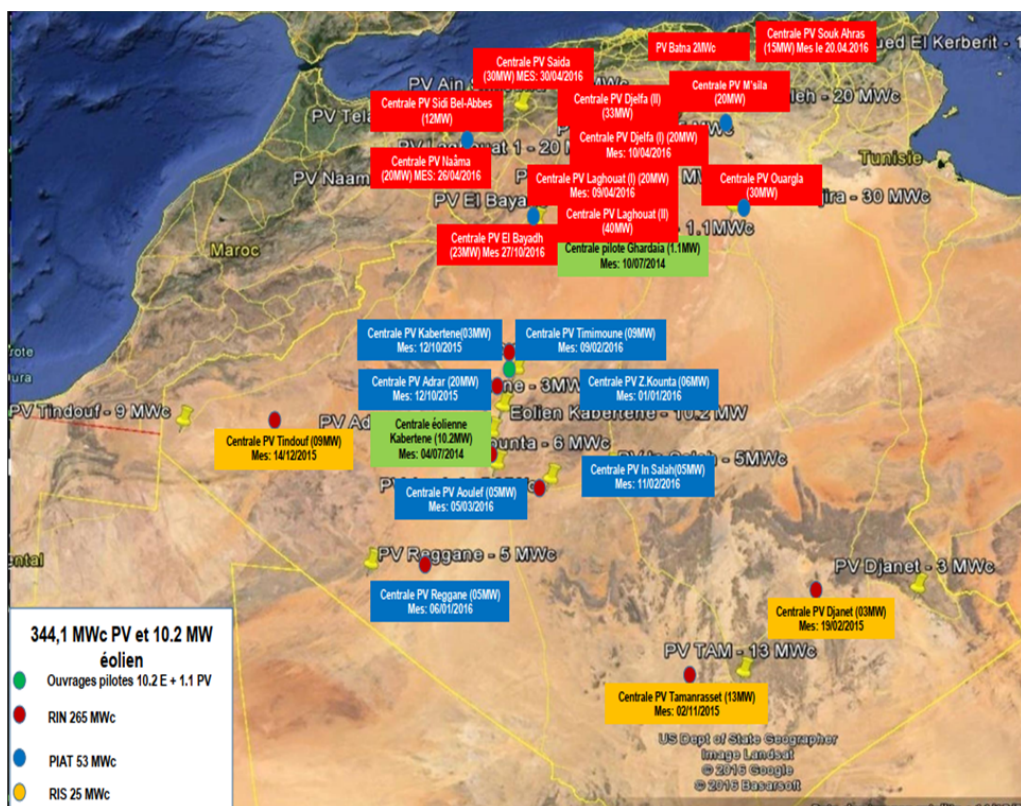


Figure 1.4 Les centrales de production d'énergie photovoltaïque et éolienne en Algérie [31].

1.5 Les corrélations spatio-temporelles

La prédiction spatio-temporelle de la production photovoltaïque ou éolien au sens de l'utilisation des mesures des centrales voisines (par exemple un réseau de capteurs) dépend de l'existence d'une corrélation spatio-temporelle entre ces centrales spatialement distribuées.

Par conséquent, une analyse doit être effectuée sur la base d'indicateurs de liens spatiaux, en plus il faut étudier les corrélations des mesures de production des différents sites, en comparant l'énergie produite au nord avec celle produite dans les hauts plateaux et le sud. On prend à titre d'exemple une station de production photovoltaïque dans Alger, Elbaidh et Adrar. Quant à la production d'énergie éolienne, nous comparons la production entre Oran, Ghardaïa et Adrar, et nous notons les facteurs qui affectent la production, qu'ils soient de type climatiques ou spatiaux.

1.5.1 Indice de Moran

La dépendance temporelle de la production d'énergie photovoltaïque ou éolienne est une caractéristique du phénomène de production, il s'agit donc d'analyser leur existence spatiale. La production photovoltaïque est une variable régionale avec autant d'observations que de sites de production.

La statistique la plus utilisée pour tester l'autocorrélation spatiale dans une série définie spatialement est celle de Moran . Cette statistique s'écrit formellement de la façon suivante pour une variable régionalisée de production Y :

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij} (Y_i - \bar{Y})(Y_j - \bar{Y})}{\sum_i (Y_i - \bar{Y})^2} \quad (1)$$

où Y est la production photovoltaïque sur un réseau discret de N sites, et w_{ij} est l'élément une matrice carrée de poids positifs de dimension N qui quantifie les influences du site j sur le site i . La matrice de poids W que nous avons retenue ici est une matrice de distance.

On trouve dans la littérature des matrices de contiguïté [32], mais elles ne sont pas adaptées au cas de la production photovoltaïque qui est faite sur des sites et non des régions entières, ce qui rend difficile la définition de voisinage.

1.5.2 Corrélogramme modifié

Les auto-corrélogrammes sont utilisés pour étudier l'existence de corrélation spatiale entre les mesures de production. Nous utilisons ici une version modifiée des méthodes classiques de calcul d'auto-corrélogramme [33]. Le principe de cette méthode peut être résumé dans les étapes suivantes :

1. Répartir les corrélations entre les paires de centrales en classes en fonction des distances séparant les sites.
2. Dans chaque classe, tester la significativité des corrélations par des tirages aléatoires de valeur de corrélation croisée de sorte à ce qu'une centrale soit utilisée une seule fois. Par exemple si la corrélation entre A et B est choisie, toutes les autres paires de combinaison intégrant A et B sont exclues de la base de tirage.
3. Répéter la procédure jusqu'à ce qu'il n'y ait aucune centrale inutilisée.
4. Déterminer le nombre de coefficients de corrélation négatifs et positifs.
5. Après un nombre suffisant de tirage, la significativité peut être assurée si on a plus de valeurs positives que négatives.

La production d'énergie éolienne est une variable régionale avec autant d'observations que de sites de production. La corrélation spatiale entre ces différentes observations fait l'objet de différents critères d'indicateurs.

1.5.3 Impact de la saisonnalité

La production photovoltaïque et éolien est caractérisée par une variabilité importante. Les principales sources de cette variabilité sont la saisonnalité et les conditions climatiques. Si les variations saisonnières sont assez bien prévisibles car liées à la course du soleil (production

photovoltaïque) et le vent (production éolien), celles liées au climat notamment à la couverture nuageuse, le sont beaucoup moins. On peut caractériser l’impact de la saisonnalité sur la production photovoltaïque par:

- une production nulle la nuit;
- un cycle journalier avec une pointe de production aux environs de midi;
- un cycle annuel avec des niveaux de production élevés l’été et qui baissent significativement l’hiver.

Les différentes variations que l’on peut observer sur la production journalière d’une centrale sont essentiellement dues aux variations météorologiques.

On peut aussi décrire l’effet de la saisonnalité sur la production éolienne à travers:

- connaître la vitesse du vent la nuit et le jour;
- connaître la vitesse du vent pendant les quatre saisons.

1.5.4 Variabilité de la production

La variabilité de la production d’une installation photovoltaïque ou d’un réseau de plusieurs installations PV peut être analysée par différents critères. On retrouve un outil d’analyse de cette variabilité appelé *variabilité relative de la production*. On définit la variabilité d’un ensemble de N installations photovoltaïque par [34]:

$$\sigma_{\Delta t}^{\Sigma N} = \frac{1}{C^{Fleet}} \sqrt{\text{Var} \left[\sum_{n=1}^N \Delta P_{\Delta t}^n \right]} \quad (2)$$

où

- C^{Fleet} : est la capacité total installée de l’ensemble des N installations photovoltaïques;

- $\Delta P_{\Delta t}^n$: est la série temporelle des évolution de la production pour la centrale n , $\Delta P_{\Delta t}^n = \{(t_1, \Delta P_{t_1, \Delta t}^n), (t_2, \Delta P_{t_2, \Delta t}^n), \dots, (t_T, \Delta P_{t_T, \Delta t}^n)\}$ et $\Delta P_{t_1, \Delta t}^n = P_{t_1}^n - P_{t_1 + \Delta t}^n$.

La variation relative ou ROV (Relative Output Variability) d'une centrale i d'un ensemble de N est définie comme étant [34]:

$$\text{ROV}_i = \frac{\sigma_{\Delta t}^{\sum N}}{\sigma_{\Delta t}^{\sum i}} \quad (3)$$

La variabilité peut être aussi analysée visuellement en s'inspirant de la série des différences proposée. On définit donc une nouvelle série des différences P_t^d qui permet de visualiser les variations de la production photovoltaïque au pas de temps k , $P_t^d = P_t - P_{t+k}$.

De la même manière, la variance de la production d'énergie éolienne ou d'un réseau de plusieurs installations d'énergie éolienne peut être analysée avec différents critères.

1.6 Conclusion

Les prévisions spatiales et temporelles sont d'une grande importance pour la production d'énergie photovoltaïque et d'énergie éolienne. Etant d'une grande variabilité spatio-temporelle, ces ressources d'énergies renouvelables nécessitent l'élaborations de modèles qui nous aident à prédire la production. On trouve dans la littérature de nombreuses méthodes de prévision de la production photovoltaïque éolienne. Ces méthodes peuvent être regroupées en trois grandes familles: les méthodes statistiques, les méthodes physiques et les méthodes hybrides. On a présenté dans ce premier chapitre plusieurs modèles, et on consacre le chapitre suivant pour l'étude des modèle spatio-temporels probabilistes, mais en donnant tout d'abord un aperçu rapide sur les modèles déterministes pour mieux comprendre la particularité des modèles probabilistes.

Chapitre 2

Modèles spatio-temporels probabilistes

2.1 Introduction

Dans ce chapitre nous allons considérer les modèles probabilistes spatio-temporels de la production photovoltaïque et éolienne, mais avant cela nous présentons d’abord dans la première section les bases des modèles dits déterministes afin de bien comprendre les différences entre les deux approches.

Les horizons de prévision considérés vont de quelques minutes à plusieurs heures. Ici, les corrélations spatio-temporelles sont exploitées afin de réduire les erreurs de prédiction par rapport aux méthodes « classiques » de la littérature. Ces méthodes *classiques* exploitent le plus souvent des dépendances temporelles ou spatiales. De plus, pour la plupart des méthodes classées comme spatio-temporelles, le spatio-temporel est souvent corrélé à la précision spatiale des prévisions météorologiques utilisées [35].

Les modèles spatio-temporels déterministes présentés dans le paragraphes suivant fournissent des prévisions qui ne permettent pas d’évaluer les incertitudes liées aux prévisions. Les prévisions probabilistes qui font l’objet de ce chapitre répondent à ce besoin. En effet, elles permettent non seulement de fournir plus d’informations sur la distribution future de la production photovoltaïque mais aussi d’évaluer les incertitudes liées à la prévision. Cette quantification des incertitudes sur les prévisions est très utile pour les différents acteurs du domaine notamment pour la prise de décision et l’évaluation des risques.

Il existe dans la littérature très peu de modèles qui exploitent les relations spatio-temporelles entre les mesures de production pour la prévision probabiliste à court-terme de la production photovoltaïque. On retrouve un modèle [36], basé sur une combinaison entre le modèle auto-régressif vectoriel et la méthode du gradient boosting [37]. Il existe un

modèle basé sur une approche non paramétrique [38], et des modèles qui fassent intervenir des champs aléatoires gaussiens [39].

2.2 Les modelés spatio-temporels déterministes

2.2.1 Les modèles spatio-temporels déterministes de la production photovoltaïque

Dans les modèles spatio-temporels déterministes chaque centrale photovoltaïque est utilisée comme un élément d'un ensemble de capteurs qui permettent de suivre les évolutions des perturbations météorologiques qui affectent la production [35]. Cela différencie cette approche de celles qui utilisent des données hors site provenant de stations météorologiques et de capteurs d'irradiation. Les prévisions sont réalisées avec des données de production et non des données d'irradiation globale permettant ainsi de ne pas introduire un niveau d'incertitude supplémentaire lors du passage de l'irradiation à la production.

Une autre approche déterministe consiste à intégrer les prévisions météorologiques issues de modèles numériques. Cette intégration permet de tenir compte des phénomènes météorologiques pour des résolutions spatiales plus importantes [40]. De plus avec les prévisions météorologiques, les horizons plus éloignés peuvent aussi être prédits et les performances pour les courts horizons sont stabilisées [41].

La densité spatiale des centrales photovoltaïque considérées peut être variable. Pour cela, nous illustrons l'utilité des modèles déterministes avec les deux cas de test d_1 et d_2 qui présentent respectivement un nombre faible de centrales photovoltaïques réparties sur une grande zone géographique et un nombre élevé d'installations photovoltaïques réparties sur une zone dense.

Le problème de la dimensionnalité et la nécessité de l'introduction d'un processus de sélection de variables sont mis en évidence à travers le cas de test d_2 . Nous présentons par la suite une méthode de sélection

de variables pour pallier à ce problème. Une transformation inverse est appliquée une fois la prévision faite pour retrouver des valeurs de production prédites [41]. L'évaluation des erreurs est faite sur ces valeurs de production.

a) Persistance et modèle autorégressif

Dans le but d'évaluer les avantages d'une approche spatio-temporelle pour la prévision photovoltaïque, nous présentons des modèles de référence pour l'analyse comparative qui n'utilisent pas les informations géographiquement distribuées. Plusieurs méthodes peuvent être utilisées pour prévoir la production photovoltaïque [42].

Le *modèle de persistance* est souvent utilisé comme référence dans la littérature pour comparer les performances des modèles avancés [10]. En effet, il est simple à implémenter, et est basé uniquement sur les données mesurées et n'implique aucun processus de modélisation. Les résultats du modèle de persistance sont facilement reproductibles. De plus, dans les applications pratiques de la prévision photovoltaïque, la persistance est souvent choisie comme un modèle de secours pour fournir des prévisions dans le cas où les modèles avancés échouent. Nous définissons ici comme «persistance» un modèle qui considère que la production d'énergie d'une installation photovoltaïque au temps $t + h$ est la même que la production de cette centrale au même moment la veille. Cette approche ne tient pas compte des données hors site. Malgré sa popularité en tant que modèle de référence dans la littérature, sa performance globale est assez faible [10]. Pour tenir compte des différents facteurs qui affectent la production photovoltaïque, on pourrait ajuster la persistance en fonction des valeurs observées durant le jour en cours. Toutefois, cela implique déjà une certaine manipulation de données, et différentes options peuvent être envisagées. Pour éviter d'obtenir des résultats trop optimistes à partir d'une méthode spatio-temporelle, il est également nécessaire d'utiliser un modèle de «référence avancé» qui présente des performances de pointe et une complexité raisonnable afin que les résul-

tats puissent être facilement reproduits. Pour cela, nous considérons le *modèle autorégressif* (AR) défini comme suit:

$$\hat{P}_{t+h/t}^x = \hat{\beta}_h^0 + \sum_{l=0}^L \hat{\beta}_h^l P_{t-1}^x \quad (4)$$

où P_t^x est la production de la centrale x au temps t et $\hat{P}_{t+h/t}^x$ est la prévision de la production pour l'horizon h . L'ordre optimal L est très important pour la qualité de l'estimation car il implique un compromis entre biais et variance. Cet ordre maximal est choisi par minimisation du critère de l'AIC (Akaike Information Criterion) [10]. En règle générale l'AIC se calcule par maximum de vraisemblance suivant la formule:

$$\text{AIC} = -2 \log \tilde{L} + 2k \quad (5)$$

avec \tilde{L} la vraisemblance maximisée et k le nombre de paramètres du modèle.

Le modèle optimal est celui avec l'AIC le plus faible. Dans le cas du modèle AR, la vraisemblance peut être remplacée lorsque n représente la taille des données par la variance des innovations σ_L^2 , soit [10]:

$$\text{AIC} = n(\log \sigma_L^2 + 1) + 2(L + 1) \quad (6)$$

Les modèles de persistance et AR ont été appliqués au jeu de données d_1 avec des échantillons d'apprentissage et de test couvrant respectivement 15 mois et 5 mois. Les prévisions sont mises à jour au pas de temps de 15 minutes. La figure 2.1 présente l'erreur quadratique moyenne RMSE normalisée pour les modèles AR et persistance pour les centrales du jeu de données d_1 en fonction de l'horizon de prévision. Cette figure montre que le meilleur modèle est le modèle AR qui présente les niveaux de RMSE les plus bas. Les valeurs de biais et de MAE sont aussi plus basses pour le modèle AR que pour la persistance. Ce modèle servira à évaluer la contribution de l'intégration d'informations supplémentaires relatives aux centrales avoisinantes.

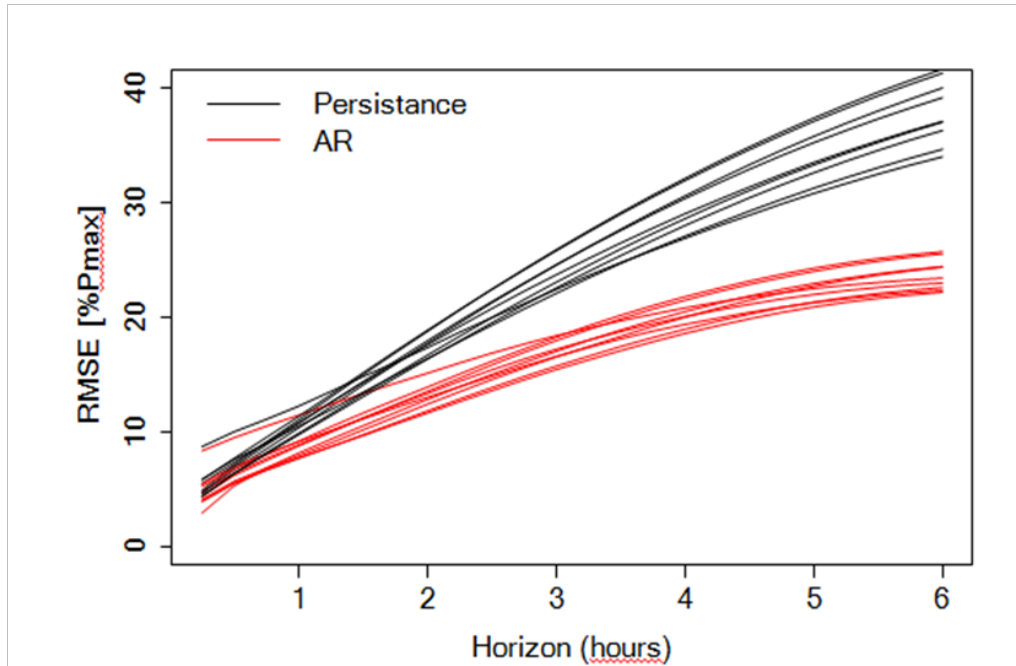


Figure 2.1 Comparaison des valeurs de RMSE normalisées des modèles AR (rouge) et persistance (noir). Le pas de temps est 15 min. Jeu de données d_1 [43].

b) Le modèle de forêts aléatoires

Une forêt aléatoire est l'agrégation d'une collection d'arbres aléatoires. Le nom forêt aléatoire vient du fait que les prédicteurs individuels sont des prédicteurs par arbre et de l'introduction de l'aléatoire dans le choix des variables de division et des échantillons. Les forêts aléatoires ont été développées par L. Breiman et font partie des méthodes déterministes les plus performantes que l'on trouve dans la littérature pour prévoir la production photovoltaïque [44].

Nous présentons ici le principe d'estimation de cette méthode et son application à la prévision. Dans ce paragraphe la variable aléatoire Y représente la production photovoltaïque, ses réalisations sont les mesures de production à chaque instant et X représente les variables explicatives. Le principe de ces méthodes est de construire une collection de prédicteurs et d'agréger ensuite l'ensemble de leurs prédictions. Dans le cadre d'une régression, si on dispose de q prédicteurs individuels qui

fournissent chacun une prévision \hat{Y}_t , agréger leur prédictions revient à faire une moyenne $1/q \sum_{t=1}^q \hat{Y}_t$.

Les forêts aléatoires sont caractérisées par [44]:

- important de prédicteurs individuels (d’arbres);
- la création pour chaque arbre d’un échantillon destiné aux tests appelé échantillon «out-of-bag»;
- le choix pour tous les arbres d’un paramètre correspondant à la taille du sous-échantillon de variables tiré aléatoirement à chaque nœud de chaque arbre;
- une variable de division choisie pour chaque arbre parmi le sous-échantillon précédemment décrit;
- l’absence d’élagage des arbres.

Pour décrire le fonction des modèles de forêts aléatoires désignons par θ le vecteur de paramètres qui détermine la construction d’un arbre (les variables de division à chaque nœud par exemple), et on notera $T(\theta)$ l’arbre correspondant. La prévision par les forêts aléatoires pour une nouvelle observation $X = x_0$ se fait suivant les étapes suivantes:

1. Construire K arbres $T(\theta_t), t=1, \dots, K$ comme décrit précédemment.
2. Faire passer x_0 dans l’arbre et conserver toutes les observations de la feuille terminale dans laquelle il tombe.
3. Calculer pour chaque arbre t des poids $w_i(x_0, \theta_t)$ tel que $w_i(x_0, \theta_t) = \frac{1}{k_t}$ si l’observation x_i de l’échantillon d’apprentissage fait partie des k_t points du nœud terminal contenant x_0 , sinon $w_i(x_0, \theta_t) = 0$.
4. La prévision pour un arbre t est alors

$$\hat{\mu}(x_0) = \sum_{i=0}^n w_i(x_0, \theta_t) Y_i \tag{7}$$

On en déduit la prévision par les forêts aléatoires en moyennant sur tous les arbres:

$$\hat{\mu}(x_0) = \sum_{i=0}^n \bar{w}_i(x_0) Y_i \quad (8)$$

avec $\bar{w}_i(x_0) = \frac{1}{k} \sum_{t=1}^k w_i(x_0, \theta_t)$.

On vient ainsi de construire une approximation de l'espérance conditionnelle $\mathbb{E}(Y|X)$ par une somme pondérée sur toutes les observations. Les pondérations, qui varient avec les covariables sont au centre de la démarche de prévision. Plusieurs études sont consacrées à la compréhension du lien entre ces pondérations et la distribution conditionnelle. Y. Lin et Y. Jeon ont montré que les poids ont tendance à être d'autant plus importants que la distribution conditionnelle de $(Y|X = X_i)$ est similaire à celle de $(Y|X)$ [45].

2.2.2 Les modèles spatio-temporels déterministes de la production éolien

L'énergie éolienne va devoir s'adapter à un nouveau mode de valorisation. Pour les producteurs, il faudra nécessairement anticiper la vente d'électricité sur le marché. Il existe donc un besoin énorme de prévisions précises. Dans ce paragraphe nous présentons quelques modèles statistiques de prévisions du productible de parcs éoliens. La prévision de la production des parcs éoliens n'est cependant pas nouvelle [46, 47, 48, 49]. Elle est établie depuis des décennies au moyen de modèles et de techniques différents.

Pour prévoir la production de parcs éoliens à l'aide de données météorologiques à court terme (de 24 à 47 heures), il existe deux approches différentes dans la littérature: l'approche directe qui consiste à prévoir directement la production à partir des entrées (principalement des variables météorologiques) et l'approche indirecte qui prévoit d'abord le vent sur le site à partir des données météorologiques puis le transforme en production. C'est une méthode en une étape contre une en deux étapes qui peuvent être résumées dans les figures 2.2 et 2.3.

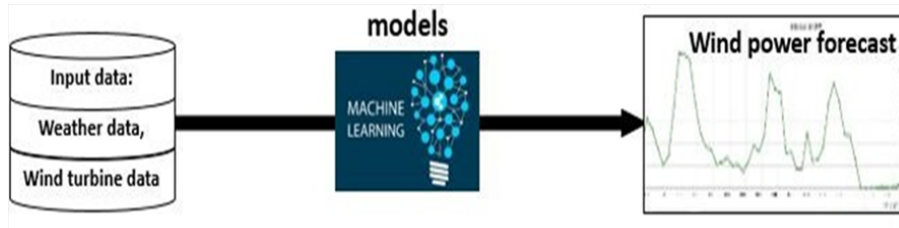


Figure 2.2 Approche directe de la prévision de la production éolienne [50].



Figure 2.3 Approche indirecte de la prévision de la production éolienne [50]

La première approche est plus simple par ce qu'elle est constituée d'un seul modèle basé sur un apprentissage direct. Cependant elle n'exploite pas les mesures réelles des vitesses du vent sur les sites. L'approche consiste à prédire la production en se basant uniquement sur les prévisions météorologiques. L'approche indirecte de son côté exploite les mesures réelles du vent qui apporte une information en plus et permet d'améliorer les prévisions météorologiques du vent avant de les convertir en prévisions de production. Cependant elle a des limites notamment si les mesures du vent par les anémomètres sont très variables (avec beaucoup d'incertitudes par exemple en cas de dysfonctionnement ou de présence de neige ou de gel sur les anémomètres). Dans ce cas la courbe de puissance qui résulte des mesures réelles du vent et de la production réelle est très incertaine et moins lisse. Par conséquent la deuxième phase de la traduction du vent prédit en prévision de production apporte plus d'incertitudes qu'elle n'améliore les prévisions de la production. Dans ce cas l'approche indirecte peut être moins avantageuse.

a) *Les données du modèle GFS*

Le Global Forecast System (GFS) est un modèle de prévisions météorologiques produit par les “National Centers for Environmental Prediction” (NCEP) dépendant du NWS (National Weather Service) des USA. Des centaines de variables atmosphériques et de terrain sont disponibles, depuis les températures, les vents, les précipitations jusqu’à l’humidité au sol et la concentration de l’ozone atmosphérique [51]. Le monde entier est couvert par le modèle. Jusqu’au mois de mars 2015 GFS fournissait des prévisions tri-horaires avec une résolution de 0.5° en latitude et en longitude (environ $56\text{km} \times 36\text{km}$) avec un horizon de prévision allant jusqu’à dix jours.

Variables	Descriptions	Unités
U_10m	Composante est-ouest du vent à 10m du sol	m/s
V_10m	Composante nord-sud du vent à 10m du sol	m/s
U_80m	Composante est-ouest du vent à 80m du sol	m/s
V_80m	Composante nord-sud du vent à 80m du sol	m/s
U_100m	Composante est-ouest du vent à 100m du sol	m/s
V_100m	Composante nord-sud du vent à 100m du sol	m/s
T_2m	Température à 2m	Kelvin
Pr_80m	Pression à 80m	Pa
RH_2m	Humidité relative à 2m	%
GUST	Rafale de vent au sol	m/s

Tableau 2.1 Les données du modèle GFS [50].

b) *Les données du modèle ECMWF:*

Le ECMWF (European Center for Medium-Range Weather Forecasts) est une organisation intergouvernementale indépendante financée par 34 Etats européens. Ces données sont intégralement mises à la disposition des services météorologiques nationaux des Etats-membres. Le Centre propose également un catalogue de produits de prévision qui peuvent être achetés. Le modèle ECMWF fournit des prévisions horaires

avec une résolution de 0.125° en latitude et en longitude (environ $13.8 \text{ km} \times 8.9 \text{ km}$) et quatre runs par jour (00h, 06h, 12h et 18h en UTC). Les prévisions de chaque run débutent 3h après l'heure du run pour un horizon de prévision pouvant aller jusqu'à dix jours [52]. Ce modèle fournit entre autres les prévisions des composantes est-ouest et nord-sud du vent à 10 m et 100 m au-dessus du sol.

Variables	Descriptions	Unités
U_10m	Composante est-ouest du vent à 10m du sol	m/s
V_10m	Composante nord-sud du vent à 10m du sol	m/s
U_100m	Composante est-ouest du vent à 100m du sol	m/s
V_100m	Composante nord-sud du vent à 100m du sol	m/s
T_2m	Température à 2m du sol	deg C
Pr	Pression à la surface	Pa
TP	Précipitation à la surface	m

Tableau 2.2 Les données du modèle ECMWF [50].

Dans la suite de ce manuscrit, nous noterons U la composante est-ouest du vent, V la composante nord-sud du vent, Pr la pression et T la température. La vitesse du vent sera notée $W = \sqrt{U^2 + V^2}$. La figure 2.4 résume le mécanisme d'acquisition et d'acheminement des données

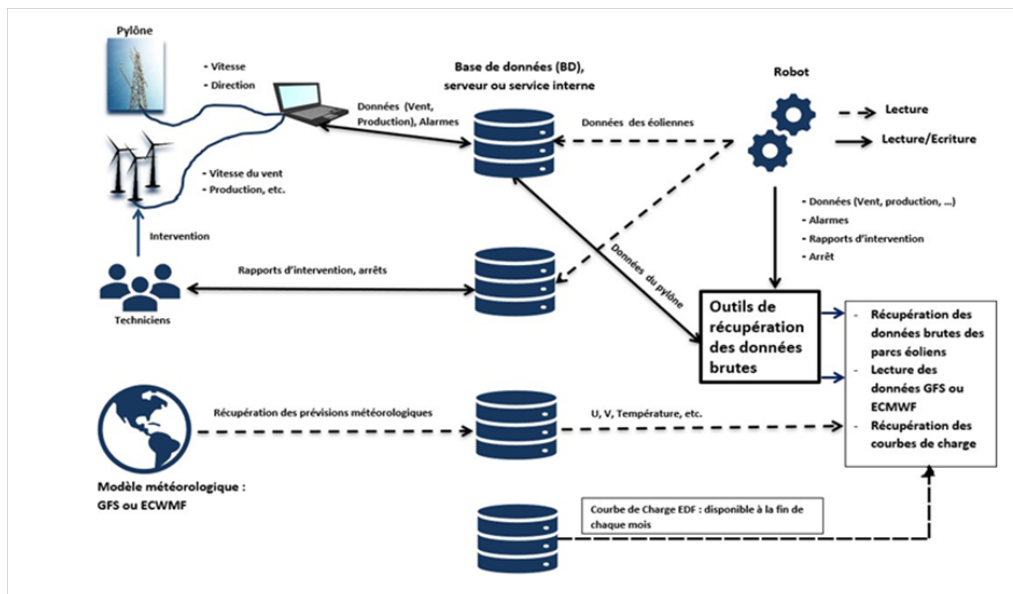


Figure 2.4 Schéma récapitulatif de l'acheminement des données [50].

2.3 Modèles spatio-temporels probabilistes

Les modèles spatio-temporels présentés dans la section précédente fournissent des prévisions déterministes qui ne permettent pas d'évaluer les incertitudes liées aux prévisions. Les prévisions probabilistes répondent à ce besoin. En effet, elles permettent non seulement de fournir plus d'informations sur la distribution future de la production mais aussi d'évaluer les incertitudes liées à la prévision. Cette quantification des incertitudes sur les prévisions est très utile pour les différents acteurs du domaine notamment pour la prise de décision et l'évaluation des risques.

2.3.1 Modèles probabilistes pour la prévision de la production photovoltaïque

Dans cette section nous présentons des méthodes de prévision probabilistes à court terme qui exploites l'information disponible à partir d'installations photovoltaïques à grande échelle. L'objectif est d'améliorer la prévisibilité pour les horizons compris entre 0 et 6 heures. Les prévisions attendues sont donc les distributions futures de la production photovoltaïque. Les prévisions météorologiques de type NWP sont utilisées comme informations complémentaires dans les modèles probabilistes présentés pour construire les densités de probabilité les plus proches possibles des réalisations. L'utilisation de ces prévisions permet aussi d'avoir un meilleur couplage entre les prévisions spatio-temporelles à très court terme et les prévisions classiques faites pour la journée à venir ($J+1$). Les performances des modèles sont évaluées avec les critères appropriés des évaluations probabilistes sur le jeu de données d_2 qui comporte un grand nombre d'installations photovoltaïques. Les problèmes de dimensionnalité et de parcimonie des modèles dus à la diversité des sources de données disponibles (données historiques de chaque centrale, prévisions NWP) sont traités par des méthodes de sélection de variables [39].

Le modèle de référence utilisé est un modèle de prévision probabiliste très courant dans la littérature de la prévision de la production

photovoltaïque: l'estimation par noyau de la densité (Kernel Density Estimation, ou KDE) [20, 38, 39]. Le modèle KDE n'exploite que les corrélations temporelles pour fournir des prévisions probabilistes de la production.

a) *L'estimation par noyau (KDE)*

La méthode d'estimation par noyau KDE est une méthode d'estimation non paramétrique de la densité. Les KDE permettent une meilleure réduction des erreurs d'estimation en comparaison aux méthodes paramétriques car elles n'admettent pas d'hypothèses sur la distribution sous-jacente au phénomène estimé. Le problème de minimisation du KDE consiste à fournir une estimation de la densité de probabilité f d'une variable aléatoire X . L'estimateur des noyaux multidimensionnel (de taille n) s'écrit :

$$\hat{f}(x) = \frac{1}{N |H|} \sum_{i=1}^N K(H^{-1}(x - x_i)) \quad (9)$$

où x représente le point d'évaluation de l'estimateur, et $x_i, i = 1, \dots, N$ sont les données. H est une matrice $n \times n$ appelée *matrice bandwidth* ou de "lissage", $|H|$ étant son déterminant. K est le noyau choisi.

Le noyau K et la matrice de lissage H sont les deux paramètres à déterminer pour mettre en œuvre un modèle de KDE. L'impact du type de noyau sur la qualité de l'estimation est faible mais les noyaux gaussiens nécessitent d'importantes capacités de calcul [54, 54]. Il est préférable donc de choisir des noyaux d'Epanechnikov dont la fonction s'écrit:

$$K(u) = \frac{3}{4\sqrt{5}} \left(1 - \frac{u^2}{5} \right), u \in [-1, 1] \quad (10)$$

Le passage à la version multi-variée donne le produit $K(u) = \prod_{j=1}^n K(u_j)$.

La matrice de lissage H est le paramètre le plus important de l'estimation KDE car elle a une grande influence sur la qualité de l'estimation. Il existe plusieurs méthodes pour choisir la matrice de lissage optimale.

Nous utilisons ici la méthode de validation croisée [56].

Soit $Y \in \mathbb{R}^p$ la variable aléatoire dont les réalisations sont la production photovoltaïque d'une centrale, $X \in \mathbb{R}^p$ les variables explicatives. La prévision de la production consiste à calculer la densité de probabilité de la variable conditionnée $Y_{t+k}|X_t$ où t est l'instant auquel la prévision est faite et k représente l'horizon de prévision. Cette densité s'obtient par:

$$f_{Y_{t+k}|X_t} = \frac{f_{Y_{t+k}, X_t}}{f_{X_t}} \quad (11)$$

La densité estimée se retrouve donc par:

$$\hat{f}_{Y_{t+k}|X_t} = \frac{1}{|H|} \sum_{i=1}^N w(x, x_i) K(H^{-1}(y - y_i)) \quad (12)$$

avec

$$w(x, x_i) = \frac{K(H^{-1}(y - y_i))}{\sum_{j=1}^N K(H^{-1}(x - x_j))} \quad (13)$$

b) La régression quantile

Lorsqu'on s'intéresse à un échantillon de réalisations d'une variable aléatoire réelle, la moyenne n'est pas toujours une statistique satisfaisante à elle seule. Une autre statistique intéressante est le quantile d'ordre α qui est défini par:

$$q_\alpha(Y) = \mathbb{F}^{-1}(\alpha) = \inf \{y, \mathbb{F}(y) \geq \alpha\} \quad (14)$$

lorsque Y a pour fonction de répartition \mathbb{F} [57]. La régression quantile est basée sur le fait que le quantile d'ordre α est solution du problème de minimisation suivant:

$$q_\alpha(Y) = \arg \min_g \mathbb{E}[\rho_\alpha(Y - g(X))] \quad (15)$$

avec $\rho_\alpha(u) = u(\alpha - \mathbf{1}_{\{u < 0\}})$ une fonction de perte appelée *pinball loss* [58]. En étendant cette approche au quantile conditionnel d'ordre α , on retrouve le problème d'estimation:

$$q_\alpha(Y|X) = \arg \min_g \mathbb{E}[\rho_\alpha(Y - g(X))|X = x] \quad (16)$$

La régression quantile permet d'évaluer comment les quantiles conditionnels de la variable d'intérêt se déforment en fonction des régresseurs X . De plus, avec la fonction pinball loss, seul le signe des écarts importe pour la minimisation. Cela se traduit par une pénalisation plus faible des très grands écarts donc une meilleure robustesse aux valeurs extrêmes ou aberrantes [59]. Dans le cas simple où la relation entre Y et X est linéaire ($Y = X'\beta + \varepsilon$, $\beta \in \mathbb{R}^p$) on construit l'estimateur:

$$\hat{\beta}_\alpha = \arg \min_{\beta} \sum_{i=1}^n \rho_\alpha(Y_i - X'_i \beta) \quad (17)$$

Cette équation n'a pas de solution explicite, le programme doit donc être résolu numériquement. Un problème important dans cette résolution numérique est que la fonction ρ_α n'est ni dérivable en 0, ni strictement convexe. Les algorithmes standards tels que celui de Newton Raphson ne peuvent être utilisés ici. Il faut alors reformuler le problème un programme d'optimisation linéaire qui peut être résolu par la méthode du simplexe dans le cas de petits échantillons ou des méthodes de point intérieur dans le cas de grands échantillons [60].

c) Extension des modèles : cas de données de grande dimension

La prévision spatio-temporelle pour les cas avec un nombre important de centrales photovoltaïques pose le problème du choix des variables d'entrée appropriées pour les modèles impliqués. Les entrées possibles incluent la série de production du site pour lequel la prévision est faite, les mesures des sites voisins et leurs séries temporellement retardées respectives ainsi que les données météorologiques. De plus, les centrales photovoltaïques peuvent être réparties sur une large surface de plusieurs kilomètres carrés, ce qui requiert de choisir adéquatement les points de grille pouvant être pris en considération. Toutes ces informations représentent une quantité importante de variables d'entrée potentielles pour les modèles de prévision et soulèvent la question de leur dimensionnalité et aussi de la parcimonie. Il faut alors faire recourir aux méthodes de sélection de variables adaptées aux modèles précédemment définis.

La méthode de sélection de variables adoptée est basée sur le critère d'information mutuelle. Cette technique a été utilisée avec succès pour sélectionner des variables pour la prévision de l'énergie éolienne [61]. La sélection de variables suivant le critère de l'information mutuelle peut être utilisée pour déterminer les variables à implémenter dans le modèle KDE.

Soit X la variable aléatoire représentant la météorologie, Y celle de la production photovoltaïque, f_X et f_Y leurs fonctions de densité de probabilité respective. L'information mutuelle permet d'évaluer la distance entre la fonction de densité de probabilité de la loi jointe $f_{X,Y}$ (entre la production et la variable météorologique), et celle de la loi jointe $f_X \cdot f_Y$. L'égalité de ces deux densités signifie que X et Y sont indépendants. De plus, une valeur élevée de l'information mutuelle traduit une forte dépendance entre X et Y . Cette évaluation de distance est faite pour chacune des variables météorologiques. Les variables météorologiques sont ensuite classées en fonction de la valeur de l'information mutuelle calculée. L'information mutuelle moyenne pour les variables Y et X s'obtient par la formule:

$$I(X, Y) = \int f_{X,Y} \log\left(\frac{f_{X,Y}}{f_X f_Y}\right) \quad (18)$$

L'information mutuelle entre X et Y est bornée par le minimum des entropies de X et Y [62]. Cette valeur seuil est utilisée pour normaliser les valeurs du critère d'information mutuel qui à leur tours sont calculées pour les prévisions NWP afin de choisir les plus adéquates pour la mise en œuvre des KDE.

2.3.2 Modèles probabilistes pour la prévision de la production éolienne

Dans un souci d'amélioration des prévisions de la production éolienne, il est important de prendre en compte les aspects métiers notamment de la météorologie. A partir des prévisions météorologiques un certain nombre de phénomènes peuvent être modélisés par post-engineering pour servir de covariables et potentiellement améliorer les prévisions de la production. Dans la plupart des études anciennes de prévision de la production éolienne, les prévisions sont faites sans tenir compte convenablement des dépendances spatio-temporelles observées dans le domaine. Cependant depuis quelques années, la structure spatio-temporelle commence à être intégrée dans la prévision éolienne [47]. Nous sommes intéressés d'abord à l'apport de la dynamique spatio-temporelle du vent et à l'impact la turbulence sur la prévision de la production éolienne. Ensuite nous prenons en compte la sélection des points en se basant sur une mesure d'importance des points de grille.

a) L'impact des dérivées spatio-temporelles du vent

La modélisation de la dynamique spatio-temporelle du vent dans la prévision de la production éolienne n'est pas nouvelle. Plusieurs techniques de modélisation existent dans la littérature. L'analyse et la modélisation spatio-temporelle pour la prévision court terme ont été déjà étudiées par Ben Miloud F. et al. [63] avec des prévisions de la production de parcs éoliens situés au Kaberten (75 km au nord de la wilaya d'Adrar) et à Zaouiat Kounta (80 km au sud d'Adrar). Les auteurs ont démontré une structure spatio-temporelle des erreurs de prévision de la production. Ils ont étudié l'impact de la vitesse et de la direction sur la nature et la forme de la structure et ils ont proposé un modèle pour capturer cette structure. Après un clustering sur les deux parcs, les ACFs (Auto-Correlation Function) et CCFs (Cross-Correlation Function) ont permis d'étudier la dépendance intra-groupe et inter-groupe respectivement pour démontrer la corrélation temporelle et spatiale. Les

données sur la vitesse et la direction du vent ont été fournies par la station météorologique d'Adrar située à la latitude $27^{\circ}40'$ N et $8^{\circ}06'$ W de longitude, et à une altitude de 263 m au-dessus de la mer mar. Les données ont été mesurées périodiquement toutes les 3 heures régulièrement à des hauteurs de 10 mètres au-dessus du sol et pendant la période entre 1995 et 1999. La fonction de probabilité de Weibull a été utilisée pour estimer le climat éolien régional de la région d'Adrar. Les valeurs finales obtenues pour les paramètres de forme et d'échelle sont respectivement de 2,05 et 6,7 m/s. La haute valeur de forme fait intervenir des vents réguliers, concentrés autour d'une valeur moyenne. Avec la valeur du facteur d'échelle enregistrée, on peut conclure qu'Adrar présente un bon état en terme de ressource éolienne pour accueillir un parc éolien.

La figure 2.5 montre la fréquence de la vitesse du vent et le diagramme de la rose des vents. On remarque que la courbe de la distribution de Weibull ne correspondait pas parfaitement à la vitesse du vent mesurée sur l'histogramme. Ceci est principalement dû à la faible valeur de densité de fréquence estimée à 1m/s. Concernant le diagramme de la rose des vents, on peut noter que le vent dominant la direction dans la région d'Adrar vient du secteur nord-est.

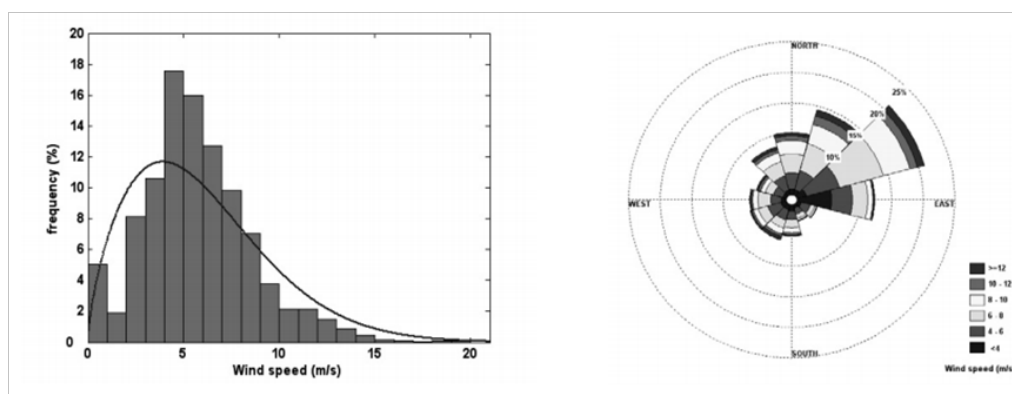


Figure 2.5 La distribution de Weibull ($A=0.7$ m/s et $k=2.17$), l'histogramme et la rose des vents à 10 m de hauteur [64].

Les données de vent correspondant aux 5 années de mesures présentées ci-dessus ont été analysées par le WASP (Programme d'analyse et

d'application du vent). C'est un outil puissant qui a été largement utilisé pour prédire la ressource éolienne. Il a été développé par le Laboratoire national de Risø et qui est basé sur un écoulement linéaire maquette [65]. Le logiciel WASP prend également en compte des données relatives à:

- un modèle numérique de la zone,
- un modèle de variation de rugosité,
- un modèle pour les obstacles voisins.

La carte numérique du terrain est dérivée des données gratuites NASA SRTM (Shuttle Radar topographic Mission). Elle est basée sur une résolution de 90 m [65]. En plus du soulagement, la rugosité est un paramètre qui influence la vitesse du vent. Selon la nature du terrain de surface de cette région, deux classes de la rugosité peuvent être distinguées:

- la zone désertique : 0,001.
- la ville : 0,5.

La cartographie des ressources éoliennes de la région à une hauteur de 80 mètres est réalisée par le logiciel WASP avec une résolution de 300 mètres et pour toute la zone d'intérêt. Les ressources éoliennes de la région sont montrées sur la figure 2.6. On constate que la valeur de la vitesse du vent est comprise entre 8,75 m/s et 9,38 m/s. D'après la carte obtenue, nous sélectionnons deux sites venteux, Kaberten situé au nord et Zouiat Kounta, situé dans le sud de la ville d'Adrar.

b) Modélisation des variations spatio-temporelles

Les variables temporelles météorologiques ont été calculées et intégrées dans l'entrée du modèle. Si le modèle est linéaire, alors les différences sont directement prises en compte par le modèle. Cependant, le modèle est non linéaire, c'est pourquoi la dynamique spatio-temporelle du vent est utilisée comme entrée du modèle.

Considérons par exemple la composante est-ouest du vent à 10m, au point de grille i , à l'instant t noté $U_{10m}(i, t)$. Les variations temporelles sont définies par $U_{10m}(i, t) - U_{10m}(i, t - 1)$. La même chose est faite pour toutes les variables météorologiques. Le principe est résumé sur la figure 2.7 où $X(i, t)$ représente les données météorologiques en un point de grille i à l'instant t . La généralisation avec les autres points de grille se fait de la même manière.

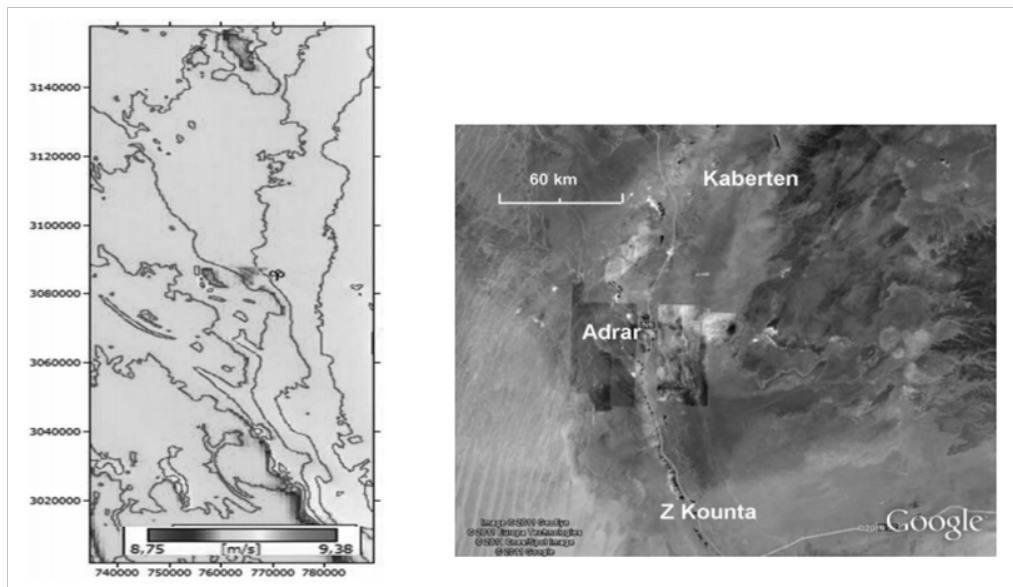


Figure 2.6 La carte des vitesses des vents (à gauche) et la localisations des sites (à droite).

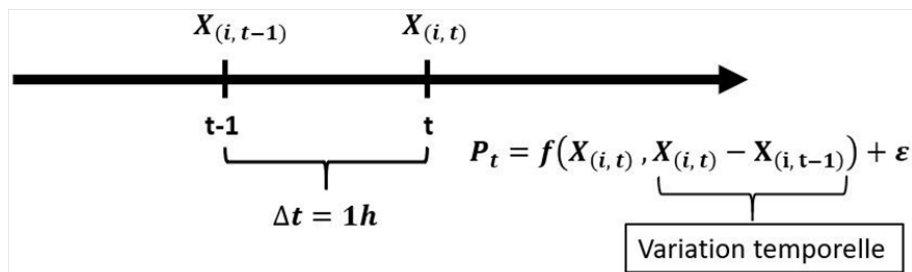


Figure 2.7 Modélisation de la variation temporelle

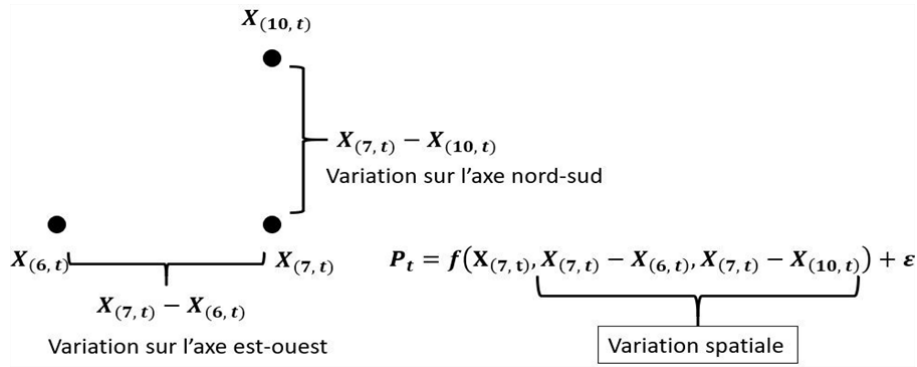


Figure 2.8 Modélisation de la variation spatiale

Pour modéliser la dynamique spatiale du vent en chaque point du réseau à l'instant t , les écarts verticaux et horizontaux entre ce point et les deux autres points du réseau sont calculés pour les composantes du vent (U , V , W et Dir) en haut sur l'axe vertical et à gauche sur l'axe horizontal en même temps. Par exemple, pour U_{10m} au point de grille 6 à l'instant t on a $(U_{10m}(6, t) - U_{10m}(7, t), U_{10m}(6, t) - U_{10m}(10, t))$ comme variations spatiales. La figure 2.8 illustre cet exemple. Le calcul sur les autres points de grille se généralise de la même manière. Les différences spatiales et temporelles ont été incorporées comme données d'entrée dans le modèle pour étudier l'effet des changements spatio-temporels sur les prévisions de production d'énergie éolienne.

c) Turbulence du vent et prévision éolienne

La turbulence se caractérise par un champ de vitesses dont les directions, les sens et les vitesses des particules qui le composent ne présentent aucune similarité : à un intervalle de temps au même endroit, ou bien en se décalant dans l'espace au même instant on ne retrouve pas de symétries de la vitesse d'une particule. Dans l'énergie éolienne, la turbulence est évaluée par son intensité. L'intensité de la turbulence (IT) est calculée en divisant l'écart type d'une mesure de la vitesse du vent sur des intervalles de 10 minutes par la vitesse moyenne du vent [66]:

$$IT = \frac{\sigma_w}{\bar{W}} \quad (19)$$

où σ_w est l'écart type de la vitesse du vent par rapport à la vitesse moyenne du vent et \bar{W} est la vitesse moyenne du vent.

Cependant les modèles météorologiques fournissent en général des prévisions de fréquence horaire rendant impossible le calcul de la turbulence telle que définie précédemment. Il existe une autre définition basée sur les vitesses du vent à différents endroits au même instant t . Dans ce cas, pour chaque point de grille, σ_w est prise comme étant l'écart type des vitesses du vent sur seize points de grille et \bar{W} est la moyenne estimée du vent par le modèle météorologique au point de grille en question. C'est cette définition qui nous permet de calculer la turbulence en se servant du vent aux des points de grille et de l'utiliser en entrée supplémentaire pour la prévision de la production éolienne. Cela fait intervenir le modèle indirect avec les forêts aléatoires. On s'est restreint par simplicité au parc éolien n° 1. Les résultats étant les mêmes sur les autres sites éoliens.

d) Incertitudes liées à la prévision de la production

L'analyse des erreurs de prévision est un aspect important pour évaluer les performances d'un modèle de prévision. Une bonne connaissance des erreurs de prévision peut aider à améliorer la qualité d'un modèle de prévision afin de réduire les risques sur la vente d'énergie. Les mesures d'erreurs classiques (NMAE et NRMSE) ne prennent pas en compte le décalage horizontal ou la distance temporelle entre les prévisions et les observations. C'est pourquoi, il faut également étudié les erreurs temporelles de la prévision éolienne.

Les modèles de prévision éolienne qu'on a rencontré dans les paragraphes précédents, présentent deux sources d'incertitudes: l'incertitude liée à la prévision du vent qui découle de la forêt aléatoire et des données météorologiques et l'incertitude liée à la courbe de puissance estimée par une spline. Il faut s'appuyé sur ces deux incertitudes pour construire l'intervalle de prévision associé aux prévisions de production éolienne dans le cadre d'un modèle indirect schématisé dans la figure 2.9.

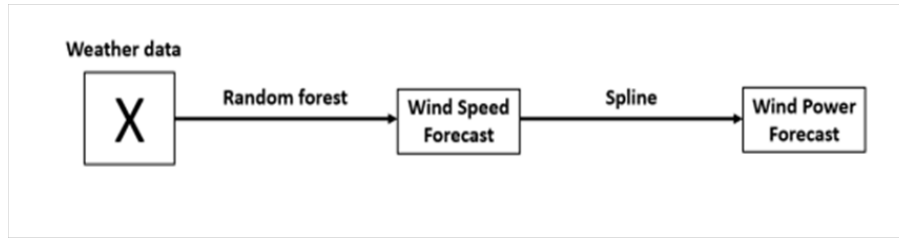


Figure 2.9 Modèle de prévision indirect de la production [50].

2.4 Conclusion

Dans la première partie de ce chapitre on a présenté les bases des modèles spatio-temporels déterministes utilisés pour la prévision de la production photovoltaïque et éolienne. On a vu que ces modèles déterministes ne permettent pas d'évaluer les incertitudes liées aux prévisions. Ainsi, la deuxième partie de ce chapitre avait comme objet les modèles de prévision probabilistes qui répondent à ce besoin. Ces modèles permettent non seulement de fournir plus d'informations sur la distribution future de la production photovoltaïque mais aussi d'évaluer les incertitudes liées à la prévision. Cette quantification des incertitudes sur les prévisions est très utile pour les différents acteurs du domaine notamment pour la prise de décision et l'évaluation des risques. Un jeu de données réelles avec un nombre élevé d'installations photovoltaïques a été considéré pour illustrer le potentiel de l'approche spatio-temporelle probabiliste. Nous avons aussi considéré l'apport des variations spatio-temporelles sur la prévision du productible de parcs éoliens et nous avons montré que la modélisation des variations spatio-temporelles permet de réduire les erreurs de prévision. L'étude de la turbulence du vent, accessible à travers les modèles météorologiques, montre que la turbulence n'est pas très importante dans la prévision éolienne et que l'intensité de turbulence est très faible à cause d'une forte corrélation des vitesses du vent d'un point de grille à un autre.

Chapitre 3

Analyse comparative et perspectives

3.1 Introduction

Dans ce dernier chapitre, nous allons mener une analyse comparative évaluant les modèles spatio-temporels des prédictions à court terme présentés dans ce mémoire. Nous allons considérer leurs capacités de prédiction et de leurs impacts, que ce soit dans l'amélioration de la production ou dans l'identification des lacunes afin de les réduire dans la production de l'énergie éolien ou la production d'énergie photovoltaïque. Nous concluons par donner des perspectives liées aux efforts nécessaires pour améliorer les modèles de prévision dans le domaine des énergies renouvelables, et des suggestions pour surmonter les difficultés rencontrées dans cette étude. Ce chapitre contient les paragraphes principaux suivants:

- Une analyse comparative: qui traite les performances des différents modèles présentés dans les chapitres précédents.
- Perspectives: qui donne des suggestions concernant l'utilisation de modèles de probabilité dans les prévisions de la production d'énergie.

3.2 Analyse comparative

3.2.1 Critères d'évaluation des prévisions probabilistes

L'évaluation des prévisions probabilistes nécessite des critères spécifiques qui prennent en compte les différents points de la densité et non seulement la moyenne. Les prévisions fournies par les modèles KDE seront évaluées par des critères d'évaluation liés aux quantiles. Ces

critères d'évaluation des quantiles des distributions prévues sont très utilisés dans la littérature [67].

Soit \hat{q}_{t+h}^α le quantile conditionnel de niveau α estimé à l'instant t pour l'horizon h . Les principaux critères d'évaluation utilisés sont:

1) la fiabilité d'un quantile estimé: on la définit comme la proportion excédante constatée en prévision pour le quantile

$$\text{rel}_h^\alpha = \alpha - \frac{1}{n} \sum_{t=}^n m_{\hat{P}_{t+h} \leq \hat{q}_{t+h}^\alpha} \quad (20)$$

où $m_{\hat{P}_{t+h} \leq \hat{q}_{t+h}^\alpha}$ est la proportion de données constatées en dessous du quantile estimé.

2) la statistique de Kolmogorov-Smirnov KS: qui est la plus grande différence en valeur absolue entre la proportion théorique d'excédance et celle observée pour la prévision, (voir figure 3.1).

3) Le MAEP (Mean Absolute Excess Probability): qui est l'aire entre la courbe de la proportion d'excédance théorique et celle obtenue en prévision, (voir figure 3.1).

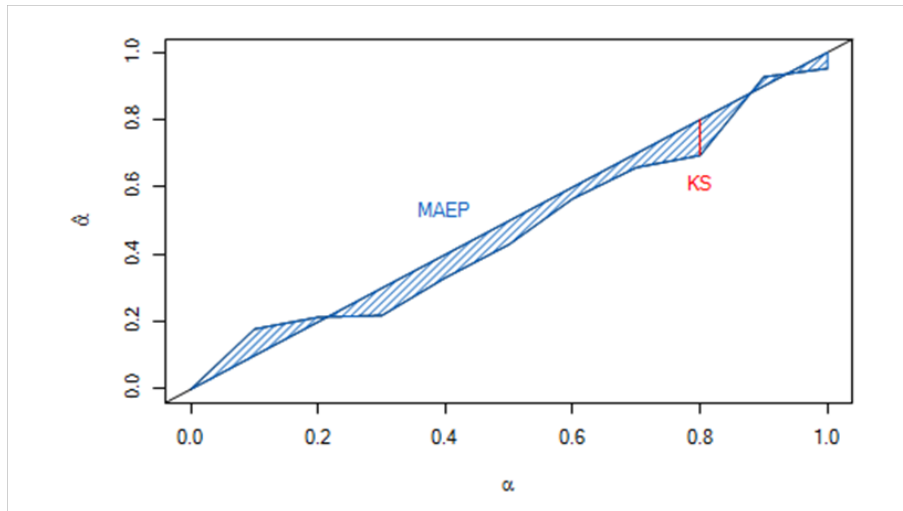


Figure 3.1 Les critères de KS et de MAEP [67].

3.2.2 Évaluation des méthodes de sélection des variables

La procédure de sélection de variable permet de choisir les variables pertinentes parmi de nombreuses entrées, mesures et prévisions issues de modèles NWP disponibles dans les cas où l'on dispose d'un grand nombre d'installations photovoltaïques. On peut comparer les valeurs du critère d'information mutuelles normalisées entre la production photovoltaïque de quatre centrales photovoltaïques et les variables NWP. L'information mutuelle ne doit pas être calculée pour la variable de rayonnement solaire à la surface ou sur plan incliné, puisque le rayonnement solaire est à la base du processus de production photovoltaïque. La variable de rayonnement solaire à la surface est intégrée au modèle. L'étude comparative montre que les deux variables les plus importantes sont la température et l'humidité relative, suivies par la direction du vent. L'importance du niveau de précipitation est la plus faible. Le modèle KDE de référence sera alors multivarié avec les quatre variables exogènes suivantes :

- le rayonnement solaire net au sol (TSR) ;
- la température à 2 m du sol (2T) ;
- l'humidité relative (RH) ;
- la direction du vent (10U/V).

L'efficacité de la prévision par la méthode du KDE avec sélection de variables en amont par le critère de l'information mutuelle est aussi évaluée. Pour cela, la qualité des quantiles prévus est examinée visuellement à la fois au sens des critères d'évaluation probabiliste de prévision. L'analyse des données liées à la production prédite avec le modèle KDE multivarié pour deux centrales montre que les intervalles inter-quantiles

sont larges traduisant une sur-estimation des densités pour les niveaux de production les plus élevés. De plus cette taille d'intervalles de confiance ne permet pas de juger efficacement de la fiabilité des estimations.

La figure 3.2 présente pour les déciles (intervalles prédictifs), la fiabilité (reliability) ou l'écart observé par rapport à la couverture nominale pour les horizons de 3h. La fiabilité parfaite signifie qu'il n'y a pas d'écart et est représentée par la ligne horizontale noire ($y = 0$) sur la figure. Deux écarts sont tracés pour le modèle KDE représentant deux matrices de lissage différentes H1, H2. Ces deux matrices de lissage sont respectivement les matrices optimales des méthodes *unbiased cross validation* et *smooth cross-validation* [56]. La figure montre que selon les valeurs des matrices de lissage, les densités prédictives peuvent être surestimées (ligne rouge pour KDE/H1) ou sous-estimées (ligne verte pour KDE/H2). Les valeurs de précision des quantiles (*sharpness*) sont présentées sur la figure 3.3 pour les mêmes intervalles de prévision que celui de la fiabilité. La figure montre que la *sharpness* augmente avec le quantile nominal (taux de couverture).

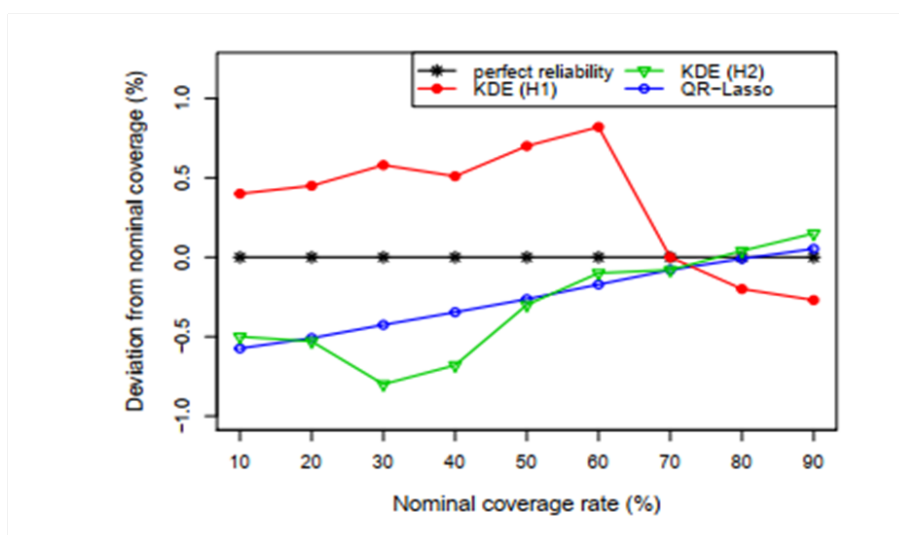


Figure 3.2 Fiabilité des quantiles estimés pour 3 heures d'horizon. Pour le modèle KDE, deux matrices de lissage ont été utilisées (courbes rouges et vertes) pour montrer l'impact de la matrice de lissage sur la fiabilité.

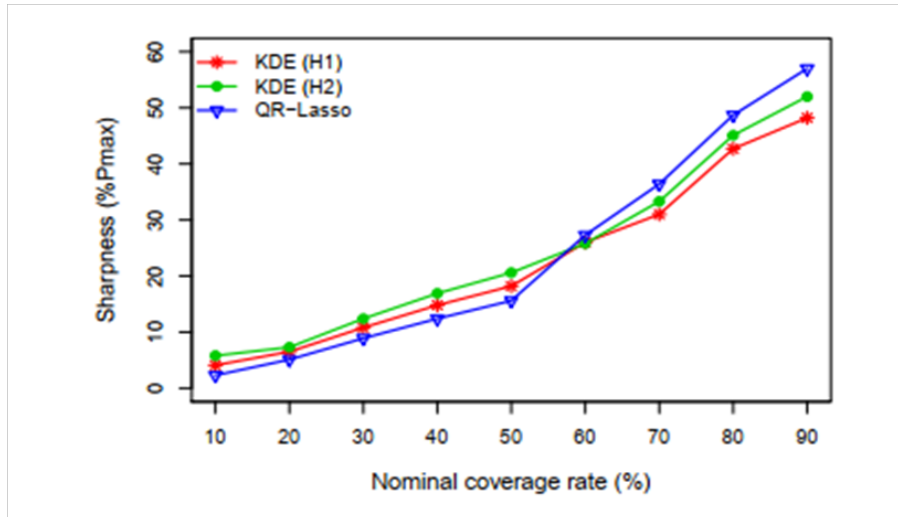


Figure 3.3 Sharpness des densités prévues pour des horizons 3h. Les courbes rouges et vertes représentent les modèles KDE avec deux cas de matrice de lissage.

3.2.3 Evaluation de la production de l'énergie éolienne

L'évaluation du potentiel éolien de la région d'Adrar a été faite à l'aide du logiciel WASP [64]. Selon la carte des ressources éoliennes obtenue, deux sites (Kaberten et Zouiat Kounta) qui se situent à proximité d'un axe 220/30 transformateur KV ont été sélectionnés. En termes de ressources éoliennes, les cartes des vents montrent que la région d'Adrar est le meilleur endroit pour installer des éoliennes, où la vitesse du vent est d'environ 9 m/s à 80 m/s. Les résultats montrent que la perte de sillage de la première région est supérieure à la seconde. Ceci indique que les générateurs d'éoliennes ont une meilleure position dans la deuxième zone. L'AEP net a été calculé pour chaque site et il a été constaté que dans le site de Zaouiat Kounta le net électrique de la production par an est d'environ 40 GWh et que les heures à pleine charge sont d'environ 4040 h. De plus, la région dispose de tous les arguments qui plaident en faveur d'une installation éolienne, à savoir accessibilité, faible densité de population, espaces nus importants et absence de zones protégées. Les éoliennes, les paramètres de Weibull, la vitesse moyenne du vent et le vent moyen unitaire densité de puis-

sance pour les deux sites sont données dans le tableau 3.1.

Kaberten site results							
	Height, m a.g.l.	Scale factor, m/s	Shape factor	V m/s	P W/m ²	Net AEP GWh	Wake loss [%]
Turbine1	250	10,2	2,69	9,08	687	8,519	0,01
Turbine 2	245	10,2	2,69	9,08	686	8,513	0,03
Turbine 3	249	10,2	2,69	9,07	686	8,508	0,03
Turbine 4	247	10,2	2,69	9,07	686	8,510	0,03
Turbine 5	248	10,2	2,69	9,07	686	8,508	0,01
Zouia Kounta site results							
	Height, m a.g.l.	Scale factor, m/s	Shape factor	V m/s	P W/m ²	Net AEP GWh	Wake loss [%]
Turbine1	269	10,3	2,69	9,18	710	8,661	0,02
Turbine 2	275	10,4	2,69	9,22	720	8,721	0,03
Turbine 3	277	10,4	2,69	9,21	718	8,713	0,02
Turbine 4	280	10,3	2,69	9,18	711	8,661	0,04
Turbine 5	280	10,4	2,68	9,23	725	8,736	0,03

Tableau 3.1 Énergie produite annuellement par chaque éolienne [64].

Les éoliennes installées sur le site de Kaberten ont les mêmes résultats car elles sont positionnées au même taille. Dans le deuxième domaine, l'énergie annuelle unitaire estimée respective produite diffère d'une machine à l'autre, à cause de la différence d'altitude. Les faibles pertes sont données par WASP et dépendent essentiellement de la position d'une machine par rapport à une autre. Ce facteur est petit, donc les turbines ont le meilleure position sur le site. Cependant, l'AEP (production nette d'énergie) nette doit être calculée après déduction de toutes les pertes qui se produisent pendant les phases de production et de transport de l'électricité produite (perte de sillage, disponibilité, pertes des lignes de transmission). La ligne de transmission utilisée dans la région d'Adrar est de 30 KV, donc conformément au résultat donné par T.R. Ackermann les pertes de transmission peuvent être supposées égales à 4 % [64]. Les pertes de disponibilité sont donnés par le constructeur et la valeur est fixée à 3%. Les valeurs de pertes faibles et totales, le Brut, l'AEP net, le facteur de capacité et une heure de pleine charge sont reporté au tableau 3.3 pour les deux sites.

Site	Average speed [m/s]	Gross AEP [GWh]	Weak loss [%]	Total loss [%]	Net AEP [GWh]	Capacity factor [%]	Full load hours [h]
Kaberten	9.08	42.56	0.02	7,07	39.57	45	3957
Z. Kounta	9.20	43.5	0.03	7,03	40.44	46	4044

Tableau 3.2 Résultats de la production électrique nette des deux sites sélectionnés [64].

Les résultats montrent que les deux sites sont propices à l'installation d'un parc éolien. Cependant, on peut s'attendre à un la planification d'un parc éolien sur Zouiat Kounta, où les productions électriques peuvent atteindre 40 GWh dans l'année et avec un facteur de capacité d'environ 45 %. Ces résultats peuvent être considérés comme prometteurs et montrent que ce site est adapté à l'énergie éolienne.

3.3 Perspectives

Une première piste pour améliorer les modèles de prévision probabilistes sera certainement liée au choix des variables d'entrée. Un tel modèle serait capable de choisir entre les mesures sur sites, les mesures de centrales voisines, les prévisions NWP, les mesures de station météo et les images satellites. Les données adéquates en fonction de l'horizon doivent être implémentées pour fournir des prévisions de la distribution future de la production photovoltaïque et éolien . Ce choix serait fait de façon automatique à l'intérieur du modèle par des méthodes de sélection de variables appropriées éliminant ainsi non seulement la possibilité de perte d'informations (par un choix subjectif/aléatoire des sources de données utilisées) mais aussi les problèmes de dimension du modèle.

Une perspective complémentaire pour un tel modèle serait de fournir des prévisions de production photovoltaïque ou éolien pour des horizons plus longs pouvant atteindre une journée. Du point de vue liée à l'utilisation des images satellites, plusieurs pistes peuvent être étudiées. Une première serait d'identifier les nuages susceptibles d'affecter

la production photovoltaïque, ainsi que la vitesse du vent ou la pression atmosphérique susceptible d'affecter la production d'énergie éolienne et d'extraire à partir des images les informations de vitesse et direction de ces nuages. Ces informations seraient ensuite intégrées dans le modèle global avec choix automatique des informations pertinentes. Une seconde piste d'amélioration liée aux images satellites serait d'évaluer les incertitudes associées aux informations extraites des images satellites et d'étudier leur impact et les erreurs qu'elles introduisent dans la prévision.

Conclusion Générale

La prévision à court terme de la production photovoltaïque et éolien est nécessaire pour le contrôle de la variance due à l’intermittence de la production, et pour maintenir de l’équilibre entre offre et demande d’électricité et la gestion des réserves. C’est aussi très utile pour que les producteurs qui participent aux marchés de l’électricité et pour la planification des opérations de maintenance et de gestion de la production. Il existe dans la littérature de nombreuses modèles pour prédire la production de l’énergie photovoltaïque et éolien pour de divers horizons.

Notre travail porte sur la prévision à court terme (de moins de 6 heures) de la production photovoltaïque et éolien basé sur l’exploitation des relations spatio-temporelles entre les différents sites de production. Ces relations sont implémentées dans des modèles probabilistes et l’évaluation de ces modèles fut l’objectif principale de notre étude. Cela nécessite de connaître l’importance de prévoir la production d’un site spécifique en utilisant des sites de production voisins comme sources informations complémentaires.

Dans le premier chapitre, nous avons présenté l’état de l’art de la prévision de production énergétique photovoltaïque et éolienne. Nous avons aussi présenté les différentes étapes de stationnarisation et les corrélations spatio-temporelles.

Nous avons commencé ce deuxième chapitre, avec une introduction sur les modèles spatio-temporels déterministes de la production d’énergie photovoltaïque et éolienne, notamment le modèle autorégressif, les modèles de persistance et le modèle de forêts aléatoires. Puisque les modèles spatio-temporels déterministes fournissent des prévisions qui ne permettent pas d’évaluer les incertitudes liées aux prévisions, nous avons opté pour les modèles de prévisions probabilistes qui font l’objet de ce travail et qui répondent à ce besoin. Les modèles spatio-temporels probabilistes que nous avons considérés pour la production

photovoltaïque sont: le modèle de l'estimation par noyau de la densité KDE, le modèle de la régression quantile et nous avons discuté les extensions de ces modèles pour le cas de données de grande dimension. En ce qui concerne les modèles probabilistes pour la prévision spatio-temporelle de la production éolien, nous avons considéré la modélisation des variations spatio-temporelles, et l'effet de la turbulence du vent sur les prévisions éoliennes et les incertitudes qui interviennent.

Dans le troisième chapitre, nous avons évalué les modèles spatio-temporels des prédictions à court terme présentés et nous avons effectué une analyse comparative qui traite les performances des différents modèles. Ensuite nous avons donné des perspectives et des suggestions concernant l'utilisation et l'amélioration des modèles probabilistes dans les prévisions de production d'énergie.

Ce travail nous a permis de noter qu'il y a une grande importance des prévisions météorologiques dans la production d'énergie, et nous avons également noté qu'avant d'installer une ferme de production d'énergie, il est nécessaire de connaître toutes les prévisions météorologiques pour cette zone (température, vent, pression de l'air, etc). Nous avons également remarqué que le développement et l'amélioration de la production dépendent fortement de ces prévisions pour minimiser les coûts.

Notre pays, riche en sources pour la production de ces énergies renouvelables, a certainement intérêt dans tous ce qui améliore la qualité des prévisions. Les modèles spatio-temporels probabilistes, étant un outil puissant dans ce domaine, sont alors des approches essentielles susceptibles d'être développées afin d'améliorer la production de l'énergie.

Nous avons rencontré plusieurs obstacles dans cette recherche, notamment le manque d'informations sur les prévisions à court terme. Nous avons constaté qu'il existe très peu de documents de recherches en Algérie sur ce sujet.

En conclusion, nous avons constaté que cette axe de recherche est d'une grande importance, en particulier pour ceux qui s'intéressent au domaine de la prévision de la production des énergies renouvelables.

Bibliographie

1. REN21 Renewables Global Status Report (GSR), 2017 edition, <<http://www.ren21/status-of-renewables/Global-status-report>>.
2. Global Wind Energy Council, Global Wind Statistics 2016, <http://gwec.net/wp-content/uploads/vip/GWEC_pRs-tats2016_EN_WEB.pdf>.
3. Jerez, Sonia, et al. "The impact of climate change on photovoltaic power generation in Europe." *Nature communications* 6.1 (2015): 1-8.
4. Moemken, Julia, et al. "Wind speed and wind energy potentials in EURO-CORDEX ensemble simulations: evaluation, bias-correction and future changes." *EGU General Assembly Conference Abstracts*. 2017.
5. Tastu, Julija, et al. "Probabilistic forecasts of wind power generation accounting for geographically dispersed information." *IEEE Transactions on Smart Grid* 5.1 (2013): 480-489.
6. Lave, Matthew, and Jan Kleissl. "Cloud speed impact on solar variability scaling—Application to the wavelet variability model." *Solar Energy* 91 (2013): 11-21.
7. Golestaneh, Faranak, Hoay Beng Gooi, and Pierre Pinson. "Generation and evaluation of space–time trajectories of photovoltaic power." *Applied Energy* 176 (2016): 80-91.
8. Antonanzas, Javier, et al. "Review of photovoltaic power forecasting." *Solar energy* 136 (2016): 78-111.

9. Kostylev, Vladimir, and Alexandre Pavlovski. "Solar power forecasting performance—towards industry standards." *1st international workshop on the integration of solar power into power systems, Aarhus, Denmark*. 2011.
10. Inman, Rich H., Hugo TC Pedro, and Carlos FM Coimbra. "Solar forecasting methods for renewable energy integration." *Progress in energy and combustion science* 39.6 (2013): 535-576.
11. Shuman, Frederick G. "Numerical weather prediction." *Bulletin of the American Meteorological Society* 59.1 (1978): 5-17.
12. Pelland, Sophie, George Galanis, and George Kallos. "Solar and photovoltaic forecasting through post-processing of the Global Environmental Multiscale numerical weather prediction model." *Progress in photovoltaics: Research and Applications* 21.3 (2013): 284-296.
13. Perez, Richard, et al. "Validation of short and medium term operational solar radiation forecasts in the US." *Solar Energy* 84.12 (2010): 2161-2172.
14. Perez, Richard, et al. "Forecasting solar radiation—Preliminary evaluation of an approach based upon the national forecast database." *Solar Energy* 81.6 (2007): 809-812.
15. Numerical weather prediction (NWP), wikipedia, <https://en.wikipedia.org/wiki/Numerical_weather_prediction>.
16. Rigollier, Christelle, Olivier Bauer, and Lucien Wald. "On the clear sky model of the ESRA-European Solar Radiation Atlas with respect to the Heliosat method." *Solar energy* 68.1 (2000): 33-48.

17. Engerer, N. A., and F. P. Mills. "KPV: A clear-sky index for photovoltaics." *Solar energy* 105 (2014): 679-693.
18. Hamilton, James Douglas. *Time Series Analysis*. Princeton University Press, 1997.
19. Pedro, Hugo TC, and Carlos FM Coimbra. "Assessment of forecasting techniques for solar power production with no exogenous inputs." *Solar Energy* 86.7 (2012): 2017-2028.
20. Bacher, Peder, Henrik Madsen, and Henrik Aalborg Nielsen. "Online short-term solar power forecasting." *Solar energy* 83.10 (2009): 1772-1783.
21. Monteiro, Claudio, et al. "Short-term power forecasting model for photovoltaic plants based on historical similarity." *Energies* 6.5 (2013): 2624-2643.
22. Simon Haykin. *Neural Networks : A Comprehensive Foundation*. Prentice HallPTR, Upper Saddle River, NJ, USA, 2nd edition, 1998.
23. Tao, Cai, Duan Shanxu, and Chen Changsong. "Forecasting power output for grid-connected photovoltaic power system without using solar radiation measurement." *The 2nd International Symposium on Power Electronics for Distributed Generation Systems*. IEEE, 2010.
24. Nello Cristianini and John Shawe-Taylor. *An Introduction to Support Vector Machines: And Other Kernel-based Learning Methods*. Cambridge University Press, New York, NY, USA, 2000.
25. Bouzerdoum, Moufida, Adel Mellit, and A. Massi Pavan. "A hybrid model (SARIMA-SVM) for short-term power forecasting of a small-scale grid-connected photovoltaic plant." *Solar energy* 98 (2013): 226-235.

26. Li, Jiaming, et al. "Machine learning for solar irradiance forecasting of photovoltaic system." *Renewable energy* 90 (2016): 542-553.
27. Huang, Jing, and Matthew Perry. "A semi-empirical approach using gradient boosting and k-nearest neighbors regression for GEFCom2014 probabilistic solar power forecasting." *International Journal of Forecasting* 32.3 (2016): 1081-1086.
28. Yang, Qingyu, Dou An, and Yuanli Cai. "A Novel Evolution Kalman Filter Algorithm for Short-Term Climate Prediction." *Asian Journal of Control* 18.1 (2016): 400-405.
29. Giebel, Gregor, et al. "The state-of-the-art in short-term prediction of wind power. A literature overview." (2011). <<https://www.osti.gov/etdeweb/servlets/purl/1011554>>.
30. Costa, Alexandre, et al. "A review on the young history of the wind power short-term prediction." *Renewable and Sustainable Energy Reviews* 12.6 (2008): 1725-1744.
31. Boulakhras Chahar, Rencontre Nationale avec MICLAT, Juin 2018, Sonelgaz. <<https://www.interieur.gov.dz/images/Projets-raliss-dans-le-cadre-du-programma-national-des-nergies-renouvelables-.pdf>>.
32. Moran, Patrick AP. "The interpretation of statistical maps." *Journal of the Royal Statistical Society. Series B (Methodological)* 10.2 (1948): 243-251.
33. Koenig, Walter D. "Spatial autocorrelation of ecological phenomena." *Trends in Ecology & Evolution* 14.1 (1999): 22-26.
34. Hoff, Thomas E., and Richard Perez. "Quantifying PV power output variability." *Solar Energy* 84.10 (2010): 1782-1793.

35. Lonij, Vincent PA, et al. "Intra-hour forecasts of solar power production using measurements from a network of irradiance sensors." *Solar energy* 97 (2013): 58-66.
36. Andrade, José R., and Ricardo J. Bessa. "Improving renewable energy forecasting with a grid of numerical weather predictions." *IEEE Transactions on Sustainable Energy* 8.4 (2017): 1571-1580.
37. Bessa, Ricardo J., et al. "Probabilistic solar power forecasting in smart grids using distributed information." *International Journal of Electrical Power & Energy Systems* 72 (2015): 16-23.
38. Golestaneh, Faranak, Pierre Pinson, and Hoay Beng Gooi. "Very short-term nonparametric probabilistic forecasting of renewable energy generation-With application to solar energy." *IEEE Transactions on Power Systems* 31.5 (2016): 3850-3863.
39. Zhang, Bei, Payman Dehghanian, and Mladen Kezunovic. "Spatial-temporal solar power forecast through use of Gaussian conditional random fields." *2016 IEEE power and energy society general meeting (PESGM)*. IEEE, 2016.
40. Patrick, Joshua D., Jane L. Harvill, and Clifford W. Hansen. "A semiparametric spatio-temporal model for solar irradiance data." *Renewable Energy* 87 (2016): 15-30.
41. Yang, Chen, Anupam A. Thatte, and Le Xie. "Multitime-scale data-driven spatio-temporal forecast of photovoltaic generation." *IEEE Transactions on Sustainable Energy* 6.1 (2014): 104-112.
42. Bracale, Antonio, et al. "A Bayesian-based approach for a short-term steady-state forecast of a smart grid." *IEEE Transactions on Smart Grid* 4.4 (2013): 1760-1771.

43. Agoua, Xwégnon. “Développement de méthodes spatio-temporelles pour la prévision à court terme de la production photovoltaïque.” PhD diss., Paris Sciences et Lettres (ComUE), 2017. p 49.
44. Breiman, Leo. “Random forests." *Machine learning* 45.1 (2001): 5-32.
45. Lin, Yi, and Yongho Jeon. “Random forests and adaptive nearest neighbors." *Journal of the American Statistical Association* 101.474 (2006): 578-590.
46. Juban, Jérémie, Lionel Fugon, and Georges Kariniotakis. “Uncertainty estimation of wind power forecasts: Comparison of probabilistic modelling approaches." *European Wind Energy Conference & Exhibition EWEC 2008*. EWEC, 2008.
47. Tastu, Julija, et al. “Spatio-temporal analysis and modeling of short-term wind power forecast errors." *Wind Energy* 14.1 (2011): 43-60.
48. Kusiak, Andrew, Haiyang Zheng, and Zhe Song. “Wind farm power prediction: a data-mining approach." *Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology* 12.3 (2009): 275-293.
49. Fugon, Lionel, Jérémie Juban, and Georges Kariniotakis. “Data mining for wind power forecasting." *European Wind Energy Conference & Exhibition EWEC 2008*. EWEC, 2008.
50. Dione, Mamadou. *Prévision court terme de la production éolienne par Machine learning*. Diss. Institut polytechnique de Paris, 2020.
51. National Weather Service, NOAA Operational Model Archive and Distribution System. <<https://nomads.ncep.noaa.gov/>>

52. Wedam, Garrett B., Lynn A. McMurdie, and Clifford F. Mass. "Comparison of model forecast skill of sea level pressure along the east and west coasts of the United States." *Weather and forecasting* 24.3 (2009): 843-854.
53. Zhang, Yao, and Jianxue Wang. "K-nearest neighbors and a kernel density estimator for GEFCom2014 probabilistic wind power forecasting." *International Journal of forecasting* 32.3 (2016): 1074-1080.
54. Mugdadi, Abdel Razzaq, and Ibrahim A. Ahmad. "A bandwidth selection for kernel density estimation of functions of random variables." *Computational Statistics & Data Analysis* 47.1 (2004): 49-62.
55. O'Brien, Travis A., et al. "A fast and objective multidimensional kernel density estimation method: fastKDE." *Computational Statistics & Data Analysis* 101 (2016): 148-160.
56. Duong, Tarn, and Martin L. Hazelton. "Cross-validation bandwidth matrices for multivariate kernel density estimation." *Scandinavian Journal of Statistics* 32.3 (2005): 485-506.
57. Roger Koenker. *Quantile Regression*. Cambridge University Press, 2005.
58. Dantzig, George Bernard. *Linear Programming and Extensions*. Princeton University Press, 1991.
59. Xavier D'Haultfoeuile & Pauline Givord. *La régression quantile en pratique*. Document de travail INSEE, Janvier 2013.
60. F. Glineur. *Etude des méthodes de point intérieur appliquées à la programmation linéaire et à la programmation semidéfinie*. <<https://perso.uclouvain.be/francois.glineur/oldwww/courses/inma2471/ReferenceIPM.pdf>>.

61. Juban, Jeremie, Nils Siebert, and George N. Kariniotakis. "Probabilistic short-term wind power forecasting for the optimal management of wind generation." *2007 IEEE Lausanne Power Tech.* IEEE, 2007.
62. Thomas, M. T. C. A. J., and A. Thomas Joy. *Elements of information theory.* Wiley-Interscience, 2006.
63. Miloud, F. Sebaa-Ben, and R. Aissaoui. "Etude du potentiel éolien d'Adrar Sélection de sites pour la ferme éolienne de 10 MW." *Séminaire Méditerranéen En Energie Eolienne* 13 (2010): 295-300.
64. Djamai, Mourad, and Nachida KASBADJI Merzouk. "Wind farm feasibility study and site selection in Adrar, Algeria." *Energy Procedia* 6 (2011): 136-142.
65. Dehmas, Djamila Abdeslame, et al. "On the use of wind energy to power reverse osmosis desalination plant: A case study from Ténès (Algeria)." *Renewable and Sustainable Energy Reviews* 15.2 (2011): 956-963.
66. Göçmen, Tuhfe, and Gregor Giebel. "Estimation of turbulence intensity using rotor effective wind speed in Lillgrund and Horns Rev-I offshore wind farms." *Renewable energy* 99 (2016): 524-532.
67. Team, R. Core. "R: A language and environment for statistical computing." (2013): 201.