



**HAL**  
open science

# Prévision multi-horizon de l'éclairage global horizontal pour la gestion intelligente du réseau électrique de distribution en région Occitanie

Shab Gbemou

► **To cite this version:**

Shab Gbemou. Prévision multi-horizon de l'éclairage global horizontal pour la gestion intelligente du réseau électrique de distribution en région Occitanie. Autre. Université de Perpignan, 2021. Français. NNT : 2021PERP0050 . tel-03641953

**HAL Id: tel-03641953**

**<https://theses.hal.science/tel-03641953>**

Submitted on 14 Apr 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THÈSE

Pour obtenir le grade de  
Docteur

Délivré par  
**UNIVERSITÉ DE PERPIGNAN VIA DOMITIA**

Préparée au sein de l'école doctorale 305 : Énergie et  
Environnement

Et de l'unité de recherche PROMES-CNRS UPR 8521

Spécialité : **Sciences de l'ingénieur**

Présentée par **Shab GBEMOU**

**Prévision multi-horizon de l'éclairage global  
horizontal pour la gestion intelligente du réseau  
électrique de distribution en région Occitanie**

Soutenue le **16 décembre 2021** devant le jury composé de :

<b>M. Philippe LAURET</b> Professeur, Université de La Réunion	Rapporteur
<b>M. Ted SOUBDHAN</b> Professeur, Université des Antilles	Rapporteur
<b>M. Gilles NOTTON</b> Maître de Conférences, Université de Corse Pascal Paoli	Examineur
<b>M. Hervé GUEGUEN</b> Professeur, CentraleSupélec	Examineur
<b>M. Éric BIDEAUX</b> Professeur, INSA Lyon	Examineur
<b>M. Laurent THOMAS</b> Professeur, Université de Perpignan Via Domitia	Examineur
<b>M. Stéphane GRIEU</b> Professeur, Université de Perpignan Via Domitia	Directeur de thèse
<b>M. Stéphane THIL</b> Maître de Conférences, Université de Perpignan Via Domitia	Co-directeur de thèse





« C'est dans la pratique qu'il faut que l'homme prouve la vérité. »  
Karl Marx



# Dédicaces

À mes parents et à Kéval.



## Remerciements

La thèse de doctorat est un travail s'inscrivant dans la durée, et pour cette raison, constitue le fil conducteur d'une tranche de vie de son auteur, parfois au crépuscule de la candeur étudiante, et souvent à l'aube de la maturité scientifique. De nombreuses personnes se retrouvent ainsi de manière fortuite ou non, pour le pire ou le meilleur, entre le doctorant et son doctorat. Ce sont certaines de ces personnes que j'aimerais mettre en avant dans ces remerciements.

Je tiens tout d'abord à remercier particulièrement Stéphane Grieu, mon directeur de thèse, et Stéphane Thil, mon co-directeur de thèse, qui m'ont encadré tout au long de cette thèse et qui m'ont fait partager leurs brillantes intuitions. Qu'ils soient également remerciés pour leur disponibilité permanente et pour les nombreux encouragements qu'ils m'ont prodigués. Ils ont permis d'améliorer significativement la qualité du présent mémoire, tant sur le plan scientifique que sur les plans sémantique et syntaxique. Merci de votre encadrement de haute qualité, de votre patience et de votre rigueur.

J'adresse mes remerciements à l'endroit de Philippe Lauret, Professeur à l'Université de La Réunion, et Ted Soubdhan, Professeur à l'Université des Antilles, pour l'honneur qu'ils m'ont fait en acceptant d'être rapporteurs de cette thèse. Je leur suis reconnaissant d'avoir examiné avec attention mon travail et d'avoir apporté des commentaires précieux qui permettront d'enrichir les travaux entrepris au cours de cette thèse.

Je remercie également Gilles Notton, Maître de Conférences à l'Université de Corse Pascal Paoli, Hervé Gueguen, Professeur à CentraleSupélec et Éric Bideaux, professeur à INSA Lyon d'avoir accepté d'être examinateurs de ma thèse et pour la qualité des échanges lors de la soutenance.

Je suis très reconnaissant envers Laurent Thomas, professeur à l'Université de Perpignan Via Domitia, d'avoir accepté de présider le jury de cette thèse.

Je remercie Hanany de m'avoir fait partager ses connaissances en programmation. J'adresse un grand merci à tous mes camarades du laboratoire pour leur disponibilité et leur sympathie au laboratoire comme en dehors. Merci pour tous ces bons moments partagés ensemble. Il me sera très difficile de citer tout le monde car c'est grâce à l'aide de nombreuses personnes que j'ai pu mener cette thèse à son terme. Néanmoins, je tiens à citer les incontournables : Nouha, Julien, Romain, Anis, Youssef, Hasan, Kadar, Danielle, Aïssatou, Antonella. Vous avez été de véritables soutiens émotionnels et même des partenaires de jeux pour d'autres. Mention spéciale à tous les membres de l'équipe de foot du laboratoire dont le capitaine est Martin.

Je ne saurais oublier ma famille, mes parents et mon petit frère, qui m'ont toujours soutenu et respecté mes choix. C'est en grande partie grâce à eux si je suis arrivé jusque-là. Qu'ils trouvent dans cette réussite le résultat de leur confiance. Un clin d'œil à ma famille en Suisse dont le courage et la détermination m'inspirent chaque jour. Enfin, je voudrais conclure en remerciant Merveille pour son soutien continu et pour son énergie communicative qui me pousse chaque jour à accomplir ce que je n'aurais pas imaginé sans elle. Je suis très heureux que nos routes se soient croisées pour ne faire plus qu'une.



# Table des matières

<b>Table des figures</b>	<b>xiii</b>
<b>Liste des tableaux</b>	<b>xvii</b>
<b>Nomenclature</b>	<b>xix</b>
<b>Introduction générale</b>	<b>1</b>
<b>1 Contexte et enjeux</b>	<b>5</b>
1.1 Gestion intelligente des réseaux électriques de distribution	5
1.1.1 Enjeux climatiques	6
1.1.2 Évolution de la consommation énergétique mondiale	6
1.1.3 Épuisement des combustibles fossiles	7
1.1.4 Essor des énergies renouvelables	7
1.1.5 Engagement de la France dans la lutte contre le réchauffement climatique	8
1.1.6 Le solaire photovoltaïque	9
1.1.6.1 Évolution de la production photovoltaïque à l'échelle mondiale	9
1.1.6.2 Pénétration de la production PV dans les réseaux électriques de distribution	10
1.1.7 Mutation du réseau électrique de distribution	11
1.1.8 Intérêt de la prévision	12
1.1.9 Le projet « Smart Occitania »	13
1.1.9.1 Présentation du projet	13
1.1.9.2 Cas d'étude	14
1.1.9.3 Stratégie proposée par PROMES-CNRS	15
1.1.9.4 Résultats et discussion	15
1.2 L'éclairement global horizontal	17
1.2.1 Fondamentaux sur le rayonnement électromagnétique	17
1.2.2 Le rayonnement solaire	17
1.2.3 Modèles ciel clair	19
1.2.3.1 L'éclairement solaire extra-atmosphérique	20
1.2.3.2 La masse d'air optique relative	21
1.2.4 L'indice ciel clair	21
1.2.5 L'indice de clarté	21
1.2.6 Mesures du GHI	22
1.2.6.1 Site de Font-Romeu-Odeillo-Via	22
1.2.6.2 Site de Perpignan	23
1.2.6.3 Site de Villeneuve-de-la-Raho	23
1.2.6.4 Contrôle qualité des mesures du GHI	25
1.2.6.5 Variabilité des données	26
1.3 État de l'art des modèles de prévision du GHI	28
1.3.1 Modèles de persistance	29



1.3.1.1	Persistence sur le GHI	29
1.3.1.2	Persistence sur l'indice ciel clair	29
1.3.2	Modèles de prévision localisée du GHI	29
1.3.2.1	Prévision par caméra à grand angle de champ	29
1.3.2.2	Prévision par imagerie satellitaire	30
1.3.2.3	Modèles numériques de prévision météorologique	31
1.3.2.4	Modèles statistiques	32
1.3.3	Prévision spatio-temporelle du GHI	32
1.4	Critères de performance	33
1.4.1	La racine carrée de l'erreur quadratique moyenne normalisée	33
1.4.2	L'erreur absolue moyenne dynamique	33
1.4.3	Le <i>skill score</i>	34
1.4.4	L' <i>interval score</i>	34
1.5	Conclusion	35
<b>2</b>	<b>Prévision du GHI par apprentissage automatique</b>	<b>37</b>
2.1	Introduction	37
2.2	Prévision multi-horizon	38
2.2.1	Stratégie itérative	40
2.2.2	Stratégie directe	40
2.2.3	Stratégie directe-itérative	40
2.3	Description des modèles développés	41
2.3.1	Les processus gaussiens	41
2.3.1.1	Définition	41
2.3.1.2	La régression par processus gaussien	41
2.3.1.3	Les fonctions de covariance	42
2.3.1.4	Algorithme d'apprentissage	45
2.3.2	Les machines à vecteurs de support	46
2.3.2.1	Régression par machine à vecteurs de support	47
2.3.2.2	Modèle LS-SVR	48
2.3.2.3	Fonction de covariance	48
2.3.3	Les réseaux de neurones artificiels	49
2.3.3.1	Le perceptron multicouche	49
2.3.3.2	Le réseau LSTM	50
2.3.3.3	Algorithme d'apprentissage	51
2.4	Résultats	52
2.5	Conclusion	54
<b>3</b>	<b>Prévision multi-horizon du GHI par GPR</b>	<b>59</b>
3.1	Introduction	59
3.2	Modèles GPR utilisant le temps en entrée	60
3.2.1	Étude comparative des modèles GPR utilisant le temps en entrée fondés sur des noyaux de base et des noyaux combinés	60
3.2.2	Modèle GPR adapté à la prévision, utilisant le temps en entrée, pour des données couvrant plusieurs années	64
3.3	Modèles GPR utilisant un historique de GHI en entrée	65
3.4	Résultats et discussion	66
3.4.1	Modèles GPR multi-horizons : temps ou observations en entrée ?	66
3.4.2	Modèles GPR multi-horizons ou mono-horizons ?	68
3.4.3	Contribution du noyau périodique annuel	69
3.5	Conclusion	76

<b>4</b>	<b>Vers la prévision spatio-temporelle du GHI par GPR</b>	<b>77</b>
4.1	Introduction . . . . .	77
4.2	État de l'art . . . . .	78
4.2.1	Les méthodes physiques . . . . .	78
4.2.2	Les méthodes statistiques . . . . .	78
4.2.2.1	Les approches classiques . . . . .	78
4.2.2.2	Le krigeage . . . . .	79
4.3	Instrumentation et données disponibles . . . . .	82
4.4	Processus gaussien spatio-temporel . . . . .	85
4.4.1	Notations . . . . .	85
4.4.2	Propriétés des noyaux spatio-temporels . . . . .	85
4.4.3	Noyau spatio-temporel utilisé . . . . .	86
4.5	Résultats et discussion . . . . .	86
4.5.1	Évaluation des modèles . . . . .	87
4.5.2	Journées-types . . . . .	88
4.6	Conclusion . . . . .	91
	<b>Conclusion générale</b>	<b>99</b>
	<b>Bibliographie</b>	<b>101</b>



# Table des figures

1	Mix électrique de la France en 2019. . . . .	2
1.1	Évolution de la puissance photovoltaïque installée dans le monde entre 2010 et 2020. . . . .	9
1.2	Puissance photovoltaïque installée en France par région au 31/12/2020. . . . .	10
1.3	Carte des projets de démonstrateurs de réseaux électriques intelligents portés par ENEDIS. . . . .	13
1.4	Le cas d'étude. . . . .	14
1.5	Aperçu de deux pyréliomètres montés sur traqueurs solaires. . . . .	18
1.6	Deux RSI installés au laboratoire PROMES-CNRS à Perpignan. . . . .	19
1.7	Illustration de six jours de données GHI issues de mesures réalisées au laboratoire PROMES-CNRS à Perpignan. . . . .	22
1.8	Aperçu d'un pyranomètre CM5 de Kipp & Zonen. . . . .	23
1.9	Données GHI couvrant les années 2018 et 2019 tirées de mesures réalisées au laboratoire PROMES-CNRS à Font-Romeu-Odeillo-Via. . . . .	23
1.10	Données GHI couvrant les années 2016 et 2017 tirées de mesures réalisées au laboratoire PROMES-CNRS à Perpignan. . . . .	24
1.11	Les 16 pyranomètres installés à Villeneuve-de-la-Raho. . . . .	24
1.12	La caméra et l'un des pyranomètres installés à Villeneuve-de-la-Raho. . . . .	25
1.13	Répartition des situations atmosphériques selon l'indice ciel clair à Font-Romeu-Odeillo-Via (01/01/2018 – 31/12/2019) et Perpignan (01/01/2016 – 31/12/2017). . . . .	27
1.14	Approches pour la prévision de la ressource solaire, en fonction de l'horizon spatio-temporel. . . . .	30
1.15	<i>PROMES sky imager</i> installé sur le toit du grand four solaire à Font-Romeu-Odeillo-Via et quelques images HDR acquises. . . . .	31
2.1	Différence entre les dynamiques du GHI échantillonné à 10 min et à 1 h. . . . .	38
2.2	Représentation graphique des noyaux de base usuels et d'un noyau combiné avec les valeurs de leurs hyperparamètres et exemples de réalisations. . . . .	44
2.3	Modèle de prévision GPR fondé sur le noyau $k_{SE-ARD}$ . . . . .	45
2.4	Modèle de prévision GPR fondé sur le noyau $k_{RQ-ARD}$ . . . . .	46
2.5	Modèle de prévision LS-SVR fondé sur le noyau $k_{RBF}$ . . . . .	49
2.6	Modèle de prévision MLP : la topologie du réseau est définie comme une combinaison de couches cachées et de couches <i>dropout</i> . La couche de sortie fournit la prévision. $[GHI(t - \Delta t), GHI(t - 2\Delta t), \dots, GHI(t - D\Delta t)]^T$ est un vecteur d'observations et $\Delta t = 10\text{min}$ est le pas de temps. . . . .	50
2.7	Modèle de prévision LSTM : la topologie du réseau est définie comme une combinaison de couches LSTM et de couches <i>dropout</i> . La couche entièrement connectée fournit la prévision. $[GHI(t - \Delta t), GHI(t - 2\Delta t), \dots, GHI(t - D\Delta t)]^T$ est le vecteur d'observations et $\Delta t = 10\text{min}$ est le pas de temps. . . . .	51
2.8	Valeurs des critères de performance pour tous les modèles considérés dans l'étude comparative en fonction de l'horizon de prévision. . . . .	52

2.9	Exemples de prévisions à 10 min du GHI obtenues grâce aux modèles considérés dans l'étude comparative. . . . .	55
2.10	Exemples de prévisions à 1 h du GHI obtenues grâce aux modèles considérés dans l'étude comparative. . . . .	56
2.11	Exemples de prévision à 3 h du GHI obtenues grâce aux modèles considérés dans l'étude comparative. . . . .	56
2.12	Exemples de prévision à 5 h du GHI obtenues grâce aux modèles considérés dans l'étude comparative. . . . .	57
2.13	Exemples de prévision à 24 h du GHI obtenues grâce aux modèles considérés dans l'étude comparative. . . . .	57
3.1	Stratégie itérative . . . . .	60
3.2	Stratégie directe . . . . .	60
3.3	Erreur nRMSE en fonction de l'horizon de prévision pour le modèle de persistance sur le GHI et les modèles GPR avec différents noyaux de base et noyaux combinés. . . . .	62
3.4	Décomposition du GHI journalier. . . . .	63
3.5	Données du GHI couvrant les années 2016 et 2017. Une composante périodique saisonnière apparaît lorsqu'on considère des données sur plusieurs années. Les mesures ont été réalisées, à l'aide d'un pyranomètre à bande rotative (RSI), au laboratoire PROMES-CNRS à Perpignan. . . . .	64
3.6	Modèle de prévision GPR fondé sur le noyau $k_{\text{Per} \times \text{RQ}}$ . . . . .	65
3.7	Modèle de prévision GPR fondé sur le noyau $k_{\text{Per} \times \text{RQ} + \text{Per}}$ . . . . .	65
3.8	Modèle de prévision GPR à un pas utilisant des observations en entrée, donnant un modèle multi-horizon par itération. . . . .	66
3.9	Modèle de prévision GPR mono-horizon utilisant des observations en entrée. . . . .	66
3.10	Valeurs des critères de performance pour les modèles GPR utilisant le temps en entrée et utilisant des observations du GHI. . . . .	67
3.11	Valeurs des critères de performance pour les modèles GPR utilisant le temps en entrée et pour les modèles GPR mono-horizons utilisant des observations du GHI. . . . .	69
3.12	Valeurs des critères de performance pour les modèles GPR utilisant le temps en entrée. . . . .	70
3.13	Exemples de prévision du GHI à 10 min sur sept jours avec le modèle de persistance sur l'indice ciel clair et les modèles GPR. . . . .	71
3.14	Exemples de prévision du GHI à 1 h sur sept jours avec le modèle de persistance sur l'indice ciel clair et les modèles GPR. . . . .	72
3.15	Exemples de prévision du GHI à 3 h sur sept jours avec le modèle de persistance sur l'indice ciel clair et les modèles GPR. . . . .	73
3.16	Exemples de prévision du GHI à 5 h sur sept jours avec le modèle de persistance sur l'indice ciel clair et les modèles GPR. . . . .	74
3.17	Exemples de prévision du GHI à 24 h sur sept jours avec le modèle de persistance sur l'indice ciel clair et les modèles GPR. . . . .	75
4.1	Carte montrant la position des 16 pyranomètres installés à Villeneuve-de-la-Raho et la grille de $100 \times 100$ considérée pour les prévisions. . . . .	83
4.2	Pyranomètres aux points C, D et K, installés à Villeneuve-de-la-Raho sur des lampadaires et un poteau électrique. . . . .	83
4.3	Mesures du GHI sur quelques jours réalisées à l'aide des 16 pyranomètres installés à Villeneuve-de-la-Raho. . . . .	84
4.4	Exemple de mesures du GHI à Odeillo, avec 4 pyranomètres espacés de 50 m à 100 m et un échantillonnage de 1 s. . . . .	84
4.5	Valeurs des critères de performance pour les modèles GPR spatio-temporels développés en considérant les données de test couvrant quinze jours (du 16/08/2019 au 30/08/2019). . . . .	87

4.6	Mesures du GHI (16 pyranomètres installés à Villeneuve-de-la-Raho) au cours de la journée « ciel clair » du 16/08/2019. . . . .	88
4.7	Mesures du GHI (16 pyranomètres installés à Villeneuve-de-la-Raho) au cours de la journée partiellement couverte du 21/08/2019. . . . .	88
4.8	Mesures du GHI (16 pyranomètres installés à Villeneuve-de-la-Raho) au cours de la journée complètement couverte du 27/08/2019. . . . .	89
4.9	Valeurs des critères de performance pour la journée ciel clair du 16/08/2019. . . . .	89
4.10	Valeurs des critères de performance pour la journée partiellement couverte du 21/08/2019. . . . .	90
4.11	Valeurs des critères de performance pour la journée complètement couverte du 27/08/2019. . . . .	91
4.12	Exemples de prévision à 10 min obtenue grâce aux modèles GPR testés. . . . .	93
4.13	Exemples de prévision à 1 h obtenue grâce aux modèles GPR testés. . . . .	94
4.14	Exemples de prévision à 3 h obtenue grâce aux modèles GPR testés. . . . .	95
4.15	Exemples de prévision à 5 h obtenue grâce aux modèles GPR testés. . . . .	96
4.16	Exemples de prévision à 24 h obtenue grâce aux modèles GPR testés. . . . .	97



# Liste des tableaux

1.1	Caractéristiques des sites étudiés. . . . .	22
1.2	Tests de contrôle qualité réalisés sur les mesures du GHI. . . . .	26
1.3	Résultats de l'analyse de la qualité des mesures du GHI. . . . .	26
1.4	Variabilité du GHI. . . . .	27
2.1	Récentes études comparatives des outils de l'apprentissage automatique. . . . .	39
2.2	Durées d'apprentissage et temps d'exécution des modèles. . . . .	54
4.1	Coordonnées GPS des 16 pyranomètres installés à Villeneuve-de-la-Raho. . . . .	82





# Nomenclature

ANN	Réseaux de neurones artificiels ( <i>artificial neural networks</i> )	[-]
BT	Basse tension	[-]
CWC	<i>Coverage width-based criterion</i>	[-]
$D_{UA}$	Distance moyenne Terre-Soleil	[m]
DHI	Éclairement diffus horizontal ( <i>diffuse horizontal irradiance</i> )	[Wm <sup>-2</sup> ]
DMAE	Erreur absolue moyenne dynamique ( <i>dynamic mean absolute error</i> )	[-]
DNI	Éclairement normal direct ( <i>direct normal irradiance</i> )	[Wm <sup>-2</sup> ]
EIA	Administration américaine d'information sur l'énergie ( <i>U.S. Energy Information Administration</i> )	[-]
GES	Gaz à effet de serre	[-]
GHI	Éclairement global horizontal ( <i>global horizontal irradiance</i> )	[Wm <sup>-2</sup> ]
GHI <sub>cs</sub>	Éclairement global horizontal par ciel clair	[Wm <sup>-2</sup> ]
GIEC	Groupe d'experts intergouvernemental sur l'évolution du climat	[-]
GP	Processus gaussien ( <i>Gaussian process</i> )	[-]
GPR	Régression non paramétrique par processus gaussien ( <i>Gaussian process regression</i> )	[-]
HDR	<i>High dynamic range</i>	[-]
HTA	Haute tension A	[-]
$I_0$	Éclairement solaire extra-atmosphérique	[-]
$I_{UA}$	Éclairement solaire extra-atmosphérique incident à la distance moyenne Terre-Solaire	[Wm <sup>-2</sup> ]
$k_c$	Indice ciel clair	[-]
$k_t$	Indice de clarté	[-]
$k_{Lin}$	Noyau linéaire ( <i>linear kernel</i> )	[-]
$k_{SE}$	Noyau exponentiel quadratique ( <i>squared exponential kernel</i> )	[-]
$k_{RQ}$	Noyau quadratique rationnel ( <i>rational quadratic kernel</i> )	[-]
$k_{WN}$	Noyau bruit blanc ( <i>white noise kernel</i> )	[-]
$k_E$	Noyau exponentiel ( <i>exponential kernel</i> )	[-]
$k_{M_{3/2}}$	Noyau Matérn d'ordre $\frac{3}{2}$	[-]
$k_{M_{5/2}}$	Noyau Matérn d'ordre $\frac{5}{2}$	[-]
$k_{Per \times RQ}$	Noyau quasi-périodique résultant du produit d'un noyau périodique et d'un noyau quadratique rationnel	[-]
$k_{Per \times SE}$	Noyau quasi-périodique résultant du produit d'un noyau périodique et d'un noyau exponentiel quadratique	[-]
$k_{Per \times E}$	Noyau quasi-périodique résultant du produit d'un noyau périodique et d'un noyau exponentiel	[-]
$k_{Per \times M_{3/2}}$	Noyau quasi-périodique résultant du produit d'un noyau périodique et d'un noyau Matérn- $\frac{3}{2}$	[-]
$k_{Per \times M_{5/2}}$	Noyau quasi-périodique résultant du produit d'un noyau périodique et d'un noyau Matérn- $\frac{5}{2}$	[-]

$k_{\text{Per}+\text{RQ}}$	Noyau quasi-périodique résultant de la somme d'un noyau périodique et d'un noyau quadratique rationnel	[–]
$k_{\text{Per}+\text{SE}}$	Noyau quasi-périodique résultant de la somme d'un noyau périodique et d'un noyau exponentiel quadratique	[–]
$k_{\text{Per}+\text{E}}$	Noyau quasi-périodique résultant de la somme d'un noyau périodique et d'un noyau exponentiel	[–]
$k_{\text{Per}+\text{M}_{3/2}}$	Noyau quasi-périodique résultant de la somme d'un noyau périodique et d'un noyau Matérn- $\frac{3}{2}$	[–]
$k_{\text{Per}+\text{M}_{5/2}}$	Noyau quasi-périodique résultant de la somme d'un noyau périodique et d'un noyau Matérn- $\frac{5}{2}$	[–]
$k\text{NN}$	$k$ -plus proches voisins ( <i>k-nearest neighbours</i> )	[–]
$m$	Masse d'air optique relative	[m]
MINLP	Programmation non linéaire mixte en nombres entiers ( <i>mixed integer nonlinear programming</i> )	[–]
MLP	Perceptron multicouche ( <i>multilayer perceptron</i> )	[–]
nRMSE	Erreur quadratique moyenne normalisée ( <i>normalized root mean square error</i> )	[–]
NWP	Modèle numérique de prévision météorologique ( <i>numerical weather prediction</i> )	[–]
PPE	Programmation pluriannuelle de l'énergie	[–]
PSI	PROMES Sky Imager	[–]
PV	Photovoltaïque	[–]
RSI	Pyranomètre à bande rotative ( <i>rotating shadowband irradiator</i> )	[–]
SG	Solar Geometry	[–]
SG2	Solar Geometry 2	[–]
SNBC	Stratégie nationale bas-carbone	[–]
SVM	Machine à vecteurs de support ( <i>support vector machines</i> )	[–]
SVR	Régression par machine à vecteurs de support ( <i>support vector regression</i> )	[–]
$T_L$	Trouble de Linke	[–]
$\theta_z$	Angle entre le Soleil et le zénith	[–]
TOR	Tout ou rien	[–]
TSI	Total Sky Imager	[–]

# Introduction générale

En France comme à l'échelle mondiale, plus de 70 % des émissions de gaz à effet de serre (GES) produites proviennent de la production et de la consommation d'énergie. Ces émissions, en perpétuelle augmentation pour des raisons essentiellement anthropiques, sont à l'origine de plusieurs problèmes environnementaux et obligent tous les gouvernements à s'engager dans une politique de développement durable. Ainsi, à l'instar de la France, les autres pays de l'Union européenne (UE), en signant l'Accord de Paris en décembre 2015 [1], se sont engagés à atteindre la neutralité carbone à l'horizon 2050 afin de respecter leurs engagements dans la lutte contre le changement climatique. Pour espérer atteindre cet objectif ambitieux, l'UE vise comme palier intermédiaire une baisse drastique de ses émissions de GES d'au moins 55 % en 2030 par rapport au niveau de 1990 [2]. À cet égard, le gouvernement français a établi une feuille de route, la stratégie nationale bas-carbone (SNBC) [3], instaurée par la loi du 17 août 2015 relative à la transition énergétique pour la croissance verte, pour piloter la politique d'atténuation du changement climatique de la France. Cette stratégie met au cœur des préoccupations le secteur de l'énergie. Il repose ainsi en grande partie sur la décarbonation de ce secteur et une amélioration de l'efficacité énergétique.

Pour parvenir à la neutralité carbone à l'horizon 2050, une profonde restructuration de l'ensemble des secteurs d'activités, notamment le secteur énergétique, est nécessaire. En effet, cela implique des bouleversements dans le secteur des transports, de l'industrie et du bâtiment aujourd'hui encore très dépendants des énergies fossiles qui sont responsables de la majeure partie des émissions de GES. Par ailleurs, dans un contexte d'augmentation de la demande mondiale d'énergie, les réserves de combustibles fossiles se raréfient au fil des années. Ainsi, pour décarboner le secteur des transports par exemple, on assiste de nos jours au recours à l'électricité pour remplacer les énergies fossiles. Il faudra donc répondre à la demande d'énergie sans cesse croissante et aux préoccupations environnementales par une production d'énergie bas-carbone, en particulier d'électricité. En effet, l'un des objectifs majeurs de la SNBC est une production d'énergie avec zéro émission à l'horizon 2050. Cela passera par une démocratisation de l'utilisation des énergies renouvelables, fortement plébiscitées pour leur faible empreinte écologique. En 2020, le Conseil Européen a fixé comme objectif d'accroître à 32 % la part des énergies renouvelables dans le mix énergétique de l'UE à l'horizon 2030 [2]. Ainsi, depuis quelques années, la part des énergies renouvelables dans le mix énergétique mondial, en particulier celle de l'énergie solaire, connaît une très forte progression. En 2019, selon le gestionnaire du réseau de transport d'électricité en France (RTE), la part des énergies renouvelables dans le mix électrique de la France était de 22 % (Figure 1) avec une part du solaire essentiellement d'origine photovoltaïque (PV) de 2,2 % [4]. Les estimations annoncent une forte croissance du solaire PV en France et dans le monde dans les années à venir.

Dans un contexte où les énergies renouvelables jouent un rôle prépondérant dans la réussite de la transition énergétique, le développement de nouvelles stratégies de gestion des réseaux électriques de distribution est nécessaire [5, 6, 7, 8]. En effet, les réseaux électriques, conçus à la base pour une production centralisée et des flux d'énergie unidirectionnels, font face à un déploiement à grande échelle d'une production décentralisée, en particulier le solaire PV. Du fait de sa nature très variable, due essentiellement aux conditions météorologiques, une pénétration élevée de la production PV pourrait entraîner des problèmes tels que des fluctuations de tension et entraver le maintien de l'équilibre entre offre et demande. Évoluer vers plus d'intelligence dans la gestion des réseaux électriques doit permettre :

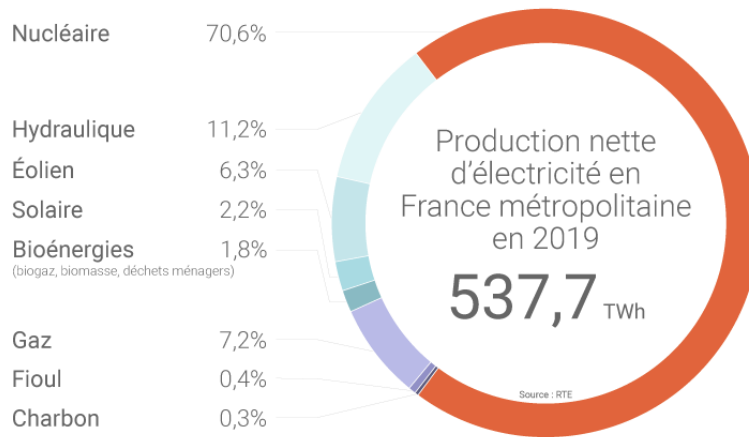


FIGURE 1 – Mix électrique de la France en 2019 [4].

- une anticipation des problèmes qui pourraient se poser à cause de la pénétration de la production décentralisée de nature intermittente ;
- une amélioration de l'observabilité des réseaux électriques ;
- une garantie de sécurité et de qualité de service.

Cette thèse s'inscrit dans le cadre du projet « Smart Occitania » visant à développer les algorithmes nécessaires à la gestion prévisionnelle du réseau électrique de distribution, en présence d'une production décentralisée et intermittente, en particulier le solaire photovoltaïque, en région Occitanie. Le projet « Smart Occitania » fait partie des nombreux projets de démonstrateurs de réseaux électriques intelligents portés par ENEDIS (le gestionnaire du réseau électrique de distribution en France) ayant pour objectif le développement d'outils permettant, d'une part, d'augmenter la capacité d'accueil du réseau électrique de distribution sans construction de nouvelles infrastructures et favorisant, d'autre part, la pénétration de la production d'énergie renouvelable qui ne cesse d'augmenter du fait de la dynamique insufflée par la transition énergétique. La prévision de la production PV se trouve au cœur du développement des stratégies de gestion intelligente des réseaux électriques de distribution qui font face à un déploiement massif du solaire PV, car elle permet aux gestionnaires de mieux appréhender l'intermittence du solaire PV et aide à la réussite de sa pénétration à grande échelle. Ainsi, parmi les différents axes de recherche du projet, la prévision multi-horizon de la production photovoltaïque est essentielle en ce qu'elle permettrait une amélioration de la gestion du réseau électrique de distribution. La prévision de la production PV découle directement de celle de l'éclairement global horizontal (ou GHI, pour *global horizontal irradiance*), qui est sujet à une variabilité à différentes échelles spatiales et temporelles. De ce fait, les travaux présentés dans ce manuscrit se focalisent sur la prévision localisée et spatio-temporelle du GHI à des horizons de temps infra-horaires et infra-journaliers en vue de prévoir la production PV à des fins de gestion intelligente du réseau électrique de distribution en région Occitanie faisant face à une pénétration à grande échelle de la production PV.

L'objectif global de cette thèse est de prévoir le GHI pour le contrôle-commande du réseau électrique de distribution en région Occitanie. Pour cela, nous nous sommes posés quelques questions qui seront étudiées dans les différents chapitres de ce manuscrit.

- Quel type de modèle pour la prévision du GHI ?
- Quel outil de l'apprentissage automatique pour la prévision localisée du GHI ?
- Quelle entrée pour les modèles fondés sur la régression non paramétrique par processus gaussien pour la prévision localisée du GHI ?
- Quelle fonction de covariance pour la prévision spatio-temporelle du GHI par régression non paramétrique par processus gaussien ?

Ce manuscrit s'articule autour de quatre chapitres. Le premier chapitre situe le contexte et fait le choix du type de modèle pour la prévision du GHI. Tout d'abord, les raisons sous-jacentes de la transition énergétique et les défis posés par une pénétration massive de la production d'énergie renouvelable dans le réseau électrique de distribution, ainsi que les solutions pour faire face à ces défis, sont abordés. On y notera que la prévision de la production photovoltaïque est un élément clé des stratégies de gestion intelligente des réseaux électriques de distribution. Une présentation du projet « Smart Occitania » et de ses différents enjeux est également faite. Ensuite, les jeux de données du GHI utilisés dans cette thèse sont présentés après un rapide rappel des composantes de l'éclairage solaire et des instruments de mesure associés. Enfin, un état de l'art des modèles de prévision du GHI est proposé. Une analyse des différentes catégories de modèle existantes dans la littérature est faite et un type de modèle est choisi pour la prévision du GHI.

Le deuxième chapitre traite d'une étude comparative de modèles statistiques de prévision fondés sur l'apprentissage automatique dans le but de choisir l'outil le plus adapté à la prévision localisée du GHI. Après une description des outils considérés dans l'étude comparative, à savoir la régression non paramétrique par processus gaussien, la régression par machines à vecteurs de support et la régression par réseaux de neurones artificiels, les résultats obtenus sont présentés. Une comparaison de ces résultats sur la base des critères de performance choisis et quelque exemples de prévisions obtenus grâce aux modèles développés est faite. Enfin, l'outil le plus adapté à la prévision du GHI au regard des exigences de notre application est choisi.

Le troisième chapitre a pour objectif d'identifier l'entrée adéquate pour les modèles fondés sur la régression non paramétrique par processus gaussien pour la prévision localisée du GHI. En effet, un choix naturel est de prendre un historique d'observations du GHI, ce qui permet de mettre en évidence les corrélations. On y verra qu'il est également possible de prendre le temps comme entrée afin de tirer profit des connaissances *a priori* sur la dynamique du GHI pour développer un modèle approprié. Une étude comparative de ces différents modèles GPR utilisant le temps en entrée et utilisant les observations est menée. Ensuite, une fonction de covariance – la fonction de covariance, également appelée noyau est l'élément clé des modèles GPR – adaptée à la prévision localisée du GHI est identifiée.

Le quatrième chapitre traite de la prévision spatio-temporelle du GHI par régression non paramétrique par processus gaussien. Ce chapitre est un premier pas vers la prévision spatio-temporelle du GHI. Il cherche à identifier un noyau spatial à associer au noyau temporel identifié au [Chapitre 3](#) afin de développer des modèles GPR de prévision spatio-temporelle du GHI. Pour ce faire, une étude comparative de plusieurs noyaux spatiaux considérés est menée.

Les différents travaux menés au cours de cette thèse ont donné lieu aux publications et communications suivantes.

1. Publications dans des revues internationales à comité de lecture :

- Shab Gbémou, Julien Eynard, Stéphane Thil, Emmanuel Guillot, Stéphane Grieu. *A comparative study of machine learning-based methods for global horizontal irradiance forecasting*. Dans : *Energies*, vol. 14, n° 11, 2021.
- Shab Gbémou, Stéphane Thil, Julien Eynard, Stéphane Grieu. *Time or historical observations as model input for global horizontal irradiance forecasting using Gaussian process regression ? En cours de rédaction*.

2. Communications dans des conférences internationales :

- Shab Gbémou, Hanany Tolba, Stéphane Thil, Stéphane Grieu. *Global horizontal irradiance forecasting using online sparse Gaussian process regression based on quasiperiodic kernels*. Dans : *International Conference on Environment and Electrical Engineering IEEEIC 2019*.
- Shab Gbémou, Hanany Tolba, Julien Nou, Julien Eynard, Stéphane Thil, Stéphane Grieu. *Clustering and intraday forecasting of global horizontal irradiance using machine learning methods*. Dans : *International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems ECOS 2020*.



# Chapitre 1

## Contexte et enjeux

Ce premier chapitre présente la démarche scientifique suivie au cours de cette thèse et les fondements des travaux traitant du développement d’algorithmes de prévision de l’éclairement global horizontal. Prévoir le GHI a pour finalité de prévoir la production photovoltaïque et, dans le cadre du projet « Smart Occitania », porté par ENEDIS, le développement d’algorithmes pour la gestion intelligente du réseau électrique de distribution en région Occitanie. Cette meilleure gestion permettra, d’une part, de limiter les contraintes, principalement en tension, qui apparaissent sur le réseau basse tension du fait de la pénétration de la production décentralisée, en particulier le solaire photovoltaïque, et, d’autre part, de favoriser la pénétration de cette production, participant ainsi aux efforts de décarbonation énergétique.

Ce chapitre s’articule en quatre parties : la [Section 1.1](#) présente les motivations derrière l’essor rapide des énergies renouvelables et les défis posés par leur pénétration dans les réseaux électriques de distribution. Cette section montre aussi la nécessité de développer des stratégies de gestion intelligente pour relever ces défis et présente le projet « Smart Occitania », ses objectifs et les résultats obtenus. La [Section 1.2](#) présente les différentes composantes de l’éclairement solaire et les jeux de données GHI utilisés dans cette thèse. Un état de l’art des techniques de prévision du GHI est proposé dans la [Section 1.3](#). La [Section 1.4](#) décrit les critères de performance utilisés pour évaluer les modèles de prévision développés dans cette thèse.

### 1.1 Vers une gestion intelligente des réseaux électriques de distribution

Une conséquence directe de la croissance démographique et du développement économique est la hausse continue de la demande en énergie. Cette dernière est, depuis des décennies, satisfaite par l’utilisation des combustibles fossiles responsables de fortes émissions de gaz à effet de serre. À l’heure de la transition énergétique partout dans le monde, il faut répondre à la consommation énergétique, entre autres, par une production d’énergie bas-carbone. Face à la raréfaction des ressources énergétiques non renouvelables et aux enjeux climatiques, le recours aux énergies renouvelables s’impose comme une évidence dans la lutte contre le réchauffement climatique et la protection de l’environnement.

Le réseau électrique, depuis fort longtemps basé sur un fonctionnement unidirectionnel, a pour rôle d’acheminer l’électricité depuis les grands centres de production (centrales nucléaires ou thermiques) vers les sites de consommation et d’assurer à tout moment l’équilibre entre production et consommation. De nos jours, les réseaux électriques sont confrontés à une évolution des usages de l’énergie électrique et au développement de la production d’électricité à partir des énergies renouvelables, de nature intermittente. Pour s’adapter à cette évolution, les réseaux électriques doivent se doter de nouvelles technologies garantissant une fourniture d’électricité efficace et sécurisée.

Dans la suite, nous exposerons les raisons qui poussent à avoir recours aux énergies renouvelables et les défis liés à leur pénétration dans les réseaux électriques de distribution. Par ailleurs, nous nous



attarderons sur le solaire photovoltaïque et sur la nécessité de développer des outils de prévision de la ressource solaire, qui est la pierre angulaire de nos travaux.

### 1.1.1 Enjeux climatiques

De nos jours, il est indéniable que les activités anthropiques constituent la principale cause du réchauffement climatique observé un peu partout dans le monde [9, 10]. La plupart des grands organismes scientifiques, de la NASA (*National Aeronautics and Space Administration*) à l'Organisation météorologique mondiale, partagent ce point de vue. Le réchauffement climatique est un phénomène généralisé à tous les continents. On observe déjà partout dans le monde les effets désastreux du réchauffement climatique tels que des conditions météorologiques extrêmes, l'extinction de certaines espèces, l'accélération de la fonte des glaces et l'augmentation du niveau des eaux.

Face à ce contexte environnemental, le Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC) a été créé en 1988 dans le but de fournir des évaluations détaillées de l'état des connaissances scientifiques, techniques et socioéconomiques en lien avec le réchauffement climatique, ses causes et ses répercussions potentielles. Il identifie également les stratégies d'atténuation du réchauffement climatique et les possibilités d'adaptation aux changements attendus. Dans son rapport de 2014, il confirme que le climat se réchauffe et alerte sur les incidences d'évènements climatiques extrêmes survenus récemment [11]. En France, plusieurs équipes de recherche travaillent sur ces sujets, impliquant des centaines de scientifiques, certains d'entre eux contribuant à l'élaboration des rapports du GIEC.

Ce réchauffement climatique est essentiellement dû à la concentration des GES dans l'atmosphère. Selon le GIEC, les émissions anthropiques de GES ont fortement augmenté entre 1970 et 2010 avec une hausse importante vers la fin de cette période. Malgré le nombre croissant de politiques mises en œuvre pour l'atténuation du réchauffement climatique, les émissions anthropiques totales de GES ont été les plus importantes de toute l'histoire de l'humanité entre 2000 et 2010. Le dioxyde de carbone (CO<sub>2</sub>) représente à lui seul près de 70 % des émissions de GES. Sa concentration a augmenté de 40 % par rapport à son niveau de l'ère préindustrielle [12]. Les émissions de CO<sub>2</sub> imputables à l'usage des combustibles fossiles ont contribué à hauteur de 78 % aux émissions mondiales de GES entre 1970 et 2010 [11]. La production d'énergie en hausse d'année en année, en réponse à des besoins énergétiques croissants, est toujours largement tributaire des combustibles fossiles.

### 1.1.2 Évolution de la consommation énergétique mondiale

Au vu de ce qui précède, il est évident que le réchauffement climatique et le secteur énergétique sont étroitement liés. Depuis la révolution industrielle, les progrès scientifiques et technologiques ont entraîné une nette amélioration des conditions de vie dans le monde, mais aussi une hausse de la consommation d'énergie. En effet, la demande énergétique mondiale a progressé de 109 % de 1973 à 2018 [13]. Dans son rapport de 2019 sur la consommation d'énergie dans le monde [14], l'Administration américaine d'information sur l'énergie (EIA, pour *U.S. Energy Information Administration*) prévoit une augmentation de la consommation énergétique mondiale de 50 % entre 2018 et 2050. Cette progression de la demande énergétique mondiale est imputable à la croissance démographique et économique. En effet, la majeure partie de cette croissance provient des pays en voie de développement où l'on remarque une forte croissance économique, notamment les pays d'Afrique et d'Asie. Les pays développés continuent de consommer une énorme quantité d'énergie, tandis que la demande énergétique augmente rapidement dans les pays en voie de développement. Par exemple, la Chine est le premier consommateur d'électricité dans le monde. Par ailleurs, l'EIA estime dans son rapport publié en 2019 [14] une hausse de la production d'électricité de 79 % entre 2018 et 2050 pour couvrir les besoins en électricité qui subiront également une forte croissance avec l'augmentation de la population et les progrès dans tous les secteurs à travers le monde.

Les estimations annoncent une consommation d'énergie croissante partout dans le monde. Face à l'épuisement des ressources énergétiques fossiles et aux préoccupations environnementales actuelles, il

faudra se tourner vers d'autres ressources plus respectueuses de l'environnement. Ainsi, pour répondre à cette croissance rapide de la demande énergétique, les énergies renouvelables, y compris l'énergie solaire, connaîtront un essor considérable d'ici à 2050. En effet, l'EIA prévoit que la consommation d'énergie renouvelable augmentera de 3,1 % par an sur cette période de temps, contre une croissance annuelle de 0,6 % pour le pétrole, de 0,4 % pour le charbon et de 1,1 % pour le gaz naturel [14].

### 1.1.3 Épuisement des combustibles fossiles

Les combustibles fossiles, malgré leur forte contribution au réchauffement climatique de la planète, sont toujours très utilisés dans le monde pour la production d'énergie. En effet, ils couvrent aujourd'hui plus de 80 % des besoins énergétiques mondiaux. Les combustibles fossiles sont issus de roches formées par fossilisation d'espèces organiques mortes depuis plusieurs millions d'années. Ils sont donc non renouvelables dans la mesure où ils ne peuvent pas être reconstitués à l'échelle du temps humain. Malgré l'utilisation des énergies renouvelables, de nombreux secteurs restent toujours dépendants des combustibles fossiles, comme le secteur des transports dont l'énergie est issue à plus de 90 % du pétrole. En 2017, les combustibles fossiles ont contribué à hauteur de 41 % à la production nette totale d'énergie en Europe [15]. Cependant, la disponibilité des réserves devient une importante source de préoccupation. Au rythme actuel de consommation, le pétrole sera le premier combustible fossile dont il deviendra impératif de se passer [16]. En 2016, le pétrole a fourni près de 33 % de l'approvisionnement total en énergie primaire dans le monde. Selon les estimations, il resterait encore trente ans de réserves prouvées de pétrole [17, 18]. Pour le gaz naturel, il y aurait encore une cinquantaine d'années de réserves [17, 18]. Le charbon pourrait, quant à lui, être encore exploité pendant deux siècles [17, 18]. En additionnant l'ensemble des réserves prouvées (et non extraites aujourd'hui), on pense qu'il reste un peu plus de deux cents ans de consommation. Par conséquent, la communauté scientifique est unanime sur l'épuisement des combustibles fossiles [12, 19, 20]. Toutefois, il se pourrait que ces combustibles soient disponibles un peu plus longtemps que prévu, car les estimations se basent sur les réserves prouvées, c'est-à-dire celles que l'on est sûr de pouvoir exploiter. Cependant, ces données sont à relativiser, car elles ne prennent pas en compte la croissance de la consommation alors qu'il est clair qu'elle va considérablement augmenter. La croissance de la consommation implique depuis le début de l'ère industrielle une demande en énergie amplifiée par la croissance démographique et économique dans le monde. Or, on estime qu'en 2030 les combustibles fossiles devraient encore représenter près de 80 % du mix énergétique. Les enjeux actuels seront amplifiés à cette date. Il est clair que les ressources énergétiques fossiles existantes, en quantité limitée, ne pourront pas répondre aux besoins croissants de la population mondiale. Toutefois, elles sont en quantité suffisante pour que leur combustion soit à l'origine d'un réchauffement climatique dangereux pour la planète.

La France est le 7<sup>e</sup> plus important pays consommateur d'énergie dans le monde. Le pétrole, le charbon et le gaz naturel sont encore utilisés en France, mais très peu dans la production d'électricité. L'usage du pétrole est nettement dominant dans les transports et la production de chaleur, pour les particuliers comme pour les industries. Les centrales thermiques basées sur la combustion du charbon, du fioul et du gaz jouent encore un rôle essentiel, car elles fournissent le complément d'électricité capable de faire face aux pics de consommation imprévus et de plus en plus fréquents, comme lors des vagues de froid.

Depuis quelques années, les coûts d'extraction des combustibles fossiles connaissent une tendance haussière. Ainsi, face au déclin des gisements de combustibles fossiles et à l'augmentation de leurs coûts d'extraction, il s'avère important de se tourner vers d'autres ressources énergétiques non limitées. Dans un contexte de demande énergétique en excès, les énergies renouvelables constituent une alternative viable pour la production d'énergie propre et durable.

### 1.1.4 Essor des énergies renouvelables

Face à l'épuisement des ressources énergétiques fossiles, au réchauffement climatique et à la dégradation de l'environnement, les sociétés humaines cherchent de plus en plus à modifier leurs

modes de production et de consommation de façon à être plus écologiques. Les énergies renouvelables, également désignées par les termes « énergies vertes » ou « énergies propres », deviennent un élément majeur des stratégies des gouvernements à travers le monde en matière de transition énergétique et de développement durable grâce à leur faible impact environnemental et à leur statut de ressources illimitées à l'échelle du temps humain. Il existe plusieurs types d'énergies renouvelables telles que : l'énergie solaire, l'énergie éolienne, l'énergie hydraulique, la biomasse et la géothermie. Leur exploitation génère peu de polluants, comparativement aux énergies fossiles, ce qui fait de ces énergies un élément clé dans la lutte contre le réchauffement climatique.

La part des énergies renouvelables dans la consommation finale d'énergie mondiale était de 17,9 % en 2018, pour une part dans la production d'électricité de 26,4 % [21]. En 2020, selon l'Agence Internationale de l'Énergie, 90 % de la puissance électrique nouvellement installée revient aux énergies renouvelables, en particulier l'éolien, le solaire et l'hydraulique [22]. Elle prévoit une forte croissance de la part des énergies renouvelables dans le mix énergétique mondial au cours des années à venir.

L'utilisation à grande échelle des énergies renouvelables contribue non seulement à l'atténuation des problèmes environnementaux, mais aussi à la diminution des effets de la pollution de l'air et du réchauffement climatique sur la santé. L'atténuation des dommages sanitaires peut, dans certains cas, plus que compenser les coûts d'une politique climatique. Par exemple, une étude menée aux États-Unis a montré que les avantages économiques liés à l'utilisation des énergies renouvelables pour la santé sont d'environ 60 % plus élevées que les subventions à l'énergie éolienne [23].

### 1.1.5 Engagement de la France dans la lutte contre le réchauffement climatique

Le réchauffement climatique trouve sa cause dans la production de GES dont environ 70 % résulte de la consommation d'énergie fossile. Pour limiter son impact sur nos sociétés, les pays du monde se sont engagés par l'Accord de Paris [1] à réduire considérablement leurs émissions de GES. En cohérence avec ses engagements internationaux et européens, la France mène une politique nationale de lutte contre le réchauffement climatique. Ainsi, le gouvernement français a adopté la loi Énergie-Climat publiée au Journal Officiel du 9 novembre 2019 face à l'urgence écologique et climatique. Elle renforce, actualise et complète les objectifs de la loi relative à la transition énergétique pour la croissance verte de 2015. Les objectifs visés par la loi Énergie-Climat sont :

- une réduction d'au moins 40 % des émissions de GES en 2030 par rapport au niveau de 1990 ;
- une division des émissions par au moins 6 d'ici 2050 par rapport au niveau de 1990 ;
- la fermeture des dernières centrales à charbon en 2022 ;
- une réduction d'environ 50 % de la consommation énergétique finale en 2050 par rapport au niveau de 2012, avec comme objectif intermédiaire une baisse de 20 % en 2030 ;
- une réduction d'au moins 40 % de la consommation énergétique primaire fossile en 2030 par rapport à 2012 ;
- une part des énergies renouvelables dans le mix énergétique d'au moins 33 % en 2030.

Ces objectifs trouvent leurs déclinaisons dans plusieurs outils de planification nationale, définis dans ce cadre législatif. Les deux principaux portant sur l'énergie et le climat sont la stratégie nationale bas-carbone (SNBC) et la programmation pluriannuelle de l'énergie (PPE). La SNBC définit les grandes lignes des politiques transversales et sectorielles permettant d'atteindre les objectifs de réduction des émissions de GES à moyen et long termes. La PPE est un outil de pilotage qui dessine le chemin à suivre au cours de la décennie à venir afin d'atteindre la neutralité carbone à l'horizon 2050. Elle fixe les actions prioritaires des pouvoirs publics dans le domaine de l'énergie afin d'atteindre les objectifs de la loi Énergie-Climat en cohérence avec la SNBC. Elle contient plusieurs volets relatifs, entre autres, à la sécurité de l'approvisionnement, à l'amélioration de l'efficacité énergétique, à la popularisation de l'exploitation des énergies renouvelables et au développement des réseaux électriques intelligents.

### 1.1.6 Le solaire photovoltaïque

Parmi les énergies renouvelables utilisées pour la production d'énergie, on remarque un attrait particulier pour l'énergie solaire. On distingue deux grandes technologies exploitant l'énergie solaire à des fins de production de l'électricité : le solaire photovoltaïque et le solaire thermodynamique (solaire à concentration). Le solaire PV est, de loin, le plus répandu. Son développement rapide s'explique par les subventions des états et par son potentiel de déploiement à grande échelle : avec une surface photovoltaïque de 100 000 km<sup>2</sup> (soit la superficie de l'Islande), la totalité des besoins mondiaux en électricité seraient couverts [24].

#### 1.1.6.1 Évolution de la production photovoltaïque à l'échelle mondiale

L'énergie solaire devient de plus en plus utilisée dans le monde pour produire de l'électricité ou pour chauffer et dessaler l'eau. Les systèmes photovoltaïques, également appelés cellules solaires, sont des dispositifs électroniques qui convertissent directement la lumière du Soleil en électricité. La cellule solaire a été inventée en 1954 dans les laboratoires Bell aux États-Unis. Les panneaux solaires ont une durée de vie d'environ trente ans et se déclinent en différentes teintes selon le type de matériau utilisé pour leur fabrication. Aujourd'hui, le solaire PV est l'une des technologies à énergie renouvelable qui connaît la croissance la plus rapide et joue un rôle majeur dans la transition énergétique mondiale.

Les installations solaires photovoltaïques peuvent être combinées pour fournir de l'électricité à des fins commerciales, ou disposées en plus petites configurations pour les mini-réseaux ou l'usage personnel. L'utilisation de l'énergie solaire photovoltaïque pour alimenter des mini-réseaux est un excellent moyen d'offrir l'accès à l'électricité aux personnes qui ne vivent pas à proximité des réseaux électriques de distribution, en particulier dans les pays en voie de développement disposant d'excellentes ressources en énergie solaire.

Les gestionnaires des réseaux électriques de distribution prévoient une forte augmentation de la pénétration de la production décentralisée, en particulier le solaire PV, dans les réseaux. En effet, la puissance PV installée dans le monde est passée de 40 GW en 2010 à 707 GW en 2020 (Figure 1.1) et les estimations annoncent une forte croissance dans les années à venir [25].

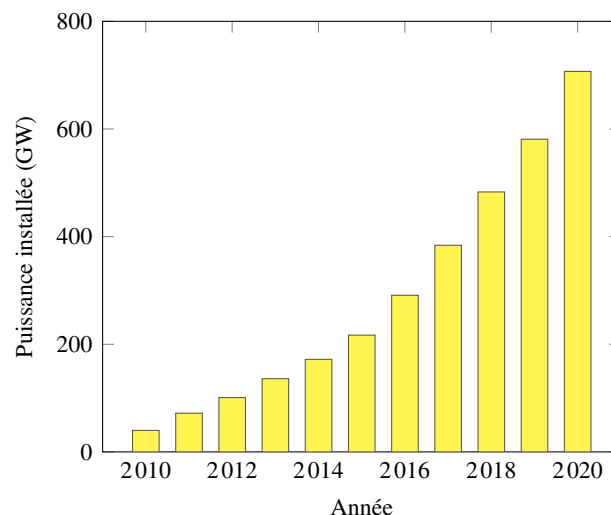


FIGURE 1.1 – Évolution de la puissance photovoltaïque installée dans le monde entre 2010 et 2020 [25].

En 2020, la puissance PV installée en France s'élevait à près de 10 GW et une carte de localisation géographique (Figure 1.2) nous montre que la production PV se développe rapidement dans le sud de la France qui est une zone à fort potentiel [26]. En effet, les régions Nouvelle-Aquitaine et Occitanie constituent les plus importants pôles de production PV.

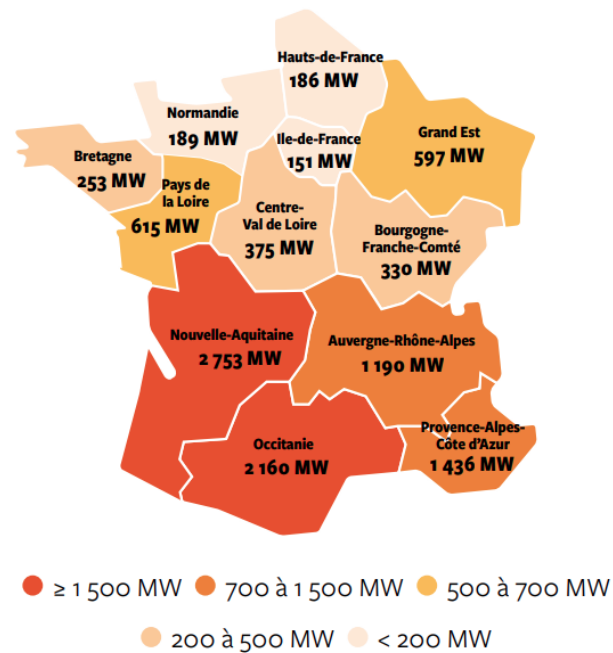


FIGURE 1.2 – Puissance photovoltaïque installée en France par région au 31/12/2020 [26].

### 1.1.6.2 Pénétration de la production PV dans les réseaux électriques de distribution

Les gestionnaires des réseaux électriques de distribution ont l'obligation contractuelle de fournir en permanence l'électricité à leurs clients tout en respectant un certain nombre de contraintes [27].

- Contrainte d'intensité : l'intensité du courant est réglementée par des normes. Les câbles électriques ont une intensité de courant maximale prédéfinie qui varie en fonction de la température et selon qu'il s'agit d'une installation souterraine ou aérienne.
- Contrainte en tension : la tension fluctue en fonction des utilisateurs qui sont raccordés au réseau et de sa configuration. Cependant, de légères fluctuations par rapport à la valeur nominale sont autorisées. Le projet « Smart Occitania » – les travaux présentés dans ce manuscrit s'inscrivent dans le projet – a notamment pour objectif le développement d'algorithmes permettant de lever ce type de contrainte.
- Contrainte de gradient : le gradient de chute de tension est la chute de tension supplémentaire générée en un point du réseau si 1 kW monophasé est rajouté en ce même point. Afin de garantir une qualité de tension correcte chez les consommateurs, le gradient doit être inférieur à une valeur donnée.

La décarbonation du système électrique repose sur une utilisation massive des énergies renouvelables telles que l'énergie éolienne et l'énergie solaire photovoltaïque. Cette transition vers un système comportant une part croissante d'électricité renouvelable intermittente soulève de nouveaux enjeux. Alors que la France s'oriente vers un mix électrique avec une part renouvelable d'environ 40 % en 2030, la pénétration des énergies renouvelables, en particulier le solaire PV, devient une préoccupation majeure. De nos jours, avec le déploiement du solaire PV, les consommateurs raccordés au réseau électrique de distribution deviennent aussi producteurs et injectent de manière irrégulière [28] une puissance dans le réseau. *Quels sont les impacts de cette production décentralisée sur le réseau électrique de distribution, conçu à l'origine pour assurer un transit radial et unidirectionnel des flux de puissance des centres de production vers les consommateurs ?* Cette question a fait l'objet de plusieurs études dans la littérature [29, 30, 31, 32, 33, 34, 5, 35].

- Impact sur le transit de puissance : l'injection d'une production décentralisée dans le réseau électrique de distribution produit un apport de puissance au point d'injection qui peut engendrer des

problèmes de congestion locale et une inversion des flux de puissance du réseau de distribution vers celui de transport. Ainsi, les appareils de protection utilisés dans les réseaux doivent être bidirectionnels.

- Impact sur la qualité de tension : l'utilisation des câbles électriques pour transporter l'électricité est responsable d'une chute de tension. Au fur et à mesure que l'on s'éloigne du poste source, la tension diminue. Cependant, le gestionnaire doit être capable de maintenir la tension dans les plages définies. L'injection d'une production décentralisée peut donc engendrer un dépassement de la contrainte en tension autorisée. Ceci va modifier le plan de tension au point de raccordement et peut aussi affecter la qualité de l'onde de tension.
- Impact sur les courants de court-circuit : l'injection d'une production décentralisée affecte les courants de court-circuit ainsi que la puissance de court-circuit en modifiant l'impédance globale du réseau. Des conséquences sur le plan de protection s'ensuivent. En effet, le dimensionnement des seuils de protection est modifié. Cela entraîne une mauvaise sélectivité des protections.
- Impact sur la stabilité du système : en France, les producteurs étant quasi-exclusivement de type photovoltaïque, la puissance injectée sur un départ basse tension est synchrone. L'injection d'une production décentralisée de type synchrone change le temps critique d'élimination d'un défaut.
- Impact sur l'observabilité et la contrôlabilité du système : le réseau de distribution, pensé à l'origine comme un élément passif, devient actif avec la pénétration d'une production décentralisée intermittente. Il faudra surveiller l'état du réseau électrique à tout moment afin d'anticiper les incidents et faciliter la prise de décision pour optimiser le réseau.
- Impact sur la continuité de service : l'intermittence de la production décentralisée peut également conduire à une indisponibilité en cas de sollicitation et ainsi occasionner une coupure dans la fourniture d'électricité par manque de puissance disponible.

Ce foisonnement de producteurs décentralisés fait émerger de nouveaux défis pour les gestionnaires des réseaux électriques de distribution. Ainsi, afin de permettre à ces derniers d'assurer et de garantir une très bonne qualité de service, une mutation du réseau est nécessaire.

### 1.1.7 Mutation du réseau électrique de distribution

La pénétration des énergies renouvelables dans le réseau électrique représente un défi majeur pour leur développement, particulièrement pour le réseau électrique de distribution français auquel est raccordée aujourd'hui plus de 95 % de la puissance photovoltaïque produite en France. En réponse à ces nouvelles problématiques, une mutation du réseau électrique de distribution et de son mode de gestion est aujourd'hui nécessaire. Cette mutation doit contribuer à augmenter la capacité d'accueil du réseau, sans renforcement des infrastructures, et à limiter les contraintes en tension qui commencent à apparaître sur le réseau électrique de distribution basse tension du fait du déploiement d'une production décentralisée. En effet, le réseau électrique de distribution basse tension fait aujourd'hui face à des contraintes que l'on observait principalement sur le réseau HTA (Haute tension A), du fait de la multiplication des demandes de raccordement des producteurs PV de petite puissance en basse tension, ainsi que des nouveaux usages de l'énergie électrique. De ce fait, on assiste au développement de réseaux électriques intelligents qui associent à l'infrastructure du système électrique les technologies de l'information et de la communication afin d'ajuster en temps réel les flux entre producteurs et consommateurs. Entre autres technologies de l'information et de la communication, les réseaux électriques intelligents s'appuient sur :

- les outils de prévision et de modélisation permettant le développement de stratégies de gestion prévisionnelle ;
- les technologies de stockage pour lisser les productions renouvelables intermittentes et gérer efficacement les pics de consommation ;



- des compteurs communicants pouvant recevoir et envoyer des données et des ordres sans l'intervention physique d'un technicien et permettant de proposer de nouvelles offres tarifaires.

Les réseaux intelligents doivent permettre d'atteindre les objectifs fixés en matière de transition énergétique et doivent répondre à plusieurs enjeux.

- Flexibilité : ils permettent de gérer plus finement l'équilibre entre production et consommation.
- Fiabilité : ils améliorent l'efficacité énergétique et la sécurité de l'approvisionnement en énergie électrique.
- Accessibilité : ils favorisent la pénétration des énergies renouvelables sur l'ensemble du réseau.
- Économie : ils permettent, grâce à une meilleure gestion du système, une diminution des coûts de production et de consommation et, ainsi, de faire des économies d'énergie.

Le gestionnaire du réseau électrique de distribution français ENEDIS fait le pari du développement d'outils logiciels afin de lever les contraintes en tension qui pèsent sur le réseau électrique de distribution par le biais de plusieurs projets de démonstrateurs de réseaux électriques intelligents pour la mise en œuvre de stratégies de gestion prévisionnelle (Figure 1.3). Ces outils doivent permettre à ENEDIS, d'une part, d'augmenter la capacité d'accueil du réseau électrique de distribution sans construction de nouvelles infrastructures et, d'autre part, d'assurer le maintien de l'équilibre entre production et consommation.

### 1.1.8 Intérêt de la prévision

La prévision de la ressource solaire a plusieurs applications. Les prévisions à des horizons de temps allant de quelques minutes à plusieurs jours sont utilisées pour le contrôle-commande des centrales solaires, l'optimisation des transactions sur les marchés de l'énergie et la gestion des réseaux électriques.

La prévision de l'éclairement normal direct (ou DNI, pour *direct normal irradiance*) aide les gestionnaires des centrales solaires à concentration dans l'optimisation du contrôle-commande de la centrale [36]. En effet, la prévision du DNI permet l'optimisation des séquences de démarrage et d'arrêt de la turbine, permettant de faire face à des conditions de ressource solaire variables. Elle permet également une régulation en temps réel du débit du fluide caloporteur afin d'optimiser les performances de la centrale.

La prévision de l'éclairement global horizontal à des fins de prévision de la production PV permet aux gestionnaires de centrales solaires photovoltaïques équipées de systèmes de stockage de lisser la production électrique en gérant efficacement les systèmes de stockage. Elle leur permet aussi de délivrer une puissance moins incertaine et ainsi d'éviter les pénalités éventuelles liées aux écarts entre la production réelle et prévue. La prévision permet également aux gestionnaires de centrales solaires PV de garantir une continuité de service, en particulier dans les zones non interconnectées [37].

La variabilité de la production électrique des centrales solaires photovoltaïques a pour conséquence une volatilité croissante des prix infra-journaliers de l'énergie. La bourse européenne de l'électricité EPEX SPOT SE est une bourse de l'électricité au comptant pour plusieurs pays européens, dont la France. Elle gère également, pour ces pays, les marchés infra-journaliers. Ces marchés, organisés sur le principe de la négociation continue, sont aujourd'hui en plein essor et contribuent à la régulation du réseau électrique de distribution. En effet, ils garantissent aux producteurs d'électricité de ne vendre leur production que lorsque cela est nécessaire. Les contrats pour des volumes horaires d'énergie électrique – ces volumes sont en forte augmentation – peuvent être négociés, en France, jusqu'à 30 minutes avant livraison, voire jusqu'à 15 minutes en Allemagne, en Belgique et aux Pays-Bas. Les transactions transfrontalières peuvent être complétées jusqu'à 60 minutes avant livraison. Ainsi, les solutions de prévision pourraient contribuer aux activités de marché et autoriser, pour un producteur d'électricité, un surplus de rémunération.

Le complément de rémunération est un mécanisme de soutien qui rend obligatoire, pour les producteurs, la vente de l'électricité produite sur les marchés de l'énergie. Il s'agit d'une prime versée à

ces producteurs qui complète les revenus que génère la vente de l'électricité. Sur ces marchés, les ordres de vente sont établis à l'aide de prévisions de production. Si le volume d'électricité réellement produit diffère du volume vendu, le producteur reçoit des pénalités financières afin que soient régularisés les écarts. Pour un producteur, ces pénalités peuvent se chiffrer à plusieurs milliers d'euros par an et par centrale. Par conséquent, les solutions de prévision permettent, grâce à la qualité des prévisions infra-horaires et infra-journalières de la production électrique qu'elles fournissent, de limiter les pénalités financières reçues par les producteurs concernés.

Une autre mécanique pouvant avantageusement tirer profit, dans sa mise en œuvre, de prévisions infra-horaires et infra-journalières est l'ajustement à la baisse. Ainsi, un producteur pourrait ajuster à la baisse la production de ses centrales (voire, grâce à ces prévisions, décider de leur arrêt) et, ainsi, contribuer au nécessaire équilibre du système électrique. Les producteurs proposant ce type de flexibilité sont rémunérés.

La gestion du réseau électrique s'effectue en temps réel. Le gestionnaire du réseau doit en permanence assurer l'équilibre entre offre et demande. Cependant, l'augmentation de la part des énergies renouvelables intermittentes dans le mix électrique, en particulier le solaire PV, devient préoccupante pour les gestionnaires. Ainsi, afin d'assurer la stabilité du réseau électrique et une disponibilité optimale de l'énergie électrique, tout changement brusque de la production électrique doit être anticipé le plus longtemps à l'avance et le plus précisément possible. Il est donc nécessaire au gestionnaire du réseau d'avoir des prévisions fiables de la production électrique à tout instant. Les prévisions sont notamment très utiles pour les gestionnaires des réseaux électriques insulaires où l'absence d'interconnexion avec des réseaux voisins nécessite un pilotage intelligent des réserves et des différentes sources de production. Ainsi, la prévision s'impose comme un outil très important dans la mise en œuvre des stratégies de gestion des réseaux électriques. Elle gagne encore plus en importance dans la transformation des réseaux électriques actuels vers des réseaux électriques intelligents, dont l'objectif est d'ajuster en temps réel les flux entre producteurs et consommateurs.

## 1.1.9 Le projet « Smart Occitania »

### 1.1.9.1 Présentation du projet

Le gestionnaire du réseau électrique de distribution français ENEDIS fonde son activité de R&D sur plusieurs projets de démonstrateurs de réseaux électriques intelligents pour, notamment, la mise en œuvre de stratégies de gestion prévisionnelle. Début 2021, le site d'ENEDIS répertorie, en France métropolitaine, dix projets terminés et dix projets en cours. La [Figure 1.3](#) précise la localisation de ces démonstrateurs.

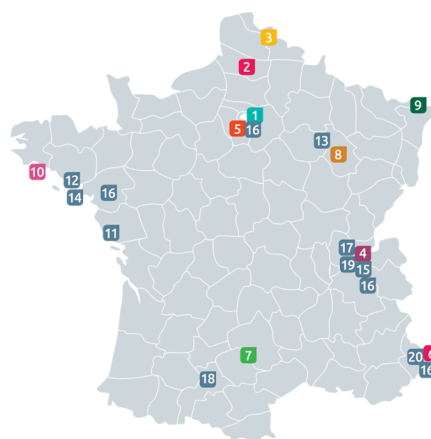


FIGURE 1.3 – Carte des projets de démonstrateurs de réseaux électriques intelligents portés par ENEDIS. Le projet ADEME « Smart Occitania » (2017-2021), auquel a participé PROMES-CNRS, est le projet 7 [38].



Les travaux présentés dans ce manuscrit – une bourse de thèse octroyée par la Région Occitanie en a permis la réalisation – sont connexes au projet « Smart Occitania » (repéré par le chiffre 7 sur la figure susmentionnée), visant à apporter la preuve de concept d’un réseau électrique intelligent en milieu périurbain/rural. Le projet, porté par ENEDIS et financé par l’Agence de la Transition Écologique (anciennement Agence De l’Environnement et de la Maîtrise de l’Energie ou ADEME), a impliqué des laboratoires de recherche et des industriels, parmi lesquels l’Institut de Recherche en Informatique de Toulouse (IRIT), PROMES-CNRS (UPR 8521), ACTIA Télécom et le groupe CAHORS. Le projet a débuté en mars 2017 et s’est terminé en mai 2021, pour un budget d’environ 8 millions d’euros. La contribution principale du laboratoire PROMES-CNRS est le développement d’algorithmes fondés sur la théorie de la commande prédictive (ou MPC, pour *Model-based Predictive Control*), tirant profit de « flexibilités », pour la gestion prévisionnelle de réseaux électriques de distribution basse tension faisant face (ou qui auront à faire face dans les années à venir) à une importante production décentralisée (thèse de Nouha Dkhili, soutenue le 9 décembre 2020 à Perpignan) [6]. Ce développement répond à un besoin exprimé par ENEDIS : disposer de solutions logicielles à coût calculatoire maîtrisé favorisant la pénétration au sein des réseaux électriques de distribution basse tension d’une production décentralisée (en particulier une production solaire photovoltaïque) diffuse et intermittente, limitant les problèmes opérationnels que rencontrent les opérateurs de ces réseaux et garantissant une qualité de service suffisante.

### 1.1.9.2 Cas d’étude

Un cas d’étude (Figure 1.4) – un quartier résidentiel périurbain d’environ 120 habitations (pour onze producteurs solaires photovoltaïques) situé à Villeneuve-de-la-Raho, près de Perpignan, et deux flexibilités (un méthaniseur et un château d’eau) – a permis la validation des algorithmes développés. Le contrôleur opère au niveau du transformateur HTA/BT et tire profit d’un méthaniseur qui injecte une production électrique dans le réseau et d’un château d’eau (une station de pompage) dont les pompes consomment de l’énergie électrique pour alimenter un réservoir et répondre à la demande locale en eau. La prise de décision est fondée sur la connaissance du comportement du système et sur des prévisions. Le pas de temps est de 10 min. Dans la mesure où les flexibilités considérées ne sont pas physiquement connectées au réseau – aucun réseau électrique de distribution basse tension en région Occitanie ne fait face à un important déploiement d’installations solaires photovoltaïques et ne dispose par ailleurs des flexibilités en question –, le cas d’étude a été simulé. Deux configurations distinctes ont été considérées : l’une permettant d’évaluer la capacité de la stratégie proposée à réduire l’écart entre production et consommation (120 habitations et 50 installations solaires photovoltaïques d’une puissance unitaire de 4kWc), l’autre provoquant d’importants dépassements de tension (600 habitations et 200 installations solaires photovoltaïques d’une puissance unitaire de 4kWc).



FIGURE 1.4 – Le cas d’étude : un réseau électrique de distribution basse tension (Villeneuve-de-la-Raho, près de Perpignan), un méthaniseur et un château d’eau (une station de pompage).

L’instrumentation du quartier – un réseau de pyranomètres pour la mesure distribuée de l’éclairage global horizontal et une caméra à grand angle de champ générant des images du ciel à haute dynamique – est succinctement présentée à la [Sous-sous-section 1.2.6.3](#) et détaillée au [Chapitre 4](#).

### 1.1.9.3 Stratégie proposée par PROMES-CNRS

La stratégie proposée est telle qu'à chaque pas de temps, le modèle du réseau électrique de distribution basse tension est alimenté par des données provenant à la fois du module de prévision et des instruments de mesure. Grâce à ce modèle sont évaluées les contraintes en tension ; ces informations sont transmises à l'algorithme d'optimisation qui tente de trouver une solution : les valeurs de consigne des deux flexibilités. Une fois le problème d'optimisation résolu, la première valeur de consigne est implémentée au prochain pas de temps. Le processus se répète à chaque pas de temps [39]. La commande prédictive repose, d'une part, sur l'utilisation d'un modèle dynamique du système à contrôler permettant de simuler son comportement futur et, d'autre part, sur la résolution, sur une fenêtre glissante, d'un problème d'optimisation contraint. La force d'une telle stratégie réside dans sa capacité à prendre en considération, en temps réel, les prévisions infra-journalières des perturbations affectant la stabilité du réseau électrique de distribution basse tension considéré, pour une meilleure anticipation des contraintes émergentes. Trois grandeurs stochastiques ont été prédites pour la mise en œuvre de la stratégie proposée : (1) la charge du réseau électrique de distribution basse tension considéré, (2) la production électrique des installations solaires photovoltaïques – les prévisions de la production électrique sont obtenues à partir de prévisions de l'éclairement global horizontal (la prévision du GHI fait l'objet des travaux de thèse présentés dans ce manuscrit) – et (3) la demande en eau. Des algorithmes pour la prévision infra-journalière (jusqu'à 24h) de ces grandeurs ont donc été développés par PROMES-CNRS. La charge du réseau est la demande en énergie électrique agrégée à l'échelle du quartier résidentiel périurbain considéré. Les données de production électrique sont elles aussi agrégées. Enfin, la demande en eau correspond au débit d'eau sortant du réservoir du château d'eau ; ce dernier influe sur le niveau d'eau stockée et, de fait, sur le fonctionnement de l'installation [40].

Le choix d'une méthode de prévision étant conditionné par les exigences de l'application considérée, il était important de garder à l'esprit les objectifs du projet « Smart Occitania », tels que décidés par ENEDIS, à savoir développer des algorithmes pour la mise en œuvre d'une stratégie de contrôle capable de fonctionner en temps réel, ce qui impliquait la recherche d'un compromis entre performance et coût calculatoire. Ainsi, PROMES-CNRS a fait le choix d'une régression non paramétrique par processus Gaussien (ou GPR, pour *Gaussian process regression*). Les modèles GPR développés, bien que requérant une mise à jour de leurs paramètres à chaque pas de temps, sont adaptés à l'application considérée. Toutefois, le contrôleur prédictif est dépendant d'un accès en temps réel aux mesures permettant cette mise à jour. Par conséquent, une instrumentation et une infrastructure de communication garantissant un transfert de données rapide et fiable est nécessaire. Par ailleurs, les intervalles de confiance associés aux prévisions GPR ont été mis à profit pour améliorer la robustesse du contrôleur prédictif aux erreurs de prévision.

L'apport des travaux réalisés par PROMES-CNRS réside principalement dans le traitement du cas d'étude, impliquant des flexibilités jusqu'alors inexploitées par ENEDIS à des fins de régulation de la tension, mais également dans la gestion du fonctionnement tout-ou-rien du château d'eau sans avoir recours à la programmation non linéaire mixte en nombres entiers (ou MINLP, pour *Mixed Integer Nonlinear Programming*) ou à une relaxation [39, 40].

### 1.1.9.4 Résultats et discussion

Pour ce qui est de la prévision infra-journalière de la charge du réseau, le modèle GPR développé est performant pour des horizons de prévision courts (l'erreur quadratique moyenne normalisée est égale à 9% pour un horizon de 10min). Cependant, l'erreur de prévision se dégrade rapidement à mesure que l'horizon augmente. Par ailleurs, il convient de noter que le modèle GPR développé – le noyau périodique, le noyau exponentiel quadratique et le noyau quadratique rationnel sont associés – est capable de fournir des prévisions satisfaisantes en ayant été entraîné avec seulement deux semaines de données. Cela diminue non seulement le coût calculatoire associé au développement du modèle mais permet aussi une adaptation relativement rapide à une éventuelle reconfiguration du réseau électrique

de distribution ou à la disponibilité de nouveaux jeux de données.

En raison de la nature régulière des données de demande en eau, la combinaison simple du noyau périodique et du noyau quadratique rationnel s'avère suffisante pour capturer plutôt fidèlement la dynamique infra-journalière de la série temporelle considérée. En effet, pour un horizon de prévision de 10 min, l'erreur quadratique moyenne normalisée est égale à 9 %. Cette erreur se stabilise rapidement autour de 13 % pour des horizons plus lointains.

Pour ce qui est de la prévision infra-journalière de l'éclairement global horizontal, le modèle GPR développé – le noyau périodique et le noyau quadratique rationnel sont associés – est performant pour des horizons de temps courts (l'erreur quadratique moyenne normalisée est égale à 15 % pour un horizon de 10 min). Toutefois, les performances du modèle se dégradent à mesure que l'horizon augmente, l'erreur quadratique moyenne normalisée se stabilisant à environ 32 % pour des horizons supérieurs à 6 h. En effet, les fluctuations infra-journalières deviennent de plus en plus difficiles à appréhender.

Concernant la stratégie proposée par PROMES-CNRS, les résultats obtenus en simulation témoignent de la capacité du contrôleur prédictif développé à gérer efficacement les flux d'énergie au sein du réseau de distribution basse tension considéré [39]. Ainsi, le contrôleur réduit l'écart entre offre et demande et minimise les dépassements de tension. Il limite par ailleurs les phénomènes de refoulement vers les réseaux moyenne tension (HTA) et représente un pas en avant vers la possibilité de faire fonctionner les réseaux électriques de distribution basse tension en mode îloté, à condition qu'un moyen de stockage d'énergie approprié (et d'une capacité suffisante) soit disponible. Les résultats obtenus montrent également que la stabilité des réseaux électriques de distribution basse tension tels que celui considéré dans le cas d'étude (Villeneuve-de-la-Raho) et, de fait, la qualité de service sont très loin d'être menacées par le taux actuel de pénétration du solaire photovoltaïque. Les réseaux de distribution basse tension étant traditionnellement surdimensionnés, leur capacité à absorber une production décentralisée, sans impact significatif sur leur stabilité, en particulier en matière de fluctuations de tension, est importante. Toutefois, le déploiement à grande échelle des technologies tirant profit des énergies renouvelables pourrait changer la donne.

Les résultats obtenus en simulation montrent par ailleurs que les erreurs de prévision ne provoquent pas de dégradation significative des performances du contrôleur prédictif. Toutefois, afin de prévenir une éventuelle dégradation de ses performances, ce dernier a été modifié afin que soit améliorée sa robustesse aux erreurs de prévision par la prise en compte des intervalles de confiance associés aux prévisions GPR (l'approche choisie est le problème dit « min-max ») [40]. De plus, pour la mise en œuvre d'une stratégie telle que celle que propose PROMES-CNRS pour la gestion prévisionnelle des réseaux électriques de distribution basse tension, le choix du pas de temps est très important. En effet, il convient de rechercher un compromis entre coût calculatoire et granularité des modèles afin d'appréhender au mieux les phénomènes – les dynamiques sont rapides – observés sur le réseau. La mise en œuvre d'une telle stratégie est par ailleurs dépendante d'un accès en temps réel à des mesures, cet accès étant contraint par des considérations techniques. Des solutions existent, en particulier grâce au déploiement, au cours des dernières années, d'une infrastructure avancée de mesure. Le pas de temps considéré – 10 min – résulte d'un compromis : il permet la réalisation des calculs nécessaires à l'obtention des prévisions et à la résolution du problème d'optimisation ; toutefois, ce pas de temps limite la capacité de la stratégie proposée par PROMES-CNRS à appréhender efficacement les dynamiques rapides des phénomènes observés sur les réseaux électriques et rend difficile certaines interventions. La stratégie proposée peut être vue comme une stratégie de contrôle de haut niveau, qu'il convient d'associer à des outils de planification à plus long terme et à des méthodes d'exploitation de bas niveau, telles que les méthodes qui relèvent du génie électrotechnique et qui permettent de réagir à des phénomènes électriques rapides.

Enfin, les travaux réalisés par PROMES-CNRS dans le cadre du projet « Smart Occitania » mettent en lumière les limitations du duo de flexibilités (un méthaniseur et un château d'eau). Tout d'abord, le fonctionnement en tout-ou-rien du château d'eau augmente la complexité calculatoire du problème d'optimisation et s'avère être un handicap pour les applications temps réel. De plus, ce fonctionnement

aggrave les fluctuations de tension et réduit la durée de vie des différents équipements. En outre, les flexibilités considérées doivent pouvoir adapter leurs marges de manœuvre – à coût minimal – à une demande croissante en énergie électrique et à la pénétration d'une production décentralisée au sein des réseaux électriques de distribution en plein essor. Enfin, les flexibilités considérées ne sont pas facilement « extensibles », en particulier le méthaniseur dont la source d'énergie repose sur un procédé organique complexe.

## 1.2 L'éclairage global horizontal

Cette section introduit l'éclairage global horizontal au sol (ou GHI, pour *global horizontal irradiance*), dont la prévision est l'objet des travaux présentés dans ce manuscrit. Après un rapide rappel des fondamentaux sur le rayonnement électromagnétique et le rayonnement solaire, nous présenterons les sites étudiés et les instruments de mesure de l'éclairage utilisés.

### 1.2.1 Fondamentaux sur le rayonnement électromagnétique

Une onde électromagnétique est composée d'un champ électrique et d'un champ magnétique qui oscillent perpendiculairement entre eux et à l'axe de propagation de l'onde. La longueur d'onde  $\lambda$  définie par la distance entre deux maxima de la sinusoïde est liée à la période  $T$  par l'équation (1.1), où  $c$  est la vitesse de propagation de l'onde :

$$c = \frac{\lambda}{T} \quad (1.1)$$

Dans la définition des termes énergétiques utilisés en transfert radiatif, on retrouve la notion d'angle solide définie à partir d'une sphère unitaire. La surface de cette dernière représente l'angle solide entier  $\Omega = 4\pi$ . L'angle solide élémentaire  $d\Omega$  dans la direction définie par l'angle  $\theta$  par rapport à la normale et l'angle  $\phi$  dans le plan de la surface de la sphère représente la fraction de sphère dans cette direction. Il est donné par l'équation (1.2) et son unité est le stéradian (sr) :

$$d\Omega = \sin \theta d\theta d\phi \quad (1.2)$$

Quand une onde électromagnétique de longueur d'onde  $\lambda$  rencontre de la matière, une partie de l'onde est réfléchiée, une autre partie est absorbée par la matière et le reste est transmis ou diffusé. Les paramètres utilisés pour décrire ces phénomènes sont :

- le facteur d'absorption  $\beta_\lambda$  défini comme le ratio entre le rayonnement absorbé et le rayonnement incident ;
- le facteur de réflexion ou albédo monochromatique  $\delta_\lambda$  défini comme le ratio entre le rayonnement réfléchi et le rayonnement incident ;
- le facteur de transmission  $\tau_\lambda$  défini comme le ratio entre le rayonnement transmis et le rayonnement incident.

### 1.2.2 Le rayonnement solaire

Le rayonnement solaire désigne l'ensemble des ondes électromagnétiques émises par le Soleil. On note une différence entre le rayonnement au sommet de l'atmosphère et celui au niveau du sol. En effet, une partie de ce rayonnement est absorbée par les gaz présents dans l'atmosphère. Ces gaz ont des facteurs d'absorption et de transmission qui dépendent de la longueur d'onde. Ce phénomène explique une partie de la variabilité du rayonnement solaire puisque la composition atmosphérique est variable dans le temps et dans l'espace.

Un nuage est une masse visible constituée d'une grande quantité de gouttelettes d'eau (parfois de cristaux de glace associés à des aérosols chimiques ou à des minéraux) en suspension dans l'atmosphère

au-dessus de la surface d'une planète. La plupart des nuages sont situés dans la troposphère. Selon leur altitude et leur composition, ils présentent diverses formes et propriétés. L'impact des nuages sur le rayonnement solaire reçu au sol est difficile à décrire de façon quantitative. En effet, les nuages d'un même type peuvent avoir un effet d'atténuation différent en fonction de leurs propriétés macroscopiques et microscopiques (géométrie, épaisseur optique, quantité d'eau, etc.). Cependant, les nuages ont généralement un effet atténuant sur le rayonnement reçu au sol puisqu'ils réfléchissent une partie de la lumière incidente vers l'espace.

## NOTE

*La troposphère est la première couche de l'atmosphère terrestre comprise entre la surface du globe et la stratosphère (seconde couche de l'atmosphère terrestre). Sa limite supérieure, encore appelée tropopause, se situe à une altitude d'environ 8 à 15 km selon la latitude et la saison. La troposphère représente environ 80 % de la masse totale de l'atmosphère. On y trouve la plupart des phénomènes météorologiques.* ◇

Le rayonnement solaire arrivant au sol dépend de la position du Soleil, du rayonnement extra-atmosphérique et des caractéristiques atmosphériques propres au site étudié. Au rayonnement solaire extra-atmosphérique, c'est-à-dire le rayonnement électromagnétique émis par le Soleil arrivant en incidence normale sur la couche externe de l'atmosphère, est associé un éclairement énergétique.

## NOTE

*L'éclairement énergétique, exprimé en  $W \cdot m^{-2}$ , est défini comme la densité du flux énergétique incident par rapport à l'aire en un point d'une surface réelle ou fictive. Comme il n'y a pas de confusion possible avec l'éclairement lumineux, exprimé en  $lm \cdot m^{-2}$ , nous parlerons simplement d'éclairement dans la suite.* ◇

L'éclairement solaire se décline en trois composantes.

- L'éclairement normal direct (ou DNI, pour *direct normal irradiance*) : il s'agit du rayonnement arrivant en incidence normale depuis le Soleil sur un plan perpendiculaire à la direction du rayon. Le pyrhéliomètre est l'instrument de mesure le plus couramment utilisé pour la mesure du DNI. Il se présente sous la forme d'un tube dont les dimensions définissent l'angle d'ouverture et au fond duquel est placée une thermopile. Afin de mesurer le flux solaire direct, le pyrhéliomètre doit être aligné dans l'axe du Soleil. Il doit alors être monté sur un traqueur solaire, si l'on souhaite automatiser l'acquisition du DNI (Figure 1.5).



FIGURE 1.5 – Aperçu de deux pyrhéliomètres montés sur traqueurs solaires.

- L'éclairement diffus horizontal (ou DHI, pour *diffuse horizontal irradiance*) : il s'agit du rayonnement reçu par l'ensemble des autres chemins possibles sur un plan horizontal (diffusion de l'atmosphère, réflexion simple ou multiple avec les nuages, l'atmosphère elle-même, le sol ou



des objets de l'environnement). Le DHI se mesure généralement à l'aide d'un pyranomètre. Tout comme le pyréliomètre, le pyranomètre utilise une thermopile pour convertir le flux solaire en flux thermique puis en un signal électrique. Pour la mesure du DHI, le pyranomètre est généralement installé sur un traqueur solaire équipé d'une boule d'ombrage masquant le Soleil.

- L'éclairement global horizontal : il s'agit de la combinaison de l'éclairement diffus horizontal et de la partie horizontale de l'éclairement normal direct. À partir de la connaissance du DNI, du DHI et de la position du soleil, on déduit le GHI donné par l'équation (1.3) :

$$\text{GHI} = \text{DNI} \cdot \cos \theta_z + \text{DHI} \quad (1.3)$$

où  $\theta_z$  est l'angle entre le Soleil et le zénith. Cependant, en pratique, le GHI se mesure à l'aide d'un pyranomètre pour lequel le Soleil est apparent. Un traqueur solaire n'est donc pas nécessaire pour en faire la mesure.

Pour une mesure simultanée de ces trois composantes, l'instrument le plus couramment utilisé est le pyranomètre à bande rotative (ou RSI, pour *rotating shadowband irradiator*). Cet instrument est équipé d'une photodiode montée horizontalement et permettant la mesure du GHI (Figure 1.6). Une bande rotative vient alors occulter la photodiode périodiquement afin de mesurer le DHI. À partir des mesures du DHI et du GHI, et de la connaissance de la position du Soleil, le DNI peut alors être estimé (équation (1.3)). Cet instrument a donc l'avantage de fournir plusieurs mesures de l'éclairement, contrairement au pyréliomètre. Le fait d'avoir une simple bande rotative au lieu d'un traqueur solaire sur deux axes réduit également la complexité du système tout en améliorant sa robustesse. En outre, des études ont montré que le RSI était nettement moins sensible aux salissures que le pyréliomètre.



FIGURE 1.6 – Deux RSI installés au laboratoire PROMES-CNRS à Perpignan.

#### NOTE

Une caméra à grand angle de champ peut potentiellement être utilisée comme instrument de mesure du DNI, du DHI et du GHI [41, 42]. ◇

L'éclairement global horizontal au sol peut être décomposé en deux parties.

- Une composante déterministe liée à la géométrie Terre-Soleil et à l'atmosphère sans nuage : l'éclairement par ciel clair  $\text{GHI}_{\text{cs}}$ .
- Une composante stochastique liée à la couverture nuageuse : l'indice ciel clair  $k_c$ .

### 1.2.3 Modèles ciel clair

Les modèles ciel clair permettent d'estimer l'éclairement au niveau du sol en conditions ciel clair, c'est-à-dire en l'absence de nuage devant le Soleil, en fonction de l'angle entre le Soleil et le zénith,

de l'altitude du site et de divers paramètres atmosphériques. Il existe un vaste choix de modèles ciel clair dans la littérature, dont la plupart sont passés en revue dans [43, 44]. L'approche proposée dans [44] pour élaborer un modèle DNI ciel clair précis, qui combine le modèle ciel clair d'Ineichen et Perez [45] avec une persistance sur le coefficient de turbidité, a été adaptée au GHI pour obtenir un modèle GHI ciel clair. Dans [44], les auteurs ont prouvé que cette approche est capable de fournir des prévisions infra-journalières précises du DNI par ciel clair. Le GHI en conditions ciel clair est donné par :

$$\text{GHI}_{\text{cs}}(t) = c_1 \cdot I_0 \cdot \cos(\theta_z(t)) \cdot \exp(-c_2 \cdot m(t) \cdot (f_{h1} + f_{h2}(T_L(t - \Delta t) - 1))) + 0,01 \cdot (m(t))^{1,8} \quad (1.4)$$

où  $t$  est le temps,  $I_0$  est l'éclairement solaire extra-atmosphérique,  $\theta_z$  est l'angle entre le Soleil et le zénith,  $m$  est la masse d'air optique relative,  $T_L$  est le trouble atmosphérique introduit par Linke [46] et  $f_{h1}$ ,  $f_{h2}$ ,  $c_1$ , et  $c_2$  sont des paramètres. Les paramètres  $f_{h1}$ ,  $f_{h2}$ ,  $c_1$  et  $c_2$  sont respectivement donnés par :

$$f_{h1} = \exp\left(-\frac{h}{8000}\right) \quad (1.5)$$

$$f_{h2} = \exp\left(-\frac{h}{1250}\right) \quad (1.6)$$

$$c_1 = 5,09 \times 10^{-5}h + 0,868 \quad (1.7)$$

$$c_2 = 3,92 \times 10^{-5}h + 0,0387 \quad (1.8)$$

où  $h$  est l'altitude. Le coefficient  $T_L$  correspond au nombre d'atmosphères sans aérosols ou vapeur d'eau qu'il faudrait superposer pour aboutir à l'atténuation constatée. Sa valeur est autour de 3 atmosphères dans la plupart des régions d'Europe, mais elle peut monter jusqu'à 7 atmosphères dans les zones très polluées.

#### NOTE

*Il faut noter que si, en conditions ciel clair, il n'y a pas de nuage voilant le Soleil au moment de la mesure (du point de vue de l'observateur), cela ne veut pas dire qu'il n'y a aucun nuage dans le ciel (toujours du point de vue de l'observateur). Dans ce dernier cas, on parle plutôt de « ciel dégagé ».* ◇

### 1.2.3.1 L'éclairement solaire extra-atmosphérique

L'éclairement solaire extra-atmosphérique  $I_0$  est l'éclairement arrivant à incidence normale sur la couche externe de l'atmosphère. En raison de l'orbite elliptique de la Terre, cet éclairement n'est pas constant et varie au cours de l'année. Toutefois, lorsque la distance Terre-Soleil n'est pas connue, l'éclairement  $I_0$  peut être estimé à l'aide d'une formule simplifiée, comme par exemple :

$$I_0 = \left(\frac{D_0}{D_{\text{UA}}}\right)^2 I_{\text{UA}} \quad (1.9)$$

où  $I_0$  est l'éclairement solaire extra-atmosphérique incident à la distance Terre-Soleil  $D_0$ ,  $D_{\text{UA}}$  est la distance moyenne Terre-Soleil ( $D_{\text{UA}} = 1 \text{ UA} \simeq 1,5 \times 10^8 \text{ km}$ ) et  $I_{\text{UA}}$  est l'éclairement solaire extra-atmosphérique incident à la distance moyenne Terre-Soleil, également appelé constante solaire.

L'estimation de la constante solaire a fait l'objet de nombreuses missions satellitaires, en raison de son importance dans le développement de modèles climatiques [47, 48]. Plusieurs mesures peuvent ainsi être trouvées dans la littérature [49, 50]. Cependant, une réévaluation de la valeur de la constante solaire en utilisant une base de données couvrant 42 ans a été effectuée récemment et a permis d'obtenir  $I_{\text{UA}} = (1361,1 \pm 0,5) \text{ W} \cdot \text{m}^{-2}$  [51].

Pour le calcul du rapport  $\frac{D_0}{D_{\text{UA}}}$ , l'algorithme SG2 (pour *Solar Geometry 2*) qui est précis et rapide a été choisi [52]. Cet algorithme permettant de calculer la distance Terre-Soleil en temps réel est une version évoluée de l'algorithme SG (pour *Solar Geometry*) développé dans le cadre de l'Atlas européen du rayonnement solaire [53, 54].

### 1.2.3.2 La masse d'air optique relative

La masse d'air optique relative, notée  $m$ , est définie comme le rapport entre la distance parcourue par la lumière à travers l'atmosphère et la distance que parcourrait ce faisceau au travers d'une atmosphère standard, au niveau de la mer et avec le Soleil au zénith. La position du Soleil est donc nécessaire pour calculer cette masse d'air. Ainsi, afin de prendre en compte les effets de la courbure de la Terre, de l'altitude du site et des densités variables de l'atmosphère sur le parcours moyen du rayonnement solaire, plusieurs formulations différentes de  $m$  existent dans la littérature [55]. Cependant, pour le développement du modèle GHI ciel clair, la formule de Kasten et Young [56], également utilisée dans [44] pour l'élaboration d'un modèle DNI ciel clair, a été choisie :

$$m(t) = \frac{1}{\cos(\theta_z(t)) + \frac{a_1}{(a_2 - \theta_z(t))^{a_3}}} \quad (1.10)$$

où  $t$  est le temps,  $a_1$ ,  $a_2$  et  $a_3$  sont des paramètres ajustables et  $\theta_z$  est l'angle entre le Soleil et le zénith. Les paramètres  $a_1$ ,  $a_2$  et  $a_3$  sont respectivement égaux à :

$$a_1 = 0,50572 \quad a_2 = 96,07995 \quad a_3 = 1,6364 \quad (1.11)$$

#### NOTE

*Il convient de noter que les conditions atmosphériques rencontrées sur les sites où ont été réalisées les mesures du GHI exploitées dans cette thèse sont proches de l'atmosphère normalisée utilisée pour définir les paramètres ajustables de l'équation (1.10).*  $\diamond$

### 1.2.4 L'indice ciel clair

L'indice ciel clair, noté  $k_c$ , est défini comme le rapport entre l'éclairement global horizontal observé au sol (GHI) et l'éclairement global horizontal en conditions ciel clair (GHI<sub>cs</sub>) :

$$k_c(t) = \frac{\text{GHI}(t)}{\text{GHI}_{cs}(t)} \quad (1.12)$$

où  $t$  est le temps.  $k_c$  indique la présence ou non d'un nuage devant le Soleil. Il varie entre 0 et 1, où  $k_c = 0$  signifie qu'un nuage épais occulte le Soleil tandis que  $k_c = 1$  signifie l'absence de nuage devant le Soleil. Enfin, les cas intermédiaires sont tels que des nuages fins viennent voiler le Soleil.

#### NOTE

*L'indice ciel clair représente les variations du GHI liées à la couverture nuageuse. La décomposition de l'éclairement global horizontal en éclairement par ciel clair et indice ciel clair constitue le point de départ de certains modèles de prévision du GHI. En effet, dans ce cas, la prévision du GHI revient donc à faire une prévision de  $k_c$  et de GHI<sub>cs</sub> :*

$$\text{GHI}(t + \Delta t) = k_c(t + \Delta t) \cdot \text{GHI}_{cs}(t + \Delta t) \quad (1.13)$$

*où  $\Delta t$  est l'horizon de prévision. En étudiant l'indice ciel clair plutôt que le GHI, on s'affranchit des cycles connus de l'éclairement pour se concentrer uniquement sur les variations liées aux conditions atmosphériques. Néanmoins, cette approche requiert une certaine vigilance dans le choix du modèle ciel clair.*  $\diamond$

### 1.2.5 L'indice de clarté

L'indice de clarté, noté  $k_t$ , correspond au rapport entre l'éclairement global horizontal observé au sol et l'éclairement solaire extra-atmosphérique sur un plan horizontal :

$$k_t(t) = \frac{\text{GHI}(t)}{I_0 \cos(\theta_z(t))} \quad (1.14)$$



où  $t$  est le temps,  $I_0$  est l'éclairement solaire extra-atmosphérique et  $\theta_z$  est l'angle entre le Soleil et le zénith. L'indice de clarté rend compte de l'atténuation de l'éclairement solaire due à la traversée de l'atmosphère et des fluctuations dues aux nuages. Son calcul ne nécessite pas l'utilisation d'un modèle ciel clair. Par définition, l'indice ciel clair est toujours supérieur à l'indice de clarté. Certains modèles de prévision sont fondés sur l'indice de clarté (voir la [Sous-sous-section 1.3.2.4](#)).

### 1.2.6 Mesures du GHI

Les données utilisées dans cette étude proviennent de mesures du GHI effectuées au laboratoire PROMES-CNRS, situé sur deux sites dans la région Occitanie à savoir : Font-Romeu-Odeillo-Via et Perpignan. Outre les données de ces deux sites, nous avons aussi utilisé des mesures du GHI provenant d'un réseau électrique basse tension (une zone résidentielle) instrumenté par PROMES-CNRS et ENEDIS situé à Villeneuve-de-la-Raho, non loin de Perpignan. Les caractéristiques des trois sites sont présentées dans le [Tableau 1.1](#).

TABLEAU 1.1 – Caractéristiques des sites étudiés.

Caractéristiques	Font-Romeu-Odeillo-Via	Perpignan	Villeneuve-de-la-Raho
Localisation	42°29'36" N / 2°1'47" E	42°39'53" N / 2°54'37" E	42°37'59" N / 2°55'21" E
Altitude	1650 m	50 m	50 m
Instrument de mesure du GHI	Pyranomètre	RSI	Pyranomètre
Période d'échantillonnage	10 min	10 min	10 min
Jeu de données	Du 01/01/2018 au 31/12/2019	Du 01/01/2016 au 31/12/2017	Du 01/08/2019 au 31/07/2020

Les données sont échantillonnées avec un pas de temps de 10 min. Une illustration de six jours de données GHI (du 22/11/2017 au 27/11/2017) issues de mesures réalisées sur le site de Perpignan est présentée sur la [Figure 1.7](#).

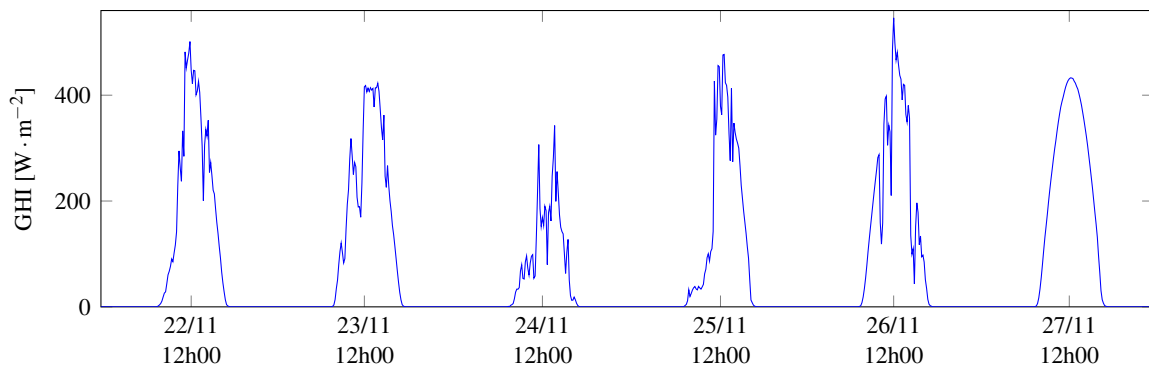


FIGURE 1.7 – Illustration de six jours de données GHI (du 22/11/2017 au 27/11/2017). Les mesures du GHI ont été réalisées au laboratoire PROMES-CNRS, situé à Perpignan, en région Occitanie (sud de la France).

#### 1.2.6.1 Site de Font-Romeu-Odeillo-Via

Situé à haute altitude (1650m), le site de Font-Romeu-Odeillo-Via (le grand four solaire) est caractérisé par un climat montagneux relativement sec avec une influence méditerranéenne. Les étés

sont chauds et les hivers glacés avec des précipitations neigeuses. En outre, des orages violents s'abattent occasionnellement sur la région. Les données sont collectées à l'aide d'un pyranomètre Kipp & Zonen CM5 (Figure 1.8).

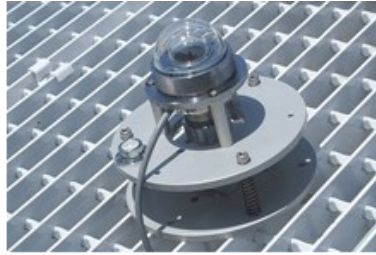


FIGURE 1.8 – Aperçu d'un pyranomètre CM5 de Kipp & Zonen [57].

Les données couvrant les années 2018 et 2019, tirées des mesures du GHI réalisées sur ce site, ont été utilisées dans le Chapitre 2 pour l'étude comparative des modèles de prévision fondés sur l'apprentissage automatique. La Figure 1.9 présente une carte des données couvrant les deux années.

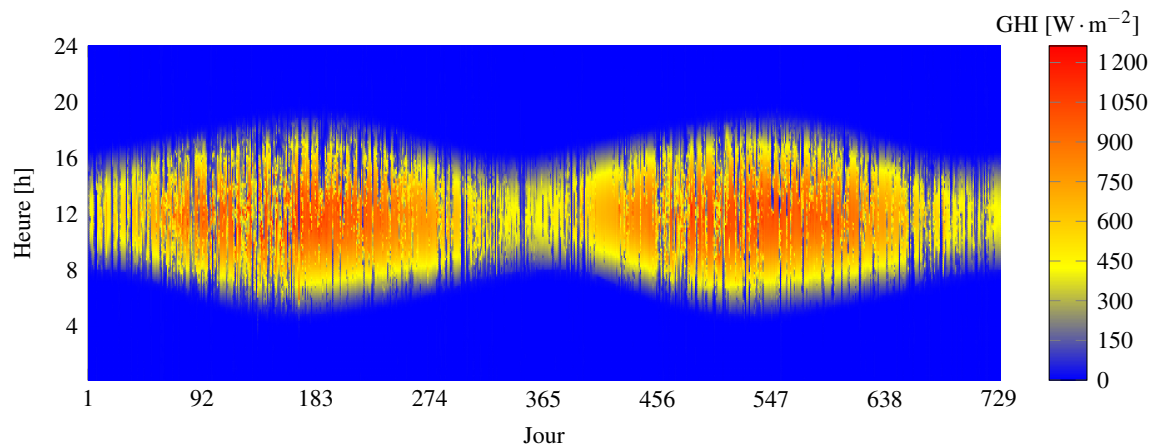


FIGURE 1.9 – Données GHI couvrant les années 2018 et 2019. Elles proviennent de mesures réalisées à l'aide d'un pyranomètre CM5 de Kipp & Zonen au laboratoire PROMES-CNRS, situé à Font-Romeu-Odeillo-Via, en région Occitanie (sud de la France).

### 1.2.6.2 Site de Perpignan

Les données issues du site de Perpignan (Tecnosud) sont collectées à l'aide d'un RSI (Figure 1.6). La ville de Perpignan bénéficie d'un climat méditerranéen caractérisé par des étés très chauds et des hivers assez doux. La région est souvent soumise à un vent de direction nord-ouest, la Tramontane. Les données couvrant les années 2016 et 2017 ont été utilisées dans le Chapitre 3. La Figure 1.10 montre une carte des données couvrant les deux années.

### 1.2.6.3 Site de Villeneuve-de-la-Raho

Des mesures du GHI provenant d'un réseau électrique basse tension (une zone résidentielle) situé à Villeneuve-de-la-Raho ont aussi été utilisées au cours de nos travaux. PROMES-CNRS ayant fait le choix d'une prévision de la production électrique des installations solaires photovoltaïques à partir de prévisions du GHI, la zone résidentielle a été instrumentée par un réseau de pyranomètres et une caméra à grand angle de champ. Les mesures distribuées du GHI et les observations atmosphériques réalisées grâce à la caméra sont à l'origine du développement d'algorithmes pour la prévision dans le temps et l'espace de la ressource solaire présentés dans ce manuscrit (Chapitre 4), à des fins de prévision

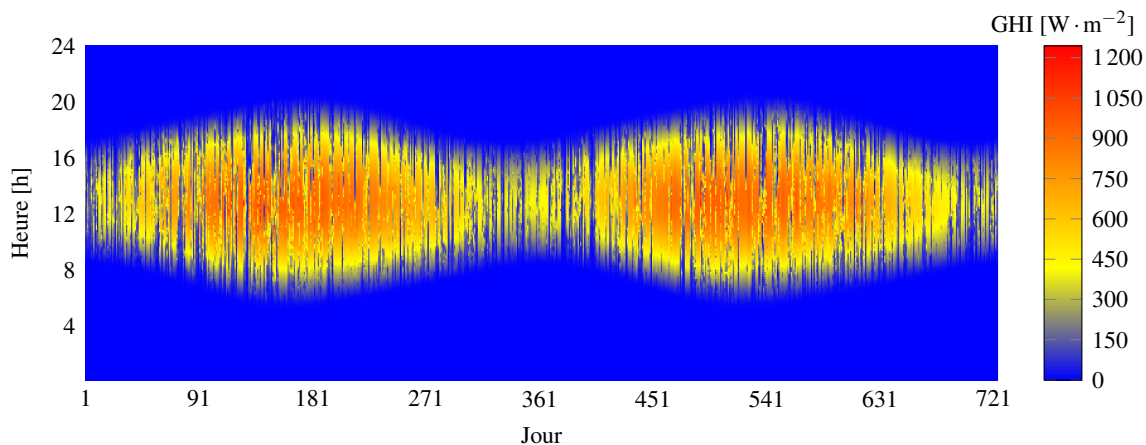


FIGURE 1.10 – Données GHI couvrant les années 2016 et 2017. Elles proviennent de mesures réalisées à l'aide d'un RSI au laboratoire PROMES-CNRS, situé à Perpignan, en région Occitanie (sud de la France).

de la production électrique des installations solaires photovoltaïques équipant la zone considérée. Un réseau de 16 instruments pour la mesure distribuée du GHI – des pyranomètres SP Lite 2 de la marque Kipp & Zonen – a été installé au sein et aux alentours de la zone résidentielle choisie sur des supports déjà existants (lampadaire ou poteau électrique). La Figure 1.11 donne la répartition dans l'espace des instruments.



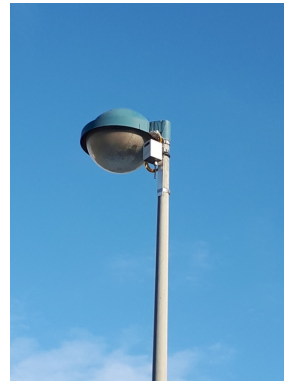
FIGURE 1.11 – Les 16 pyranomètres installés à Villeneuve-de-la-Raho. Chaque emplacement est repéré sur la carte par une lettre majuscule (A à P). La zone résidentielle est le triangle ABC et la caméra est installée au point D.

Une caméra à grand angle de champ complète le réseau de pyranomètres. Cette dernière est installée sur le toit du poste Estagnol (Figure 1.12a), au sein de la zone résidentielle, et fournit des observations atmosphériques. L'instrument, développé par PROMES-CNRS et commercialisé par la société PROMECA, est équipé d'un ordinateur embarqué de dernière génération permettant la mise en œuvre en temps réel de solutions logicielles avancées et d'un système optique optimisé (couche antireflet, revêtement hydrophobe et unité de ventilation). La sous-couche logicielle qui accompagne l'instrument est composée, d'une part, d'un algorithme pour la génération d'images à haute dynamique préservant l'information dans la région circumsolaire – cette information est cruciale pour la prévision à court terme de la ressource solaire – et, d'autre part, d'un algorithme d'autocalibration géométrique

fondé sur la comparaison entre la position du Soleil tel que détecté sur l'image et sa position théorique. La [Figure 1.12b](#) montre une vue du pyranomètre D sur un lampadaire.



(a) La caméra à grand angle de champ installée à Villeneuve-de-la-Raho.



(b) Le pyranomètre au point D, installé sur un lampadaire.

FIGURE 1.12 – La caméra et l'un des pyranomètres installés à Villeneuve-de-la-Raho.

#### NOTE

*Rappelons que l'objectif est de prévoir le GHI (prévision localisée puis spatio-temporelle) en vue de prévoir la production PV à des fins de gestion intelligente du réseau électrique de distribution en région Occitanie et la zone d'expérimentation est le quartier résidentiel situé à Villeneuve-de-la-Raho. Cependant, l'instrumentation de cette zone a pris un certain temps. Par ailleurs, nous avons fait face à des problèmes de transmission des données et de stockage qui ont été résolus. Il a donc fallu attendre longtemps avant d'avoir une base de donnée consolidée. Les données du site de Font-Romeu-Odeillo-Via étant disponibles et propres, elles ont été utilisées pour mener l'étude comparative des outils d'intelligence artificielle au [Chapitre 2](#) afin de dégager la méthode adaptée à la prévision du GHI à l'aune des exigences de l'application visée. Le site Perpignan est situé juste à côté du cas d'étude. Pour des raisons de commodité, les données issues de ce site ont été utilisées pour le développement des modèles de prévision localisée du GHI au [Chapitre 3](#).* ◇

#### 1.2.6.4 Contrôle qualité des mesures du GHI

La qualité des modèles de prévision, surtout ceux fondés sur l'analyse d'une base de données et sur l'apprentissage automatique, est liée à celles des données. Un contrôle qualité des observations du GHI s'avère donc nécessaire afin de détecter et de corriger d'éventuelles anomalies. L'utilisation des mesures sans correction de ces défauts peut potentiellement conduire à des résultats de prévision erronés. Le contrôle qualité des mesures de la ressource solaire comprend divers tests visant à établir des limites entre lesquelles les données sont dites acceptables. Ces tests permettent de détecter des mesures physiquement impossibles ou une récurrence d'événements rares qui proviendraient, par exemple, d'un problème lors de l'acquisition.

Dans [58], les différentes façons d'encadrer des données météorologiques et de la ressource solaire afin de s'assurer de leur qualité ont été exposées en fonction de la résolution temporelle. Les auteurs se sont inspirés des travaux de [59] et [60]. Pour des mesures infra-horaires, la procédure de contrôle qualité est celle typiquement adoptée par WRMC-BSRN (*World Radiation Monitoring Center - Baseline Surface Radiation Network*) sur des données à pas de temps de 1 min [58, 61]. Les tests de contrôle qualité effectués sur les données GHI à pas de temps de 10 min utilisées dans cette thèse sont présentés dans le [Tableau 1.2](#) :

- le premier test permet de détecter les mesures physiquement impossibles ;
- le second test porte sur la détection des observations rares ;

— le dernier test est un test croisé entre l'éclairement global horizontal et l'éclairement diffus horizontal afin de vérifier leur cohérence physique.

TABLEAU 1.2 – Tests de contrôle qualité réalisés sur les mesures du GHI. Il s'agit des tests proposés par [58, 61] pour des données infra-horaires. DHI est l'éclairement diffus horizontal,  $I_0$  est l'éclairement solaire extra-atmosphérique et  $\theta_z$  est l'angle entre le Soleil et le zénith.

Test	Condition
Mesure physiquement possible	$0 < \text{GHI} < \min(1,2 \cdot I_0; 1,5 \cdot I_0 \cdot (\cos(\theta_z))^{1,2} + 100)$
Observation rare	$0 < \text{GHI} < 1,2 \cdot I_0 \cdot (\cos(\theta_z))^{1,2} + 50$
Conditions sur le ratio $\frac{\text{DHI}}{\text{GHI}}$	$\begin{cases} \frac{\text{DHI}}{\text{GHI}} < 1,05 & \text{pour } \theta_z < 75^\circ, \text{ GHI} > 50 \text{ W} \cdot \text{m}^{-2} \\ \frac{\text{DHI}}{\text{GHI}} < 1,10 & \text{pour } 75^\circ < \theta_z < 93^\circ, \text{ GHI} > 50 \text{ W} \cdot \text{m}^{-2} \end{cases}$

Seules les données correspondant à des instants pour lesquels l'angle entre le Soleil et le zénith  $\theta_z$  est inférieur à  $90^\circ$  sont analysées (les données des nuits sont supprimées). En effet, le GHI est très facile à prévoir la nuit ; il est égal à 0. Les résultats de l'analyse de la qualité des différents jeux de données utilisés dans cette thèse sont présentés dans le [Tableau 1.3](#). Les données qui ne réussissent pas au moins un des tests susmentionnées sont considérées comme aberrantes. Le test croisé entre le GHI et le DHI n'a pas pu être réalisé sur les mesures de Villeneuve-de-la-Raho car nous ne disposons pas de mesure du DHI sur ce site.

TABLEAU 1.3 – Résultats de l'analyse de la qualité des mesures du GHI.

Site	Nombre de points de mesure	Données aberrantes	Données manquantes
Font-Romeu-Odeillo-Via (01/01/2018 – 31/12/2019)	52 769	Non	318 (0,6 %)
Perpignan (01/01/2016 – 31/12/2017)	52 524	Non	557 (1,06 %)
Villeneuve-de-la-Raho (01/08/2019 – 31/08/2019)	35 200	Non	Non
Villeneuve-de-la-Raho (01/09/2019 – 31/07/2020)	385 634	5,9 %	35,78 %

Le [Tableau 1.3](#) indique une très grande qualité des données des sites de Font-Romeu-Odeillo-Via et de Perpignan. On ne détecte aucune donnée aberrante sur ces deux sites. Comme la part de données manquantes dans les jeux de données issus de ces deux sites (0,6 % de l'ensemble des données de Font-Romeu-Odeillo-Via et 1,06 % de l'ensemble des données de Perpignan) est insignifiante et n'altère pas la représentativité de ces données, aucune méthode de traitement des données manquantes n'a été utilisée. Concernant le site de Villeneuve-de-la-Raho, on constate une bonne qualité des mesures réalisées au cours du mois d'août 2019 qui seront utilisées dans le [Chapitre 4](#) pour la prévision spatio-temporelle du GHI. Les données couvrant la période du 01/09/2019 au 31/07/2020 sont inutilisables : on note des données aberrantes (5,9 %) et une part très importante de données manquantes (35,78 %). Au-delà de la qualité des mesures, une caractérisation des conditions météorologiques des sites de mesure est importante. Par exemple, des modèles de prévision développés pour un site dont les journées « ciel clair » sont prédominantes auront de bonnes performances.

### 1.2.6.5 Variabilité des données

Cette section est consacrée à la caractérisation mathématique des conditions météorologiques des sites de mesure. Un coefficient de variabilité permettant d'estimer le degré de variation des mesures du



GHI est calculé pour chaque site. Il s'agit d'évaluer les variations que subissent les données au cours du temps. Dans le cas du GHI, la variabilité est directement liée aux conditions météorologiques du site de mesure. Après quantification de cette variabilité, il est possible de répondre à la question suivante : *à partir de la connaissance de la variabilité des données, quelle méthode de prévision choisir ?*

D'après l'étude de la variabilité des données d'éclairement solaire réalisée par [62] en testant divers paramètres statistiques issus de l'économétrie, il en ressort que le paramètre le plus adapté au GHI est le *mean absolute log return* (MALR) :

$$\text{MALR} = \frac{1}{n} \sum_{i=1}^{n-1} |\ln(k_c(i+1)) - \ln(k_c(i))| \quad (1.15)$$

où  $n$  est le nombre de données et  $k_c$  est l'indice ciel clair. Les coefficients de variabilité obtenus pour les jeux de données de Font-Romeu-Odeillo-Via et de Perpignan sont présentés dans le [Tableau 1.4](#).

TABLEAU 1.4 – Variabilité du GHI.

Site	MALR
Font-Romeu-Odeillo-Via	0,47
Perpignan	0,38

On remarque que la variabilité du GHI est plus marquée à Font-Romeu-Odeillo-Via qu'elle ne l'est à Perpignan. Les données de Villeneuve-de-la-Raho subissent les mêmes variations au cours du temps que celles de Perpignan car ces deux sites sont très proches. Cependant, la variabilité spatiale des mesures du site de Villeneuve-de-la-Raho sera caractérisée dans le [Chapitre 4](#).

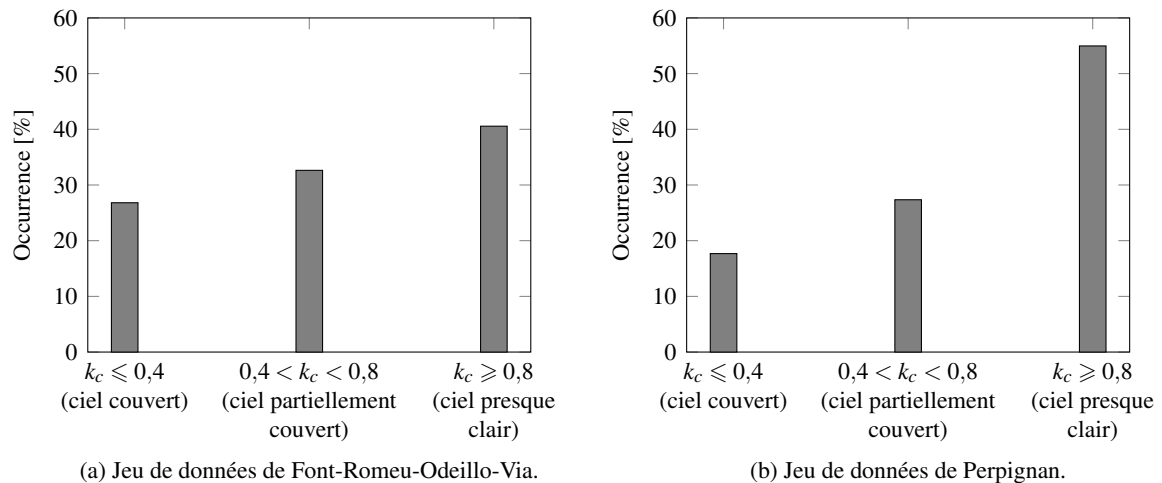


FIGURE 1.13 – Répartition des situations atmosphériques selon l'indice ciel clair à Font-Romeu-Odeillo-Via (01/01/2018 – 31/12/2019) et Perpignan (01/01/2016 – 31/12/2017).

Les résultats du coefficient MALR pour les sites de Font-Romeu-Odeillo-Via et Perpignan sont confirmés par la répartition des situations atmosphériques selon l'indice ciel clair présentée sur la [Figure 1.13](#). Une journée presque ciel clair est caractérisée par un indice ciel clair supérieur ou égal à 0,8, tandis qu'une journée complètement couverte est caractérisée par un indice ciel clair inférieur ou égal à 0,4. Une journée partiellement couverte est quant à elle caractérisée par un indice ciel clair compris entre 0,4 et 0,8. En examinant les jeux de données couvrant respectivement deux années issus de mesures réalisés à Font-Romeu-Odeillo-Via et Perpignan, on remarque une proportion de journées presque ciel clair plus importante à Perpignan qu'à Font-Romeu-Odeillo-Via ([Figure 1.13](#)).

### 1.3 État de l'art des modèles de prévision du GHI

La prévision de la ressource solaire est une étape importante pour relever les défis posés par la pénétration de la production PV dans les réseaux électriques de distribution. Lorsque l'on parle de prévision, il est important d'en définir les caractéristiques spatio-temporelles ; d'autant que, selon ces caractéristiques, certaines approches sont plus appropriées que d'autres.

L'horizon de prévision temporel correspond à la durée séparant le moment actuel du moment auquel la prévision est souhaitée. Dans le secteur énergétique, le « très court terme » fait référence à un horizon infra-horaire et le « court terme » à un horizon infra-journalier. Au-delà, il s'agit de prévisions journalières, saisonnières ou annuelles. Le pas de résolution temporel correspond à la période d'échantillonnage de la prévision. On peut par exemple faire une prévision allant jusqu'à 1 h (l'horizon) avec un pas de temps de 5, 10 ou 20 min (le pas de résolution).

L'horizon spatial est lié à la surface sur laquelle la prévision est faite. Le pas de résolution spatial correspond à la surface d'échantillonnage de la prévision. Même si la majorité des travaux existants dans la littérature scientifique traite de la prévision localisée, il est important, dans certains cas, de s'intéresser à la validité spatiale de cette prévision : une prévision peut être précise en un point d'une centrale solaire PV sans pour autant l'être sur tout le champ solaire. Ainsi, on peut par exemple faire une prévision sur une surface de  $10\text{km}^2$  (l'horizon) découpée en éléments de  $1\text{km}^2$  (le pas de résolution).

Dans le cas d'une prévision en temps réel, on peut définir une durée séparant deux mises à jour de la prévision, qui doit forcément être inférieure à l'horizon temporel choisi. Cette durée est indépendante de la résolution temporelle. En effet, il est possible de rafraîchir la prévision dès qu'une nouvelle mesure est obtenue, d'autant plus qu'une vitesse de rafraîchissement élevée va permettre de réduire l'erreur de prévision. Cependant, la vitesse de rafraîchissement de la prévision reste contrainte par le temps d'exécution du modèle de prévision. De manière générale, on cherchera donc à s'approcher au maximum du temps d'exécution du modèle de prévision afin de minimiser les erreurs de prévision tout en continuant d'opérer en temps réel.

Le choix d'une méthode de prévision est lié aux besoins du gestionnaire du réseau électrique de distribution, aux horizons de prévision envisagés et aux données disponibles. Ces trois aspects sont étroitement liés. En effet, les besoins du gestionnaire déterminent les horizons intéressants et donc les données à utiliser pour fournir les prévisions. On peut ainsi faire une classification des applications liées aux prévisions en fonction des horizons [63, 64].

- Les prévisions infra-horaires : elles couvrent des horizons de temps allant de quelques minutes à 1 h avec une granularité de quelques secondes à 10 min. Elles offrent la possibilité de prendre des décisions en lien avec la régulation et la distribution en temps réel.
- Les prévisions infra-journalières : elles couvrent les horizons de temps qui vont de 1 h à 6 h avec une granularité horaire. Elles sont utilisées dans le suivi des charges et dans l'actualisation du planning d'utilisation des moyens de production.
- Les prévisions à moyen terme : elles ont un horizon de prévision allant d'un jour à quelques jours avec une granularité horaire. Elles sont exploitées dans le choix des unités de production afin de minimiser les coûts liés aux temps de redémarrage et d'arrêt. Elles permettent aussi l'optimisation des moyens de production et de stockage.

La littérature scientifique regorge plusieurs méthodes de prévision du GHI et de la production PV qui peuvent être regroupées en deux grandes familles [65, 66] : les modèles de prévision localisée et les modèles de prévision spatio-temporelle. Cependant, commençons tout d'abord par la présentation des modèles de prévision les plus simples (mais néanmoins efficaces dans certaines conditions) : les modèles de persistance.

### 1.3.1 Modèles de persistance

Cette sous-section décrit les modèles de persistance classiques largement utilisés dans les études de prévision du GHI comme modèles de référence, à savoir la persistance sur le GHI et la persistance sur l'indice ciel clair.

#### 1.3.1.1 Persistance sur le GHI

La persistance sur le GHI est sans doute le modèle de prévision le plus simple. Elle consiste à supposer une invariance du GHI sur l'horizon de temps considéré. Le modèle de persistance sur le GHI s'exprime par :

$$\text{GHI}(t + \Delta t) = \text{GHI}(t) \quad (1.16)$$

où GHI est l'éclairement global horizontal,  $t$  correspond à l'instant de mesure et  $\Delta t$  correspond à l'horizon de prévision. Ce modèle ne nécessite aucune connaissance particulière de l'atmosphère et requiert uniquement un accès en temps réel à la mesure du GHI. Il est très efficace pour des horizons courts, ou lorsque la variabilité du GHI est faible sur l'horizon de prévision considéré.

#### 1.3.1.2 Persistance sur l'indice ciel clair

La persistance sur le GHI trouve ses limites lorsque la variabilité du signal est grande sur l'horizon de prévision considéré. Une part de cette variabilité provient du mouvement apparent du Soleil ; ce mouvement étant précisément connu, il est possible de corriger son effet en utilisant un modèle ciel clair. La persistance sur l'indice ciel clair est donnée par [67] :

$$k_c(t + \Delta t) = k_c(t) \quad (1.17)$$

où  $k_c$  est l'indice ciel clair,  $t$  est l'instant de mesure et  $\Delta t$  est l'horizon de prévision. En remplaçant l'indice ciel clair par son expression (équation (1.12)) dans l'équation (1.17), on obtient la prévision du GHI par la persistance sur l'indice ciel clair :

$$\text{GHI}(t + \Delta t) = \frac{\text{GHI}(t)}{\text{GHI}_{cs}(t)} \cdot \text{GHI}_{cs}(t + \Delta t) \quad (1.18)$$

Le modèle de persistance sur l'indice ciel clair – ce modèle est utilisé dans ce manuscrit comme modèle de référence – est généralement plus performant que la persistance sur le GHI. Il nécessite non seulement un accès en temps réel à la mesure du GHI, mais aussi un modèle ciel clair permettant de prévoir l'éclairement global horizontal en conditions ciel clair. Il faut donc être vigilant dans le choix du modèle ciel clair, car les erreurs commises sur la prévision de  $\text{GHI}_{cs}$  peuvent rapidement devenir compromettantes.

### 1.3.2 Modèles de prévision localisée du GHI

Diverses approches permettant d'effectuer la prévision localisée du GHI existent la littérature [65, 63, 68, 69, 70, 71, 72, 73] : les caméras à grand angle de champ, l'imagerie satellitaire et les modèles statistiques. La Figure 1.14 résume le type d'approche selon l'horizon spatio-temporel souhaité.

#### 1.3.2.1 Prévision par caméra à grand angle de champ

À une échelle infra-horaire, les caméras à grand angle de champ sont les systèmes les plus adaptés à la prévision de la ressource solaire [75, 76]. Grâce à leur haute résolution spatio-temporelle (Figure 1.14), elles permettent de caractériser finement la situation nuageuse au-dessus d'un site donné. Ces caméras offrent ainsi la possibilité d'anticiper les variations brutales de la ressource solaire. Elles peuvent être utilisées pour prévoir le DNI [77, 78, 79, 80, 81, 82, 83, 76] ou le GHI [84, 85, 86, 87, 88, 89, 90, 91, 92, 93].



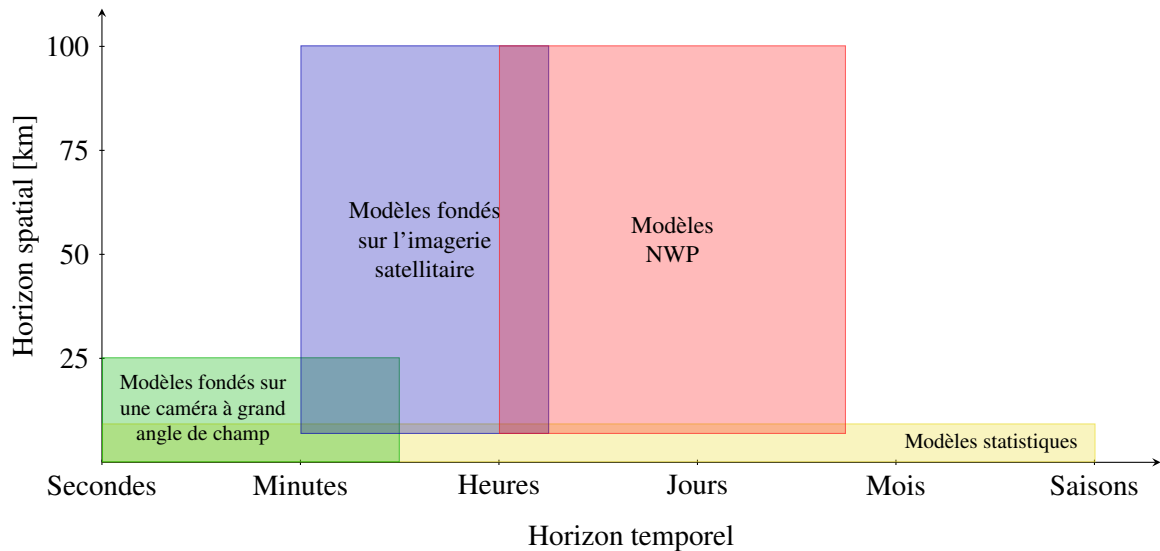


FIGURE 1.14 – Approches pour la prévision de la ressource solaire, en fonction de l'horizon spatio-temporel (adapté de [74]).

La prévision de la ressource solaire par caméra à grand angle de champ comporte deux parties : la prévision de l'éclairement par ciel clair et la prévision de l'indice ciel clair. À partir d'une séquence d'images du ciel, il est possible de prévoir l'évolution des perturbations atmosphériques et donc d'en déduire l'indice ciel clair. L'horizon de prévision est limité par l'angle de champ de la caméra, la vitesse des nuages et la position du Soleil sur les images. La prévision de l'indice ciel clair s'effectue en plusieurs étapes simples :

- segmentation nuages / ciel dégagé ;
- estimation de la vitesse et direction des nuages ;
- prévision de la position des nuages à l'horizon de prévision considéré.

Pourtant, la caméra à grand angle de champ reste peu utilisée dans la littérature pour la prévision de la ressource solaire. Si la première étape susmentionnée est assez bien maîtrisée [94, 95], les deux dernières le sont moins. La principale difficulté réside dans l'estimation et la prévision du mouvement des masses nuageuses. La procédure peut fournir de bonnes prévisions dans le cas de nuages approximativement à la même altitude et se déplaçant globalement par une simple translation. Cependant, les choses se compliquent lorsqu'il y a plusieurs couches de nuages, à différentes altitudes, prenant des directions et des vitesses différentes ou dans le cas de transformations non rigides de la couverture nuageuse.

Il existe plusieurs systèmes permettant d'obtenir des images du ciel tels que des solutions commerciales et des systèmes développés en laboratoire. Parmi les solutions commerciales existantes, le *Total Sky Imager* (TSI) développé par *Yankee Environmental Systems* est le plus connu [81, 86, 96, 78, 79]. On pourra consulter [97] et les références qui y sont citées pour un tour d'horizon des systèmes existants, ainsi que de leurs limitations. Pour pallier les limitations des systèmes existants et avoir un plein contrôle sur les réglages, certains laboratoires développent leur propre caméra adaptée à leurs besoins [36, 98, 87]. Le laboratoire PROMES-CNRS a aussi développé sa propre solution matérielle, le PSI (pour *PROMES-CNRS Sky Imager*). La génération actuelle, le PSI4, permet de générer une image HDR (pour *High Dynamic Range*) de 1,5 MP (1300 × 1216) toutes les 10 s (voir Figure 1.15).

### 1.3.2.2 Prévision par imagerie satellitaire

Pour des horizons infra-journaliers, l'imagerie satellitaire permet d'avoir de bonnes prévisions de la ressource solaire [75, 99, 100]. En effet, les images satellites offrent une vision globale du mouvement



FIGURE 1.15 – *PROMES sky imager* installé sur le toit du grand four solaire à Font-Romeu-Odeillo-Via et quelques images HDR acquises.

des masses nuageuses et permettent d'observer la couverture nuageuse avec une résolution temporelle élevée [101, 102]. De plus, elles sont collectées à différentes longueurs d'onde, permettant une mesure spectrale des interactions entre le rayonnement solaire et l'atmosphère. Ces mesures de luminance, couplées à des équations de transfert radiatif appliquées aux différentes couches de l'atmosphère, permettent d'évaluer la ressource solaire sur l'ensemble du globe [65, 103].

Une autre méthode permettant d'estimer la ressource solaire consiste à calibrer l'intensité des pixels des images satellites avec les mesures d'éclairement faites au sol en utilisant des modèles physiques [104, 105] ou statistiques [106, 107, 108]. Les informations sur le mouvement des nuages déduites en appliquant un algorithme de détermination des mouvements nuageux aux images satellites [109, 110] peuvent être extrapolées aux quelques heures à venir allant jusqu'à 6 h. Elles permettent d'anticiper les variations du rayonnement au cours des heures à venir [111]. La ressource solaire prédite est alors déduite de ces dernières [112, 113]. Pour effectuer la prévision, il est généralement supposé que la profondeur optique des nuages et leur vitesse de déplacement sont constantes, et que les masses nuageuses ne subissent que des déformations rigides.

### 1.3.2.3 Modèles numériques de prévision météorologique

La prévision à long terme du GHI (au-delà de l'échelle infra-journalière) se fait généralement à l'aide de modèles numériques de prévision météorologique (ou NWP, pour *numerical weather prediction*) [114, 115]. Une étude approfondie de ces modèles peut être trouvée dans [116]. Ces modèles permettent de faire une modélisation fine de l'atmosphère et ainsi prédire sur plusieurs jours l'évolution de facteurs météorologiques (température, vents, pressions, etc.). La performance d'un modèle NWP est fortement liée à la bonne connaissance de l'état initial de l'atmosphère et des lois physiques gouvernant son évolution [65]. Parmi les modèles NWP les plus connus, nous pouvons citer le Weather and Research Forecasting, le Global Forecast System, l'European Centre for Medium-Range Weather Forecast et le modèle North American Mesoscale [117, 65, 118]. Ces modèles diffèrent les uns des autres par leur résolution spatiale, leurs paramètres d'entrée et, surtout, les modèles physiques sous-jacents. Une fois l'évolution de l'atmosphère connue, un modèle de transfert radiatif permet d'en déduire l'évolution de la ressource solaire.

Ces modèles présentent plusieurs limitations. La principale concerne les temps de calcul et de mise à jour souvent très longs. En effet, ces modèles produisent des centaines de paramètres à chaque exécution. Malgré les avancées dans le domaine pour élaborer des modèles plus rapides en réduisant le nombre de paramètres estimés, les temps de calcul et de mise à jour restent des freins importants pour l'utilisation de ces modèles dans le cadre de la prévision à court terme. Une autre limitation est le faible pas de résolution spatiale des modèles NWP impliquant une précision grossière sur une

évaluation locale de la variabilité de la ressource solaire. Par conséquent, les sorties des modèles NWP sont souvent couplées à des observations d'images satellites ou de caméras hémisphériques [65]. Outre le couplage de ces différentes techniques, des post-traitements des sorties peuvent être effectués par filtre de Kalman ou à l'aide des modèles statistiques pour affiner les prévisions initialement obtenues [119, 120, 121].

#### 1.3.2.4 Modèles statistiques

Pour la prévision de la ressource solaire à des horizons infra-horaires et infra-journaliers, les modèles de prévision statistiques élaborés à partir d'un historique de mesures effectuées sur un site donné sont les plus utilisés. Ils regroupent les approches classiques de modélisation des séries temporelles, telles que le modèle de persistance et les modèles autorégressifs [122, 123, 124, 125, 126], et les outils fondés sur l'intelligence artificielle, telles que les arbres de régression [126], les  $k$  plus proches voisins (kNN) [127, 128, 129], les réseaux de neurones artificiels (ANN) [67, 130, 131, 127, 132, 133, 129], la régression par machines à vecteurs de support (SVR) [134, 67, 135, 131, 127, 136] et la régression par processus gaussien (GPR) [137, 138, 67, 131, 139].

Le GHI est non stationnaire, principalement en raison de la présence d'une périodicité journalière et saisonnière (si des données couvrant plusieurs années sont considérées). Étant donné que la plupart des approches classiques de séries temporelles sont bien adaptées aux signaux stationnaires ou quasi-stationnaires, plusieurs techniques de prétraitement des données du GHI existent, telles que l'utilisation de l'indice de clarté [65] et de l'indice ciel clair [76]. Grâce à leur capacité à modéliser des relations complexes entre les entrées et les sorties, les approches fondées sur l'intelligence artificielle, qui peuvent modéliser le GHI sans aucune étape de prétraitement spécifique, apparaissent comme la meilleure alternative aux approches classiques de séries temporelles. En conséquence, les outils d'apprentissage automatique sont devenus incontournables dans les études de prévision du GHI et de la production PV. De nombreux auteurs ont montré qu'ils surpassent les approches classiques [67, 140, 131, 137].

Bien que les méthodes fondées sur l'intelligence artificielle arrivent à apprendre le comportement stochastique de la couverture nuageuse au-dessus d'un site donné à travers l'historique de mesures, elles ne peuvent pas anticiper certains phénomènes dynamiques tels que le mouvement des nuages qui créent des changements soudains dans le GHI [75]. Cette limitation pourrait mettre en avant les méthodes de prévision par caméra hémisphérique ou par imagerie satellitaire. Cependant, ces modèles sont complexes et présentent une plusieurs limitations liées à leurs résolutions spatiale et temporelle, à l'incertitude des paramètres d'entrée et aux hypothèses simplificatrices des modèles physiques sous-jacents [75]. Par conséquent, l'utilisation des modèles statistiques pour la prévision de la ressource solaire est prépondérante dans la littérature.

Les modèles statistiques peuvent également être alimentés par des données d'entrée exogènes telles que les prévisions NWP pour améliorer les prévisions [141]. Ils peuvent aussi être combinés entre eux ou avec les autres approches de prévision afin d'élaborer un modèle plus performant [141, 142, 143]. Ces différentes combinaisons sont regroupées sous le nom de modèles « hybrides » dans la littérature. Pour un tour d'horizon des différentes combinaisons de modèles de prévision existantes dans la littérature, on pourra consulter [144].

### 1.3.3 Prévision spatio-temporelle du GHI

L'interpolation spatiale est un ensemble des techniques visant à estimer la valeur d'un phénomène physique donné en des sites non échantillonnés d'un domaine d'étude à partir des données connues aux sites d'observation. Pour la prévision dans le temps et dans l'espace du GHI, plusieurs approches existent dans la littérature. Les approches les plus utilisées pour la prévision de la ressource solaire sont : STARMA [145], STARIMA [145] et le krigeage [146]. Nous reviendrons plus en détails sur ces approches dans le [Chapitre 4](#).

## NOTE

Les modèles de prévision présentés dans ce manuscrit ont été développés en utilisant un serveur de calcul équipé d'un processeur Intel® Xeon® E7-4890 v2 @ 2,80 GHz. Plusieurs boîtes à outils ont été utilisés. Les modèles GPR ont été développés à l'aide de la boîte à outils Gaussian Processes for Machine Learning Matlab Toolbox [147], le modèle LS-SVR à l'aide de la LS-SVM Matlab Toolbox [148] et les modèles ANN à l'aide de la Deep Learning Matlab Toolbox. ◇

## 1.4 Critères de performance

La littérature scientifique regorge divers métriques pour évaluer la capacité de généralisation d'un modèle de prévision. Cette section présente les critères utilisés pour mesurer la précision des modèles de prévision développés dans ce manuscrit, à savoir la racine carrée de l'erreur quadratique moyenne normalisée (ou nRMSE, pour *normalized root mean square error*), l'erreur absolue moyenne dynamique (ou DMAE, pour *dynamic mean absolute error*), le score de compétence (ou SS, pour *skill score*) et le critère IS (*interval score*) évaluant la qualité des prévisions à partir des intervalles de confiances qui y sont associés.

### 1.4.1 La racine carrée de l'erreur quadratique moyenne normalisée

La racine carrée de l'erreur quadratique moyenne (ou RMSE, pour *root mean square error*) est la métrique couramment utilisée dans la littérature scientifique pour évaluer les performances d'un modèle de prévision, car elle permet d'avoir une bonne lisibilité ainsi qu'une bonne évaluation de la précision des modèles [149]. Afin d'avoir des résultats faciles à interpréter, le RMSE est souvent normalisé. Dans notre cas, nous avons choisi la version classique de normalisation du RMSE dans laquelle on divise le RMSE par la moyenne des données. Le nRMSE se calcule par :

$$\text{nRMSE} = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}}{\frac{1}{n} \sum_{i=1}^n y_i} \quad (1.19)$$

où  $\mathbf{y} \in \mathbb{R}^{n \times 1}$  est le vecteur de données de test et  $\hat{\mathbf{y}} \in \mathbb{R}^{n \times 1}$  est le vecteur de prévisions.

Le nRMSE pénalise les grands écarts entre les prévisions et les mesures sans quantifier les erreurs dues aux décalages temporels qui peuvent conduire à la mauvaise estimation des pics ou des rampes. Afin d'évaluer simultanément l'erreur temporelle et l'erreur en amplitude, l'erreur absolue moyenne dynamique (DMAE) [150] est aussi utilisée pour évaluer les performances des modèles de prévision développés.

### 1.4.2 L'erreur absolue moyenne dynamique

Le critère DMAE est fondé sur une méthodologie d'alignement de deux séries temporelles par transformation de l'axe temporel [151, 152]. L'alignement optimal est réalisé tout en minimisant l'erreur absolue moyenne normalisée (ou nMAE, pour *normalized mean absolute error*) entre la série temporelle alignée (les prévisions) et la série temporelle de référence (les observations). L'erreur absolue moyenne est la valeur absolue de l'écart moyen entre les mesures et les prévisions. Sa version normalisée, notée nMAE, se calcule par :

$$\text{nMAE} = \frac{\sum_{i=1}^n |\hat{y}_i - y_i|}{\sum_{i=1}^n y_i} \quad (1.20)$$

où  $\mathbf{y} \in \mathbb{R}^{n \times 1}$  est le vecteur des observations et  $\hat{\mathbf{y}} \in \mathbb{R}^{n \times 1}$  est le vecteur des prévisions.

La déformation de l'axe temporel nécessaire pour aligner les deux séries est mesurée grâce à l'indice de distorsion temporelle (ou TDI, pour *temporal distortion index*) [151]. On pourra consulter [150, 151, 152, 153] et les références qui y sont citées pour plus de détails sur l'algorithme de calcul de l'indice de distorsion temporelle. Le TDI fait partie des nouvelles métriques introduites dans [153] pour évaluer les performances des modèles de prévision de la ressource solaire et palier les limitations des métriques usuelles tels que le RMSE et le MAE.

Le critère DMAE évalue simultanément l'erreur temporelle et l'erreur en amplitude entre les prévisions et les observations. Ce critère est un compromis entre l'indice de distorsion temporelle et l'erreur absolue moyenne. Il se calcule par :

$$\text{DMAE} = \int_0^c \text{nMAE}(t_{di}) f(t_{di}; \lambda, c) dt_{di} \quad (c, \lambda) > 0 \quad (1.21)$$

où  $\text{nMAE}(t_{di})$  est l'erreur absolue moyenne normalisée entre les prévisions et les mesures après alignement de ces deux séries temporelles avec une distorsion temporelle égale à  $t_{di}$ ,  $f(t_{di}; \lambda, c)$  est la densité conditionnelle d'une loi exponentielle de paramètre  $\lambda$  et  $c$  est la distorsion temporelle maximale autorisée. La densité conditionnelle est donnée par :

$$f(t_{di}; \lambda, c) = \begin{cases} \frac{\lambda \exp(-\lambda t_{di})}{1 - \exp(-\lambda c)} & \text{si } t_{di} \leq c \\ 0 & \text{si } t_{di} \geq c \end{cases} \quad (\lambda, c) > 0 \quad (1.22)$$

En d'autres termes, le DMAE considère les erreurs absolues moyennes entre les séries de prévisions alignées et les mesures pour toutes les valeurs de TDI comprises entre 0 et  $c$ . Le critère DMAE inclut également la densité conditionnelle d'une loi exponentielle de paramètre  $\lambda$  déterminant la région des valeurs de DMAE où les erreurs associées seront plus pénalisées. Dans [150], pour une évaluation juste des modèles de prévision dans le domaine des énergies renouvelables, les auteurs recommandent une valeur maximale de 10 % pour le paramètre  $c$  et une valeur de 0,1 pour  $\lambda$ . Ici, nous avons choisi  $\lambda = 0,1$  et  $c = 5\%$ .

### 1.4.3 Le skill score

La précision des prévisions de la ressource solaire varie non seulement selon l'horizon de temps, mais aussi selon l'endroit considéré [72]. Ainsi, le nRMSE d'un modèle de prévision dépend fortement des conditions atmosphériques de l'endroit considéré. Pour démontrer la qualité des performances d'un modèle de prévision, son nRMSE doit être comparé au nRMSE d'un modèle de référence (ici, la persistance sur l'indice ciel clair est utilisée comme référence). Ce modèle de référence quantifie la prévisibilité du GHI sur un site particulier. Le *skill score* [154], noté SS, est le critère généralement utilisé pour comparer les performances d'un modèle par rapport à celles d'un modèle de référence :

$$\text{SS} = \left( 1 - \frac{\text{nRMSE}_{\text{Modèle}}}{\text{nRMSE}_{\text{Référence}}} \right) \cdot 100 \quad (1.23)$$

où  $\text{nRMSE}_{\text{Modèle}}$  est le nRMSE du modèle de prévision et  $\text{nRMSE}_{\text{Référence}}$  est le nRMSE de la persistance sur l'indice ciel clair. Un score négatif indique que le modèle développé est moins performant que la persistance sur l'indice ciel clair, tandis qu'un score positif démontre que le modèle développé est meilleur que la persistance sur l'indice ciel clair.

### 1.4.4 L'interval score

Pour évaluer la qualité des prévisions à partir des intervalles de confiance qui y sont associés, nous utilisons le score propre IS [155] défini par :

$$\text{IS} = \frac{1}{n} \sum_{i=1}^n \left( (U_i - L_i) + \frac{2}{\alpha} \max \{L_i - y_i, y_i - U_i, 0\} \right) \quad (1.24)$$



où  $\mathbf{y} \in \mathbb{R}^{n \times 1}$  est le vecteur des observations,  $\mathbf{L} \in \mathbb{R}^{n \times 1}$  est le vecteur des bornes inférieures des intervalles de confiance,  $\mathbf{U} \in \mathbb{R}^{n \times 1}$  est le vecteur des bornes supérieures des intervalles de confiance et  $\alpha = 1 - \mu$ , avec  $\mu$  la valeur nominale associée aux intervalles de confiance. Ici,  $\mu = 0,95$ , donc  $\alpha = 0,05$ . Ce critère privilégie les intervalles de confiance étroits et pénalise (avec un terme de pénalité dépendant de  $\alpha$ ) les prévisions pour lesquelles l'observation est en dehors des intervalles de confiance.

## 1.5 Conclusion

Au cours de ce premier chapitre, le contexte dans lequel s'inscrivent les travaux réalisés dans ce manuscrit a été présenté. Les objectifs et les résultats du projet « Smart Occitania »— le projet traite du développement d'un démonstrateur de réseau électrique intelligent en milieu périurbain/rural en région Occitanie — ont également été présentés. Cette thèse s'inscrit dans le cadre de ce projet. Les travaux entrepris par PROMES-CNRS ont pour objectif le développement d'une stratégie de gestion prévisionnelle d'un réseau électrique de distribution basse tension. Ce chapitre a permis de mieux saisir la nécessaire mutation du réseau électrique de distribution qui fait face au déploiement d'une production décentralisée, en particulier du solaire photovoltaïque.

On retiendra que la prévision du GHI à des fins de prévision de la production PV est un élément clé du développement d'une stratégie de gestion prévisionnelle du réseau électrique de distribution. Un état de l'art des modèles de prévision du GHI a été présenté à la [Section 1.3](#). Contrairement aux modèles numériques de prévision météorologique et aux modèles fondés sur une caméra à grand angle de champ ou sur l'imagerie satellitaire, les modèles de prévision statistiques fondés sur l'analyse d'une base de données et sur l'apprentissage automatique ont l'avantage de pouvoir fournir des prévisions à divers horizons de temps et sont sans doute les plus utilisés pour la prévision du GHI. Nous avons également vu que le GHI peut se décomposer en un éclaircissement par ciel clair et un indice ciel clair. Cette décomposition est le point de départ de certains modèles de prévision. Cependant, les outils de l'apprentissage automatique peuvent prévoir directement le GHI, c'est-à-dire sans passer par l'étape de décomposition du GHI où un modèle ciel clair est utilisé pour prévoir l'éclaircissement par ciel clair. Ainsi, en utilisant ces outils, on s'affranchit de l'utilisation d'un modèle ciel clair dont le choix requiert une certaine vigilance, car les erreurs commises sur la prévision de l'éclaircissement par ciel clair peuvent rapidement devenir préjudiciables. Notre choix de modèle se porte donc sur les modèles fondés sur l'apprentissage automatique.

Le [Chapitre 2](#) fait une étude comparative des modèles de prévision fondés sur l'apprentissage automatique afin d'identifier la méthode la plus adaptée à la prévision du GHI. Le [Chapitre 3](#) fait le choix de l'entrée adéquate pour les modèles fondés sur la régression non paramétrique par processus gaussien dédiés à la prévision du GHI. Le [Chapitre 4](#) traite de la prévision spatio-temporelle du GHI. Il s'agit d'un premier pas vers la prévision, dans le temps et l'espace, du GHI où l'on cherche à identifier un noyau spatial à associer au noyau identifié au [Chapitre 3](#) pour la prévision dans le temps et dans l'espace du GHI par modèles fondés sur la régression non paramétrique par processus gaussien.



## Chapitre 2

# Étude comparative des modèles fondés sur l'apprentissage automatique pour la prévision multi-horizon de l'éclairement global horizontal

### 2.1 Introduction

Grâce à l'état de l'art des modèles de prévision du GHI proposé dans le chapitre précédent, on comprend que les modèles de prévision statistiques, fondés sur l'analyse d'une base de données et l'apprentissage automatique, sont très utilisés pour la prévision localisée du GHI. Bien que ces modèles ne puissent pas prévoir avec succès l'occurrence de phénomènes atmosphériques (tels que le passage d'un nuage devant le Soleil, par exemple), ils permettent de réaliser des prévisions à divers horizons de temps (voir la [Figure 1.14](#) à la [Sous-section 1.3.2](#) du [Chapitre 1](#)).

Parmi les outils de l'apprentissage automatique utilisés pour la prévision du GHI, les plus populaires sont : les réseaux de neurones artificiels (ou ANN, pour *artificial neural networks*), les machines à vecteurs de support (ou SVM, pour *support vector machines*) et les processus gaussiens (ou GP pour *Gaussian processes*). Ils ont été largement étudiés dans la littérature [[67](#), [131](#), [137](#), [140](#)] et ont prouvé leur capacité à fournir de très bonnes prévisions. Plusieurs études comparatives existent dans la littérature [[67](#), [140](#)], mais les auteurs ont souvent du mal à départager ces outils et à donner des recommandations claires quant à celui utiliser pour obtenir les meilleures prévisions. En effet, concernant la prévision localisée du GHI par des outils de l'apprentissage automatique, plusieurs points de divergence existent encore.

- Dans la plupart des études, les auteurs utilisent des données échantillonnées à 1 h ou plus ([Tableau 2.1](#)). Ce pas de temps entraîne une simplification importante de la dynamique du GHI, comme illustré à la [Figure 2.1](#), ce qui facilite les prévisions.
- Les méthodes ne sont pas toujours bien optimisées et calibrées au GHI. Par exemple, en utilisant la méthode de régression par processus gaussien, le noyau isotrope par défaut (c'est-à-dire le noyau SE, pour *squared exponential*) est généralement le plus utilisé [[71](#), [67](#)]. Cependant, un noyau particulièrement adapté à l'application en question a une influence significative sur les résultats [[137](#)].
- Concernant les variables d'entrée, dans certaines études, les auteurs ne considèrent que des données endogènes, tandis que d'autres utilisent en plus des variables exogènes, ce qui ne permet pas de faire des comparaisons justes et équitables entre les méthodes.
- Certains auteurs prévoient directement le GHI, d'autres l'indice ciel clair (en utilisant des modèles ciel clair comme étapes de pré- et post-traitement) ou même la production photovoltaïque.



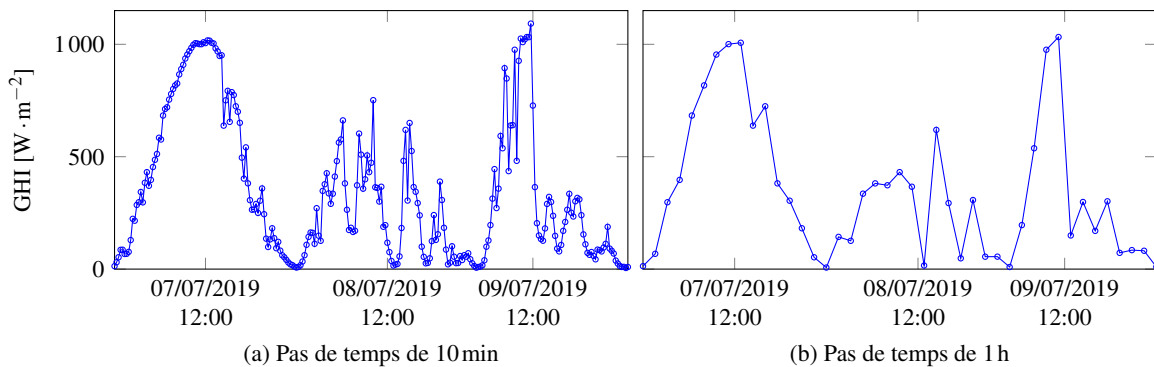


FIGURE 2.1 – Différence entre les dynamiques du GHI échantillonné à 10 min et à 1 h. Les mesures ont été réalisées au laboratoire PROMES-CNRS, situé à Font-Romeu-Odeillo-Via, en région Occitanie.

Par conséquent, bien que la prévision du GHI soit au cœur de plusieurs études dans la littérature scientifique, une étude comparative approfondie des modèles fondés sur les outils de l'apprentissage automatique utilisant uniquement des données endogènes, sans aucune étape de prétraitement, pour la prévision localisée du GHI à des horizons infra-horaires et infra-journaliers est, à notre connaissance, inexistante. En effet, une question reste toujours sans réponse : *quel outil de l'apprentissage automatique pour la prévision multi-horizon du GHI ?* Pour répondre à cette question, une étude comparative impliquant trois outils (GPR, SVR et ANN) a été menée afin de choisir l'outil adapté à notre application. En effet, la stratégie de contrôle prédictif développée dans le cadre du projet « Smart Occitania » peut être mise en œuvre en temps réel, avec un pas de temps de 10 min. La prise de décision est fondée sur la connaissance du comportement du système et sur des prévisions de grandeurs tels que la production photovoltaïque influant sur la dynamique du système tels à horizons de temps allant de 10 min à 24 h à pas de temps de 10 min (144 horizons de temps au total).

Les modèles sont développés en utilisant un jeu de données couvrant 2018 et 2019 issu de mesures réalisées sur le site de Font-Romeu-Odeillo-Via (voir la [Sous-sous-section 1.2.6.1](#) du [Chapitre 1](#) pour la description de ces données). Les données de l'année 2018 sont utilisées pour l'apprentissage des modèles et celles de 2019 pour le test. Les modèles sont évalués sur des données couvrant une année afin de mettre en évidence leur robustesse face à toutes les conditions atmosphériques. Afin de faciliter l'apprentissage des modèles, les données sont normalisées en utilisant la méthode de normalisation min-max.

NOTE

La normalisation min-max d'une variable  $X$  s'effectue par :

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \quad (2.1)$$

où  $X'$  est la variable normalisée,  $X_{min}$  est le minimum et  $X_{max}$  est le maximum.  $\diamond$

La [Section 2.2](#) définit la prévision multi-horizon et expose les différentes stratégies de développement de modèles de prévision multi-horizons. La [Section 2.3](#) décrit les modèles fondés sur les trois outils de l'apprentissage automatique développés. La [Section 2.4](#) présente les résultats de prévision.

## 2.2 Prévision multi-horizon

Dans notre application, le gestionnaire du réseau électrique de distribution a besoin de prévisions du GHI à des horizons de temps allant de 10 min à 24 h, avec un pas de temps de 10 min. Les modèles développés doivent donc être capable de fournir des prévisions multi-horizons. La prévision multi-horizon consiste à prévoir plusieurs observations futures d'une série temporelle à partir de la

TABLEAU 2.1 – Récentes études comparatives des outils de l'apprentissage automatique.

Auteurs	Méthodes de prévision	Pas de temps	Horizons de prévision	Variables d'entrée	Variables prédites	Critères de performance	Jeu de données
Lubbe et al. (2020) [138]	GPR	1 h	1 h	GHI, DNI, DHI, UVA, UVB, température ambiante, pression barométrique, humidité relative, vitesse du vent, direction du vent	GHI	RMSE	Données couvrant une période de 14 j ; entraînement sur 9 j et test sur 5 j
Tolba et al. (2020) [137]	Persistance, GPR	30 min	30 min à 48 h	Temps	GHI	nRMSE	Deux jeux de données, chacun couvrant une période de 45 j ; entraînement sur 30 j et test sur 15 j
Sharifzadeh et al. (2019) [140]	ANN, SVR, GPR	1 h	1 h à 6 h	Production PV, température, DHI, DNI	Production PV	MSE	De 1985 à 2014
Lauret et al. (2015) [67]	Persistance, persistance sur l'indice ciel clair, AR, MLP, SVR, GPR	1 h	1 h à 6 h	Indice ciel clair	Indice ciel clair	nRMSE, nMAE, nMBE	Données provenant de 3 sites ; entraînement sur 1 an et test sur 1 an

connaissance d'un historique d'observations. Plusieurs stratégies de développement de modèles de prévision multi-horizon existent dans la littérature.

### 2.2.1 Stratégie itérative

Elle consiste à élaborer un modèle de prévision à un pas et à itérer les prévisions de ce modèle jusqu'à l'obtention de la prévision à l'horizon de temps considéré [156]. Tout d'abord, le modèle fournit une prévision à un pas. Ensuite, cette valeur prévue est introduite dans les données d'entrée pour produire la prévision au pas suivant et ainsi de suite jusqu'à l'horizon de temps souhaité. L'avantage de cette stratégie est qu'on a un seul modèle à entraîner. Cependant, cette stratégie souffre, surtout à long terme, d'une accumulation d'erreurs en itérant les prévisions qui affecte la capacité de généralisation du modèle [156, 157].

### 2.2.2 Stratégie directe

Elle consiste à prévoir indépendamment les valeurs de la grandeur étudiée aux horizons de temps considérés [156, 157]. Il est donc nécessaire de développer plusieurs modèles mono-horizons. Cette stratégie permet d'éviter l'accumulation d'erreurs constatée dans la stratégie itérative, mais le développement de modèles mono-horizons peut être astreignant en termes de temps de calcul lorsque de nombreux horizons de temps sont considérés. De plus, étant donné que les modèles sont entraînés de manière indépendante, la stratégie directe ne permet pas de prendre en compte les dépendances pouvant exister entre les prévisions, ce qui peut être préjudiciable à la précision des prévisions.

### 2.2.3 Stratégie directe-itérative

Cette stratégie est une combinaison des stratégies directe et itérative. Des modèles mono-horizons sont développés et, à chaque horizon de temps considéré, le modèle donnant la prévision à cet horizon de temps intègre dans ses données d'entrée la prévision du modèle mono-horizon dédié à l'horizon de temps précédent [156]. Cette stratégie permet de surmonter les limitations des stratégies directe et itérative.

Outre ces trois stratégies à sortie unique, on peut aussi développer des modèles de prévision à plusieurs sorties (chaque sortie correspondant à chaque horizon de temps considéré) [156, 157]. Cela permet d'éviter l'hypothèse d'indépendance faite par la stratégie directe ainsi que l'accumulation d'erreurs dont souffre la stratégie itérative. Toutefois, un tel modèle est complexe à élaborer. Une combinaison de la stratégie directe et de celle à plusieurs sorties peut aussi être trouvée dans la littérature [156].

Dans les études de prévision du GHI, l'utilisation de la stratégie directe est très courante [67, 140]. Par ailleurs, dans certaines études, la période d'échantillonnage des données est généralement modifiée pour correspondre à chaque horizon de temps considéré, ce qui peut entraîner une simplification significative de la dynamique du GHI [131]. Cependant, dans notre cas où le coût calculatoire des algorithmes doit être maîtrisé et des horizons de temps allant de 10 min à 24h, avec un pas de temps de 10 min, sont nécessaires, développer des modèles de prévision mono-horizons reviendrait à élaborer 144 modèles, ce qui est exigeant en termes de temps calcul. Par conséquent, pour le développement des modèles de prévision multi-horizons dans ce chapitre, la stratégie itérative est adoptée. D'après l'étude comparative des différentes stratégies de développement de modèles de prévision multi-horizons menée dans [156], cette stratégie s'avère être la plus rapide et la plus simple, surtout dans le cas où de nombreux horizons de temps sont considérés. Pour notre étude comparative, les modèles de prévision à un pas (ici à  $\Delta t = 10$  min) fondés sur les outils GPR, SVR et ANN sont développés et les prévisions sont itérées jusqu'à un horizon de 24h.

## 2.3 Description des modèles développés

Cette section présente les trois différents outils de l'apprentissage automatique utilisés pour la prévision du GHI, à savoir : la régression par processus gaussien, la régression par machines à vecteurs de support et la régression par réseaux de neurones artificiels. Pour la description de ces outils, nous considérons un ensemble de données d'apprentissage  $\mathcal{D} = \{(\mathbf{x}_i, y_i), 1 \leq i \leq n\}$ , où  $\mathbf{x}_i \in \mathbb{R}^D$  est un vecteur d'observations passées du GHI,  $y_i \in \mathbb{R}$  est la valeur du GHI au pas de temps suivant, correspondant donc à la prévision à un pas du GHI, et  $n$  est le nombre d'échantillons. Nous regroupons tous les vecteurs d'entrée  $\mathbf{x}_i$  dans une matrice  $\mathbf{X} \in \mathbb{R}^{n \times D}$  et dans un vecteur  $\mathbf{y} \in \mathbb{R}^n$  toutes les sorties correspondantes  $y_i$ .

### 2.3.1 Les processus gaussiens

#### 2.3.1.1 Définition

Un processus gaussien (GP) est un processus aléatoire dont toutes les lois fini-dimensionnelles sont gaussiennes. Pour  $f$  un processus gaussien, on écrira  $f(\mathbf{x}) \sim \mathcal{GP}(\boldsymbol{\mu}(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$ , où  $\mathbf{x}$  et  $\mathbf{x}'$  sont des variables d'entrée,  $\boldsymbol{\mu}(\mathbf{x}) = \mathbb{E}[f(\mathbf{x})]$  est la fonction moyenne et  $k(\mathbf{x}, \mathbf{x}') = \mathbb{E}[(f(\mathbf{x}) - \boldsymbol{\mu}(\mathbf{x}))(f(\mathbf{x}') - \boldsymbol{\mu}(\mathbf{x}'))^\top]$  est la fonction de covariance, également appelée noyau. Un GP est complètement défini par sa fonction moyenne et sa fonction de covariance. Pour des raisons de simplicité et puisque le noyau suffit généralement à encoder les connaissances *a priori* sur la fonction à apprendre, la fonction moyenne est souvent considérée nulle.

#### 2.3.1.2 La régression par processus gaussien

Considérons le modèle de régression standard :

$$y = f(\mathbf{x}) + \varepsilon \quad (2.2)$$

où  $\mathbf{x}$  est une variable d'entrée,  $y$  une variable de sortie,  $f$  est la fonction de régression et  $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$  est un bruit gaussien indépendant et identiquement distribué.

La méthode de régression par processus gaussien est une approche bayésienne non paramétrique [147] permettant de résoudre des problèmes de régression. Elle consiste à construire une distribution gaussienne *a priori* sur les fonctions de régression à partir des données d'apprentissage  $\mathcal{D}$ .

NOTE

*L'objectif d'un modèle de régression est de modéliser la relation existante entre les données d'entrée et celles de sortie à valeurs réelles. La régression non paramétrique est une forme de régression dans laquelle la fonction d'estimation ne prend pas de forme prédéterminée, mais est construite à partir des informations provenant des données. Les méthodes paramétriques font l'hypothèse que le modèle peut être caractérisé par un vecteur de paramètres de dimension finie. Les modèles non paramétriques permettent d'éviter les choix souvent restrictifs des modèles paramétriques en définissant une distribution a priori sur des espaces de fonctions de dimension infinie.*  $\diamond$

Considérons  $\mathbf{X}_*$  une matrice de vecteurs d'entrée des données de test. En supposant que  $f(\mathbf{x}) \sim \mathcal{GP}(\boldsymbol{\mu}(\mathbf{x}), k(\mathbf{x}, \mathbf{x}'))$  et  $\varepsilon \sim \mathcal{N}(0, \sigma_\varepsilon^2)$ , le vecteur d'observations  $\mathbf{y}$  et les images des entrées de l'ensemble des données de test  $\mathbf{X}_*$  par la fonction  $f_*$  suivent une distribution gaussienne conjointe. Par conséquent, la densité prédictive postérieure est également gaussienne [147] :

$$\mathbf{f}_* | \mathbf{X}, \mathbf{y}, \mathbf{X}_* \sim \mathcal{N}(\boldsymbol{\mu}_*, \boldsymbol{\sigma}_*^2) \quad (2.3)$$

où la moyenne prédite  $\boldsymbol{\mu}_*$  et la variance prédictive  $\boldsymbol{\sigma}_*^2$  sont respectivement données par :

$$\boldsymbol{\mu}_* = \boldsymbol{\mu}(\mathbf{X}_*) + \mathbf{K}_*^\top (\mathbf{K} + \sigma_\varepsilon^2 \mathbf{I})^{-1} (\mathbf{y} - \boldsymbol{\mu}(\mathbf{X})) \quad (2.4)$$

$$\boldsymbol{\sigma}_*^2 = \mathbf{K}_{**} - \mathbf{K}_*^\top (\mathbf{K} + \sigma_\varepsilon^2 \mathbf{I})^{-1} \mathbf{K}_* \quad (2.5)$$

avec  $\mathbf{f}_* = f(\mathbf{X}_*)$ ,  $\mathbf{K}_* = k(\mathbf{X}, \mathbf{X}_*)$ ,  $\mathbf{K}_{**} = k(\mathbf{X}_*, \mathbf{X}_*)$ , et  $\mathbf{K} = k(\mathbf{X}, \mathbf{X})$ . La moyenne indique la valeur prédite et la variance permet de déterminer l'intervalle de confiance autour de la valeur prédite.

Voyons ce que donne la moyenne prédictive dans le cas d'un seul vecteur d'entrée de test  $\mathbf{x}_*$ . La moyenne prédite (équation (2.4)) devient :

$$\boldsymbol{\mu}_* = \boldsymbol{\mu}(\mathbf{x}_*) + \mathbf{k}_*^\top (\mathbf{K} + \sigma_\varepsilon^2 \mathbf{I})^{-1} (\mathbf{y} - \boldsymbol{\mu}(\mathbf{X})) \quad (2.6)$$

où  $\mathbf{k}_* = [k(\mathbf{x}_1, \mathbf{x}_*) \quad k(\mathbf{x}_2, \mathbf{x}_*) \quad \dots \quad k(\mathbf{x}_n, \mathbf{x}_*)]^\top$ . Ainsi, dans ce cas, la moyenne prédite  $\boldsymbol{\mu}_*$  (équation 2.6) peut s'écrire comme une combinaison linéaire de fonctions noyau, chacune centrée sur un point d'apprentissage :

$$\boldsymbol{\mu}_* = \boldsymbol{\mu}(\mathbf{x}_*) + \sum_{i=1}^n \alpha_i k(\mathbf{x}_i, \mathbf{x}_*) \quad (2.7)$$

où  $\boldsymbol{\alpha} = (\mathbf{K} + \sigma_\varepsilon^2 \mathbf{I})^{-1} (\mathbf{y} - \boldsymbol{\mu}(\mathbf{X}))$ . Les coefficients  $\alpha_i$ , également appelés paramètres, sont mis à jour chaque fois qu'une nouvelle observation est ajoutée à l'ensemble d'observations, contrairement aux paramètres du noyau, appelés hyperparamètres, qui ne changent pas une fois l'apprentissage terminé.

La régression par processus gaussien fournit naturellement une distribution prédictive, à partir de laquelle les intervalles de confiance à 95 % sont obtenus :

$$\text{IC} = \boldsymbol{\mu}_* \pm 1,96 \sqrt{\sigma_*^2} \quad (2.8)$$

où IC sont les limites de l'intervalle de confiance,  $\boldsymbol{\mu}_*$  est la moyenne prédictive et  $\sigma_*^2$  est la variance prédictive.

### 2.3.1.3 Les fonctions de covariance

Un noyau formule les hypothèses relatives à la fonction sous-jacente à apprendre et quantifie la similarité entre les valeurs de la fonction pour différentes variables d'entrée  $\mathbf{x}$  et  $\mathbf{x}'$  [147, 158]. Ces hypothèses peuvent être, par exemple, à quel point la fonction est régulière ou si elle est périodique. Le choix du noyau a un impact crucial sur la performance d'un modèle GPR [137]. Toute fonction peut être utilisée comme noyau tant que la matrice de covariance résultante est semi-définie positive.

**2.3.1.3.1 Noyaux de base** Pour bien apercevoir l'impact d'un noyau sur la performance d'un modèle GPR, examinons les structures de distributions *a priori* encodées par les noyaux de base. Pour des variables d'entrée  $x, x' \in \mathbb{R}$ , quelques noyaux de base couramment utilisés sont décrits ci-après. Pour une liste exhaustive des noyaux, on pourra consulter [147].

— Le noyau bruit blanc (ou WN, pour *white noise*), noté  $k_{\text{WN}}$ , est donné par :

$$k_{\text{WN}}(x, x') = \sigma^2 \delta(x - x') \quad (2.9)$$

où  $\sigma$  est l'amplitude. Lors de la modélisation par GPR, un bruit blanc additif gaussien – l'estimation de la variance du bruit est faite à partir des données – est ajouté au noyau (équation (2.2)).

— Le noyau linéaire (ou Lin, pour *linear*), noté  $k_{\text{Lin}}$ , est donné par :

$$k_{\text{Lin}}(x, x') = \sigma^2 (x - \ell)(x' - \ell) \quad (2.10)$$

où  $\sigma$  est l'amplitude et  $\ell$  le paramètre de corrélation. Ce noyau permet de modéliser des fonctions linéaires.

— Le noyau exponentiel quadratique (ou SE, pour *squared exponential*), noté  $k_{\text{SE}}$ , est donné par :

$$k_{\text{SE}}(x, x') = \sigma^2 \exp\left(-\frac{(x - x')^2}{2\ell^2}\right) \quad (2.11)$$

où  $\sigma$  est l'amplitude et  $\ell$  est le paramètre de corrélation. Ce noyau permet de modéliser des fonctions régulières. C'est la fonction de covariance de base couramment choisie par défaut, car il est parcimonieux et ses paramètres à estimer sont faciles à interpréter. De plus, il peut être considéré comme un noyau universel, capable d'apprendre le comportement de n'importe quelle fonction continue avec suffisamment de données, sous certaines conditions qui sont étudiées dans [159].

- Le noyau quadratique rationnel (ou RQ, pour *rational quadratic*), noté  $k_{\text{RQ}}$ , est donné par :

$$k_{\text{RQ}}(x, x') = \sigma^2 \left( 1 + \frac{(x - x')^2}{2\alpha\ell^2} \right)^{-\alpha} \quad (2.12)$$

où  $\sigma$  est l'amplitude et  $\ell$  est le paramètre de corrélation et  $\alpha$  est le paramètre de pondération des variations à grande et à petite échelles. Ce noyau peut être considéré comme une somme infinie de noyaux SE avec différents paramètres de corrélation tirés d'une distribution bêta [147]. Il permet de modéliser des fonctions variant de façon régulière sur plusieurs échelles.

- Les noyaux de la classe Matérn sont donnés par :

$$k_{\text{M}_\nu}(x, x') = \sigma^2 \frac{1}{2^{\nu-1}\Gamma(\nu)} \left( \sqrt{2\nu} \frac{|x - x'|}{\ell} \right)^\nu \cdot \mathcal{B}_\nu \left( \sqrt{2\nu} \frac{|x - x'|}{\ell} \right) \quad (2.13)$$

où  $\sigma$  est l'amplitude,  $\ell$  est le paramètre de corrélation,  $\Gamma$  est la fonction Gamma et  $\mathcal{B}_\nu$  est la fonction de Bessel modifiée de seconde espèce d'ordre  $\nu$ . Le paramètre  $\nu$  contrôle le degré de régularité du processus gaussien résultant. Pour  $\nu = \frac{1}{2}$ , le noyau exponentiel (ou E, pour *exponential*), noté  $k_{\text{E}}$ , est donné par :

$$k_{\text{E}}(x, x') = \sigma^2 \exp\left(-\frac{|x - x'|}{\ell}\right) \quad (2.14)$$

où  $\sigma$  est l'amplitude et  $\ell$  est le paramètre de corrélation. C'est le noyau le moins régulier de la classe Matérn. Deux autres noyaux de base de la classe Matérn souvent utilisés correspondant respectivement à  $\nu = \frac{3}{2}$  et  $\nu = \frac{5}{2}$ , respectivement notés  $k_{\text{M}_{3/2}}$  et  $k_{\text{M}_{5/2}}$ , sont donnés par :

$$k_{\text{M}_{3/2}}(x, x') = \sigma^2 \left( 1 + \sqrt{3} \frac{|x - x'|}{\ell} \right) \cdot \exp\left(-\sqrt{3} \frac{|x - x'|}{\ell}\right) \quad (2.15)$$

$$k_{\text{M}_{5/2}}(x, x') = \sigma^2 \left( 1 + \sqrt{5} \frac{|x - x'|}{\ell} + \sqrt{5} \frac{(x - x')^2}{3\ell^2} \right) \cdot \exp\left(-\sqrt{5} \frac{|x - x'|}{\ell}\right) \quad (2.16)$$

où  $\sigma$  est l'amplitude et  $\ell$  est le paramètre de corrélation. Ces noyaux sont définis du moins régulier au plus régulier.

- Le noyau périodique (ou Per, pour *periodic*), noté  $k_{\text{Per}}$ , est donné par :

$$k_{\text{Per}}(x, x') = \sigma^2 \exp\left(-\frac{2 \sin^2\left(\frac{\pi(x - x')}{P}\right)}{\ell^2}\right) \quad (2.17)$$

où  $\sigma$  est l'amplitude,  $\ell$  est le paramètre de corrélation et  $P$  est la période. Ce noyau permet de modéliser des fonctions ayant des structures répétitives.

La Figure 2.2 présente une représentation graphique des noyaux de base usuels  $k_{\text{Lin}}$ ,  $k_{\text{SE}}$ ,  $k_{\text{Per}}$ ,  $k_{\text{E}}$ ,  $k_{\text{M}_{3/2}}$  et  $k_{\text{M}_{5/2}}$ , de leurs hyperparamètres et des exemples de réalisations. Le noyau  $k_{\text{SE}}$  est représenté avec différents paramètres pour montrer l'influence des hyperparamètres sur la forme de la fonction à apprendre. Les noyaux de base peuvent être combinés par addition ou multiplication afin de créer des fonctions de covariances adaptées à l'application en question. Des exemples de fonctions pouvant

être modélisées par le noyau quasi-périodique  $k_{\text{Per} \times \text{RQ}}$  – ce noyau est utilisé dans le [Chapitre 3](#) pour modéliser le GHI – sont également présentés sur la [Figure 2.2](#). Le noyau  $k_{\text{Per} \times \text{RQ}}$  est donné par :

$$k_{\text{Per} \times \text{RQ}} = \sigma^2 \exp \left( -\frac{2 \sin^2 \left( \frac{\pi(x-x')}{P} \right)}{\ell_1^2} \right) \cdot \left( 1 + \frac{(x-x')^2}{2\alpha\ell_2^2} \right)^{-\alpha} \quad (2.18)$$

où  $\sigma$  est l'amplitude,  $\ell_1$  et  $\ell_2$  sont les paramètres de corrélation,  $P$  est la période et  $\alpha$  est le paramètre de pondération des variations multi-échelles.

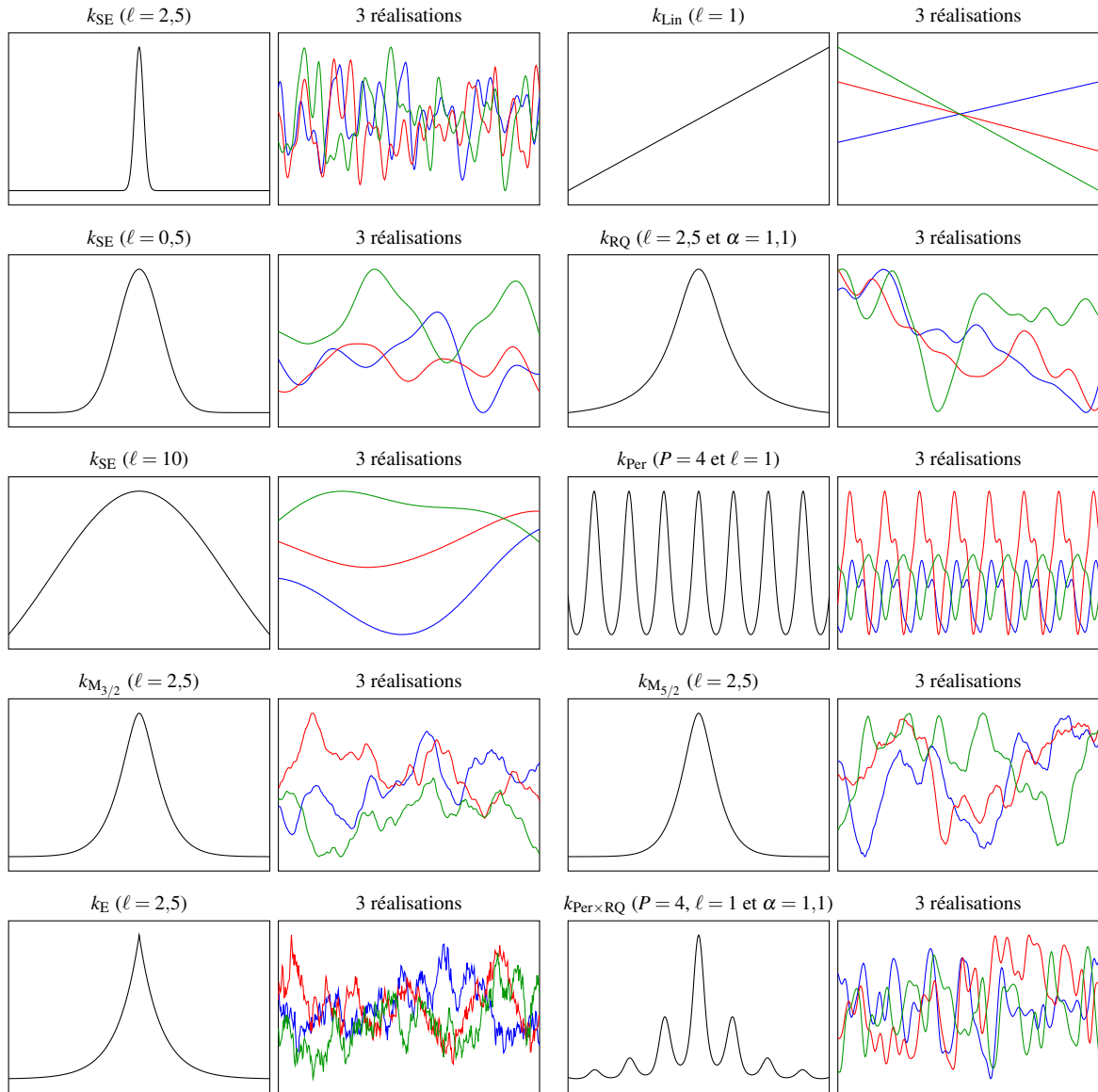


FIGURE 2.2 – Représentation graphique des noyaux de base usuels et d'un noyau combiné avec les valeurs de leurs hyperparamètres et exemples de réalisations.  $k_{\text{SE}}$  est le noyau exponentiel quadratique,  $k_{\text{Lin}}$  est le noyau linéaire,  $k_{\text{RQ}}$  est le noyau quadratique rationnel,  $k_{\text{Per}}$  est le noyau périodique,  $k_{\text{E}}$ ,  $k_{\text{M}_{3/2}}$  et  $k_{\text{M}_{5/2}}$  sont des noyaux de la classe Matérn d'ordre  $\frac{1}{2}$ ,  $\frac{3}{2}$  et  $\frac{5}{2}$  respectivement.  $k_{\text{Per} \times \text{RQ}}$  est le noyau combiné résultant du produit d'un noyau périodique et d'un noyau quadratique rationnel.  $P$  est la période,  $\ell$  est le paramètre de corrélation et  $\alpha$  est le paramètre de pondération des variations à grande et à petite échelles.



**2.3.1.3.2 Noyaux utilisés** Concernant la prévision du GHI par GPR, les auteurs choisissent généralement des noyaux isotropes [71, 67]. Cependant, en utilisant un historique d'observations comme entrées du modèle, la fonction à apprendre fait face à une variable d'entrée multidimensionnelle qui ne varie pas forcément de manière isotrope suivant chaque dimension du vecteur d'entrée. Il serait alors préférable d'utiliser des noyaux qui définissent un paramètre de corrélation spécifique et approprié à chaque dimension du vecteur d'entrée. C'est pourquoi les noyaux avec détermination implicite automatique du niveau de pertinence de chaque dimension du vecteur d'entrée (ou ARD, pour *automatic relevance determination*) peuvent être utilisés pour modéliser des fonctions dont l'entrée est multidimensionnelle. De plus, des structures additives peuvent être développées pour les noyaux ARD afin de modéliser les corrélations entre les dimensions d'entrée et ainsi améliorer la capacité de généralisation des modèles, surtout pour les horizons à long terme [160]. Afin de choisir les noyaux ARD adaptés à notre étude, plusieurs noyaux ont été testés et les plus performants ont été retenus. Ces derniers sont décrits ci-après.

- Le noyau exponentiel quadratique avec ARD, noté  $k_{SE-ARD}$ , est exprimé comme un produit de noyaux SE avec différents paramètres de corrélation :

$$k_{SE-ARD}(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{1}{2} \sum_{d=1}^D \frac{(x_d - x'_d)^2}{\ell_d^2}\right) \quad (2.19)$$

où  $\sigma$  est l'amplitude et  $\ell_1, \ell_2, \dots, \ell_D > 0$  sont les paramètres de corrélation qui contrôlent la variation de la fonction suivant chaque dimension du vecteur d'entrée. Ce noyau est une fonction de covariance à entrée multidimensionnelle couramment choisie par défaut, car il est parcimonieux et ses paramètres à estimer sont faciles à interpréter.

- Le noyau quadratique rationnel avec ARD, noté  $k_{RQ-ARD}$ , est donné par :

$$k_{RQ-ARD}(\mathbf{x}, \mathbf{x}') = \sigma^2 \left(1 + \frac{1}{2\alpha} \sum_{d=1}^D \frac{(x_d - x'_d)^2}{\ell_d^2}\right)^{-\alpha} \quad (2.20)$$

où  $\alpha > 0$  est un paramètre de pondération des variations à grande et à petite échelles. Le noyau RQ peut être considéré comme une somme infinie de noyaux SE avec différents paramètres de corrélation [147] et modélise les fonctions qui varient de façon régulière sur plusieurs échelles. De manière analogue, le noyau RQ-ARD, qui permet de modéliser des fonctions présentant des variations multi-échelles suivant chaque dimension du vecteur d'entrée, peut être considéré comme une somme infinie de noyaux SE-ARD.

Les Figures 2.3 et 2.4 présentent le principe de la prévision du GHI par les modèles GPR fondés sur les noyaux  $k_{SE-ARD}$  et  $k_{RQ-ARD}$ .

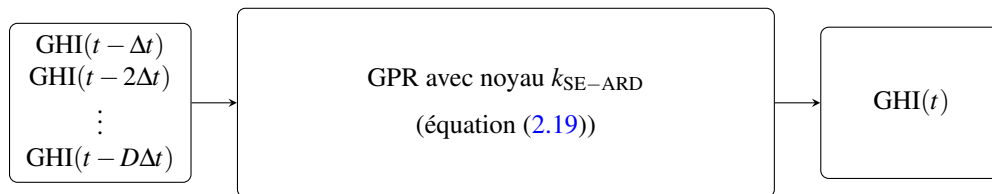


FIGURE 2.3 – Modèle de prévision GPR fondé sur le noyau  $k_{SE-ARD}$ , où  $[\text{GHI}(t - \Delta t), \text{GHI}(t - 2\Delta t), \dots, \text{GHI}(t - D\Delta t)]^T$  est un vecteur d'observations et  $\Delta t = 10$  min est le pas de temps.

### 2.3.1.4 Algorithme d'apprentissage

Les hyperparamètres  $\theta$  d'un modèle GPR, qui regroupent les paramètres impliqués dans le noyau et la variance du bruit, doivent être déduits des données d'entraînement. En pratique, ils sont généralement



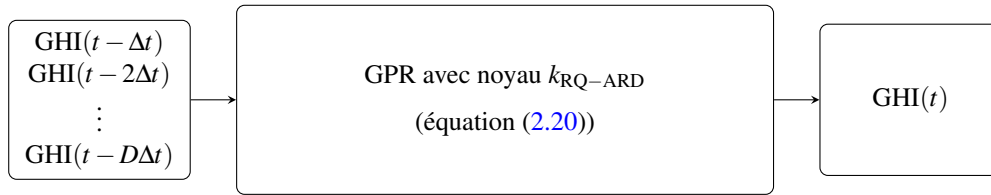


FIGURE 2.4 – Modèle de prévision GPR fondé sur le noyau  $k_{\text{RQ-ARD}}$ , où  $[\text{GHI}(t - \Delta t), \text{GHI}(t - 2\Delta t), \dots, \text{GHI}(t - D\Delta t)]^T$  est un vecteur d'observations et  $\Delta t = 10 \text{ min}$  est le pas de temps.

estimés en maximisant le logarithme de la vraisemblance marginale donné par :

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{1}{2} \left( \mathbf{y}^T (\mathbf{K} + \sigma_\varepsilon^2 \mathbf{I})^{-1} \mathbf{y} + \log(\det(\mathbf{K} + \sigma_\varepsilon^2 \mathbf{I})) + n \log(2\pi) \right) \quad (2.21)$$

Cette fonction  $\mathcal{L}(\boldsymbol{\theta})$  est généralement non convexe par rapport aux hyperparamètres. L'espace du problème peut avoir de nombreux optima locaux et les hyperparamètres optimisés, qui dépendent de leur initialisation, peuvent ne pas être l'optimum global [147]. Une approche classique pour pallier ce problème est d'utiliser divers points de départ choisis aléatoirement à partir d'une distribution préalable appropriée. Une étude traitant de l'impact des distributions *a priori* des hyperparamètres initiaux sur les performances des modèles GPR a été menée dans [161] et les auteurs sont arrivés à la conclusion que des hypothèses simples telles que la distribution uniforme dans un intervalle de valeurs approprié peuvent être suffisants dans l'optique d'une modélisation par GPR en termes de prévisibilité.

Pour l'initialisation des hyperparamètres  $\boldsymbol{\theta}$ , les choix suivants ont été faits : l'hyperparamètre  $\sigma^2$  a été initialisé à la variance des données d'apprentissage et les autres hyperparamètres sont tirés aléatoirement d'une distribution uniforme  $\theta_i \sim \mathcal{U}(0, 1)$ .

### 2.3.2 Les machines à vecteurs de support

Les machines à vecteurs de support (ou SVM, pour *support vector machines*), développées au cours des dernières décennies [162, 163, 164, 165, 166], sont des systèmes d'apprentissage initialement définis pour la discrimination (prévision d'une variable qualitative binaire). Elles sont basées sur la recherche de l'hyperplan de marge optimale qui, lorsque c'est possible, classe ou sépare correctement les données tout en étant le plus éloigné possible de toutes les observations. Le principe est donc de trouver un classifieur, ou une fonction de discrimination, dont la capacité de généralisation est la plus grande possible. L'algorithme SVM, qui est une généralisation non linéaire de la méthode développée dans [167], est basé sur un algorithme d'apprentissage issu de la théorie de l'optimisation et ancré dans le cadre de l'apprentissage statistique [162, 168]. Les SVM, développées initialement pour la reconnaissance de formes [165], présentent également de bonnes performances pour la modélisation des séries temporelles et ont été appliquées avec succès aux problèmes de régression et de prévision des séries temporelles [164, 166].

Le principe de base des SVM est de ramener le problème de discrimination à celui, linéaire, de la recherche d'un hyperplan optimal. Deux astuces permettent d'atteindre cet objectif :

- la première consiste à définir l'hyperplan comme solution d'un problème d'optimisation sous contraintes dont la fonction objectif ne s'exprime qu'à l'aide de produits scalaires entre vecteurs et dans lequel le nombre de contraintes « actives » ou « vecteurs supports » contrôle la complexité du modèle ;
- le passage à la recherche de surfaces séparatrices non linéaires est obtenu par l'introduction d'une fonction de covariance ou noyau induisant implicitement une transformation non linéaire des données vers un espace intermédiaire de plus grande dimension. D'où l'appellation couramment rencontrée de machine à noyau (ou *kernel machine*).

### 2.3.2.1 Régression par machine à vecteurs de support

La régression par machine à vecteurs de support (SVR) ramène la matrice d'entrée  $\mathbf{X}$  vers un espace intermédiaire de plus grande dimension  $F$  par le biais d'une fonction non linéaire  $\phi$  afin d'effectuer une régression linéaire dans cet espace. La fonction de régression linéaire  $f$  dans  $F$  est donnée par :

$$f(\mathbf{x}) = \mathbf{w}^\top \phi(\mathbf{x}) + b \quad (2.22)$$

où le poids  $\mathbf{w}$  et le biais  $b$  sont obtenus en résolvant un problème d'optimisation convexe.

Dans [162], le but de la régression par machines à vecteurs de support est de trouver une fonction  $f$  qui soit aussi plate (hyperplan) que possible et qui présente une erreur maximale  $\varepsilon$  par rapport aux sorties  $y_i$  pour toutes les entrées des données d'apprentissage. Ce problème peut être formulé comme un problème d'optimisation convexe [166]. En outre, des variables muettes  $\xi_i, \xi_i^*$  sont introduites pour relaxer le problème d'optimisation [162].

Les équations (2.23) et (2.24) décrivent le problème d'optimisation convexe comme suit :

$$\min_{\mathbf{w}, \xi^{(*)}} \left( \frac{1}{2} \mathbf{w}^\top \mathbf{w} + C \sum_{i=1}^n (\xi_i + \xi_i^*) \right) \quad (2.23)$$

où  $C > 0$  détermine la marge, soumis aux contraintes suivantes :

$$\forall i, \begin{cases} y_i - (\mathbf{w}^\top \phi(\mathbf{x}_i) + b) \leq \varepsilon + \xi_i \\ (\mathbf{w}^\top \phi(\mathbf{x}_i) + b) - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \quad (2.24)$$

Il a été démontré dans [166] que le problème d'optimisation peut être résolu plus facilement dans sa formulation duale en utilisant des multiplicateurs de Lagrange. La forme duale de ce problème est :

$$\min_{\alpha^{(*)}} \left( \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) k(\mathbf{x}_i, \mathbf{x}_j) + \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) - \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) \right) \quad (2.25)$$

soumis à :

$$\forall i, \begin{cases} 0 \leq \alpha_i, \alpha_i^* \leq C \\ \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \end{cases} \quad (2.26)$$

où  $\alpha_i, \alpha_i^*$  sont des multiplicateurs de Lagrange et  $k$  est une fonction noyau, définie comme suit :

$$k(\mathbf{x}, \mathbf{x}') = \phi^\top(\mathbf{x}) \phi(\mathbf{x}') \quad (2.27)$$

Les poids  $\mathbf{w}$  sont définis par :

$$\mathbf{w} = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \phi(\mathbf{x}_i) \quad (2.28)$$

Ainsi, la fonction de régression est donnée par :

$$f(\mathbf{X}) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) k(\mathbf{x}_i, \mathbf{X}) + b \quad (2.29)$$

Les algorithmes d'apprentissage doivent résoudre le problème d'optimisation (équations (2.23) et (2.24)), qui devient plus complexe et nécessite beaucoup de temps de calcul lorsqu'un grand ensemble de données est considéré. Afin de réduire la complexité algorithmique, une méthode équivalente de régression par machines à vecteur de support couplée aux moindres carrés (ou LS-SVR, pour *Least Squares Support Vector Regression*) a été introduite [169]. Dans cette méthode, les contraintes d'inégalité sont remplacées par des contraintes d'égalité, ce qui réduit considérablement le temps de calcul.

### 2.3.2.2 Modèle LS-SVR

La fonction de perte  $J$  est définie par :

$$J(\mathbf{w}, \mathbf{e}) = \frac{1}{2} \mathbf{w}^\top \mathbf{w} + \frac{1}{2} \gamma \sum_{i=1}^n e_i^2 \quad (2.30)$$

où  $\gamma$  est une constante et  $e_i$  est l'erreur. Le problème d'optimisation peut être décrit par :

$$\min_{\mathbf{w}, \mathbf{e}} J(\mathbf{w}, \mathbf{e}) \quad (2.31)$$

où  $\gamma$  est soumis à :

$$\forall i, \quad y_i = \mathbf{w}^\top \phi(\mathbf{x}_i) + b + e_i \quad (2.32)$$

où  $\phi$  est la fonction non linéaire dans  $F$  et  $b$  est le biais.

Le lagrangien du problème d'optimisation est donné par :

$$L = J(\mathbf{w}, \mathbf{e}) - \sum_{i=1}^n \alpha_i [\mathbf{w}^\top \phi(\mathbf{x}_i) + b + e_i - y_i] \quad (2.33)$$

où  $\alpha_i$  est un multiplicateur de Lagrange. Les dérivées partielles de  $L$  sont données par :

$$\begin{cases} \frac{\partial L}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^n \alpha_i \phi(\mathbf{x}_i) \\ \frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^n \alpha_i = 0 \\ \frac{\partial L}{\partial e_i} = 0 \Rightarrow \alpha_i = \gamma e_i \\ \frac{\partial L}{\partial \alpha_i} = 0 \Rightarrow \mathbf{w}^\top \phi(\mathbf{x}_i) + b + e_i - y_i = 0 \end{cases} \quad (2.34)$$

En éliminant  $\mathbf{w}$  et  $e_i$  de l'équation (2.34) nous pouvons formuler le problème sous forme matricielle :

$$\begin{bmatrix} 0 & \mathbb{1}_n^\top \\ \mathbb{1}_n & \mathbf{K} + \frac{1}{\gamma} \mathbf{I} \end{bmatrix} \begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix} \quad (2.35)$$

où :

$$\mathbf{K}(i, j) = k(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^\top \phi(\mathbf{x}_j) \quad (2.36)$$

Donc, en utilisant l'équation (2.34), la fonction de régression du modèle LS-SVR est :

$$\mathbf{y} = f(\mathbf{X}) = \sum_{i=1}^n \alpha_i k(\mathbf{X}, \mathbf{x}_i) + b \quad (2.37)$$

### 2.3.2.3 Fonction de covariance

Parmi les noyaux qui peuvent être utilisés dans la régression par machines à vecteurs de support (noyaux linéaire, polynomial et à fonction de base radiale), nous avons opté pour le noyau à fonction de base radiale (ou RBF, pour *radial basis function*) pour modéliser et prévoir le GHI. Ce noyau, plus flexible que les noyaux linéaire et polynomial, est capable de représenter des relations complexes entre les données et est adapté au GHI. Le noyau RBF est donné par :

$$k_{\text{RBF}}(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\delta^2}\right) \quad (2.38)$$

où  $\delta$  est le paramètre de corrélation. Ici,  $\delta$  et  $\gamma$  sont estimés en utilisant la méthode de recherche par grille *gridsearch*. Le principe de la prévision du GHI par modèle LS-SVR est présenté sur la [Figure 2.5](#).

Pour le LS-SVR, la méthode développée dans [170] pour la construction d'intervalles de confiance est utilisée pour obtenir les intervalles de confiance.

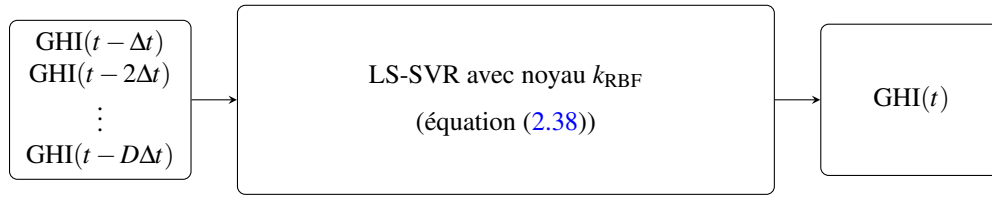


FIGURE 2.5 – Modèle de prévision LS-SVR fondé sur le noyau  $k_{\text{RBF}}$ , où  $[\text{GHI}(t - \Delta t), \text{GHI}(t - 2\Delta t), \dots, \text{GHI}(t - D\Delta t)]^T$  est un vecteur d'observations et  $\Delta t = 10 \text{ min}$  est le pas de temps.

### 2.3.3 Les réseaux de neurones artificiels

Les réseaux de neurones artificiels (ou ANN, pour *artificial neural networks*) sont des approches connexionnistes inspirées du fonctionnement des neurones biologiques, largement utilisées pour la prévision de séries temporelles [171], en particulier pour la prévision de la ressource solaire [172, 133, 67, 132]. En effet, les ANN sont des outils puissants permettant, notamment, de résoudre des problèmes de régression. Un réseau de neurones artificiels est entièrement défini par son architecture (ou topologie), ses paramètres (poids synaptiques et biais) et l'algorithme d'apprentissage utilisé.

Si les réseaux de neurones à propagation avant ont été les plus utilisés au cours des dernières décennies pour la prévision des séries temporelles, les réseaux de neurones récurrents sont aujourd'hui largement utilisés en raison de leur capacité à apprendre les dépendances temporelles entre des données [173]. Les réseaux de neurones convolutifs (ou CNN, pour *convolutional neural networks*), souvent utilisés pour la classification d'images et la résolution de problèmes de vision par ordinateur, ont la capacité d'apprendre et d'extraire automatiquement des caractéristiques de données d'entrée brutes [173]. Ainsi, ils peuvent également être utilisés pour la prévision de séries temporelles. Ces réseaux peuvent être combinés pour construire des modèles exploitant leurs caractéristiques ; il est par exemple possible d'associer réseaux CNN et LSTM pour la prévision d'une série temporelle [174].

Dans cette étude comparative, un perceptron multicouche (ou MLP, pour *multilayer perceptron*), c'est-à-dire un réseau de neurones à propagation avant, et un réseau de neurones LSTM sont utilisés pour prévoir le GHI. Les deux réseaux sont décrits dans la suite. Les réseaux de neurones MLP et LSTM ont été largement utilisés pour la prévision des séries temporelles et ont démontré de bonnes performances [175, 171]. Étant donné que notre objectif n'est pas de faire une étude exhaustive sur les ANN, les modèles combinant différents types de réseaux ne sont pas considérés ici.

#### 2.3.3.1 Le perceptron multicouche

Le perceptron multicouche est un réseau de neurones artificiels à propagation avant, historiquement utilisé pour l'approximation de fonctions et la prévision des séries temporelles. Dans un MLP avec  $\ell$  couches cachées, il n'y a pas de rétroaction et l'information circule via les connexions, dans une seule direction, de la couche d'entrée jusqu'à la couche de sortie. Pour la couche d'entrée,  $h^0(\mathbf{x}) = \mathbf{x}$ , où  $\mathbf{x}$  est un vecteur. Pour les couches cachées du réseau, avec  $L = 1, \dots, \ell$  :

$$a^L(\mathbf{x}) = \mathbf{b}^L + \mathbf{W}^L h^{L-1}(\mathbf{x}) \quad (2.39)$$

et :

$$h^L(\mathbf{x}) = \phi(a^L(\mathbf{x})) \quad (2.40)$$

où  $\mathbf{b}^L$  est le vecteur des biais de la couche  $L$ ,  $\mathbf{W}^L$  est la matrice des poids de la couche  $L$  et  $\phi$  est la fonction d'activation.

La fonction sigmoïde est historiquement la fonction d'activation la plus utilisée pour les neurones des couches cachées car elle est différentiable et permet de conserver des valeurs dans l'intervalle  $[0, 1]$ . Cependant, l'utilisation de cette fonction est problématique car son gradient est très proche de 0 lorsque

$|\mathbf{x}|$  n'est pas proche de 0. Étant donné que cela peut causer des problèmes lors de l'apprentissage, en particulier avec les architectures profondes, la fonction sigmoïde a été supplantée par l'unité linéaire rectifiée (ou ReLU). Comme caractéristique intéressante, la fonction ReLU, qui n'est pas différentiable en 0, a un effet de sparsification. Elle est utilisée dans notre réseau. Ainsi :

$$\phi(a^L(\mathbf{x})) = \max(0, a^L(\mathbf{x})) \quad (2.41)$$

Pour la couche de sortie, avec  $L = \ell + 1$  :

$$a^{\ell+1}(\mathbf{x}) = \mathbf{b}^{\ell+1} + \mathbf{W}^{\ell+1}h^\ell(\mathbf{x}) \quad (2.42)$$

et :

$$h^{\ell+1}(\mathbf{x}) = \psi(a^{\ell+1}(\mathbf{x})) \quad (2.43)$$

où  $\mathbf{b}^{\ell+1}$  est le vecteur des biais de la couche  $\ell + 1$ ,  $\mathbf{W}^{\ell+1}$  est la matrice des poids de la couche  $\ell + 1$  et  $\psi$  est la fonction d'activation. L'architecture du modèle MLP utilisé pour la prévision du GHI est présentée sur la [Figure 2.6](#).

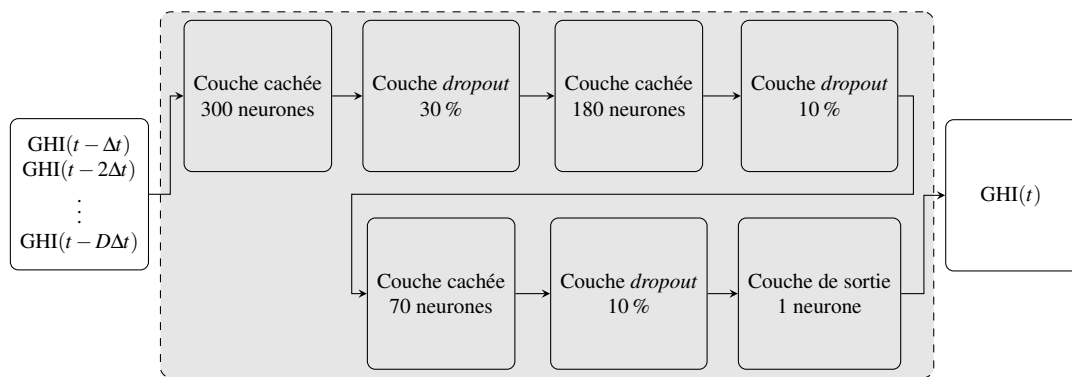


FIGURE 2.6 – Modèle de prévision MLP : la topologie du réseau est définie comme une combinaison de couches cachées et de couches *dropout*. La couche de sortie fournit la prévision.  $[\text{GHI}(t - \Delta t), \text{GHI}(t - 2\Delta t), \dots, \text{GHI}(t - D\Delta t)]^\top$  est un vecteur d'observations et  $\Delta t = 10$  min est le pas de temps.

### 2.3.3.2 Le réseau LSTM

Les réseaux LSTM (pour *long short-term memory*), introduits par [176], sont des réseaux de neurones récurrents développés pour surmonter le problème de disparition du gradient qui se produit lors de l'entraînement des réseaux de neurones récurrents traditionnels [177, 178]. Les réseaux LSTM sont capables d'apprendre des séquences d'observations. Cela fait d'eux des réseaux de neurones adaptés à la prévision des séries temporelles. Initialement, une cellule LSTM était composée d'une cellule mémoire, d'une porte d'entrée et d'une porte de sortie. Plus tard, la porte d'oubli a été introduite pour permettre au LSTM de réinitialiser son propre état [179]. Désignons par  $i$ ,  $f$ ,  $g$  et  $o$  la porte d'entrée, la porte d'oubli, le nœud d'entrée et la porte de sortie respectivement. Ces portes et nœuds sont respectivement définis par :

$$\mathbf{g}^k = \phi(\mathbf{W}^{gx}\mathbf{x}^k + \mathbf{W}^{gh}\mathbf{h}^{k-1} + \mathbf{b}_g) \quad (2.44)$$

$$\mathbf{i}^k = \sigma(\mathbf{W}^{ix}\mathbf{x}^k + \mathbf{W}^{ih}\mathbf{h}^{k-1} + \mathbf{b}_i) \quad (2.45)$$

$$\mathbf{f}^k = \sigma(\mathbf{W}^{fx}\mathbf{x}^k + \mathbf{W}^{fh}\mathbf{h}^{k-1} + \mathbf{b}_f) \quad (2.46)$$

où  $\mathbf{x}^k$  est le vecteur d'entrée,  $\mathbf{h}^{k-1}$  est l'état caché au pas de temps précédent,  $\mathbf{W}^{gx}$ ,  $\mathbf{W}^{ix}$  et  $\mathbf{W}^{fx}$  sont les poids d'entrée,  $\mathbf{W}^{gh}$ ,  $\mathbf{W}^{ih}$  et  $\mathbf{W}^{fh}$  sont les poids récurrents,  $\mathbf{b}_g$ ,  $\mathbf{b}_i$ ,  $\mathbf{b}_f$  sont les biais,  $\sigma$  est la fonction sigmoïde et  $\phi$  est la fonction tangente hyperbolique.

Au cœur de chaque cellule de mémoire, il y a un nœud à activation linéaire :

$$\mathbf{s}^k = \mathbf{g}^k \odot \mathbf{i}^k + \mathbf{s}^{k-1} \odot \mathbf{f}^k \quad (2.47)$$

où  $\odot$  est le produit d'Hadamard.

La porte de sortie est décrite par :

$$\mathbf{o}^k = \sigma(\mathbf{W}^{ox} \mathbf{x}^k + \mathbf{W}^{oh} \mathbf{h}^{k-1} + \mathbf{b}_o) \quad (2.48)$$

et :

$$\mathbf{h}^k = \phi(\mathbf{s}^k) \odot \mathbf{o}^k \quad (2.49)$$

L'architecture du modèle LSTM utilisé pour la prévision du GHI est présentée sur la [Figure 2.7](#).

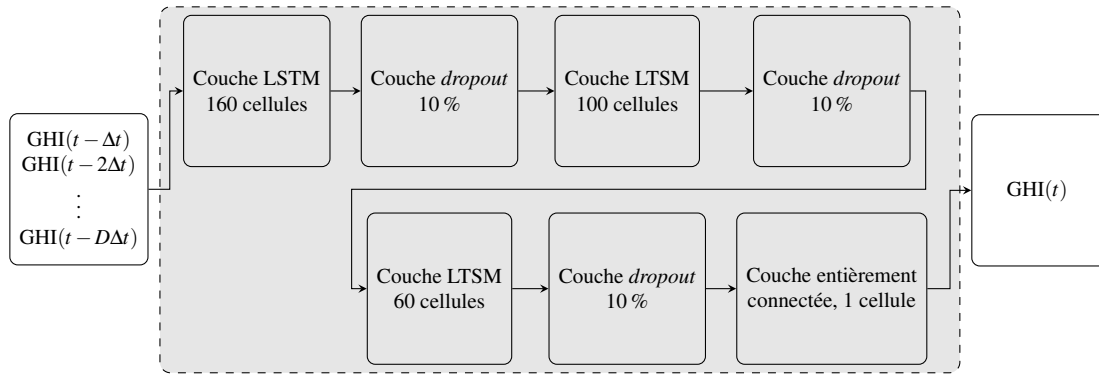


FIGURE 2.7 – Modèle de prévision LSTM : la topologie du réseau est définie comme une combinaison de couches LSTM et de couches *dropout*. La couche entièrement connectée fournit la prévision.  $[\text{GHI}(t - \Delta t), \text{GHI}(t - 2\Delta t), \dots, \text{GHI}(t - D\Delta t)]^T$  est le vecteur d'observations et  $\Delta t = 10$  min est le pas de temps.

En ce qui concerne les modèles ANN, la méthode *bootstrap* [180, 181] qui permet d'entraîner des réseaux de neurones avec différentes initialisations de paramètres afin d'estimer l'incertitude du modèle est utilisée pour quantifier l'incertitude associée aux prévisions (30 répétitions ont été réalisées pour les modèles MLP et LSTM).

### 2.3.3.3 Algorithme d'apprentissage

Adam (pour *adaptive moment estimation*), un algorithme d'optimisation à taux d'apprentissage adaptatif bien connu [182], a été utilisé pour entraîner les réseaux de neurones MLP et LSTM. Adam effectue une optimisation de premier ordre basée sur le gradient de fonctions objectifs stochastiques, à partir des estimations adaptatives des moments d'ordre inférieur. Il est efficace d'un point de vue calculatoire, nécessite très peu de mémoire et est adapté aux problèmes impliquant une grande quantité de données ou de paramètres. Il est également adapté aux problèmes avec des gradients très bruités. Enfin, les hyperparamètres ont des interprétations intuitives et leur réglage est plutôt simple [182].

La méthode du *dropout* [183] consiste à supprimer aléatoirement (avec une probabilité prédéfinie) les entrées d'une couche, qui peuvent être des variables d'entrée du jeu de données ou des activations d'une couche précédente. Cela a pour effet de simuler un grand nombre de réseaux avec différentes topologies et, par conséquent, de rendre les neurones généralement plus robustes aux entrées. Un tel mécanisme améliore considérablement la capacité de généralisation des réseaux de neurones profonds, même si son coût calculatoire est faible, et réduit le risque de surapprentissage (ce qui conduit à une meilleure capacité de généralisation). C'est pourquoi la méthode du *dropout* est aujourd'hui la méthode de régularisation la plus populaire. Son principal inconvénient est de ralentir la phase d'apprentissage et, généralement, d'augmenter le nombre d'itérations.

Pour les modèles MLP et LSTM, nous avons fait de notre mieux pour optimiser les résultats, et après avoir testé de nombreuses architectures, nous avons retenu les topologies présentées sur les Figures 2.6 et 2.7.

## 2.4 Résultats

Cette section présente les résultats des différents modèles développés. La capacité de généralisation des modèles est évaluée en comparant les prévisions sur divers horizons de temps allant de 10min à 24h, couvrant ainsi les échelles infra-horaire et infra-journalière. Pour évaluer la précision des prévisions, les critères mentionnés à la Section 1.4 du Chapitre 1, à savoir la racine carrée de l'erreur quadratique moyenne normalisée (nRMSE), l'erreur absolue moyenne dynamique (DMAE), l'*interval score* (IS) et le score de compétence (SS), sont calculés pour chaque horizon de temps et pour chaque modèle. Les résultats pour les horizons de prévision considérés sont présentés sur la Figure 2.8.

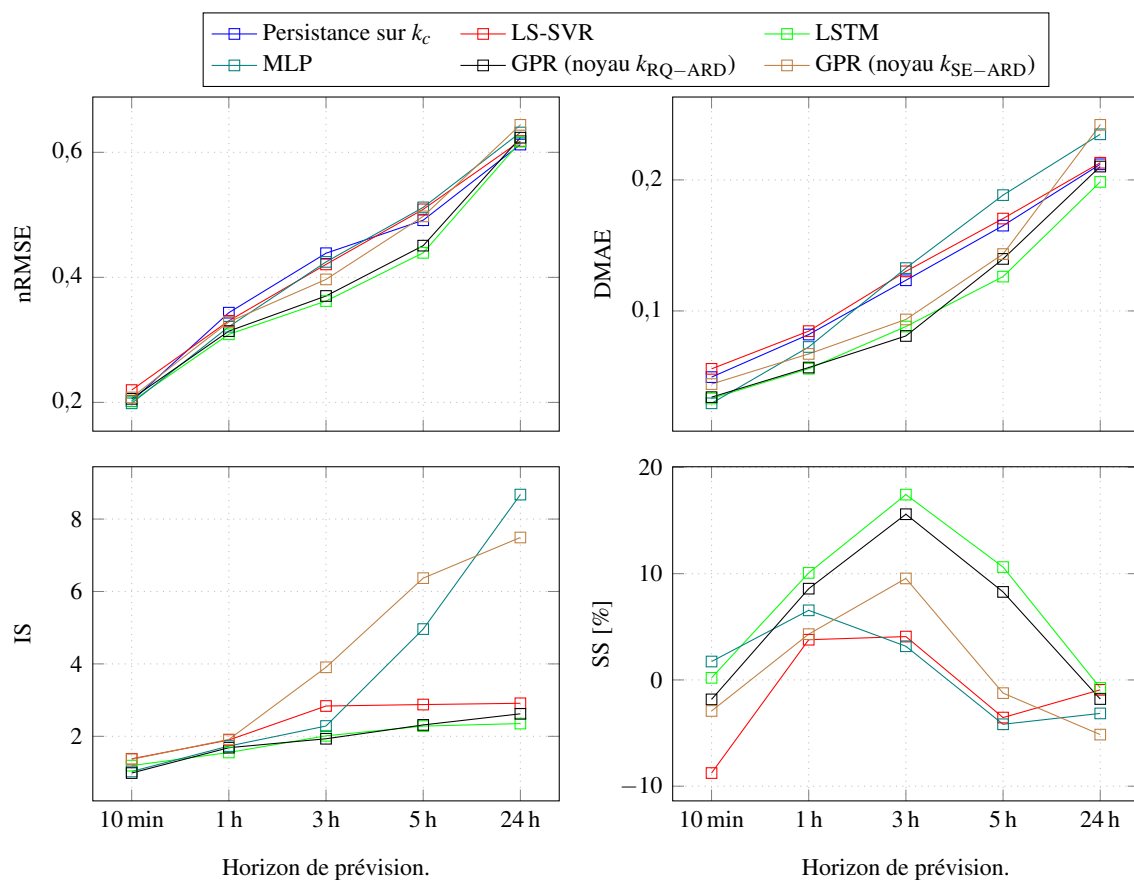


FIGURE 2.8 – Valeurs des critères de performance pour tous les modèles considérés dans l'étude comparative en fonction de l'horizon de prévision. Ces valeurs sont calculées sur toute la période de test (année 2019).

Intéressons-nous dans un premier temps à l'horizon de prévision de 10 min. Comme on peut le voir sur la Figure 2.8, selon les valeurs de nRMSE, tous les modèles développés donnent des résultats quasiment similaires. Les modèles ANN ont néanmoins un léger avantage, tandis que le modèle LS-SVR produit le moins bon résultat (nRMSE = 21,97%). D'après le score de compétence SS, on remarque que les performances de la persistance sur l'indice ciel clair sont pratiquement similaires à celles des modèles ANN et légèrement supérieures à celles des modèles GPR et du modèle LS-SVR; ce qui met en lumière la principale propriété de la persistance, qui est d'être efficace pour des horizons courts. Cependant, il existe des décalages temporels entre les prévisions fournies par le modèle de



persistance sur l'indice ciel clair et les données de test (Figures 2.9–2.13) qui se reflètent dans les valeurs de DMAE. En effet, sur la base de ce critère, la persistance et le modèle LS-SVR donnent les moins bons résultats. Les valeurs du critère IS pour les modèles fondés sur l'apprentissage automatique sont également présentés sur la Figure 2.8. Tous ces modèles donnent une probabilité de couverture élevée pour un horizon de temps de 10 min, avec un avantage pour le modèle GPR fondé sur le noyau  $k_{RQ-ARD}$  qui, de plus, a des intervalles de confiance étroits, comme le montre sa valeur de IS.

Les performances de tous les modèles se dégradent à mesure que l'horizon de temps augmente. En effet, les modèles ont alors du mal à appréhender les fluctuations infra-journalières. L'augmentation des valeurs des critères nRMSE, DMAE et IS témoigne de ces effets. Pour des horizons temps infra-journaliers de 1 h et 3 h, le modèle de persistance sur l'indice ciel clair produit un niveau d'erreur plus élevé que tous les autres modèles, selon les valeurs de nRMSE (Figure 2.8). Pour un horizon de temps de 5 h, il produit de meilleurs résultats que les modèles LS-SVR, MLP et GPR fondé sur le noyau  $k_{SE-ARD}$ . Pour un horizon de temps de 24 h, en regardant le score de compétence, on remarque que le modèle de persistance sur l'indice ciel clair prend le dessus sur tous les autres modèles. En effet, à cet horizon, la persistance sur l'indice ciel clair pondère le GHI obtenu par modèle ciel clair des vingt-quatre prochaines heures par l'indice ciel clair de la présente journée, tandis que les modèles fondés sur l'apprentissage automatique souffrent d'une accumulation des erreurs de prévision en itérant les prévisions des modèles de prévision à un pas. Cependant, la Figure 2.8 montre que, si on considère tous les horizons de temps infra-journaliers (1 h, 3 h et 24 h), on constate que le modèle LSTM et le modèle GPR fondé sur le noyau  $k_{RQ-ARD}$  surpassent tous les autres modèles.

Quelques exemples de prévisions pour des horizons de temps allant de 10 min à 24 h sont présentés sur les Figures 2.9–2.13. Une période de six jours tirés du jeu de données de test (du 28/11/2019 au 03/12/2019), comportant à la fois des jours clairs et des jours nuageux, a été choisie. Pour un horizon de temps de 10 min, le modèle de persistance sur l'indice ciel clair fournit de bonnes prévisions, mais lorsque l'horizon augmente, ses performances se dégradent rapidement. Pour un horizon de temps 24 h, en regardant, par exemple, la prévision du deuxième jour par la persistance sur l'indice ciel clair sur la Figure 2.13, on voit bien l'effet de pondération du GHI ciel clair des vingt-quatre prochaines heures par l'indice ciel clair de la précédente journée. Si l'on observe la Figure 2.9, pour un horizon de temps 10 min, tous les modèles suivent assez bien les données de test, à l'exception du modèle LS-SVR qui ne prédit pas bien les fluctuations lors des jours couverts. Lorsque l'horizon de temps augmente, les performances de tous les modèles de prévision se dégradent. Cependant, les figures montrent que le modèle LSTM et le modèle GPR fondé sur le noyau  $k_{RQ-ARD}$  surpassent les autres modèles. Cela est particulièrement évident sur les Figures 2.10–2.12, pour les horizons de temps infra-journaliers, où malgré la dégradation de leurs performances, ces deux modèles suivent assez bien les données de test. À mesure que l'horizon de temps augmente, l'accumulation des erreurs de prévision en itérant des prévisions à un pas devient préjudiciable aux modèles fondés sur l'apprentissage automatique, ceci est surtout visible sur la Figure 2.13, pour un horizon de temps de 24 h. Les modèles ANN présentent également de larges intervalles de confiance à long terme, ce qui se voit dans leurs prévisions affichées sur la Figure 2.13.

Globalement, sur les Figures 2.9–2.12, on note que le modèle LSTM donne de meilleurs résultats que les autres modèles pour des journées nuageuses, tandis que le modèle GPR fondé sur le noyau  $k_{RQ-ARD}$  donne les meilleurs résultats pour des jours clairs. Les modèles MLP et GPR fondé sur le noyau  $k_{SE-ARD}$  n'arrivent pas à suivre les données lorsque l'horizon de temps augmente. Le modèle LS-SVR converge vers une fonction lisse à long terme. Ceci n'est pas surprenant, car son noyau est adapté à la représentation de fonctions lisses et ne peut pas modéliser correctement les variations multi-échelles du GHI. Parmi les modèles ANN, le modèle LSTM, qui peut apprendre les dépendances à long terme dans les données, est plus performant que le modèle MLP. Les grands intervalles de confiance, lorsque l'horizon de prévision augmente, démontrent la dégradation des performances des modèles.

En examinant les performances des modèles sur la base des critères (Figure 2.8) et des exemples de prévisions présentés sur les Figures 2.9–2.13, on constate donc que les deux meilleurs modèles pour la



prévision du GHI sont le modèle GPR fondé sur le noyau  $k_{RQ-ARD}$  et le modèle LSTM. En effet, ces deux modèles sont plus performants que les autres modèles considérés dans cette étude comparative. Cependant, comme le modèle LSTM et le modèle GPR fondé sur le noyau  $k_{RQ-ARD}$  donnent des résultats similaires, il n'est guère facile de les départager.

Le [Tableau 2.2](#) présente les temps d'apprentissage des modèles, le nombre de paramètres impliqués dans chaque modèle et le temps nécessaire pour effectuer une prévision pour chaque modèle. Bien que de nombreux paramètres soient impliqués dans les modèles ANN, ils sont rapides à entraîner et leurs prévisions sont presque instantanées. En ce qui concerne les modèles GPR, la phase d'entraînement est très longue et ils nécessitent plus de temps pour fournir des prévisions. En effet, la production d'une prévision par un modèle GPR nécessite l'inversion d'une matrice de dimension égale à la taille des données d'entraînement. Cependant, il existe dans la littérature de nombreuses approches [184] qui peuvent être utilisées pour réduire la complexité calculatoire et le besoin en mémoire tout en préservant les performances, dans l'optique d'une implémentation *in situ*.

TABLEAU 2.2 – Durées d'apprentissage et temps d'exécution des modèles.

Modèle	Durée d'apprentissage (min)	Nombre de paramètres	Temps d'exécution pour une prévision (s)
LS-SVR	133	26457	4,12
LSTM	20	127459	0,05
MLP	10	58635	0,02
GPR fondé sur $k_{RQ-ARD}$	209	26457	12,59
GPR fondé sur $k_{SE-ARD}$	200	26457	12,59

Le principal objectif de cette étude comparative était de faire le choix d'une méthode pour la prévision du GHI. Le cas d'étude situé à Villeneuve-de-la-Raho, à proximité de Perpignan, présente plus de jours clairs que de jours nuageux. De plus, le choix d'un modèle de prévision doit être en adéquation avec les exigences de l'application visée. En effet, le projet de démonstrateur de réseau électrique intelligent en milieu périurbain/rural « Smart Occitania » traite du développement d'une stratégie de contrôle à coût calculatoire maîtrisé pour la gestion prédictive d'un réseau électrique de distribution basse tension. Par conséquent, malgré un coût calculatoire plus élevé, du fait des caractéristiques climatiques du cas d'étude, notre choix s'est porté sur la régression non paramétrique par processus gaussien. En effet, d'après l'étude de la variabilité des données réalisée à la [Sous-sous-section 1.2.6.5](#) du [Chapitre 1](#), les journées « ciel clair » sont prédominants dans les jeux de données de Perpignan et Villeneuve-de-la-Raho (cas d'étude). Par ailleurs, les intervalles de confiance associés aux prévisions que fournissent naturellement les modèles GPR (il n'est pas nécessaire, contrairement aux modèles LSTM de réaliser des simulations de Monte-Carlo) ont permis d'améliorer la robustesse du contrôleur prédictif développé. Il faut noter aussi que les modèles GPR sont capables de fournir des prévisions satisfaisantes avec peu de données d'entraînement. Ces modèles sont donc plus flexibles et peuvent s'adapter facilement à une reconfiguration du réseau ou à la disponibilité de nouveaux jeux de données. Les processus gaussiens peuvent également être utilisés pour prévoir le GHI dans le temps et l'espace ([Chapitre 4](#)).

## 2.5 Conclusion

Les modèles de prévision statistiques fondés sur l'analyse d'une base de données et l'apprentissage automatique sont très utilisés pour la prévision localisée du GHI. Afin d'identifier une méthode pour notre cas d'étude, une étude comparative des modèles phares a été menée. Sur la base des critères et des exemples de prévisions présentés sur les [Figures 2.9–2.13](#) pour les horizons de prévision considérés, les modèles qui émergent sont le modèle LSTM et le modèle GPR fondé sur le noyau  $k_{RQ-ARD}$ .

Cependant, en raison des exigences de notre application, notre choix s'est porté sur le modèle GPR. En voyant le temps d'exécution des modèles (Tableau 2.2), l'on pourrait plutôt être amené à choisir le modèle LSTM mais il existe aujourd'hui plusieurs approches permettant de réduire considérablement le temps d'exécution des modèles GPR tout en préservant les performances. De plus les modèles GPR fournissent naturellement des intervalles de confiance – les intervalles de confiance associés aux prévisions que fournissent naturellement les modèles GPR ont permis d'améliorer la robustesse du contrôleur prédictif développé – et s'adaptent facilement même lorsque les données d'entraînement sont en nombre limité.

Néanmoins, ces modèles développées ici souffrent d'une accumulation des erreurs lors de l'itération des prédictions à un pas. Dans le Chapitre 3, nous verrons qu'en prenant le temps en entrée du modèle GPR, on peut prendre en considération les comportements du GHI en formulant des biais inductifs et ainsi construire des modèles plus performants.

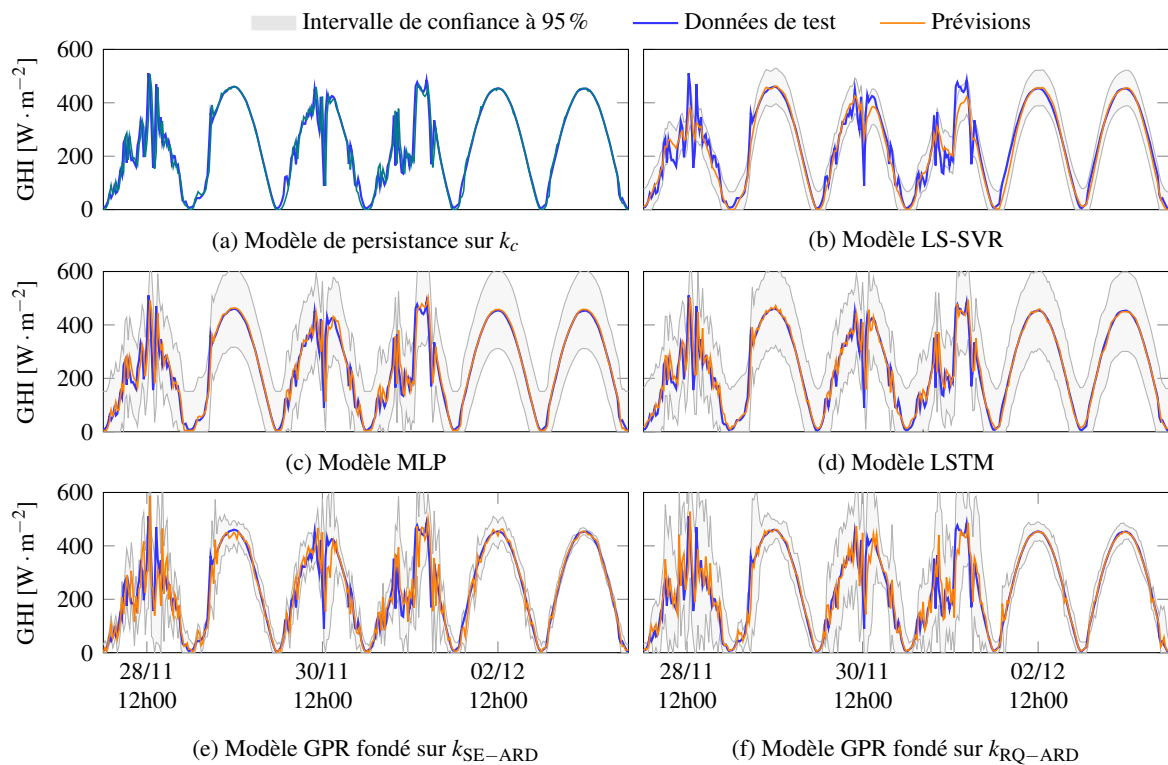


FIGURE 2.9 – Exemples de prévisions à 10 min du GHI obtenues grâce aux modèles considérés dans l'étude comparative.

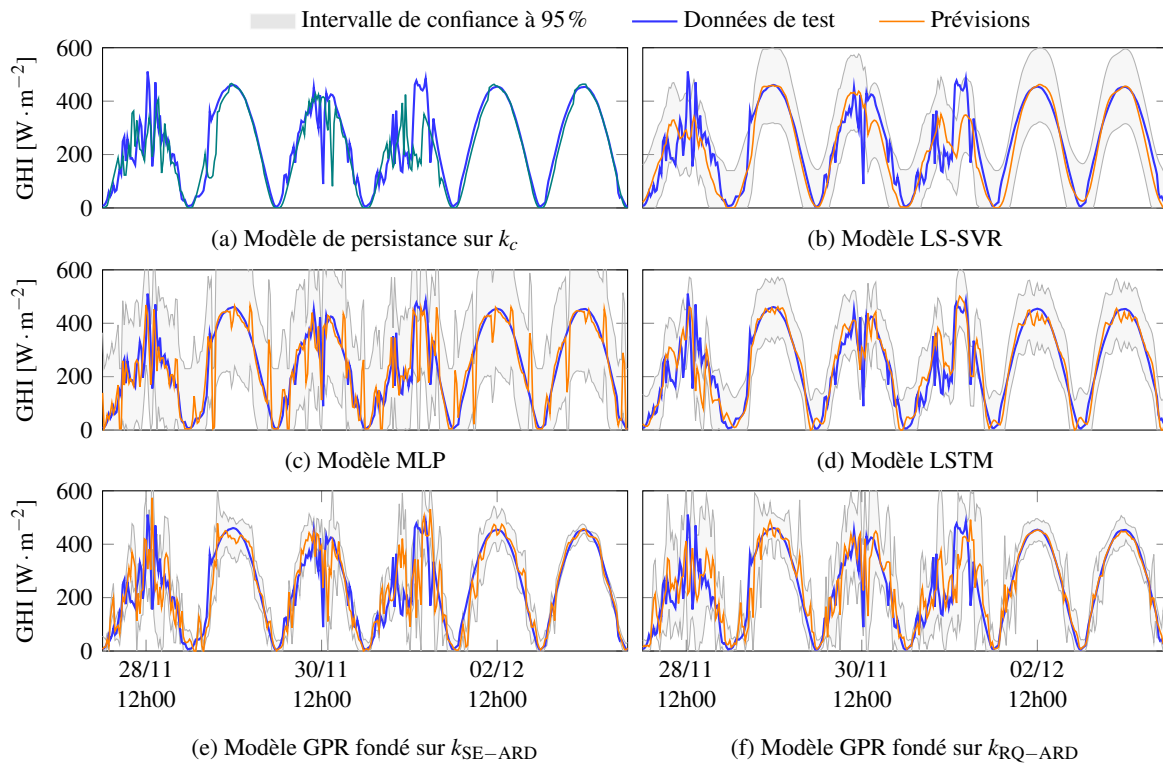


FIGURE 2.10 – Exemples de prévisions à 1 h du GHI obtenues grâce aux modèles considérés dans l'étude comparative.

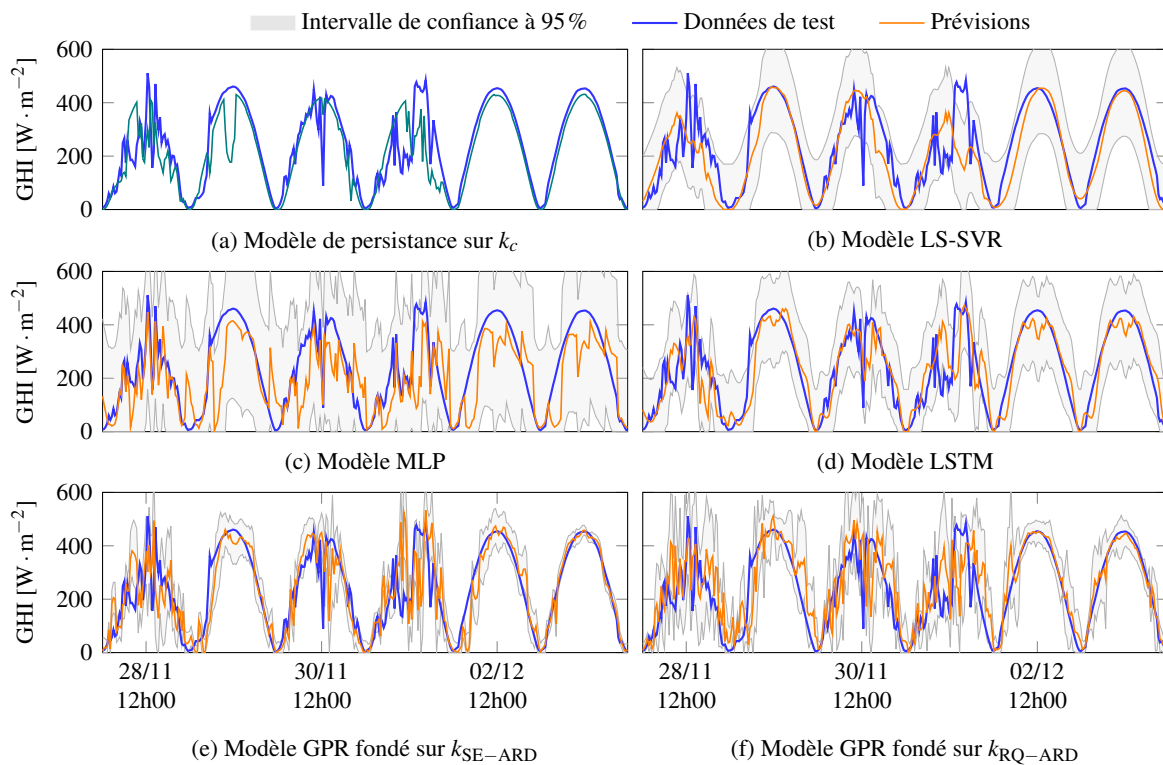


FIGURE 2.11 – Exemples de prévision à 3 h du GHI obtenues grâce aux modèles considérés dans l'étude comparative.

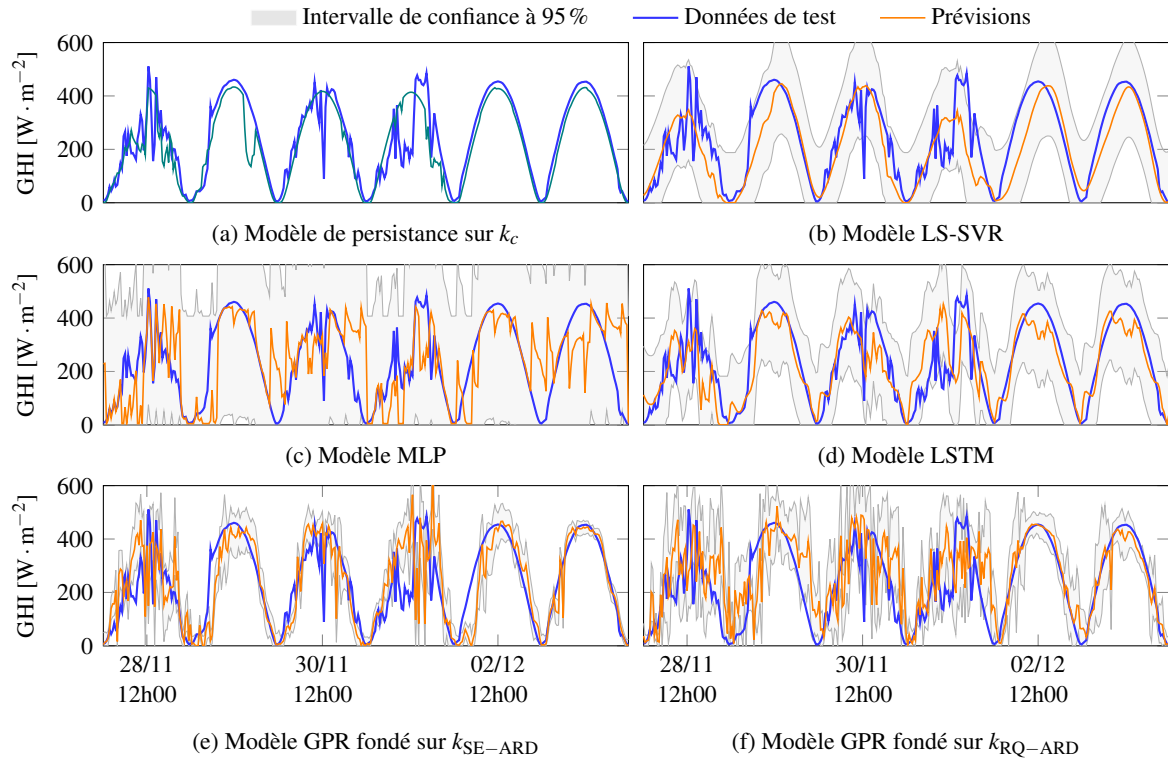


FIGURE 2.12 – Exemples de prévision à 5 h du GHI obtenues grâce aux modèles considérés dans l'étude comparative.

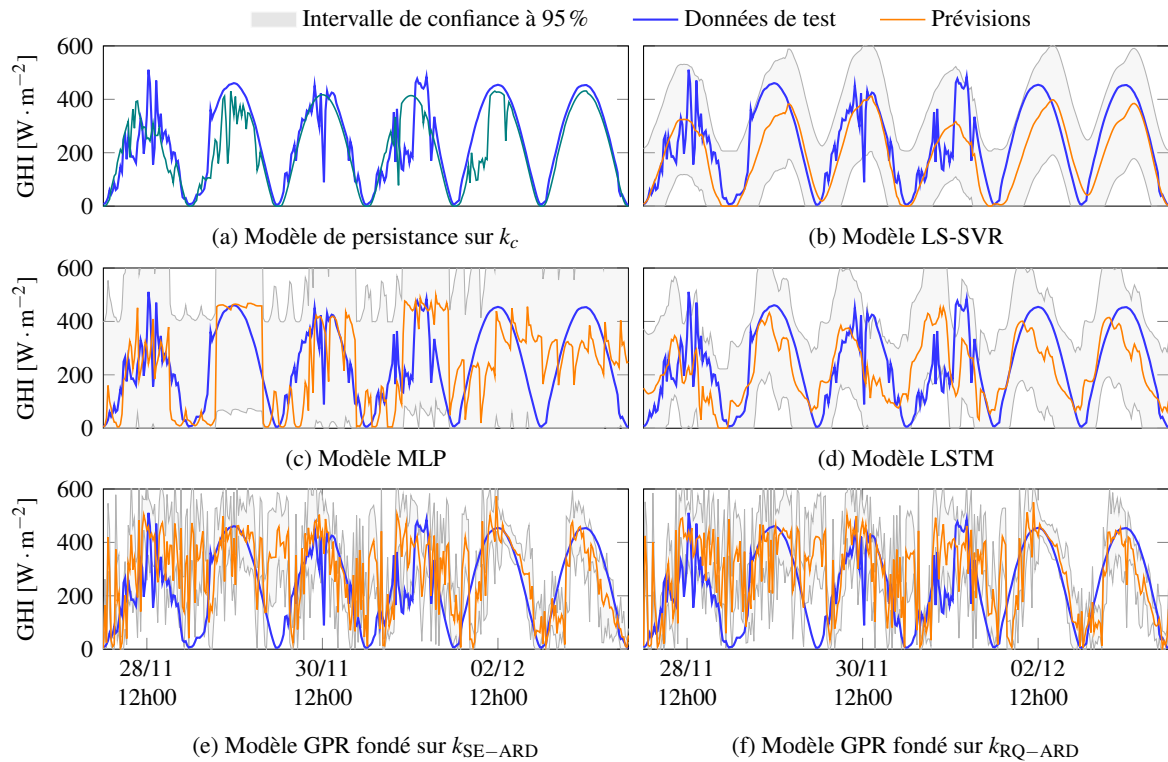


FIGURE 2.13 – Exemples de prévision à 24 h du GHI obtenues grâce aux modèles considérés dans l'étude comparative.



## Chapitre 3

# Prévision multi-horizon de l'éclairement global horizontal par régression non paramétrique par processus gaussien

### 3.1 Introduction

Dans le chapitre précédent, une étude comparative des modèles statistiques de prévision fondés sur l'analyse d'une base de données et l'apprentissage automatique a été menée, dans le but de choisir un outil adapté à la prévision localisée du GHI. Ces modèles ont été développés en utilisant un historique d'observations du GHI comme entrée. Après analyse des résultats et au regard des exigences de notre application, la régression non paramétrique par processus gaussien a été choisie. En effet, cette thèse traitant de la prévision de l'éclairement global horizontal à des fins de prévision de la production photovoltaïque fait partie des travaux entrepris par PROMES-CNRS dans le cadre du projet « Smart Occitania ». Ces travaux ont pour objectif le développement d'algorithmes à coût calculatoire maîtrisé, pouvant être implémentés en temps réel, pour la gestion prévisionnelle d'un réseau électrique de distribution basse tension qui fait face au déploiement d'une production décentralisée, en particulier du solaire photovoltaïque. Nous avons également vu également que, pour ce type de régression, le choix du noyau est primordial. En revanche, nous ne nous sommes pas posé de questions vis-à-vis de l'entrée : considérer un historique d'observations permettant de mettre en évidence les corrélations semble, en effet, naturel. Toutefois, comme nous allons le voir dans ce chapitre, il est également possible de considérer le temps en entrée, ce qui a deux effets notables : tout d'abord, cela permet de construire un noyau particulièrement adapté au GHI, grâce aux biais inductifs connus sur son évolution au cours du temps ; ensuite, le modèle ainsi obtenu est par construction multi-horizon, puisqu'il peut fournir une prévision à n'importe quel instant.

Ce chapitre a pour objectif de répondre à la question suivante : *quelle entrée pour les modèles fondés sur la régression non paramétrique par processus gaussien pour la prévision du GHI ?* Ainsi, ce chapitre traite de la comparaison entre modèles GPR utilisant le temps en entrée et utilisant un historique d'observation du GHI. Une fois entraînés, les modèles GPR fondés sur le temps fournissent des prévisions du GHI à n'importe quel horizon de temps. Parmi les modèles utilisant un historique d'observations du GHI en entrée, on distingue des modèles multi-horizons et des modèles mono-horizons. Deux stratégies sont ainsi employées pour obtenir une prévision à un horizon  $k\Delta t$  à partir d'observations.

- La stratégie itérative, illustrée par la [Figure 3.1](#), qui utilise un modèle de prévision à un pas.
- La stratégie directe, illustrée par la [Figure 3.2](#), qui utilise un modèle de prévision à l'horizon considéré.

Par conséquent, en considérant un historique d'observations du GHI comme entrée, une nouvelle question voit le jour : *faut-il faire le choix de modèles multi-horizons ou de modèles mono-horizons ?*

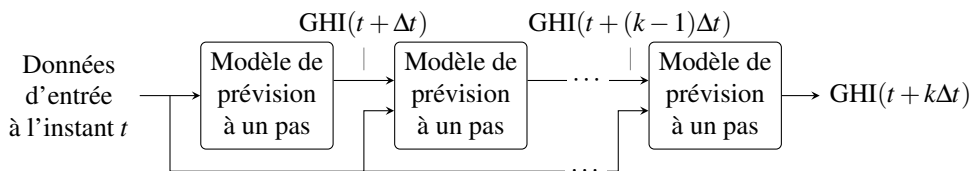


FIGURE 3.1 – Stratégie itérative

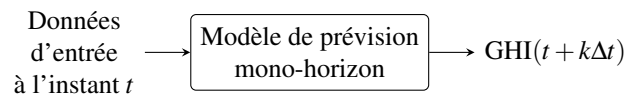


FIGURE 3.2 – Stratégie directe

Une comparaison entre ces différents modèles sera effectuée dans ce chapitre. Enfin, un noyau adapté à la prévision temporelle du GHI sera identifié. Ce noyau sera associé à un noyau spatial pour la prévision spatio-temporelle du GHI dans le [Chapitre 4](#).

Les modèles sont développés en utilisant le jeu de données couvrant les années 2016 et 2017 tiré des mesures du GHI effectuées sur le site de Perpignan (pour plus de détails sur ces données, voir la [Sous-sous-section 1.2.6.2 du Chapitre 1](#)). Les données sont échantillonnées à pas de 10 min et les horizons de prévision considérés vont de 10 min à 24h. Les modèles sont entraînés avec les données de 2016 et sont évalués avec celles de 2017, permettant d'évaluer la robustesse des modèles face à toutes les conditions atmosphériques possibles au cours d'une année.

La [Section 3.2](#) présente les modèles GPR utilisant le temps en entrée. Les hypothèses sous-jacentes aux noyaux choisis sont également détaillées. La [Section 3.3](#) est consacrée à la description des modèles GPR utilisant comme entrée un historique d'observations du GHI. Dans la [Section 3.4](#), sont effectués une analyse des résultats et un choix du noyau adapté à la prévision temporelle du GHI.

## 3.2 Modèles GPR utilisant le temps en entrée

On rappelle qu'un processus gaussien est complètement défini par sa fonction moyenne et sa fonction de covariance ou noyau. Le choix d'un noyau – le noyau est l'élément clé d'un modèle GPR – est fondé sur les hypothèses à propos de la fonction à apprendre. Les noyaux de base – quelques noyaux de base sont définis dans le [Chapitre 2](#) – peuvent être combinés par addition ou multiplication afin de créer des fonctions de covariance adaptées à l'application en question. L'addition et la multiplication sont des opérations simples permettant de garder la principale propriété d'un noyau qui est d'être semi-défini positif. Une liste complète de noyaux de base peut être trouvée dans [\[147\]](#).

### 3.2.1 Étude comparative des modèles GPR utilisant le temps en entrée fondés sur des noyaux de base et des noyaux combinés

Afin d'identifier le noyau le plus adapté à la prévision du GHI, une étude comparative des modèles GPR fondés sur des noyaux de base et des noyaux combinés définis comme des sommes ou produits de noyaux de base a été menée au sein de notre équipe [\[137\]](#). Deux jeux de données couvrant chacun une période de 45 jours tirés de mesures du GHI réalisées à Perpignan à l'aide d'un RSI (voir la [Sous-sous-section 1.2.6.2 du Chapitre 1](#)) ont été utilisés. Ces jeux de données ont été nommés « données d'hiver » (du 01/11/2015 au 15/12/2015) et « données d'été » (du 05/06/2015 au 19/07/2015). Le GHI est plus facile à prévoir en été puisque les journées « ciel clair » y sont fréquentes, tandis que les données d'hiver ont permis de mettre en évidence la robustesse des modèles GPR face à des conditions atmosphériques complexes. En effet, les journées partiellement couvertes et complètement couvertes sont fréquentes en hiver, ce qui rend difficile la prévision du GHI. Pour ces deux jeux de données, les



30 premiers jours ont été utilisés pour l'entraînement des modèles et les 15 derniers jours pour évaluer leur capacité de généralisation. Les données ont été échantillonnées à un pas de temps de 30 min et divers horizons de prévision allant de 30 min à 48 h ont été considérés par les auteurs [137]. L'erreur quadratique moyenne normalisée nRMSE des modèles GPR en fonction de l'horizon de prévision est présentée sur la [Figure 3.3a](#) pour les données d'été et sur la [Figure 3.3b](#) pour les données d'hiver.

Les noyaux de base considérés sont :  $k_{\text{Per}}$ ,  $k_{\text{E}}$ ,  $k_{\text{SE}}$ ,  $k_{\text{RQ}}$ ,  $k_{\text{M}_{3/2}}$  et  $k_{\text{M}_{5/2}}$  (se référer au [Paragraphe 2.3.1.3.1](#) du [Chapitre 2](#) pour la formulation de ces noyaux). Les noyaux quasi-périodiques résultant du produit d'un noyau périodique et d'un noyau non périodique sont définis par :

$$k_{\text{Per} \times \text{SE}} = \sigma^2 \exp \left( -\frac{2 \sin^2 \left( \frac{\pi(x-x')}{P} \right)}{\ell_1^2} - \frac{(x-x')^2}{2\ell_2^2} \right) \quad (3.1)$$

$$k_{\text{Per} \times \text{RQ}} = \sigma^2 \exp \left( -\frac{2 \sin^2 \left( \frac{\pi(x-x')}{P} \right)}{\ell_1^2} \right) \cdot \left( 1 + \frac{(x-x')^2}{2\alpha\ell_2^2} \right)^{-\alpha} \quad (3.2)$$

$$k_{\text{Per} \times \text{E}} = \sigma^2 \exp \left( -\frac{2 \sin^2 \left( \frac{\pi(x-x')}{P} \right)}{\ell_1^2} - \frac{|x-x'|}{\ell_2} \right) \quad (3.3)$$

$$k_{\text{Per} \times \text{M}_{3/2}} = \sigma^2 \exp \left( -\frac{2 \sin^2 \left( \frac{\pi(x-x')}{P} \right)}{\ell_1^2} - \sqrt{3} \frac{|x-x'|}{\ell_2} \right) \cdot \left( 1 + \sqrt{3} \frac{|x-x'|}{\ell_2} \right) \quad (3.4)$$

$$k_{\text{Per} \times \text{M}_{5/2}} = \sigma^2 \exp \left( -\frac{2 \sin^2 \left( \frac{\pi(x-x')}{P} \right)}{\ell_1^2} - \sqrt{5} \frac{|x-x'|}{\ell_2} \right) \cdot \left( 1 + \sqrt{5} \frac{|x-x'|}{\ell_2} + \sqrt{5} \frac{(x-x')^2}{3\ell_2^2} \right) \quad (3.5)$$

où  $\sigma$  est l'amplitude du signal,  $\ell_1$  et  $\ell_2$  sont les paramètres de corrélation,  $P$  est la période et  $\alpha$  est le paramètre de pondération des variations multi-échelles.

Les noyaux quasi-périodiques résultant de la somme d'un noyau périodique et d'un noyau non périodique sont définis par :

$$k_{\text{Per} + \text{SE}} = \sigma_1^2 \exp \left( -\frac{2 \sin^2 \left( \frac{\pi(x-x')}{P} \right)}{\ell_1^2} \right) + \sigma_2^2 \exp \left( -\frac{(x-x')^2}{2\ell_2^2} \right) \quad (3.6)$$

$$k_{\text{Per} + \text{RQ}} = \sigma_1^2 \exp \left( -\frac{2 \sin^2 \left( \frac{\pi(x-x')}{P} \right)}{\ell_1^2} \right) + \sigma_2^2 \left( 1 + \frac{(x-x')^2}{2\alpha\ell_2^2} \right)^{-\alpha} \quad (3.7)$$

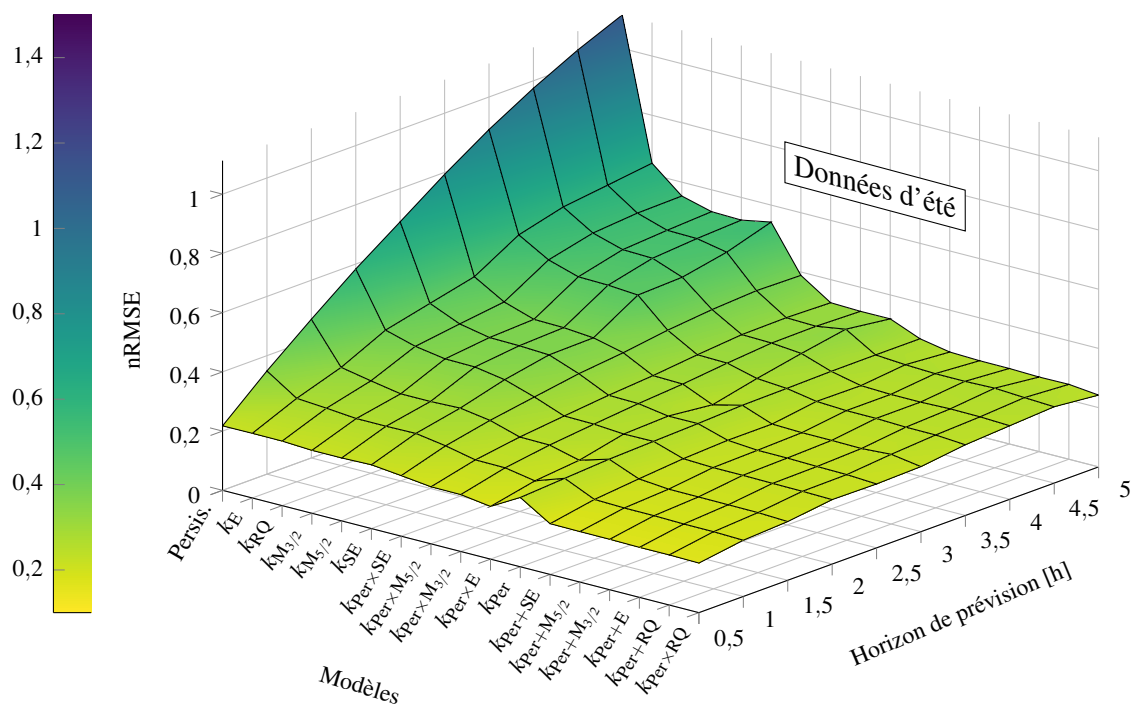
$$k_{\text{Per} + \text{E}} = \sigma_1^2 \exp \left( -\frac{2 \sin^2 \left( \frac{\pi(x-x')}{P} \right)}{\ell_1^2} \right) + \sigma_2^2 \exp \left( -\frac{|x-x'|}{\ell_2} \right) \quad (3.8)$$

$$k_{\text{Per} + \text{M}_{3/2}} = \sigma_1^2 \exp \left( -\frac{2 \sin^2 \left( \frac{\pi(x-x')}{P} \right)}{\ell_1^2} \right) + \sigma_2^2 \left( 1 + \sqrt{3} \frac{|x-x'|}{\ell_2} \right) \cdot \exp \left( -\sqrt{3} \frac{|x-x'|}{\ell_2} \right) \quad (3.9)$$

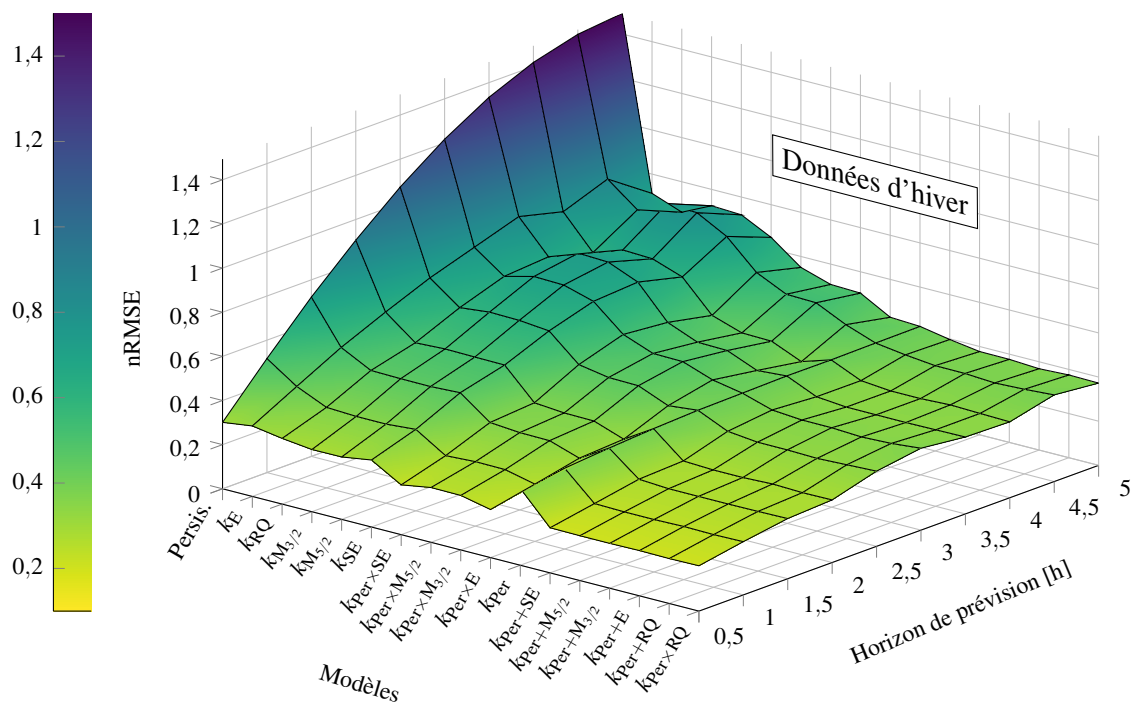
$$k_{\text{Per} + \text{M}_{5/2}} = \sigma_1^2 \exp \left( -\frac{2 \sin^2 \left( \frac{\pi(x-x')}{P} \right)}{\ell_1^2} \right) + \sigma_2^2 \left( 1 + \sqrt{5} \frac{|x-x'|}{\ell_2} + \sqrt{5} \frac{(x-x')^2}{3\ell_2^2} \right) \cdot \exp \left( -\sqrt{5} \frac{|x-x'|}{\ell_2} \right) \quad (3.10)$$

où  $\sigma_1$  et  $\sigma_2$  sont les amplitudes des signaux,  $\ell_1$  et  $\ell_2$  sont les paramètres de corrélation,  $P$  est la période et  $\alpha$  est le paramètre de pondération des variations multi-échelles.





(a) Résultats pour les données d'été, du 05/06/2015 au 19/07/2015, couvrant une période de 45 jours dont les 15 derniers jours sont utilisés pour l'évaluation de la performance des modèles.



(b) Résultats pour les données d'hiver, du 01/11/2015 au 15/12/2015, couvrant une période de 45 jours dont les 15 derniers jours sont utilisés pour l'évaluation de la performance des modèles.

FIGURE 3.3 – Erreur nRMSE en fonction de l'horizon de prévision pour le modèle de persistance sur le GHI et les modèles GPR avec différents noyaux de base et noyaux combinés [137].

Les résultats ont montré que tous les modèles GPR surpassent la persistance sur le GHI, considérée dans l'étude comparative comme modèle de référence. En ce qui concerne les modèles GPR fondés sur des noyaux de base, aucun modèle ne sort du lot. En examinant les valeurs des erreurs sur les Figures 3.3a et 3.3b, en fonction de l'horizon de prévision, on constate qu'ils fournissent tous alternativement de bons résultats. En particulier, on peut voir que le noyau  $k_{SE}$ , généralement choisi par défaut pour prévoir le GHI, ne produit pas forcément les meilleurs résultats. Le noyau périodique  $k_{Per}$  produit tout simplement un signal périodique. Il peut être assimilé à un modèle « ciel clair » dont les paramètres ont été ajustés par apprentissage. En conséquence, les modèles GPR fondés sur ce noyau reproduisent simplement de manière récurrente la même courbe en forme de cloche et donnent pratiquement les mêmes nRMSE lorsque l'horizon de temps augmente (Figures 3.3a et 3.3b). Bien que ces modèles GPR puissent produire de bonnes prévisions pendant des journées « ciel clair », ils sont incapables d'appréhender la variabilité du GHI induite par les perturbations atmosphériques.

Cependant, les modèles GPR fondés sur des noyaux quasi-périodiques – ces noyaux combinent un noyau périodique et un noyau non périodique – possèdent l'avantage du noyau périodique, tout en parvenant à prévoir les variations du GHI dues aux perturbations atmosphériques. Les modèles GPR fondés sur les noyaux quasi-périodiques surpassent non seulement la persistance sur le GHI, mais aussi les modèles GPR fondés sur les noyaux de base (Figures 3.3a et 3.3b). Bien que tous les modèles GPR donnent globalement des résultats similaires à horizon de temps infra-horaire (30 min) – les noyaux quasi-périodiques ont néanmoins un léger avantage –, lorsque l'horizon augmente, la supériorité des noyaux quasi-périodiques devient manifeste. Les Figures 3.3a et 3.3b montrent que les modèles fondés sur le noyau quasi-périodique  $k_{Per \times RQ}$  (équation (3.2)) sont les plus performants.

Par conséquent, le noyau retenu pour l'étude comparative est  $k_{Per \times RQ}$ . Ce noyau tire profit des connaissances *a priori* sur la dynamique du GHI. En effet, le GHI journalier peut être décomposé comme suit (Figure 3.4) : une composante périodique modélisée par le noyau  $k_{Per}$ , une composante stochastique due aux conditions atmosphériques modélisée par le noyau  $k_{RQ}$  et un bruit blanc. Dans [138], les auteurs ont confirmé que les noyaux quasi-périodiques sont adaptés à la prévision du GHI, en particulier le noyau  $k_{Per \times RQ}$ .

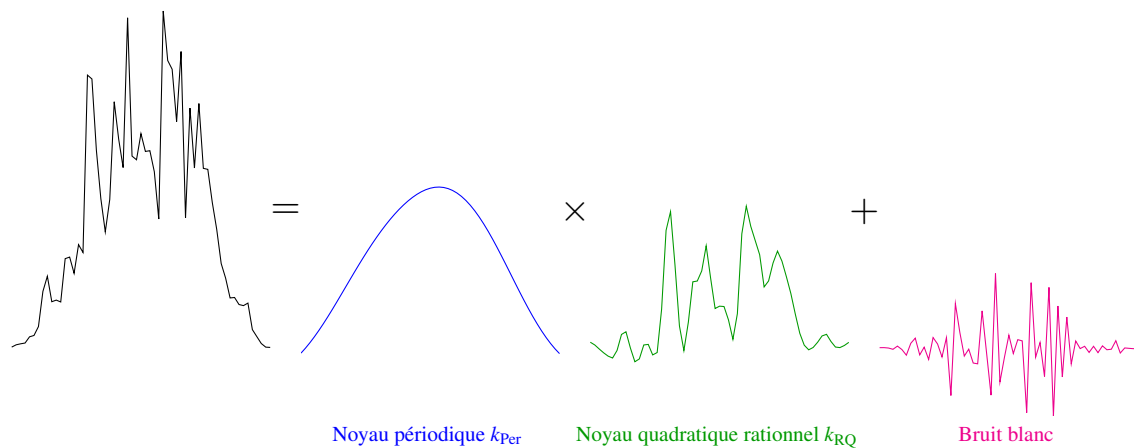


FIGURE 3.4 – Décomposition du GHI journalier. La composante périodique, liée à la géométrie Terre-Soleil, est modélisée par le noyau périodique  $k_{Per}$  et celle stochastique, due aux perturbations atmosphériques, par le noyau quadratique rationnel  $k_{RQ}$ .

Les paramètres du noyau  $k_{Per \times RQ}$  (équation (3.2)) et la variance du bruit, également appelés hyperparamètres, sont estimés en maximisant le logarithme de la vraisemblance marginale donné par l'équation. Pour leur initialisation, les choix suivants ont été faits :

- l'hyperparamètre  $\sigma^2$  a été initialisé à la variance des données d'apprentissage ;
- la période  $P$  a été initialisée à un jour indiquant la périodicité journalière du GHI ;
- les autres hyperparamètres ont été tirés aléatoirement d'une distribution uniforme  $\mathcal{U}(0, 1)$ .

### 3.2.2 Modèle GPR adapté à la prévision, utilisant le temps en entrée, pour des données couvrant plusieurs années

Si l'on considère des données du GHI couvrant des années – dans notre cas, deux ans de données sont considérés – on note l'existence d'une composante périodique annuelle, illustrée sur la [Figure 3.5](#), qui peut être incorporée au modèle GPR fondé sur le noyau quasi-périodique  $k_{\text{Per} \times \text{RQ}}$  par l'ajout d'un second noyau périodique [185]. Par conséquent, un noyau approprié pour la prévision du GHI

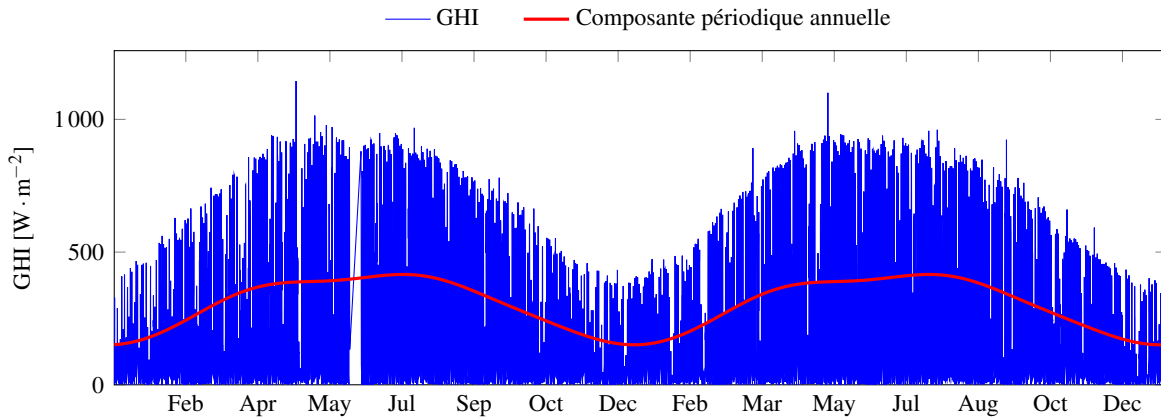


FIGURE 3.5 – Données du GHI couvrant les années 2016 et 2017. Une composante périodique saisonnière apparaît lorsqu'on considère des données sur plusieurs années. Les mesures ont été réalisées, à l'aide d'un pyranomètre à bande rotative (RSI), au laboratoire PROMES-CNRS à Perpignan.

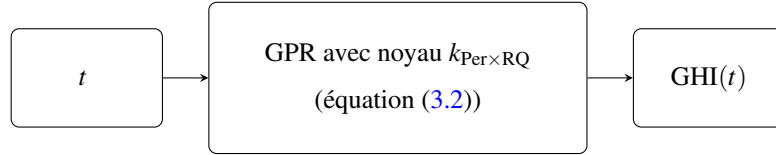
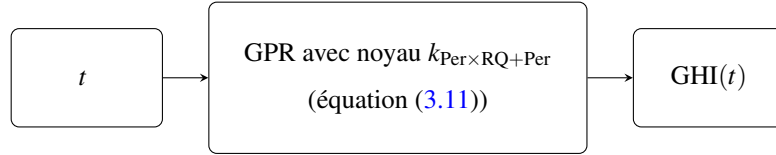
par modèle GPR utilisant le temps en entrée, lorsque des données couvrant plusieurs années sont considérées, est donné par :

$$k_{\text{Per} \times \text{RQ} + \text{Per}} = \sigma_1^2 \exp \left( -\frac{2 \sin^2 \left( \frac{\pi(x-x')}{P_1} \right)}{\ell_1^2} \right) \times \left( 1 + \frac{(x-x')^2}{2\alpha\ell_2^2} \right)^{-\alpha} + \sigma_2^2 \exp \left( -\frac{2 \sin^2 \left( \frac{\pi(x-x')}{P_2} \right)}{\ell_3^2} \right) \quad (3.11)$$

où  $\sigma_1$  et  $\sigma_2$  sont les amplitudes des signaux,  $\ell_1$ ,  $\ell_2$  et  $\ell_3$  sont les paramètres de corrélation,  $P_1$  et  $P_2$  sont des périodes,  $\alpha$  est un paramètre de pondération relatif aux variations multi-échelles. Pour l'initialisation des hyperparamètres, les choix suivants ont été faits :

- les hyperparamètres  $\sigma_1^2$  et  $\sigma_2^2$  ont été initialisés à la variance des données d'apprentissage ;
- les périodes  $P_1$  et  $P_2$  ont été respectivement initialisées à un jour et 365 jours, indiquant respectivement les périodicités journalière et annuelle du GHI ;
- les autres hyperparamètres sont tirés aléatoirement d'une distribution uniforme  $\mathcal{U}(0, 1)$ .

Les modèles GPR utilisant le temps en entrée considérés dans notre étude sont le modèle GPR fondé sur le noyau  $k_{\text{Per} \times \text{RQ}}$  et le modèle GPR fondé sur le noyau  $k_{\text{Per} \times \text{RQ} + \text{Per}}$ . Une comparaison entre ces deux modèles est également faite, mettant en évidence l'apport du noyau périodique modélisant la composante périodique saisonnière du GHI. Ces modèles GPR peuvent être utilisés pour prévoir le GHI à n'importe quel horizon de temps, une fois la phase d'apprentissage terminée [137, 185]. Il suffit de leur fournir l'instant correspondant à l'horizon de temps souhaité pour obtenir la prévision du GHI à cet instant. Ce sont naturellement des modèles de prévision multi-horizons. Les Figures 3.6 et 3.7 présentent le principe de la prévision du GHI par les modèles GPR fondés sur les noyaux  $k_{\text{Per} \times \text{RQ}}$  et  $k_{\text{Per} \times \text{RQ} + \text{Per}}$ .

FIGURE 3.6 – Modèle de prévision GPR fondé sur le noyau  $k_{\text{Per} \times \text{RQ}}$ , où  $t$  est le temps.FIGURE 3.7 – Modèle de prévision GPR fondé sur le noyau  $k_{\text{Per} \times \text{RQ} + \text{Per}}$ , où  $t$  est le temps.

### 3.3 Modèles GPR utilisant un historique d'observations du GHI en entrée

Les modèles GPR utilisant des observations en entrée ont été étudiés dans le chapitre précédent. En effet, dans l'étude comparative des modèles de prévision fondés sur l'apprentissage automatique présentée au [Chapitre 2](#), des modèles GPR multi-horizons utilisant des observations du GHI en entrée ont été développés. Les noyaux choisis, à savoir  $k_{\text{SE-ARD}}$  et  $k_{\text{RQ-ARD}}$ , sont également utilisés dans ce présent chapitre. En effet, l'utilisation d'un historique d'observations comme entrée du modèle implique une variable d'entrée multidimensionnelle qui ne varie pas forcément de manière isotrope suivant chacune des dimensions du vecteur d'entrée. Pour passer en revue toutes les approches possibles dans le cadre des modèles GPR utilisant des observations en entrée, des modèles multi-horizons et spécifiques par horizon de prévision fondés sur ces noyaux sont développés. Outre ces deux noyaux, les autres noyaux de la classe Matérn avec ARD sont également utilisés. Les noyaux ARD considérés sont définis par :

$$k_{\text{SE-ARD}}(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{1}{2} \sum_{d=1}^D \frac{(x_d - x'_d)^2}{\ell_d^2}\right) \quad (3.12)$$

$$k_{\text{RQ-ARD}}(\mathbf{x}, \mathbf{x}') = \sigma^2 \left(1 + \frac{1}{2\alpha} \sum_{d=1}^D \frac{(x_d - x'_d)^2}{\ell_d^2}\right)^{-\alpha} \quad (3.13)$$

$$k_{\text{M}_{1/2}\text{-ARD}}(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\sum_{d=1}^D \frac{|x_d - x'_d|}{\ell_d}\right) \quad (3.14)$$

$$k_{\text{M}_{3/2}\text{-ARD}}(\mathbf{x}, \mathbf{x}') = \sigma^2 \left(1 + \sqrt{3} \sum_{d=1}^D \frac{|x_d - x'_d|}{\ell_d}\right) \cdot \exp\left(-\sqrt{3} \sum_{d=1}^D \frac{|x_d - x'_d|}{\ell_d}\right) \quad (3.15)$$

$$k_{\text{M}_{5/2}\text{-ARD}}(\mathbf{x}, \mathbf{x}') = \sigma^2 \left(1 + \sqrt{5} \sum_{d=1}^D \frac{|x_d - x'_d|}{\ell_d} + \sqrt{5} \sum_{d=1}^D \frac{(x_d - x'_d)^2}{3\ell_d^2}\right) \cdot \exp\left(-\sqrt{5} \sum_{d=1}^D \frac{|x_d - x'_d|}{\ell_d}\right) \quad (3.16)$$

où  $\sigma$  est l'amplitude du signal,  $\alpha$  est un paramètre de pondération relatif aux variations multi-échelles, et  $\ell_1, \ell_2, \dots, \ell_D$  sont les paramètres de corrélation.

Les Figures 3.8 et 3.9 présentent la prévision du GHI par les modèles GPR multi-horizons (itération d'une prévision à un pas) et mono-horizons fondés sur le noyau  $k_{\text{SE-ARD}}$ . Le principe est similaire pour les autres noyaux ARD considérés.

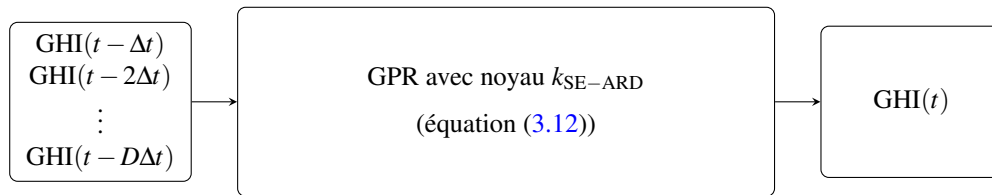


FIGURE 3.8 – Modèle de prévision GPR à un pas utilisant des observations en entrée, donnant un modèle multi-horizon par itération.  $GHI(t - \Delta t)$ ,  $GHI(t - 2\Delta t)$ ,  $\dots$ ,  $GHI(t - D\Delta t)$  sont des observations et  $\Delta t = 10$  min est le pas de temps.

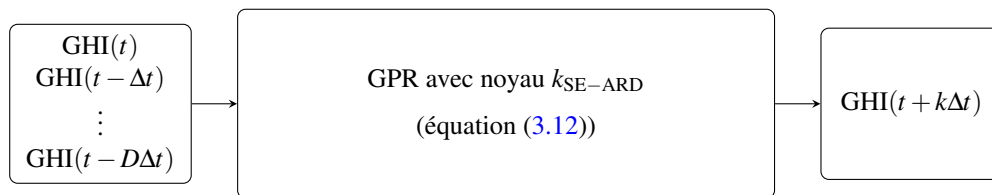


FIGURE 3.9 – Modèle de prévision GPR mono-horizon utilisant des observations en entrée, où  $GHI(t - \Delta t)$ ,  $GHI(t - 2\Delta t)$ ,  $\dots$ ,  $GHI(t - D\Delta t)$  sont des observations et  $\Delta t = 10$  min est le pas de temps.

## 3.4 Résultats et discussion

Cette section présente les prévisions du GHI obtenues grâce aux différents modèles GPR considérés. Les performances des modèles sont évaluées à l'aide des critères décrits dans la [Section 1.4 du Chapitre 1](#), à savoir l'erreur quadratique moyenne normalisée (nRMSE), l'erreur absolue moyenne dynamique (DMAE), l'*interval score* (IS) et le score de compétence (SS). Le modèle de persistance sur l'indice ciel clair est utilisé ici comme modèle de référence. Ces critères sont calculés pour chacun des modèles aux horizons considérés (allant de 10 min à 24 h).

Pour un horizon de temps de 10 min, il n'y a pas de différence entre les modèles GPR multi-horizons et les modèles GPR mono-horizons, puisque les prévisions multi-horizons sont obtenues en itérant les prévisions à un pas. Quelques exemples de prévision du GHI sur sept jours (du 20/11/2017 au 26/11/2017) aux horizons considérés sont présentées sur les [Figures 3.13 – 3.17](#).

### 3.4.1 Modèles GPR multi-horizons : temps ou observations en entrée ?

Dans cette section, les modèles GPR multi-horizons utilisant le temps en entrée sont comparés à ceux utilisant un historique d'observations du GHI afin de répondre à la question : *quelle entrée pour les modèles fondés sur la régression non paramétrique par processus gaussien pour la prévision du GHI ?* Il faut noter que les modèles GPR fondés sur le temps sont naturellement multi-horizons. Les valeurs des critères de performance de ces modèles sont présentées sur la [Figure 3.10](#).

Intéressons-nous dans un premier temps à l'horizon de prévision de 10 min. D'après la [Figure 3.10](#), on remarque que tous les modèles GPR multi-horizons et le modèle de persistance sur l'indice ciel clair donnent des résultats similaires. On constate que la persistance sur l'indice ciel clair est même légèrement plus performante que certains modèles GPR, ce qui confirme son efficacité pour des horizons très courts (nRMSE = 19,62 %). Néanmoins, les deux modèles GPR utilisant le temps en entrée ont un léger avantage sur tous les autres. Ces deux modèles donnent des résultats semblables (nRMSE = 18,05% pour le modèle GPR fondé sur le noyau  $k_{Per \times RQ + Per}$  et nRMSE = 18,11% pour le modèle GPR fondé sur le noyau  $k_{Per \times RQ}$ ). On observe également des résultats similaires pour tous les modèles GPR, selon les valeurs de l'*interval score* qui évalue la qualité des prévisions à partir des intervalles de confiance associés. La majorité des données de test se retrouvent dans les intervalles

de confiance associés aux prévisions obtenues par les modèles GPR multi-horizons. En examinant les exemples de prévision du GHI à 10min montrés sur la [Figure 3.13](#), les évolutions temporelles de toutes les prévisions sont semblables. Toutefois, on note que le modèle GPR fondé sur le noyau  $k_{SE-ARD}$  fournit le moins bon résultat. L'analyse des valeurs des critères de performance confirme ce constat. On en déduit que ce noyau, généralement choisi par défaut, ne garantit pas forcément de bons résultats.

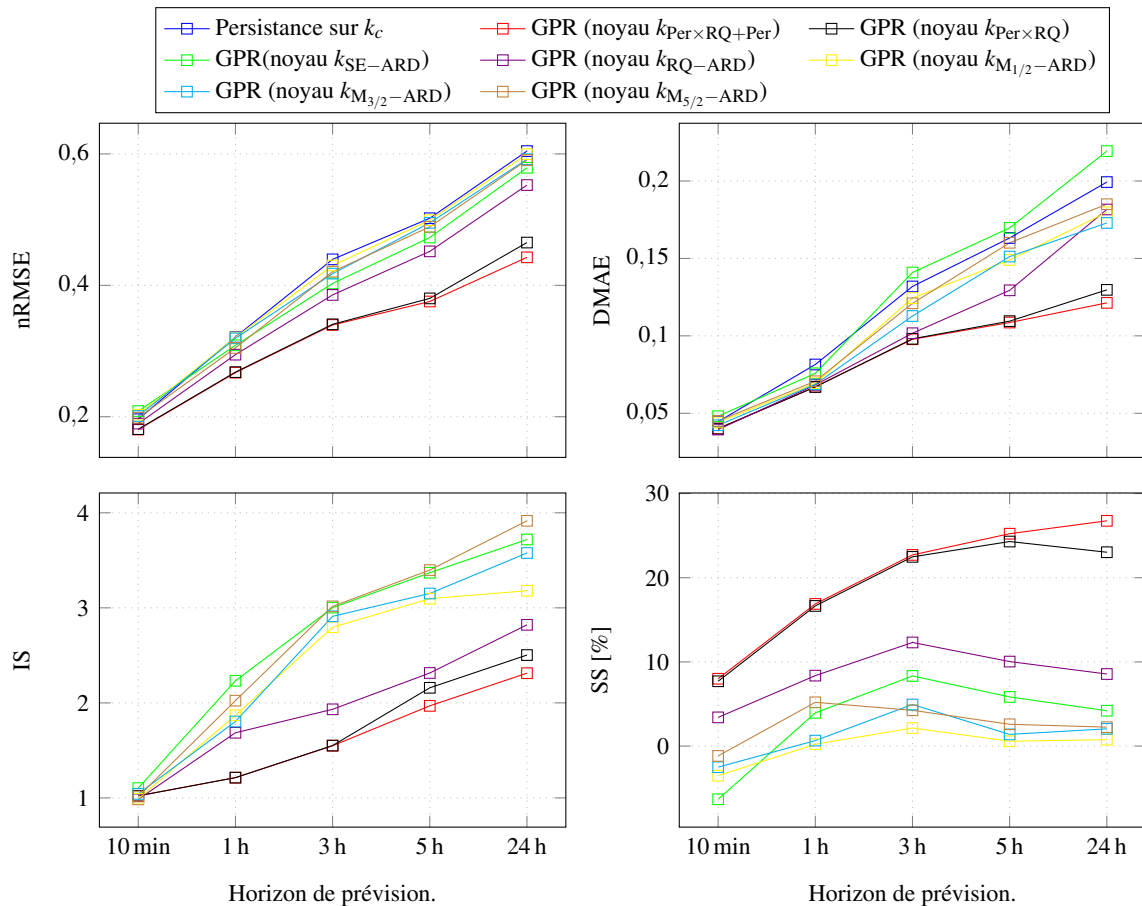


FIGURE 3.10 – Valeurs des critères de performance pour les modèles GPR utilisant le temps en entrée et utilisant des observations du GHI. Ces valeurs sont calculées en considérant toutes les données de test (année 2017).

Lorsque l'horizon de prévision augmente, les performances de tous les modèles se dégradent. On note une forte dégradation des performances du modèle de persistance sur l'indice ciel clair qui produit les moins bons résultats pour tous les horizons infra-journaliers ([Figure 3.10](#)). Bien que la persistance sur le GHI ait été améliorée par l'utilisation d'un modèle « ciel clair », son caractère persistant lui devient préjudiciable lorsque l'horizon de temps augmente. Par ailleurs, à mesure que l'horizon de temps augmente, l'accumulation des erreurs en itérant les prévisions à un pas affecte les performances des modèles GPR multi-horizons utilisant des observations en entrée. Parmi ces modèles, le modèle fondé sur le noyau  $k_{RQ-ARD}$  est celui qui produit les meilleurs résultats aux horizons de temps considérés. Quant aux modèles GPR utilisant le temps en entrée, leur supériorité par rapport aux autres modèles multi-horizons, y compris la persistance sur l'indice ciel clair, devient évidente à mesure que l'horizon de temps augmente. En observant l'évolution temporelle des prévisions du GHI pour des horizons infra-journaliers sur les [Figures 3.14 – 3.17](#), on constate une dégradation rapide des performances des modèles GPR multi-horizons utilisant des observations en entrée qui produisent de nombreuses fluctuations surtout à long terme. Ces fluctuations observées lorsque l'horizon de temps augmente, peinent à expliquer celles du GHI. Concernant les deux modèles GPR utilisant le temps



en entrée, ils produisent des signaux périodiques ajustés aux données à long terme. En effet, lorsque l'horizon de temps augmente, ces modèles convergent vers les connaissances *a priori* sur le GHI formulées par le choix des noyaux. Sur les Figures 3.14 – 3.17, lorsque l'horizon de temps augmente, on constate également des décalages temporels entre les prévisions fournies par la persistance sur l'indice ciel clair et les données de test, ainsi qu'entre le modèle GPR multi-horizon fondé sur le noyau  $k_{SE-ARD}$  et ces mêmes données de test. Ces décalages se retrouvent dans les valeurs de l'erreur DMAE (Figure 3.10). En effet, sur la base de ce critère, la persistance sur l'indice ciel clair et le modèle GPR multi-horizon fondé sur le noyau  $k_{SE-ARD}$  produisent les moins bons résultats pour tous les horizons de temps considérés.

Ces modèles de prévision sont développés dans l'optique de l'implémentation *in situ* d'une stratégie de gestion prévisionnelle d'un réseau électrique de distribution basse tension (projet « Smart Occitania »). La stratégie de contrôle développée présente un caractère prédictif et doit être mise en œuvre en temps réel, avec un pas de temps de 10 min. La prise de décision est fondée sur la connaissance du comportement du système et sur des prévisions de la production PV. Pour ce faire, il faut fournir au contrôleur, à chaque pas de temps, des prévisions sur les 24h à venir à pas de temps de 10 min, ce qui fait en tout 144 horizons de prévision. La prévision du GHI par modèle GPR nécessite l'inversion d'une matrice de dimension égale à la taille des données d'entraînement. En considérant des modèles GPR multi-horizons utilisant des observations du GHI, à chaque pas de temps, il faudra inverser 144 matrices pour obtenir les prévisions (itération des prévisions d'un modèle de prévision un pas), tandis qu'en considérant les modèles multi-horizons utilisant le temps en entrée, il faudra inverser qu'une seule matrice. On précise que le temps d'exécution nécessaire à l'obtention d'une prévision pour les modèles GPR avec les données d'entraînement considérés est d'environ 12s en utilisant un serveur de calcul équipé d'un processeur Intel® Xeon® E7-4890 v2 @ 2,80 GHz.

On conclut que les modèles GPR fondés sur le temps, qui sont non seulement plus performants que les modèles GPR multi-horizons mais ont aussi un coût calculatoire plus faible, sont plus adaptés à notre application. Cependant, il est également possible de développer des modèles GPR utilisant des observations spécifiques pour chaque horizon de temps considéré. Dans la section suivante les modèles GPR utilisant le temps en entrée seront comparés aux modèles GPR mono-horizons.

### 3.4.2 Modèles GPR multi-horizons ou mono-horizons ?

Dans cette section, les modèles GPR multi-horizons utilisant le temps en entrée sont comparés aux modèles GPR mono-horizons utilisant des observations du GHI afin de répondre à la question : *faut-il faire le choix de modèles multi-horizons ou de modèles mono-horizons ?* Les valeurs des critères de performance de ces différents modèles sont présentées sur la Figure 3.11.

Comme on peut le voir sur la Figure 3.11 et sur les Figures 3.13–3.17, les modèles GPR mono-horizons sont les plus performants pour tous les horizons de temps considérés, à l'exception de l'horizon de temps infra-horaire (10min) où les modèles GPR multi-horizons utilisant le temps en entrée prennent légèrement le dessus. Dans le cadre de notre application (projet « Smart occitania »), développer des modèles GPR mono-horizons reviendrait à développer 144 modèles (144 horizons de prévision sont nécessaires). Par ailleurs, pour obtenir une prévision, par exemple, à un horizon de temps de 24h (sachant que pour notre application, nous avons besoin des prévisions avec un pas de temps 10min), à l'aide des modèles GPR mono-horizons, il faudrait inverser 144 matrices. Ces modèles mono-horizons engendrent donc un coût calculatoire très important qui entraverait l'implémentation en temps réel des algorithmes. Par conséquent, dans le cadre de notre application, nous nous tournons vers des modèles GPR multi-horizons utilisant le temps en entrée. Notons toutefois que dans le cas où peu d'horizons de prévision sont considérés ou dès lors qu'il n'y a pas de contrainte temps réel, les modèles GPR mono-horizons sont recommandés.

À la suite des études comparatives des modèles GPR utilisant le temps en entrée et utilisant des observations, on voit que les modèles GPR fondés sur le temps sont plus performants que les modèles GPR multi-horizons utilisant des observations et ont un faible coût calculatoire par rapport aux modèles GPR mono-horizons dans le cadre de notre application. Deux modèles GPR utilisant le temps en entrée

ont été développés : le modèle GPR fondé sur le noyau  $k_{\text{Per} \times \text{RQ} + \text{Per}}$  et le modèle GPR fondé sur le noyau  $k_{\text{Per} \times \text{RQ}}$ . Il convient d'analyser les performances de ces deux modèles afin d'évaluer l'apport du noyau périodique annuel ajouté au noyau  $k_{\text{Per} \times \text{RQ}}$ .

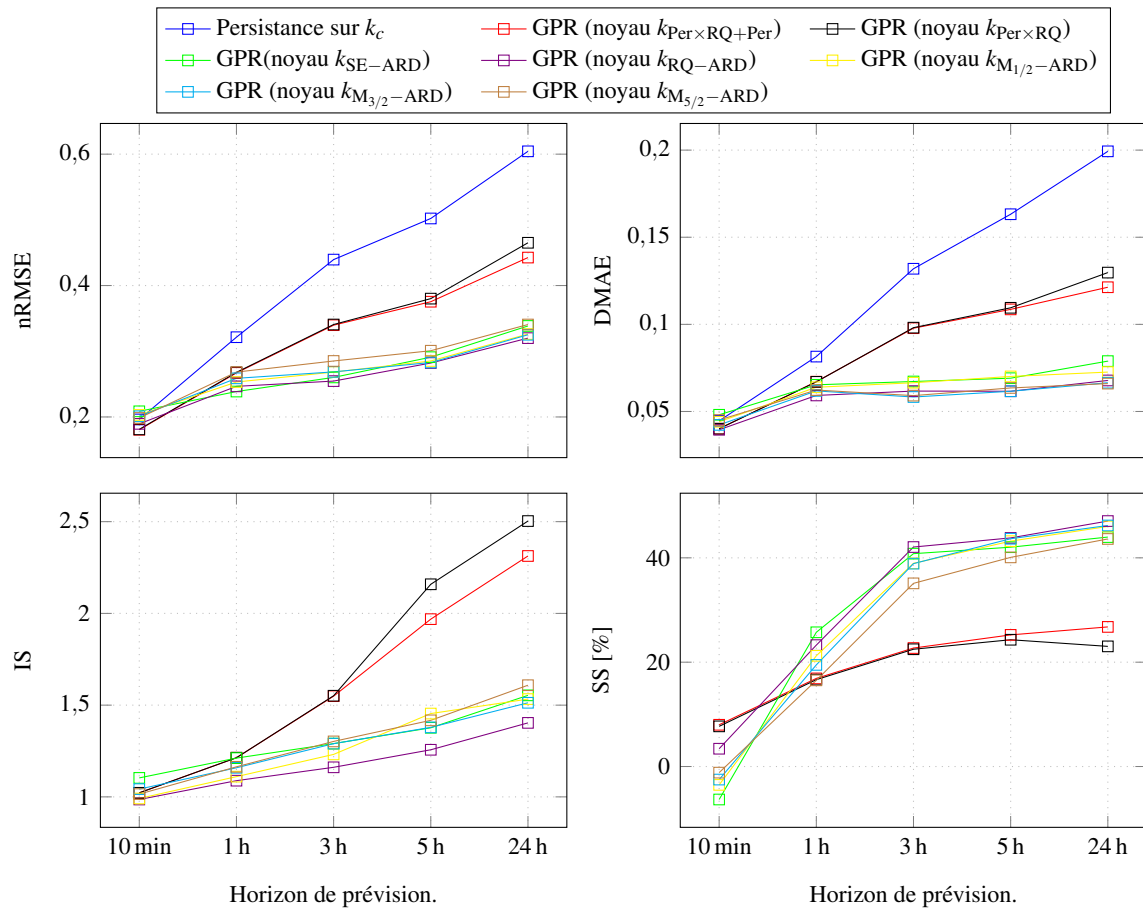


FIGURE 3.11 – Valeurs des critères de performance pour les modèles GPR utilisant le temps en entrée et pour les modèles GPR mono-horizons utilisant des observations du GHI. Ces valeurs sont calculées en considérant toutes les données de test (année 2017).

### 3.4.3 Contribution du noyau périodique annuel

Dans cette section, une étude comparative des modèles GPR fondés sur le temps est menée afin d'évaluer l'apport du noyau périodique annuel. Les valeurs des critères de performance pour ces modèles sont présentées sur la Figure 3.12. Pour un horizon de temps de 10 min, ces deux modèles ont des performances similaires et surpassent légèrement la persistance sur l'indice ciel clair. Les performances des modèles utilisant le temps en entrée (modèles GPR fondés sur le noyau  $k_{\text{Per} \times \text{RQ}}$  et sur le noyau  $k_{\text{Per} \times \text{RQ} + \text{Per}}$ ) restent similaires jusqu'à un horizon de temps de 3 h, mais à partir de 5 h on constate une légère supériorité du modèle GPR fondé sur le noyau  $k_{\text{Per} \times \text{RQ} + \text{Per}}$ . On note là l'importance du noyau modélisant la composante saisonnière du GHI qui améliore sensiblement les performances du noyau  $k_{\text{Per} \times \text{RQ}}$ , surtout à long terme. L'addition d'un noyau périodique modélisant la composante annuelle du GHI complexifie le modèle GPR. Cependant, le modèle fondé sur le noyau  $k_{\text{Per} \times \text{RQ} + \text{Per}}$  ne se détache pas clairement du modèle GPR fondé sur le noyau  $k_{\text{Per} \times \text{RQ}}$ . La légère supériorité du modèle GPR fondé sur le noyau  $k_{\text{Per} \times \text{RQ} + \text{Per}}$  nous indique que le gain associé au noyau périodique annuel pourrait être plus important pour des horizons de plusieurs jours, ou dans le cas où des données couvrant plusieurs années sont considérées. Lorsque l'horizon de temps augmente, l'augmentation des valeurs du critère IS (Figure 3.12) des modèles GPR multi-horizons témoigne de la dégradation de



leurs performances. Néanmoins, parmi les modèles multi-horizons, le modèle GPR fondé sur le noyau  $k_{\text{Per} \times \text{RQ} + \text{Per}}$  utilisant le temps en entrée produit les meilleurs résultats pour tous les horizons de temps considérés. Sur la Figure 3.17, pour un horizon de 24h, le modèle GPR fondé sur le noyau  $k_{\text{Per} \times \text{RQ} + \text{Per}}$  s'adapte mieux aux données. Le noyau périodique modélisant la composante annuelle du GHI permet à ce modèle de produire un signal périodique avec une bonne amplitude à long terme.

## NOTE

Sans données pour mettre à jour les prévisions, un modèle GPR donne simplement la valeur moyenne apprise lors de l'entraînement. Pour un modèle avec un noyau quasi-périodique, cette valeur moyenne est un signal périodique (un modèle ciel clair, en quelque sorte). Ainsi, si on demandait une prévision à un horizon très long (quelques mois par exemple), le modèle GPR fondé sur le noyau  $k_{\text{Per} \times \text{RQ} + \text{Per}}$  serait largement avantage, puisque son « modèle ciel clair » inclut également une composante périodique annuelle.  $\diamond$

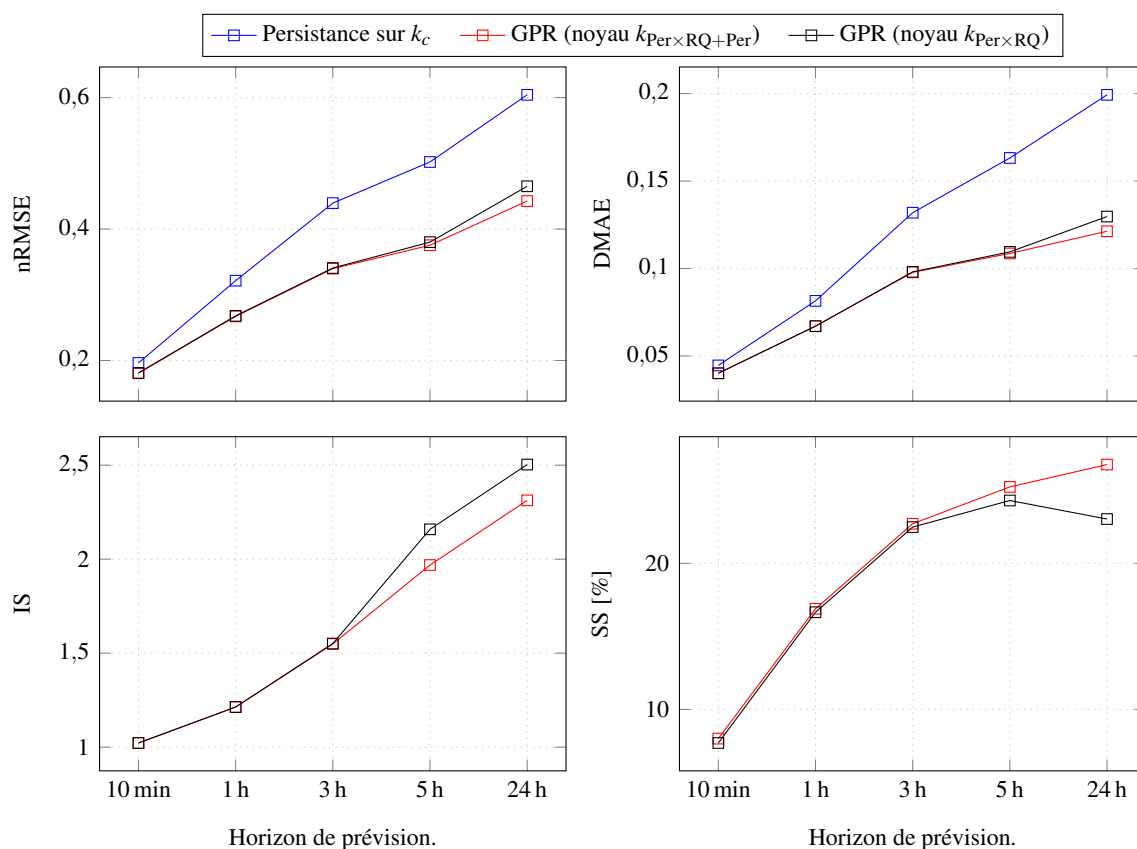


FIGURE 3.12 – Valeurs des critères de performance pour les modèles GPR utilisant le temps en entrée. Ces valeurs sont calculées en considérant toutes les données de test (année 2017).

Si le nombre d'horizons de temps est limité, les modèles GPR mono-horizons sont les plus adaptés, mais si le nombre d'horizon est important, il vaut mieux se tourner vers les modèles GPR multi-horizons et, parmi ces modèles, ceux utilisant le temps en entrée sont les plus appropriés. Pour notre application où 144 horizons de prévision sont considérés, notre choix se porte donc sur les modèles GPR utilisant le temps en entrée. Il faut aussi noter que les deux modèles utilisant le temps en entrée donnent des résultats similaires pour des horizons infra-horaires et infra-journaliers mais, pour un horizon de 24h, on note une légère supériorité du modèle GPR fondé sur le noyau  $k_{\text{Per} \times \text{RQ} + \text{Per}}$ . On en déduit que si l'on dispose de données couvrant quelques semaines, le noyau  $k_{\text{Per} \times \text{RQ}}$  est le meilleur choix, mais pour des données couvrant des années, le noyau  $k_{\text{Per} \times \text{RQ} + \text{Per}}$  est le plus adapté.

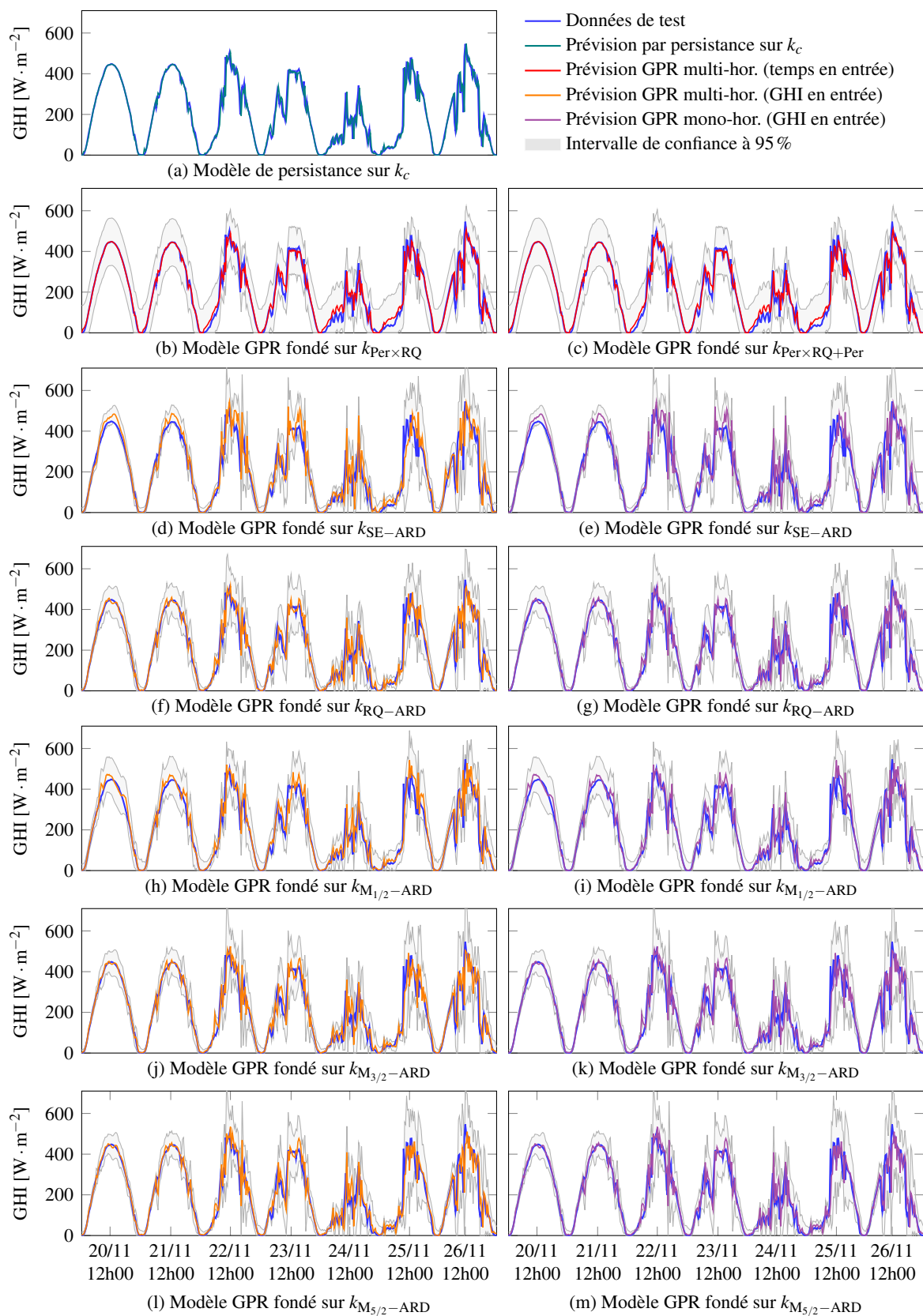


FIGURE 3.13 – Exemples de prévision du GHI à 10 min sur sept jours avec le modèle de persistance sur l'indice ciel clair et les modèles GPR.

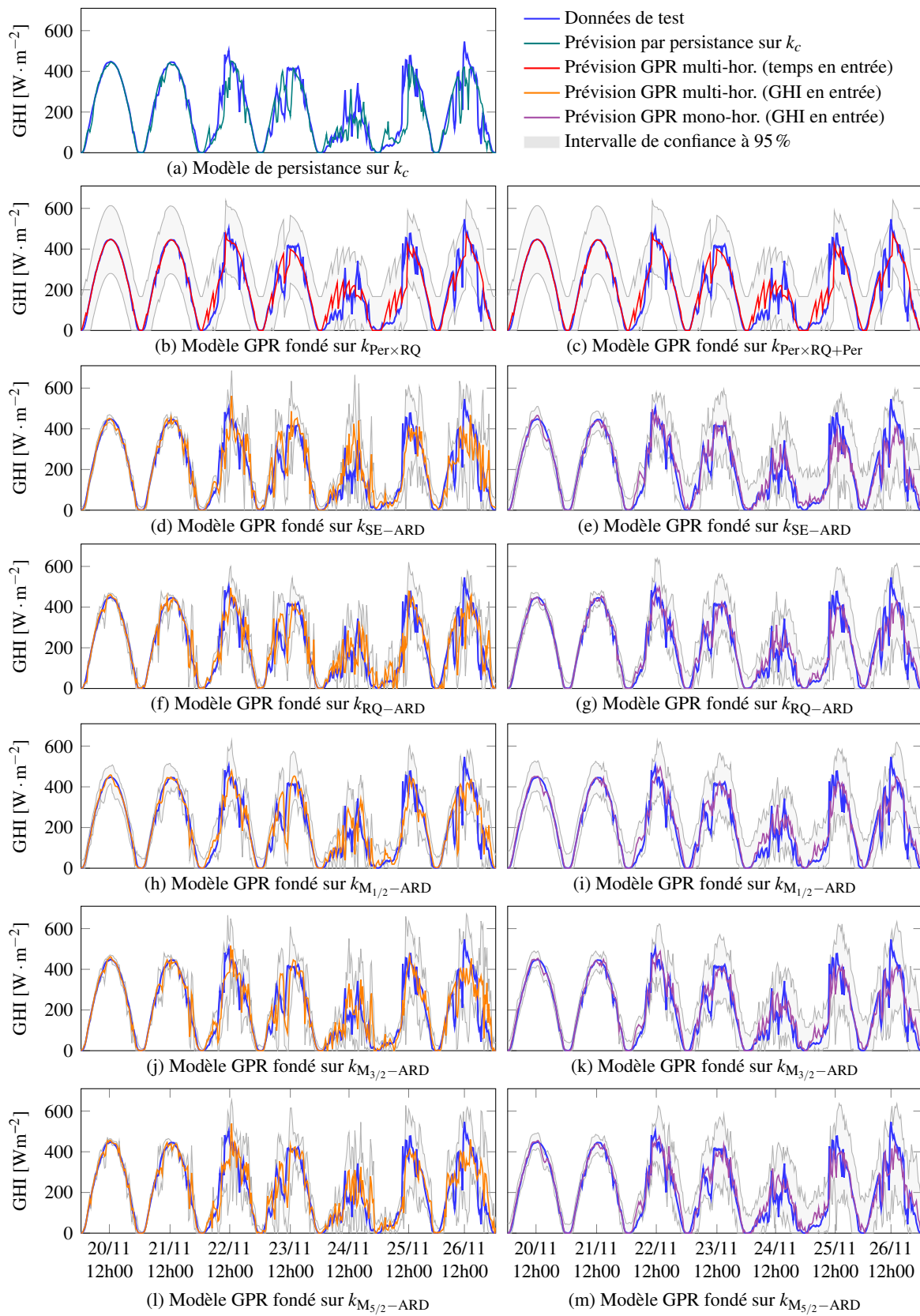


FIGURE 3.14 – Exemples de prévision du GHI à 1 h sur sept jours avec le modèle de persistance sur l'indice ciel clair et les modèles GPR.

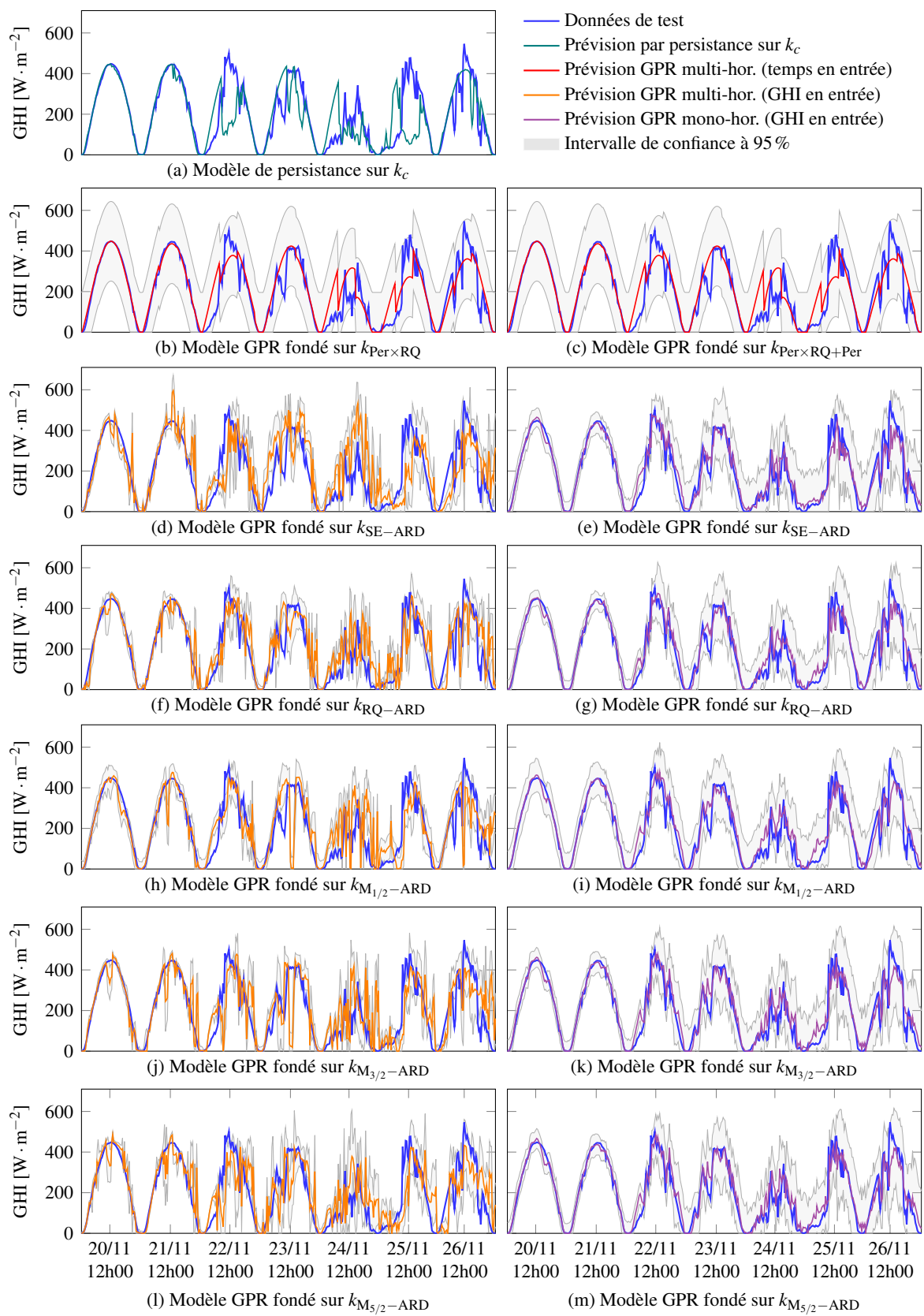


FIGURE 3.15 – Exemples de prévision du GHI à 3 h sur sept jours avec le modèle de persistance sur l'indice ciel clair et les modèles GPR.

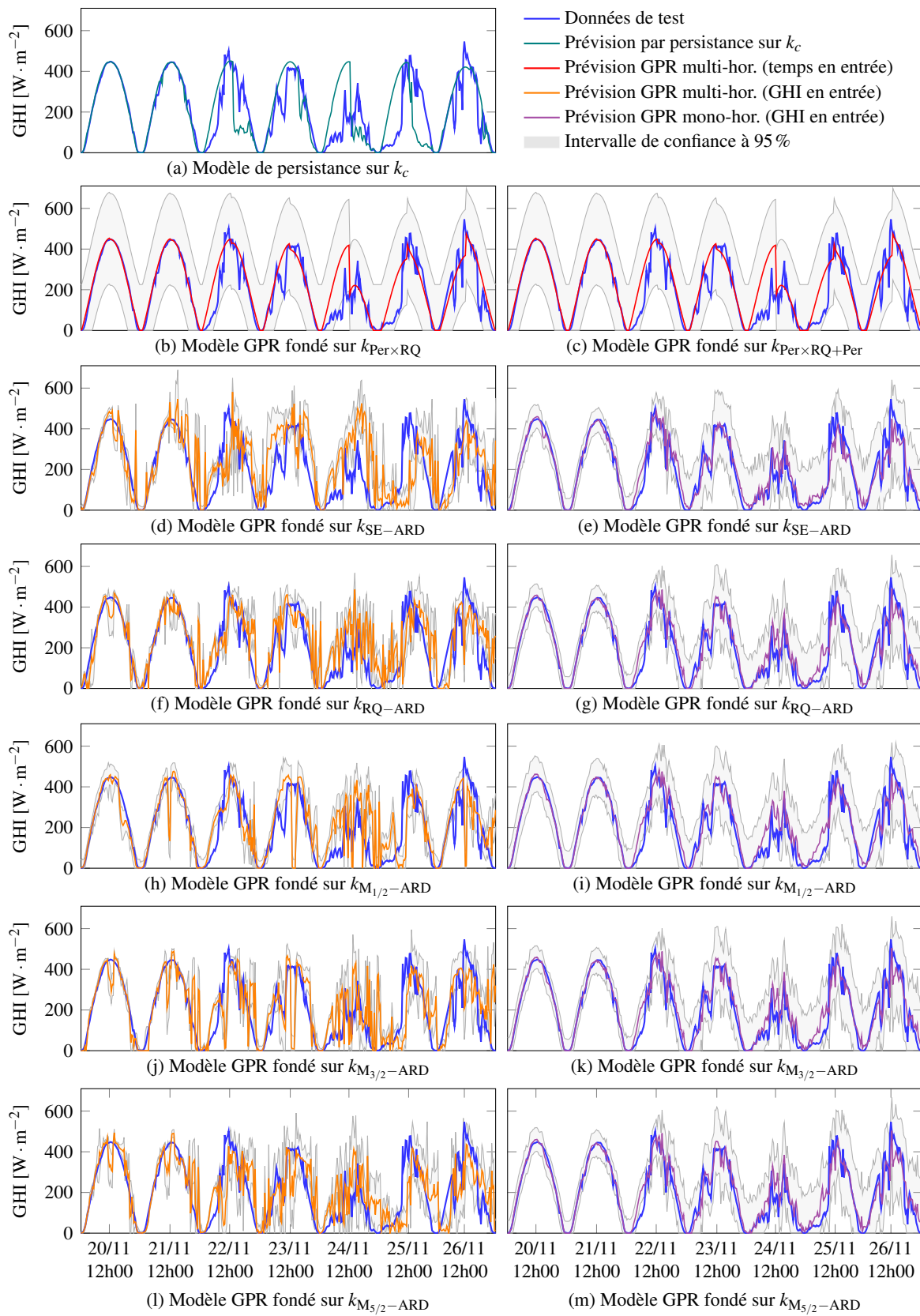


FIGURE 3.16 – Exemples de prévision du GHI à 5 h sur sept jours avec le modèle de persistance sur l'indice ciel clair et les modèles GPR.



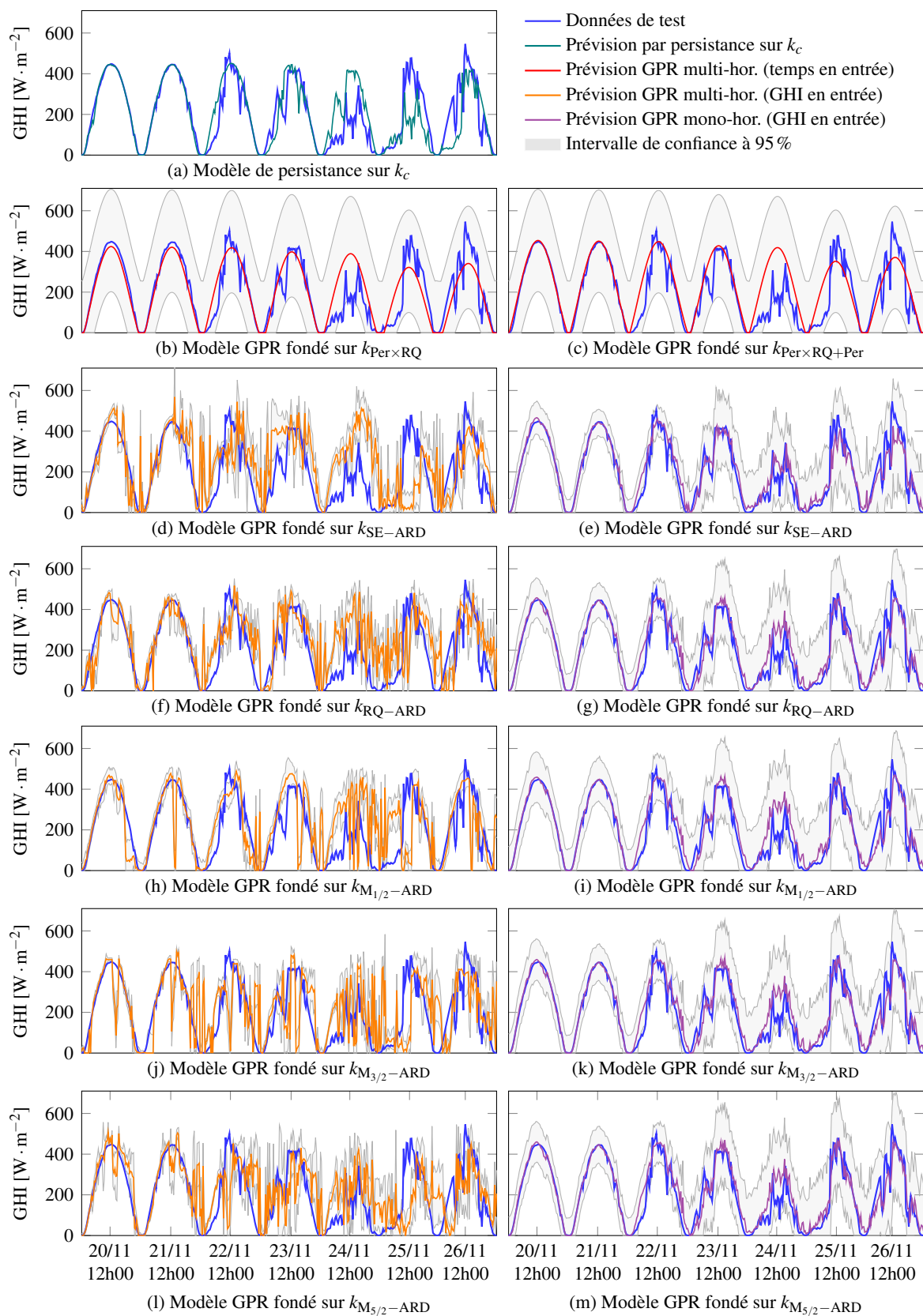


FIGURE 3.17 – Exemples de prévision du GHI à 24 h sur sept jours avec le modèle de persistance sur l'indice ciel clair et les modèles GPR.

### 3.5 Conclusion

Le [Chapitre 1](#) a permis de choisir la régression non paramétrique par processus gaussien comme outil de prévision du GHI pour notre application (projet ADEME « Smart Occitania »). Les modèles GPR peuvent être développés en utilisant soit le temps en entrée, soit des observations. Ce chapitre a tranché sur l'entrée à utiliser afin d'obtenir de bons résultats tout en prenant en compte les exigences de notre application. Une étude comparative entre les modèles GPR utilisant le temps en entrée et ceux utilisant des observations a été menée. Les modèles GPR utilisant le temps en entrée sont naturellement des modèles de prévision multi-horizons, tandis que parmi ceux utilisant des observations, on distingue des modèles multi-horizons et des modèles mono-horizons.

Les modèles mono-horizons produisent de meilleurs résultats que les modèles multi-horizons. Cependant, développer des modèles mono-horizons est contraignant en termes de coût calculatoire dès lors que plusieurs horizons sont considérés, comme dans notre cas où 144 horizons sont considérés. Notre choix s'est par conséquent porté sur des modèles multi-horizons et parmi ces modèles, le modèle GPR fondé sur le noyau  $k_{\text{Per} \times \text{RQ} + \text{Per}}$  utilisant le temps en entrée est celui qui produit les meilleurs résultats. Par ailleurs, pour développer des modèles mono-horizons, les horizons de temps doivent être connus à l'avance, alors que modèle GPR fondé sur le noyau  $k_{\text{Per} \times \text{RQ} + \text{Per}}$ , une fois entraîné, peut fournir des prévisions à n'importe quel horizon de temps. Toutefois, en l'absence d'un jeu de données couvrant des années, le noyau  $k_{\text{Per} \times \text{RQ}}$  est recommandé.

Dans la mesure où le jeu de données utilisé pour la prévision spatio-temporelle du GHI ne couvre que quelques semaines, le noyau dédié à la prévision temporelle du GHI qui sera associé à un noyau spatial est le noyau  $k_{\text{Per} \times \text{RQ}}$  ([Chapitre 4](#)).

## Chapitre 4

# Vers la prévision spatio-temporelle de l'éclairement global horizontal par régression non paramétrique par processus gaussien

### 4.1 Introduction

Dans le chapitre précédent, une étude comparative de modèles GPR utilisant en entrée le temps ou des observations a été menée afin de trancher sur l'entrée à utiliser pour obtenir de bonnes prévisions du GHI. Les modèles utilisant le temps en entrée, par nature multi-horizons, se sont avérés les plus adaptés à la prévision du GHI, à l'aune des exigences de notre application. En effet, prévoir le GHI a pour finalité de prévoir la production PV et, dans le cadre du projet « Smart Occitania », porté par le gestionnaire du réseau électrique de distribution français ENEDIS, le développement d'algorithmes pour la gestion prévisionnelle du réseau électrique de distribution basse tension en région Occitanie. Ces algorithmes, dont le coût calculatoire est maîtrisé dans le but de pouvoir les implémenter en temps réel, favorisent la pénétration d'une production décentralisée (en particulier une production solaire photovoltaïque) sans construction de nouvelles infrastructures, tout en garantissant une qualité de service suffisante. Nous avons également vu qu'en considérant le temps comme entrée des modèles GPR pour la prévision du GHI, on tire profit des biais inductifs connus sur son évolution au cours du temps pour construire un noyau approprié. Ainsi, un noyau adapté à la prévision localisée du GHI a été identifié : le noyau  $k_{\text{Per} \times \text{RQ}}$ , recommandé en l'absence d'un jeu de données couvrant une période suffisamment longue (plus d'une année). Dans le cas contraire, il est possible de compléter ce noyau par une composante périodique annuelle, en considérant le noyau  $k_{\text{Per} \times \text{RQ} + \text{Per}}$ , ce qui permet une légère amélioration des résultats.

Dans ce chapitre, nous cherchons à identifier un noyau spatio-temporel dans le but de développer des modèles de prévision fondés sur la régression non paramétrique par processus gaussien. Ce travail n'est qu'un petit pas vers la prévision dans le temps et dans l'espace du GHI. Pour le développement des modèles GPR de prévision spatio-temporels, PROMES-CNRS a instrumenté une zone résidentielle par un réseau de pyranomètres et une caméra à grand angle de champ à Villeneuve-de-la-Raho (voir la [Sous-sous-section 1.2.6.3](#) du [Chapitre 1](#)). À partir de la disponibilité et de la variabilité du GHI dans le temps et dans l'espace, il est possible de déduire la production électrique des installations photovoltaïques dans la zone résidentielle. Les modèles de prévision spatio-temporels existant dans la littérature scientifique sont passés en revue dans la [Section 4.2](#). L'instrumentation installée au sein du quartier résidentiel à Villeneuve-de-la-Raho est présentée dans la [Section 4.3](#). La [Section 4.4](#) présente un processus gaussien spatio-temporel et les hypothèses posées lors de la construction des noyaux spatio-temporels. La [Section 4.5](#) est consacrée à l'analyse des résultats.



## 4.2 État de l'art

Un processus spatio-temporel  $\{z(s, t) : s \in D_s \subset \mathbb{R}^d, t \in D_t \subset \mathbb{R}\}$  définit un phénomène qui évolue dans le temps et dans l'espace. De tels processus sont présents dans de nombreux domaines des sciences de l'environnement telles que la biologie, la géophysique, l'hydrologie et la géologie [186]. Comme exemples, nous pouvons citer les concentrations de polluants atmosphériques, les recueils météorologiques, les données de précipitations, les paramètres hydrologiques et la ressource solaire qui sont caractérisés par des variabilités spatiale et temporelle. La modélisation spatio-temporelle est une étape importante pour comprendre l'évolution de ces phénomènes. Les méthodes permettant de décrire les mécanismes régissant les processus spatio-temporels se regroupent en deux grandes catégories dans la littérature scientifique : les méthodes physiques et les méthodes statistiques. Ces méthodes sont passées en revue dans cette section.

### 4.2.1 Les méthodes physiques

Les modèles physiques reposent sur des lois ayant une signification physique explicite. Par exemple, il existe le modèle EPIC (*environmental policy integrated climate*) qui peut simuler la croissance des cultures en fonction de l'état du sol, des informations météorologiques, de l'irrigation et d'autres données relatives aux cultures [187]. En fonction des conditions météorologiques prévues pour les prochaines journées ou prochains mois, ce modèle peut prévoir le rendement des cultures pour la prochaine saison. Pour la prévision spatio-temporelle de la ressource solaire, les méthodes physiques généralement utilisés reposent sur les modèles numériques de prévision météorologique (NWP) [186, 188, 189].

Les modèles NWP utilisent les lois physiques gouvernant les changements d'état de l'atmosphère pour prévoir l'évolution des facteurs météorologiques. Leur précision est liée à la bonne connaissance de l'état de l'atmosphère et à la structure des modèles physiques sous-jacents. Les conditions initiales ne sont souvent pas optimales en raison d'erreurs de mesure. Cependant, elles peuvent être affinées à l'aide d'observations par des techniques d'assimilation de données [190]. La connaissance actuellement limitée des mécanismes derrière les changements d'état de l'atmosphère entraîne souvent des biais dans les résultats de prévision. Les modèles NWP ont connu des progrès, principalement grâce à l'amélioration des techniques de résolution des équations mathématiques sous-jacentes [189]. Un autre progrès concerne l'amélioration de leur résolution spatiale. Ces progrès vont de pair avec une grande disponibilité de données et des moyens de calcul plus conséquents. Cette évolution est décrite comme la révolution tranquille des modèles NWP dans [188]. Cependant, le coût calculatoire de ces modèles reste toujours très élevé et est un frein à leur mise en œuvre dans un contexte de prévision infra-horaire [189]. À cet égard, les modèles statistiques présentent un avantage car ils peuvent fournir des prévisions sans avoir à résoudre des équations mathématiques complexes.

### 4.2.2 Les méthodes statistiques

De nombreux modèles spatio-temporels sont fondés sur les corrélations entre les différentes variables observées, ce qui promeut l'application de modèles statistiques. Récemment, les chercheurs se sont tournés vers la statistique spatio-temporelle pour la modélisation d'un processus spatio-temporel [191, 192]. Les méthodes statistiques regroupent les approches classiques de modélisation de séries temporelles telles que les modèles autorégressifs spatio-temporels et les méthodes géostatistiques telles que le krigeage [186]. En général, les modèles statistiques sont plus faciles à mettre en œuvre et permettent de s'affranchir du problème de complexité informatique qui affecte les modèles physiques.

#### 4.2.2.1 Les approches classiques

Le modèle vectoriel autorégressif (ou VAR, pour *vector autoregressive*), qui est une extension du modèle autorégressif univarié (ou AR, pour *autoregressive*), est l'un des modèles les plus simples à

utiliser pour l'analyse de séries temporelles multivariées. Il s'est avéré particulièrement adapté à la modélisation du comportement dynamique des séries temporelles dans le domaine des finances. Il fournit de meilleures prévisions que les modèles AR. Le modèle VAR peut également être utilisé pour fournir des prévisions dans le temps en tenant compte des dépendances spatio-temporelles entre les données. Dans [193], le modèle VAR spatio-temporel a été utilisé pour la prévision de concentrations en monoxyde de carbone enregistrées sur quatre sites et pour la prévision de la vitesse du vent sur onze sites. Les résultats ont montré que ce modèle surpasse le modèle de persistance et le modèle AR. Pour la prévision du vent, les auteurs ont également noté une meilleure performance du modèle VAR, surtout aux endroits où les capteurs sont alignés selon la direction du vent [193]. Dans [194], le modèle VAR a été utilisé pour la prévision du GHI à un horizon de temps de 5 min à partir de mesures provenant de treize capteurs. Les résultats ont montré une meilleure performance de ce modèle par rapport à la persistance. La théorie autour du modèle VAR et de sa variante VARMA (*vector autoregressive moving average*), qui est une extension du modèle ARMA, peut être trouvée dans [195].

Des cas particuliers des modèles VARMA tels que le modèle STARMA (*space-time autoregressive moving average*) et sa variante STARIMA (*space-time autoregressive integrated moving average*), proposés par [196, 197, 198, 199, 200], sont également utilisés pour la prévision spatio-temporelle. Dans [201], le modèle STARMA a été utilisé pour évaluer la variabilité de la production électrique d'un réseau de panneaux solaires photovoltaïques. Les auteurs ont modélisé le GHI en considérant une grille carrée de  $10\text{ km} \times 10\text{ km}$  en utilisant des données provenant de dix sites [201]. Les paramètres des modèles ne sont pas liés à un site spécifique, mais dépendent simplement des vecteurs de distances entre les sites, permettant ainsi de définir des matrices de dépendance structurées [202]. Ils ont constaté qu'un processus autorégressif spatio-temporel du premier ordre fournit un ajustement adéquat aux données. Dans [203], les auteurs ont proposé une nouvelle approche basée sur le modèle STARIMA pour la prévision du GHI à partir de mesures et d'images satellitaires. L'intérêt majeur de cette étude se trouve dans la combinaison d'informations temporelles avec les informations spatiales des pixels HelioClim-3 entourant le site de mesure. Les cartes HelioClim-3 fournissent une grande quantité d'informations spatio-temporelles selon la région traitée dans l'étude. Les auteurs ont donc créé une stratégie de sélection de la région à étudier permettant d'obtenir les paramètres spatiaux à inclure dans les données d'entrée du modèle. Les résultats ont montré que l'intégration d'informations spatio-temporelles provenant d'images satellitaires dans le modèle STARIMA améliore les résultats [203].

Ces familles de modèles VAR, VARMA et VARIMA ne permettent d'avoir des prévisions qu'en des sites où il existe des mesures. Elles relient simplement la valeur actuelle de chaque site aux observations passées des sites voisins. De plus, elles n'arrivent à capter que les relations linéaires entre les données. Par ailleurs, ces méthodes sont généralement limitées à la modélisation d'un faible volume de données. Dans le cas de la prévision de la ressource solaire, des prévisions hors site d'observation peuvent être importantes pour une meilleure gestion de l'intégration de la production solaire photovoltaïque dans un réseau électrique de distribution. D'autres modèles spatio-temporels tels que le krigeage sont alors envisageables. Le krigeage est un outil adapté à la prévision spatio-temporelle, si l'on veut connaître des valeurs du phénomène en des sites non observés.

#### 4.2.2.2 Le krigeage

La régression non paramétrique par processus gaussien est largement utilisée en géostatistique pour modéliser des fonctions à entrée bidimensionnelle ou tridimensionnelle où elle est connue sous le nom de krigeage. Le krigeage est une méthode d'interpolation spatiale qui tient compte de la structure spatiale des données. Le Krigeage produit de meilleurs résultats que les méthodes déterministes d'interpolation spatiale qui se basent généralement sur des propriétés géométriques, sans tenir compte du phénomène physique étudié [204]. Il doit son nom au chercheur français Matheron [146] qui a défini le formalisme mathématique de cette méthode d'interpolation spatiale. Matheron s'est sans doute inspiré des travaux de l'ingénieur minier sud-africain Krige [205] qui a été le précurseur de cette méthode. D'autres chercheurs en météorologie se sont également inspirés des travaux de Krige pour établir les fondements théoriques de cette méthode. Cependant, le formalisme proposé par Matheron

est le plus adopté par la communauté scientifique [206]. Contrairement aux approches classiques citées ci-avant, le krigeage prend en compte la structure de dépendance spatiale des données et permet de prévoir la valeur de la variable régionalisée étudiée en un site non échantillonné. Le modèle de base du krigeage a recours au même formalisme qu'un modèle GPR, c'est-à-dire qu'il est défini par une fonction moyenne et une fonction de covariance ou noyau. Des exemples de travaux traitant de la prévision spatiale du GHI par krigeage peuvent être trouvés dans [207, 208, 209, 210]. Dans ces travaux, on note que le krigeage a donné de bons résultats à la fois pour des sites observés et pour les sites non observés. Ces études ont été réalisées dans divers pays soumis à différents climats et les auteurs arrivent à la même conclusion que le krigeage reste le meilleur outil d'interpolation spatiale du GHI.

Le krigeage spatio-temporel est la généralisation du krigeage traditionnel spatial à des données spatio-temporelles. Sa capacité à prévoir en des sites non observés apporte des solutions à de nombreux problèmes dans divers domaines [211]. Le krigeage spatio-temporel est fréquemment utilisé dans diverses disciplines tels que l'agriculture [212], la recherche environnementale [213, 214], la géologie [215] et la géostatistique [216]. Le krigeage spatio-temporel est donc l'outil approprié à la prévision de la ressource solaire, surtout lorsque des prévisions en des sites non observés sont nécessaires. Les méthodes de krigeage spatio-temporel développées pour la prévision de la ressource solaire sont passées en revue dans [217]. Diverses formes de krigeage spatio-temporel sont utilisées pour la prévision de la ressource solaire [217] :

- le krigeage simple qui suppose que la moyenne du processus dans l'espace et dans le temps est connue ;
- le krigeage ordinaire qui suppose que la moyenne du processus est inconnue, mais constante ;
- le krigeage universel qui considère que la moyenne du processus inconnue est une fonction des variables d'entrée, par exemple, la longitude et la latitude.

Les corrélations dans l'espace et dans le temps sont modélisées par une fonction de covariance spatio-temporelle et les paramètres sont généralement déterminés par maximisation de la vraisemblance [218]. Dans la construction du noyau, certaines hypothèses simplificatrices, telles que la stationnarité, la séparabilité et l'isotropie sont souvent considérées. Par analogie aux processus gaussiens, la fonction de covariance spatio-temporelle a un impact crucial sur les performances du modèle. Les noyaux spatio-temporels existants sont passés en revue dans [219]. Cependant, la plupart des études sur ces fonctions de covariance restent purement théoriques. En pratique, pour modéliser les processus spatio-temporels, la plupart des auteurs considèrent des noyaux séparables, car ces structures de covariance sont faciles à utiliser. Néanmoins, ces noyaux peuvent ne pas modéliser correctement certaines données météorologiques telles que la vitesse du vent [220, 221].

Les fonctions de covariance non séparables et non stationnaires ont été étudiées par quelques chercheurs [222, 223, 220, 224, 221, 225]. Des travaux sur la prévision spatio-temporelle du GHI par krigeage spatio-temporel peuvent être trouvés dans [226, 227]. Étant donné que la principale source de variabilité spatio-temporelle de la ressource solaire est liée au passage des nuages au-dessus de zone étudiée, des noyaux spatio-temporels anisotropes prenant en compte le déplacement des nuages ont été développés dans la littérature. Dans [220], une fonction de covariance lagrangienne anisotrope a été développée pour la prévision de données météorologiques. Il a été démontré que le noyau lagrangien améliore la prévision de données météorologiques, telle que la vitesse du vent [220] et la ressource solaire [227]. La fonction de covariance lagrangienne a été utilisée dans de nombreuses études sur la prévision de la ressource solaire pour tenir compte de l'anisotropie due aux mouvements des nuages [228, 229, 230]. Dans [228], un noyau obtenu par produit d'un noyau lagrangien et d'un noyau temporel a été considéré. Dans [229] et [230], un noyau similaire, exprimé comme produit d'une fonction de covariance lagrangienne (noyau lagrangien proposé par [231] en supposant une vitesse de déplacement des nuages constante) et d'un noyau temporel, a été considéré. Cependant, leur méthode de krigeage spatio-temporel anisotrope a conduit à de moins bons résultats de prévision sur certains jours que la méthode de krigeage fondée sur un noyau spatial simple. En effet, les noyaux spatio-temporels anisotropes proposés par [228, 229, 230] sont réduits à des noyaux séparables lorsque

le terme lagrangien est négligeable, ce qui se produit lorsque le mouvement des nuages est négligeable ou si les nuages varient rapidement de sorte que leur direction principale n'est pas détectable et, par conséquent, la fonction de covariance est pratiquement réduite à un noyau isotrope. Dans ces situations, le modèle anisotrope proposé par [228, 229, 230] est moins performant, probablement parce que l'hypothèse de séparabilité est très discutable. Dans [226], un noyau spatio-temporel stationnaire isotrope – la stationnarité et l'isotropie spatiales ont été obtenues par des déformations de l'espace géographique [232] – a été considéré pour la prévision du GHI.

Un défi dans l'application du krigeage spatio-temporel, surtout celui anisotrope, à la prévision de la ressource solaire est la maîtrise du coût calculatoire. Pour palier cet inconvénient, les auteurs quantifient les distances de décorrélation spatiale et temporelle [228, 233, 234, 235]. La détermination de la zone de corrélation spatiale et temporelle permet de réduire le coût calculatoire, car pour la prévision en un point, seules les données situées dans la zone de corrélation sont introduites dans le modèle de krigeage. L'état de l'art de l'estimation des distances de décorrélation peut être trouvé dans [217]. Dans [204], les corrélations spatio-temporelles des données GHI sont analysées et le krigeage spatio-temporel a été utilisé pour la modélisation du GHI. Un noyau spatio-temporel non séparable anisotrope a été proposé et une nouvelle approche d'estimation des distances de décorrélation spatiale et temporelle a été développée. Cette approche consiste à calculer des longueurs de décorrélation distinctes dans la direction du mouvement des nuages ainsi que dans d'autres directions, y compris la direction du vent. Les résultats ont montré une supériorité du krigeage spatio-temporel anisotrope par rapport au modèle de persistance [204].

Dans le cadre du projet « Smart Occitania », les horizons de prévision nécessaires vont de l'infra-horaire (10min) au journalier (24h). À travers ce projet, le gestionnaire du réseau électrique de distribution français ENEDIS souhaite disposer de solutions logicielles à coût calculatoire maîtrisé garantissant une qualité de service suffisante et favorisant la pénétration au sein des réseaux électriques de distribution basse tension d'une production décentralisée (en particulier une production solaire photovoltaïque) diffuse et intermittente. Ainsi, les prévisions dans le temps et dans l'espace du GHI – les prévisions de la production électrique des installations solaires photovoltaïques sont obtenues à partir de prévisions du GHI – en des sites de mesure et en des sites non échantillonnés de la zone résidentielle (Villeneuve-de-la-Raho) servant de cas d'étude sont importantes afin de déduire la production électrique des installations solaires photovoltaïques dans la zone résidentielle et d'anticiper un déploiement massif de la production PV. Les modèles physiques dont le coût calculatoire reste toujours très élevé malgré leurs progrès ces dernières années [189] ne peuvent pas être utilisés dans notre cas. Contrairement à ces modèles, les modèles statistiques ont l'avantage de pouvoir fournir des prévisions à divers horizons de temps sans avoir à résoudre des équations mathématiques complexes. Les modèles statistiques regroupent les approches classiques (les modèles VAR, VARMA, VARIMA et leurs dérivés) et le krigeage. Les approches classiques ne permettent d'avoir des prévisions qu'en des sites de mesure. Par conséquent, dans cette thèse, le krigeage spatio-temporel ayant prouvé son efficacité pour la prévision spatio-temporelle dans divers domaines, que ce soit en des sites de mesure ou en des sites non observés, sera utilisé pour la prévision dans le temps et dans l'espace du GHI. Divers noyaux spatio-temporels non-séparables et anisotropes prenant en compte le déplacement des nuages ont développés dans la littérature scientifique pour la prévision spatio-temporelle de la ressource solaire [228, 229, 230]. Ces noyaux se réduisent à des fonctions de covariance séparables lorsque le mouvement des nuages est négligeable ou si les nuages varient rapidement de sorte que leur direction principale n'est pas détectable. Les mesures du GHI issues du quartier résidentiel à Villeneuve-de-la-Raho sont échantillonnées avec un pas de temps de 10min. Ce pas de temps limite la capacité d'appréhender correctement l'influence des perturbations atmosphériques sur la variabilité du GHI dans la zone considérée (voir la [Section 4.3](#)). Ainsi, pour la prévision dans le temps et dans l'espace du GHI, un noyau séparable formé par produit d'un noyau spatial à identifier et du noyau temporel identifié au [Chapitre 3](#) sera considéré. Même si l'hypothèse de séparabilité reste discutable, elle permet de réduire le coût calculatoire et d'avoir de premiers résultats. De plus, dans les travaux sur la prévision de la ressource solaire par krigeage spatio-temporel, le noyau temporel considéré dans

l'élaboration du noyau spatio-temporel est généralement un noyau exponentiel quadratique. Dans cette thèse, un noyau temporel particulièrement adapté au GHI est utilisé.

### 4.3 Instrumentation et données disponibles

Dans le cadre du projet « Smart Occitania », le gestionnaire du réseau électrique de distribution français ENEDIS et le laboratoire PROMES-CNRS ont fait le choix d'une zone résidentielle d'environ 120 habitations (avec 11 producteurs PV) à quelques kilomètres de Perpignan, à Villeneuve-de-la-Raho (voir la [Sous-sous-section 1.2.6.3 du Chapitre 1](#)), pour la validation des algorithmes développés. PROMES-CNRS ayant fait le choix d'une prévision de la production électrique des installations solaires photovoltaïques à partir de prévisions de l'éclairement global horizontal, la zone résidentielle a été instrumentée avec un réseau de 16 pyranomètres SP Lite 2 de la marque Kipp & Zonen, afin d'obtenir une mesure distribuée du GHI, et une caméra à grand angle de champ fournissant des observations atmosphériques. Le [Tableau 4.1](#) répertorie les coordonnées GPS des pyranomètres installés au sein et aux alentours de la zone résidentielle (voir [Figure 4.1](#)).

TABLEAU 4.1 – Coordonnées GPS des 16 pyranomètres installés à Villeneuve-de-la-Raho.

Pyranomètre	Latitude	Longitude
A	42,640 806° N	2,909 631° E
B	42,642 661° N	2,912 237° E
C	42,638 505° N	2,915 209° E
D	42,640 678° N	2,912 036° E
E	42,631 491° N	2,915 837° E
F	42,634 169° N	2,918 813° E
G	42,639 929° N	2,920 987° E
H	42,637 320° N	2,923 010° E
I	42,633 142° N	2,922 450° E
J	42,643 701° N	2,882 073° E
K	42,654 698° N	2,894 584° E
L	42,653 674° N	2,901 226° E
M	42,657 415° N	2,914 155° E
N	42,636 605° N	2,897 229° E
O	42,643 064° N	2,900 490° E
P	42,640 715° N	2,905 635° E

La [Figure 4.1](#) montre la répartition dans l'espace des pyranomètres et la grille considérée pour les prévisions. La position des capteurs a été décidée par PROMES-CNRS selon les possibilités d'installation, la position des installations solaires photovoltaïques et les vents dominants. Ces derniers influent sur le déplacement des masses nuageuses et, par conséquent, sur la disponibilité de la ressource solaire. Instrumenter par des pyranomètres une zone plus étendue que la zone résidentielle considérée permet d'anticiper les fluctuations de la ressource solaire dues aux perturbations atmosphériques. La zone d'instrumentation s'étend sur 4 km<sup>2</sup>.

Les pyranomètres ont été installés sur des supports existants (lampadaires ou poteaux électriques). La [Figure 4.2](#) montre une vue des pyranomètres aux points C, D et K, installés respectivement sur des lampadaires et un poteau électrique. Les capteurs transmettent leurs informations sous forme analogique à un module de codage et de transmission (MCT) par radiofréquences (RF) qui permet d'envoyer les mesures afin qu'elles puissent être archivées et consultées. L'émission et la réception RF utilisent le réseau Sigfox. Le MCT convertit la mesure analogique en valeur numérique grâce à un convertisseur analogique numérique. Ce dernier doit avoir une sensibilité en rapport avec la sensibilité des capteurs qui varie selon le capteur de 60 à 100  $\mu\text{V} \cdot \text{m}^2 \cdot \text{W}^{-1}$ . L'information est transmise sous



la forme d'une tension électrique, permettant de mesurer un éclairage global horizontal entre 0 et  $1500 \text{ W} \cdot \text{m}^{-2}$ . La tension mesurée maximale peut donc varier selon le capteur entre 90 et 150 mV pour cette plage de mesures. Le système MCT est alimenté par une batterie interne.

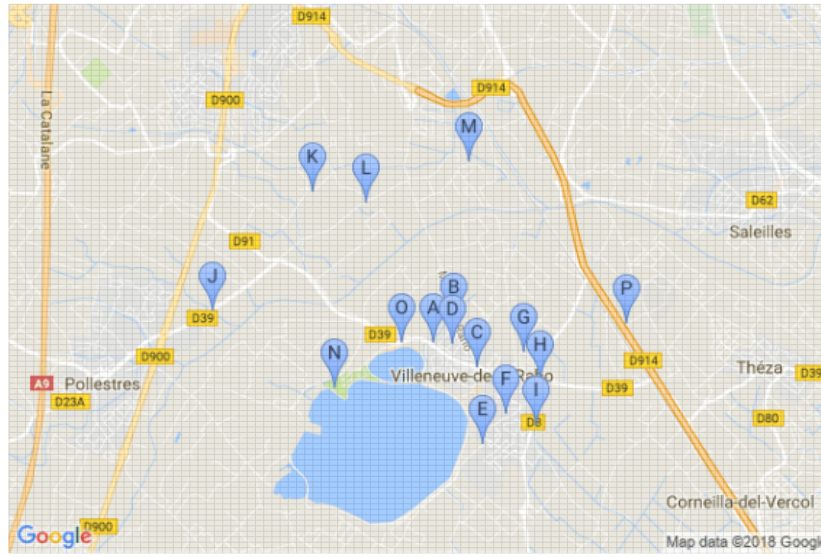


FIGURE 4.1 – Carte montrant la position des 16 pyranomètres installés à Villeneuve-de-la-Raho et la grille de  $100 \times 100$  considérée pour les prévisions.



FIGURE 4.2 – Pyranomètres aux points C, D et K, installés à Villeneuve-de-la-Raho sur des lampadaires et un poteau électrique.

Les mesures du GHI provenant de la zone résidentielle sont utilisées pour la prévision spatio-temporelle du GHI. Quelques jours de mesures sont présentés sur la Figure 4.3. Dans [236], la variabilité spatiale et temporelle de la ressource solaire a été étudiée en utilisant une base de données couvrant vingt ans. Ainsi, un coefficient de variabilité spatiale, noté ici VS, a été défini [236] :

$$VS = \frac{\sqrt{\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n (y_{ij} - \bar{y}_j)^2}}{\frac{1}{n} \sum_{j=1}^n \bar{y}_j} \quad (4.1)$$

où  $\bar{\mathbf{y}} \in \mathbb{R}^{n \times 1}$  est le vecteur des moyennes des données issues des  $m$  capteurs et  $\mathbf{y}_i \in \mathbb{R}^{n \times 1}$  est le vecteur des données issues de mesures réalisées avec le capteur  $i$ . En calculant ce critère de variabilité spatiale pour les sites de Villeneuve-de-la-Raho et Font-Romeu-Odeillo-Via, nous trouvons respectivement 0,32 et 0,91. On note une faible variabilité spatiale des données de Villeneuve-de-la-Raho, qui pourrait

s'expliquer par l'échantillonnage de 10 min, sans doute insuffisant pour distinguer la variabilité due aux phénomènes atmosphériques sur une zone couvrant  $4 \text{ km}^2$  (d'autant que les pyranomètres sont concentrés sur une zone plus réduite). À Font-Romeu-Odeillo-Via, les mesures du GHI échantillonnées avec un pas de temps de 1 s sont réalisées avec un réseau de quatre pyranomètres espacés de 50 m à 100 m. Cet échantillonnage temporel très fin permet de distinguer la variabilité spatiale due aux phénomènes atmosphériques. Une illustration de ces phénomènes est montrée sur la Figure 4.4. On remarque des chutes du GHI successivement au niveau de chaque capteur qui reflètent le passage, par exemple, d'un nuage au-dessus de la zone et ces événements durent 2 min à 5 min. Cette faible variabilité spatiale des données issues de notre cas d'étude a une incidence directe sur la qualité de l'interpolation spatiale produite par les modèles spatio-temporels développés.

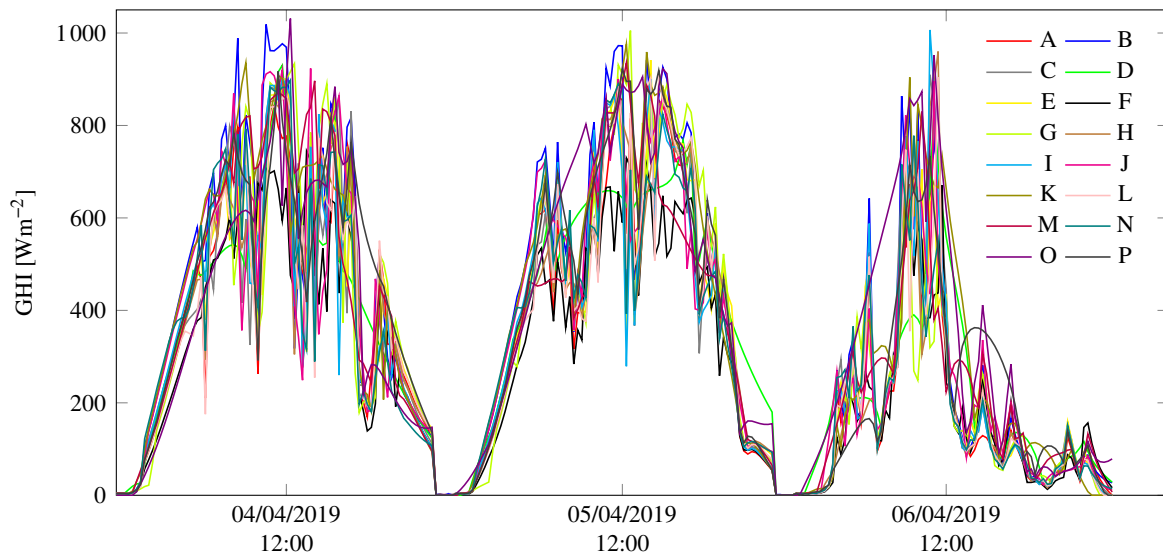


FIGURE 4.3 – Mesures du GHI sur quelques jours réalisées à l'aide des 16 pyranomètres installés à Villeneuve-de-la-Raho.

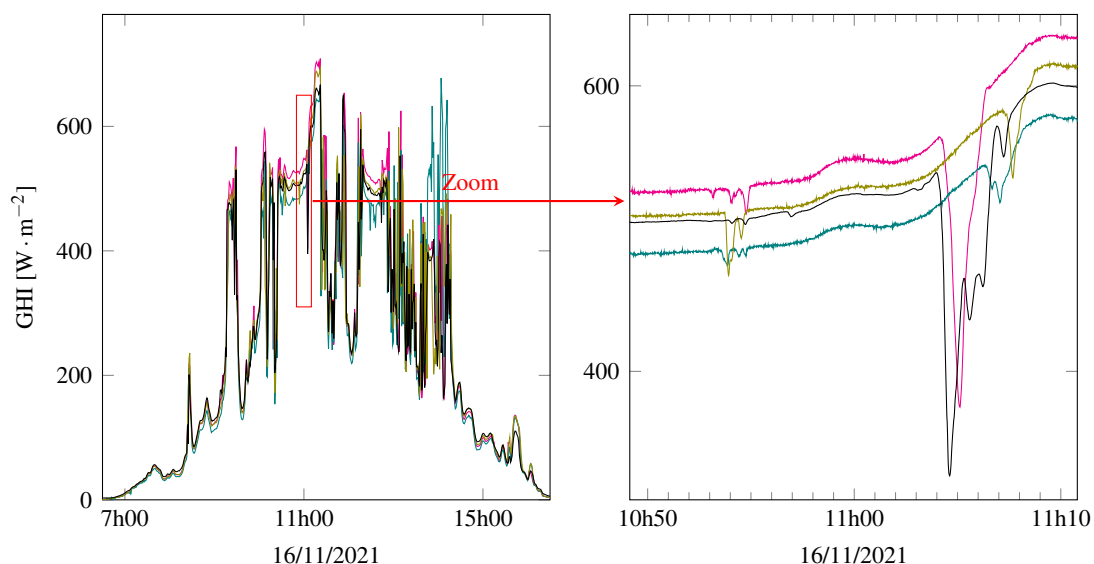


FIGURE 4.4 – Exemple de mesures du GHI à Odeillo, avec 4 pyranomètres espacés de 50 m à 100 m et un échantillonnage de 1 s.

Afin d'identifier le noyau spatio-temporel, une étude comparative impliquant plusieurs noyaux spatiaux ARD (en raison d'une coordonnée spatiale bidimensionnelle) et le noyau temporel  $k_{\text{Per} \times \text{RQ}}$

(identifié au chapitre précédent) est menée. Chaque pyranomètre est localisé par sa longitude et sa latitude. Les données sont échantillonnées à un pas de temps de 10 min et les horizons de prévision considérés vont de 10 min à 24 h. Les données couvrant les 30 premiers jours du mois d'août 2019 sont considérées : les 15 premiers jours sont utilisés pour entraîner les modèles et les 15 derniers jours pour évaluer leurs performances.

## 4.4 Processus gaussien spatio-temporel

### 4.4.1 Notations

Un processus gaussien est complètement défini par sa fonction moyenne et par sa fonction de covariance. Pour  $f$  un processus gaussien spatio-temporel, on écrira  $f(\mathbf{s}, t) \sim \mathcal{GP}(\mu(\mathbf{s}, t), k((\mathbf{s}, t), (\mathbf{s}', t')))$ , où  $(\mathbf{s}, t)$  et  $(\mathbf{s}', t')$  sont deux variables d'entrée localisées par  $\mathbf{s}$  et  $\mathbf{s}'$  aux instants  $t$  et  $t'$ ,  $\mu(\mathbf{s}, t)$  est la fonction moyenne et  $k((\mathbf{s}, t), (\mathbf{s}', t'))$  est la fonction de covariance.

### 4.4.2 Propriétés des noyaux spatio-temporels

Pour un processus spatio-temporel  $\{z(\mathbf{s}, t) : \mathbf{s} \in D_s \subset \mathbb{R}^d, t \in D_t \subset \mathbb{R}\}$ , une fonction de covariance est dite spatio-temporellement stationnaire si  $\forall \mathbf{s}, \mathbf{s}' \in D_s, t, t' \in D_t$ , il existe une fonction  $f_1 : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}$  telle que :

$$k((\mathbf{s}, t), (\mathbf{s}', t')) = f_1(\mathbf{s} - \mathbf{s}', t - t') \quad (4.2)$$

Une fonction de covariance est spatio-temporellement isotrope si  $\forall \mathbf{s}, \mathbf{s}' \in D_s, t, t' \in D_t$ , il existe une fonction  $f_2 : \mathbb{R}^d \times \mathbb{R} \rightarrow \mathbb{R}$  telle que :

$$k((\mathbf{s}, t), (\mathbf{s}', t')) = f_2(\|\mathbf{s} - \mathbf{s}'\|, |t - t'|) \quad (4.3)$$

Un noyau spatio-temporel est complètement symétrique si  $\forall \mathbf{s}, \mathbf{s}' \in D_s, t, t' \in D_t$  :

$$k((\mathbf{s}, t), (\mathbf{s}', t')) = k((\mathbf{s}', t'), (\mathbf{s}, t)) = k((\mathbf{s}', t), (\mathbf{s}, t')) = k((\mathbf{s}, t'), (\mathbf{s}', t)) \quad (4.4)$$

Un noyau spatio-temporel séparable est donné par :

$$k((\mathbf{s}, t), (\mathbf{s}', t')) = k_s(\mathbf{s}, \mathbf{s}') \cdot k_t(t, t') \quad \forall \mathbf{s}, \mathbf{s}' \in D_s, t, t' \in D_t \quad (4.5)$$

ou

$$k((\mathbf{s}, t), (\mathbf{s}', t')) = k_s(\mathbf{s}, \mathbf{s}') + k_t(t, t') \quad \forall \mathbf{s}, \mathbf{s}' \in D_s, t, t' \in D_t \quad (4.6)$$

où  $k_s(\mathbf{s}, \mathbf{s}')$  est un noyau spatial et  $k_t(t, t')$  est un noyau temporel. Dans l'équation (4.6), même si les fonctions  $k_s$  et  $k_t$  sont strictement définies positives, le noyau  $k$  peut ne pas l'être. Il faut noter que le noyau spatio-temporel séparable est un cas spécial d'un noyau spatio-temporel complètement symétrique. Par conséquent, si le noyau n'est pas complètement symétrique, il ne peut pas être séparable. Les données environnementales ne sont généralement pas séparables en raison des interactions complexes entre le temps et l'espace. Cependant, cette hypothèse de séparabilité, qui n'est vraisemblablement pas vérifiée, permet d'obtenir des résultats préliminaires sans avoir à recourir à des méthodes d'approximation pour l'inversion des matrices (voir, par exemple, les travaux de [237]). En effet, cette hypothèse permet le développement de modèles GPR spatio-temporels adaptés à de grands jeux de données tirant profit des théorèmes de Kronecker [238].

NOTE

Par exemple, pour le calcul de la matrice de covariance  $K$ , en utilisant l'équation (4.5), il suffit de calculer les matrices  $K_s$  et  $K_t$  et, ensuite, d'utiliser le produit de Kronecker [238] :

$$K = K_s \otimes K_t \quad \iff \quad K^{-1} = K_s^{-1} \otimes K_t^{-1} \quad (4.7)$$

Cette approche réduit considérablement le coût calculatoire.  $\diamond$



### 4.4.3 Noyau spatio-temporel utilisé

Pour le développement de modèles GPR spatio-temporels, nous utilisons des noyaux séparables (équation (4.5)) associant le noyau temporel  $k_{\text{Per} \times \text{RQ}}$  identifié au [Chapitre 3](#) et un noyau spatial que nous cherchons à identifier. Le noyau  $k_{\text{Per} \times \text{RQ}}$ , recommandé en présence d'un jeu de données couvrant quelques semaines, est donné par :

$$k_{\text{Per} \times \text{RQ}}(t, t') = \sigma^2 \exp \left( -\frac{2 \sin^2 \left( \frac{\pi(t-t')}{P} \right)}{\ell_1^2} \right) \cdot \left( 1 + \frac{(t-t')^2}{2\alpha\ell_2^2} \right)^{-\alpha} \quad (4.8)$$

où  $t$  est le temps,  $\sigma$  est l'amplitude du signal,  $\ell_1$  et  $\ell_2$  sont les paramètres de corrélation,  $P$  est la période et  $\alpha$  est un paramètre de pondération des variations multi-échelles. La coordonnée spatiale  $\mathbf{s}$  étant bidimensionnelle, pour modéliser la structure de corrélation spatiale des données, les noyaux avec ARD définissant automatiquement un paramètre de corrélation suivant chacune des dimensions du vecteur d'entrée sont utilisés. Les noyaux avec ARD considérés, à savoir  $k_{\text{SE-ARD}}$ ,  $k_{\text{RQ-ARD}}$ ,  $k_{\text{M}_{1/2}\text{-ARD}}$ ,  $k_{\text{M}_{3/2}\text{-ARD}}$  et  $k_{\text{M}_{5/2}\text{-ARD}}$ , sont respectivement définis par :

$$k_{\text{SE-ARD}}(\mathbf{s}, \mathbf{s}') = \sigma^2 \exp \left( -\frac{1}{2} \sum_{i=1}^2 \frac{(s_i - s'_i)^2}{\ell_i^2} \right) \quad (4.9)$$

$$k_{\text{RQ-ARD}}(\mathbf{s}, \mathbf{s}') = \sigma^2 \left( 1 + \frac{1}{2\alpha} \sum_{i=1}^2 \frac{(s_i - s'_i)^2}{\ell_i^2} \right)^{-\alpha} \quad (4.10)$$

$$k_{\text{M}_{1/2}\text{-ARD}}(\mathbf{s}, \mathbf{s}') = \sigma^2 \exp \left( -\sum_{i=1}^2 \frac{|s_i - s'_i|}{\ell_i} \right) \quad (4.11)$$

$$k_{\text{M}_{3/2}\text{-ARD}}(\mathbf{s}, \mathbf{s}') = \sigma^2 \left( 1 + \sqrt{3} \sum_{i=1}^2 \frac{|s_i - s'_i|}{\ell_i} \right) \cdot \exp \left( -\sqrt{3} \sum_{i=1}^2 \frac{|s_i - s'_i|}{\ell_i} \right) \quad (4.12)$$

$$k_{\text{M}_{5/2}\text{-ARD}}(\mathbf{s}, \mathbf{s}') = \sigma^2 \left( 1 + \sqrt{5} \sum_{i=1}^2 \frac{|s_i - s'_i|}{\ell_i} + \sqrt{5} \sum_{i=1}^2 \frac{(s_i - s'_i)^2}{3\ell_i^2} \right) \cdot \exp \left( -\sqrt{5} \sum_{i=1}^2 \frac{|s_i - s'_i|}{\ell_i} \right) \quad (4.13)$$

où  $\sigma$  est l'amplitude du signal,  $\ell_1$  et  $\ell_2$  sont les paramètres de corrélation,  $\mathbf{s}$  et  $\mathbf{s}'$  sont des coordonnées spatiales, et  $\alpha$  est un paramètre de pondération des variations multi-échelles. Une étude comparative des modèles GPR associant le noyau temporel  $k_{\text{Per} \times \text{RQ}}$  et ces noyaux spatiaux est présentée dans ce chapitre.

## 4.5 Résultats et discussion

Cette section présente les résultats obtenus grâce aux modèles de prévision spatio-temporels développés. Les performances des modèles sont évaluées à l'aide des critères définis à la [Section 1.4](#) du [Chapitre 1](#) tels que la racine carrée de l'erreur quadratique moyenne normalisée (nRMSE), l'erreur absolue moyenne dynamique (DMAE), l'*interval score* (IS) et le *skill score* (SS). Pour ne pas surcharger la présentation des résultats, cinq horizons de temps allant de 10min à 24h sont considérés. Tout d'abord, une comparaison des performances des différents modèles GPR spatio-temporels développés est effectuée en analysant les résultats sur les 15 jours de test. Ensuite, nous considérerons les cas particuliers d'une journée ciel clair, d'une journée partiellement couverte et d'une journée complètement couverte. Pour la prévision dans l'espace, nous avons considéré une grille de  $100 \times 100$ , ce qui fournit une prévision tous les 30m en longitude et en latitude, à laquelle nous avons ajouté la position des 16 pyranomètres. La [Figure 4.1](#) montre la position des pyranomètres et la grille utilisée.

## NOTE

L'apprentissage des modèles GPR a été réalisé à partir de mesures provenant de tous les pyranomètres installés au sein du cas d'étude (16 capteurs). Toutefois, initialement, nous souhaitions utiliser des données ne provenant que d'une partie de ces pyranomètres pour l'apprentissage et valider les structures de corrélation spatiale avec des données provenant des pyranomètres restants. Malheureusement, en raison de la faible granularité de l'instrumentation, réduire la quantité de données utilisées lors de l'apprentissage des modèles GPR a conduit à une dégradation des résultats.  $\diamond$

## 4.5.1 Évaluation des modèles

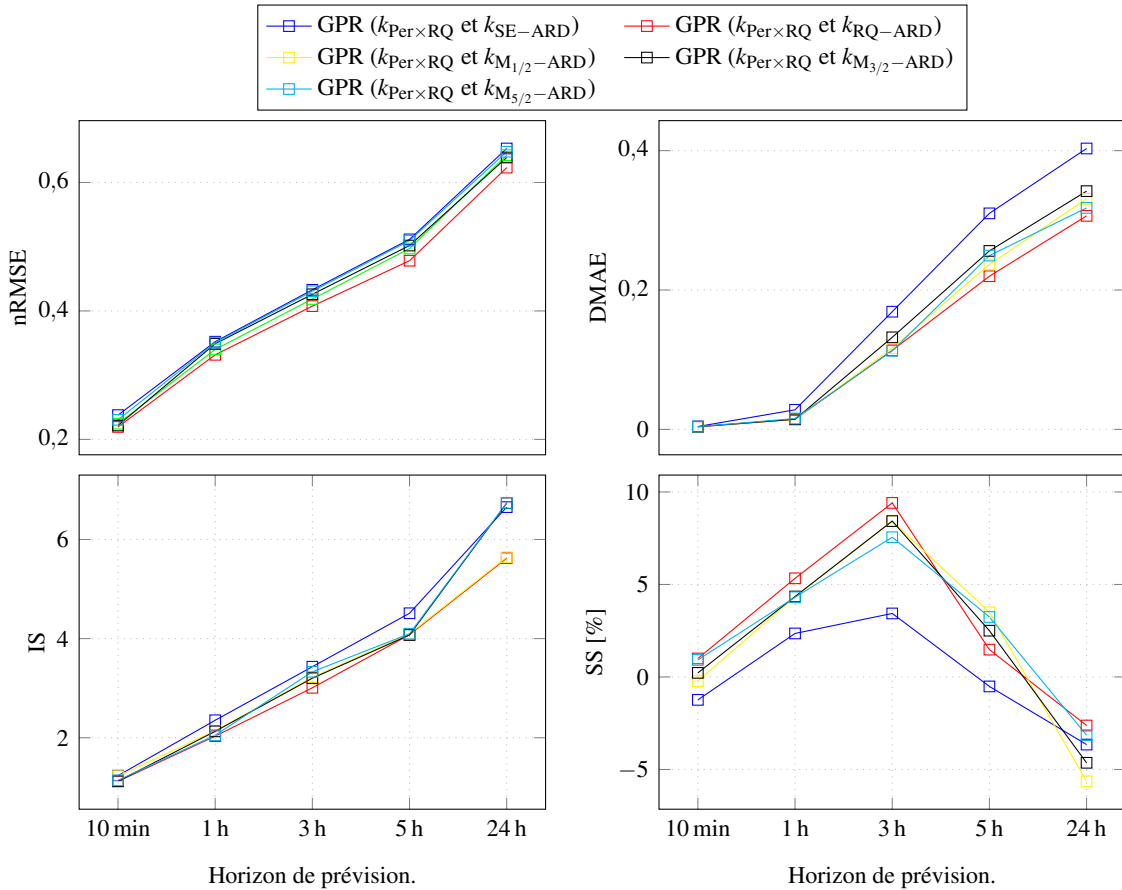


FIGURE 4.5 – Valeurs des critères de performance pour les modèles GPR spatio-temporels développés en considérant les données de test couvrant quinze jours (du 16/08/2019 au 30/08/2019).

Les valeurs des critères de performance pour les modèles GPR développés sont présentées sur la Figure 4.5. En examinant les valeurs du nRMSE et du DMAE, nous ne pouvons pas tirer de conclusion consolidée quant à la pertinence d'un noyau spatial car ces erreurs, qui combinent « erreurs de prévision dans le temps » et « erreurs de prévision dans l'espace », sont calculées aux 16 points de mesure. Par ailleurs, pour tous les horizons de temps considérés, les erreurs commises aux points de mesure sont sensiblement identiques pour tous les modèles GPR développés, à l'exception du modèle GPR associant noyau spatial  $k_{\text{SE-ARD}}$  et noyau temporel  $k_{\text{Per} \times \text{RQ}}$  dont le DMAE augmente rapidement lorsque l'horizon de temps augmente. Selon le score de compétence, pour un horizon de temps de 10 min, la persistance sur l'indice ciel clair a des performances légèrement supérieures au modèle GPR associant noyau spatial  $k_{\text{SE-ARD}}$  et noyau temporel  $k_{\text{Per} \times \text{RQ}}$ , et modèle GPR associant noyau spatial  $k_{\text{M}_{1/2}\text{-ARD}}$  et noyau temporel  $k_{\text{Per} \times \text{RQ}}$ . Pour un horizon de temps de 24 h, les performances de ce modèle de référence sont supérieures à celles des modèles GPR développés.

Pour mieux évaluer les structures de corrélation spatiale, nous nous focaliserons sur les résultats

obtenus pour quelques journées-types : la journée ciel clair du 16/08/2019, la journée partiellement couverte du 21/08/2019 et la journée complètement couverte du 27/08/2019.

NOTE

*Rappelons qu'une journée presque ciel clair est caractérisée par un indice ciel clair supérieur ou égal à 0,8, tandis qu'une journée complètement couverte est caractérisée par un indice ciel clair inférieur ou égal à 0,4. Une journée partiellement couverte est caractérisée par un indice ciel clair compris entre 0,4 et 0,8.* ◇

#### 4.5.2 Journées-types

Examinons les résultats de la prévision spatio-temporelle du GHI obtenus pour les journées susmentionnées. Les données utilisées au cours de la phase de test sont respectivement présentées sur les Figures 4.6, 4.7 et 4.8. Les valeurs des critères de performance pour les modèles GPR développés sont présentées sur la Figure 4.9 pour la journée ciel clair, sur la Figure 4.10 pour la journée partiellement couverte et sur la Figure 4.11 pour la journée complètement couverte.

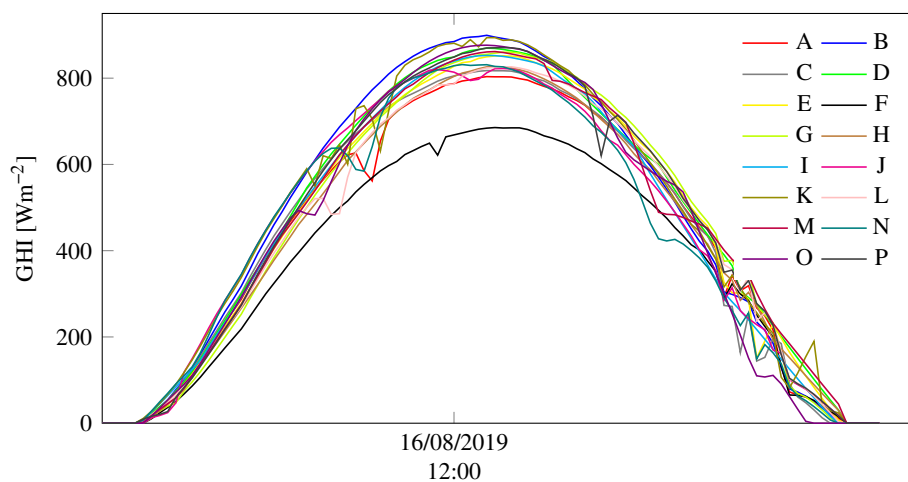


FIGURE 4.6 – Mesures du GHI (16 pyranomètres installés à Villeneuve-de-la-Raho) au cours de la journée « ciel clair » du 16/08/2019.

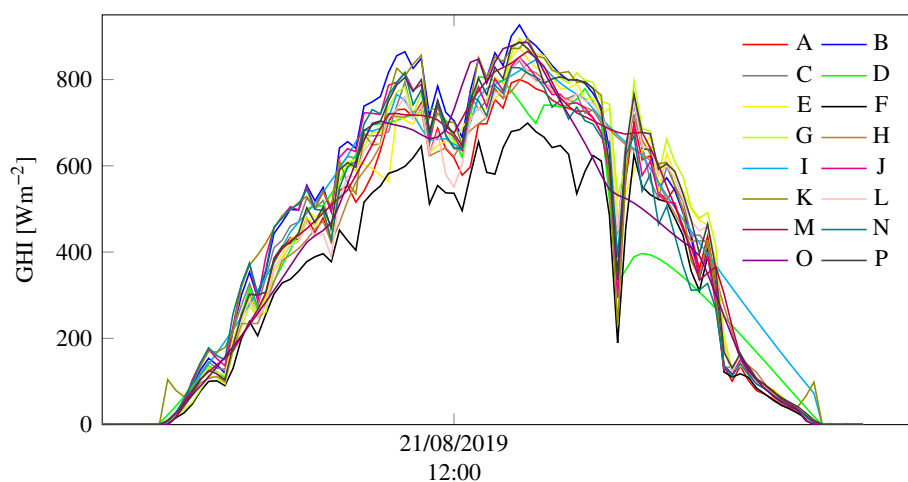


FIGURE 4.7 – Mesures du GHI (16 pyranomètres installés à Villeneuve-de-la-Raho) au cours de la journée partiellement couverte du 21/08/2019.

Afin d'évaluer la pertinence des structures de corrélation spatiale considérées, quelques exemples de prévision spatio-temporelle du GHI lors de la journée partiellement couverte sont présentés sur

les Figures 4.12–4.16 (les exemples de prévision pour les autres journées-types ne sont pas présentés ici car ils n'apportent pas d'information supplémentaire quant à la pertinence de tel ou tel noyau spatial pour la prévision dans le temps et l'espace du GHI). Sur ces figures, les 16 pyranomètres sont représentés par des points dont la couleur reflète la valeur du GHI mesuré à l'horizon de prévision considéré ; à l'exception des points de mesure, la couleur indique la valeur des prévisions.

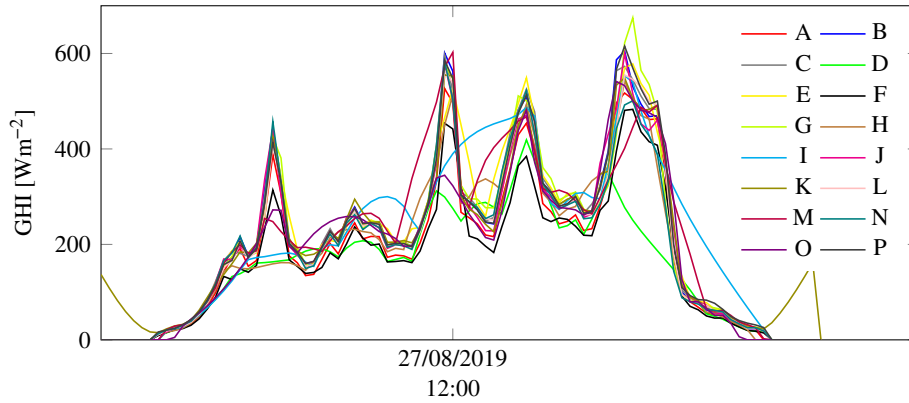


FIGURE 4.8 – Mesures du GHI (16 pyranomètres installés à Villeneuve-de-la-Raho) au cours de la journée complètement couverte du 27/08/2019.

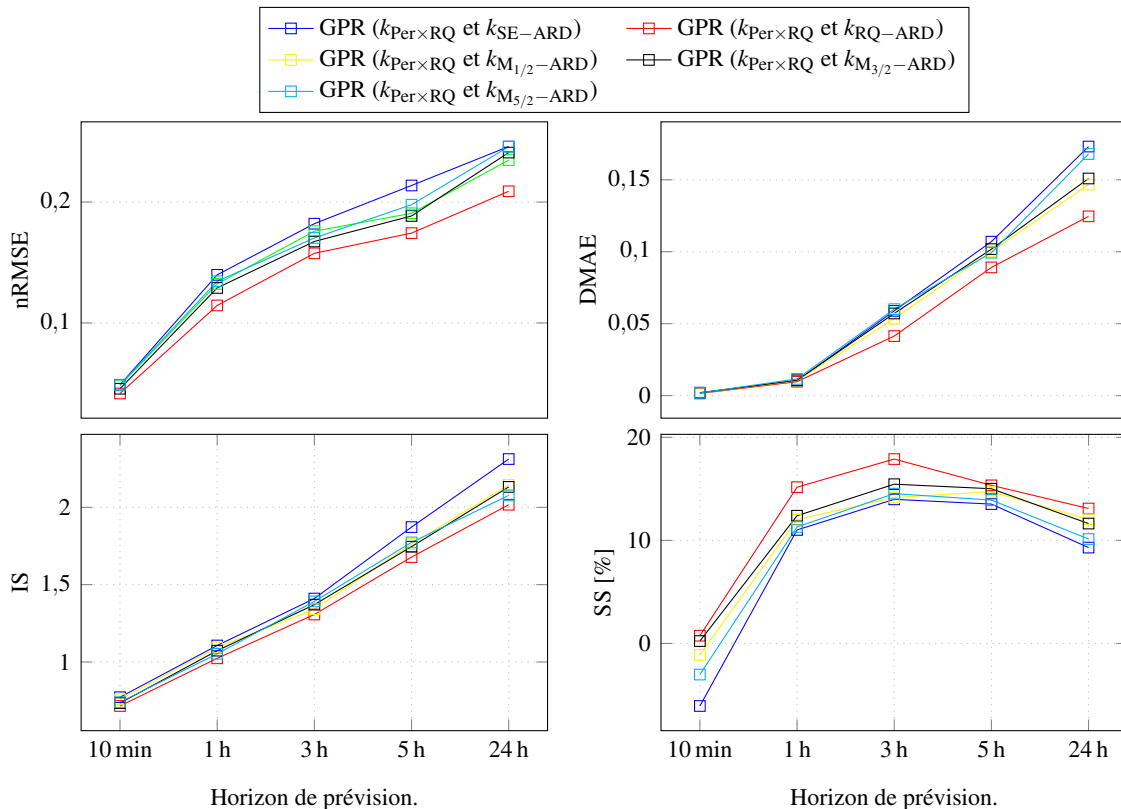


FIGURE 4.9 – Valeurs des critères de performance pour la journée ciel clair du 16/08/2019.

En examinant, d'une part, les valeurs des critères de performance pour les modèles GPR lors de la journée partiellement couverte (Figure 4.10) et, d'autre part, les prévisions obtenues (Figures 4.12–4.16), on peut noter que pour tous les horizons de temps considérés, le modèle GPR associant noyau temporel  $k_{Per \times RQ}$  et noyau spatial  $k_{SE-ARD}$  produit les moins bons résultats. En effet, sur ces figures, on constate que ce modèle produit des prévisions du GHI qui tendent vers zéro dès lors que l'on

s'éloigne des points de mesure, ce qui n'est bien évidemment pas le résultat souhaité. Par exemple, sur la Figure 4.12a, le modèle GPR associant noyau temporel  $k_{\text{Per}\times\text{RQ}}$  et noyau spatial  $k_{\text{SE-ARD}}$  produit une prévision du GHI à 13h59 qui tend vers zéro dès lors que l'on s'éloigne des points de mesure alors que le GHI mesuré est proche de  $1000 \text{ W} \cdot \text{m}^{-2}$ . Ce constat est le même quelle que soit la journée-type considérée.

NOTE

Notons qu'en l'absence de données pour la mise à jour des modèles, les prévisions obtenues vont tendre vers une valeur moyenne, qui est a priori nulle. Le fait que le modèle GPR associant noyau temporel  $k_{\text{Per}\times\text{RQ}}$  et noyau spatial  $k_{\text{SE-ARD}}$  produise des prévisions qui tendent vers zéro dès lors que l'on s'éloigne des points de mesure semble indiquer que l'estimation des paramètres de corrélation  $\ell_1$  et  $\ell_2$  du noyau spatial produit de petites valeurs, contrairement aux autres modèles GPR développés.  $\diamond$

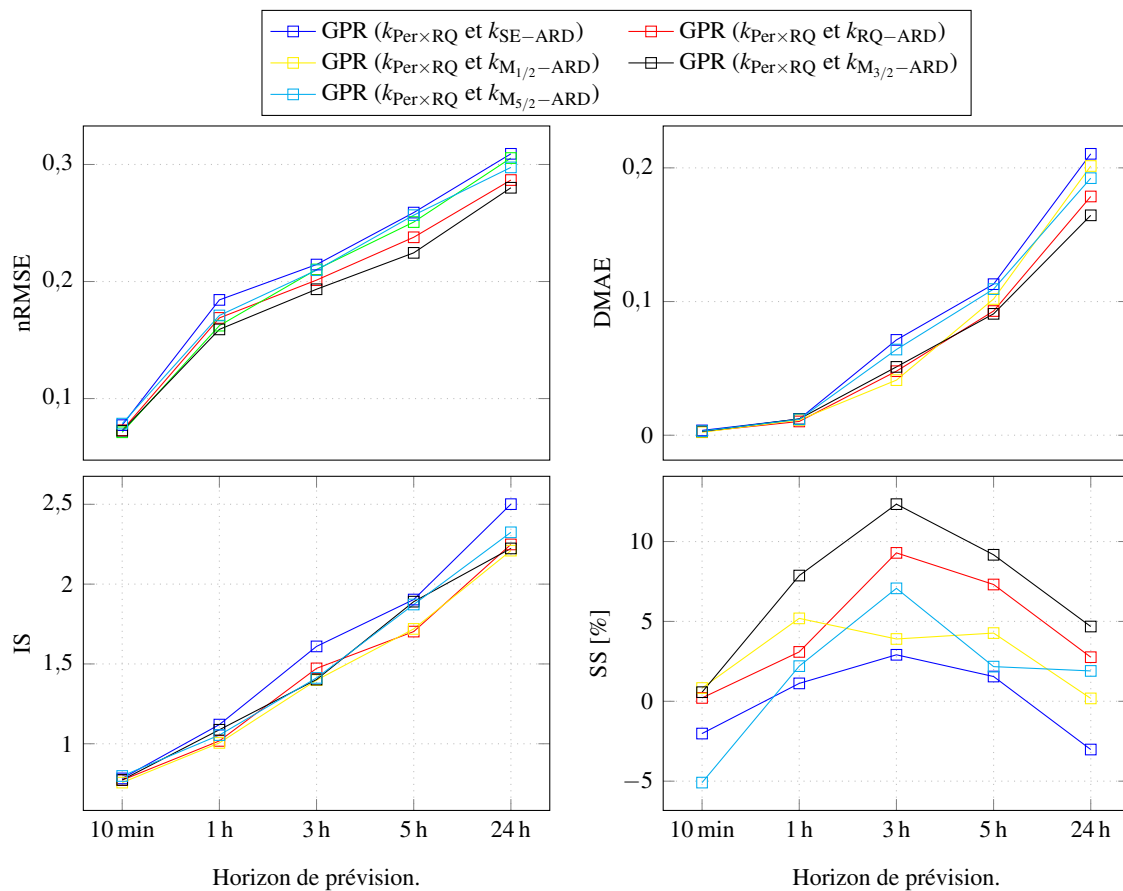


FIGURE 4.10 – Valeurs des critères de performance pour la journée partiellement couverte du 21/08/2019.

Pour ce qui est des autres modèles GPR développés, à savoir le modèle associant noyau temporel  $k_{\text{Per}\times\text{RQ}}$  et noyau spatial  $k_{\text{RQ-ARD}}$  et les modèles associant noyau temporel  $k_{\text{Per}\times\text{RQ}}$  et noyau spatial de la classe Matérn ( $k_{\text{M}_{1/2-ARD}}$ ,  $k_{\text{M}_{3/2-ARD}}$  ou  $k_{\text{M}_{5/2-ARD}}$ ), les résultats obtenus sont similaires tant pour l'erreur de prévision commise aux points de mesure (nRMSE) que pour l'interpolation spatiale réalisée. Les prévisions du GHI fournies par ces modèles semblent indiquer des résultats cohérents (Figures 4.12–4.16) sur la totalité de la zone d'intérêt (Villeneuve-de-la-Raho). Étant donné que des mesures issues de tous les pyranomètres (16 capteurs) ont été considérées lors de l'entraînement des modèles, nous ne pouvons pas, malheureusement, évaluer précisément la qualité de l'interpolation spatiale et ainsi distinguer les modèles développés. On note également des prévisions faiblement variables dans l'espace, ce qui n'est pas surprenant étant donné que les mesures utilisées présentent une faible variabilité spatiale. Lorsque l'horizon de temps augmente, les prévisions fournies par les

modèles se dégradent : quel que soit le modèle considéré, on peut constater sur les cartes présentées des prévisions éloignées des mesures.

## 4.6 Conclusion

Ce chapitre traite d'une première étape vers la prévision dans le temps et l'espace du GHI par la mise en œuvre d'une régression non paramétrique par processus gaussien. Des modèles GPR ont été développés en supposant une séparabilité entre l'espace et le temps. Le noyau temporel qui a été considéré est le noyau  $k_{\text{Per} \times \text{RQ}}$  identifié au Chapitre 3 ; ce noyau est adapté à des jeux de données couvrant quelques semaines. Une étude comparative des modèles GPR associant noyau temporel  $k_{\text{Per} \times \text{RQ}}$  et différents noyaux spatiaux a été menée. L'analyse des résultats et l'examen de quelques cas d'étude (une journée ciel clair, une journée partiellement couverte et une journée complètement couverte) ont montré qu'il convient d'écarter le modèle GPR associant noyau temporel  $k_{\text{Per} \times \text{RQ}}$  et noyau spatial  $k_{\text{SE-ARD}}$ . En effet, ce modèle produit les moins bons résultats selon les critères de performance utilisés (nRMSE, DMAE, IS et SS) et produit des prévisions spatiales qui tendent vers 0 lorsqu'on s'éloigne des points de mesure. En ce qui concerne les autres modèles GPR développés, à savoir le modèle GPR associant noyau temporel  $k_{\text{Per} \times \text{RQ}}$  et noyau spatial  $k_{\text{RQ-ARD}}$  et les modèles GPR associant noyau temporel  $k_{\text{Per} \times \text{RQ}}$  et les noyaux de la classe Matérn, il est délicat de tirer des conclusions consolidées sans procéder à des expérimentations supplémentaires. Ces modèles donnent des résultats similaires qui semblent correctes, sans qu'aucun modèle ne se dégage clairement.

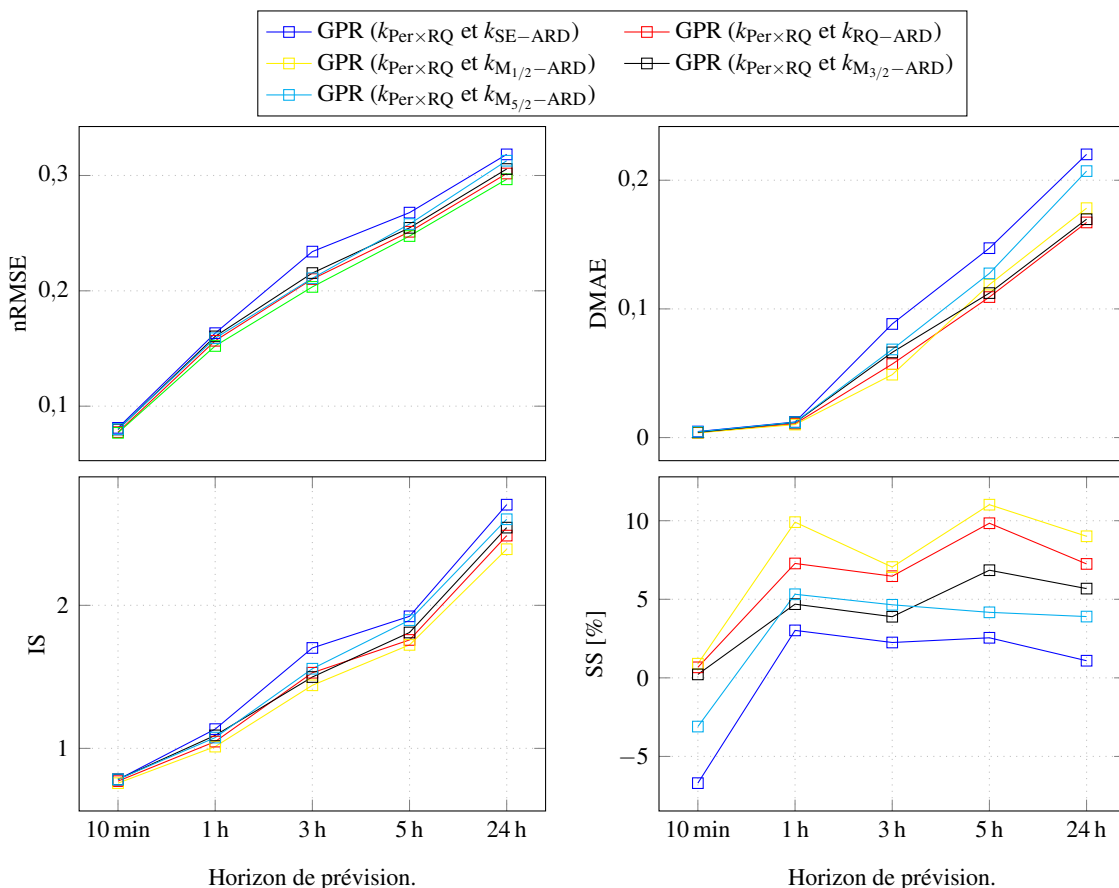


FIGURE 4.11 – Valeurs des critères de performance pour la journée complètement couverte du 27/08/2019.

Cette étude a mis en lumière les limitations des données utilisées. En effet, on note des mesures aberrantes et beaucoup de données manquantes dans les mesures réalisées à Villeneuve-de-la-Raho

(seules les données du mois d'août 2019 ont passé les tests de qualités), ce qui limite l'étude sur un jeu de données très restreint. Par ailleurs, on remarque également une faible granularité spatiale et les mesures ont été réalisées avec un échantillonnage temporel grossier qui ne permet pas de distinguer correctement les effets des phénomènes atmosphériques sur la variabilité spatiale du GHI ; ce qui affecte directement la qualité de l'interpolation spatiale fournie par les modèles GPR spatio-temporels développés. En raison de cette faible granularité spatiale, les données provenant de tous les capteurs ont été utilisées lors de la phase d'entraînement des modèles.

Cette étude préliminaire a permis d'aboutir à des résultats qui peuvent être améliorés. Dans un premier temps, augmenter la granularité de l'instrumentation au sein de la zone d'étude et utiliser un temps d'échantillonnage plus fin pourrait permettre d'évaluer plus finement les prévisions spatiales fournies par les modèles. Ensuite, il est possible de revenir sur certaines hypothèses utilisées pour la construction des noyaux. D'une part, considérer une fonction moyenne non nulle (égale, par exemple, à l'indice ciel clair) pourrait être intéressant car cela éviterait que les prévisions tendent vers 0 lorsque l'on s'éloigne des points de mesure. D'autre part, des structures de noyau plus complexes pourraient être considérées, notamment en ne supposant plus la séparabilité entre le temps et l'espace : bien qu'impliquant une complexité calculatoire plus élevée, qui pourrait entraver l'implémentation en temps réel des algorithmes, cela permettrait une modélisation plus réaliste du comportement spatio-temporel du GHI.



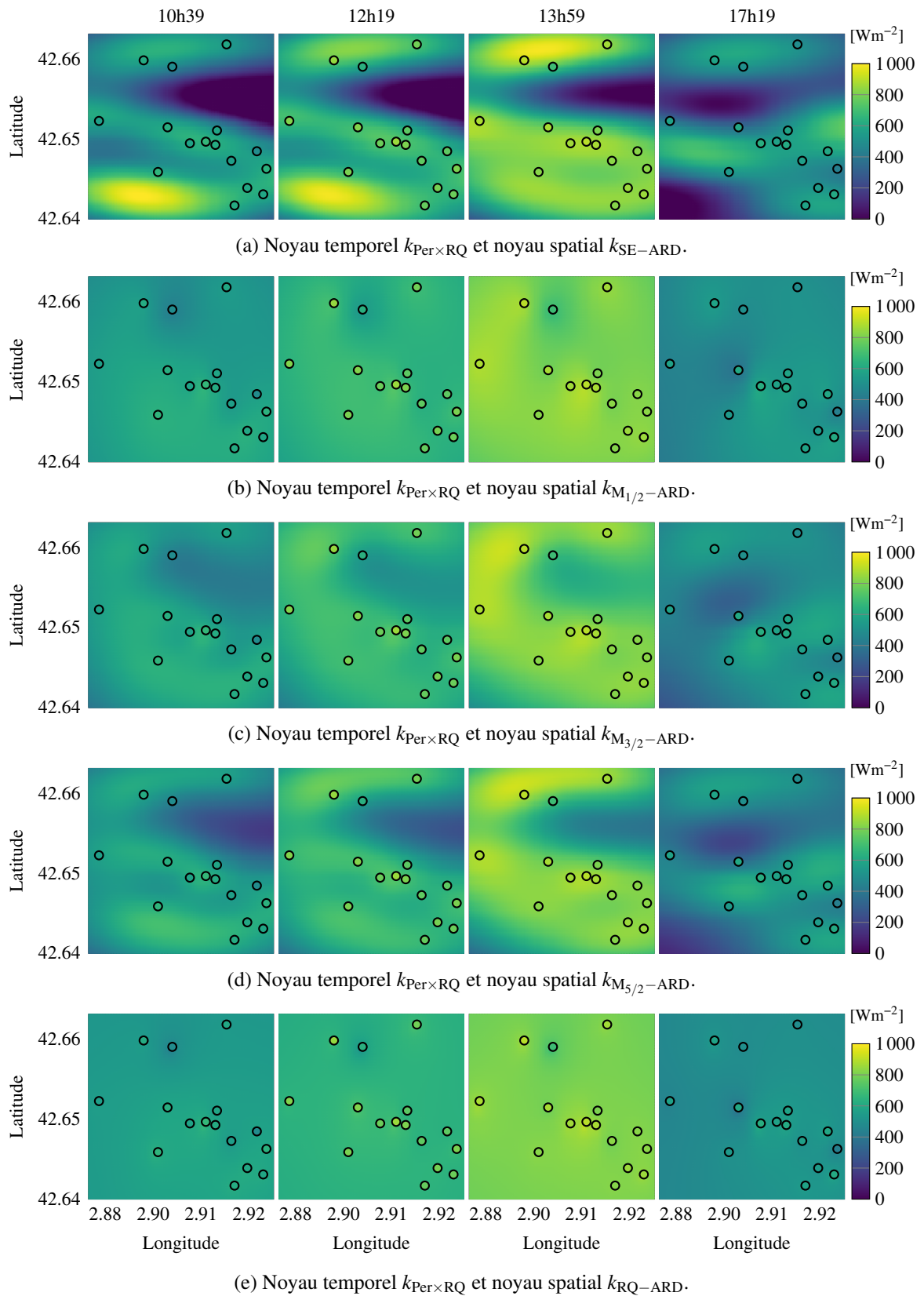


FIGURE 4.12 – Exemples de prévision à 10 min obtenue grâce aux modèles GPR testés.



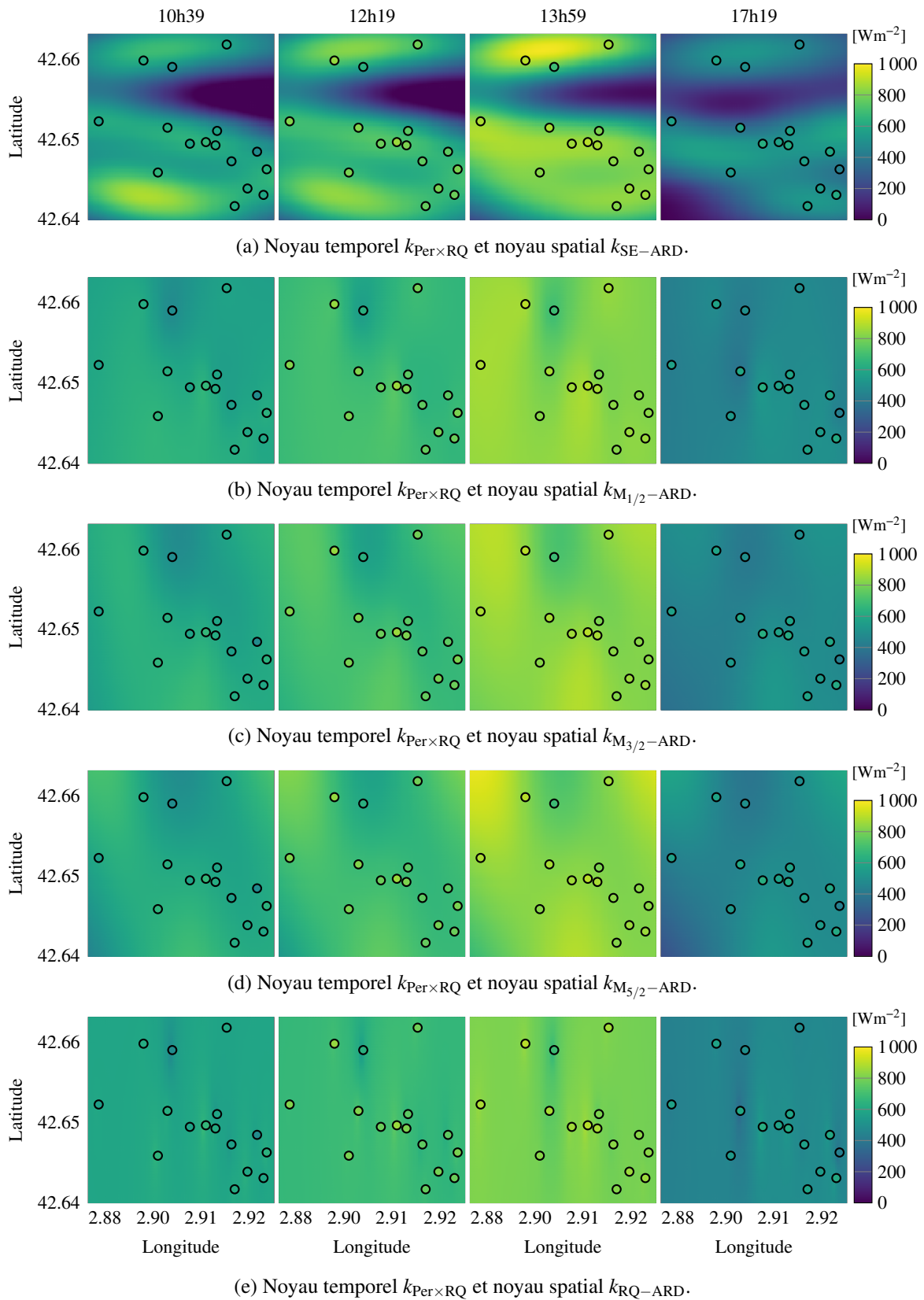


FIGURE 4.13 – Exemples de prévision à 1 h obtenue grâce aux modèles GPR testés.

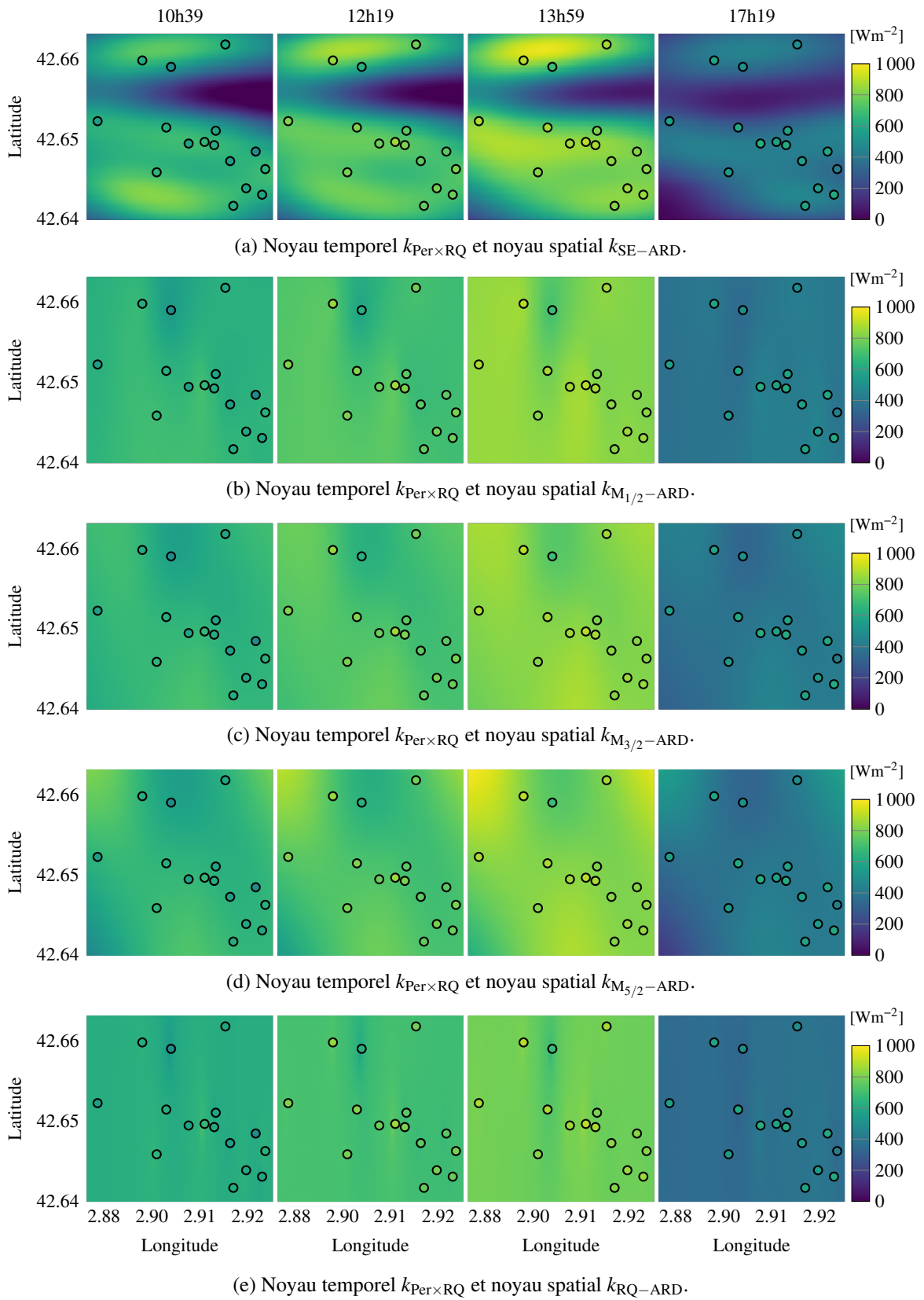


FIGURE 4.14 – Exemples de prévision à 3 h obtenue grâce aux modèles GPR testés.

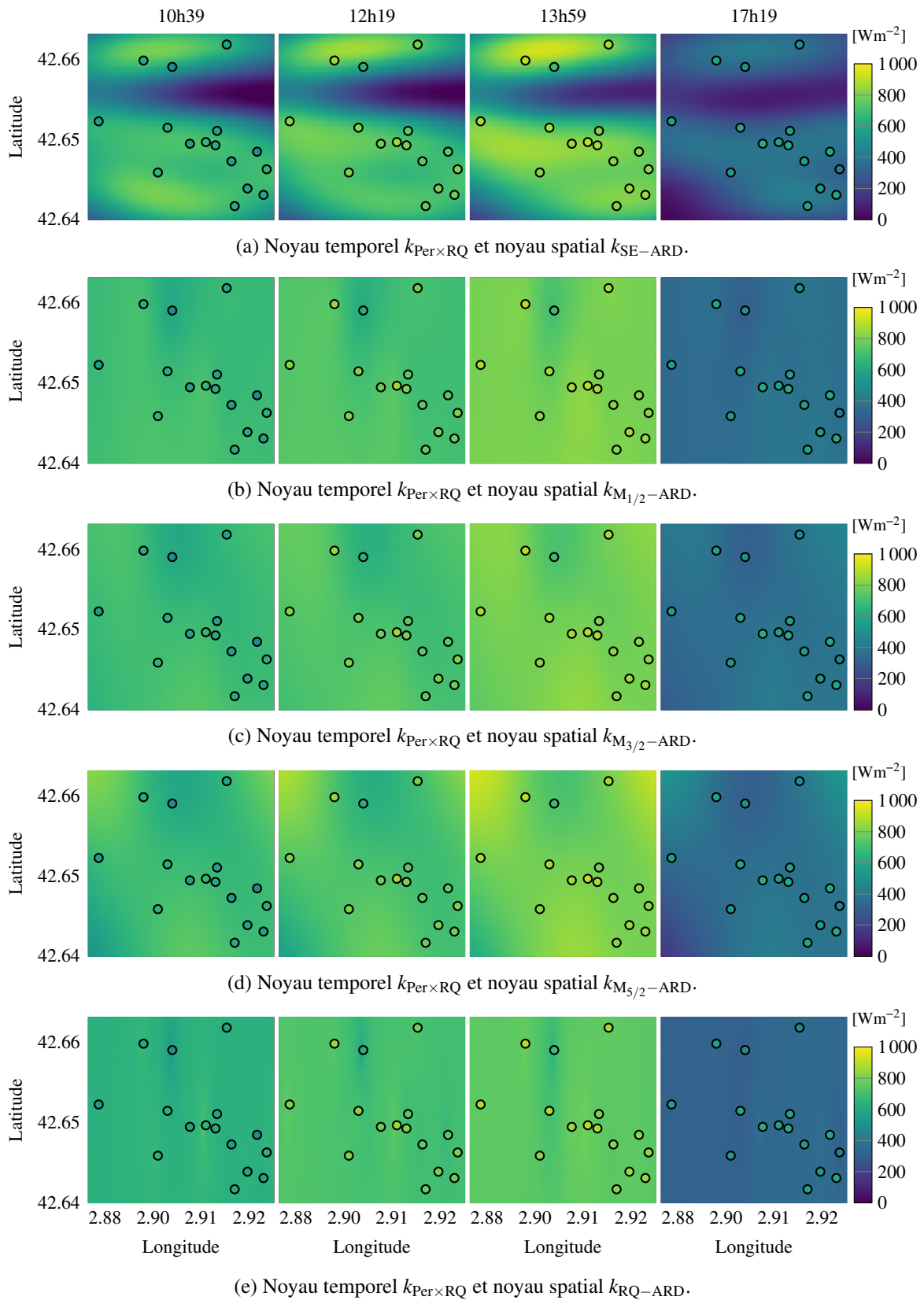


FIGURE 4.15 – Exemples de prévision à 5 h obtenue grâce aux modèles GPR testés.

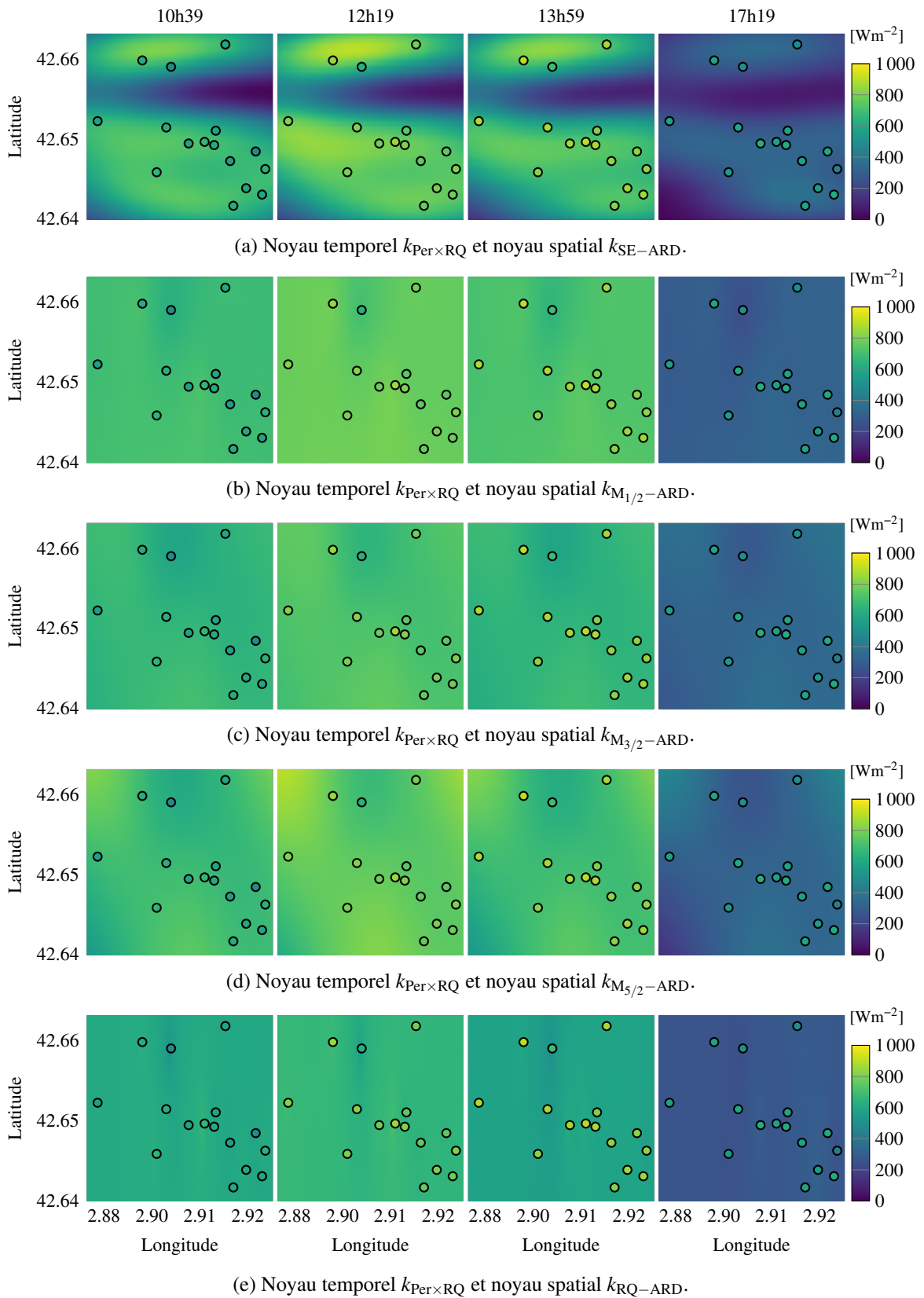


FIGURE 4.16 – Exemples de prévision à 24 h obtenue grâce aux modèles GPR testés.



# Conclusion générale

L'objectif de ce travail de thèse réalisé dans le cadre du projet « Smart Occitania » était de développer des modèles pour la prévision localisée du GHI puis des modèles pour sa prévision dans le temps et l'espace à des fins de gestion intelligente du réseau électrique de distribution en région Occitanie. Pour y parvenir, nous nous sommes posés un certain nombre de questions. L'état de l'art des modèles de prévision du GHI nous a amené à faire le choix de l'apprentissage automatique. Ces modèles statistiques, capables de prévisions multi-horizons, permettent de s'affranchir d'un modèle « ciel clair ». Néanmoins, il s'est avéré nécessaire d'identifier l'outil de l'apprentissage automatique le plus adapté à la prévision GHI. Pour le contrôle-commande du réseau électrique de distribution basse tension, le gestionnaire a besoin de prévision à des horizons de temps allant de l'infra-horaire à l'infra-journalier à pas de temps 10 min. Différents jeux de données du GHI échantillonnées à 10 min ont été utilisés et des horizons de temps allant de 10 min à 24 h ont été considérés.

L'outil de l'apprentissage automatique le plus adapté à la prévision du GHI a été choisi au [Chapitre 2](#). Une étude comparative des outils phares, à savoir la régression non paramétrique par processus gaussien, la régression par machines à vecteurs de support et la régression par réseaux de neurones artificiels a été menée. La persistance sur l'indice ciel clair a été utilisée comme modèle de référence. Les résultats ont montré qu'à horizon de temps infra-horaire (10 min), tous les modèles fondés sur l'apprentissage automatique et le modèle de persistance sur l'indice ciel clair donnent des résultats similaires en termes de nRMSE. À mesure que l'horizon de temps augmente, nous avons constaté la supériorité des modèles fondés sur l'apprentissage automatique. Toutefois, les modèles les plus performants sont le modèle GPR fondé sur le noyau  $k_{\text{RQ-ARD}}$  et le modèle LSTM (*long short-term memory*). Notre choix s'est porté sur le modèle GPR, à l'aune des exigences de notre application. Par ailleurs, les modèles GPR fournissent naturellement des intervalles de confiance associés aux prévisions qui ont permis d'améliorer la robustesse du contrôleur prédictif développé dans le cadre du projet « Smart Occitania ».

Un processus gaussien est complètement défini par sa fonction moyenne et sa fonction de covariance, également appelé noyau. Au [Chapitre 2](#), nous avons vu que le noyau qui encode les connaissances *a priori* sur la fonction à apprendre est l'élément clé d'un modèle GPR. La question de l'entrée des modèles GPR dédiés à la prévision du GHI a été posée au [Chapitre 3](#). Un choix naturel est de considérer un historique d'observations du GHI. Cependant, comme nous l'avons vu, il est également possible de considérer le temps en entrée. Cela a permis de construire un noyau particulièrement adapté au GHI. Le modèle GPR fondé sur ce noyau est par nature multi-horizon, en ce sens qu'il peut fournir une prévision du GHI à n'importe quel horizon de temps. Il s'agit du noyau quasi-périodique  $k_{\text{Per} \times \text{RQ} + \text{Per}}$  adapté à des données couvrant plusieurs années. Pour des données couvrant seulement quelques semaines, le noyau  $k_{\text{Per} \times \text{RQ}}$  est recommandé. Ces noyaux tirent profit des biais inductifs connus sur la dynamique du GHI. Parmi les modèles GPR utilisant des observations en entrée, on distingue des modèles multi-horizons et des modèles mono-horizons. Les résultats ont montré que les modèles mono-horizons prennent le dessus sur les modèles multi-horizons. Cependant, dans le cas où de nombreux horizons sont considérés – dans notre cas 144 horizons de prévisions sont nécessaires –, il est préférable de se tourner vers les modèles multi-horizons. Parmi ces derniers, les modèles GPR utilisant le temps en entrée sont les plus performants.

Un noyau temporel ayant été identifié au [Chapitre 3](#), nous avons cherché à identifier un noyau spatial à associer à ce noyau afin de développer des modèles GPR pour la prévision spatio-temporelle du



GHI (Chapitre 4). Une étude comparative impliquant divers noyaux spatiaux avec *automatic relevance determination* (ARD) a été menée. Les résultats obtenus ont été analysés pour trois cas particuliers. Les résultats ont montré que l'on peut écarter le modèle GPR associant noyau temporel  $k_{\text{Per} \times \text{RQ}}$  et noyau spatial  $k_{\text{SE-ARD}}$ . Ce modèle a produit les moins bons résultats pour tous les horizons de temps considérés. Quant aux autres modèles GPR développés, à savoir le modèle GPR associant noyau temporel  $k_{\text{Per} \times \text{RQ}}$  et noyau spatial  $k_{\text{RQ-ARD}}$  et les modèles GPR associant noyau temporel  $k_{\text{Per} \times \text{RQ}}$  et noyaux spatiaux de la classe Matérn, nous ne sommes pas arrivés à conclure. Ils ont produit des résultats similaires. Cette étude a permis d'aboutir à des résultats préliminaires, qui peuvent être améliorés.

En perspective, dans un premier temps, augmenter la granularité de l'instrumentation installée au sein de la zone d'étude pourrait permettre d'évaluer plus finement les prévisions spatio-temporelles fournies par les modèles. Des prétraitements appropriés aux données, notamment, pour rendre réaliste l'hypothèse de séparabilité pourraient être envisagés. Ensuite, il est possible de revenir sur certaines hypothèses utilisées pour la construction des noyaux. D'une part, considérer une fonction moyenne non nulle (égale, par exemple, à l'indice ciel clair) pourrait être intéressant car cela éviterait que les prévisions tendent vers 0 lorsque l'on s'éloigne des points de mesure. D'autre part, des structures de noyau plus complexes pourraient être considérées, notamment en ne postulant plus la séparabilité entre le temps et l'espace : bien qu'impliquant une complexité calculatoire plus élevée, qui pourrait entraver l'implémentation en temps réel des algorithmes, cela permettrait une modélisation plus réaliste du comportement spatio-temporel du GHI. Enfin, les prévisions pourraient être améliorées en tirant profit d'informations provenant des images du ciel à haute dynamique prises par la caméra à grand angle de champ installée à Villeneuve-de-la-Raho. Ces images fournissent des informations sur la couverture nuageuse qui pourront être prises en compte dans les modèles afin d'anticiper les chutes brutales du GHI.



# Bibliographie

- [1] UNITED NATIONS. *Paris Agreement*. United Nations Treaty Collection, chapter XXVII, 7.d. 2015.
- [2] *2030 climate & energy framework*. <https://ec.europa.eu/clima/policies/strategies/2030>. 2021.
- [3] *Stratégie nationale bas-carbone*. Rapp. tech. Ministère de la transition écologique et solidaire, 2020.
- [4] *Bilan électrique 2019*. Rapp. tech. Gestionnaire du Réseau de Transport d'Electricité (RTE), 2020.
- [5] T. OLOWU, A. SUNDARARAJAN, M. MOGHADDAMI et A. SARWAT. "Future Challenges and Mitigation Methods for High Photovoltaic Penetration : A Survey". Dans : *Energies* 11.7 (2018), p. 1782. ISSN : 1996-1073. DOI : [10.3390/en11071782](https://doi.org/10.3390/en11071782).
- [6] N. DKHILI. "A model-based predictive control strategy for low-voltage power distribution grids with prolific distributed generation in Occitania". Thèse de doct. Perpignan, France : Université de Perpignan Via Domitia, 2020.
- [7] N. DKHILI, J. EYNARD, S. THIL et S. GRIEU. "A survey of modelling and smart management tools for power grids with prolific distributed generation". Dans : *Sustainable Energy, Grids and Networks* 21 (2020), p. 100284. ISSN : 2352-4677. DOI : <https://doi.org/10.1016/j.segan.2019.100284>.
- [8] S. A. ALEEM, S. M. S. HUSSAIN et T. S. USTUN. "A Review of Strategies to Increase PV Penetration Level in Smart Grids". Dans : *Energies* 13.3 (2020). ISSN : 1996-1073. DOI : [10.3390/en13030636](https://doi.org/10.3390/en13030636).
- [9] M. L. BRUSSEAU, M. RAMIREZ-ANDREOTTA, I. L. PEPPER et J. MAXIMILLIAN. "Chapter 26 - Environmental Impacts on Human Health and Well-Being". Dans : *Environmental and Pollution Science (Third Edition)*. Sous la dir. de M. L. BRUSSEAU, I. L. PEPPER et C. P. GERBA. Third Edition. Academic Press, 2019, p. 477-499. ISBN : 978-0-12-814719-1. DOI : <https://doi.org/10.1016/B978-0-12-814719-1.00026-4>.
- [10] A. B. A. BOXALL. "Global Climate Change and Environmental Toxicology". Dans : *Encyclopedia of Toxicology (Third Edition)*. Sous la dir. de P. WEXLER. Third Edition. Oxford : Academic Press, 2014, p. 736-740. ISBN : 978-0-12-386455-0. DOI : <https://doi.org/10.1016/B978-0-12-386454-3.00208-6>.
- [11] *Changements climatiques 2014 : L'atténuation du changement climatique*. Rapp. tech. Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC), 2014.
- [12] M. HÖÖK et X. TANG. "Depletion of fossil fuels and anthropogenic climate change—A review". Dans : *Energy Policy* 52 (2013). Special Section : Transition Pathways to a Low Carbon Economy, p. 797-809. ISSN : 0301-4215. DOI : <https://doi.org/10.1016/j.enpol.2012.10.046>.
- [13] *Ressources et consommation énergétiques mondiales*. [https://fr.wikipedia.org/wiki/Ressources\\_et\\_consommation\\_energetiques\\_mondiales](https://fr.wikipedia.org/wiki/Ressources_et_consommation_energetiques_mondiales). 2021.

- [14] *International Energy Outlook 2019*. Rapp. tech. U.S. Energy Information Administration, 2019.
- [15] *Electricity in Europe 2017*. Rapp. tech. European Network of Transmission System Operators for Electricity, 2017.
- [16] D. FANTAZZINI, M. HÖÖK et A. ANGELANTONI. “Global oil risks in the early 21st century”. Dans : *Energy Policy* 39.12 (2011). Clean Cooking Fuels and Technologies in Developing Economies, p. 7865-7873. ISSN : 0301-4215. DOI : <https://doi.org/10.1016/j.enpol.2011.09.035>.
- [17] S. SHAFIEE et E. TOPAL. “When will fossil fuel reserves be diminished?” Dans : *Energy Policy* 37.1 (2009), p. 181-189. ISSN : 0301-4215. DOI : <https://doi.org/10.1016/j.enpol.2008.08.016>.
- [18] B. R. SINGH et O. SINGH. “Global Trends of Fossil Fuel Reserves and Climate Change in the 21st Century”. Dans : *Fossil Fuel and the Environment*. Sous la dir. de S. KHAN. Rijeka : IntechOpen, 2012. Chap. 8. DOI : [10.5772/38655](https://doi.org/10.5772/38655).
- [19] S. SORRELL, J. SPEIRS, R. BENTLEY, A. BRANDT et R. MILLER. “Global oil depletion : A review of the evidence”. Dans : *Energy Policy* 38.9 (2010). Special Section on Carbon Emissions and Carbon Management in Cities with Regular Papers, p. 5290-5295. ISSN : 0301-4215. DOI : <https://doi.org/10.1016/j.enpol.2010.04.046>.
- [20] R. W. BENTLEY. “Global oil & gas depletion : an overview”. Dans : *Energy Policy* 30.3 (2002), p. 189-205. ISSN : 0301-4215. DOI : [https://doi.org/10.1016/S0301-4215\(01\)00144-6](https://doi.org/10.1016/S0301-4215(01)00144-6).
- [21] *Énergie renouvelable*. [https://fr.wikipedia.org/wiki/Énergie\\_renouvelable](https://fr.wikipedia.org/wiki/Énergie_renouvelable). 2021.
- [22] *Renewables 2020 : Analysis and forecast to 2025*. Rapp. tech. U.S. Energy Information Administration, 2020.
- [23] J. BUONOCORE. “Climate policy not so costly”. Dans : *Nature Climate Change* 4.10 (2014), p. 861-862. ISSN : 1758-6798. DOI : [10.1038/nclimate2391](https://doi.org/10.1038/nclimate2391).
- [24] Y.-B. CHENG, A. PASCOE, F. HUANG et Y. PENG. “Print flexible solar cells”. Dans : *Nature* 539.7630 (2016), p. 488-489. ISSN : 1476-4687. DOI : [10.1038/539488a](https://doi.org/10.1038/539488a).
- [25] *Solar PV global capacity*. <https://ourworldindata.org/grapher/installed-solar-pv-capacity?tab=table>. 2021.
- [26] *Panorama de l'électricité renouvelable au 31 décembre 2020*. Rapp. tech. Gestionnaire du Réseau de Transport d'Electricité (RTE), 2021.
- [27] ENEDIS. *Principes d'étude et de développement du réseau pour le raccordement des clients et producteurs BT*. Rapp. tech. 2018.
- [28] J. M. VINDEL et J. POLO. “Intermittency and variability of daily solar irradiation”. Dans : *Atmospheric Research* 143 (2014), p. 313-327. ISSN : 0169-8095. DOI : <https://doi.org/10.1016/j.atmosres.2014.03.001>.
- [29] V. COURTECUISSÉ. “Supervision d'une centrale multisources à base d'éoliennes et de stockage d'énergie connectée au réseau électrique”. Thèse de doct. Arts et Métiers ParisTech, 2008.
- [30] W. VICENTE. “Modélisation des réseaux de distribution sous incertitudes”. Thèse de doct. Sept. 2012.
- [31] ALVAREZ-HÉRAULT, MARIE-CÉCILE. “ARCHITECTURES DES RÉSEAUX DE DISTRIBUTION DU FUTUR EN PRÉSENCE DE PRODUCTION DÉCENTRALISÉE”. Thèse de doct. Institut National Polytechnique de Grenoble - INPG, 2009.

- [32] P. P. BARKER et R. W. D. MELLO. “Determining the impact of distributed generation on power systems. I. Radial distribution systems”. Dans : *2000 Power Engineering Society Summer Meeting (Cat. No.00CH37134)*. T. 3. 2000, p. 1645-1656. DOI : [10.1109/PSS.2000.868775](https://doi.org/10.1109/PSS.2000.868775).
- [33] M. S. ELNOZAHY et M. M. A. SALAMA. “Technical impacts of grid-connected photovoltaic systems on electrical networks—A review”. Dans : *Journal of Renewable and Sustainable Energy* 5.3 (2013), p. 032702. DOI : [10.1063/1.4808264](https://doi.org/10.1063/1.4808264). eprint : <https://doi.org/10.1063/1.4808264>.
- [34] M. A. ELTAWIL et Z. ZHAO. “Grid-connected photovoltaic power systems : Technical and potential problems - A review”. Dans : *Renewable and Sustainable Energy Reviews* 14.1 (2010), p. 112-129. ISSN : 1364-0321. DOI : <https://doi.org/10.1016/j.rser.2009.07.015>.
- [35] K. N. NWAIGWE, P. MUTABILWA et E. DINTWA. “An overview of solar power (PV systems) integration into electricity grids”. Dans : *Materials Science for Energy Technologies* 2.3 (2019), 629-633. ISSN : 2589-2991. DOI : <https://doi.org/10.1016/j.mset.2019.07.002>.
- [36] R. CHAUVIN. “Evaluation de la ressource solaire pour la gestion optimisée de centrales CSP”. Thèse de doct. Université de Perpignan Via Domitia, 2016.
- [37] H. M. DIAGNE. “Gestion intelligente du réseau électrique réunionnais. Prévion de la ressource solaire en milieu insulaire”. Thèse de doct. Université de la Réunion, 2015.
- [38] *Carte des projets de démonstrateurs portés par ENEDIS*. <http://smartgrids-enedis.fr/>. 2021.
- [39] N. DKHILI, D. SALAS, J. EYNARD, S. THIL et S. GRIEU. “Innovative Application of Model-Based Predictive Control for Low-Voltage Power Distribution Grids with Significant Distributed Generation”. Dans : *Energies* 14.6 (2021). ISSN : 1996-1073. DOI : [10.3390/en14061773](https://doi.org/10.3390/en14061773).
- [40] N. DKHILI, J. EYNARD, S. THIL et S. GRIEU. “Resilient Predictive Control Coupled with a Worst-Case Scenario Approach for a Distributed-Generation-Rich Power Distribution Grid”. Dans : *Clean Technologies* 3.3 (2021), p. 629-655. ISSN : 2571-8797. DOI : [10.3390/cleantechnol3030038](https://doi.org/10.3390/cleantechnol3030038).
- [41] R. CHAUVIN, J. NOU, S. THIL et S. GRIEU. *System for measuring components of solar radiation*. Patent WO2019053232. 2019.
- [42] R. CHAUVIN, J. NOU, S. THIL et S. GRIEU. *Système de mesure des composantes du rayonnement solaire*. Brevet FR1758549. 2019.
- [43] F. ANTONANZAS-TORRES, R. URRACA, J. POLO, O. PERPIÑÁN-LAMIGUEIRO et R. ESCOBAR. “Clear sky solar irradiance models : A review of seventy models”. Dans : *Renewable and Sustainable Energy Reviews* 107 (2019), 374-387. ISSN : 1364-0321. DOI : <https://doi.org/10.1016/j.rser.2019.02.032>.
- [44] R. CHAUVIN, J. NOU, J. EYNARD, S. THIL et S. GRIEU. “A new approach to the real-time assessment and intraday forecasting of clear-sky direct normal irradiance”. Dans : *Solar Energy* 167 (2018), p. 35-51. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2018.02.027>.
- [45] P. INEICHEN et R. PEREZ. “A new airmass independent formulation for the Linke turbidity coefficient”. Dans : *Solar Energy* 73.3 (2002), 151-157. ISSN : 0038-092X. DOI : [https://doi.org/10.1016/S0038-092X\(02\)00045-2](https://doi.org/10.1016/S0038-092X(02)00045-2).
- [46] F. LINKE. “Transmissions-Koeffizient und Trübungs faktor”. Dans : *Beiträge zur Physik der Atmosphäre* 10 (1922), p. 91-103.
- [47] G. KOPP. “An assessment of the solar irradiance record for climate studies”. Dans : *J. Space Weather Space Clim.* 4 (2014), A14. DOI : [10.1051/swsc/2014012](https://doi.org/10.1051/swsc/2014012).

- [48] J. L. LEAN et D. H. RIND. “How will Earth’s surface temperature change in future decades ?” Dans : *Geophysical Research Letters* 36.15 (2009). DOI : <https://doi.org/10.1029/2009GL038932>. eprint : <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2009GL038932>.
- [49] H. LI, Y. LIAN, X. WANG, W. MA et L. ZHAO. “Solar constant values for estimating solar radiation”. Dans : *Energy* 36.3 (2011), p. 1785-1789. ISSN : 0360-5442. DOI : <https://doi.org/10.1016/j.energy.2010.12.050>.
- [50] G. KOPP et J. L. LEAN. “A new, lower value of total solar irradiance : Evidence and climate significance”. Dans : *Geophysical Research Letters* 38.1 (2011). DOI : <https://doi.org/10.1029/2010GL045777>. eprint : <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2010GL045777>.
- [51] C. A. GUEYMARD. “A reevaluation of the solar constant based on a 42-year total solar irradiance time series and a reconciliation of spaceborne observations”. Dans : *Solar Energy* 168 (2018). Advances in Solar Resource Assessment and Forecasting, p. 2-9. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2018.04.001>.
- [52] P. BLANC et L. WALD. “The SG2 algorithm for a fast and accurate computation of the position of the Sun for multi-decadal time period”. Dans : *Solar Energy* 86.10 (2012), p. 3072-3083. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2012.07.018>.
- [53] C. RIGOLLIER, O. BAUER et L. WALD. “On the clear sky model of the ESRA – European Solar Radiation Atlas – with respect to the heliosat method”. Dans : *Solar Energy* 68.1 (2000), p. 33-48. ISSN : 0038-092X. DOI : [https://doi.org/10.1016/S0038-092X\(99\)00055-9](https://doi.org/10.1016/S0038-092X(99)00055-9).
- [54] L. WALD. “Solar radiation energy (fundamentals).” Dans : sous la dir. de J. BLANCO et S. MALATO. Encyclopedia of Life Support Systems (EOLSS), 2007.
- [55] Í. RAPP-ARRARÁS et J. M. DOMINGO-SANTOS. “Functional forms for approximating the relative optical air mass”. Dans : *Journal of Geophysical Research : Atmospheres* 116.D24 (2011). DOI : <https://doi.org/10.1029/2011JD016706>. eprint : <https://agupubs.onlinelibrary.wiley.com/doi/pdf/10.1029/2011JD016706>.
- [56] F. KASTEN et A. T. YOUNG. “Revised optical air mass tables and approximation formula”. Dans : *Applied Optics* 28.22 (1989), p. 4735-4738. DOI : [10.1364/AO.28.004735](https://doi.org/10.1364/AO.28.004735).
- [57] *KIPP & ZONEN CM5*. <https://www.kippzonen.com/News/963/46-year-old-CM5.2021>.
- [58] B. ESPINAR, P. BLANC, L. WALD, C. HOYER-KLICK, M. SCHROEDTER-HOMSCHEIDT et T. WANDERER. “Controlling the quality of measurements for meteorological variables and solar radiation. From sub-hourly to monthly average time periods”. Dans : avr. 2012.
- [59] M. GEIGER, L. DIABATÉ, L. MÉNARD et L. WALD. “A web service for controlling the quality of measurements of global solar irradiation”. Dans : *Solar Energy* 73.6 (2002), p. 475-480. ISSN : 0038-092X. DOI : [https://doi.org/10.1016/S0038-092X\(02\)00121-4](https://doi.org/10.1016/S0038-092X(02)00121-4).
- [60] T. MUNEER et F. FAIROOZ. “Quality control of solar radiation and sunshine measurements – lessons learnt from processing worldwide databases”. Dans : *Building Services Engineering Research and Technology* 23.3 (2002), p. 151-166. DOI : [10.1191/0143624402bt038oa](https://doi.org/10.1191/0143624402bt038oa). eprint : <https://doi.org/10.1191/0143624402bt038oa>.
- [61] A. ROESCH, M. WILD, A. OHMURA, E. G. DUTTON, C. N. LONG et T. ZHANG. “Assessment of BSRN radiation records for the computation of monthly means”. Dans : *Atmospheric Measurement Techniques* 4.2 (2011), p. 339-354. DOI : [10.5194/amt-4-339-2011](https://doi.org/10.5194/amt-4-339-2011).
- [62] C. VOYANT, T. SOUBDHAN, P. LAURET, M. DAVID et M. MUSELLI. “Statistical parameters as a means to a priori assess the accuracy of solar forecasting models”. Dans : *Energy* 90 (2015), p. 671-679. ISSN : 0360-5442. DOI : <https://doi.org/10.1016/j.energy.2015.07.089>.

- [63] M. DIAGNE, M. DAVID, P. LAURET, J. BOLAND et N. SCHMUTZ. “Review of solar irradiance forecasting methods and a proposition for small-scale insular grids”. Dans : *Renewable and Sustainable Energy Reviews* 27 (2013), p. 65-76. ISSN : 1364-0321.
- [64] G. NOTTON et C. VOYANT. “Chapter 3 - Forecasting of Intermittent Solar Energy Resource”. Dans : *Advances in Renewable Energies and Power Technologies*. Sous la dir. d’I. YAHYAOU. Elsevier, 2018, p. 77-14. ISBN : 978-0-12-812959-3. DOI : <https://doi.org/10.1016/B978-0-12-812959-3.00003-4>.
- [65] R. H. INMAN, H. T. C. PEDRO et C. F. M. COIMBRA. “Solar forecasting methods for renewable energy integration”. Dans : *Progress in Energy and Combustion Science* 39.6 (2013), p. 535-576. ISSN : 0360-1285. DOI : <https://doi.org/10.1016/j.pecs.2013.06.002>.
- [66] R. AHMED, V. SREERAM, Y. MISHRA et M. D. ARIF. “A review and evaluation of the state-of-the-art in PV solar power forecasting : Techniques and optimization”. Dans : *Renewable and Sustainable Energy Reviews* 124 (2020), p. 109792. ISSN : 1364-0321. DOI : <https://doi.org/10.1016/j.rser.2020.109792>.
- [67] P. LAURET, C. VOYANT, T. SOUBDHAN, M. DAVID et P. POGGI. “A benchmarking of machine learning techniques for solar radiation forecasting in an insular context”. Dans : *Solar Energy* 112 (2015), p. 446-457. ISSN : 0038-092X.
- [68] J. ANTONANZAS, N. OSORIO, R. ESCOBAR, R. URRACA, F. J. M. de PISON et F. ANTONANZAS-TORRES. “Review of photovoltaic power forecasting”. Dans : *Solar Energy* 136 (2016), p. 78-111. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2016.06.069>.
- [69] S. SOBRI, S. KOOHI-KAMALI et N. A. RAHIM. “Solar photovoltaic generation forecasting methods : A review”. Dans : *Energy Conversion and Management* 156 (2018), p. 459-497. ISSN : 0196-8904. DOI : <https://doi.org/10.1016/j.enconman.2017.11.019>.
- [70] H. T. C. PEDRO et C. F. M. COIMBRA. “Assessment of forecasting techniques for solar power production with no exogenous inputs”. Dans : *Solar Energy* 86.7 (2012), p. 2017-2028. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2012.04.004>.
- [71] C. VOYANT, G. NOTTON, S. KALOGIROU, M.-L. NIVET, C. PAOLI, F. MOTTE et A. FOUILLOY. “Machine learning methods for solar radiation forecasting : A review”. Dans : *Renewable Energy* 105 (2017), p. 569-582. ISSN : 0960-1481. DOI : <https://doi.org/10.1016/j.renene.2016.12.095>.
- [72] G. M. YAGLI, D. YANG et D. SRINIVASAN. “Automatic hourly solar forecasting using machine learning models”. Dans : *Renewable and Sustainable Energy Reviews* 105 (2019), p. 487-498. ISSN : 1364-0321. DOI : <https://doi.org/10.1016/j.rser.2019.02.006>.
- [73] A. MELLIT, A. MASSI PAVAN, E. OGLIARI, S. LEVA et V. LUGHI. “Advanced Methods for Photovoltaic Output Power Forecasting : A Review”. Dans : *Applied Sciences* 10.2 (2020). ISSN : 2076-3417. DOI : [10.3390/app10020487](https://doi.org/10.3390/app10020487).
- [74] L. RAMÍREZ et J. M. VINDEL. “13 - Forecasting and nowcasting of DNI for concentrating solar thermal systems”. Dans : *Advances in Concentrating Solar Thermal Research and Technology*. Sous la dir. de M. J. BLANCO et L. R. SANTIGOSA. Woodhead Publishing Series in Energy. Woodhead Publishing, 2017, p. 293-310. ISBN : 978-0-08-100516-3. DOI : <https://doi.org/10.1016/B978-0-08-100516-3.00013-7>.
- [75] E. LORENZ et D. HEINEMANN. “Prediction of solar irradiance and photovoltaic power”. Dans : *Comprehensive Renewable Energy*. Sous la dir. d’A. SAYIGH. Oxford : Elsevier, 2012, p. 239-292. ISBN : 978-0-08-087873-7. DOI : <https://doi.org/10.1016/B978-0-08-087872-0.00114-1>.



- [76] J. NOU, R. CHAUVIN, J. EYNARD, S. THIL et S. GRIEU. “Towards the intrahour forecasting of direct normal irradiance using sky-imaging data”. Dans : *Heliyon* 4.4 (2018), e00598. ISSN : 2405-8440. DOI : <https://doi.org/10.1016/j.heliyon.2018.e00598>.
- [77] Y. CHU, H. T. C. PEDRO, L. NONNENMACHER, R. H. INMAN, Z. LIAO et C. F. M. COIMBRA. “A Smart Image-Based Cloud Detection System for Intrahour Solar Irradiance Forecasts”. Dans : *Journal of Atmospheric and Oceanic Technology* 31.9 (2014), p. 1995-2007.
- [78] R. MARQUEZ et C. F. M. COIMBRA. “Intra-hour DNI forecasting based on cloud tracking image analysis”. Dans : *Solar Energy* 91 (2013), p. 327-336. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2012.09.018>.
- [79] S. QUESADA-RUIZ, Y. CHU, J. TOVAR-PESCADOR, H. T. C. PEDRO et C. F. M. COIMBRA. “Cloud-tracking methodology for intra-hour DNI forecasting”. Dans : *Solar Energy* 102 (2014), p. 267-275. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2014.01.030>.
- [80] Y. CHU, H. T. C. PEDRO, M. LI et C. F. M. COIMBRA. “Real-time forecasting of solar irradiance ramps with smart image processing”. Dans : *Solar Energy* 114 (2015), p. 91-104. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2015.01.024>.
- [81] Y. CHU, H. T. C. PEDRO et C. F. M. COIMBRA. “Hybrid intra-hour DNI forecasts with sky image processing enhanced by stochastic learning”. Dans : *Solar Energy* 98 (2013), p. 592-603. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2013.10.020>.
- [82] M. LI, Y. CHU, H. T. C. PEDRO et C. F. M. COIMBRA. “Quantitative evaluation of the impact of cloud transmittance and cloud velocity on the accuracy of short-term DNI forecasts”. Dans : *Renewable Energy* 86 (2016), p. 1362-1371. ISSN : 0960-1481. DOI : <https://doi.org/10.1016/j.renene.2015.09.058>.
- [83] V. BONE, J. PIDGEON, M. KEARNEY et A. VEERARAGAVAN. “Intra-hour direct normal irradiance forecasting through adaptive clear-sky modelling and cloud tracking”. Dans : *Solar Energy* 159 (2018), p. 852-867. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2017.10.037>.
- [84] S. R. WEST, D. ROWE, S. SAYEEF et A. BERRY. “Short-term irradiance forecasting using skycams : Motivation and development”. Dans : *Solar Energy* 110 (2014), p. 188-207. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2014.08.038>.
- [85] D. BERNECKER, C. RIESS, E. ANGELOPOULOU et J. HORNEGGER. “Continuous short-term irradiance forecasts using sky images”. Dans : *Solar Energy* 110 (2014), p. 303-315. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2014.09.005>.
- [86] C. W. CHOW, B. URQUHART, M. LAVE, A. DOMINGUEZ, J. KLEISSL, J. SHIELDS et B. WASHOM. “Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed”. Dans : *Solar Energy* 85 (2011), p. 2881-2893.
- [87] H. YANG, B. KURTZ, D. NGUYEN, B. URQUHART, C. W. CHOW, M. GHONIMA et J. KLEISSL. “Solar irradiance forecasting using a ground-based sky imager developed at UC San Diego”. Dans : *Solar Energy* 103 (2014), p. 502-524. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2014.02.044>.
- [88] C.-L. FU et H.-Y. CHENG. “Predicting solar irradiance with all-sky image features via regression”. Dans : *Solar Energy* 97 (2013), p. 537-550. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2013.09.016>.
- [89] T. SCHMIDT, J. KALISCH, E. LORENZ et D. HEINEMANN. “Evaluating the spatio-temporal performance of sky-imager-based solar irradiance analysis and forecasts”. Dans : *Atmospheric Chemistry and Physics* 16.5 (2016), p. 3399-3412. DOI : [10.5194/acp-16-3399-2016](https://doi.org/10.5194/acp-16-3399-2016).

- [90] K. STEFFERUD, J. KLEISSL et J. SCHOENE. “Solar forecasting and variability analyses using sky camera cloud detection and motion vectors”. Dans : *2012 IEEE Power and Energy Society General Meeting*. 2012, p. 1-6. DOI : [10.1109/PESGM.2012.6345434](https://doi.org/10.1109/PESGM.2012.6345434).
- [91] M. CALDAS et R. ALONSO-SUÁREZ. “Very short-term solar irradiance forecast using all-sky imaging and real-time irradiance measurements”. Dans : *Renewable Energy* 143 (2019), p. 1643-1658. ISSN : 0960-1481. DOI : <https://doi.org/10.1016/j.renene.2019.05.069>.
- [92] A. AL-LAHAM, O. THEEB, K. ELALEM, T. A. ALSHAWI et S. A. ALSHEBEILI. “Sky Imager-Based Forecast of Solar Irradiance Using Machine Learning”. Dans : *Electronics* 9.10 (2020). ISSN : 2079-9292. DOI : [10.3390/electronics9101700](https://doi.org/10.3390/electronics9101700).
- [93] C. FENG et J. ZHANG. “SolarNet : A sky image-based deep convolutional neural network for intra-hour solar forecasting”. Dans : *Solar Energy* 204 (2020), p. 71-78. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2020.03.083>.
- [94] R. CHAUVIN, J. NOU, S. THIL, A. TRAORÉ et S. GRIEU. “Cloud Detection Methodology Based on a Sky-imaging System”. Dans : *Energy Procedia* 69 (2015). International Conference on Concentrating Solar Power and Chemical Energy Systems, SolarPACES 2014, p. 1970-1980. ISSN : 1876-6102. DOI : <https://doi.org/10.1016/j.egypro.2015.03.198>.
- [95] Y. KAROUT, J. EYNARD, S. THIL et S. GRIEU. “Cloud/sky segmentation from ground camera based on supervised machine learning approach”. Dans : *33rd International Conference on Efficiency, Cost, Optimization, Simulation and Environmental Impact of Energy Systems (ECOS 2020)*. Osaka, Japan, 2020.
- [96] J. ALONSO, F. J. BATLLES, G. LÓPEZ et A. TERNERO. “Sky camera imagery processing based on a sky classification using radiometric data”. Dans : *Energy* 68 (2014), p. 599-608. ISSN : 0360-5442. DOI : <https://doi.org/10.1016/j.energy.2014.02.035>.
- [97] R. CHAUVIN, J. NOU, S. THIL et S. GRIEU. “Modelling the clear-sky intensity distribution using a sky imager”. Dans : *Solar Energy* 119 (2015), p. 1-17. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2015.06.026>.
- [98] A. CAZORLA, C. HUSILLOS, M. ANTÓN et L. ALADOS-ARBOLEDAS. “Multi-exposure adaptive threshold technique for cloud detection with sky imagers”. Dans : *Solar Energy* 114 (2015), p. 268-277. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2015.02.006>.
- [99] P. WANG, R. VAN WESTRHENEN, J. F. MEIRINK, S. VAN DER VEEN et W. KNAP. “Surface solar radiation forecasts by advecting cloud physical properties derived from Meteosat Second Generation observations”. Dans : *Solar Energy* 177 (2019), p. 47-58. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2018.10.073>.
- [100] V. KALLIO-MYERS, A. RIIHELÄ, P. LAHTINEN et A. LINDFORS. “Global horizontal irradiance forecast for Finland based on geostationary weather satellite data”. Dans : *Solar Energy* 198 (2020), p. 68-80. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2020.01.008>.
- [101] R. PEREZ, K. MOORE, S. WILCOX, D. RENNÉ et A. ZELENKA. “Forecasting solar radiation – Preliminary evaluation of an approach based upon the national forecast database”. Dans : *Solar Energy* 81.6 (2007), p. 809-812. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2006.09.009>.
- [102] R. PEREZ, S. KIVALOV, J. SCHLEMMER, K. HEMKER, D. RENNÉ et T. E. HOFF. “Validation of short and medium term operational solar radiation forecasts in the US”. Dans : *Solar Energy* 84.12 (2010), p. 2161-2172. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2010.08.014>.



- [103] D. CANO, J. M. MONGET, M. ALBUISSON, H. GUILLARD, N. REGAS et L. WALD. "A method for the determination of the global solar radiation from meteorological satellite data". Dans : *Solar Energy* 37.1 (1986), p. 31-39. ISSN : 0038-092X. DOI : [https://doi.org/10.1016/0038-092X\(86\)90104-0](https://doi.org/10.1016/0038-092X(86)90104-0).
- [104] A. A. LACIS et J. HANSEN. "A Parameterization for the Absorption of Solar Radiation in the Earth's Atmosphere". Dans : *Journal of Atmospheric Sciences* 31.1 (1974), p. 118-133.
- [105] W. MÖSER et E. RASCHKE. "Incident Solar Radiation over Europe Estimated from METEOSAT Data". Dans : *Journal of Applied Meteorology and Climatology* 23.1 (1984), p. 166-170.
- [106] C. G. JUSTUS, M. V. PARIS et J. D. TARPLEY. "Satellite-measured insolation in the United States, Mexico, and South America". Dans : *Remote Sensing of Environment* 20.1 (1986), p. 57-83. ISSN : 0034-4257. DOI : [https://doi.org/10.1016/0034-4257\(86\)90014-3](https://doi.org/10.1016/0034-4257(86)90014-3).
- [107] J. D. TARPLEY. "Estimating Incident Solar Radiation at the Surface from Geostationary Satellite Data". Dans : *Journal of Applied Meteorology and Climatology* 18.9 (1979), p. 1172-1181.
- [108] L. MAZORRA AGUIAR, B. PEREIRA, M. DAVID, F. DÍAZ et P. LAURET. "Use of satellite data to improve solar radiation forecasting with Bayesian Artificial Neural Networks". Dans : *Solar Energy* 122 (2015), p. 1309-1324. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2015.10.041>.
- [109] S. CROS, O. LIANDRAT, N. SÉBASTIEN et N. SCHMUTZ. "Extracting cloud motion vectors from satellite images for solar power forecasting". Dans : *2014 IEEE Geoscience and Remote Sensing Symposium*. 2014, p. 4123-4126. DOI : [10.1109/IGARSS.2014.6947394](https://doi.org/10.1109/IGARSS.2014.6947394).
- [110] A. HAMMER, D. HEINEMANN, E. LORENZ et B. LÜCKEHE. "Short-term forecasting of solar radiation : a statistical approach using satellite data". Dans : *Solar Energy* 67.1 (1999), p. 139-150. ISSN : 0038-092X. DOI : [https://doi.org/10.1016/S0038-092X\(00\)00038-4](https://doi.org/10.1016/S0038-092X(00)00038-4).
- [111] T. M. HAMILL et T. NEHRKORN. "A Short-Term Cloud Forecast Scheme Using Cross Correlations". Dans : *Weather and Forecasting* 8.4 (1993), p. 401-411.
- [112] R. PEREZ, P. INEICHEN, K. MOORE, M. KMIETIK, C. CHAIN, R. GEORGE et F. VIGNOLA. "A new operational model for satellite-derived irradiances : description and validation". Dans : *Solar Energy* 73.5 (2002), p. 307-317. ISSN : 0038-092X. DOI : [https://doi.org/10.1016/S0038-092X\(02\)00122-6](https://doi.org/10.1016/S0038-092X(02)00122-6).
- [113] R. PEREZ, P. INEICHEN, M. KMIETIK, K. MOORE, D. RENNE et R. GEORGE. "Producing satellite-derived irradiances in complex arid terrain". Dans : *Solar Energy* 77.4 (2004). The American Solar Energy Society's Solar 2003 Special Issue, p. 367-371. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2003.12.016>.
- [114] P. MATHIESEN, C. COLLIER et J. KLEISSL. "A high-resolution, cloud-assimilating numerical weather prediction model for solar irradiance forecasting". Dans : *Solar Energy* 92 (2013), p. 47-61. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2013.02.018>.
- [115] E. LORENZ, J. HURKA, D. HEINEMANN et H. G. BEYER. "Irradiance Forecasting for the Power Prediction of Grid-Connected Photovoltaic Systems". Dans : *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 2.1 (2009), p. 2-10. ISSN : 2151-1535. DOI : [10.1109/JSTARS.2009.2020300](https://doi.org/10.1109/JSTARS.2009.2020300).
- [116] E. LORENZ, J. KÜHNERT et D. HEINEMANN. "Overview of Irradiance and Photovoltaic Power Prediction". Dans : *Weather Matters for Energy*. Sous la dir. d'A. TROCCOLI, L. DUBUS et S. E. HAUPT. New York, NY : Springer New York, 2014, p. 429-454. ISBN : 978-1-4614-9221-4. DOI : [10.1007/978-1-4614-9221-4\\_21](https://doi.org/10.1007/978-1-4614-9221-4_21).
- [117] P. LAURET, E. LORENZ et M. DAVID. "Solar Forecasting in a Challenging Insular Context". Dans : *Atmosphere* 7.2 (2016). ISSN : 2073-4433. DOI : [10.3390/atmos7020018](https://doi.org/10.3390/atmos7020018).

- [118] R. PEREZ, E. LORENZ, S. PELLAND, M. BEAUHARNOIS, G. VAN KNOWE, K. HEMKER, D. HEINEMANN, J. REMUND, S. C. MÜLLER, W. TRAUNMÜLLER, G. STEINMAUER, D. POZO, J. A. RUIZ-ARIAS, V. LARA-FANEGO, L. RAMIREZ-SANTIGOSA, M. GASTON-ROMERO et L. M. POMARES. “Comparison of numerical weather prediction solar irradiance forecasts in the US, Canada and Europe”. Dans : *Solar Energy* 94 (2013), p. 305-326. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2013.05.005>.
- [119] M. DIAGNE, M. DAVID, J. BOLAND, N. SCHMUTZ et P. LAURET. “Post-processing of solar irradiance forecasts from WRF model at Reunion Island”. Dans : *Solar Energy* 105 (2014), p. 99-108. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2014.03.016>.
- [120] P. MATHIESEN et J. KLEISSL. “Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States”. Dans : *Solar Energy* 85.5 (2011), p. 967-977. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2011.02.013>.
- [121] C. CHEN, S. DUAN, T. CAI et B. LIU. “Online 24-h solar power forecasting based on weather type classification using artificial neural network”. Dans : *Solar Energy* 85.11 (2011), p. 2856-2870. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2011.08.027>.
- [122] C. VOYANT, C. PAOLI, M. MUSELLI et M.-L. NIVET. “Multi-horizon solar radiation forecasting for Mediterranean locations using time series models”. Dans : *Renewable and Sustainable Energy Reviews* 28 (2013), p. 44-52. ISSN : 1364-0321. DOI : <https://doi.org/10.1016/j.rser.2013.07.058>.
- [123] B. BELMAHDI, M. LOUZAZNI et A. E. BOUARDI. “One month-ahead forecasting of mean daily global solar radiation using time series models”. Dans : *Optik* 219 (2020), p. 165207. ISSN : 0030-4026. DOI : <https://doi.org/10.1016/j.ijleo.2020.165207>.
- [124] J. BOLAND, M. DAVID et P. LAURET. “Short term solar radiation forecasting : Island versus continental sites”. Dans : *Energy* 113 (2016), p. 186-192. ISSN : 0360-5442. DOI : <https://doi.org/10.1016/j.energy.2016.06.139>.
- [125] M. DAVID, F. RAMAHATANA, P. J. TROMBE et P. LAURET. “Probabilistic forecasting of the solar irradiance with recursive ARMA and GARCH models”. Dans : *Solar Energy* 133 (2016), p. 55-72. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2016.03.064>.
- [126] A. FOUILLOY, C. VOYANT, G. NOTTON, F. MOTTE, C. PAOLI, M.-L. NIVET, E. GUILLOT et J.-L. DUCHAUD. “Solar irradiation prediction with machine learning : Forecasting models selection method depending on weather variability”. Dans : *Energy* 165 (2018), p. 620-629. ISSN : 0360-5442. DOI : <https://doi.org/10.1016/j.energy.2018.09.116>.
- [127] Ü. AĞBULUT, A. E. GÜREL et Y. BIÇEN. “Prediction of daily global solar radiation using different machine learning algorithms : Evaluation and comparison”. Dans : *Renewable and Sustainable Energy Reviews* 135 (2021), p. 110114. ISSN : 1364-0321. DOI : <https://doi.org/10.1016/j.rser.2020.110114>.
- [128] H. T. C. PEDRO et C. F. M. COIMBRA. “Nearest-neighbor methodology for prediction of intra-hour global horizontal and direct normal irradiances”. Dans : *Renewable Energy* 80 (2015), p. 770-782. ISSN : 0960-1481. DOI : <https://doi.org/10.1016/j.renene.2015.02.061>.
- [129] H. T. C. PEDRO et C. F. M. COIMBRA. “Short-term irradiance forecastability for various solar micro-climates”. Dans : *Solar Energy* 122 (2015), p. 587-602. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2015.09.031>.
- [130] A. INANLOUGANJI, A. T. REDDY et S. KATIPAMULA. “Evaluation of regression and neural network models for solar forecasting over different short-term horizons”. Dans : *Science and Technology for the Built Environment* 24.9 (2018), p. 1004-1013. ISSN : 2374-474X.

- [131] S. GBÉMOU, J. EYNARD, S. THIL, E. GUILLOT et S. GRIEU. “A Comparative Study of Machine Learning-Based Methods for Global Horizontal Irradiance Forecasting”. Dans : *Energies* 14.11 (2021). ISSN : 1996-1073. DOI : [10.3390/en14113192](https://doi.org/10.3390/en14113192).
- [132] L. BENALI, G. NOTTON, A. FOUILLOY, C. VOYANT et R. DIZENE. “Solar radiation forecasting using artificial neural network and random forest methods : Application to normal beam, horizontal diffuse and global components”. Dans : *Renewable Energy* 132 (2019), p. 871-884. ISSN : 0960-1481.
- [133] B. AMROUCHE et X. L. PIVERT. “Artificial neural network based daily local forecasting for global solar radiation”. Dans : *Applied Energy* 130 (2014), p. 333-341. ISSN : 0306-2619.
- [134] A. ZENDEHBOUDI, M. A. BASEER et R. SAIDUR. “Application of support vector machine models for forecasting solar and wind energy resources : A review”. Dans : *Journal of Cleaner Production* 199 (2018), p. 272-285. ISSN : 0959-6526. DOI : <https://doi.org/10.1016/j.jclepro.2018.07.164>.
- [135] J. LI, J. K. WARD, J. TONG, L. COLLINS et G. PLATT. “Machine learning for solar irradiance forecasting of photovoltaic system”. Dans : *Renewable Energy* 90 (2016), p. 542-553. ISSN : 0960-1481. DOI : <https://doi.org/10.1016/j.renene.2015.12.069>.
- [136] K. Y. BAE, H. S. JANG et D. K. SUNG. “Hourly Solar Irradiance Prediction Based on Support Vector Machine and Its Error Analysis”. Dans : *IEEE Transactions on Power Systems* 32.2 (2017), p. 935-945. ISSN : 1558-0679. DOI : [10.1109/TPWRS.2016.2569608](https://doi.org/10.1109/TPWRS.2016.2569608).
- [137] H. TOLBA, N. DKHILI, J. NOU, J. EYNARD, S. THIL et S. GRIEU. “Multi-Horizon Forecasting of Global Horizontal Irradiance Using Online Gaussian Process Regression : A Kernel Study”. Dans : *Energies* 13.16 (2020). ISSN : 1996-1073. DOI : [10.3390/en13164184](https://doi.org/10.3390/en13164184).
- [138] F. LUBBE, J. MARITZ et T. HARMS. “Evaluating the Potential of Gaussian Process Regression for Solar Radiation Forecasting : A Case Study”. Dans : *Energies* 13.20 (2020). ISSN : 1996-1073. DOI : [10.3390/en13205509](https://doi.org/10.3390/en13205509).
- [139] A. ROHANI, M. TAKI et M. ABDOLLAHPOUR. “A novel soft computing model (Gaussian process regression with K-fold cross validation) for daily and monthly solar radiation forecasting (Part : I)”. Dans : *Renewable Energy* 115 (2018), p. 411-422. ISSN : 0960-1481. DOI : <https://doi.org/10.1016/j.renene.2017.08.061>.
- [140] M. SHARIFZADEH, A. SIKINIOTI-LOCK et N. SHAH. “Machine-learning methods for integrated renewable power generation : A comparative study of artificial neural networks, support vector regression, and Gaussian process regression”. Dans : *Renewable and Sustainable Energy Reviews* 108 (2019), 513-538. ISSN : 1364-0321. DOI : <https://doi.org/10.1016/j.rser.2019.03.040>.
- [141] Y. GALA, Á. FERNÁNDEZ, J. DÍAZ et J. R. DORRONSORO. “Hybrid machine learning forecasting of solar radiation values”. Dans : *Neurocomputing* 176 (2016), p. 48-59. ISSN : 0925-2312. DOI : <https://doi.org/10.1016/j.neucom.2015.02.078>.
- [142] N. BIGDELI, M. SALEHI BORUJENI et K. AFSHAR. “Time series analysis and short-term forecasting of solar irradiation, a new hybrid approach”. Dans : *Swarm and Evolutionary Computation* 34 (2017), p. 75-88. ISSN : 2210-6502. DOI : <https://doi.org/10.1016/j.swevo.2016.12.004>.
- [143] F. O. HOCAOGLU et F. SERTTAS. “A novel hybrid (Mycielski-Markov) model for hourly solar radiation forecasting”. Dans : *Renewable Energy* 108 (2017), p. 635-643. ISSN : 0960-1481. DOI : <https://doi.org/10.1016/j.renene.2016.08.058>.
- [144] M. GUERMOUI, F. MELGANI, K. GAIRAA et M. L. MEKHALFI. “A comprehensive review of hybrid models for solar radiation forecasting”. Dans : *Journal of Cleaner Production* 258 (2020), p. 120357. ISSN : 0959-6526. DOI : <https://doi.org/10.1016/j.jclepro.2020.120357>.

- [145] Y. KAMARIANAKIS et P. PRASTACOS. “Forecasting Traffic Flow Conditions in an Urban Network : Comparison of Multivariate and Univariate Approaches”. Dans : *Transportation Research Record* 1857.1 (2003), p. 74-84. DOI : [10.3141/1857-09](https://doi.org/10.3141/1857-09). eprint : <https://doi.org/10.3141/1857-09>.
- [146] G. MATHERON. “Principles of geostatistics”. Dans : *Economic Geology* 58.8 (déc. 1963), p. 1246-1266. ISSN : 0361-0128. DOI : [10.2113/gsecongeo.58.8.1246](https://pubs.geoscienceworld.org/economicgeology/article-pdf/58/8/1246/3481854/1246.pdf). eprint : <https://pubs.geoscienceworld.org/economicgeology/article-pdf/58/8/1246/3481854/1246.pdf>.
- [147] C. E. RASMUSSEN et C. K. I. WILLIAMS. *Gaussian processes for machine learning*. MIT Press, 2006.
- [148] J. A. K. SUYKENS, T. VAN GESTEL, J. DE BRABANTER, B. DE MOOR et J. VANDEWALLE. *Least Squares Support Vector Machines*. WORLD SCIENTIFIC, 2002. DOI : [10.1142/5089](https://www.worldscientific.com/doi/pdf/10.1142/5089). eprint : <https://www.worldscientific.com/doi/pdf/10.1142/5089>.
- [149] T. E. HOFF, R. PEREZ, J. KLEISSL, D. RENNE et J. STEIN. “Reporting of irradiance modeling relative prediction errors”. Dans : *Progress in Photovoltaics : Research and Applications* 21.7 (2013), p. 1514-1519. DOI : <https://doi.org/10.1002/pip.2225>. eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1002/pip.2225>.
- [150] L. FRÍAS-PAREDES, F. MALLOR, M. GASTÓN-ROMEO et T. LEÓN. “Dynamic mean absolute error as new measure for assessing forecasting errors”. Dans : *Energy Conversion and Management* 162 (2018), p. 176-188. ISSN : 0196-8904. DOI : <https://doi.org/10.1016/j.enconman.2018.02.030>.
- [151] L. FRÍAS-PAREDES, F. MALLOR, T. LEÓN et M. GASTÓN-ROMEO. “Introducing the Temporal Distortion Index to perform a bidimensional analysis of renewable energy forecast”. Dans : *Energy* 94 (2016), p. 180-194. ISSN : 0360-5442. DOI : <https://doi.org/10.1016/j.energy.2015.10.093>.
- [152] L. FRÍAS-PAREDES, F. MALLOR, M. GASTÓN-ROMEO et T. LEÓN. “Assessing energy forecasting inaccuracy by simultaneously considering temporal and absolute errors”. Dans : *Energy Conversion and Management* 142 (2017), p. 533-546. ISSN : 0196-8904. DOI : <https://doi.org/10.1016/j.enconman.2017.03.056>.
- [153] L. VALLANCE, B. CHARBONNIER, N. PAUL, S. DUBOST et P. BLANC. “Towards a standardized procedure to assess solar forecast accuracy : A new ramp and time alignment metric”. Dans : *Solar Energy* 150 (2017), p. 408-422. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2017.04.064>.
- [154] D. YANG. “Making reference solar forecasts with climatology, persistence, and their optimal convex combination”. Dans : *Solar Energy* 193 (2019), p. 981-985. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2019.10.006>.
- [155] P. LAURET, M. DAVID et P. PINSON. “Verification of solar irradiance probabilistic forecasts”. Dans : *Solar Energy* 194 (2019), p. 254-271. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2019.10.041>.
- [156] G. BONTEMPI, S. BEN TAIEB et Y.-A. LE BORGNE. “Machine Learning Strategies for Time Series Forecasting”. Dans : *Business Intelligence : Second European Summer School, eBISS 2012, Brussels, Belgium, July 15-21, 2012, Tutorial Lectures*. Sous la dir. de M.-A. AUFAURE et E. ZIMÁNYI. Berlin, Heidelberg : Springer Berlin Heidelberg, 2013, p. 62-77. ISBN : 978-3-642-36318-4. DOI : [10.1007/978-3-642-36318-4\\_3](https://doi.org/10.1007/978-3-642-36318-4_3).
- [157] S. B. TAIEB, A. SORJAMAA et G. BONTEMPI. “Multiple-output modeling for multi-step-ahead time series forecasting”. Dans : *Neurocomputing* 73.10 (2010). Subspace Learning / Selected papers from the European Symposium on Time Series Prediction, p. 1950-1957. ISSN : 0925-2312. DOI : <https://doi.org/10.1016/j.neucom.2009.11.030>.



- [158] S. ROBERTS, M. OSBORNE, M. EBDEN, S. REECE, N. GIBSON et S. AIGRAIN. “Gaussian processes for time-series modelling”. Dans : *Philosophical Transactions of the Royal Society A : Mathematical, Physical and Engineering Sciences* 371.1984 (2013), p. 20110550. DOI : [10.1098/rsta.2011.0550](https://doi.org/10.1098/rsta.2011.0550). eprint : <https://royalsocietypublishing.org/doi/pdf/10.1098/rsta.2011.0550>.
- [159] C. MICCHELLI, Y. XU et H. ZHANG. “Universal Kernels”. Dans : *Mathematics* 7 (2006), p. 2651-2667.
- [160] D. K. DUVENAUD, H. NICKISCH et C. E. RASMUSSEN. “Additive Gaussian Processes”. Dans : *Advances in Neural Information Processing Systems 24*. Sous la dir. de J. SHAWE-TAYLOR, R. S. ZEMEL, P. L. BARTLETT, F. PEREIRA et K. Q. WEINBERGER. Curran Associates, Inc., 2011, p. 226-234.
- [161] Z. CHEN et B. WANG. “How priors of initial hyperparameters affect Gaussian process regression models”. Dans : *Neurocomputing* 275 (2018), p. 1702-1710. ISSN : 0925-2312. DOI : <https://doi.org/10.1016/j.neucom.2017.10.028>.
- [162] V. N. VAPNIK. *The Nature of Statistical Learning Theory*. Berlin, Heidelberg : Springer-Verlag, 1995. ISBN : 0-387-94559-8.
- [163] V. N. VAPNIK, S. E. GOLOWICH et A. J. SMOLA. “Support Vector Method for Function Approximation, Regression Estimation and Signal Processing”. Dans : *Advances in Neural Information Processing Systems 9*. Sous la dir. de M. C. MOZER, M. I. JORDAN et T. PETSCH. MIT Press, 1997, p. 281-287.
- [164] H. DRUCKER, C. J. C. BURGESS, L. KAUFMAN, A. J. SMOLA et V. N. VAPNIK. “Support Vector Regression Machines”. Dans : *Advances in Neural Information Processing Systems 9*. Sous la dir. de M. C. MOZER, M. I. JORDAN et T. PETSCH. MIT Press, 1997, p. 155-161.
- [165] C. J. C. BURGESS. “A Tutorial on Support Vector Machines for Pattern Recognition”. Dans : *Data Mining and Knowledge Discovery 2.2* (1998), p. 121-167. ISSN : 1573-756X. DOI : [10.1023/A:1009715923555](https://doi.org/10.1023/A:1009715923555).
- [166] A. J. SMOLA et B. SCHÖLKOPF. “A tutorial on support vector regression”. Dans : *Statistics and Computing* 14.3 (2004), p. 199-222. ISSN : 1573-1375. DOI : [10.1023/B:STCO.0000035301.49549.88](https://doi.org/10.1023/B:STCO.0000035301.49549.88).
- [167] V. N. VAPNIK et A. LERNER. “Pattern Recognition using Generalized Portrait Method”. Dans : *Automation and Remote Control* 24 (1963).
- [168] N. CRISTIANINI et J. SHAWE-TAYLOR. *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press, 2000. DOI : [10.1017/CB09780511801389](https://doi.org/10.1017/CB09780511801389).
- [169] J. SUYKEN, G. HORVÁTH, S. BASU, C. MICCHELLI et J. VANDEWALLE. *Advances in learning theory : methods, models and applications*. T. 190. IOS Press Amsterdam, 2003.
- [170] K. DE BRABANTER, J. DE BRABANTER, J. A. K. SUYKENS et B. DE MOOR. “Approximate Confidence and Prediction Intervals for Least Squares Support Vector Regression”. Dans : *IEEE Transactions on Neural Networks* 22.1 (2011), p. 110-120. ISSN : 1941-0093. DOI : [10.1109/TNN.2010.2087769](https://doi.org/10.1109/TNN.2010.2087769).
- [171] G. P. ZHANG. “Neural Networks for Time-Series Forecasting”. Dans : *Handbook of Natural Computing*. Sous la dir. de G. ROZENBERG, T. BÄCK et J. N. KOK. Berlin, Heidelberg : Springer Berlin Heidelberg, 2012, p. 461-477. ISBN : 978-3-540-92910-9. DOI : [10.1007/978-3-540-92910-9\\_14](https://doi.org/10.1007/978-3-540-92910-9_14).
- [172] K. S. REDDY et M. RANJAN. “Solar resource estimation using artificial neural networks and comparison with other correlation models”. Dans : *Energy Conversion and Management* 44.15 (2003), p. 2519-2530. ISSN : 0196-8904.

- [173] J. BROWNLEE. *Deep Learning for Time Series Forecasting : Predict the Future with MLPs, CNNs and LSTMs in Python*. Machine Learning Mastery, 2018.
- [174] S. GHIMIRE, R. C. DEO, N. RAJ et J. MI. “Deep solar radiation forecasting with convolutional neural network and long short-term memory network algorithms”. Dans : *Applied Energy* 253 (2019), p. 113541. ISSN : 0306-2619. DOI : <https://doi.org/10.1016/j.apenergy.2019.113541>.
- [175] S. SIAMI-NAMINI, N. TAVAKOLI et A. SIAMI NAMIN. “A Comparison of ARIMA and LSTM in Forecasting Time Series”. Dans : *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. 2018, p. 1394-1401. DOI : [10.1109/ICMLA.2018.00227](https://doi.org/10.1109/ICMLA.2018.00227).
- [176] S. HOCHREITER et J. SCHMIDHUBER. “Long Short-Term Memory”. Dans : *Neural Computation* 9.8 (1997), p. 1735-1780. ISSN : 0899-7667. DOI : [10.1162/neco.1997.9.8.1735](https://doi.org/10.1162/neco.1997.9.8.1735).
- [177] R. PASCANU, T. MIKOLOV et Y. BENGIO. “On the difficulty of training recurrent neural networks”. Dans : *Proceedings of the 30th International Conference on Machine Learning*. Sous la dir. de S. DASGUPTA et D. MCALLESTER. T. 28. Proceedings of Machine Learning Research 3. Atlanta, GA, USA : PMLR, 2013, p. 1310-1318.
- [178] S. HOCHREITER. “The Vanishing Gradient Problem during Learning Recurrent Neural Nets and Problem Solutions”. Dans : *International Journal of Uncertainty Fuzziness and Knowledge-Based Systems* 6.2 (1998), p. 107-116. ISSN : 0218-4885. DOI : [10.1142/S0218488598000094](https://doi.org/10.1142/S0218488598000094).
- [179] F. A. GERS, J. SCHMIDHUBER et F. CUMMINS. “Learning to Forget : Continual Prediction with LSTM”. Dans : *Neural Computation* 12.10 (2000), p. 2451-2471. DOI : [10.1162/089976600300015015](https://doi.org/10.1162/089976600300015015).
- [180] A. KHOSRAVI, S. NAHAVANDI, D. C. CREIGHTON et A. F. ATIYA. “Comprehensive Review of Neural Network-Based Prediction Intervals and New Advances”. Dans : *IEEE Transactions on Neural Networks* 22 (2011), p. 1341-1356.
- [181] T. PEARCE, A. BRINTRUP, M. ZAKI et A. NEELY. “High-Quality Prediction Intervals for Deep Learning : A Distribution-Free, Ensembled Approach”. Dans : *Proceedings of the 35th International Conference on Machine Learning*. Sous la dir. de J. DY et A. KRAUSE. T. 80. Proceedings of Machine Learning Research. Stockholmsmässan, Stockholm Sweden : PMLR, 2018, p. 4075-4084.
- [182] D. P. KINGMA et J. BA. “Adam : A Method for Stochastic Optimization”. Dans : *3rd International Conference on Learning Representations, ICLR 2015*. San Diego, CA, USA, mai 2015.
- [183] G. E. HINTON, N. SRIVASTAVA, A. KRIZHEVSKY, I. SUTSKEVER et R. SALAKHUTDINOV. “Improving neural networks by preventing co-adaptation of feature detectors”. Dans : *Computing Research Repository (CoRR)* abs/1207.0580 (jan. 2012).
- [184] J. QUIÑONERO-CANDELA et C. E. RASMUSSEN. “A Unifying View of Sparse Approximate Gaussian Process Regression”. Dans : *Journal of Machine Learning Research* 6 (2005), p. 1939-1959. ISSN : 1532-4435.
- [185] S. GBÉMOU, H. TOLBA, S. THIL et S. GRIEU. “Global horizontal irradiance forecasting using online sparse Gaussian process regression based on quasiperiodic kernels”. Dans : *2019 IEEE International Conference on Environment and Electrical Engineering and 2019 IEEE Industrial and Commercial Power Systems Europe (EEEIC/ICPS Europe)*. 2019, p. 1-6.
- [186] L. XU, N. CHEN, Z. CHEN, C. ZHANG et H. YU. “Spatiotemporal forecasting in earth system science : Methods, uncertainties, predictability and future directions”. Dans : *Earth-Science Reviews* 222 (2021), p. 103828. ISSN : 0012-8252. DOI : <https://doi.org/10.1016/j.earscirev.2021.103828>.

- [187] J. R. WILLIAMS, C. A. JONES, J. R. KINIRY et D. A. SPANEL. “The EPIC Crop Growth Model”. Dans : *Transactions of the ASAE* 32.2 (1989), p. 497-0511. ISSN : 0001-2351. DOI : [10.13031/2013.31032](https://doi.org/10.13031/2013.31032).
- [188] P. BAUER, A. THORPE et G. BRUNET. “The quiet revolution of numerical weather prediction”. Dans : *Nature* 525.7567 (2015), p. 47-55. ISSN : 1476-4687. DOI : [10.1038/nature14956](https://doi.org/10.1038/nature14956).
- [189] C. SWEENEY, R. J. BESSA, J. BROWELL et P. PINSON. “The future of forecasting for renewable energy”. Dans : *WIREs Energy and Environment* 9.2 (2020), e365. DOI : <https://doi.org/10.1002/wene.365>. eprint : <https://wires.onlinelibrary.wiley.com/doi/pdf/10.1002/wene.365>.
- [190] L. XU, P. ABBASZADEH, H. MORADKHANI, N. CHEN et X. ZHANG. “Continental drought monitoring using satellite soil moisture, data assimilation and an integrated drought index”. Dans : *Remote Sensing of Environment* 250 (2020), p. 112028. ISSN : 0034-4257. DOI : <https://doi.org/10.1016/j.rse.2020.112028>.
- [191] B. FINKENSTADT, L. HELD et V. ISHAM. *Statistical Methods for Spatio-Temporal Systems*. Chapman et Hall/CRC, 2006.
- [192] N. CRESSIE et C. WIKLE. *Statistics for spatio-temporal data*. Wiley, 2011.
- [193] X. LUNA et M. GENTON. “Predictive Spatio-Temporal Models for Spatially Sparse Environmental Data”. Dans : *Statistica Sinica* 15 (avr. 2005), p. 547-568.
- [194] D. YANG, Z. DONG, T. REINDL, P. JURUTITJAROEN et W. M. WALSH. “Solar irradiance forecasting using spatio-temporal empirical kriging and vector autoregressive models with parameter shrinkage”. Dans : *Solar Energy* 103 (2014), p. 550-562. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2014.01.024>.
- [195] H. LUTKEPOHL. *New Introduction to Multiple Time Series Analysis*. Berlin, Heidelberg : Springer, 2005.
- [196] P. E. PFEIFER et S. E. BODILY. “A test of space-time arma modelling and forecasting of hotel data”. Dans : *Journal of Forecasting* 9 (1990), p. 255-272.
- [197] P. E. PFEIFER et S. J. DEUTSCH. “Identification and Interpretation of First Order Space-Time ARMA Models”. Dans : *Technometrics* 22.3 (1980), p. 397-408. ISSN : 0040-1706. DOI : [10.2307/1268325](https://doi.org/10.2307/1268325).
- [198] P. E. PFEIFER et S. J. DEUTSCH. “A STARIMA Model-Building Procedure with Application to Description and Regional Forecasting”. Dans : *Transactions of the Institute of British Geographers* 5.3 (1980), p. 330-349. DOI : [10.2307/621846](https://doi.org/10.2307/621846).
- [199] P. E. PFEIFER et S. J. DEUTSCH. “A Three-Stage Iterative Procedure for Space-Time Modeling”. Dans : *Technometrics* 22.1 (1980), p. 35-47. ISSN : 0040-1706. DOI : [10.2307/1268381](https://doi.org/10.2307/1268381).
- [200] S. J. DEUTSCH et P. E. PFEIFER. “Space-Time ARMA Modeling with Contemporaneously Correlated Innovations”. Dans : *Technometrics* 23.4 (1981), p. 401-409. ISSN : 0040-1706. DOI : [10.2307/1268231](https://doi.org/10.2307/1268231).
- [201] C. A. GLASBEY et D. J. ALLCROFT. “A spatiotemporal auto-regressive moving average model for solar radiation”. Dans : *Journal of the Royal Statistical Society : Series C (Applied Statistics)* 57.3 (2008), p. 343-355. DOI : <https://doi.org/10.1111/j.1467-9876.2007.00617.x>. eprint : <https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9876.2007.00617.x>.
- [202] R. J. BENNETT. *Spatial time series : analysis-forecasting-control*. English. London : Pion, 1979.



- [203] R. DAMBREVILLE, P. BLANC, J. CHANUSSOT et D. BOLDO. "Very short term forecasting of the Global Horizontal Irradiance using a spatio-temporal autoregressive model". Dans : *Renewable Energy* 72 (2014), p. 291-300. ISSN : 0960-1481. DOI : <https://doi.org/10.1016/j.renene.2014.07.012>.
- [204] M. JAMALY et J. KLEISSL. "Spatiotemporal interpolation and forecast of irradiance data using Kriging". Dans : *Solar Energy* 158 (2017), p. 407-423. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2017.09.057>.
- [205] D. G. KRIGE. "A statistical approach to some basic mine valuation problems on the Witwatersrand". Dans : *Journal of the Southern African Institute of Mining and Metallurgy* 52.6 (1951), p. 119-139. DOI : [10.10520/AJA0038223X\\_4792](https://doi.org/10.10520/AJA0038223X_4792).
- [206] N. CRESSIE. "The origins of kriging". Dans : *Mathematical Geology* 22.3 (1990), p. 239-252. ISSN : 1573-8868. DOI : [10.1007/BF00889887](https://doi.org/10.1007/BF00889887).
- [207] S. REHMAN et S. G. GHORI. "Spatial estimation of global solar radiation using geostatistics". Dans : *Renewable Energy* 21.3 (2000), p. 583-605. ISSN : 0960-1481. DOI : [https://doi.org/10.1016/S0960-1481\(00\)00078-1](https://doi.org/10.1016/S0960-1481(00)00078-1).
- [208] R. RIGHINI, H. G. GALLEGOS et C. RAICHIJK. "Approach to drawing new global solar irradiation contour maps for Argentina". Dans : *Renewable Energy* 30.8 (2005), p. 1241-1255. ISSN : 0960-1481. DOI : <https://doi.org/10.1016/j.renene.2004.10.010>.
- [209] J. A. RUIZ-ARIAS, D. POZO-VÁZQUEZ, F. J. SANTOS-ALAMILLOS, V. LARA-FANEGO et J. TOVAR-PESCADOR. "A topographic geostatistical approach for mapping monthly mean values of daily global solar radiation : A case study in southern Spain". Dans : *Agricultural and Forest Meteorology* 151.12 (2011), p. 1812-1822. ISSN : 0168-1923. DOI : <https://doi.org/10.1016/j.agrformet.2011.07.021>.
- [210] L. BECHINI, G. DUCCO, M. DONATELLI et A. STEIN. "Modelling, interpolation and stochastic simulation in space and time of global solar radiation". Dans : *Agriculture, Ecosystems & Environment* 81.1 (2000), p. 29-42. ISSN : 0167-8809. DOI : [https://doi.org/10.1016/S0167-8809\(00\)00170-5](https://doi.org/10.1016/S0167-8809(00)00170-5).
- [211] G. CHAOJUN, D. YANG, P. JIRUTITIJARON, W. M. WALSH et T. REINDL. "Spatial Load Forecasting With Communication Failure Using Time-Forward Kriging". Dans : *IEEE Transactions on Power Systems* 29.6 (2014), p. 2875-2882. DOI : [10.1109/TPWRS.2014.2308537](https://doi.org/10.1109/TPWRS.2014.2308537).
- [212] N. CRESSIE et J. J. MAJURE. "Spatio-Temporal Statistical Modeling of Livestock Waste in Streams". Dans : *Journal of Agricultural, Biological, and Environmental Statistics* 2.1 (1997), p. 24-47. ISSN : 1085-7117. DOI : [10.2307/1400639](https://doi.org/10.2307/1400639).
- [213] M. I. HARTFIELD et R. F. GUNST. "Identification of Model Components for a Class of Continuous Spatiotemporal Models". Dans : *Journal of Agricultural, Biological, and Environmental Statistics* 8.1 (2003), p. 105-121. ISSN : 1085-7117.
- [214] S. LOPHAVEN, J. CARSTENSEN et H. ROOTZÉN. "Space-time modeling of environmental monitoring data". Dans : *Environmental and Ecological Statistics* 11.3 (2004), p. 237-256. ISSN : 1573-3009. DOI : [10.1023/B:EEST.0000038014.91328.58](https://doi.org/10.1023/B:EEST.0000038014.91328.58).
- [215] R. A. BILONICK. "The space-time distribution of sulfate deposition in the northeastern United States". Dans : *Atmospheric Environment* (1967) 19.11 (1985), p. 1829-1845. ISSN : 0004-6981. DOI : [https://doi.org/10.1016/0004-6981\(85\)90009-5](https://doi.org/10.1016/0004-6981(85)90009-5).
- [216] S. ROUHANI et T. J. HALL. "Space-Time Kriging of Groundwater Data". Dans : *Geostatistics*. Sous la dir. de M. ARMSTRONG. Dordrecht : Springer Netherlands, 1989, p. 639-650. ISBN : 978-94-015-6844-9.
- [217] R. PEREZ, M. DAVID, T. E. HOFF, M. JAMALY, S. KIVALOV, J. KLEISSL, P. LAURET et M. PEREZ. "Spatial and Temporal Variability of Solar Energy". Dans : *Foundations and Trends in Renewable Energy* 1.1 (2016), p. 1-44. ISSN : 2328-8892. DOI : [10.1561/2700000006](https://doi.org/10.1561/2700000006).

- [218] D. L. ZIMMERMAN. “Computationally efficient restricted maximum likelihood estimation of generalized covariance functions”. Dans : *Mathematical Geology* 21.7 (1989), p. 655-672. ISSN : 1573-8868. DOI : [10.1007/BF00893314](https://doi.org/10.1007/BF00893314).
- [219] M. SHERMAN. *Spatial Statistics and Spatio-Temporal Data : Covariance Functions and Directional Properties*. Wiley Series on Probability et Statistics, 2011. ISBN : 978-0-470-97492-6.
- [220] T. GNEITING, M. GENTON et P. GUTTORP. “Geostatistical Space-Time Models, Stationarity, Separability and Full Symmetry”. Dans : *Statistical Methods for Spatio-temporal Systems* (2007), p. 151-175. DOI : [10.1201/9781420011050.ch4](https://doi.org/10.1201/9781420011050.ch4).
- [221] T. GNEITING. “Nonseparable, Stationary Covariance Functions for Space-Time Data”. Dans : *Journal of the American Statistical Association* 97.458 (2002), p. 590-600. ISSN : 0162-1459.
- [222] E. PORCU, P. GREGORI et J. MATEU. “Nonseparable stationary anisotropic space-time covariance functions”. Dans : *Stochastic Environmental Research and Risk Assessment* 21.2 (2006), p. 113-122. ISSN : 1436-3259. DOI : [10.1007/s00477-006-0048-3](https://doi.org/10.1007/s00477-006-0048-3).
- [223] E. PORCU, J. MATEU et F. SAURA. “New classes of covariance and spectral density functions for spatio-temporal modelling”. Dans : *Stochastic Environmental Research and Risk Assessment* 22.1 (2008), p. 65-79. ISSN : 1436-3259. DOI : [10.1007/s00477-007-0160-z](https://doi.org/10.1007/s00477-007-0160-z).
- [224] N. CRESSIE et H.-C. HUANG. “Classes of Nonseparable, Spatio-Temporal Stationary Covariance Functions”. Dans : *Journal of the American Statistical Association* 94.448 (1999), p. 1330-1340. ISSN : 0162-1459. DOI : [10.2307/2669946](https://doi.org/10.2307/2669946).
- [225] M. W. MITCHELL, M. G. GENTON et M. L. GUMPERTZ. “Testing for separability of space-time covariances”. Dans : *Environmetrics* 16.8 (2005), p. 819-831. DOI : <https://doi.org/10.1002/env.737>. eprint : <https://onlinelibrary.wiley.com/doi/pdf/10.1002/env.737>.
- [226] D. YANG, C. GU, Z. DONG, P. JIRUTITIJAROEN, N. CHEN et W. M. WALSH. “Solar irradiance forecasting using spatial-temporal covariance structures and time-forward kriging”. Dans : *Renewable Energy* 60 (2013), p. 235-245. ISSN : 0960-1481. DOI : <https://doi.org/10.1016/j.renene.2013.05.030>.
- [227] A. W. ARYAPUTERA, D. YANG, L. ZHAO et W. M. WALSH. “Very short-term irradiance forecasting at unobserved locations using spatio-temporal kriging”. Dans : *Solar Energy* 122 (2015), p. 1266-1278. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2015.10.023>.
- [228] V. P. A. LONIJ, A. E. BROOKS, A. D. CRONIN, M. LEUTHOLD et K. KOCH. “Intra-hour forecasts of solar power production using measurements from a network of irradiance sensors”. Dans : *Solar Energy* 97 (2013), p. 58-66. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2013.08.002>.
- [229] T. INOUE, T. SASAKI et T. WASHIO. “Spatio-temporal Kriging of solar radiation incorporating direction and speed of cloud movement”. Dans : *26th Annual Conference of the Japanese Society for Artificial Intelligence*. 2012.
- [230] K. SHINOZAKI, N. YAMAKAWA, T. SASAKI et T. INOUE. “Areal Solar Irradiance Estimated by Sparsely Distributed Observations of Solar Radiation”. Dans : *IEEE Transactions on Power Systems* 31.1 (2016), p. 35-42. DOI : [10.1109/TPWRS.2015.2393636](https://doi.org/10.1109/TPWRS.2015.2393636).
- [231] M. SCHLATHER. “Some covariance models based on normal scale mixtures”. Dans : *Bernoulli* 16.3 (2010), p. 780-797. DOI : [10.3150/09-BEJ226](https://doi.org/10.3150/09-BEJ226).
- [232] P. D. SAMPSON et P. GUTTORP. “Nonparametric Estimation of Nonstationary Spatial Covariance Structure”. Dans : *Journal of the American Statistical Association* 87.417 (1992), p. 108-119. ISSN : 0162-1459. DOI : [10.2307/2290458](https://doi.org/10.2307/2290458).

- [233] R. PEREZ, S. KIVALOV, J. SCHLEMMER, K. HEMKER et T. E. HOFF. “Short-term irradiance variability : Preliminary estimation of station pair correlation as a function of distance”. Dans : *Solar Energy* 86.8 (2012). Progress in Solar Energy 3, p. 2170-2176. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2012.02.027>.
- [234] M. LAVE et J. KLEISSL. “Cloud speed impact on solar variability scaling – Application to the wavelet variability model”. Dans : *Solar Energy* 91 (2013), p. 11-21. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2013.01.023>.
- [235] M. J. R. PEREZ et V. M. FTHENAKIS. “On the spatial decorrelation of stochastic solar resource variability at long timescales”. Dans : *Solar Energy* 117 (2015), p. 46-58. ISSN : 0038-092X. DOI : <https://doi.org/10.1016/j.solener.2015.04.020>.
- [236] A. HABTE, M. SENGUPTA, C. GUEYMARD, A. GOLNAS et Y. XIE. “Long-term spatial and temporal solar resource variability over America using the NSRDB version 3 (1998–2017)”. Dans : *Renewable and Sustainable Energy Reviews* 134 (2020), p. 110285. ISSN : 1364-0321. DOI : <https://doi.org/10.1016/j.rser.2020.110285>.
- [237] G. PLEISS. “A scalable and flexible framework for Gaussian processes via matrix-vector multiplication”. Thèse de doct. Cornell University, 2020.
- [238] W.-H. STEEB et Y. HARDY. *Matrix Calculus and Kronecker Product*. WORLD SCIENTIFIC, 2011. DOI : [10.1142/8030](https://doi.org/10.1142/8030). eprint : <https://www.worldscientific.com/doi/pdf/10.1142/8030>.

## Résumé

La pénétration des énergies renouvelables intermittentes, en particulier du solaire photovoltaïque, dans le réseau électrique de distribution ne cesse d'augmenter du fait de la dynamique insufflée par la transition énergétique. Une mutation du réseau électrique de distribution est donc devenue nécessaire. Cette mutation doit permettre d'augmenter la capacité d'accueil du réseau sans renforcement des infrastructures et de limiter les contraintes, en particulier en tension, qui commencent à apparaître sur le réseau électrique de distribution basse tension.

Cette thèse s'inscrit dans le cadre du projet de démonstrateur de réseau électrique intelligent en milieu périurbain/rural « Smart Occitania », traitant du développement d'une stratégie de contrôle à coût calculatoire maîtrisé pour la gestion prédictive d'un réseau électrique de distribution basse tension. La stratégie de contrôle implique des prévisions infra-horaires et infra-journalières de grandeurs stochastiques, telles que la production solaire photovoltaïque. Ces travaux de thèse traitent de la prévision multi-horizon, dans un premier temps localisée, puis spatio-temporelle, de l'éclairement global horizontal (ou GHI, pour *global horizontal irradiance*) à des fins de prévision de la production solaire photovoltaïque.

Tout d'abord, une étude comparative des outils de l'apprentissage automatique a été menée dans le but de choisir l'outil le plus adapté à la prévision du GHI. Les résultats ont montré que la régression non paramétrique par processus gaussien (ou GPR, pour *Gaussian process regression*) est l'outil le plus adapté à notre application. Ensuite, le choix de l'entrée des modèles GPR a été étudié. En effet, pour le GPR, il est possible d'utiliser le temps ou un historique d'observations comme entrée. Une étude comparative a été menée, mettant en lumière qu'au regard de notre application, le temps était l'entrée adéquate. Par ailleurs, les biais inductifs connus sur la dynamique du GHI ont été utilisés afin de construire un noyau adapté. Enfin, une étude préliminaire a été menée, visant à déterminer le noyau spatial à associer au noyau temporel pour la prévision dans le temps et l'espace du GHI.

Mots clés : éclairage global horizontal, régression non paramétrique par processus gaussien, noyau, prévision, apprentissage automatique, production photovoltaïque, transition énergétique.

## Abstract

The penetration of intermittent renewable energies, in particular solar photovoltaic (PV), in the power distribution grid is constantly increasing due to the dynamics instilled by the energy transition. A change of the power distribution grid has thus become necessary. This mutation must allow to increase the capacity of the network without reinforcing the infrastructures and to limit the constraints, in particular voltage constraints, which start to appear on the low voltage power distribution grid.

This thesis is part of the project Smart Occitania, which deals with the development of a control strategy with controlled computing cost for the predictive management of a low voltage power distribution grid. The control strategy involves intrahour and intraday forecasts of stochastic quantities, such as solar photovoltaic generation. This thesis deals with the temporal and spatio-temporal forecasting of global horizontal irradiance (GHI), in order to forecast PV power generation.

First, a comparative study of machine learning tools was conducted in order to choose the most suitable method for GHI forecasting. The results showed that non-parametric Gaussian process regression (GPR) is the most suitable tool for our application. Then, the choice of the input of the GPR models has been studied. Indeed, for GPR, time or historical observations can be used as input. A comparative study was conducted, highlighting that, with respect to our application, time as an adequate input. Furthermore, the known inductive biases on the GHI dynamics have been used to build a suitable kernel. Finally, a preliminary study has been conducted, so as to determine the spatial kernel to be associated with the temporal kernel for spatio-temporal forecasting of GHI.

Keywords : global horizontal irradiance, Gaussian process regression, kernel, forecasting, machine learning, photovoltaic power generation, energy transition.