



HAL
open science

Prévision du rayonnement solaire global par télédétection pour la gestion de la production d'énergie photovoltaïque

Romain Dambreville

► **To cite this version:**

Romain Dambreville. Prévision du rayonnement solaire global par télédétection pour la gestion de la production d'énergie photovoltaïque. Traitement du signal et de l'image [eess.SP]. Université de Grenoble, 2014. Français. NNT: 2014GRENT078 . tel-01130251

HAL Id: tel-01130251

<https://theses.hal.science/tel-01130251>

Submitted on 11 Mar 2015

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.

UNIVERSITÉ DE GRENOBLE
ÉCOLE DOCTORALE EEATS
Électronique Électrotechnique Traitement du Signal

T H È S E

pour obtenir le titre de

docteur en sciences

de l'Université de Grenoble

Mention : SIGNAL, IMAGE, PAROLE, TÉLÉCOMS

Présentée et soutenue par

Romain DAMBREVILLE

**Prévision du rayonnement solaire global par télédétection pour
la gestion de la production d'énergie photovoltaïque.**

Thèse dirigée par Jocelyn CHANUSSOT, Co-dirigée par Giorgio A.
LICCIARDI et Philippe BLANC,

préparée au laboratoire Grenoble Image Parole Signal Automatique
(Gipsa-lab)

soutenue le 16 Octobre 2014

Jury :

<i>Rapporteurs :</i>	Marie CHABERT	-	IRIT, INP-ENSEEIH
	Philippe LAURET	-	Laboratoire PIMENT, Uni. de la Réunion
<i>Directeur :</i>	Jocelyn CHANUSSOT	-	Gipsa-Lab, Grenoble-INP
<i>Co-directeurs :</i>	Giorgio Antonino LICCIARDI	-	Gipsa-Lab, Grenoble-INP
	Philippe BLANC	-	Mines-ParisTech
<i>Encadrant</i>	Stephanie DUBOST	-	EDF R&D
<i>Président :</i>	Nicolas PAPARODITIS	-	Laboratoire MATIS, IGN
<i>Examinateur :</i>	Dominique DUMORTIER	-	ENTPE

Remerciements

J'aimerais tout d'abord remercier mon directeur de thèse, M. Jocelyn Chanussot, pour ses conseils avisés et pour m'avoir fait confiance depuis le début. Il m'a permis de découvrir le monde de la recherche par ses cours et son accompagnement lors de mes stages ; puis de m'y épanouir au cours de mon doctorat. Je remercie également l'ensemble des personnes qui m'ont accompagné dans mes travaux au cours de ces trois dernières années, M. Giorgio Licciardi, pour sa disponibilité et sa patience en particulier lors de nos premiers échanges scientifiques. Mme Stephanie Dubost qui a porté le projet depuis son origine et qui a toujours montré une motivation vivifiante lors de nos entretiens. Je tiens à remercier tout particulièrement M. Philippe Blanc, entré dans cette aventure par pure curiosité scientifique, il a su trouver les mots pour me mettre en confiance et me permettre d'avancer lors des moments clés, sa présence en tant que co-directeur de thèse lors de ma soutenance fut un honneur pour moi et un juste reflet de nos collaborations pendant ces trois ans. Il m'est également impossible de ne pas citer Dr. Didier Boldo et M. Alexandre Girard dont les apports scientifiques et techniques ont été inestimables au cours de mes travaux. Je remercie enfin l'ensemble des personnes ayant contribué à l'avancement de ce projet, M. Bruno Charbonnier, M. Laurent Dubus, M. Christophe Chaussin, M. Martial Haeffelin, M. Jordi Badossa, M. Dominique Demengel, M. Guy d'Urso, et beaucoup d'autres encore...

J'hésite à placer mes proches dans cette section, ce que je leur dois s'inscrit bien au delà des remerciements, je leur dois tout. J'ai traversé trois années de thèse qui sont également trois années de vie, avec ce qu'elles comportent de joies, de tristesses, d'échecs, de réussites... Pendant ces trois ans ils ont été mes appuis, ceux sur lesquels j'ai batti tout le reste, ceux qui ont permis que j'arrive jusqu'ici. Donc pour tout, merci à vous : Maman, Papa, Charles, Stéphanie, Mathias, Adèle, Noémie, Jérôme, Thomas, Thomas, Floriane, Sam, Nico, Quentin, Pef, Mugi, Koho, Cresc, Niko, Cindy, Céline, Flo, Manu, Raphael, Ernesto, Clément, Clémentine, Edouard, Jessica, Thomas P., Manon, Julien, Antoine ...

Table des matières

Table des acronymes	xvii
Introduction	1
1 Le rayonnement global : généralités	5
1.1 Fondamentaux sur le rayonnement	6
1.1.1 Définitions	6
1.1.2 Le rayonnement solaire	7
1.1.3 Les nuages	8
1.2 Mesures	10
1.2.1 Site d'étude : le SIRTA	10
1.2.2 Capteurs pyranométriques	12
1.2.3 Contrôle de qualité et bilan des données	15
1.3 Fondamentaux pour la prévision du GHI	19
1.3.1 Modèle ciel clair	19
1.3.2 Index de ciel clair K_c	20
1.3.3 Modèle de référence : la persistance	20
1.3.4 Métriques	21
1.4 Caractérisation du site d'étude	23
1.4.1 Variabilité	23
1.4.2 Résolution temporelle	26
1.5 Conclusion	26
2 Estimation par utilisation d'imageurs hémisphériques du ciel depuis le sol	29
2.1 Etat de l'art des imageurs au sol	29

2.1.1	Capteurs	29
2.1.1.1	Le Total Sky Imager	30
2.1.2	Traitements	31
2.1.2.1	Segmentation et classification de la couverture nuageuse	31
2.1.2.2	Vers la prévision du GHI	33
2.1.3	Conclusion et positionnement des travaux	37
2.2	Estimation du GHI à partir d'images au sol	38
2.2.1	Présentation de la méthode générale	38
2.2.2	Pré-traitement des images	38
2.2.3	Liste des caractéristiques disponibles	39
2.2.3.1	Caractéristiques classiques	41
2.2.3.2	Caractéristiques de texture	41
2.2.3.3	Caractéristique de la couverture nuageuse	44
2.2.3.4	Critère de présence du rayonnement direct	48
2.2.4	Analyse des entrées	49
2.2.5	Réseau de neurones artificiels	51
2.2.6	Construction des bases d'apprentissage/test	51
2.2.7	Résultats	54
2.2.7.1	Méthodologie	54
2.2.7.2	Analyse des résultats	55
2.3	Conclusion	58
3	Imagerie satellitaire pour la prévision du GHI	59
3.1	Introduction	59
3.2	Etat de l'art	60
3.2.1	Acquisition des images	60
3.2.2	Traitement	62

3.2.2.1	Conversion	62
3.2.2.2	Prévision	63
3.2.2.3	Travaux publiés et positionnement des travaux	63
3.2.3	HélioClim-3	64
3.2.3.1	Heliosat : la méthode	64
3.2.3.2	Hélioclim : les différentes base de données	65
3.3	Prévision du GHI par modèle de prévision linéaire spatio-temporel	67
3.3.1	Performances d'HélioClim-3 sur le site du SIRTAs	67
3.3.1.1	Localisation des zones d'intérêts de l'image	68
3.3.2	Les modèles Autoregressifs (AR)	70
3.3.2.1	Choix de l'ordre temporel	72
3.3.3	Définition des modèles testés	73
3.3.4	Résultats	74
3.3.4.1	Analyse quantitative	75
3.3.4.2	Analyse qualitative	76
3.3.5	Conclusion	78
3.4	Prévision par calcul de champ de vecteur	81
3.4.1	Méthodologie	81
3.4.1.1	Le calcul du champs de vecteurs	81
3.4.1.2	Zone d'intérêt	85
3.4.1.3	Interprétation de la zone d'intérêt	86
3.4.2	Résultats	86
3.4.2.1	Analyse quantitative	86
3.4.2.2	Analyse qualitative	88
3.5	Conclusion	92
4	Fusion d'information pour la prévision du GHI	95

4.1	Méthodologie	96
4.2	Résultats expérimentaux	97
4.2.1	Procédé expérimental	97
4.2.2	Analyse quantitative	98
4.2.3	Analyse qualitative	100
4.3	Conclusion	102
	Conclusion	105
A	Tests sur le ratio pour la méthode de segmentation de l'image hémisphérique	109
B	Exemples de profils journaliers de prévision	115
C	Publications de l'auteur	119
C.1	Publication en rapport avec la thèse	119
C.1.1	Publication en revue internationale	119
C.1.1.1	Publication acceptée	119
C.1.1.2	Publication soumise	119
C.1.2	Publication en conférence internationale	119
C.2	Brevet	119
C.3	Autre publication hors du contexte de la thèse	120
	Bibliographie	125

Table des figures

1	Les quatre grandes familles de modèles de prévision (figure inspirée de [Dia+13]).	3
1.1	Représentation d'une onde électromagnétique	6
1.2	Classification du spectre électromagnétique	6
1.3	Solid angle definition [Mal05]	7
1.4	Spectre du rayonnement solaire	8
1.5	Bandes d'absorption de l'atmosphère	9
1.6	Bilan du rayonnement solaire ¹	9
1.7	Classification des types de nuages [Mal05]	10
1.8	Le SIRTa en image ²	11
1.9	Rayonnement direct et diffus : différents chemins	12
1.10	Les angles couramment utilisés pour le rayonnement solaire : θ_z est l'angle zénithal, θ_i est l'angle d'élévation solaire, γ est l'angle azimutal	12
1.11	Station automatique Kipp&Zonen 2AP	13
1.12	Pyrhéliomètre CHP 1	13
1.13	Pyranomètre CMP 22	13
1.14	Exemple de mesures sur une journée	14
1.15	Mesures de GHI sur les années 2011 et 2012 au SIRTa.	16
1.16	Détection des conditions extrêmes sur le rayonnement global en 2011/2012 au SIRTa.	17
1.17	Détection des conditions extrêmes sur le rayonnement diffus en 2011/2012 au SIRTa.	17
1.18	Détection des conditions extrêmes sur le rayonnement direct en 2011/2012 au SIRTa.	18
1.19	Modèle ciel clair ESRA pour l'année 2011 évalué sur le site du SIRTa.	19

1. <http://earthobservatory.nasa.gov/Features/EnergyBalance/page1.php>

2. <http://sirta.ipsl.fr/>

1.20	Présentation de l'index Kc sur une journée de mesure	20
1.21	Modèle de persistance appliqué directement au GHI pour une journée ciel clair. .	21
1.22	Exemple d'erreur d'estimation d'un signal de rayonnement	23
1.23	Répartition des journées de 2011 dans le plan (μ_{Kc}, V)	24
1.24	Exemple de profil pour chaque classe, le nom des classes est explicité dans le tableau 1.8	25
1.25	Impact de la résolution temporelle. (a) échantillonnage : 1 minute, (b) moyenne sur 15 minutes, (c) moyenne sur 30 minutes, (d) moyenne sur 60 minutes	26
2.1	TSI 880	30
2.2	Image brute et image traitée	30
2.3	Image brute	32
2.4	Image seuillée : bleu = ciel clair ; blanc = nuages clairs ; gris = nuages opaques	32
2.5	Vue d'ensemble de la méthode d'estimation	39
2.6	Disponibilité des images du TSI sur l'année 2011, les points rouges représentent les images disponibles.	40
2.7	Image brute	40
2.8	Image masquée	40
2.9	Information apportée par la dissymétrie sur une distribution donnée	41
2.10	Valeurs des pixels	43
2.11	Image correspondante	43
2.12	Combinaisons possibles Référence/voisin	43
2.13	MCOO pour l'image exemple Fig. 2.11	43
2.14	Set d'images test pour la comparaison des méthodes de segmentation	45
2.15	Test de seuil pour la méthode de segmentation par différence : Seuil = 65	46
2.16	Test de seuil pour la méthode de segmentation par ratio : Seuil =0.65	47
2.17	Effet de blanchissement de la bande noire en présence de rayonnement direct . .	48
2.18	Matrice d'inter-corrélation des entrées	50

2.19	Modèle d'un neurone	52
2.20	Structure d'un réseau de neurones en boucle ouverte	52
2.21	Apprentissage à arrêt prématuré	52
2.22	Processus de création des bases de données pour l'apprentissage des réseaux de neurone	53
2.23	Profil journaliers pour chacune des classes	58
3.1	Canal visible de Meteosat 9	61
3.2	Canal Infrarouge de Meteosat 9	61
3.3	Rappel sur les relations entre l'observation depuis le satellite et le rayonnement perçu au sol. Image extraite de [IPC13]. La lettre "a" désigne l'atmosphère, la lettre "g" désigne le sol (ground).	62
3.4	Nuage de points entre K_{CHC3} et K_c , le résultat de la régression polynomiale d'ordre 4 est représenté par la courbe. Données impliquant l'année 2011 uniquement.	68
3.5	Sensibilité de la carte de corrélation en fonction de la variable utilisée; (a) : Corrélation sur le GHI; (b) : Corrélation sur l'index de ciel clair K_c ; (c) : Corrélation sur le différentiel de K_C ;	69
3.6	Cartes de corrélations utilisant les séries temporelles de Δ_{K_c} : (a) Décalage 0; (b) Décalage 1 (15 min); (c) Décalage 2 (30 min); (d) Décalage 3 (45 min); (e) Décalage 4 (60 min); (f) Région équivalente sur google map	70
3.7	(a) Autocorrélation partielle (PACF); (b) RMSE sur le résultat de la prévision	73
3.8	Localisation des entrées sélectionnées en fonction de l'horizon de prévision sur les cartes de corrélation de Δ_{K_c} . L'horizon de prévision h est directement liée au décalage appliqué aux cartes de corrélation : (a) $h=0$, estimation; (b) $h=15$ min, filtrage moyen 2×2 ; (c) $h=30$ min, filtrage moyen 3×3 ; (d) $h=45$ min, filtrage moyen 4×4 ; (e) $h=60$ min, filtrage moyen 5×5	75
3.9	Evolution de la RMSE en fonction de l'horizon de prévision pour chacun des 4 modèles testés.	77
3.10	Profil journaliers pour chacune des classes pour une prévision à 30 minutes	80
3.11	Illustration de l'algorithme de block matching	83

3.12	Exemple 1 : application du block matching sur deux images HC-3 consécutives. (a) Image au temps t-1; (b) Image au temps t; (c) image au temps t avec le champ de vecteur résultant.	84
3.13	Exemple 2 : application du block matching sur deux images HC-3 consécutives. (a) Image au temps t-1; (b) Image au temps t; (c) image au temps t avec le champ de vecteur résultant.	84
3.14	La zone d'intérêt est définie à partir du pixel idéal P_i , en amont du site P_s par rapport au vecteur calculé \mathbf{V}	85
3.15	Exemple d'application en cas réel	85
3.16	Profils journaliers pour chacune des classes pour une prévision à 30 minutes . .	90
3.17	Zoom sur un profil journalier pour une prévision à 30 minutes. Le modèle pré- senté donne non seulement une prévision mais également un "encadrement es- timé" de la prévision.	91
3.18	Fonction d'intercorrélation entre la prévision et la mesure pour trois modèles différents. Les prévisions sont effectuées pour un horizon de prévision de 30 minutes ($h=2$).	91
3.19	Fonction d'intercorrélation entre la prévision et la mesure pour trois modèles différents. Les prévisions sont effectuées pour un horizon de prévision de 45 minutes ($h=3$).	92
3.20	Evolution de la RMSE en fonction de l'horizon de prévision pour différents modèles.	93
4.1	Selection des entrées pour le modèle P_{TSI90} : (a) ZI définie sur l'image HC-3 d'après le vecteur vitesse; (b) Cône de l'image au sol sélectionné à partir du vecteur	96
4.2	Selection des entrées pour le modèle $P_{TSIcône}$: (a) ZI définie sur l'image HC-3 d'après le vecteur vitesse; (b) Cone de l'image au sol sélectionné en fonction du vecteur vitesse (bleu) et de l'angle azimutal solaire (rouge).	97
4.3	Structure du modèle P_{pyr}	98
4.4	Structure des modèles P_{TSI} , P_{TSI90} et $P_{TSIcône}$	98
4.5	Fonction d'intercorrélation entre les modèles et la mesure pour un horizon $h =$ 15 minutes.	101
4.6	Fonction d'intercorrélation entre les modèles et la mesure pour un horizon $h =$ 30 minutes.	101

4.7	Fonction d'intercorrélation entre les modèles et la mesure pour un horizon $h = 45$ minutes.	102
4.8	Profil journalier, $h=15\text{min}$, sur ce profil, le "déphasage" du modèle P_{pyr} lié à l'effet de persistance est très marqué par rapport aux autres modèles, en particulier dans la zone pointé dans l'ellipse noire.	103
4.9	Profil journalier, $h=15\text{min}$, les ellipses noires indiquent les endroits où le modèle $P_{TSI_{cone}}$ permet une modification positive par rapport à l'ensemble des autres modèles.	103
A.1	Test de seuil pour la méthode de segmentation par ratio : Seuil =0.4	110
A.2	Test de seuil pour la méthode de segmentation par ratio : Seuil =0.55	111
A.3	Test de seuil pour la méthode de segmentation par ratio : Seuil =0.6	112
A.4	Test de seuil pour la méthode de segmentation par ratio : Seuil =0.65	113
A.5	Test de seuil pour la méthode de segmentation par ratio : Seuil =0.8	114
B.1	Profil journalier, $h=15$ minutes.	115
B.2	Profil journalier, $h=15$ minutes.	116
B.3	Profil journalier, $h=30$ minutes.	116
B.4	Profil journalier, $h=30$ minutes.	117
B.5	Profil journalier, $h=45$ minutes.	117
B.6	Profil journalier, $h=45$ minutes.	118

Liste des tableaux

1.1	Caractéristiques du pyréliomètre CHP 1	13
1.2	Caractéristiques du pyranomètre CMP 22	13
1.3	Conditions physiquement possibles, S_0 représente la constante solaire ajustée pour la distance Terre-Soleil. μ est le cosinus de l'angle solaire zénithal $\cos(\theta_z)$	15
1.4	Conditions extrêmement rares, S_0 représente la constante solaire ajustée pour la distance Terre-Soleil. μ est le cosinus de l'angle solaire zénithal $\cos(\theta_z)$	15
1.5	Validations croisées.	15
1.6	Tests de qualité des données de rayonnement. Les pourcentages sont donnés par rapport au nombre de mesures de jour (angle zénithal $< 90^\circ$) soit un total de 507159 points de mesures sur 2 ans.	16
1.7	Résultats pour chacune des estimations	22
1.8	Résultats de classification des journées 2011 et 2012 en pourcentage (total de journées = 699)	25
2.1	Caractéristiques techniques du TSI	31
2.2	Classes utilisées pour classification par Heinle <i>et al.</i> [HMS10] (cf chap1)	32
2.3	Comparaison des résultats de classification : Les éléments de la ligne i et colonne j représentent le pourcentage de la classe i identifié comme classe j pour chacune des méthodes : H Heinle <i>et al.</i> [HMS10], K Kazantzidis <i>et al.</i> [Kaz+12], L Liu <i>et al.</i> [Liu+13]. (Cf tableau 2.2)	32
2.4	Récapitulatif des travaux de Chow <i>et al.</i> [Cho+11]	34
2.5	Récapitulatif des travaux de Marquez <i>et al.</i> [MC13a]	35
2.6	Récapitulatif des travaux de Fu <i>et al.</i> [FC13]	36
2.7	MCOO normalisée pour l'exemple :	43
2.8	Paramètres des méthodes de segmentation utilisées	45
2.9	Bilan des entrées brutes	49
2.10	Entrées sélectionnées après l'étude de la matrice d'inter-corrélation (Fig. 2.18)	49

2.11	Coefficient de corrélation entre chaque entrée et le GHI	54
2.12	Tableau des résultats de RMSE	56
2.13	Tableau des résultats de MBE	56
2.14	Tableau des résultats de coefficient de détermination	57
2.15	Résultats de RMSE par classe	57
2.16	Résultats de coefficient de détermination par classe	57
2.17	Résultats de MBE par classe	58
3.1	Les satellite de Meteosat	60
3.2	Tableau récapitulatif des travaux de la littérature	64
3.3	Performances d'HC-3 sur le site du SIRTA, mesures faites pour un angle solaire zénithal $< 85^\circ$	68
3.4	RMSE (W/m^2) en fonction de l'horizon et du modèle	76
3.5	R^2 en fonction de l'horizon et du modèle	76
3.6	MBE	77
3.7	Résultats par classes : RMSE	78
3.8	Résultats par classes : MBE	79
3.9	Résultats par classes : Coefficient de détermination	79
3.10	Choix des paramètres de l'algorithme	86
3.11	RMSE en fonction de l'horizon et du modèle	87
3.12	MBE en fonction de l'horizon et du modèle	87
3.13	R^2 en fonction de l'horizon et du modèle	88
3.14	Résultats en RMSE détaillés par classes	89
3.15	Résultats en MBE détaillés par classes	89
3.16	Résultats par classes : Coefficient de détermination	90
4.1	Caractéristiques du TSI retenues au chapitre 2	98
4.2	RMSE en Wm^{-2} en fonction du modèle et de l'horizon de prévision h	99

4.3	MBE en Wm^{-2} en fonction du modèle et de l'horizon de prévision	99
4.4	Coefficient de détermination R^2 en fonction du modèle et de l'horizon de prévision h	100

Table des acronymes

OMM	Organisation Météorologique Mondiale
GHI	<i>Global Horizontal Irradiance</i>
DNI	<i>Direct Normal Irradiance</i>
DHI	<i>Diffuse Horizontal Irradiance</i>
SIRTA	Site Instrumental de Recherche par Télédétection Atmosphérique
TSI	Total Sky Imager

Introduction

L'histoire de l'énergie photovoltaïque commence en 1839 lorsqu'Edmond Becquerel, un physicien français, observe la production d'électricité par certains matériaux lorsqu'ils sont soumis à la lumière du soleil. Einstein étudie plus tard ce phénomène et publie en 1905 un article au sujet de l'effet photo-électrique [Ein05] pour lequel il reçoit le prix Nobel en 1921. Cependant, l'énergie solaire est jugée trop coûteuse malgré les travaux de Robert Millikan et sa "cellule solaire" (1916). Elle est mise de côté pendant presque 40 ans. En 1960, la course à l'espace fait ressurgir la technologie solaire sur le devant de la scène : c'est une solution particulièrement adaptée à l'alimentation des premiers satellites. La recherche dans le domaine de l'énergie photovoltaïque bénéficie alors des fonds de la course à l'espace ce qui l'amène à devenir en 1970, avec la crise énergétique, une solution alternative. Dans le contexte énergétique actuel, l'énergie photovoltaïque est une solution viable et promue par de nombreux pays en reconversion énergétique.

L'exploitation de l'énergie solaire soulève cependant certains défis techniques. Au même titre que l'énergie éolienne, l'énergie photovoltaïque est une énergie dite intermittente : sa disponibilité varie fortement sans possibilité de contrôle. L'intégration de telles sources sur le réseau électrique engendre des problématiques pour le maintien de l'équilibre consommation/production. Ainsi, afin de limiter les risques, la loi autorise un taux de pénétration des énergies intermittentes sur les réseaux de 30% d'après l'arrêté du 23 Avril 2008³. Si ce seuil ne pose pas encore directement de contrainte sur les installations programmées en France métropolitaine, il en est autrement pour les réseaux peu ou non inter-connectés comme ceux des îles. D'après un état des lieux fait par le syndicat des énergies renouvelables⁴ les 30% avait déjà été atteint en Corse, en Guadeloupe et à la Réunion au 30 Juin 2012 alors que la Guyane présentait déjà un taux de pénétration de 22%.

Les principales pistes envisagées pour pallier cette limitation sont les suivantes :

- Le stockage : le stockage de l'énergie permettrait de stabiliser la production des centrales photovoltaïques en stockant l'électricité lors des périodes de forte production pour la restituer pendant les chutes de production. Cependant, le stockage d'importantes quantités d'énergie reste un challenge à l'heure actuelle. Des solutions existent comme les STEP (Stations de Transfert d'Énergie par Pompage-turbinage) mais elles restent peu accessibles. Aussi le développement des batteries à grande capacité de stockage est l'une des pistes les plus intéressantes.
- La prévision : la prévision de la production des énergies intermittentes permettrait d'anticiper la disponibilité des sources de production et donc de faciliter la gestion du réseau. C'est dans cette seconde voie que s'inscrivent les travaux menés dans cette thèse.

La production d'énergie électrique par un système photovoltaïque est dépendante de nombreux facteurs. L'inclinaison des panneaux, leur technologie et leur température influent sur la

3. <http://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000018698004>

4. http://www.enr.fr/docs/2012165422_notePV062012BAT2.pdf

production mais le rayonnement solaire reçu reste le facteur principal de variation. On choisit dans le cadre de cette thèse de prévoir le rayonnement solaire plutôt que la production elle-même. De ce fait, les travaux développés visent à produire une prévision non dépendante de conditions matérielles locales, et donc plus facilement transposable à d'autres sites ou moyens de production (solaire thermique, énergie solaire concentrée (ESC), etc). Cette approche sous-entend le besoin futur d'une étape de conversion entre le rayonnement solaire reçu par les panneaux et la production en sortie en fonction des caractéristiques des panneaux installés. On définit comme variable principale d'intérêt le GHI (Global Horizontal Irradiance), ou rayonnement global horizontal perçu au niveau du sol.

La prévision de GHI est un domaine de recherche très actif à l'heure actuelle pour les raisons évoquées précédemment. On distingue quatre grandes familles de méthode selon les données utilisées et l'horizon de prévision souhaité :

1. Méthodes statistiques : les méthodes statistiques s'appuient sur les séries temporelles d'un ou de plusieurs capteurs de GHI afin de construire des modèles statistiques de prévision. Ces méthodes supposent l'existence de séquences répétitives dans la série temporelle que l'on peut apprendre afin de les prévoir par la suite. On retrouve dans cette famille de nombreuses méthodes de régression (linéaires et non linéaires) existantes en traitement du signal (réseaux de neurones, modèles auto-régressifs, ...). L'horizon de prévision atteint par ce type de méthode est très dépendant de l'échantillonnage choisi pour la série temporelle.
2. Modèle de prévision météorologique numérique : les modèles numériques utilisent des algorithmes de résolution numérique des équations météorologiques assimilant les données météorologiques acquises par des stations de mesures réparties sur le globe. Ils fournissent ainsi des prévisions de nombreuses variables qualifiant l'état de l'atmosphère (température, pression, vent, etc.) pour chaque point de la grille spatiale de prévision. Selon les modèles, la résolution de cette grille peut être plus ou moins fine (10 km pour le modèle ALADIN de MétéoFrance), cependant elle reste inférieure à la résolution des images satellitaires dans la majorité des cas. Les prévisions nécessitent un temps de calcul important, la résolution temporelle des prévisions est de l'ordre de 3 à 6 heures. Ces modèles permettent une prévision du GHI à long terme (plusieurs jours) via différents traitements statistiques des sorties obtenues.
3. Imagerie satellitaire : les satellites météorologiques géostationnaires fournissent des images de l'atmosphère et de la terre sur l'ensemble du globe avec une résolution temporelle intra-horaire (15 minutes pour Meteosat seconde génération). Ces caractéristiques font de l'imagerie satellitaire un outil de premier ordre pour la description et le suivi de phénomènes météorologiques de moyenne et grande échelle. Ainsi, de nombreuses méthodes de prévision du GHI profitent de ces images depuis 1980. L'horizon de prévision atteint est généralement de l'ordre de quelques heures.
4. Imagerie hémisphérique : les images hémisphériques sont des images du ciel prises depuis le sol via une caméra possédant un angle de vision de 180° . Les travaux impliquant ces images ont émergé il y a une vingtaine d'années avec l'apparition de capteurs de plus en plus performants. Ces images permettent une description très fine de la couverture nuageuse avec une résolution temporelle de l'ordre de la seconde. Elles sont particuliè-

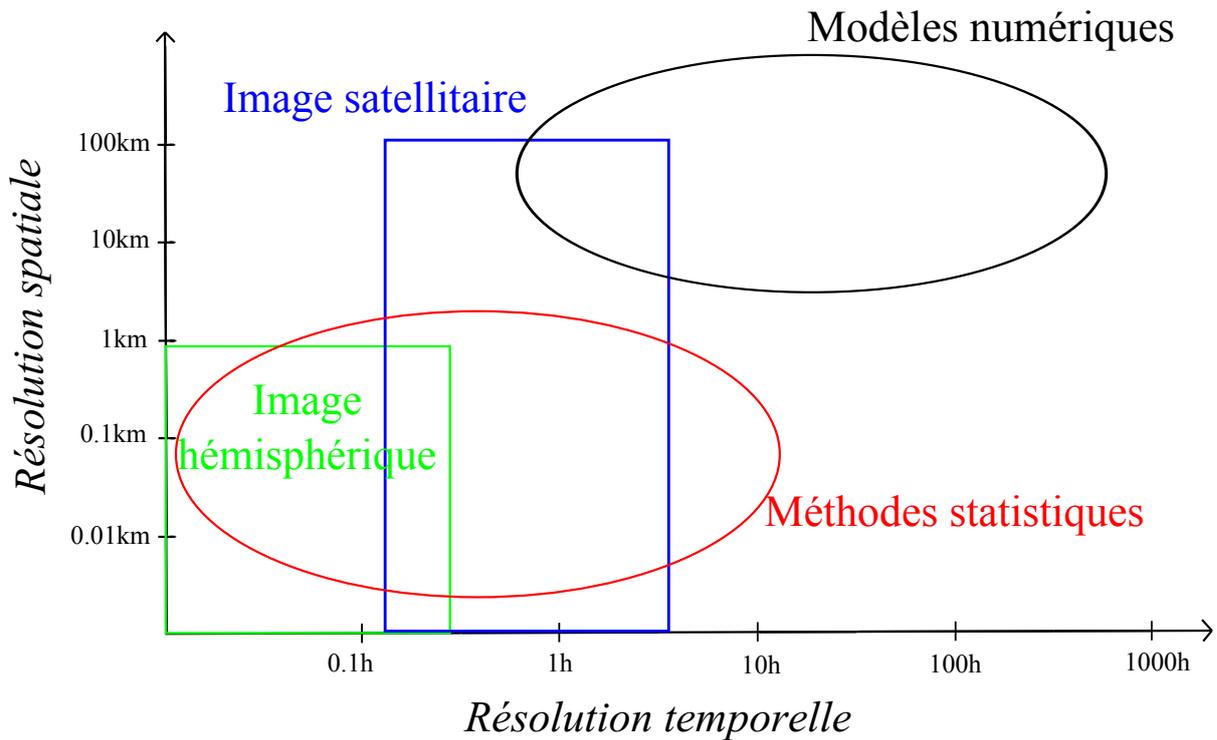


FIGURE 1 – Les quatre grandes familles de modèles de prévision (figure inspirée de [Dia+13]).

rement adaptées à l'estimation et à la prévision à très court terme (jusqu'à 30 minutes) du GHI.

La figure 1 résume la répartition de ces familles selon l'horizon de prévision et la résolution qu'ils offrent. Ces quatre familles ont évolué de façon indépendante améliorant leurs qualités respectives. Cependant, de récentes études rendent les frontières moins hermétiques. Par exemple, [MCK13] propose une méthode basée sur un modèle numérique de prévision mais intégrant des données satellitaires concernant la couverture nuageuse. Ces travaux ouvrent la voie à de nouvelles techniques, qui permettront d'une part d'élargir les horizons de prévisions initialement atteints par les méthodes indépendantes mais également de combler leurs lacunes. La problématique développée au cours de cette thèse s'inscrit dans cette dynamique : comment peut-on joindre les données satellitaires et les données in-situ (pyranomètre, images hémisphériques) dans le cadre d'une prévision intra-horaire ? Il s'agira dans un premier temps de caractériser l'apport de chacune de ces sources d'information afin de les utiliser conjointement par la suite. Nous pourrons alors juger de la pertinence d'une telle fusion en fonction de l'horizon de prévision. Le manuscrit s'articule ainsi en quatre étapes :

Le premier chapitre permet d'introduire les notions fondamentales à l'étude du GHI. On décrira les processus physiques théoriques expliquant ses divers degrés de variabilité. Ce sera alors l'occasion de rappeler les différentes interactions entre le sommet de l'atmosphère et le sol. Nous présenterons ensuite le site d'étude retenu dans le cadre de la thèse, ainsi que les capteurs utilisés. Les données obtenues seront analysées afin d'établir un jeu de données valides permettant de tester les modèles développés par la suite. Enfin, on établira un modèle de

référence pour la prévision, le modèle de persistance, avant d'ouvrir la discussion sur quelques observations qualitatives concernant l'étude des séries temporelles de GHI.

Le second chapitre est consacré à l'étude des images hémisphériques. On y trouvera en première partie un état de l'art des travaux récents utilisant ces images pour diverses applications, de la segmentation de couverture nuageuse à la prévision de GHI à très court terme. La suite du chapitre présente une méthode originale d'estimation du GHI. Cette méthode s'appuie sur l'utilisation d'un réseau de neurones permettant d'effectuer une régression entre les caractéristiques de l'image et le GHI à estimer. L'enjeu principal de l'étude réside donc dans le choix des caractéristiques extraites des images définissant ainsi les variables pertinentes pour l'étude du GHI par imagerie hémisphérique.

Le troisième chapitre concerne l'ensemble des travaux impliquant l'imagerie satellitaire. Après un état de l'art, nous présenterons deux approches construites à partir d'images satellitaires transformées en carte de GHI via une méthode de conversion externe (Héliosat-2). La première peut s'apparenter à une méthode "hybride" au sens où elle intègre des données sol (pyranomètre) et des données de l'image satellitaire dans un modèle autoregressif spatio-temporel. C'est une méthode robuste, nécessitant une base d'apprentissage permettant d'assimiler les conditions météorologiques "moyennes" pour améliorer la prévision. La seconde méthode développée se veut plus locale et fait intervenir un vecteur de déplacement calculé à partir des dernières images reçues. Elle développe ainsi plus pleinement le potentiel des images satellitaires mais ne bénéficie pas de l'information prise au sol. La comparaison des résultats de ces deux modèles permettra d'établir un premier bilan concernant la "fusion" d'information entre le satellite et le sol.

Le quatrième chapitre rassemble les connaissances acquises au cours des chapitres précédents pour proposer une première méthode de fusion des informations. La méthode développée consiste à améliorer les performances d'une méthode satellitaire en y intégrant de l'information provenant de l'image hémisphérique. On s'appuiera pour cela sur les résultats des chapitres 2 et 3. Les résultats obtenus permettront d'identifier l'intérêt d'une telle fusion.

Enfin, la conclusion fera le bilan des résultats obtenus afin de dégager les enjeux principaux relatifs à la fusion d'information pour la prévision à très court terme de GHI. Nous ouvrirons ensuite la discussion sur les perspectives de futurs travaux en rapport avec les conclusions retenues.

Le rayonnement global : généralités

Sommaire

1.1	Fondamentaux sur le rayonnement	6
1.1.1	Définitions	6
1.1.2	Le rayonnement solaire	7
1.1.3	Les nuages	8
1.2	Mesures	10
1.2.1	Site d'étude : le SIRTA	10
1.2.2	Capteurs pyranométriques	12
1.2.3	Contrôle de qualité et bilan des données	15
1.3	Fondamentaux pour la prévision du GHI	19
1.3.1	Modèle ciel clair	19
1.3.2	Index de ciel clair Kc	20
1.3.3	Modèle de référence : la persistance	20
1.3.4	Métriques	21
1.4	Caractérisation du site d'étude	23
1.4.1	Variabilité	23
1.4.2	Résolution temporelle	26
1.5	Conclusion	26

Ce chapitre introduit l'éclairement global horizontal au sol, variable principale de cette thèse, qui sera noté GHI (Global Horizontal Irradiance). Après un rapide rappel des fondamentaux sur le rayonnement électromagnétique et les transferts radiatifs, on présentera dans un cadre plus pratique les capteurs de rayonnement utilisés ainsi que le site de mesures. On donnera ensuite les notions élémentaires nécessaires à l'estimation ou la prévision du GHI tels que les métriques ou l'index de ciel clair Kc qui seront utilisés pour proposer une classification des types de journées observées sur le site. Enfin, une discussion sur la résolution temporelle des séries de mesure permettra de mettre en avant les problématiques liées à l'évaluation d'un modèle d'estimation ou de prévision du GHI.

1.1 Fondamentaux sur le rayonnement

1.1.1 Définitions

Ondes Électromagnétiques : une onde électromagnétique est composée d'un champ électrique et d'un champ magnétique. Ces deux champs oscillent perpendiculairement entre eux et perpendiculairement à l'axe de propagation de l'onde (Fig. 1.1). La longueur d'onde λ est définie par la distance entre 2 maxima de la sinusoïde, elle est liée à la période T par l'équation 1.1 où c est la vitesse de propagation de l'onde. La fréquence f est l'inverse de la période. Elle est communément utilisée pour décrire les ondes électromagnétiques et les classer dans différentes catégories (Fig. 1.2).

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

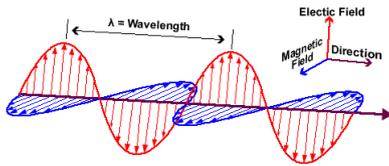


FIGURE 1.1 – Représentation d'une onde électromagnétique

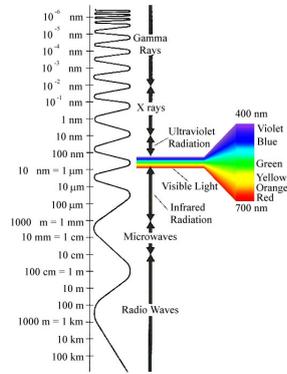


FIGURE 1.2 – Classification du spectre électromagnétique

Angle solide : la notion d'angle solide intervient dans la définition des termes énergétiques utilisés en transfert radiatif. Il est défini à partir d'une sphère unitaire centrée en M (Fig. 1.3). L'angle solide élémentaire ($d\Omega$) dans la direction définie par l'angle θ par rapport à la normale \vec{n} et l'angle ϕ dans le plan de la surface de la sphère représente la fraction de sphère dans cette direction. L'angle solide entier correspond à la surface de la sphère unité, i.e $\Omega = 4\pi$. L'unité d'angle solide est le stéradian (sr).

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

De la luminance au GHI : la luminance monochromatique $L_f(\theta, \phi)$ est la puissance électromagnétique d'une onde de fréquence f se propageant à travers l'angle solide élémentaire dans la direction définie par (θ, ϕ) . Son unité est le $W.m^{-2}.sr^{-1}.Hz^{-1}$. La luminance $l(\theta, \phi)$ est obtenue par intégration de cette quantité sur l'ensemble du spectre solaire, elle est donnée en $W.m^{-2}.sr^{-1}$. Enfin, l'éclairement L est calculé par intégration de la luminance sur l'hémisphère

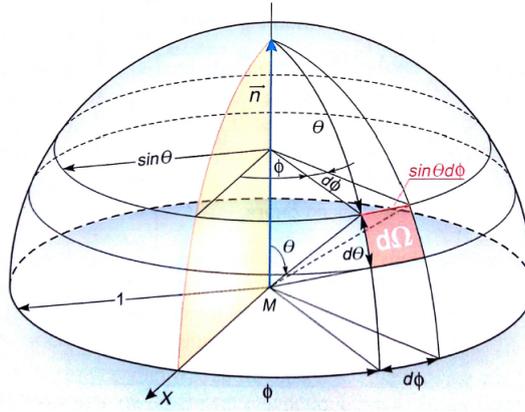


FIGURE 1.3 – Solid angle definition [Mal05]

exposé, dans le cas d'un rayonnement isotropique, l'éclairement est donnée par l'équation 1.3. Le GHI correspond à l'éclairement solaire perçu au niveau du sol sur une surface horizontale.

$$L = \int_{\phi=0}^{2\pi} \int_{\theta=0}^{\pi/2} l(\theta, \phi) \cos \theta d\Omega = \pi l \quad (1.3)$$

Interaction avec la matière : quand une onde électromagnétique de longueur d'onde λ rencontre de la matière, une partie de l'onde est réfléchi, une autre absorbée par la matière et le reste est transmis ou diffusé. Plusieurs paramètres sont définis pour chaque longueur d'onde pour décrire ces phénomènes et caractériser la matière :

- Le facteur d'absorption β_λ est le ratio entre le rayonnement absorbé et le rayonnement incident
- Le facteur de réflexion ou albédo monochromatique δ_λ est le ratio entre le rayonnement réfléchi et le rayonnement incident
- Le facteur de transmission τ_λ est le ratio entre le rayonnement transmis et le rayonnement incident

1.1.2 Le rayonnement solaire

Le spectre solaire Le soleil peut être assimilé à un corps noir qui émet un rayonnement électromagnétique dont le spectre est présenté en figure 1.4. Sur cette figure, on note la différence entre le rayonnement arrivant au sommet de l'atmosphère et le rayonnement arrivant au niveau du sol. Cette différence est liée aux bandes d'absorption des gaz de l'atmosphère. En effet, les gaz ont des facteurs d'absorption et de transmission qui dépendent beaucoup de la fréquence des ondes (Fig. 1.5). On remarque notamment que la vapeur d'eau absorbe une grande partie de la partie des bandes de l'infrarouge thermique reçu. Ce phénomène explique une partie de la variabilité du rayonnement puisque la composition atmosphérique est variable

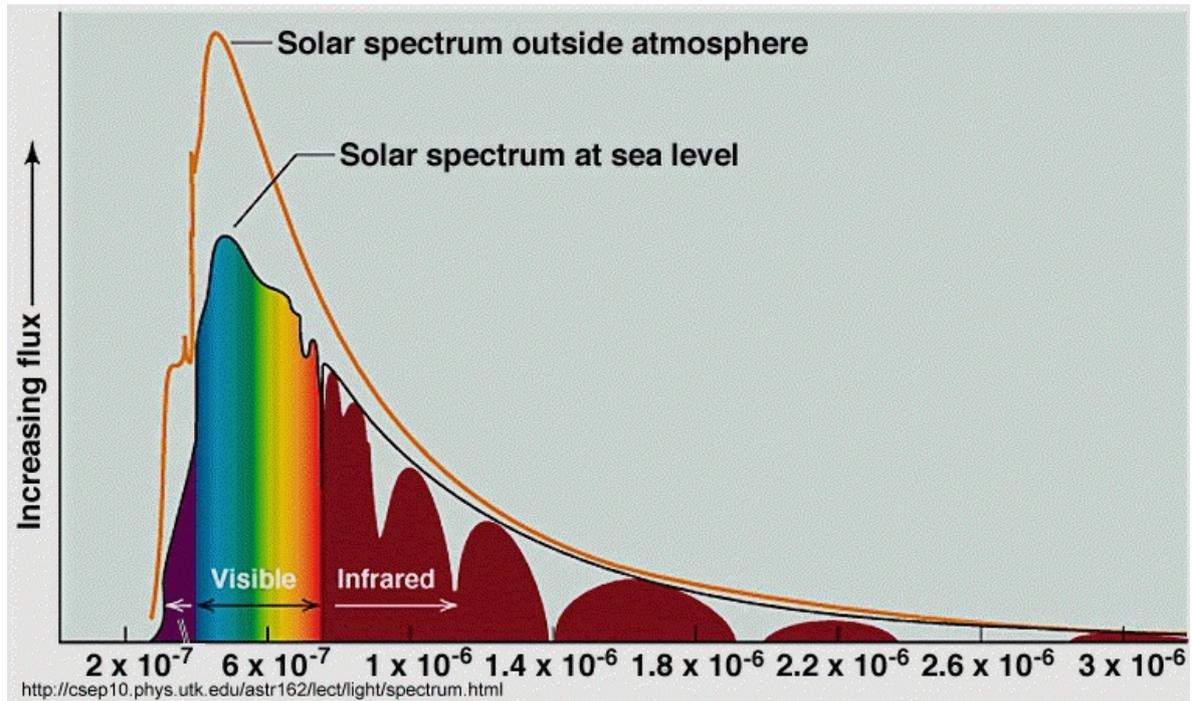


FIGURE 1.4 – Spectre du rayonnement solaire

dans le temps et dans l'espace.

Transferts moyens la moyenne annuelle de rayonnement arrivant au sommet de l'atmosphère est la constante solaire S ($1368 \text{ W}\cdot\text{m}^{-2}$) [KL11]. Il s'agit de la quantité de rayonnement électromagnétique qui serait reçue sur un plan perpendiculaire aux rayons, à une distance de 1 unité astronomique (UA). En considérant la rotation de la Terre et l'angle avec lequel les rayons atteignent l'atmosphère selon l'endroit considéré, le rayonnement solaire moyen est estimé à un quart de la constante solaire ($340 \text{ W}/\text{m}^2$). C'est la valeur de référence pour la figure 1.6 qui présente un bilan énergétique global du rayonnement solaire sur Terre. En moyenne, seul 48% du rayonnement incident parvient au sol, l'atmosphère et les nuages retenant l'autre moitié soit par réflexion (23%), soit par absorption (23%). L'étude de la couverture nuageuse et des paramètres de l'atmosphère joue donc un rôle capital pour l'étude du rayonnement global reçu au sol, ce qui constitue l'enjeu principal de cette étude avec une approche beaucoup plus locale.

1.1.3 Les nuages

Les nuages sont constitués de petites gouttes d'eau et de cristaux de glace en suspension dans l'air. La plupart des nuages sont situés dans la troposphère; ils se forment par condensation habituellement au niveau des courants ascendants. Selon leur altitude et leur composition, ils présentent diverses formes et propriétés. L'Organisation Météorologique Mondiale (WMO)

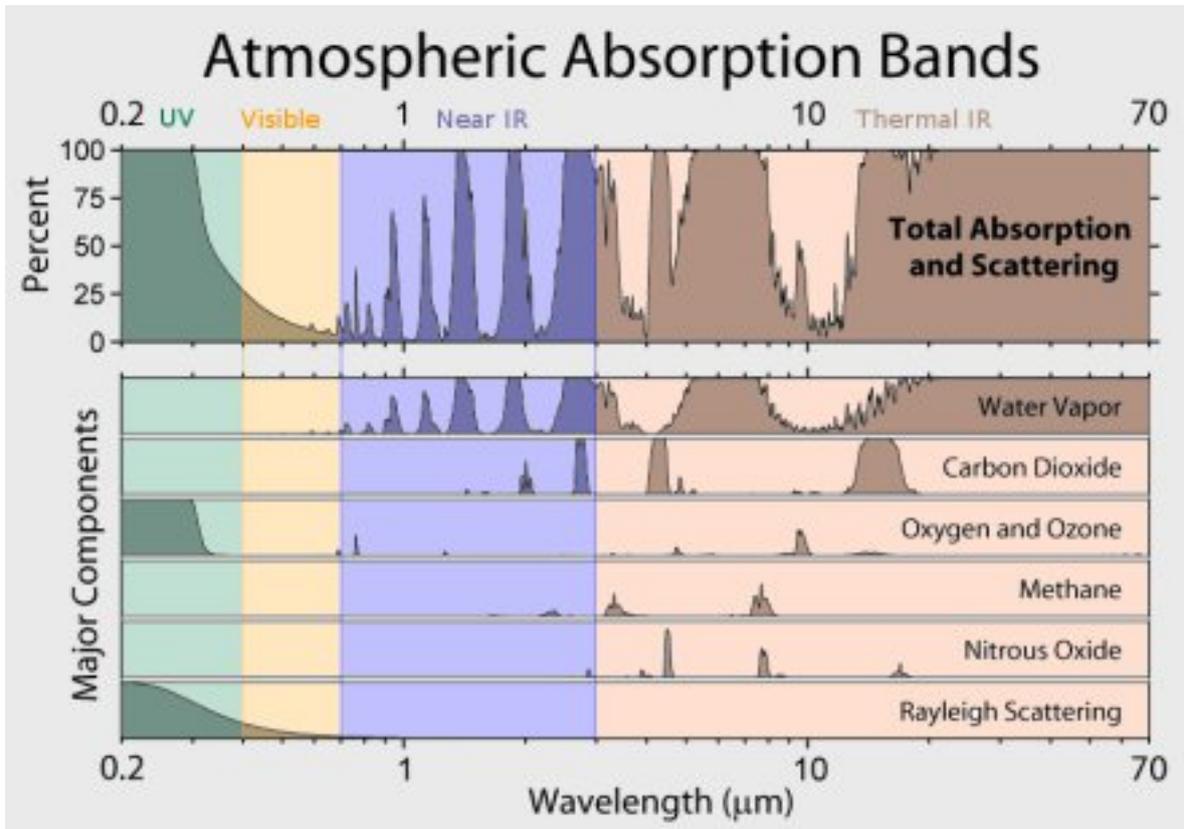


FIGURE 1.5 – Bandes d’absorption de l’atmosphère

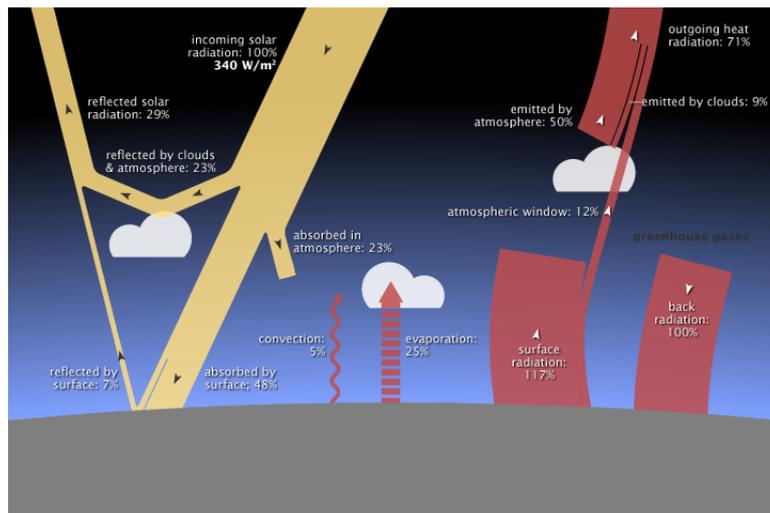


FIGURE 1.6 – Bilan du rayonnement solaire¹

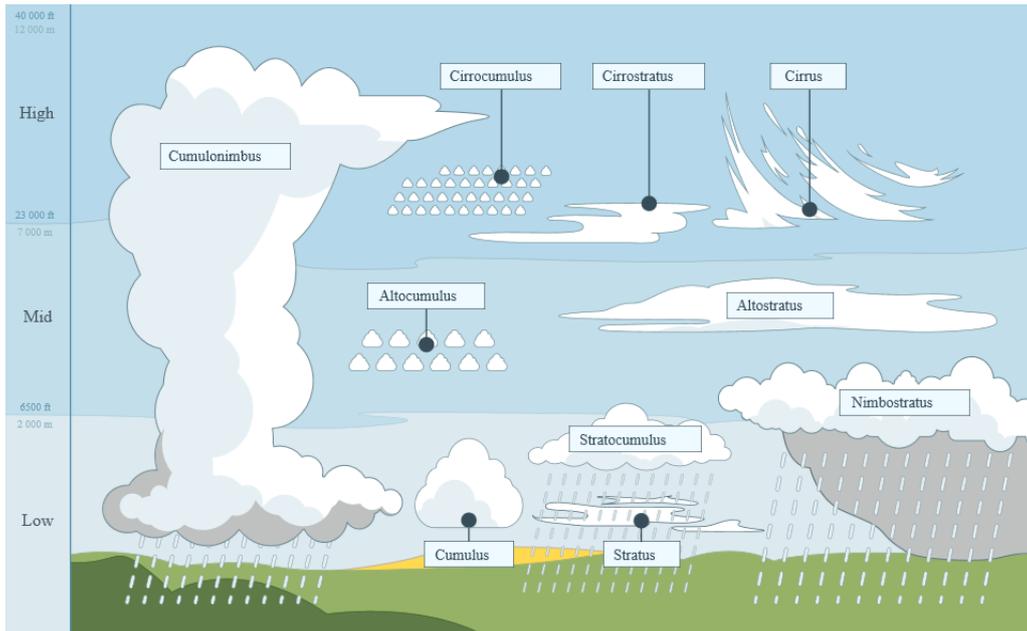


FIGURE 1.7 – Classification des types de nuages [Mal05]

propose une classification des nuages en 10 principales classes (Fig. 1.7) :

L'impact des nuages sur le rayonnement perçu 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 liquide, etc). Cependant les nuages ont généralement un effet réducteur sur le rayonnement reçu au sol puisqu'ils réfléchissent une partie de la lumière incidente vers l'espace. Diverses approches seront abordées au cours de cette étude pour caractériser de façon quantitative l'effet de la couverture nuageuse sur le rayonnement au sol.

1.2 Mesures

1.2.1 Site d'étude : le SIRTa

L'ensemble des données de rayonnement utilisées dans cette étude a été acquis sur le site du SIRTa (Site Instrumental de Recherche par Télédétection Atmosphérique). Situé sur le plateau de Palaiseau², ce site dispose de nombreux capteurs nécessaires à l'étude de l'atmosphère et plus particulièrement du rayonnement incident au sol (Fig. 1.8). Il fait parti du BSRN (Baseline Radiation Surface Network), un réseau regroupant 40 stations de mesure dans le monde et garantissant un haut niveau de qualité et de suivi des instruments de mesure. Les capteurs impliqués dans cette étude sont présentés dans la suite.

2. (48.7130 N, 2.2080 E), 156m)



FIGURE 1.8 – Le SIRTA en image³

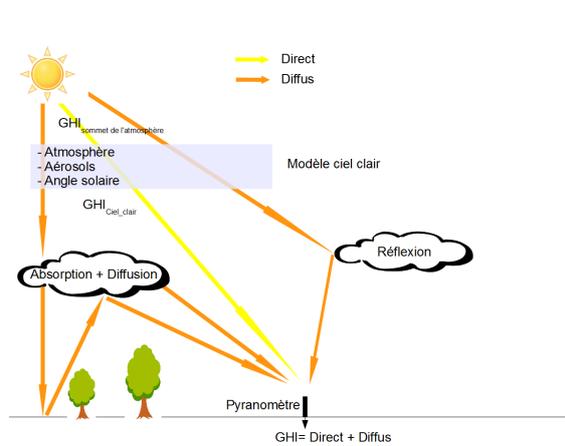


FIGURE 1.9 – Rayonnement direct et diffus : différents chemins

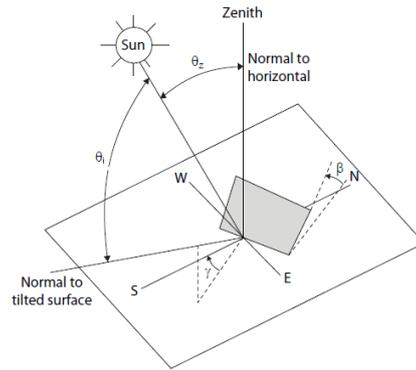


FIGURE 1.10 – Les angles couramment utilisés pour le rayonnement solaire : θ_z est l'angle zénithal, θ_i est l'angle d'élévation solaire, γ est l'angle azimutal

1.2.2 Capteurs pyranométriques

Les radiations à faibles longueurs d'onde ($0.28\text{-}5\ \mu\text{m}$) peuvent être séparées en trois composantes (Fig. 1.9) :

- Le rayonnement direct en incidence normale (DNI) : il s'agit de la mesure du rayonnement arrivant par un chemin direct depuis le soleil sur un plan perpendiculaire à la direction du rayon. L'OMM recommande l'utilisation d'un pyréliomètre pour sa mesure.
- Le rayonnement diffus horizontal (DHI) : il s'agit de la mesure 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 rayonnement global horizontal (GHI) : il s'agit de la combinaison du rayonnement diffus et de la partie horizontale du rayonnement direct. Il peut être calculé à partir des précédents par l'équation 1.4 où θ_z représente l'angle zénithal décrit en Fig. 1.10.

$$GHI = \cos(\theta_z) \cdot DNI + DHI \quad (1.4)$$

La zone 2 du SIRTA (située sur un toit) est équipée d'une station automatique Kipp&Zonen 2AP (Fig. 1.11) depuis Septembre 2002⁴. Cette station est constituée d'un "sun tracker" lui-même équipé de trois capteurs Kipp&Zonen :

- Un pyréliomètre pour la mesure du rayonnement normal direct (Fig. 1.12).
- Un pyranomètre pour la mesure du rayonnement horizontal global (Fig. 1.13).
- Un pyranomètre masqué pour la mesure du rayonnement horizontal diffus.

Les spécifications de ces capteurs sont données dans les tableaux⁵ 1.1 et 1.2. Conformément à la norme BSRN, ils présentent les meilleures performances actuellement disponibles. Les

4. <http://sirta.ipsl.fr/index.php/instruments/instruments-radiometers/136-sirta-instruments-radiometers-fluxmeters>

5. <http://www.kippzonen.com/Product/>.

Spectral range	200 to 4000 nm
Sensitivity	7 to 14 $\mu V/W/m^2$
Response time	<5s
Field of view	$5 \pm 0.2^\circ$
Operational temperature range	-40 to 80 °C
Maximum solar irradiance	4000 W/m^2
International standards (ISO)	First class

TABLE 1.1 – Caractéristiques du pyréliomètre CHP 1

Spectral range	200 to 3600 nm
Sensitivity	7 to 14 $\mu V/W/m^2$
Response time	<5s
Field of view	180°
Operational temperature range	-40 to 80 °C
Maximum solar irradiance	4000 W/m^2
International standards (ISO)	First class

TABLE 1.2 – Caractéristiques du pyranomètre CMP 22

mesures sont effectuées toutes les minutes et correspondent à une moyenne de 60 mesures (1 par seconde). Un exemple de journée de mesure est donné en figure 1.14.



FIGURE 1.11 – Station automatique Kipp&Zonen 2AP



FIGURE 1.12 – Pyréliomètre CHP 1



FIGURE 1.13 – Pyranomètre CMP 22

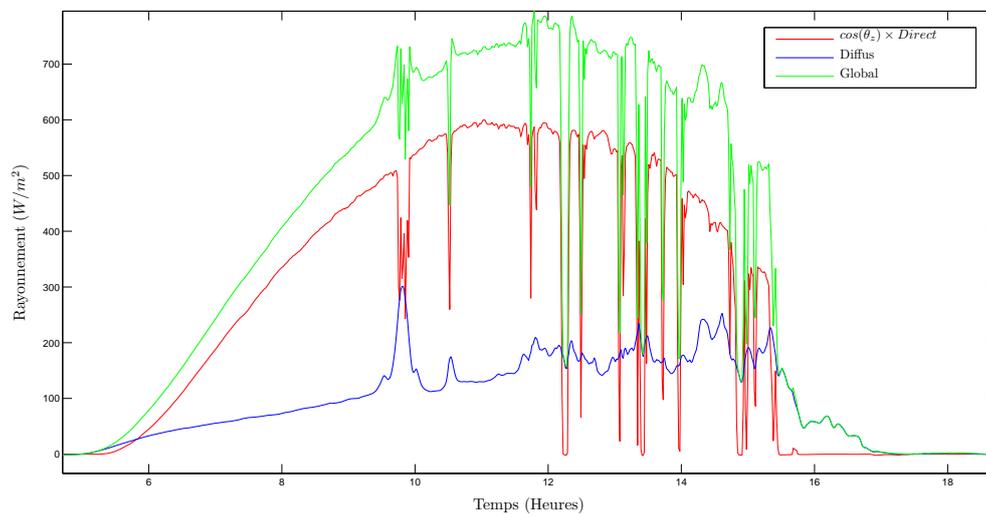


FIGURE 1.14 – Exemple de mesures sur une journée

Paramètre	Borne inférieure	Borne supérieure
Global	$-4 W/m^2$	$1.5 \cdot S_0 \mu^{1.2} + 100 W/m^2$
Diffus	$-4 W/m^2$	$0.95 \cdot S_0 \mu^{1.2} + 100 W/m^2$
Direct	$-4 W/m^2$	$S_0 \mu^{1.2}$

TABLE 1.3 – Conditions physiquement possibles, S_0 représente la constante solaire ajustée pour la distance Terre-Soleil. μ est le cosinus de l'angle solaire zénithal $\cos(\theta_z)$.

Paramètre	Borne inférieure	Borne supérieure
Global	$-2 W/m^2$	$1.2 S_0 \mu^{1.2} + 50 W/m^2$
Diffus	$-2 W/m^2$	$0.75 S_0 \mu^{1.2} + 30 W/m^2$
Direct	$-2 W/m^2$	$0.95 \cdot S_0 \mu^{1.2} + 10 W/m^2$

TABLE 1.4 – Conditions extrêmement rares, S_0 représente la constante solaire ajustée pour la distance Terre-Soleil. μ est le cosinus de l'angle solaire zénithal $\cos(\theta_z)$.

1.2.3 Contrôle de qualité et bilan des données

Les mesures de rayonnement effectuées sur un site du BSRN passent plusieurs tests de qualité [Roe+11]. Ces tests sont décrits dans les tableaux 1.3 à 1.5. Il s'agit principalement de vérifier d'une part la présence des données mais aussi leur cohérence physique :

- Le premier test porte sur la possibilité physique des mesures faites (Tableau 1.3).
- Le second test porte sur les conditions dites extrêmes : elles sont tolérées dans la mesure où leur apparition reste exceptionnelle (Tableau 1.4).
- Enfin le dernier test est un test croisé entre les différentes mesures de rayonnement (direct/diffus/global) afin de vérifier leur cohérence physique (Tableau 1.5).

Dans le cadre de cette thèse, seules les données acquises en 2011 et 2012 sur le site du SIRTa ont été utilisées. La figure 1.15 illustre les deux cycles annuels de GHI mesuré. On observe parfois des bandes sombres indiquant un manque de données. Les tests précédemment présentés ont été appliqués sur ces données, les résultats sont synthétisés dans le tableau 1.6 et ne tiennent pas compte des 32 jours de données manquantes sur les 2 années. Ce tableau indique effectivement une très grande qualité des données : on ne remarque quasiment aucune détection sur les tests croisés et aucune donnée physiquement impossible ou très peu dans le cas du capteur de rayonnement direct. Enfin, on note également très peu de "conditions extrêmes" (<0.5% Tableau 1.6). Selon les capteurs, ces conditions ne semblent pas être les mêmes. Les figures 1.16 à 1.18 présentent la localisation temporelle des mesures détectées comme extrêmes

Test	Condition
Test 1	Global/(Diffus + Direct) = $1.0 \pm 8\%$ for GLOB $> 50 W/m^2$, $\theta_z < 75^\circ$ Global/(Diffus + Direct) = $1.0 \pm 15\%$ for GLOB $> 50 W/m^2$, $75^\circ < \theta_z < 93^\circ$
Test 2	Diffus/Global < 1.05 for Global $> 50 W/m^2$, $\theta_z < 75^\circ$ Diffus/Global $< 1.0 \pm 10\%$ for Global $> 50 W/m^2$, $75^\circ < \theta_z < 93^\circ$

TABLE 1.5 – Validations croisées.

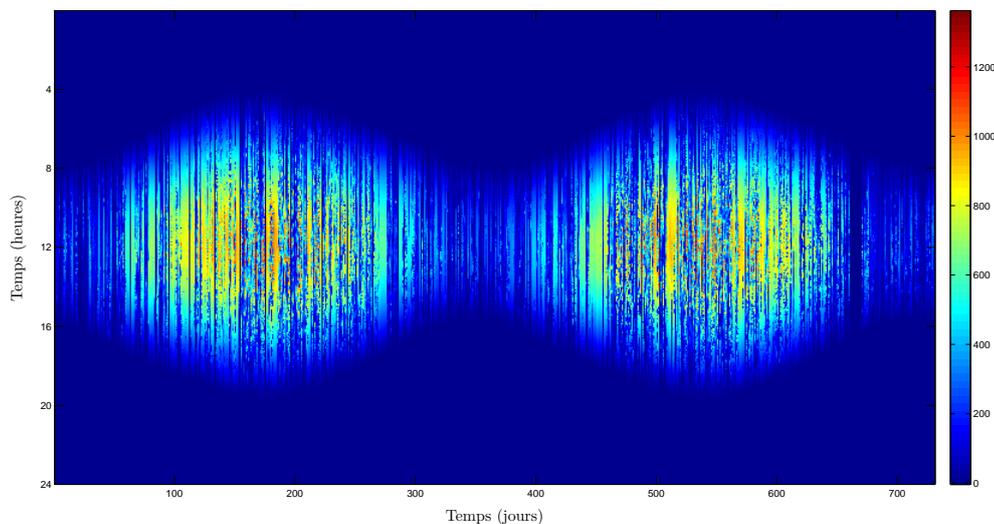


FIGURE 1.15 – Mesures de GHI sur les années 2011 et 2012 au Sirta.

	Direct	Diffus	Global
Physiquement impossible	379 (<0.1%)	0 (0%)	0 (0%)
Condition extrêmes	1954 (0.3%)	1501 (0.3%)	2062 (0.4%)
Test1	0 (0%)		
Test2	20 (<0.1%)		

TABLE 1.6 – Tests de qualité des données de rayonnement. Les pourcentages sont donnés par rapport au nombre de mesures de jour (angle zénithal <math><90^\circ</math>) soit un total de 507159 points de mesures sur 2 ans.

selon les tests du tableau 1.4 pour chaque capteur. Pour le rayonnement global, les erreurs apparaissent majoritairement la nuit, où des valeurs inférieures à -2 W/m^2 sont enregistrées. Pour le diffus, il s'agit plus ou moins du même cas mais on remarque quelques points en journée, avec une concentration particulière sur quelques journées de 2012. Enfin, les conditions extrêmes pour le rayonnement direct sont surtout observées pour les rayonnements très forts, la plupart du temps en milieu de journée en été.

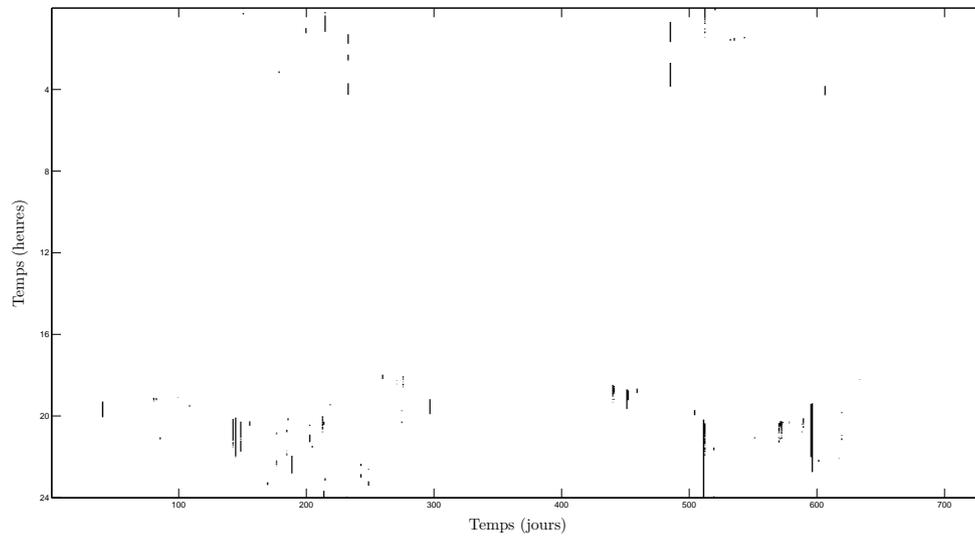


FIGURE 1.16 – Détection des conditions extrêmes sur le rayonnement global en 2011/2012 au SIRTA.

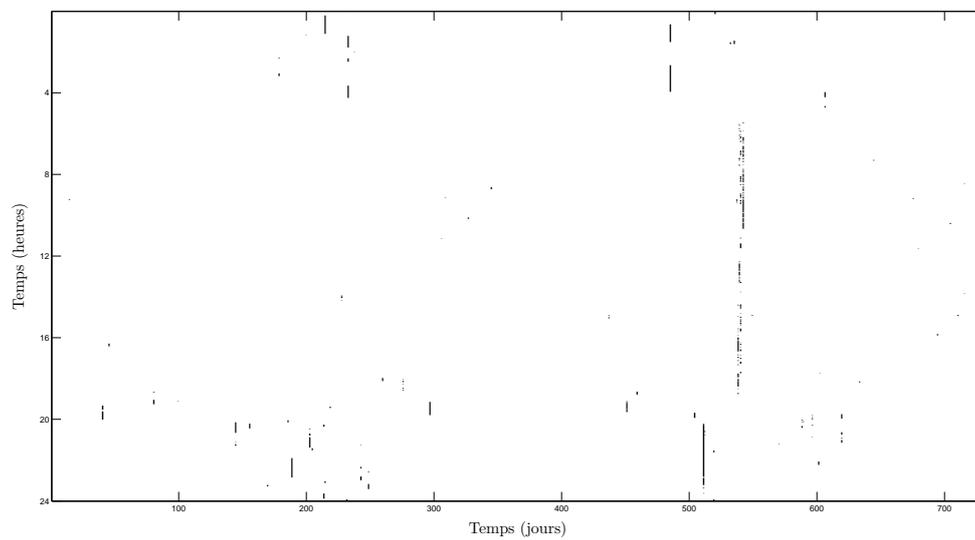


FIGURE 1.17 – Détection des conditions extrêmes sur le rayonnement diffus en 2011/2012 au SIRTA.

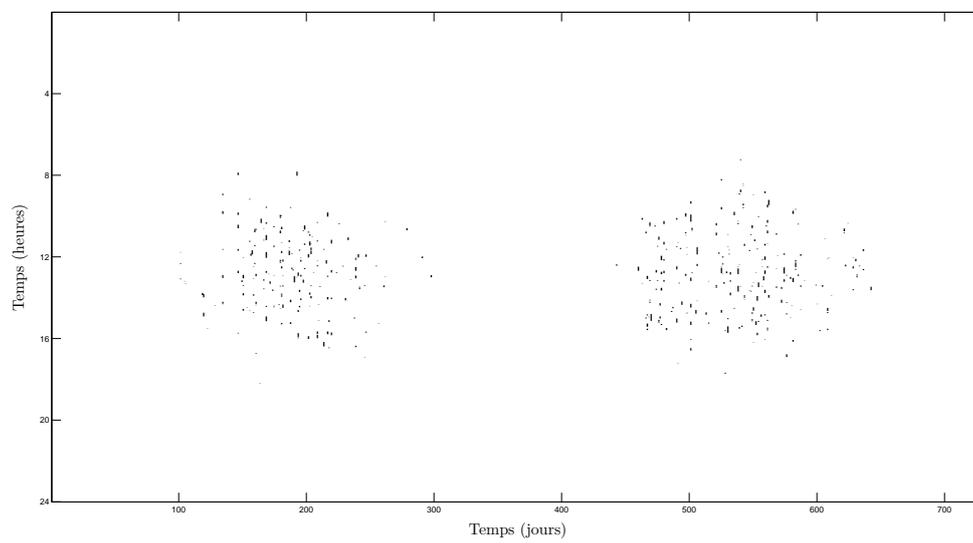


FIGURE 1.18 – Détection des conditions extrêmes sur le rayonnement direct en 2011/2012 au SIRTA.

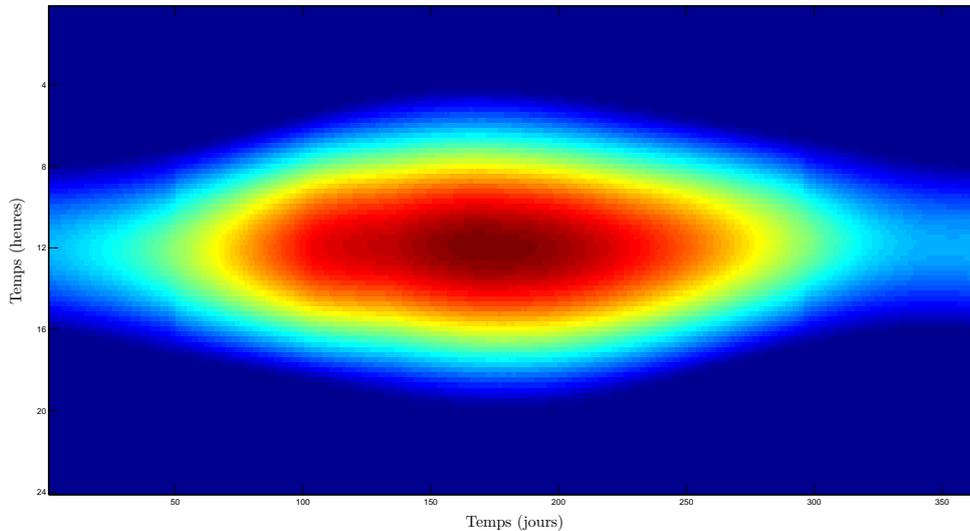


FIGURE 1.19 – Modèle ciel clair ESRA pour l'année 2011 évalué sur le site du SIRTA.

1.3 Fondamentaux pour la prévision du GHI

1.3.1 Modèle ciel clair

On définit habituellement deux types de variabilités pour le rayonnement au sol :

- La variabilité liée à la géométrie Terre - Soleil (rayonnement hors atmosphère) et à l'atmosphère sans nuage : le "ciel clair".
- La variabilité liée à la couverture nuageuse.

Le rayonnement solaire par ciel clair est le rayonnement observé lorsqu'il n'y a aucun nuage. Il dépend donc de paramètres géométriques (distance Terre/Soleil, altitude) mais aussi de paramètres de l'atmosphère, puisque même sans nuages, les gaz et les particules en suspension, les aérosols, constituant l'atmosphère ont des interactions avec le rayonnement solaire qui dépendent des caractéristiques des particules. Plusieurs modèles ont été présentés dans la littérature pour caractériser le rayonnement "ciel clair" ([Gue08], [Mue+04], [Lef+13], etc.), une étude comparative a été publiée en 2006 [Ine06] puis en 2013 [IPC13]. On retiendra ici le modèle dit ESRA (European Solar Radiation Atlas) [RBW00] pour des raisons qui seront développées plus tard (section 3.2.3). Ce modèle tient compte à la fois des paramètres géométriques mais également de la variation de composition de l'atmosphère par l'intermédiaire du facteur de trouble de Linke (Linke Turbidity factor). Ce facteur inclut la réflexion du rayonnement par les aérosols mais aussi son absorption par les gaz de l'atmosphère. Ainsi, il caractérise l'atténuation du rayonnement par l'atmosphère : plus il est grand, plus l'atmosphère ciel clair atténue le rayonnement incident. Il est établi pour le modèle ESRA à partir d'une base de données de Trouble de Linke en moyenne mensuelle climatologique. En pratique, le modèle ciel clair présente les cycles journaliers (jour/nuit) et annuels comme illustré sur la figure 1.19.

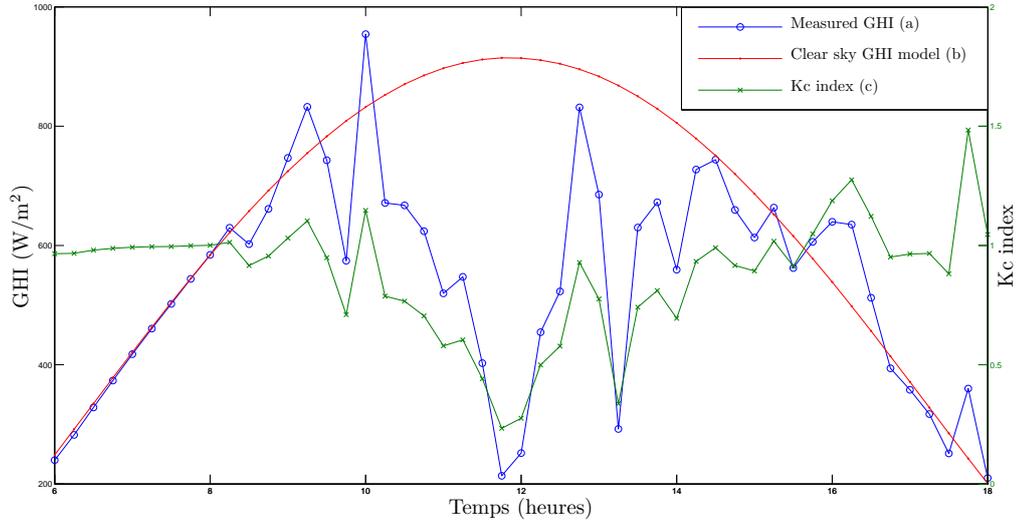


FIGURE 1.20 – Présentation de l'index Kc sur une journée de mesure

1.3.2 Index de ciel clair Kc

A partir du modèle ciel clair, on définit l'index Kc comme le rapport entre le rayonnement mesuré et le modèle ciel clair (1.5). Cet index sans unité représente les variations liées uniquement à la couverture nuageuse sur site. On obtient une variable théoriquement bornée entre 0 et 1 (Fig. 1.20), même si des valeurs supérieures à 1 sont parfois observées dans certaines conditions particulières. Il peut s'agir d'un effet lentille ou des nuages [Gue93], ou d'une sous-estimation du modèle de ciel clair utilisé. En étudiant cette variable plutôt que le GHI, on s'affranchit des cycles connus du rayonnement pour se concentrer uniquement sur les variations liées à l'environnement direct du capteur. Dans le cas d'une journée ciel clair, on obtient une variable constante et égale à 1 toute la journée. Cet indice est donc généralement préféré au GHI dans la plupart des études sur la prévision ou l'estimation locale du GHI.

$$Kc = \frac{GHI}{GHI_{CielClair}} \quad (1.5)$$

1.3.3 Modèle de référence : la persistance

Le modèle de persistance est reconnu comme la référence des méthodes de prévision du GHI pour sa simplicité et son universalité. Il s'appuie sur l'hypothèse de persistance de la mesure d'un instant à un autre de proche en proche. La prévision d'une variable temporelle $X(t)$ par persistance consiste à écrire l'équation suivante :

$$\hat{X}(t+1) = X(t) \quad (1.6)$$

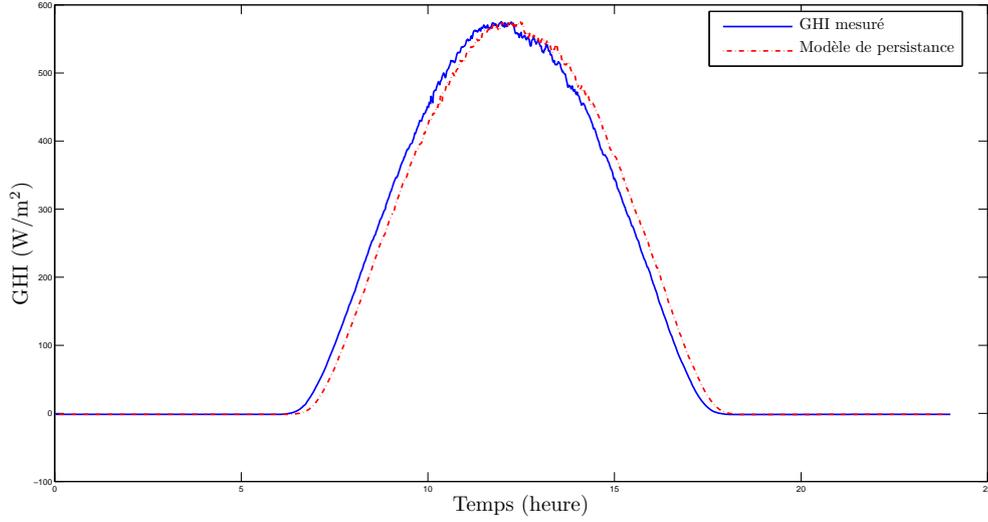


FIGURE 1.21 – Modèle de persistance appliqué directement au GHI pour une journée ciel clair.

L'application de ce modèle à la prévision du GHI nécessite l'utilisation de l'index K_c . En effet, en appliquant le modèle directement au GHI on obtiendrait une prévision toujours en décalage avec la mesure à cause de la variation journalière connue (Fig. 1.21). La prévision de GHI par persistance s'écrit donc avec les équations suivantes :

$$\begin{aligned} \hat{GHI}(t+1) &= \hat{K}_c(t+1) \times GHI_{CielClair}(t+1) \\ \hat{K}_c(t+1) &= K_c(t) \end{aligned} \quad (1.7)$$

Cette méthode de référence a l'avantage d'être très performante dès que les conditions météorologiques sont assez stables. Les scores de persistance sont d'ailleurs très importants pour évaluer et comparer des méthodes de prévision développées et testées sur des jeux de données différents. Les scores du modèle de persistance donnent d'une part une référence pour la prévision mais aussi une information sur la difficulté de celle-ci, ou du moins sur la variabilité relative des jeux de données. Un bon score du modèle de persistance indique un site où les conditions météorologiques sont relativement stables et vice versa [MC13b].

La section suivante présente les différentes métriques utilisées pour évaluer les méthodes développées par rapport à la référence.

1.3.4 Métriques

Soit la série temporelle X composée de N éléments $x(k)$, on note \hat{X} une estimation de cette série par un modèle quelconque. On peut alors évaluer la qualité de ce modèle en utilisant différents scores. Dans le cas de l'étude du GHI, ces scores sont calculés pour des angles solaires zénithaux supérieurs à 80° afin de ne pas prendre en compte les mesures de nuit et les angles

rasants. Les métriques les plus courantes s'expriment ainsi :

$$\text{MBE (Mean Biased Error)} = \frac{1}{N} \sum_{k=1}^N \hat{x}(k) - x(k) \quad (1.8)$$

$$\text{RMSE (Root Mean Square Error)} = \sqrt{\frac{1}{N} \sum_{k=1}^N (\hat{x}(k) - x(k))^2} \quad (1.9)$$

$$R^2(\text{Coefficient of determination}) = 1 - \frac{\text{Var}(\hat{X}_k - X_k)}{\text{Var}(X_k)} \quad (1.10)$$

Ces scores mettent chacun l'accent sur des caractéristiques particulières du modèle de prévision ou d'estimation. Il est important de bien comprendre le sens de chacune de ces informations pour pouvoir ensuite les utiliser dans l'évaluation comparative de modèles. Les figures 1.22.a à 1.22.c présentent des exemples d'estimations permettant de mettre en évidence l'intérêt de chacun de ces scores. Ces trois figures illustrent trois estimations différentes d'une même journée de mesure de rayonnement global. Le premier cas correspond à un simple décalage d'amplitude de la courbe de mesure, le second à l'ajout d'un signal de bruit blanc centré, et le troisième au modèle de persistance brute (sans index Kc). Les scores obtenus dans chacun des cas sont donnés dans le tableau 1.7. Ces cas très simples permettent de mettre en évidence certaines propriétés des métriques présentées :

- La MBE donne une indication sur un décalage systématique du modèle. Elle permet de détecter un biais. Elle ne donne aucune indication sur les autres types d'erreurs puisqu'elle est nulle pour les cas (b) et (c).
- La RMSE donne une indication plus générale de l'erreur. En revanche son interprétation doit être faite avec beaucoup de précaution. En effet, on notera qu'elle englobe tous les types d'erreur. D'après les résultats du tableau 1.7, elle est impactée par le biais (cas (a)), par le bruit (cas (b)) et aussi par les décalage temporels (cas (c)), c'est pourquoi elle est souvent utilisée comme référence pour caractériser un modèle. Il est toutefois important de remarquer ici que le meilleur modèle d'après la RMSE est le modèle (b). Or on sait que le modèle (a) contient lui la courbe exacte à une addition près, ce qui est beaucoup plus intéressant et facile à extraire. L'interprétation de la RMSE doit donc toujours se faire en rapport aux autres métriques pour discerner les différents cas d'erreurs possibles.
- Le coefficient de détermination permet d'avoir une indication sur la variabilité expliquée par le modèle. On dit d'un modèle qu'il caractérise parfaitement un phénomène lorsque ce coefficient tends vers 1. Les modèles (a) et (b) bien qu'erronnés subissent des variations qui sont cohérentes avec les variations des mesures observées, en revanche le modèle (c) du fait du décalage temporel important, subit des variations qui ne sont plus en lien avec les fluctuations des mesures, d'où un coefficient de détermination moindre.

TABLE 1.7 – Résultats pour chacune des estimations

	MBE	RMSE	R^2
Mesure - 20	20	20	1
Mesure + bruit centré ($\sigma = 0.05$)	0	5	0.99
Persistance	0	43	0.88

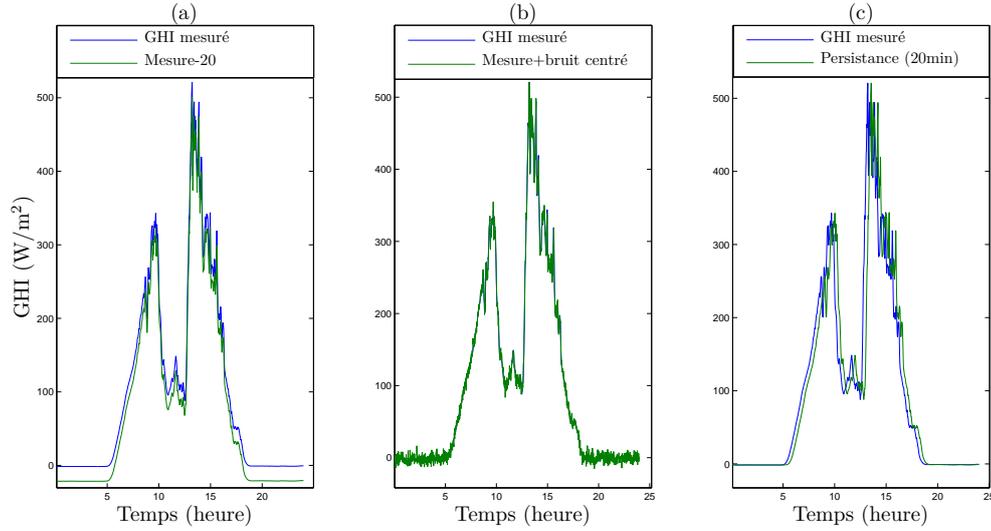


FIGURE 1.22 – Exemple d'erreur d'estimation d'un signal de rayonnement

A partir de ces métriques absolues, on peut définir des métriques relatives à partir d'une valeur de référence. Très souvent la moyenne du rayonnement mesuré sur l'ensemble de mesure considéré est choisi comme référence. On peut ainsi définir la RMSE relative (RMSE_r) ou la MBE relative (MBE_r) en divisant ces deux scores par la valeur de référence. On obtient alors des valeurs un peu plus indépendantes du site étudié. Toutefois, pour comparer des méthodes ayant été évaluées sur des jeux de données différents, il est intéressant de se référer à des scores prenant en compte les scores de persistance. Comme on l'a souligné précédemment, les scores de persistance portent en eux une information sur la difficulté de la prévision en fonction du site. Une étude complète sur ces scores est décrite dans [MC13b]. Nous retiendrons particulièrement le score $RMSE_{perf}$ décrit par l'équation 1.11, qui est le rapport entre la RMSE et la RMSE du modèle de persistance $RMSE_p$. Un score de 1 indique un prévision parfaite, 0 une prévision de performance égale à la persistance, et enfin, un score négatif indique un score inférieur à celui de la persistance.

$$RMSE_{perf} = 1 - \frac{RMSE}{RMSE_p} \quad (1.11)$$

1.4 Caractérisation du site d'étude

1.4.1 Variabilité

Au delà de la qualité des mesures, il est intéressant de savoir caractériser le site de mesure de façon simple. Il s'agit de savoir s'il existe une météorologie prédominante ou non, et si oui laquelle. Par exemple, une étude de prévision sur un site bénéficiant d'une météorologie très souvent dégagée aboutira à de meilleures performances. On propose pour cela de faire une classification assez simple des types de jours observés, basée sur les mesures journalières

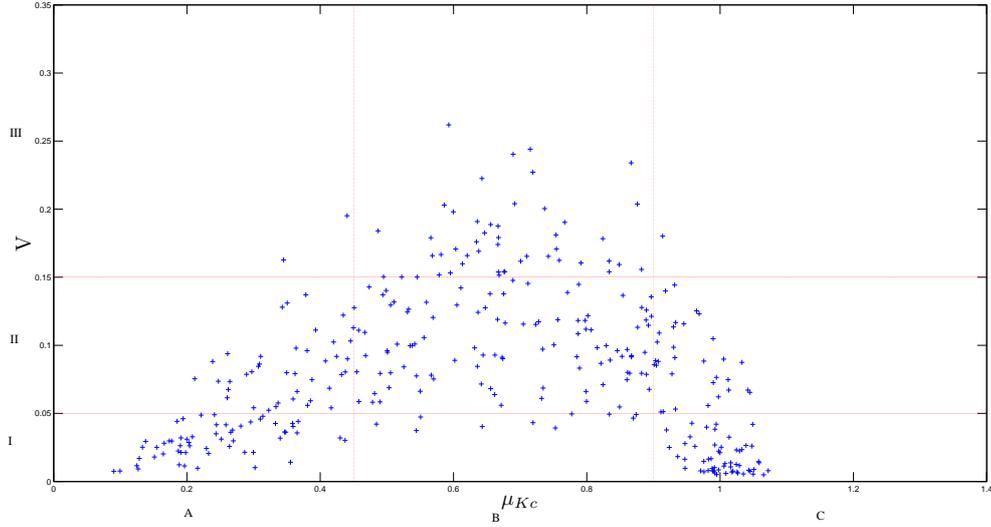


FIGURE 1.23 – Répartition des journées de 2011 dans le plan (μ_{Kc}, V)

moyennées sur 15 minutes. Afin de garder une classification simple, 2 critères ont été retenus :

- La moyenne de l'index Kc sur la journée μ_{Kc} décrite par l'équation 1.12 où N_{jour} indique le nombre de points de mesures hormis la nuit ($\theta_z > 90^\circ$).
- Une métrique V [MC13b] adaptée à la description de la variabilité "de proche en proche" permettant de mieux décrire la difficulté de la prévision (Eq. 1.13)

$$\mu_{Kc} = \frac{1}{N_{jour}} \times \sum_{k=1}^{N_{jour}} Kc(k) \quad (1.12)$$

$$V = \sqrt{\frac{1}{N_{jour}} \times \sum_{k=1}^{N_{jour}} (\delta_{Kc}(k))^2} \quad (1.13)$$

$$\delta_{Kc}(k) = Kc(k+1) - Kc(k)$$

En utilisant ces deux critères on obtient une répartition des journées présentée en figure 1.23. Il existe différentes méthodes pour la classification d'un tel nuage de points, cependant il s'agit ici simplement d'extraire les tendances générales, c'est pourquoi une simple classification arbitraire a été effectuée comme indiqué en figure 1.23. Cette classification s'appuie sur des seuils arbitraires définis sur chacune des métriques afin représenter au mieux la répartition des points observés. Elle divise l'espace de décision en 9 classes décrivant à la fois le niveau moyen de la journée et sa variabilité. Le niveau moyen est classé de A à C ou A correspond à une moyenne de Kc faible et C une moyenne forte. La variabilité est décrite de la même façon par les niveau I à III. On obtient donc 9 classes de AI à CIII couvrant l'ensemble des données. La figure 1.24 présente un exemple de journée de chaque classe. Les résultats de classification pour les deux années étudiées sont répertoriés dans le tableau 1.8. D'après ce tableau, le site du SIRTa est relativement difficile dans la mesure où plus de la moitié des jours sont considérés comme variables. On note également plus de journées couvertes (classe A) que de journées ciel clair (classe C), plus faciles à prévoir. Cette classification sera réutilisée plus tard dans le développement des jeux de données (section 2.2.6).

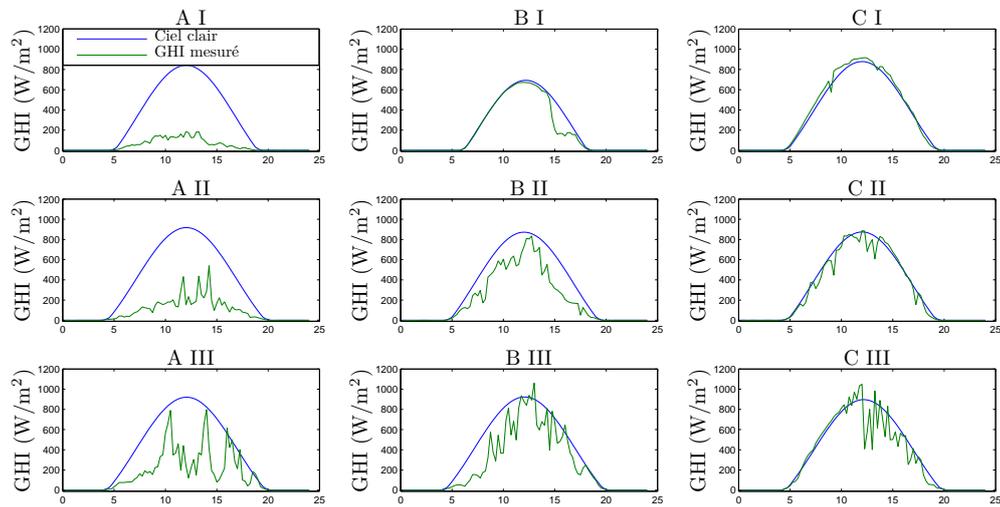


FIGURE 1.24 – Exemple de profil pour chaque classe, le nom des classes est explicité dans le tableau 1.8

TABLE 1.8 – Résultats de classification des journées 2011 et 2012 en pourcentage (total de journées = 699)

	A : couvert 31%	B : mitigé 46%	C : clair 23%
I : peu variable 33%	16	3	14
II : variable 52%	14	29	9
III : très variable 15%	1	14	1

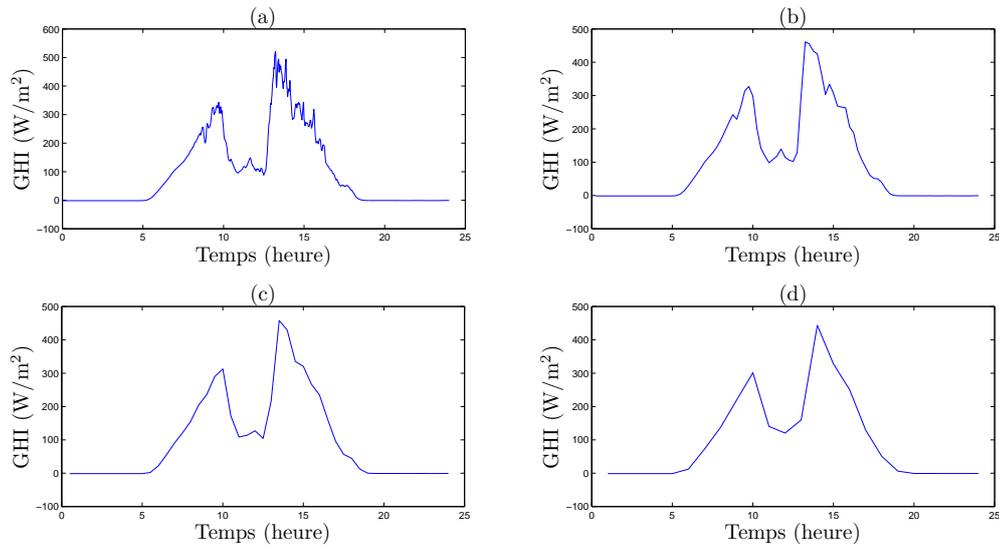


FIGURE 1.25 – Impact de la résolution temporelle. (a) échantillonnage : 1 minute, (b) moyenne sur 15 minutes, (c) moyenne sur 30 minutes, (d) moyenne sur 60 minutes

1.4.2 Résolution temporelle

Les capteurs présentés précédemment (section 1.2.2) fournissent des séries temporelles de données avec une période d'échantillonnage d'une minute. Ce paragraphe étudie les conséquences du choix de la résolution temporelle. La résolution temporelle de la série doit parfois être ajuster pour la "synchroniser" avec d'autres capteurs (caméra hémisphérique, satellite,...). Ce changement de résolution par moyenne a d'abord un impact sur le type de variations étudiées. La figure 1.25 illustre la même journée avec différentes résolutions temporelles. Avec l'augmentation de la période apparaît un effet de lissage classique lié au filtrage des hautes fréquences par l'opération de moyennage. Il est important de noter que les phénomènes lissés ici correspondent à la réponse du capteur à des phénomènes observables (petits nuages, etc.) et non uniquement à un bruit quelconque. Aussi faut-il choisir la résolution temporelle en fonction de la sensibilité du capteur que l'on souhaite utiliser pour la prévision et du besoin. Par exemple, une caméra au sol filmant le ciel sera particulièrement sensible aux petits nuages passant sur le site, il est donc important de garder dans la série temporelle de mesure l'impact de ces nuages. Dans le cas contraire, les variations observées dans l'image auront de la difficulté, voire de l'impossibilité, à observer la correspondance dans la série de mesure.

1.5 Conclusion

Tout d'abord, ce chapitre a permis de présenter le GHI sur le plan théorique et pratique. Les mesures présentées seront utilisées dans les chapitres suivants pour évaluer les modèles construits pour l'estimation et la prévision du GHI. La caractérisation de la couche nuageuse

présente sur le site a un impact très important sur les variations du GHI par rapport au modèle "ciel clair". Dans les chapitres suivants, on tentera d'extraire le maximum d'information sur la couverture nuageuse à partir d'images prises depuis le sol, d'images satellitaires, puis en les fusionnant.

Estimation par utilisation d'imageurs hémisphériques du ciel depuis le sol

Sommaire

2.1	Etat de l'art des imageurs au sol	29
2.1.1	Capteurs	29
2.1.2	Traitements	31
2.1.3	Conclusion et positionnement des travaux	37
2.2	Estimation du GHI à partir d'images au sol	38
2.2.1	Présentation de la méthode générale	38
2.2.2	Pré-traitement des images	38
2.2.3	Liste des caractéristiques disponibles	39
2.2.4	Analyse des entrées	49
2.2.5	Réseau de neurones artificiels	51
2.2.6	Construction des bases d'apprentissage/test	51
2.2.7	Résultats	54
2.3	Conclusion	58

Ce chapitre traite de l'utilisation d'images du ciel prises depuis le sol (qu'on appellera images hémisphériques) pour l'estimation et la prévision du GHI. Dans un premier temps, on présentera les capteurs habituellement rencontrés dans la littérature et utilisés pour cette étude. Une présentation synthétique de l'état de l'art sur l'utilisation de ces outils permettra ensuite d'en dégager les principaux avantages et inconvénients. Enfin on développera la méthode choisie ici pour tirer parti de ces images, dans un premier temps en estimation du GHI, puis dans un objectif de prévision à très court terme de cette variable.

2.1 Etat de l'art des imageurs au sol

2.1.1 Capteurs

L'étude de la couverture nuageuse à partir d'images au sol peut se faire à l'aide d'un quelconque appareil photo du commerce. Toutefois, il existe des systèmes développés par des



FIGURE 2.1 – TSI 880

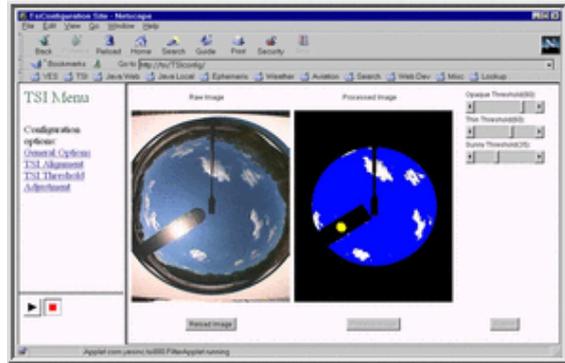


FIGURE 2.2 – Image brute et image traitée

industriels ou des universitaires qui proposent des caractéristiques particulièrement avantageuses pour le suivi régulier de la voûte de jour. Ces caractéristiques sont les suivantes :

- Champs de vision permettant l’acquisition d’images hémisphériques du ciel
- Communication des données
- Résistance aux conditions climatiques extérieures (pluie, vent, etc)

On retiendra parmi les instruments décrits dans la bibliographie les deux suivants : le Whole Sky imager (WSI) et le Total Sky Imager (TSI). Ces deux instruments ont été analysés et comparés dans [SLT01]. Le paragraphe suivant s’intéresse plus précisément au TSI : ce capteur est installé sur le site du SIRTA proche du point de mesure pyranométrique. Le TSI reste la référence industrielle même si d’autres dispositifs à moindre coût commencent à émerger [Gau+12].

2.1.1.1 Le Total Sky Imager

Le TSI est le dispositif le plus représenté dans la littérature [Cho+11], [Pfi+03], [SLT01], [MC13a]. Développé par la société Yankee Environmental Systems (YES) dans le Massachusetts, le TSI 880 (Fig. 2.1) est le dernier modèle de la gamme depuis 2006. Ses caractéristiques sont fournies dans le tableau 2.1. Le TSI est constitué d’un capteur CCD (Charged Coupled Device) observant un miroir hémisphérique en rotation synchronisée avec l’angle azimutal solaire. Une bande noire sur le miroir atténue très fortement le rayonnement normal direct pour protéger le capteur. Le TSI dispose d’une fonction intégrée permettant la segmentation des images, le calcul de fraction nuageuse et l’affichage des images traitées ou non (Fig. 2.2). Plusieurs paramètres de ces fonctions sont laissés à la main de l’utilisateur comme l’angle zénithal minimum pour l’enregistrement ou les paramètres de calibration de la segmentation de couverture nuageuse. De façon générale, les résultats de segmentation du TSI ne sont performants que pour des angles zénithaux inférieurs à 80° [SLT01] ; les images ne sont donc traitées que pour un champ de vision de 160° . Le TSI a trois principaux défauts :

- Le miroir est sensible à l’humidité et peut être endommagé dans certaines conditions climatiques.
- Du fait de la bande noire protectrice et du placement de la caméra au dessus du miroir, 8% du champ de vision inutilisable [LSC+06].
- C’est le dispositif le plus cher.

TABLE 2.1 – Caractéristiques techniques du TSI

Image resolution	640 x 480 color, 24-bit JPEG format
Image typical size	30 KB
Sampling rate	Variable, with max of 30 sec
Operating temperature	-40 ° C to +44 ° C
Software	None required for immediate real time display ; uses Internet Explorer or Netscape Browsers on MS-Windows, Mac, UNIX
Data telemetry	LAN Ethernet (TCP/IP), telephone modem (PPP) or data storage module option (for off grid site)

2.1.2 Traitements

2.1.2.1 Segmentation et classification de la couverture nuageuse

L'une des premières motivations pour l'utilisation d'images hémisphériques a été la mesure automatique de fraction nuageuse. Plusieurs méthodes de segmentation puis de classification sont alors apparues. L'idée principale pour la segmentation de telles images repose sur le ratio des canaux rouge et bleu de l'image. En effet, alors que le ciel clair diffuse plus de bleu que de rouge, les nuages ont eux un coefficient de transmission assez similaire pour ces deux bandes spectrales [Pet06]. En utilisant ce ratio Rouge/Bleu, on peut alors séparer le ciel clair des nuages à l'aide d'un seuil dont la valeur est proche de 0.6 selon [LSC+06]. Un seuil sur la valeur du Rouge - Bleu a également été proposé par [HMS10] en 2010 dans une étude utilisant un TSI. D'autres travaux ont montré plus tard de meilleures performances en faisant varier la valeur du seuil en fonction de certaines zone de l'image (circumsolaire, zénithale, horizon, etc) [Gau+12] [MC13a], et parfois en utilisant des seuils plus complexes sur plusieurs canaux [Kaz+12]. La figure 2.4 présente un exemple classique de segmentation utilisant le ratio de Rouge/Bleu.

Les images hémisphériques permettent également d'effectuer une classification des types de nuages observés. Heinle *et al.* [HMS10] proposent une classification basée sur les classes définies par l'organisation mondiale de la météorologie. Quelques modifications ont cependant été apportées : à partir des 10 classes d'origines (section 1.4.1), 7 ont été retenues en regroupant certaines classes considérées trop proches comme indiqué dans le tableau 2.2. Ces travaux ont ensuite été repris par Kazantzidis *et al.* en 2012 avec des critères de segmentation différents, puis par Liu *et al.* dans une approche impliquant l'étude de séquences d'images sous forme de tenseur. Les résultats obtenus pour chacune de ces méthodes sont rapportés dans le tableau 2.3. Ces résultats ont été établis sur des jeux de données différents, leur comparaison directe reste donc difficile. Toutefois, ils semblent montrer que l'intégration d'information temporelle permet une amélioration importante des scores de reconnaissance de classes. Ces études, bien qu'indirectement liées à la problématique principale de la thèse, représentent une très importante source d'inspiration pour l'utilisation d'images au sol dans un objectif de caractérisation de la couverture nuageuse. Des approches plus directes vers la prévision du GHI sont présentées dans la section suivante.



FIGURE 2.3 – Image brute

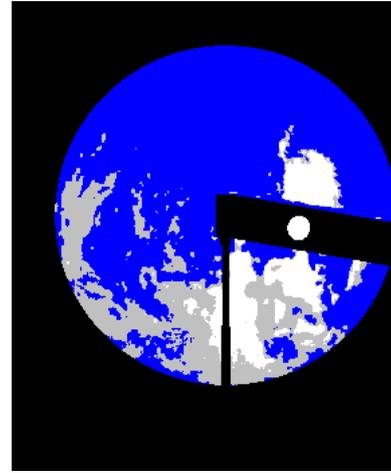


FIGURE 2.4 – Image seuillée : bleu = ciel clair ; blanc = nuages clairs ; gris = nuages opaques

TABLE 2.2 – Classes utilisées pour classification par Heinle *et al.* [HMS10] (cf chap1)

Numéro	Classe
1.	Cumulus
2.	Cirrus et Cirrostratus
3.	Cirrocumulus et Altopcumulus
4.	Ciel clair
5.	Stratocumulus
6.	Status et Altostratus
7.	Cumulonimbus et Nimbostratus

TABLE 2.3 – Comparaison des résultats de classification : Les éléments de la ligne i et colonne j représentent le pourcentage de la classe i identifié comme classe j pour chacune des méthodes : H Heinle *et al.* [HMS10], K Kazantzidis *et al.* [Kaz+12], L Liu *et al.* [Liu+13]. (Cf tableau 2.2)

Classes	1			2			3			4			5			6			7		
	H	K	L	H	K	L	H	K	L	H	K	L	H	K	L	H	K	L	H	K	L
1.	67	92	96	16	5	0	15	3	4	0	0	0	0	0	0	0	1	0	2	0	0
2.	6	1.8	0	79	95	92	0	3	8	15	0	0	0	0	0	0	0	0	0	0	0
3.	11	9	5	3	8	0	81	78	95	0	0	0	2	5	0	3	0	0	0	0	0
4.	6	0	0	18	5	0	0	0	0	76	95	100	0	0	0	0	0	0	0	0	0
5.	0	0	0	0	0	0	33	0	0	0	0	0	41	93	90	26	2	10	0	5	0
6.	0	0	0	0	0	0	10	0	0	0	0	0	0	7	9	90	93	86	0	0	5
7.	0	0	0	0	0	0	14	0	0	0	0	0	0	17	9	0	0	5	86	83	86

2.1.2.2 Vers la prévision du GHI

Depuis l'apparition des imageurs dans la littérature, de nombreuses études se sont penchées sur leur utilité quant à la prévision du rayonnement perçu au sol, dans le cadre général de la prévision de production d'énergie photovoltaïque. Ces travaux sont en lien étroit avec ceux présentés sur la classification de la couverture nuageuse et en sont souvent très inspirés. En effet, les différentes classes de nuages ont potentiellement un effet différent sur le rayonnement perçu au sol. Ainsi, les paramètres permettant d'identifier ces classes sont généralement pertinents pour l'estimation du GHI. Il n'y a alors plus qu'un pas vers la prévision, comblé soit par une étape de traitement d'image (calcul de champ de vecteurs, cf chapitre 3), soit par l'utilisation d'algorithmes mathématiques (régression). Ces travaux constituent l'état de l'art le plus proche du modèle qui sera présenté par la suite. Il est donc nécessaire d'approfondir davantage leur démarche respective. Les travaux de Chow *et al.* [Cho+11] ont posé les premières pierres pour la prévision du GHI. Ils proposent d'utiliser une méthode de traitement d'image (Cross-correlation Method) pour calculer un champ de vecteurs représentatifs du mouvement observé dans l'image. En utilisant ces vecteurs, il s'agit alors d'anticiper la future image par translation de la couverture nuageuse, puis d'appliquer les méthodes de segmentation pour identifier les nuages dans cette image projetée. La dernière étape consiste à évaluer l'impact des nuages sur le GHI. C'est sur ce dernier point que la méthode a ses points faibles, ne proposant que deux niveaux de prévision, avec ou sans nuages. On notera également de nombreuses restrictions sur les conditions d'application de l'algorithme de calcul du champ de vecteurs. C'est une stratégie similaire qui est proposée par Marquez *et al.* [MC13a] mais pour la prévision du rayonnement direct. La méthode est en effet plus adaptée à la prévision du DNI, le champ de vecteurs permettant de prévoir notamment les probabilités d'obstruction du soleil dans l'image. Les résultats obtenus sont alors très satisfaisants et pointent déjà les limites de l'image au sol en terme d'horizon de prévision à une vingtaine de minutes. Le lien avec le GHI reste lui plus compliqué à établir. Fu *et al.* [FC13] proposent d'établir ce lien au moyen d'un apprentissage sur des caractéristiques de l'image. La prévision ne requiert alors plus de champs de vecteurs mais une base d'apprentissage et des critères adaptés. Quatorze caractéristiques sont extraites pour chaque image, mais seules les plus corrélées sont utilisées pour l'apprentissage du modèle de prévision. On obtient un modèle différent pour chaque horizon de prévision variant entre 5 et 15 minutes. Les résultats démontrent l'intérêt des modèles à apprentissage pour la conversion d'information entre image et GHI. Ces trois approches différentes sont résumées dans les tableaux 2.4 à 2.6.

TABLE 2.4 – Récapitulatif des travaux de Chow *et al.* [Cho+11]

Instrument	TSI-440 (Yankee Environmental Systems)
Période d’échantillonnage	30s
Localisation	University of California San Diego
Variable d’étude	GHI
Méthodologie	<ul style="list-style-type: none"> – Segmentation : librairie de "Ciel clair" + "sunshine parameter". – Projection des ombres des nuages à partir d’une altitude de la base des nuages donnée par un ceilomètre. – Mouvement des nuages : Cross-correlation Method sur le canal rouge. – Prévision : advection de l’image en suivant le vecteur déterminé.
Restriction/Hypothèse	<ul style="list-style-type: none"> – Image acquise pour : angle d’élévation solaire $>3^\circ$ – Zone traitée : angle zénithal $<65^\circ$ – Hypothèse : altitude de la base des nuages constante. – Atténuation due à un nuage : 60% du ciel clair, choix arbitraire. – Situations écartées : jours complètement clairs ou couverts, couches multiples, déformation rapide, formation ou évaporation de nuages...
Résultat	<ul style="list-style-type: none"> – Estimation : 70% de bonne reconnaissance (ciel clair vs ciel couvert) – Prévision : Pas de scores concernant directement le GHI – Évaluation de la reconstruction de l’image après translation par divers scores pour des horizons variant de 30 secondes à 5 minutes (résultats globalement satisfaisants) – Pas de prévision de GHI.

TABLE 2.5 – Récapitulatif des travaux de Marquez *et al.* [MC13a]

Instrument	TSI-880 (Yankee Environmental Systems)
Période d'échantillonnage	1 min
Localisation	University of California
Variable d'étude	DNI
Méthodologie	<ul style="list-style-type: none"> – Masque des images (bande noire + environnement) – Transformation en grille rectangulaire (image originellement sphérique) – Calcul du champ de vecteurs : Particule Image Velocimetry (PIV) [MC04] – Classification K-means du champs de vecteur en 2 classes, choix de la classe de moyenne la plus élevée comme vecteur de l'image – Segmentation : Seuillage hybride [LLY11] – Découpage de zones d'intérêt dans la direction opposée au vecteur – Prévission : $P = DNI_{max} \times (1 - \text{Fraction nuageuse}(\text{zone d'intérêt}))$
Restrictions/Hypothèses	RAS
Résultats	<ul style="list-style-type: none"> – Prévission par TSI estimée possible jusqu'à 12-18 minutes – $RMSE_{perf}$ entre 20% et 40%, évolution des résultats cohérente avec l'augmentation de l'horizon de prévission

TABLE 2.6 – Récapitulatif des travaux de Fu *et al.* [FC13]

Instrument	All sky camera (Santa barbara Instrument Group : www.sbig.com)
Période d’échantillonnage	10s
Localisation	Industrial Technology Research Institute, No. 195, Sec. 4, Chung Hsing Rd., Chutung, Hsinchu 31040, Taiwan
Variable d’étude	GHI
Méthodologie	<ul style="list-style-type: none"> – Extraction de 14 caractéristiques de l’image – Segmentation : Ratio R/B classique comme décrit par [LSC+06] – Construction d’un modèle de régression sur une base d’apprentissage (polynôme de régression) – Utilisation d’une variante de l’index Kc à la place de l’apprentissage direct du GHI – Prévision via régression
Restriction/Hypothèses	<ul style="list-style-type: none"> – Utilisation d’artefacts propre à la caméra
Résultats	<ul style="list-style-type: none"> – $RMSE_r$ à 15 minutes : 27,5% – MAE_r à 15 minutes : 24%

2.1.3 Conclusion et positionnement des travaux

La bibliographie a démontré l'intérêt de l'image au sol pour la caractérisation de la couverture nuageuse dans un premier temps (classification) et donc pour l'estimation et la prévision du rayonnement perçu au sol. Ces images apportent une information très résolue qui permet de différencier des phénomènes fins et d'accéder à des prévisions de déplacement à très petite échelle en contrepartie d'une limitation de l'horizon de prévision à une vingtaine de minutes dépendant de l'altitude et de la vitesse des nuages. En revanche la conversion d'information entre l'image et le rayonnement reste un enjeu capital. Si le DNI semble plus facile à expliquer grâce aux informations sur la région circumsolaire, l'obtention du GHI semble plus complexe. Les premières études abordant le sujet ont montré des résultats intéressants en utilisant un modèle de régression. Le choix des critères d'entrée et la construction des bases de données deviennent alors des enjeux capitaux du modèle. C'est dans cet axe d'étude que s'inscrit le modèle présenté dans ce chapitre. En s'inspirant des travaux réalisés en classification, on étudiera la pertinence de diverses caractéristiques de l'image pour l'estimation du GHI via un modèle de régression non linéaire : le réseau de neurones artificiel.

2.2 Estimation du GHI à partir d’images au sol

Cette section présente des travaux originaux développés au cours de la thèse. Comme indiqué précédemment, l’étude s’inscrit dans la tendance donnée par l’état de l’art sur l’estimation de GHI par image au sol développée antérieurement aux travaux de Marquez *et al.* [MC13a] et de Fu *et al.* [FC13]. Les résultats de ces travaux ont fait l’objet d’une publication soumise au journal "Sustainable Energy".

2.2.1 Présentation de la méthode générale

Nous disposons d’informations de l’image hémisphérique que nous souhaitons utiliser pour estimer le rayonnement. Il s’agit alors de déterminer quelles caractéristiques de l’image sont réellement pertinentes, et ensuite, comment sont-elles liées à la variable d’intérêt. Si la première partie relève du traitement d’images pour l’extraction des caractéristiques, la seconde peut être identifiée comme une étape de régression. Il existe de multiples méthodes pour résoudre les problèmes de régression ([SL12], [Hue+96]), le choix s’est porté ici sur l’utilisation d’un réseau de neurones artificiels (ANN) (cf 2.2.5). La figure 2.5 donne une vue d’ensemble de la méthode. Les parties 2.2.3 à 2.2.4 décrivent l’extraction des entrées du réseau de neurone à partir des images brutes. Les réseaux de neurones sont ensuite présentés en section 2.2.5 suivis par la méthode de création des différentes bases de données nécessaires (section 2.2.6). Enfin, les résultats sont détaillés en section 2.2.7.

2.2.2 Pré-traitement des images

Les images utilisées proviennent d’un TSI-880 installé sur le site du SIRTÀ à moins de 10 mètres du capteur de rayonnement de référence. Les images brutes sont acquises toutes les minutes. Ce sont des images classiques à trois canaux RVB (Rouge/Vert/Bleu) avec une résolution de 480×640 pixels. De la même manière que pour les données de rayonnement, la disponibilité des images sur l’année d’étude a été testée. Les résultats sont visibles en figure 2.6. Globalement, sur les plages horaires correspondant à un angle solaire zénithal inférieur à 75° les images sont disponibles à plus de 99% (2151 images manquantes sur 176609 concernées). Ces images ne sont pas exploitables directement pour l’étude de la couverture nuageuse en raison de différents éléments gênant l’interprétation des images (Fig. 2.7) :

- la bande noire ;
- le bras de la caméra ;
- l’environnement du dispositif.

Ces différents éléments ont donc été masqués afin de pouvoir extraire des informations ne concernant que la couverture nuageuse. Dans le cas des deux derniers éléments, un masque fixe réalisé manuellement suffit. En revanche, la bande noire provient d’une bande de scotch apposée sur le miroir rotatif, sa position dans l’image dépend donc de l’angle solaire azimutal. Pour masquer cette bande, un masque a été créé sur une position fixe, puis mis en rotation en fonction de l’angle solaire. Le résultat du masque total est présenté par la figure 2.8. Aucune

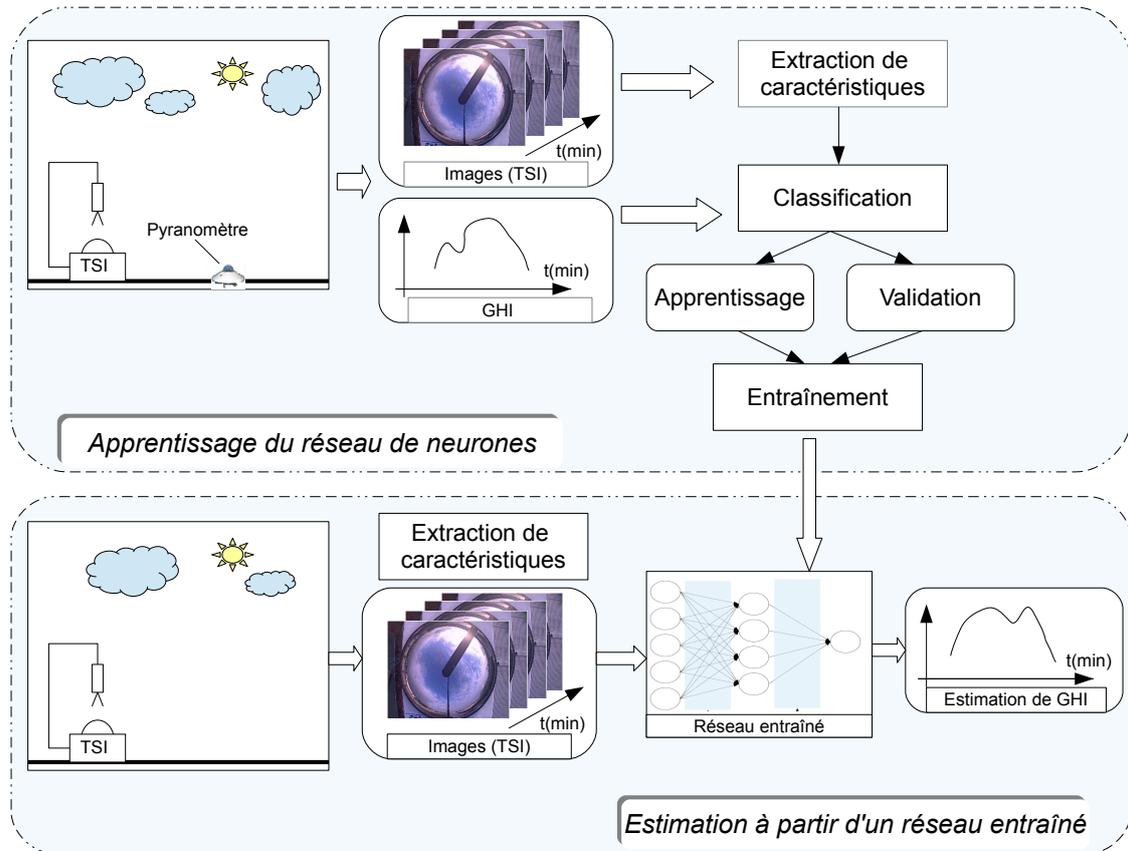


FIGURE 2.5 – Vue d'ensemble de la méthode d'estimation

autre modification n'est effectuée sur l'image.

2.2.3 Liste des caractéristiques disponibles

Il existe de multiples façons d'extraire de l'information d'une image. La liste de caractéristiques présentée ci-dessous est inspirée des travaux cités précédemment en classification de couverture nuageuse ([HMS10] [CS08]) et en estimation de GHI par image au sol [FC13]. La plupart des statistiques utilisées sont applicables pour un seul canal à la fois. Dans le but d'être le plus exhaustif possible, elles ont été calculées pour chacun des 3 canaux (RVB) de l'image, ainsi que pour un canal unique S obtenu en sommant ces 3 canaux. Ainsi l'apport de chaque canal pourra-t-il être étudié *a posteriori*. L'objectif sera alors de ne retenir parmi toutes ces possibilités que celles qui apportent une information pertinente pour l'estimation du rayonnement. Pour l'ensemble des définitions, la lettre C représente un des canaux et renvoie donc à l'ensemble $\{R,V,B,S\}$. Par ailleurs, seuls les pixels non masqués sont pris en compte dans tous les cas, ainsi la valeur N_{pix} représente le nombre total de pixels non masqués dans l'image.

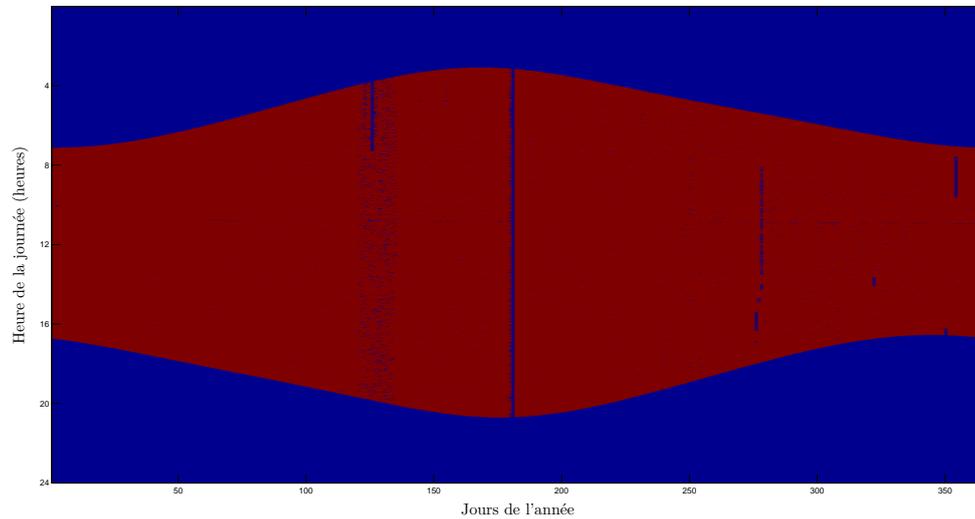


FIGURE 2.6 – Disponibilité des images du TSI sur l'année 2011, les points rouges représentent les images disponibles.

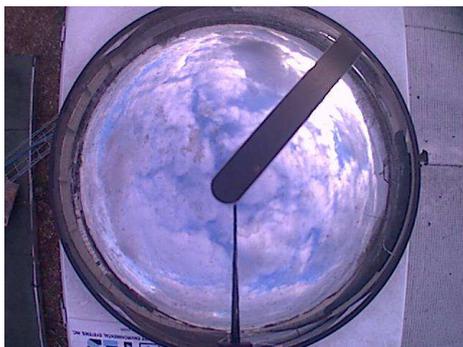


FIGURE 2.7 – Image brute



FIGURE 2.8 – Image masquée

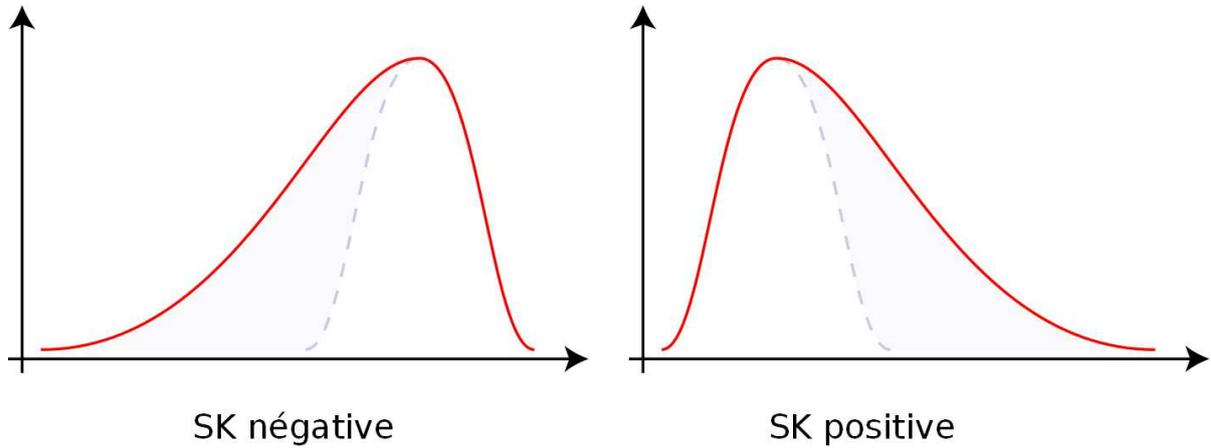


FIGURE 2.9 – Information apportée par la dissymétrie sur une distribution donnée

2.2.3.1 Caractéristiques classiques

La moyenne M donne la valeur moyenne des pixels hors masque de ces images. La comparaison de la moyenne de chaque canal apporte une information intéressante sur la situation globale.

$$MC = \frac{1}{N_{pix}} \sum_{i=1}^{N_{pix}} C(i) \quad (2.1)$$

L'écart type (STD) caractérise la distribution du niveau des pixels dans l'image autour de la moyenne. Il permet par exemple de différencier un ciel complètement homogène d'un ciel avec beaucoup de perturbation. Il est calculé pour chaque canal selon l'équation :

$$STDC = \sqrt{\frac{1}{N_{pix}-1} \sum_{i=1}^{N_{pix}} (C(i) - MC)^2} \quad (2.2)$$

La dissymétrie (SK) (Skewness(SK) en anglais) est un paramètre de forme de la distribution du niveau des pixels (Fig.2.9). Elle est calculée pour chaque canal selon l'équation :

$$SKC = \frac{1}{N_{pix}-1} \sum_{i=1}^{N_{pix}} \left(\frac{C(i)-MC}{STDC} \right)^3 \quad (2.3)$$

2.2.3.2 Caractéristiques de texture

Nous nous intéressons ensuite des caractéristiques de textures de l'image. Comme indiqué par Heinle *et al.* [HMS10], nous utilisons pour cela la matrice de co-occurrence des niveau de gris

(MCOO) [HSD73]. Cette matrice permet de caractériser la périodicité ainsi que la directivité des textures. Elle s'appuie sur la relation liant deux pixels, un pixel de "référence" et un pixel dit "voisin". On peut par exemple choisir pour pixel voisin celui situé à l'est de la référence, relation qui peut alors s'écrire (1,0) ce qui signifie 1 pixel dans la direction x horizontale et 0 pixel dans la direction y verticale. Le principe consiste à compter le nombre d'occurrence de chaque combinaison possible de couple {référence,voisin} pour une direction donnée dans une matrice. On s'appuiera pour l'explication de la MCOO sur l'image exemple Fig. 2.11 ¹. L'image exemple ayant des valeurs comprises entre 0 et 3, l'ensemble des combinaisons possibles (dans la direction (1,0)) peut être représenté dans une matrice de taille 4 × 4 comme illustré par la figure 2.12. On obtient ainsi pour l'exemple (Fig. 2.11) et dans la direction (1,0) la MCOO présentée en figure 2.13. On peut y lire que la valeur 2 a 3 fois la valeur 2 comme voisin de droite. Les caractéristiques de textures voulues nécessitent l'utilisation d'une matrice symétrique. On additionne donc les matrices de direction (1,0) et (-1,0) afin d'obtenir une matrice symétrique caractérisant les variations horizontales. Il est également plus commode de travailler avec des probabilité d'apparition d'un couple plus qu'avec un décompte, c'est pourquoi on normalise les valeurs obtenues par le nombre maximal d'occurrences possibles (lié à la taille de l'image). On obtient alors une valeur $P_{i,j}$ donnant la probabilité d'apparition du couple pointé par la ligne i et la colonne j. De la même manière on obtient une MCOO verticale que l'on peut combiner avec la matrice horizontale en faisant une moyenne pour obtenir une MCOO unique pour l'image. C'est à partir de cette matrice que sont calculées les différentes mesures de texture. Toutefois, pour pallier le problème de la bande noire et du masque, ces mesures n'ont pas été calculées directement pour toute l'image mais pour plusieurs sous-images de taille 20 × 20 ne comportant aucun pixel masqué. La moyenne de ces mesures est ensuite retenue comme valeur pour l'image entière. Les équations suivantes décrivent les caractéristiques extraites de la MCOO :

Contraste (CON)

$$CON = \sum_{i,j=0}^{N-1} P_{i,j}(i-j)^2 \quad (2.4)$$

Dissemblance (DIS)

$$DIS = \sum_{i,j=0}^{N-1} P_{i,j}|i-j| \quad (2.5)$$

Homogénéité (HOM)

$$HOM = \sum_{i,j=0}^{N-1} \frac{P_{i,j}}{1+(i-j)^2} \quad (2.6)$$

1. http://www.fp.ucalgary.ca/mhallbey/what_is_texture.htm

FIGURE 2.10 – Valeurs des pixels

0	0	1	1
0	0	1	1
0	2	2	2
2	2	3	3

FIGURE 2.11 – Image correspondante



FIGURE 2.12 – Combinaisons possibles Référence/voisin

Voisin \ Référence	0	1	2	3
0	0,0	0,1	0,2	0,3
1	1,0	1,1	1,2	1,3
2	2,0	2,1	2,2	2,3
3	3,0	3,1	3,2	3,3

FIGURE 2.13 – MCOO pour l'image exemple Fig. 2.11

4	2	1	0
2	4	0	0
1	0	6	1
0	0	1	2

TABLE 2.7 – MCOO normalisée pour l'exemple :

0.16	0.08	0.04	0
0.08	0.16	0	0
0.04	0	0.25	0.04
0	0	0.04	0.08

2.2.3.3 Caractéristique de la couverture nuageuse

Les statistiques citées précédemment sont des critères classiques utilisés dans de nombreuses applications en traitement d’image. D’autres caractéristiques plus spécialisées peuvent également être extraites. Il s’agit des caractéristiques basées sur la segmentation de la couverture nuageuse évoquée dans l’état de l’art. A partir de l’image segmentée, il est possible d’extraire au moins deux critères simples, le ratio de ciel couvert et le nombre de nuages détectés. L’état de l’art propose de multiples pistes pour la segmentation de la couverture nuageuse. Les critères souhaités étant assez généraux, nous nous attacherons à choisir une méthode simple et robuste. Afin de choisir le critère de segmentation, quelques tests ont été effectués pour comparer les méthodes les plus classiques :

- Seuil sur le ratio R/B [LSC+06].
- Seuil sur la différence R - B [HMS10].

Il n’existe pas de règle absolue pour juger de la qualité d’une segmentation de nuage, les résultats sont toujours comparés à l’estimation faite par un observateur. C’est donc en comparant visuellement les résultats de segmentation à l’image d’origine que l’on jugera la méthode et/ou les paramètres testés. Afin de tester la robustesse des méthodes de segmentation, les images tests sont choisies de façon à être représentatives des différentes conditions d’illumination possible. On choisit pour cela trois critères :

- L’angle solaire : *Haut* ou *rasant*.
- La présence de nuage devant le soleil : *nuage* ou *soleil*.
- Ciel *uni* ou *fractionné*.

Un ensemble de test de 8 images présentées en figure 2.14 est ainsi défini, ces images représentent l’ensemble des situations observables. Les seuils donnés dans les références respectives de chaque méthode dépendent de la caméra utilisée. Une étude préalable visant à la définition des seuils de chaque méthode a été menée. L’étude concernant la méthode de ratio est détaillée ci-dessous, les autres méthodes ont été calibrées de la même manière.

Définition du seuil pour la méthode R/B : La valeur conseillée [LSC+06] du seuil pour la méthode du ratio R/B est de 0.6. Afin de vérifier la validité de cette valeur pour les images du capteur utilisé ici, une série de tests a été effectuée avec des valeurs de seuils avoisinant 0.6. Pour chacun de ces tests, les 8 images tests sont représentées avec un contour de segmentation rouge représentant la frontière des nuages selon le seuil utilisé (Fig. A.1 à A.5). Un seuil trop faible conduit à une segmentation de l’image entière en tant que nuage (Fig. A.1) tandis que le phénomène inverse se produit pour un seuil pris trop grand (Fig. A.5) où l’ensemble de l’image est toujours considéré comme ciel clair. Le compromis se situe donc bien autour de la valeur 0.6. Notons toutefois deux problèmes :

- Pour les angles rasants, le "halo de diffusion" est considéré comme un nuage. Cette zone de mauvaise classification est d’autant plus large que le seuil sur le ratio R/B est faible.
- Certains nuages ne sont pas correctement identifiés lorsque le ratio augmente, notamment sous condition d’angle solaire proche du zénith.

Après une analyse visuelle détaillée des résultats présentés, un seuil de 0.65 a été choisi pour cette méthode. Cette valeur semble être le meilleur compromis étant donnés les problèmes

FIGURE 2.14 – Set d'images test pour la comparaison des méthodes de segmentation

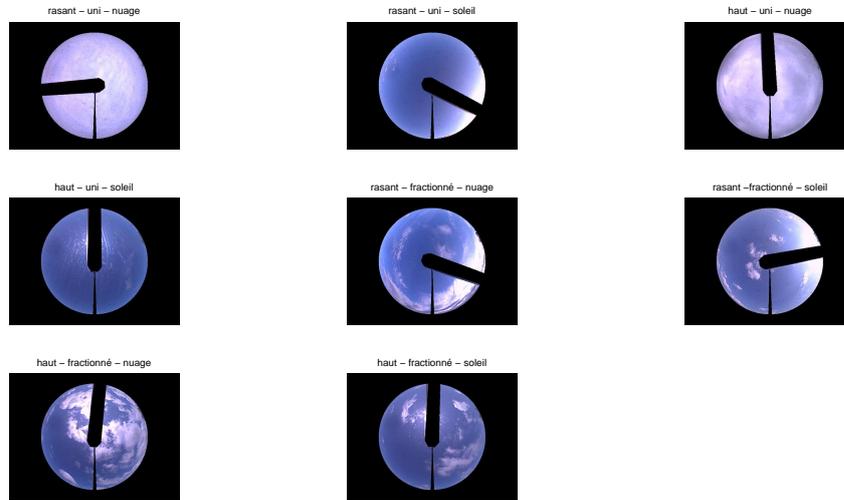


TABLE 2.8 – Paramètres des méthodes de segmentation utilisées

Nom	Formulation	Seuil
Ratio	$R/B > S_R$	$S_R = 0.65$
Différence	$B-R > S_D$	$S_D = 65$

évoqués ci-dessus.

De la même manière, nous fixons les seuils pour la méthode utilisant la différence R-B. Les résultats sont synthétisés dans le tableau 2.8. Les résultats obtenus avec les seuils choisis sont donnés en figure 2.15 et A.4. De manière générale, les résultats sont similaires et assez difficiles à départager. La méthode utilisant la différence semble toutefois introduire plus de bruit notamment dans le cas de couverture de type "uni". C'est pourquoi nous lui préférons le ratio R/B d'ailleurs plus couramment utilisé dans la littérature. Au vu des différents cas de segmentation, l'une des pistes pour améliorer cette étape serait de modéliser l'impact de l'éclairement sur l'image pour pouvoir corriger l'image avant de la segmenter.

Définition des caractéristiques nuageuses : A partir de la carte segmentée, deux caractéristiques principales sont définies :

- L'indice de couverture nuageuse : il est calculé en divisant le nombre de pixels nuageux par le nombre total de pixels de ciel (après masquage de l'image) :

$$ICC = N_{pixnuage}/N_{pixtotal} \quad (2.7)$$

- Le nombre de nuages Nb_{nuage} : il est calculé en utilisant une fonction de labélisation des objets segmentés (connexité 8) puis un décompte. Ce nombre permet de différencier les

FIGURE 2.15 – Test de seuil pour la méthode de segmentation par différence : Seuil = 65

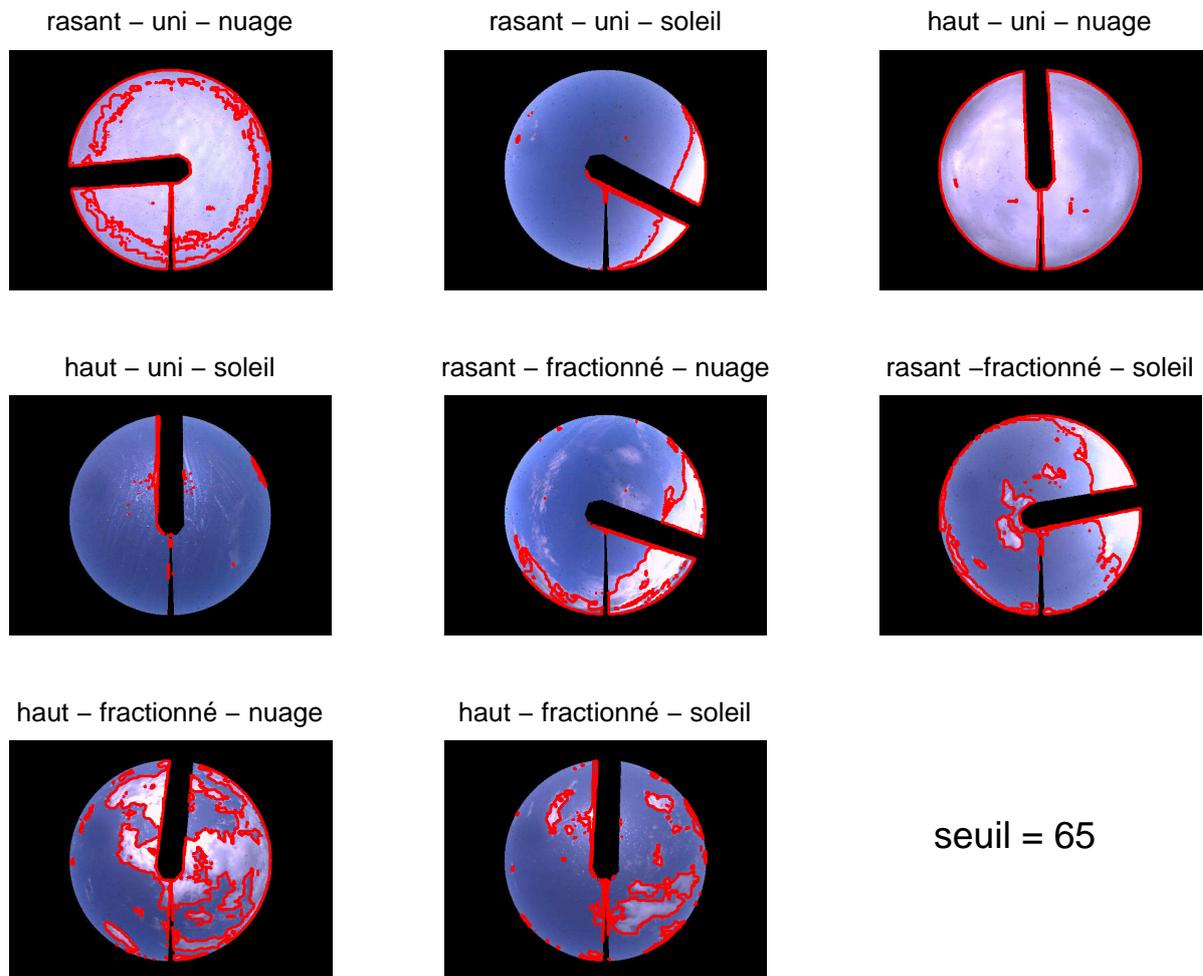


FIGURE 2.16 – Test de seuil pour la méthode de segmentation par ratio : Seuil = 0.65

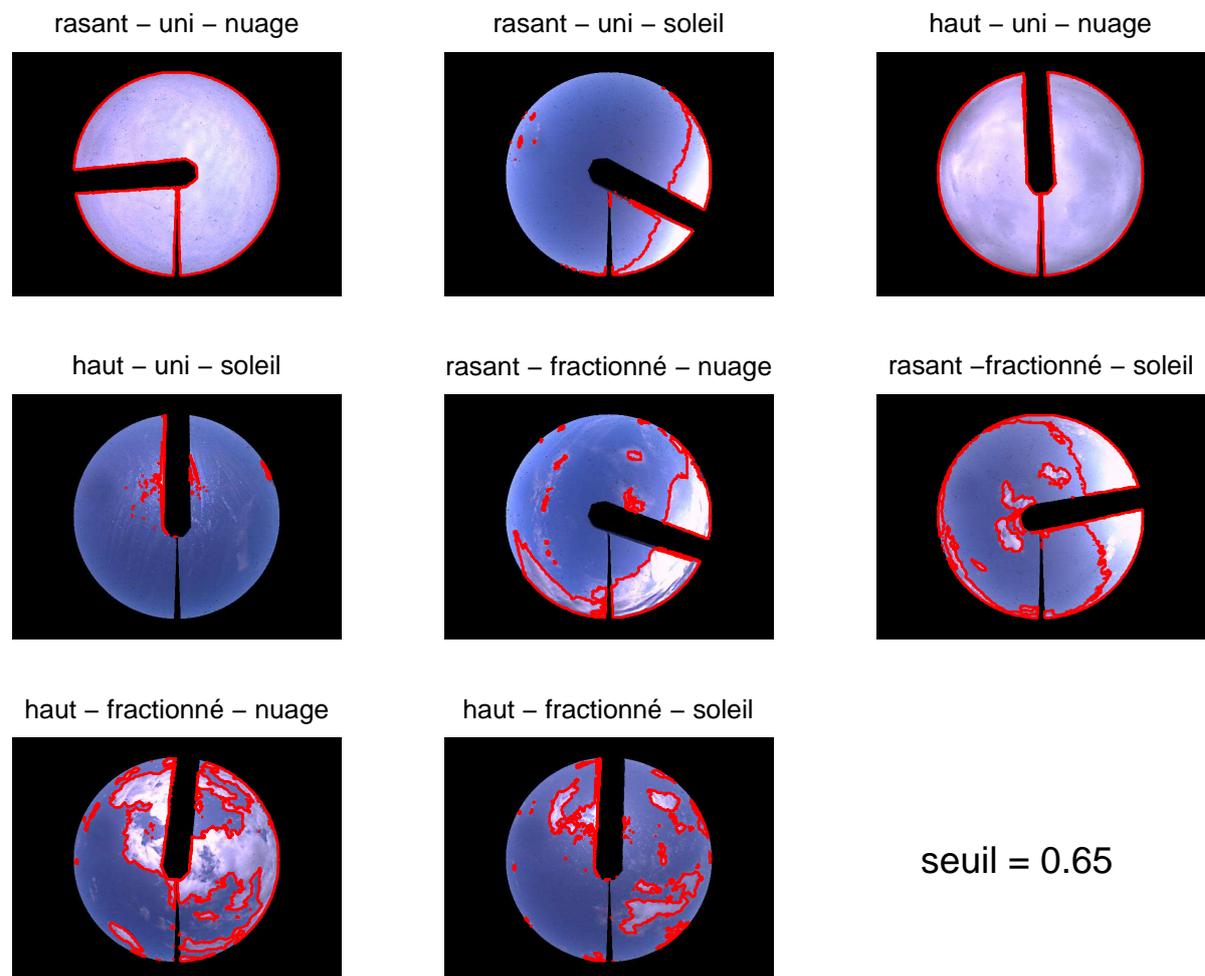
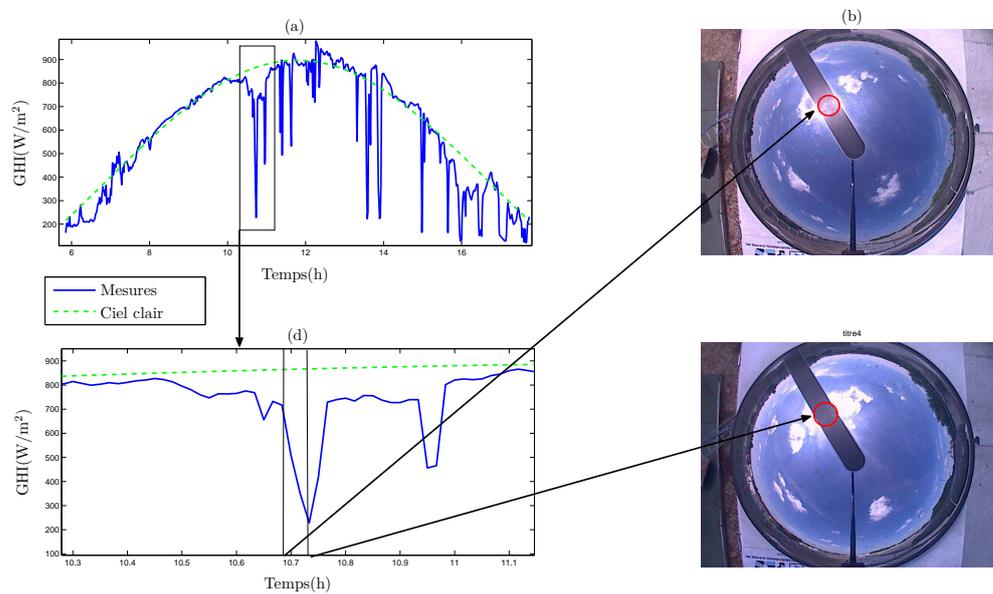


FIGURE 2.17 – Effet de blanchissement de la bande noire en présence de rayonnement direct



cas de ciel partiellement couvert de petits nuages des cas de ciel présentant un ou deux nuages importants.

2.2.3.4 Critère de présence du rayonnement direct

La bande noire apposée sur le miroir a pour objectif de masquer le rayonnement direct du soleil afin de protéger les cellules de la caméra sur le long terme. Toutefois, une observation approfondie a permis de mettre en évidence un comportement particulier : en présence de rayonnement direct, la zone circumsolaire de la bande noire subit un effet de blanchissement lié à la réflexion spéculaire du rayonnement direct. Ce phénomène n’apparaît pas lorsque le rayonnement provient uniquement de sources diffuses (Fig. 2.17). Pour avoir une observation indirecte de l’occultation ou non de la zone circumsolaire, le niveau moyen de la zone circumsolaire de la bande noire (entourée en rouge sur la figure 2.17.b) est extrait et utilisé comme un critère additionnel *Ban*. S’agissant d’un effet de blanchiment, les trois canaux se comportent de façon similaire, seul le niveau du canal rouge a été extrait apportant ainsi une information très liée à la présence de rayonnement direct, composante déterminante pour l’estimation du GHI.

Bilan : Nous retenons finalement un ensemble de 20 critères extraits pour chaque image rassemblés dans le tableau 2.9

TABLE 2.9 – Bilan des entrées brutes

Numéro	Critère	Numéro	Critère
1	MR	11	SkB
2	MG	12	SkS
3	MB	13	Con
4	MS	14	Hom
5	StdR	15	Dis
6	StdG	16	Ent
7	StdB	17	Ener
8	StdS	18	Icc
9	SkR	19	Nbnuage
10	SkG	20	Ban

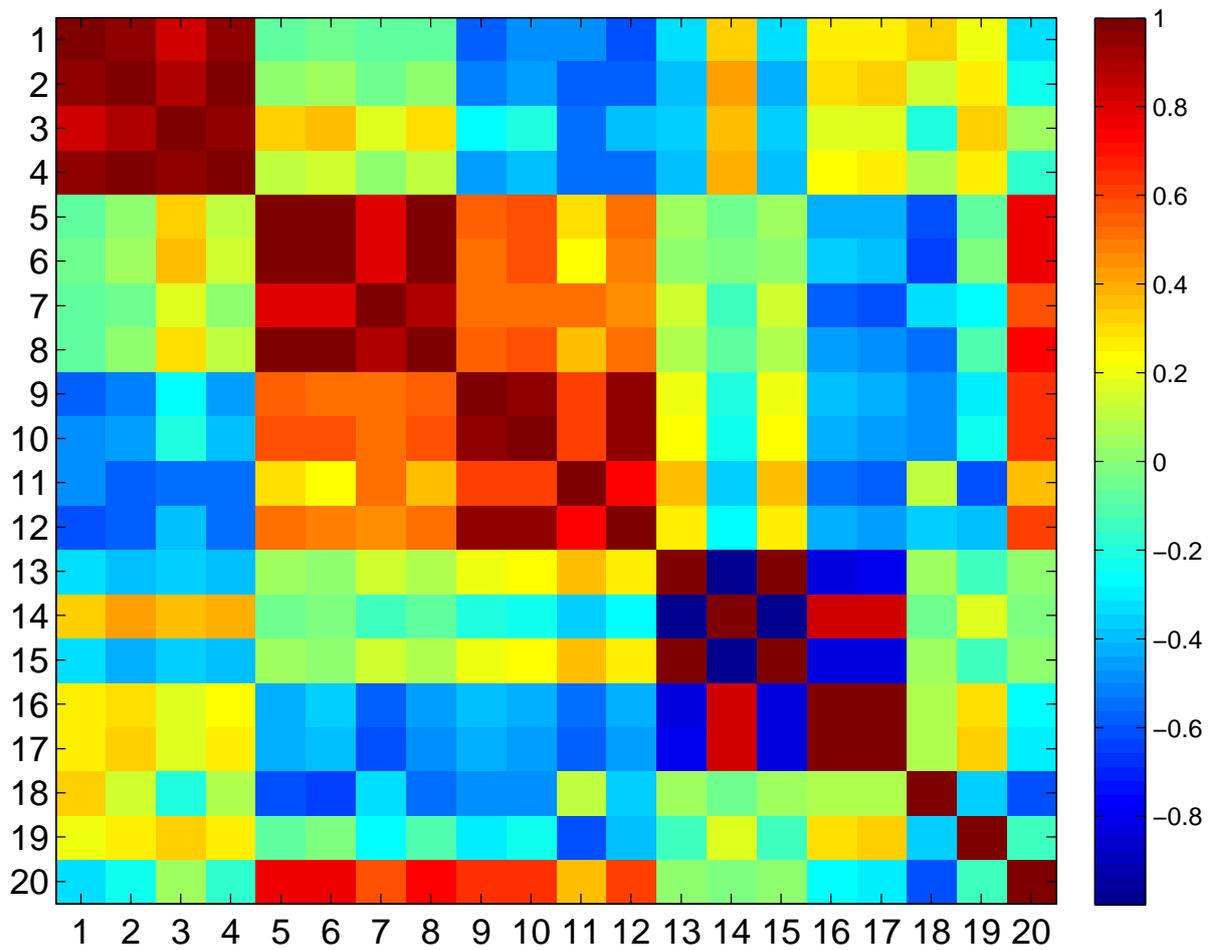
TABLE 2.10 – Entrées sélectionnées après l'étude de la matrice d'inter-corrélation (Fig. 2.18)

Numéro	Critère	Numéro	Critère
1	MR	7	Con
2	MB	8	Ent
3	StdR	9	Icc
4	StdB	10	Nbnuage
5	SkR	11	Ban
6	SkB		

2.2.4 Analyse des entrées

Ce paragraphe analyse la pertinence des entrées choisies et propose un premier tri des données. Les 20 entrées ont été extraites d'après des considérations bibliographiques sur l'interprétation de l'image en vue de son utilisation pour l'estimation du GHI. Nous nous interrogeons sur la redondance des critères sélectionnés ainsi que sur leur pertinence quant à l'estimation du rayonnement. Pour répondre à la première question, nous observons la matrice d'intercorrélation des entrées : si deux entrées sont extrêmement corrélées, elles apportent alors potentiellement la même information. Cette matrice est présentée en figure 2.18. Il apparaît alors clairement que les canaux rouge, vert et S (somme) sont très corrélés entre eux quels que soient les critères pris en comptes (moyenne, écart type ou skewness). Cela s'explique par l'impact similaire sur les canaux rouge et vert des nuages. D'après ces observations, il semble raisonnable de ne garder qu'un seul de ces canaux pour la suite de l'étude sans perdre d'information. Nous garderons donc de façon arbitraire le canal rouge en supprimant les entrées 2,4,6,8,10 et 12. Les critères de textures sont également redondants, le contraste (Con) est extrêmement corrélé à la dissemblance (par construction) et anti-corrélé à l'homogénéité, l'entropie est corrélée à l'énergie, les entrées 14 et 15 et 17 sont donc supprimés pour ne garder que le contraste et l'entropie. Finalement un ensemble réduit à 11 entrées est présenté dans le tableau 2.10.

FIGURE 2.18 – Matrice d'inter-corrélation des entrées



2.2.5 Réseau de neurones artificiels

Les réseaux de neurones sont apparus en télédétection dans les années 1990 et se sont montrés très performants par rapport aux algorithmes statistiques pour l'analyse de problèmes multi-sources [BSE90]. Leur capacité de généralisation dans des environnements bruités en font des techniques robustes dans les cas où les données d'entrées sont incomplètes ou biaisées [FS98]. De plus, les ANN sont flexibles et peuvent combiner différents types de données [BS97]. Ces qualités font des ANN de très bons outils pour estimer le GHI à partir de caractéristiques d'images au sol. D'un point de vue pratique, un réseau de neurones est un modèle mathématique. L'information est traitée par un groupe inter-connecté de neurones artificiels utilisant une approche calculatoire connexionniste [Bis95]. Parmi les différentes topologies de réseau de neurones, nous utilisons ici le Perceptron Multi-Couches (PMC) appartenant à la famille des réseaux en boucle ouverte (feedforward). Dans un réseau en boucle ouverte, chaque neurone est connecté à l'ensemble des neurones de la couche précédente (2.20), l'information se dirige dans une seule direction vers la sortie. Chaque neurone - aussi appelé nœud - dispose d'une fonction d'activation non linéaire f activée par la somme pondérée des entrées (Fig. 2.19). La phase d'apprentissage d'un PMC consiste à optimiser les poids associés aux connexions du réseau en minimisant l'erreur en sortie [Bis95]. Cette opération est souvent effectuée par des méthodes supervisées comme les approches de rétro-propagations. Nous choisissons dans le cadre de cette thèse la méthode d'apprentissage dite "scaled conjugate gradient" (SCG). Cette configuration du réseau de neurones est très adaptée au problème de régression [Hsu+92], toutefois son utilisation inadaptée peut mener à de mauvais résultats notamment liés aux problèmes de sur-apprentissage. Il y a sur-apprentissage lorsque la phase d'apprentissage conduit à un réseau tellement proche des données d'apprentissage que toute capacité de généralisation sur d'autres données est perdue. Par conséquent, deux problèmes majeures sont à traiter lors de l'utilisation de PMC :

- La définition d'un critère d'arrêt de l'apprentissage
- Le choix d'un nombre de neurones dans le(s) couche(s) cachée(s) de la topologie.

Afin d'éviter les problèmes de sur-apprentissage, nous effectuons un apprentissage à arrêt prématuré [Bis95]. La phase d'apprentissage implique alors également une base de validation contenant des éléments indépendants de la base d'apprentissage. A chaque nouvelle itération, les paramètres du réseau sont toujours ajustés pour réduire l'erreur en sortie sur la base d'apprentissage, mais en parallèle, on évalue également les performances du réseau ainsi paramétré sur la base de validation. La phase d'apprentissage s'arrête lorsque l'erreur minimale sur la base de validation est atteinte (Fig. 2.21). En utilisant cette méthode, l'erreur finale sur la base d'apprentissage peut être plus grande mais la capacité de généralisation du réseau est améliorée. Concernant le second problème, il est d'usage de minimiser l'erreur de sortie en fonction du nombre de neurones cachés à l'aide d'une série de tests.

2.2.6 Construction des bases d'apprentissage/test

D'après la description du processus d'apprentissage utilisé pour le réseau de neurones, trois bases de données indépendantes sont nécessaires :

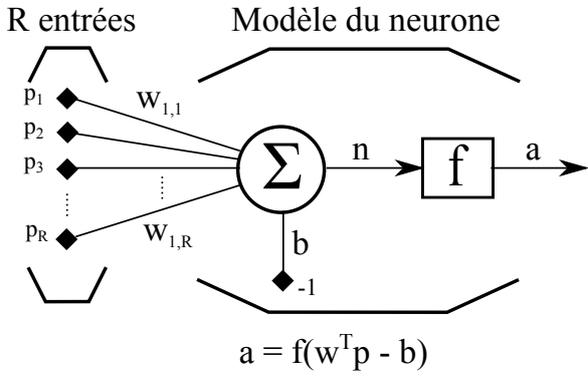


FIGURE 2.19 – Modèle d'un neurone

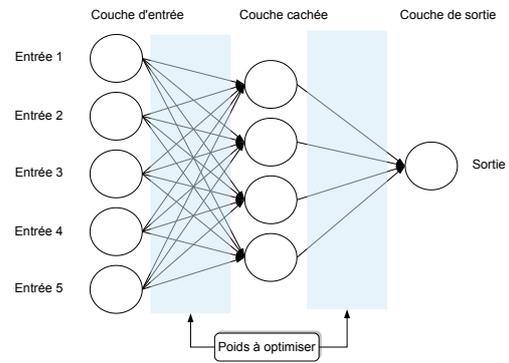


FIGURE 2.20 – Structure d'un réseau de neurones en boucle ouverte

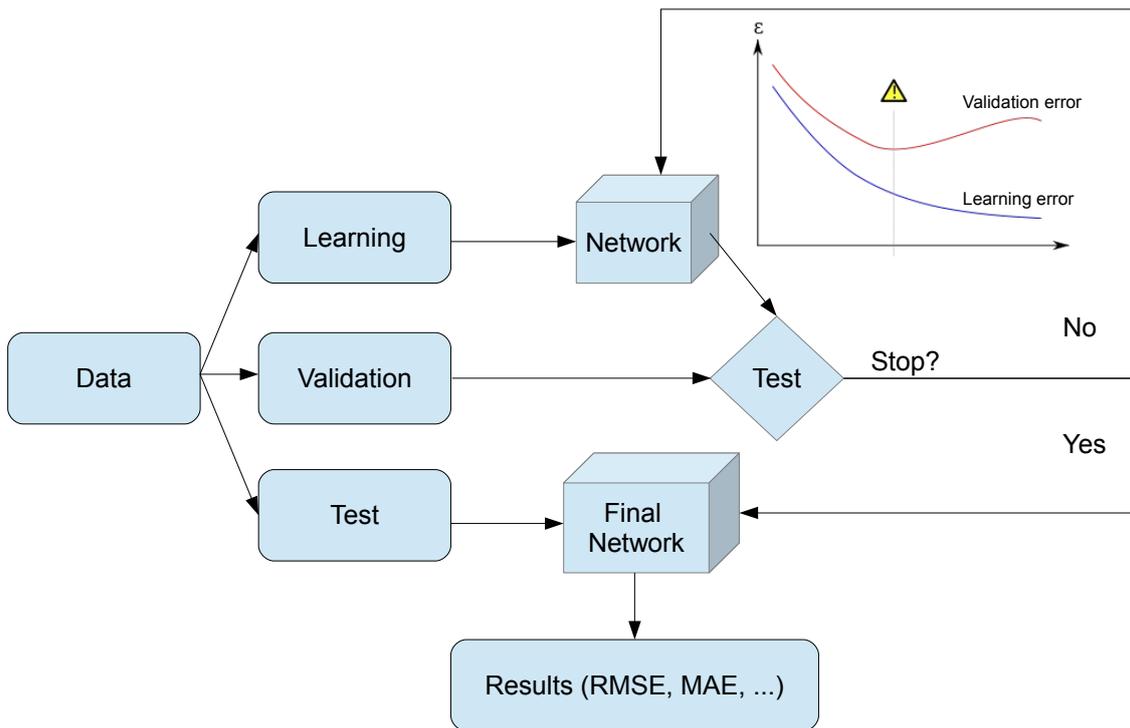


FIGURE 2.21 – Apprentissage à arrêt prématuré

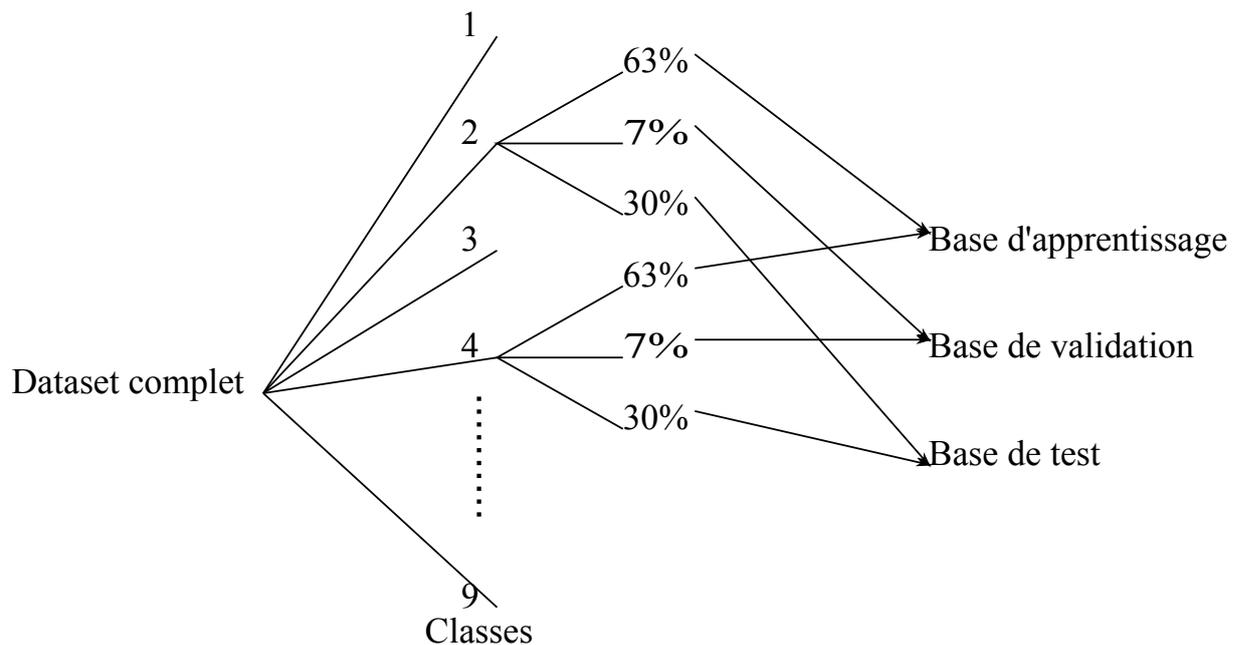


FIGURE 2.22 – Processus de création des bases de données pour l'apprentissage des réseaux de neurone

- La base d'apprentissage (App) : elle permet au réseau d'apprendre ses paramètres.
- La base de validation (Val) : elle permet de tester les capacités de généralisation du réseau pour décider de l'arrêt de l'apprentissage.
- La base de test (Test) : elle permet de tester les performances du réseau entraîné sur des données n'ayant jamais été impliquées ni dans l'apprentissage ni dans la validation.

Afin de valider les résultats obtenus, le modèle doit être testé dans toutes les conditions (tout type de temps). Ainsi faut-il que la base de test comprenne suffisamment d'éléments différents. En revanche, les performances du réseau ne pourront être optimales si la base de test contient des éléments trop éloignés de ceux appris dans la base d'apprentissage. Par conséquent, il est important que toutes les bases de données (App, Val, Test) soient suffisamment représentatives de l'ensemble des situations observables. Pour remplir cette condition, on utilise la classification des journées présentée au premier chapitre (section 1.4.1) en s'assurant que chaque classe soit bien représentée dans chacune des trois bases. Par ailleurs, il est d'usage de garder une majorité de données pour la base d'apprentissage : le nombre d'éléments qu'elle contient détermine non seulement le nombre de neurones maximal de la structure mais aussi la qualité de l'apprentissage dans une certaine mesure. Nous choisissons ici de manière empirique un taux de 63% pour la base d'apprentissage, 7% pour la base de validation et enfin, 30% des données sont consacrées à la base de test pour l'évaluation des résultats. Sur le plan pratique, dans un premier temps les trois bases pour chacune des neuf classes de journées sont créées suivant les taux donnés ci-dessus. Toutes les bases d'apprentissage sont ensuite rassemblées entre elles (de même pour les bases de validation et de test) pour créer les trois bases finales comme indiqué sur la figure 2.22.

TABLE 2.11 – Coefficient de corrélation entre chaque entrée et le GHI

1	2	3	4	5	6	7	8	9	10	11
Ban	SkR	SkB	MR	Icc	StdB	StdR	MB	Ent	Con	Nbnu
0.82	0.76	0.73	-0.66	-0.63	0.57	0.56	-0.44	-0.17	0.12	0.12

2.2.7 Résultats

2.2.7.1 Méthodologie

L’apprentissage et le test des réseaux de neurone a été fait grâce au logiciel Stuttgart Neural Network Simulator (SNNS) [Zel+94]. Ce logiciel se présente sous la forme de plusieurs fenêtre permettant notamment :

- La création de la structure du réseau.
- Le réglage des paramètres et de la fonction d’apprentissage, nous choisissons la fonction SCG.
- La visualisation graphique de la courbe de RMSE sur les bases d’apprentissage et de validation (utilisée pour l’apprentissage à arrêt prématuré).
- L’enregistrement dans un fichier texte du réseau entraîné et/ou des résultats sur la base souhaitée.

Le logiciel requiert toutefois des variables d’entrées comprises entre -1 et 1 et délivre des variables de sortie comprises entre 0 et 1. Les variables d’entrées sont normalisées par l’équation 2.8 où $Entree$ correspond à un vecteur contenant un certain nombre d’échantillon d’une des variables d’entrées sélectionnées et $Entree_{norm}$ la version normalisée et ainsi prête à être utilisée par SNNS. Concernant la sortie, l’index Kc est utilisé comme indiqué dans le chapitre 1 (section 1.3.2). Cet index est naturellement compris entre 0 et 2 dans des conditions normales, sa version normalisée est donc simplement le résultat d’une division par 2.

$$Entree_{norm} = \frac{Entree - moyenne(Entree)}{\max(|Entree - moyenne(Entree)|)} \quad (2.8)$$

L’apprentissage d’un réseau s’effectue selon le procédé suivant :

- Cinq apprentissages complets (sans arrêt prématuré) avec des initialisations aléatoires du réseau.
- Définition du point d’arrêt par la méthode d’arrêt prématuré.
- Apprentissage avec un arrêt prématuré (défini par le minimum trouvé).
- Évaluation du réseau sur la base de test et enfin enregistrement les résultats.

En suivant ce procédé, la recherche de la structure optimale devient alors le cœur de la phase expérimentale. Il s’agit de définir l’influence du nombre d’entrées d’une part, et du nombre de neurones de la couche cachée de l’autre. Pour ce faire, on effectue de multiples apprentissages en faisant varier ces deux paramètres. L’augmentation du nombre d’entrées a été effectué arbitrairement en suivant un ordre décroissant de corrélation avec la variable d’intérêt, le GHI. Cet ordre est présenté dans le tableau 2.11. Le nombre de neurones dans la couche cachée varie lui entre 1 et 15 par pas de 5 afin de limiter le nombre de tests requis. Les résultats de cette expérience sont présentés dans la section suivante.

2.2.7.2 Analyse des résultats

Analyse Quantitative Les tableaux 2.12 à 2.14 synthétisent les résultats des multiples expériences réalisées en rassemblant les trois critères retenus dans le chapitre 1. Les résultats de RMSE sont les plus aptes à qualifier la qualité du réseau étant donné que la fonction d'apprentissage du réseau de neurones s'appuie sur la minimisation d'une fonction de coût qui est elle-même l'erreur quadratique en sortie. Les résultats de RMSE varient entre $83.9W/m^2$ et $99.6W/m^2$ ce qui correspond à une évolution en valeur relative non négligeable de 27.1% à 32.3% (en référence à la moyenne annuelle de $306 W/m^2$). Il y a donc un réel intérêt à tester les multiples combinaisons de structures pour améliorer les résultats. D'après l'analyse du tableau 2.12, les moins bons résultats sont obtenus pour un nombre d'entrées allant de 1 à 3 ce qui suppose un manque d'information prévisible. Il est toutefois intéressant d'observer que la diminution de la RMSE ne va pas de paire avec l'augmentation du nombre d'entrées. En effet, le nombre d'entrées optimal semble être atteint pour cinq entrées. Avant ce stade, chaque ajout semble diminuer la RMSE, prouvant ainsi sa pertinence. En revanche, après 5 entrées, on obtient des résultats moins bons et assez stagnant (entre 89 et $93 W/m^2$) quel que soit le nombre d'entrées ou la structure choisie. Il apparaît donc que les entrées supérieures au rang 5 n'apportent pas de réelle information supplémentaire pour l'estimation du GHI. Ces résultats sont confirmés par l'étude des scores de coefficient de détermination (tableau 2.14). La même tendance est observée avec une amplitude moindre puisque les résultats varient entre 0.86 et 0.90. Le tableau de MBE doit être étudié avec précaution, la MBE peut être considéré comme un score auxiliaire, il n'est pas directement visé par la méthode utilisée ici qui se focalise sur la RMSE. Aussi, les résultats sont-ils compris entre -9.83 et $9.61 W/m^2$ ce qui reste relativement faible (2%) en comparaison des scores de RMSE avoisinant les 30%. Notons toutefois que les scores obtenus pour 5 entrées sont relativement faibles, ce qui conforte l'hypothèse préalablement établie. Du point de vue de la structure du réseau, les résultats sont relativement difficiles à exploiter. Le nombre optimal de neurones ne semble pas être lié aux nombre d'entrées, les résultats semblent en effet assez indépendants les uns des autres. On retiendra donc qu'une étude sur la structure du réseau est toujours nécessaire lorsqu'on choisit une approche par réseau de neurones.

Analyse Qualitative D'après le tableau de RMSE (tableau 2.12) le réseau optimal choisi possède 5 entrées et 1 neurone dans la couche cachée. Nous étudions les résultats classe par classe pour mieux interpréter les erreurs du modèle. Les tableaux 2.15 à 2.17 présentent l'ensemble des scores pour chaque classe. La figure 2.23 donne un exemple de profil journalier pour chaque classe. La RMSE semble évoluer avec le niveau moyen de l'index K_c des journées (classes A,B,C). Des RMSE faibles sont observées pour les journées couvertes alors que les plus fortes sont enregistrées pour les journées relativement claires. En effet, l'impact d'une erreur d'estimation a tendance à être plus important pour les journées claires car les valeurs de GHI y sont beaucoup plus élevées. En revanche l'évolution du coefficient de détermination est elle plus dépendantes de la variabilité de la journée. Les journées peu variables obtiennent des scores faibles alors que de très bonnes performances sont observées pour les journées présentant un profil très perturbé. Ces résultats sont confirmés par les courbes de la figure 2.23

TABLE 2.12 – Tableau des résultats de RMSE

Nb neurones cachés		1	5	10	15
Nb entrées					
1		99.6	99.1	98.5	97.4
2		98.9	98.1	98.6	97.9
3		97.7	90.9	93.1	96.9
4		87.4	98.2	95.3	95.9
5		83.9	87.5	88.5	85.2
6		92.3	84.2	85.2	89.9
7		92.2	85.6	92.5	89.2
8		90.5	95.6	92.4	90.6
9		93.9	90.8	86.6	90.4
10		90.7	91.9	90.6	90.6
11		90	92.4	90.3	90.6

TABLE 2.13 – Tableau des résultats de MBE

Nb neurones cachés		1	5	10	15
Nb entrées					
1		6.93	9.61	7.93	8.80
2		5.83	5.3	10.70	11
3		0.12	-0.14	2.34	-1.59
4		2.33	7.3	2.36	5.2
5		-2.54	-1.43	-1.32	3.33
6		0.94	1.82	-5.76	-8.77
7		0.97	-0.95	1.26	5.18
8		-6.46	9.15	2.74	-2.10
9		3.59	1.92	-2.11	-3.53
10		0.56	-9.83	0.23	-0.79
11		-2.69	-8.72	-0.17	-1.78

TABLE 2.14 – Tableau des résultats de coefficient de détermination

Nb entrées	Nb neurones cachés			
	1	5	10	15
1	0.86	0.86	0.86	0.86
2	0.86	0.86	0.86	0.86
3	0.86	0.88	0.88	0.86
4	0.89	0.86	0.87	0.87
5	0.90	0.89	0.89	0.90
6	0.88	0.90	0.90	0.88
7	0.88	0.89	0.88	0.89
8	0.88	0.87	0.88	0.88
9	0.87	0.88	0.89	0.88
10	0.88	0.88	0.88	0.88
11	0.88	0.88	0.88	0.88

TABLE 2.15 – Résultats de RMSE par classe

	I : peu variable	II : variable	III : très variable
A : couvert	33	85	99
B : mitigé	47	87	123
C : clair	36	75	98

où la cohérence globale de l'estimation est meilleure lorsque les phénomènes observés ont une grande amplitude. Ces importantes variations sont principalement dues aux fluctuations du rayonnement direct, en d'autre terme au passage de nuages dans l'axe soleil/camera. Dans ces conditions, le ciel observé par la caméra présente des nuages blancs ou gris sur un fond bleu. Ces conditions particulièrement contrastées forment des conditions idéale pour l'étude par traitement d'image. Par ailleurs, la caractéristique Ban extraite de la bande noire est très sensible aux atténuations du rayonnement direct, permettant de meilleures performances dans ce cas. En revanche dans le cas de journées couvertes, il n'y a que peu de contraste dans l'image, il s'agit globalement de faibles nuances de gris. Dans ces conditions, les caractéristiques extraites ne permettent plus d'expliquer les variations du GHI (mauvais coefficient de détermination). Pour conclure, la méthode est intrinsèquement performantes pour caractériser les fortes variations de GHI, en revanche elle donne peu de précision sur les faibles fluctuations.

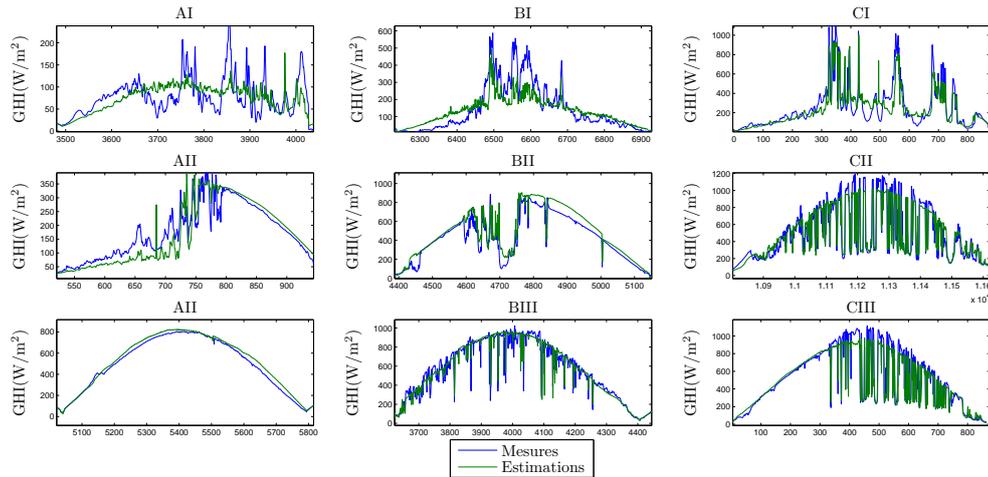
TABLE 2.16 – Résultats de coefficient de détermination par classe

	I : peu variable	II : variable	III : très variable
A : couvert	0.50	0.62	0.79
B : mitigé	0.95	0.87	0.80
C : clair	0.99	0.92	0.91

TABLE 2.17 – Résultats de MBE par classe

	I : peu variable	II : variable	III : très variable
A : couvert	-1.3	-1	13.2
B : mitigé	12.8	0.17	8.7
C : clair	-22	-9	13

FIGURE 2.23 – Profil journaliers pour chacune des classes



2.3 Conclusion

L’analyse de la bibliographie a montré l’intérêt de l’étude des images au sol pour la caractérisation de la couverture nuageuse. De nombreuses études ont permis de mieux segmenter les nuages et de trouver les caractéristiques adaptées à la classification des types de couverture. En s’appuyant sur ces éléments et indications bibliographiques nous avons montré l’intérêt de ces mêmes images pour l’estimation du GHI. L’hypothèse formulée ici est la suivante : les nuages ont un impact fort sur le GHI. Si nous disposons d’outils pour les caractériser, alors ces mêmes outils peuvent aussi servir à l’estimation du GHI. Cette hypothèse a été ici validée via l’apprentissage d’un réseau de neurones alimenté par des caractéristiques choisies et discutées issues de l’analyse des images hémisphériques du ciel acquises par le TSI. Les résultats ont permis de définir un set de caractéristiques optimales pour l’estimation du GHI parmi l’ensemble des caractéristiques extraites préalablement. On obtient alors une estimation très satisfaisante, avec des scores de 27% de RMSE relative.

Imagerie satellitaire pour la prévision du GHI

Sommaire

3.1	Introduction	59
3.2	Etat de l'art	60
3.2.1	Acquisition des images	60
3.2.2	Traitement	62
3.2.3	HélioClim-3	64
3.3	Prévision du GHI par modèle de prévision linéaire spatio-temporel	67
3.3.1	Performances d'HélioClim-3 sur le site du SIRT	67
3.3.2	Les modèles Autoregressifs (AR)	70
3.3.3	Définition des modèles testés	73
3.3.4	Résultats	74
3.3.5	Conclusion	78
3.4	Prévision par calcul de champ de vecteur	81
3.4.1	Méthodologie	81
3.4.2	Résultats	86
3.5	Conclusion	92

3.1 Introduction

Les images provenant des satellites géostationnaires d'observation sont probablement les données les plus adaptées à la prévision à court terme (de l'ordre de l'heure) de GHI. Elles couvrent la totalité du globe terrestre avec des résolutions spatiale et temporelle plus fines que celles obtenues en sortie des modèles numériques de prévision. Le développement des capteurs embarqués au cours des dernières années a permis une plus large utilisation de ces images notamment grâce à l'amélioration des résolutions spatio-temporelles et spectrales. Ce chapitre est consacré à l'ensemble des travaux développés à partir de ces images. En première partie, nous présenterons une étude bibliographique des méthodes de prévision à court terme à partir d'imagerie satellitaire géostationnaire. La seconde partie décrit une première approche globale permettant de mettre en évidence l'apport de l'information spatio-temporelle par rapport à d'autres modèles plus simples. Enfin, la troisième partie décrit une approche plus locale

TABLE 3.1 – Les satellite de Meteosat

Satellite	Longitude	Date de lancement	Usage
Meteosat 7 (MFG)	57°E	1997 / Océan indien depuis 2006	Couverture de l'océan Indien, Imagerie temps-réel
Meteosat 8 (MSG)	3.5°E	2002	Service back up pour 0°E et RSS
Meteosat 9 (MSG)	9.5°E	2005	Rapide scan service depuis 2013
Meteosat 10 (MSG)	0°E	2012	Image temps-réel du capteur SEVIRI

s'appuyant sur l'extraction de vecteur de mouvement directement issus des images satellites précédant l'instant de la prévision. Ces deux dernières parties ont fait l'objet de publications dans le journal international Renewable Energy pour la première (en cours de publication), par la conférence IREC¹ 2014 pour la seconde [Dam+14].

3.2 Etat de l'art

3.2.1 Acquisition des images

Les images utilisées dans ce chapitre proviennent du satellite Meteosat 10 qui est un satellite géostationnaire appartenant à la famille des Meteosat. Le projet Meteosat est dirigé par EUMETSAT (European Organization for the Exploitation of Meteorological Satellites) et l'agence spatiale européenne (ESA). Quatre satellites sont actuellement actifs et couvrent différentes parties du globe (Tableau 3.1). A l'heure actuelle, il existe deux générations de satellites, une première (MFG) dont le dernier représentant est Meteosat 7 sur l'Océan Indien et une seconde génération (MSG) regroupant les satellites Meteosat 8 à 10. Les satellites de seconde génération, dont Meteosat 10, sont équipés entre autres du capteur SEVIRI (Spinning Enhanced Visible and InfraRed Imager) qui permet l'acquisition d'images terrestres dans douze bandes spectrales alors que le capteur de première génération n'en autorisait que trois. Sur les douze canaux, huit concernent l'infrarouge thermique et quatre le domaine visible dont un en haute résolution (HRV). Ce dernier dispose d'une résolution spatiale au NADIR de 1 km tandis que les autres canaux ont une résolution de 3 km au NADIR. Des exemples d'images brutes sont donnés en figures 3.1 et 3.2. Sur ces images, la valeur de chaque pixel correspond à la réflectance perçue par le satellite.

1. International Renewable Energy Congress

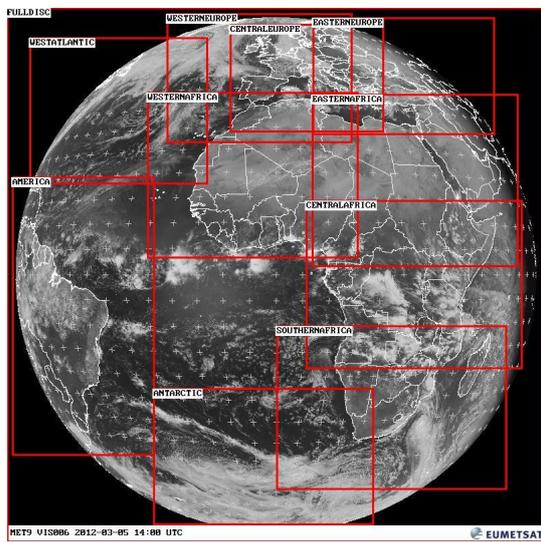


FIGURE 3.1 – Canal visible de Meteosat 9

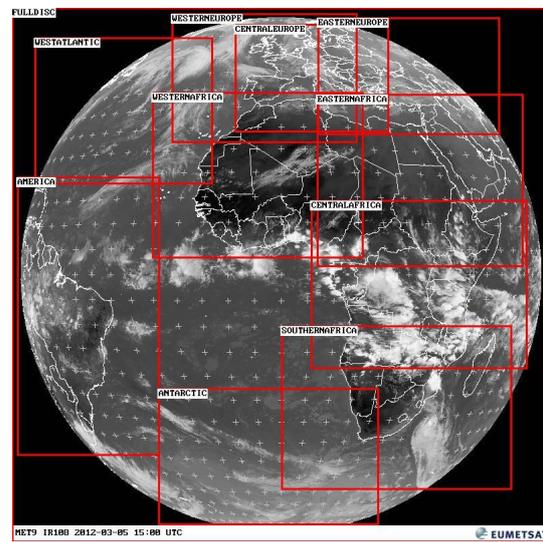


FIGURE 3.2 – Canal Infrarouge de Meteosat 9

3.2.2 Traitement

La prévision de GHI à partir d'images satellitaires peut se décomposer en deux étapes principales :

- Conversion : Il s'agit d'interpréter les valeurs des pixels de l'image satellite brute en estimation de GHI.
- Prévision : Il s'agit d'anticiper le déplacement des masses nuageuses observées.

Ces deux étapes sont généralement traitées séparément pour la plupart des méthodes proposées dans la bibliographie (sauf [MPC13]). Elles sont décrites dans la suite.

3.2.2.1 Conversion

Afin de mieux comprendre l'étape de conversion, rappelons le lien entre l'observation depuis le satellite et le rayonnement perçu au sol. La figure 3.3 présente les flux interagissant dans la compréhension du phénomène. Sur cette figure, le GHI mesuré correspond à la somme du rayonnement incident au sol I_{in}^g et du rayonnement réfléchi par celui-ci I_{out}^g tandis que le signal perçu par le satellite est lié à la valeur de I_{out}^{TOA} ². La relation liant ces variables (3.1) fait intervenir le rayonnement incident au sommet de l'atmosphère I_{in}^{TOA} , le rayonnement absorbé I_{in}^a et le rayonnement réfléchi par l'atmosphère et les nuages (I_{out}^a). Si le premier (I_{in}^{TOA}) est une variable relativement aisée à calculer, les suivantes justifient une partie de la bibliographie concernant l'estimation du GHI à partir d'image satellite.

$$I_{in}^g + I_{out}^g = I_{in}^{TOA} - I_{in}^a - I_{out}^a \quad (3.1)$$

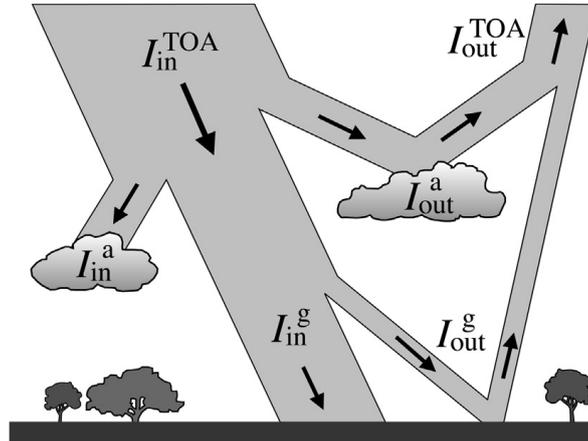


FIGURE 3.3 – Rappel sur les relations entre l'observation depuis le satellite et le rayonnement perçu au sol. Image extraite de [IPC13]. La lettre "a" désigne l'atmosphère, la lettre "g" désigne le sol (ground).

Deux types de méthode sont alors envisagées pour établir le lien entre l'observation satellitaire et le GHI : les modèles physiques [NRF93a] et les modèles statistiques ou empiriques

2. TOA = Top Of Atmosphere

[NRF93b]. Les modèles physiques ont l'avantage d'être généraux et peuvent s'appliquer directement. Ils requièrent en contrepartie des mesures précises sur la composition de l'atmosphère et une calibration fine des images satellites. Au contraire, les méthodes statistiques basées sur des régressions nécessitent un historique de données mais sont moins lourdes en calcul et plus robustes. On retiendra parmi les modèles physiques ceux de Gautier-Diak-Masse [GDM80], de Marullo-Dalu-Viola [MDV87], de Möser-Raschke model [MR84] et enfin celui de Dedieu-Deschamps-Kerr [DDK87]. Du côté des modèles statistiques, les plus connus sont ceux de Hay-Hanson [HH78], le modèle opérationnel de Perez [Per+02] et enfin la méthode Heliosat de Cano [Can+86]. Les références données ici correspondent aux premières publications pour chacun des modèles, il y a eu ensuite de multiples évolutions dans chacun des cas. On trouvera une présentation plus détaillée de ces modèles dans la revue proposée par Inman [IPC13] dont ce paragraphe est inspiré.

3.2.2.2 Prévision

L'intérêt principal de l'imagerie satellitaire géostationnaire pour la prévision du GHI réside dans sa capacité à décrire de manière dense dans le temps et l'espace la couverture nuageuse sur de larges zones avec une résolution spatiale de 3 à 10 km et une résolution temporelle de 15 min. La couverture nuageuse représente le principal facteur de variabilité intra-journalière du GHI pour la majorité des sites. C'est la raison pour laquelle la capacité à prévoir son mouvements conduit à un potentiel de prévision très intéressant. L'image satellitaire permet de localiser précisément la position des principales masses nuageuses en temps réel. En effet, la plupart des nuages ont une réflectivité supérieure à celle de la terre ou de la mer, ce qui les rend facilement identifiable dans l'image. Si leur vitesse de déplacement est connue, il est possible d'obtenir une prévision du futur positionnement des nuages et donc calculer le rayonnement au sol résultant via une des méthode de conversion citées auparavant. L'enjeu principal de l'étape de prévision est donc d'obtenir la future configuration de la couverture nuageuse. Cette information peut être extraite des prévisions météorologiques de vents ou même d'observations mesurées sur site. Toutefois, la vitesse et la direction des vents dépend fortement de l'altitude. Cette méthode requiert donc une donnée supplémentaire sur l'altitude des nuages. Or cette altitude n'est pas facile à obtenir, le satellite n'ayant accès qu'à la couche supérieure du nuage. Il est donc généralement préférable d'utiliser les séries d'images elles-même pour évaluer le mouvement des masses nuageuses. En observant l'évolution de la position des nuages dans le temps, leur vitesse est obtenue grâce à des algorithmes de traitement d'image classiques. Parmi les méthodes de traitement d'image possibles, les plus couramment utilisées sont : le *blockmatching*, le flot optique et la technique connue sous le nom anglais "particle image velocimetry" (PIV) que l'on pourrait traduire par "vélocimétrie des images des particules".

3.2.2.3 Travaux publiés et positionnement des travaux

Le tableau 3.2 résume les choix concernant ces deux problèmes dans les publications récentes sur le sujet :

TABLE 3.2 – Tableau récapitulatif des travaux de la littérature

Méthode	Conversion	Prévision
Hammer 1999 [Ham+99]	Heliosat	block matching
Perez 2002	Perez operational model	block matching
Marquez 2013 [MPC13]	Neural network (statistical)	PIV
Peng 2013 [Pen+13]	Support Vector machine (SVM)	flot optique

Les méthodes décrites dans ce tableau effectuent d’abord une étape de prévision par traitement d’image puis une conversion menant à une prévision de GHI. Les algorithmes de conversion récents ont un temps de calcul qui permet la production de carte de GHI en temps réel au moins sur des zones réduites. Nous nous interrogeons sur la possibilité d’inverser les deux étapes et d’effectuer une prévision directement issue des cartes de GHI produites. L’intérêt de cette approche est multiple :

- S’affranchir du savoir faire nécessaire à la bonne conduite de l’étape de conversion.
- Travailler avec des données d’entrées directement liées à la variable que l’on souhaite prévoir. Cela permet notamment d’ouvrir le champs des méthodes possibles pour la prévision.

En contrepartie, nous incluons les erreurs liées à la méthode d’estimation dans le modèle de prévision. C’est sur cette alternative que s’articulent les travaux présentés au cours de ce chapitre. La méthode de conversion Héliosat-2 a été choisie pour ses résultats, mais également pour la disponibilité des cartes de GHI résultantes par l’intermédiaire de la base de données HélioClim-3. Ce sont les cartes extraites de cette base que l’on notera carte HC-3 qui seront utilisées comme uniques données d’entrées satellitaires tout au long de ce chapitre. HC-3 et la méthode Héliosat-2 sur laquelle elle s’appuie sont décrites plus amplement dans le paragraphe suivant.

3.2.3 HélioClim-3

3.2.3.1 Heliosat : la méthode

La méthode Heliosat permet la conversion d’images satellitaires dans le domaine de l’optique visible en estimations d’irradiation au niveau du sol. Sa première version été publiée en 1986 par *Cano et al.* [Can+86]. Plusieurs corrections du modèle ont alors été proposées dans la littérature [Mou+89], [Obr90], [Zel+92], [BCH96], [Fon+97], [Ieh+97], [IP99]. L’expérience apportée par ces publications a permis d’enrichir une nouvelle méthode, appelée Heliosat-2 [RLW04] répondant à un besoin d’uniformité dans les résultats présentés par ces études. La dernière version de la méthode, Héliosat-4 a été présentée par *Oumbe et al.* [Oum+09] et propose une estimation distincte des rayonnement direct et diffus. En raison de la jeunesse de la méthode Heliosat-4 en mode opérationnel, c’est la méthode Héliosat-2 qui sera utilisée ici et présentée ci-dessous.

Méthode opérationnelle : Ce paragraphe reprend les grandes lignes de la méthode Héliosat-2, décrite dans la publication de *Rigollier et al.* [RLW04]. Dans un premier temps les valeurs brutes des pixels de l'image sont converties en radiances suivant le modèle proposé par *Rigollier et al.* [Rig+02]. L'étape la plus importante de la méthode réside alors dans le calcul de l'indice de couverture nuageuse $n^t(i, j)$ pour chaque pixel à la position (i,j) dans l'image, décrit par l'équation (3.2) où :

- $\rho^t(i, j)$ est la réflectance, ou albedo apparent obtenu à partir de la valeur de radiance du pixel
- $\rho_g^t(i, j)$ est l'albedo du sol, sous condition de ciel clair. Il est obtenu par une analyse statistique sur fenêtre glissante (typiquement les 15 à 30 derniers jours). Cette analyse statistique consiste à déterminer la réflectance minimale observée dans le passé proche et de la considérer comme étant observée par condition de ciel clair modélisé par le modèle ESRA.
- $\rho_{cloud}^t(i, j)$ est l'albedo apparent pour un nuage, obtenu à partir de l'albedo des nuages les plus clairs et contraint selon le procédé décrit par [Bau96]
- θ_S est l'angle solaire zénithal
- $L^t(i, j)$ est la radiance observée
- I_{0met} est l'irradiance totale observée par les capteurs de Météosat dans le domaine visible.

$$\begin{aligned} n^t(i, j) &= [\rho^t(i, j) - \rho_g^t(i, j)] / [\rho_{cloud}^t(i, j) - \rho_g^t(i, j)] \\ \rho^t(i, j) &= \pi L^t(i, j) / I_{0met} \epsilon(t) \cos \theta_S(t, i, j) \end{aligned} \quad (3.2)$$

Cet indice $n^t(i, j)$ est ensuite mis en relation avec l'indice de ciel clair Kc (section 1.3.2) suivant les équations 3.3. Multiplier cet indice par le modèle ciel clair ESRA permet alors de produire l'estimation de GHI attendue.

$$\begin{aligned} n^t < -0.2 &\quad \Rightarrow & Kc = 1.2 \\ -0.2 < n^t < 0.8 &\quad \Rightarrow & Kc = 1 - n \\ 0.8 < n^t < 1.1 &\quad \Rightarrow & Kc = 2.0667 - 3.6667n^t + 1.6667(n^t)^2 \\ n^t > 1.1 &\quad \Rightarrow & Kc = 0.05 \end{aligned} \quad (3.3)$$

Les résultats de cette méthode ont été testés sur 35 sites en Europe avec des erreurs variant entre 18% et 45% de RMSE relative selon les mois testés pour une estimation en irradiation horaire de GHI.

3.2.3.2 Hélioclim : les différentes base de données

Le projet Hélioclim a été lancé en 1997 par MINES ParisTech [Bla+11]. L'objectif principal de ce projet est de pouvoir fournir des données de rayonnement en surface pour une grande zone géographique sur un intervalle temporel long afin de mener des études à long terme. La première base de données Hélioclim-1 (HC-1) propose des valeurs moyennes journalières pour les zones Europe, Afrique et Océan Atlantique de 1985 à 2005. Elle a été créée à partir des images des satellites MFG d'EUMETSAT traitées par la méthode Héliosat originale [Can+86]. Profitant des meilleures performances des satellites de seconde génération et

des avancements techniques de la méthode Héliosat-2, la nouvelle base de données HélioClim-3 bénéficie de meilleures résolutions temporelle (15 minutes) et spatiale (3 km au NADIR) mais aussi de meilleures performances. Elle couvre la période 2004 à aujourd'hui. Les images peuvent également être traitées en temps réel grâce à la station de réception METEOSAT acquise en 2003 par MINES ParisTech. Les cartes sont donc calculées toutes les 15 minutes et peuvent être moyennées pour fournir des des irradiances demi-horaires, journalière et plus (mensuelles ou annuelles).

3.3 Prédiction du GHI par modèle de prédiction linéaire spatio-temporel

Cette section présente un travail original visant à mettre en évidence la pertinence de l'information spatio-temporelle apportée par les cartes HC-3 pour la prédiction à très court terme. L'utilisation des cartes HC-3 homogènes à la variable d'intérêt rend l'utilisation de modèles de prédiction linéaire tels que les modèles autoregressifs (AR) possible. L'objectif principal est de quantifier l'apport de l'information spatio-temporelle pour des horizons de prédiction allant de 15 minutes à 1 heure mais aussi d'établir un premier modèle global. Ce modèle est dit "global" car il ne fait pas appel à une étape de "prédiction" au sens où on l'a décrite précédemment. Il ne nécessite aucun calcul de vecteur, en revanche il s'appuie sur une étude préalable définissant des zones d'intérêt d'après une étude statistique des phénomènes les plus courants. Nous commencerons par étudier la localisation de ces zones d'intérêt en fonction de l'horizon de prédiction (section 3.3.1.1). Nous présenterons ensuite brièvement le modèle AR qui sera utilisé pour réaliser la prédiction (section 3.3.2). La partie expérimentale permet la quantification de l'apport d'information grâce à la comparaison de quatre modèles de complexité croissante allant du modèle de persistance (section 1.3.3) au modèle de prédiction spatio-temporel (section 3.3.3). Les résultats et discussions sont développés dans la section 3.3.4.

3.3.1 Performances d'HélioClim-3 sur le site du SIRTA

Avant d'utiliser HC-3 pour la prédiction nous nous intéressons aux résultats d'estimation effectivement obtenus sur le site du SIRTA³. Grâce aux positions géolocalisées du site de mesure, une large zone de la carte HC-3 centrée sur le site a été extraite pour deux années consécutives (2011 et 2012) dans le cadre de la thèse. La série temporelle d'estimation du pixel HC-3 couvrant le site de mesure a été comparée à la série de mesures au sol. Les résultats sont donnés dans le tableau 3.3. Les performances sont cohérentes avec les performances attendues d'HC-3 à cette résolution temporelle, à savoir environ 25% de RMSE relative et un biais négatif de $20W/m^2$ [Bla+11]. Les pourcentages de RMSE relative donnés ici font référence à la valeur moyenne du GHI sur l'ensemble des mesures prises en compte ($306W/m^2$ en 2011 et $291W/m^2$ en 2012). Ces scores établissent le seuil le minimum des performances que l'on peut attendre en prédiction à partir de ces images si on ne considère pas de processus de calibration par rapport au site. En revanche il est possible de diminuer cette erreur en calculant une courbe de calibration à partir d'un historique de mesure sur site. Nous utilisons alors l'année 2011 pour la calibration et l'année 2012 pour tester les modèles de prédiction.

Comme indiqué au chapitre 1 (section 1.3.2), il est plus intéressant de travailler avec l'index de ciel clair Kc pour s'affranchir des phénomènes climatiques connus. Aussi, la calibration a été effectuée sur l'index de ciel clair Kc . L'enjeu est de trouver une régression permettant de diminuer l'écart entre l'index $K_{c_{HC3}}$ issu de HC-3 et l'index Kc issu de la mesure au sol. La figure 3.4 illustre le polynôme d'ordre 4 retenu pour effectuer une correction simple du nuage

3. (48.7130 N, 2.2080 E), 156m)

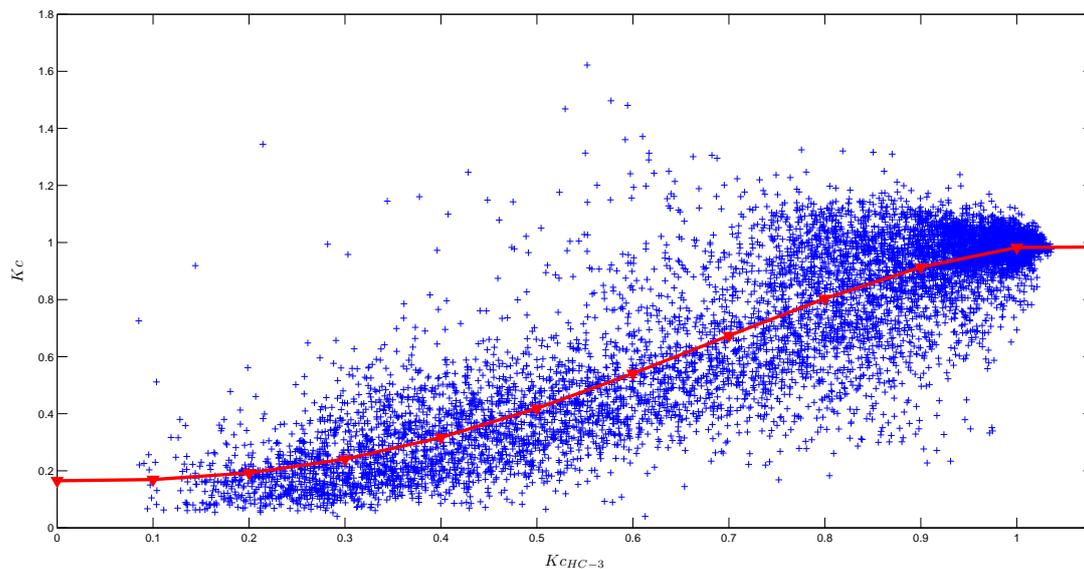


FIGURE 3.4 – Nuage de points entre K_{cHC3} et K_c , le résultat de la régression polynomiale d'ordre 4 est représenté par la courbe. Données impliquant l'année 2011 uniquement.

de point. D'autres solutions sont également proposées dans la bibliographie [Sch+11] [VBP13]. Cette correction établie à partir de l'année 2011 permet d'améliorer les performances d'HC-3 sur l'année 2012 comme le montre le tableau 3.3 où tous les scores sont améliorés par la version calibrée.

3.3.1.1 Localisation des zones d'intérêts de l'image

Nous cherchons dans un premier temps à identifier les zones d'intérêt de la carte Hc-3 en fonction de l'horizon de prévision. A chaque pixel (i,j) de la carte HC-3, correspond une série temporelle $HC3_{i,j}(t)$ donnant les valeurs de GHI estimées en ce point par HC-3 sur une période donnée. Nous calculons un coefficient de corrélation $r_{GHI,HC3_{i,j}}(\tau)$ entre cette série temporelle et la série temporelle de mesure $GHI(t)$ en suivant l'équation 3.4. τ représente

TABLE 3.3 – Performances d'HC-3 sur le site du SIRTA, mesures faites pour un angle solaire zénithal $< 85^\circ$

Année	Points de mesures	RMSE en W/m^2 (rRMSE)	MBE en W/m^2	R^2
2011	15128	82 (26%)	-18	0.89
2012	15539	88(30%)	-22	0.87
2012 (calibrée)	15539	79.7 (27%)	-8	0.89

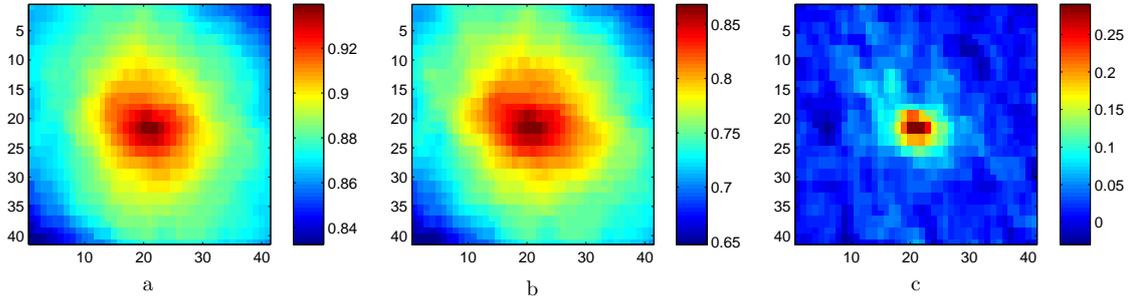


FIGURE 3.5 – Sensibilité de la carte de corrélation en fonction de la variable utilisée; (a) : Corrélation sur le GHI; (b) : Corrélation sur l'index de ciel clair Kc ; (c) : Corrélation sur le différentiel de KC ;

le décalage temporel entre les deux séries temporelles exprimé en valeur entière du pas des données HC-3, i.e. 15 minutes. La carte de corrélation correspond à la carte des coefficients de corrélation obtenus pour chacun des pixels. Elle permet de localiser les pixels les plus corrélés à la mesure faite par le pyranomètre (GHI) en fonction du décalage temporel. Lorsque que l'on calcule ces coefficients avec $\tau = 0$, le maximum de corrélation se situe bien au centre de la carte (Fig. 3.5.a). En effet la carte est par construction centrée sur le pixel couvrant le site dont la position a été obtenue par géo-référencement. Le maximum de corrélation obtenu est de 0.94 ce qui confirme les bonnes performances d'HC-3 en estimation. Toutefois, une large région autour du site est également très corrélée à la mesure sur cette carte. Dans le but de fournir une meilleure localisation graphique de l'information, deux autres cartes de corrélation ont été créées. L'une est calculée à partir de l'index de ciel clair Kc (Fig. 3.5.b), la seconde à partir du gradient temporel de l'indice de ciel clair Kc , obtenu par différence finie Δ_{Kc} (Fig. 3.5.c, eq.3.5). Si la valeur absolue des coefficients de corrélation diminue de façon importante en utilisant ces variables, elles permettent néanmoins une meilleure distinction du pixel d'intérêt dans la figure 3.5.c utilisant Δ_{Kc} . L'observation des variations du GHI semble plus discriminante que celle du GHI lui-même. C'est cette variable qui sera utilisée par la suite pour localiser les pixels d'intérêt pour la prévision. L'utilisation de carte de corrélation pour $\tau = 0$ a permis d'une part de vérifier la pertinence de ce descripteur pour retrouver les pixels d'intérêt, mais aussi d'affiner la méthode en choisissant une variable plus sensible à l'étalement spatial observé.

$$r_{GHI,HC3_{i,j}}(\tau) = \frac{\sum_{t=1}^N (HC3_{i,j}(t) - \overline{HC3}_{i,j})(GHI(t-\tau) - \overline{GHI})}{\sqrt{\sum_{t=1}^N (HC3_{i,j}(t) - \overline{HC3}_{i,j})^2} \cdot \sqrt{\sum_{t=1}^N (GHI(t-\tau) - \overline{GHI})^2}} \quad (3.4)$$

où $\overline{HC3}_{i,j}$ représente la moyenne temporelle de la série $HC3_{i,j}(t)$.

$$\Delta_{Kc} = Kc(t+1) - Kc(t), t = 2 \dots N \quad (3.5)$$

Nous calculons à présent les cartes de corrélations en faisant varier τ de 0 à 4 afin de localiser les zones d'intérêt pour des prévisions allant de 15 à 60 minutes. Nous utiliserons par la suite la notation h pour indiquer l'horizon de prévision normalisé par la résolution temporelle

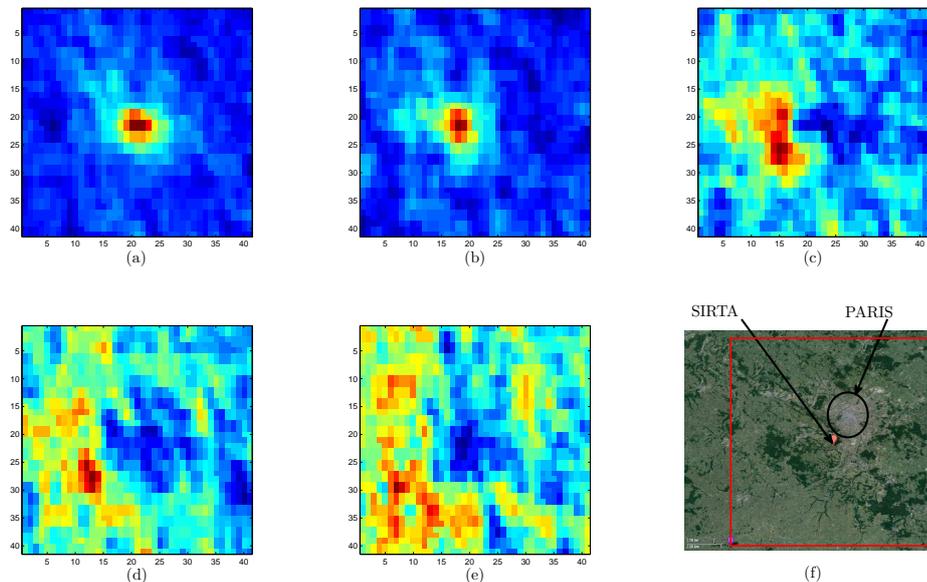


FIGURE 3.6 – Cartes de corrélations utilisant les séries temporelles de Δ_{K_c} : (a) Décalage 0 ; (b) Décalage 1 (15 min) ; (c) Décalage 2 (30 min) ; (d) Décalage 3 (45 min) ; (e) Décalage 4 (60 min) ; (f) Région équivalente sur google map

des cartes HC-3. Un horizon $h=2$ correspond à une prévision à 30 minutes. Les cartes ainsi calculées sont présentées en figure 3.6. Ces cartes permettent d'observer la position des pixels les plus intéressants pour chacun des horizons de prévision. Les pixels les plus intéressants restent très proches du site pour des horizons faibles (15-min) et s'éloignent progressivement lorsque le décalage augmente. Ce phénomène traduit l'effet d'advection caractérisant le mouvement moyen des nuages arrivant sur le site. Le décalage temporel que l'on souhaite combler en faisant de la prévision peut être ici traduit en décalage spatial sous hypothèse de conservation des masses nuageuses. La position des pixels d'intérêt ainsi définie sous entend un vecteur moyen de vent sur la période utilisée pour le calcul de ces cartes (l'année 2011). On remarque sur ces cartes une domination de la direction ouest correspondant en effet à la direction privilégiée des vents sur la région du SIRTA. La "zone d'intérêt" regroupant les pixels de forte corrélation est relativement petite et dense pour des horizons faibles, puis s'élargit entre le nord ouest et le sud-ouest lorsque h augmente. Ce sont ces zones qui seront prises en compte par la suite dans le modèle de prévision linéaire.

3.3.2 Les modèles Autoregressifs (AR)

On se propose d'utiliser un modèle de régression afin de profiter de l'information mise en évidence par les cartes de corrélation dans un modèle de prévision. Parmi les différents algorithmes de régression, nous avons choisi d'utiliser un modèle autorégressif (AR). Les modèles AR sont considérés comme des modèles de référence pour la prévision dans la bibliographie

[Aka69]. Ils possèdent en effet de multiples avantages :

- Implémentation simple
- Temps de calcul très rapide
- Résultats robustes

Il existe bien entendu des algorithmes de régression présentant de meilleures performances finales, mais l'utilisation de modèles de référence tels que les AR permet une meilleure mise en valeur de la contribution de chaque entrée. Par ailleurs, alors que les entrées utilisées au chapitre 2 étaient hétérogènes, l'utilisation des cartes HC-3 assure un jeu de variables d'entrées limité à des valeurs homogènes. Dans ce cadre, le modèle AR semble une solution plus appropriée au but recherché : établir un premier pas dans l'utilisation des cartes HC-3 pour la prévision de GHI. Un modèle autoregressif classique pour la prévision d'une variable temporelle X_t s'écrit sous la forme :

$$X(t) = \sum_{m=1}^M a_m \cdot X(t-m) + a_0 \quad (3.6)$$

La principale difficulté lors de la construction d'un tel modèle réside dans le choix de l'ordre M [Aka69] du processus AR, correspondant au nombre M de décalages du signal X à prendre en compte. Cet ordre correspond à la longueur de l'historique que l'on intègre dans le calcul de la prévision. Dans le cadre de ce chapitre, l'objectif est d'étendre ce modèle afin d'intégrer non seulement un historique de mesure, mais aussi des valeurs provenant de la carte HC-3. Dans un premier temps, nous développerons un modèle AR simple sur les mesures au sol. Ce modèle permettra d'aborder les problèmes de construction des modélisations ARs et servira de référence à d'autres modèles plus complexes par la suite. Nous cherchons donc à construire un modèle AR appliqué aux mesures qui s'écrit :

$$\hat{Kc}(t) = \sum_{m=1}^M a_m \cdot Kc(t-m) + a_0 \quad (3.7)$$

Nous utiliserons le modèle ciel clair ESRA pour effectuer la conversion de l'index de ciel clair Kc en GHI. La résolution d'un modèle AR est un problème classique [Vai07] et bien connu. Elle se résume dans les étapes suivantes :

Dans la recherche d'une solution optimale pour les coefficients du modèle, on cherche à minimiser, en fonction de chacun des coefficients, l'erreur quadratique moyenne (EQM) :

$$\begin{aligned} EQM &= E[(Kc(t) - \hat{Kc}(t))^2] \\ &= E[(Kc(t) - \sum_{m=1}^M a_m \cdot Kc(t-m) + a_0)^2] \end{aligned} \quad (3.8)$$

Pour trouver le minimum de cette fonction, on cherche le point d'annulation de la dérivée partielle pour chaque coefficient a_m : $\frac{\partial EQM}{\partial a_m}$:

$$\begin{aligned} \frac{\partial EQM}{\partial a_m} &= 0 \\ E[Kc(t-m)(2(Kc(t) - \sum_{k=1}^M a_k Kc(t-k)))] &= 0 \\ 2E[(Kc(t) - \sum_{k=1}^M a_k Kc(t-k))Kc(t-m)] &= 0 \\ E[Kc(t)Kc(t-m)] - \sum_{k=1}^M a_k E[Kc(t-k)Kc(t-m)] &= 0 \end{aligned} \quad (3.9)$$

En utilisant le changement de variable suivant :

$$T = t - k \quad (3.10)$$

On obtient :

$$E[Kc(t)Kc(t-m)] - \sum_{k=1}^M a_k E[Kc(T)Kc(T-(m-k))] = 0 \quad (3.11)$$

Or par définition, le coefficient d'autocorrélation $r_{Kc}(m)$ pour un décalage de m peut s'écrire :

$$r_{Kc}(m) = E[Kc(t)Kc(y-m)] \quad (3.12)$$

ce qui conduit au système d'équation suivant :

$$r_{Kc}(m) - \sum_{k=1}^M a_k r_{Kc}(m-k) = 0 \quad m = 1..M \quad (3.13)$$

On peut représenter ce système de façon matricielle sous la forme suivante :

$$R_M \times a = R_m \quad (3.14)$$

où R_M est la matrice d'autocorrélation de taille $N \times M$ avec $N \geq M$, \mathbf{a} le vecteur des coefficients que l'on cherche, et R_m le vecteur des coefficients d'autocorrélation de $r_{Kc}(m)$ de taille $M \times 1$. Pour illustration, ce système est écrit en équation 3.15 pour un ordre $M=3$:

$$\begin{bmatrix} r_{Kc}(0) & r_{Kc}(1) & r_{Kc}(2) \\ r_{Kc}(1) & r_{Kc}(0) & r_{Kc}(1) \\ r_{Kc}(2) & r_{Kc}(1) & r_{Kc}(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ a_3 \end{bmatrix} = \begin{bmatrix} r_{Kc}(1) \\ r_{Kc}(2) \\ r_{Kc}(3) \end{bmatrix} \quad (3.15)$$

Ainsi, dès lors que la matrice d'autocorrélation est estimée, il est possible de trouver les coefficients a_m optimaux par inversion :

$$a = R_M^{-1} \times R_m \quad (3.16)$$

Dans le cas pratique, les coefficients de la matrice d'autocorrélation sont obtenus en suivant l'équation suivante :

$$r_{Kc}(m) = \sum_{i=1}^N Kc(i)Kc(i-m) \quad (3.17)$$

pour un signal centré et normé où N est la taille de la base de donnée disponible. Puisque la matrice d'autocorrélation est nécessaire à la résolution du modèle, il apparaît ici clairement que le modèle AR nécessite une "base d'apprentissage" pour estimer cette matrice.

3.3.2.1 Choix de l'ordre temporel

L'ordre temporel M du modèle AR est un paramètre majeur et de nombreuses approches sont développées dans la littérature pour optimiser ce choix. Toutefois, dans le contexte de cette thèse nous avons sélectionné deux approches très simples : la fonction d'autocorrélation partielle et la RMSE de la prévision résultante. L'autocorrélation partielle d'ordre k PACF(k)

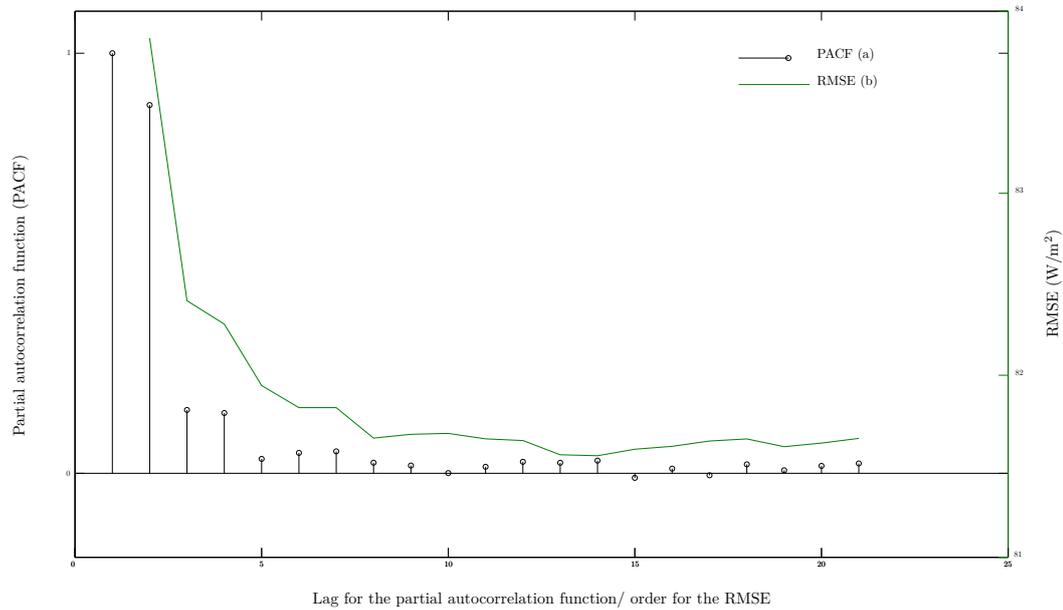


FIGURE 3.7 – (a) Autocorrélation partielle (PACF) ; (b) RMSE sur le résultat de la prévision

correspond à la corrélation entre $Kc(t)$ et $Kc(t-k)$ obtenue lorsque l'influence des variables $Kc(t-k+i)$, $i < k$ est retirée. Etant donnée une série temporelle X_t , la PACF au rang k s'écrit :

$$PACF(k) = Cor(X_{t+k} - P_{t,k}(X_{t+k}), X_t - P_{t,k}(X_t)) \text{ pour } k > 2 \quad (3.18)$$

où $P_{t,k}(X)$ désigne la projection de X sur l'espace engendré par $X_{t+1}, \dots, X_{t+k-1}$. La RMSE est quant à elle obtenue en appliquant le modèle sur l'année 2011 pour $h=1$. La figure 3.7 présente l'évolution de ces deux variables en fonction de l'ordre choisi. Les deux critères choisis présentent une évolution similaire. Le gain principal se situe entre l'ordre 1 et l'ordre 2 où l'on observe une chute de la RMSE. Ensuite, les deux fonctions montrent un palier pour les ordres 2 et 3 et se stabilisent à partir de l'ordre 6. Dans un objectif de limitation des données d'entrées, il serait envisageable de choisir un ordre 3, toutefois, afin d'optimiser les résultats, nous choisirons ici l'ordre 5 pour la suite de l'étude. L'ordre temporel de l'étude est donc choisi uniquement sur étude de la série temporelle de mesure au sol, et non sur l'étude des cartes HC-3. C'est un choix justifié puisque les cartes HC-3 utilisent la même variable (GHI) dans une zone relativement proche du site de mesure.

3.3.3 Définition des modèles testés

Afin de pouvoir estimer plus précisément l'apport de chaque nouvelle information, plusieurs modèles AR ont été construits. Chaque nouveau modèle inclut une information supplémentaire en entrée afin d'évaluer l'impact dans les résultats. Les modèles mis en œuvre sont décrits ci-dessous :

- Le modèle de persistance P est la référence décrite dans le chapitre 1.

$$P : \hat{K}c^P(t+h) = Kc(t), t = 1..T \quad (3.19)$$

- Le modèle A_t est un modèle de prévision linéaire simple sur la série temporelle Kc_t . Il permet d'évaluer l'intérêt du modèle de prévision linéaire et donc de l'historique de mesure du capteur. Il s'écrit :

$$\hat{K}c^{A_t}(t+h) = \sum_{m=0}^M (\alpha_m \cdot Kc(t-m)) + \gamma \quad (3.20)$$

- Le modèle A_{st} est le premier modèle à prendre en compte l'information provenant d'HC-3. Il s'agit d'un modèle de prévision linéaire prenant en compte à la fois la série temporelle de mesure, mais aussi les 9 pixels HC-3 environnant le site de mesure. La localisation des pixels est donc toujours la même quel que soit l'horizon de prévision. Ce modèle permet d'établir une référence pour les modèles spatio-temporels. Il s'écrit :

$$\hat{K}c^{A_{st}}(t+h) = \sum_{m=0}^M (\alpha_m \cdot Kc(t-m)) + \sum_{p=1}^9 \sum_{m=0}^M (\beta_{p,m} \cdot Kc_{HC3}(t-m,p)) + \gamma \quad (3.21)$$

- Enfin A_{st2} prend également en compte l'information provenant d'HC-3, mais de façon guidée par les cartes de corrélation présentées dans la section précédente 3.3.1.1. Cela permet d'évaluer l'intérêt de cette méthode de sélection des zones par rapport à la zone fixe du modèle A_{st} . Ainsi le choix des pixels inclus dans la prévision dépend de l'horizon de prévision visé. Toutefois, afin de garder un modèle comparable au modèle A_{st} , le nombre de pixels est restreint à 9. Ce choix implique quelques modifications : selon l'horizon de prévision, il est parfois difficile de contenir les zones intéressantes selon la carte de corrélation dans 9 pixels. Aussi, les zones d'intérêt ont-elles été moyennées spatialement selon l'horizon de prévision afin de couvrir de larges zones tout en ne gardant que 9 points *in fine*. Les zones ainsi définies sont présentées en figure 3.8. Sur cette figure, les entrées sélectionnées en fonction de l'horizon de prévision sont présentées dans un carré rouge. La carte présentée en arrière plan est la carte de corrélation entre les mesures au sol et la carte HC-3 décrite en section 3.3.1.1. Le décalage appliqué pour la corrélation est ici directement lié à l'horizon de prévision h . Pour un horizon $h=2$ (30 min), nous choisirons les entrées sur la carte de corrélation calculée pour τ égal à 2.

3.3.4 Résultats

Les modèles présentés précédemment ont été entraînés (matrices de corrélations, définition des coefficients du modèle de prévision linéaire) sur l'année 2011 et testés sur l'année 2012. Les données provenant du pyranomètre ont été moyennées préalablement de façon à atteindre la résolution temporelle des cartes HC-3. Les horizons de prévision testés sont compris entre 15 et 60 minutes ($h \in [1;4]$). Chaque modèle a donc été entraîné 4 fois pour s'adapter aux horizons de prévision. Les résultats obtenus sont présentés ci dessous, avec les critères de qualités sélectionnées au chapitre 1.

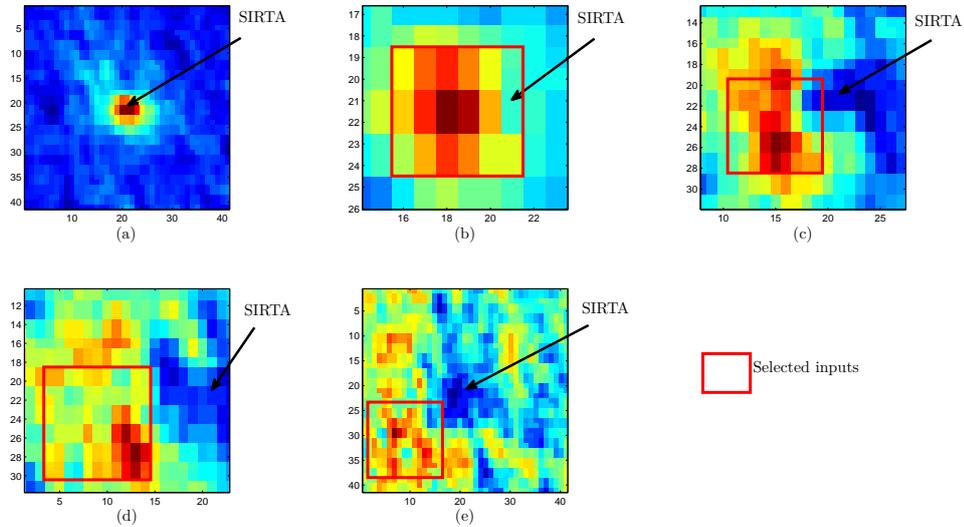


FIGURE 3.8 – Localisation des entrées sélectionnées en fonction de l’horizon de prévision sur les cartes de corrélation de Δ_{Kc} . L’horizon de prévision h est directement liée au décalage appliqué aux cartes de corrélation : (a) $h=0$, estimation ; (b) $h=15$ min, filtrage moyen 2×2 ; (c) $h=30$ min, filtrage moyen 3×3 ; (d) $h=45$ min, filtrage moyen 4×4 ; (e) $h=60$ min, filtrage moyen 5×5

3.3.4.1 Analyse quantitative

Les tableaux 3.4 à 3.6 regroupent les critères de qualité obtenus pour chacun des modèles en fonction de l’horizon de prévision. Nous donnons également une présentation plus visuelle de l’évolution de la RMSE en fonction de l’horizon de prévision et du modèle en figure 3.9. Ces résultats peuvent être analysés en fonction de l’apport de chaque nouvelle information :

Apport du modèle linéaire : La comparaison des résultats du modèle A_t à ceux du modèle de persistance permet d’évaluer les performances initiales du processus du prévision linéaire. Le modèle donne des résultats bien meilleurs que la persistance pour tous les horizons de prévision (Fig. 3.9). Il est toutefois important de remarquer que l’amélioration est d’autant plus importante que l’horizon de prévision est grand. Cet effet est principalement du à l’hypothèse de persistance : plus l’horizon de prévision est grand, moins l’hypothèse est vérifiée. Par conséquent, les performances du modèle P se dégradent rapidement avec l’horizon de prévision. En revanche cette dégradation n’est pas aussi importante avec le modèle A_{st} qui lui s’appuie sur une hypothèse de prévision linéaire de la mesure en fonction de son historique.

Apport de la carte HC-3 simple : Il s’agit cette fois-ci de comparer les modèles A_t et A_{st} qui ne diffèrent l’un de l’autre que par l’intégration des données spatio-temporelles de la carte Hc3. Nous observons alors une amélioration des résultats plus importante pour les horizons

de prévisions faibles. L'information spatiale ajoutée est extraite des 9 pixels environnant le site de mesure, soit une zone de plus ou moins 10 km^2 . Cette zone apporte de l'information utile pour une prévision à très court terme (jusqu'à 30 minutes), en revanche elle semble trop restreinte pour apporter les mêmes gains de performances à plus long terme. En d'autres termes, les nuages susceptibles d'affecter le site à des horizons plus lointain sont situés en dehors de la zone d'influence de la pondération linéaire. Les résultats tendent donc vers le modèle A_t lorsque l'horizon de prévision augmente.

Apport de la méthode de sélection des entrées Hc3 : Enfin la dernière information apportée concerne la localisation des pixels d'influence en fonction de l'horizon de prévision. Les performances à très court terme ($h=1$) de A_{st2} sont assez similaires à celles de A_{st} . En effet, à cet horizon, la zone sélectionnée par rapport à la carte de corrélation est très semblable à la zone de référence prise pour le modèle A_{st} . En revanche, l'amélioration est plus significative lorsque l'horizon de prévision augmente. Cela met en évidence l'importance du choix judicieux des entrées spatiales à prendre en compte dans le modèle de prévision. Ce choix dépendant de l'horizon permet de stabiliser le gain du modèle spatio-temporel par rapport au modèle A_t pour tout les horizons testés. En effet, nous observons une amélioration quasiment constante de 11 W/m^2 entre les modèles A_t et A_{st2} .

TABLE 3.4 – RMSE (W/m^2) en fonction de l'horizon et du modèle

Horizon \ Model	P0	At	Ast	Ast2
15 min	94	86	78	75
30 min	118	100	94	87
45 min	130	107	103	94
60 min	140	111	108	99

TABLE 3.5 – R^2 en fonction de l'horizon et du modèle

Horizon \ Model	P0	At	Ast	Ast2
15 min	0.84	0.87	0.89	0.90
30 min	0.75	0.82	0.84	0.87
45 min	0.70	0.80	0.81	0.84
60 min	0.66	0.78	0.79	0.83

3.3.4.2 Analyse qualitative

Pour une analyse plus détaillée, nous observons les résultats obtenus en fonction des classes définies au chapitre 1. Le tableau 3.9 présentent les résultats de RMSE en fonction de l'horizon de prévision et du modèle utilisé pour chaque classe. Si de manière générale on observe bien la même tendance que pour les scores globaux, quelques points particuliers sont toutefois à noter :

TABLE 3.6 – MBE

Horizon \ Model	P0	At	Ast	Ast2
15 min	0	-2	-2	-3
30 min	0	-3	-2	-5
45 min	1	-3	-3	-6
60 min	2	-4	-3	-5

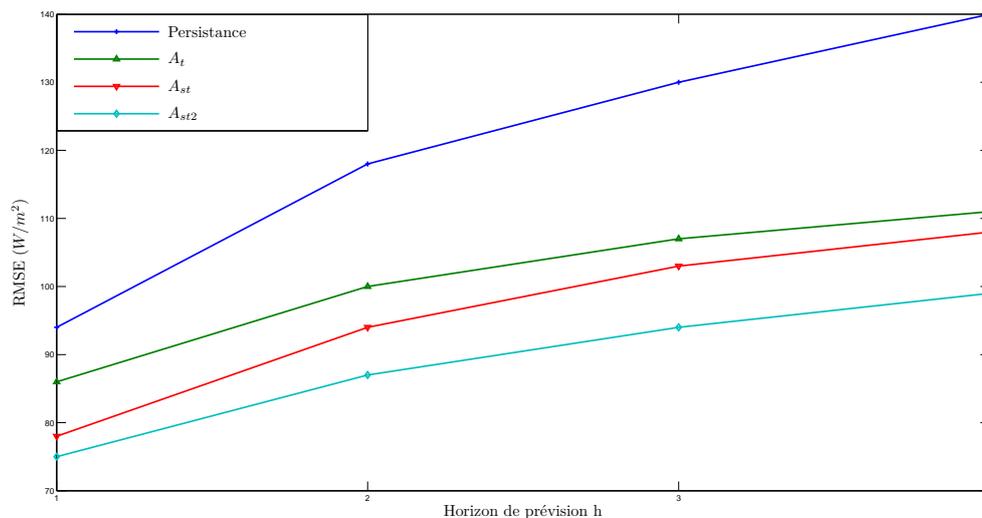


FIGURE 3.9 – Evolution de la RMSE en fonction de l'horizon de prévision pour chacun des 4 modèles testés.

Les classes peu variables : Les scores de la classe I (peu variable) sont toujours meilleurs. Cette classe compte en effet les jours les plus faciles à prévoir puisqu'ils ne présentent que peu de variations. A l'intérieur de cette classe, les meilleurs scores sont obtenus lorsque le niveau moyen de K_c est élevé (CI), ce qui correspond aux journées dites "ciel clair". Ces journées ne présentent aucun phénomène perturbateur d'où les bonnes performances de l'ensemble des modèles. En revanche, si la classe A1 obtient de meilleurs résultats que B1, c'est uniquement grâce à une plus faible valeur moyenne de K_c impactant intrinsèquement la RMSE.

Gains en fonction de la classe : Alors que les résultats globaux portaient à croire que les performances s'améliorent toujours avec la complexité du modèle, ce n'est pas toujours le cas selon la classe considérée. De manière générale, le modèle de persistance est toujours plus performant lorsqu'il y a peu de variations dans la journée (classe I). L'apport de la complexité des modèles se fait principalement sur les journées à forte variabilité (classe III) où les limites du modèle de persistance sont atteintes. Pour un horizon de prévision de 60 minutes par exemple, le gain est de $-5 W/m^2$ pour la classe ciel clair (CI) entre le modèle de persistance et

TABLE 3.7 – Résultats par classes : RMSE

Horizon \ Model		P			At			Ast			Ast2		
		A	B	C	A	B	C	A	B	C	A	B	C
15 min	I	24	29	13	33	35	17	32	26	20	32	25	18
	II	84	98	85	132	171	77	104	132	73	106	138	72
	III	136	163	241	123	241	207	126	188	181	113	191	192
30 min	I	33	45	17	42	53	22	41	40	26	38	41	19
	II	108	124	100	137	177	86	114	140	82	112	142	80
	III	146	193	223	130	243	179	136	197	183	115	196	166
45 min	I	39	59	20	47	64	26	47	50	29	43	57	23
	II	114	138	101	137	181	88	118	146	87	109	142	83
	III	159	206	201	141	242	179	145	202	170	120	193	179
60 min	I	43	72	23	53	76	32	52	64	33	47	73	28
	II	119	147	105	137	182	90	119	152	89	111	142	86
	III	166	212	204	147	237	171	146	202	170	125	188	168

le modèle de A_{st2} alors que ce gain est de $36 W/m^2$ pour la classe CIII très variable. Or, nous savons d'après le chapitre I que les journées variables sont majoritaires sur le site du SIRTa, ce qui explique l'impact de ces effets sur le score global.

Gain en fonction de l'horizon de prévision : Si les améliorations les plus importantes sont bien observées pour la classe C, nous nous intéressons également au gain obtenu en fonction de l'horizon de prévision. Il apparaît que le gain le plus important est obtenu pour un horizon de prévision de 30 minutes ($57 W/m^2$) entre les modèles P et A_{st2} , (Tableau 3.7). Au delà de cet horizon, les résultats sont toujours meilleurs pour le modèle A_{st2} , mais le gain obtenu est moins important. Cette information est très certainement liée à l'étalement de la zone d'intérêt sur les cartes de corrélations lorsque l'horizon de prévision augmente. Cet étalement dilue l'information utile.

La figure 3.10 présente un profil journalier par classe pour un horizon de prévision de 30 minutes. Les courbes observées confirment les résultats précédemment décrits d'après l'analyse de tableau de RMSE. En revanche, ces courbes montrent que le gain obtenu pour les journées à très forte variabilité s'expliquent plus par un effet de lissage des variations extrêmes que par une réelle anticipation de ces fortes variations.

3.3.5 Conclusion

Cette section présente une première étape pour la prévision du GHI en utilisant les cartes d'estimations HC-3. L'utilisation de cartes de corrélation entre les pixels de HC-3 et la mesure au sol a permis une fois pour toute de localiser la zone la plus propice à établir un lien de régression linéaire avec la prévision pour un horizon de temps donné. Les zones d'intérêt

TABLE 3.8 – Résultats par classes : MBE

Horizon	Model	P			At			Ast			Ast2		
		A	B	C	A	B	C	A	B	C	A	B	C
15 min	I	0	0	0	-3	1	3	-1	1	4	-1	1	2
	II	0	0	0	7	16	2	5	11	3	6	12	1
	III	2	0	0	-4	39	5	-5	29	14	-6	30	13
30 min	I	0	1	0.5	-5	2	6	-3	1	6	-2	2	3
	II	0	0	0	5	16	4	3	10	5	3	11	3
	III	3	1	0.6	-6	38	7	-7	27	19	-7	27	14
45 min	I	0	2	2	-6	2	7	-5	31	7	-4	42	4
	II	1	1	0	3	15	5	0	9	5	1	9	3
	III	5	2	2	-7	37	9	-	25	24	-8	23	23
60 min	I	0	4	3	-8	3	10	-6	3	9	-4	4	6
	II	1	2	2	0	15	7	-1	9	7	-1	8	5
	III	7	4	3	-	36	12	-	24	27	-8	22	24

TABLE 3.9 – Résultats par classes : Coefficient de détermination

Horizon	Model	P			At			Ast			Ast2		
		A	B	C	A	B	C	A	B	C	A	B	C
15 min	I	93	99	99	88	99	99	89	99	99	89	99	99
	II	82	93	96	59	78	97	74	87	97	73	86	97
	III	65	81	73	71	61	80	70	76	85	75	75	83
30 min	I	88	98	99	83	98	99	83	99	99	84	99	99
	II	72	88	95	55	77	96	69	85	96	70	85	97
	III	60	73	77	68	60	85	65	73	85	75	73	87
45 min	I	83	97	99	78	97	99	77	98	99	79	97	99
	II	68	85	95	55	76	96	66	84	96	71	84	96
	III	52	69	78	62	60	85	60	71	87	72	74	85
60 min	I	78	96	99	72	96	99	72	97	99	73	96	99
	II	65	83	94	54	75	96	65	82	96	69	84	96
	III	47	67	80	59	61	86	59	71	87	70	75	87

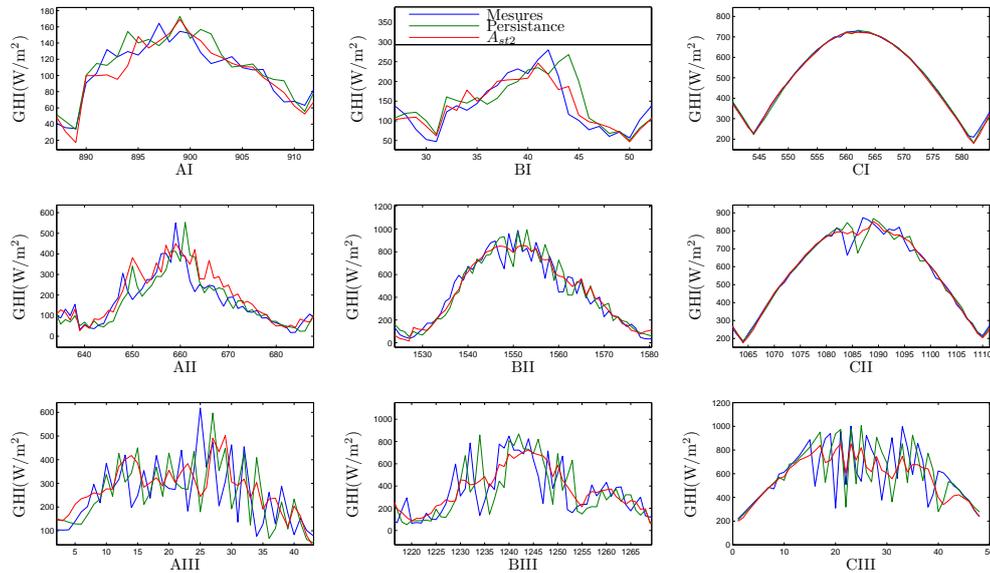


FIGURE 3.10 – Profil journaliers pour chacune des classes pour une prévision à 30 minutes

ainsi définies ont ensuite été intégrées dans un modèle de prévision linéaire combinant l'historique de mesure du capteur au sol et l'information spatio-temporelle des cartes HC-3. Le modèle construit a été entraîné sur l'année 2011 et testé sur l'année 2012. Plusieurs modèles de complexité croissante ont ensuite été développés afin de quantifier l'impact de chaque apport d'information sur les résultats de prévision pour des horizons allant de 15 minutes à 60 minutes. Les résultats obtenus confirment la pertinence du modèle spatio-temporel relativement à l'ensemble des autres modèles testés. Toutefois, l'intérêt d'un tel modèle est d'autant plus important que l'horizon de prévision est grand. D'après l'analyse détaillée des critères de qualité obtenus, les principales améliorations du modèle sont observées pour les journées présentant une variabilité importante où les limites de l'hypothèse de persistance sont atteintes. En revanche, d'après l'observation des profils journaliers de prévision, ces améliorations sont plus souvent dues à un effet de lissage qu'à une réelle anticipation des variations. Cela constitue la limite d'un modèle construit sur des zones d'intérêt moyennes. Afin d'obtenir des résultats plus proches des variations réellement observées, nous proposons une nouvelle approche permettant la définition de zone d'intérêt plus localisée dans la section suivante. On utilisera la méthode Ast2 définie dans cette section comme référence, cependant, rappelons que cette méthode nécessite une mesure du GHI au sol en temps réel.

3.4 Prévision par calcul de champ de vecteur

Il s'agit dans cette section d'utiliser la résolution spatio-temporelle fine de l'imagerie satellitaire pour établir un modèle de prévision en fonction de la situation météorologique en cours. Dans le modèle précédant (*Ast2*), les zones d'intérêt choisies comme entrées pour le modèle de prévision linéaire étaient dépendantes de l'horizon de prévision, mais pas des instants considérés. Dans cette nouvelle approche, nous tenterons d'estimer la direction et la vitesse des vents en action à l'aide d'un algorithme de traitement d'images afin de redéfinir pour chaque prévision une nouvelle zone d'intérêt. Cette méthode se rapproche ainsi des méthodes présentées dans la littérature avec une étape de "prévision" plus classique. La méthodologie employée est décrite en section 3.4.1.

3.4.1 Méthodologie

La méthode proposée ici se présente en trois étapes clés :

- Obtention d'un vecteur de déplacement ;
- Définition d'une Zone d'Intérêt (ZI) ;
- Interprétation de la ZI pour la prévision.

3.4.1.1 Le calcul du champs de vecteurs

La première étape de cet algorithme consiste à définir un vecteur de déplacement représentatif du mouvement des masses nuageuses de la scène observée. Plusieurs solutions sont alors envisageables :

1. Les modèles météorologiques : ils fournissent des prévisions de champs de vecteurs en fonction de l'altitude.
2. Mesure sur site : le site du SIRTA est également équipé d'instruments de mesure de vitesse et de direction des vent.
3. Imagerie satellitaire : il est possible d'obtenir des champs de vecteurs en utilisant des techniques de traitement d'image sur les séries temporelles d'images satellitaires.

Les deux premières options ont l'avantage de fournir directement un vecteur de vent sans étapes de calcul intermédiaires. En revanche elles requièrent une information supplémentaire sur l'altitude des nuages afin de choisir la vitesse correspondante dans le profil de vent donné. En utilisant la troisième option, nous obtenons directement un vecteur représentatif du mouvement des masses nuageuses intéressantes, puisque les nuages sont les seuls éléments de contraste dont se sert l'algorithme pour calculer le vecteur. Pour ces raisons, nous avons choisi d'utiliser la troisième méthode. Par ailleurs, en restreignant les informations nécessaires en entrées aux seules cartes Hc3, nous développons une méthode facilement transposable sur d'autres sites sans installations supplémentaires.

De nombreuses méthodes de traitement d'images permettent le calcul de vecteur de déplacement d'objet entre deux images consécutives. Les méthodes les plus utilisées en traitement

d'image satellitaires sont les trois suivantes :

- Le flot optique [HS81]
- Vélocimétrie par image de particule [AW10]
- Le block matching [GWE09]

Nous avons retenu la méthode de *block matching* pour ses bons résultats démontrés [Ham+99] mais aussi pour la simplicité de ses paramètres. Nous montrerons par la suite que leur choix peut être fait d'après des considérations physiques des phénomènes observés.

Block matching Le *block matching* est un algorithme de traitement d'image permettant d'estimer un champ de vecteur de déplacement entre deux images consécutives. Supposons que l'on souhaite estimer le déplacement entre les images A et B (Fig.3.11). La première étape consiste à diviser chacune des images en blocs de taille $b \times b$ pixels. Il s'agit alors à retrouver pour chaque bloc de l'image A son correspondant dans l'image B, d'où l'appellation "block matching" ou correspondance de bloc. Une fois le bloc correspondant trouvé, le vecteur de déplacement est défini à partir des coordonnées respectives de chacun des blocs. Si les blocs définis ne se recouvrent jamais, le nombre de vecteurs ainsi obtenus est dépendant de la taille des blocs. En revanche, si nous permettons un certain recouvrement ρ lors de la définition des blocs, nous pouvons obtenir jusqu'à un vecteur par pixel de l'image. La difficulté de la méthode réside dans la recherche du bloc correspondant. Dans le cadre de cet algorithme, la notion de correspondance est exprimée par le coefficient de corrélation entre deux blocs de pixels. Aussi, s'agit-il de rechercher dans l'image B le bloc ayant le maximum de corrélation avec celui de l'image A dont nous cherchons le déplacement. Afin de limiter le temps de calcul et ne pas calculer de coefficient de corrélation pour l'ensemble des blocs à chaque itération, nous définissons un horizon de recherche hr . De ce fait, afin de trouver un correspondant au bloc de position (i, j) dans l'image A, nous n'effectuerons la recherche que dans une zone $(i \pm hr, j \pm hr)$. Cet algorithme possède donc trois paramètres à régler : b , ρ et hr permettant d'ajuster le compromis qualité/temps de calcul de la méthode :

- La taille des blocs \mathbf{b} : la taille des blocs doit être choisie en fonction de la taille moyenne des éléments stables de l'image. L'utilisation d'un coefficient de corrélation dans la méthode repose sur l'hypothèse de translation des objets de l'image sans modification de leur forme en présence éventuelle de bruit de mesure blanc. Dans le cas des nuages que l'on observe sur les cartes satellitaires, cette hypothèse n'est pas tout à fait vérifiée : avec une résolution temporelle de 15 minutes, les détails de bordures de nuage et même le relief des zones intérieures peuvent varier légèrement. En revanche la forme globale est relativement stable. Il est donc plus avantageux d'utiliser des blocs ayant une taille suffisante pour englober des formes larges et plus stables. Par ailleurs la taille des blocs est également conditionnée par la taille totale de l'image : à cause des effets de bords, la zone exploitable de l'image peut être extrêmement restreinte si b est très grand. D'après ces considérations, la taille des blocs a été fixée à $\mathbf{b=20}$ (près de 60km) suite à une analyse qualitative des résultats obtenus sur une série de tests.
- Le taux de recouvrement ρ : le taux de recouvrement permet d'ajuster la résolution du champ de vecteurs résultant. Avec un fort taux de recouvrement la résolution obtenue est très fine mais en contrepartie le temps de calcul est relativement long. Par ailleurs, la résolution du champ de vecteur est également liée à b , en d'autre terme, si la taille

du bloc est petite, la résolution du champ de vecteur sera naturellement fine également. Il s'agit donc de fixer une résolution souhaitée et d'ajuster ensuite le coefficient ρ par rapport à b afin d'obtenir le résultat souhaité. Dans le cas de l'observation de masses nuageuses, cette question est liée à la distance pour laquelle il est possible de considérer des caractéristiques homogènes pour les vitesses de vent. Nous avons retenu une valeur $\rho = 50\%$, équivalant à l'estimation d'un vecteur de déplacement tous les 10 pixels.

- L'horizon de recherche hr : l'horizon de recherche permet principalement de limiter le temps de calcul pour chaque itération. Il est fixé à partir du déplacement maximal attendu entre 2 images. Il est donc lié à la résolution spatio-temporelle de la série d'images, mais aussi à la vélocité des phénomènes observés. Dans le cas présent, un horizon hr de 10 pixels permet de suivre des nuages se déplaçant jusqu'à environ 140 km/h, ce qui couvre largement les conditions normales estimées. Cette restriction de la zone de recherche permet également d'éviter les résultats aberrants.

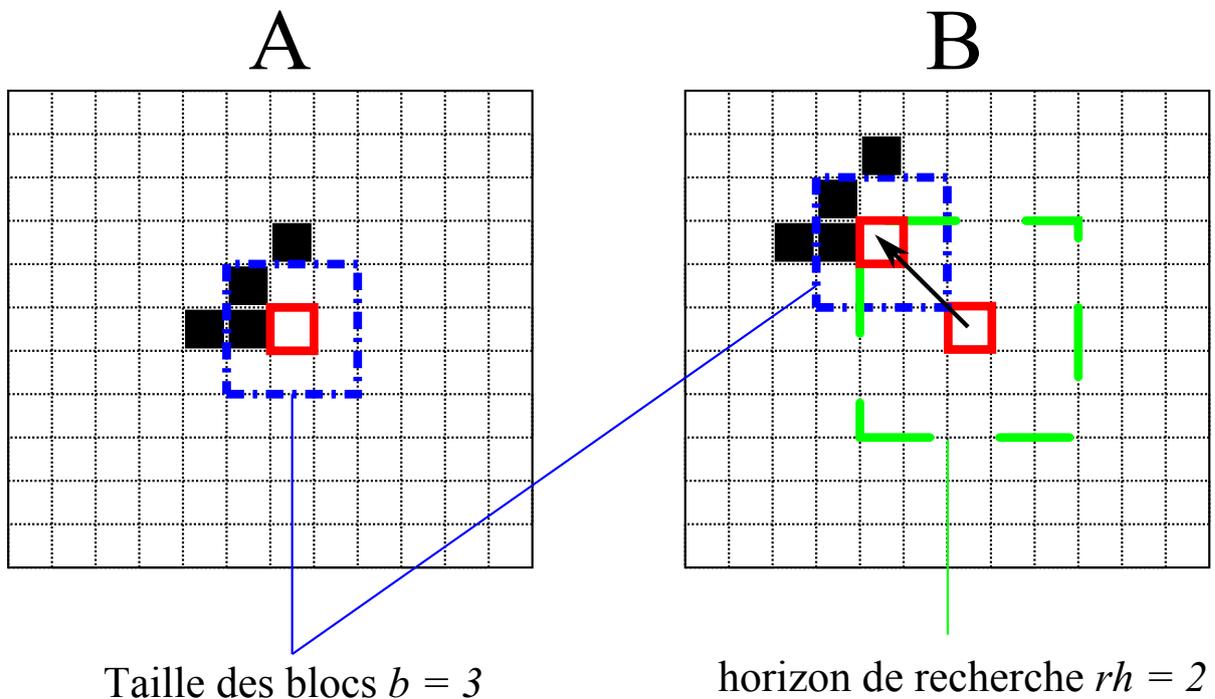


FIGURE 3.11 – Illustration de l'algorithme de block matching

L'application de cet algorithme sur les images HC-3 avec les paramètres $b=20$, $\rho=50\%$ et $hr=10$ conduit aux résultats illustrés en figure 3.12 et 3.13. Nous n'obtenons pas toujours les neufs vecteurs souhaités, lorsqu'il existe plusieurs maxima pour la recherche d'un bloc correspondant, nous choisissons de ne pas calculer le vecteur. Cela se produit généralement dans les zones homogènes ou lors de cas particuliers assez rares. Si les champs obtenus sont généralement homogènes (Fig.3.12), une certaine disparité est parfois observée parmi les vecteur calculés proches du site (Fig.3.13). Afin de définir un vecteur de déplacement unique pour la suite de l'étude, nous sélectionnons les 9 vecteurs les plus proches du site d'étude pour en prendre la moyenne. Cette méthode permet d'obtenir des résultats plus robustes aux éventuelles erreurs de l'algorithme.

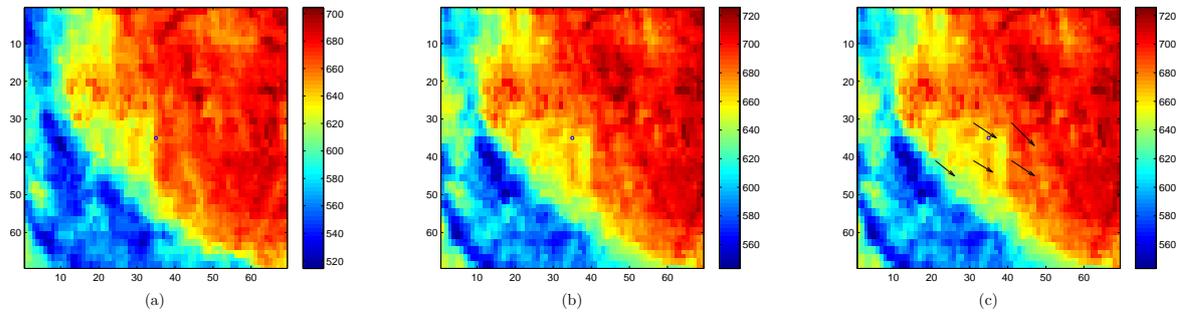


FIGURE 3.12 – Exemple 1 : application du block matching sur deux images HC-3 consécutives. (a) Image au temps $t-1$; (b) Image au temps t ; (c) image au temps t avec le champ de vecteur résultant.

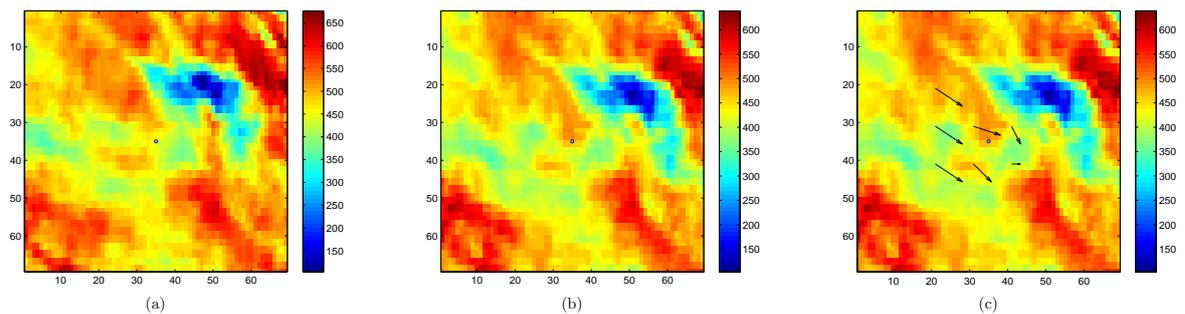


FIGURE 3.13 – Exemple 2 : application du block matching sur deux images HC-3 consécutives. (a) Image au temps $t-1$; (b) Image au temps t ; (c) image au temps t avec le champ de vecteur résultant.

3.4.1.2 Zone d'intérêt

En supposant des conditions idéales, à savoir :

- Estimation parfaite du mouvement
- Persistance du champ de vent jusqu'à l'horizon de prévision visé

alors la prévision consisterait simplement à prendre le pixel P_i se trouvant en amont du site en suivant le vecteur de mouvement (Fig.3.14). Cependant, ces hypothèses ne sont pas vérifiées. L'algorithme de *block matching* ne permet pas toujours de retrouver une estimation correcte du mouvement observé. L'estimation du mouvement est notamment difficile lorsqu'il y a peu d'objets stables dans l'image : lorsqu'il n'y a pas de masses nuageuses importantes par exemple. Par ailleurs, même si le vecteur calculé est réellement représentatif de mouvement observé au temps t , il peut y avoir une évolution de ce vecteur entre les temps t et $t+h$, où h représente l'horizon de prévision. Afin d'intégrer ces incertitudes dans la méthode, nous définissons une zone d'intérêt (ZI) autour du pixel P_i (Fig.3.14). La taille de la zone d'intérêt est établie à

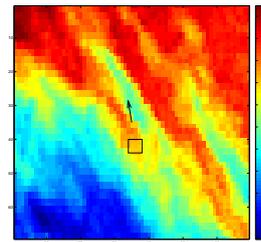
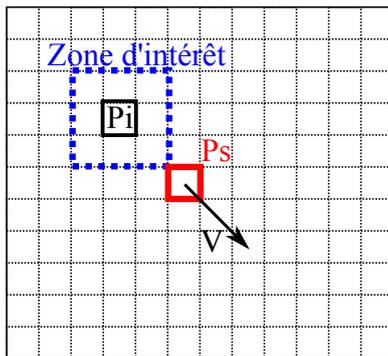


FIGURE 3.15 – Exemple d'application en cas réel

FIGURE 3.14 – La zone d'intérêt est définie à partir du pixel idéal P_i , en amont du site P_s par rapport au vecteur calculé \mathbf{V}

partir de l'équation :

$$t_{ZI} = 1 + 2 * h \quad (3.22)$$

Elle est dépendante de l'horizon de prévision. En effet, plus l'horizon de prévision est grand, moins l'hypothèse de persistance du vecteur calculé est probable. Il est également envisageable de faire varier la taille de la zone d'intérêt en fonction d'un indice de confiance sur le vecteur calculé. Un tel indice est toutefois relativement difficile à obtenir, nous ne retenons pas cette option dans un premier temps. La taille t_{ZI} ainsi définie prend bien un seul pixel Hc3 en condition d'estimation ($h=0$), mais peut considérer une zone de 9×9 pixels pour une prévision à 60 min ($h=4$).

TABLE 3.10 – Choix des paramètres de l'algorithme

Paramètre	Valeur
<i>Block matching</i>	
Taille bloc	tb = 20 pixels
Horizon de recherche	hr = 10 pixels
Taux de recouvrement	$\rho = 50\%$
<i>Zone d'intérêt</i>	
Taille de la zone	$t_{ZI} = 1 + 2 * h$

3.4.1.3 Interprétation de la zone d'intérêt

La zone d'intérêt extraite contient l'information utile à la prévision. En revanche nous disposons potentiellement d'autant de prévisions différentes qu'il y a de pixels dans la ZI. Il s'agit alors de choisir une méthode pour déterminer une prévision unique à partir de cet ensemble de valeurs. Plusieurs solutions simples sont alors envisageables :

- P_{id} : Ne considérer que le pixel central en supposant des conditions idéales de prévision
- P_{moy} : Utiliser la moyenne de l'ensemble des valeurs de la ZI.
- P_{med} : Utiliser la médiane de l'ensemble des valeurs de la ZI.

Ces trois solutions permettent de donner une prévision nominale du GHI pour un horizon de prévision donné. La zone d'intérêt peut être de plus exploitée afin de déterminer une prévision de type probabiliste complémentaire à la prévision déterministe, à partir de l'histogramme des valeurs de Kc sur la ZI. En particulier, la valeur minimale P_{min} et maximale P_{max} de cette zone peuvent être interprétées comme un encadrement de la prévision donnée. Nous disposons alors non seulement d'une prévision mais également d'informations quantitatives liées à la précision attendue de la prévision. Plus l'écart entre ces bornes extrêmes est grand, plus le risque d'erreur est important. Physiquement un tel écart reflète une zone à forte variabilité laissant présager une prévision au point Ps plus difficile.

3.4.2 Résultats

Afin de garder des résultats comparables à ceux des modèles de prévision linéaire présentés dans la section précédente, les modèles P_{id} , P_{moy} et P_{med} ont également été testés sur l'année 2012. Cela permet par ailleurs de bénéficier de la correction faite en section 3.3.1 sur l'année 2011. Les paramètres choisis pour ces résultats sont rappelés dans le tableau 3.10

3.4.2.1 Analyse quantitative

Les trois modèles définis en section 3.4.1.3 ont été testés pour des horizons de prévision allant de 15 minutes à 45 minutes. La carte HC-3 dont nous disposons pour l'étude a une taille de 69×69 pixels ce qui limite les horizons possibles pour cette méthode : les vecteurs calculés pointent des régions non couvertes lorsque l'horizon de prévision et la vitesse des vents sont

importants. Les résultats globaux sur l'année 2012 sont exposés dans les tableaux 3.11 à 3.13. Les performances du modèle Ast2 de la section précédente ont été ajoutés pour faciliter la comparaison.

Interprétation de la ZI : Il n'y a que très peu de différence entre les modèles P_{id} , P_{moy} et P_{med} . Quelle que soit la métrique considérée, les résultats tendent à se dissocier lorsque l'horizon de prévision augmente. Pour des horizons de prévision faibles (h petit), la manière d'interpréter la ZI n'a pas de réel impact sur le score de prévision. Dans ces conditions, la taille de la ZI t_{ZI} est très faible (3×3 pour $h=1$) ce qui explique en partie les scores observés. En revanche pour un horizon de prévision de 45 minutes, la ZI représente 7×7 pixels, ce qui semble suffisant pour engendrer une dissociation notable entre les trois modèles. Le modèle P_{moy} est alors le plus robuste des trois, confirmant l'intérêt de la prise en compte d'une zone d'intérêt par rapport au modèle idéal. Pour les raisons citées ci-dessus, nous choisissons les résultats du modèle P_{moy} pour la suite de l'analyse des résultats.

Comparaison avec le modèle Ast2 : De façon générale la méthode développée dans cette section aboutit à des résultats moins performants que ceux du modèle A_{st2} . La comparaison des résultats de RMSE doit se faire néanmoins en rapport avec la MBE observée (cf chap1). Contrairement au modèle A_{st2} , le modèle P_{moy} ne prend pas en compte les mesures passées du pyranomètre, il ne s'appuie que sur les estimations d'HC-3. Or ces estimations, même corrigées (section 3.3.1), présentent un biais négatif que l'on retrouve ici. Les meilleures performances du modèle A_{st2} sont principalement dues à cette différence majeure. D'ailleurs, lorsque l'horizon de prévision augmente et que l'influence de la persistance diminue, les résultats deviennent comparables entre les deux modèles toutes métriques confondues.

TABLE 3.11 – RMSE en fonction de l'horizon et du modèle

Horizon \ Modèle	P_{id}	P_{moy}	P_{med}	A_{st2}
15 min	86	86	86	75
30 min	93	92	92	87
45 min	100	97	98	94

TABLE 3.12 – MBE en fonction de l'horizon et du modèle

Horizon \ Modèle	P_{id}	P_{moy}	P_{med}	A_{st2}
15 min	-10	-9	-10	-3
30 min	-9	-8	-8	-5
45 min	-7	-6	-6	-6

TABLE 3.13 – R^2 en fonction de l'horizon et du modèle

Horizon \ Modèle	P_{id}	P_{moy}	P_{med}	A_{st2}
15 min	0.87	0.87	0.87	0.90
30 min	0.84	0.85	0.85	0.87
45 min	0.82	0.83	0.82	0.84

3.4.2.2 Analyse qualitative

Les résultats par classes de la méthode P_{moy} sont donnés dans les tableaux 3.14 à 3.16. Ces résultats font apparaître de nouveau les tendances relatives précédemment observées en section 3.3.4.2. Cependant les erreurs liées au biais d'Hc3 et l'absence de prise en compte de la persistance ont de lourdes conséquences sur les scores présentés. Les qualités du modèle présenté ici ne sont donc pas directement visibles à travers les métriques classiques utilisées jusqu'à présent. Observons alors les profils journaliers des différents modèles de prévision d'abord pour chacune des classes (Fig.3.16) puis pour une journée choisie pour sa forte variabilité (Fig.3.17). Ces figures présentent à la fois les courbes issues de la méthode développée dans cette section mais également la prévision du modèle A_{st2} pour comparaison. Pour faciliter les commentaires, nous nous appuyerons sur les courbes de la figure 3.17 pour plus de visibilité. Deux faits importants sont remarquables :

Indication de bornes extrêmes : Les bornes P_{min} et P_{max} déterminées en même temps que la prévision P_{moy} apportent une information supplémentaire à la prévision que ne peut fournir le modèle A_{st2} . Ces indications permettent de mieux comprendre les erreurs de prévision faites. Si la mesure est comprise entre ces bornes, alors la ZI contient bien l'information nécessaire à la prévision. Dans le cas contraire, soit la ZI a été mal positionnée (mauvais calcul de vecteur) soit elle ne permet pas de décrire les phénomènes à l'œuvre (limite de résolution spatiale du satellite). Dans un cadre applicatif, il peut être très intéressant de fournir une borne minimale à la prévision.

Absence de "l'effet de persistance" : Le modèle P_{moy} semble plus "en phase" avec les mesures que le modèle A_{st2} . Les variations brutales du GHI sont toutes anticipées par P_{moy} ou au moins par ses bornes ce qui confirme la présence de l'information "de prévision" dans la ZI. Afin d'illustrer de manière plus rigoureuse ce phénomène, observons la fonction d'intercorrélation entre la prévision et la mesure effective avec différents décalage temporels (Fig 3.18). Sur cette figure, les fonctions d'intercorrélation ont été tracées pour les modèles de persistance, P_{moy} et A_{st2} . Le modèle de persistance permet d'avoir une référence, il présente un maximum pour un décalage de -2 cohérent avec l'horizon de prévision $h=2$. Le score obtenu en 0 correspond au coefficient de corrélation entre la mesure et la prévision à l'horizon de prévision considéré. Si les scores des modèles A_{st2} et P_{moy} sont très similaires en 0, il est important de noter le décalage pour lequel le maximum est atteint. Le modèle P_{moy} atteint son maximum en 0, il est bien "en phase" avec la mesure que l'on cherche à prévoir. En revanche

TABLE 3.14 – Résultats en RMSE détaillés par classes

Horizon	Modèle	P_{moy}			
		Kc moyen	A	B	C
		Variabilité			
15 min	I		50	41	50
	II		80	96	82
	III		86	122	NaN
30 min	I		52	50	51
	II		87	103	86
	III		96	131	NaN
45 min	I		51	62	55
	II		93	109	90
	III		110	138	NaN

TABLE 3.15 – Résultats en MBE détaillés par classes

Horizon	Modèle	P_{moy}			
		Kc moyen	A	B	C
		Variabilité			
15 min	I		-15	11	24
	II		-28	-15	12
	III		-35	-22	NaN
30 min	I		-15	17	26
	II		-29	-12	13
	III		-33	-20	NaN
45 min	I		-14	24	30
	II		-29	-9	17
	III		-40	-18	NaN

le modèle A_{st2} atteint lui son maximum en -2, il présente un certain "effet de persistance". Cet effet permet d'améliorer grandement les performances du modèle comme nous l'avons vu précédemment mais crée aussi un effet de retard préjudiciable, notamment lorsque l'horizon de prévision augmente. Ainsi, alors que les scores étaient similaires pour un horizon de prévision de 30 minutes, le modèle P_{moy} obtient une meilleure corrélation en 0 pour un horizon de prévision de 45 minutes (Fig. 3.19). Dans ce cas, l'effet de persistance du modèle A_{st2} est toujours présent mais moins marqué, le maximum est atteint en 0 mais reste plus faible que celui obtenu avec P_{moy} . Ainsi au delà de 30 minutes, les effets positifs de la persistance ne sont plus prédominants sur les performances des modèles de prévision.

TABLE 3.16 – Résultats par classes : Coefficient de détermination

Horizon	Modèle	P_{moy}				
		Variabilité	Kc moyen	A	B	C
15 min	I			0.21	0.97	0.94
	II			0.64	0.80	0.84
	III			0.74	0.71	NaN
30 min	I			0.16	0.96	0.94
	II			0.57	0.77	0.82
	III			0.66	0.66	NaN
45 min	I			0.17	0.95	0.94
	II			0.50	0.74	0.80
	III			0.56	0.63	NaN

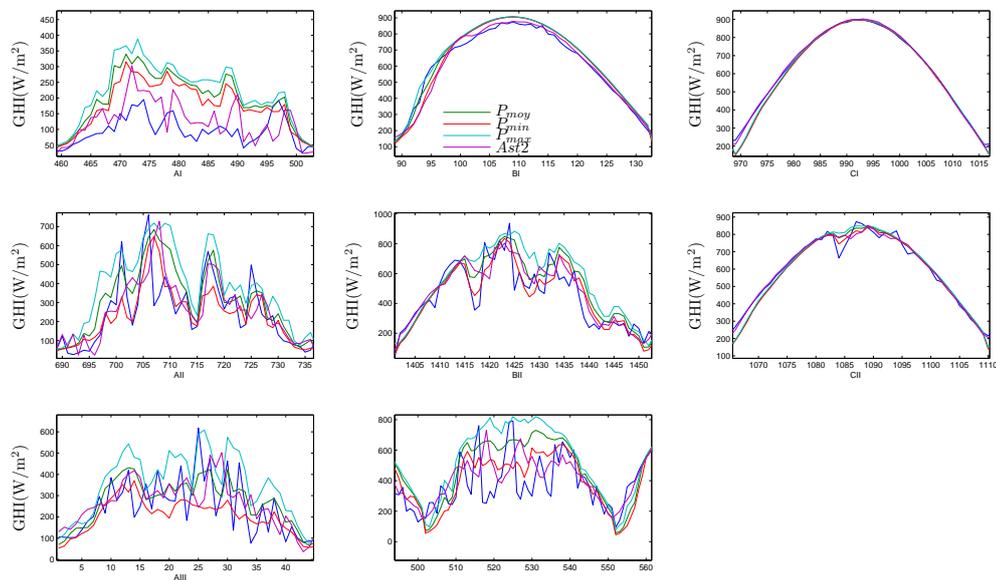


FIGURE 3.16 – Profils journaliers pour chacune des classes pour une prévision à 30 minutes

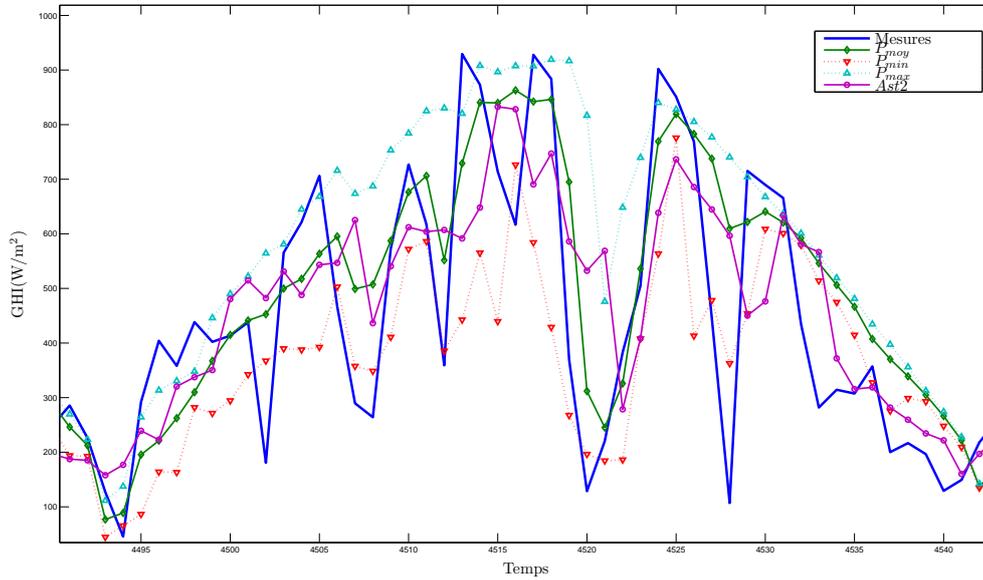


FIGURE 3.17 – Zoom sur un profil journalier pour une prédiction à 30 minutes. Le modèle présenté donne non seulement une prédiction mais également un "encadrement estimé" de la prédiction.

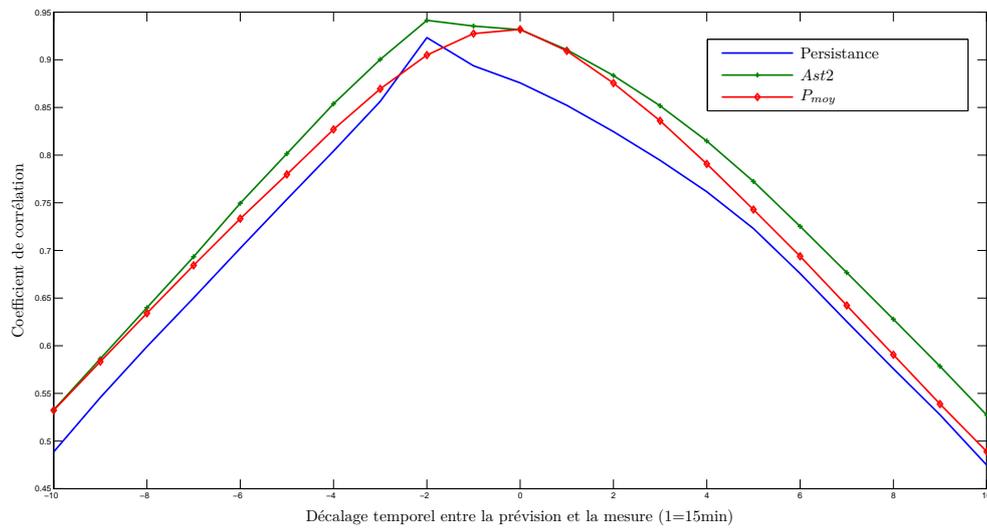


FIGURE 3.18 – Fonction d'intercorrélation entre la prédiction et la mesure pour trois modèles différents. Les prévisions sont effectuées pour un horizon de prédiction de 30 minutes ($h=2$).

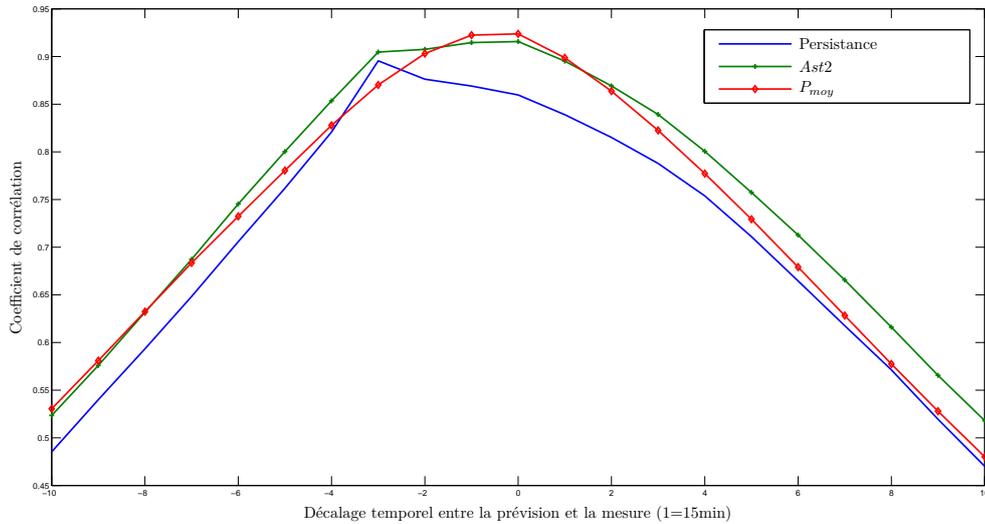


FIGURE 3.19 – Fonction d’intercorrélation entre la prévision et la mesure pour trois modèles différents. Les prévisions sont effectuées pour un horizon de prévision de 45 minutes ($h=3$).

3.5 Conclusion

Les résultats obtenus au cours de ce chapitre confirment de manière générale l’intérêt de l’utilisation des images satellitaires dans un objectif de prévision à très court terme. La figure 3.20 montre l’évolution de la RMSE pour les principaux modèles développés dans ce chapitre. Cette figure illustre les enjeux de la prévision par imagerie satellitaire. En effet, comme nous l’avons décrit précédemment, deux problématiques doivent être traitées, d’une part l’interprétation de l’image satellitaire en GHI (conversion), et de l’autre, l’anticipation du mouvement des masses nuageuse (prévision). L’impact de chacun de ces problèmes sur les performances de l’algorithme dépend de l’horizon de prévision. Pour des horizons de prévision faibles (15 minutes par exemple), l’erreur d’anticipation est négligeable devant l’erreur de conversion de l’image satellite : le modèle *Ast* qui n’a qu’une faible prise en compte du mouvement des nuages a de meilleures performances que le modèle P_{moy} . La prise en compte de l’historique de mesure et le processus d’apprentissage lui confèrent une meilleure capacité à faire le lien entre l’image et la mesure au sol contrairement à P_{moy} . Au contraire, lorsque l’horizon de prévision augmente (au delà de 30 minutes), les modèles incluant une étape d’anticipation sont meilleurs : l’erreur d’interprétation devient faible devant l’erreur d’anticipation. Le modèle *Ast2*, qui bénéficie à la fois de la force de l’apprentissage mais aussi d’une certaine prise en compte moyenne du mouvement principal des masses nuageuse donne les meilleures performances globales pour des horizons de prévisions allant jusqu’à 45 minutes. Pour des horizons plus grands, le modèle P_{moy} , plus performant en terme d’anticipation, semble être le plus approprié d’après la tendance des courbes de la figure 3.20. Les résultats n’ont pu être calculés au cours de la thèse pour de tels horizons de prévision pour des raisons de taille de la zone HC-3 extraite. D’après ces résultats, plusieurs pistes sont envisageables pour la suite.

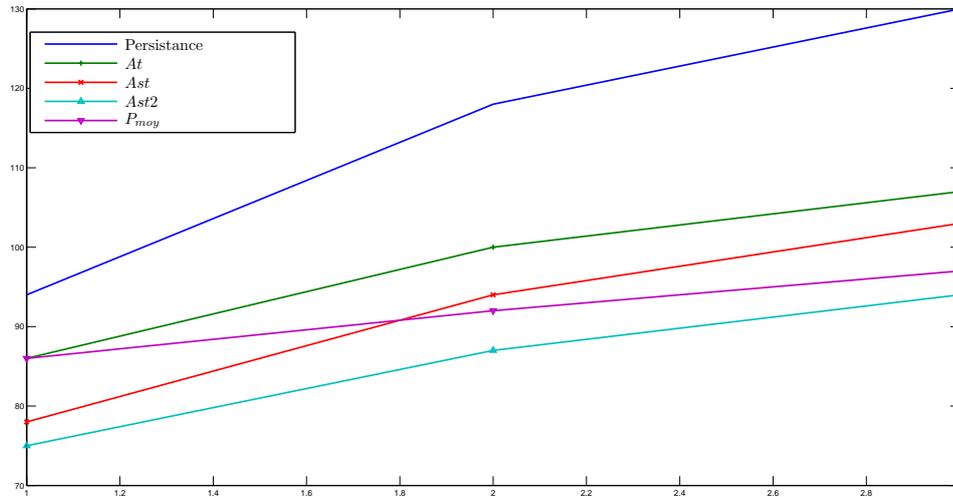


FIGURE 3.20 – Evolution de la RMSE en fonction de l’horizon de prévision pour différents modèles.

Les deux modèles de références A_{st2} et P_{moy} sont complémentaires. Le premier bénéficie de la robustesse de l’apprentissage et de l’effet de persistance, le second apporte des informations très riches et plus en phase avec les variations réelles de la mesure grâce à une meilleure prise en compte du mouvement observé. Deux pistes sont donc envisageables :

- Intégrer des informations provenant du sol dans la méthode P_{moy} pour améliorer ses performances à très court terme
- Intégrer la prise en compte d’un champ de vecteur dans la méthode utilisant les modèles AR.

La première solution est développée dans le chapitre 4 de cette thèse, la seconde reste au stade de perspective dans le cadre de ces travaux.

Fusion d'information pour la prévision du GHI

Sommaire

4.1	Méthodologie	96
4.2	Résultats expérimentaux	97
4.2.1	Procédé expérimental	97
4.2.2	Analyse quantitative	98
4.2.3	Analyse qualitative	100
4.3	Conclusion	102

La méthode P_{moy} développée dans le chapitre précédent présente des caractéristiques de prévision très intéressantes notamment en ce qui concerne la prévision de la variabilité du GHI. En revanche l'analyse de résultats a montré certaines faiblesses notamment sur l'étape de conversion. Dans ce chapitre, l'enjeu principal sera de combler cette lacune en intégrant des informations provenant des capteurs au sol. En effet nous savons depuis le chapitre 3 que l'intégration de l'historique de mesure du pyranomètre permet d'améliorer cette étape et améliore les performances à très court terme. Nous cherchons donc à vérifier cette hypothèse, mais également à évaluer le potentiel de l'imagerie hémisphérique pour cette application. La méthode d'estimation développée dans le chapitre 2 a permis de sélectionner cinq caractéristiques d'image du TSI permettant de décrire le GHI estimé. Au même titre que la mesure du pyranomètre, ces caractéristiques peuvent potentiellement améliorer les performances de prévision grâce à leur description plus fine de la couverture nuageuse présente sur le site. Les méthodes développées dans ce chapitre s'inscrivent ainsi dans une volonté de fusion de données entre :

- des mesures pyranométriques précises mais ponctuelles ;
- des images hémisphériques du ciel permettant une appréhension spatiale localisée très résolue spatialement et temporellement, en lien avec la variabilité du GHI induit par les nuages ;
- des images satellites offrant une observation du GHI dense et plus étendue , à des échelles spatiales plus larges.

La méthodologie appliquée est décrite en section 4.1. Elle s'appuie principalement sur l'utilisation de réseaux de neurones artificiels présentés précédemment au chapitre 2. Le procédé expérimental ainsi que les résultats sont présentés en section 4.2.

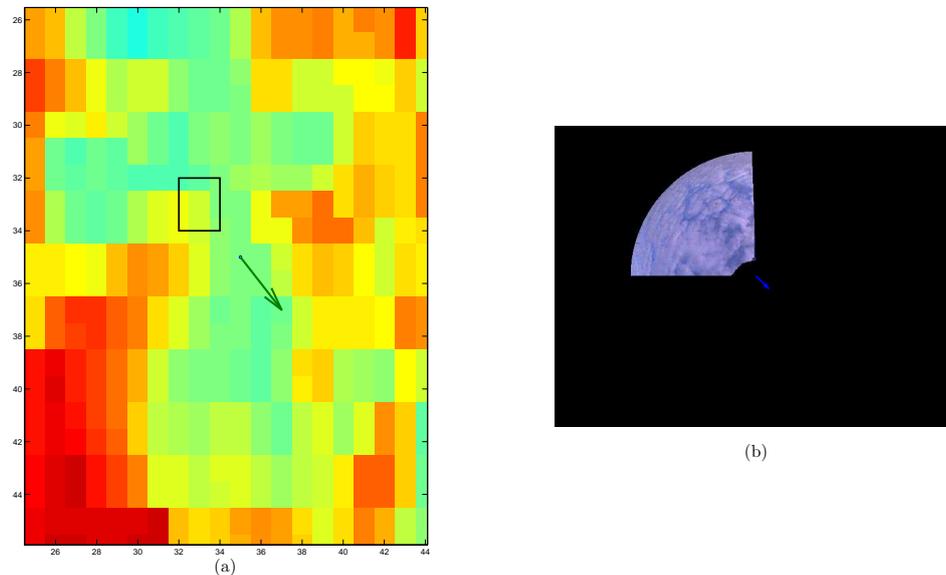


FIGURE 4.1 – Selection des entrées pour le modèle P_{TSI90} : (a) ZI définie sur l'image HC-3 d'après le vecteur vitesse ; (b) Cône de l'image au sol sélectionné à partir du vecteur

4.1 Méthodologie

Nous souhaitons ainsi fusionner des informations provenant de la méthode P_{moy} et des caractéristiques de l'image provenant du TSI au même instant afin de produire une prévision améliorée de GHI. Comme indiqué au chapitre 2 (section 2.2.5), les réseaux de neurones artificiels sont des outils bien adaptés aux problèmes de régression non linéaires de ce type. Parmi les caractéristiques sélectionnées au chapitre 2, nous ne conservons pas l'information "Ban" (tableau 4.1) pour la prévision. Cette information est extrêmement liée à la présence de rayonnement direct, elle est donc importante pour l'estimation, mais induit un effet de persistance très important en prévision. Par ailleurs, nous disposons également grâce aux travaux du chapitre 3 d'informations sur la direction des vents. Il est donc possible d'identifier dans l'image au sol des informations plus ciblées en fonction de la direction des vents observée. Cette zone est définie en isolant un cône de 90° centré sur la direction opposée au vecteur de déplacement calculé dans le chapitre précédent par block matching sur les images satellite. La figure 4.1 illustre ce procédé. Cependant, en utilisant cette technique, la position du soleil dans l'image n'est pas prise en compte. En choisissant la position de la zone uniquement en fonction du vecteur vitesse, nous supposons que le soleil est toujours au centre de l'image, ce qui est naturellement faux. Nous définissons alors une nouvelle zone d'intérêt dépendante à la fois de la vitesse calculée des vents, mais aussi de l'angle azimutal du soleil. De cette façon la zone ciblée contient les nuages susceptibles de passer entre le soleil et le capteur dans les instants futurs. Le procédé utilisé est décrit en figure 4.2, toutefois, la zone ainsi obtenue est de taille variable contrairement aux zones décrites précédemment.

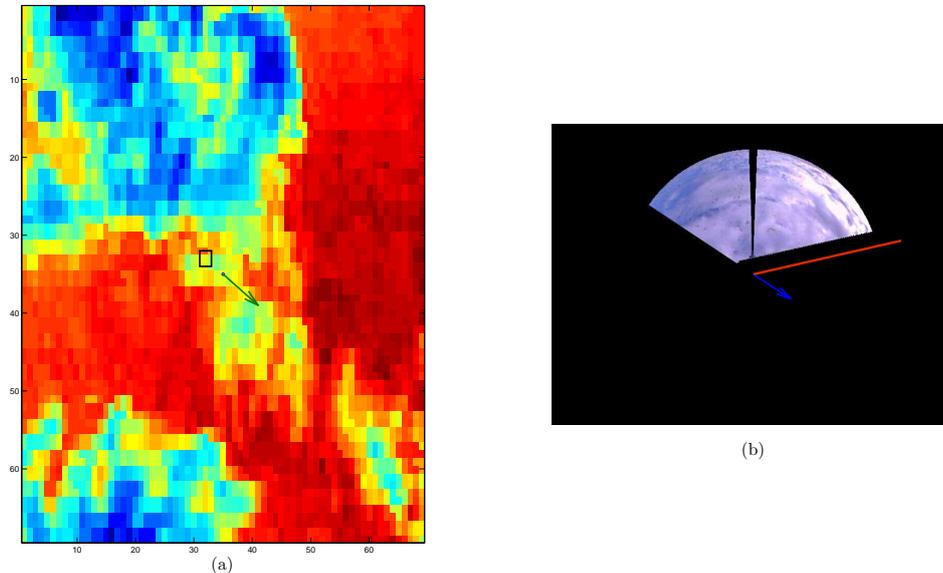


FIGURE 4.2 – Selection des entrées pour le modèle $P_{TSI_{cone}}$: (a) ZI définie sur l'image HC-3 d'après le vecteur vitesse ; (b) Cone de l'image au sol sélectionné en fonction du vecteur vitesse (bleu) et de l'angle azimutal solaire (rouge).

Nous disposons ainsi de plusieurs degrés de complexité qui justifient la construction parallèle des modèles suivants :

- P_{moy} est le modèle satellite de référence développé dans le chapitre précédent.
- P_{pyr} est un réseau à deux entrées fusionnant la prévision $P_{moy}(t+h)$ et la mesure du pyranomètre GHI(t).
- P_{TSI} est un réseau à cinq entrées fusionnant la prévision $P_{moy}(t+h)$ et les quatre caractéristiques extraites de l'image hémisphérique du TSI retenues au chapitre 2 prises au temps t. Elles sont rappelées dans le tableau 4.1.
- P_{TSI90} est un réseau à cinq entrées fusionnant la prévision $P_{moy}(t+h)$ et les quatre caractéristiques extraites uniquement dans un cone de 90° du TSI défini par le vecteur de déplacement des nuages (Fig.4.1).
- $P_{TSI_{cone}}$ est un réseau à cinq entrées fusionnant la prévision $P_{moy}(t+h)$ et les quatre caractéristiques extraites uniquement dans le cone du TSI défini par le vecteur de déplacement des nuages et l'angle azimutal solaire (Fig.4.2).

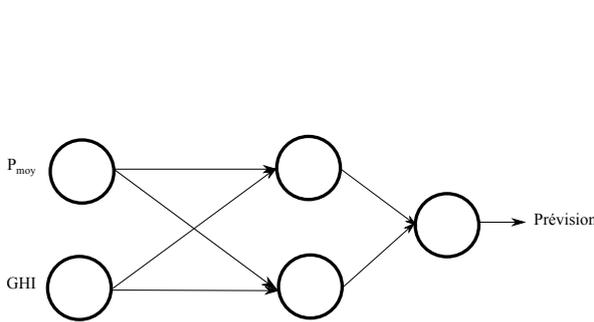
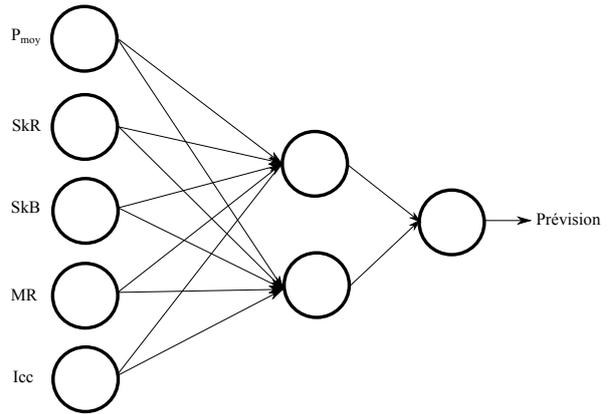
4.2 Résultats expérimentaux

4.2.1 Procédé expérimental

La construction des bases d'apprentissage, de validation et de test est exactement la même que celle effectuée au cours du chapitre 2 et décrite dans la section 2.2.6. Les images du

TABLE 4.1 – Caractéristiques du TSI retenues au chapitre 2 .

Nom	Description
Ban	Moyenne du canal rouge dans la zone circumsolaire sur la bande noire de protection du TSI.
SkR	Skewness du canal rouge.
SkB	Skewness du canal bleu.
MR	Moyenne du canal rouge.
Icc	Indice de couverture nuageuse.

FIGURE 4.3 – Structure du modèle P_{pyr} FIGURE 4.4 – Structure des modèles P_{TSI} , P_{TSI90} et $P_{TSIcône}$

TSI ne sont disponibles dans le cadre de la thèse que pour l'année 2011. La méthode P_{moy} appliquée dans le chapitre précédent sur l'année 2012 et bénéficiant de la calibration d'HC-3 effectuée sur l'année 2012 est ici appliquée à l'année 2011 avec une calibration faite sur l'année 2012 par le même procédé (régression polynomiale d'ordre 4). Cette méthode sert de référence à l'expérience menée dans ce chapitre sur l'introduction des données acquises sur site. La structure du réseau de neurones pour chacun des modèles P_{pyr} , P_{TSI} , P_{TSI90} et $P_{TSIcône}$ a été optimisée via une approche expérimentale telle que celle développée dans le chapitre 2. Les structures retenues sont représentées sur les figures 4.3 et 4.4.

4.2.2 Analyse quantitative

Les résultats présentés dans les tableaux 4.2 à 4.4 ont été calculés sur la base de test définie au chapitre 2. Cette base a été construite de façon à être représentative de l'ensemble des situations observées sur l'année 2011 via la classification effectuée au chapitre 1 (section 1.4.1). En première observation, l'intégration d'information *in-situ* via cette méthode n'apporte que des effets positifs. Les performances initiales de la méthode satellitaire P_{moy} sont toutes améliorées quel que soit l'horizon temporel ou le critère de qualité concerné (excepté pour la méthode $P_{TSIcône}$ qui sera commentée plus tard). Ces améliorations semblent être principalement dues à la réduction du biais étant donné l'amélioration des biais obtenus. Les

coefficients de détermination sont relativement insensibles aux différences introduites dans les modèles.

Intérêt de l'image au sol : la comparaison directe des modèles P_{pyr} et P_{TSI} permet d'évaluer l'intérêt de l'utilisation d'un imageur hémisphérique par rapport à un pyranomètre classique pour la prévision. Deux phénomènes sont remarquables dans ces résultats :

- La réduction de MBE observée de façon générale lors de l'intégration de données in-situ est beaucoup plus importante pour le modèle P_{TSI} qui annule quasiment le biais du modèle satellitaire.
- Le gain en RMSE est dépendant de l'horizon de prévision pour le modèle P_{pyr} , il est plus important pour une prévision à 15 minutes que pour des horizons plus grand. Le gain du modèle P_{TSI} semble lui plus stable par rapport à l'horizon de prévision : il oscille autour de 5% de la RMSE du modèle P_{moy} .

Ces deux informations combinées laissent penser que si le premier modèle bénéficie principalement de l'effet positif de la persistance, le second permet, quant à lui, une réelle amélioration des capacités de conversion du modèle de prévision.

Apport de la sélection de zone : D'après l'ensemble des résultats obtenus, la sélection de zone plus précises de l'image au sol semble dégrader les performances globales de l'algorithme. L'analyse quantitative donne peu d'informations sur les raisons de cette dégradation. Cependant, de meilleurs résultats sont observés pour la méthode P_{TSI90} par rapport à la méthode $P_{TSI_{cone}}$ ce qui pourrait indiquer que la définition d'une zone plus stable (taille fixe) est primordiale pour le succès du réseau de neurones.

TABLE 4.2 – RMSE en Wm^{-2} en fonction du modèle et de l'horizon de prévision h

Modèle \ h (min)	15	30	45
P_{moy}	87	94	102
P_{pyr}	83	92	99
P_{TSI}	82	90	97
P_{TSI90}	87	94	101
$P_{TSI_{cone}}$	89	96	104

TABLE 4.3 – MBE en Wm^{-2} en fonction du modèle et de l'horizon de prévision

Modèle \ h (min)	15	30	45
P_{moy}	-13	-12	-12
P_{pyr}	-9	-10	-11
P_{TSI}	-1	-1	-4
P_{TSI90}	-2	-3	-4
$P_{TSI_{cone}}$	1.2	-6	-6

TABLE 4.4 – Coefficient de détermination R^2 en fonction du modèle et de l'horizon de prévision h

Modèle \ h (min)	15	30	45
P_{moy}	0.87	0.84	0.81
P_{pyr}	0.88	0.85	0.82
P_{TSI}	0.88	0.85	0.83
P_{TSI90}	0.86	0.84	0.82
$P_{TSIcône}$	0.86	0.83	0.80

4.2.3 Analyse qualitative

Nous avons vu dans le chapitre 3 que l'apport de mesures du pyranomètre impliquait un "effet de persistance". Observons alors les fonctions d'intercorrélation entre la mesure et les différents modèles afin d'observer la présence de cet effet dans la prévision. Ces courbes sont données dans les figures 4.5 à 4.7. Il existe bien un effet de persistance pour le modèle P_{pyr} dont le maximum de corrélation est atteint pour un retard correspondant à l'horizon de prévision. En revanche, les autres modèles, faisant intervenir les statistiques de l'image du TSI, ne présentent pas cet effet. Pour un horizon de prévision de 15 minutes (Fig. 4.5), ces modèles sont bien en phase avec la mesure et avec le modèle satellitaire original quand le modèle P_{pyr} présente un important déphasage. Pour les horizons de 30 et 45 minutes, nous observons ensuite un "tassement" du maximum entre les retards 0 et -1 qui traduit le manque de portée de l'image au sol pour atteindre ces horizons. De façon plus globale, le coefficient de corrélation en 0 est toujours plus important pour les modèles incluant des données *in-situ* mettant en évidence un effet de persistance non désiré.

Ces résultats sont également visibles sur les profils journaliers de prévision. La figure 4.8 expose un exemple de profil où l'effet de persistance observé pour le modèle P_{pyr} est identifiable pour une prévision à 15 minutes. Par ailleurs, si les résultats quantitatifs et les fonctions d'intercorrélation ne semblaient pas indiquer de quelconque avantage à la définition de zone dans l'image du TSI, certains profils journaliers prouvent le contraire. Par exemple la figure B.1 illustre deux situations où le modèle $P_{TSIcône}$ permet de prévoir une chute de rayonnement qu'aucun autre modèle n'était en mesure de prévoir. La raison des performances moindres de ce modèle s'explique par son manque de robustesse. Les zones définies rendent le modèle dépendant du calcul de vecteur mais entraînent également un apprentissage plus délicat : en donnant les statistiques globale de l'image nous attendons de l'ANN qu'il corrige le biais de la prévision satellite en fonction des conditions observées sur site, ce qu'il réalise correctement. En revanche, en définissant des zones dans l'image, nous tentons d'intégrer une nouvelle information de prévision physique dans le modèle. Or les statistiques de l'image du TSI telles qu'on les a définies semblent trop instables pour réaliser cette étape de façon robuste. La distorsion de l'image et l'absence de prise en compte de l'altitude de la couverture nuageuse sont autant de paramètres qui rendent l'utilisation des statistiques de nuages très instables dans le temps. Il existe des cas où les conditions sont réunies pour obtenir de très bons résultats

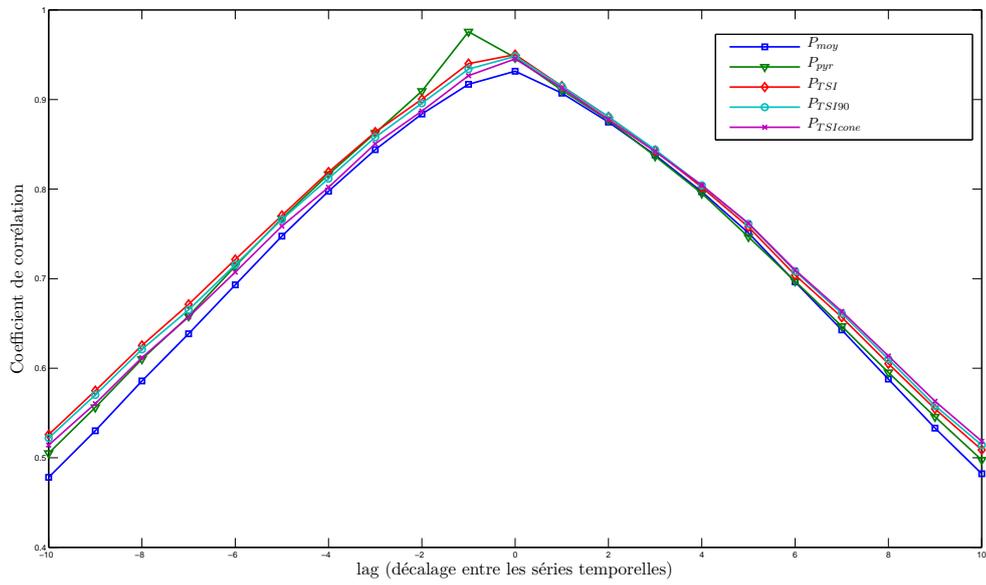


FIGURE 4.5 – Fonction d’intercorrélacion entre les modèles et la mesure pour un horizon $h = 15$ minutes.

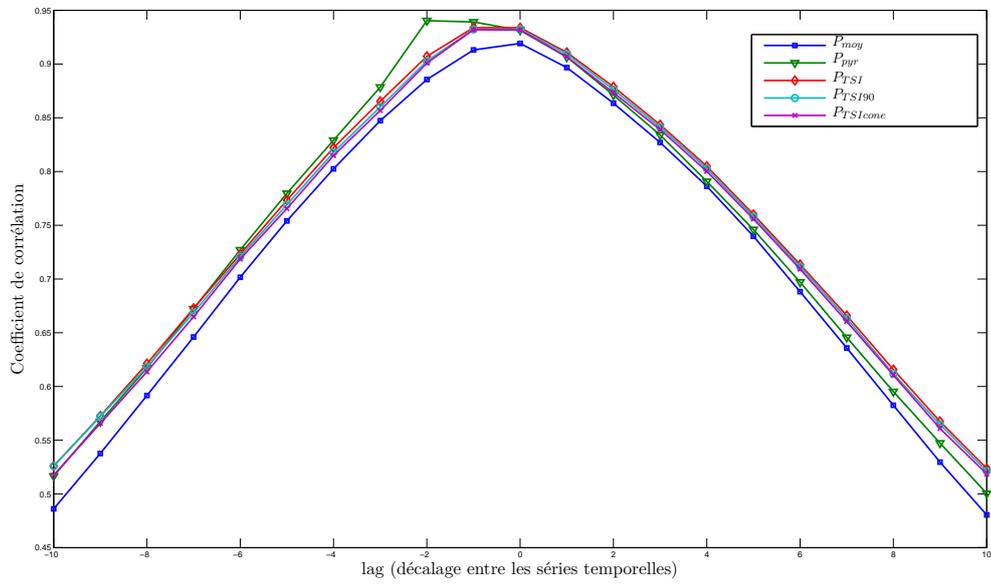


FIGURE 4.6 – Fonction d’intercorrélacion entre les modèles et la mesure pour un horizon $h = 30$ minutes.

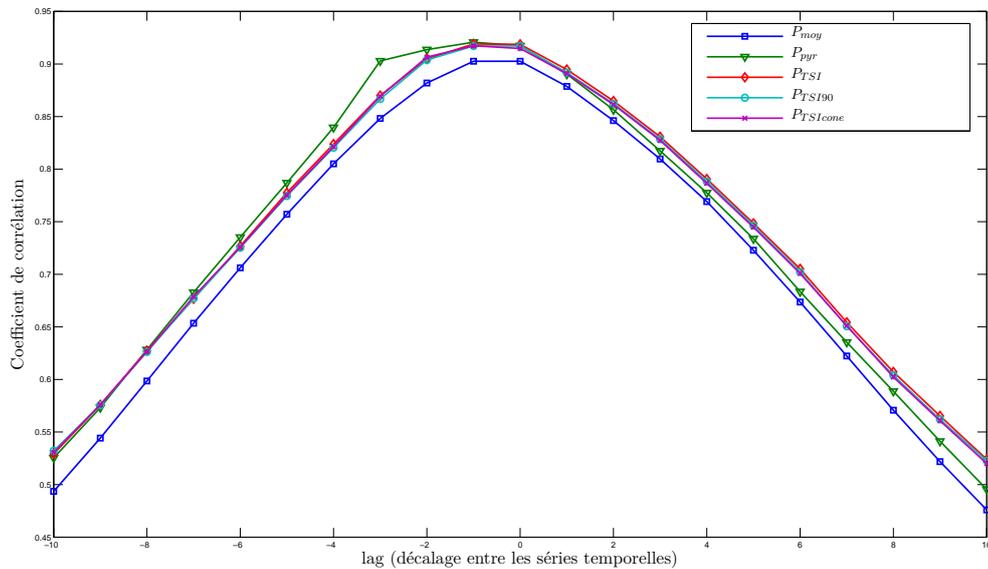


FIGURE 4.7 – Fonction d'intercorrélation entre les modèles et la mesure pour un horizon $h = 45$ minutes.

(Fig.B.1) mais le manque de robustesse affecte très fortement l'ensemble des résultats. D'autres profils journaliers sont présentés en annexe (B) pour mieux illustrer l'ensemble des horizons de prévision testés.

4.3 Conclusion

En utilisant les connaissances acquises au chapitre 2 sur les caractéristiques de l'image au sol et celles du chapitre 3 pour la prévision, nous avons développé un modèle de fusion permettant de palier les faiblesses du modèle satellitaire initial du chapitre 3. Par ailleurs, la création de modèles de complexité croissante a permis de mettre en évidence les limites de l'utilisation directe de l'image du TSI pour la prévision. Si l'ajout de statistiques globales de l'image a permis une réduction significative de la MBE et une amélioration moyenne du modèle, ce n'est pas le cas pour les modèles plus complexes faisant intervenir une sélection de zone d'intérêt. Ces modèles, qui présentent tout de même des résultats très prometteurs, nécessitent cependant un degré de précision supplémentaire dans les entrées sélectionnées pour être plus performant. Il s'agirait notamment d'inclure dans le processus la prise en compte de l'altitude des nuages et leur position réelle par rapport au capteur afin de donner une meilleure signification physique aux caractéristiques extraites. Ces informations peuvent être acquises mais requièrent des installations et/ou des calculs supplémentaires dont nous ne disposons pas dans le cadre de cette thèse. Cependant, cette première étude permet d'établir un premier modèle de fusion et ouvre la voie pour la création de modèles plus perfectionnés permettant la fusion d'information entre l'imagerie satellitaire et l'imagerie hémisphérique.

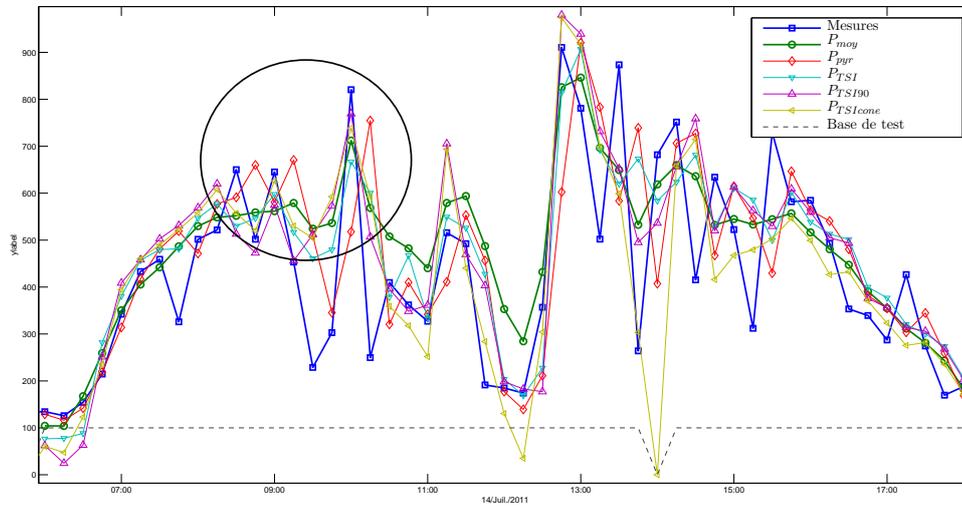


FIGURE 4.8 – Profil journalier, $h=15\text{min}$, sur ce profil, le "déphasage" du modèle P_{pyr} lié à l'effet de persistance est très marqué par rapport aux autres modèles, en particulier dans la zone pointé dans l'ellipse noire.

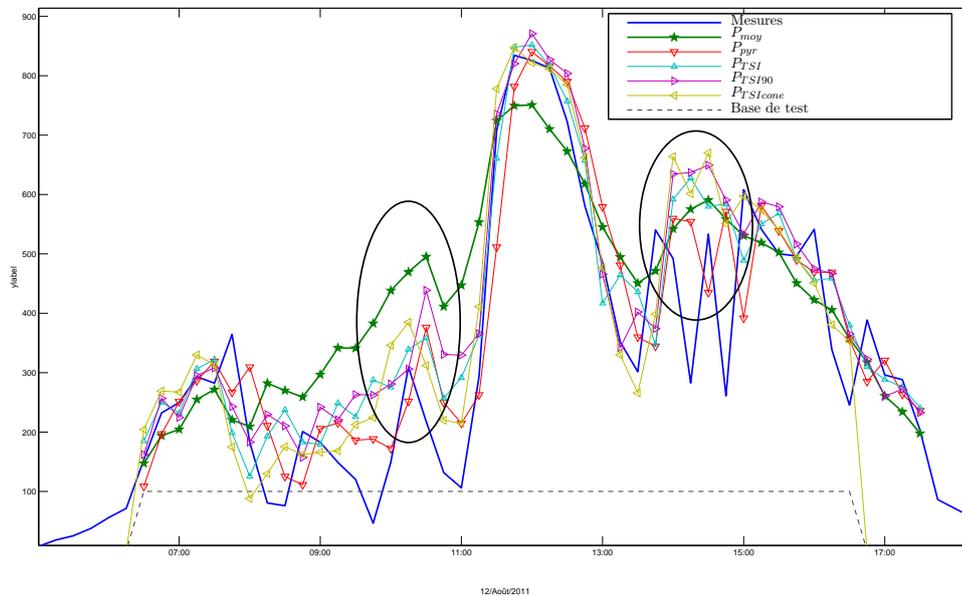


FIGURE 4.9 – Profil journalier, $h=15\text{min}$, les ellipses noires indiquent les endroits où le modèle $P_{TSI\text{cone}}$ permet une modification positive par rapport à l'ensemble des autres modèles.

Conclusion et perspectives

Cette thèse porte sur la prévision à très court terme (intra-horaire) du rayonnement solaire incident au sol sur un plan horizontal par l'intermédiaire d'imagerie satellitaire et/ou hémisphérique vue du sol. L'objectif principal était de fusionner les deux types d'information afin d'améliorer les performances de prévision. Dans un premier temps, chacun des capteurs a été étudié séparément. Un modèle d'estimation du GHI à partir d'images hémisphériques d'un TSI a été développé dans le chapitre 2. Cette étude a permis d'identifier les caractéristiques de l'image pertinentes pour l'étude du GHI mais aussi d'établir des bases de données et un cadre de travail pour l'utilisation de réseaux de neurones artificiels. Le chapitre 3 présente des modèles de prévision à très court terme à partir de données satellitaires. Deux modèles ont ainsi été construits et validés. Le premier modèle est un modèle autoregressif spatio-temporel qui s'appuie sur des séries temporelles spatialement étendue et résolue sur des zones d'intérêt fixes dans le voisinage immédiat du site d'intérêt. Il intègre également un historique de mesures au sol. Le deuxième modèle s'appuie, quant à lui, sur des zones d'intérêt sélectionnées en fonction des mouvements observés par traitement d'image sur les séries temporelles d'images satellitaires. Ces modèles ont permis d'une part de valider l'utilisation de cartes Hélioclim-3 comme seule source d'entrées à la place des images satellites brutes. Cette approche constitue une innovation en regard de la bibliographie et a fait l'objet d'une publication acceptée dans un journal international à comité de lecture. D'autre part, ils ont permis d'établir une référence pour la prévision satellitaire et de cibler les inconvénients/avantages de ces modèles en fonction de la méthode utilisée. Enfin, le chapitre 4 présente une méthode fusionnant la prévision satellitaire et les informations provenant de l'imagerie hémisphérique en utilisant la structure précédemment établie des réseaux de neurones. Nous avons alors montré l'intérêt d'une telle fusion en comparant l'impact de l'intégration de données du TSI par rapport à l'intégration de données du capteur de rayonnement lui-même. Alors que l'ajout de données du pyranomètre implique généralement un "effet de persistance" limitant quelque peu l'intérêt de la fusion d'information, l'utilisation de caractéristiques de l'image hémisphérique permet d'améliorer les performances de la méthode satellitaire sans impact sur les capacités de prévision intrinsèques à la méthode satellitaire. Pour résumer, nous disposons de trois sources d'information, le pyranomètre, la base de données satellitaire HC-3 et le TSI que nous souhaitons fusionner dans le but de fournir une prévision du GHI améliorée. La fusion d'image au sens direct du terme semble difficile étant données les écarts de résolutions à la fois temporelle et spatiale entre les deux types d'image. Par ailleurs, cette approche aurait nécessité de nombreux paramètres physiques précis supplémentaires tels que l'altitude des nuages et leur épaisseur optique qui sont des paramètres relativement difficiles à obtenir. C'est donc une approche différente qui a été adoptée dans cette thèse : identifier les caractéristiques fortes de chaque information afin de pouvoir les utiliser de façon complémentaire. Il s'agit donc d'une fusion d'information extraite des trois sources. L'imagerie satellitaire permet d'avoir une observation large des phénomènes météorologique permettant leur prévision. En revanche, elle ne dispose pas d'une résolution spatiale suffisante pour pouvoir estimer l'impact de ces phénomènes sur la mesure d'un capteur ponctuel comme le pyranomètre. Au contraire, le TSI

dispose de cette information mais reste limité en terme d'horizon de prévision par son champs de vision restreint. Les résultats obtenus au chapitre 4 montrent l'intérêt de la combinaison de ces deux outils pour une prévision à court terme du GHI.

L'ensemble de ces travaux a été mené et présenté dans cette thèse de façon progressive, ce qui permet de quantifier l'apport de chaque changement de méthode ou de nature des données d'entrée comme les mesures pyranométriques ponctuelles, les images hémisphériques du ciel, ou encore, les séries temporelles de cartes de GHI issues de HC-3 centrées sur des zones d'intérêt fixe ou variables. Cette approche permet de quantifier les améliorations relatives tout en identifiant les sources d'erreurs afin de dégager les pistes d'amélioration. Sur le plan industriel, elle pourrait permettre de proposer une réponse adaptée aux installations existantes pour le client. Ainsi, il est possible de fournir une prévision en fonction des données effectivement disponibles en opération : données issues de satellite, station pyranométrique, TSI ou autres. Les résultats obtenus par les modèles plus élaborés peuvent désormais justifier de façon quantitative l'installation de capteurs *in-situ* pour de meilleures performances à très court terme.

Pour conclure, nous retiendrons en tant que résultats principaux de cette thèse les points suivants :

- L'image hémisphérique du TSI permet d'estimer le GHI malgré l'obstruction de la bande protectrice.
- Les images HC-3 peuvent être utilisées comme seules sources d'information satellitaire pour la prévision du GHI, y compris pour le calcul de mouvement des masses nuageuses.
- L'utilisation d'images hémisphériques est préférable à celle de pyranomètres dans le cadre d'une fusion d'information pour la prévision car elle permet d'éviter l'"effet de persistance" caractéristique.

Les travaux réalisés ont également ouvert la voie à d'autres pistes. A la suite de l'étude des images hémisphériques, l'intégration des travaux réalisés en prévision du rayonnement direct dans la littérature permettrait d'affiner les performances à très court terme en indiquant une probabilité de présence ou d'atténuation du rayonnement direct. Par ailleurs, ces travaux font souvent appel à une étape de calcul de champ de vecteur réalisé sur la série d'images hémisphériques. La comparaison des champs de vecteurs obtenus de cette façon avec ceux obtenus par traitement d'images satellitaires permettrait d'une part d'affiner la résolution des champs, mais également de détecter les situations de couches multiples de nuages souvent sources d'erreurs pour la plupart des méthodes.

Du point de vue de l'imagerie satellitaire, une perspective d'amélioration de fusion des méthodes proposées dans le chapitre 3 consisterait à trouver une solution pour bénéficier à la fois de la robustesse de l'apprentissage du modèle autoregressif et des performances d'adaptation de la méthode utilisant les champs de vecteurs. Nous avons en effet proposé premièrement un modèle linéaire utilisant une zone d'intérêt fixe mais dont les coefficients sont appris, puis un autre modèle dont la localisation de la zone d'intérêt s'adapte à la situation avec une transposition effectuée par une simple moyenne. Il s'agit donc de développer un modèle utilisant les informations en temps réel pour ajuster à la fois la zone d'intérêt mais aussi les coefficients linéaires appliqués. Par ailleurs, la reprise des méthodes développées à une résolution tempo-

relle plus faible avec, par exemple, des irradiations à 30 minutes permettrait d'améliorer les résultats pour des horizons de prévision plus importants.

Enfin, les méthodes développées au cours de de cette thèse ont été comparées à une méthode servant actuellement de référence dans la bibliographie actuelle, la persistance sur l'indice de ciel clair. Toutefois, leur confrontation à d'autres méthodes de la bibliographie permettrait mieux juger de leurs forces et faiblesses sur les mêmes jeux de données et les mêmes méthodologie de validation consistant à distinguer absolument les données d'apprentissage des données de validation. C'est pourquoi il pourra être intéressant de confronter les différentes méthodes proposées en participant à des activités internationales de comparaison de méthodes de prévision de la ressource solaire telles que celles proposées par la tâche 46 du programme *Solar Heating & Cooling*(SHC)¹. Ce sera aussi l'occasion de contribuer en proposant les nouvelles approches d'évaluation de la qualité de la prévision exposées dans la thèse comme celle consistant à décomposer l'analyse suivant une classification des différents types de journées proposée dans le chapitre 1 ou encore celle proposant l'analyse des intercorrelations entre les modèles et la variable d'intérêt qui a permis la caractérisation de l'effet de persistance dans les chapitre 3 et 4.

1. <http://task46.iea-shc.org/subtasks>

Tests sur le ratio pour la méthode de segmentation de l'image hémisphérique

Les figures suivantes présentent les résultats de segmentation de l'image en fonction du seuil choisi pour la méthode de segmentation R/B décrite dans la section 2.2.3.3.

FIGURE A.1 – Test de seuil pour la méthode de segmentation par ratio : Seuil = 0.4

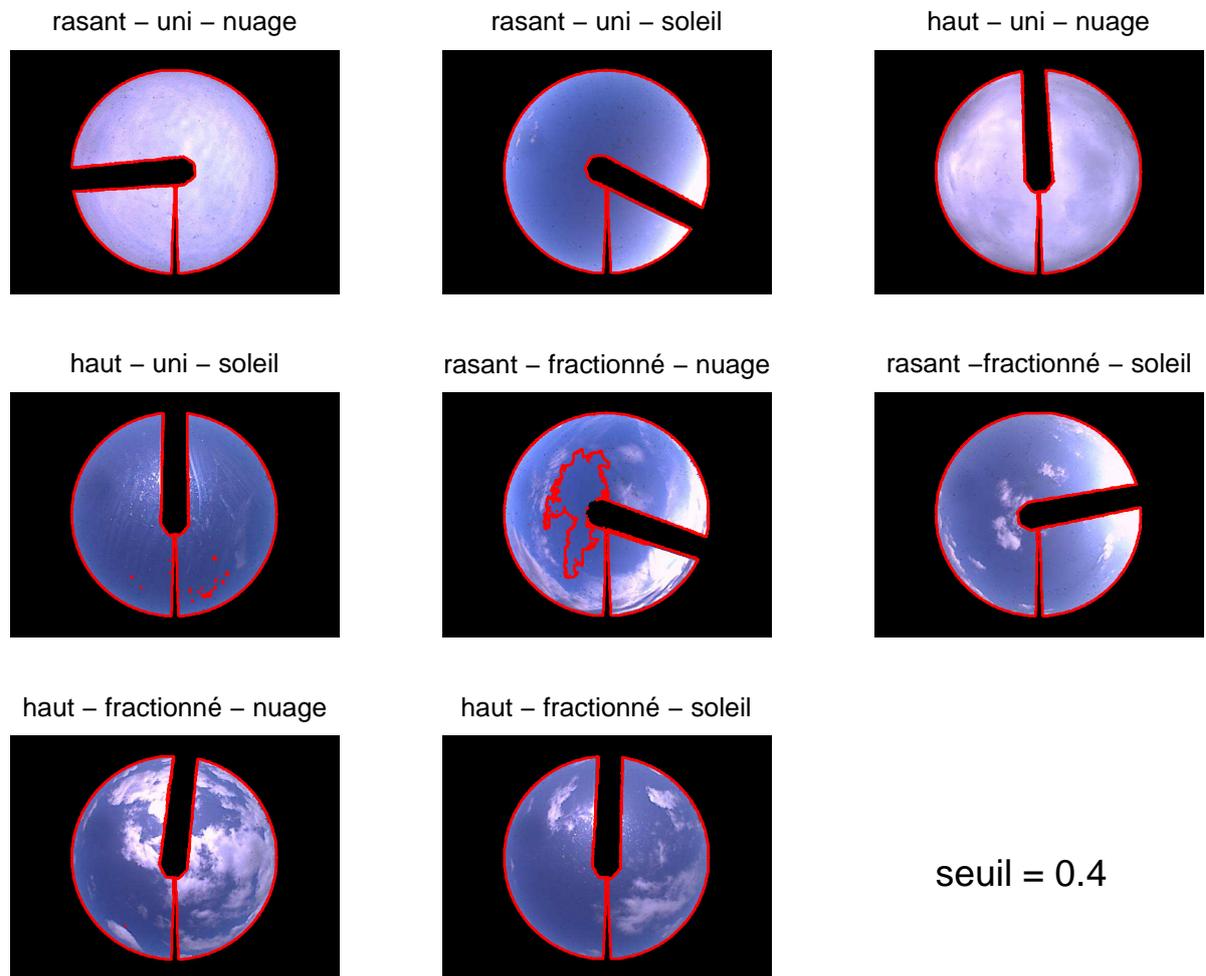


FIGURE A.2 – Test de seuil pour la méthode de segmentation par ratio : Seuil = 0.55

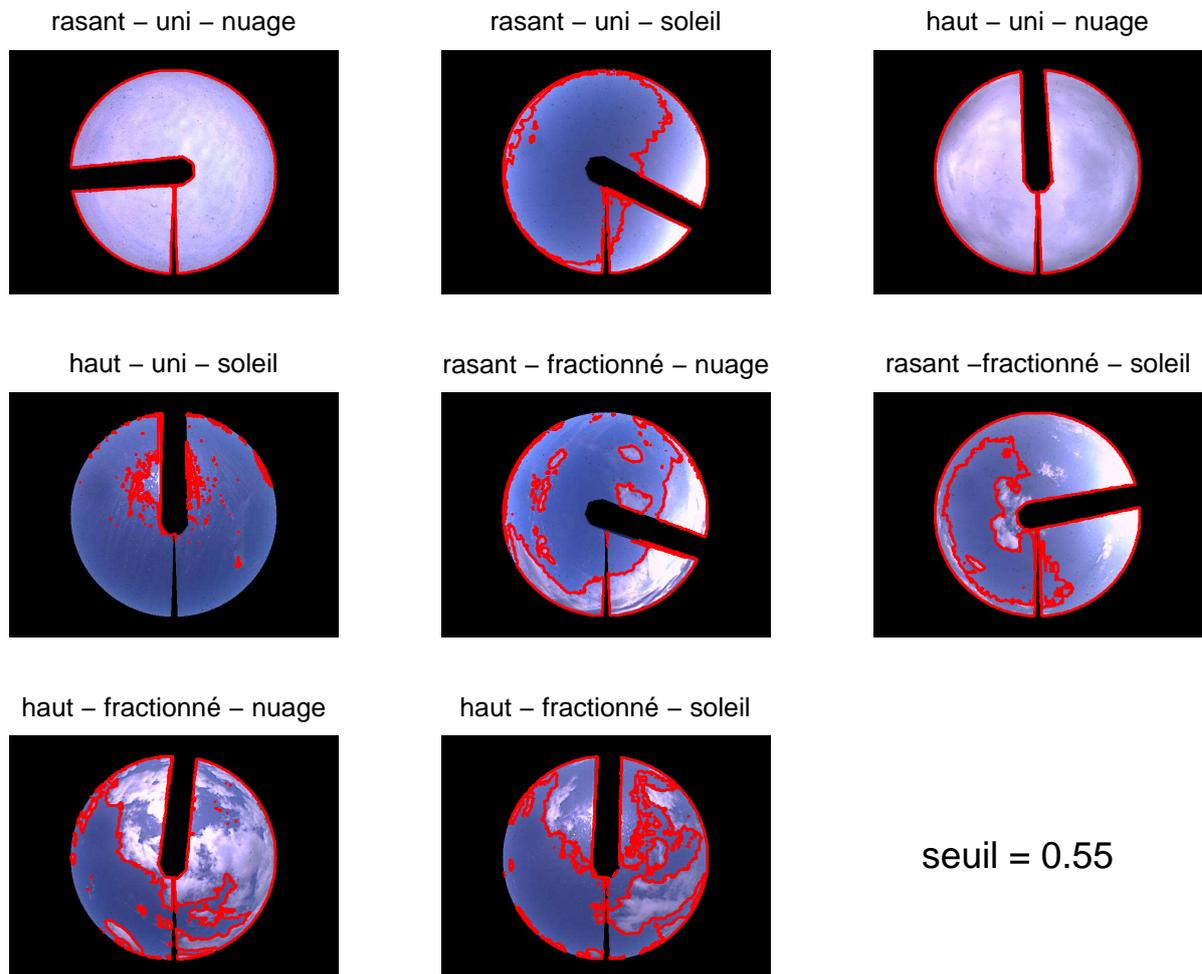


FIGURE A.3 – Test de seuil pour la méthode de segmentation par ratio : Seuil =0.6

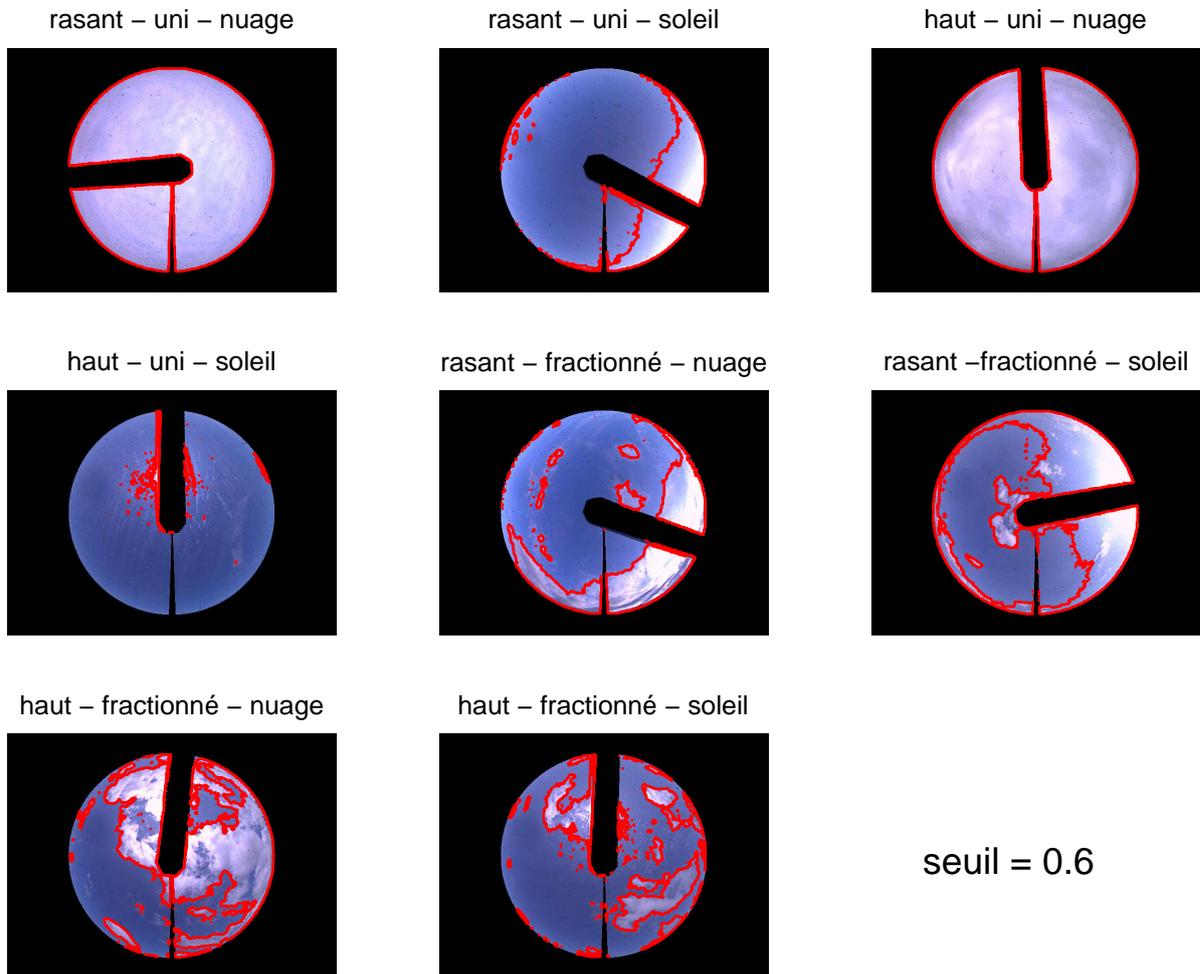


FIGURE A.4 – Test de seuil pour la méthode de segmentation par ratio : Seuil = 0.65

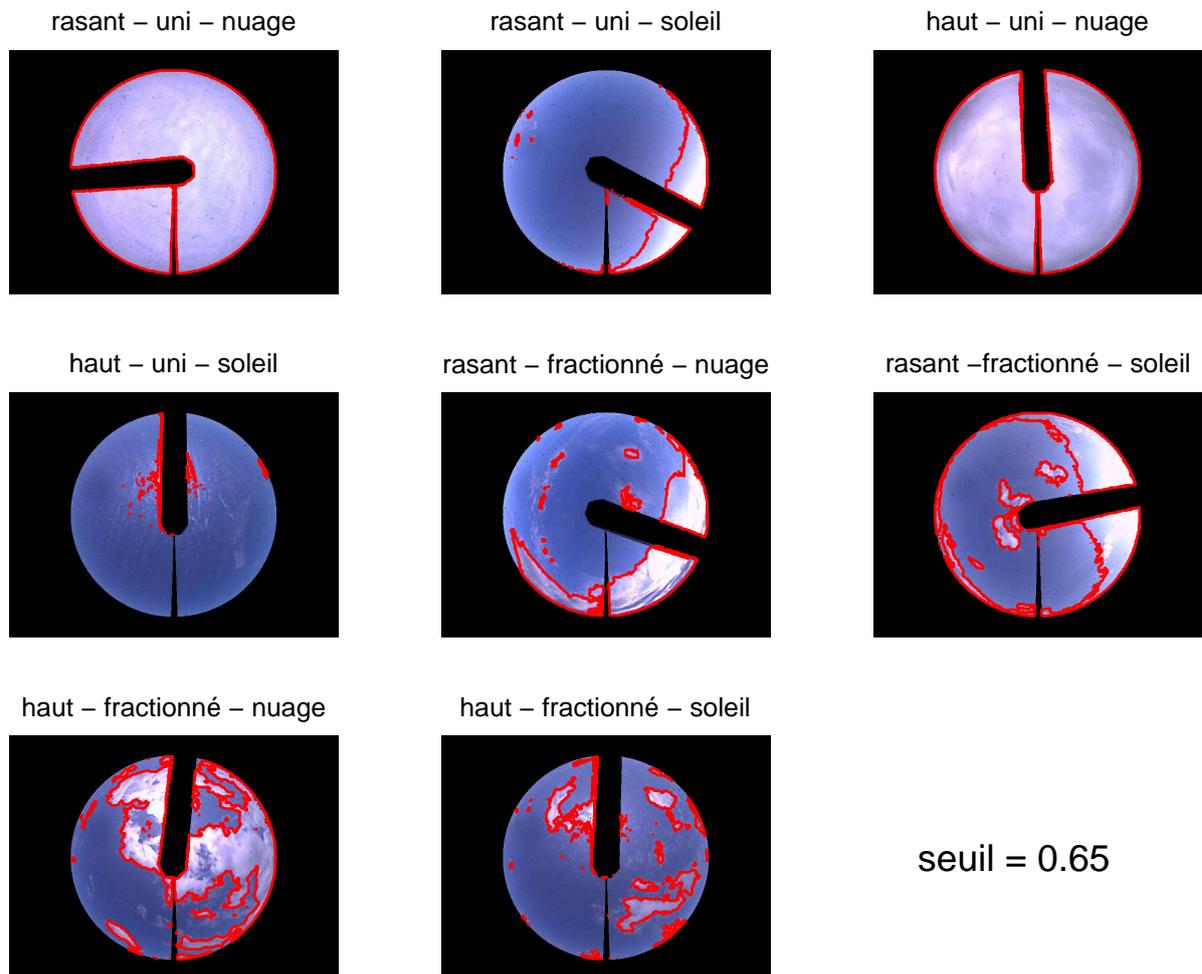
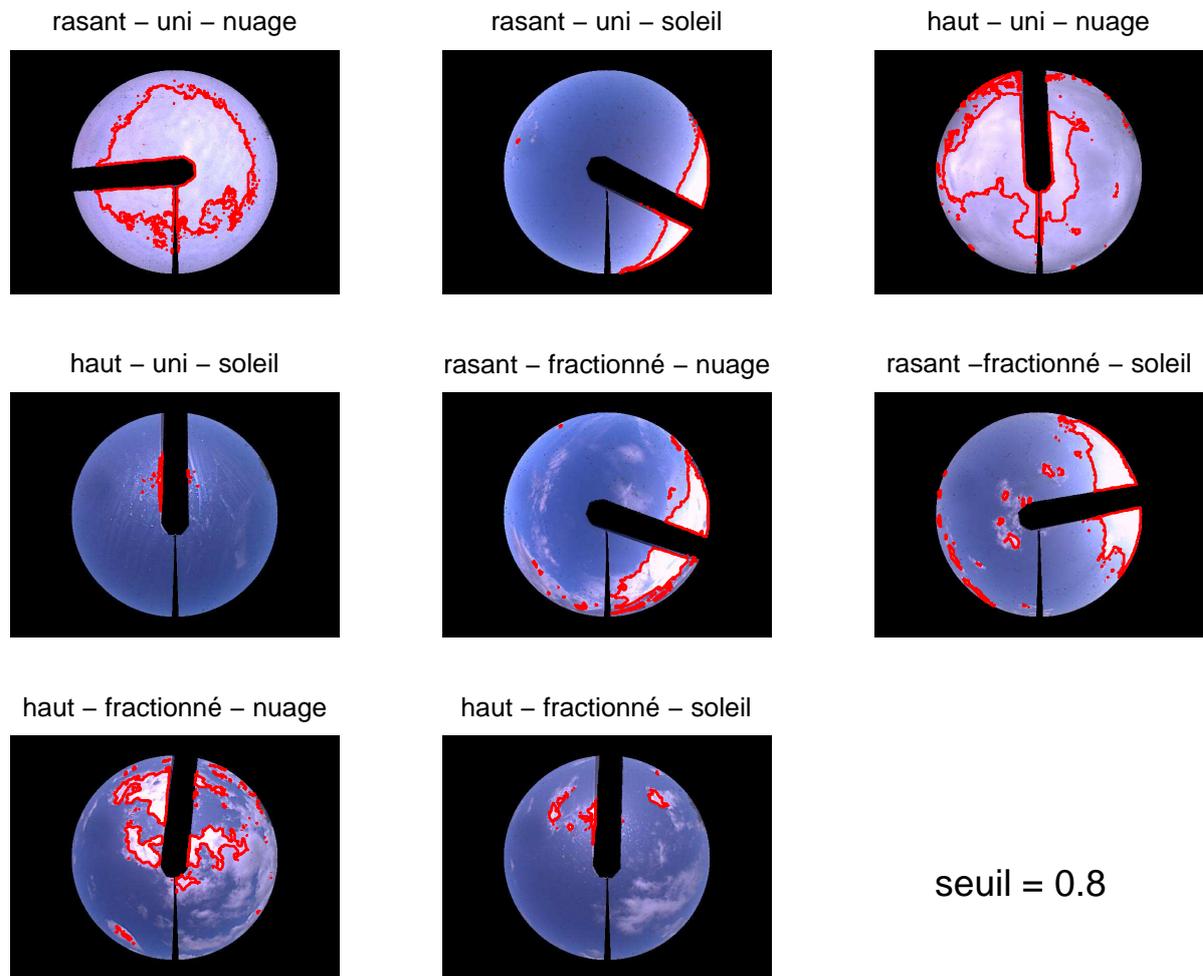


FIGURE A.5 – Test de seuil pour la méthode de segmentation par ratio : Seuil = 0.8



Exemples de profils journaliers de prévision

Les figures suivantes présentent des profils journaliers permettant de comparer les résultats du modèle P_{moy} aux différents modèles développés dans le chapitre 4.

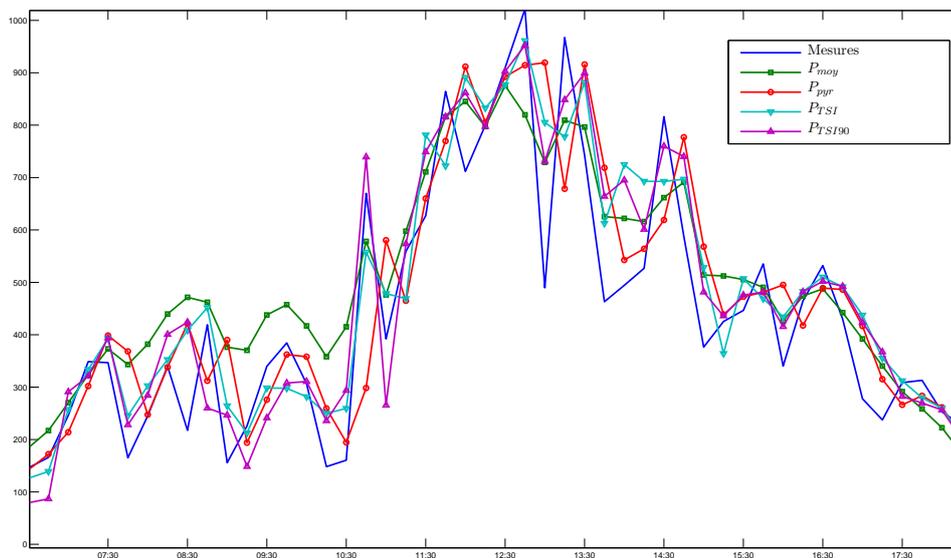


FIGURE B.1 – Profil journalier, h=15 minutes.

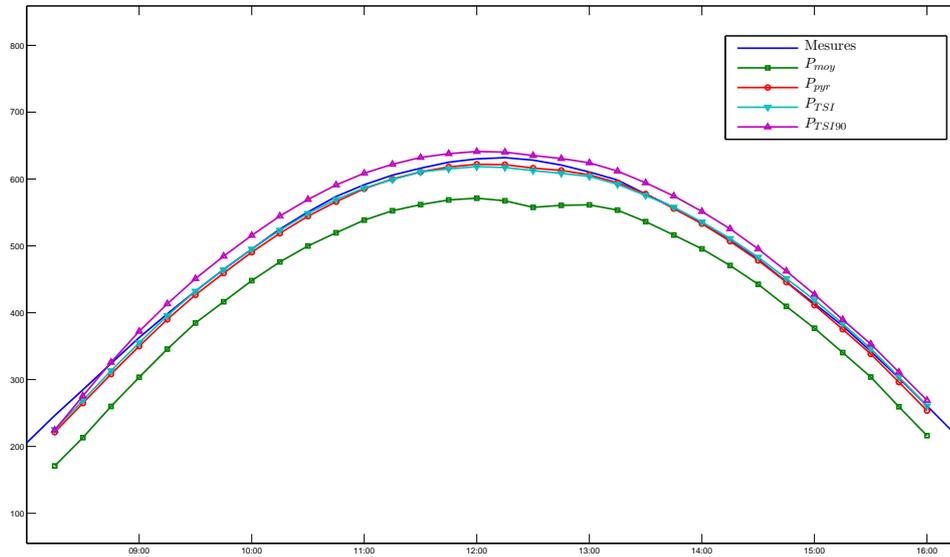


FIGURE B.2 – Profil journalier, h=15 minutes.

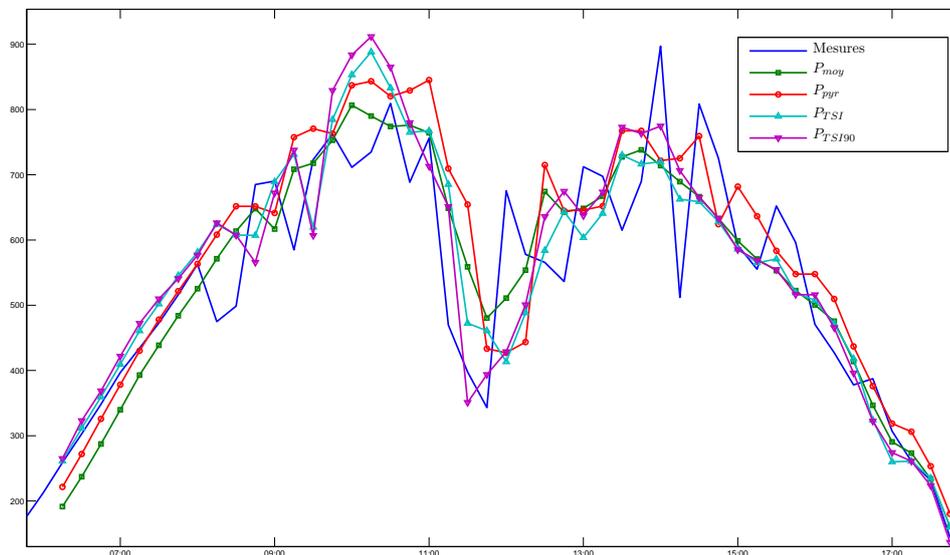


FIGURE B.3 – Profil journalier, h=30 minutes.

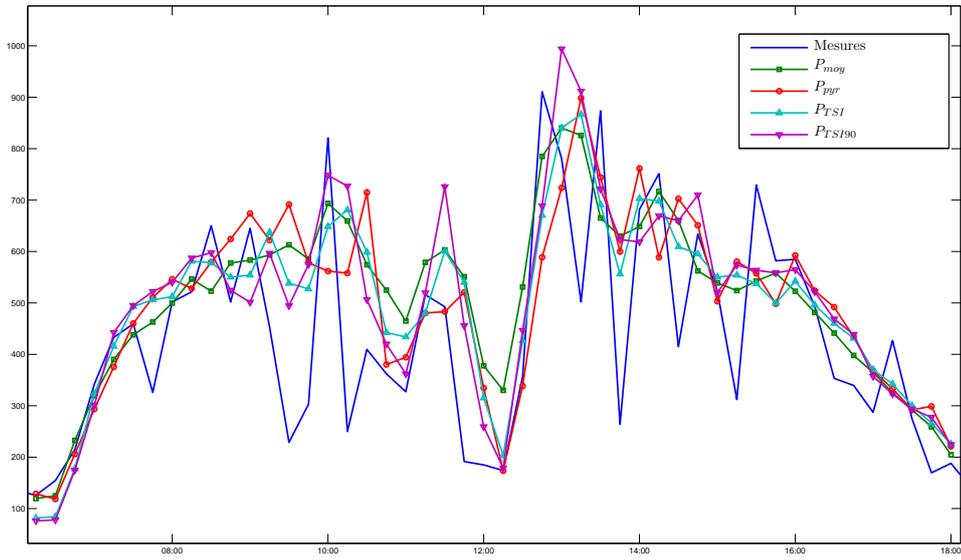


FIGURE B.4 – Profil journalier, h=30 minutes.

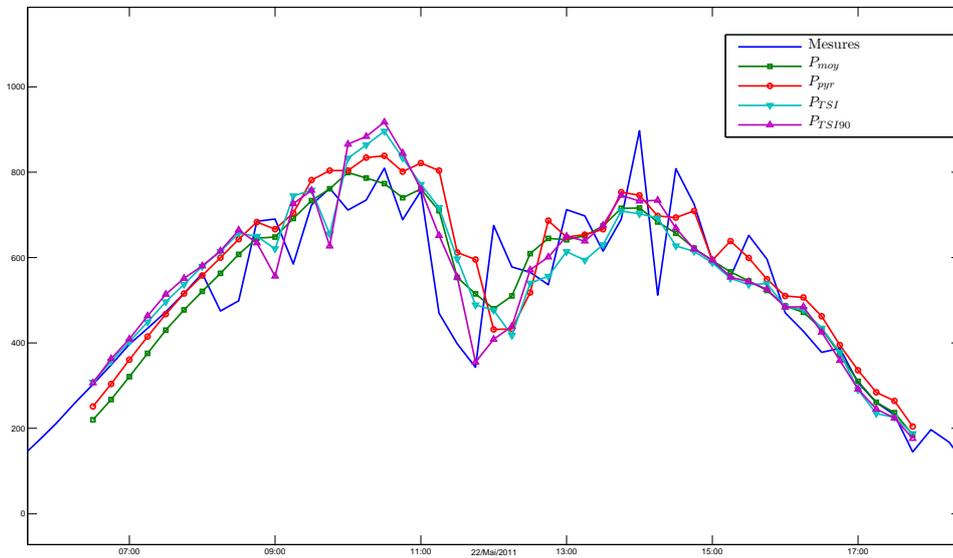
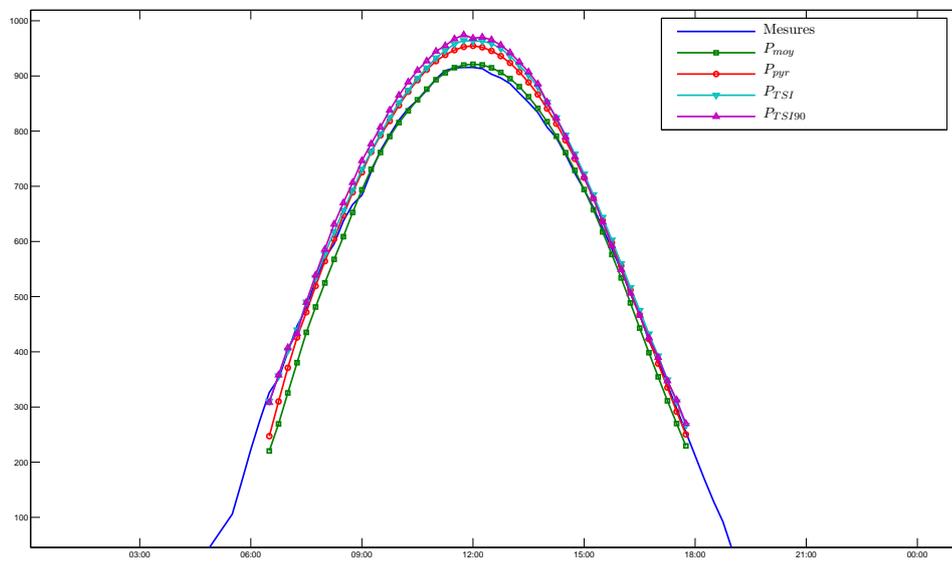


FIGURE B.5 – Profil journalier, h=45 minutes.

FIGURE B.6 – Profil journalier, $h=45$ minutes.

Publications de l'auteur

C.1 Publication en rapport avec la thèse

C.1.1 Publication en revue internationale

C.1.1.1 Publication acceptée

- R. Dambreville, P. Blanc, J. Chanussot, D. Boldo. Very short term forecasting of the Global Horizontal Irradiance using a Spatio-temporal Autoregressive Model. Accepted at *Renewable Energy*
- G. A. Licciardi, R. Dambreville, J. Chanussot. SPATIO-TEMPORAL PATTERN RECOGNITION AND NONLINEAR PCA FOR GLOBAL HORIZONTAL IRRADIANCE FORECASTING. Accepted at *Geoscience and Remote Sensing Letters*

C.1.1.2 Publication soumise

- R. Dambreville, P. Blanc, J. Chanussot, G. A. Licciardi, D. Boldo, S. Dubost. Data fusion for the enhancement of satellite based surface solar irradiance forecasting. Submitted at *Geoscience and Remote Sensing Letters*

C.1.2 Publication en conférence internationale

- R. Dambreville, P. Blanc, J. Chanussot, D. Boldo. Very short term forecasting of the Global Horizontal Irradiance through Helioclim maps. In Renewable Energy Congress (IREC), 2014 5th International, pp. 1-6. IEEE, 2014.

C.2 Brevet

- R. Dambreville, S. Dubost, D. Boldo, P. Blanc, J. Chanussot. Préviation photovoltaïque très court terme utilisant des données satellitaires. Dépôt le 24 Mars 2014.

C.3 Autre publication hors du contexte de la thèse

- A. Villa, J. Chanussot, J. A. Benediktsson, C. Jutten, R Dambreville. (2013). Unsupervised methods for the classification of hyperspectral images with low spatial resolution. *Pattern Recognition*, 46(6), 1556-1568.

Bibliographie

- [Aka69] Hirotugu AKAIKE. “Fitting autoregressive models for prediction”. Dans : *Annals of the institute of Statistical Mathematics* 21.1 (1969), p. 243–247 (cf. p. 71).
- [AW10] Ronald J ADRIAN et Jerry WESTERWEEL. *Particle image velocimetry*. T. 30. Cambridge University Press, 2010 (cf. p. 82).
- [Bau96] Olivier BAUER. “Les échanges océan-atmosphère dans l’Atlantique subtropical nord-est : apports de Meteosat”. Thèse de doct. 1996 (cf. p. 65).
- [BCH96] Hans Georg BEYER, Claudio COSTANZO et Detlev HEINEMANN. “Modifications of the Heliosat procedure for irradiance estimates from satellite images”. Dans : *Solar Energy* 56.3 (1996), p. 207–212 (cf. p. 64).
- [Bis95] Christopher M BISHOP. *Neural networks for pattern recognition*. Oxford university press, 1995 (cf. p. 51).
- [Bla+11] Philippe BLANC et al. “The HelioClim project : Surface solar irradiance data for climate applications”. Dans : *Remote Sensing* 3.2 (2011), p. 343–361 (cf. p. 65, 67).
- [BS97] JA BENEDIKTSSON et JR SVEINSSON. “Feature extraction for multisource data classification with artificial neural networks”. Dans : *International Journal of Remote Sensing* 18.4 (1997), p. 727–740 (cf. p. 51).
- [BSE90] JONA BENEDIKTSSON, Philip H SWAIN et Okan K ERSOY. “Neural network approaches versus statistical methods in classification of multisource remote sensing data”. Dans : *IEEE Transactions on geoscience and remote sensing* 28.4 (1990), p. 540–552 (cf. p. 51).
- [Can+86] Daniel CANO et al. “A method for the determination of the global solar radiation from meteorological satellite data”. Dans : *Solar Energy* 37.1 (1986), p. 31–39 (cf. p. 63–65).
- [Cho+11] Chi Wai CHOW et al. “Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed”. Dans : *Solar Energy* 85.11 (2011), p. 2881–2893 (cf. p. 30, 33, 34).
- [CS08] Josep CALBO et Jeff SABBURG. “Feature extraction from whole-sky ground-based images for cloud-type recognition.” Dans : *Journal of Atmospheric & Oceanic Technology* 25.1 (2008) (cf. p. 39).
- [Dam+14] Romain DAMBREVILLE et al. “Very short term forecasting of the Global Horizontal Irradiance through Helioclim maps”. Dans : *Renewable Energy Congress (IREC), 2014 5th International*. IEEE. 2014, p. 1–6 (cf. p. 60).
- [DDK87] G DEDIEU, PY DESCHAMPS et YH KERR. “Satellite estimation of solar irradiance at the surface of the earth and of surface albedo using a physical model applied to Metcosat Data”. Dans : *Journal of Climate and Applied Meteorology* 26.1 (1987), p. 79–87 (cf. p. 63).

- [Dia+13] Maimouna DIAGNE et al. “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 (cf. p. 3).
- [Ein05] Albert EINSTEIN. “The photoelectric effect”. Dans : *Ann. Phys* 17 (1905), p. 132 (cf. p. 1).
- [FC13] Chia-Lin FU et Hsu-Yung CHENG. “Predicting solar irradiance with all-sky image features via regression”. Dans : *Solar Energy* 97 (2013), p. 537–550 (cf. p. 33, 36, 38, 39).
- [Fon+97] M FONTOYNONT et al. “Satel-Light”. Dans : *Processing of Meteosat data for the production of high quality daylight and solar radiation data available on a World Wide Web Internet server, Mid-term progress report JOR3-CT* (1997), p. 95–0041 (cf. p. 64).
- [FS98] Fabio Del FRATE et Giovanni SCHIAVON. “A combined natural orthogonal functions/neural network technique for the radiometric estimation of atmospheric profiles”. Dans : *Radio Science* 33.2 (1998), p. 405–410 (cf. p. 51).
- [Gau+12] Charlotte GAUCHET et al. “Surface solar irradiance estimation with low-cost fish-eye camera”. Dans : *Workshop on Remote Sensing Measurements for Renewable Energy*. 2012 (cf. p. 30, 31).
- [GDM80] Catherine GAUTIER, Georges DIAK et Serge MASSE. “A simple physical model to estimate incident solar radiation at the surface from GOES satellite data”. Dans : *Journal of Applied Meteorology* 19.8 (1980), p. 1005–1012 (cf. p. 63).
- [Gue08] Christian A GUEYMARD. “REST2 : High-performance solar radiation model for cloudless-sky irradiance, illuminance, and photosynthetically active radiation—Validation with a benchmark dataset”. Dans : *Solar Energy* 82.3 (2008), p. 272–285 (cf. p. 19).
- [Gue93] Christian GUEYMARD. “Critical analysis and performance assessment of clear sky solar irradiance models using theoretical and measured data”. Dans : *Solar Energy* 51.2 (1993), p. 121–138 (cf. p. 20).
- [GWE09] Rafael C GONZALEZ, Richard E WOODS et Steven L EDDINS. *Digital image processing using MATLAB*. T. 2. Gatesmark Publishing Knoxville, 2009 (cf. p. 82).
- [Ham+99] Annette HAMMER et al. “Short-term forecasting of solar radiation : a statistical approach using satellite data”. Dans : *Solar Energy* 67.1 (1999), p. 139–150 (cf. p. 64, 82).
- [HH78] JE HAY et KJ HANSON. “Satellite-based methodology for determining solar irradiance at the ocean surface during gate”. Dans : *Bulletin of the American Meteorological Society*. T. 59. 11. AMER METEOROLOGICAL SOC 45 BEACON ST, BOSTON, MA 02108-3693. 1978, p. 1549–1549 (cf. p. 63).
- [HMS10] Anna HEINLE, Andreas MACKE et Anand SRIVASTAV. “Automatic cloud classification of whole sky images”. Dans : *Atmospheric Measurement Techniques Discussions* 3.1 (2010), p. 269–299 (cf. p. 31, 32, 39, 41, 44).

- [HS81] Berthold K HORN et Brian G SCHUNCK. “Determining optical flow”. Dans : *1981 Technical Symposium East*. International Society for Optics et Photonics. 1981, p. 319–331 (cf. p. 82).
- [HSD73] Robert M HARALICK, Karthikeyan SHANMUGAM et Its’ Hak DINSTEIN. “Textural features for image classification”. Dans : *Systems, Man and Cybernetics, IEEE Transactions on* 6 (1973), p. 610–621 (cf. p. 42).
- [Hsu+92] Shin-Yi HSU et al. “Comparative analysis of five neural network models”. Dans : *Remote Sensing Reviews* 6.1 (1992), p. 319–329 (cf. p. 51).
- [Hue+96] Sylvie HUET et al. *Statistical tools for nonlinear regression*. Springer, 1996 (cf. p. 38).
- [Ieh+97] A IEHLÉ et al. “Meteosat : A valuable tool for agro-meteorology”. Dans : *Final report for the European Commission, Joint Research Centre, Ispra, Italy* (1997) (cf. p. 64).
- [Ine06] Pierre INEICHEN. “Comparison of eight clear sky broadband models against 16 independent data banks”. Dans : *Solar Energy* 80.4 (2006), p. 468–478 (cf. p. 19).
- [IP99] Pierre INEICHEN et Richard PEREZ. “Derivation of cloud index from geostationary satellites and application to the production of solar irradiance and daylight illuminance data”. Dans : *Theoretical and Applied Climatology* 64.1-2 (1999), p. 119–130 (cf. p. 64).
- [IPC13] Rich H INMAN, Hugo TC PEDRO et Carlos FM COIMBRA. “Solar forecasting methods for renewable energy integration”. Dans : *Progress in Energy and Combustion Science* 39.6 (2013), p. 535–576 (cf. p. 19, 62, 63).
- [Kaz+12] A KAZANTZIDIS et al. “Cloud detection and classification with the use of whole-sky ground-based images”. Dans : *Atmospheric Research* 113 (2012), p. 80–88 (cf. p. 31, 32).
- [KL11] Greg KOPP et Judith L LEAN. “A new, lower value of total solar irradiance : Evidence and climate significance”. Dans : *Geophysical Research Letters* 38.1 (2011) (cf. p. 8).
- [Lef+13] Mireille LEFEVRE et al. “McClear : a new model estimating downwelling solar radiation at ground level in clear-sky conditions”. Dans : *Atmospheric Measurement Techniques Discussions* 6.2 (2013), p. 3367–3405 (cf. p. 19).
- [Liu+13] Shuang LIU et al. “Tensor Ensemble of Ground-Based Cloud Sequences : Its Modeling, Classification, and Synthesis”. Dans : (2013) (cf. p. 32).
- [LLY11] Qingyong LI, Weitao LU et Jun YANG. “A Hybrid Thresholding Algorithm for Cloud Detection on Ground-Based Color Images.” Dans : *Journal of Atmospheric & Oceanic Technology* 28.10 (2011) (cf. p. 35).
- [LSC+06] Charles N LONG, JM SABBURG, J CALBÓ et al. “Retrieving cloud characteristics from ground-based daytime color all-sky images.” Dans : *Journal of Atmospheric & Oceanic Technology* 23.5 (2006) (cf. p. 30, 31, 36, 44).
- [Mal05] S MALARDEL. *Fondamentaux de météorologie : à l’école du temps*. 2005 (cf. p. 7, 10).

- [MC04] N MORI et K CHANG. “Introduction to MPIV-PIV toolbox in MATLAB-version 0.965”. Dans : *See http://www.oceanwave.jp/software/mpiv_doc/index.html* (2004) (cf. p. 35).
- [MC13a] Ricardo MARQUEZ et Carlos FM COIMBRA. “Intra-hour DNI forecasting based on cloud tracking image analysis”. Dans : *Solar Energy* 91 (2013), p. 327–336 (cf. p. 30, 31, 33, 35, 38).
- [MC13b] Ricardo MARQUEZ et Carlos FM COIMBRA. “Proposed Metric for Evaluation of Solar Forecasting Models”. Dans : *Journal of solar energy engineering* 135.1 (2013) (cf. p. 21, 23, 24).
- [MCK13] Patrick MATHIESEN, Craig COLLIER et Jan KLEISSL. “A high-resolution, cloud-assimilating numerical weather prediction model for solar irradiance forecasting”. Dans : *Solar Energy* 92 (2013), p. 47–61 (cf. p. 3).
- [MDV87] S MARULLO, G DALU et A VIOLA. “Incident short-wave radiation at the surface from meteosat data”. Dans : *Il Nuovo Cimento C* 10.1 (1987), p. 77–90 (cf. p. 63).
- [Mou+89] Guy MOUSSU et al. “A method for the mapping of the apparent ground brightness using visible images from geostationary satellites”. Dans : *Remote Sensing* 10.7 (1989), p. 1207–1225 (cf. p. 64).
- [MPC13] Ricardo MARQUEZ, Hugo TC PEDRO et Carlos FM COIMBRA. “Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to ANNs”. Dans : *Solar Energy* 92 (2013), p. 176–188 (cf. p. 62, 64).
- [MR84] Werner MÖSER et Ehrhard RASCHKE. “Incident solar radiation over Europe estimated from METEOSAT data”. Dans : *Journal of Climate and Applied Meteorology* 23.1 (1984), p. 166–170 (cf. p. 63).
- [Mue+04] RW MUELLER et al. “Rethinking satellite-based solar irradiance modelling : The SOLIS clear-sky module”. Dans : *Remote sensing of Environment* 91.2 (2004), p. 160–174 (cf. p. 19).
- [NRF93a] M NOIA, CF RATTO et R FESTA. “Solar irradiance estimation from geostationary satellite data : I. Statistical models”. Dans : *Solar Energy* 51.6 (1993), p. 449–456 (cf. p. 62).
- [NRF93b] M NOIA, CF RATTO et R FESTA. “Solar irradiance estimation from geostationary satellite data : II. Physical models”. Dans : *Solar Energy* 51.6 (1993), p. 457–465 (cf. p. 63).
- [Obr90] Dominique OBRECHT. “Météorologie solaire et images satellitaires : cartographie du rayonnement solaire, détermination de l’albédo des sols et évaluation de l’enneigement”. Thèse de doct. Université de Nice Sophia-Antipolis, 1990 (cf. p. 64).
- [Oum+09] Armel OUMBE et al. “A new method for estimating solar energy resource”. Dans : *Proceedings of the ISRSE 33* (2009) (cf. p. 64).
- [Pen+13] Zhenzhou PENG et al. “Solar irradiance forecast system based on geostationary satellite”. Dans : *Smart Grid Communications (SmartGridComm), 2013 IEEE International Conference on*. IEEE. 2013, p. 708–713 (cf. p. 64).

- [Per+02] Richard PEREZ et al. “A new operational model for satellite-derived irradiances : description and validation”. Dans : *Solar Energy* 73.5 (2002), p. 307–317 (cf. p. 63).
- [Pet06] Grant William PETTY. *A first course in atmospheric radiation*. Sundog Pub, 2006 (cf. p. 31).
- [Pfi+03] G PFISTER et al. “Cloud coverage based on all-sky imaging and its impact on surface solar irradiance.” Dans : *Journal of Applied Meteorology* 42.10 (2003) (cf. p. 30).
- [RBW00] Christelle RIGOLLIER, Olivier BAUER et Lucien WALD. “RADIATION ATLAS — WITH RESPECT TO THE HELIOSAT METHOD”. Dans : *Solar Energy* 68.1 (2000), p. 33–48 (cf. p. 19).
- [Rig+02] Christelle RIGOLLIER et al. “The Operational Calibration of Images Taken in the Visible Channel of the Meteosat Series of Satellites.” Dans : *Journal of Atmospheric & Oceanic Technology* 19.9 (2002) (cf. p. 65).
- [RLW04] Christelle RIGOLLIER, Mireille LEFÈVRE et Lucien WALD. “The method Heliosat-2 for deriving shortwave solar radiation from satellite images”. Dans : *Solar Energy* 77.2 (2004), p. 159–169 (cf. p. 64, 65).
- [Roe+11] A. ROESCH et al. “Assessment of BSRN radiation records for the computation of monthly means”. Dans : *Atmospheric Measurement Techniques* 4 (2011), p. 339–354 (cf. p. 15).
- [Sch+11] Kathrin SCHUMANN et al. “Improving satellite-derived solar resource analysis with parallel ground-based measurements”. Dans : *ISES Solar World Congress, Kassel, Germany*. 2011 (cf. p. 68).
- [SL12] George AF SEBER et Alan J LEE. *Linear regression analysis*. T. 936. John Wiley & Sons, 2012 (cf. p. 38).
- [SLT01] DW SLATER, CN LONG et TP TOOMAN. “Total sky imager/whole sky imager cloud fraction comparison”. Dans : *Eleventh ARM Science Team Meeting Proceedings, Atlanta, Georgia*. 2001 (cf. p. 30).
- [Vai07] PP VAIDYANATHAN. “The theory of linear prediction”. Dans : *Synthesis Lectures on Signal Processing* 2.1 (2007), p. 1–184 (cf. p. 71).
- [VBP13] Christophe VERNAY, Philippe BLANC et Sébastien PITAVAL. “Characterizing measurements campaigns for an innovative calibration approach of the global horizontal irradiation estimated by HelioClim-3”. Dans : *Renewable Energy* 57 (2013), p. 339–347 (cf. p. 68).
- [Zel+92] A ZELENKA et al. “Techniques for supplementing solar radiation network data”. Dans : *Report IEA Task 9* (1992) (cf. p. 64).
- [Zel+94] Andreas ZELL et al. “SNNS (stuttgart neural network simulator)”. Dans : *Neural Network Simulation Environments*. Springer, 1994, p. 165–186 (cf. p. 54).

Résumé — L'intégration des énergies intermittentes sur le réseau électrique existant soulève des problèmes de stabilité de la balance consommation/production. Afin de limiter les risques, la loi autorise un taux de pénétration maximum de ces énergies de 30% de la puissance installée. Afin de pallier cette limitation, deux pistes sont envisagées, le stockage d'énergie et la prévision de production. Les travaux menés dans cette thèse s'inscrivent dans la seconde catégorie et se concentrent sur la prévision du rayonnement global au sol, ressource principale des systèmes de production d'énergie photovoltaïque. Dans l'objectif d'une prévision à très court terme (infra-horaire), la problématique développée concerne la fusion d'informations provenant d'une part d'observations satellitaires et de l'autre d'observations in-situ telles que des images hémisphériques. L'approche suivie se veut progressive et s'articule autour de 4 grand axes. Le premier chapitre énonce les définitions et les enjeux principaux liés à l'étude du GHI en décrivant les différents facteurs physiques ayant des interactions sur sa mesure. Le second chapitre permet d'évaluer le potentiel des images hémisphériques pour l'estimation du rayonnement global. On y développe une méthode d'estimation du GHI basée sur l'utilisation d'un réseau de neurones artificiels permettant d'effectuer une régression non linéaire entre des caractéristiques choisies de l'image et le rayonnement mesuré sur site par un pyranomètre. Le troisième chapitre concerne l'utilisation des images satellitaires pour la prévision à très court terme du rayonnement global. L'originalité des méthodes proposées provient de l'utilisation en entrées de cartes de rayonnement dérivées des images satellitaires via une méthode externe. L'utilisation de ces cartes de rayonnement permet la construction de modèles linéaires simples (modèles auto-régressifs) grâce à l'homogénéité des données utilisées. Cependant, leur pertinence pour le calcul de champ de vecteurs a également été prouvé et validé dans un second modèle. La comparaison des deux modèles ainsi créés à partir d'imagerie satellitaire a permis de caractériser les forces et faiblesses de ces modèles. L'intérêt de l'observation satellitaire réside dans l'anticipation du déplacement des masses nuageuse. Cependant, un biais non systématique persiste dans la conversion des valeurs des pixels en rayonnement. Ce biais est inhérent à la résolution spatio-temporelle de ces observations. Étant donné ces considérations, le chapitre 4 présente alors une méthode d'intégration des données acquises par l'imagerie hémisphérique, disposant une meilleure résolution spatio-temporelle, dans les résultats de prévision satellitaires précédemment évoqués. On joint alors les caractéristiques retenues au chapitre 2 dans un réseau de neurone avec la prévision pour produire une prévision dont le biais est largement réduit. L'utilisation de caractéristiques dérivées de l'imagerie hémisphérique à la place de simple mesures du pyranomètre est justifiée par l'effet de persistance non souhaité apportées par ces dernières. Ainsi, après avoir étudié chaque source d'information séparément, on a pu développer un modèle leur permettant de se compléter.

Mots clés : Réseau de neurones artificiels, traitement d'image, apprentissage, rayonnement solaire, télédétection, image satellite.

Abstract — To handle the integration of intermittent energies to the existing grid, managers require more and more accurate tools to forecast the primary resources. This thesis focuses on the very short term forecast of the global horizontal irradiance (GHI), main input for most photovoltaic power plant. We chose to use both ground based images from hemispherical cameras and satellite images to provide a forecasting tool. In the first hand we present a novel approach to estimate the GHI using ground based images. On second hand, we propose several satellite based methods to forecast the GHI up to one hour. Finally, we develop a new method allowing us to merge both data in order to benefit from their respective advantages. All the methods were tested against real data acquired on the SIRTAsite, Polytechnique campus.

Keywords : Artificial neural networks, image processing, machine learning, solar irradiance, remote sensing, satellite images.

Grenoble Images Paroles Signal et Automatique, 11 Rue des Mathématiques, 38400
Saint-Martin-d'Hères, France
GRENOBLE