



**HAL**  
open science

# Synergie des mesures pyranométriques et des images hémisphériques in-situ avec des images satellites météorologiques pour la prévision photovoltaïque

Loïc Vallance

► **To cite this version:**

Loïc Vallance. Synergie des mesures pyranométriques et des images hémisphériques in-situ avec des images satellites météorologiques pour la prévision photovoltaïque. *Météorologie*. Université Paris sciences et lettres, 2018. Français. NNT : 2018PSLEM064 . tel-02097021

**HAL Id: tel-02097021**

**<https://pastel.hal.science/tel-02097021v1>**

Submitted on 11 Apr 2019

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

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



**THÈSE DE DOCTORAT**  
**DE L'UNIVERSITÉ PSL**

Préparée à MINES ParisTech

**Synergie des mesures pyranométriques et des images hémisphériques *in situ* avec des images satellites météorologiques pour la prévision photovoltaïque**

Soutenue par

**Loïc VALLANCE**

Le 9 novembre 2018

Ecole doctorale n° 432

**Sciences des métiers  
de l'ingénieur**

Spécialité

**Énergétique et procédés**

**Composition du jury :**

|  |                           |
|--|---------------------------|
| Jocelyn, CHANUSSOT<br>Professeur, Université de Grenoble | <i>Président du jury</i>  |
| Philippe, LAURET<br>Professeur, Université de la Réunion | <i>Rapporteur</i>         |
| Giuseppe Marco, TINA<br>Professeur, Université de Catane | <i>Rapporteur</i>         |
| Jordi, BADOSA<br>Docteur, LMD – école polytechnique      | <i>Examineur</i>          |
| Stéphanie, DUBOST<br>Docteur, EDF R&D                    | <i>Examineur</i>          |
| Bruno, CHARBONNIER<br>Ingénieur, EDF R&D                 | <i>Invité</i>             |
| Philippe, BLANC<br>Professeur, Mines ParisTech           | <i>Directeur de thèse</i> |







---

# Remerciements

---

Mes remerciements s'adressent avant tout à M. Philippe Blanc du centre Observation Impacts Énergie (O.I.E) de Mines ParisTech, dont le rôle de directeur de thèse a été exercé avec brio. Ses compétences scientifiques indéniables, son implication pédagogique passionnée et sa profonde bienveillance ont éclairé mon apprentissage et mon expérience tout au long de ces trois années.

Je tiens à remercier l'ensemble des membres du jury pour m'avoir fait l'honneur d'assister à ma soutenance de thèse. Les deux rapporteurs de cette thèse, M. Philippe Lauret et M. Giuseppe Marco Tina, ainsi que les examinateurs externes M. Jocelyn Chanussot et M. Jordi Badosa pour leur évaluation minutieuse de mon travail et pour avoir animé des discussions à la fois passionnantes et pertinentes à la suite de mon exposé.

Je remercie également l'équipe d'encadrement industriel à EDF qui m'a accompagné dans la réalisation de ces travaux. M. Bruno Charbonnier pour son suivi assidu, ses conseils avisés et pour m'avoir embarqué dans cette aventure dès mon stage d'ingénieur. Merci à M<sup>me</sup> Stéphanie Dubost qui porte le projet et qui a su me connecter aux réalités opérationnelles. M. Dominique Demengel pour m'avoir fourni des données *in situ* de qualité et m'avoir aidé à comprendre les subtilités de leur acquisition. Je remercie également M. Nicolas Paul, M. Vincent Le Guen et M. Paul Kassis pour leurs contributions aux discussions d'avancement de ma thèse.

Le centre de recherche O.I.E de Mines ParisTech m'a accueilli de manière très amicale et j'en remercie tous ses membres. Je remercie mesdames et messieurs Lucien Wald, Thierry Ranchin, Yves-Marie Saint-Drenan, Lionel Menard, Benoit Gschwind, Paula Perez-Lopez, Isabelle Blanc, Mireille Lefevre et Thibaut Barbier pour leurs conseils scientifiques justes et leur bonne humeur. Merci à Sandra Hassan pour son aide administrative. Merci aux doctorants d'O.I.E, Marc Cloarec, Mélodie Trolliet, Romain Besseau, Benoit Tournadre, Hadrien Verbois ainsi qu'aux doctorants du centre voisin PERSEE qui ont été d'un grand soutien au quotidien et qui ont contribué au bien-être ambiant.

Il m'est impossible de ne pas citer mes proches, même si ce que je leur dois dépasse largement le cadre de ces remerciements. Je remercie mes amis de l'ENSEEIH, toujours fidèles et enjoués. Merci à ma famille, mes parents, David, Marie, Nelly et Charlotte pour leur soutien inconditionnel et immuable.

---

# Table des matières

---

|  |           |
|--|-----------|
| <b>TABLE DES MATIÈRES</b> .....  | <b>6</b>  |
| <b>NOMENCLATURE</b> .....  | <b>9</b>  |
| <b>CHAPITRE I - INTRODUCTION</b> .....   | <b>11</b> |
| I.1 CONTEXTE ET ENJEUX .....   | 11        |
| I.2 LA NATURE VARIABLE DE L'ÉNERGIE PHOTOVOLTAÏQUE.....  | 13        |
| I.3 INTÉRÊT DE LA PRÉVISION PHOTOVOLTAÏQUE POUR LE RÉSEAU ÉLECTRIQUE.....  | 16        |
| I.4 HORIZON ET RÉOLUTION TEMPORELLE DE PRÉVISION .....   | 18        |
| I.5 PRÉVISION LOCALE ET AGRÉGÉE.....   | 19        |
| I.6 MÉTHODES DE PRÉVISION EXISTANTES EN FONCTION DES ÉCHELLES SPATIALES ET TEMPORELLES CONSIDÉRÉES .....   | 21        |
| I.7 OBJECTIFS ET DÉMARCHÉ DE LA THÈSE .....  | 22        |
| I.8 STRUCTURE DU MANUSCRIT .....   | 23        |
| <b>CHAPITRE II - ANALYSE DES PERFORMANCES DES PRÉVISIONS SOLAIRES</b> .....  | <b>25</b> |
| II.1 INTRODUCTION À L'ANALYSE DES PERFORMANCES DANS LE DOMAINE DE LA PRÉVISION SOLAIRE .....   | 25        |
| II.2 LA MESURE PYRANOMÉTRIQUE <i>IN SITU</i> , UNE MESURE DE RÉFÉRENCE POUR L'ANALYSE DES PERFORMANCES .....   | 27        |
| II.2.1 <i>Moyens de mesures du rayonnement solaire existants</i> .....   | 27        |
| II.2.2 <i>Instrumentation déployée dans le cadre de l'étude</i> .....  | 29        |
| II.2.3 <i>Analyses de qualité des mesures pyranométriques</i> .....  | 30        |
| II.2.4 <i>Visualisations des données favorisant l'identification d'anomalies sur la mesure in situ</i> .....   | 31        |
| II.3 LIMITATION DES MÉTHODES STANDARDS D'ÉVALUATION DES PERFORMANCES DES PRÉVISIONS .....  | 33        |
| II.3.1 <i>Métriques usuelles</i> .....   | 34        |
| II.3.2 <i>Pratiques pertinentes pour l'application des métriques statistiques</i> .....  | 35        |
| II.3.3 <i>Le cas de la métrique RMSE</i> .....   | 38        |
| II.3.4 <i>Rappel des attentes des utilisateurs</i> .....   | 41        |
| II.4 PRÉSENTATION DE MÉTHODES EXPLORATOIRES QUI TENTENT DE COMBLER LE MANQUE LIÉ AUX MÉTHODES EXISTANTES ET EN RÉPONSE AUX BESOINS UTILISATEURS..... | 42        |
| II.4.1 <i>Suivi des rampes</i> .....   | 42        |
| II.4.2 <i>Erreur temporelle : le Temporal Distortion Index (TDI)</i> .....   | 44        |
| II.5 MÉTRIQUES PROPOSÉES.....  | 47        |
| II.5.1 <i>Quantifier la capacité de la prévision à suivre les rampes</i> .....   | 47        |
| II.5.2 <i>Quantifier l'alignement temporel de la prévision par rapport à la mesure</i> .....   | 51        |
| II.6 CARACTÉRISATION CATÉGORIELLE DES PERFORMANCES : VERS UNE INDÉPENDANCE DE L'ENSEMBLE DES DONNÉES UTILISÉES .....                                 | 54        |
| II.6.1 <i>Classification d'instant</i> .....   | 54        |
| II.6.2 <i>Classifications de jours</i> .....   | 58        |
| II.6.3 <i>Autres classifications pertinentes</i> .....   | 61        |
| II.7 PRÉSENTATION DE L'ANALYSE DES PERFORMANCES POUR TROIS MÉTHODES DE PRÉVISION DE RÉFÉRENCE .....  | 61        |
| II.7.1 <i>Suggestion de représentation graphique des résultats et métriques sélectionnées</i> .....  | 61        |
| II.7.2 <i>Trois méthodes de références utilisées pour l'illustration des avantages et inconvénients des métriques sélectionnées</i> .....            | 62        |
| II.7.3 <i>Résultats globaux et par catégorie</i> .....   | 64        |
| II.8 CONCLUSIONS DU CHAPITRE.....  | 67        |
| <b>CHAPITRE III - OBSERVATION DES COMPLÉMENTARITÉS DES SOURCES D'INFORMATION</b> .....   | <b>69</b> |
| III.1 COMPLEXITÉ DE LA SCÈNE 3D ET DE SES OBSERVATIONS .....   | 69        |
| III.1.1 <i>Une vision partielle des nuages pour chaque point d'observation</i> .....   | 69        |
| III.1.2 <i>Taille de la couverture nuageuse perçue par l'image satellite</i> .....   | 70        |
| III.1.3 <i>Image hémisphérique : la distorsion optique au service d'une vision périphérique du ciel</i> .....  | 72        |

|   |   |            |
|---|---|------------|
| III.1.4   | <i>Perception d'une grille horizontale par les deux types d'images en fonction de son altitude imposée.</i> | 74         |
| III.1.5   | <i>Effet parallaxe.</i>   | 75         |
| III.2   | DIFFÉRENCE DE RÉOLUTION SPATIALE ENTRE IMAGES HÉMISPHERIQUES ET IMAGES SATELLITES                           | 77         |
| III.2.1   | <i>Résolution spatiale et taille du pixel.</i>  | 77         |
| III.2.2   | <i>Taille du pixel satellite.</i>   | 78         |
| III.2.3   | <i>Taille du pixel de l'image hémisphérique.</i>  | 79         |
| III.3   | DIFFÉRENCE DE RÉOLUTION TEMPORELLE DES INFORMATIONS   | 79         |
| III.3.1   | <i>Instant d'acquisition des pixels de l'image satellite</i>  | 79         |
| III.3.2   | <i>Résolution temporelle des informations in situ.</i>  | 80         |
| III.4   | OBSERVATION DES COMPLÉMENTARITÉS DES DIFFÉRENTES SOURCES D'INFORMATION                                      | 81         |
| III.4.1   | <i>Ciel clair.</i>  | 81         |
| III.4.2   | <i>Ciel couvert.</i>  | 85         |
| III.4.3   | <i>Ciel variable.</i>   | 87         |
| III.4.4   | <i>Ressemblance multiéchelle de la couverture nuageuse.</i>   | 90         |
| III.4.5   | <i>Nets dépassements de la courbe de ciel clair.</i>  | 91         |
| III.4.6   | <i>Situations propices à l'exploitation de complémentarités.</i>  | 95         |
| III.4.7   | <i>Anomalies ou événements limitant l'exploitation des données.</i>   | 99         |
| III.5   | VERS L'EXPLOITATION DES COMPLÉMENTARITÉS  | 101        |
| <b>CHAPITRE IV - SIMULATEUR GÉOMÉTRIQUE 3D : VALIDATION SPATIALE DE LA GÉOLOCALISATION.....</b> |   | <b>103</b> |
| IV.1  | L'ESTIMATION DE L'OMBRE PORTÉE DES NUAGES : UNE ÉTAPE DU PROCESSUS PRÉVISIONNEL                             | 103        |
| IV.2  | MÉTHODES DE GÉOLOCALISATION DES NUAGES PAR PHOTOGRAMMÉTRIE  | 106        |
| IV.3  | OBJECTIFS D'UNE SIMULATION DE L'APPROCHE  | 108        |
| IV.4  | DESCRIPTION DU SIMULATEUR   | 109        |
| IV.4.1  | <i>Choix de la représentation 3D des nuages.</i>  | 109        |
| IV.4.2  | <i>Génération d'images synthétiques à partir d'une situation nuageuse donnée.</i>                           | 110        |
| IV.4.3  | <i>Application de la méthode de géolocalisation des nuages en 3D.</i>                                       | 114        |
| IV.4.4  | <i>Comparaison des ombres portées au sol.</i>   | 118        |
| IV.5  | ANALYSE DES RÉSULTATS   | 120        |
| IV.5.1  | <i>Analyse statique.</i>  | 120        |
| IV.5.1  | <i>Analyse dynamique.</i>   | 125        |
| IV.6  | CONCLUSIONS DU CHAPITRE   | 133        |
| <b>CHAPITRE V - APPLICATION DE LA GÉOLOCALISATION 3D DES NUAGES AUX DONNÉES RÉELLES .....</b>   |   | <b>135</b> |
| V.1   | RECHERCHE D'UN RÉFÉRENTIEL COMMUN.....  | 135        |
| V.1.1   | <i>Rappel du référentiel géométrique commun : l'espace 3D.</i>  | 136        |
| V.1.2   | <i>Référentiel temporel commun : harmonisation des pas de temps à la minute.</i>                            | 136        |
| V.2   | SPÉCIFICITÉS DE LA GÉOLOCALISATION DES NUAGES À PARTIR DES DONNÉES RÉELLES                                  | 138        |
| V.2.1   | <i>Segmentation des images.</i>   | 138        |
| V.2.2   | <i>Les trois versions de la méthode de géolocalisation dans le cas réel.</i>                                | 144        |
| V.2.3   | <i>Analyse qualitative de la géolocalisation sur le jour d'exemple.</i>                                     | 147        |
| V.3   | ESTIMATION DU RAYONNEMENT SOLAIRE SPATIALEMENT ÉTENDU   | 152        |
| V.3.1   | <i>Relation a priori.</i>   | 152        |
| V.3.2   | <i>Utilisation de l'indice de ciel clair HelioClim.</i>   | 153        |
| V.3.3   | <i>Régression.</i>  | 154        |
| V.4   | PRÉVISION DU RAYONNEMENT SOLAIRE SPATIALEMENT ÉTENDU  | 157        |
| V.4.1   | <i>Le principe.</i>   | 157        |
| V.4.2   | <i>Propagation des volumes potentiellement nuageux.</i>   | 159        |
| V.4.3   | <i>Résultats sur le jour d'exemple.</i>   | 160        |
| V.5   | CONCLUSIONS DU CHAPITRE   | 164        |
| <b>CHAPITRE VI - CONCLUSIONS ET PERSPECTIVES .....</b>  |   | <b>167</b> |
| VI.1  | CONCLUSIONS   | 167        |
| VI.2  | PERSPECTIVES  | 170        |
| VI.2.1  | <i>L'analyse des performances des méthodes de prévision.</i>  | 170        |
| VI.2.2  | <i>Le simulateur géométrique.</i>   | 171        |
| VI.2.3  | <i>L'application de la fusion pour une géolocalisation des nuages dans le cas réel.</i>                     | 172        |

---

|  |            |
|--|------------|
| <b>ANNEXE A1 - ESTIMATION DU MOUVEMENT DES NUAGES.....</b>                           | <b>175</b> |
| A1.1. TECHNIQUES D'ESTIMATION DU MOUVEMENT .....                                     | 175        |
| A1.1.1. <i>Appariement par corrélation</i> .....                                     | 175        |
| A1.1.2. <i>Flot optique</i> .....  | 176        |
| A1.2. BANC D'ESSAI POUR L'ESTIMATION DES CMV .....                                   | 177        |
| A1.2.1. <i>Un mouvement artificiel des nuages imposé aux images satellites</i> ..... | 177        |
| A1.2.2. <i>Imposé aux images hémisphériques</i> .....                                | 180        |
| <b>ANNEXE A2 - SUR-ÉCHANTILLONNAGE TEMPOREL DES IMAGES SATELLITES.....</b>           | <b>183</b> |
| A2.1. MÉTHODE DE SUR-ÉCHANTILLONNAGE CAUSALE.....                                    | 184        |
| A2.2. MÉTHODE DE SUR-ÉCHANTILLONNAGE NON CAUSALE .....                               | 185        |
| <b>RÉFÉRENCES .....</b>  | <b>187</b> |

# Nomenclature

## Liste des principaux opérateurs

| Opérateur | Signification       |
|-----------|---------------------|
| $ X $     | Valeur absolue de X |
| $\bar{X}$ | Valeur moyenne de X |
| $\hat{X}$ | Valeur estimée de X |

## Liste des principaux acronymes

| Acronyme | Signification anglaise                                    | Signification française   |
|----------|---|---|
| CAMS     | <i>Copernicus Atmospheric Monitoring Services</i>         | Services Copernicus de surveillance de l'atmosphère               |
| CBH      | <i>Cloud Base Height</i>                                  | Altitude de la base des nuages                                    |
| CMV      | <i>Cloud Motion Vectors</i>                               | Vecteurs de déplacement des nuages                                |
| CTH      | <i>Cloud Top Height</i>                                   | Altitude du sommet des nuages                                     |
| DHI      | <i>Diffuse Horizontal Irradiance</i>                      | Eclairement diffus intégré sur plan horizontal                    |
| DNI      | <i>Direct Normal Irradiance</i>                           | Eclairement direct en incidence normale                           |
| DTW      | <i>Dynamic Time Warping</i>                               | Adaptation temporelle dynamique                                   |
| ECMWF    | <i>European Centre for Medium-Range Weather Forecasts</i> | Centre européen pour les prévisions météorologiques à moyen terme |
| ESRA     | <i>European Solar Radiation Atlas</i>                     | Atlas solaire européen  |
| GHI      | <i>Global Horizontal Irradiance</i>                       | Eclairement global sur plan horizontal                            |
| GSD      | <i>Granulometric Size Distribution</i>                    | Distribution de la taille granulométrique                         |
| HRV      | <i>High-Resolution Visible</i>                            | Image satellite du visible à haute résolution                     |
| IEA      | <i>International Energy Agency</i>                        | Agence internationale de l'énergie                                |
| KNMI     | <i>Royal Netherlands Meteorological Institute</i>         | Institut royal météorologique des Pays-Bas                        |
| KSI      | <i>Kolmogorov–Smirnov test Integral</i>                   | Test integral de Kolmogorov–Smirnov                               |
| MAE      | <i>Mean Absolute Error</i>                                | Erreur absolue moyenne  |
| MBE      | <i>Mean Bias Error</i>                                    | Biais   |
| MSG      | <i>Meteosat Second Generation</i>                         | Meteosat de seconde génération                                    |
| NWP      | <i>Numerical Weather Prediction</i>                       | Prévision météorologique numérique                                |

|          |  |   |
|----------|--|---|
| ramp-MAD | <i>Ramp Mean Absolute Deviation</i>                  | Moyenne des valeurs absolues des déviations de rampe    |
| RMSE     | <i>Root Mean Square Error</i>                        | Racine carrée de l'erreur quadratique moyenne           |
| SEVIRI   | <i>Spinning Enhanced Visible and InfraRed Imager</i> | Imageur visible et infrarouge tournant embarqué sur MSG |
| TDI      | <i>Temporal Distortion Index</i>                     | Indice de distorsion temporelle                         |
| TDM      | <i>Temporal Distortion Mix</i>                       | Mix de distorsion temporelle                            |
| TOA      | <i>Top-of-Atmosphere Irradiance</i>                  | Eclairement au sommet de l'atmosphère                   |
| UE       | <i>European Union</i>                                | Union Européenne  |
| WMO      | <i>World Meteorological Organization</i>             | Organisation météorologique mondiale                    |

## Liste des principaux symboles

| Symbole             | Description   | Unité                                 |
|---------------------|---|---------------------------------------|
| G                   | Eclairement global sur plan horizontal (GHI)                    | [W/m <sup>2</sup> ]                   |
| G <sub>b</sub>      | Eclairement direct en incidence normale (DNI)                   | [W/m <sup>2</sup> ]                   |
| G <sub>d</sub>      | Eclairement diffus sur plan horizontal (DHI)                    | [W/m <sup>2</sup> ]                   |
| γ <sub>s</sub>      | Angle d'élévation solaire au-dessus de l'horizon                | [°]                                   |
| θ <sub>s</sub>      | Angle zénithal solaire  | [°]                                   |
| k <sub>c</sub>      | Indice de ciel clair  | —                                     |
| k <sub>d</sub>      | Fraction diffuse  | —                                     |
| k <sub>t</sub>      | Indice de clarté  | —                                     |
| G <sub>0</sub>      | Eclairement au sommet de l'atmosphère (TOA)                     | [W/m <sup>2</sup> ]                   |
| ρ                   | Coefficient de corrélation de Pearson                           | —                                     |
| R <sup>2</sup>      | Coefficient de détermination                                    | —                                     |
| S <sub>RMSE</sub>   | Score de Marquez et Coimbra (2012) appliqué sur la RMSE         | —                                     |
| G <sup>CLS</sup>    | Eclairement global théorique en l'absence de nuages             | [W/m <sup>2</sup> ]                   |
| ε                   | Paramètre de sensibilité de l'algorithme <i>Swinging Door</i>   | [W/m <sup>2</sup> ]                   |
| TDI                 | Indice de distorsion temporelle                                 | —                                     |
| ramp <sub>MAD</sub> | Moyenne des valeurs absolues des déviations de rampe (ramp-MAD) | [W.m <sup>-2</sup> .h <sup>-1</sup> ] |
| TDM                 | Mix de distorsion temporelle                                    | —                                     |
| γ <sub>B</sub>      | Fonction d'ouverture de taille B                                | —                                     |

---

# Chapitre I - Introduction

---

## I.1 Contexte et enjeux

La consommation d'énergie primaire dans le monde a plus que doublé depuis 1970 selon les estimations de l'Agence Internationale de l'Énergie (IEA - International Energy Agency, 2017a) pour atteindre, en 2016, un équivalent de 13,7 milliards de tonnes de pétrole par an. Les projections pour les décennies à venir sont délicates mais cette tendance ne devrait pas s'inverser (IEA, 2017b), en particulier au vu des besoins énergétiques croissants des pays émergents et de la croissance démographique mondiale. Aujourd'hui, plus de 80% de la consommation d'énergie primaire mondiale provient des ressources fossiles, dont 32% de consommation de pétrole.

Les chocs pétroliers de 1973 et 1979 ont illustré l'importance de la dépendance énergétique à l'échelle nationale de certains pays. De telles menaces de rupture d'approvisionnement en énergie et l'impact des fortes variations de prix sur les balances des paiements ont encouragé les gouvernements à définir des politiques énergétiques afin de proposer des stratégies d'évolution du secteur énergétique à moyen et long terme.

Lors des décennies qui vont suivre, des préoccupations environnementales liées au réchauffement climatique prennent progressivement de l'importance dans la conscience collective, donnant lieu notamment à la formation du Groupe d'experts Intergouvernemental sur l'Évolution du Climat (GIEC) en 1988, et à une succession de négociations internationales annuelles sur le climat des pays membres de la Convention-Cadre des Nations Unies sur les Changements Climatiques (CCNUCC). Le rapport de 2014 du GIEC confirme par exemple l'influence de l'homme sur le climat, souligne le caractère inédit de l'augmentation des émissions anthropiques de gaz à effet de serre depuis le début de l'ère industrielle et observe les probables répercussions de ce dérèglement climatique sur les systèmes humains et naturels. Dans l'édition 2016 de son mémento sur l'énergie, le CEA indique une augmentation de 56% des gaz à effet de serre depuis 1990, issus, pour 65%, de la combustion des énergies fossiles.

La 21<sup>e</sup> « conférence des parties » (COP) s'est tenue à Paris en décembre 2015, rassemblant les représentants de 196 pays. Ledit « Accord de Paris » qui en a résulté a comme principal objectif de maintenir l'augmentation de la température mondiale « nettement en dessous » de 2°C d'ici à 2100, comparé aux niveaux préindustriels (The Paris Agreement, 2015). Le Rapport Stern (2006) indique que le coût de l'inaction face aux changements climatiques à venir pourrait rapidement atteindre 5 à 20% du PIB annuel mondial, encourageant par conséquent un investissement dans la lutte contre le dérègle-

ment du climat dont le coût moyen pourrait se limiter à 1% du PIB annuel mondial (Cany, 2017).

Afin de satisfaire les objectifs fixés par les signataires, des efforts sont nécessaires pour assurer cette transition dans de nombreux secteurs et en particulier dans celui de la génération d'énergie électrique. Le secteur électrique au sein de l'Union européenne (UE) a en effet un potentiel de réduction des gaz à effet de serre prometteur puisqu'elles pourraient être « quasiment annulées à l'horizon 2050 » (Cany, 2017), avec des efforts suffisants. Comme l'indique la figure 1, la production d'électricité à l'échelle mondiale est à 66% d'origine fossile en 2015 et a été multipliée par quatre par rapport à 1970.

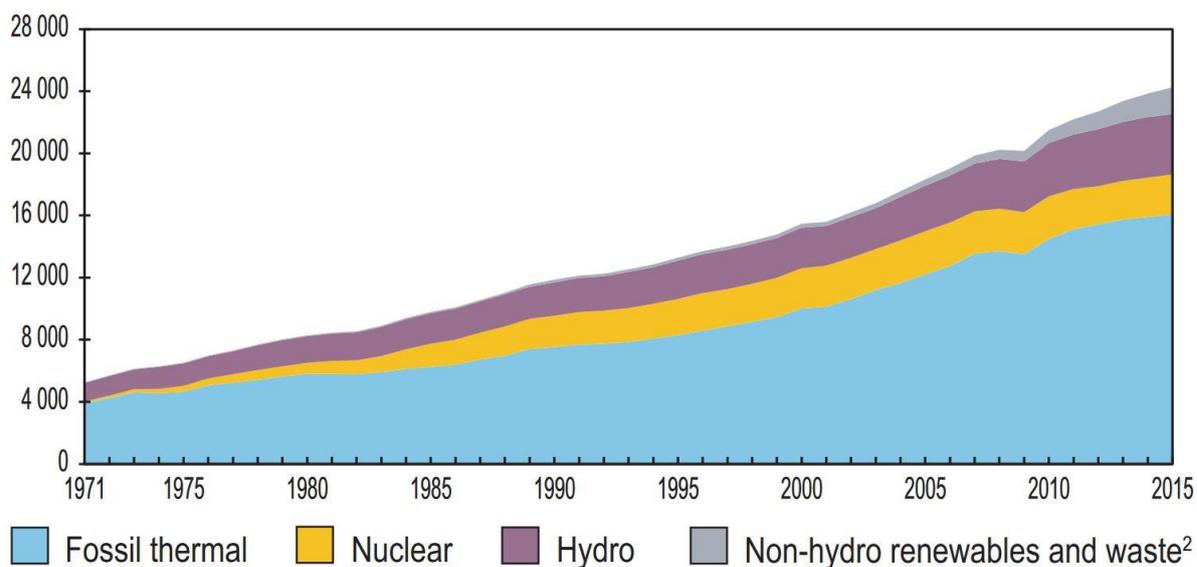


Figure 1 – Évolution de la production d'électricité mondiale (en TWh). *Source* : Key world energy statistics – IEA, 2017.

Au sein de l'UE, l'industrie de l'énergie est le secteur le plus émetteur de gaz à effet de serre en 2014 avec une contribution de 29%, devant la part de 21% en provenance du secteur des transports (Ministère de l'Écologie, 2017). La mutation attendue des véhicules thermiques en véhicules électriques va sans doute accentuer cette contribution. Face à ce constat, l'UE prolonge ses objectifs de l'horizon 2020 à l'horizon 2030 via le Paquet Energie-Climat : 40% de réduction des émissions de gaz à effet de serre par rapport à 1990, une augmentation de 27% de la part des renouvelables dans la consommation énergétique finale brute et une augmentation de 27% de l'efficacité énergétique (Ministère de l'Écologie, 2017).

La France avec sa spécificité nucléaire émet avec des proportions différentes selon les secteurs d'activité : 9% sont attribuées à celui de l'énergie contre 29% pour les transports (Ministère de l'Écologie, 2017). Via l'article L100-4 de la loi relative à la transition énergétique pour la croissance verte, le gouvernement français s'engage à « porter la part des énergies renouvelables [...] à 32 % de la consommation finale brute d'énergie en 2030 » et 40% de la production d'électricité.

De tels engagements en termes d'augmentation de la part des renouvelables dans le mix énergétique suscitent des interrogations économiques et technologiques liées à leur usage dans une infrastructure énergétique existante. Ainsi, l'exploitation de l'énergie solaire, au même titre que l'énergie éolienne, soulève notamment des défis liés à la nature variable des processus météorologiques mis en jeu. Ce caractère dit « intermittent » de ces sources d'énergie constitue une difficulté conséquente dans la gestion du réseau électrique à l'échelle d'un pays, où l'équilibre entre la production et la consommation d'électricité est primordial. Cette intermittence soulève aussi des difficultés dans la gestion indépendante de parcs photovoltaïques ou éoliens qui peuvent alors être couplés à des moyens de stockage ou de production complémentaires pour délivrer une puissance moins variable et incertaine. Dans cette thèse, l'intérêt se porte spécifiquement sur la production d'électricité issue directement de la conversion de l'énergie solaire par effet photovoltaïque.

## I.2 La nature variable de l'énergie photovoltaïque

L'effet photovoltaïque caractérise un phénomène physique de conversion d'énergie propre à certains matériaux semi-conducteurs qui génèrent un courant électrique par exposition à la lumière. Dans la suite de ce document, les termes rayonnement solaire et éclairement seront utilisés de manière équivalente et sont définis comme l'intégration de la luminance – puissance électromagnétique à travers un angle solide infinitésimal – sur l'hémisphère exposé. De cette manière, l'éclairement représente une puissance par unité de surface, exprimé en  $W/m^2$ . La production électrique d'une cellule photovoltaïque dépend des caractéristiques de conception de la cellule qui lui attribueront un certain rendement – rendement qui sera notamment susceptible de varier en fonction de la température de la cellule et de l'éclairement incident (Beyer et al., 2004). Cependant, la variable fondamentale qui régit les fluctuations de la production électrique est le rayonnement solaire incident correspondant à la sensibilité spectrale de la cellule. Les influences de la température de cette dernière et de la distribution spectrale du rayonnement solaire incident sur la production photovoltaïque ne seront pas à l'étude dans cette thèse : seul le rayonnement total (c.-à-d. intégré sur l'ensemble du spectre solaire) sera ainsi considéré.

Plus précisément, le rayonnement solaire total incident sur plan horizontal, noté GHI (*Global Horizontal Irradiance* – éclairement global sur plan horizontal) est une variable essentielle pour l'exploitation de l'énergie solaire par la conversion photovoltaïque. Cette technologie de conversion peut être utilisée sur une grande variété de supports, directement sur les toitures de bâtiments ou parkings existants, sur des supports adaptant leurs orientations pour maximiser la production, mais aussi moins couramment sur des façades, des routes, des voitures, des drones, etc. Ainsi, le fait d'analyser le rayonnement solaire global sur un plan horizontal et sa variabilité rend l'étude indépendante de l'orientation et de la sensibilité spectrale des cellules et assure une certaine transposabilité des méthodes développées à différents sites et cas d'études.

Au-delà de son intérêt pour l'énergie solaire, ce rayonnement solaire incident au sol a été reconnu par le Système mondial d'observation du climat (GCOS, *Global Climate Observing System*) comme l'une des variables climatiques essentielles pour la caractérisation de l'état du système climatique mondial et pour la surveillance du climat à long terme (*'Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC'*, 2010).

De manière schématique, le GHI se décompose en une partie directe (BHI – *Beam Horizontal Irradiance*) et une partie diffuse (DHI, *Diffuse horizontal irradiance*) comme illustré sur la figure 2.

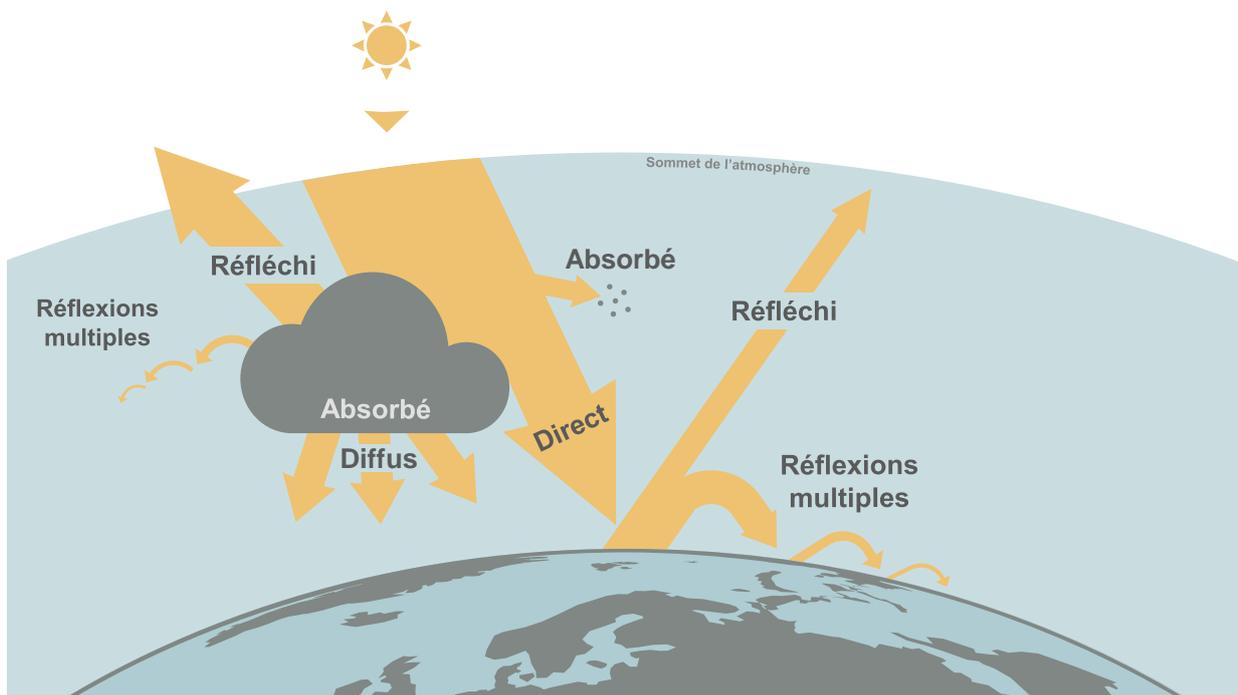


Figure 2 – Rayonnement solaire dans l'atmosphère.

La variabilité du rayonnement solaire au sol s'exerce à différentes échelles temporelles, et chaque échelle de variabilité peut être associée à un ou des phénomène(s) physique(s).

Les mouvements complexes des nuages et des constituants atmosphériques comme la vapeur d'eau ou les aérosols génèrent une forte variabilité intrajournalière du rayonnement solaire incident. Dans l'étude du rayonnement solaire, la décomposition en une partie attribuée aux effets des nuages et une partie associée aux variations des constituants de l'atmosphère non nuageuse est très largement utilisée. Cependant, Oumbe et al. (2014) soulignent le très faible couplage qui s'exerce entre ces deux composantes, justifiant l'hypothèse de leur indépendance et de séparabilité généralement utilisée en pratique. Ces échelles de variabilité vont nous intéresser dans cette thèse et nous chercherons à les prévoir. En particulier, les nuages seront au cœur de l'analyse puisqu'ils sont le principal facteur influant le rayonnement solaire perçu au sol (Dambreville, 2014), à l'exception des zones où la fréquence de jours sans nuages est très importante et pour lesquelles les aérosols et le contenu en vapeur d'eau jouent alors des rôles dominants.

La rotation terrestre autour de son axe des pôles géographiques engendre une variabilité journalière à l'origine de l'alternance jour/nuit et la révolution de la Terre autour du Soleil apporte une contribution saisonnière à ces variations du rayonnement au sol. Enfin, sur des échelles annuelles et pluriannuelles, les changements météorologiques et climatiques mènent une tendance ou des oscillations lentes. Les mouvements de rotation et de révolution de la Terre étant précisément connus (l'erreur sur le calcul rapide des angles solaires comme celui proposé par Blanc et Wald (2012) est de l'ordre du millième de degrés), il est possible d'établir des modèles théoriques du rayonnement solaire hors atmosphère très précis dans le passé aussi bien que dans le futur.

En réalité, il existe des couplages entre les différentes échelles temporelles de variabilité. Bengulescu (2017) propose d'utiliser la transformée adaptative de Hilbert-Huang comme outil d'analyse des séries temporelles du rayonnement solaire, dont la nature de ces dernières s'avère être non linéaire et non stationnaire. Ces travaux ont notamment démontré l'existence d'une modulation en amplitude de la variabilité stochastique à hautes fréquences du rayonnement par le cycle annuel qui, lui, est de nature déterministe.

Il est possible d'évaluer l'éclairement incident au sol en absence de nuages en tenant compte de la transparence ou trouble de l'atmosphère sensible à la composition de cette dernière, notamment en vapeur d'eau ou en aérosols. Ces modèles sont dénommés modèles d'éclairement par condition de « ciel clair ». Il existe une grande variété de modèles permettant un calcul de ces profils journaliers d'éclairement par ciel clair et de nombreuses comparaisons des modèles existants ont été proposées par la communauté scientifique (Badescu et al., 2013; Engerer et Mills, 2015; Gueymard, 2012; Ineichen, 2016, 2006; Inman et al., 2013). Dans notre étude, nous avons choisi d'utiliser le modèle ESRA (*European Solar Radiation Atlas*) pour sa robustesse et sa possibilité d'effectuer le calcul rétrospectivement, en temps réel ainsi qu'en prévision. En effet, ce modèle d'éclairement par ciel clair ESRA est un modèle de calcul empirique alimenté par une base de données climatologique de moyennes mensuelles d'indice de trouble de Linke (Remund et al., 2003). Cependant, il ne prend en compte ni la variabilité interannuelle ni la variabilité inframensuelle du trouble de Linke.

Le modèle McClear (Lefèvre et al., 2013) est un modèle physique exploitant les propriétés optiques des aérosols, le contenu en vapeur d'eau et en ozone dans l'atmosphère estimés dans le cadre des différents projets européens MACC (Kaiser et al., 2012; Peuch et al., 2009) et actuellement de CAMS (*Copernicus Atmospheric Monitoring Services*) par approximation à l'aide d'abaques précalculés du code de transfert radiatif libRadtran (Mayer et Kylling, 2005). Cependant, l'absence actuelle de disponibilité en temps réel ou en prévision de ce modèle constitue un point discriminant pour notre étude : pour des contraintes techniques du moment, le modèle délivre des estimations depuis 2004 jusqu'à maintenant avec deux jours de délai. Or, l'ensemble des données sur les propriétés des aérosols et des contenus de la colonne d'atmosphère en vapeur d'eau et d'ozone utilisé par McClear étant dans l'absolu disponible en analyse et en prévision, il n'y a pas d'impossibilité à voir émerger dans le futur une version de McClear accessible en prévision jusqu'à deux à trois jours en avance.

### I.3 Intérêt de la prévision photovoltaïque pour le réseau électrique

L'étude de la stabilité dynamique de réseaux de grande envergure nécessite la connaissance d'un grand nombre de paramètres influençant grandement leur comportement futur. Or l'énergie solaire photovoltaïque (PV) connaît une forte croissance en termes de puissance installée dans le monde (voir figure 3). Selon Du et al. (2017), cette technologie devient d'ores et déjà compétitive sur les marchés de l'énergie, ce qui laisse à penser que cette croissance ne va pas s'essouffler dans les prochaines années. En France, le principal gestionnaire du réseau de distribution est l'entreprise ENEDIS (ex. ERDF) et a pour mission de distribuer l'électricité aux consommateurs de l'échelle régionale à l'échelle locale, via les lignes de moyennes et basses tensions en assurant l'équilibre entre l'offre et la demande d'électricité. À ces échelles, la variabilité et l'incertitude de la production croissante d'énergie photovoltaïque sont un défi pour les gestionnaires du réseau en termes de qualité d'alimentation (courant, tension, fréquence) et du maintien des coûts des opérations d'ajustement en temps réel entre l'offre et la demande (Bletterie et Pfajfar, 2007; Ela et al., 2013).

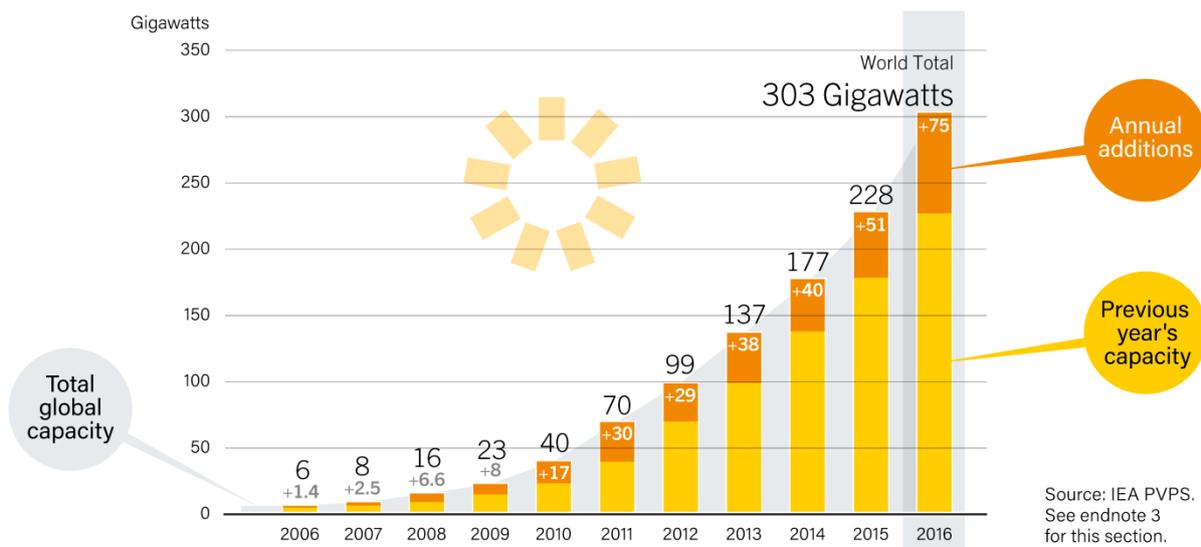


Figure 3 – Evolution de la puissance photovoltaïque installée dans le monde (en GW). Source : REN21, 2017.

De manière à atténuer les effets déstabilisants de la variabilité PV et ainsi assurer la fiabilité des réseaux électriques existants, plusieurs stratégies peuvent être adoptées. Ces dernières dépendent de l'infrastructure du réseau, des interconnexions disponibles qui rendent possible les transactions sur les marchés de l'énergie, de la possibilité d'effacement de certaines unités (Lew et al., 2013), du taux de pénétration des énergies intermittentes et de la capacité de stockage (stations de transfert d'énergie par pompage – STEP, batteries à hydrogène, au lithium, au plomb, au sel, à flux organique, volant d'inertie, etc.). Ela et al. (2013) proposent différentes stratégies de réduction des déséquilibres en évaluant leur impact sur le coût de production et l'équilibre consommation-production en Arizona. Lorsque peu d'énergies renouvelables variables sont installées sur un territoire,

les gestionnaires de réseau se limitent souvent aux prévisions sur la demande en électricité pour orchestrer et planifier les unités de production à moindres coûts grâce à des algorithmes d'optimisation. Cependant, l'augmentation de la part des énergies renouvelables intermittentes dans le mix devient particulièrement préoccupante, notamment sur certains réseaux insulaires (Diagne et al., 2012, 2013) où l'absence – ou l'insuffisance – d'interconnexion avec des réseaux voisins nécessite un pilotage intelligent des réserves et/ou des autres sources de production. Dans de telles conditions, il est non seulement nécessaire d'avoir des prévisions de leur production, mais il faut surtout qu'elles soient fiables à tout instant pour le gestionnaire réseau, c'est-à-dire de bonne qualité.

L'écart entre la production et la consommation d'électricité à chaque instant doit être maintenu à son niveau le plus bas à chaque instant, afin d'assurer un état stable du réseau et une « disponibilité optimale de l'énergie électrique » (Jeanjean et De Metz-noblat, 1997). Ainsi, tout changement brusque et de grande amplitude de la production électrique doit être anticipé le plus longtemps à l'avance et le plus précisément possible. Cependant, de manière générale, la qualité des prévisions décroît lorsque l'horizon de prévision (*cf.* paragraphe 1.4) augmente, avec une pente plus raide pour les méthodes « naïves » telle que la persistance présentée dans la suite de ce mémoire (Pelland et al., 2013). Ainsi, tous les horizons de prévision ont leur intérêt puisque le gestionnaire de réseau doit anticiper dès que possible et avec précision ces événements. Cependant, la moindre erreur sur l'instant d'occurrence prévu de ce que l'on appellera par la suite une « rampe » aura pour conséquence d'engendrer un écart instantané important entre production et consommation. Par ailleurs, Andersson et al. (2005) pointe notamment un risque de pannes en cascade engendrées par des occurrences concomitantes d'événements extrêmes antagonistes. Des exemples de telles pannes ont eu lieu notamment le 14 août 2003 dans plusieurs états des États-Unis et au Canada, le 23 septembre 2003 à l'Est du Danemark et au Sud de la Suède (privant 4 millions de consommateurs d'électricité) ou encore le 28 septembre 2003 en Italie.

Finalement, la prévision s'est imposée comme un outil très important pour la mise en œuvre des différentes stratégies d'équilibre, et sa qualité assure une meilleure rentabilité et la maîtrise des risques. L'IEA (2014) prédit que l'énergie PV continuera son expansion pour atteindre 16% de la production mondiale d'électricité en 2050, ce qui laisse penser que la prévision solaire continuera de gagner en importance dans la transformation des réseaux actuels vers des réseaux électriques dits plus « intelligents » (certains parlent de *Smart-Grid*), dont l'objectif est de gagner en efficacité grâce à un ajustement des flux électriques en production comme en consommation, le tout en temps réel.

La prévision de l'éclairement solaire a également d'autres applications potentielles que la gestion du réseau électrique : l'optimisation d'un parc photovoltaïque afin de délivrer une puissance moins incertaine, la prévision du rayonnement direct pour l'optimisation du pilotage de centrales à concentration solaire visant notamment à adapter les débits des fluides caloporteurs, à arbitrer entre la production électrique et le stockage thermique ou encore procéder à des défocalisations pour éviter l'échauffement des zones de concentra-

tion en cas d'éclairement dépassant sa capacité (objet du projet *DNICast*<sup>1</sup>). Bien entendu d'autres applications de cette prévision existent comme la prévision pour les systèmes de chauffe-eau solaires, pour l'agriculture (photosynthèse) ou encore pour des applications de santé (rayonnements ultra-violet). Cependant pour cette étude nous nous limiterons à l'application de systèmes photovoltaïques connectés à un réseau électrique.

## I.4 Horizon et résolution temporelle de prévision

Il est essentiel de différencier les termes horizon et résolution temporelle de prévision, bien qu'ils puissent se superposer pour certaines prévisions. L'horizon correspond à la période de temps sur laquelle porte la prévision et sa résolution temporelle représente en quelque sorte sa granularité.

Dans la section précédente, nous avons noté que, d'un point de vue conceptuel, tous les horizons de prévision peuvent avoir leur intérêt pour les gestionnaires du réseau électrique. Cependant, ces gestionnaires ont à leur disposition des moyens et méthodes limités pour agir de façon à rétablir ou assurer l'équilibre, avec des caractéristiques de délais de mise en œuvre qui leur sont propres. Le gestionnaire du réseau doit anticiper les situations de contraintes que peut engendrer la production PV (ou éolienne) si elle devient déséquilibrée par rapport à la consommation. Une telle anticipation permet de ne pas sur-solliciter les appareils chargés du rétablissement l'équilibre et ainsi d'optimiser la durée de vie de ces appareils. Voyant (2011) récapitule les différents horizons d'intérêt du point de vue du gestionnaire de réseau en fonction des sources de production correspondantes ou des possibilités de transactions sur les marchés de l'énergie pour le cas de la Corse. La figure 4 illustre cette correspondance mais définit également les tranches d'horizons très courts, courts et moyens termes que l'on reprendra dans ce manuscrit.

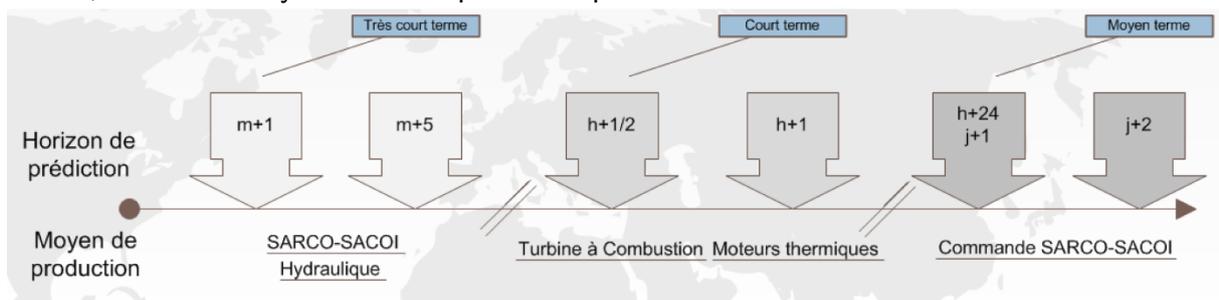


Figure 4 – Horizons de prévision et moyens de production associés. Les lettres *m*, *h*, et *j* indiquent l'horizon de prévision en termes respectivement de minutes, d'heures et de jours. [Source](#) : Voyant (2011).

Le tableau 1 résume les indications de Kostylev et Pavlovski (2011) en ce qui concerne les résolutions temporelles et horizons des prévisions les plus souvent demandées par les industriels, avec les événements d'intérêt associés. Chaque type d'opération est associé à une échelle temporelle allant le plus souvent de quelques secondes à quelques jours (Mills, 2010; Sayeef et al., 2012), et nécessite une prévision de la production PV de qualité avec la plus faible incertitude possible.

<sup>1</sup> <http://www.dnicast-project.net/>

| Horizon                   | Résolution temporelle | Utilité  |
|---------------------------|-----------------------|--|
| 15 min – 2 h              | 30 s – 5 min          | Événements extrêmes (rampes), variabilité impactant les opérations   |
| 1 h – 6 h                 | 15 min – 1 h          | Suivi de la courbe de charge   |
| 1 jour – 3 jours          | 1 h – 3 h             | Démarrage des unités de production, planification de la transmission et ventes sur les marchés du jour suivant                 |
| 1 semaine – 2 mois        | 1 jour                | Couverture, planification, optimisation des actifs   |
| 1 année – quelques années | 1 jour – 1 an         | Analyse de séries temporelles à long terme, évaluation des ressources, sélection de sites et dossiers de financement bancaires |

Tableau 1 – Événements d'intérêt pour les industriels avec leurs horizons et pas de prévision associés. [Source](#) : inspiré de Kostylev et Pavlovski (2011).

La variabilité qui intervient à chacune de ces échelles n'a pas la même amplitude. La particularité du rayonnement solaire est qu'il est potentiellement extrêmement variable à très court terme, et donc plus dur à prévoir avec précision, principalement à cause des couvertures nuageuses de tailles, de vitesses, de nature (gouttelettes d'eau, cristaux de glace, etc.) et d'épaisseurs optiques variables pouvant notamment occulter à des degrés divers la composante directe (et circumsolaire) du rayonnement. La résolution temporelle de la prévision va limiter de ce fait le type de variations prévisibles. Yang et al. (2017) ont récemment suggéré une approche permettant de « réconcilier », c'est-à-dire de fusionner, des prévisions dont l'horizon ou la résolution temporelle sont différents.

## I.5 Prévision locale et agrégée

De la même manière que la prévision s'effectue à différentes échelles temporelles, chaque échelle spatiale a son intérêt, avec bien entendu des couplages spatiotemporels entre ces dernières. Selon que l'utilisateur des prévisions est un producteur, un gestionnaire de réseau, un responsable d'équilibre, un agrégateur, etc., l'échelle spatiale de la prévision va différer. Un producteur ou un gestionnaire de réseau va en effet s'intéresser à la maille locale (prévision de chaque parc) alors qu'un agrégateur ou un responsable d'équilibre va s'intéresser à la maille agrégée (qui dépend alors de l'étendue du périmètre).

Les prévisions solaires à l'échelle locale ont des empreintes spatiales très variables allant d'un simple instrument pyranométrique *in situ* à des parcs photovoltaïques de plusieurs dizaines d'hectares. Ces types de prévisions, qui seront à l'étude dans ce manuscrit, s'incluent dans une stratégie partant de prévisions locales pour faire émerger le comportement global des fluctuations du solaire sur le réseau. Cette approche est difficile à mettre en œuvre en pratique, car elle nécessite beaucoup d'efforts de calcul et une connaissance de la production de toutes les fermes solaires dans la région d'intérêt – ce qui

n'est généralement pas le cas. Elle est cependant la plus à même de prévoir des variations brusques de la production globale ; selon Da Silva Fonseca et al. (2014), elle a montré de meilleures performances par rapport à une prévision régionale issue des modèles numériques météorologiques à grande échelle.

Une autre approche consiste donc s'intéresser à un comportement agrégé de la production, en profitant d'un effet dit de « foisonnement » (Antonanzas et al., 2016; Lorenz et al., 2011; Mills et Wiser, 2010), c'est-à-dire de lissage des variabilités temporelles par agrégation multisites, qui intervient grâce à une décroissance de la corrélation intersites en fonction de leur éloignement, comme le montre la figure 5. La difficulté de cette approche réside dans le fait que la production agrégée sur un large périmètre n'est souvent pas mesurée, la comparaison s'effectue donc par extrapolation d'un sous-ensemble de parcs photovoltaïques constituant un échantillonnage statistique que l'on juge représentatif.

Cependant, les contraintes du réseau ne sont pas uniquement globales, c'est pourquoi le gestionnaire de réseau français adopte une stratégie hybride, en considérant à la fois les prévisions locales et régionales. Une telle stratégie a également pour but de combiner différentes prévisions afin d'en tirer le meilleur résultat (Yang et al., 2017a). De plus, des approches telles que celle proposée par Saint-Drenan et al. (2018) permettent d'estimer la production d'énergie photovoltaïque agrégée à partir d'informations météorologiques locales et avec relativement peu d'informations sur les installations photovoltaïques elles-mêmes (notamment l'orientation de milliers de panneaux en Allemagne permet la construction d'une distribution statistique appliquée en Europe).

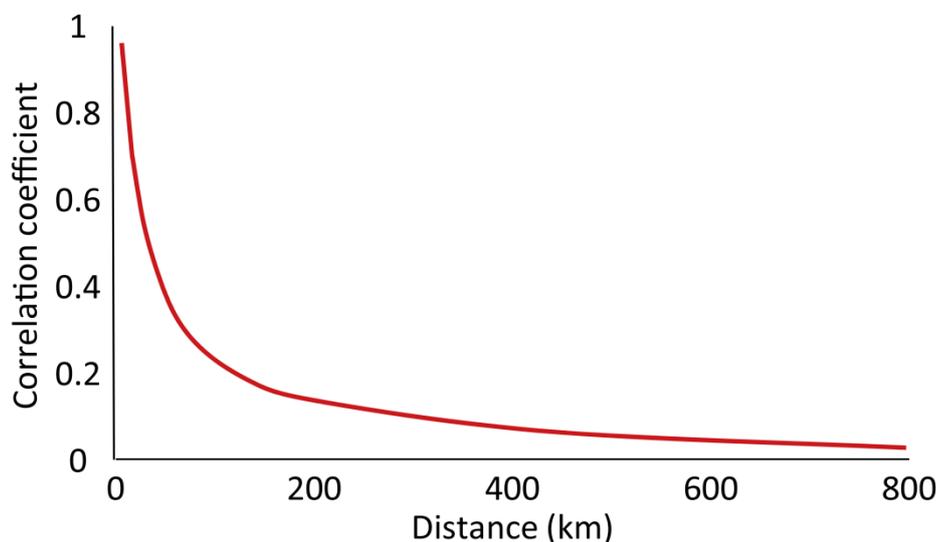


Figure 5 – Coefficient de corrélation des erreurs de prévision de deux systèmes en Allemagne en fonction de la distance qui les sépare. [Source](#) : Antonanzas et al. (2016), inspiré de Lorenz et al. (2008).

## I.6 Méthodes de prévision existantes en fonction des échelles spatiales et temporelles considérées

Plusieurs pistes prometteuses ont déjà été à l'étude, reposant sur des domaines de connaissances variés tels que la modélisation physique des processus météorologiques et du transfert radiatif dans l'atmosphère ou encore l'utilisation de processus autorégressifs, l'intelligence artificielle ou encore le traitement d'images et du signal, dans un but d'améliorer les capacités prédictives des modèles existants pour différents horizons de prévision. La résolution numérique des équations de la physique atmosphérique permet de simuler des systèmes météorologiques donnant lieu aux prévisions numériques météorologiques (NWP, *numerical weather prediction*). Ce type de prévision est très utilisé, voire incontournable, pour des horizons supérieurs à la journée. Cependant, pour de telles approches, des erreurs aléatoires et des biais importants persistent (Lorenz et al., 2009; Mathiesen et Kleissl, 2011; Perez et al., 2013, 2010; Sengupta et al., 2015). Pour les modèles globaux comme IFS (ECMWF) ou GFS (NCEP), les résolutions spatiales sont actuellement au mieux de l'ordre de la dizaine de kilomètres et les résolutions temporelles de l'ordre de 1 à 3 heures, ce qui est trop peu résolu pour des prévisions à court et à très court terme – c'est-à-dire de quelques minutes à quelques heures. Pour des horizons typiquement inférieurs à 5 heures, les images provenant de satellites météorologiques géostationnaires sont susceptibles d'améliorer les erreurs de prévision (Perez et al., 2010). En effet, la résolution spatiale de l'ordre du kilomètre et leur disponibilité toutes les dizaines ou quinzaines de minutes se prêtent complètement aux horizons allant de quelques minutes à quelques heures.

Pour des horizons temporels plus courts, il est également possible d'utiliser des caméras hémisphériques au sol permettant l'observation de l'état du ciel au-dessus d'un site donné (Fu et Cheng, 2013; Gauchet et al., 2012; Marquez et Coimbra, 2013), afin de pallier notamment le déficit de résolution subkilométrique des satellites météorologiques. Comme nous le verrons en détail par la suite, ces images ont une haute résolution spatiale et temporelle (de l'ordre de la seconde), mais admettent des limitations dues à leur vision limitée du ciel (un rayon de l'ordre de la dizaine de kilomètres, en fonction de l'altitude des nuages et de l'orographie – cf. Chapitre III), réduisant l'horizon de prévision à quelques minutes, en fonction de la vitesse de déplacement des nuages.

Enfin, les mesures pyranométriques délivrent une donnée locale – *in situ* – du rayonnement perçu au niveau du sol. Des études ont été menées pour tenter d'établir une prévision en considérant uniquement ces données passées du rayonnement incident (processus autorégressifs, apprentissage statistique, etc.), mais leurs performances souffrent d'un manque de description spatiale de la répartition des nuages (Chu et al., 2015). La figure 6 présente le domaine d'applicabilité typique de ces différentes sources d'information (avec les méthodes de prévision généralement associées) en fonction de leur résolution spatiale et temporelle.

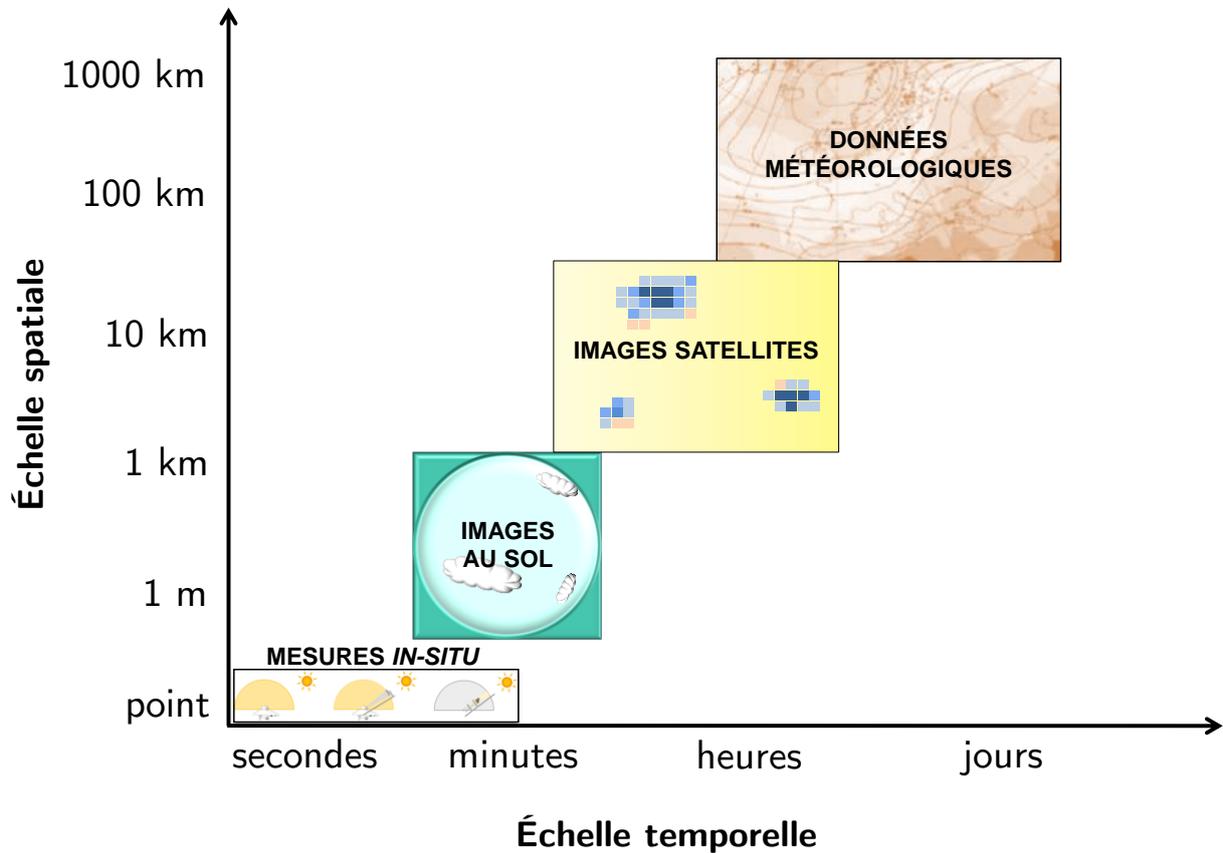


Figure 6 - Illustration de différentes sources d'information utilisées dans les méthodes de prévision existantes selon leurs échelles spatiales et temporelles. [Source](#) : inspiré de Sengupta et al. (2015).

## I.7 Objectifs et démarche de la thèse

Bien que ces différentes sources de données semblent bien proposer une partition des échelles spatiales et temporelles où elles sont les plus pertinentes, il est naturel d'envisager des synergies entre certaines d'entre elles, permettant d'améliorer la qualité des prévisions résultantes.

L'utilisation des mesures *in situ* remontées en temps réel pour la correction des erreurs locales des prévisions des modèles météorologiques (Diagne et al., 2013; Lynch et al., 2015), ou la rétroaction de la mesure pyranométrique dans les prévisions à base d'images satellites (Blanc et al., 2017b) ou d'images au sol (Bernecker et al., 2014; Cheng, 2016) peut être vue comme une synergie.

À part peut-être la thèse prospective de Dambreville (2014), peu d'études se sont intéressées à la complémentarité entre les caméras hémisphériques et les estimations par satellite, avec ou sans rétroaction des mesures *in situ*. Ces deux sources d'observation directe, toutes deux sous forme de séquences d'images de luminance du « même ciel », sont très différentes sur les plans de leurs résolutions spatiale et temporelle, de leurs couvertures spatiales, de leurs points de vue ou même de leur sensibilité spectrale.

Ces différences nous engagent à chercher à en exploiter les complémentarités avec ou sans des capteurs pyranométriques afin d'élaborer des schémas de prévision performants de l'éclairement solaire incident à hautes résolutions temporelle et spatiale et pour des horizons allant du très court terme au court terme.

Cette recherche est résumée par le titre et l'objectif de ma thèse : *synergie des mesures pyranométriques et des images hémisphériques in situ avec des images satellites météorologiques pour la prévision photovoltaïque.*

Les questions sous-jacentes à cet objectif global sont alors les suivantes :

*Existe-t-il effectivement une telle synergie ?* Pour tenter d'y répondre, nous allons identifier les forces et les faiblesses qu'offre chaque source d'information indépendamment, ainsi que les complémentarités susceptibles d'être exploitées lors d'un processus de fusion.

*Peut-on effectivement exploiter cette synergie et quelle méthode de fusion entreprendre ?* Les différences entre les deux sources d'images de luminance du ciel énoncées précédemment sont tellement marquées qu'elles constituent en même temps autant de verrous à la mise en place de méthodes de fusion.

*Quels bénéfices en tirer pour l'estimation et la prévision de la ressource solaire ?* Les avantages à tirer des méthodes proposées peuvent être observés qualitativement sur des exemples types, identifiés par des simulations ou quantifiés grâce à des comparaisons avec des mesures *in situ*.

## **I.8 Structure du manuscrit**

Ces questions sous-jacentes sont les composantes fondatrices qui structurent ce manuscrit. Le Chapitre II s'intéresse à l'évaluation des performances des prévisions, puisqu'il est essentiel d'avoir des outils permettant de quantifier et de juger leurs apports. Suite à l'établissement de ces prérequis en termes d'analyse des performances, nous proposons dans le Chapitre III une identification des complémentarités des différentes sources d'information afin d'en dégager les potentialités de synergies ainsi que les obstacles associés à leur développement. Enfin, une nouvelle approche proposée dans cette étude tente de profiter de la différence de point de vue des images satellites et des images hémisphériques afin d'en extraire une meilleure connaissance de la scène nuageuse. Dans le Chapitre IV, nous commencerons par évaluer l'apport d'une telle approche à l'aide d'un simulateur géométrique en trois dimensions, qui autorise toutes sortes d'analyses. Enfin, le Chapitre V sera consacré à l'utilisation et l'analyse de cette nouvelle approche sur des cas concrets avec des images et des mesures pyranométriques de situations réelles. Des conclusions seront dressées au Chapitre VI avec la description de perspectives sur la suite de cette thèse.



---

# Chapitre II - Analyse des performances des prévisions solaires

---

## II.1 Introduction à l'analyse des performances dans le domaine de la prévision solaire

Le présent chapitre se focalise sur un aspect essentiel du domaine de la prévision solaire parfois négligé : qu'est-ce qui fait la qualité d'une prévision ? Quels outils et procédures appliquer afin d'extraire des informations quantitatives relatives aux performances des prévisions adaptées à leur application pratique ? Comme l'expliquent Kostylev et Pavlovski (2011), les tests de performance des prévisions sont souvent entrepris en s'appuyant sur des pratiques analytiques établies sans compréhension claire de la pertinence des métriques pour une application donnée. Comme nous l'avons vu dans l'introduction, les objectifs des gestionnaires sont multiples et requièrent des prévisions à différents horizons. Les inadéquations entre la prévision et la production effectivement constatée peuvent avoir des conséquences plus ou moins lourdes pour le réseau électrique ou sur la gestion des installations photovoltaïques. Du et al. (2017) ont identifié quatre étapes clés qui permettront d'assurer un coût minimal à l'utilisateur final : (1) des données de mesure représentatives et de haute qualité, (2) des techniques et modèles de prévisions de qualité, (3) une évaluation significative de la performance prévisionnelle et (4) une communication efficace et efficiente de l'information de prévision à l'utilisateur. De plus, Zhang et al. (2013) pointent du doigt la difficulté de l'évaluation des performances qui découle du constat que chaque chercheur ou expert apporte des résultats selon sa propre métrique. En l'absence de l'implémentation de la méthode, des données ou de la manière dont sont calculés les différents scores, les résultats des méthodes de prévision sont difficiles à exploiter et à comparer. Ainsi, face à ce constat, plusieurs tentatives de standardisation de l'analyse des performances ont été suggérées (Beyer et al. 2009; Kostylev et Pavlovski 2011). Ces propositions d'analyses standardisées des performances peuvent cependant être complétées par les plus récentes avancées sur le sujet dont certaines seront mises en avant dans ce chapitre.

L'analyse des performances de prévisions ponctuelles du rayonnement global (intégré sur tout le spectre) sans considération sur celles relatives aux prévisions probabilistes, de distributions angulaires ou spectrales. Elle s'effectue usuellement par la comparaison entre des prévisions faites dans le passé avec des mesures effectives. Différentes quantités statistiques sont ensuite calculées, à partir desquelles sont déduites lesdites performances. Quelle que soit l'application, l'objectif de la prévision et ses propriétés requises sont im-

portants à définir. Les propriétés recherchées peuvent être diverses, comme la corrélation, l'alignement temporel, la distribution statistique, le suivi des rampes, etc., mais peuvent également concerner le type de ciel à mieux respecter – par exemple, les ciels sans nuages pour les systèmes à concentration –, ou encore concerner la validité de la prévision sur une zone géographique donnée.

L'approche la plus couramment répandue dans la communauté scientifique est l'utilisation de métriques statistiques – définies au paragraphe II.3.1 – quantifiant l'erreur globale qu'effectue la prévision par rapport à une mesure considérée comme référence. Le paragraphe II.2 de ce chapitre traite des mesures du rayonnement solaire utilisées pour la comparaison, ainsi que des méthodes d'analyse de leur qualité justifiant leur utilisation en tant que référence, c'est-à-dire de « vérité terrain ». Cependant, comme nous l'illustrerons dans le paragraphe II.3.3, l'utilisation de ces métriques statistiques, bien que donnant une indication globale des performances en termes d'erreur moyenne commise par la prévision, admet certaines limitations et ne suffit pas à transcrire complètement le comportement des méthodes de prévision. De nombreux chercheurs se sont penchés sur cette problématique au cours des dernières années, ce qui a conduit à la proposition de nombreuses métriques pour la prévision solaire. Une liste non exhaustive des métriques existantes et de leur utilité est proposée au paragraphe II.4.

Par ailleurs, les échanges que nous avons pu avoir avec des utilisateurs de prévisions solaires ont fait émerger deux concepts signifiants pour leur réalité dans la gestion du réseau électrique. Le concept du suivi de rampes évoqué dans le chapitre d'introduction est en effet un aspect à prendre en compte, mais la notion d'alignement temporel a également son importance. Si une variation brusque est prévue, mais avec une avance ou un retard, même limités, l'erreur instantanée commise par la prévision peut être conséquente. Deux nouvelles métriques sont ainsi proposées au paragraphe II.5 ce chapitre afin de quantifier ces deux notions.

Outre le choix des indicateurs de performance et les problématiques associées à leur analyse, le choix du jeu de données sur lequel appliquer ces indicateurs est lui aussi primordial. Ainsi, appliquer de telles métriques sur des classes spécifiques – les jours dont la couverture nuageuse est variable par exemple – peut apporter un complément à l'analyse principale et peut permettre de se détacher des spécificités du jeu de données utilisé. Par exemple, calculer les performances d'une prévision sur un ensemble de journées majoritairement sans nuages donnera évidemment de meilleurs scores qu'une analyse similaire sur des jours plus variables, du fait de la bonne prévisibilité du rayonnement par ciel clair. Différentes stratégies de classification des données seront exposées et discutées au paragraphe II.6 afin d'obtenir une analyse catégorielle des performances. Enfin, les résultats des nouvelles métriques proposées et de métriques plus conventionnelles appliquées à des méthodes de prévision solaire de référence, et ce, sur différents jeux et classes de données, seront présentés au paragraphe II.7.

## II.2 La mesure pyranométrique *in situ*, une mesure de référence pour l'analyse des performances

### II.2.1 Moyens de mesures du rayonnement solaire existants

La mesure de l'éclairement est essentielle pour l'analyse de la ressource solaire, la construction d'un projet photovoltaïque, la gestion de l'exploitation ou pour développer et tester des modèles d'estimation ou de prévision de cette ressource. Tout instrument de mesure a une précision limitée, et donc, une incertitude intrinsèquement associée. L'essentiel est alors de déterminer et de quantifier l'importance qu'a cette incertitude au regard des quantités mises en jeu et des échelles considérées.

Dans le cadre de la prévision solaire, deux catégories de capteurs sont utilisées : les capteurs de rayonnement solaire et la mesure de production de cellule(s) photovoltaïque(s) de référence. Seuls les capteurs d'éclairement et leurs incertitudes associées seront énumérés ici pour leur indépendance aux caractéristiques techniques des panneaux solaires, ainsi que par soucis de cohérence avec le reste du document qui traite principalement du rayonnement solaire total. Comme introduit dans le premier chapitre, le rayonnement global GHI s'exprime de la manière suivante :

$$G = G_b + G_d \quad (1)$$

où  $G$  est l'éclairement global sur plan horizontal (GHI, exprimé en  $W/m^2$ ),  $G_b$  est l'éclairement direct sur le plan horizontal (BHI, exprimé en  $W/m^2$ ),  $G_d$  est l'éclairement diffus intégré sur le plan horizontal (DHI, exprimé en  $W/m^2$ ). Par ailleurs, certaines études utilisent la variable de rayonnement global sur plan incliné (notée POA ou GTI, *Plane-Of-Array irradiance* ou *Global Tilted Irradiance*), qui est une variable essentielle pour effectuer la conversion de l'éclairement en production électrique photovoltaïque.

En fonction de son inclinaison, le pyranomètre est l'instrument de mesure du GHI – s'il se trouve sur un plan horizontal – ou du GTI – lorsqu'il est adossé au plan des panneaux solaires. Le pyréliomètre est celui qui permet l'acquisition du rayonnement direct en incidence normale (DNI, *Direct Normal Irradiance*). Enfin, le pyranomètre couplé à un disque ou une bande d'ombrage mesure le DHI. Le tableau 2 reprend les différents instruments de mesure du rayonnement solaire cités précédemment avec leur support d'intégration angulaire de la luminance du ciel, leur installation la plus courante ainsi qu'un exemple d'illustration du matériel.

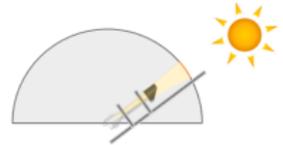
| Type de radiomètre  | Composante mesurée | Champ de vision | Installation   | Illustration   |
|---|--------------------|-----------------|--|--|
| Pyranomètre   | GHI                | $2\pi$ sr       | Monté sur une surface horizontale stable et libre d'obstructions locales   |   |
| Pyrhéliomètre   | DNI                | 5 à 6°          | Monté sur traqueur solaire automatique pour l'alignement avec le disque solaire  |   |
| Pyranomètre couplé à un disque ou une bande obstruant le Soleil | DHI                | $2\pi$ sr       | Monté sur un traqueur solaire automatique équipé d'un mécanisme d'ombrage ou sur une plate-forme à bande ajustée manuellement pour bloquer le DNI de la surface du détecteur |   |
| Pyranomètre   | POA ou GTI         | $2\pi$ sr       | Monté sur une surface de même inclinaison que les panneaux solaires  |  |

Tableau 2 – Instruments de mesure associés aux différentes composantes du rayonnement solaire. Source : Sengupta et al. (2015).

Les pyranomètres et pyrhéliomètres utilisent le plus souvent un détecteur thermoélectrique (thermopile) ou photoélectrique pour convertir la puissance solaire ( $W/m^2$ ) en un signal électrique proportionnel. Ces deux types de détecteurs ont cependant des caractéristiques différentes : les thermopiles ont une sensibilité spectrale couvrant la totalité du spectre solaire de manière quasi uniforme et ont une réponse temporelle de quelques secondes, alors que les photodiodes sont spectralement plus sélectives et non-uniformes mais ont une réponse temporelle plus courte (quelques millisecondes). Les incertitudes de la mesure, avec des instruments propres correctement étalonnés et sous des conditions d'utilisation nominales, sont de l'ordre de  $\pm 5\%$  pour le GHI,  $\pm 3\%$  pour le DNI et  $\pm 7\%$  pour le DHI (Kleissl, 2013). Typiquement, les performances de ces instruments décroissent de 0,5% à 1,0% par an et nécessitent une maintenance et un étalonnage réguliers.

Ces instruments de mesure sont notamment utilisés par les réseaux de capteurs qui délivrent des données de rayonnement solaire. Parmi ces réseaux, citons les stations du réseau BSRN<sup>2</sup> (*Baseline Surface Radiation Network*) initié par Ohmura et al. (1998) et qui sont notamment utilisées à des fins de validation des méthodes Heliosat (Qu et al., 2016; Vernay et al., 2013).

<sup>2</sup> Une liste des stations est disponible à l'adresse suivante :

<https://www.pangaea.de/ddi?request=bsrn/BSRNEvent&format=html&title=BSRN+Stations>

## II.2.2 Instrumentation déployée dans le cadre de l'étude

EDF a installé une instrumentation sur plusieurs centrales photovoltaïques dans le département de la Vendée. Les deux sites qui sont à l'étude dans cette thèse, appelés par la suite sites n°1 et n°2, sont séparés d'une distance d'environ 65 km, approximativement à la même latitude. Le site n°1 se situe sur la côte alors que le site n°2 est plus reculé dans les terres. L'étude se limite à la période du 01/08/2014 au 31/12/2015 (*i.e.* 517 jours) pour laquelle toutes les données citées ci-dessous sont disponibles.

L'ensemble des appareils de mesure installés sur les deux sites sont représentés sur la figure 7. Dans cette étude, des pyranomètres SPN1 (Badosa et al., 2014; Delta-Devices, 2011; Vuilleumier et al., 2017) et des caméras hémisphériques (ASC pour *All-Sky Camera*) sont essentiellement utilisés. La station météorologique permettant la mesure notamment de la pression atmosphérique, de la température à 2 m et de l'humidité relative peut être utile pour mieux comprendre les conditions d'obtention des mesures. Les pyranomètres SPN1, dans cette chaîne d'acquisition, fournissent des mesures du rayonnement global et diffus sur plan horizontal toutes les minutes, qui sont le fruit d'une intégration des mesures instantanées relevées lors de la dernière minute.



Figure 7 – Photos des appareils de mesure qui équipent le site n°1 (a) et le site n°2 (b). ASC : caméra hémisphérique ; pyranomètre SPN1 : instrument de mesure du rayonnement global et diffus.

*N.B. : Un Pyrgéomètre est un radiomètre hémisphérique mesurant le flux radiatif infrarouge incident à la surface de la Terre dans toutes les directions. Ce type d'instrumentation ne sera pas utilisé dans le cadre de cette étude.*

Les caméras ASC intègrent un objectif de type hémisphérique avec une focale courte de 1.4 mm : elles permettent de faire des acquisitions du ciel avec un angle de vue de 185°. Les images, de dimensions 640 x 480 pixels, sont disponibles suivant un pas de 10 s. Les paramètres d'acquisition comme le temps d'exposition, l'ouverture, le gain, etc. varient automatiquement en fonction des conditions d'illumination mais ne sont hélas pas disponibles.

### II.2.3 Analyses de qualité des mesures pyranométriques

Bien qu'une incertitude soit indiquée dans les documentations techniques des instruments installés, il convient d'analyser les mesures pour en détecter d'éventuelles « anomalies », c'est-à-dire des mesures physiquement impossibles ou une récurrence d'évènements rares qui suggèreraient un problème lors de l'acquisition ou de l'interprétation des mesures. Il est important de détecter et éliminer – ou corriger – ces mesures physiquement impossibles ou hautement improbables car leur utilisation peut potentiellement conduire à des résultats et des conclusions erronés. Cependant, des tests excessivement restrictifs risquent de détériorer la représentativité des données restantes. Les points de mesure exclus au milieu d'une journée, ou initialement manquants dans les données, peuvent être abordés de différentes façons. En effet, pour certains algorithmes, les données manquantes dans une série temporelle peuvent causer différents problèmes. Une solution consiste à essayer de compléter ces données manquantes par des données interpolées ou, si une partie importante de la journée est absente, exclure ce jour de l'étude. L'impact de ces données manquantes sur le calcul des moyennes d'éclairement mensuel a été traité par Peruchena et Amores (2017). Dans cette étude, nous excluons les jours pour lesquels il manque plus de 10% des données de jour, ou qui ont été considérés comme incorrects après avoir appliqué l'analyse de qualité expliquée ci-dessous. Ce seuil n'exclut aucune journée pour le site n°1 mais exclut 31 jours sur le site n°2, dont 17 sont des journées totalement absentes des mesures de rayonnement sur ce site.

Trois tests conçus pour les mesures de radiation à la minute du réseau BSRN (Baseline Surface Radiation Network) sont appliqués sur nos mesures SPN1, comme l'ont suggéré Roesch et al. (2010). Par conséquent, les mesures GHI et DHI sont rejetées si elles ne passent pas les tests du tableau 3.

| Nom du test                    | Paramètre             | Condition 1                                  | Condition 2   |
|--------------------------------|-----------------------|--|---|
| Physiquement possible          | GHI                   | $G > -4 \text{ W.m}^{-2}$                    | $G < 1.5 S_0 \mu^{1.2} + 100 \text{ W.m}^{-2}$          |
| Événement rare                 | GHI                   | $G > -2 \text{ W.m}^{-2}$                    | $G < 1.2 S_0 \mu^{1.2} + 50 \text{ W.m}^{-2}$           |
| Condition sur le ratio DHI/GHI | $k_d = \frac{G_d}{G}$ | $k_d < 1.05$<br>Avec : $\theta_s < 75^\circ$ | $k_d < 1.10$<br>Avec : $75^\circ < \theta_s < 93^\circ$ |

Tableau 3 – Tests proposés par Roesch et al. (2010).  $\theta_s$  est l'angle zénithal solaire,  $S_0$  est la constante solaire ajustée par la distance Terre-Soleil et  $\mu$  est le cosinus de l'angle zénithal solaire.

Seules les données correspondant à des instants dont l'élévation solaire dépasse  $7^\circ$  sont analysées, ce qui représente 57.4% de l'ensemble des données. Les résultats de cette analyse de qualité appliquée aux données du site n°1 sont présentés dans la figure 8 dans une représentation  $(k_d, k_t)$ , où  $k_d = \frac{G_b}{G}$  est la fraction diffuse et  $k_t = \frac{G}{G_0}$  est l'indice de clarté (où  $G_0$  est le rayonnement au sommet de l'atmosphère TOA, déterminé par Blanc et Wald (2012)). Les différents tests excluent moins de 0.02% de l'ensemble des données, ce qui ne perturbe pas la représentativité des mesures. Des résultats similaires sont obtenus

nus pour le site n°2. D'autres analyses de qualité ont été proposées dans la littérature, notamment celles de Younes et al. (2005) utilisées par Ruiz-Arias et al. (2010) mais dont les tests – excepté pour le seuil sur l'élévation solaire – sont souvent trop restrictifs. Pour de faibles élévations solaires, les points du graphe ( $k_d, k_t$ ) sont sensibles à un léger biais du GHI ou du DHI. Or l'incertitude de la mesure étant de  $\pm 8\% \pm 10 \text{ W/m}^2$  (Badosa et al., 2014), nous avons privilégié les tests de Roesch et al. (2010) qui ont d'ailleurs été adoptés par Long et Dutton (2010).

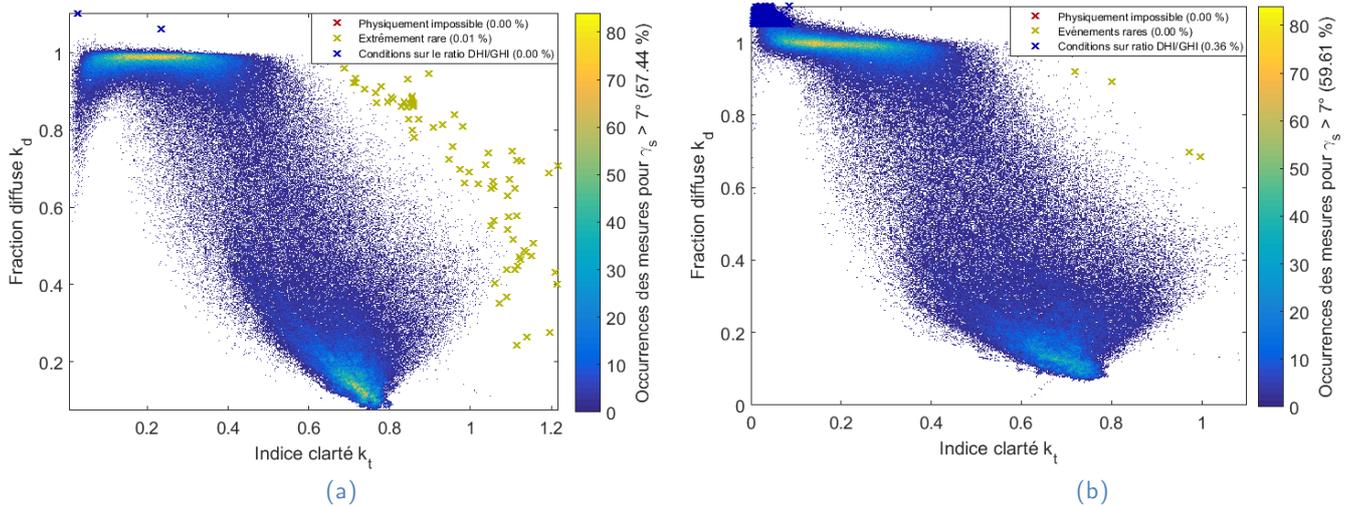


Figure 8 – Résultats de l'analyse de qualité sur le site n°1 (a) et n°2 (b) en appliquant les tests suggérés par Roesch et al. (2010).

Les analyses de qualité mentionnées précédemment sont des outils pertinents permettant d'identifier de potentielles anomalies sur les mesures, mais il est préférable de ne pas s'en contenter et d'y ajouter une analyse visuelle des mesures.

## II.2.4 Visualisations des données favorisant l'identification d'anomalies sur la mesure *in situ*

Plusieurs représentations des données facilitent la détection d'anomalies. Une représentation usuelle qui aide à la visualisation des données manquantes et de certaines incohérences sur les données est la représentation en image indiquant les instants de la journée en fonction de la date de chaque donnée, telle qu'illustrée sur la figure 9. À l'aide de ce graphe, il est ainsi possible de vérifier que l'horodatage des données corresponde bien au temps universel UT, qu'il n'y a pas de saut lors d'un changement d'heure en été ou en hiver ou encore que l'unité est effectivement en  $\text{W/m}^2$ . La figure 9.b illustre les trous qui apparaissent dans les données d'origine en blanc, représentant 5.02% des données pour le site n°2 contre 0.05% pour le site n°1 (figure 9.a). En dehors des données manquantes, aucune anomalie significative n'est détectée sur ces données.

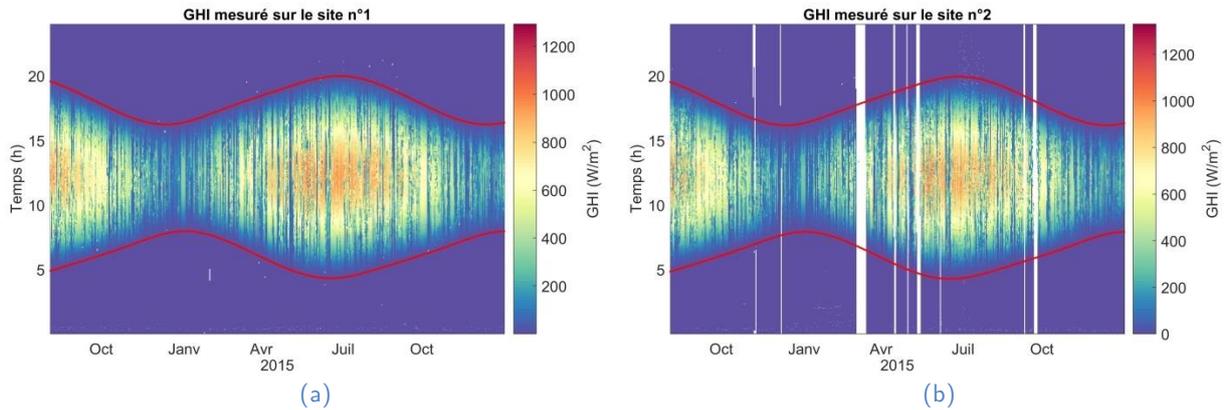


Figure 9 – Représentation des données du site n°1 (a) et du site n°2 (b) selon l'heure en ordonnée et la journée en abscisse. Le tracé rouge représente la limite d'angle d'élévation solaire nul ( $\gamma_s = 0$ ).

La figure 10 est une représentation des mesures de l'indice de clarté  $k_t$  – défini au paragraphe précédent – selon les azimuts et élévations solaires correspondantes. La ligne rouge représente l'horizon, calculé à l'aide du modèle numérique de terrain à 100 m de résolution et 10 m de précision recueilli par interférométrie radar lors de la mission Shuttle Radar Topography Mission (SRTM)<sup>3</sup> en 2000 (Jarvis et al., 2008). Une différence entre les deux sites qui est révélée par cette représentation est la présence de très faibles valeurs de l'indice  $k_t$  pour les élévations entre 0° de 10° sur le site n°2 (figure 10.b). Dans un environnement parfait, les mesures indiqueraient ainsi que pour les très faibles élévations solaires il serait très fréquent d'observer un ciel couvert au-dessus de ce site. Cependant, si l'on se rapporte à la figure 7.b, la caméra ASC horizontale et le pyrgéomètre CGR4 sont disposés respectivement aux azimuts 273° et 93° par rapport au pyranomètre SPN1 horizontal. Ces instruments étant approximativement à la même altitude et relativement proches les uns des autres, nous attribuerons ce comportement des élévations inférieures à 10° aux phénomènes d'ombrage des instruments voisins. Une attention particulière devra être portée lors de l'utilisation de ces données de faible élévation solaire sur le site n°2.

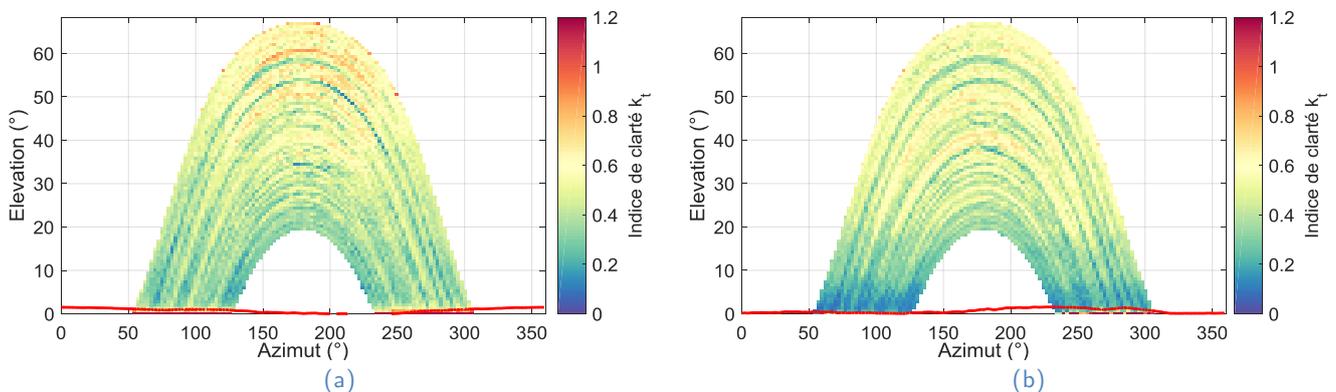


Figure 10 – Représentation des données du site n°1 (a) et du site n°2 (b) selon l'azimut solaire (en abscisse) et l'élévation solaire (en ordonnée) correspondant à l'instant d'acquisition de ces premières.

<sup>3</sup> Page de la NASA dédiée à ce format : <https://www2.jpl.nasa.gov/srtm/>

## II.3 Limitation des méthodes standards d'évaluation des performances des prévisions

L'examen des avantages relatifs des différentes méthodes de prévision n'est pas simple. Les résultats dépendent notamment de la situation géographique, de la période de l'année et du climat local. Aussi, différentes métriques sont utilisées par des auteurs pour caractériser les performances de leurs prévisions. Ce paragraphe recense les métriques statistiques les plus utilisées par la communauté scientifique (II.3.1), les pratiques recommandées pour appliquer ces métriques (II.3.2) ainsi que des exemples concrets illustrant les limitations de la métrique RMSE (*Root Mean Square Error*, II.3.3). Enfin, la partie II.3.4 résumera certaines attentes utilisateurs que nous tenterons de satisfaire dans cette étude.

De manière analogue à une mesure physique, une métrique vise à communiquer sur une caractéristique, un attribut ou un comportement de la méthode considérée, afin d'en faciliter la comparaison quantifiée avec d'autres méthodes existantes. Comme définie dans le paragraphe II.7.2.a, la méthode de prévision de référence qui exploite la forte composante déterministe de la variabilité du rayonnement solaire induite par la trajectoire quotidienne du Soleil par condition de ciel clair est appelée méthode par persistance. Les scores des métriques présentées dans ce paragraphe sont alors fréquemment indiqués à la fois pour la méthode à évaluer et pour cette méthode par persistance. À noter que cette méthode par persistance est pertinente pour des comparaisons de prévisions à courts et très courts termes, mais la comparaison perd de son sens pour des horizons de prévision plus importants. En effet, l'autocorrélation temporelle de la mesure du rayonnement solaire décroît naturellement avec le temps, rendant l'hypothèse de persistance du rayonnement à quelques heures peu valable. Inversement, pour des horizons très courts, l'autocorrélation est suffisamment importante pour que la persistance à très court terme génère des résultats statistiques très bons, ce qui en fait une méthode difficile à « battre » au sens des métriques couramment utilisées.

Enfin, de manière comparable à la visualisation des données du paragraphe II.2.4 ayant pour but d'attester de la qualité de la mesure *in situ*, l'analyse quantitative de la qualité des prévisions par le biais de scores doit être complétée par une analyse visuelle des séries temporelles sur des situations typiques, comme recommandé notamment par Beyer et al. (2009).

### II.3.1 Métriques usuelles

La catégorie de métriques dont les concepteurs de prévisions ont le plus recours est celle des métriques statistiques. La métrique la plus largement utilisée est la racine carrée de l'erreur quadratique moyenne (RMSE, *Root Mean Square Error*) et se calcule par la formule suivante :

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{G}_i - G_i)^2} \quad (2)$$

où  $\hat{G}$  représente le rayonnement solaire prévu et  $G$  est la mesure de référence. Le score délivré par cette métrique pénalise les grands écarts à la mesure.

Tout aussi communément utilisée, la métrique d'erreur absolue moyenne (MAE, *Mean Absolute Error*) est adaptée à l'évaluation uniforme des erreurs de prévision, et s'écrit :

$$\text{MAE} = \frac{1}{N} \sum_{i=1}^N |\hat{G}_i - G_i| \quad (3)$$

Les mesures statistiques traditionnelles pour caractériser la qualité d'un modèle comprennent aussi le coefficient de détermination qui compare la variance des erreurs à la variance des données modélisées :

$$R^2 = 1 - \frac{\sum_{i=1}^N (G_i - G_i)^2}{\sum_{i=1}^N (\bar{G} - G_i)^2} \quad (4)$$

où  $\bar{G}$  est la moyenne des mesures.

Fréquemment utilisé en appui au coefficient de détermination, le coefficient de corrélation de Pearson est défini par :

$$\rho = \frac{\text{Cov}(G, \hat{G})}{\sqrt{\text{Var}(\hat{G})\text{Var}(G)}} \quad (5)$$

où Cov est la fonction de covariance.

Afin de caractériser les erreurs systématiques d'une prévision, c'est-à-dire sa justesse, la norme ISO 21748 (Organisation internationale de normalisation) recommande le recours au biais (aussi noté MBE, *Mean Bias Error*), décrit par la formule suivante :

$$\text{Biais} = \frac{1}{N} \sum_{i=1}^N (\hat{G}_i - G_i) \quad (6)$$

Cette part systématique de l'erreur quantifiée par le biais est incluse dans la RMSE d'après la relation suivante :

$$\text{RMSE}^2 = \text{Biais}^2 + \sigma^2 \quad (7)$$

où  $\sigma$  est l'écart type, qui mesure la dispersion de la prévision, c'est-à-dire son infidélité selon les termes des normes ISO. Ainsi, l'utilisation conjointe du biais avec la RMSE apporte une information redondante. À l'inverse, le biais et l'écart type forment un couple de métriques complémentaires.

Enfin, de manière analogue au biais, certaines métriques recherchent plus la représentativité statistique que la concomitance des événements. Dans cette approche, nommons par exemple le KSI (*Kolmogorov–Smirnov test Integral*) (Espinar et al., 2009). Le KSI est défini comme la moyenne des maxima des valeurs absolues des différences entre les fonctions de répartition des deux ensembles de données sur 100 intervalles réguliers entre leurs minimum et maximum. Il s'agit donc d'une métrique de similarité statistique, qui admet cependant une dépendance vis-à-vis du temps d'intégration du rayonnement et du nombre d'intervalles imposé.

## **II.3.2 Pratiques pertinentes pour l'application des métriques statistiques**

Dans l'optique d'améliorer la comparabilité des résultats de validation des performances des prévisions, il est important d'adopter certaines pratiques qui sont listées – de manière non exhaustive – dans la suite de cette section.

### **II.3.2.a Scores relatifs et données diurnes**

Dans de nombreux domaines, il est d'usage d'utiliser les métriques énumérées précédemment avec leur score relatif associé – une méthode qui est souvent suffisante et satisfaisante. Cependant, dans le domaine de l'énergie solaire, cette méthode apporte une certaine confusion en partie à cause de l'évolution diurne de la courbe d'éclairement solaire incidente au niveau du sol. En effet, les mesures nocturnes ne présentent pas d'intérêt particulier. De plus, l'hétérogénéité en latitude des durées de jour va induire un certain biais artificiel dans l'interprétation des métriques, puisqu'un calcul de scores incluant les données nocturnes va favoriser les lieux qui connaissent le plus d'heures de nuit sur une période non entière annuellement. Ces instants de mesures sont donc souvent écartés de l'analyse en imposant un seuil sur l'élévation solaire. De manière générale, qu'il ait été choisi d'inclure ou de séparer les données jour/nuit, ou de ne considérer que les valeurs de GHI supérieures à un certain seuil (Antonanzas et al., 2016), il est impératif d'indiquer clairement la période sur laquelle a été évaluée la métrique.

Comme le soulignent Hoff et al. (2013), le calcul des erreurs relatives est souvent plébiscité dans l'industrie, mais celles-ci ne font pas encore consensus. Une erreur relative est constituée d'un score absolu d'une métrique d'erreur (i.e. RMSE ou MAE) qui est divisé par un terme de normalisation. Ce terme peut être la moyenne du jeu de données considéré ou encore la puissance pic. Cette dernière est utilisée lors de l'analyse de production électrique, la puissance pic peut être soit la puissance de la centrale considérée soit un terme arbitraire de normalisation de  $1000 \text{ W/m}^2$ . Le terme normalisateur peut varier suivant les études, il est donc essentiel de préciser la normalisation effectuée.

Dans la suite de ce document, lorsque des scores seront affichés, ils seront calculés sur les instants dont l'élévation solaire dépasse les  $7^\circ$ , seuil suggéré par Ruiz-Arias et al. (2010). Dans le cas de scores relatifs, le terme de normalisation sera la moyenne des valeurs sur cette période.

### II.3.2.b Comparaison à la persistance (Variabilité & incertitude)

L'une des difficultés majeures dans la caractérisation des performances est la dépendance aux données sur lesquelles sont testées les méthodes de prévision. Cette dépendance est l'un des facteurs qui rend difficile la comparaison des scores inter-études. L'une des stratégies qui ont été proposées consiste à comparer systématiquement les scores des métriques statistiques d'une méthode de prévision à la méthode de référence : la persistance. Le lecteur est invité à se rendre au paragraphe II.7.2.a pour une description détaillée de cette méthode de référence.

Parfois appelé *skill score*, parfois nommé *forecast skill*, la métrique  $s$  proposée par Marquez et Coimbra (2012) utilise le rapport entre l'incertitude ( $U$ ) et la variabilité ( $V$ ) de la mesure pour en extraire la métrique suivante :

$$s = 1 - \frac{U}{V} \quad (8)$$

avec :

$$U = \sqrt{\frac{1}{N} \sum_{i=1}^N \left( \frac{\hat{G}_i - G_i}{G_i^{\text{CLS}}} \right)^2} \quad (9)$$

$$V = \sqrt{\frac{1}{N} \sum_{i=1}^N \left( \frac{G_i}{G_i^{\text{CLS}}} - \frac{G_{i-1}}{G_{i-1}^{\text{CLS}}} \right)^2} \quad (10)$$

où  $G^{\text{CLS}}$  est la courbe d'éclairement par ciel clair.

Marquez et Coimbra suggèrent d'en moyenniser la valeur grâce à un calcul sur une fenêtre glissante. Cette métrique répond alors à l'exigence d'une indépendance des métriques proposées vis-à-vis des conditions climatiques des sites d'études. Ainsi, si la fenêtre contient un grand nombre de journées de ciel clair, l'incertitude  $U$  et la variabilité  $V$  de la période seront faibles, le rapport des deux grandeurs informe sur l'erreur de prévision au regard de la variabilité observée. Marquez et Coimbra indiquent que cette métrique  $s$  peut être approché par :

$$s_{\text{RMSE}} \approx 1 - \frac{\text{RMSE}}{\text{RMSE}_p} \quad (11)$$

où  $\text{RMSE}_p$  est la RMSE calculée entre la persistance et la mesure, la persistance apportant l'information de la variabilité. Dans sa forme initiale, la faiblesse de cette métrique réside dans sa dépendance au modèle de ciel clair choisi, puisque la grande variété de résultats de ces modèles va donc influencer sur les résultats de  $s$ . Yang et al. (2018) notent également que pour la forme suggérée par l'équation (11), calculer la RMSE pour la persistance du GHI plutôt que de l'indice de ciel clair conduirait à un score apparemment meilleur. Pour toute utilisation de ce score, il sera ainsi nécessaire d'indiquer le modèle de ciel clair utilisé et/ou la méthode de persistance appliquée.

### II.3.2.c Agrégation temporelle (moyenne temporelle)

La manière dont sont agrégées les mesures temporellement a également son importance. Chaque source d'information a une résolution temporelle qui lui est propre ; or le sur-échantillonnage ou le sous-échantillonnage des informations a des conséquences sur l'analyse. La technique la plus couramment utilisée pour le sous-échantillonnage d'une information pyranométrique initialement échantillonnée à quelques secondes est d'effectuer une moyenne des mesures précédentes de la manière suivante :

$$\forall i, \quad G_i^{SE} = \frac{1}{N} \sum_{k=1}^N G_{i-k} \quad (12)$$

où  $N$  est le rapport du nombre d'indices de la série d'origine  $G$  et de celui de la série sous-échantillonnée  $G^{SE}$ . Une autre approche moins courante consiste à simplement prélever à intervalle régulier la valeur du signal d'origine. Au vu de la variabilité de la mesure à la minute, il n'est généralement pas recommandé d'opter pour cet échantillonnage discret.

Un effet bien connu des experts du domaine est la réduction des erreurs statistiques d'amplitude lorsqu'elles sont calculées sur les échantillons moyennés sur des pas de temps plus larges (Kaur et al., 2016). Cette caractéristique ne doit pas être négligée dans la comparaison de différentes études. L'horizon de prévision et son pas de temps sont donc intrinsèquement liés aux résultats de la méthode de prévision évaluée. Yang et al. (2017b) proposent une méthode permettant d'intégrer des prévisions dont les pas de temps et les horizons diffèrent. Cette méthode agrège les différentes séries temporelles en une série temporelle typiquement au pas de temps 6 h, en prenant en considération les différences intrinsèques de chaque modèle pour s'assurer de la cohérence de la réconciliation plutôt que de les sommer simplement.

### II.3.2.d Agrégation spatiale

De la même manière que l'agrégation temporelle a tendance à réduire les erreurs résultantes, l'agrégation spatiale est elle aussi une technique diminuant l'erreur de l'analyse. Les prévisions régionales agrégées ne peuvent donc pas être directement comparées aux prévisions locales. De façon similaire à la conciliation temporelle des prévisions disparates, Yang et al. (2017a) ont proposé une approche permettant de réconcilier les prévisions locales avec des prévisions régionales pour une utilisation conjointe. Les auteurs analysent différents estimateurs régressifs utilisant la structure hiérarchique du réseau électrique.

### II.3.2.e Périodes de test

Certains auteurs privilégient une analyse des performances sur une longue période de test, ce qui se justifie par une meilleure représentativité des résultats. Cependant, la disponibilité des données et leur qualité ne sont pas toujours assurées sur de longues périodes, et les effets spécifiques à certains climats vont avoir une incidence sur les résultats et ne seront donc pas uniquement révélateurs des caractéristiques de la méthode évaluée. Aussi, nombre de méthodes fonctionnent par apprentissage, pour lesquelles une séparation des données est nécessaire en une partie dédiée à cet apprentissage, et une partie, tota-

lement indépendante, réservée à l'évaluation des performances. Cette pratique limite aussi la période disponible pour l'évaluation des différentes méthodes.

D'autres périodes de test peuvent être envisagées, par exemple en appliquant la procédure de caractérisation des performances uniquement sur les jours de ciel clair (Mellit et Pavan, 2010), de ciel couvert (Lonij et al., 2013) ou selon différentes classes de journées (Dambreville, 2014). Une telle classification des jours permet d'identifier les situations sur lesquelles la méthode évaluée est moins performante. Cette stratégie peut être utilisée en classifiant les instants ou les saisons ou selon d'autres critères, comme nous le verrons dans la partie II.6.

### **II.3.2.f Localisation et spécificités du site d'étude**

Finalement, la dépendance à la localisation et aux installations déployées sur le site à l'étude est un problème qui limite la possibilité de comparaison des études entre elles. La stratégie la plus courante est d'appliquer l'analyse sur le plus de sites possible avec une certaine diversité de climats représentés. Cependant le nombre de stations de mesures de qualité est limité, et chaque ajout de site d'étude nécessite une analyse préliminaire sur la qualité des données et les conditions d'acquisition de celles-ci qui peuvent s'avérer chronophages. Dans notre étude, nous avons privilégié une approche d'évaluation des performances à la fois globalement sur l'ensemble des données, mais également catégorielle sur des classes de données, ce qui sera explicité plus en détail dans la partie II.6. Cette approche a pour objectif d'aller vers une plus grande indépendance des scores affichés vis-à-vis des données d'entrée et donc du site étudié.

## **II.3.3 Le cas de la métrique RMSE**

Comme nous l'avons déjà souligné, un aspect limitant de la RMSE – et qui s'étend aussi aux autres métriques existantes – est la difficulté de comparer les scores entre les différentes études. La spécificité de la RMSE avec sa pénalisation quadratique des erreurs de grande amplitude a pour conséquence de privilégier des comportements lisses qui prennent peu de « risque » d'un point de vue de la prévision. Cependant, dans notre étude de la variabilité à très court terme, les comportements de basses fréquences des séries temporelles à l'étude sont principalement sous-tendus par la composante « ciel clair » de l'éclairement. L'intérêt se porte donc plus sur le suivi des variations brusques, ce qui n'est pas tout à fait bien caractérisé par la RMSE. Minimiser ce score dans un processus d'optimisation peut encourager le choix de réduire les composantes de hautes fréquences de la prévision résultante. Dans ce paragraphe, trois méthodes de prévision sont utilisées à des fins illustratives. Une description plus précise de leur implémentation est disponible au paragraphe II.7.2.

Afin de mieux comprendre cette limitation de la RMSE, intéressons-nous à l'exemple concret de la figure 11. La sous-figure (a) représente l'évolution du rayonnement solaire mesuré au 05/07/2015 sur le site n°1 (en noir), superposé avec la courbe journalière cor-

respondant à la prévision météorologique (en vert) délivrée par l'organisme ECMWF (centre européen pour les prévisions météorologiques à moyen terme) avec un pas horaire de 3 h puis sur-échantillonné à 15 min tel que décrit au paragraphe II.7.2.c. La particularité de cette prévision est d'être émise à minuit pour toute la journée. Du fait de la faible résolution spatiale (environ 12 km) et temporelle (3 h) de la donnée météorologique d'origine, la série temporelle résultante est extrêmement lisse et est donc incapable de détecter le contenu à haute fréquence du rayonnement solaire, qui est capturé par le pyranomètre au sol. Pour le jour en question, une telle prévision affiche un score de RMSE de  $110 \text{ W/m}^2$  (26%).

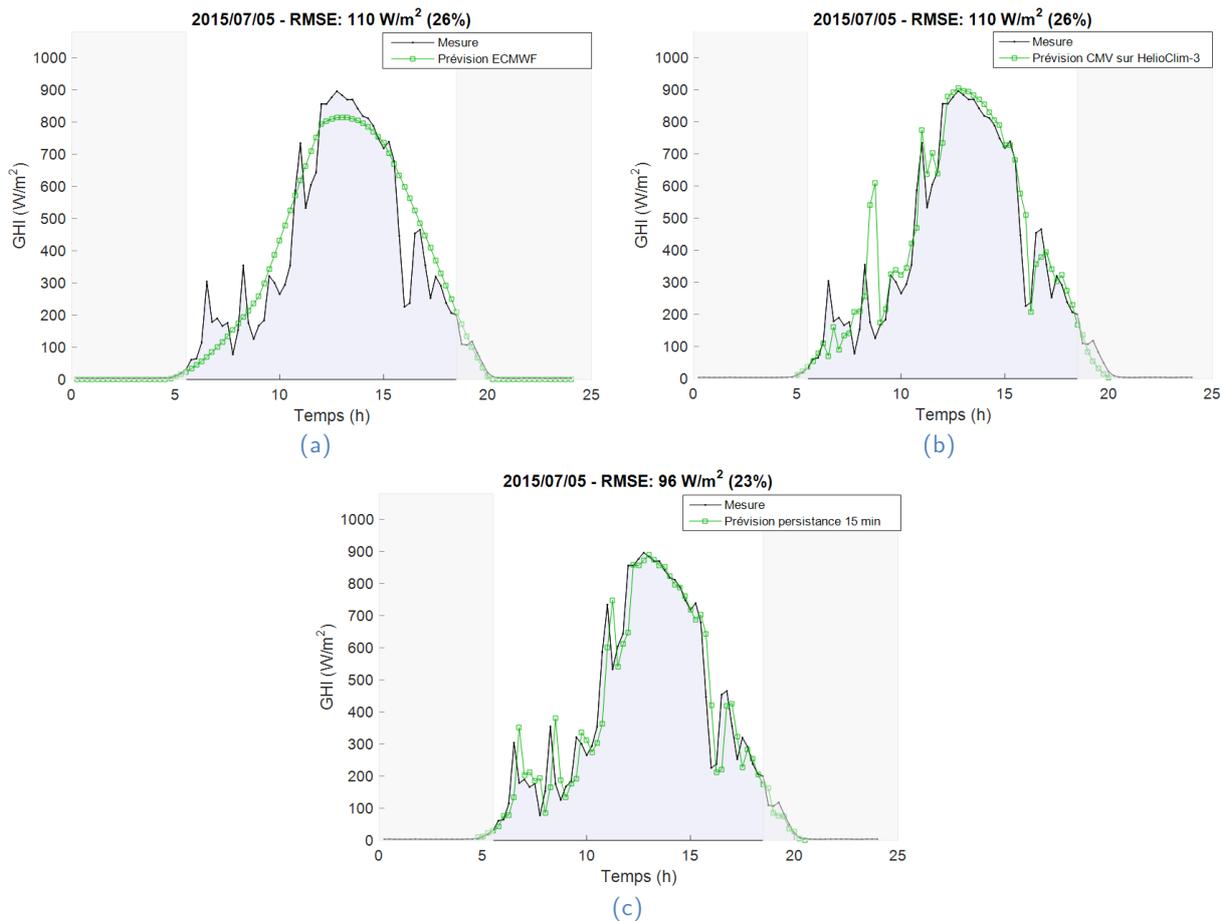


Figure 11 - Exemple de la journée du 05/07/2015 sur le site n°1 pour trois méthodes de prévision de référence (en vert) : la prévision ECMWF interpolée à 15 min (a), une prévision satellite simple utilisant les images HelioClim (b) et la persistance de l'indice de ciel clair (c). Sur les trois sous-figures, la mesure est représentée en noir.

La figure 11.b représente en vert une prévision satellitaire (cf. II.7.2.b pour le détail de la méthode), qui repose sur l'estimation du mouvement des nuages (*Cloud Motion Vectors*, CMV<sup>4</sup>) entre deux images successives, puis sur une hypothèse de persistance de ces vecteurs déplacements au pas de temps suivant. Les images utilisées pour générer cette prévision proviennent de la base de données HelioClim-3 de cartes de rayonnement solaire, issues de la conversion des images satellites du visible par la méthode Heliosat-2 (Lefevre

<sup>4</sup> Une description de ces méthodes d'estimation des CMV est proposée à l'annexe A1. La méthode de prévision appelée « CMV sur HelioClim-3 » est décrite au paragraphe II.7.2.b.

et al., 2002; Rigollier et al., 2004) Pour tout le manuscrit, HelioClim fera référence aux images de cette base HelioClim-3.

En suivant l'hypothèse de persistance des vecteurs déplacements, il est ainsi possible de fournir une première prévision du rayonnement solaire au pas de temps et à l'horizon 15 min, en prenant le pixel qui va arriver sur le site d'étude en prolongeant les vecteurs déplacements. Cette méthode de prévision a des caractéristiques qui lui sont propres : elle a tendance à mieux suivre les fluctuations du rayonnement solaire mais la limite de résolution spatiale des images satellites (environ 3 km au nadir) et les effets parallaxe (cf. partie III.1.5) peuvent expliquer en partie les erreurs restantes. De manière analogue à la prévision ECMWF, la RMSE sur ce jour particulier est de  $110 \text{ W/m}^2$  (26%). Or les comportements des deux séries temporelles sont radicalement différents. Cette erreur de RMSE s'explique en partie par la forte pénalisation de l'erreur de prévision commise à 8:45. La surestimation de ce pic de rayonnement incident est le résultat d'une mauvaise localisation d'une trouée dans la couverture nuageuse passant au-dessus du site d'intérêt à cet instant, ce qui est particulièrement difficile à anticiper uniquement avec l'information satellite à cause du phénomène de parallaxe, de la résolution spatiale limitée de ce type d'images et de la déformation des nuages qui n'est pas prise en compte ici.

Enfin, la figure 11.c affiche l'évolution de la prévision par persistance (en vert), qui profite du modèle théorique de ciel clair pour générer une prévision « naïve » des fluctuations du GHI. Une telle méthode est caractérisée par un suivi quasi parfait des variations de la mesure (en noir) mais systématiquement en retard d'un pas de temps de 15 min. Si l'on se cantonne au score de RMSE, cette méthode simple semble avoir de meilleures performances ( $96 \text{ W/m}^2$ , 23%) sur ce jour que la méthode météorologique ou la méthode satellite choisie.

En dépit des scores très proches délivrés par la RMSE pour ces trois méthodes de prévision que l'on considère comme référence, les comportements de ces trois séries temporelles prévues sont intrinsèquement différents par construction. Parmi ces comportements, certaines caractéristiques sont particulièrement intéressantes pour les utilisateurs des prévisions et méritent d'être mises en avant lors de l'analyse des performances. Les attentes des utilisateurs des prévisions sont décrites dans le paragraphe suivant.

Ce type de limitations est aussi présent sur les autres métriques très utilisées comme la MAE ou le biais. En effet, chaque métrique est conçue pour indiquer une information spécifique et n'est pas suffisante en soi. L'approche suggérée dans cette étude, et détaillée au paragraphe II.7, consiste à privilégier une analyse multicritère de manière à décrire le comportement des prévisions à l'aide d'un panel de métriques préalablement choisies, plutôt que de se reposer sur une unique métrique qui ne sera jamais parfaite.

### II.3.4 Rappel des attentes des utilisateurs

Florita et al. (2013) remarquent un attrait récent des opérateurs réseau vers une meilleure compréhension de l'impact des événements extrêmes sur le système électrique. En particulier, l'accent est mis sur la caractérisation des « grandes rampes » – ou fortes et brusques variations – afin d'en étudier l'incidence sur la fiabilité et la rentabilité du système. Notamment, les impacts d'une accumulation de rampes sur le réseau électrique sont indéniables selon Kudo (2015). L'étude propose une analyse du lien entre les rampes extrêmes constatées et les données météorologiques correspondantes, sur deux exemples de journées.

Comme nous le constaterons au paragraphe II.4.1, le concept de rampes a fait l'objet de nombreuses tentatives de détection et de caractérisation. Dans le contexte prédictif, il est fréquent de mal estimer ces fortes fluctuations, que ce soit dans son amplitude, sa durée ou son instant d'occurrence. En effet, les modèles météorologiques peuvent facilement dévier spatialement, engendrant alors une prévision correcte mais décalée temporellement. De même, les prévisions qui reposent sur l'hypothèse d'advection monocouche des nuages (sur des images satellites ou des images hémisphériques) peuvent être source d'erreur d'amplitude ou temporelle dans la description des fortes fluctuations – l'effet parallaxe étant un facteur détériorant davantage la qualité de ces prévisions.

L'objectif des prévisions de court-termes est en premier lieu une aide à la gestion de la variabilité de la production solaire afin de fournir le délai et la visibilité suffisante sur les événements à venir (Du et al., 2017). Les opérateurs réseau tentent d'exploiter ces prévisions dans leurs stratégies décisionnelles qui cherchent une meilleure rentabilité tout en assurant des niveaux de fiabilité très élevés de leurs services. Pour que cette information prédictive constitue effectivement une aide à la décision, il est primordial que les erreurs de prévision n'engendrent que peu de conséquences économiques pour l'ensemble du système. Or, lorsqu'une grande variation de production intervient, si elle n'a pas été prévue ou si elle a été prévue trop en avance ou trop en retard, il y a des conséquences sur l'optimisation économique recherchée.

Du point de vue du réseau électrique, la variable qui fait foi est l'écart entre la production et la consommation d'électricité à tout instant, de manière à assurer un voltage, un courant et une fréquence à l'intérieur des intervalles requis. Ainsi, bien que la RMSE soit souvent utilisée pour sa pénalisation des larges erreurs, elle ne décrit pas totalement la capacité à détecter ces événements extrêmes au bon moment.

## II.4 Présentation de méthodes exploratoires qui tentent de combler le manque lié aux méthodes existantes et en réponse aux besoins utilisateurs

Ce paragraphe traite des méthodes exploratoires sélectionnées qui entendent préciser et diversifier l'information, avec souvent une intention d'harmoniser les métriques de performances avec les besoins des utilisateurs exprimés précédemment. Deux concepts phares sont notamment mis en avant :

- le premier a d'abord été exploité dans le contexte éolien mais commence à trouver sa place dans le domaine solaire, il s'agit de la propension d'une méthode de prévision à suivre les "rampes". Comme indiqué dans le paragraphe II.3.4, cette notion suscite un intérêt croissant pour la gestion de la puissance photovoltaïque délivrée sur le réseau. La méthode exploratoire qui est d'un intérêt particulier pour notre étude repose sur l'algorithme *Swinging Door*, qui sera exploité lors de la proposition d'une nouvelle métrique sur les rampes au paragraphe II.5.1.
- La notion d'erreur temporelle suggère de ne plus considérer uniquement l'erreur comme une erreur d'amplitude mais plutôt comme la concomitance d'une erreur temporelle et d'une erreur d'amplitude. Cette approche suggérée par Frías-Paredes et al. (2016) sera utilisée dans la proposition de la nouvelle métrique du TDM (*Temporal Distortion Mix*) au paragraphe II.5.2.

### II.4.1 Suivi des rampes

Les évènements extrêmes s'avèrent être une préoccupation majeure des utilisateurs des prévisions, ce qui explique en partie l'émergence de nouvelles recherches sur la caractérisation des dénommées « rampes » afin d'en faciliter la gestion. La notion de rampe a été notamment définie en 2007 dans le cadre de la prévision éolienne par (Cutler et al., 2007) comme étant un changement de la puissance de sortie ayant une amplitude suffisamment élevée sur une courte période. Cette définition, peu précise, a été l'objet de plusieurs tentatives d'améliorations. Ainsi, Ferreira et al. (2011) ont-ils proposé une série de quatre définitions d'une « rampe » dans le cas de l'énergie éolienne en se basant sur trois paramètres intrinsèques des rampes : leur amplitude, leur durée et leur direction. Des seuils sur ces paramètres doivent ainsi être identifiés de manière à détecter une rampe à partir d'une certaine valeur. Ces seuils sont souvent directement liés aux caractéristiques spécifiques du parc étudié, telles que la puissance nominale installée. Cependant, cette puissance nominale peut être amenée à changer en cas de postes défaillants ou de maintenance (Kamath, 2010) et n'est pas forcément connue, c'est pourquoi d'autres définitions de rampes doivent être envisagées.

Une autre approche consiste à considérer tous les changements d'amplitude pour un pas de temps donné. Avec cette approche, il est possible de calculer la pente typiquement observée : le *Ramp Rate* (Kleissl, 2013; Kudo, 2015; Lave et Kleissl, 2013). Ce *Ramp Rate* est uniquement déterminé sur la base de la série temporelle. Les probabilités de détection (POD), la probabilité de fausse alarme (FAR) et un indice d'amplitude (RMI) (Chu et al., 2015) permettent aussi d'évaluer la qualité de la prédiction des rampes en comparaison avec la mesure. Ces probabilités nécessitent la définition d'un seuil qui conditionne la détection ou non d'une rampe. Ces métriques sont prometteuses mais ont l'inconvénient de ne détecter que des rampes pour un pas de temps fixe. Or les rampes qui semblent avoir un impact fort sur le réseau peuvent avoir différentes durées.

Un algorithme d'approximation d'une courbe par morceaux suivant des pas de temps irréguliers a été proposé par Bristol en 1990 pour une application de compression de données. Cette méthode, appelée « *Swinging door* » – ou porte battante en français –, est réapparue en 2013 dans un article de Florita et al. en démontrant l'intérêt d'une telle méthode pour l'identification de rampes dans les séries temporelles de puissances électriques produites par des systèmes solaires ou éoliens. Cette méthode, illustrée par la figure 12, identifie les points de départ et d'arrivée des rampes en se basant sur un unique paramètre  $\epsilon$ . Ce paramètre « d'approximation » détermine la sensibilité de l'algorithme aux petites variations. L'opération réalisée est une sorte d'approximation par morceaux irréguliers de la courbe initiale. Lorsque  $\epsilon$  est faible, une grande quantité de rampes est potentiellement détectée alors que pour un  $\epsilon$  plus grand, seules les rampes les plus importantes sont détectées. Ce paramètre est donc central puisqu'il régit la détection des rampes telle qu'expliquée ci-dessous.

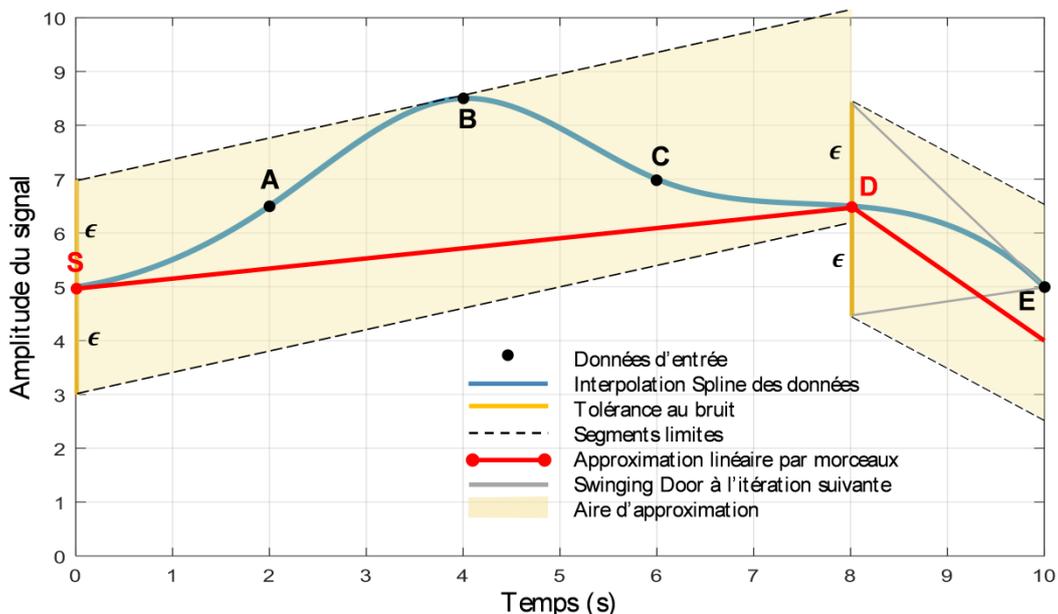


Figure 12 – Algorithme "Swinging door" pour l'extraction de rampes contenues dans une série temporelle. Source : inspiré de Florita et al. (2013).

L'algorithme démarre au point initial  $S$  puis détermine le segment le plus long possible respectant la condition d'une erreur d'approximation entre le segment et la courbe d'au plus  $\epsilon$ . Dès que l'erreur entre le segment et la série temporelle d'origine dépasse  $\epsilon$  – comme c'est le cas pour le segment [SE] sur la figure 12 – l'algorithme réitère en prenant en entrée le point de mesure précédent – c'est-à-dire le point  $D$ .

Une optimisation de cet algorithme a été proposée par Cui et al. (2017, 2016, 2015) en fusionnant des rampes adjacentes de même direction. Ces articles introduisent trois définitions d'une rampe significative avec des seuils reposant sur la puissance nominale installée et sur la durée des rampes. Quatre définitions de rampes significatives ont également été proposées par Zhang et al. (2014) pour les installations éoliennes, se reposant là encore sur les caractéristiques en puissance des installations. Dans l'optique d'introduire une métrique qui puisse être universelle, peut-être sera-t-il préférable de ne pas se reposer sur la puissance installée ? Cependant, la notion de rampe significative est pertinente dans l'étude de la production photovoltaïque. En effet, certaines rampes pouvant être détectées dans la mesure du rayonnement solaire par un seul pyranomètre peuvent finalement être peu perceptibles dans la puissance de sortie. En effet, la variabilité temporelle peut être grandement réduite, par foisonnement (*i.e.* décorrélation partielle), lors de l'agrégation spatiale de la production de systèmes photovoltaïques géographiquement répartis (Hoff R., 2012; Hoff et Perez, 2010; Perez et al., 2016). Une définition d'une rampe significative pourrait s'appuyer sur les rampes qui se retrouvent effectivement dans la puissance finalement délivrée, compte tenu de ces effets de foisonnement. Cette puissance délivrée peut par exemple être estimée au préalable à l'aide d'un modèle simulant la variabilité qu'une centrale solaire atteindrait en cumulé. À cet effet, le modèle de variabilité en ondelettes (WVM, *Wavelet Variability Model*), présenté par Lave et al. (2013), se base sur une mesure ponctuelle du rayonnement, la densité spatiale de panneaux photovoltaïques et la vitesse des nuages de la journée pour en estimer la production agrégée sur le support spatial.

#### **II.4.2 Erreur temporelle : le *Temporal Distortion Index* (TDI)**

Frías-Paredes et al. (2016) ont suggéré une approche différente des analyses des performances des méthodes de prévision. L'idée est de décomposer l'erreur de prévision en une part de « distorsion temporelle » et une part d'erreur d'amplitude. Cette distorsion est caractérisée par un indice de distorsion temporel (TDI, *Temporal Distortion Index*) qui quantifie l'erreur temporelle qui est faite par la prévision. Cette distorsion est quantifiée grâce à un algorithme classique de *Dynamic Time Warping* (DTW) (Sakoe et Chiba, 1978). Ce type d'algorithme fait partie de la famille de la programmation dynamique, dont le principe consiste à décomposer un problème complexe en problèmes locaux plus simples. L'approche permet d'aligner temporellement la série temporelle de la prévision sur celle de la référence, de manière à appliquer dans un second temps les métriques statistiques quantifiant l'erreur d'amplitude : une fois temporellement alignée, le TDI est estimé et l'erreur d'amplitude est évaluée, par exemple par la MAE (Frías-Paredes et al., 2017, 2016).

L'idée de l'algorithme DTW est d'associer les indices de la série temporelle de test  $T = (T_1, T_2, \dots, T_N)$  avec la série temporelle de référence  $R = (R_1, R_2, \dots, R_N)$ , où  $N$  est la longueur de la série test, considérée comme égale à celle de référence pour simplifier l'explication (le raisonnement reste valide pour des longueurs différentes). Cet appariement s'effectue en associant une distance locale à chaque couple d'indices possible entre la série de test  $T$  et la série de référence  $R$ . La distance proposée par Frías-Parades est l'erreur absolue instantanée entre les couples  $(T_i, R_j)_{0 < i < N, 0 < j < N}$  :

$$\forall (i,j) \in [1:N]^2, d(T_i, R_j) = |T_i - R_j| \quad (13)$$

D'autres distances ont été proposées par la communauté scientifique dans le but d'affiner cette attribution des points de la série test aux points de la série de référence. Macrae et Dixon (2010) ont suggéré l'utilisation d'une corrélation locale afin de caractériser la distance entre les deux séries temporelles de manière plus robuste, par la formule suivante :

$$\forall (i,j) \in [1:N]^2, d_w(T_i, R_j) = 1 - \sum_{k \in [-w,w]} \frac{\langle T_{i-k}, R_{j-k} \rangle}{\|T_{i-k}\| \|R_{j-k}\|} \quad (14)$$

où  $w$  définit la taille de la fenêtre sur laquelle calculer cette distance. Le choix de la distance dans le calcul du TDI sera discuté lors de la présentation de nouvelles métriques, au paragraphe II.5.2.

Le calcul de ces distances donne lieu à une matrice dite de coûts cumulés  $D$  qui indique la somme des distances selon un motif d'évolution récursive (*step pattern*) prédéfini :

$$\forall (i,j) [1:N]^2, D(i,j) = d(T_i, R_j) + \min \left\{ \begin{array}{l} D(i-1, j) \\ D(i, j-1) \\ D(i-1, j-1) \end{array} \right\} \quad (15)$$

Plusieurs motifs d'évolution de cette étape de programmation dynamique ont été évalués par Frías-Paredes et al. (2017, 2016) afin de contraindre plus ou moins les mouvements possibles du chemin optimal. Dans cette étude nous nous limitons au motif standard décrit dans l'équation (15) qui ne privilégie pas de chemin a priori.

L'étape suivante consiste à identifier le chemin  $w$  qui minimise le coût total entre les indices des deux séries temporelles :

$$w = ((i_1, j_1), \dots, (i_l, j_l), \dots, (i_k, j_k)), \text{ avec } k \in \mathbb{N} \quad (16)$$

La fonction de coût total que l'on cherche à minimiser est alors définie comme la somme des distances locales entre les deux séries calculées sur la séquence de points  $(i_l, j_l)$  avec  $l \in [1, k]$ .

Un exemple de chemin optimal détecté par l'algorithme est proposé en figure 13.b pour les séries temporelles artificielles proposées par Frías-Paredes et al. (2016) sur la figure 13.a. Une fois le chemin optimal trouvé, le TDI est défini comme l'aire entre le chemin

optimal et le chemin identité (en rouge sur la figure 13.b), normalisée par l'aire sous la droite identité, afin d'en extraire un pourcentage de la distorsion temporelle maximale :

$$\text{TDI} = \frac{1}{N^2} \sum_{l=1}^{k-1} |(i_{l+1} - i_l)(i_{l+1} + i_l - j_{l+1} - j_l)| \quad (17)$$

Cette formulation du TDI sous la forme d'une somme de Riemann décrite par Vallance et al. (2017) découle du développement de la forme intégrale du TDI proposée initialement par Frías-Paredes et al. (2016).

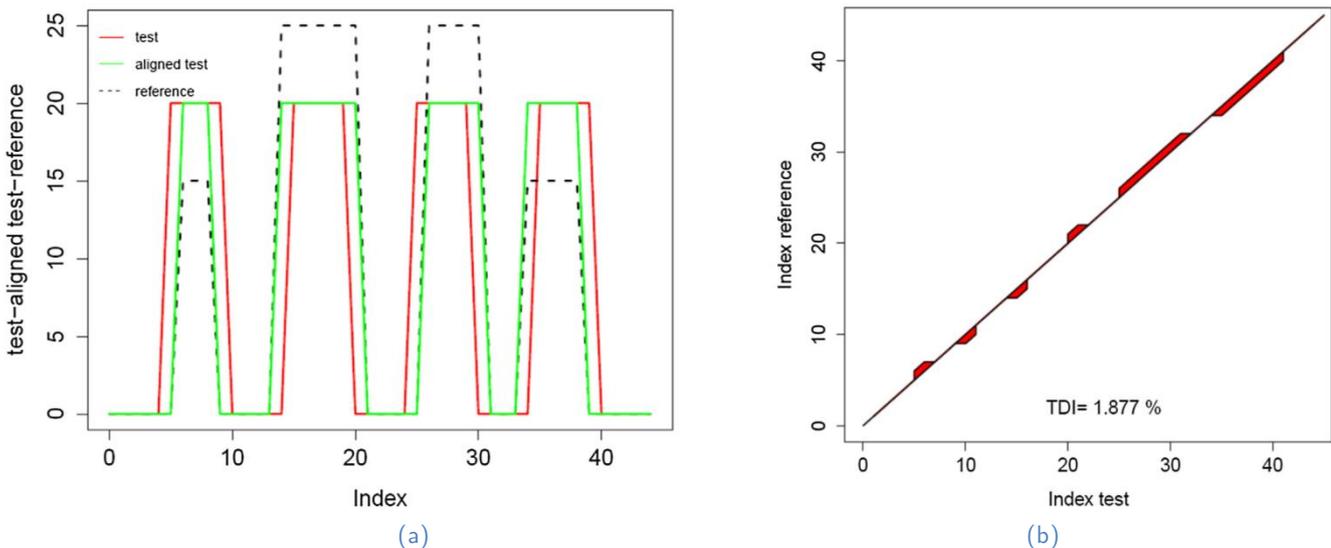


Figure 13 - Calcul du TDI : (a) séries temporelles artificiellement générées et (b) délimitation de l'aire entre le chemin optimal et l'identité. [Source](#) : Frías-Paredes et al. (2016).

L'algorithme DTW permet donc d'associer les indices de la série de test à ceux de la série de référence. Cependant, ce regroupement, qui dépend de la distance  $d$  et du motif d'évolution choisis, est subjectif. L'appariement indiciel perd en effet du sens lorsque l'on essaye d'identifier la distorsion temporelle entre deux signaux très différents par nature. Frías-Paredes et al. (2017) suggèrent un choix du motif récursif permettant de minimiser la MAE entre la série alignée et la série de référence : un exemple de résultat résultant de ce choix est disponible sur la figure 14. Intéressons-nous aux événements entre les heures 45 et 50. Avec la distance  $d$  et le motif récursif choisis, la variation de la prévision sur cet intervalle est considérée comme une erreur temporelle et est associée au mauvais événement. De plus, lorsqu'un événement n'apparaît que dans l'une des deux séries temporelles, à quel indice faut-il l'associer ? Pour tenter de limiter ces défauts de l'analyse de la distorsion temporelle, nous recommandons d'utiliser la distance énoncée à l'équation (14). De même, dans l'optique d'utiliser cet algorithme dans un cadre standardisé de l'analyse des performances, nous préconisons l'utilisation du motif récursif le plus simple, suivant l'équation (15), qui ne privilégie pas de chemin a priori.

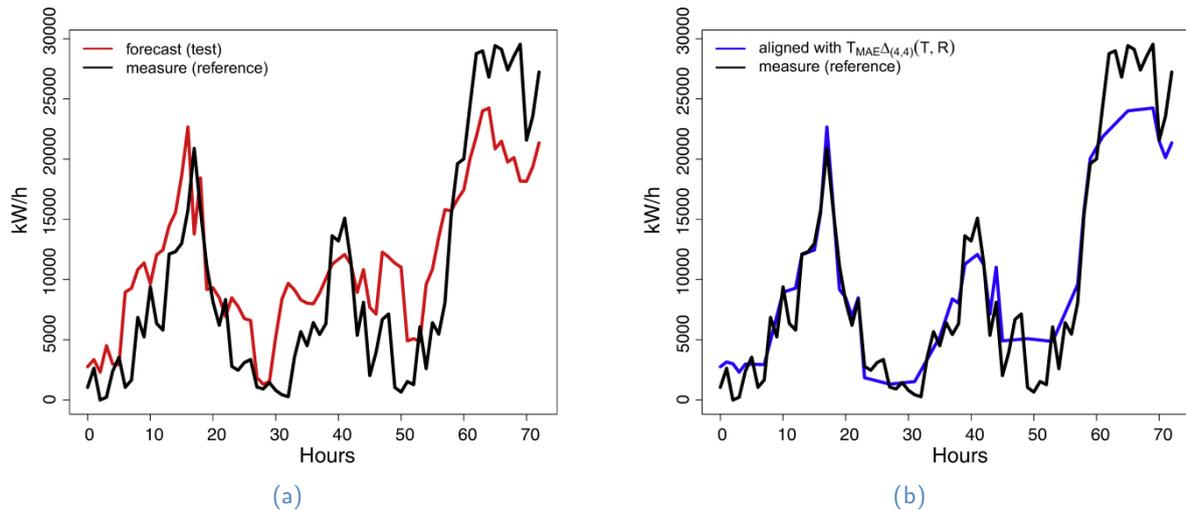


Figure 14 – (a) Evolution de l'énergie éolienne mesurée (noir) et prévue (rouge) sur trois jours. (b) prévision alignée sur la mesure (bleu). [Source](#) : Frías-Paredes et al. (2017).

## II.5 Métriques proposées

Florita et al. (2013) identifient l'inadéquation des métriques statistiques utilisées par la communauté scientifique avec les besoins des opérateurs système, où les facteurs les plus critiques du point de vue utilisateur ne sont souvent pas pris en compte. Deux aspects essentiels ont été identifiés comme étant en accord avec la réalité de la gestion de la variabilité de la production solaire : la capacité d'une méthode de prévision à appréhender les fortes variations de la mesure, et sa capacité à établir une prévision alignée temporellement avec la mesure. La première métrique proposée exploite la décomposition en rampes délivrée par l'algorithme *Swinging Door* – décrite au paragraphe II.4.1 – appliquée sur la série temporelle prévue ainsi que celle mesurée, dans le but d'en quantifier l'écart entre les pentes détectées. La seconde repose sur l'approche proposée dans le calcul du TDI – décrit au paragraphe II.4.2 – afin d'en créer un score capable de quantifier la tendance d'avance ou de retard de la méthode de prévision testée.

### II.5.1 Quantifier la capacité de la prévision à suivre les rampes

Comme discuté au paragraphe II.4.1, la détection de rampes dites "significatives" pour l'analyse a connu un intérêt croissant ces dernières années. La métrique que nous avons proposée (Vallance et al., 2017) vise à quantifier la capacité d'une méthode de prévision à capturer les fluctuations significatives du rayonnement solaire. La significativité des rampes détectées par l'algorithme *Swinging Door* est maîtrisée par son unique mètre  $\epsilon$ , ce qui en fait un algorithme adapté à notre analyse. L'idée sous-jacente est de prendre en compte uniquement les fluctuations susceptibles d'avoir un impact sur le système photovoltaïque. Ainsi, dans un premier temps, la métrique proposée décompose séparément les séries temporelles de la mesure et de la prévision en leur approximation respective par morceaux, issues de l'algorithme *Swinging Door*.

Un exemple d'une telle décomposition est proposée sur la figure 15.a, où les séries temporelles de la mesure et de la prévision à partir des données HelioClim (cf. II.7.2.b) sont représentées avec leurs aires d'approximation – constituée d'une séquence de parallélogrammes de largeur  $\epsilon$  – issues de cette décomposition pour  $\epsilon = 90 \text{ W/m}^2$  arbitrairement choisi. Ensuite, l'évolution temporelle des pentes des rampes extraites est représentée sur la figure 15.b. La déviation absolue moyenne des rampes, notée ramp-MAD, est définie par l'aire normalisée entre les deux séries de pentes détectées par l'algorithme *Swinging Door*, exprimée en  $\text{W}\cdot\text{m}^{-2}\cdot\text{h}^{-1}$ , et représentant un équivalent de la différence d'amplitude moyenne horaire entre les deux séries de rampes :

$$\text{ramp}_{\text{MAD}} = \frac{1}{t_{\text{max}} - t_{\text{min}}} \int_{t_{\text{min}}}^{t_{\text{max}}} |SD_{\epsilon}(T(t)) - SD_{\epsilon}(R(t))| dt \quad (18)$$

où  $SD_{\epsilon}$  donne les pentes des rampes extraites par l'algorithme *Swinging Door* de paramètre  $\epsilon$ ,  $t_{\text{max}}$  et  $t_{\text{min}}$  sont les bornes de la période considérée.

Lorsque les rampes de la mesure et de la prévision vont dans la même direction, l'écart de pente est faible, ce qui correspond à une aire plus petite sur la représentation  $SD(t)$  et donc à un score faible. Sur la figure 15, les pentes de la mesure et de la prévision sont dans des directions opposées entre 10:00 et 10:15, ce qui contribue à l'augmentation de l'erreur de suivi de rampes.

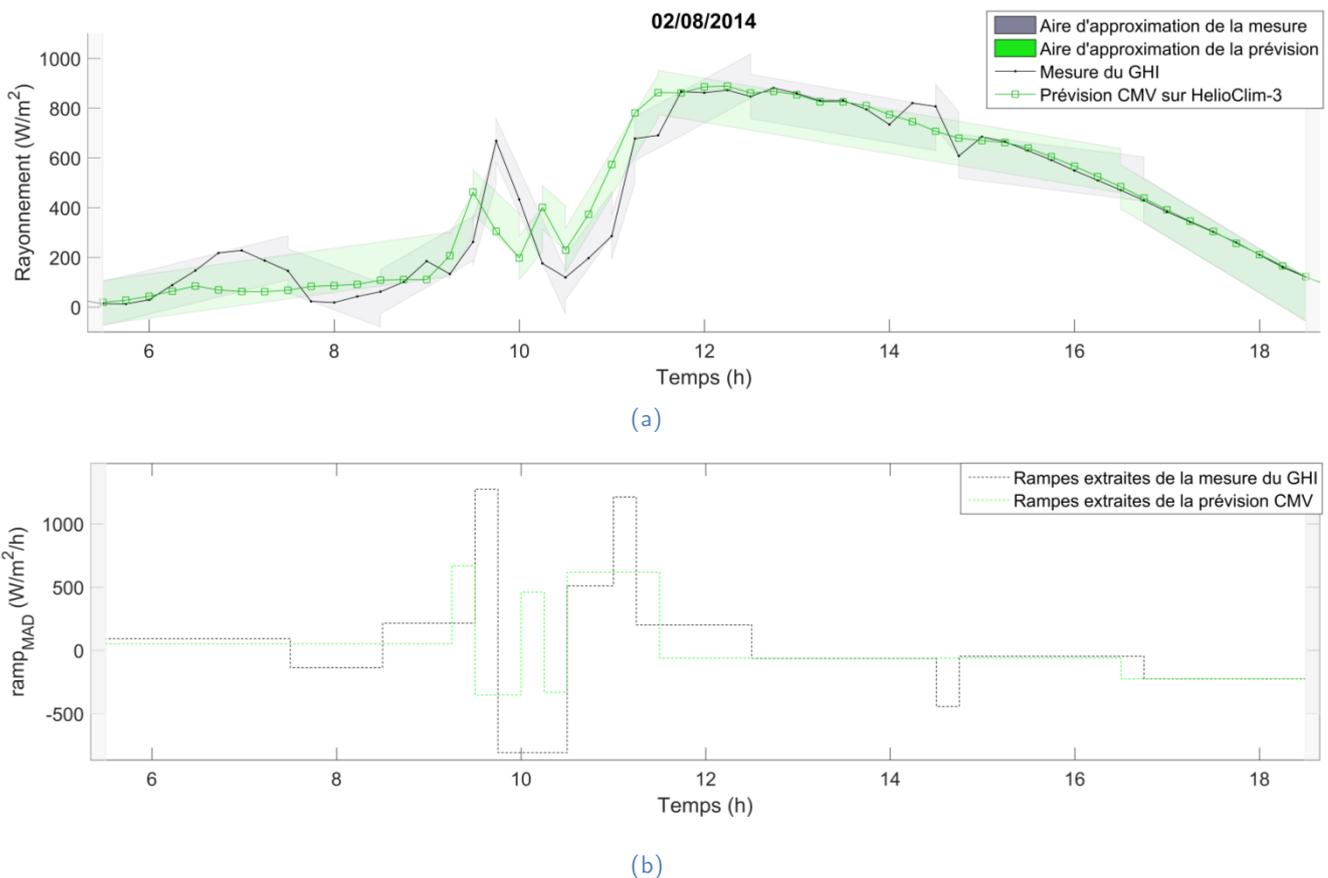


Figure 15 – Détection de rampes issues de l'algorithme *Swinging Door* appliqué aux mesures et de la méthode CMV sur les cartes HelioClim (a) avec leur représentation de la pente en fonction du temps (b).

Le paramètre  $\epsilon$  définit la sensibilité d'amplitude aux fluctuations de l'éclairement dans l'algorithme *Swinging Door*. Néanmoins, compte tenu des effets saisonniers, une variation importante de l'éclairement en été peut avoir une amplitude plus importante qu'une variation en hiver. Considérer un paramètre fixe  $\epsilon$  tout au long de l'année conduirait, en moyenne, à moins de rampes détectées en hiver qu'en été. Par conséquent, nous proposons de définir  $\epsilon$  par jour comme une fraction constante du maximum quotidien du rayonnement dans des conditions de ciel clair. Cette fraction est un nouveau paramètre, noté  $\tau_{CLS}$ , compris entre 0 et 1 :

$$\epsilon(d) = \tau_{CLS} \max_{\text{day } d} G^{CLS} \tag{19}$$

La figure 16 montre un exemple d'approximation par morceaux faite par cet algorithme pour la journée 2014/08/10 afin de mieux comprendre l'effet de  $\tau_{CLS}$ . Ce paramètre doit être réglé en fonction de l'amplitude des rampes que l'on souhaite détecter, ce qui dépend de l'utilisation finale. Ce paramètre définit la frontière entre les fluctuations significatives et les autres. Le pourcentage de rampes détectées en fonction de la MAE relative engendrée par l'approximation est représenté sur la figure 17. Le choix de  $\tau_{CLS}$  est en effet un compromis à trouver entre ces deux grandeurs : le pourcentage de rampes détectées et la MAE relative résiduelle.

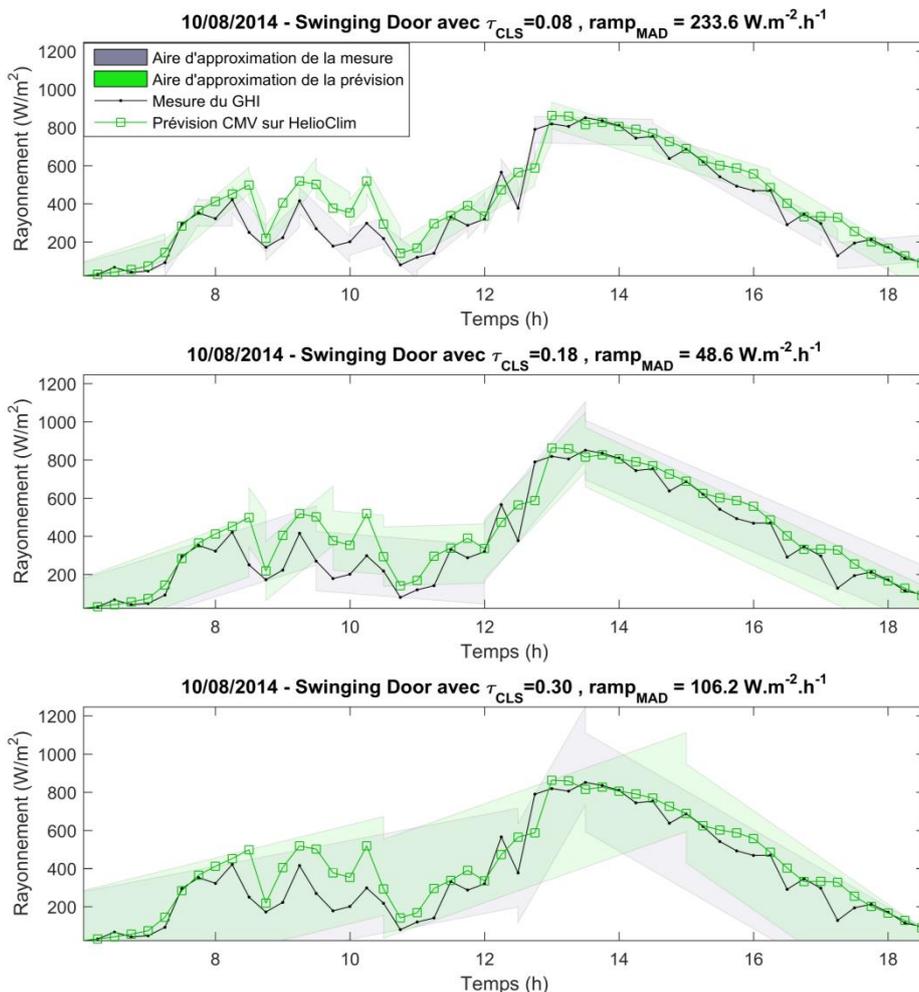


Figure 16 – L'algorithme *Swinging Door* appliqué le jour 2014/08/10 pour différentes valeurs de  $\tau_{CLS}$ .

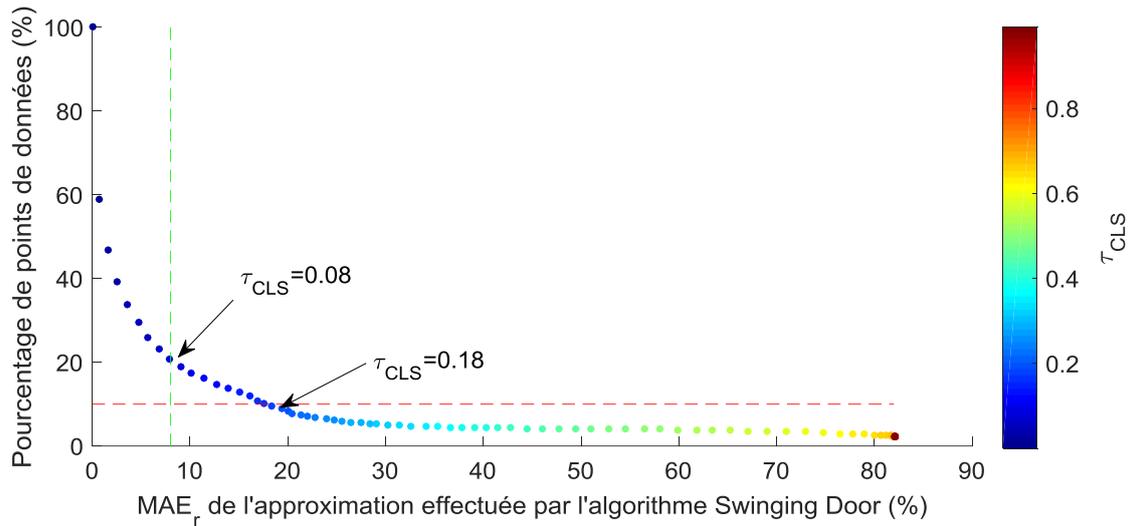


Figure 17 – Pourcentage de rampes détectées dans la série temporelle de la mesure via l'algorithme Swinging Door en fonction de la MAE relative résultante, calculée entre la mesure et son approximation par morceaux.

L'instrument pyranométrique dans notre cas est un SPN1 avec une incertitude sur les lectures individuelles de  $\pm 8\% \pm 10 \text{ W/m}^2$  (Badosa et al., 2014). Nous nous attendons à ce que l'algorithme Swinging Door effectue une approximation de la série temporelle de rayonnement en rampes « significatives ». Lorsque le paramètre  $\tau_{CLS}$  est trop petit, la détection de rampes résultante est trop sensible à l'incertitude de mesure. Par conséquent, nous recommandons de définir  $\tau_{CLS}$  de manière à obtenir une erreur d'approximation de la rampe au moins supérieure à l'incertitude. Inversement, comme l'illustre la figure 16, pour  $\tau_{CLS} = 0.3$ , une valeur trop élevée de  $\tau_{CLS}$  conduirait à des rampes détectées qui ne représentent pas fidèlement des variations significatives du rayonnement. Dans cette étude, afin de caractériser la variabilité « significative », nous avons fixé  $\tau_{CLS} = 18\%$ , ce qui correspond à 10% des rampes détectées. D'autres définitions de rampes significatives sont disponibles dans la littérature (Zhang et al., 2014). Cependant, ces définitions dépendent généralement d'un pourcentage de la puissance installée de l'installation, qui n'est pas toujours connue. Ainsi, le paramètre  $\tau_{CLS}$  n'est pas à optimiser mais à choisir en fonction du contexte.

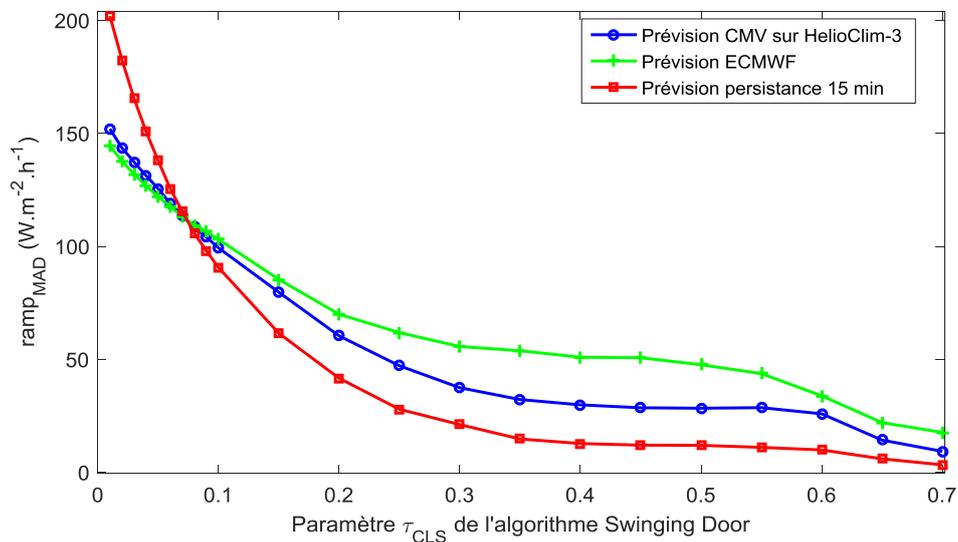


Figure 18 - Evolution de la métrique de rampe en fonction du paramètre  $\tau_{CLS}$ .

La figure 18 montre l'évolution de la métrique ramp-MAD en fonction de  $\tau_{CLS}$  pour un pas de temps de 15 min pour le site n°1 (des résultats similaires sont obtenus pour le site n°2). Comme indiqué sur ce graphique, en dessous d'une sensibilité de 7% du rayonnement maximal quotidien par ciel clair, la méthode de persistance présente des erreurs de rampe plus élevées que la méthode CMV et une erreur de rampe plus élevée que la méthode ECMWF. Cependant, la persistance est plus apte à prévoir à 15 min des rampes que les méthodes CMV et ECMWF lorsque le paramètre de sensibilité est supérieur à 7%. En effet, la persistance identifie les rampes, mais seulement après qu'elles se soient réellement produites, ce qui signifie que le  $\tau_{CLS}$  choisi permet l'association des rampes avec un certain décalage temporel. Cette métrique seule n'est pas suffisante pour adresser la problématique du décalage temporel. Ainsi, le principe de calcul du TDI sera exploité au paragraphe II.5.2 pour proposer un critère complémentaire.

De plus, afin de permettre une détection des rampes qui ne soit pas parfaitement en phase entre les deux séries chronologiques, il est important de permettre un certain niveau d'approximation. Ce comportement explique pourquoi la métrique  $ramp_{MAD}$  est plus élevée pour la persistance avec un  $\tau_{CLS}$  bas mais diminue à mesure que le signal est simplifié par l'algorithme *Swinging Door*. Avec le paramètre de sensibilité choisi  $\tau_{CLS} = 18\%$ , la figure 18 montre que la méthode de persistance a une erreur de rampe inférieure à celle de la méthode CMV et qui a une valeur d'erreur de rampe elle-même inférieure à la prévision ECMWF. Ces caractéristiques sont cohérentes avec le comportement de ces trois méthodes de prévision vues au paragraphe II.3.3.

## II.5.2 Quantifier l'alignement temporel de la prévision par rapport à la mesure

Quantifier la tendance à générer une prévision alignée temporellement (i.e. ni en retard, ni en avance) sur la mesure est une problématique qui peut être approchée de plusieurs manières. Les gestionnaires de réseaux portent en effet une attention particulière à la minimisation de l'écart entre consommation et production, et un décalage temporel dans la prévision – qu'elle soit en avance ou en retard – peut avoir des conséquences néfastes sur le réseau en cas d'événement extrême. Fournir une prévision qui est systématiquement mal alignée temporellement pourrait même engendrer des coûts plus importants qu'une gestion sans prévision. Face à cette problématique, il est donc nécessaire d'avoir des outils permettant d'indiquer l'avance ou le retard des prévisions. La connaissance de l'alignement temporel des prévisions est essentielle non seulement pour les utilisateurs des prévisions solaires (gestionnaires de réseau ou de parcs photovoltaïques), mais également pour les concepteurs des prévisions. En effet, il est fréquent que des erreurs d'indice se glissent dans l'implémentation de telles méthodes, ou que les effets parallaxes non corrigés se traduisent par des décalages.

Ainsi qu'expliqué dans le paragraphe II.4.2, Frías-Paredes et al. (2016) ont proposé un indice permettant de quantifier la distorsion temporelle apparente entre deux séries temporelles. Nous avons repris cette approche en y ajoutant une décomposition de cette distorsion estimée en deux parties, une partie représentant la contribution « en avance » de cette distorsion et une partie celle « en retard ». De cette décomposition nous proposons un indice signé négatif si la prévision est en avance, 0 si la prévision est alignée temporellement et positif si la prévision est en retard. Une valeur de 100 % (respectivement -100%) signifie que lorsqu'un décalage temporel est détecté, il est systématiquement en retard (respectivement en avance). Ces scores indiquent la composition – ou mix – de la distorsion temporelle estimée en termes de retard et d'avance. D'où le nom de cet indice : TDM (*Temporal Distortion Mix*).

La nécessité d'une distinction des composantes « avance » et « retard » de la distorsion temporelle est illustrée sur la figure 19. La sous-figure (a) présente une prévision par persistance qui est par construction systématiquement en retard, et la sous-figure (b) affiche la prévision utilisant la mesure des déplacements des nuages par HelioClim qui est mieux alignée temporellement mais qui manque d'une description précise du contenu à haute fréquence de la mesure. Cependant, les valeurs du TDI calculées sont sensiblement les mêmes pour ces deux jours. De manière à distinguer ces deux comportements différents – qui ont des implications différentes du point de vue opérationnel comme du point de vue de la conception des méthodes –, l'approche algorithmique du DTW exprimée dans la partie II.4.2 est reprise afin d'extraire le chemin optimal qui associe les indices de la série de la mesure avec ceux de la série de la prévision. Le calcul du *TDI* est effectué sous les deux conditions  $i \geq j$  et  $i \leq j$ , pour lesquelles nous utiliserons les notations  $TDI_{advance}$  et  $TDI_{late}$  :

$$TDI_{advance} = \frac{1}{N^2} \sum_{\substack{l=1 \\ i_l \geq j_l \\ i_{l+1} \geq j_{l+1}}}^{k-1} |(i_{l+1} - i_l)(i_{l+1} + i_l - j_{l+1} - j_l)| \quad (20)$$

$$TDI_{late} = \frac{1}{N^2} \sum_{\substack{l=1 \\ i_l \leq j_l \\ i_{l+1} \leq j_{l+1}}}^{k-1} |(i_{l+1} - i_l)(i_{l+1} + i_l - j_{l+1} - j_l)| \quad (21)$$

Ces deux composantes sont respectivement proportionnelles à la partie située au-dessous et au-dessus de la courbe d'identité de la zone rouge de la figure 13.b. Les composantes « avance » et « retard » sont définies en pourcentage du TDI total :

$$Adv = \frac{TDI_{advance}}{TDI} \quad \text{et} \quad Late = \frac{TDI_{late}}{TDI} \quad (22)$$

Ces composantes « avance » et « retard » sont complémentaires et peuvent être considérées comme les deux contributions de la distorsion temporelle globale. Par conséquent, nous définissons le mélange de distorsion temporelle TDM comme étant :

$$\text{TDM} = 2 \cdot \text{Late} - 1 = 1 - 2 \cdot \text{Adv} \quad (23)$$

Le TDM est exprimé en pourcentage signé, où la valeur -100% (respectivement + 100%) décrit une prévision systématique à l'avance (respectivement en retard).

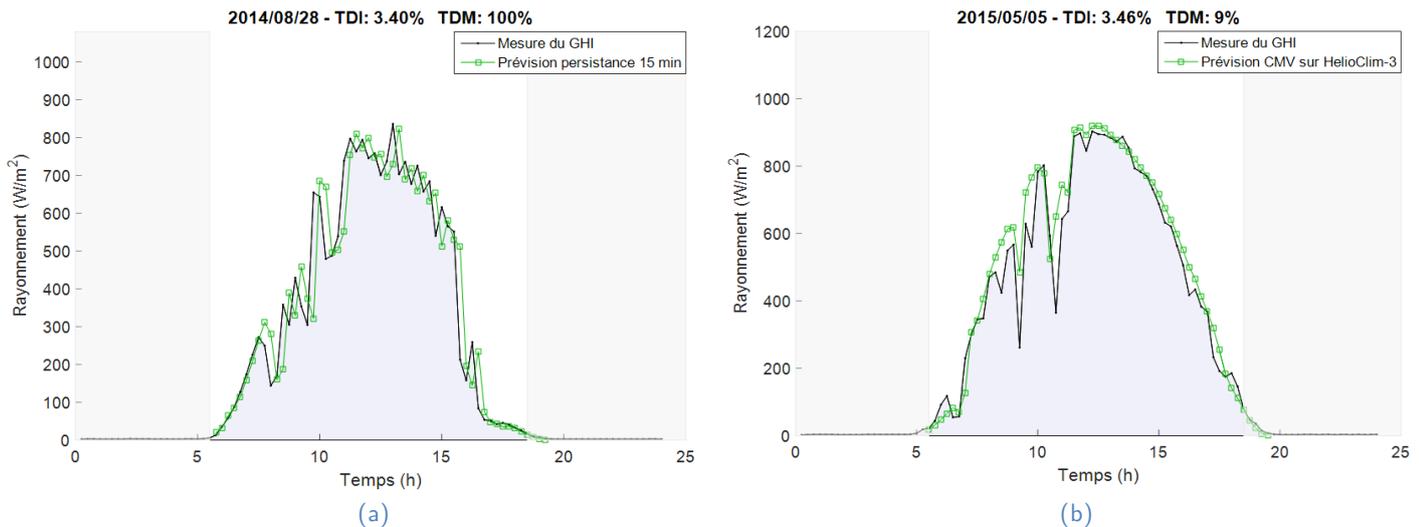


Figure 19 – Exemples de deux jours donnant la même valeur TDI entre les données mesurées et les prévisions. La prévision sur la sous-figure de gauche (a) est toujours en retard alors que la courbe de droite (b) montre une prévision qui admet un meilleur alignement temporel.

Lors du calcul du TDI, d'autres distances peuvent être utilisées pour associer des paires de points de données. En fait, une inadéquation dans l'utilisation de la distance euclidienne a été observée lors du calcul du TDM : pour une prévision qui a un équilibre entre ses composantes d'avance et de retard, la série temporelle alignée résultante peut ne pas avoir un TDM nul. Comme exprimé dans la partie II.4.2, la corrélation locale (14) proposée par Macrae et Dixon (2010) est capable de résoudre cette anomalie d'alignement temporel. Nous choisissons pour cette étude une taille de fenêtre  $w = 2$  pour le calcul de cette distance, ce qui semblait le plus à même de satisfaire la condition suivante : détecter 100% de retard sur la persistance et ne pas perturber l'équilibre avance-retard lors de l'alignement de la série de la prévision sur la mesure (c'est-à-dire  $\text{TDM} = 0$ ).

## **II.6 Caractérisation catégorielle des performances : vers une indépendance de l'ensemble des données utilisées**

La complexité de la dynamique atmosphérique induit une variabilité de la météorologie d'un instant, d'un jour, d'une saison et d'un endroit à un autre. Dans la majorité des publications approfondissant les domaines d'estimation ou de prévision de la ressource solaire, les positions géographiques des lieux d'étude sont indiquées. C'est un facteur évident dont dépendent les résultats de ces études, et les conclusions sont donc généralement conditionnées à cette localisation. Cependant, les résultats sont également dépendants de la fréquence d'occurrence des différentes situations dans le jeu de données utilisé. Par exemple, analyser les performances d'une méthode de prévision sur un jeu de données où le ciel était dégagé la plupart du temps donnera des résultats plus favorables que la même analyse effectuée sur un jeu de données admettant une plus grande variabilité. Cette dépendance des résultats d'études aux données est une problématique importante qui rend compliquée, voire impossible, l'extraction de conclusions sur les performances des méthodes proposées. Ainsi, bien que les métriques énoncées dans ce chapitre donnent des informations globales pertinentes lorsqu'elles sont calculées sur l'ensemble du jeu des données, une analyse plus fine des résultats est possible en indiquant ces scores sur des classes prédéfinies. Par exemple, analyser les performances des prévisions sur un site particulier ou un ensemble limité de données en distinguant les classes de situations peut permettre d'estimer les performances de ces prévisions sur un autre site ou un autre jeu de données. Cette estimation pour un autre site peut s'effectuer en analysant l'occurrence de ces situations puis en pondérant les résultats du premier site par la différence d'occurrence. Nous énumérerons de façon non exhaustive plusieurs stratégies de classification des données, qu'il s'agisse de regrouper les données dont les instants sont dans des situations similaires (II.6.1), ou par type de journées (II.6.2), ou autre (II.6.3).

L'analyse catégorielle des performances a pour ambition de rendre la comparaison inter-études plus légitime, mais elle prétend également contribuer à une clarification des circonstances atmosphérique sous lesquelles la méthode testée est plus ou moins performante. Les classifications proposées sont élaborées à différentes échelles temporelles. Le choix du nombre et du type de classes à distinguer dépendra à la fois du cadre de l'étude et de la résolution temporelle des données fournies. En effet, des données horaires ne permettent pas de capturer la haute variabilité temporelle du rayonnement, toute tentative d'en extraire une classe d'instant variables s'annonce donc vaine.

### **II.6.1 Classification d'instant**

La classification des instants est la catégorisation la plus fine que l'on puisse effectuer. Des détections d'instant de ciel sans nuages ont été proposées par Lefèvre et al. (2013), ou encore par Reno et Hansen (2016), classifiant chaque instant dans deux classes. Une autre classification a été proposée par Tina et al. (2012) autorisant une distinction circonstancielle en quatre classes : ciel clair, légèrement couvert, couvert et variable. Cet

l'algorithme est basé sur un «indice de sérénité»  $k$  et son écart absolu local MF à la moyenne mobile MA :

$$k(t) = \frac{|G(t) - G^{CLS}(t)|}{G^{CLS}(t)} \quad (24)$$

$$MA(t) = \frac{1}{N_m} \sum_{s=0}^{N_m-1} k(t-s) \quad (25)$$

$$MF(t) = \frac{1}{N_m} \sum_{s=0}^{N_m-1} |MA(t) - k(t-s)| \quad (26)$$

$k$  correspond à l'écart de rayonnement relatif à la courbe de ciel clair et MF décrit sa variabilité en utilisant les  $N_m$  instants précédents. Il est important de noter que cet algorithme ne nécessite que des données présentes et passées, sur quelques minutes. Il est donc applicable en analyse de données historiques mais aussi en temps réel.

La figure 20 schématise l'arbre de décision à l'origine de cette classification. Pour notre étude, nous avons choisi d'appliquer cette méthode à des mesures de 1 min afin de capturer les fluctuations rapides du rayonnement, qui seraient invisibles en moyennant à des pas de temps plus élevés. La période d'échantillonnage est importante pour cette classification car un pas de temps plus long conduirait à des courbes beaucoup plus lisses, les rendant plus difficiles à classer. D'ailleurs, Ransome et Funtan (2005) considèrent que les données horaires ne devraient pas être utilisées pour caractériser correctement la performance d'une méthode de prévision, car elles sont insuffisantes pour décrire la dynamique de la variabilité infrajournalière.

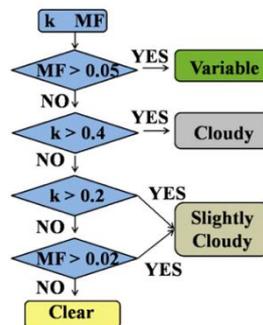


Figure 20 – Arbre de décision de l'algorithme de classification des instants proposé par Tina et al. (2012).

La classification d'origine proposée par Tina et al. (2012) s'appliquait sur des mesures pyranométriques au pas de temps 10 min avec une taille de fenêtre glissante  $N_m = 3$ , soit 30 min. Dans l'optique d'appliquer cette méthode à nos données à 1 min de résolution, nous avons choisi d'adapter uniquement la taille de la fenêtre glissante  $N_m$ , les seuils de la figure 20 étant conservés à l'identique.

Une base de données de situations étiquetée de référence est alors nécessaire afin d'adapter la fenêtre glissante de la classification sur le pas de temps 1 min. Ainsi, une interface utilisateur a été créée dans le but de faire étiqueter par des experts du rayonne-

ment solaire du centre OIE et d'EDF R&D des instants aléatoirement choisis sur la période du 01/08/2014 au 31/12/2015. Cinq experts du rayonnement solaire par site ont contribué à cet étiquetage en classifiant 1000 instants chacun selon l'une des cinq classes suivantes : "Ciel clair" (absence de nuages), "Légèrement couvert" (une faible nébulosité avec un faible impact sur GHI), "Couvert" (faible GHI et faible variabilité), "Variable" (forte variabilité autour de l'instant mesuré) et une classe échappatoire "Je ne sais pas". L'interface de la figure 21 propose à chaque expert de sélectionner la classe correspondante à la situation nuageuse de l'instant sélectionné en se basant sur les fluctuations locales du rayonnement. L'image hémisphérique correspondante à l'instant à étiqueter est également montrée aux experts afin de lever certaines ambiguïtés pouvant survenir lors de l'analyse de la courbe d'éclairement uniquement.

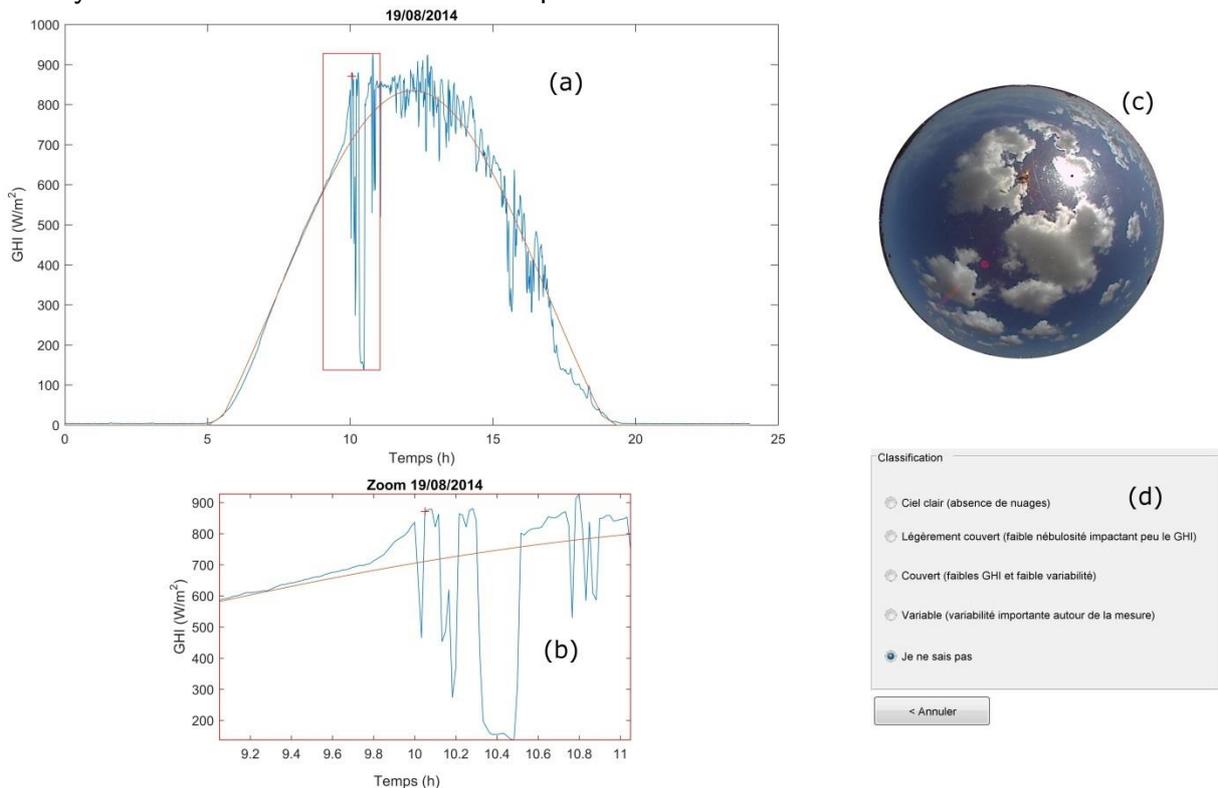


Figure 21 – Interface utilisateur développée afin de recueillir une base de données étiquetée selon cinq classes (d) à partir de la courbe de mesure du GHI (a), un zoom autour de l'instant à classifier (b) et l'image hémisphérique correspondante (c).

Sur chacun des jeux d'instant à étiqueter, 20 instants sont communs à tous les experts de chaque site de manière à évaluer la fiabilité de cette phase d'étiquetage et la cohérence entre les experts. La figure 21 montre les étiquettes associées à chacun de ces instants communs choisis au hasard. Sur 40% de ces instants, les experts sont unanimes. Les autres instants sont sujets à des divergences lors de cette étape d'étiquetage. En outre, nous remarquons que la frontière entre les définitions "Couvert" et "Variable" est quelque peu vague pour les experts, ainsi que pour les classes "Légèrement couvert" et "Ciel clair". La fiabilité de ces étiquettes, dont la figure 21 donne un aperçu, est considérée comme suffisante pour notre étude.

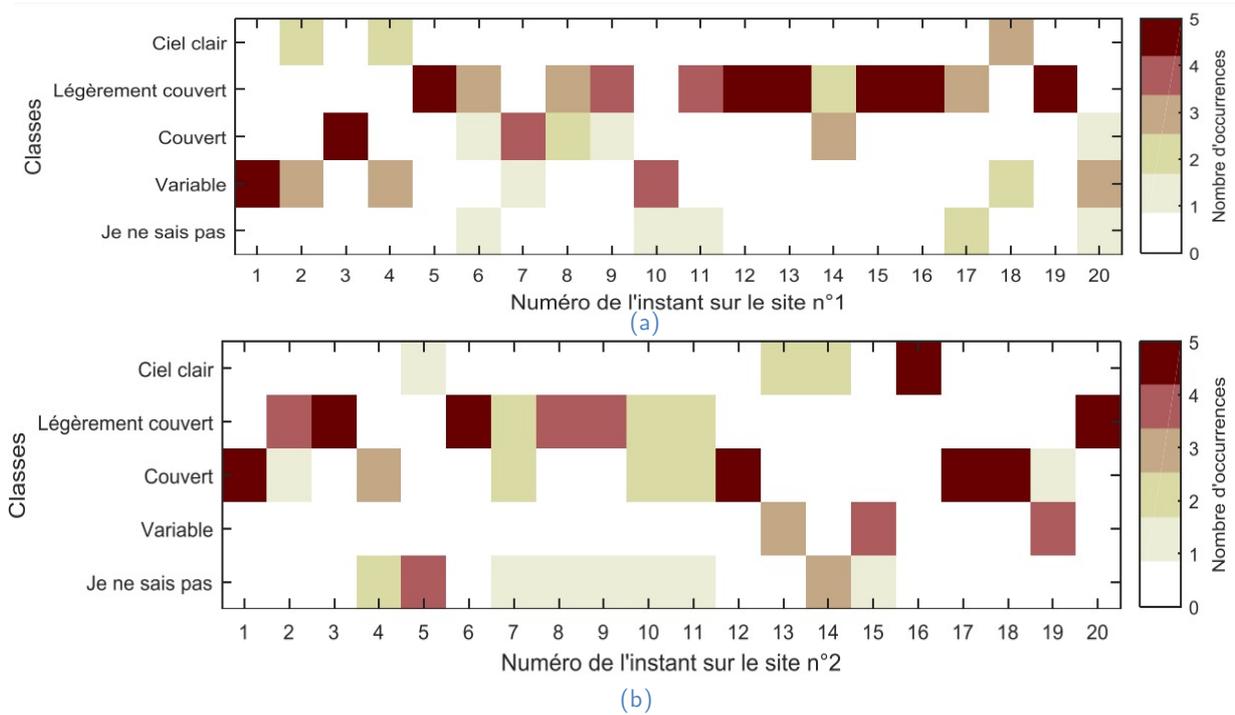


Figure 22 – Distribution des étiquettes pour les 20 instants communs aux cinq experts du site n°1 (a) et pour les 20 instants communs aux cinq experts du site n°2 (b).

| (%)                | Ciel clair<br>Tina et al. | Couvert<br>Tina et al. | Légèrement couvert<br>Tina et al. | Variable<br>Tina et al. |
|--------------------|---------------------------|------------------------|-----------------------------------|-------------------------|
| Ciel clair         | 87                        | 0                      | 13                                | 0                       |
| Couvert            | 0                         | 86                     | 3                                 | 11                      |
| Légèrement couvert | 52                        | 4                      | 29                                | 15                      |
| Variable           | 7                         | 12                     | 13                                | 68                      |
| « Je ne sais pas » | 17                        | 29                     | 18                                | 36                      |

Tableau 4 – Matrice de confusion issue de la classification des instants de Tina et al. pour  $N_m = 4$  au regard des instants étiquetés, sur des mesures du GHI au pas de temps 1 min.

Le tableau 4 est un exemple de matrice de confusion où chaque colonne de la matrice représente le nombre d'occurrences de la classe de la classification de Tina, tandis que chaque ligne représente le nombre d'occurrences des classes étiquetées par les experts. Le critère simple choisi pour déterminer la fenêtre glissante est le pourcentage de données sur la diagonale de cette matrice globale de confusion. Dans le contexte de cette classification, l'algorithme est appliqué a posteriori aux données mesurées, de sorte que nous pourrions utiliser une fenêtre glissante présentant les pas de temps passés et futurs. Cependant, puisqu'il peut être pratique d'avoir une méthode causale (i.e. applicable en temps réel) pour la prévision solaire, la fenêtre glissante causale de la forme  $[t - N_m; t]$  proposée par Tina est retenue dans cette analyse. Le pourcentage de données tombant sur la diagonale est maximisé pour  $N_m = 4$ . Cette valeur sera utilisée dans la suite afin d'assurer la meilleure correspondance avec les instants étiquetés. La précision totale de la classification vis-à-vis de celle de Tina, définie comme le nombre d'instants classifiés similairement divisé par le nombre d'instants, vaut 67,5%.

Les instants étiquetés par les experts comme "Légèrement couvert" sont considérés comme "Ciel clair" par l'algorithme de Tina et al. 52% du temps et seulement 29% sont correctement classés. Cette différence peut s'expliquer par la difficulté de différencier un ciel clair parfait d'un moment légèrement couvert basé uniquement sur l'évolution du GHI. En effet, comme observé pendant la phase d'étiquetage de la base de données, certains jours clairs comportaient des écarts par rapport à la courbe du ciel clair comparables à ceux causés par une couverture nuageuse légère en dehors de la région circumsolaire : la distinction entre ces deux classes n'est en fait visible qu'en observant les images hémisphériques associées à chaque instant. À l'exception de la classe "Légèrement couvert", la matrice de confusion affiche des valeurs relativement élevées sur la diagonale.

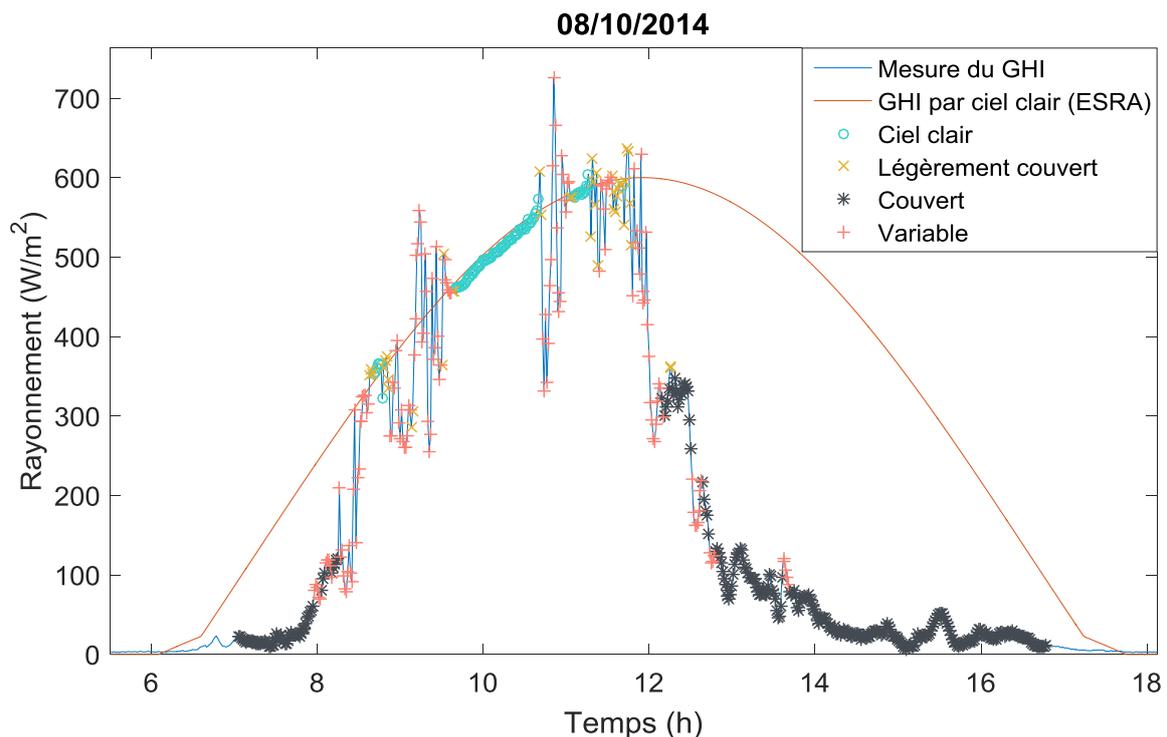


Figure 23 – Exemple résultant de la classification des instants en quatre classes appliquée le 2014/10/08 sur le site n°1.

Un exemple de résultats de cette classification est présenté à la figure 23. Nous limitons notre étude à cette classification car elle sert uniquement à valider les performances des méthodes de prévision sur ces différentes classes. L'imperfection de la classification des instants "Légèrement couverts" doit être gardée à l'esprit lors de l'analyse des résultats. Enfin, une classification plus efficace pourrait être obtenue en utilisant les caractéristiques des sous-ensembles étiquetés des experts en utilisant, par exemple, la logique floue.

## II.6.2 Classifications de jours

À une autre échelle de temps, une pratique de plus en plus considérée repose sur une classification des jours. À l'origine, la classification des instants de Tina et al. (2012) a été proposée afin d'attribuer un pourcentage de ces quatre classes à chaque jour. Une classification des jours peut donc aisément être exprimée en appliquant des seuils sur les proportions de ces classes. Une autre approche a été proposée par Gastón-Romeo et al. (2011). Cette classification repose sur une fonction dite de distribution de la taille granu-

lométrie (GSD, *Granulometric Size Distribution*) des courbes de rayonnement direct (DNI). Cette classification peut être envisagée sur les courbes GHI mais la variabilité est plus marquée sur le DNI. Plus formellement, il s'agit de construire la fonction  $F_{DNI}(\lambda)$  en appliquant une fonction d'ouverture à la courbe d'une longueur variable  $\lambda$  de la manière suivante :

$$F_{G_b}(\lambda) = 1 - \frac{m(\lambda)}{m(0)} \text{ avec } \begin{cases} m(\lambda) = \int \gamma_{\lambda B}(G_b)(t)dt \\ B = [-1, +1] \end{cases} \quad (27)$$

où  $\gamma_B$  une fonction d'ouverture, définie par :

$$\begin{aligned} \gamma_B &= \delta_B(\epsilon_B(G_b)) \\ \epsilon_B(G_b)(t) &= \min_{b \in B} G_b(t + b) \\ \delta_B(G_b)(t) &= \max_{b \in B} G_b(t + b) \end{aligned} \quad (28)$$

avec  $\epsilon_B$  la fonction d'érosion et  $\delta_B$  la fonction de dilatation de taille  $B$ .

La fonction résultante  $F_{DNI}$  d'une courbe de ciel clair sera donc proche de zéro pour de faibles valeurs de  $\lambda$  puis atteindra 1 (cf. Figure 24.1). Pour une courbe dotée d'une forte variabilité à l'inverse, la fonction  $F_{DNI}$  approchera plus rapidement la valeur 1 lorsque  $\lambda$  augmente (cf. Figure 24.2).

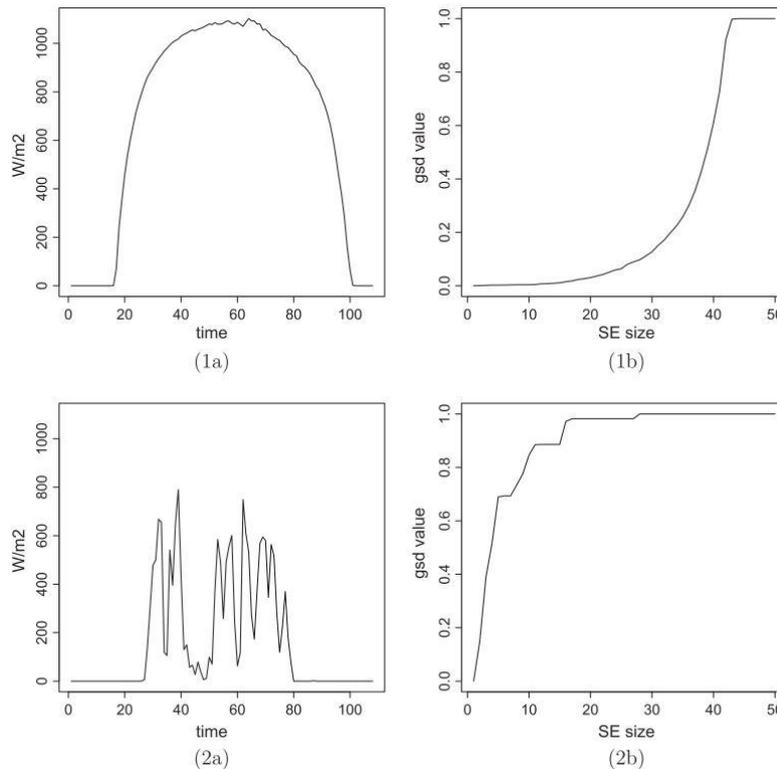


Figure 24 – Courbe de rayonnement direct par ciel clair (1a) et sa fonction GSD associée (1b). Courbe de rayonnement direct d'un jour très variable (2a) et les valeurs de sa GSD associée (2b). [Source](#) : (Gastón-Romeo et al., 2011).

La GSD  $F_{DNI}$  décrit la texture morphologique – appelée granulométrie – de la courbe du DNI en observant la courbe restante après « filtrage » de l'original avec la fonction d'ouverture de taille  $\lambda$ . Une GSD est calculée par jour. L'algorithme k-médoïdes (Van der Laan et al., 2003) est ensuite appliqué au groupe de courbes GSD sur la période d'intérêt.

L'algorithme k-médoïdes est similaire à l'algorithme k-means mais utilise des points de données représentatifs – les médoïdes – comme centre de chaque cluster. Gastón-Romeo et al. ont suggéré un regroupement en quatre classes, permettant la classification de chaque jour en fonction de leur variabilité infrajournalière. Cependant, la classe «ciel clair» qui en résulte contient des jours avec une variabilité notable et nous proposons donc un léger changement dans la procédure pour assurer une classe de ciel clair plus restrictive.

Afin d'extraire une classe plus « sélective », nous proposons une première classification préliminaire avec un nombre plus élevé de classes (50 dans notre cas, ce paramètre doit être adapté à chaque site). Ensuite, nous mettons de côté la classe avec la médoïde la plus basse (la classe ciel clair) et nous appliquons une deuxième fois l'algorithme k-médoïdes, avec 2 classes, sur les classes restantes. Un exemple de classification résultante des courbes GSD est proposé sur la figure 25. Des exemples de courbes de DNI résultant de cette étape de classification se trouvent sur la figure 26. Bien sûr, un nombre plus élevé de classes peut être sélectionné au cours de ce processus, mais pour notre cas d'étude, nous avons choisi de classer les jours en trois classes seulement. La méthode présentée dans ce paragraphe a l'avantage d'être indépendante de tout seuil autre que le nombre de classes et est donc applicable à tout autre site. Seule la sélection des jours de ciel clair nécessitera d'imposer un paramètre. D'autres classifications de journées ont également été proposées dans la littérature, comme celle de Dambreville (2014) qui utilise des seuils arbitraires sur la variabilité journalière et l'indice de ciel clair moyen, ou celles de Harrouni (2008), Harrouni et al. (2005) ou encore de Maafi et Harrouni (2003) basées sur des descripteurs fractals.

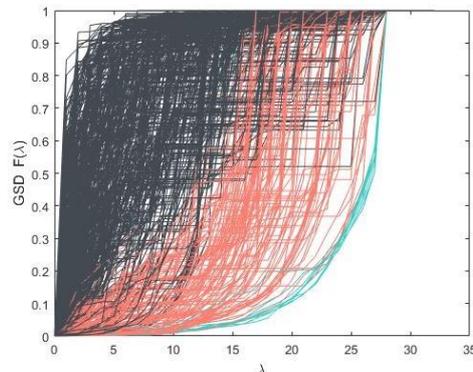


Figure 25 – Classification en trois classes de distributions de taille granulométrique (GSD).

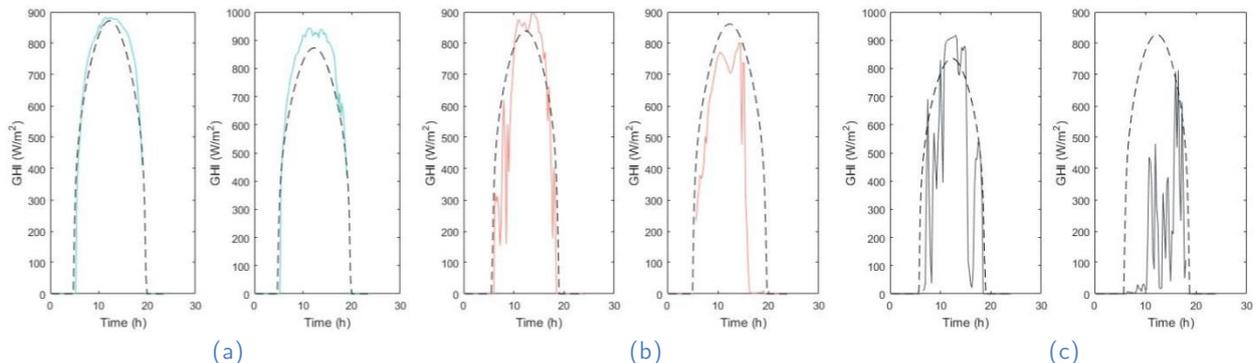


Figure 26 - Exemples de courbes de GHI classées par la méthode proposée selon 3 classes: (a) faible, (b) moyenne et (c) grande variabilité infrajournalière (la courbe en pointillés est la courbe de ciel clair).

### II.6.3 Autres classifications pertinentes

D'autres outils d'analyse peuvent être appropriés selon l'utilisation. Il peut par exemple y avoir un intérêt à classifier par saison, demi-journée, par élévation solaire ou par localisation afin d'identifier certaines caractéristiques propres à la prévision ou au site d'étude. En définitive, pour obtenir la caractérisation des performances la plus complète possible, il est nécessaire à la fois d'avoir à disposition des métriques permettant de mesurer les caractéristiques souhaitées, mais il est aussi nécessaire d'appliquer ces métriques sur différentes classes de données. L'objectif est d'être le moins dépendant possible de la composition des données, et ainsi indiquer des scores par catégories d'instant, de jours, de saisons ou autre plutôt que d'indiquer un unique score pour l'ensemble des données. En effet, cette approche a pour objectif d'analyser les erreurs de la méthode de prévision de manière détaillée, mais également de transposer les performances à d'autres sites ne présentant pas forcément le même nombre d'occurrence de situations par classe.

## II.7 Présentation de l'analyse des performances pour trois méthodes de prévision de référence

Cette partie entend illustrer l'analyse des performances décrite précédemment en l'appliquant à trois méthodes considérées comme référence (II.7.2) et en affichant les scores des métriques sélectionnées sous la forme d'un graphe radar (II.7.1). Les classifications d'instant et de jours de la partie II.6 sont utilisées pour extraire ces mêmes scores et graphes associés à chaque classe.

### II.7.1 Suggestion de représentation graphique des résultats et métriques sélectionnées

Nous avons adopté une représentation des résultats sous la forme d'un graphique radar pour faciliter la comparaison des performances des méthodes de prévision. Nous associons les six mesures suivantes à chaque branche du graphique radar : la RMSE relative, la MAE relative et le biais relatif sont tracés en bas; le ramp-MAD, le TDI et le TDM en haut. Cependant, puisque le biais et le TDM peuvent être négatifs, nous affichons leur valeur absolue avec un marqueur associé ('-' pour le négatif et '+' pour le positif). Avec cette représentation radar, la meilleure méthode de prévision – selon les métriques choisies – est celle qui a les valeurs les plus proches du centre du graphique. En particulier, les valeurs maximales de chaque métrique sont ajustées pour faciliter la lecture des résultats et leur éventuelle comparaison à d'autres méthodes.

## II.7.2 Trois méthodes de références utilisées pour l'illustration des avantages et inconvénients des métriques sélectionnées

Trois méthodes de prévisions ont été utilisées pour l'illustration des limitations de la RMSE au paragraphe II.3.3. Nous allons désormais apporter quelques précisions sur l'implémentation et la conception de ces méthodes afin de clarifier la compréhension des différents graphiques proposés dans le prochain paragraphe II.7.3. Ces trois méthodes sont dénommées prévision par persistance, prévision par CMV appliquée à HelioClim et prévision ECMWF.

### II.7.2.a Méthode de prévision par persistance

Lors de l'étude de la variabilité infrajournalière de l'éclairement, il est désormais classique de s'intéresser non pas à prédire l'éclairement mais plutôt l'indice de « ciel clair », noté  $k_c$ , qui a pour objectif de s'affranchir le plus possible des variations intra-journalières relativement lentes d'éclairement par ciel clair connues a priori avec une bonne de précision – et donc de se concentrer sur la variabilité infrajournalière du rayonnement solaire :

$$k_c = \frac{G}{G^{CLS}} \quad (29)$$

où  $G$  désigne le GHI et  $G^{CLS}$  la courbe théorique par ciel clair.

La persistance désigne de manière générale la technique consistant à considérer la valeur d'une variable présente comme une estimation de sa future valeur. Bien que la persistance sur le GHI soit parfois utilisée par la communauté scientifique, l'expression "persistance à  $\Delta t$  min" dénommera dans la suite la méthode de persistance appliquée à l'indice de ciel clair  $K_c$ , telle que :

$$\widehat{G}(t + \Delta t|t) = k_c(t) \cdot G^{CLS}(t + \Delta t) \quad (30)$$

où  $\widehat{G}(t + \Delta t|t)$  désigne la série temporelle prévue par cette méthode de persistance. Pour générer les résultats de cette méthode dans le cadre de cette étude, le modèle de ciel clair climatologique ESRA sera systématiquement utilisé.

### II.7.2.b Méthode de prévision par CMV sur HelioClim

Les méthodes de prévision satellitaires sont des approches reconnues comme étant performantes sur des horizons de prévision de 15 min à quelques heures (Dambreville, 2014). La popularité de ces méthodes est en partie due à leur praticité, leur couverture presque mondiale et leur faible coût de mise en œuvre. Les satellites météorologiques géostationnaires des familles METEOSAT, GOES, FENG YUN et HIMAWARI sont largement utilisés afin d'en dériver des estimations du rayonnement incident au sol (Blanc et al., 2017b). Que ce soit des approches physiques modélisant le transfert radiatif de l'atmosphère ou des approches empiriques ou semi-empiriques qui se basent sur un indice nuageux, leur résultat délivrent une estimation du rayonnement solaire au pas de temps 15 à 30 min et avec une résolution spatiale de 1 à 10 km. Dans le cadre de l'élaboration de cette méthode de référence, nous utilisons la base de données HelioClim-3 élaborée à partir de la méthode Heliosat-2 (Lefevre et al., 2002; Rigollier et al., 2004).

Les méthodes appelées *Cloud Motion Vectors* (CMV) tentent de profiter conjointement de l'information spatiale et temporelle de ces séries temporelles d'images. L'idée directrice est d'estimer le mouvement d'advection des masses nuageuses entre deux images successives. Les différentes techniques de calcul des CMV existantes pour la prévision satellitaire sont énumérées et discutées par Blanc et al. (2017). Pour notre étude, l'algorithme de flot optique développé par Liu (2009) dans le cadre de sa thèse a été choisi pour sa rapidité d'exécution, sa performance et la disponibilité de son implémentation<sup>5</sup>.

La dernière étape consiste à faire l'hypothèse de persistance de ces vecteurs de déplacements des nuages estimés, dont la validité décroît avec l'accroissement de l'horizon de prévision. La méthode que nous considérons pour l'illustration consiste à identifier le pixel – ou une valeur interpolée – qui va « atteindre » le site d'étude sous cette hypothèse de persistance des vitesses de déplacement des nuages. L'estimation de la future valeur du rayonnement incident est donc directement celle du pixel – ou de la valeur interpolée – de la dernière image HelioClim disponible. Cette méthode profite d'une meilleure localisation des nuages et d'une meilleure résolution spatiale et temporelle comparée aux prévisions météorologiques telle que celle proposée au paragraphe II.7.2.c. Cependant, leur limite de résolution spatiale et temporelle, les erreurs d'estimation de la ressource, de l'estimation des mouvements nuageux ou les erreurs de parallaxe vont dégrader et limiter les performances de ce type de prévisions.

### II.7.2.c Méthode de prévision ECMWF

Un modèle de prévision météorologique numérique de ECMWF (centre européen pour les prévisions météorologiques à moyen terme) appelé *Integrated Forecasting System (IFS)* (Morcrette et al., 2008) est utilisé. La prévision ECMWF est calculée tous les jours, à minuit et 12:00 UT, et donne les valeurs prévisionnelles pour le lendemain à un pas de temps de 3 h. Comme nous le verrons au paragraphe V.1.2, il existe plusieurs approches pour la comparaison de données n'ayant pas le même pas de temps. Cependant, interpoler simplement cette prévision initialement au pas de temps 3 h sur un pas de temps 15 min apporte des erreurs directement liées à la forme de la courbe du GHI par ciel clair qui n'est pas prise en compte dans l'interpolation. Ici, afin de ramener la courbe prévue par ECMWF à un pas de temps de 15 min, une interpolation linéaire est appliquée sur le paramètre  $k_t = \frac{\text{GHI}}{\text{TOA}}$  multiplié ensuite par le TOA. Ainsi, cette technique permet de s'affranchir des erreurs principales d'interpolation associées à la variation connue du TOA.

---

<sup>5</sup> <https://people.csail.mit.edu/celiu/OpticalFlow/>

### II.7.3 Résultats globaux et par catégorie

La procédure de caractérisation des performances exprimée dans ce chapitre a été appliquée aux trois méthodes décrites au paragraphe II.7.2. Ces méthodes ont été sélectionnées afin de mettre en lumière certaines caractéristiques bien connues des experts du domaine, et qui ont un intérêt particulier dans cette analyse.

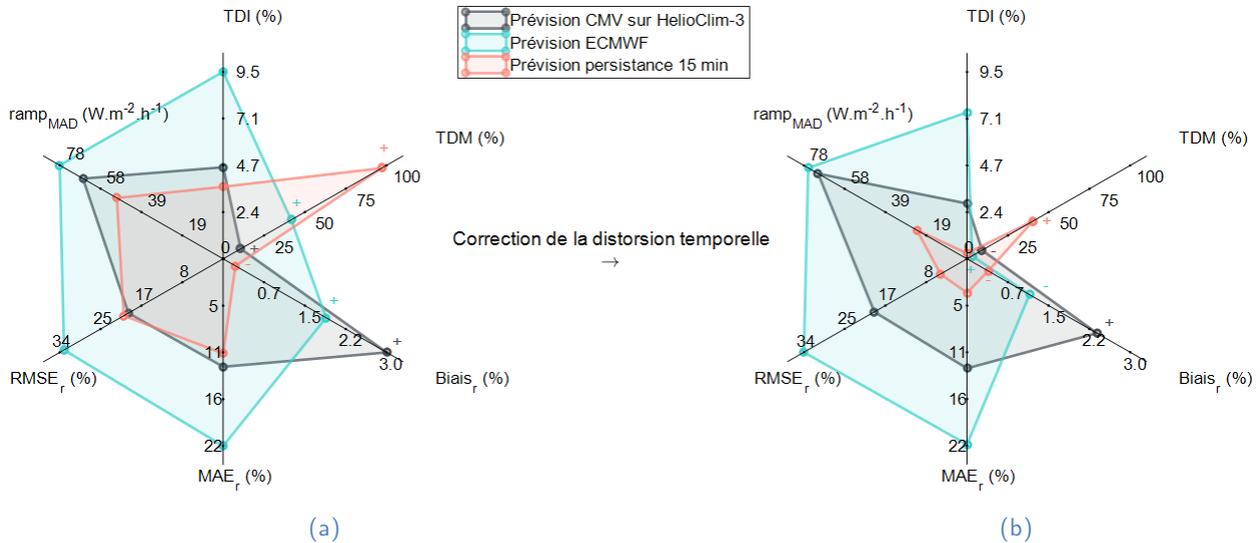


Figure 27 – Performances globales de la prévision par persistance 15 min, de la méthode CMV basée sur les images HelioClim et de la prévision interpolée ECMWF sur 15 min avant (a) et après (b) alignement des séries de prévisions pour le site n°1.

Figure 27.a illustre l'ensemble des résultats de chaque méthode de prévision selon les six métriques sélectionnées pour l'ensemble de la période du 01/08/2014 au 31/12/2015. Pour les métriques classiques de RMSE, MAE et Biais, la méthode de persistance est légèrement meilleure que la simple méthode CMV, qui à son tour fonctionne mieux que la prévision ECMWF. Le TDI donne des informations sur la distorsion temporelle globale entre les séries temporelles prévues et les mesures. Cette distorsion peut être provoquée par différents comportements de prévision, un décalage temporel ou une approximation qui est, en moyenne, alignée temporellement (*i.e.* une faible valeur du TDM). Selon les scores représentés sur la figure 27, la persistance est à 100% en retard, ce qui est cohérent avec sa définition; la méthode CMV donne une prévision alignée temporellement et la prévision ECMWF est en retard de 36%. La métrique ramp-MAD décrit la capacité de la prévision à suivre les pentes des rampes significatives. Comme mentionné ci-dessus, nous avons utilisé  $\tau_{CLS}$  comme un paramètre de sensibilité de l'algorithme *Swinging Door*. Sur la figure 27.a, la méthode de persistance a un meilleur ramp-MAD que la méthode CMV et la prévision ECMWF, similairement au cas présenté à la figure 18. La figure 27.b affiche les résultats de la série de prévisions alignées selon le calcul TDI choisi. La persistance alignée donne de très bons résultats, mais pas parfaits, puisque la persistance est appliquée à l'indice de ciel clair  $k_c$  et non au GHI. La correction temporelle est satisfaisante puisque le TDI de la persistance alignée est proche de zéro. Notons aussi que l'alignement temporel diminue significativement le TDM. La figure 27 montre que cette correction de la distorsion temporelle peut entraîner une nette amélioration des performances de persistance et une légère amélioration du ECMWF et de la méthode CMV.

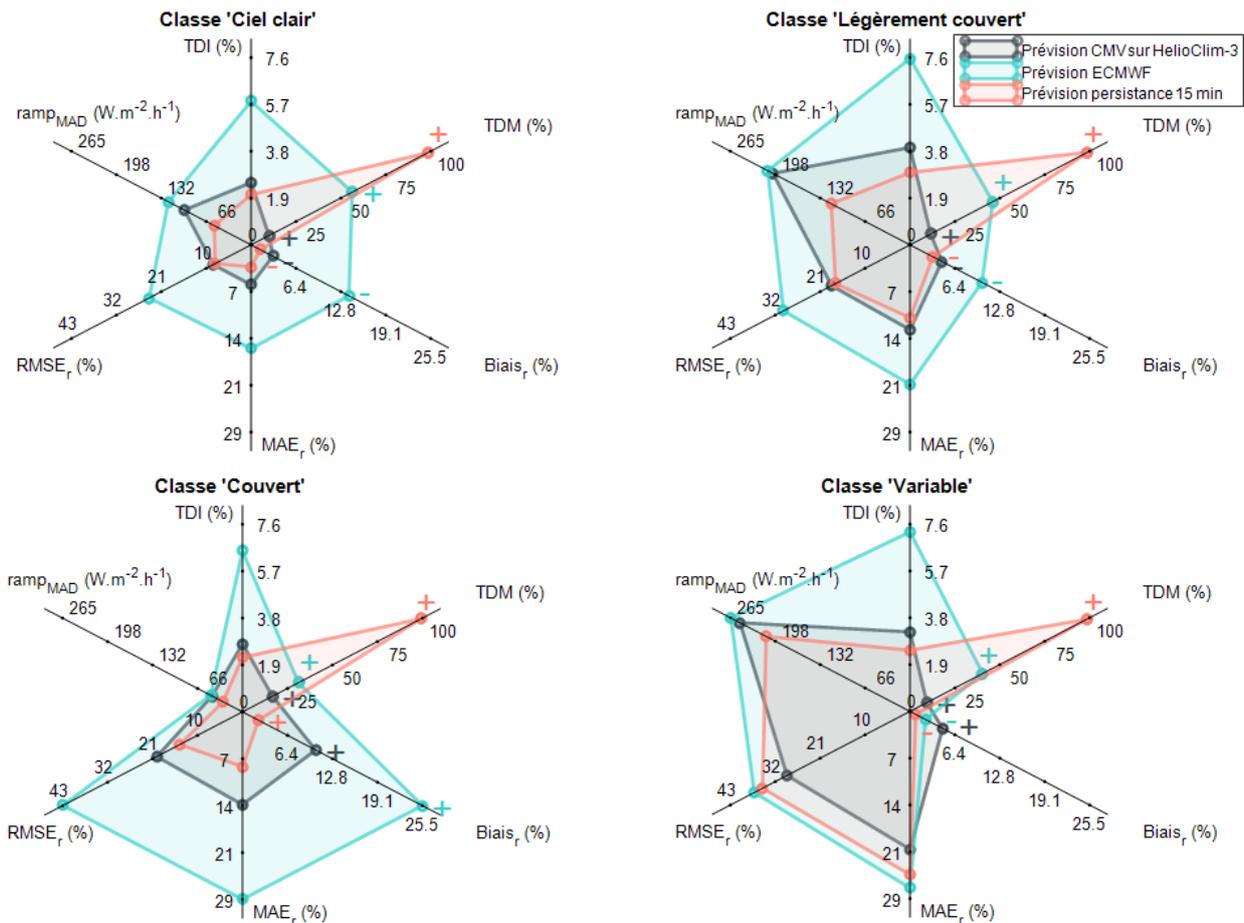


Figure 28 – Performance de chaque classe à partir de la classification des instants pour la prévision par persistance 15 min, de la méthode CMV basée sur les images Helioclim et de la prévision interpolée ECMWF de 15 min.

La figure 28 affiche les scores des métriques sélectionnées pour chaque classe de la classification des instants (*cf.* paragraphe II.6.1). Afin d'obtenir les classes correspondantes à un pas de temps de 15 min, nous prenons l'étiquette la plus commune des 15 précédents instants classifiés au pas de temps 1 min en calculant le premier mode (c'est-à-dire la valeur la plus fréquente). Le calcul des trois métriques supérieures du graphe radar est adapté afin d'attribuer des valeurs à une série d'instants classifiés. Nous ne prenons en compte que les instants consécutifs car le TDI et le ramp-MAD sont intrinsèquement cumulatifs dans leurs définitions. Pour autoriser la comparaison des scores entre les classes, toute la période d'étude est utilisée pour le calcul des scores relatifs. Sans surprise, ECMWF donne les moins bons résultats des trois méthodes de prévision utilisées à des fins d'illustration.

Cette classification nous aide à comprendre dans quelles circonstances chaque méthode fonctionne le mieux. Par exemple, la prévision ECMWF manque d'une description précise de la couverture nuageuse et produit donc de très mauvais résultats pour les instants classifiés comme "couvert". La classe *clear-sky* affiche un TDM non nul, qui est dû aux petites variations encore présentes dans la sortie de l'étape de classification et aux inexactitudes potentielles dans le modèle de ciel clair. En effet, le TDI et le TDM sont sensibles

à une mauvaise estimation de la courbe de ciel clair. Cependant, selon la définition du TDM, son analyse devrait être faite en conjonction avec celle du TDI. De plus, l'algorithme qui calcule le TDI tend à associer à tort certains instants proches lors d'une faible variabilité autour de la courbe ciel clair, ce qui explique sa légère augmentation pour les instants légèrement nuageux. Notons enfin que le meilleur score de TDI est atteint pour les instants ciel clair et couvert lorsque la variabilité temporelle est plus faible.

La façon dont le ramp-MAD est calculé dépend de l'amplitude des rampes considérées et de la résolution temporelle de la prévision. Les résultats sont également susceptibles de varier pour des horizons de prévision plus élevés. Comme le montre la figure 28, quelle que soit la classe considérée, la méthode de persistance suit mieux les rampes que les méthodes CMV et ECMWF. La valeur de ramp-MAD est la plus faible pour les instants de ciel couvert (ce qui est cohérent car ils ont des pentes très faibles) et est légèrement plus élevée pour les ciels clairs (correspondant aux pentes de la courbe du ciel dégagé). Les instants variables et légèrement couverts ont des valeurs de rampe assez proches puisque ces instants sont soumis aux mêmes ordres de grandeur de variabilité. Une classification des instants plus précise conduirait certainement à des valeurs de ramp-MAD plus petites (*i.e.* meilleures) pour la classe légèrement couvert. De la même manière, une classe de ciel clair plus restreinte pourrait conduire à une meilleure description de la performance sous un ciel dégagé.

La portée de cette étude est limitée à la variabilité infrajournalière du rayonnement solaire, mais d'autres classifications de résultats par jour, mois, saison ou année pourraient être considérées avec la même approche. Cette étape de classification est essentielle pour mieux comprendre les sources d'erreurs de prévision.

## II.8 Conclusions du chapitre

En conclusion, ce chapitre a eu pour objectif d'étudier la manière dont est réalisée usuellement l'analyse des performances dans le domaine de la prévision solaire.

Dans le cas des prévisions locales, la mesure pyranométrique s'est imposée comme référence dans la comparaison *a posteriori* des performances de méthodes de prévision. Après un bref descriptif des moyens de mesures du rayonnement solaire existant, une analyse de la qualité de nos mesures a permis de s'assurer de leur plausibilité vis-à-vis des quantités physiques étudiées.

Certaines limitations des métriques usuelles ont été illustrées sur des exemples concrets de journées, avec notamment la particularité de la RMSE ayant tendance à privilégier des prévisions contenant très peu de variabilité hautes fréquences. Les utilisateurs des prévisions à très court terme semblent avoir cependant un intérêt pour ce contenu hautes fréquences, et plus particulièrement sur la notion de restitution des rampes principales de la mesure au bon moment.

En réponse à ce constat, deux nouveaux outils ont été proposés dans ce chapitre dans l'objectif de compléter l'information délivrée par les métriques les plus utilisées. Une métrique nommée ramp-MAD tente de quantifier la capacité d'une méthode de prévision à restituer les rampes de la mesure, dont le niveau d'approximation est contrôlé par un unique paramètre. La métrique TDM s'appuie sur la notion de distorsion temporelle proposée antérieurement par Frías-Paredes et al. (2016) pour quantifier la proportion de cette distorsion en avance ou en retard par rapport à la mesure.

Ensuite, afin de compléter cette analyse des performances, nous soutenons l'importance d'y adosser une caractérisation catégorielle de ces performances. Qu'il s'agisse d'une classification d'instant, de journées ou de tout autre support temporel ou d'autres variables, ces techniques affinent l'analyse, permettent de faciliter la compréhension de certains résultats globaux et vont vers une plus grande indépendance des données de l'étude.

L'analyse des performances suggérée ici a été appliquée sur des méthodes de prévision dont les comportements sont bien connus des experts de la prévision solaire, et leurs caractéristiques sont en phase avec les résultats des nouvelles métriques, sur les horizons et pas de temps considérés, au lieu étudié. Elle n'a donc pas vocation à être exhaustive mais plutôt démonstrative. Les deux nouvelles métriques ont été publiées par Vallance et al. (2017). Un effort collaboratif avec Dazhi Yang de l'institut SIMTech de Singapour a été entamé dans l'optique d'une dissémination large des codes de ces métriques en Open Source dans plusieurs langages, et ainsi faciliter l'utilisation et l'appropriation de ces métriques par la communauté scientifique. Ces métriques font aussi l'objet d'une étude menée par le DLR où elles sont utilisées et analysées sur différentes méthodes de prévision sous différents horizons et lieux d'étude.

Par l'introduction de ces nouvelles métriques, l'accent est mis sur la qualité de la restitution et de la prévision des rampes tant sur leurs amplitudes que sur leurs instants d'occurrence. Afin de développer de nouvelles approches de prévision dont l'objectif est d'apporter des améliorations au regard de ces nouvelles métriques, nous avons envisagé d'exploiter non pas une seule source d'informations (mesures pyranométriques, images hémisphériques *in situ* ou images satellites) mais la synergie de plusieurs d'entre elles.

Le chapitre qui suit identifiera les complémentarités observables qui sont attendues lors d'une utilisation conjointe de ces trois sources d'informations à disposition.

---

# Chapitre III - Observation des complémentarités des sources d'information

---

En l'état actuel de la recherche dans le domaine, l'élaboration de méthodes de prévision de rayonnement ou de production PV à très court terme se base principalement sur trois sources d'information : des images issues de satellites météorologiques géostationnaires, des images issues de caméras hémisphériques scrutant le ciel avec un angle solide de  $2\pi$  sr sur les sites à étudier et des mesures pyranométriques *in situ* représentant en temps réel l'évolution du rayonnement incident en un point. Chaque méthode de prévision a des résultats issus de l'analyse des performances qui lui sont propres, dont certaines méthodes ont été présentées et discutées au chapitre II. Certaines caractéristiques de ces méthodes de prévision sont intrinsèquement liées aux données qu'elles utilisent. Ce chapitre est dédié à la description des caractéristiques propres à chaque source d'information, ainsi qu'à leurs aspects complémentaires suggérant la possibilité d'exploiter ces synergies dans un contexte prévisionnel. Les trois sources d'information à l'étude sont des observations de la dynamique de la même scène nuageuse en présence. Chacune apporte une partie de l'information sous une forme et avec une perception différentes de la même réalité. En particulier, l'accent sera mis sur les différences et complémentarités entre les deux sources d'images (caméra hémisphérique et satellite météorologique géostationnaire) se dessinant par comparaison des deux.

## III.1 Complexité de la scène 3D et de ses observations

L'image satellite et l'image hémisphérique du ciel constituent deux acquisitions quasi instantanées d'une même scène en trois dimensions. La complexité des scènes nuageuses (formes tridimensionnelles complexes, frontières floues et fractales), qui peuvent survenir ne peut pas être décrite parfaitement par une unique observation. Ce paragraphe vise à énumérer les caractéristiques de la géométrie en présence.

### III.1.1 Une vision partielle des nuages pour chaque point d'observation

L'une des différences fondamentales entre les deux types d'images est leur différence de points de vue induisant des phénomènes d'occlusion inévitables du champ de vision. La figure 29 illustre cette différence à l'aide d'une situation artificiellement générée depuis le logiciel Blender©. La figure 29.a affiche la scène qui comporte deux nuages de tailles différentes ainsi qu'une sphère rouge et une sphère bleue. La figure 29.b représente le point de vue du satellite, avec toutefois une meilleure résolution spatiale pour l'illustration. Sur cette image simulée, les deux nuages sont visibles mais apparaissent collés et seule la

sphère rouge est visible. La sphère bleue est occultée par le nuage 1. La figure 29.c représente ce qui serait alors perçu au même instant sur l'image hémisphérique, la caméra hémisphérique étant représentée par une demi-sphère sur les images (a) et (b). Outre la distorsion optique de l'image, il apparaît de ce point de vue que les deux nuages sont effectivement distincts et seule la sphère bleue est visible, la rouge étant quant à elle occultée par le nuage 1. Un autre aspect intéressant et visible sur ces images est la difficulté d'estimer la taille d'un objet avec une seule image en l'absence d'information sur son altitude : l'objet en gris sur l'image hémisphérique paraît plus gros que la sphère bleue, alors qu'il est en réalité cinq fois plus petit, mais nettement plus proche de la caméra.

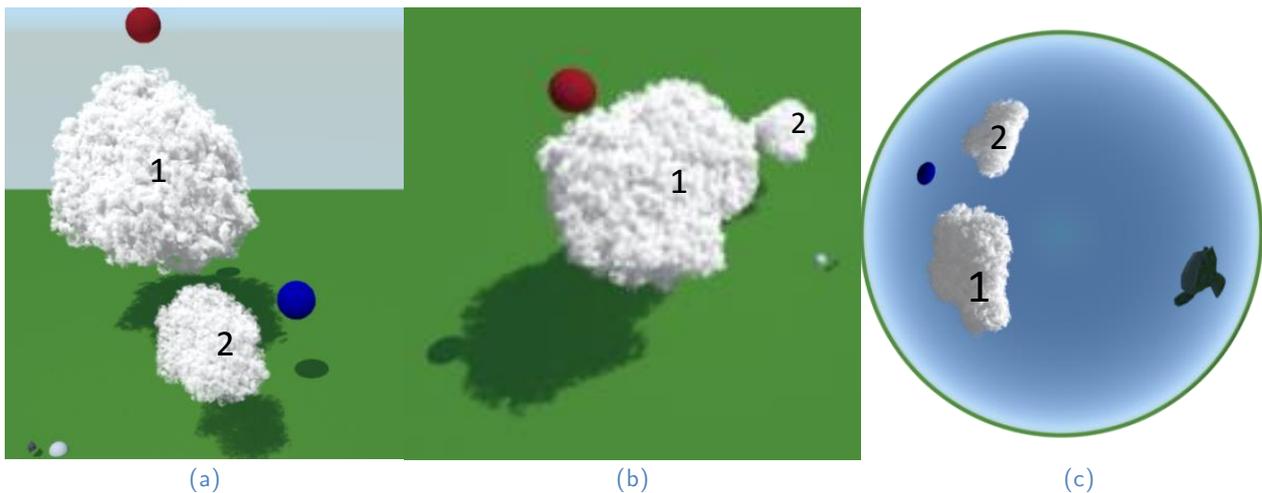


Figure 29 – Scène nuageuse artificielle (a) illustrant les différences de perception d'une même scène entre le point de vue satellitaire (b) et le point de vue de la caméra hémisphérique au sol (c). La demi-sphère au sol sur les images (a) et (b) représente la caméra hémisphérique.

### III.1.2 Taille de la couverture nuageuse perçue par l'image satellite

Les satellites météorologiques géostationnaires constituent une famille de satellites qui restent par définition au-dessus du même point géographique en suivant la rotation de la Terre. Une grande variété de capteurs est embarquée dans ces satellites, dont des capteurs optiques à grande focale autorisant des observations régulières de l'atmosphère et de la Terre sur le long terme et sur une large partie du globe. La figure 30 proposée par Blanc et al. (2017) décrit la couverture de la surface terrestre observée par chaque satellite des familles de satellites METEOSAT, GOES, FENG YUN et HIMAWARI. Ainsi, la couverture typique des images issues de ces satellites est le tiers de la surface terrestre. À l'exception des régions polaires, ces familles de satellites observent la surface terrestre avec une résolution spatiale kilométrique et une résolution temporelle intra-horaire.

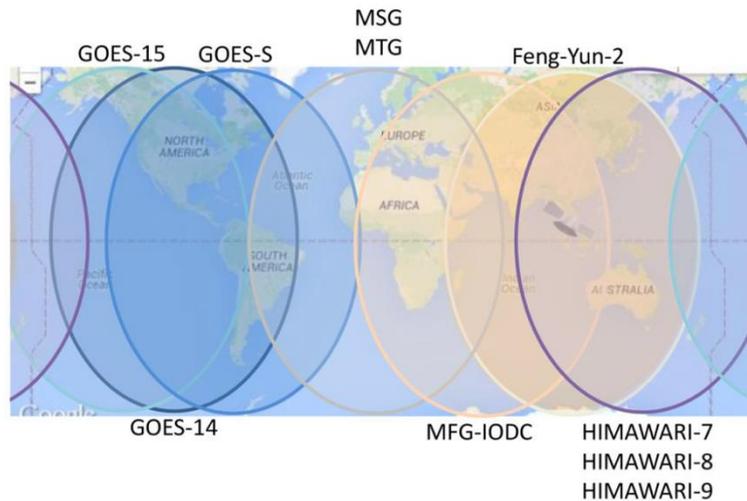


Figure 30 – Champ de vision des principales familles de satellites météorologiques géostationnaires (GOES, Meteosat première, deuxième et troisième génération, Feng-Yun et Himawari) pour une couverture mondiale. Source : Blanc et al. (2017) à partir des informations délivrées par OMM OSCAR<sup>6</sup>.

Dans cette étude, plusieurs produits issus directement ou indirectement du satellite Météosat Seconde Génération (MSG) ont été utilisés. Les deux canaux du visible utilisés dans la méthode Heliosat-2 ainsi que le canal infrarouge exploité par le KNMI pour extraire l'information d'altitude du sommet des nuages (CTH, *Cloud Top Height*) recouvrent le même disque complet du globe, tel qu'illustré à la figure 31. Cependant, les limites de débit de données confinent les images visibles à haute résolution (HRV, *high-resolution visible*) à la moitié de la Terre dans une direction Est-Ouest. Les images HRV ont une résolution spatiale au nadir de 1 km contre 3 km pour les autres canaux. La couverture exacte de la Terre est programmable, sur la figure 31 elle recouvre la moitié du disque avec deux parties distinctes, l'une centrée sur l'Europe et l'autre centrée sur l'Afrique de l'Est.

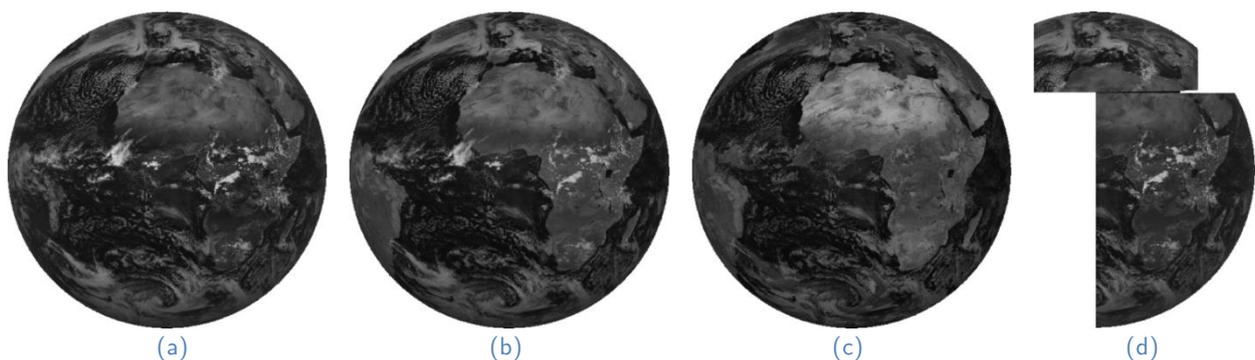


Figure 31 – Couverture MSG pour le canal 1 visible centré sur  $0.635 \mu\text{m}$  (a) et le canal 2 visible centré sur  $0.81 \mu\text{m}$  (b) utilisés dans la méthode Heliosat-2, la couverture MSG pour le canal 3 infrarouge centré sur  $1.64 \mu\text{m}$  (c) utilisé pour générer les images CTH du KNMI, et le canal 12 haute résolution visible (HRV) à large bande spectrale (de  $0.4$  à  $1.1 \mu\text{m}$ ). Source : Eumetsat<sup>7</sup>.

<sup>6</sup> OSCAR: Observing Systems Capability Analysis and Review tool de l'Organisation Météorologique Mondiale (<http://www.wmo-sat.info/oscar/>)

<sup>7</sup> EUMETSAT : European Organisation for the Exploitation of Meteorological Satellites (<https://www.eumetsat.int/>)

Ainsi, la large couverture géographique des images satellites météorologiques constitue un avantage considérable pour les estimations et les prévisions solaires sur les gammes d'horizons allant de 15 min à quelques heures. Cependant, en fonction de la position géographique du site sur lequel il est requis d'avoir une prévision ou une estimation du GHI, certains décalages spatiaux peuvent intervenir, ce sera l'objet du paragraphe III.1.5. Aussi, les déficits de résolutions spatiales et temporelles sont également source d'erreur, des observations de ces spécificités sont proposées dans la suite du chapitre. Dans toute la suite, l'expression images satellites fera référence aux images HelioClim, CTH et/ou HRV.

### III.1.3 Image hémisphérique : la distorsion optique au service d'une vision périphérique du ciel

La géométrie de la scène nuageuse capturée sur une image hémisphérique est particulière. La distorsion optique de l'objectif hémisphérique de très grand angle permet une acquisition de la totalité du ciel au-dessus de site d'étude sur une seule image. Ainsi, l'angle d'élévation (i.e. angle par rapport à l'horizon) correspondant à chaque pixel de l'image est illustré par les lignes de niveau en surimpression de la figure 32.a, le centre de l'image correspondant à un angle d'élévation de  $90^\circ$  (zénith). La figure 32.b, quant à elle, représente un exemple de positions du Soleil à chaque heure pleine pour l'année 2015 sur le site n°1. Des boucles à temps universel constant sont affichées, le haut de ces boucles représente les positions du Soleil pour les jours d'été, et le bas pour les jours d'hiver.

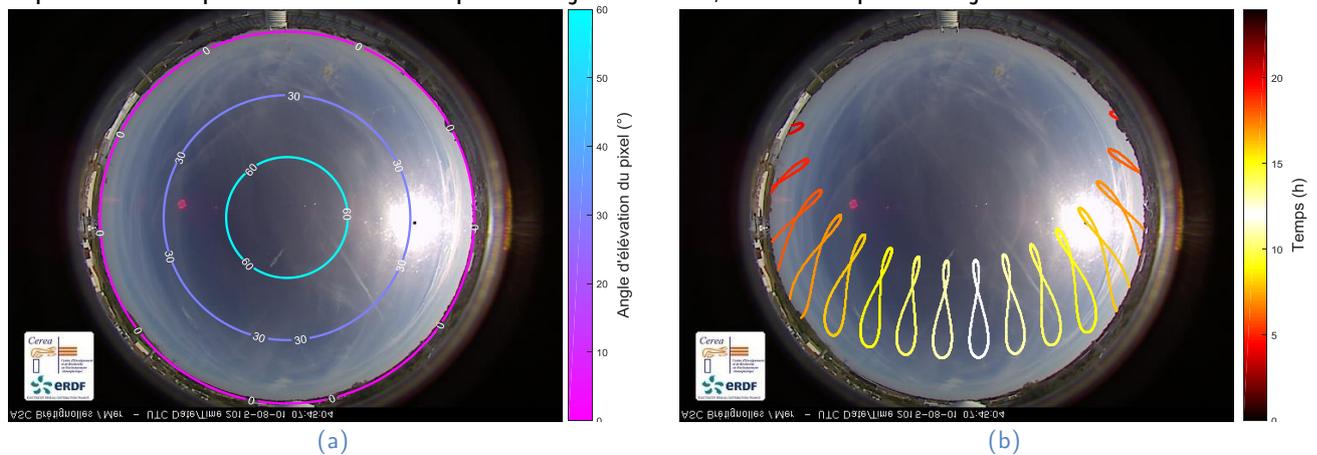


Figure 32 – Exemple d'image acquise sur le site n°1 à 8 h UT le 01/08/2015, avec en superposition les lignes de niveau de l'angle d'élévation de chaque pixel (a) et la position du Soleil en fonction de l'heure de la journée pour toute l'année 2015 (b).

La particularité des prévisions solaires reposant sur ce type d'images est qu'elles se limitent à des horizons de prévisions très courts (Kazantzidis et al., 2017). En effet, les nuages qui vont passer au-dessus du site dans quelques heures se trouvent, en général, au-delà de l'horizon de la caméra. L'étendue de la couverture nuageuse visible depuis le sol dépend de l'altitude des nuages. La figure 33.a représente la zone de l'image hémisphérique pour laquelle l'angle d'élévation (angle entre le plan horizontal et le point observé) est supérieur à  $15^\circ$ . La figure 33.b délimite la zone correspondante sur l'image satellite en fonction de l'altitude considérée.

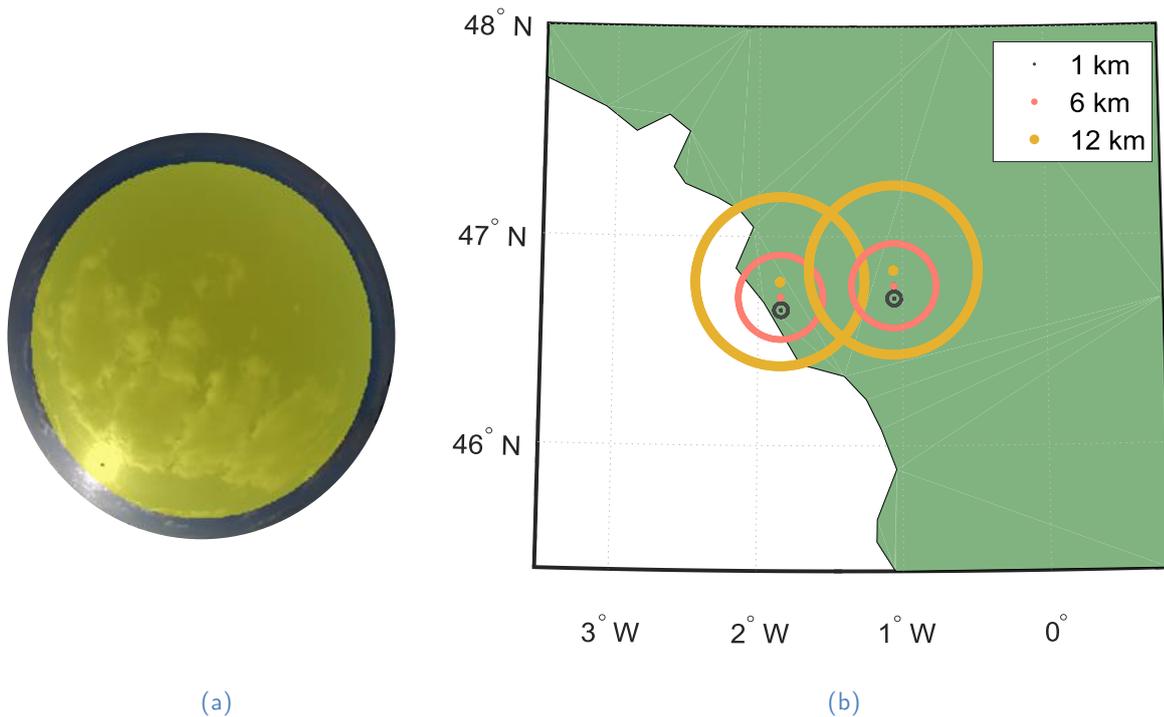


Figure 33 – Zone de l'image hémisphérique (en jaune) dont l'angle d'élévation est supérieur à  $15^\circ$  (a) et la zone de visée correspondante du point de vue du satellite (b) pour trois altitudes de nuages : 1 km, 6 km et 12 km.

Les centres des cercles correspondants aux nuages pour différentes altitudes sur la figure 33.b ne correspondent pas à la localisation des sites. La raison de ces décentrages provient de l'effet parallaxe qui existe entre le point de vue de la caméra et le point de vue satellitaire au point ( $0^\circ\text{N}$ ,  $0^\circ\text{E}$ , 36000 km). Une explication de cet effet de parallaxe est proposée au paragraphe III.1.5.

Les intersections des cercles de même altitude des deux sites suggèreraient une utilisation conjointe des images hémisphériques des deux sites. Cependant, les nuages d'altitude supérieure à 6 km – souvent des cirrus – ne sont pas facilement identifiables sur les deux types d'images, en particulier pour des angles d'élévation faibles du point de vue du sol. De plus, la majorité des nuages épais se trouve en dessous de 7 km d'altitude d'après la classification de l'Organisation Météorologique Mondiale (WMO), donc il est difficilement envisageable de recouper les informations des deux sites de manière directe.

L'une des difficultés qui transparait déjà dans ce cas simple est la nécessité d'avoir une estimation fiable de l'altitude des nuages pour être à même de compenser ces phénomènes de parallaxe rendant difficile la mise en correspondance de la perception des nuages pour les deux sources d'observation.

### III.1.4 Perception d'une grille horizontale par les deux types d'images en fonction de son altitude imposée

L'illustration proposée à la figure 34 décrit une partie de la complexité liée à la géométrie 3D observée de deux points de vue et liée à la géométrie hémisphérique de la caméra. Dans cette situation complètement artificielle, nous réduisons la diversité des configurations nuageuses possibles à une couche nuageuse d'épaisseur infinitésimale et de taille 40 km par 40 km située à la verticale du site n°1. Cette couche nuageuse est matérialisée par une grille suivant un pas de 5 km de côté de manière à visualiser les déformations de la grille dans les deux projections : satellitaire et hémisphérique. Sur la première colonne de la figure 34, l'altitude de la grille artificielle est imposée à 1 km. Dans ce cas, du point de vue satellitaire (deuxième ligne) le site se trouve à peu près au centre de la grille : les effets de parallaxe sont plus limités pour des altitudes de 1 km, compte tenu de la distance angulaire entre le site d'étude et la position de MSG. Ainsi pour un nuage d'altitude 1 km, la présence d'un nuage au-dessus du site peut être estimée à l'aide de l'image satellite en prenant directement le pixel correspondant aux coordonnées du site. Pour ce qui est de la perception de cette grille depuis le point de vue du sol au site n°1 (dernière ligne), les quatre dalles centrales (i.e. environ 10 km de diamètre) représentent la majeure partie de ce qui est visible sur l'image.

Lorsque la grille s'éloigne du sol (colonnes de gauche à droite sur la figure 34), un effet de glissement de la grille vers le Nord est observé sur les images satellitaires – ce phénomène est l'effet parallaxe et sera discuté au paragraphe III.1.5 –, et la perception de la grille du point de vue de la caméra est de moins en moins distordue à mesure que l'altitude de la grille augmente. Cette évolution de la taille perçue par les pixels de l'image hémisphérique sera traitée au paragraphe 0.

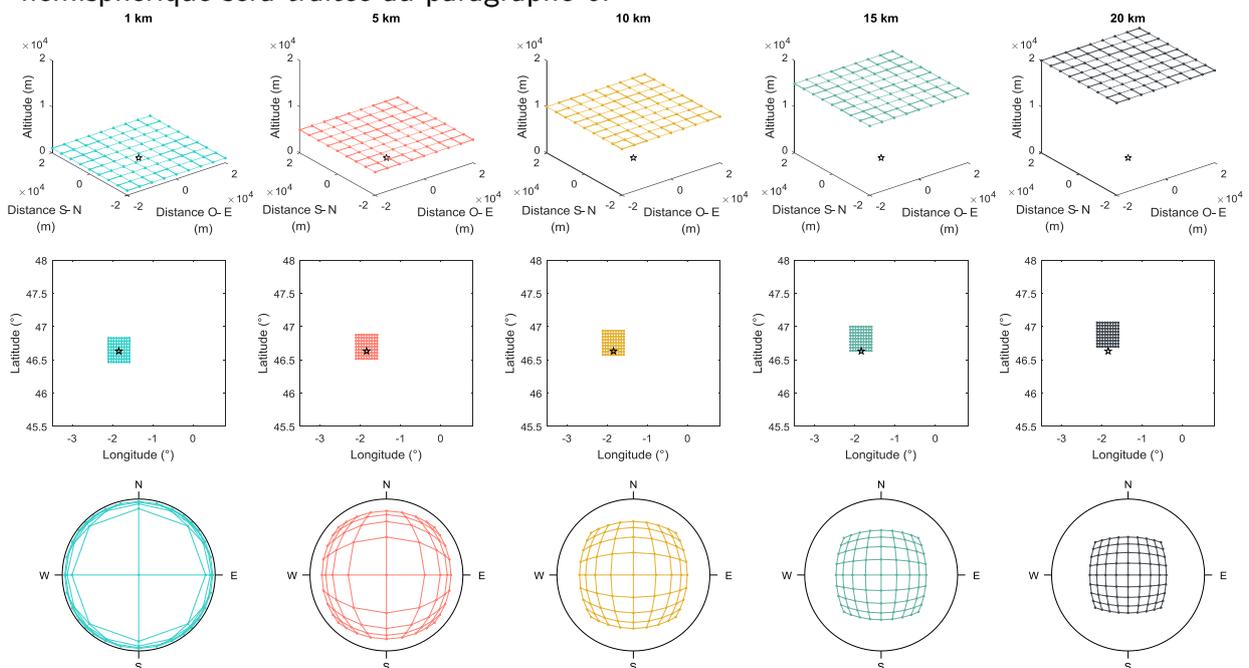


Figure 34 – Illustration des différences de perception d'une même grille de taille 40 km x 40 km et de pas 5 km positionnée à différentes altitudes (première ligne, dans le repère local Est-Nord-Zénith) du point de vue satellitaire (deuxième ligne, dans le repère latitude-longitude) et du point de vue du sol (dernière ligne, dans le repère de la caméra hémisphérique). Seul le site n°1 est représenté sur cette image.

L'objectif de cette illustration est de pointer l'importance qu'occupe l'altitude des nuages dans la problématique qui nous intéresse, au-delà des problèmes d'occlusion. Plusieurs aspects ne sont pas considérés dans la représentation précédente. Par exemple, l'épaisseur des nuages et leur géométrie en trois dimensions compliquent la perception de la scène par seulement deux points de vue.

### III.1.5 Effet parallaxe

De manière générale, la parallaxe est l'effet du changement de position d'un observateur sur la perception de l'objet éloigné observé. Ici, l'effet parallaxe intervient à la fois sur la différence de point de vue du satellite et de la direction d'éclairage du Soleil, et à la fois entre les points de vue de la caméra et de celle du satellite. Ces effets sont illustrés sur les figures 35 et 36 en prenant l'exemple de la position angulaire du Soleil le 23 juin 2015 à midi (UT) où l'effet a beaucoup d'influence puisque proche du solstice d'été. La figure 35 indique que, du fait de cet effet parallaxe, les nuages de l'hémisphère Nord (resp. Sud) sont perçus par le satellite décalés vers au Nord (resp. Sud). Un site peut ainsi être considéré comme couvert par des nuages depuis le satellite alors qu'il est en fait ensoleillé, et inversement.

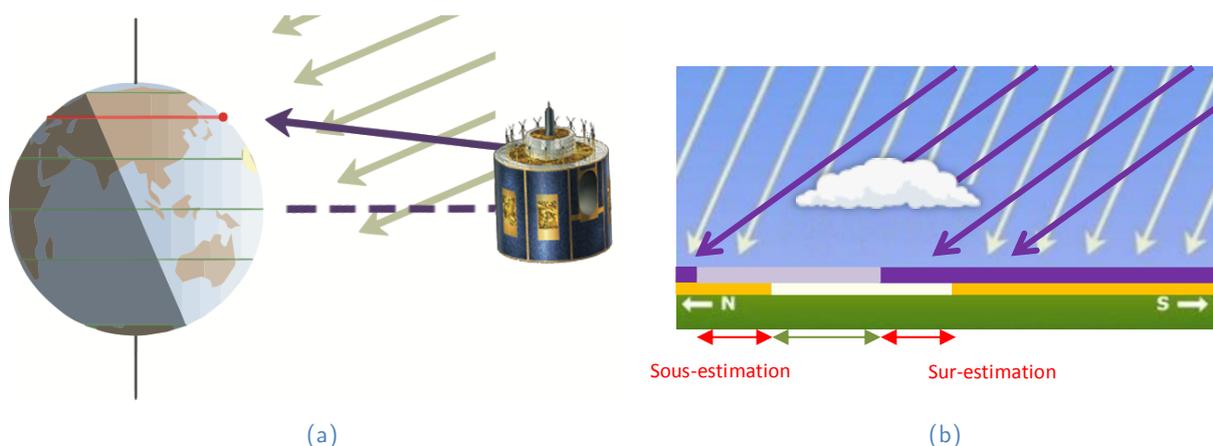


Figure 35 – Explication du phénomène de parallaxe : vue schématisée dans le référentiel géocentrique (a) et terrestre (b). Dans le référentiel terrestre, l'angle des rayons de visées du satellite MSG (en violet) fait qu'ils atteignent le sol avec un angle d'élévation de  $36.4^\circ$ . La situation proposée correspond à celle du solstice d'été (càd. le 20 ou 21 juin). Schéma inspiré des animations de <http://astro.unl.edu>.

Cet effet est observable sur la figure 36. La représentation de l'ensemble des informations pour un instant donné telle que proposée sur cette figure sera utilisée de façon récurrente dans la partie III.4. En haut à droite se trouve l'image hémisphérique orientée de la même manière que les images satellites, l'instant de l'image hémisphérique étant 4 min en avance par rapport à l'instant d'acquisition de l'image satellite afin d'assurer une correspondance temporelle des observations (se référer à III.3.1 pour plus de détails). À droite, une projection de cette image dans un plan horizontal arbitrairement imposé à 2 km est proposée à des fins illustratives sur une grille de  $500 \times 500$  px. En pointillés, les cercles concentriques de l'image hémisphérique d'origine correspondent aux angles d'élévation de  $30^\circ$  et  $60^\circ$ . Ils sont représentés sur l'image projetée. En haut à droite,

l'image HRV de résolution 1 km au nadir codée sur 8 bits. Pour la suite, deux marqueurs différents représenteront les sites à l'étude sur les images satellites. Le symbole ☆ décrira la position du site n°1, et le symbole □ celle du site n°2. En bas à droite est représentée l'image HelioClim correspondante de résolution 3 km au nadir, le plus souvent exprimée en indice de ciel clair  $k_c$  de manière à s'affranchir des évolutions du rayonnement par ciel clair. Enfin, en bas à gauche l'évolution quotidienne du rayonnement mesuré par le pyranomètre du site est proposée. Les courbes théoriques de "ciel clair" du GHI et du DHI, respectivement en jaune et en violet, sont issues du modèle ESRA (cf. Chapitre I). Les évolutions du GHI en bleu et du DHI en orange sont affichées au pas de temps 1 min, et l'instant d'intérêt (correspondant à l'acquisition de l'image hémisphérique du dessus) est matérialisé par une barre verticale rouge.

Sur cette figure 36, à 7h11 le centre de l'image hémisphérique indique la présence d'un nuage à la verticale du site. À ce même instant, la mesure pyranométrique est typiquement celle d'un instant couvert, puisqu'elle vaut moins de la moitié de la valeur théorique par ciel clair et qu'elle n'admet qu'une très faible variabilité temporelle locale. Cependant, qu'il s'agisse de l'image HelioClim ou de l'image HRV, l'information satellitaire ne perçoit pas le site comme étant couvert par des nuages. Ces effets parallaxes dépendent de l'altitude, de l'épaisseur et du nombre de couches de nuages présentes.

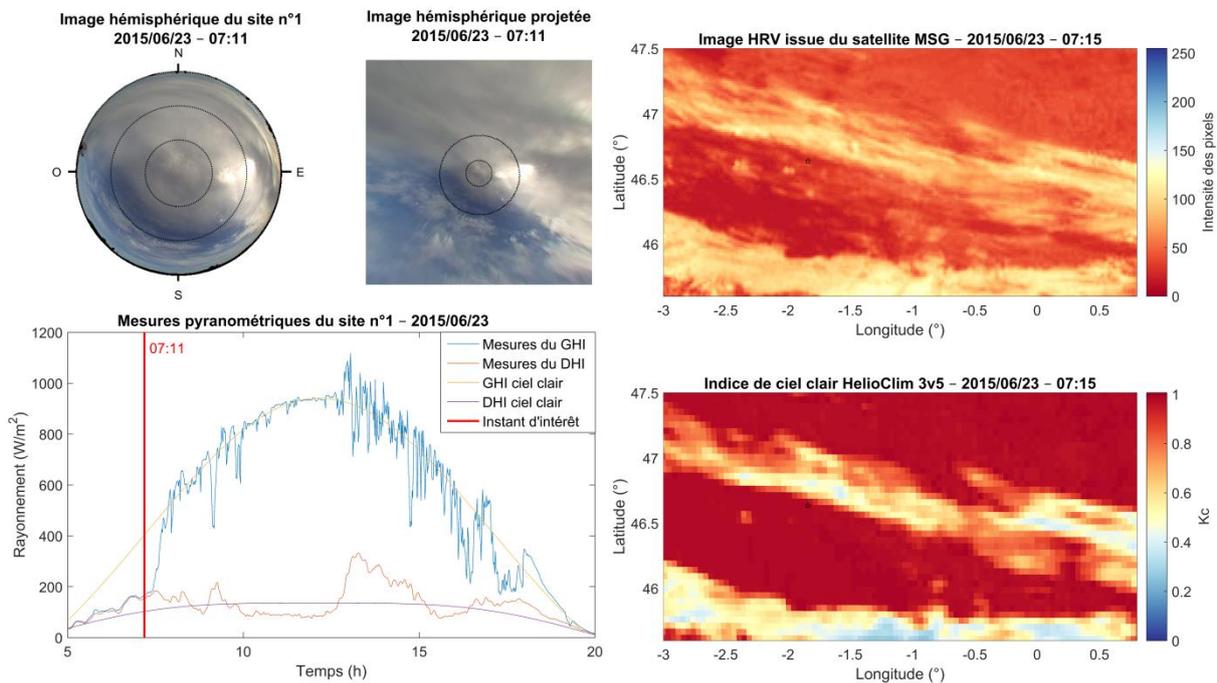


Figure 36 – Exemple concret d'observation de l'effet parallaxe.

La figure 37 représente l'altitude du sommet des nuages estimée par les images CTH du KNMI. L'altitude du sommet du nuage qui se trouve légèrement au Nord du site à 7h11 est d'environ 4 km. Ainsi du point de vue du satellite, le décalage spatial du nuage est perçu vers le Nord. Sachant que l'angle d'élévation du satellite MSG dans le référentiel du site n°1 est de  $36.4^\circ$ , il est possible d'estimer grossièrement ce décalage, équiva-

lent au produit de l'altitude estimée du nuage et de la tangente de cet angle d'élévation de MSG. Dans ce cas, la parallaxe entre le point de vue du sol et le point de vue du satellite est de l'ordre de 3 km. Cependant, l'évolution journalière de la position du Soleil va générer un déplacement de l'ombre portée du nuage dans le sens inverse de la course du Soleil (c'est-à-dire d'Ouest en Est). Ce phénomène est lui aussi induit par l'effet de parallaxe, mais il intervient cette fois-ci à cause de la différence de point de vue entre le satellite et le Soleil.

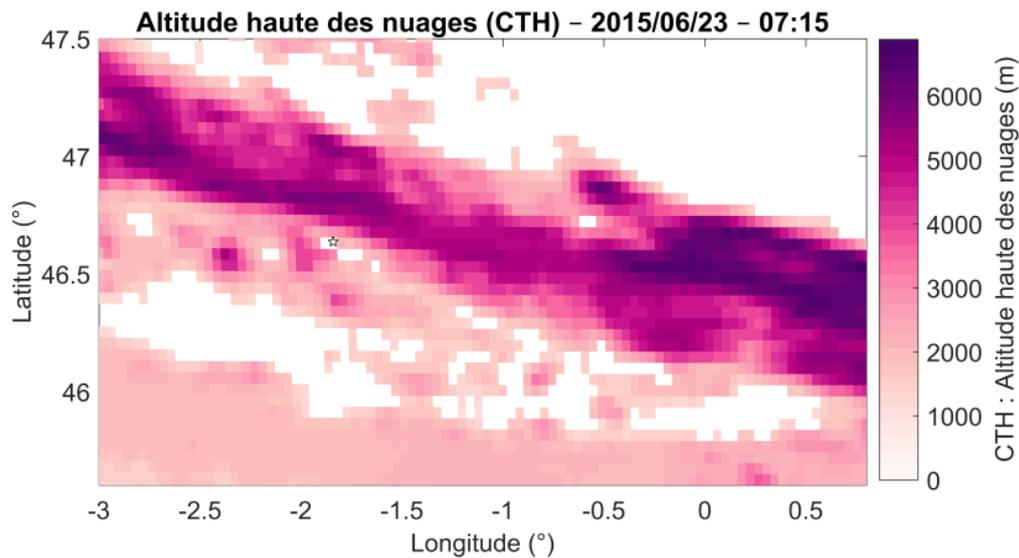


Figure 37 – Altitude du sommet des nuages (CTH) provenant du KNMI correspondant à la situation illustrée sur la figure 36.

## III.2 Différence de résolution spatiale entre images hémisphériques et images satellites

### III.2.1 Résolution spatiale et taille du pixel

De manière générale, la résolution définit la faculté du système d'acquisition d'images à distinguer les plus petits détails de la scène observée. Dans cette étude, l'expression résolution spatiale sera utilisée en référence à la taille du pixel uniquement, de façon à ce que plus la taille du pixel est petite, plus la résolution spatiale est grande. Cette notion est cependant plus complexe. Selon la définition du Larousse elle est "*la propriété d'un système d'observation à permettre la vision distincte de deux points rapprochés*". Selon cette définition, la résolution spatiale de capteurs optiques dépend non seulement de la taille du pixel, mais aussi de la tâche image et du bruit de mesure. Nous ignorerons à ce stade ces deux paramètres dans notre analyse puisque le noyau de convolution des capteurs embarqués dans les satellites météorologiques est généralement de rayon très faible.

### III.2.2 Taille du pixel satellite

L'acquisition des images satellites MSG s'opère par révolution du satellite – pour le scan des lignes d'Est en Ouest – et rotation du miroir interne – pour le passage à la ligne supérieure. Ces deux rotations effectuent un balayage angulaire régulier selon les deux axes. La rotondité de la Terre a une influence importante sur la résolution spatiale de l'image acquise, qui augmente du centre (nadir) à la bordure de l'image tel qu'illustré sur la figure 38.

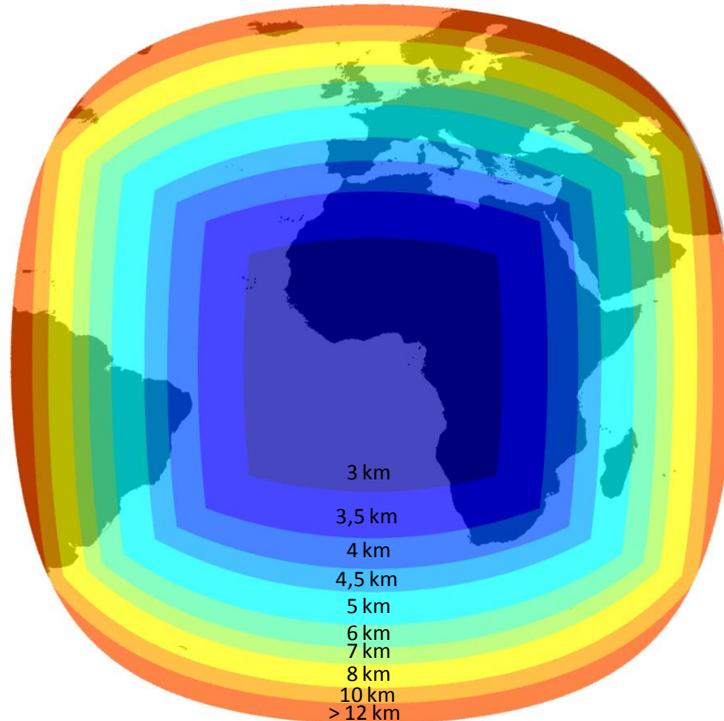


Figure 38 – Évolution de la résolution spatiale des images satellites MSG dont la résolution spatiale au nadir est de 3 km (tous les canaux excepté le canal 12 visible à haute résolution HRV), [Source](#) : Mines ParisTech.

L'angle de balayage de l'instrument SEVIRI (*Spinning Enhanced Visible and InfraRed Imager*) embarqué dans MSG est de  $18^\circ$ , et la longueur focale du système optique est de 5367 mm. Ainsi, l'altitude des nuages observés n'a que très peu d'influence sur la résolution spatiale de ceux-ci. Pour un nuage au nadir d'altitude 20 km, l'écart de taille du pixel central comparé à ce même pixel en l'absence de nuages est de l'ordre du mètre. Cette propriété des images satellites ne se vérifie pas pour les caméras hémisphériques puisqu'ils ont des focales nettement plus petites et sont, proportionnellement, plus proches des nuages.

Les huit canaux IR thermiques et les trois canaux solaires ont un pas d'échantillonnage de 3 km au nadir. Cependant, le canal visible à haute résolution (HRV) fournit des images avec un échantillonnage de 1 km au nadir. Ainsi, l'évolution de la résolution spatiale pour les images HRV est semblable à celle de la figure 38 mais avec le tiers des valeurs présentées.

### III.2.3 Taille du pixel de l'image hémisphérique

La taille du pixel de l'image hémisphérique dépend principalement de l'altitude de la base du nuage observé (altitude de la base puisqu'observés depuis le dessous) et de la position du pixel dans l'image (dépendance vis-à-vis de l'angle entre le pixel et le zénith, appelé angle zénithal du pixel – PZA, *pixel zenithal angle*). La figure 39 décrit l'évolution de la longueur des pixels pour trois altitudes données (1 km, 5 km et 10 km) selon la position du pixel dans l'image. La résolution des pixels est proportionnelle à l'altitude du nuage observé et dépend de l'angle zénithal de chaque pixel.

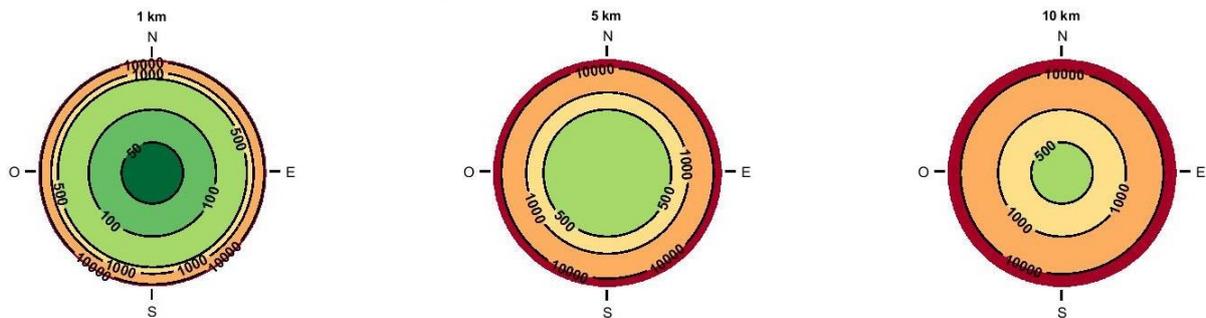


Figure 39 – Lignes de niveaux des longueurs des pixels (en m) de l'image hémisphérique en fonction de l'altitude de la base du nuage considéré. Le pixel central de l'image pour ces altitudes est respectivement de longueur 40 m, 200 m et 400 m aux altitudes 1 km, 5 km et 10 km.

## III.3 Différence de résolution temporelle des informations

### III.3.1 Instant d'acquisition des pixels de l'image satellite

Le principe d'acquisition des images satellites issues du satellite MSG est décrit dans la figure 40. Cette acquisition est réalisée en utilisant la combinaison de la révolution du satellite et de la rotation des miroirs de balayage, un processus pas à pas qui induit un écart d'instant d'acquisition entre les pixels de l'image finale. Les images sont prises du Sud au Nord et d'Est en Ouest.

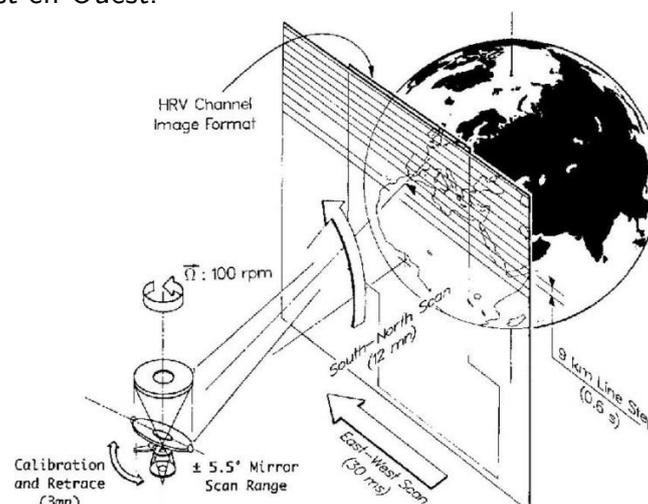


Figure 40 – Principe d'acquisition d'une image par le satellite Météosat Seconde Génération (MSG).  
Source : Müller et al. (2013).

Les informations fournies par EUMETSAT ont permis de générer la figure 41. L'imagerie à 12 canaux du satellite, appelé SEVIRI, observe le disque complet de la Terre avec un cycle de répétition de 15 minutes. L'inscription en rouge correspond à l'instant d'acquisition de la ligne de pixels recouvrant les sites à l'étude. Aussi, afin de s'assurer de la cohérence d'une comparaison avec les images hémisphériques, l'instant d'acquisition des extrémités des images satellites sont affichées sur la figure 41. L'instant  $t$  associé à chaque image satellite étant l'instant de fin d'acquisition, la comparaison avec les informations *in situ* se fera en prenant l'instant  $t - 4$  min comme référence au sol, avec une erreur acceptable de moins de 10 s.

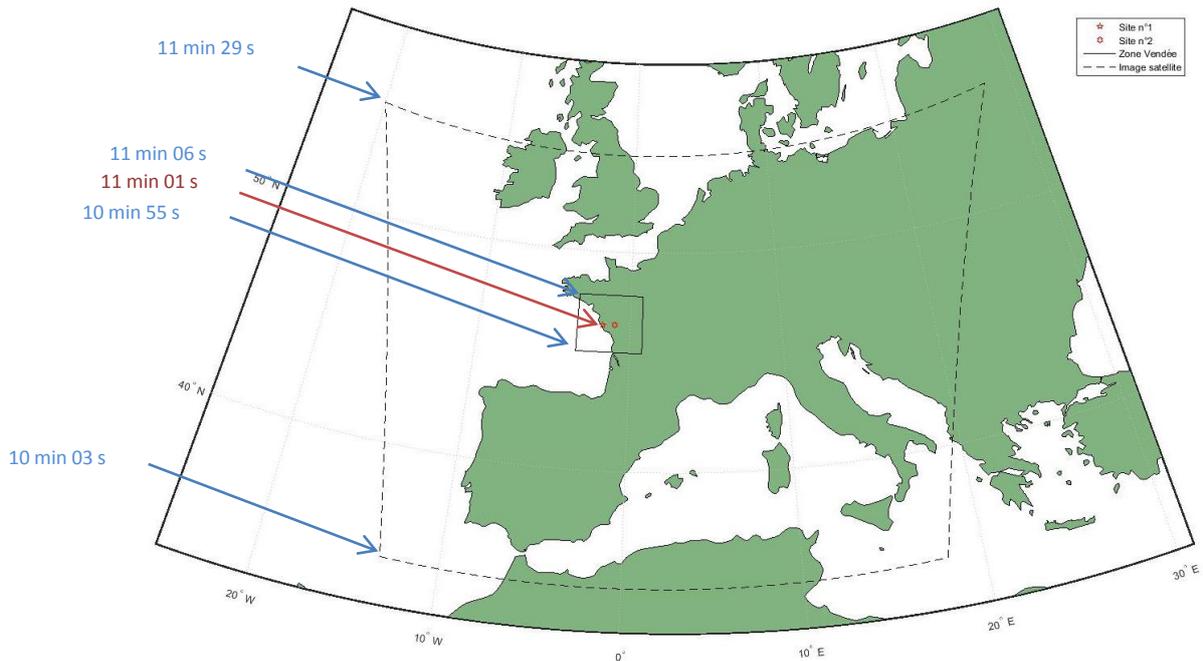


Figure 41 – Détermination de l'instant d'acquisition des lignes de l'image satellite. En rouge la durée entre le début d'acquisition de l'image et l'acquisition de la ligne qui contient les deux sites d'étude.

### III.3.2 Résolution temporelle des informations *in situ*

L'instrumentation *in situ* délivre les mesures du GHI et du DHI qui sont stockées au pas de temps 1 min par intégration sur les derniers échantillons recueillis. La prise de photo de l'imagerie hémisphérique est quant à elle paramétrée sur un pas de temps 10 s. Cette différence de résolution temporelle avec les images satellites peut être assimilée de trois manières : (1) pour les observations de la suite de ce chapitre et pour toute analyse au pas de temps 15 min, seule l'image obtenue instantanément 4 min avant l'image satellite sera exploitée ; (2) pour une analyse des pas de temps plus courts, des techniques de sur-échantillonnage temporel des images satellites peuvent être utiles pour l'utilisation de l'ensemble des images hémisphériques (*cf.* annexe A2) ; (3) l'agrégation temporelle des images hémisphériques pour se ramener à un pas de temps de 15 min est une méthode qui a été écartée puisqu'elle floute l'information sur la position des nuages et leur texture, réduisant considérablement leur potentiel d'exploitation.

## III.4 Observation des complémentarités des différentes sources d'information

Les nuages sont perçus différemment du point de vue de chaque source d'information. Ces observations tentent d'illustrer – de manière non exhaustive – certaines particularités directement issues des caractéristiques des sources d'information en présence, mais également certaines complémentarités qui suggèrent de tirer parti d'une utilisation conjointe de ces informations.

### III.4.1 Ciel clair

La situation de ciel clair est souvent considérée comme le cas le plus simple à estimer ou à prévoir. Cependant, certaines caractéristiques des sources d'information et de la physique atmosphérique peuvent rendre cette tâche complexe.

En effet, la définition même d'un instant de ciel clair comme étant l'absence de nuages peut être ambiguë à cause de la définition d'un nuage qui peut être évasive. Calbó et al. (2017) ont dédié une publication scientifique au sujet de la frontière floue entre les définitions de nuages et d'aérosols, qui font toutes deux référence au même phénomène physique : des particules en suspension dans l'air. Cette distinction se justifie par leurs différences de compositions, de taille et de nombre de particules (i.e. différences de propriétés optiques).

Certaines des situations proposées dans ce paragraphe s'inscrivent dans la zone appelée « crépusculaire » – *twilight* – par Koren et al. (2007), qui décrit une "ceinture de formation et d'évaporation de fragments de nuages et d'aérosols hydratés s'étendant à des dizaines de kilomètres des nuages dans la zone dite «sans nuage»". Comme nous l'avons expliqué dans l'introduction, certains modèles de ciel clair, tel que le produit McClear, incorpore les informations sur les aérosols dans le modèle de ciel clair.

Certaines pistes d'explications ont été mentionnées, les observations des instants de ciel clair se focaliseront sur les situations décrites dans le tableau suivant, où 1 correspond à un instant perçu comme ciel clair selon chaque source :

| Description de la situation  | Figure associée          | Mesure du GHI | Image au sol | Image HRV | Image HelioClim |
|--|--------------------------|---------------|--------------|-----------|-----------------|
| Toutes les sources d'information s'accordent sur le ciel clair   | Figure 42                | 1             | 1            | 1         | 1               |
| La mesure dévie de la courbe de ciel clair sans nuage visible  | Figure 43                | 0             | 1            | 1         | 1               |
| Des nuages sont visibles à l'horizon mais aucun impact sur la mesure   | Figure 44                | 1             | 0            | 1         | 1               |
| Image satellite dégagée, mais du point de vue du sol le ciel est couvert : 2 possibilités, soit effet parallaxe soit nuage plus petit qu'1 km. | Parallaxe<br>Figure 36   | 0             | 0            | 1         | 1               |
|  | Petit nuage<br>Figure 45 |               |              |           |                 |

La figure 42 est une situation proche du cas idéal d'un jour de ciel clair. À l'instant considéré, aucun nuage n'est apparent sur les trois images. L'image HRV affiche ainsi la silhouette de la côte atlantique aux abords de la région de Vendée. Dans cette situation, la qualité des estimations ou prévisions du rayonnement solaire repose principalement sur la capacité à estimer ou prévoir les concentrations des différents constituants atmosphériques qui ont un impact sur le rayonnement solaire incident au sol, et ainsi obtenir une courbe d'éclairement par condition de ciel clair proche de la mesure. D'après la figure 42, le modèle climatologique ESRA semble convenir aux conditions atmosphériques de la journée du 02/09/2014. Cependant certaines légères déviations peuvent survenir lorsque le trouble de Linke effectif s'éloigne de celui estimé dans le calcul du modèle ESRA. Par exemple, les valeurs affichées sur la courbe de la mesure de la figure 43 sont systématiquement plus petites que les valeurs estimées par le modèle ESRA, ce qui suggère une sous-estimation de la valeur de ce trouble.

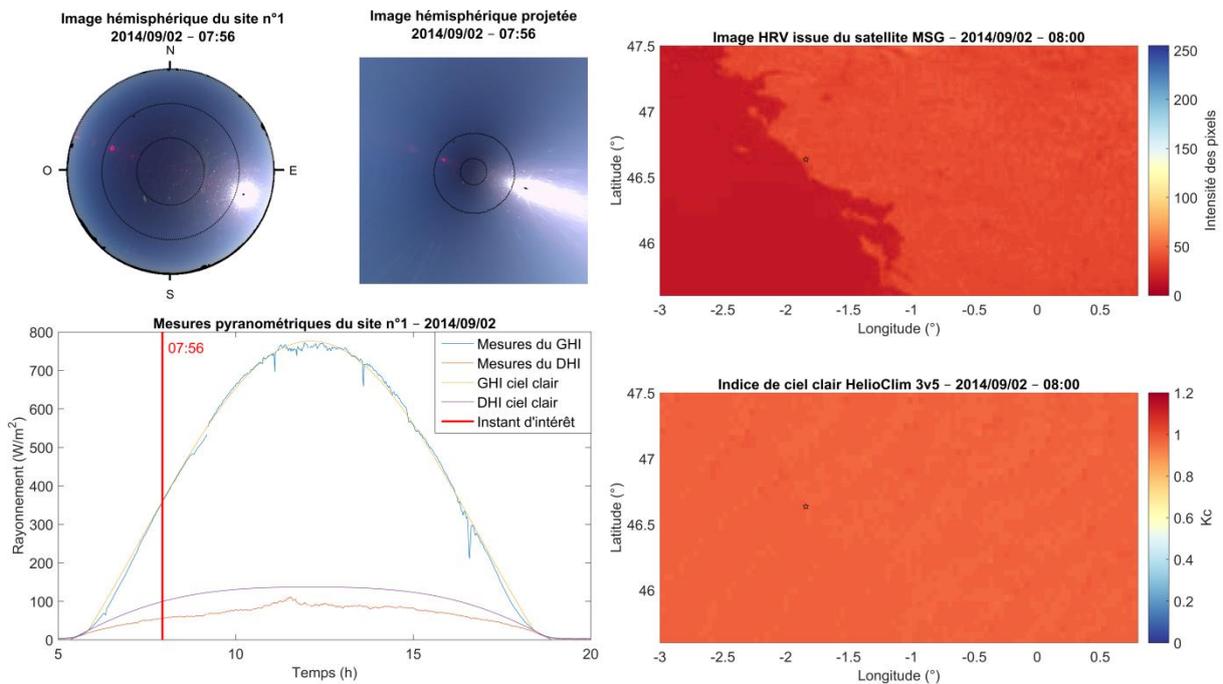


Figure 42 – Instant perçu comme ciel clair par toutes les sources d'information. Aucun nuage n'est visible sur les deux types d'images et la mesure suit parfaitement la courbe théorique d'éclairement par ciel clair du modèle d'ESRA.

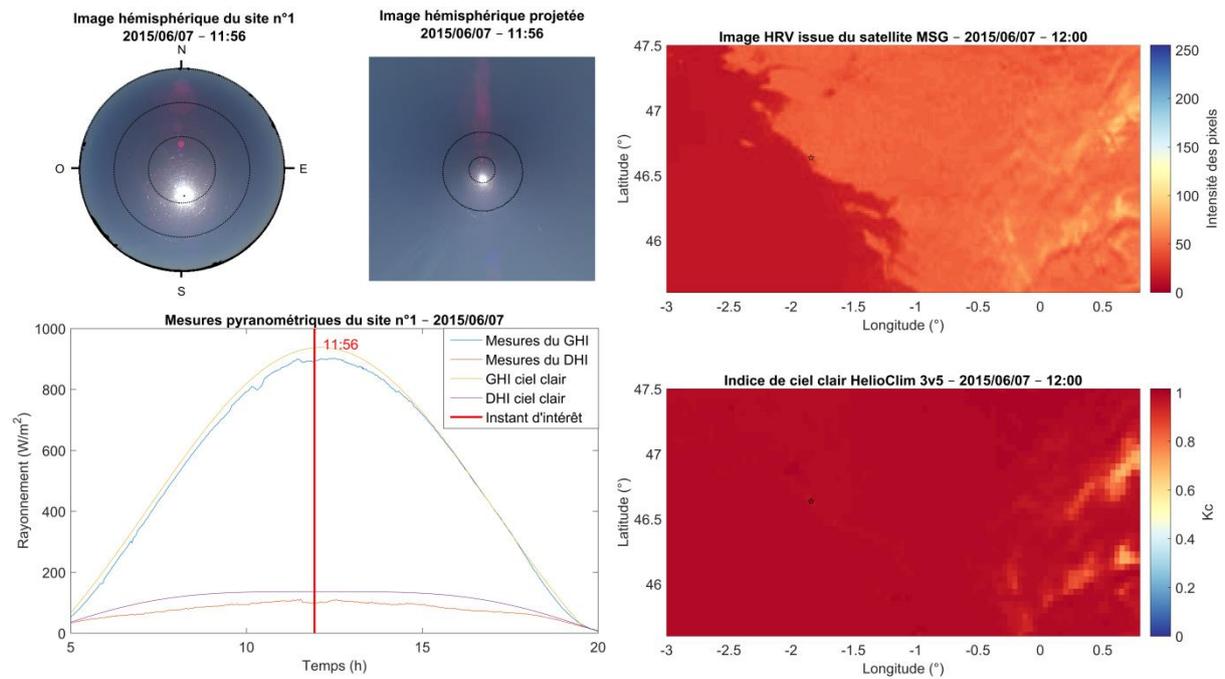


Figure 43 – Instant perçu comme ciel clair par toutes les images : aucun nuage n'est visible sur les images satellites comme sur l'image hémisphérique. Cependant la mesure s'éloigne légèrement de la courbe théorique d'éclairement par ciel clair du modèle d'ESRA.

La définition du « ciel clair » comme une absence de nuages n'attribuerait pas l'instant de la figure 44 à cette catégorie. En effet, l'image hémisphérique indique la présence de nuages dans plusieurs secteurs du ciel, essentiellement dans sa partie Ouest. Cependant, l'observation de la mesure indique que la courbe d'éclairement par ciel clair est presque parfaitement atteinte sur une période de 30 min. Du point de vue satellitaire, le site n°1 peut également être considéré comme ciel clair. Cette situation illustre la difficulté d'estimer le rayonnement diffusé par l'environnement nuageux et les autres constituants atmosphériques.

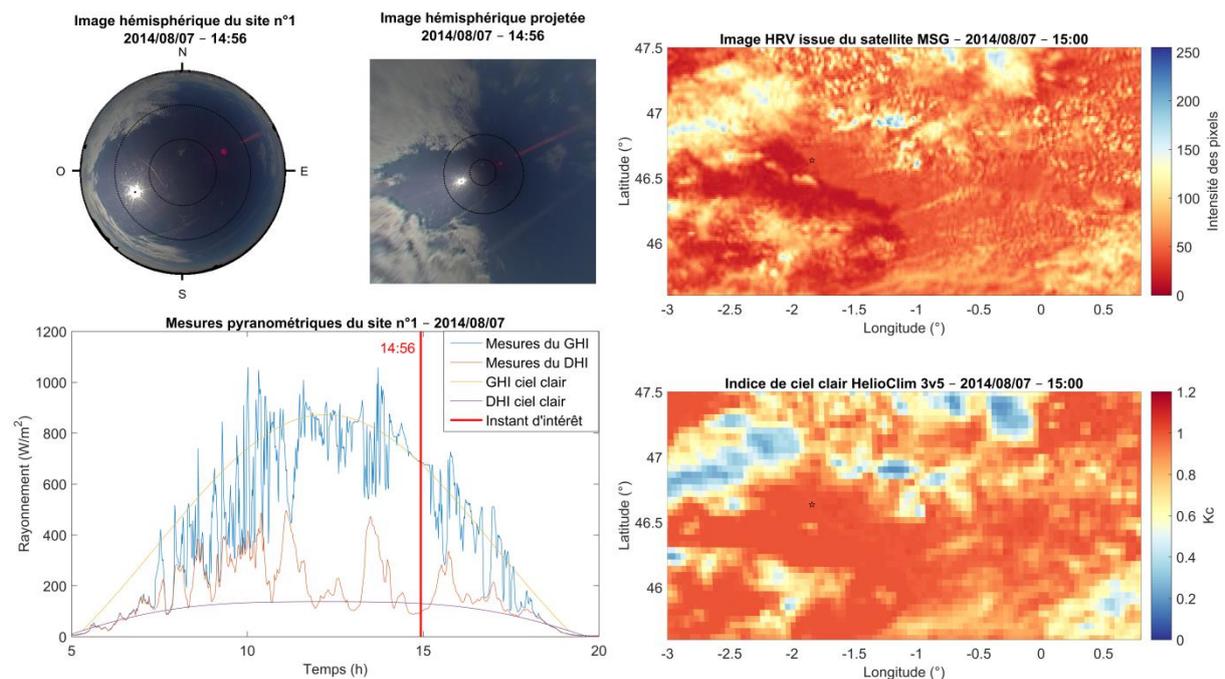


Figure 44 – Instant perçu comme ciel clair du point de vue satellitaire et de la mesure, mais des nuages sont présents sur l'image hémisphérique.

Le déficit de résolution spatiale des images satellites est illustré par la situation de la figure 45, où le point de vue satellitaire n'est pas en mesure d'estimer et encore moins de prévoir la baisse de rayonnement observée à 13h26 le 01/08/2014 sur le site n°1. Ces fluctuations d'amplitude moyenne autour de 70 % de la valeur par ciel clair peuvent être attribuées aux nuages filandreux observés sur l'image hémisphérique, dont la taille et la transparence en font des objets difficiles à percevoir avec la résolution spatiale limitée de l'observation satellitaire.

Un lien entre variabilité spatiale et variabilité temporelle peut être observé ici, où les petits amas nuageux viennent atténuer tour à tour le rayonnement direct reçu par le pyranomètre. Ce nuage a été identifié comme étant un cirrocumulus, grâce à son aspect granuleux et son altitude estimée par l'image CTH est de plus de 6 km. Ces nuages sont principalement constitués de cristaux de glace.

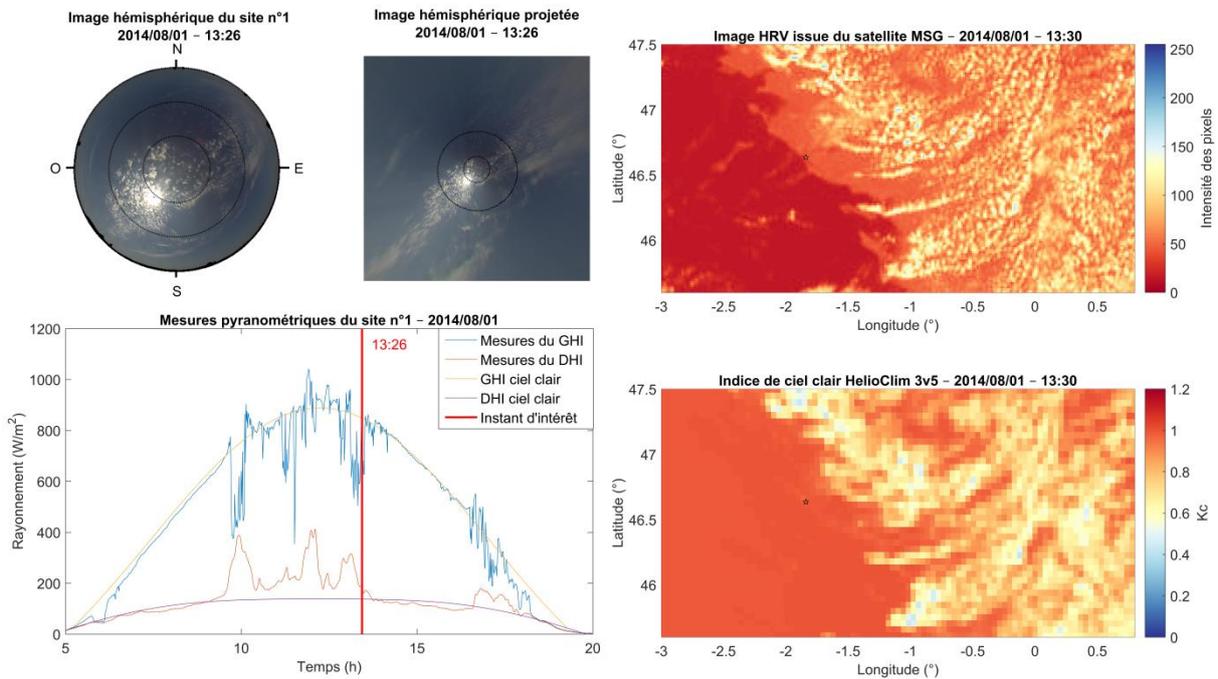


Figure 45 – Des nuages de très petites tailles se trouvent dans la zone circumsolaire, générant une variabilité moyenne du rayonnement au sol. L'image hémisphérique délivre une bonne perception de ce type de nuages, mais leur taille inférieure aux résolutions des images satellites en font des objets invisibles du point de vue satellitaire.

### III.4.2 Ciel couvert

Typiquement, un instant correspondant à un ciel couvert admet un rayonnement global et direct mesuré plus faible par rapport à sa valeur théorique par ciel clair et une variabilité temporelle locale plutôt faible. Une situation faisant l'unanimité des observations des sources d'informations est proposée sur la figure 46. Tous les pixels de l'image hémisphérique perçoivent le nuage dense présent au-dessus du site, les images satellites indiquent nettement la présence du nuage sur le site et ses environs, et la mesure du GHI égale celle du DHI, atteignant 10 % de sa valeur théorique par ciel clair et ne variant que très légèrement dans le voisinage de l'instant considéré.

Dans un contexte prédictif, toutes les sources d'information s'accordent sur la prévision d'un instant couvert à court et très court terme. Cependant, une prévision précise du rayonnement diffus ou de pics de rayonnement direct n'est pas chose aisée. Par exemple, le pic de rayonnement intervenant à 16 h est complexe à prévoir : il s'agit d'un trou de petite taille dans la couverture nuageuse dense qui, du point de vue du sol, va laisser entrevoir le Soleil. Or, pour prévoir un tel évènement au bon moment, il est nécessaire d'avoir une bonne estimation du déplacement des nuages (CMV), une bonne résolution spatiale et une bonne estimation de la parallaxe (donc de l'altitude des nuages).

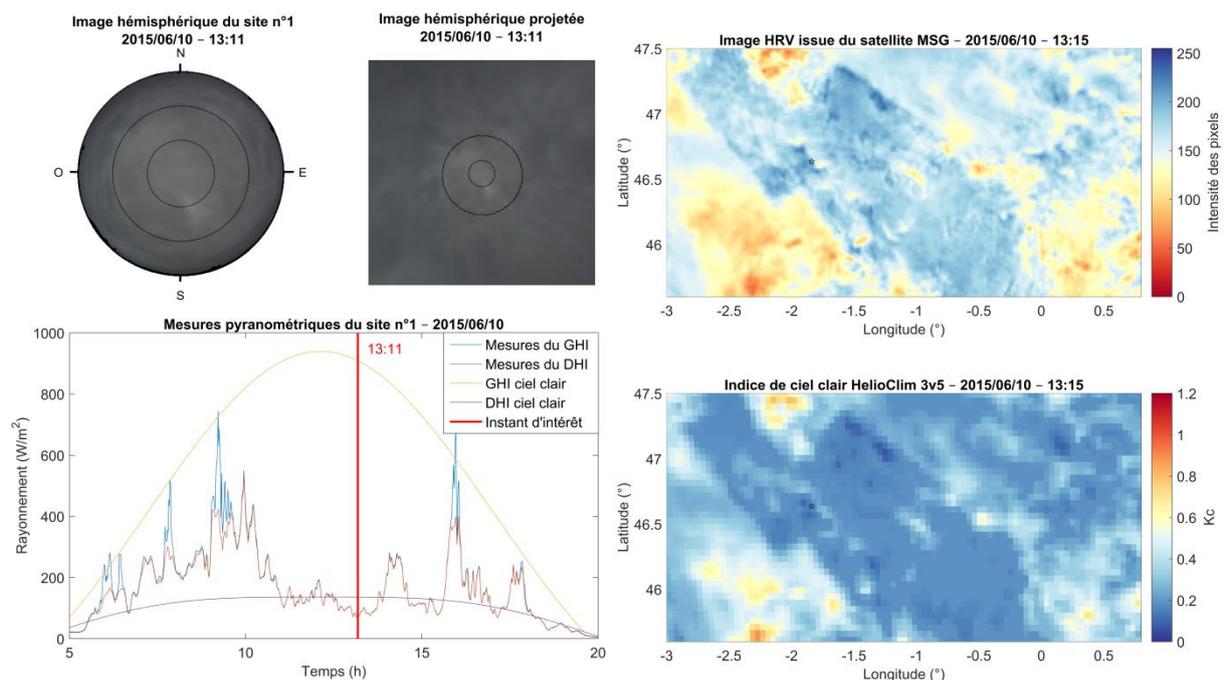


Figure 46 – Exemple d'un instant totalement couvert par des nuages ne laissant passer que 10 % du rayonnement théorique par condition de ciel clair. Toutes les sources d'information sont en accord.

La situation illustrée par la figure 47 présente également les caractéristiques d'un ciel couvert. Pourtant, une partie de l'image hémisphérique observe bien le ciel et non des nuages. La couverture nuageuse est localisée au-dessus du site n°1, selon le point de vue du satellite. Ainsi, cette figure et la figure 44 laissent penser que la présence ou non de nuages dans la zone circumsolaire joue un rôle prédominant dans l'atténuation du rayonnement solaire incident au sol. Blanc et al. (2014) soulignent le rôle significatif que joue cette zone sur l'évolution des composantes du rayonnement solaire incident au sol que sont les éclairagements direct et diffus circumsolaire.

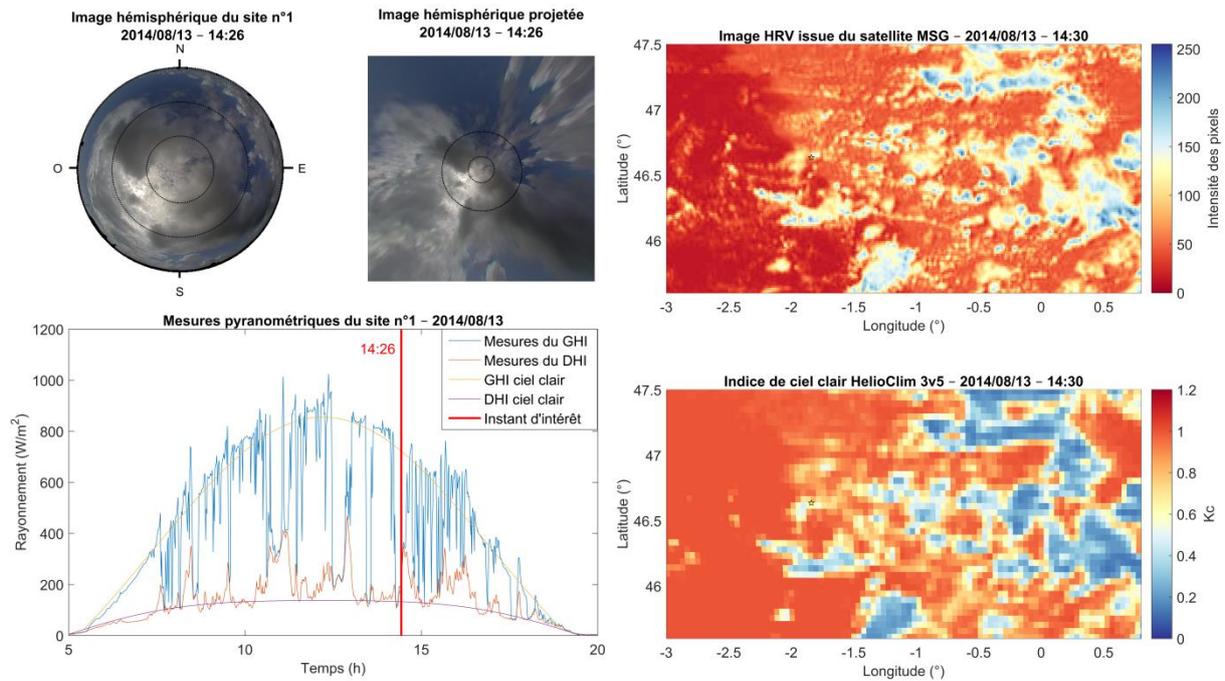


Figure 47 – Situation où le ciel est considéré comme couvert par toutes les sources d'information bien que le ciel soit visible sur l'image hémisphérique et que la couverture du nuage atténuant l'éclairement ne couvre qu'une portion limitée de la zone au-dessus du site n°1.

### III.4.3 Ciel variable

L'une des spécificités du rayonnement solaire incident au sol est sa variabilité à haute fréquence temporelle que génèrent les nuages par obstructions successives des composantes directe et diffuse circumsolaire du rayonnement. L'anisotropie de la composante diffuse de la luminance solaire joue également un rôle non négligeable sur cette variabilité.

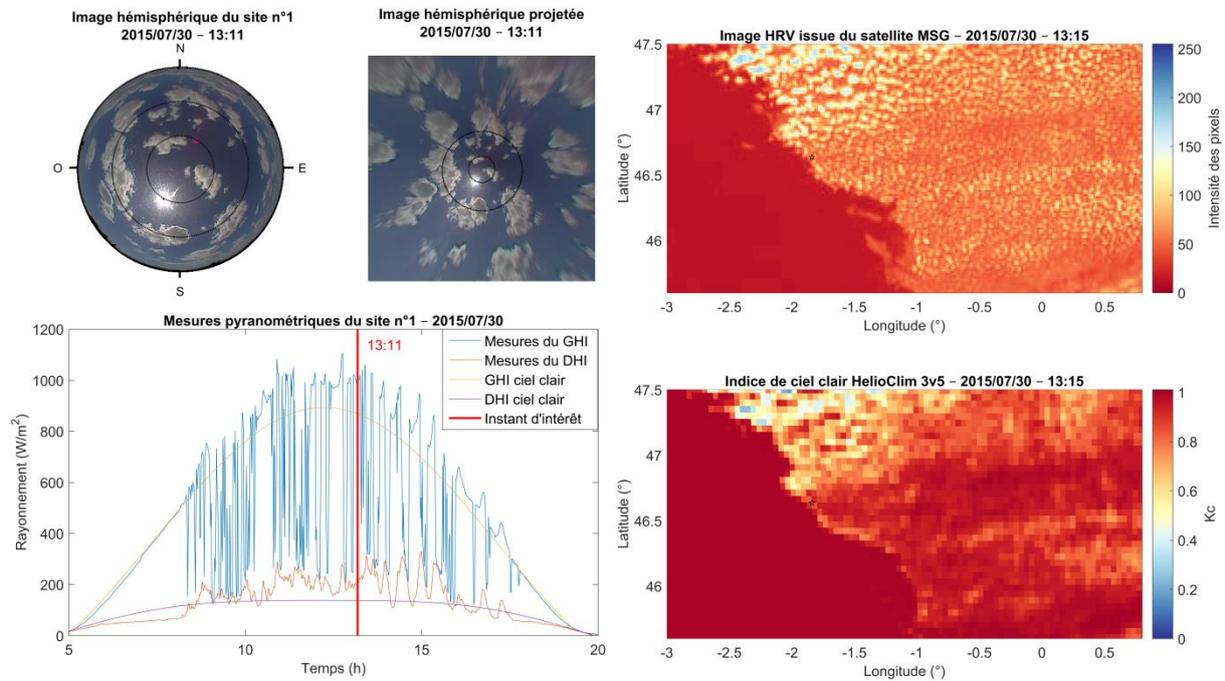


Figure 48 – Haute variabilité temporelle de la mesure avec forte amplitude. Nuages fractionnés sur l'image hémisphérique, légèrement perçus par HRV mais information floue sur HelioClim

Un exemple de situation affichant une telle variabilité de grande amplitude est proposé à la figure 48. Sur l'image hémisphérique, des nuages de tailles variables sont répartis dans le ciel au-dessus du site n°1. Les mouvements et l'évolution de cette couverture nuageuse fractionnée engendrent une forte variabilité temporelle du rayonnement, principalement entraînée par l'évolution « en dent-de-scie » de sa composante directe et les pics marqués de la composante diffuse suite à de complexes et fugaces réflexions multiples. Du point de vue satellitaire, l'image HRV perçoit la « granularité » de la couverture nuageuse, mais, compte tenu de la vitesse de déplacement des nuages, sa résolution temporelle ne permet pas de prévoir précisément la variabilité observée sur la mesure au pas de temps 1 min. L'image HelioClim est directement une estimation du rayonnement solaire sur le site, mais sa résolution spatiale étant trois fois moindre que celle de l'image HRV : uniquement le contenu de plus large échelle peut être perçu.

Pour illustrer les conséquences de la limite de résolution spatiale et temporelle des images HelioClim, la figure 49 donne l'évolution de l'estimation – par interpolation linéaire – de l'estimation issue d'HelioClim suivant son pas de temps d'origine de 15 min. Cette série temporelle est juxtaposée avec celles issue des mesures *in situ* au pas de temps 1 min et 15 min, pour la journée du 30/07/2015. La comparaison de ces courbes, notamment celles suivant le même pas de 15 min, montre clairement que la résolution spatiale et temporelle d'HelioClim ne permet pas de capter cette haute variabilité spatiale et temporelle du rayonnement, et constitue donc une source d'erreur pour l'estimation et la prévision solaire.

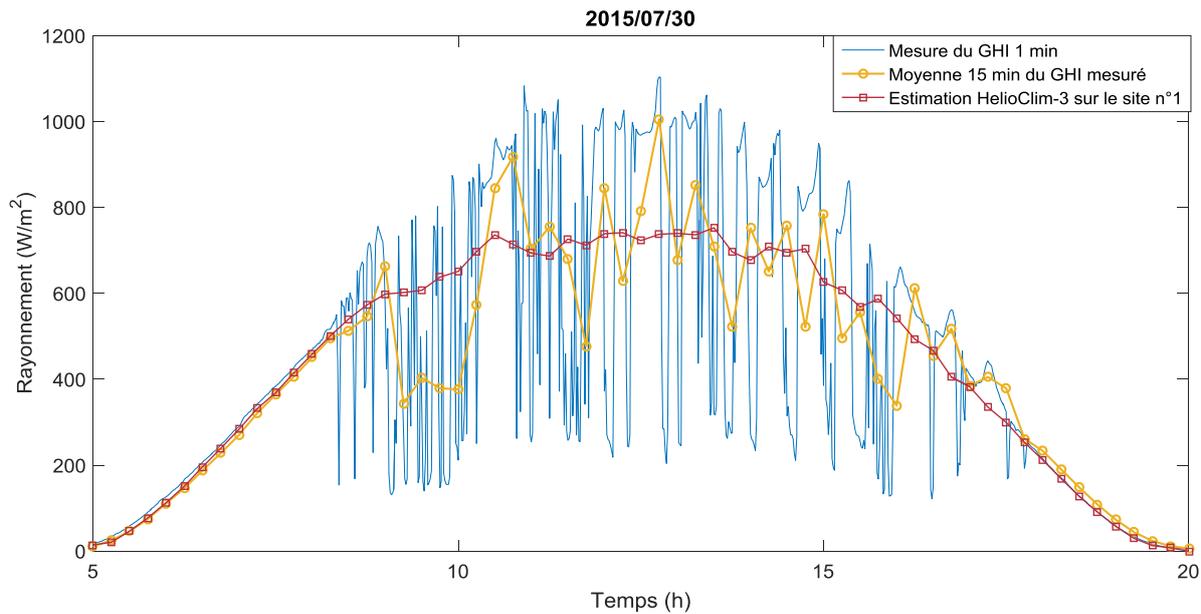


Figure 49 – Conséquences des limites de résolutions spatiales et temporelles d'HelioClim sur le contenu haute fréquence non perçu.

Du point de vue prévisionnel, la particularité de ce jour est la présence d'un mouvement des nuages du Nord-Est vers l'Ouest, avec une déformation des nuages allant jusqu'à leur disparition lors de leur arrivée au-dessus de l'océan Atlantique. Dans ce type de situations, plusieurs facteurs rendent la prévision précise du rayonnement solaire complexe même sur de courts horizons. Notamment, l'hypothèse de mouvement simplement advectif des nuages sans déformation, ni formation / disparition est de moins en moins vérifiée à mesure que l'horizon de prévision augmente. De plus, des estimations précises de l'altitude et du mouvement des nuages sont nécessaires pour prévoir toutes les rampes de pentes positives et négatives résultantes sur le rayonnement solaire incident au sol. Le dépassement récurrent de la courbe de ciel clair tout au long de la journée fera l'objet du paragraphe III.4.5.

La figure 50 propose une situation dont la couverture nuageuse s'avère complexe à modéliser dynamiquement. En particulier, les amplitudes des variations dépendent de l'atténuation qui en est faite par chaque partie des nuages.

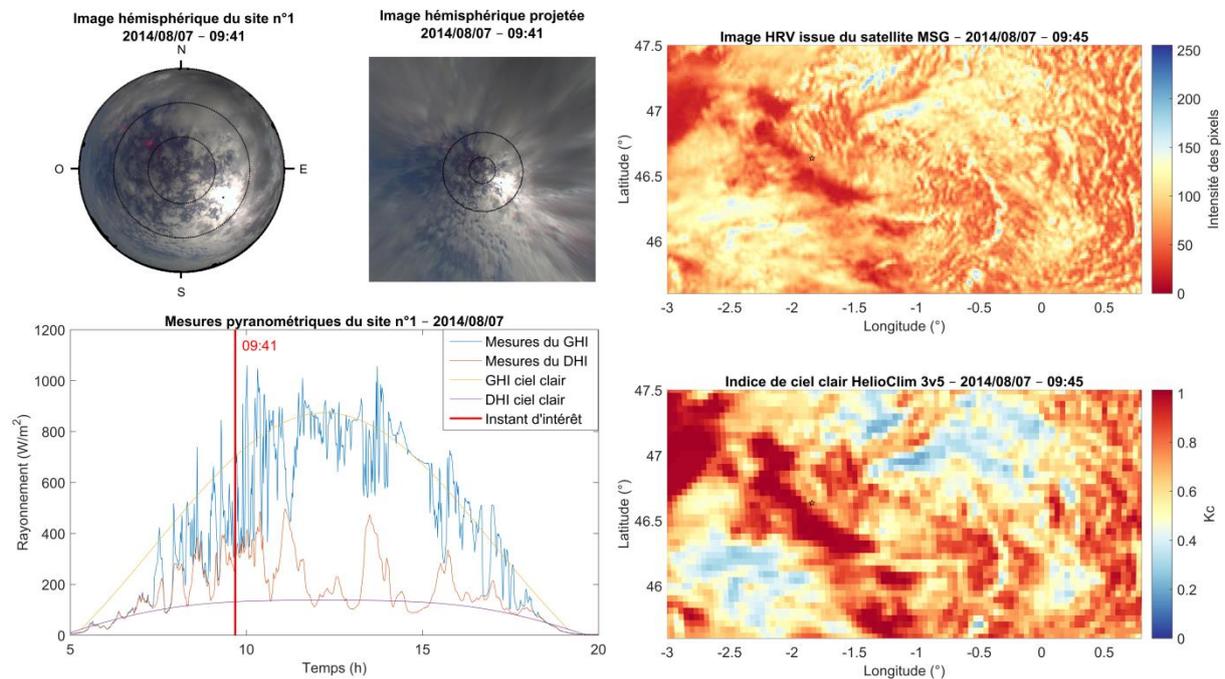


Figure 50 – Situation illustrative de la complexité de la couverture nuageuse susceptible d'intervenir sur le territoire côtier français.

Enfin, dans certains cas la variabilité spatiale des images HRV ne suffit pas non plus, comme le montre la figure 51 pour laquelle la forte variabilité temporelle de la mesure ne peut être complètement caractérisée par l'information satellitaire, même à l'aide du canal de haute résolution du visible HRV.

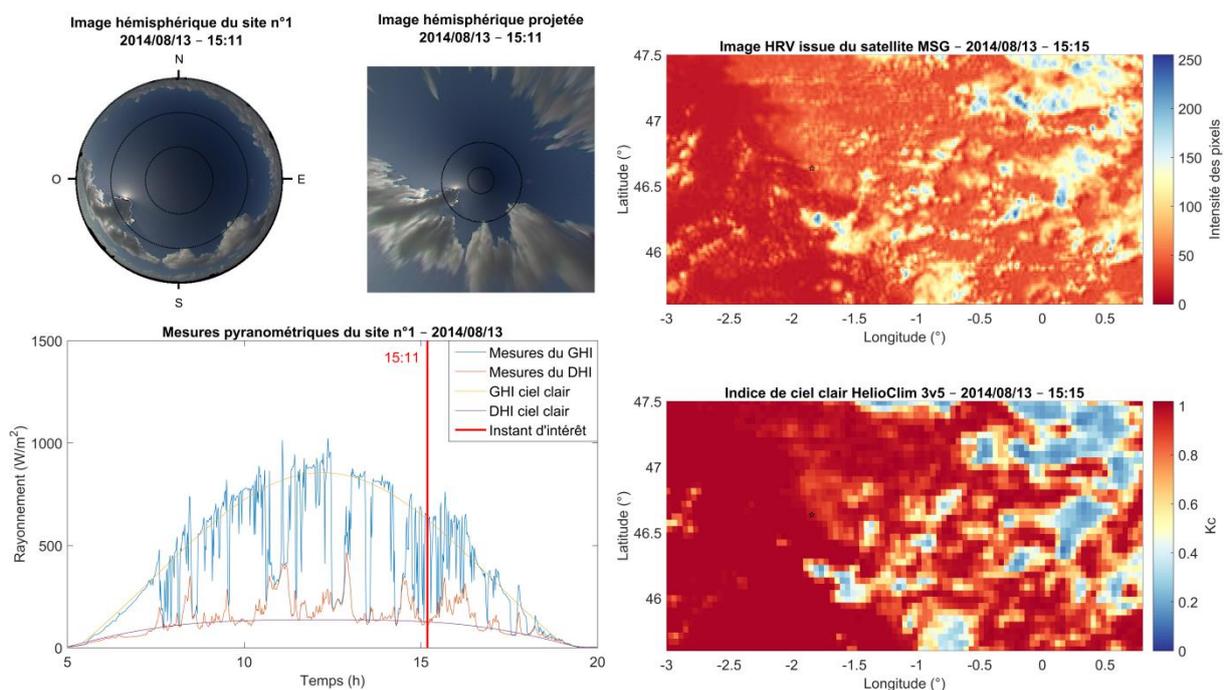


Figure 51 – Haute variabilité temporelle, variabilité invisible sur HelioClim, peu visible sur HRV.

### III.4.4 Ressemblance multiéchelle de la couverture nuageuse

L'Organisation Mondiale de la Météorologie (WMO) liste une grande diversité de types de nuages<sup>8</sup>, avec des altitudes et des caractéristiques de formes typiques. Certains ont la particularité d'avoir un aspect semblable à plusieurs échelles spatiales d'observation (autosimilarité). La figure 52 illustre une situation dont la couverture nuageuse a été identifiée comme étant de type cirrostratus. Ce type de nuage a la particularité de s'étendre de manière stratiforme en haute altitude, avec des propriétés de transparence générant un haut niveau de rayonnement diffus, tel qu'illustré sur la mesure du DHI (en orange) conjointement à une baisse importante de la composante directe.

Sur l'image hémisphérique, le Soleil apparaît en « transparence » et génère un halo lumineux dans la zone circumsolaire – typique de ce type de nuages. Enfin, du point de vue satellitaire, l'image HRV n'affiche que très peu de contenu hautes fréquences et l'image HelioClim indique une valeur moyenne de l'indice de ciel clair (c.-à-d 0.6). Ce déficit de contenu de haute fréquence spatiale de ce type de nuage indiqué par les trois images se répercute sur la mesure du GHI autour de l'instant considéré.

La principale difficulté qu'évoque ce type de situation est l'estimation du rayonnement solaire à l'aide uniquement de l'image hémisphérique. Le point de vue satellitaire semble très approprié pour la prévision solaire en présence de ce type de nuages, grâce à la ressemblance multiéchelle des propriétés du cirrostratus.

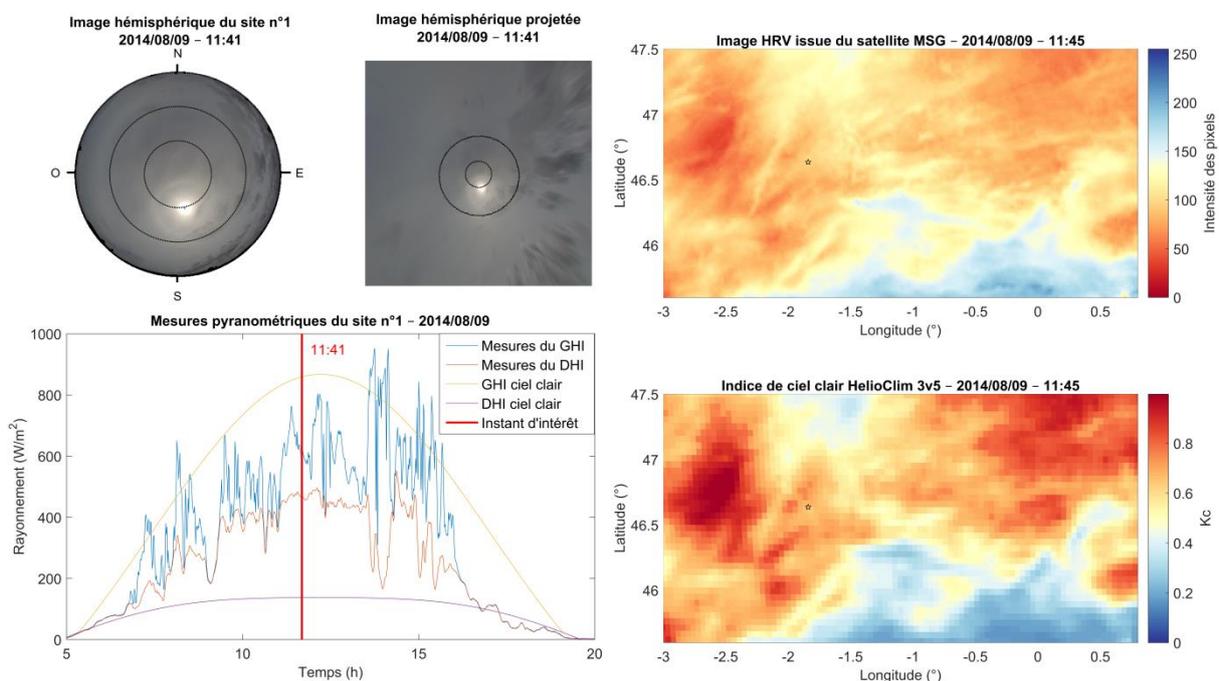


Figure 52 – Exemple de la présence d'un cirrostratus observé par toutes les sources.

<sup>8</sup> Atlas international des nuages de l'organisation météorologique mondiale – <https://cloudatlas.wmo.int>

À nouveau dans la catégorie des nuages de haute altitude, la figure 53 affiche une situation où la présence d'un nuage de type cirrus semi-transparent est observable sur les trois images. L'aspect filandreux se retrouve à plusieurs échelles spatiales : sur l'image hémisphérique, sur l'image HRV et dans une moindre mesure sur l'image HelioClim. L'effet de cette couverture nuageuse sur le rayonnement au sol est très caractéristique : il induit des légères fluctuations autour de la courbe de ciel clair. Au regard du rayonnement diffus mesuré, l'atténuation du rayonnement direct par ce nuage semi-transparent est en partie compensée par l'augmentation du rayonnement diffus.

Les similitudes multiéchelles vont dans le sens d'une description fractale des nuages telle que suggérée par Lovejoy et Mandelbrot (1985), où la forme des nuages est invariante par changement d'échelle. Une telle propriété pourrait ainsi être exploitée afin d'établir des prévisions aux échelles spatiales plus petites, et de manière corollaire avec des pas de temps plus fins.

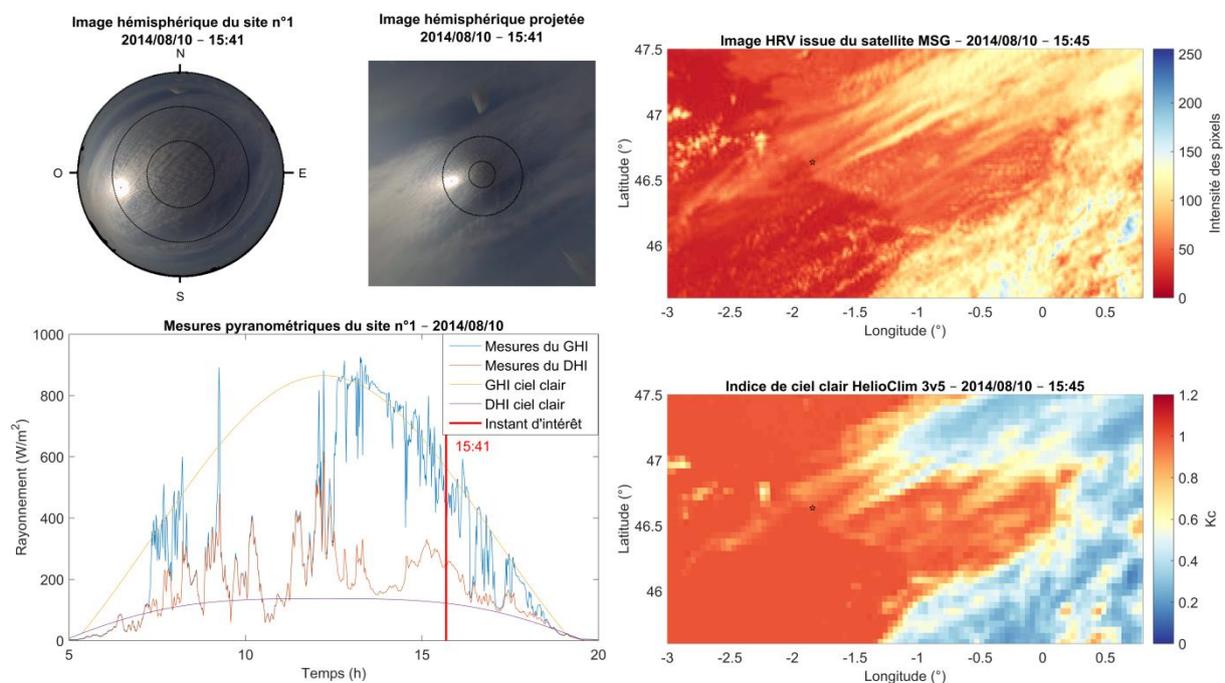


Figure 53 – Texture filandreuse des nuages perceptible sur les 3 images. Faible variabilité de la mesure autour de la courbe de ciel clair.

### III.4.5 Nets dépassements de la courbe de ciel clair

Un phénomène qui se produit notamment régulièrement sur les sites à l'étude et qui rencontre un intérêt croissant dans la communauté scientifique est nommé de nombreuses manières : *over irradiance*, *overshoot*, *cloud edge*, *cloud enhancement*, *cloud lensing*, *irradiance spikes caused by broken clouds*, *silver lining*, *super irradiance*, *cumulus solar-irradiance reflection*, *cloud gap effect*, *irradiance enhancement*. Ces dénominations se réfèrent toutes à un accroissement rapide et important de la composante diffuse du rayonnement causé par les nuages entourant le disque solaire non obstrué. Cet accroissement du rayonnement diffus vient donc s'ajouter au rayonnement direct, ce qui a pour effet une

amplification du rayonnement global pouvant ainsi dépasser de 30% – dans le cas extrême – la valeur au sommet de l'atmosphère sur les données au pas de temps 1 min (Gueymard, 2017). Cette dernière étude propose une analyse détaillée de ce phénomène à l'aide de 10 ans de données de mesures au pas de temps 1 min et 1 s. Nous décidons de nommer ces évènements « suréclairement ».

L'étude de ce phénomène a une importance du point de vue opérationnel, notamment pour le dimensionnement optimal des onduleurs des systèmes PV (Luoma et al., 2012). Par ailleurs, Chicco et al. (2016) soulignent-ils l'effet négatif qui est induit par ces effets sur les performances des prévisions des puissances produites par des systèmes photovoltaïques dont les onduleurs, non dimensionnés pour ces niveaux de suréclairements, induisent des écrêtements fugaces mais importants de la puissance en sortie.

La figure 54 présente une situation pour laquelle ce phénomène de suréclairement intervient sur une période d'une heure. Le cumulus longiforme qui apparaît sur l'image hémisphérique et sur les images satellites a une frontière située dans la zone circumsolaire. L'analyse précise de ce phénomène dans cette situation n'est pas simple, mais il repose certainement sur la conjonction de plusieurs effets d'accroissement de rayonnement en lien avec la diffusion et réflexion de celui-ci par les constituants atmosphériques – nuages inclus – et par l'albédo du sol. De ces effets multiples résulte une augmentation du diffus difficile à modéliser, estimer et donc à prévoir en se basant uniquement sur ces informations.

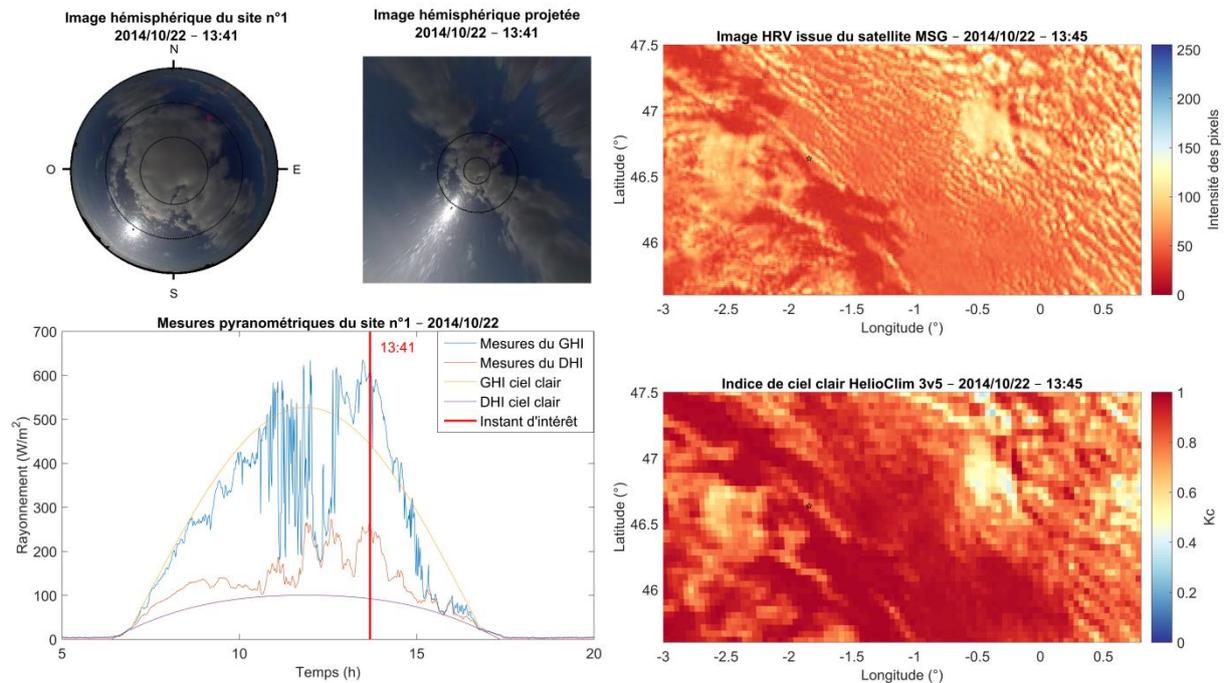


Figure 54 – Observation du phénomène de suréclairement persistant sur une heure. Illustration de la complexité de l'analyse du phénomène avec les moyens d'observation disponibles.

Toujours dans l'observation du phénomène de suréclairement, la figure 55 tente d'illustrer une difficulté supplémentaire de ce type d'évènements : les nuages observés dans l'image hémisphérique – qui semblent contribuer grandement à ces phénomènes – ne sont pas identifiables sur les images satellites, qu'il s'agisse de l'image HRV ou de l'image HelioClim. La présence de nuages fractionnés favorise l'apparition de ces dépassements très nets du rayonnement théorique par ciel clair. Dans cette situation, l'effet a perduré pendant 30 min.

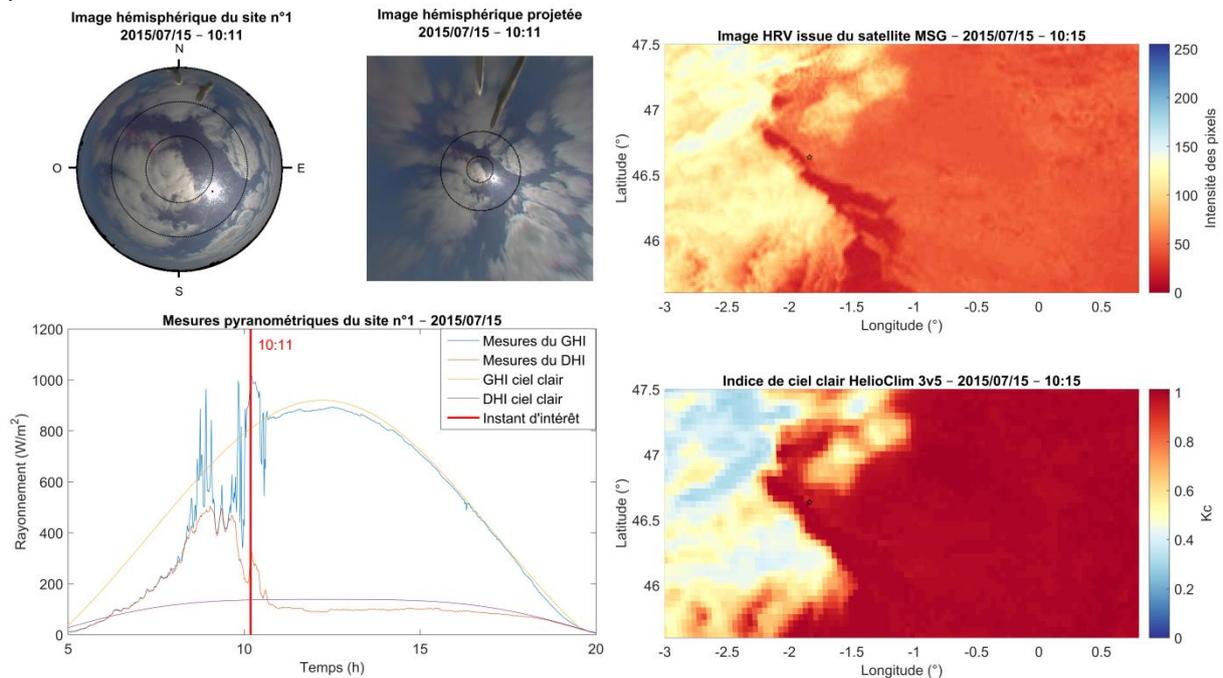


Figure 55 – Observation d'un phénomène de suréclairement alors que rien n'est visible avec l'information satellitaire uniquement.

Enfin, les instants des deux maxima du GHI mesuré sur les sites n°1 et n°2 sont présentés par les figures 56 et 57. Ils correspondent respectivement à un rayonnement incident mesuré à  $1294 \text{ W/m}^2$  et  $1330 \text{ W/m}^2$ , soit 109% et 118% des valeurs correspondantes hors de l'atmosphère. Dans les deux cas, les situations nuageuses sont relativement complexes et entraînent une haute variabilité temporelle sur la mesure. La variabilité spatiale à haute fréquence des nuages est peu visible avec l'image HRV et totalement imperceptible avec l'image HelioClim.

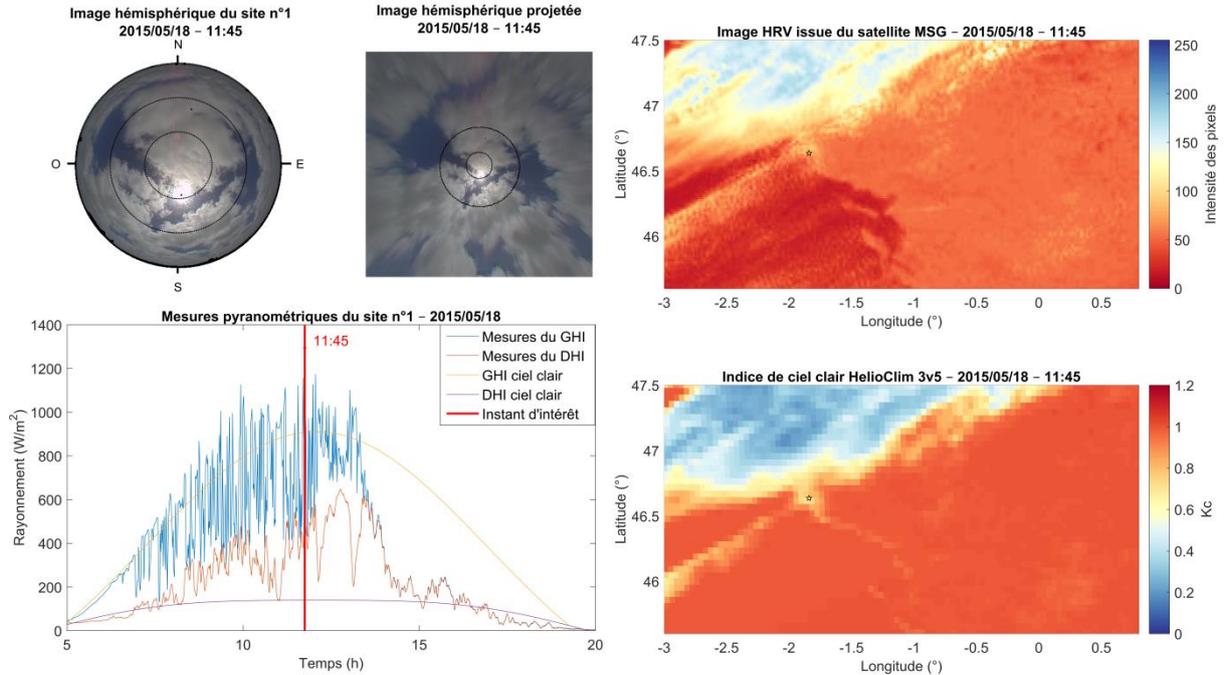


Figure 56 – Maximum de rayonnement atteint sur le site n°1 à 11h45 le 18/05/2015 pour atteindre  $1294 \text{ W/m}^2$  sur les mesures du GHI au pas de temps 1 min. Ce dépassement correspond à un dépassement de 9% de la valeur théorique au sommet de l'atmosphère.

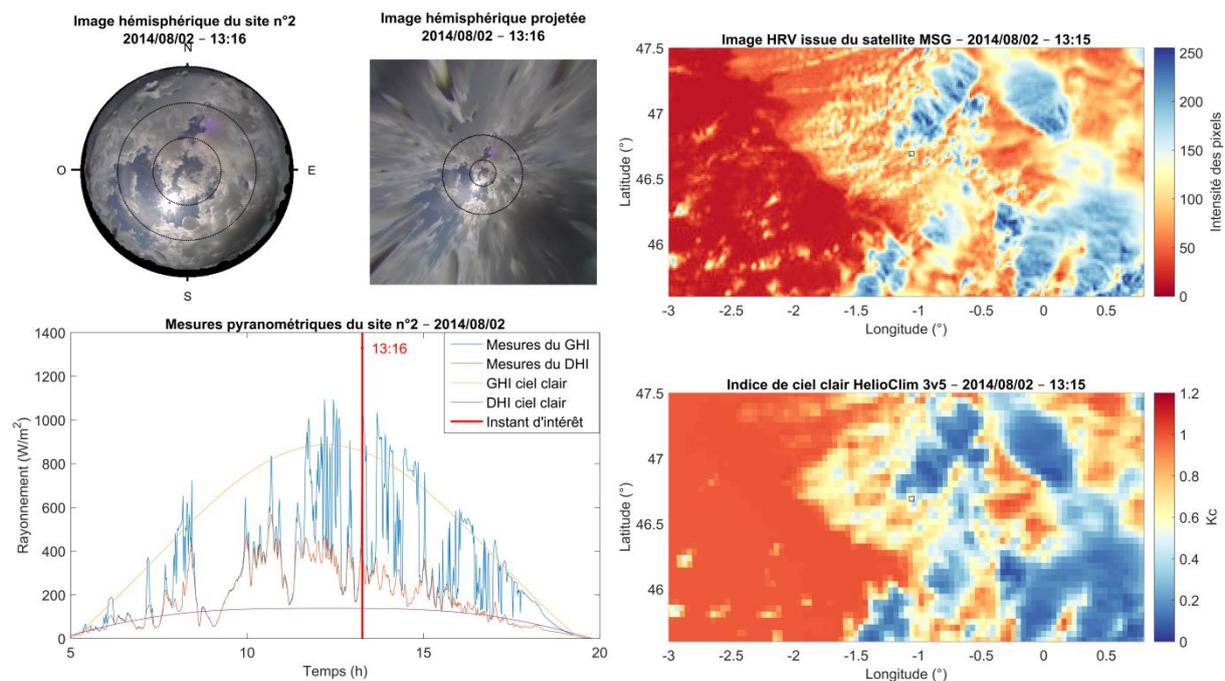


Figure 57 – Maximum de rayonnement atteint sur le site n°2 à 11h45 le 02/08/2014 pour atteindre  $1330 \text{ W/m}^2$  sur les mesures du GHI au pas de temps 1 min. Ce dépassement correspond à un dépassement de 18% de la valeur théorique au sommet de l'atmosphère.

### III.4.6 Situations propices à l'exploitation de complémentarités

Certaines situations mettent en évidence des complémentarités entre les sources d'information, et suggèrent qu'une exploitation de ces complémentarités peut être bénéfique pour l'application aux estimations et prévisions solaires.

La situation de la figure 58 présente un nuage dont l'épaisseur semble suffisamment fine pour envisager une correspondance entre l'image hémisphérique et l'image HRV. Une telle tentative a été réalisée, mais de nombreux facteurs limitent la mise en œuvre de cette mise en correspondance :

- La couverture nuageuse doit être monocouche puisque les points de vue sont opposés ;
- Dans le cas où la couverture nuageuse a une texture autosimilaire, l'association de l'image hémisphérique avec l'image satellite est délicate

Nous reviendrons sur ces difficultés au paragraphe III.5.

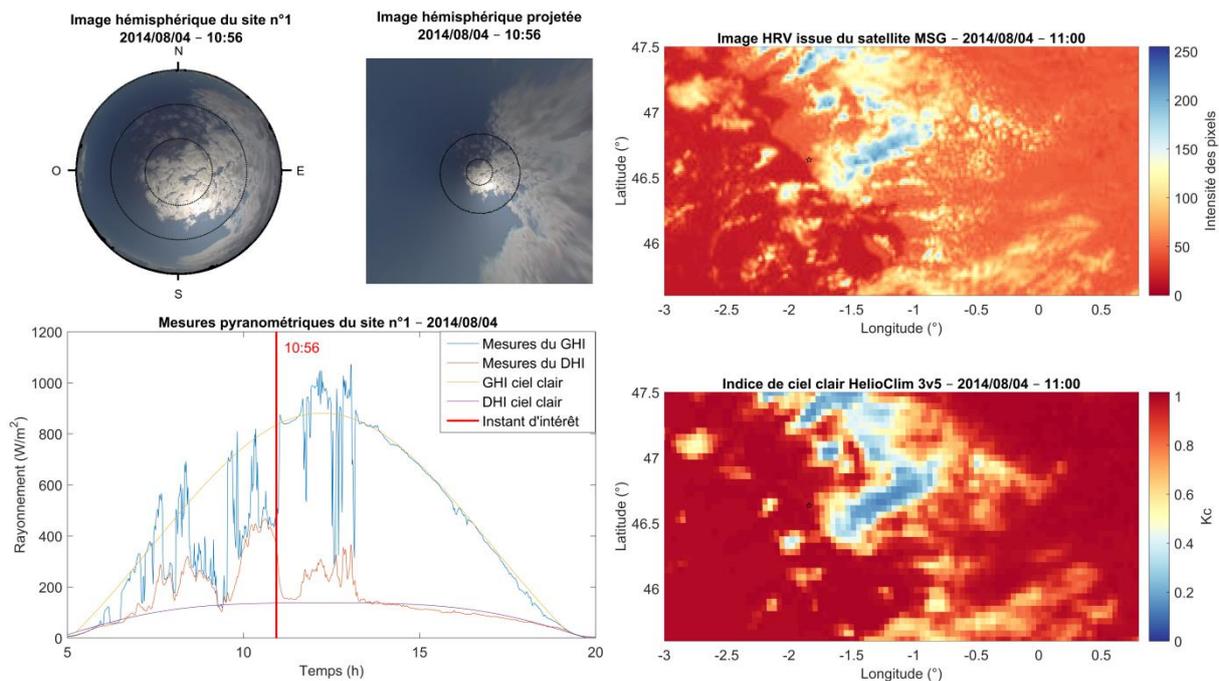


Figure 58 – Instant précédent un passage au ciel clair, couche nuageuse fine, silhouette de la frontière du nuage perceptible sur l'image HRV.

La matinée du 17/08/2014 sur le site n°1 permet d'illustrer les avantages et inconvénients de chaque source d'information. Les figures 59, 60 et 61 affichent les différentes observations respectivement à 9h, 10h et 11h.

L'évolution des masses nuageuses est ainsi proposée selon les différents points de vue. Les images satellites de la figure 59 indiquent qu'un nuage est présent au-dessus du site à 9h, bloquant la composante directe du rayonnement incident. Sur l'image hémisphérique, seul ce nuage est visible, occupant la totalité de l'hémisphère observé. En se basant uniquement sur ce type d'images, il est ainsi impossible de prévoir la situation décrite sur la figure 60 où une couverture nuageuse fractionnée apparaît une heure plus tard – couverture qui n'est pas encore visible à 9h du point de vue du sol.

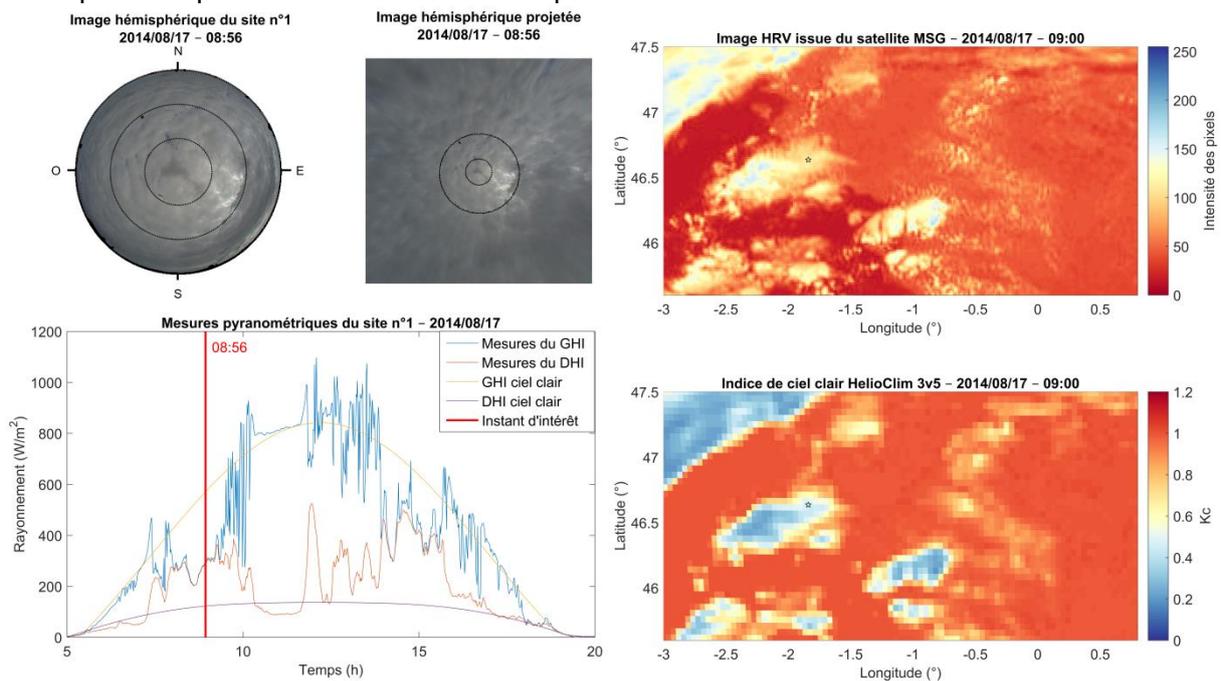


Figure 59 – Ciel couvert observé sur toutes les sources d'informations.

Ensuite, à 10h sur la figure 60, la masse nuageuse perçue du point de vue satellitaire s'est déplacée de telle manière que sa frontière s'approche du site d'intérêt. Or en observant l'image hémisphérique, cette frontière est en fait constituée de petits amas nuageux spatialement répartis. Ce sont ces petits nuages qui font obstacle au rayonnement direct du Soleil, générant ainsi la haute variabilité temporelle de la mesure autour de l'instant d'intérêt. L'image hémisphérique permet donc une bonne description de la variabilité spatiale à haute résolution, et donc de la variabilité temporelle à très court terme. Dans cette situation, la masse nuageuse au sud de l'image hémisphérique correspond à la masse nuageuse couvrant le site une heure avant. Cependant, là encore, il n'est pas possible de prévoir la situation de ciel clair qui apparaît 20 min plus tard en se basant uniquement sur l'image hémisphérique, du fait de sa vision limitée du ciel par l'horizon.

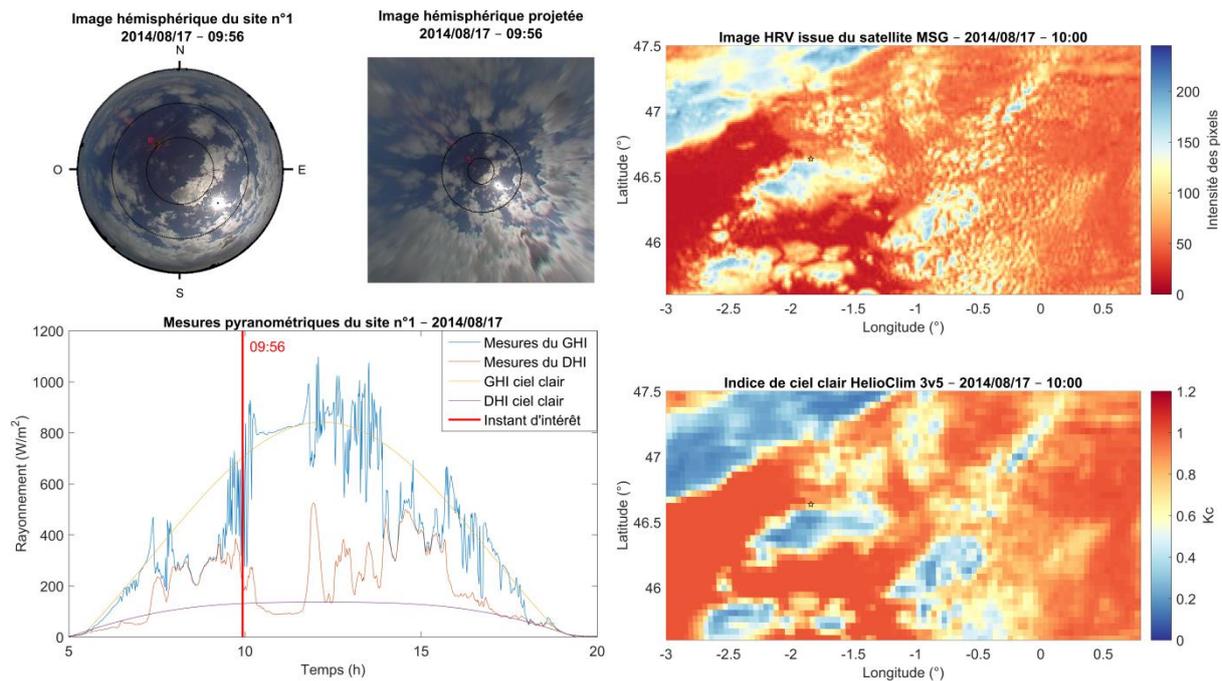


Figure 60 – Nuages fractionnés générant une grande variabilité temporelle de la mesure.

Enfin, la situation à 11h de la figure 61 affiche un ciel clair quasi parfait du point de vue de toutes les sources d'information. En conclusion, la dynamique du rayonnement solaire à basse fréquence est plutôt bien décrite par l'information satellitaire, mais son déficit de résolutions spatiale et temporelle dégrade sensiblement sa capacité à représenter la variabilité de haute fréquence temporelle induite par une couverture nuageuse en mouvement et finement fractionnée. Leur large couverture spatiale permet cependant d'effectuer des prévisions à plus long terme avec toutefois un effet parallaxe, source d'erreurs notamment temporelle qui aura des conséquences mesurables par le TDI : ce point sera discuté en détail dans le chapitre IV.

L'image HelioClim délivre une estimation du rayonnement solaire ce qui facilite son utilisation dans le contexte prédictif. À l'inverse, l'image hémisphérique permet une bonne description de la couverture nuageuse au-dessus du site, propice à une estimation et une prévision à très court terme des variations brusques. Cependant, sa vision limitée du ciel limite intrinsèquement les horizons de prévision possibles.

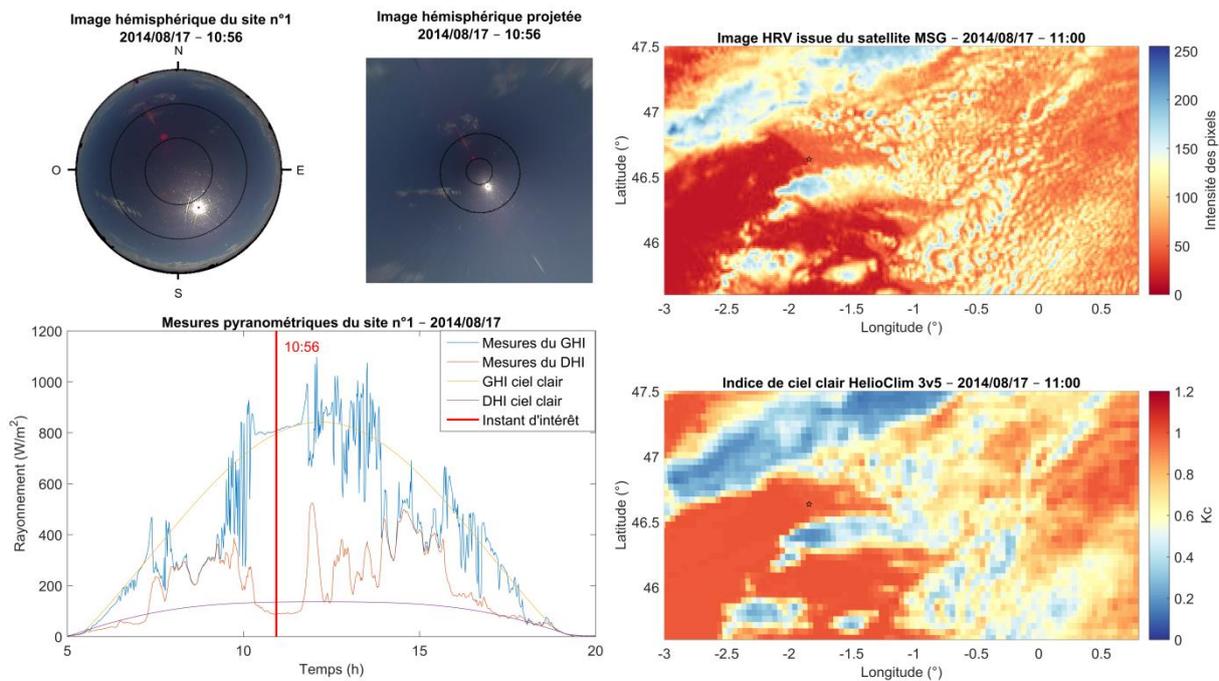


Figure 61 – Ciel clair observé sur toutes les sources d'information.

### III.4.7 Anomalies ou événements limitant l'exploitation des données

Ce paragraphe recense plusieurs artefacts, événements ou anomalies limitant l'exploitation des données ou qui sont susceptible de générer de mauvaises interprétations de celles-ci.

Ces exemples particuliers montrent que l'exploitation des différentes sources d'observations présentées ici soulève non seulement des difficultés théoriques complexes (effets de parallaxe, de suréclairement, de déficit de résolutions spatiales et temporelles, etc.) mais aussi des difficultés pratiques de mesures dans des conditions réelles qu'il ne faut pas oublier ou négliger.

Tout d'abord, les événements météorologiques tels que la pluie, la grêle ou la neige peuvent altérer l'exploitation des instrumentations *in situ*. Sur l'image hémisphérique, la figure 62 affiche un exemple d'instant où les gouttes de pluie viennent localement entraver la visibilité de la scène. Ces gouttes peuvent rester plusieurs heures sur l'objectif de la caméra et détériorent la qualité des algorithmes de détection de nuages ou d'estimation du rayonnement.

Des algorithmes de détection de ces gouttes d'eau ont par exemple été proposés par Roser et Geiger (2009) avec des tentatives de suppression de ces artefacts par You et Tan (2013). Sur les deux sites, aucun épisode de neige n'a été détecté sur la période d'étude. Or, la neige est un problème fréquent dans les analyses *in situ* de données pyranométriques, d'images hémisphériques ou de production photovoltaïque puisque les instruments recouverts de neige ne délivrent plus l'information désirée.

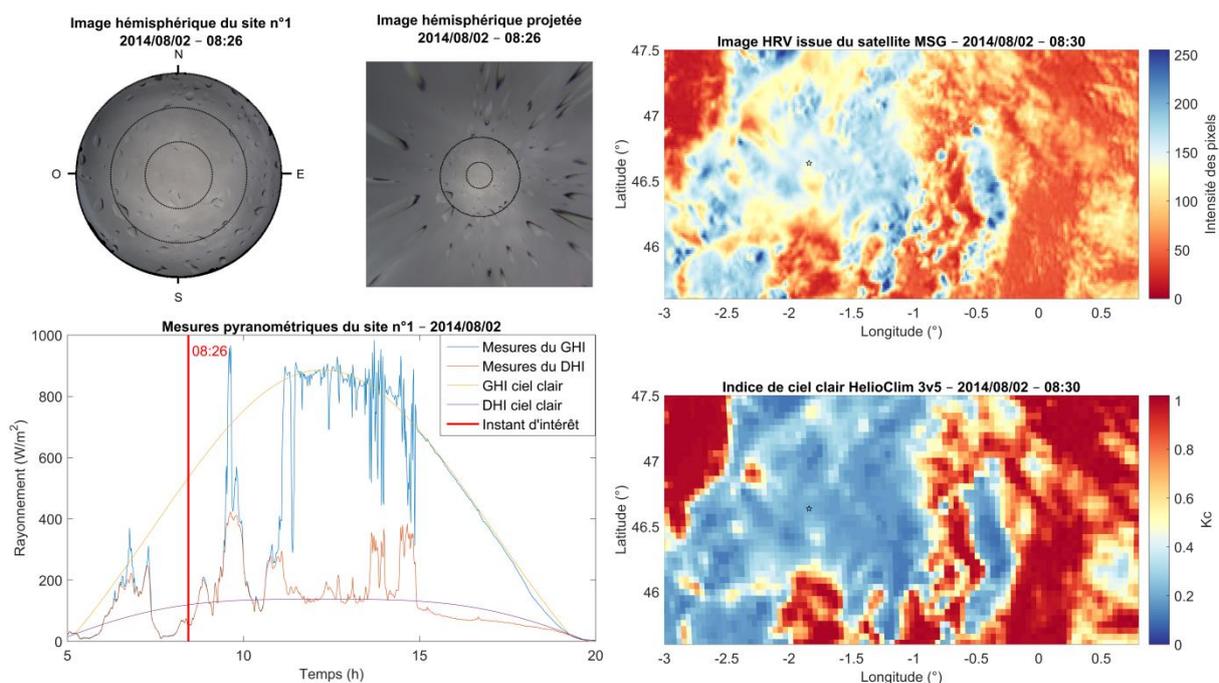


Figure 62 – Exemple d'un instant couvert pour lequel l'effet des gouttes d'eau détériore la perception de l'image hémisphérique. Avant 10h, trou dans la couche nuageuse qui génère un phénomène de suréclairement.

Qu'elles soient les conséquences ou non de précipitations, des salissures et poussières s'installent progressivement sur l'objectif de la caméra hémisphérique, ainsi que sur les autres instruments de mesure *in situ*. Une illustration de ce phénomène est proposée figure 63. Ces salissures peuvent perdurer plusieurs jours en attendant les prochaines intempéries ou le prochain nettoyage de l'instrumentation lors de la maintenance.

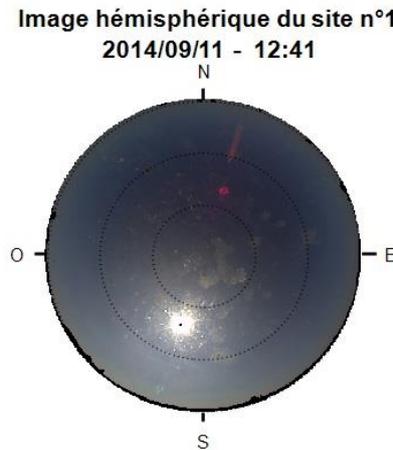


Figure 63 – Observation de salissures liées aux intempéries précédentes.

D'autres évènements liés à la faune locale peuvent survenir. La figure 64 illustre les obstructions du champ visuel de la caméra par des araignées et des oiseaux. De la même manière, et certainement plus difficiles à détecter, des oiseaux peuvent se percher sur l'instrumentation voisine et perturber les mesures. L'exemple de la figure 65 est typique de l'impact d'un oiseau posté sur un pyranomètre. La journée du 06/06/2015 est une journée où le ciel est complètement dégagé. Or entre 13h15 et 13h30, la mesure chute brutalement. L'image hémisphérique correspondant à cet instant renseigne sur la présence d'un oiseau sur le capteur SPN1 voisin.

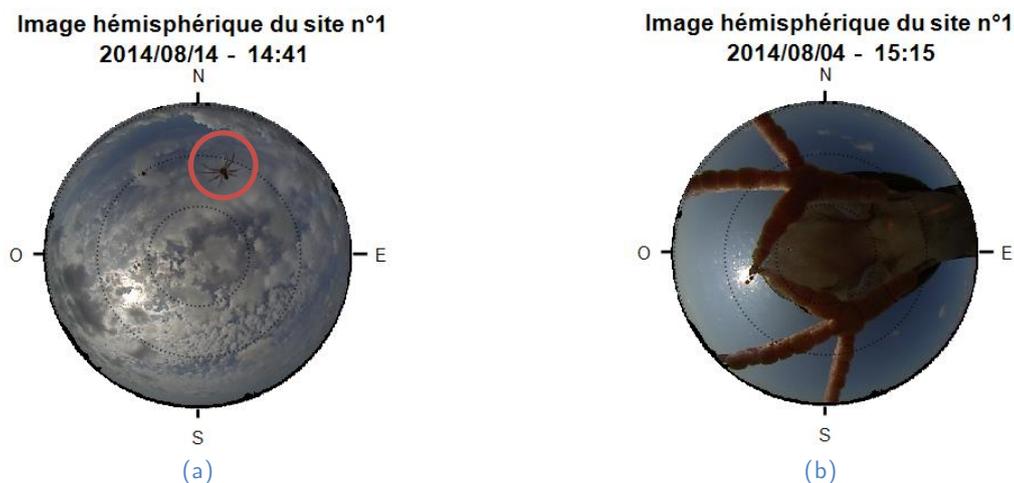


Figure 64 – Observation de la faune prenant support sur la caméra hémisphérique. Exemple d'une araignée (a) et d'un oiseau (b) observés sur le site n°1.

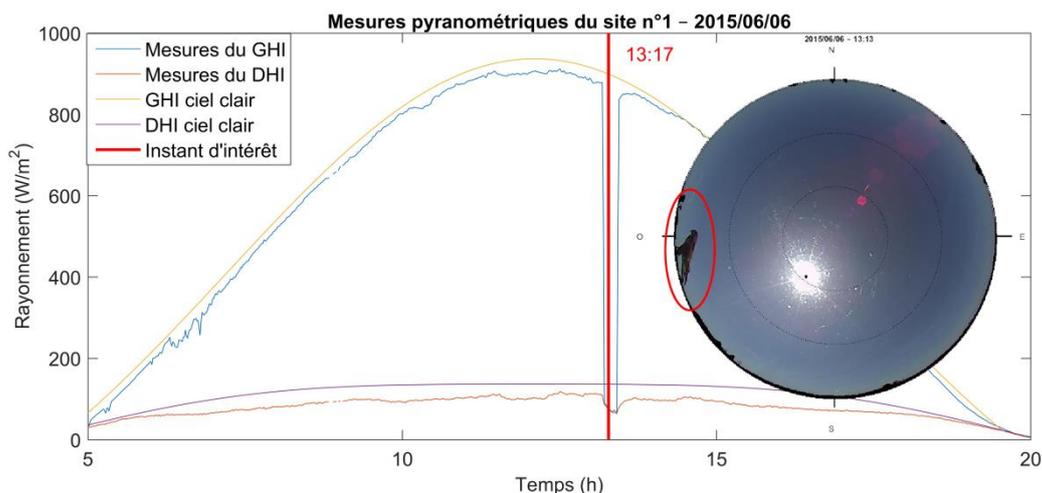


Figure 65 – Impact de la présence d'un oiseau sur le capteur pyranométrique lors d'une journée de ciel clair.

Finalement, plusieurs évènements sont susceptibles d'affecter la qualité des observations et des mesures *in situ*. Certains évènements peuvent également altérer l'information satellitaire, comme des déviations dans l'orientation, des perturbations ou des images manquantes. Cependant ces évènements sont moins fréquents et vite maîtrisés par les entités responsables de la qualité du service délivré par les satellites géostationnaires.

### III.5 Vers l'exploitation des complémentarités

Ce chapitre a été l'occasion d'observer les points communs et les différences des sources d'information à disposition. Aussi, des potentiels de complémentarités émergeant de ces observations ont été identifiés. En particulier, profiter de la fine résolution spatiale et temporelle des observations *in situ*, de la large couverture nuageuse perçue par le satellite, de la différence de point de vue des deux imageurs sont des pistes prometteuses dans la recherche de synergies. Ces synergies seront exploitées par l'utilisation de techniques de fusion de données.

La définition de la fusion de données proposée par Wald (1999) et qui sera retenue dans la suite de ce document est la suivante : « la fusion de données constitue un cadre formel dans lequel s'expriment les moyens et techniques permettant l'alliance des données provenant de sources diverses. Elle vise à l'obtention d'information de plus grande qualité ; la définition exacte de « plus grande qualité » dépendra de l'application ».

Dans notre cas, comme nous l'avons souligné dans le chapitre II, que l'on considère l'estimation ou la prévision du rayonnement solaire, une plus grande qualité s'exprimera en termes d'alignement temporel, de suivi des rampes et d'erreur statistique vis-à-vis de la mesure.

Au préalable, afin d'allier les données de sources diverses, il est essentiel de trouver un référentiel commun avant toute opération de fusion. Dans notre étude, ces référentiels

géométrique et temporel communs doivent être définis afin d'envisager l'élaboration de schémas de fusion.

Une première idée d'exploitation des complémentarités, qui ont été identifiées dans ce chapitre, a été de considérer les images satellites et les images hémisphériques comme des informations en partie redondantes de la géométrie 2D des nuages au-dessus des sites d'étude dans le plan horizontal. L'objectif était de pallier le déficit de résolution spatiale des images satellites en incorporant les détails de l'image hémisphérique, en espérant affiner suffisamment le contour des nuages localement au-dessus de chaque site pour améliorer la qualité du rayonnement solaire estimé et/ou prévu. Cette idée – qui a été explorée mais qui n'est pas détaillée dans ce document – s'est heurtée à la complexité de la géométrie des deux types d'images et des nuages observés, avec en particulier leur différence de point de vue qui rend l'alignement des deux types d'images compliqué.

L'étude des nuages en deux dimensions a ainsi rapidement trouvé ses limites dans la recherche d'une exploitation des complémentarités des sources d'information disponibles. La différence de point de vue, qui était considérée comme une limitation de l'approche 2D, est la base de l'approche qui sera développée dans les deux prochains chapitres : l'idée sous-jacente consiste à profiter justement de cette différence de point de vue pour estimer la géolocalisation 3D des nuages par fusion des données contenues dans les deux types d'images. Dans tout le manuscrit, l'expression géolocalisation des nuages fera référence à la fusion de données dans l'espace 3D.

En effet, l'espace 3D s'avère être un référentiel géométrique commun privilégié autorisant l'alliance des données présentées dans ce chapitre. Cependant, les deux types d'images ont des résolutions temporelles eux aussi différentes : il sera également nécessaire de trouver un référentiel temporel commun.

Face aux complexités géométriques de l'observation des scènes nuageuses, un simulateur a été conçu afin de mieux comprendre les potentialités qu'offre le référentiel 3D en termes d'estimation de la géolocalisation des nuages dans l'espace. Le chapitre IV décrit le principe de la méthode de géolocalisation proposée qui exploite la différence de point de vue des deux types d'images, puis présente les résultats issus du simulateur. La nécessité d'un référentiel temporel commun n'interviendra qu'au chapitre V, où la géolocalisation des nuages sera appliquée aux données réelles puis exploitée à des fins d'estimation et de prévision de la ressource solaire.

---

# Chapitre IV - Simulateur géométrique 3D : validation spatiale de la géolocalisation

---

Les observations du chapitre III ont notamment permis d'illustrer l'importance de la contribution des nuages sur la variabilité temporelle à hautes fréquences du rayonnement solaire. En particulier, leurs altitudes, leurs formes et leurs déplacements sont à l'origine de la majorité des variations brusques du rayonnement direct lorsque ceux-ci obstruent le disque et l'auréole solaires. La recherche d'une prévision précise de ces variations passe ainsi par une géolocalisation dynamique précise de la couverture nuageuse.

Ce chapitre se concentre sur l'une des complémentarités exprimées au chapitre précédent qui n'a pas encore été caractérisée par la communauté scientifique : la complémentarité des points de vue entre celui du capteur embarqué par le satellite géostationnaire et celui de la caméra hémisphérique, donnant accès à l'information 3D sur la couverture nuageuse. L'analyse de cette complémentarité proposée est purement géométrique, sans description du transfert radiatif dans l'atmosphère au travers des nuages.

Dans un premier temps, nous exprimons le choix de se concentrer sur l'estimation du rayonnement direct spatialement étendu au paragraphe IV.1. Puis, l'approche proposée qui estime la position et la forme des nuages sera décrite au paragraphe IV.2. Le reste de ce chapitre s'intéressera à la simulation de cette approche (*cf. IV.3 et IV.4*) à l'aide de situations nuageuses artificiellement générées. Leur but est d'extraire les informations d'erreur attendues selon le site considéré sur Terre (*cf. IV.5*) et d'appréhender les avantages et les limites d'une utilisation conjointe des deux types d'images.

## IV.1 L'estimation de l'ombre portée des nuages : une étape du processus prévisionnel

Le chapitre II a fait l'objet de la proposition de deux métriques caractérisant deux notions importantes pour l'analyse de la qualité à la fois de l'estimation et de la prévision du rayonnement solaire global : leur alignement temporel et la qualité de la restitution des rampes détectées.

Cette recherche explicite de restitution des variations rapides du rayonnement solaire global nécessite de considérer les deux composantes constitutives de ce rayonnement qui sont les rayonnements direct et diffus.

Les définitions exactes des rayonnements direct et diffus circumsolaire en fonction de l'angle solide d'ouverture sont proposées par Blanc et al. (2014). Dans notre cas le rayonnement direct est considéré comme celui mesuré par le capteur pyranométrique SPN1 des sites d'études : comme l'indiquent Badosa et al. (2014), le demi-angle d'ouverture équivalent pour cet instrument dépend de manière complexe de l'angle zénithal solaire et se situe entre  $5^\circ$  et  $25^\circ$  : il comprend donc une part non négligeable du diffus circumsolaire. Le rayonnement diffus est donc le complémentaire : c'est la part du rayonnement global ne provenant pas de cette zone circumsolaire.

Ces deux composantes du rayonnement global présentent des variabilités temporelles de nature très différentes, avec, la plupart du temps, une nette prépondérance de la variabilité du direct. En effet, même si le rayonnement direct provient d'une portion d'angle solide du ciel très restreinte (en moyenne de l'ordre de 5 pour 1000 des  $2\pi$  sr de l'hémisphère du ciel), il représente en moyenne, par condition de ciel clair, entre 70 % et 80 % du rayonnement global. Son support angulaire étant faible, il est très sensible aux passages nuageux dans la zone circumsolaire et il est fréquent d'observer une extinction totale très rapide de cette composante dans ces situations. À l'inverse, la composante diffuse correspond à l'intégrale de la luminance solaire sur tout l'hémisphère hors zone circumsolaire – cette zone étant déjà comptabilisée par le rayonnement direct. Son support d'intégration angulaire étant plus grand, avec une luminance du ciel nettement moins variable et moins importante que dans la zone circumsolaire, sa variabilité temporelle est généralement plus tempérée, sur des amplitudes plus faibles, à l'exception des phénomènes fugaces de réflexions multiples entre nuages (phénomène de suréclairage, cf. paragraphe III.4.5).

Afin d'illustrer cette prépondérance de la variabilité de la composante directe dans celle du rayonnement global, nous avons analysé et comparé, sur l'ensemble des mesures pyranométriques *in situ*, les amplitudes des gradients temporels de ces deux composantes au regard de celles du rayonnement global au pas de 1 min.

La figure 66 représente les trois fonctions de répartition de ces amplitudes en échelle semi-logarithmique. Cette représentation permet ainsi de constater que 97% des variations de rayonnement entre deux instants successifs au pas de temps 1 min ont une amplitude inférieure à  $16 \text{ W/m}^2$  pour le DHI, à  $45 \text{ W/m}^2$  pour le BHI et  $50 \text{ W/m}^2$  pour le GHI. Au-delà d'une amplitude de gradient de  $30 \text{ W/m}^2/\text{min}$ , les courbes de répartition correspondant au GHI et au BHI se confondent (celle du DHI étant alors à la valeur maximale de 100 %). Cela signifie que les fortes rampes de rayonnement global sont pour la grande majorité induites par celles de sa composante directe.

Cette contribution prépondérante du rayonnement direct aux variations significatives du rayonnement global se constate aussi en considérant la métrique ramp-MAD proposée et définie au paragraphe II.5.1. En effet, cette métrique appliquée entre le BHI et le GHI d'une part et entre le DHI et le GHI d'autre part vaut respectivement  $102 \text{ W.m}^{-2}.\text{h}^{-1}$  et  $203 \text{ W.m}^{-2}.\text{h}^{-1}$ . Ceci indique que le BHI suit plus fidèlement les rampes du GHI que le DHI.

Géométriquement, la variabilité de la composante directe du rayonnement induite par les passages nuageux correspond à celle des ombres portées des nuages au sol. C'est la raison pour laquelle, dans le cadre de cette thèse, on se focalisera sur la reconstruction 3D des nuages permettant la géolocalisation des ombres portées induites par ces derniers. La précision d'une telle géolocalisation des ombres a pour objectif d'affiner la précision temporelle de l'estimation et de la prévision de leurs effets sur le rayonnement direct.

La caractérisation complexe des effets 3D de la couverture nuageuse sur la composante diffuse du rayonnement solaire dépasse le cadre de cette thèse.

Nous nous appuyerons par la suite sur l'approximation validée par Oumbe et al. (2014) concernant le découplage possible entre les effets des paramètres de l'atmosphère sans nuages sur le rayonnement – liés notamment aux aérosols, à la vapeur d'eau ou encore l'ozone – et ceux induits par les nuages. En d'autres termes, on s'intéressera à caractériser les effets des nuages sur le rayonnement direct au travers de l'indice de ciel clair qui ne dépend que des paramètres des nuages et de l'éclairement par condition de ciel clair (*i.e.* sans nuages), lui-même dépendant des paramètres de l'atmosphère non nuageuse ainsi que de l'angle zénithal solaire.

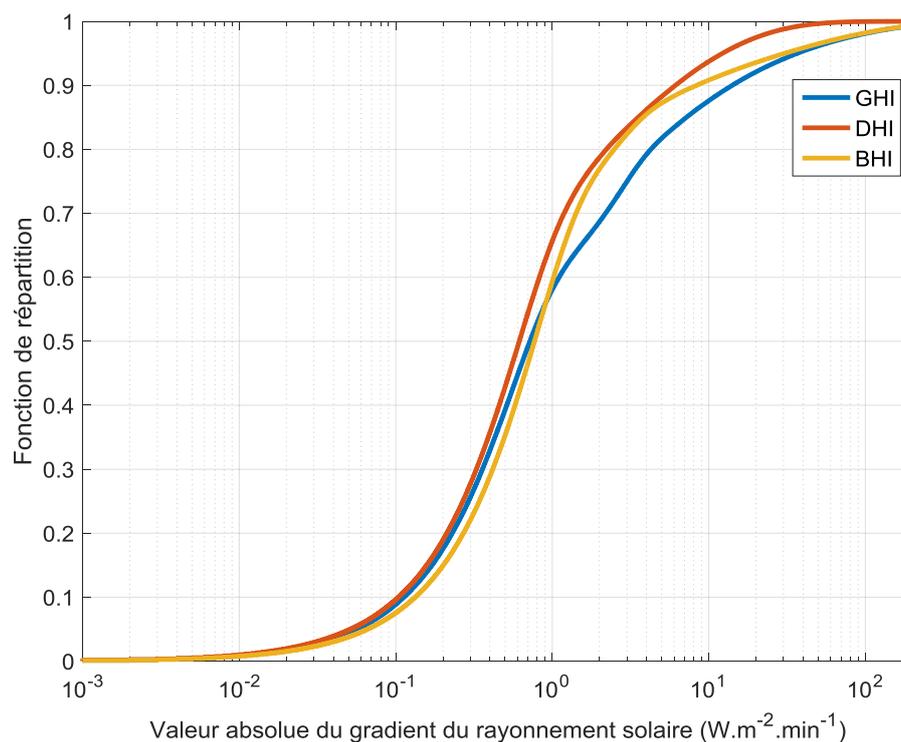


Figure 66 – Fonction de répartition des amplitudes des gradients des composantes diffuse (DHI) et directe (BHI) du rayonnement solaire global (GHI). L'axe des abscisses est échelle logarithmique.

## IV.2 Méthodes de géolocalisation des nuages par photogrammétrie

L'exploitation de la différence de point de vue de plusieurs observations d'une même scène à un instant donné par photogrammétrie est une approche qui a été envisagée à plusieurs reprises dans le contexte de la prévision solaire. À l'aide de multiples caméras hémisphériques proches de quelques centaines de mètres les unes des autres, Blanc et al. (2017) utilisent la légère différence de point de vue pour en déduire l'altitude des nuages par stéréoscopie, dans le but d'en extraire une prévision à très court terme de DNI. Cette méthode fait l'approximation de nuages plats sans épaisseur et apparie les textures semblables pour en estimer une altitude. Avec le même type de configuration instrumentale, Oberländer et al. (2015) ont proposé une méthode de géolocalisation des nuages par « sculpture de voxels » (*voxel carving*). Cette technique stéréoscopique a l'avantage de s'affranchir de cette hypothèse réductrice des nuages plats et tente plutôt d'en estimer leur épaisseur afin d'estimer l'ombre portée des nuages sur le sol. Elle consiste, après détection préalable des nuages sur chaque image, à recouper les lignes de visée des deux – ou plus – caméras hémisphériques afin d'en extraire des volumes englobant les nuages, tel qu'illustré sur la figure 67. Sans hypothèse sur l'altitude ni sur l'épaisseur des nuages en présence, cette technique permet d'augmenter très fortement la probabilité de détection des nuages. En effet, pour une détection parfaite des nuages sur les images hémisphériques, les volumes générés avec cette méthode contiennent assurément les nuages. À l'inverse, la probabilité de fausse alarme peut être grande, puisque toutes les formes et toutes les altitudes sont envisagées dans ces volumes des possibles.

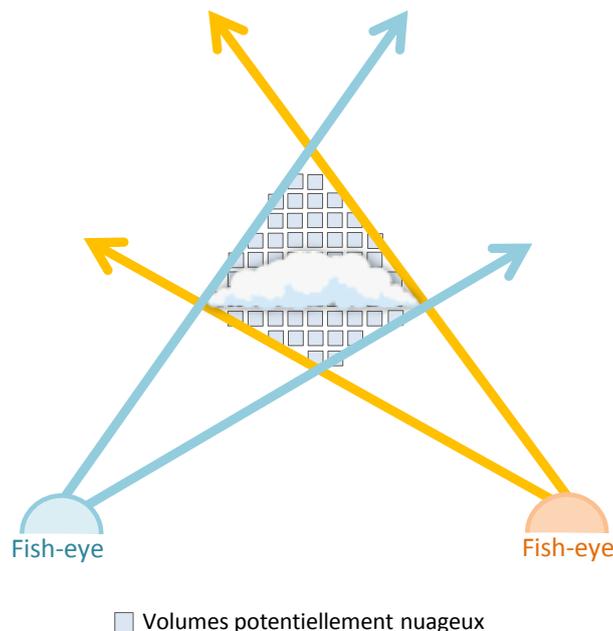


Figure 67 – Schéma explicatif de la méthode stéréoscopique de *voxel carving* proposée par Oberländer et al. (2015).

L'approche de *voxel carving* est reprise dans cette étude, mais cette fois l'intérêt se porte sur l'utilisation concomitante des images de la Terre issues du satellite géostationnaire et des images hémisphériques du ciel issues d'une unique caméra hémisphérique. Le schéma explicatif de cette méthode adaptée à ces données d'entrée est proposé sur la figure 68. Cette méthode appliquée aux images satellites et hémisphériques sera plus simplement désignée dans la suite par la « géolocalisation des nuages ».

L'utilisation de l'information satellitaire dans ce cadre a la particularité d'inclure une observation des nuages d'un tout autre point de vue que celle de la caméra hémisphérique. Le capteur SEVIRI embarqué dans le satellite MSG avec une focale de près de 5,4 m observe en effet les nuages « depuis le dessus » avec un angle d'incidence quasiment constant de  $53.6^\circ$  à l'échelle de nos deux sites, alors que la caméra hémisphérique informe sur la couverture nuageuse « depuis le dessous » dans quasiment toutes les directions du ciel, avec une focale de quelques millimètres. Compte tenu de ces grandes différences, les techniques de stéréoscopie sont plus complexes à mettre en œuvre puisque des faces différentes des nuages sont observées avec des capteurs différents.

Ainsi qu'illustrée par la figure 68, l'approche de géolocalisation des nuages consiste à « croiser » les volumes quasi cylindriques – appelés potentiellement nuageux – issus de la détection des nuages sur les images satellitaires avec les volumes coniques issus de la détection de ces derniers sur les images hémisphériques.

L'objectif des chapitres IV et V est d'évaluer cette méthode de géolocalisation des nuages – et de leurs ombres portées – vus conjointement par le satellite MSG et une caméra hémisphérique au sol, respectivement par la simulation géométrique 3D et son application sur des données réelles.

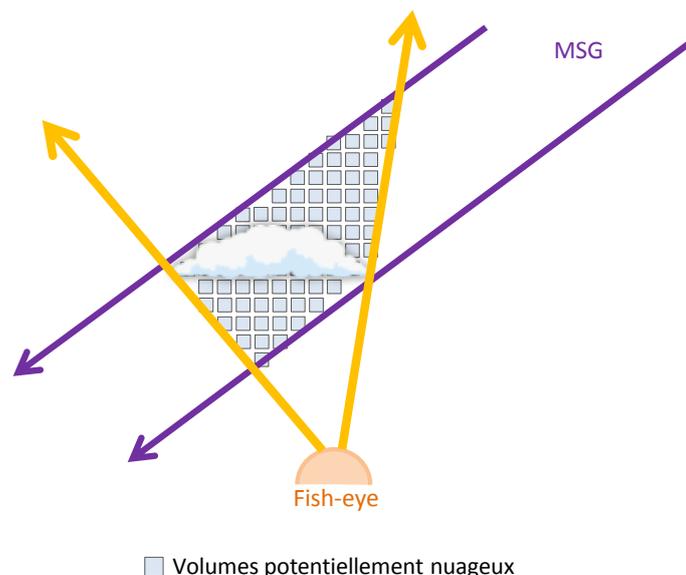


Figure 68 – Schéma explicatif de la méthode de géolocalisation des nuages exploitant les images issues d'une caméra hémisphérique et les images du satellite MSG.

Ces analyses considéreront à chaque fois trois niveaux graduels d'information sur la position des nuages pour lesquels les notations suivantes sont adoptées dans la suite de ce manuscrit :

- (1) 'SOL+SAT' fait référence à l'intersection des volumes potentiellement nuageux issus des segmentations des images satellites et hémisphériques ;
- (2) 'SOL+SAT+CTH' fait référence au niveau d'information (1) 'SOL+SAT' affiné avec l'information d'altitude du sommet des nuages ;
- (3) 'SOL+SAT+CTH+CBH' fait référence à 'SOL+SAT+CTH' affiné avec l'information d'altitude de la base des nuages (*Cloud base height*, CBH) ;

Enfin, le nombre de couches traversées dans la direction du Soleil sera noté  $N_L$ .

### IV.3 Objectifs d'une simulation de l'approche

Plusieurs facteurs rendent difficile l'évaluation des méthodes proposées sur les données réelles, principalement à cause de l'étape préliminaire de détection des nuages sur les images d'origine. Nous avons constaté au chapitre III la porosité de la frontière entre nuages, aérosols et trouble de l'atmosphère : la turbidité de l'atmosphère rend la perception des nuages, notamment à l'horizon des images hémisphériques, compliquée. Ces phénomènes complexifient donc toute tentative de segmentation des images hémisphériques afin d'en extraire les masses nuageuses. De la même manière, les effets de halo et de sur-exposition dans la zone circumsolaire affectent les résultats de la méthode implémentée. Du point de vue satellitaire, l'étape de segmentation s'avère aussi difficile. Notamment, les canaux du visible ne permettent pas toujours de distinguer les nuages d'un sol enneigé, et leur identification suppose une bonne connaissance de l'albédo du sol. Enfin, la limitation de résolution spatiale joue en défaveur de la qualité de la segmentation des images satellites. Ces particularités de l'utilisation de la méthode de géolocalisation seront exprimées plus en détail au chapitre V.

Afin de s'affranchir de ces difficultés, la simulation géométrique de l'approche est envisagée comme moyen d'évaluation *a priori*. Cet outil de simulation permet de tester la méthode expliquée au paragraphe IV.1 sur des configurations nuageuses artificiellement générées, pour un site et une configuration instrumentale donnée. Elle permet également d'évaluer la méthode en termes d'ombres portées des nuages au sol.

Cette évaluation est une information pertinente pour la prévision de la production photovoltaïque puisque l'ombrage de chaque cellule PV va jouer un rôle dans la production en sortie des panneaux qui les agrègent. Sur les deux sites à l'étude, caractériser les performances de ces méthodes en termes d'ombre portée n'est possible que par la simulation puisqu'aucun dispositif d'observation de l'ombre portée au sol n'est disponible. La plateforme solaire d'Almería dans le désert de Tabernas dans le sud de l'Espagne dispose d'une série de caméras fixées à une tour de 81 m de haut, dirigées vers le sol (Kuhn et al., 2017). Ce type d'instrumentation autorise une analyse de l'ombre estimée par

comparaison avec l'ombre effectivement observée par ces caméras, mais cette configuration est à ce jour unique, à notre connaissance. Dans ce chapitre, la simulation se contentera d'illustrer les résultats de la méthode pour le site n°1 de l'étude et sa configuration instrumentale associée. La conception de ce simulateur est expliquée au paragraphe suivant IV.4.

## IV.4 Description du simulateur

Le simulateur se décompose en deux parties distinctes : le générateur d'images qui a pour but de simuler ce que les deux imageurs – satellite et hémisphérique – sont censés percevoir d'une scène 3D donnée, et la méthode de géolocalisation des nuages en elle-même qui extrait les volumes « potentiellement nuageux » en se basant sur ces deux images simulées. La caractérisation des performances de la géolocalisation se fera par comparaison des ombres portées initiales et estimées.

### IV.4.1 Choix de la représentation 3D des nuages

Plusieurs représentations 3D peuvent être adoptées dans l'implémentation de la méthode de géolocalisation décrite au paragraphe IV.2. La plus utilisée dans le domaine de l'énergie solaire est la représentation en *voxels*. Il s'agit de quadriller l'espace 3D de manière régulière et avec une résolution prédéfinie, chaque cube élémentaire est appelé *voxel*. L'un des inconvénients majeurs de cette représentation est le temps de traitement d'une scène 3D qui peut être extrêmement élevé pour une petite taille de *voxels* et une grande scène analysée (*i.e.* un grand nombre de *voxels*).

Deux autres représentations peuvent également être envisagées :

- une représentation par nuage de points, où un certain nombre de points seront répartis aléatoirement dans chaque cube potentiellement nuageux, par une approche de type Monte-Carlo. Cette représentation permet de réduire considérablement le temps de calcul d'opérations de projection ;
- Une représentation en *metaballs* (« objets mous » en français) initialement proposé par Blinn (1982), où chaque « balle » est définie par une fonction de champ scalaire – c'est-à-dire qu'elle associe un scalaire à chaque point de l'espace – qui dépend généralement de son centre. Les volumes sont délimités en appliquant un seuil à la somme de ces fonctions. Dobashi et al. (1999) ont par exemple exploité cette représentation en *metaballs* afin de modéliser et d'animer les nuages détectés sur des images satellites.

La première méthode par *voxels* a été choisie pour la suite, pour sa simplicité de mise en œuvre. Le cadre prospectif de ces travaux de thèse autorise en effet d'avoir un temps d'exécution des algorithmes d'analyse relativement élevé car sans contrainte opérationnelle. Dans cette approche purement géométrique, chaque voxel contiendra ainsi une seule valeur binaire, valant 1 si le voxel correspond à un cube potentiellement nuageux, 0 sinon.

## IV.4.2 Génération d'images synthétiques à partir d'une situation nuageuse donnée

L'idée directrice de cette partie est de simuler ce qui serait perçu du point de vue satellitaire et du point de vue du sol pour une scène nuageuse artificielle donnée, dont la représentation est effectuée par l'attribution de la valeur 1 à un ensemble de *voxels* sur une étendue spatiale donnée.

### IV.4.2.a Situations nuageuses artificielles

La première étape du simulateur consiste à générer une scène nuageuse de manière artificielle. Dans un but d'obtenir une caractérisation des méthodes de géolocalisation des nuages la plus proche possible de cas réels typiques, nous avons envisagé la construction d'un ensemble de situations nuageuses représentatives du site étudié, tel que proposé par la figure 69. Différentes stratégies de génération de ces scènes peuvent être envisagées, nous décidons de regrouper les classes de nuages proposées par WMO<sup>9</sup> afin de simplifier l'analyse. Cette étape de génération de la scène nuageuse ne prend cependant pas en compte la transmittance des nuages ou des autres constituants atmosphériques. Elle est strictement géométrique et simplifiée à l'information de présence d'un nuage.

L'analyse exhaustive des situations nuageuses n'étant pas le sujet central de la thèse, seuls trois types de nuages ont été analysés :

- les nuages de type cumulus, dont l'intérêt porte sur l'estimation de leur position et taille ;
- les nuages de type stratus divisés, permettant de mettre en avant le gain de résolution obtenu grâce à la caméra hémisphérique ;
- les nuages stratus uniformes, qui seront principalement exploités dynamiquement afin de mettre l'accent sur la précision temporelle du passage d'un front nuageux devant le Soleil. Il peut être assimilé à un nuage de type cumulus, plat et de grande surface.

Les simulations peuvent s'effectuer pour un instant particulier – appelé analyse statique dans la suite – ou dans le cadre d'une analyse temporelle d'une séquence de situations simulées – appelé analyse dynamique. Seuls les nuages de type cumulus feront l'objet d'une analyse statique des méthodes de géolocalisation. Ces nuages seront à l'étude dans le cadre d'une évaluation dynamique du mouvement des nuages avec leur série temporelle d'ombrage du site associé.

---

<sup>9</sup> Atlas international des nuages de l'organisation météorologique mondiale – <https://cloudatlas.wmo.int>

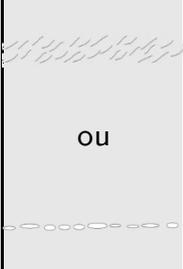
| Types de nuages  | Schéma   | Classes de nuages OMM   | Altitude                                    | Paramètres Monte-Carlo  |
|------------------|--|---|---|---|
| Cumulus          |   | Cumulus (Cu)<br>Cumulonimbus (Cb)   | 7 km<br> <br>0 km                           | <ul style="list-style-type: none"> <li>Position du nuage</li> <li>Taille du nuage</li> </ul>  |
| Stratus divisés  | <br>ou<br> | Cirrus (Ci)<br>Cirrocumulus (Cc)<br><br>Alto cumulus (Ac)<br><br>Stratocumulus (Sc) | 7 km<br> <br>5 km<br> <br>2 km<br> <br>0 km | <ul style="list-style-type: none"> <li>Position du centre de la grille</li> <li>Taille de chaque nuage</li> <li>Espaces entre les nuages</li> </ul> |
| Stratus uniforme |   | Cirrostratus (Cs)<br><br>Nimbostratus (Ns)<br><br>Altostratus (As)<br>Stratus (St)  | 7 km<br> <br>5 km<br> <br>2 km<br> <br>0 km | <ul style="list-style-type: none"> <li>Position du nuage</li> <li>Taille du nuage</li> </ul>  |

Figure 69 – Simplification de la classification des nuages de l'OMM vers 3 classes de nuages englobantes. [Source:](#) "Situations météorologiques typiques dans la région des Alpes" – MétéoSuisse.

#### IV.4.2.b Simulation des deux types d'images

##### Projections du volume nuageux dans le plan des images

À partir de chaque situation nuageuse artificiellement générée, l'ensemble des *voxels* du volume nuageux est projeté dans le plan des images. Avec la représentation des volumes nuageux en *voxels*, les coordonnées  $(x_{c_i}, y_{c_i}, z_{c_i})$  de chaque voxel  $c_i$  correspondent aux coordonnées de leur centre. Une étape préliminaire consiste donc à récupérer les coordonnées des sommets de chaque voxel, c'est-à-dire les coordonnées des points de l'isosurface du volume. Cet ensemble de points de l'espace 3D va être projeté dans le référentiel de la caméra hémisphérique et dans celui du satellite. Dans un premier temps, il est exprimé en coordonnées sphériques :

$$\forall (x, y, z) \in \mathbb{R}, \quad \begin{cases} az = \text{atan2}(y, x) \\ el = \text{atan2}(z, \sqrt{x^2 + y^2}) \\ r = \sqrt{x^2 + y^2 + z^2} \end{cases} \quad (31)$$

où  $az$  et  $el$  sont respectivement les angles d'azimut et d'élévation. Seuls ces deux angles sont exploités pour projeter ces points dans le référentiel de l'image hémisphérique en lignes  $l$  et colonnes  $c$  :

$$\begin{cases} l = \left(\frac{\pi}{2} - el\right) \cdot \cos(az) \\ c = \left(\frac{\pi}{2} - el\right) \cdot \sin(az) \end{cases} \quad (32)$$

Le principe de cette projection est illustré sur la figure 70.

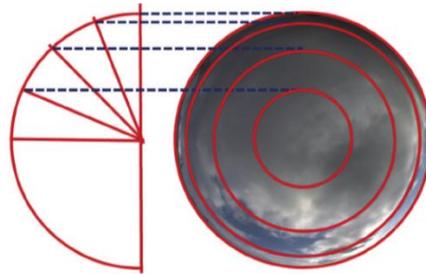


Figure 70 – Principe de la projection de l'espace 3D dans le plan image de la caméra hémisphérique.

Il est à noter que cette projection est une projection idéalisée dite « en équidistance angulaire » de la projection réelle de caméra hémisphérique<sup>10</sup>. Une calibration de la caméra utilisant *OcamCalib* (Scaramuzza et al., 2006) est décrite par Gauchet et al. (2012) et permet, avec une précision de l'ordre du dixième de degré, de se ramener à ce cas idéal.

Concernant la projection dans le repère de l'image satellite, les équations de passage du plan focal à la ligne de visée ou inversement indiquées par Müller et al. (2013) sont réutilisées. Ces équations intègrent l'altitude des nuages dans le référentiel géographique.

La figure 71 propose deux exemples de projection de scènes artificiellement générées dans le plan des deux capteurs satellite et hémisphérique. Les sous-figures (a) et (b) décrivent la projection des nuages de type *stratus divisés* respectivement de taille  $10 \times 10 \times 0.5 \text{ km}^3$  et  $1 \times 1 \times 0.5 \text{ km}^3$ . Hormis les résolutions spatiales des images qui ne sont pas encore prises en compte à ce stade, les caractéristiques de perception de la géométrie 3D par chaque image coïncident avec les observations qui en sont faites au chapitre III.

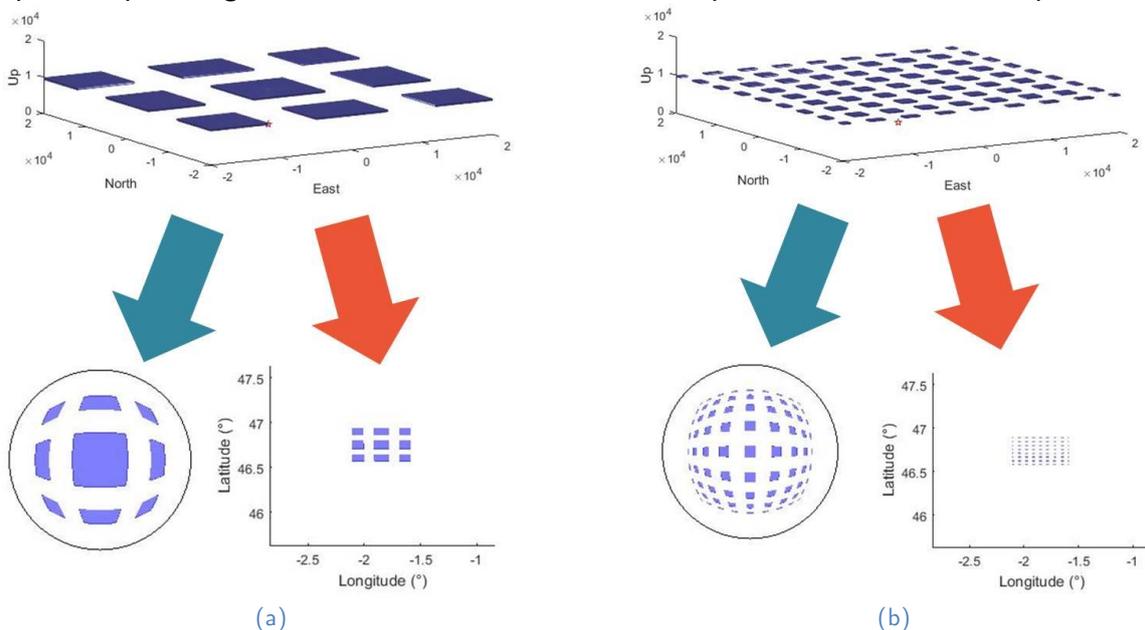


Figure 71 – Exemple de projection d'une scène 3D artificiellement générée dans le référentiel de la caméra hémisphérique et du satellite. La scène nuageuse (a) – respectivement (b) – est composée de nuages de taille  $10 \times 10 \times 0.5 \text{ km}^3$  – respect.  $1 \times 1 \times 0.5 \text{ km}^3$  – espacés de  $5 \text{ km}$  – respect.  $1.5 \text{ km}$  – dans les deux directions du plan horizontal.

<sup>10</sup> Cours de traitement des images de Télécom ParisTech dispensé par Henri Maître sur les objectifs photo Fish-Eye : [https://perso.telecom-paristech.fr/maitre/cours/projet\\_fish\\_eye/appareil\\_fish\\_eye.pdf](https://perso.telecom-paristech.fr/maitre/cours/projet_fish_eye/appareil_fish_eye.pdf)

### Prise en compte de la limite de résolution spatiale de chaque image

Afin de prendre en compte la limite de résolution spatiale des deux types d'images, la grille pixellique est définie en accord avec les caractéristiques de chaque image, puis les points de l'isosurface sont comptabilisés dans chaque pixel. Sur les deux situations précédentes, la prise en compte de la limite de résolution spatiale des deux images est proposée sur la figure 72. Cette limitation affecte particulièrement l'image satellite. À partir d'un certain espacement entre les nuages, la taille limitée des pixels empêche toute distinction des nuages proches en utilisant uniquement cette information satellitaire. La figure 72.b affiche un exemple de situation où l'information satellitaire apparaît comme une tâche floue non uniforme matérialisant la perception de cumulus de 1 km de côté et 0,5 km d'épaisseur, répartis en strate horizontale, proches de 1,5 km les uns des autres.

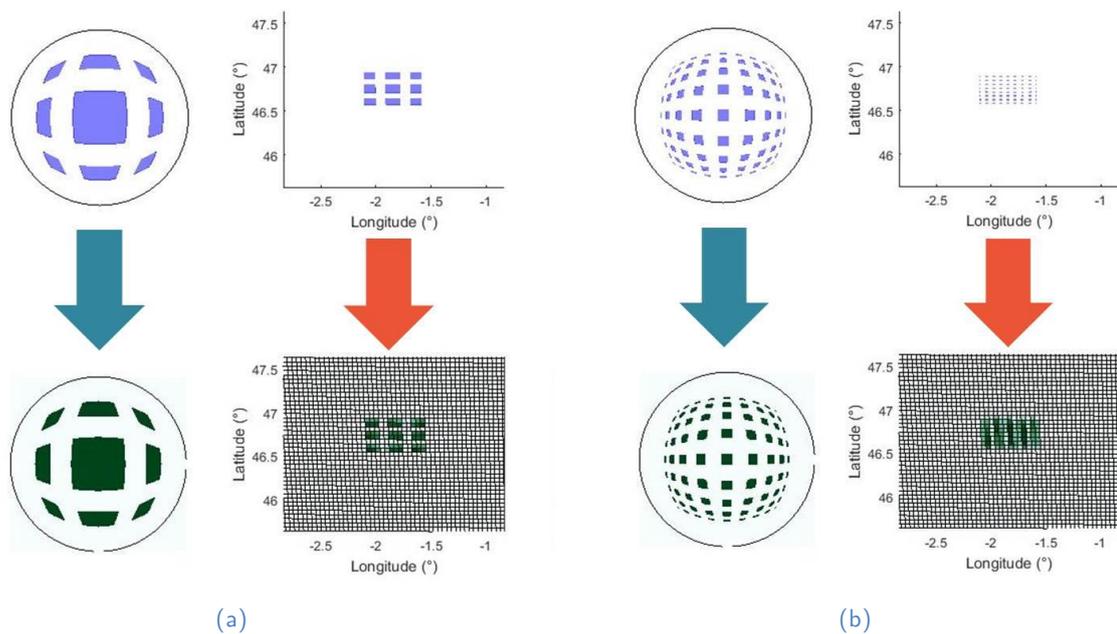


Figure 72 – Mêmes exemples que la figure 71 dont la limite de résolution spatiale des deux types d'images est prise en compte.

### **IV.4.3 Application de la méthode de géolocalisation des nuages en 3D**

#### **IV.4.3.a Segmentation des images**

À ce stade de la simulation, nous disposons de deux images synthétiques – satellite et hémisphérique – correspondant à une scène initiale artificiellement générée (cf. IV.4.2). Comme formulé au paragraphe IV.2, les différents niveaux d'information de la méthode de géolocalisation considérés passent par une étape préliminaire d'identification des masses nuageuses dans les deux types d'images. Dans ce chapitre, les images satellites et hémisphériques sont des images synthétiques dont la seule information qu'elles contiennent correspond à la présence ou non d'un nuage. La segmentation de ces images est donc triviale, tout pixel d'une valeur supérieure à zéro indique la présence d'un nuage – puisque l'on compte le nombre de points de l'isosurface arrivant sur chaque pixel. Effectuer la segmentation des deux types d'images sur les données réelles n'est pas aussi trivial et son résultat ne sera jamais parfait. Cet aspect est traité au paragraphe V.2.1. Ainsi, cette simulation géométrique de la géolocalisation 3D des nuages permet de s'affranchir des erreurs de segmentation des deux types d'images, et de se concentrer sur les performances de la méthode pour un site et une instrumentation donnée.

#### **IV.4.3.b Projection de l'information issue des images dans le volume 3D**

Suite à la segmentation des deux types d'images, aucune hypothèse sur l'altitude des nuages n'est prise pour l'instant – ce qui correspond au niveau d'information (1) 'SOL+SAT' de la méthode de géolocalisation décrite au paragraphe IV.2. Cette information de présence des nuages est alors projetée dans l'espace 3D selon chaque point de vue : pour chaque pixel identifié comme nuageux, tout voxel dont le centre est inclus dans le cône de visée se voit attribuer la valeur 1. De cette étape résulte la génération d'un « cône » de *voxels* – des deux points de vue – partant de chaque imageur et contenant le nuage. Ces volumes sont appelés « volumes potentiellement nuageux » puisqu'ils englobent le volume nuageux, mais peuvent contenir de nombreux *voxels* non nuageux.

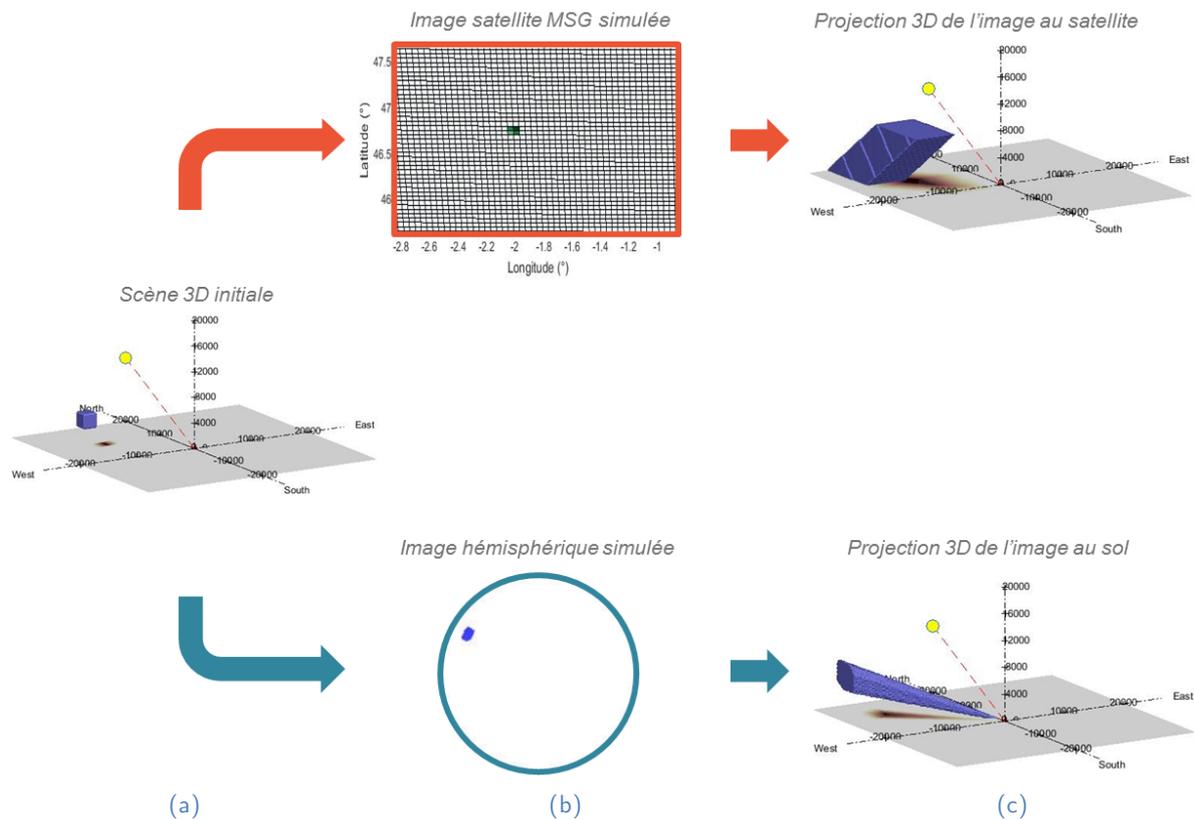
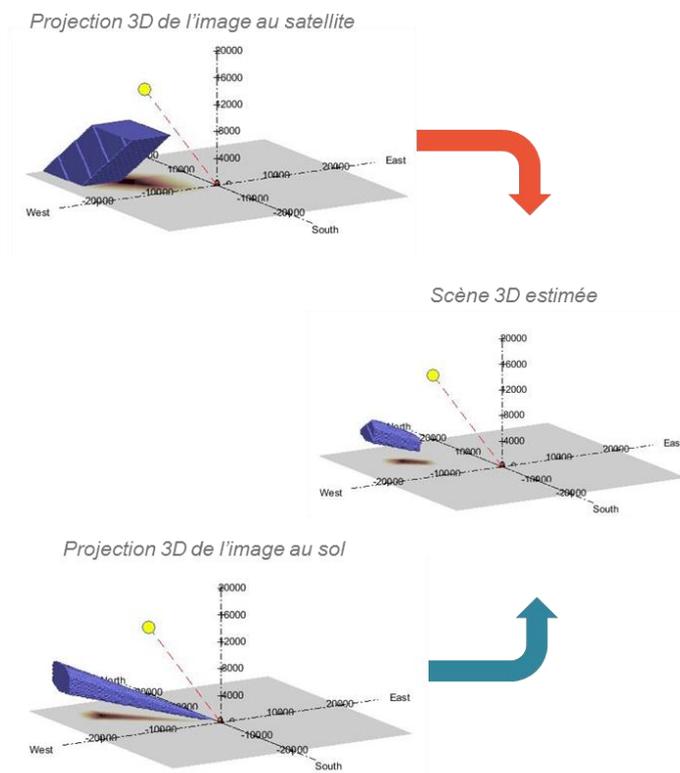


Figure 73 – Vue schématique des premières étapes du simulateur : (a) la génération d'une scène nuageuse – ici de type *cumulus* et de taille  $1 \times 1 \times 1 \text{ km}^3$ , (b) la simulation de l'image satellite (en haut) et hémisphérique (en bas) correspondant à cette scène, et (c) la projection de ces informations dans l'espace 3D en l'absence d'hypothèse sur l'altitude des nuages.

En ne considérant que l'information satellitaire, l'ensemble des positions possibles d'un nuage d'après son identification dans l'image satellite est projeté dans l'espace 3D, ce qui va générer un « cône » (quasi cylindrique) de *voxels* partant du satellite, englobant le nuage de la scène initiale et atteignant le sol. Comme illustré sur la partie haute de la figure 73, ce volume potentiellement nuageux se basant strictement sur l'information satellitaire et généré à partir de nuages synthétiques rectangulaires, s'apparente au sol à un parallélépipède. Ce volume à base rectangulaire provient de l'étape de segmentation de l'image satellite, et le parallélisme apparent des rayons d'incidence au sol vient de la prépondérance de la distance Terre-satellite devant les distances considérées (d'un rapport  $\sim 3000$ ). La partie basse de cette figure illustre le volume potentiellement nuageux se basant uniquement sur l'information de l'image hémisphérique au sol. Le « cône » contenant l'ensemble des positions possibles du nuage initial part de la caméra hémisphérique et se dirige dans la direction du nuage identifié sur l'image correspondante. Enfin, les trois volumes nuageux illustrés sur la figure 73 sont représentés avec une position du Soleil arbitraire (disque jaune) et leur ombre portée au sol correspondante, dont la colorimétrie est indexée sur le nombre de couches nuageuses traversées dans la direction du Soleil  $N_L$ .

### IV.4.3.c Intersection des volumes

À partir des deux volumes potentiellement nuageux issus respectivement de l'information satellitaire et de l'information au sol, l'opération de fusion d'information proposée consiste simplement à réaliser l'intersection de ces deux volumes. En effet, chaque cône potentiellement nuageux peut être vu comme l'ensemble des solutions de la position des nuages issu d'un seul point de vue. L'intersection de ces ensembles de positions possibles pour les nuages va affiner le volume potentiellement nuageux et donc générer une première estimation de la géolocalisation des nuages en 3D puisque les deux imageurs observent la même scène nuageuse. La figure 74 décrit cette opération d'intersection des volumes potentiellement nuageux. Notons sur cet exemple la réduction considérable des volumes potentiellement nuageux, qui se traduit au sol par une réduction de l'ombre portée, en se rapprochant ainsi de celle de la scène initiale.



#### IV.4.3.d Incorporation d'informations partielles de géolocalisation des nuages

Le volume potentiellement nuageux peut être grand comparé au volume du nuage initial. La comparaison de ces volumes nuageux potentiels et initiaux est, dans notre cas, réalisée par comparaison de leurs ombres portées au sol respectives.

Ce simulateur peut également servir à caractériser l'apport d'un ajout d'autres informations partielles sur la position des nuages à la méthode de géolocalisation de ceux-ci.

Le niveau d'information (2) 'SOL+SAT+CTH' exprimée au paragraphe IV.2 reprend la méthode telle qu'exprimée jusqu'à présent en y intégrant en outre l'information d'altitude du sommet des nuages. Des estimations de l'altitude du sommet des nuages sont effectivement proposées par divers organismes comme EUMETSAT (Stengel et al., 2014) ou KNMI (Roebeling et al., 2006) mais ont une incertitude de l'ordre de 2 km en écart type. Une description des méthodes d'estimation de l'altitude du sommet des nuages est proposée au paragraphe V.2.2.a. Dans le cadre de ce simulateur, l'hypothèse d'une connaissance parfaite de l'altitude du sommet des nuages est adoptée dans un premier temps, mais l'incertitude de cette estimation peut être prise en compte afin d'en déterminer son impact sur la reconstruction de l'ombre portée au sol.

L'information de niveau (3) 'SOL+SAT+CTH+CBH' reprend celle de niveau (2) 'SOL+SAT+CTH' en considérant parfaitement connue en outre l'information de l'altitude de la base des nuages. Une méthode sera proposée dans le chapitre V (*cf.* paragraphe V.2.2.b) afin d'extraire du croisement des dynamiques des deux types d'images une estimation de cette altitude de la base des nuages.

Une illustration de la méthode de géolocalisation exploitant les trois niveaux d'information est proposée à la figure 75.

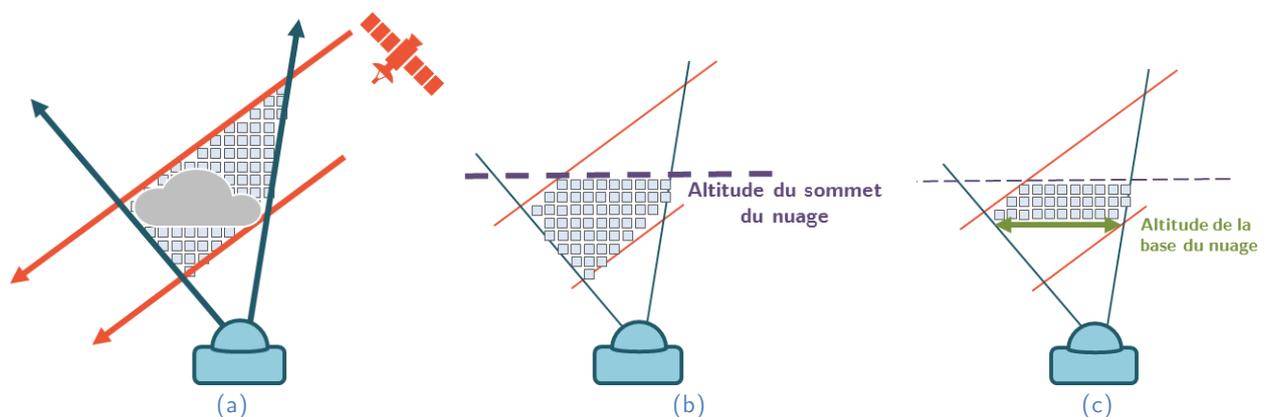


Figure 75 – Schéma des trois niveaux d'information de la méthode de géolocalisation proposée intégrant graduellement les informations sur l'altitude du sommet des nuages et de la base des nuages. Les sous-figures (a), (b) et (c) correspondent respectivement aux schémas explicatifs des niveaux d'information (1) 'SOL+SAT', (2) 'SOL+SAT+CTH', (3) 'SOL+SAT+CTH+CBH' de la méthode de géolocalisation des nuages.

#### IV.4.4 Comparaison des ombres portées au sol

Afin de caractériser les performances de la méthode proposée de géolocalisation 3D des nuages en fonction des trois niveaux d'information, l'ombre portée au sol s'avère être une information pertinente à comparer du point de vue de l'application. En effet, ce n'est pas tant la reconstruction parfaite du volume nuageux qui importe que son ombre portée au sol, dont l'estimation facilite ensuite la transposition en production instantanée d'une installation photovoltaïque par la connaissance des modules ombragés.

L'étape de comparaison des ombres portées au sol entre le volume nuageux d'origine artificiellement généré et celui résultant de l'intersection des volumes potentiellement nuageux est illustrée sur la figure 76. Une méthode simple d'extraction de l'ombre portée au sol est adoptée ici : il s'agit de sommer les couches de *voxels* potentiellement nuageux dans la direction du Soleil. Le résultat est donc une fraction de couches traversées  $N_L$  en direction du Soleil – ou un entier si la méthode d'interpolation au plus proche voisin est adoptée. Afin de quantifier la capacité de la méthode à géolocaliser l'ombre portée des nuages, une matrice de confusion (cf. tableau 5 et figure 76.e) est calculée entre les deux présences d'ombres d'origine (cf. figure 76.c) et de l'estimation (cf. figure 76.d). Dans la partie statique de l'analyse de ce simulateur (cf. paragraphe IV.5.1), il sera question en effet de présence d'ombres, c'est-à-dire du résultat logique de la condition ( $N_L > 0$ ). La partie dynamique de l'analyse de ce simulateur (cf. paragraphe IV.5.1) utilisera à la fois l'information non-binaire d'ombre et l'information binaire de présence d'ombre.

|            |                   | Scène d'origine   |                   |
|------------|-------------------|-------------------|-------------------|
|            |                   | Pixel ombragé     | Pixel non ombragé |
| Estimation | Pixel ombragé     | TP : Vrai positif | FP : Faux positif |
|            | Pixel non ombragé | FN : Faux négatif | TN : Vrai négatif |

Tableau 5 - Matrice de confusion pour l'analyse de la présence d'ombre estimée comparativement à la présence d'ombre de la scène d'origine.

À partir de cette matrice de confusion, plusieurs métriques peuvent être calculées :

- Taux de vrais positifs ou sensibilité (TPR, *true positive rate*) :

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (33)$$

- Taux de faux positifs (FPR, *false positive rate*) :

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (34)$$

- La métrique de précision, qui est en quelque sorte un taux de bonnes réponses :

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (35)$$

- Le coefficient de corrélation de Matthews (1975) qui s'avère être particulièrement approprié pour décrire des classes binaires de tailles très différentes (Boughorbel et al., 2017), contrairement à la métrique ACC. Il trouve ses valeurs entre -1 (désaccord avec l'observation) et 1 (accord avec l'observation) :

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (36)$$

Nous avons sélectionné les métriques ACC et MCC afin de quantifier les performances de la géolocalisation de l'ombre portée.

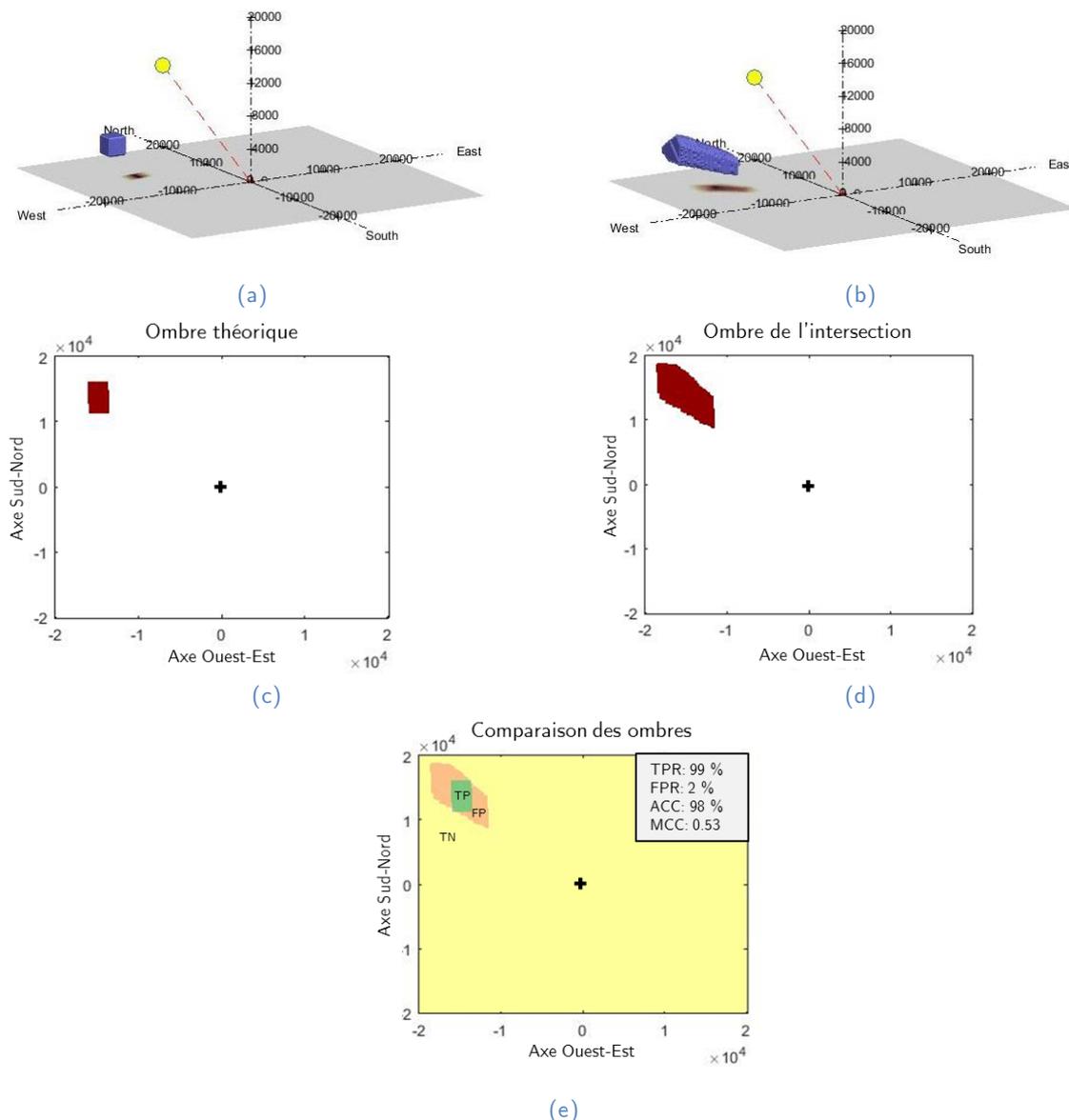


Figure 76 – Étape de comparaison des ombres portées : la matrice de confusion (e) est calculée entre les présences d'ombres (c) de la scène d'origine (a) et la présence d'ombres (d) de la scène estimée (b).

## IV.5 Analyse des résultats

### IV.5.1 Analyse statique

Désormais, pour chaque scène nuageuse artificielle et pour chaque position angulaire du Soleil, le simulateur est en mesure de calculer une valeur de la métrique ACC et une valeur de MCC, caractérisant la performance statique en termes de présence d'ombre portée de la méthode de géolocalisation des nuages considérée. Cependant, un grand nombre de configurations nuageuses peut être proposé en entrée du simulateur.

Nous nous limiterons à une analyse sur des géométries de nuages simples sous la forme de parallélépipèdes rectangles, faisant référence à la classe cumulus évoquée à la figure 69. Là encore, les paramètres de taille et de position du centre du nuage peuvent avoir tout type de valeurs. Ainsi, dans la suite de ce paragraphe, l'extraction des résultats se focalisera sur trois tailles de nuages –  $1 \times 1 \times 0.5 \text{ km}^3$ ,  $5 \times 5 \times 1 \text{ km}^3$  et  $7 \times 7 \times 1 \text{ km}^3$  – sur lesquelles les trois niveaux d'information présentés au paragraphe IV.2 seront appliqués.

Pour chacune de ces trois tailles de nuages de type cumulus, la méthode de tirage aléatoire de type Monte-Carlo est appliquée sous des lois de probabilité uniformes sur la position du centre du nuage dans les intervalles  $[-20 : 20 \text{ km}]$ ,  $[-20 : 20 \text{ km}]$  et  $[0 : 7 \text{ km}]$  respectivement en X, Y et Z (resp. directions Est, Nord et Zénith), ainsi que sur la position du Soleil. Cependant, pour obtenir des résultats réalistes, la position du Soleil est choisie aléatoirement parmi les couples d'angles d'azimut et d'élévation solaires positifs de l'année 2015 sur le site étudié. 10000 positions du centre du nuage et du Soleil sont tirées pour chaque type de nuage.

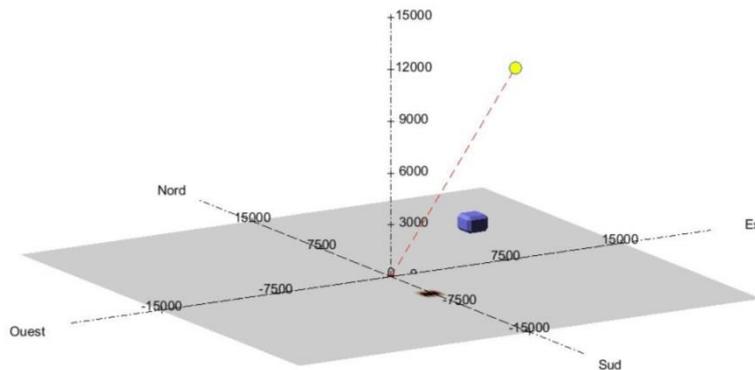


Figure 77 – Nuage de taille  $1 \times 1 \times 0.5 \text{ km}^3$  artificiellement généré dont la position de son centre ainsi que la position du Soleil sont soumises à une analyse de Monte-Carlo.

Tout d'abord, le simulateur est appliqué au nuage de taille  $1 \times 1 \times 0.5 \text{ km}^3$ , tel que représenté sur la figure 77, et en considérant uniquement le niveau d'information (1) 'SOL+SAT' dans un premier temps. Ce nuage a la particularité d'être perçu du point de vue satellitaire avec une taille inférieure à celle d'un pixel de l'image satellite (qui vaut  $3 \times 3 \text{ km}^2$  au nadir de MSG). Dans cette situation, l'image satellite apporte donc une information très grossière de géolocalisation alors que l'image hémisphérique est plus susceptible d'affiner la délimitation du nuage. L'information satellitaire est d'autant plus grossière si le nuage est localisé à la frontière de deux ou quatre pixels.

Le tirage aléatoire décrit précédemment donne ainsi 10000 valeurs de ACC et de MCC. Afin de mettre en évidence une éventuelle dépendance des résultats vis-à-vis de la position du nuage dans la scène, la métrique ACC est représentée dans le repère de l'image hémisphérique sur la figure 78, en fonction de la position du centre du nuage correspondant.

Ainsi, le centre de la sous-figure (e) correspond au Zénith et le symbole  $\square$  indique la position angulaire du satellite du point de vue du sol ( $177,6^\circ$  d'azimut et  $53,6^\circ$  d'angle zénithal). Les points cardinaux sont également indiqués. Chaque pixel contient la moyenne des valeurs de ACC dont le centre du nuage correspondant se trouve dans ce pixel, c'est-à-dire dans cet intervalle de directions.

Les valeurs de précision affichées sont dépendantes de la taille de chaque classe et donc de la grille sur laquelle est représentée chaque ombre portée. Ainsi, les valeurs de ACC peuvent être artificiellement améliorées en augmentant la taille de la grille pour la comparaison des ombres portées. Les valeurs de ACC de la figure 78 étant calculées sur la même grille, le comportement relatif à la position du centre du nuage est donc plus important ici que la valeur d'un pixel isolé. Il en sera de même pour l'affichage des valeurs des métriques ACC et MCC appliquées sur les autres tailles de nuage, la comparaison entre graphes est possible mais en aucun cas ces valeurs peuvent être comparées à des analyses régies par d'autres paramètres de grille.

Le premier constat est que la performance de la méthode de géolocalisation des nuages dépend de l'emplacement angulaire du nuage. En particulier, la tache verte diffuse autour de la direction du satellite suggère que la performance en termes de présence d'ombres portées dépend de l'écart angulaire entre la position du centre du nuage et la position du satellite. Afin de faciliter la compréhension de ce résultat, quatre scènes sont également affichées sur la figure 78, représentant l'estimation de la géolocalisation 3D d'un nuage de taille  $1 \times 1 \times 0.5 \text{ km}^3$  dont le centre a un angle d'élévation de  $30^\circ$ . La sous-figure de gauche (respectivement du haut, de droite et du bas) correspond à un nuage situé à l'Ouest (respectivement au Nord, à l'Est et au Sud). La sous-figure du bas (figure 78.d) correspond à la position d'un nuage très proche de la direction du satellite. Dans ce cas, le volume potentiellement nuageux du point de vue de l'image hémisphérique est dirigé vers le satellite, son intersection avec le volume potentiellement nuageux du point de vue satellitaire va donc être bien plus grande que pour les autres positions du nuage dans l'espace. L'ombre portée résultante est donc elle aussi très grande en moyenne en comparaison à l'ombre portée de la scène initiale.

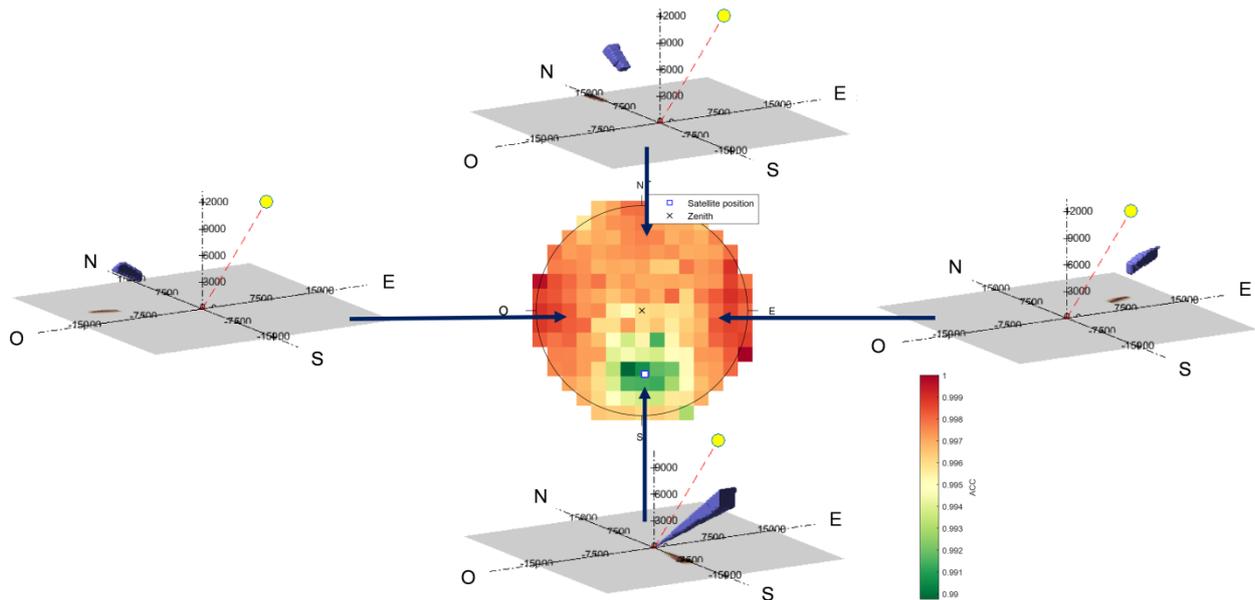


Figure 78 – Illustration et explication des résultats de la méthode de Monte-Carlo pour l'analyse du simulateur sur 10000 points avec la méthode de géolocalisation utilisant le niveau d'information (1) 'SOL+SAT' et sur un nuage de taille  $1 \times 1 \times 0.5 \text{ km}^3$ . La position du nuage et la position réaliste du Soleil sont tirées aléatoirement. Au centre, la sous-figure (e) de résultats de la métrique ACC du point de vue du sol, chaque pixel correspondant à la précision moyenne de géolocalisation de l'ombre portée d'un nuage à cet emplacement. Les sous-figures (a), (b), (c), et (d) illustrent les types de résultats obtenus après application de la géolocalisation utilisant le niveau d'information (1) 'SOL+SAT' pour un nuage situé respectivement à l'Ouest, au Nord, à l'Est et au Sud du site étudié, à une élévation de  $30^\circ$ .

Un autre effet notable sur la figure 78 est la teinte légèrement plus rouge à l'Est et à l'Ouest qu'au Nord pour des angles d'élévation faibles. Cet effet est principalement visible pour ces nuages de taille inférieure au pixel de l'image satellite. L'une des explications à cette caractéristique est que la taille d'un pixel satellitaire n'a pas la même dimension selon l'axe Ouest-Est (environ 3 km) que selon l'axe Sud-Nord (environ 5 km). Ainsi, l'intersection des volumes potentiellement nuageux pour des nuages situés à l'Est et à l'Ouest profitera d'une précision légèrement meilleure que pour des nuages situés au Sud ou au Nord.

Cette caractéristique est bien entendu dépendante de la position du site et du satellite considéré. Pour un site dont la latitude serait proche de zéro et une longitude proche de  $45^\circ$ , par exemple en Somalie, la figure de précision attendue serait proche d'une rotation de  $90^\circ$  dans le sens horaire de la figure 78, le satellite MSG se trouvant alors à l'Ouest du site et la longueur de 5 km des pixels satellites suivant l'axe Ouest-Est.

La même analyse de Monte-Carlo est appliquée au simulateur en utilisant séparément les trois premiers niveaux d'information de la méthode de géolocalisation des nuages, ainsi qu'en l'appliquant sur les nuages de dimensions  $1 \times 1 \times 0.5 \text{ km}^3$  (cf. figure 79),  $5 \times 5 \times 1 \text{ km}^3$  (cf. figure 80) et  $7 \times 7 \times 1 \text{ km}^3$  (cf. figure 81). Les sous-figures (a), (b) et (c) de ces trois figures correspondent aux résultats de la métrique de précision ACC, et les sous-figures (d), (e) et (f) aux résultats de la métrique de corrélation MCC respectivement pour les informations de niveaux (1) 'SOL+SAT', (2) 'SOL+SAT+CTH' et (3)

'SOL+SAT+CTH+CBH'. Les performances de l'information de niveau (3) sont en moyenne meilleures que celles de niveau (2), qui sont elles-mêmes meilleures que celles de niveau (1).

Les caractéristiques de l'information de niveau (1) d'une moins bonne précision et d'une moins bonne corrélation dans la zone autour du satellite ne sont respectivement peu et pas visibles pour les informations de niveaux (2) et (3). Les sous-figures (b) et (e) des figures 79, 80 et 81 affichent une moins grande dépendance des performances à la position du satellite, mais elle reste tout de même présente à cause de la partie inférieure du « cône » potentiellement nuageux. Cette caractéristique est absente des sous-figures (c) et (f). Sur les sous-figures (c), les précisions ACC de la géolocalisation sont très élevées dans toutes les directions, excepté pour les très faibles élévations, ce qui s'explique par le fait que les faces Nord, Sud, Est ou Ouest du nuage sont légèrement moins bien estimées que pour un nuage situé au Zénith. Cette caractéristique est d'autant plus marquée sur la corrélation MCC affichée sur les sous-figures (f), où le différentiel de corrélation entre la direction zénithale et les directions proches de l'horizon est plus grande pour les nuages de taille  $1 \times 1 \times 0.5 \text{ km}^3$ . En effet, la mauvaise estimation des bordures des nuages est négligeable pour les nuages de grande taille, alors qu'elle est en proportion plus impactante pour les nuages de petite taille. D'où des résultats plus isotropes pour les sous-figures 81.d, 81.e et 81.f de corrélation MCC.

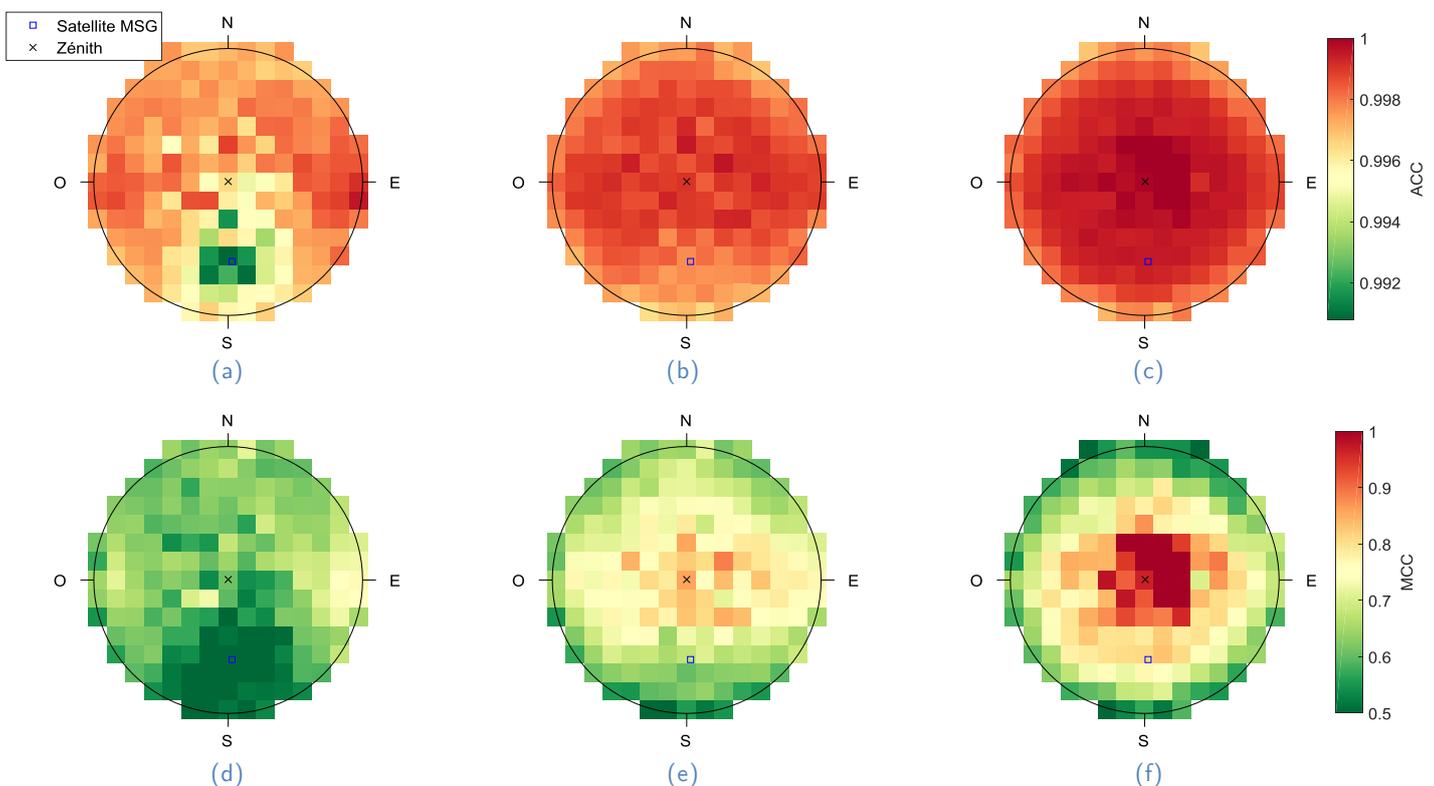


Figure 79 – Figures de résultats des performances de la méthode de géolocalisation utilisant les niveaux d'information (1) 'SOL+SAT', (2) 'SOL+SAT+CTH' et (3) 'SOL+SAT+CTH+CBH' appliquée à un nuage de dimension  $1 \times 1 \times 0.5 \text{ km}^3$  en termes de précision (ACC) et de corrélation (MCC). La colorimétrie est identique pour chaque métrique afin de faciliter la comparaison inter-figures.

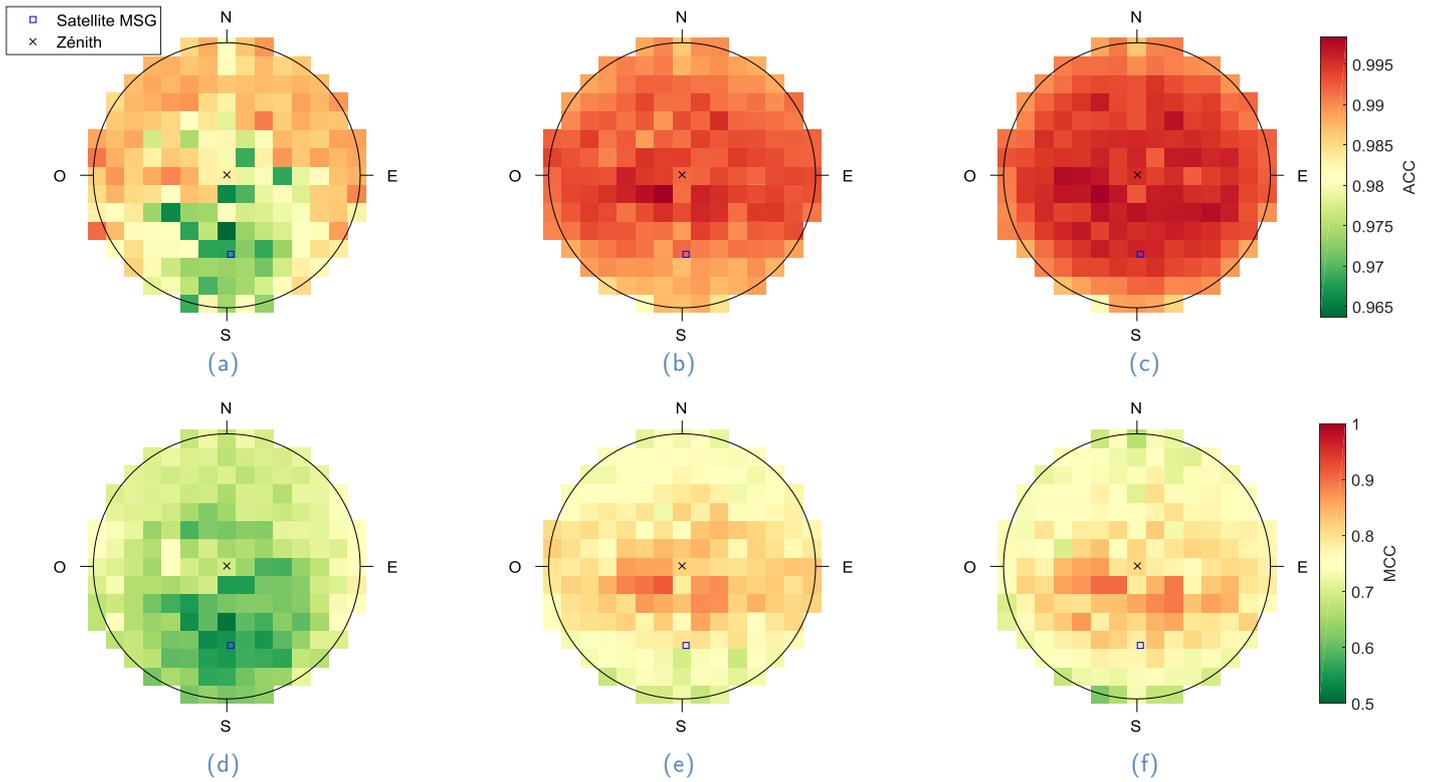


Figure 80 – Figures de résultats des performances de la méthode de géolocalisation utilisant les niveaux d'information (1) 'SOL+SAT', (2) 'SOL+SAT+CTH' et (3) 'SOL+SAT+CTH+CBH' appliquée à un nuage de dimension  $5 \times 5 \times 1 \text{ km}^3$  en termes de précision (ACC) et de corrélation (MCC). La colorimétrie est identique pour chaque métrique afin de faciliter la comparaison inter-figures.

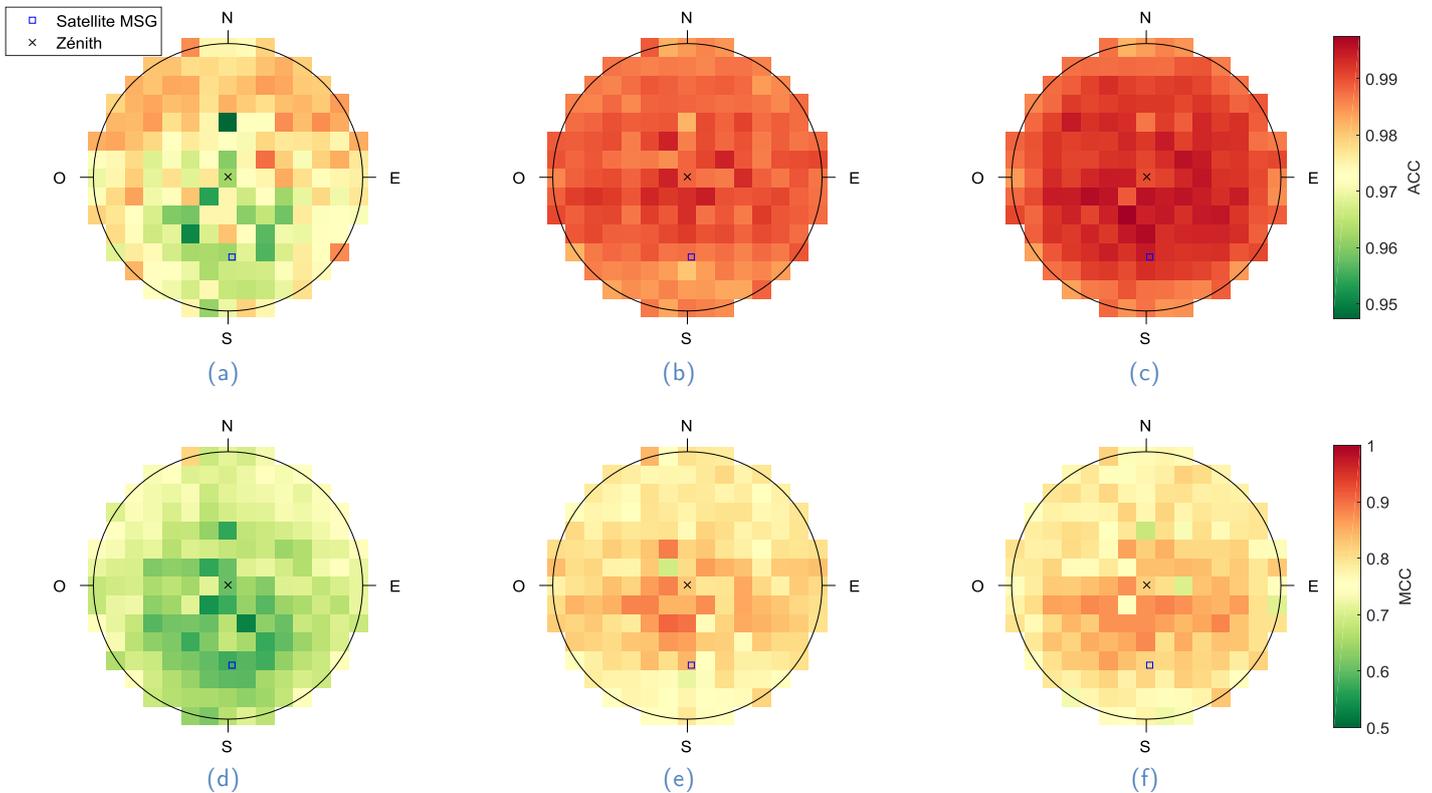


Figure 81 – Figures de résultats des performances de la méthode de géolocalisation utilisant les niveaux d'information (1) 'SOL+SAT', (2) 'SOL+SAT+CTH' et (3) 'SOL+SAT+CTH+CBH' appliquée à un nuage de dimension  $7 \times 7 \times 1 \text{ km}^3$  en termes de précision (ACC) et de corrélation (MCC). La colorimétrie est identique pour chaque métrique afin de faciliter la comparaison inter-figures.

Finalement, cette analyse statique des méthodes de géolocalisation 3D des nuages permet d'identifier certaines caractéristiques des résultats qui peuvent être attendus. En particulier la dépendance des performances du niveau d'information (1) 'SOL+SAT' à la position angulaire relative du nuage par rapport à celle du satellite a été constatée. Cette analyse suppose une identification parfaite des nuages dans les différentes images, pour un instant donné. Aussi, une amélioration de l'estimation de l'ombre segmentée des nuages est constatée lorsque les altitudes du sommet et de la base des nuages sont connus. Le paragraphe suivant poursuivra cette analyse en exploitant ce simulateur sur des nuages artificiels soumis à un mouvement arbitraire.

### IV.5.1 Analyse dynamique

L'extraction de résultats à partir du simulateur peut également s'effectuer de manière dynamique afin d'analyser les performances d'une méthode de géolocalisation 3D des nuages par comparaison de séries temporelles. Pour l'obtention de telles séries temporelles, un mouvement arbitraire est appliqué à l'une des configurations nuageuses choisies. À chaque pas de temps, il est alors possible de comparer la scène nuageuse d'origine avec celle estimée par la méthode. Cette fois, l'estimation du nombre de couches  $N_L$  potentiellement nuageuses traversées dans la direction du Soleil sera analysée dynamiquement, ce qui permet de tester la méthode dans le cas où l'identification des nuages dans les images est parfaite, le mouvement des nuages est purement advectif sans déformation de ceux-ci et sans considération sur l'étape de conversion de l'information d'ombrage en rayonnement solaire quantifié (cf. partie V.3).

#### Application d'un mouvement arbitraire aux nuages artificiellement générés

De la même manière que pour la génération des situations nuageuses, n'importe quel mouvement peut leur être appliqué. Nous avons choisi d'utiliser un mouvement simple à appliquer sur les situations nuageuses, tel que représenté sur la figure 82 : il suit une trajectoire rectiligne dans le plan horizontal d'Ouest en Est avec une vitesse de 20 km/h.

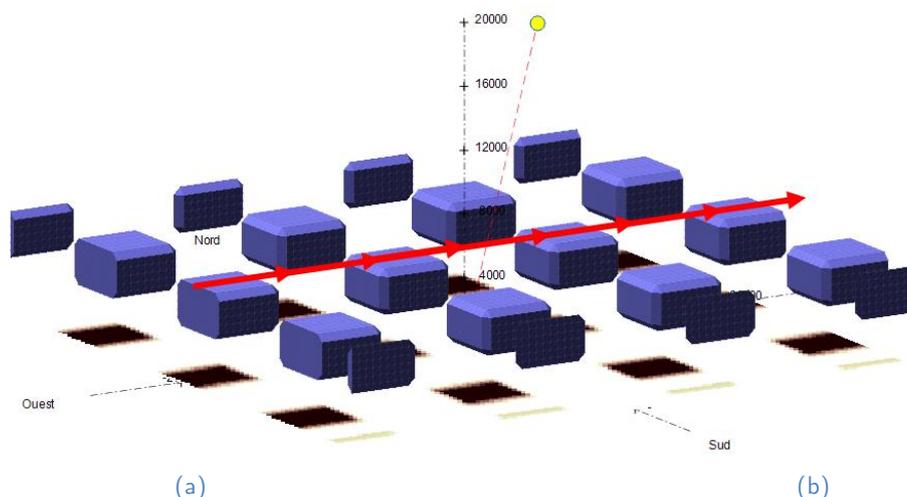


Figure 82 – Mouvements choisis pour l'application du simulateur dans le cadre d'une analyse dynamique.

### Sélection des scènes nuageuses à faire évoluer dynamiquement

Comme dans le cas statique, il existe une grande diversité de scènes nuageuses à insérer en entrée de ce simulateur. Les nuages stratocumulus de taille  $1 \times 1 \times 0.5 \text{ km}^3$ ,  $5 \times 5 \times 1 \text{ km}^3$  et  $7 \times 7 \times 1 \text{ km}^3$  et espacés de 25 km – afin d'éviter d'éventuelles interférences entre deux nuages successifs – sont soumis au mouvement décrit ci-dessus. Bien entendu, utiliser un panel de scènes nuageuses représentatif du site étudié enrichirait les résultats de cette analyse dynamique. Dans cette étude, on se contente de la sélection de ces quelques situations afin d'entrevoir les potentiels d'extraction de résultats et d'analyse par le simulateur proposé.

### Évaluation dynamique des performances de la méthode de géolocalisation

L'évaluation peut se faire de deux manières différentes : il est possible d'afficher à chaque instant – c'est-à-dire à chaque nouvelle position de la couverture nuageuse – les valeurs des métriques ACC et MCC décrites dans la partie précédente, mais il est également possible de comparer les séries temporelles d'ombres portées estimée et initiale. Ainsi, les métriques du chapitre II sont utilisées afin de caractériser les performances de la méthode vis-à-vis de son estimation ponctuelle de l'ombrage du site.

Dans l'optique de simuler dynamiquement les performances que délivre la méthode de géolocalisation des nuages intégrant progressivement les différents niveaux d'information, il est possible de fixer ou de faire évoluer la position du Soleil au cours du temps. La seconde option a été choisie afin d'intégrer dans une même analyse les effets 3D particulièrement importants aux faibles élévations solaires, et afin d'éviter une redondance des résultats. Les résultats de cette partie sont donnés sur une journée virtuelle dont la course du Soleil correspond typiquement à celle d'un jour d'été.

La notation 'SAT+CTH' est ajoutée dans cette analyse dynamique afin de mettre en évidence l'influence de l'utilisation de l'information *in situ* dans la géolocalisation. Ce niveau d'information supplémentaire exploite donc uniquement les informations susceptibles d'être obtenues du point de vue satellitaire.

Les figures 83, 84 et 85 illustrent respectivement les résultats de cette analyse dynamique des nuages de taille  $7 \times 7 \times 1 \text{ km}^3$ ,  $5 \times 5 \times 1 \text{ km}^3$  et  $1 \times 1 \times 0.5 \text{ km}^3$  soumis à un mouvement d'Ouest en Est de 20 km/h. Sur ces trois figures, la métrique de corrélation MCC est représentée sur les sous-figures (a) au cours de la journée pour les quatre niveaux d'informations de la méthode de géolocalisation. Ce graphique indique la performance globale de l'estimation de la présence d'ombre au sol. Quel que soit la taille des trois nuages simulés, cette sous-figure affiche un classement systématique des performances selon les niveaux d'information : la méthode strictement satellitaire 'SAT+CTH' est bien moins corrélée que la méthode 'SOL+SAT', elle-même légèrement moins corrélée que la méthode 'SOL+SAT+CTH'. Enfin, de manière prévisible, la méthode 'SOL+SAT+CTH+CBH' qui intègre toutes les informations obtient les meilleurs résultats de corrélation.

Ces différentes performances sont ensuite analysées en termes de séries temporelles de  $N_L$  sur le site étudié – sur les sous-figures (c) – et sur un site annexe situé à 4 km à l'Est (choix arbitraire pour l'illustration) – sur les sous-figures (e). Ces figures permettent d'abord de visualiser le lien entre variabilité spatiale et variabilité temporelle du rayonnement solaire. Par exemple, la figure 83.c suggère qu'un nuage de taille  $7 \times 7 \times 1 \text{ km}^3$  soumis à un mouvement linéaire à 20 km/h engendre une perturbation au sol pendant 21 min.

Sur les sous-figures (c) des trois figures 83, 84 et 85, seule la méthode 'SAT+CTH' n'est pas alignée temporellement avec la référence. En effet, la limite de résolution spatiale des images satellites engendre une approximation temporelle de l'estimation du nombre de couches  $N_L$  potentiellement nuageuses traversées. Cette propriété se retrouve sur les sous-figures (d), sur lesquelles le TDI est toujours le plus grand pour la méthode 'SAT+CTH'. Rappelons que pour l'analyse de tous ces graphiques radars, la valeur du TDI indique la distorsion temporelle moyenne de l'estimation par rapport à la référence, et le TDM indique le pourcentage de cette distorsion en avance (valeur négative), en retard (valeur positive) ou alignée temporellement (valeur nulle). Ainsi, les deux métriques sont à analyser ensemble. Par exemple un TDI proche de zéro signifie une distorsion faible, donc le TDM associé n'apporte peu d'information, et inversement.

Cependant, il est intéressant de remarquer que cette méthode 'SAT+CTH' admet une avance de moins en moins grande sur les fronts montants le matin, et un retard de plus en plus grand sur les fronts descendants des créneaux de l'après-midi, d'après les sous-figures (c). Ces décalages temporels d'avance le matin et de retard l'après-midi se compensent lors du calcul du TDM sur la journée, indiquant par sa valeur proche de zéro un alignement moyen journalier relativement bon.

L'utilisation de l'information issue des images hémisphériques pour l'estimation du site considéré – c'est-à-dire les méthodes 'SOL+SAT', 'SOL+SAT+CTH' et 'SOL+SAT+CTH+CBH' – assure une précision temporelle quasi parfaite, avec un TDI proche ou égal à zéro. Cette propriété d'alignement temporel découle de l'hypothèse d'une segmentation parfaite effectuée dans le cadre de ce simulateur, mais au vu des difficultés de la segmentation dans la zone circumsolaire sur les images hémisphériques réelles qui seront exprimés au chapitre V, une détérioration de cette propriété est attendue dans le cas réel.

Dans cette étude, la concordance temporelle entre les événements prévus et mesurés est importante. Nous nous sommes donc demandé si la propriété d'alignement temporel vu précédemment dans le cas d'une estimation sur le site considéré reste la même pour une estimation de  $N_L$  sur un autre site proche. Le site situé à 4 km à l'Est a été choisi et la dénomination « site annexe » sera utilisée dans la suite pour le désigner. Avec le mouvement de 20 km/h imposé d'Ouest en Est, au déplacement du Soleil près, les événements observés sur le site d'étude sont donc attendus 12 min plus tard sur le site annexe. La particularité d'estimer  $N_L$  sur le site annexe provient des volumes potentiellement nuageux qui sont calculés en partie sur les images issues du site d'étude. Ainsi, la délimita-

tion du nuage n'est pas aussi précise que pour une estimation sur le site d'étude. En particulier, même dans le cas de la méthode 'SOL+SAT+CTH+CBH' recoupant toutes les informations, les faces verticales du cube nuageux sont moins bien définies lorsqu'observées sous un autre point de vue, générant une distorsion temporelle supplémentaire sur le site annexe. Puisque dans la stratégie de construction des volumes potentiellement nuageux, il a été décidé de conserver l'ensemble des positions possibles des nuages, il est attendu d'obtenir des créneaux estimés englobant les créneaux de la référence.

D'après les sous-figures (f) des figures 83, 84 et 85, les différents niveaux d'information de la méthode de géolocalisation donnent des résultats légèrement meilleurs en termes de RMSE, de MAE, de biais et de rampes. Cependant, l'amplitude des créneaux ici n'est pas l'aspect le plus important recherché, puisqu'un post-traitement consistera à convertir ces amplitudes en rayonnement solaire. En revanche, l'approximation temporelle quantifiée par le TDI est plus grande sur le site annexe que sur le site d'étude pour toutes les méthodes intégrant l'information *in situ* et pour les trois tailles de nuages. Sur ces distorsions temporelles plus élevées, le TDM positif indique qu'elles sont majoritairement en retard. Cette caractéristique est particulièrement marquée pour la méthode 'SOL+SAT+CTH', ce qui peut s'expliquer par la partie inférieure du cône des possibles issus de l'image hémisphérique qui va entraîner une fausse détection du nuage. Le retard s'explique par la position relative du site annexe vis-à-vis du mouvement imposé et du site étudié.

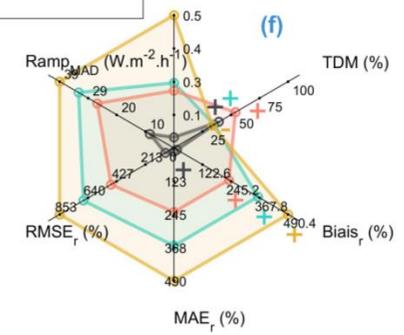
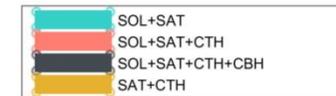
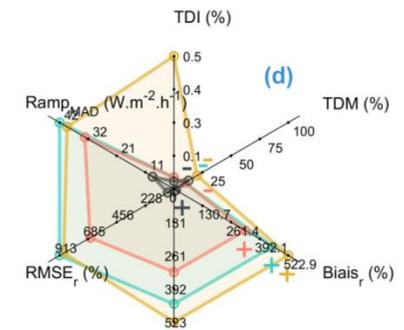
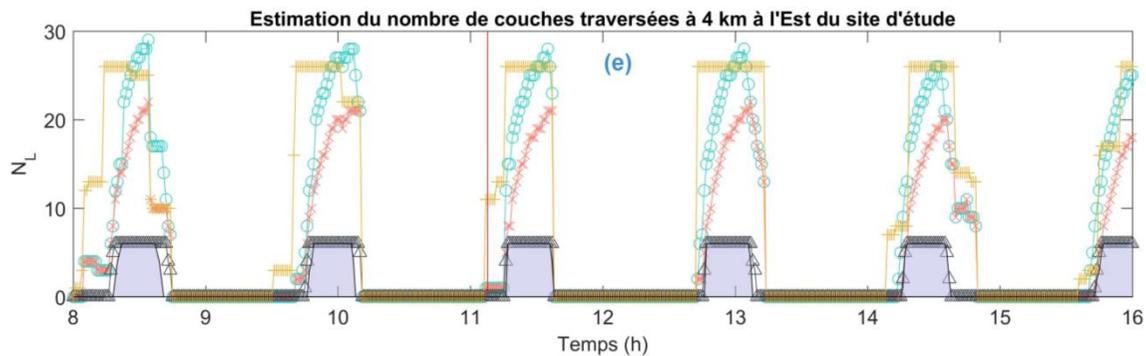
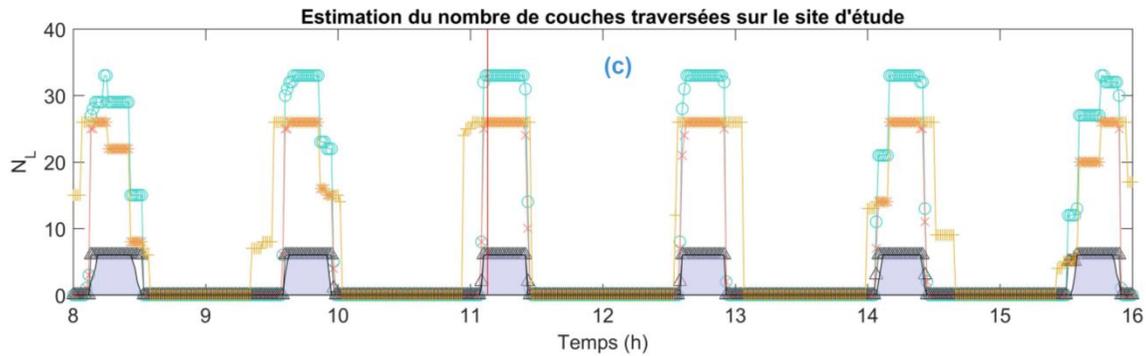
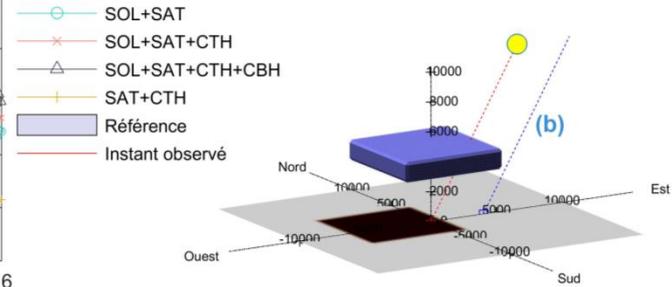
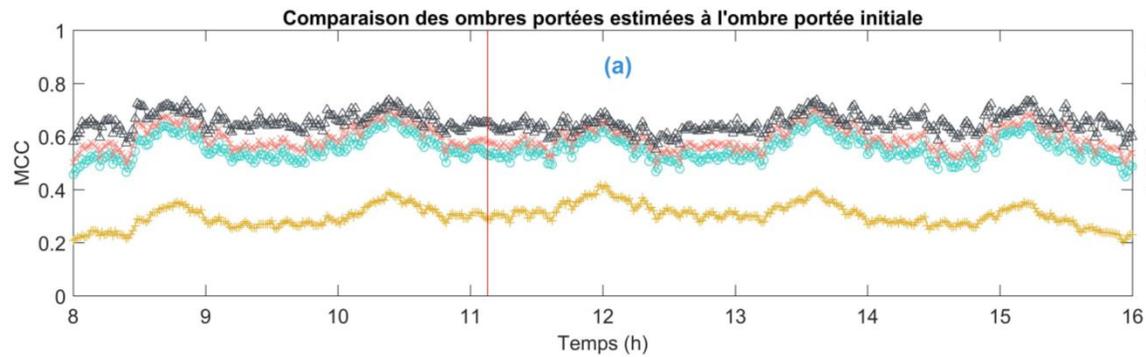


Figure 83 – Résultats de l'analyse dynamique du simulateur 3D pour un nuage de taille  $7 \times 7 \times 1 \text{ km}^3$  soumis à un mouvement de  $20 \text{ km/h}$  d'Ouest en Est. Pour chacune des trois méthodes de géolocalisation est affiché : (a) l'évolution temporelle de la métrique MCC calculée entre les ombres portées au sol, (b) une illustration d'une scène nuageuse à l'instant matérialisé en rouge, (c) l'estimation du nombre de couches  $N_L$  traversées dans la direction du Soleil pour le site et pour (e) le site annexe, ainsi que leur graphe radar respectif (d) et (f).

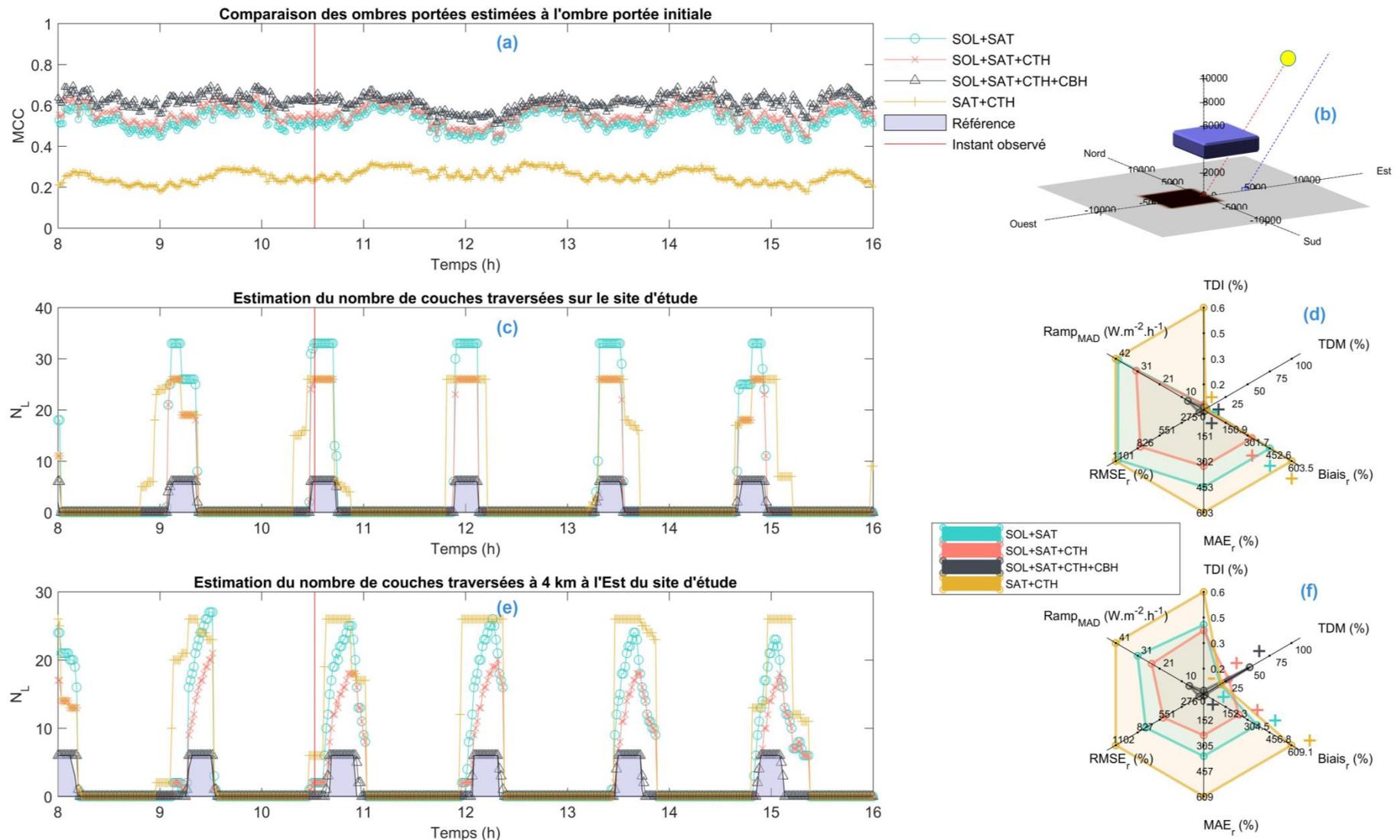


Figure 84 – Résultats de l'analyse dynamique du simulateur 3D pour un nuage de taille  $5 \times 5 \times 1 \text{ km}^3$  soumis à un mouvement de 20 km/h d'Ouest en Est. Pour chacune des trois méthodes de géolocalisation est affiché : (a) l'évolution temporelle de la métrique MCC calculée entre les ombres portées au sol, (b) une illustration d'une scène nuageuse à l'instant matérialisé en rouge, (c) l'estimation du nombre de couches  $N_L$  traversées dans la direction du Soleil pour le site et pour (e) le site annexe, ainsi que leur graphe radar respectif (d) et (f).

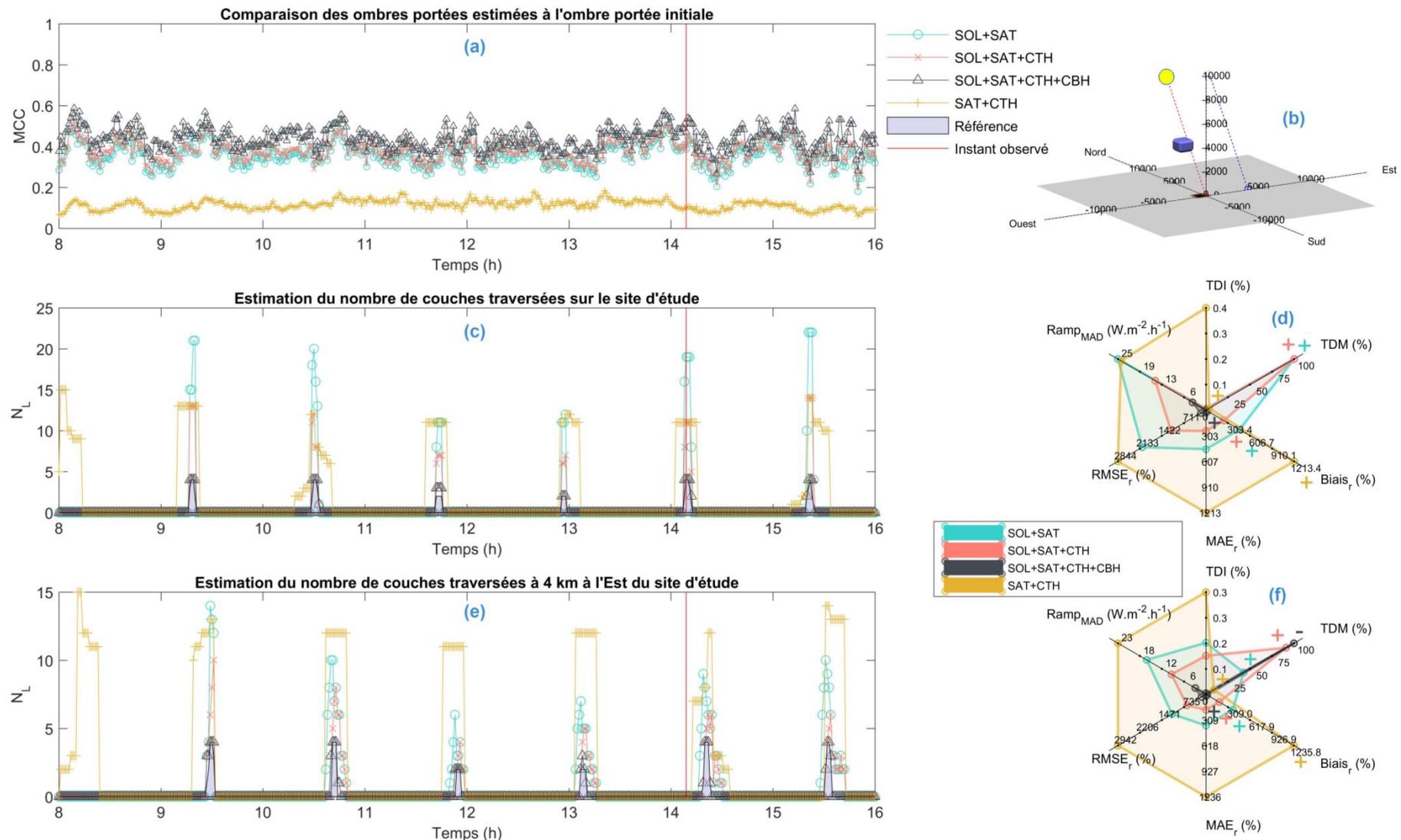


Figure 85 – Résultats de l'analyse dynamique du simulateur 3D pour un nuage de taille  $1 \times 1 \times 0,5 \text{ km}^3$  soumis à un mouvement de  $20 \text{ km/h}$  d'Ouest en Est. Pour chacune des trois méthodes de géolocalisation est affiché : (a) l'évolution temporelle de la métrique MCC calculée entre les ombres portées au sol, (b) une illustration d'une scène nuageuse à l'instant matérialisé en rouge, (c) l'estimation du nombre de couches  $N_L$  traversées dans la direction du Soleil pour le site et pour (e) le site annexe, ainsi que leur graphe radar respectif (d) et (f).

Finalement, plusieurs résultats ont été extraits du simulateur géométrique proposé. La méthode de géolocalisation de niveau d'information (1) 'SOL+SAT' exploitant uniquement le croisement des lignes de visée de l'observation satellitaire avec celles de l'observation in situ affiche des résultats simulés dépendants de l'angle relatif entre la position du nuage et celle du satellite. Cependant, ce déficit de performances lorsqu'un nuage est présent dans la zone proche de la direction satellitaire peut être réduit avec la connaissance de l'altitude du sommet et de la base des nuages. Aussi, l'analyse dynamique apporte une information sur le lien entre la précision de la localisation de l'ombre portée des nuages – quantifiée par l'analyse statique – et la précision temporelle de l'occurrence des événements – qui se rapproche de la comparaison à la mesure pyranométrique. Cette analyse autorise donc une visualisation du lien entre la variabilité spatiale et la variabilité temporelle du rayonnement solaire proche du site, et permet de mettre en évidence un inconvénient majeur des méthodes satellitaires – la limite de résolution spatiale des images satellites induisant une approximation temporelle des variations du rayonnement solaire au sol.

L'ajout d'un site annexe dans l'analyse dynamique a permis de caractériser les performances de la méthode de géolocalisation dans le cas où la caméra hémisphérique servirait à estimer le rayonnement solaire sur un site situé à 4 km dans la direction et le sens du mouvement des nuages.

## IV.6 Conclusions du chapitre

Le simulateur géométrique proposé dans ce chapitre exploite l'espace 3D comme référentiel géométrique afin d'effectuer la fusion des images synthétiques issues d'une part de la caméra hémisphérique et d'autre part du satellite géostationnaire. Statiquement, les performances en termes d'ombres portées des nuages associées à l'utilisation conjointe d'observations issues du point de vue du sol et du point de vue satellitaire ont été exposées. Le recoupement des lignes de visée de ces deux points de vue a une précision spatiale dépendante de la position relative du nuage et de sa taille. Cette partie statique du simulateur a aussi mis en évidence le gain obtenu par les connaissances des altitudes du sommet et de la base des nuages.

En dynamique, les mêmes tailles de nuages ont été évaluées et ont été soumises à un mouvement arbitraire de 20 km/h d'Est en Ouest. L'analyse des performances n'est plus spatiale et statique comme précédemment, mais temporelle et dynamique. La meilleure précision spatiale ayant été obtenue statiquement sur la méthode de géolocalisation 'SOL+SAT+CTH+CBH' a engendré systématiquement une plus petite distorsion temporelle vis-à-vis de la scène dynamique d'origine choisie. Cette amélioration n'a pas seulement été constatée sur le site où se trouve la caméra mais également pour un site arbitrairement choisi, situé à 4 km à l'Est de celui-ci.

Quelle que soit la situation nuageuse et le site choisis, le classement des niveaux d'information selon les métriques TDI, ramp-MAD, RMSE, MAE et biais reste inchangé : chaque ajout d'information dans ce simulateur améliore la géolocalisation des nuages, qui se traduit en une amélioration de leur estimation spatiale d'ombre segmentée et une amélioration de la série temporelle estimée du nombre de couches nuageuses traversées  $N_L$ . Ces améliorations sont systématiquement plus importantes pour le site d'étude que pour le site annexe (excepté pour le nuage de taille 1x1x0.5 km sur les métriques RMSE et ramp-MAD). Enfin, les méthodes utilisant l'information de niveau (2) et (3) sont toujours plus performantes que la méthode 'SAT+CTH' sur les nuages testés.

Néanmoins, la simulation suppose la réalisation d'une segmentation parfaite des deux images de nuages et ne prend nullement en compte les autres phénomènes optiques et atmosphériques qui rentrent en compte dans le cas réel. Les différentes méthodes de géolocalisation proposées dans ce chapitre seront donc appliquées sur des données réelles dans le chapitre V afin d'en extraire une estimation puis une prévision du rayonnement solaire au sol spatialement étendu.



---

# Chapitre V - Application de la géolocalisation 3D des nuages aux données réelles

---

Ce chapitre se concentre sur l'application de méthodes de géolocalisation 3D des nuages sur les données réelles de l'étude. De cette géolocalisation, une estimation et des prévisions à très court terme du rayonnement solaire seront extraites. De nouveaux défis émergent de l'application pratique des méthodes de géolocalisation des nuages par fusion et de l'étape de conversion en rayonnement solaire. En conséquence, de nombreux choix ont été effectués tout au long de ce chapitre afin de proposer une voie exploratoire, dont l'ambition est d'illustrer et d'analyser les potentialités offertes par de telles méthodes. L'objectif n'est donc pas tant de proposer une méthode définitive de prévision, mais plutôt d'emprunter une première voie – parmi d'autres – et de la tester sur une journée suffisamment diversifiée pour pouvoir discuter du gain et des limitations de l'approche.

Dans le cadre d'une application de ces méthodes à des données réelles, il sera d'abord nécessaire d'établir un référentiel temporel commun puisque les deux types d'images ont des résolutions temporelles différentes (resp. 10 s et 15 min). La partie V.1 s'attachera donc à la description d'un référentiel de l'espace-temps discret commun. Ensuite, les spécificités liées à l'utilisation des données réelles dans l'approche de géolocalisation des nuages seront exprimées dans la partie V.2, avec notamment les segmentations des nuages dans les deux types d'images et l'extraction des altitudes des nuages. Enfin, l'étape d'estimation du rayonnement solaire à partir des volumes potentiellement nuageux et l'étape de prévision du rayonnement solaire spatialement étendu (*i.e.* ne se limitant pas au point de mesure) seront traitées respectivement dans les parties V.3 et V.4.

## V.1 Recherche d'un référentiel commun

Comme il a déjà été mentionné précédemment, il est essentiel de trouver un référentiel commun pour effectuer la fusion des informations contenues dans les images des deux sources. Le paragraphe V.1.1 reviendra sur le référentiel géométrique choisi, puis une harmonisation des résolutions temporelles de sources d'informations servant à la fusion sera proposée au paragraphe V.1.2.

### V.1.1 Rappel du référentiel géométrique commun : l'espace 3D

L'espace est considéré comme référentiel géométrique commun, dans lequel le repère orthonormé  $(O, \vec{i}, \vec{j}, \vec{k})$  est utilisé, avec  $O$  l'origine du repère positionné à l'emplacement du site considéré. Les vecteurs unitaires  $\vec{i}, \vec{j}$ , et  $\vec{k}$  sont respectivement dirigés vers l'Est, vers le Nord et vers le Zénith et de leur longueur unitaire vaut 1 m. Le système de coordonnées spatiales est donc local, et seules les altitudes positives sont considérées. Dans la représentation en voxels, seules les coordonnées des centres de ceux-ci seront indiquées.

### V.1.2 Référentiel temporel commun : harmonisation des pas de temps à la minute

Dans le chapitre IV, l'analyse dynamique était possible sur n'importe quel pas de temps puisque la simulation permet de générer des images hémisphériques et satellites pour toute configuration nuageuse et toute position du Soleil souhaitées. Cependant, dans le cas réel, les images satellites MSG ne sont disponibles que toutes les 15 min – correspondant à des instants d'acquisition quasi instantanés en fonction du numéro de ligne dans l'image SEVIRI –, les images hémisphériques toutes les 10 s et les mesures pyranométriques toutes les minutes. Afin de réaliser la fusion des deux types d'images, le référentiel géométrique commun n'est pas suffisant, il est également nécessaire d'établir un référentiel temporel commun. L'objectif applicatif de cette thèse étant d'élaborer des schémas de prévision à haute résolution temporelle, nous décidons d'utiliser le plus petit pas de temps proposé par les données pyranométriques. De cette manière, seules les images hémisphériques les plus proches de la minute ronde seront considérées, et il sera ensuite nécessaire de ramener les images satellites à un pas de temps de la minute. Afin d'effectuer un tel sur-échantillonnage temporel des images satellites – qu'il s'agisse des images dans le visible ou l'infrarouge thermique ou encore des produits dérivés de ces canaux –, deux méthodes ont été identifiées :

- L'interpolation linéaire : l'obtention d'images intermédiaires peut se faire par simple somme pondérée des deux images encadrantes suivant une pondération dépendant de l'écart temporel relatif. Cette méthode n'est cependant pas causale (i.e. elle nécessite l'image satellite après l'instant de l'image intermédiaire souhaitée) et a, en outre, l'inconvénient de mal décrire la position des masses nuageuses puisqu'une sorte de moyenne des deux images est calculée ;
- L'interpolation par mesure des déplacements géométriques par flot optique. Cette méthode consiste à estimer le déplacement des nuages entre deux images successives, puis d'utiliser cette information de déplacement géométrique pour propager les nuages selon ce déplacement. La version causale de cette méthode, illustrée sur la figure 86.a, détecte le mouvement des nuages entre les deux dernières images successives précédant l'instant intermédiaire, jusqu'à la réception de la prochaine image satellite. La version non-causale (cf. figure 86.b) détecte le déplacement des nuages entre les images précédente et suivante afin de générer les images intermé-

diaires en croisant les informations de la dernière image avec celles de l'image future (approche de type *forward-backward*). Une description plus détaillée de ces deux versions de l'interpolation temporelle proposée est disponible en Annexe A2.

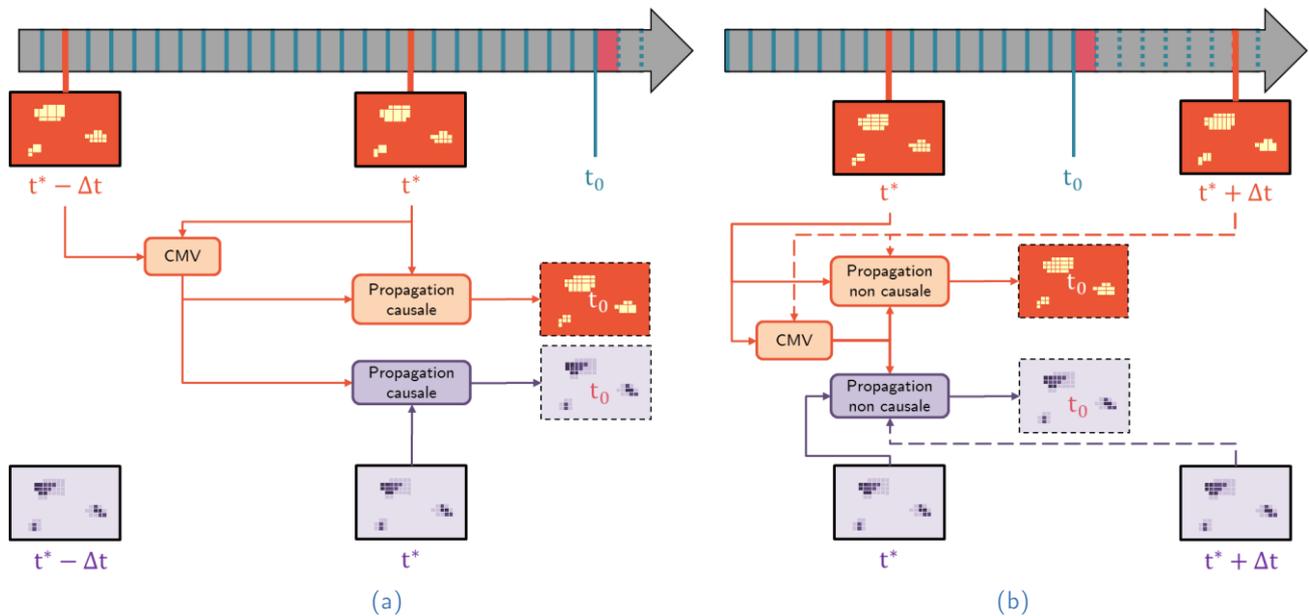


Figure 86 – Illustration des deux méthodes de sur-échantillonnage temporel des images satellites exploitant les CMV. Les images Helioclim sont schématisées en couleur orange et les images CTH (ou tout autre produit satellitaire) en violet.  $t_0$  est l'instant pour lequel l'information satellitaire est souhaitée et sera également l'instant présent (rouge) dans la suite de ce manuscrit. L'instant  $t^*$  correspond à celui de la dernière information satellitaire disponible.  $\Delta t$  est la résolution temporelle des images satellites et vaut 15 min dans notre cas. Les images Helioclim et CTH en bordures pointillées représentent les résultats des propagations causales (a) ou non-causales (b) calculés à partir des CMV estimés entre deux images Helioclim successives.

Grâce à cette dernière interpolation temporelle des images satellites, permettant d'établir des images satellites intermédiaires au pas de 1 min, il est désormais envisageable d'appliquer les méthodes de géolocalisation entre les images hémisphériques et les images satellites suivant la même granularité temporelle. Cependant, un dernier aspect est à prendre en considération dans ce référentiel temporel : l'instant d'acquisition des lignes de l'image satellite est différent pour chaque ligne de l'image satellite. Comme expliqué dans le chapitre III, l'instant d'acquisition des pixels de la zone recouvrant les deux sites à l'étude correspond à 4 min avant l'instant indiqué dans les métadonnées satellites. Ainsi, ce décalage temporel doit être considéré dans l'interpolation des données satellitaires au pas de temps 1 min.

## V.2 Spécificités de la géolocalisation des nuages à partir des données réelles

Le simulateur présenté au chapitre IV évalue les méthodes de géolocalisation 3D des nuages dans un cas où les nuages sont parfaitement identifiables dans les images artificielles. Bien qu'il soit possible de l'exploiter pour évaluer l'impact d'une erreur de segmentation des images sur les performances en termes d'ombre portée, ou encore de simuler d'autres versions des méthodes proposées, nous décidons de poursuivre l'analyse en appliquant ces méthodes sur de vraies images. Les deux spécificités de l'application de ces méthodes sur les données de l'étude qui seront traitées dans cette partie sont : la segmentation des images – décrite au paragraphe V.2.1 – et l'estimation de l'altitude du sommet et de la base des nuages – proposée au paragraphe V.2.2.

### V.2.1 Segmentation des images

L'étape qui diffère fondamentalement de la simulation est l'étape de segmentation des nuages. Que ce soit sur l'image hémisphérique ou sur l'image satellite, identifier automatiquement les nuages dans ces images est rendu difficile notamment par la complexité de la scène nuageuse et de son interaction avec le rayonnement solaire. Dans le cas de l'image hémisphérique, il s'agit de distinguer les nuages avec le ciel et le halo du Soleil ; et dans le cas satellitaire, les nuages sont à distinguer de la terre, la mer ou encore la neige.

#### V.2.1.a Segmentation des images hémisphériques

En raison de la diffusion moléculaire – appelée diffusion de Rayleigh – des courtes longueurs d'onde, celles correspondant à la couleur bleue sont plus diffusées que le reste du spectre solaire, ce qui donne au ciel clair sa couleur bleue. D'un autre côté, les nuages composés principalement d'eau et de particules de glace diffusent uniformément le rayonnement visible – appelée diffusion de Mie –, c'est pourquoi les zones nuageuses dans le ciel apparaissent souvent blanches. Ainsi, au sein de la communauté scientifique traitant de ces images hémisphériques du ciel dans les bandes spectrales du visible (et du proche infrarouge), la stratégie très largement adoptée pour la détection de nuages consiste à exploiter deux ou trois de leurs canaux rouge, vert et bleu (Chow et al., 2011; Heinle et al., 2010; Johnson et al., 1989; Long et al., 2006; Pfister et al., 2003; Roy et al., 2001).

Le plus souvent, le canal rouge est utilisé en comparaison du canal bleu car ils sont spectralement les plus discriminants entre les pixels de nuages et les pixels de ciel, mais le canal vert est parfois utilisé (Yang et al., 2016) car moins sensible au facteur de *flare*<sup>11</sup>. Dans la suite de ce paragraphe, les lettres R, G et B – pour *Red*, *Green*, *Blue* – feront respectivement référence aux canaux rouge, vert et bleu des images hémisphériques. Que ce

---

<sup>11</sup> « Aberration optique due à une diffusion parasite de la lumière à l'intérieur de l'objectif ». Source : [https://fr.wikipedia.org/wiki/Facteur\\_de\\_flare](https://fr.wikipedia.org/wiki/Facteur_de_flare)

soit sur le ratio  $R/B$ , la différence  $R-B$  ou la différence normalisée  $NRBR = (B-R)/(B+R)$  ( $NRBR$  : *Normalized Red-Blue ratio*), une méthode simple consiste à imposer un seuil arbitraire sur le résultat de ces opérations afin d'extraire les pixels nuageux (Gauchet et al., 2012; Heinle et al., 2010; Long et al., 2006; Yamashita, 2004). Heinle et al. (2010) et Kazantzidis et al. (2012) obtiennent des résultats de segmentation légèrement meilleurs avec le canal composite  $R-B$  que celui  $R/B$ , bien qu'il soit plus sujet aux bruits (Shields et al., 2009).

Le seuil imposé est cependant difficilement transposable à d'autres sites puisqu'il est dépendant des caractéristiques de l'acquisition de l'image ainsi que des conditions atmosphériques (Tapakis et Charalambides, 2013) et surtout de la position du pixel dans l'image hémisphérique. Plusieurs méthodes de seuillage automatique (Otsu, 1975; Yang et al., 2012) ou hybride ont été également proposées (Li et al., 2011), cependant deux zones de l'image restent particulièrement difficiles à segmenter :

- la zone circumsolaire (typiquement à moins de  $10$  à  $15^\circ$  de la position du Soleil) pour laquelle la luminance croît très rapidement en puissance inverse de la distance angulaire à la position Soleil (Buie et al., 2003). Lorsque la zone circumsolaire n'est pas oblitérée par des nuages ayant une épaisseur optique suffisante, cette croissance rapide de la luminance mesurée rend difficile la détection des pixels nuageux dans la zone. De plus, compte tenu de la limitation en dynamique des capteurs CCD ou CMOS des caméras numériques, les pixels de la zone circumsolaire sont très rapidement complètement saturés, empêchant toute segmentation pour ces derniers ;
- la zone proche de l'horizon correspond à des pixels du ciel acquis avec une résolution spatiale dégradée (effet *fish-eye*) correspondant à des angles solides de luminance de ciel associés à des épaisseurs optiques de l'atmosphère importantes.

Ces deux effets joints rendent difficile, même visuellement, la distinction entre un nuage dont les frontières sont floues ou une zone de luminance du ciel troublée par l'atmosphère.

Par ailleurs, la caméra hémisphérique a elle aussi une réponse colorimétrique qui n'est pas parfaitement homogène dans tout son champ hémisphérique. L'exposition de cette dernière à une source de luminance parfaitement homogène (sphère intégrante) ne générerait pas des images RGB parfaitement constantes.

Pour toutes ces raisons, la détection des nuages sur tout le champ hémisphérique peut être grandement améliorée en considérant un seuil dépendant de la position relative des pixels dans l'image et vis-à-vis du Soleil. Un tel seuillage est en fait grandement facilité par la modélisation de l'image acquise par condition de ciel clair (i.e. sans nuage) afin d'en extraire l'information de présence nuageuse par un seuillage unique sur la différence d'opérateurs colorimétriques entre l'image d'origine et l'image ciel clair modélisée correspondante.

Afin d'établir un modèle de prédiction des images par condition de ciel clair d'une caméra particulière, une sélection visuelle préalable d'images de ciel clair permet la création d'une table de correspondances (*Look-up Table*, LUT) entre la position relative des pixels dans l'image et du Soleil avec les intensités des canaux RGB (Chow et al., 2011; Long et al., 2006; Pfister et al., 2003; Shields et al., 2009). Pour cela, il est recommandé par Chauvin et al. (2015) de changer de système de coordonnées, puisque l'intensité de chaque pixel de ciel clair est notamment dépendante de l'angle qu'il forme avec le Zénith (*Pixel Zenith Angle*, PZA) et de celui qu'il forme avec le Soleil (*Solar Pixel Angle*, SPA). Un exemple de cette représentation (PZA, SPA) pour une image de ciel clair est proposé à la figure 87.b sur une grille de résolution angulaire de  $1^\circ$ . Dans cette représentation, chaque pixel est une valeur moyenne de tous les pixels de l'image d'origine ayant ces valeurs de PZA et de SPA (à  $1^\circ$  près). Ainsi, l'effet de *flare* rouge visible sur la figure 87.a ne se retrouve pas sur la représentation (PZA, SPA) puisqu'elle est moyennée (diluée) avec les valeurs proches de ces angles.

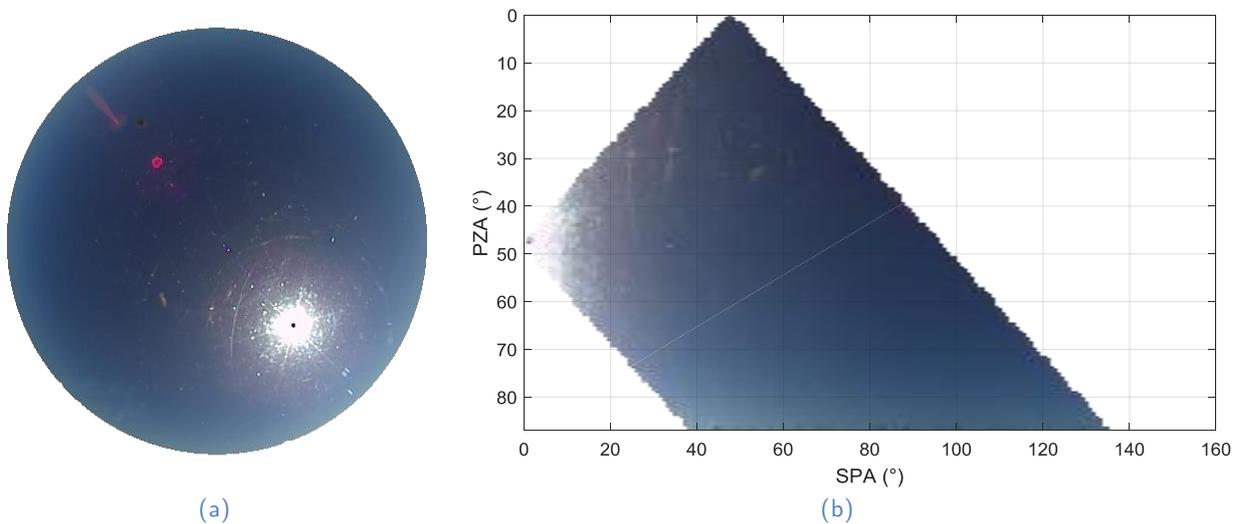


Figure 87 – Un exemple d'image de ciel clair représentée (a) dans le système de coordonnées de la caméra et (b) dans le système de coordonnées (PZA, SPA).

Dans cette étude, les 20 journées contenant le plus d'instant de ciel clair – d'après la classification d'instant proposée au paragraphe II.6.1 – sont utilisées pour chacun des deux sites afin de générer cette LUT, une pour chaque caméra. Ce jeu de données a ensuite été filtré en supprimant par inspection visuelle les images contenant des nuages. L'adaptation des résultats extraits de cette table aux zones de ciel clair apparentes le cas échéant sur l'image à segmenter a notamment été utilisée par Chauvin et al. (2015). Cette adaptation est proposée par les auteurs en adaptant une modélisation paramétrique de la luminance mesurée par ciel clair, mais peut également être envisagée plus simplement en multipliant l'image ciel clair modélisée par un facteur correctif (Ghonima et al., 2012; Shields et al., 2010), ce qui a pour but de s'adapter aux changements de turbidité de l'atmosphère.

Pour tenter de rendre compte de la variabilité de la luminance mesurée du ciel induite par un changement du trouble de l'atmosphère claire (variabilité du contenu de la colonne d'atmosphère en vapeur d'eau des épaisseurs optiques des aérosols, etc.), nous avons ajouté à ces tables de LUT une entrée supplémentaire liée au trouble de Linke  $T_L$ , en plus des entrées relatives aux paramètres angulaires PZA et SPA

L'estimation du trouble de Linke  $T_L$ , à la fois pour l'apprentissage et pour l'application de la LUT, se calcule de manière empirique d'après la relation proposée par Ineichen (2008) en fonction de la teneur en vapeur d'eau atmosphérique  $W$  et l'épaisseur optique d'aérosols AOD à 550 nm :

$$T_L \left( AOD, W, \frac{p}{p_0} \right) = 3.91 \cdot \exp \left( 0.689 \frac{p_0}{p} \right) \cdot AOD + 0.376 \cdot \ln(W) + 2 + 0.54 \cdot \frac{p_0}{p} - 0.5 \cdot \left( \frac{p_0}{p} \right)^2 + 0.16 \cdot \left( \frac{p_0}{p} \right)^3 \quad (37)$$

où  $p$  est la pression atmosphérique à l'altitude considérée et  $p_0$  la pression atmosphérique au niveau de la mer. Les variables  $W$  et AOD sont délivrées par le service *Global forecast of atmosphere composition* de CAMS (Copernicus Atmospheric Monitoring Service) et utilisées par le service CAMS Radiation McClear (Schroedter-Homscheidt, 2016) en prévision au pas de temps 3 h. La pression atmosphérique est mesurée directement sur chacun des sites au pas de temps 1 min. L'apprentissage sur les 20 jours de ciel clair délivre une LUT pour chacune des caméras, dont une illustration est proposée sur la figure 88. Le trouble de Linke a été séparé en quatre intervalles réguliers, et les angles PZA et SPA en 25 intervalles chacun. Une interpolation linéaire a été choisie afin de générer l'image de ciel clair à partir de cette LUT et des images de PZA, de SPA et de l'estimation à chaque instant du trouble de Linke de l'instant considéré.

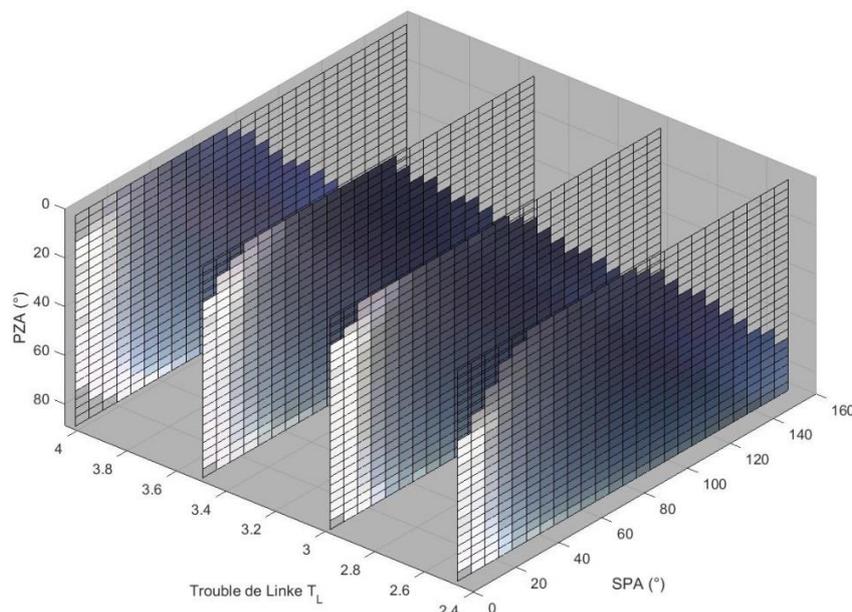


Figure 88 – Illustration de la LUT selon les trois variables d'apprentissage : PZA, SPA et  $T_L$ .

Désormais, pour chaque image hémisphérique, nous sommes en capacité de générer une image modélisée de ciel clair qui apporte l'information de la distribution anisotrope de l'intensité des pixels en l'absence de nuages. Ce modèle intègre sans toutefois les départager la répartition angulaire de la luminance du ciel ainsi que les défauts d'homogénéité de la réponse de la caméra. Des exemples de résultats extraits de la LUT sont disponibles sur la deuxième et la quatrième colonne de la figure 89 pour la journée du 22/04/2015. Dans la suite de ce chapitre, cette journée sera utilisée pour illustrer les différents comportements des estimations et prévisions issues de la méthode de géolocalisation proposée.

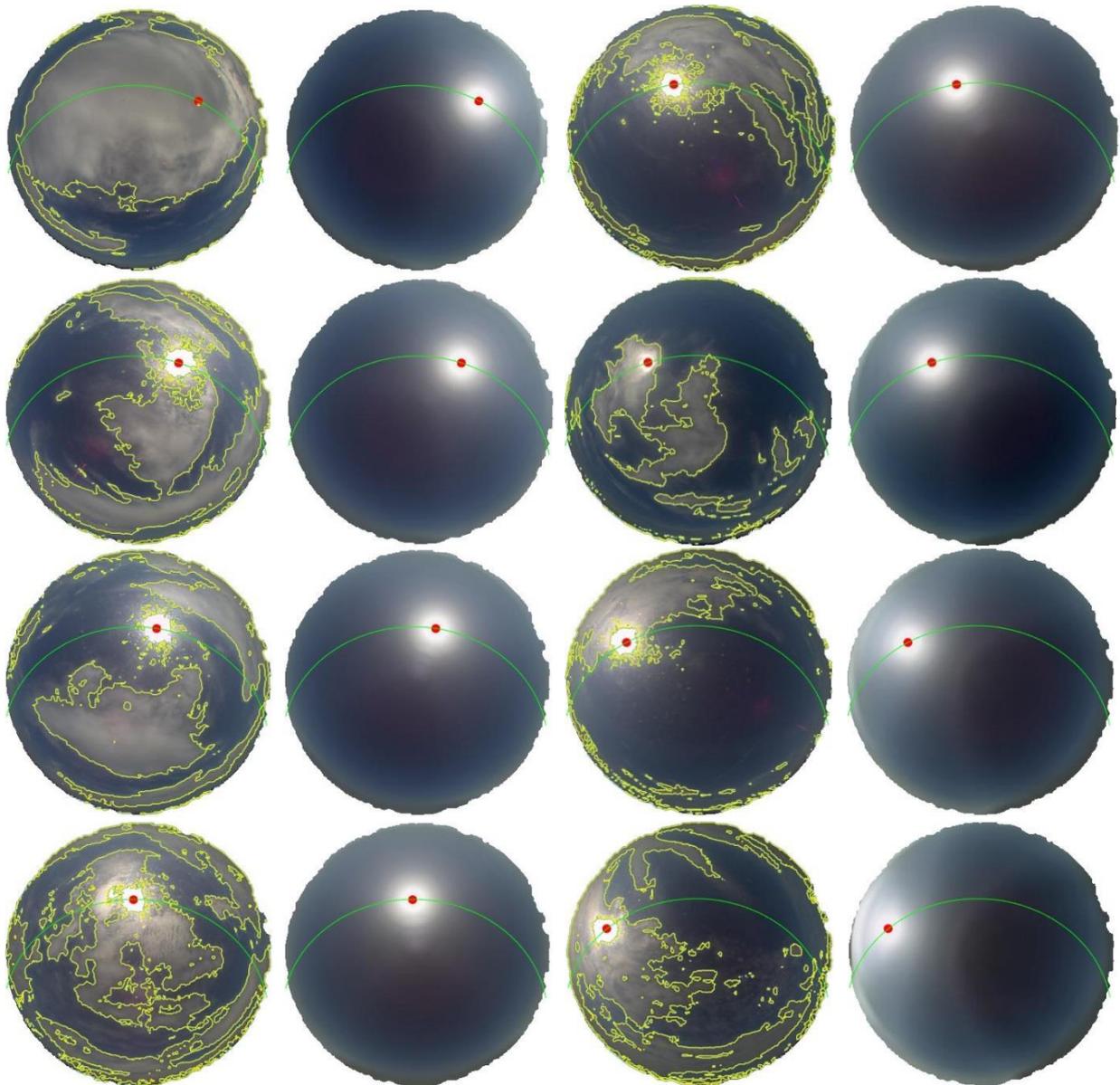


Figure 89 – Huit exemples de situations nuageuses observées par les caméras hémisphériques *in situ* lors de la journée d'exemple du 22/04/2015 avec les images issues de la LUT de ciel clair correspondantes. Ces images sont acquises toutes les heures entre 8h30 et 15h30. La segmentation des images hémisphériques est matérialisée par un contour jaune en surimpression, le Soleil est localisé sur ces images de manière théorique par le biais d'un point rouge et sa course pour cette journée apparaît en couleur verte.

À partir des images d'origine et de ciel clair, plusieurs opérateurs colorimétriques ont été proposés par la communauté scientifique dans le but de faciliter l'identification des nuages. Chauvin et al. (2015) ont suggéré d'appliquer un seuil de 0.15 sur la valeur absolue de la différence d'opérateurs NRBR de l'image d'origine et de l'image ciel clair modélisée, tel que :

$$|\text{NRBR} - \text{NRBR}^{\text{CLS}}| < 0.15 \quad (38)$$

avec

$$\text{NRBR}(R, B) = \frac{R - B}{R + B} \quad (39)$$

et l'exposant CLS indique que l'opérateur est appliqué à l'image de ciel clair modélisée.

Cet opérateur colorimétrique a l'avantage de présenter une distribution bimodale dans les ciels partiellement couverts (Chauvin, 2016), facilitant la segmentation par un unique seuil. Cependant, l'utilisation des seules différences entre les canaux rouges et bleus entraîne souvent une assimilation des *flares* rouges à des nuages. Pour faire face à ce problème spécifique à nos caméras, nous proposons l'opérateur colorimétrique suivant :

$$\text{NRBR}(\min(R - R^{\text{CLS}}, G - G^{\text{CLS}}), B - B^{\text{CLS}}) = \frac{\min(R - R^{\text{CLS}}, G - G^{\text{CLS}}) - (B - B^{\text{CLS}})}{\min(R - R^{\text{CLS}}, G - G^{\text{CLS}}) + (B - B^{\text{CLS}})} \quad (40)$$

Puisque les *flares* rouges ne sont pas décrits dans la LUT choisie, la différence  $R - R^{\text{CLS}}$  sera plus élevée dans les zones rouges de l'image que la différence  $G - G^{\text{CLS}}$ . Ainsi, le minimum de ces différences aura pour effet de passer outre cette élévation de la composante rouge. Des résultats de la segmentation sont disponibles sur la figure 89, où le seuil 0 a été choisi de manière à privilégier une segmentation correcte de la zone circumsolaire au détriment de la détection de certains nuages fins. La qualité de la segmentation est subjective et ne sera jamais parfaite, il s'agit donc de choisir un compromis.

### V.2.1.b Segmentation des images satellites

Afin d'identifier les nuages dans les images satellites, Escrig et al. (2013) proposent un seuil sur les pixels qui prend en compte l'hétérogénéité du sol en exploitant les canaux visibles et infrarouges de MSG. Avec ces mêmes canaux, Derrien et Le Gléau (2005) appliquent des tests locaux de texture afin d'en extraire le masque des nuages. Dans cette thèse, par soucis de simplicité, cette étape utilise uniquement les images d'indice de ciel clair HelioClim sur lesquelles est appliqué un seuil arbitraire. Ainsi, la détection des nuages est facilitée par l'indice de ciel clair fourni puisque la méthode Heliosat-2 compense les effets de l'albédo du sol pour ne conserver que l'influence des nuages sur le rayonnement solaire incident au sol.

Cependant, la segmentation des images satellites admet plusieurs limitations. L'une des faiblesses provient directement du modèle d'estimation solaire par imagerie satellitaire dans le domaine du visible. Son incertitude et le biais qui en résulte sont en effet accrus lorsque le sol est recouvert de neige ou, plus généralement, lorsque le sol est très réfléchissant. Perez et al. (2010) utilisent l'imagerie infrarouge afin de détecter la possible présence de neige au sol et ainsi améliorer la conversion de l'image satellite en rayonne-

ment solaire. Mais la faiblesse de la segmentation satellitaire repose surtout sur le déficit de résolution spatiale de ces images, où une décision binaire de présence de nuage sur un pixel de 3 km de large s'avère peu précise.

De plus, dans le cadre de la méthode de géolocalisation proposée, puisque l'on recoupe avec l'information au sol, une fausse alarme est préférable à une non-détection. Autrement dit, il est préférable d'être inclusif et de détecter faussement des pixels comme nuageux – ce qui sera en partie corrigé par la connaissance de l'altitude du sommet des nuages du paragraphe V.2.2.a –, plutôt que de ne pas détecter des pixels nuageux – ce qui aurait un impact conséquent lors de l'étape d'intersection des volumes potentiellement nuageux. C'est la raison pour laquelle le seuil sur l'indice de ciel clair choisi pour déterminer si l'image satellite est contaminée par des nuages est 0,98 : autrement dit si, pour un pixel donné, la valeur estimée de l'éclairement au sol est à 98% de la valeur de ciel clair théorique. Cette segmentation des images satellites peut également être effectuée directement en appliquant le masque des nuages proposé par le KNMI et inclus dans l'altitude haute des nuages ou CTH (voir section suivante).

## **V.2.2 Les trois versions de la méthode de géolocalisation dans le cas réel**

Dans le chapitre IV, l'analyse statique du simulateur a permis de mettre en évidence les avantages liés à l'ajout d'informations sur l'altitude du sommet et de la base des nuages. Au sein de cette simulation, ces altitudes étaient considérées comme connues parfaitement. Or dans le cas réel, il est nécessaire d'estimer ces deux grandeurs afin d'appliquer la méthode de géolocalisation avec les trois niveaux d'information présentés précédemment.

### **V.2.2.a Estimation de l'altitude du sommet des nuages**

Plusieurs organismes proposent des produits sur les propriétés des nuages en se basant sur différentes sources d'information. En particulier, le KNMI propose une estimation de l'altitude du sommet des nuages en se basant sur les images du proche infrarouge du satellite MSG. L'algorithme sépare dans un premier temps les pixels contaminés par des nuages des autres pixels en exploitant l'algorithme de détection des nuages développé à l'origine pour le spectroradiomètre d'imagerie à résolution modérée MODIS (Frey et al., 2008; Platnick et al., 2003). Ensuite, les propriétés physiques primaires des nuages (température au sommet, phase, épaisseur optique et tailles des gouttelettes d'eau) sont estimées de manière itérative en comparant simultanément les réflectances observées dans les canaux du visible et du proche infrarouge aux tables de correspondances (LUT) des réflectances simulées par calcul de transfert radiatif des nuages d'eau et de glace. Enfin, la détermination des autres propriétés des nuages repose principalement sur ces propriétés primaires.

Concernant la variable qui nous intéresse ici, l'estimation de l'altitude des nuages (*Cloud Top Height*, CTH) s'effectue par comparaison de la température de brillance infrarouge au sommet des nuages à un profil de température issu de modèles météorologiques. Cette technique avait été proposée par Mahani et al. (2000) en soulignant l'importance d'un recalage de type orthorectification des images satellites pour limiter l'impact de l'effet parallaxe.

L'évaluation du CTH issu du KNMI a été effectuée par Stengel et al. (2014) en comparant ce produit aux informations délivrées par les satellites CloudSat et CALIPSO. Cette étude indique un écart type de 2 km et un biais de  $-0.7$  km pour le CTH. Dans la suite de cette étude, l'information proposée par le CTH sera considérée comme correcte, bien qu'une analyse plus poussée de la sensibilité de cette variable sur nos résultats mériterait d'être développée.

Des méthodes de stéréoscopie satellitaires sont également envisageables afin d'estimer l'altitude du sommet des nuages. Campbell et al. (1996, 1998) mettaient déjà en avant la possibilité d'extraire cette altitude ainsi que les vitesses du sommet des nuages géométriquement à partir de paires d'observations satellites, puis Campbell et Holmlund (2004) ont présenté les résultats de leur méthode sur cinq paires de satellites. En 2007, Seiz et al. ont appliqué ces méthodes de détermination de l'altitude du sommet des nuages et de leur vitesse à partir des satellites Météosat de première génération 5, 7 et 8, avec, notamment, une investigation sur l'avantage de l'utilisation du canal HRV. La précision de l'estimation de l'altitude avec cette méthode est de  $\pm 1$  km. Enfin, Horváth et Davies (2001) ont appliqué ces méthodes stéréoscopiques au satellite Terra en orbite polaire avec neuf caméras embarquées, dont les caractéristiques orbitales permettent d'obtenir une précision de l'altitude du sommet des nuages de 400 m. Cependant cette information n'est pas disponible à n'importe quelle heure contrairement aux satellites géostationnaires, ce qui limite son utilité pour la prévision solaire.

### **V.2.2.b Estimation dynamique de l'altitude de la base des nuages**

De multiples sources d'informations peuvent servir de support à l'estimation de l'altitude de la base des nuages, comme par exemple des ceilomètres (Gaumet et al., 1998) ou encore des couples caméras hémisphériques utilisées en mode stéréoscopique (Blanc et al., 2017; Crispel et Roberts, 2018; Seiz, 2003; Tapakis et Charalambides, 2013). Aussi, Bertin et al. (2015) utilisent une caméra hémisphérique infrarouge pour récupérer l'altitude de la base des nuages épais – dont la profondeur optique est supérieure à 3 – en exploitant le même type de relation qu'au paragraphe V.2.2.a entre température et altitude des nuages.

En l'absence de telles configurations instrumentales, nous proposons une méthode d'estimation de l'altitude de la base des nuages exploitant la dynamique des nuages observée conjointement par la caméra hémisphérique et le satellite et dont le principe est schématisé à la figure 90. En effet, le mouvement des nuages estimé entre deux images hémisphériques successives à l'aide d'algorithmes CMV (*cf.* annexe A1) permet d'obtenir

une vitesse du nuage observé, notée  $V_{sol}$ , exprimée en pixel/s s'apparentant à une vitesse angulaire. Du point de vue satellitaire, le système a une focale et une distance d'observation de la Terre suffisamment grandes pour que la vitesse estimée soit très proche d'une vitesse de translation, notée  $V_{sat}$  et exprimée en m/s.

L'hypothèse est que le mouvement de translation du sommet du nuage est le même que celui de sa base, et que le déplacement de type angulaire perçu du point de vue du sol correspond au déplacement de la base du nuage. Une estimation de l'altitude de la base du nuage  $h_b$  à partir de ces deux vitesses a été proposée par Wang et al. (2016) et repris par Kuhn et al. (2018) :

$$h_b = \frac{V_{sat}}{V_{sol}} \cdot \frac{N}{2 \cdot \tan \theta_m} \quad (41)$$

où  $\theta_m$  est l'angle zénithal d'intégration avec lequel l'information de l'image hémisphérique est moyennée, et  $N$  correspond au diamètre du cercle défini par  $\theta_m$  en nombre de pixels, tels qu'illustrés sur la figure 90. Cette approche est plus précisément décrite dans l'article de Wang et al. (2016). Ces derniers estiment l'incertitude exprimée en RMSE de 126 m mais elle dépend de l'incertitude de l'instrument de mesure des vitesses des nuages utilisé par les auteurs. Dans notre cas, puisque la comparaison est effectuée avec une estimation satellitaire du mouvement des nuages, l'incertitude sera certainement plus grande. Aucun appareil de mesure de vitesses des nuages n'étant installé sur nos sites d'étude, l'estimation de l'incertitude de la méthode dans notre cas n'a pas été réalisée.

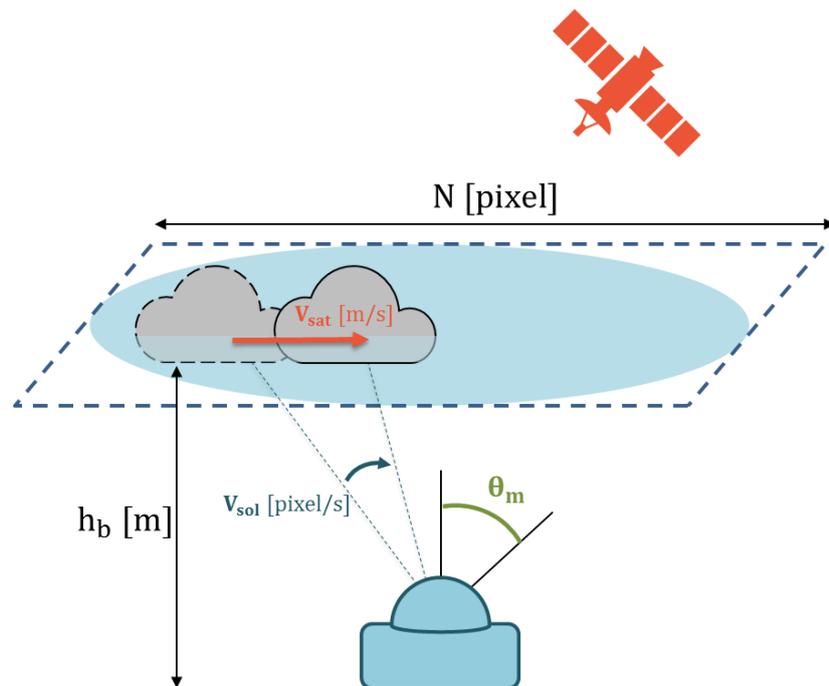


Figure 90 – Schéma explicatif de l'algorithme d'estimation de l'altitude de la base des nuages par comparaison entre le déplacement angulaire du point de vue du sol (en bleu) et le déplacement par translation du point de vue satellitaire (en orange).

### V.2.3 Analyse qualitative de la géolocalisation sur le jour d'exemple

Comme indiqué précédemment, le jour 22/04/2015 a été choisi comme jour d'exemple sur le site n°2 afin d'illustrer le comportement des méthodes proposées sur des configurations nuageuses particulières. Les illustrations de ce paragraphe ont pour objectif d'identifier certaines caractéristiques de la méthode de géolocalisation proposée sur trois périodes particulièrement intéressantes. Ces trois périodes sont délimitées sur la figure 91 en surimpression des courbes de GHI et BHI mesurés au sol sur le site n°2.

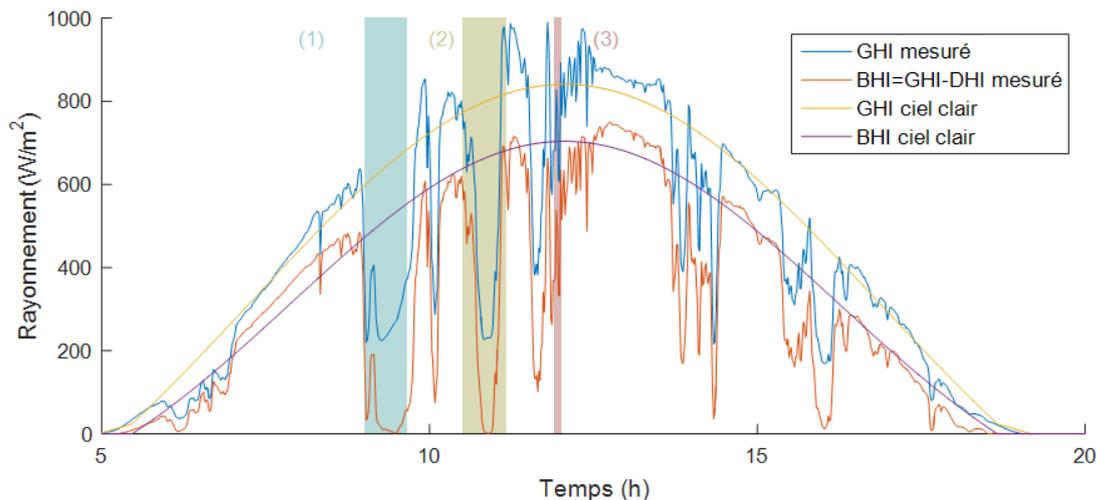


Figure 91 – Sélection des trois plages d'instants : la plage (1) s'étend de 9h à 9h40, la plage (2) de 10h30 à 11h10 et la plage (3) aux instants autour de midi. La différence des composantes globale et diffuse de la mesure pyrométrique est nommée "BHI mesuré". Les composantes de ciel clair sont issues du modèle ESRA.

Sur ces trois plages d'instants, la méthode de géolocalisation proposée avec le niveau d'information (3) SOL+SAT+CTH+CBH est appliquée. Les images satellites HelioClim et CTH utilisées pour l'estimation sont celles rapportées au pas de temps 1 min selon la propagation non-causale (cf. figure 86.b du paragraphe V.1.2). Le volume potentiellement nuageux obtenu à l'issue de cette géolocalisation est illustré sur les figures 92, 93 et 94 respectivement pour les plages d'instants (1), (2) et (3) avec les images hémisphériques, HelioClim et CTH correspondantes toutes les 10 min.

Sur la figure 92, la plage d'instants (1) correspond au passage d'un nuage conséquent dont le mouvement est globalement advectif du Sud vers le Nord. La série d'images hémisphériques de la colonne de gauche de la figure permet la délimitation du contour des nuages avec une bonne résolution spatiale. La colonne du milieu donne, pour chaque instant choisi, l'image estimée à la minute près de HelioClim et du CTH. Ces séries de deux images nous renseignent respectivement sur la présence (pour un seuil donné) et l'effet des nuages sur l'éclairement incident et sur l'altitude du sommet de ces derniers. Enfin, la colonne de droite illustre la géolocalisation des nuages issue de la méthode proposée. Cette géolocalisation est qualitativement cohérente avec les observations, elle sera analysée dans le cadre de l'estimation et de la prévision dans les paragraphes V.3 et V.4.

Dans le cas d'une estimation du rayonnement solaire direct, la zone circumsolaire contient l'essentiel des informations nécessaires. La partie saturée de cette zone peut notamment être exploitée pour affiner la segmentation de l'image dans cette zone. Par exemple, sur les images hémisphériques à 9h et 9h10 le Soleil est perçu au travers des nuages. Sur la première image, la zone saturée est détectée comme appartenant au ciel clair et dans le second cas, cette zone passe le seuil de la segmentation et est considérée comme nuageuse. Cet effet de seuil va avoir tendance à générer des rampes d'estimation de pentes plus fortes que celles de la mesure. Ces instants sont également particulièrement difficiles à prévoir puisque sans information sur la transmittance des nuages, prévoir précisément à quel moment le Soleil transpercera les nuages est une tâche difficile avec l'approche proposée.

Sur la figure 93, les images hémisphériques et satellites correspondant à la plage d'instant (2) illustrent la présence d'un nuage de plus petite taille, mais dont la délimitation est relativement bonne sur les deux segmentations. Cependant, lors de la reconstruction du volume 3D illustrée sur la partie droite de la figure, il semble que l'estimation de l'altitude de la base des nuages (CBH) soit perturbée par les nuages proches de l'horizon. Ainsi, cette mauvaise estimation du CBH ajoute une partie du volume potentiellement nuageux sous la position réelle du nuage. Cette partie du volume disparaît en partie à 10h50 pour se reformer plus tard. Ces effets sont particulièrement visibles à cause des nuages situés à l'horizon, dont l'intersection de leur cône des possibles avec l'information issue du CTH entraîne une estimation surévaluée des masses nuageuses géolocalisées. La remarque sur les difficultés liées à la transmittance des nuages dans un processus de prévision est la même qu'à la figure 92.

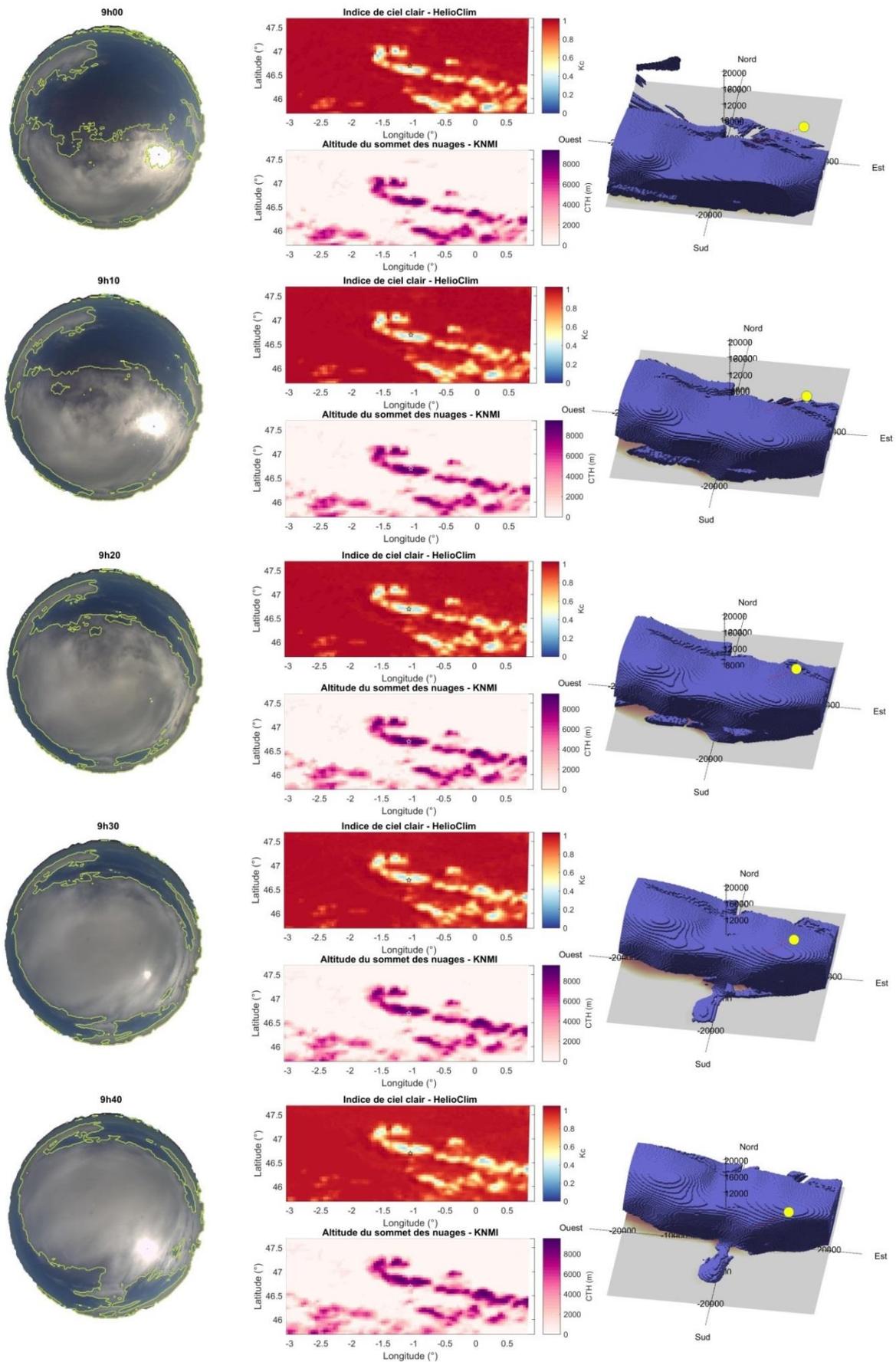


Figure 92 – Illustration d'un passage nuageux dont le mouvement est principalement advectif et pour lequel la géolocalisation proposée performe relativement bien. Cette illustration correspond à la page d'instant (1).

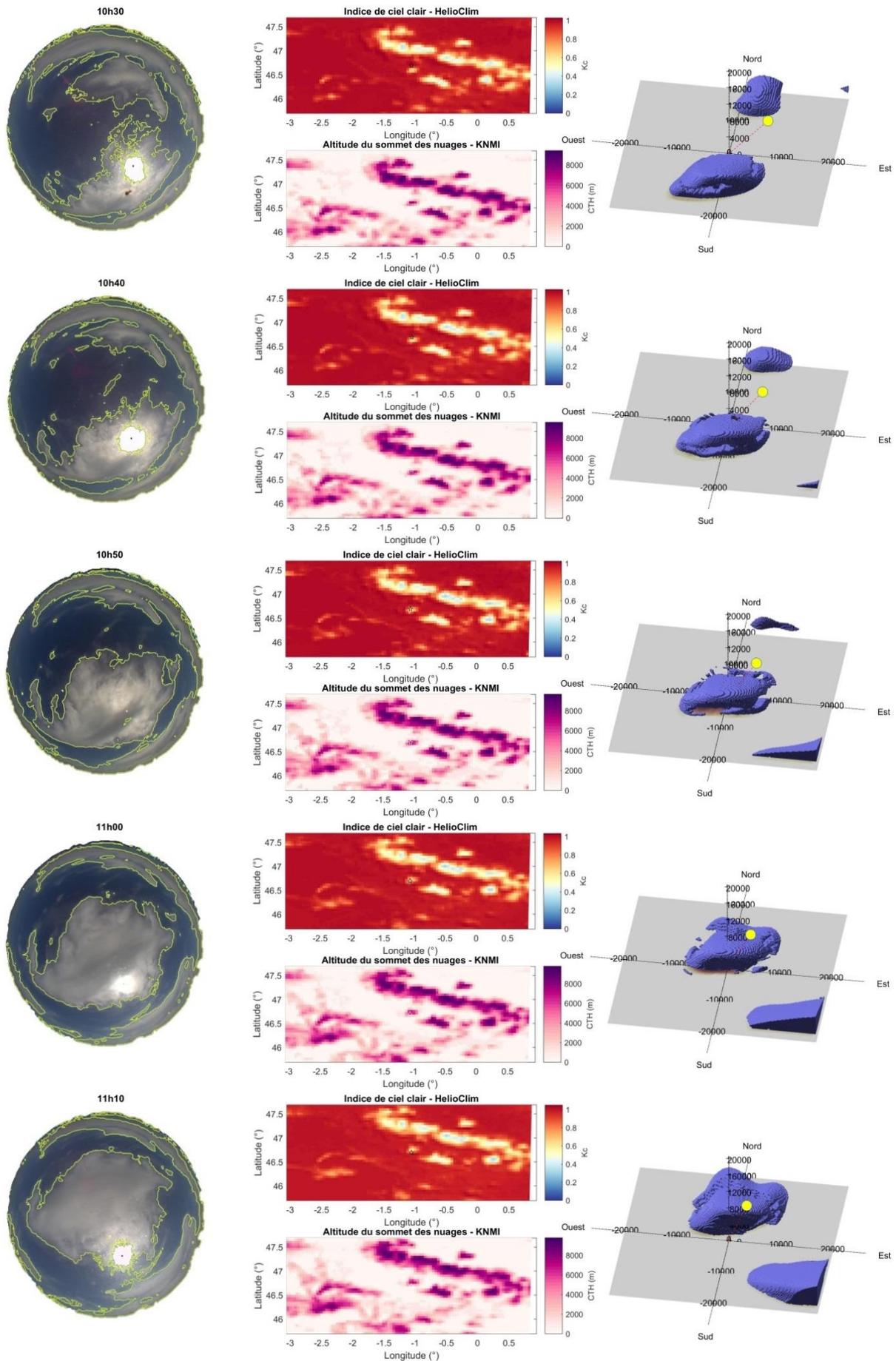


Figure 93 – Illustration d'un passage nuageux pour lequel la géolocalisation proposée affiche certaines limites. Cette illustration correspond à la plage d'instant (2).

La figure 94.a illustre le résultat de la géolocalisation 3D des nuages effectuée pour une situation nuageuse fractionnée, fine et semi-transparente. Ces nuages ont la particularité d'être détectés sur l'image hémisphérique mais ne sont pas visibles sur les informations satellitaires. Ainsi, le CTH n'est pas correct puisqu'il est estimé à une valeur proche de zéro et le CBH est lui aussi impacté puisque l'estimation du mouvement des nuages à un endroit où les nuages ne sont pas détectés est proche de zéro également. Dans ces cas où le CTH est inférieur au CBH, nous avons décidé d'imposer à ces nuages l'altitude de 2,6 km correspondant à la valeur médiane du CTH de l'ensemble de nos données. Le résultat pour ce type de situations intégrant cette condition est illustré sur la figure 94.b.

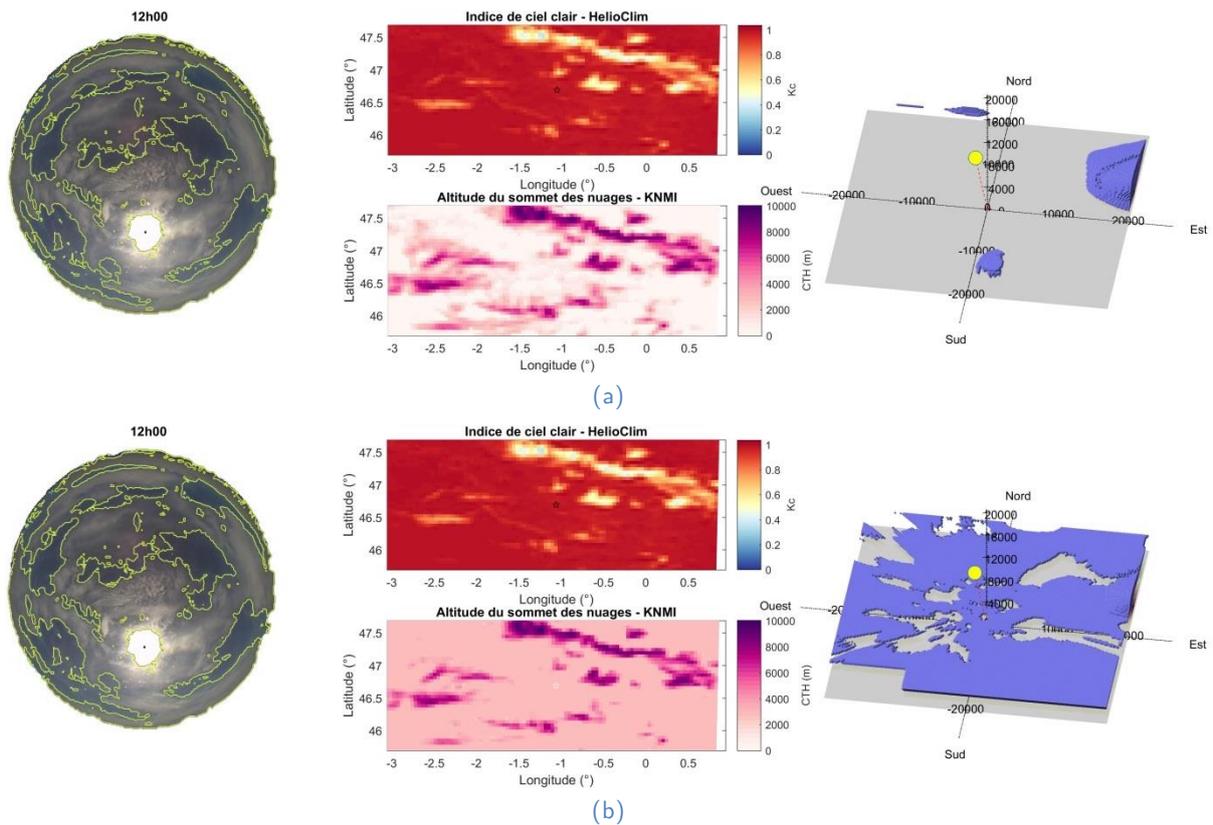


Figure 94 – Illustration d'une situation nuageuse de faible épaisseur, semi-transparente et fractionnée visible uniquement sur l'image hémisphérique. La méthode de géolocalisation telle qu'expliquée jusqu'à présent donne le volume des possibles à droite de la sous-figure (a) et la méthode utilisant l'altitude médiane des CTH en cas de conflit entre le CTH et le CBH donne celui de la sous-figure (b).

## V.3 Estimation du rayonnement solaire spatialement étendu

La partie précédente s'est focalisée sur l'étude de certaines spécificités et difficultés liées à l'application des méthodes de géolocalisation 3D des nuages sur des données réelles. Le chapitre IV traitant du simulateur géométrique de ces méthodes a permis d'extraire des résultats simulés en termes d'ombres portées des nuages. Cependant, dans le cas réel, seul le capteur pyranométrique permet l'analyse de la performance de ces méthodes. Ainsi, l'étape manquante qui va être l'objet de cette partie est la conversion de la série temporelle d'ombrage issue de la géolocalisation des nuages en série temporelle de rayonnement solaire estimé. Différentes stratégies peuvent être adoptées afin d'établir – si possible – un lien entre ces deux grandeurs. Une première estimation du rayonnement solaire direct s'effectuera par le biais d'une relation simple *a priori* avec le nombre de couches potentiellement nuageuses traversées  $N_L$  au paragraphe V.3.1. Ensuite, une méthode associant les volumes potentiellement nuageux issus des méthodes de géolocalisation des nuages avec l'estimation satellitaire de la base HelioClim est proposée au paragraphe V.3.2. Cette méthode d'estimation permet de corriger en partie les images HelioClim de l'effet parallaxe entre le point de vue du sol et celui du satellite, en exploitant l'information d'altitude du sommet des nuages proposée par le KNMI via le produit CTH (cf. paragraphe V.2.2.a). Enfin, le paragraphe V.3.3 s'intéresse à une méthode de régression se basant sur une rétroaction en temps réel de la mesure pyranométrique. Cette méthode permet d'établir un lien entre les éclaircissements solaires mesurés et les caractéristiques géométriques des ombres et des positions des nuages estimées par les méthodes de géolocalisation.

### V.3.1 Relation *a priori*

La première estimation du rayonnement solaire direct qui a été envisagée repose sur une modélisation simple de l'indice de ciel clair à partir du nombre de couches  $N_L$  traversées dans la direction du Soleil :

$$\widehat{G}_b = f\left(\frac{N_L}{N_{L\max}}\right) \cdot G_b^{\text{CLS}} \quad (42)$$

où  $N_{L\max}$  est le nombre maximal de couches nuageuses selon la maille de voxels choisie,  $\widehat{G}_b$  est l'estimation du rayonnement direct,  $G_b^{\text{CLS}}$  est le rayonnement direct par condition de ciel clair, et  $f$  est une fonction qui peut être définie de manière *a priori* ou bien déterminée par d'autres méthodes. La fonction  $f$  suivante est proposée comme fonction *a priori* :

$$\forall x \in [0; 1], \quad f(x) = 1 - x \quad (43)$$

Bien que cette formulation simple ait l'avantage de ne dépendre d'aucun autre paramètre et qu'elle soit très simple à appliquer, elle n'est reliée à aucun comportement théorique : l'épaisseur géométrique d'un nuage n'étant pas directement relié à son épaisseur optique ou à son indice de ciel clair correspondant.

### V.3.2 Utilisation de l'indice de ciel clair HelioClim

Une autre approche de l'estimation du rayonnement solaire incident consiste à exploiter celle délivrée par la base de données HelioClim en attribuant à chaque voxel potentiellement nuageux l'indice de ciel clair correspondant d'après l'estimation par la méthode Heliosat-2. Cette information est en effet très peu utilisée dans les méthodes de géolocalisation proposées jusqu'à présent, puisqu'elle sert uniquement à la segmentation des images satellites (*cf.* paragraphe V.2.1.b), qui est déjà plus ou moins présente dans l'information délivrée par le CTH. Ainsi, plutôt que de projeter l'information de présence de nuage dans l'espace 3D, la projection de l'indice de ciel clair d'HelioClim est effectuée. Chaque voxel contient désormais une valeur d'indice de ciel clair. Il reste à déterminer une manière de combiner ces indices de ciel clair dans la direction de l'éclairement solaire direct. Là encore, différentes opérations mathématiques peuvent être appliquées sur cette colonne d'indices afin d'estimer *in-fine* le rayonnement solaire incident au sol. Nous avons choisi d'utiliser la moyenne des indices de ciel clair dans la direction du Soleil intersectés avec les volumes potentiellement nuageux proposés.

Cependant, cette méthode délivre un indice de ciel clair  $k_c$  moyen pour la composante globale du rayonnement solaire. La base de données HelioClim3 utilise une procédure particulière pour évaluer les composantes directe (BHI) et diffuse (DHI) sur plan horizontal du rayonnement connaissant le GHI ainsi que l'éclairement hors atmosphère et la masse d'air correspondants. Cette procédure, empirique, est basée sur la fonction paramétrique G2 proposée par Ruiz-Arias et al. (2010).

Les résultats de l'estimation du BHI par HelioClim et de l'estimation du BHI issue de la méthode décrite précédemment – appelée « Estimation HelioClim corrigée » – sont illustrés pour la journée du 22/04/2015 à la figure 95.

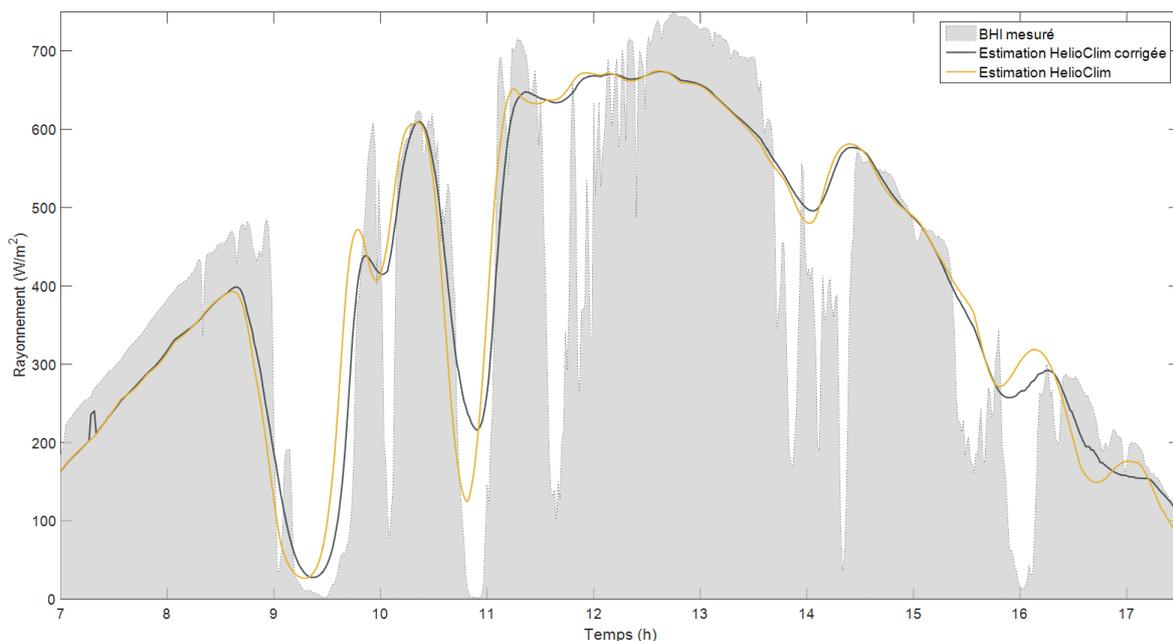


Figure 95 – Illustration sur la journée d'exemple du 22/04/2015 de l'estimation du rayonnement direct issu directement de la base HelioClim, de l'estimation corrigée par la géolocalisation des nuages via la méthode proposée (« Estimation HelioClim corrigée ») et de la mesure du BHI sur le site n°2.

Des analyses de comparaisons avec les mesures *in situ* correspondantes permettent de montrer que la correction apportée améliore sensiblement le TDI, qui passe de 5,6 % à 2,8 %. Ces opérations ont donc pour effet de corriger l'effet parallaxe induit par la différence de point de vue entre le sol et le satellite grâce à la géolocalisation des nuages réali-sée. Cependant, la limite de résolution spatiale de l'image HelioClim ne permet pas d'ajouter beaucoup d'informations hautes fréquences.

### V.3.3 Régression

La dernière approche envisagée pour extraire une estimation du rayonnement solaire incident à partir des volumes potentiellement nuageux précédemment construits est la ré-gression. Le terme de régression désigne une analyse statistique dont le but est d'établir une relation entre une variable – dite expliquée – et une (ou des) variable(s) – dite(s) explicative(s). Cette régression dans ce cas s'attachera à établir une estimation du rayon-nement solaire à partir de l'information du nombre de couches  $N_L$  nuageuses traversées dans la direction du Soleil ainsi que de la valeur moyenne de l'indice de ciel clair issue des images HelioClim. L'étape préliminaire consiste à entraîner le modèle sur des données historiques par comparaison à la mesure pyranométrique. Le lien entre ces deux grandeurs, s'il existe, dépend des paramètres intrinsèques de la représentation 3D choisie et notam-ment de la taille de chaque voxel.

Le premier modèle de régression envisagé ici, appelé régression statique, est typique-ment entraîné sur une première période, puis testé sur une autre supposée indépendante de la première. Les variables explicatives choisies pour diriger le modèle sont l'estimation proposée au paragraphe précédent V.3.2, ainsi que le nombre de couches  $N_L$  traversées dans la direction du Soleil. La structure polynomiale suivante est imposée au modèle :

$$\text{Pour } i = 1 \dots n, \quad y_i = \alpha_0 + \alpha_1 x_{i,1} + \alpha_2 x_{i,2} + \alpha_3 x_{i,1} x_{i,2} + \alpha_4 x_{i,1}^2 + \alpha_5 x_{i,2}^2 + \epsilon_i \quad (\text{a})$$

avec  $\alpha_0, \dots, \alpha_5$  les coefficients du modèle à estimer,  $x_{i,1}$  l'estimation HelioClim exploi-tant les volumes potentiellement nuageux,  $x_{i,2}$  le nombre de couches  $N_L$  traversées dans la direction du Soleil,  $\epsilon_i$  l'erreur du modèle et  $y_i$  le rayonnement solaire BHI mesurée par le pyranomètre. La représentation matricielle suivante sera utilisée dans la suite :

$$X \alpha = Y \quad (44)$$

avec  $\alpha = (\alpha_k)_{1 \leq k \leq 5}$ ,  $Y = (y_i)_{1 \leq i \leq n}$  et

$$X = \begin{pmatrix} 1 & x_{1,1} & x_{1,2} & x_{1,1}x_{1,2} & x_{1,1}^2 & x_{1,2}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n,1} & x_{n,2} & x_{n,1}x_{n,2} & x_{n,1}^2 & x_{n,2}^2 \end{pmatrix}$$

La méthode des moindres carrés est utilisée pour estimer les coefficients  $a_k$  du modèle :

$$\hat{\alpha} = (X^t X)^{-1} X^t Y \quad (45)$$

Le second modèle de régression est une régression dite dynamique (ou itérative) utilisant en temps réel la mesure pyranométrique précédente. Cette fois-ci, une version pondérée de la méthode des moindres carrés est utilisée, avec  $W$  une matrice de poids – ou matrice d'oubli – privilégiant les mesures les plus récentes dans la régression :

$$\hat{\alpha}_W = (X^t \Delta(W) X)^{-1} X^t \Delta(W) Y \quad (46)$$

avec  $\Delta(W) = \text{diag}((w_i)_{1 \leq i \leq n})$  l'opérateur de diagonalisation et  $W$  la matrice d'oubli définie arbitrairement par l'équation suivante :

$$\forall t \in [t_0 - t_w, t_0], \quad w(t) = \exp\left(\frac{t_0 - t}{5}\right) \quad (47)$$

avec  $t_w$  la taille de la fenêtre glissante et  $t_0$  l'instant présent.

Les résultats de cette régression dynamique sont affichés pour la journée du 22/04/2015 à la figure 96. Sur cette figure, l'estimation par la relation *a priori* et celle d'HelioClim corrigé sont également affichées. La relation *a priori* correspond à la courbe théorique de rayonnement direct par ciel clair lorsque l'intersection des volumes potentiellement nuageux ne détecte pas de nuage dans la direction du Soleil. Cependant de nombreux passages nuageux ne sont pas détectés. Ces défauts de détections sont directement liés aux segmentations des deux images. En particulier, le halo autour du Soleil génère souvent une saturation dans la zone circumsolaire et rend impossible la détection de nuages, particulièrement aux alentours de midi. Enfin, grâce à l'utilisation de la rétroaction de la mesure précédente, l'estimation par régression dynamique suit de très près la mesure.

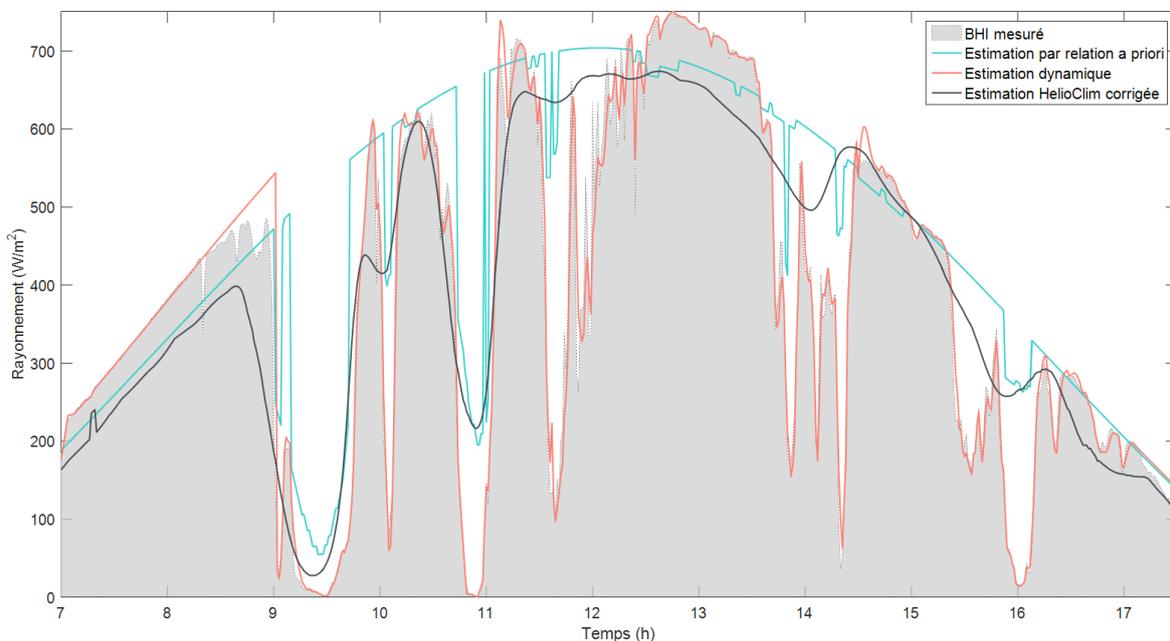


Figure 96 – Comparaison des trois méthodes de conversions du nombre de couches  $N_L$  nuageuses traversées dans la direction du Soleil en rayonnement solaire direct : l'estimation *a priori* du paragraphe V.3.1 (en turquoise), l'estimation HelioClim corrigée par le volume potentiellement nuageux du paragraphe V.3.2 (en noir), et l'estimation par régression dynamique du paragraphe V.3.3 (en rouge). Les séries temporelles correspondant aux estimations du 22/04/2015 sont présentées à titre illustratif.

Ces résultats de ces trois estimations sur cette journée d'exemple sont analysés à l'aide des outils décrits au chapitre II sur la figure 97. En termes de RMSE, de MAE ou de biais, l'estimation dynamique surpasse largement les deux autres. Le TDI indique une très faible distorsion temporelle de l'estimation dynamique par rapport à la mesure, mais cette distorsion est 15 % en retard d'après le TDM. En effet, pour tous les instants détectés comme ciel clair d'après la relation *a priori*, mais qui sont variables d'après la mesure, l'estimation dynamique effectuée en quelque sorte la persistance de l'instant précédent, d'où ce retard détecté. Cet écart de performances est attribué en grande partie à l'utilisation de la dernière mesure pyranométrique en rétroaction dans la régression dynamique.

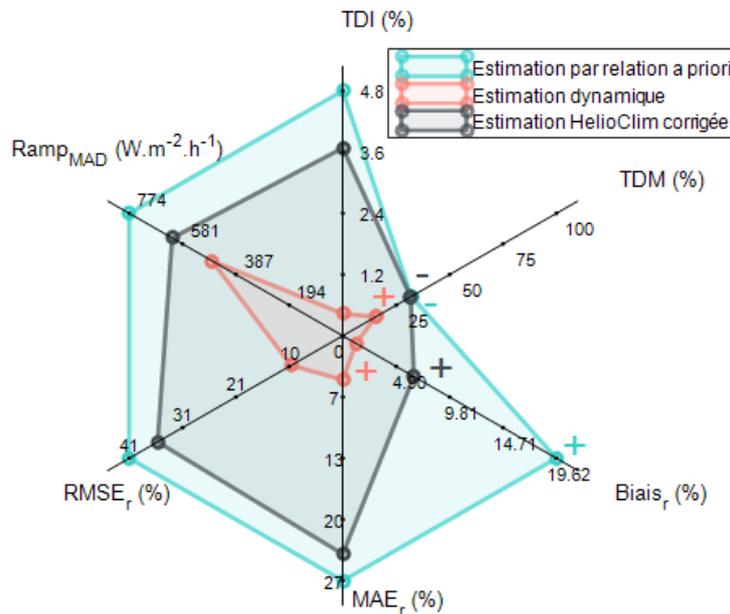


Figure 97 – Comparaison des trois méthodes de conversions du nombre de couches  $N_l$  nuageuses traversées dans la direction du Soleil en rayonnement solaire direct : l'estimation *a priori* du paragraphe V.3.1 (en turquoise), l'estimation HeliClim corrigée par le volume potentiellement nuageux du paragraphe V.3.2 (en noir), et l'estimation par régression dynamique du paragraphe V.3.3 (en rouge). Les résultats de ce graphique radar correspondent aux estimations du 22/04/2015 affichées à la figure 96.

Les deux plages (1) et (2) d'instant illustrées et commentées au paragraphe V.2.3 (*cf.* figure 91) sont analysées ici pour les résultats d'estimation : celle de 9h à 9h40 et celle de 10h30 à 11h10. Ces deux plages correspondent à des passages nuageux devant le Soleil qui ont eu la particularité d'être correctement détectés par les différentes sources d'information. Ainsi, lorsque la géolocalisation de ces nuages est correcte, le chemin optimal obtenu pour le calcul du TDI et du TDM est très proche de la bissectrice. Autrement dit, une meilleure précision temporelle est observée lorsque la géolocalisation 3D des nuages est proche de la réalité. Ceci confirme avec des données réelles des résultats conformes à la conclusion de l'étude sur le simulateur dynamique du chapitre IV.

## V.4 Prédiction du rayonnement solaire spatialement étendu

Les différentes étapes expliquées jusqu'à présent permettent d'obtenir, à partir d'une image satellite et d'une image hémisphérique, une estimation de la géolocalisation des masses nuageuses dans l'espace, puis une estimation de l'ombre portée des nuages au sol, et enfin une estimation du rayonnement solaire direct spatialement étendu. Ce paragraphe a pour objectif de proposer une manière d'exploiter ces trois étapes dans un contexte de prédiction solaire et non plus uniquement d'estimation. Le paragraphe V.4.1 explique le concept de la prédiction proposée en récapitulant les différentes étapes et leurs interactions proposées jusqu'à présent. Le paragraphe V.4.2 explique brièvement l'étape supplémentaire pour obtenir la prédiction du rayonnement solaire direct spatialement étendu, puis les résultats sur la journée d'exemple sont affichés et analysés au paragraphe V.4.3.

### V.4.1 Le principe

La figure 98 illustre la chronologie des traitements effectués menant à la prédiction proposée. La ligne de temps est affichée en haut de cette figure, pour laquelle la case rouge correspond à l'instant présent  $t_0$ , les intervalles oranges aux images satellites (HelioClim et CTH) toutes les 15 min et les intervalles bleus aux images hémisphériques toutes les minutes. Les intervalles en pointillés représentent les images futures. L'instant  $t_0 + h$  correspond à l'instant à prévoir avec l'horizon de temps  $h$  fixé à 10 min pour les besoins de l'illustration de la figure 98.

Les premières étapes du processus de prédiction consistent à estimer le mouvement des nuages entre les deux dernières images disponibles successives, du point de vue satellite et du point de vue de la caméra hémisphérique. Les CMV satellitaires sont ensuite utilisés pour propager les images HelioClim et les images de CTH de l'instant  $t^*$  (instant le plus récent de l'acquisition satellite) de manière à estimer ce que seraient ces produits satellitaires s'ils avaient été acquis à l'instant  $t_0$ , illustrées par une bordure en pointillés. Dans le cas de la prédiction, la propagation causale est donc appliquée (*cf.* figure 86.a du paragraphe V.1.2).

Dans le cas d'un développement opérationnel, il est important de noter qu'un certain laps de temps sera nécessaire à la génération des images HelioClim et CTH, de l'ordre de quelques minutes après les 15 min rondes. Ainsi, la dernière image HelioClim disponible pourrait dater de plus de 15 min, faisant elle-même référence à une prise de vue effective 4 min plus tôt. Nous estimons opérationnellement la propagation maximale de l'information satellitaire à 24 min. Cette contrainte opérationnelle ne sera pas considérée dans la suite de ce chapitre.

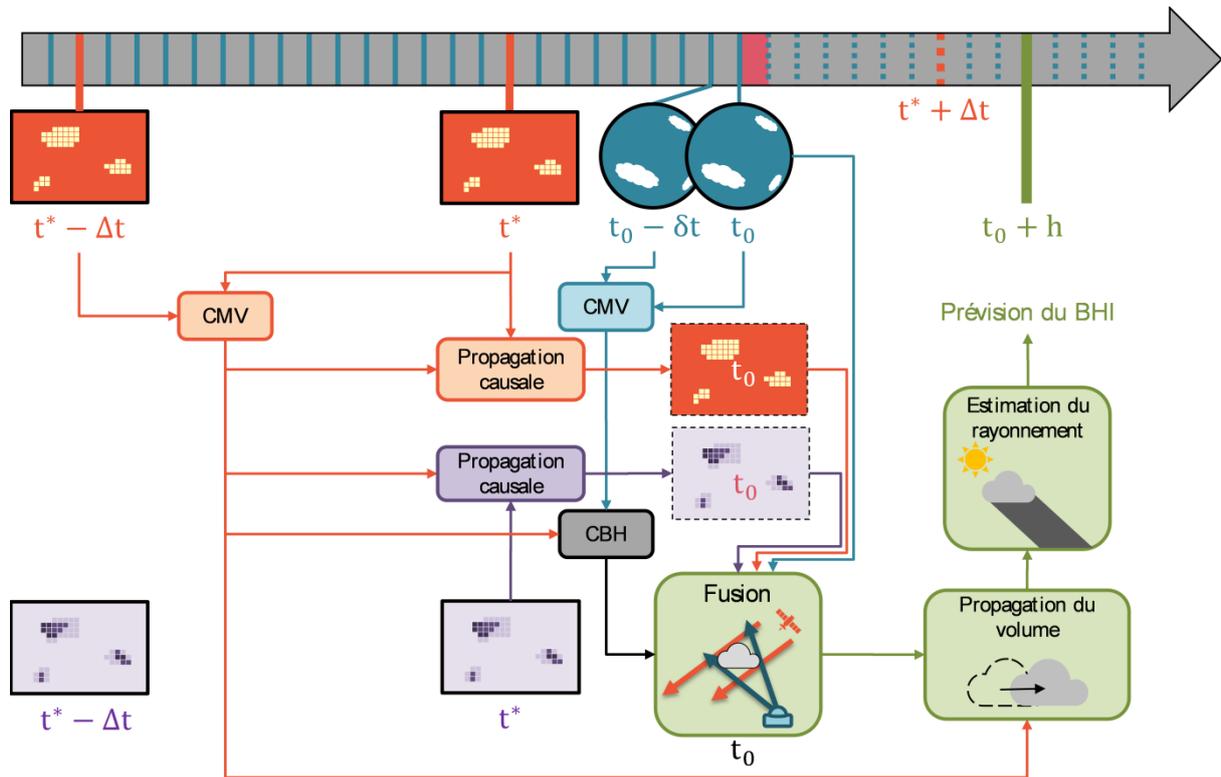


Figure 98 – Schéma récapitulatif des différents traitements réalisés au cours du processus prévisionnel proposé, accolé à un chronogramme illustrant les données temporelles utilisées pour une prévision réalisée à l'instant  $t_0$  d'horizon  $h$ . L'instant  $t^*$  correspond à l'instant de réception de la dernière image satellite généralement sur 15 min rondes.  $\Delta t$  est la résolution temporelle des images satellites de 15 min,  $\delta t$  est le pas de temps retenu pour l'étude de 1 min, c'est-à-dire qu'une image hémisphérique sur 6 est utilisée. L'instant en rouge correspond à l'instant présent  $t_0$  et les intervalles en pointillés représentent les instants d'acquisition des images futures. L'étape appelée CMV correspond à l'estimation du mouvement des nuages entre deux images successives, utilisé pour propager les informations temporellement. L'étape appelée CBH correspond à l'estimation de l'altitude de la base des nuages. En orange sont représentées les images Helioclim, en bleu les images hémisphériques et en violet les images CTH.

Pour ce qui concerne les images hémisphériques, leur pas de temps de 10 s est suffisamment court pour utiliser la dernière image disponible, sans besoin de prévision d'images futures. Les CMV issus des deux dernières images hémisphériques et des deux dernières images satellites sont exploités dans le calcul du CBH.

Enfin, l'estimation du CBH, l'image Helioclim propagée à l'instant  $t_0$ , l'image CTH propagée à l'instant  $t_0$  et l'image hémisphérique courante sont fusionnées via la méthode de géolocalisation proposée. Cette étape effectue l'intersection des différents volumes potentiellement nuageux construits à partir de chaque source d'information indépendamment. Plusieurs voies peuvent permettre la prévision du rayonnement solaire à partir du volume potentiellement nuageux final. Le processus proposé ici consiste à simplement propager le volume des possibles selon le CMV estimé du point de vue satellitaire. Cette étape suppose donc la persistance de la forme du volume potentiellement nuageux de l'instant présent à l'horizon considéré pour la prévision, ce qui peut être loin de la réalité des mouvements complexes suivant plusieurs couches de nuages avec des déformations des contours.

La nouvelle position du volume des possibles permet d'estimer la future valeur du rayonnement direct selon la méthode suggérée dans la partie V.3. La prévision à  $t_0 + h$  exploite donc les dernières images disponibles pour estimer la géolocalisation des nuages à l'instant présent  $t_0$ , information ensuite propagée dans l'espace pour réaliser la prévision de la position de la couverture nuageuse à l'horizon de prévision  $h$  puis de son rayonnement direct induit, spatialement étendu.

L'estimation des CMV est systématiquement réalisée en utilisant l'algorithme proposé par Liu (2009). Cette estimation du déplacement des nuages est décrite et discutée dans l'annexe A1. La propagation causale des images selon le champ de CMV estimés est expliquée, quant à elle, dans l'annexe A2. L'estimation du CBH est décrite au paragraphe V.2.2.b et l'estimation du rayonnement au paragraphe V.3. Le paragraphe V.4.2 suivant décrit la propagation du volume potentiellement nuageux de l'instant  $t_0$  à l'instant  $t_0 + h$  à prévoir.

#### **V.4.2 Propagation des volumes potentiellement nuageux**

La géolocalisation des nuages proposée est plus ou moins fidèle au volume réellement occupé par les nuages. Les écarts proviennent à la fois de la segmentation en deux classes ciel clair / ciel couvert qui simplifie la complexité de la définition d'un nuage en une information binaire de présence de celui-ci. Ces écarts sont également inhérents à la méthode de géolocalisation des nuages, tels qu'analysés par la simulation au chapitre IV. Les résultats de la simulation statique comme dynamique ont mis en évidence l'amélioration de géolocalisation apportée par la connaissance de l'altitude de la base et du sommet des nuages (resp. CBH et CTH).

Ainsi, le volume qui a été choisi pour la propagation est celui issu de l'intersection de tous les volumes des possibles provenant de toutes les informations considérées, soit le troisième niveau d'information 'SOL+SAT+CTH+CBH'. En effet, sans utilisation des informations d'altitude des nuages, les cônes potentiellement nuageux issus de l'image hémisphérique à l'instant  $t_0$  seraient encore très présents dans le volume final et auraient un impact non négligeable sur la qualité de la prévision. Plus le volume est affiné et proche de la réalité, plus la propagation de ce volume est juste.

Il existe un grand nombre de choix pour effectuer la propagation du volume potentiellement nuageux final. Seuls les déplacements estimés par les CMV satellitaires sont utilisés pour cette propagation. De plus, seul le vecteur déplacement moyen sur la zone étudiée est conservé et appliqué à l'ensemble du volume potentiellement nuageux de façon uniforme. Ces choix ont été faits afin de simplifier au maximum cette étape qui peut être envisagée de manière plus poussée. Toutefois, afin d'éviter de trop grandes discontinuités temporelles dans la prévision, un lissage temporel a été entrepris en utilisant la valeur médiane des déplacements estimés sur une fenêtre glissante contenant les deux derniers instants et l'instant présent.

### V.4.3 Résultats sur le jour d'exemple

La totalité des étapes illustrées sur la figure 98 est appliquée au jour d'exemple du 22/04/2015 sur le site n°2. Cependant, l'étape d'estimation du rayonnement solaire exprimée au paragraphe V.3 est difficilement applicable en l'état pour des prévisions de plusieurs minutes d'horizon. En effet, pour une prévision d'horizon  $h$ , les coefficients de régression estimés dynamiquement à l'instant  $t_0$  sont appliqués à l'instant  $t_0 + h$ . Dans le cas de l'estimation, la persistance des coefficients d'une minute à l'autre engendrait peu d'instabilité sur la sortie de la régression. Par contre, pour une prévision d'horizon de plusieurs minutes, la détermination dynamique des coefficients de régression est susceptible de s'éloigner fortement de la mesure. Ainsi, dans ce paragraphe, une approche de régression moins dynamique a été envisagée, où l'apprentissage des coefficients et l'application du modèle se font sur cette unique journée du 22/04/2015.

Bien entendu, le fait d'effectuer l'apprentissage et le test du modèle de régression sur la même période n'est pas approprié pour évaluer correctement les performances du modèle, mais l'objectif ici est plutôt de montrer un exemple de ce qui peut être extrait comme information de rayonnement solaire à partir des sorties de l'étape de géolocalisation des nuages, en mode prévision. De cette manière, les coefficients sont appliqués de façon identique sur toute la journée.

Trois horizons de prévision ont aussi été choisis : 5, 10 et 15 min. Pour chacun de ces horizons, l'estimation *a priori* (cf. V.3.1) et l'estimation décrite ci-dessus sont calculées et affichées pour la journée d'exemple sur les figures 99, 100 et 101 respectivement. Les méthodes de persistance (cf. définition au paragraphe II.7.2.a) 5, 10 et 15 min sont proposées en noir sur ces figures de manière à faciliter la comparaison avec une prévision dont le comportement est maîtrisé.

Sur chacune de ces trois figures, le comportement de la prévision sur la plage d'instant (1) de 9h à 9h40 est relativement bien déterminé. L'évènement est prévu, mais les pentes des rampes des prévisions *a priori* et par régression ne sont pas aussi abruptes que la mesure. Ce comportement s'explique principalement par l'étape de segmentation des images hémisphériques et la difficulté de quantifier l'atténuation du rayonnement direct sans information sur l'épaisseur optique des nuages. Ainsi, avant 9h, un nuage est détecté sur l'image hémisphérique mais a relativement peu d'impact sur la mesure et le Soleil reste « visible ». Ce type de comportement est moins problématique en estimation qu'en prévision puisque l'image hémisphérique informe à son instant d'acquisition sur l'occlusion effective du Soleil, grâce à la présence d'une zone saturée circumsolaire.

Certains décalages temporels peuvent être associés à l'étape de segmentation des nuages, mais il s'agit souvent d'un résultat combiné avec les erreurs d'estimations des CMV, de leur propagation – lagrangienne ou eulérienne – appliquée à tout le volume potentiellement nuageux ou encore de mauvaises estimations des altitudes et épaisseurs des nuages, en particulier pour les faibles élévations solaires. En observant les courbes des prévisions par relation *a priori* des figures 99, 100 et 101, il s'avère qu'elles sont qualitati-

vement alignées temporellement. Cette observation suggère ainsi que l'estimation des CMV est de bonne qualité et que l'hypothèse de non-déformation et d'advection des nuages semble respectée au moins sur la plage d'instant (1). En effet, une sous-estimation des CMV entraînerait une avance de la prévision de plus en plus marquée à mesure que l'horizon augmente.

Les erreurs d'amplitude associées à la conversion en rayonnement par la relation *a priori* peuvent à la fois être associées à la relation elle-même qui ne repose sur aucune réalité physique, mais aussi sur les erreurs de géolocalisation des nuages. La prévision par régression a justement un objectif de compenser les erreurs systématiques de la méthode *a priori*, mais d'autres méthodes de post-traitement dynamique corrigeraient certainement mieux les erreurs d'amplitude des prévisions.

Enfin, sur figures 99, 100 et 101, une faible variabilité des prévisions est constatée autour de 12h en sortie de l'étape de géolocalisation comparée à la mesure du BHI. Ce type de comportements est également associé en grande partie à la segmentation des images de nuages, où le seuil de décision fixe ne permet pas de capturer la variabilité spatiale contenue à l'intérieur des frontières des nuages détectés.

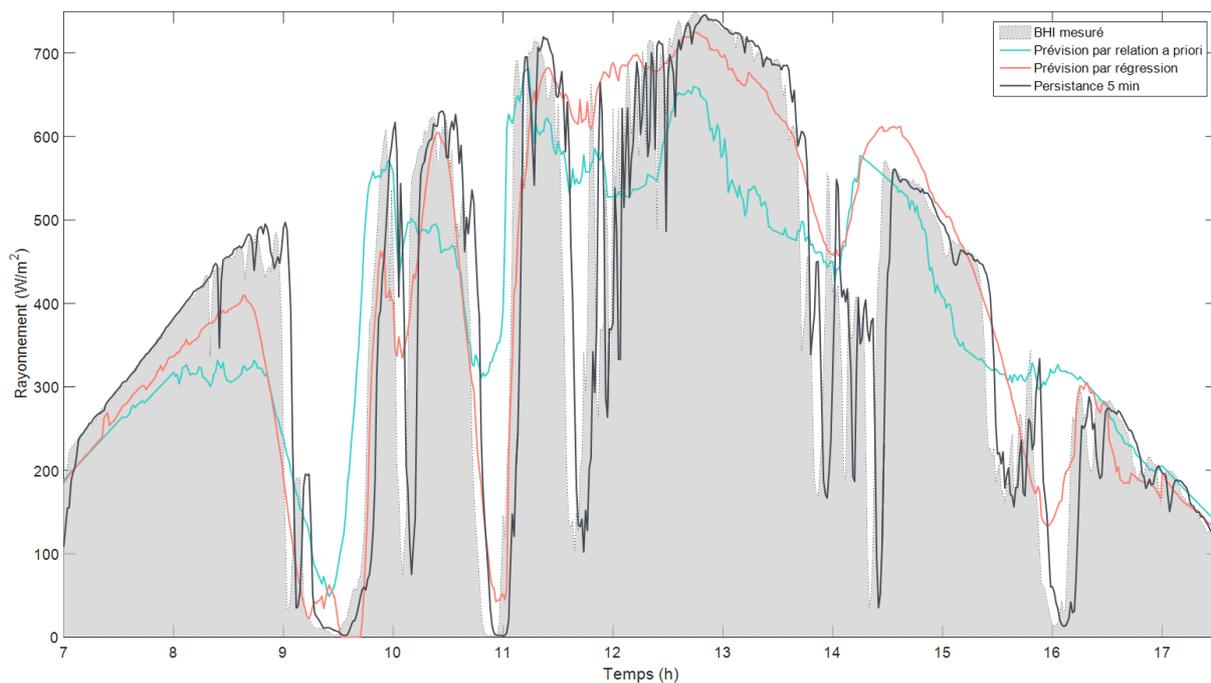


Figure 99 – Série temporelle des prévisions du BHI d'horizon 5 min à l'aide de la méthode de géolocalisation de niveau (3) 'SOL+SAT+CTH+CBH' associée à l'estimation *a priori* (en bleu) et par régression (en rouge). La méthode de prévision par persistance à l'horizon 5 min est affichée en noir. La mesure est représentée par une courbe en pointillés noirs dont l'aire sous la courbe est teintée de gris.

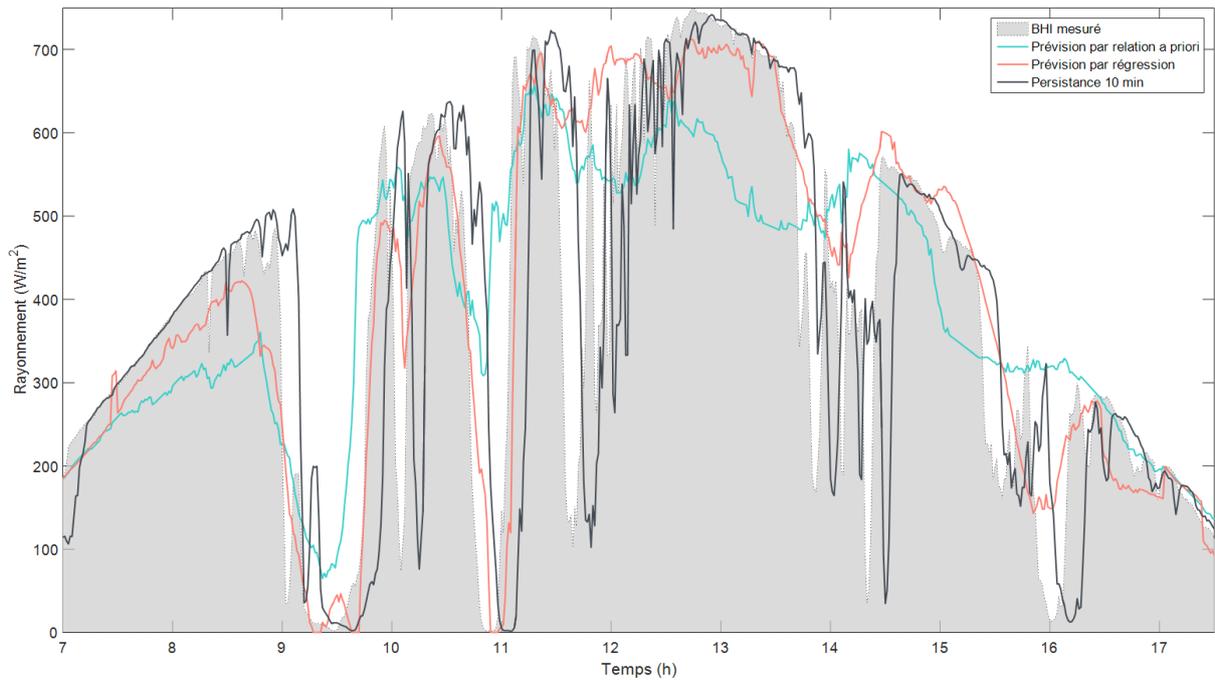


Figure 100 – Séries temporelles des prévisions du BHI d'horizon 10 min à l'aide de la méthode de géolocalisation de niveau (3) 'SOL+SAT+CTH+CBH' associée à l'estimation *a priori* (en bleu) et par régression (en rouge). La méthode de prévision par persistance à l'horizon 10 min est affichée en noir. La mesure est représentée par une courbe en pointillés noir dont l'aire sous la courbe est teintée de gris.

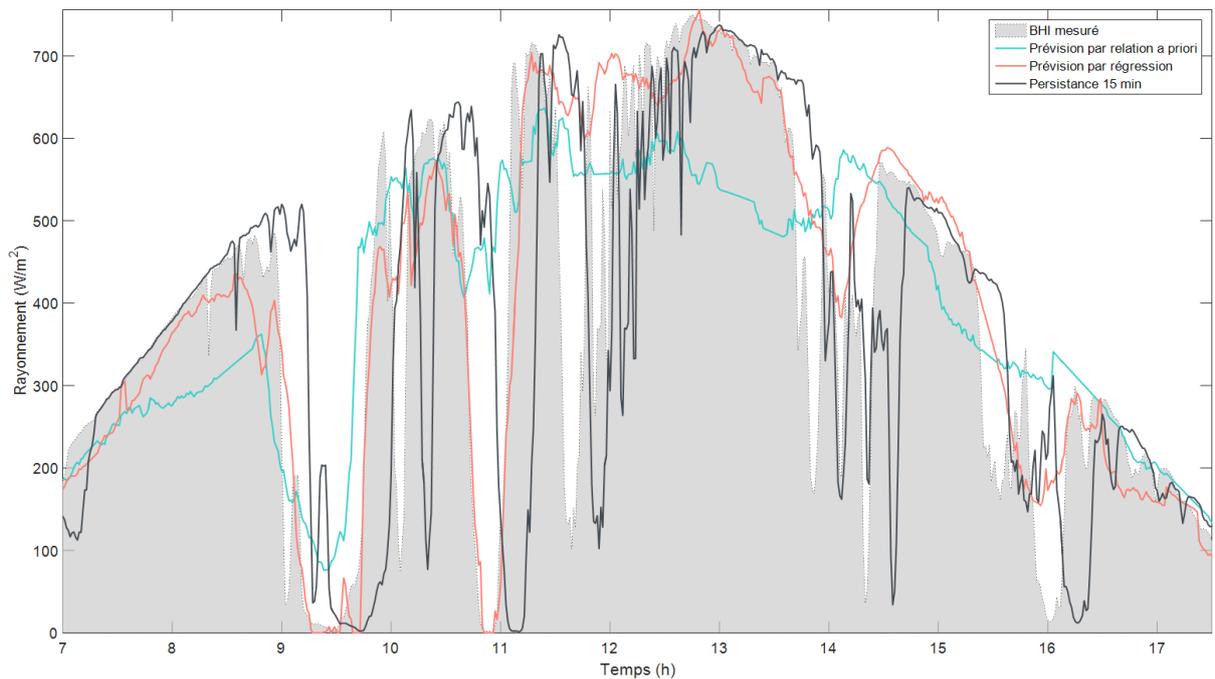


Figure 101 – Séries temporelles des prévisions du BHI d'horizon 15 min à l'aide de la méthode de géolocalisation de niveau (3) 'SOL+SAT+CTH+CBH' associée à l'estimation *a priori* (en bleu) et par régression (en rouge). La méthode de prévision par persistance à l'horizon 15 min est affichée en noir. La mesure est représentée par une courbe en pointillés noir dont l'aire sous la courbe est teintée de gris.

La figure 102 affiche les résultats des métriques décrites au chapitre II sous forme de graphe radar, appliquées aux résultats des méthodes des prévisions du BHI aux horizons 5, 10 et 15 min (cf. resp. figure 102.a, -b et -.c) pour la journée d'exemple. Ces résultats sont fondamentalement dépendants de la variabilité de cette journée, mais la persistance constitue une sorte d'étalon pour la comparaison.

D'après la figure 102.a, la prévision par régression et la persistance donnent des résultats très proches en termes de RMSE et de MAE pour l'horizon 5 min. Plus l'horizon augmente (cf. resp. figure 102.b et -.c), plus ces résultats se dégradent pour la persistance mais ceux de la prévision par régression restent à peu près les mêmes.

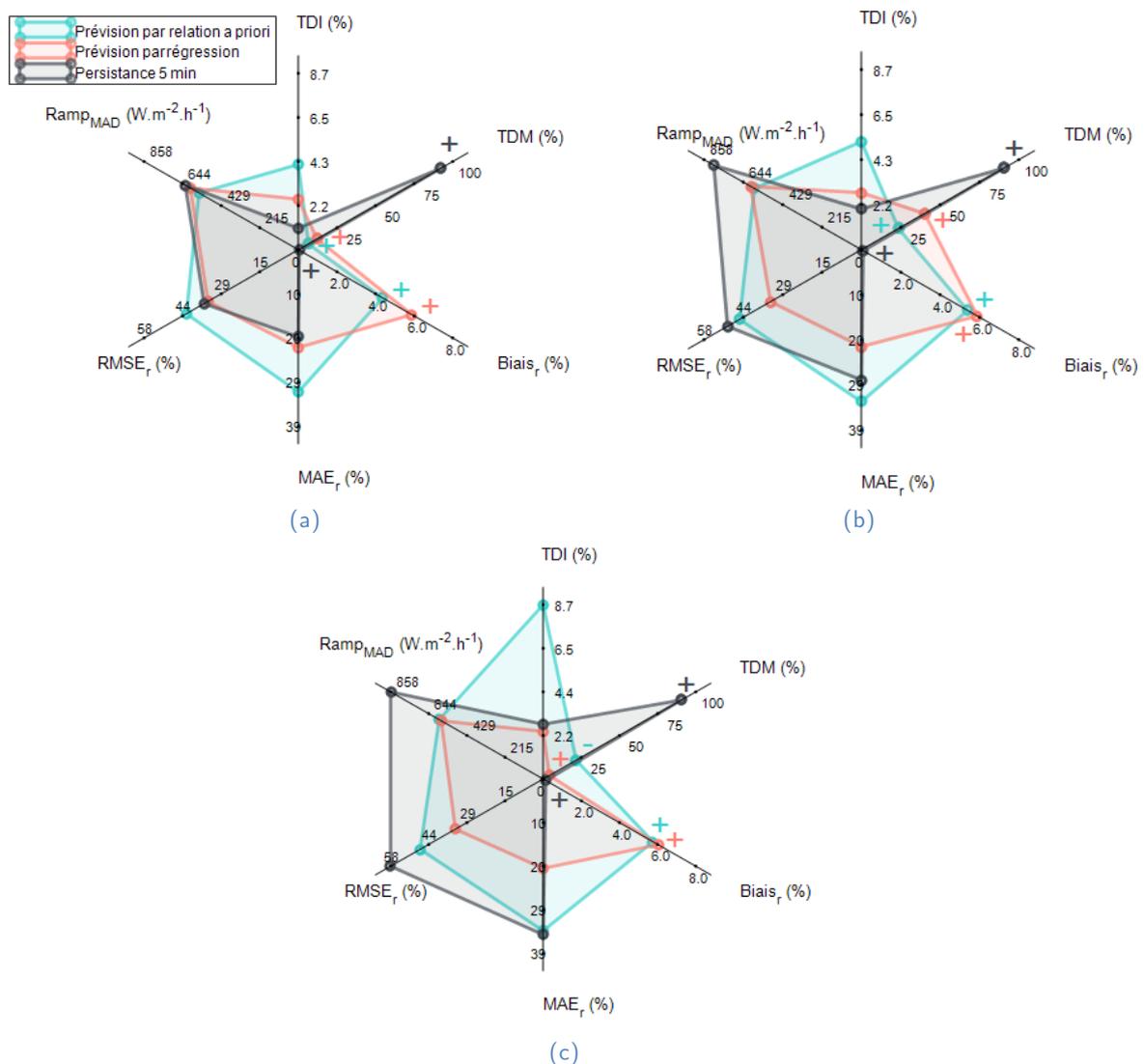


Figure 102 – Performances des prévisions du BHI d'horizon 5 min (a), 10 min (b) et 15 min (c) à l'aide de la méthode de géolocalisation de niveau (3) 'SOL+SAT+CTH+CBH' associée à l'estimation *a priori* (en bleu) et par régression (en rouge) sur la journée du 22/04/2015. La prévision par persistance 5 min (a), 10 min (b) et 15 min (c) est affichée en noir pour la comparaison.

Le biais reste nul pour la méthode par persistance puisque, par construction, elle reprend les précédentes valeurs de la mesure. Les biais pour les trois horizons de prévision sont plutôt stables pour la méthode par régression.

Enfin, le TDI quantifie la distorsion temporelle entre chaque méthode de prévision et la mesure. Cette distorsion augmente considérablement pour la prévision par relation *a priori* à mesure que l'horizon augmente. Une distorsion temporelle plus importante est également constatée pour la prévision par persistance à mesure que son horizon augmente. Cependant, la prévision par régression permet d'assurer une distorsion temporelle relativement stable avec l'horizon.

Le TDM, tel qu'attendu, indique que la persistance est systématiquement en retard. La prévision par relation *a priori* à 5 min montre un très bon alignement temporel moyen, qui se dégrade légèrement aux horizons 10 et 15 min. Concernant la prévision par régression, l'alignement temporel moyen est moins bon pour l'horizon 10 min que pour les horizons 5 et 15 min. Ces résultats du TDM sont à mettre en perspective des résultats du TDI puisqu'il s'agit du pourcentage de la distorsion temporelle attribuable à l'avance ou au retard. Ainsi, la régression diminue la distorsion temporelle par rapport à la relation *a priori*, mais cette modification influence aussi l'équilibre moyen entre le retard et l'avance. Le comportement aligné de la prévision par régression à 15 min devra être analysé plus précisément sur un support temporel bien plus étendu qu'une simple journée.

Sur ces trois graphiques (*cf.* figure 102.a, -.b et -.c), les résultats de la métrique ramp-MAD, les méthodes exploitant la géolocalisation des nuages restituent plus fidèlement les rampes que la méthode pas persistance pour les horizons 10 et 15 min. Ainsi, ce résultat suggère que le suivi des rampes de la mesure peut être amélioré par une meilleure géolocalisation des nuages dans l'espace. Ces résultats sont cohérents avec ceux issus du simulateur dynamique proposé au chapitre IV.

## V.5 Conclusions du chapitre

Ce chapitre a été l'occasion de proposer une procédure exploitant la méthode de géolocalisation des nuages par fusion proposée au chapitre IV et menant à l'estimation et à la prévision du rayonnement solaire sur le site étudié. L'espace 3D est donc le référentiel géométrique commun privilégié pour effectuer une telle fusion, mais un référentiel temporel commun au pas de temps 1 min a également été choisi de manière à établir les estimations et prévisions à haute résolution temporelle. Afin de se ramener à une donnée satellitaire au pas de temps 1 min, l'annexe A2 décrit une méthode causale et une méthode non-causale permettant de modéliser, à l'aide des CMV, les images satellites à n'importe quel instant.

Tout au long de cette procédure, une succession de choix a mené à une première estimation et une première prévision à 5, 10 et 15 min. Dans l'adaptation de la méthode de géolocalisation des nuages au cas réel, une segmentation des images hémisphériques a été proposée en se basant essentiellement sur des approches publiées. Notamment, la construction d'un abaque – ou LUT – a permis de générer des images de ciel clair pour tout instant en se basant sur la position de chaque pixel relativement au Soleil et au Zénith. Nous avons proposé une nouvelle entrée à cette LUT constituée du niveau de turbidité de

l'atmosphère via le trouble de Linke, estimé à partir des paramètres issus de CAMS. Ces images de ciel clair ont permis de faciliter et d'améliorer l'étape de segmentation des images hémisphériques. L'opérateur colorimétrique standard NRBR a aussi été adapté de manière à s'affranchir des effets de *flares* rouges sur cette segmentation. Enfin, une segmentation simple des images satellites a été proposée par simple seuillage des images HelioClim.

D'autres choix ont suivi dans les étapes d'estimation des altitudes du sommet et de la base des nuages (resp. CTH et CBH). Les images CTH fournies par le KNMI ont été utilisées comme information d'altitude du sommet des nuages bien que leur incertitude est de l'ordre de 2 km en écart type. Concernant l'estimation de l'altitude de la base des nuages, une estimation du CBH proposée par Wang et al. (2016) a été adaptée à notre étude en comparant des vitesses estimées à partir d'images hémisphériques et d'images satellites successives. Cette estimation donne une valeur de CBH unique pour toute la scène, ce qui peut être très réducteur dans le cas de nuages multicouches.

Enfin, après application de la méthode de géolocalisation des nuages, trois estimations du rayonnement solaire à partir du nombre de couches traversées  $N_L$  ont été proposées. Une méthode d'estimation par régression dynamique a donné les meilleurs résultats, mais de nombreuses autres retranscriptions en rayonnement solaire peuvent être envisagées à cette étape. La version statique de cette régression est ensuite appliquée aux volumes potentiellement nuageux qui ont été translatés du dernier déplacement estimé entre deux images satellites par les CMV satellitaires. Comme pour l'estimation, les prévisions à 5, 10 et 15 min sont évaluées pour une journée d'exemple, mais une validation plus exhaustive sur une période bien plus étendue serait pertinente afin d'extraire plus finement les gains et limitations de cette approche.



---

# Chapitre VI - Conclusions et perspectives

---

## VI.1 Conclusions

Cette thèse a été l'occasion d'explorer de nouvelles approches permettant l'exploitation de synergies entre les images hémisphériques, les images satellites et les mesures pyranométriques *in situ* pour la prévision photovoltaïque à très court terme.

Dans le contexte général de l'analyse des performances de ces prévisions solaires, deux nouvelles métriques ont été proposées au chapitre II afin de permettre, avec les approches existantes, une validation et une analyse des performances plus complète et plus adaptée aux besoins applicatifs des prévisions. Plus particulièrement, les moyens d'analyse supplémentaires proposés dans le chapitre II ont pour objet de caractériser deux aspects importants : la capacité d'une méthode de prévision à restituer les rampes et sa capacité à être alignée temporellement avec la mesure. Ces métriques ont été conçues dans le but d'affiner la compréhension de la qualité des prévisions pour les utilisateurs comme pour les concepteurs de prévisions eux-mêmes, en allant au-delà des métriques classiquement utilisées. Dans cet élan, une approche multicritères a été envisagée et apparaît comme étant appropriée à l'identification et à la quantification des avantages et inconvénients de chaque méthode, s'affranchissant de ce fait des limites de l'utilisation d'une unique métrique comme, classiquement, la racine carrée de l'erreur quadratique moyenne (RMSE) ou la moyenne des valeurs absolues des erreurs (MAE). De plus, toujours dans le cadre de l'analyse de performance, le double intérêt d'utiliser des classifications préalables d'instantanés ou de journées a été souligné : l'analyse s'en trouve d'une part affinée en fonction de classes de variabilité, et d'autre part plus indépendante des données et des conditions météorologiques associées en catégorisant l'information de manière interprétable et transposable pour d'autres lieux.

Les observations des différences entre les sources d'informations considérées au chapitre III ont permis d'identifier certaines complémentarités et difficultés liées à leur utilisation concomitante. En particulier, les résolutions spatiales et temporelles des deux types d'images sont foncièrement différentes : elles perçoivent une même scène nuageuse de deux points de vue différents avec une couverture spatiale propre à chaque image. Certains effets particulièrement importants dans l'utilisation de ces images ont été mis en évidence sur les données de l'étude, comme notamment l'effet parallaxe lié à cette différence de point de vue entre la caméra hémisphérique et le satellite géostationnaire. De ces observations a émergé la nécessité de définir préalablement un référentiel commun afin d'y opérer la fusion envisagée des deux types d'images. Le référentiel géométrique qui est mis en avant dans cette étude est l'espace 3D de la couche d'atmosphère à l'aplomb de la

zone d'intérêt et de son voisinage direct. Concernant le référentiel temporel commun, il a été montré comme essentiel d'assurer une concordance temporelle des informations à fusionner, à une résolution temporelle de l'ordre de la minute.

Les évènements d'occlusions successives du Soleil par les nuages ont un impact très important sur la variabilité temporelle du rayonnement solaire incident avec notamment des successions de rampes très marquées de ses composantes directe et circumsolaire. Ces évènements ont aussi un impact très important dans la variabilité spatiale de ces dernières composantes avec des structures très marquées correspondant aux ombres portées des nuages. Rendre compte de ces évènements d'occlusion du Soleil par les nuages et être capable de les prévoir repose en grande partie sur une détermination et une prévision précise de la géolocalisation 3D des nuages à l'horizon souhaité.

Le référentiel commun spatial et temporel des deux sources d'information étant posé, les approches de fusion proposées pour améliorer la prévision ont ainsi consisté à affiner, par la complémentarité de ces sources d'informations, l'estimation de la géolocalisation des nuages et de propager cette information à partir des champs de déplacement nuageux les plus récemment estimés.

Compte tenu de la complexité des géométries des deux types d'images des nuages, de celle de la position et des déplacements des nuages et de celle de l'éclairement du Soleil, un simulateur géométrique spécifique a été proposé et décrit dans le chapitre IV. Ce simulateur a pour objectif de caractériser les performances de méthodes de géolocalisation 3D des nuages – méthodes exploitant certaines synergies inter-images. Plus précisément, il s'agit d'un outil permettant d'analyser les performances des méthodes de géolocalisation des nuages en termes d'ombre portée au sol, sans utilisation des données réelles.

Ce simulateur a permis de mettre en évidence que la méthode « simple » de géolocalisation, réalisant « uniquement » l'intersection des informations de présence de nuages issues de chaque image dans l'espace 3D, a des performances fortement dépendantes de la position angulaire relative des nuages par rapport à la position du satellite, ainsi que de la taille des nuages. Aussi, pour des nuages de taille inférieure aux pixels de l'image satellite, les performances de cette méthode simple sont meilleures en comparaison à une utilisation seule des informations satellitaires.

De plus, le simulateur a permis de démontrer le bénéfice apporté par l'inclusion de la donnée des altitudes de la base et du sommet des nuages dans le cas statique – sur la présence des ombres – comme dans le cas dynamique – sur le nombre de couches potentiellement nuageuses traversées dans la direction du Soleil en fonction du temps. En analysant ces séries temporelles simulées, il a également été possible d'étudier cette augmentation de performances vis-à-vis des métriques proposées au chapitre II. Une meilleure géolocalisation des nuages améliore la précision temporelle de l'estimation, à la fois pour le site où se trouve la caméra, mais également pour un site éloigné de quelques kilomètres.

Les méthodes d'expression des informations dans le référentiel spatio-temporel commun ainsi que les approches de fusion proposées et analysées via le simulateur ont ensuite fait l'objet d'une application sur des données réelles du site d'étude. La description et l'analyse de ces applications font l'objet du chapitre V.

L'étape incorporant les altitudes de la base et du sommet des nuages a été adaptée au cas réel, en utilisant les images d'altitude du sommet des nuages (CTH) proposées par le KNMI, ainsi qu'en appliquant une méthode d'estimation de l'altitude de la base des nuages. Cette dernière repose sur la comparaison entre la vitesse angulaire estimée depuis le sol par l'analyse temporelle des images hémisphériques et la vitesse linéaire estimée du point de vue satellitaire.

L'autre étape qui diffère fondamentalement des simulations du chapitre IV est l'étape de segmentation des images. En effet, de par la nature floue des frontières des nuages, une segmentation des images de ces derniers est toujours imparfaite. Les images HelioClim ou CTH peuvent faciliter l'identification des nuages sur les images satellites puisque les structures du sol n'y sont pas présentes. Cependant, la limite de résolutions spatiale et temporelle inhérente à l'observation satellitaire depuis l'orbite géostationnaire peut rendre cette segmentation complexe, voire fortement dégradée. Des difficultés additionnelles s'ajoutent dans le cas des images hémisphériques, où la distorsion optique, la turbidité de l'atmosphère, la pluie, les salissures du dôme, le halo du Soleil allant jusqu'à saturer des zones entières de l'image ou encore les caractéristiques de la chaîne d'acquisition entravent l'identification des nuages par simple seuillage colorimétrique.

Les méthodes de géolocalisation des nuages telles qu'analysées dans le simulateur ont alors été appliquées sur les images réelles d'une journée d'exemple choisie pour la diversité des périodes de variabilité de la couverture nuageuse. Il ne s'agissait pas de valider complètement la méthode, mais d'évaluer dans des cas détaillés les bénéfices et les difficultés de l'approche proposée. L'étape de conversion du nombre de couches nuageuses traversées dans la direction du Soleil en rayonnement solaire a été abordée de trois façons différentes, selon que l'on considère ou pas l'utilisation conjointe de remontée en temps réel des mesures pyranométriques *in situ*. Une relation *a priori*, une correction de l'estimation HelioClim et une régression dynamique ont été ainsi évaluées. Les techniques de régression ont montré les meilleurs résultats sur la journée étudiée : une meilleure précision temporelle de l'occurrence des événements a notamment été mise en évidence, à la fois en estimation et en prévision.

L'estimation donne de meilleurs résultats que les prévisions à 5, 10, 15 et 30 min. La propagation des nuages à l'horizon souhaité est une étape qui peut s'envisager de différentes manières. Dans cette thèse, la propagation du déplacement des nuages a été réalisée de manière simple en ne considérant que l'estimation du mouvement entre les deux dernières images satellites successives. Cette étape a un impact très important sur les performances des prévisions, mais certains comportements des nuages resteront très difficiles à prévoir. L'analyse des prévisions à 5, 10 et 15 min sur la journée d'exemple a per-

mis de mettre en lumière certaines caractéristiques apportées par la méthode de géolocalisation 3D des nuages. Notamment, conformément aux résultats issus du simulateur, la meilleure géolocalisation des nuages, améliore la géolocalisation de l'ombre portée de ceux-ci et semble entraîner un alignement temporel plus marqué sur les séries temporelles d'estimation comme de prévision du rayonnement solaire direct, vis-à-vis de la mesure.

## VI.2 Perspectives

### VI.2.1 L'analyse des performances des méthodes de prévision

L'analyse des performances suggérée dans cette étude mériterait d'être appliquée à d'autres méthodes de prévision – dont le comportement est établi –, sur d'autres horizons et résolutions temporelles de prévision, ainsi que sur d'autres lieux. Cela permettra de vérifier la pertinence des métriques dans d'autres circonstances et éventuellement adapter leurs paramètres intrinsèques.

La nouvelle métrique TDM proposée dans cette thèse a été définie de manière à indiquer le pourcentage de la distorsion temporelle quantifiée par le TDI attribuable à une avance ou un retard temporel. Cependant, une autre façon de représenter cette information pourrait consister en l'introduction d'un « biais temporel » et d'un « écart type temporel » qui complèteraient le biais et l'écart type d'amplitude.

Le TDI et le TDM proposé associé sont le résultat d'un algorithme de DTW dont les paramètres de recherche fenêtrée et de distance pour le calcul du chemin optimal sont déterminants. Aussi, ces métriques admettent certaines limites d'applicabilité, notamment lorsque les deux signaux à comparer sont de variabilités temporelles très différentes. Par exemple, il est conceptuellement difficile d'attribuer une avance ou un retard à un pic de la série de test qui n'est pas présent dans la série de référence. Ces métriques calculées sur un grand jeu de données conservent leur pertinence si les deux séries temporelles ont des échelles de variabilité similaires. Ces métriques sont intéressantes par l'information qu'elles délivrent, mais il sera tout de même nécessaire d'étudier ces limitations.

Enfin, bien que déjà publiée dans le journal international *Solar Energy* (Vallance et al., 2017), l'analyse des performances proposée gagnerait à obtenir une appropriation et un consensus au sein de la communauté scientifique. Pour ce faire, une démarche de partage des codes des métriques proposées a ainsi été entamée en partenariat avec Dahzi Yang de l'institut SIMTech à Singapour. Le domaine de la prévision solaire est régulièrement enrichi par des recommandations et publications issues de rencontres d'experts mondiaux du domaine, qui donnent souvent lieu à des partenariats multiples. La tâche 16 du programme PVPS<sup>12</sup> (*Photovoltaic Power Systems Programme*) de l'IEA en est un exemple de renforcement des collaborations internationales sur le sujet. De telles initiatives pour-

---

<sup>12</sup> Les informations sur le programme de collaboration PVPS sont disponible à l'adresse : <http://iea-pvps.org>

raient être complétées par le développement d'outils collaboratifs visant à démocratiser et uniformiser les dernières avancées proposées par les experts du domaine, sur l'analyse des performances comme sur les autres aspects du domaine de la prévision solaire.

### VI.2.2 Le simulateur géométrique

Le simulateur géométrique des acquisitions conjointes d'images issues de caméras hémisphériques et de satellite géostationnaire développé dans le cadre de la thèse a été exploité sur un ensemble de situations nuageuses simples. Cependant, des simulations avec des scènes nuageuses plus représentatives au-dessus d'un lieu donné permettraient d'obtenir une information sur les performances attendues plus proche de cas réels.

Cet outil peut également intégrer toute autre source d'information, même partielle, sur la géolocalisation des nuages. L'influence d'un ajout de nouvelles sources d'informations sur les performances de la méthode, comme par exemple des images satellites géostationnaires de meilleure résolution, ou des images issues de satellites en orbite polaire, ou encore d'autres caméras hémisphériques peut être évaluée grâce à ce simulateur. La détermination de la distance optimale – au sens de la géolocalisation des nuages – entre ces différentes caméras hémisphériques est un exemple d'études rendues possibles par le biais de ce simulateur.

Enfin, une amélioration conséquente du simulateur proposé consisterait à y intégrer des calculs de transfert radiatif, tel que proposé par Kurtz et al. (2017) et Mejia (2017). Leurs contributions associent la simulation des grandes structures de la turbulence avec le code de transfert radiatif en trois dimensions SHDOM (*Spherical Harmonic Discrete Ordinate Method*) proposé par Evans (1998), pour simuler des méthodes de prévision exploitant uniquement des images hémisphériques. Une telle approche complexifie singulièrement le simulateur mais ouvre un nouveau champ de perspectives comme celle de simuler la composante diffuse du rayonnement avec notamment les effets de réflexions multiples entre les nuages pour mieux comprendre les phénomènes de suréclairement observés au paragraphe III.4.5.

Par exemple, si la simulation s'avère être suffisamment proche de la réalité, il est envisageable de l'utiliser comme banc de test pour tout type de méthode de prévision utilisant des images hémisphériques ou satellites. Apporter un jeu de données simulées réalistes (images hémisphériques, images satellites, mesures pyranométriques, ...) pourrait servir de tests standardisés dans l'analyse des performances des prévisions. Ainsi, au lieu d'analyser les performances par comparaison à des mesures – dont la fiabilité est essentielle mais pas toujours assurée – elle pourrait s'effectuer directement sur les sorties d'un tel simulateur. L'étape de classification des types d'instant ou de types de jours serait alors remplacée en imposant différents types de configurations atmosphériques.

### VI.2.3 L'application de la fusion pour une géolocalisation des nuages dans le cas réel

De la même manière que pour le simulateur, l'approche de fusion permettant une géolocalisation des nuages en 3D pourrait profiter des avancées sur les transferts radiatifs rapides. Cela permettrait d'extraire une meilleure estimation du rayonnement solaire incident au sol à partir de la description de la forme et de la position des nuages en 3D.

Les méthodes de fusion pour la géolocalisation des nuages proposées se sont révélées très sensibles à la qualité des segmentations des deux types d'images en zones nuageuses et claires. Ces segmentations ne peuvent jamais être parfaites, induisant des erreurs sur la détermination des volumes potentiels de nuages. Une première approche pourrait profiter de la cohérence temporelle des nuages entre deux instants proches pour affiner la segmentation résultante ou même affiner le volume potentiellement nuageux à partir des informations passées.

Par ailleurs, la fusion pourrait s'effectuer non plus sur la base de segmentation binaire, mais sur la base d'une logique floue – qui en est une extension. Il n'est alors plus question de présence ou non d'un nuage, mais d'un degré de certitude de présence. L'avantage d'une telle représentation de l'information, est d'autoriser la prise en compte de l'incertitude et de l'imprécision des informations à combiner, et de ne pas imposer de décision de la présence d'un nuage ou non en amont. Par exemple, il est possible d'indiquer le peu de confiance dans l'information de présence d'un nuage dans la zone circumsolaire lorsque le Soleil sature cette zone. Aussi, des lois de probabilité peuvent être définies afin d'indiquer à partir d'une information donnée le degré d'ennuagelement de chaque voxel. La méthode de géolocalisation 3D proposée dans cette thèse pourrait alors s'articuler autour de la logique floue de la façon suivante :

- L'étape de segmentation actuelle des images hémisphériques doit être légèrement modifiée en sélectionnant non plus un seuil, mais toute une série de seuils successifs sur les opérateurs colorimétriques, générant une estimation du degré d'ennuagelement. Cette information floue peut être conditionnée à un degré de certitude, par exemple défini à partir des positions angulaires relatives de chacun des pixels par rapport au Soleil, d'une part, et par rapport au zénith, d'autre part ;
- L'étape de segmentation des images satellites peut être remplacée par une exploitation non plus de l'indice de ciel clair seuillé des images HelioClim mais des images d'indices d'ennuagelement, non seuillés donc plus graduel, également disponibles avec la méthode Heliosat-2 ;
- L'information sur les altitudes du sommet et de la base des nuages peut être associée à une incertitude qui a été estimée par des études antérieures. Ainsi, les images CTH du KNMI ont une incertitude en écart type de 2 km (Stengel et al., 2014). Cette information pourrait par exemple être incorporée au sein d'une fonction de répartition a

*priori* des altitudes des nuages pour déterminer une fonction d'appartenance floue dans la direction verticale ;

- Enfin, la fusion des informations qui était opérée jusqu'à présent via une simple intersection (opérateur ET) de volumes booléens, doit être adaptée à la logique floue. Les opérateurs de la logique classique trouvent leur extension en logique floue sous deux formalismes différents : la représentation probabiliste, où l'intersection de deux variables est le produit des deux, et la représentation de Zadeh (1975), où l'intersection est le minimum des deux. Ces deux approches pourront faire l'objet d'une analyse comparée pour évaluer la plus performante.

En plus de cette sensibilité à la segmentation des zones nuageuses, la propagation du mouvement récent des nuages sur l'horizon de prévision souhaité est une étape qui, elle aussi, s'est révélée être importante pour les performances des prévisions. Dans l'annexe A2, une méthode lagrangienne et une méthode eulérienne ont été proposées. Les deux méthodes sont des approximations de la réalité et ont chacune leurs limitations. Une propagation hybride alliant les avantages des méthodes lagrangienne et eulérienne exposées dans l'annexe A2 serait une piste d'amélioration des prévisions actuelles.

L'estimation du rayonnement solaire à partir de la prévision du nombre de couches potentiellement nuageuses et de l'indice moyen de ciel clair issu d'HelioClim le long du trajet du rayonnement direct peut être améliorée de bien des manières. En particulier, la rétroaction de la mesure *in situ* améliore considérablement la précision des prévisions solaires à très court terme, mais son efficacité diminue à mesure que l'horizon de prévision augmente. Une approche de type filtrage de Kalman ayant par exemple comme vecteur d'état les paramètres issus de la prévision du volume potentiellement nuageux peut être intéressante à envisager.

L'approche proposée de la géolocalisation 3D des nuages des images hémisphérique et satellitaires ouvre une nouvelle voie prometteuse pour une amélioration sensible des prévisions locales du rayonnement solaire incident à court et très court terme. Cependant, la qualité des prévisions résultantes doit être démontrée sur un jeu de données plus étendu que l'unique journée choisie en exemple pour le chapitre V.

De plus, plusieurs verrous doivent être encore levés avant l'utilisation industrielle et opérationnelle des méthodes proposées de manière exploratoire dans cette thèse. Notamment, les temps de calcul des différents traitements aboutissant à la prévision à très court terme doivent être réduits au maximum. Aussi, la disponibilité des différentes informations en temps réel est un enjeu majeur pour la prévision opérationnelle. Par exemple, les données satellitaires – images HRV, HelioClim ou CTH – ont un temps de transmission et de traitement qui ne peuvent pas être négligés.



---

# Annexe A1 - Estimation du mouvement des nuages

---

De nombreuses techniques d'estimation du mouvement des nuages ont été proposées ces dernières décennies (Menzel, 2001), et ont été utilisées dans le domaine de la prévision solaire dès le début des années 2000 (Lorenz et al., 2004) sur des images satellites pour en extraire des prévisions allant d'une heure à cinq heures (Blanc et al., 2017b; Diagne et al., 2013; Kleissl, 2013). Le mouvement des nuages est particulièrement complexe à estimer à partir d'images. D'abord, ce sont des objets mal définis avec des formes, des frontières et une transparence qui varient dans le temps, plusieurs couches de nuages peuvent se croiser, se former ou disparaître. Les vecteurs résultants de ces techniques d'estimation du mouvement des nuages sont appelés vecteurs déplacement des nuages (CMV, *cloud motion vectors*).

## A1.1. Techniques d'estimation du mouvement

Deux catégories de techniques sont principalement utilisées :

- les techniques reposant sur une corrélation croisée normalisée – de phase ou d'amplitude – entre blocs de pixels de taille fixe de deux images successives, que l'on nommera appariement par corrélation – ou correspondance de blocs (*block-matching* en anglais) ;
- les techniques de flot optique qui sont des méthodes différentielles ;

### A1.1.1. Appariement par corrélation

Le principe de cette technique consiste à effectuer des paires de blocs de pixels semblables – au sens de la corrélation croisée – entre deux images successives afin d'en déduire le déplacement de ces blocs – appelés zones de contexte – entre ces deux instants. Pour des raisons de performances matérielles, une zone de recherche est généralement imposée telle qu'illustrée sur la figure 103. Il en résulte un champ de vecteurs dont la densité dépend de la zone de contexte et du chevauchement entre les blocs choisis. Cette technique est très largement employée sur les images satellites (Dambreville, 2014; Lorenz et al., 2004; R Perez et al., 2010; Yang et al., 2013) mais est également appliquée sur les images hémisphériques (Kazantzidis et al., 2017; Marquez et al., 2013; Peng et al., 2015; Quesada-Ruiz et al., 2014). Il est courant d'avoir recours à des techniques de filtrage temporel et/ou spatial afin de s'affranchir de valeurs aberrantes issues de cet algorithme de détection de mouvement (Chauvin, 2016).

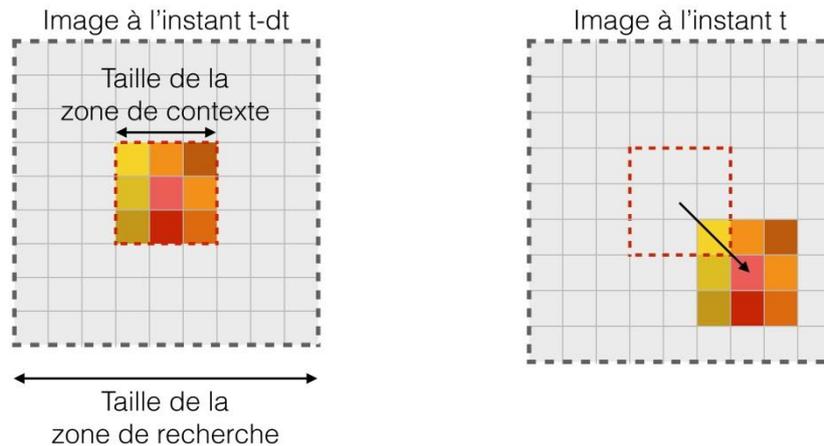


Figure 103 – Schéma explicatif de la technique d'appariement par corrélation estimant le déplacement des nuages entre deux images successives.

Une technique analogue faisant également intervenir la corrélation croisée – mais cette fois dans le domaine de Fourier – est la méthode d'alignement d'images par corrélation de phase (Cros et al., 2014a; Huang et al., 2012). Elle ne sera pas considérée dans la suite de cette annexe car trop peu utilisée dans le domaine de la prévision solaire et parce qu'elle donne les moins bons résultats dans l'évaluation d'une méthode de prévision CMV par satellite proposée par Cros et al. (2014b).

### A1.1.2. Flot optique

Le flot optique désigne la détection de mouvements d'objets sur des images successives. Sa particularité est de faire une hypothèse de conservation de la luminosité, qui suppose que la valeur des pixels doit être cohérente le long du vecteur de flux, c'est-à-dire que les différences entre deux images consécutives sont alors supposées être uniquement dues à la translation locale des nuages :

$$\frac{\delta I}{\delta x} \cdot u + \frac{\delta I}{\delta y} \cdot v + \frac{\delta I}{\delta t} = 0 \quad (48)$$

avec  $I$  l'intensité de l'image,  $x$  et  $y$  les coordonnées des pixels,  $u$  et  $v$  les composantes horizontale et verticale du vecteur déplacement. Puisque le nombre de variables est supérieur au nombre d'équations, ce qui est souvent appelé problème d'ouverture du flot optique, il est nécessaire d'apporter des contraintes supplémentaires. Par exemple, la méthode Lucas-Kanade (1981) est une approche locale du problème où l'hypothèse de conservation de la luminosité est supposée valide dans le voisinage spatial. Les vecteurs de déplacement correspondants sont ensuite déterminés localement via ces différences grâce à une approximation locale de Taylor et la méthode des moindres carrés. Ce type de résolution du flot optique a par exemple été utilisé dans le domaine de la prévision solaire par Cros et al. (2015), ou encore Wood-Bradley et al. (2012).

Dans la suite de cette annexe, l'algorithme de flot optique conçu par Liu (2009) sera utilisé comme référence. Cet algorithme<sup>13</sup> exploite la méthode de sur-relaxation successive

<sup>13</sup>Codes Matlab et C++ disponibles à l'url suivante : <https://people.csail.mit.edu/celiu/OpticalFlow/>

de résolution d'un système linéaire de la forme  $Ax = b$ , qui permet de résoudre très rapidement le problème d'ouverture grâce à un facteur de régularisation noté  $\alpha$ . Les écarts globaux aux hypothèses de conservation de luminosité et de régularité sont mesurés par la fonction énergie  $E$  de l'équation (49). L'objectif est de trouver le champ de flux optique  $u$  et  $v$  qui minimise  $E$  :

$$E(u, v) = \int \varphi(I_t(x_t, y_t) - I_{t+1}(x_t + u, y_t + v)) dp + \alpha \int \varphi(|\nabla u|^2 + |\nabla v|^2) dp \quad (49)$$

avec  $p = (x, y)$  décrit un point de l'image,  $\alpha$  le terme de régularisation,  $\nabla$  est l'opérateur du gradient et  $\varphi$  est une fonction robuste (Black et Anandan, 1996). L'approximation numérique et la résolution du problème de minimisation de l'énergie sont plus précisément décrites par Liu (2009).

## A1.2. Banc d'essai pour l'estimation des CMV

Caractériser les performances d'une détection de mouvement est souvent effectué de manière qualitative ou quantitative par la caractérisation de la sortie d'une chaîne d'opérations délivrant une prévision solaire. Concrètement, l'évolution de la RMSE entre la prévision et la mesure du rayonnement solaire en fonction de l'algorithme CMV sélectionné est une procédure qui a été utilisée (Blanc et al., 2017b). Cependant, cette technique est dépendante des différentes étapes antérieures et postérieures générant la prévision. En soumettant une image de nuages – satellitaire ou hémisphérique – à un mouvement artificiel, il est possible de caractériser l'aptitude de chaque technique de CMV à suivre le mouvement imposé, sans déformations des nuages et surtout, indépendamment des autres étapes de la prévision. Aussi, à l'aide de ce mouvement artificiel, nous pouvons ajuster les paramètres des algorithmes d'estimation du mouvement des nuages de manière à effectuer la plus faible erreur sur la détection des mouvements d'advection – qui sont d'un intérêt particulier pour notre étude. Cette technique sera appliquée au cas des images satellites au paragraphe A1.2.1. Cependant, comme nous le verrons au paragraphe A1.2.2, cette méthode ne semble pas adaptée à la géométrie des images hémisphériques. Une seconde méthode sera alors proposée afin de déterminer les paramètres optimaux des algorithmes de flot optique et d'appariement par corrélation, et finalement choisir le plus performant.

### A1.2.1. Un mouvement artificiel des nuages imposé aux images satellites

L'image HelioClim du 01/01/2015 à 12h est arbitrairement sélectionnée pour servir de base à la caractérisation des CMV. Le déplacement suivant est imposé selon les colonnes  $x$  et les lignes  $y$  de l'image satellite à chaque pas de temps de 15 min :

$$\forall t \in [0.25, 24], \quad \begin{cases} dx = -2.5 \cos\left(\frac{t}{8}\right) \\ dy = 1.5 \cos\left(\frac{t}{2}\right) \end{cases} \quad (50)$$

avec  $t$  exprimé en heures. Du fait de la différence de taille des pixels en longitude et latitude, les déplacements en  $x$  et  $y$  correspondent à une même vitesse maximale de 30

km/h. L'évolution temporelle du déplacement imposé en x (respectivement en y) est représentée sur la figure 104.a (respectivement figure 104.b) en bleu.

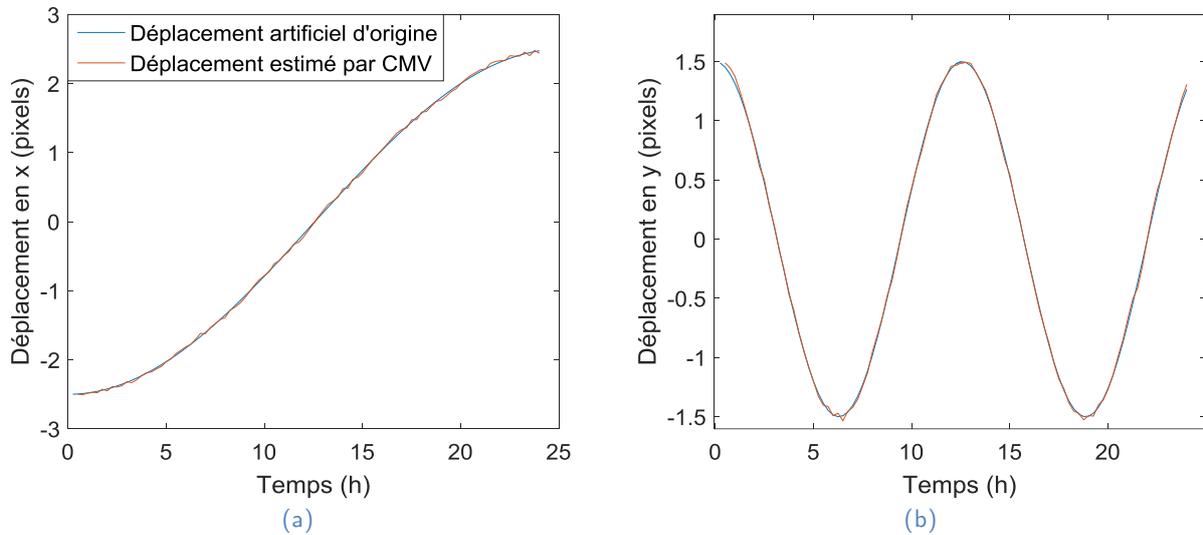


Figure 104 – Illustration du déplacement initialement imposé à l'image satellite – en bleu – à chaque pas de temps, comparé à son estimation obtenue sur le site n°1 issue de l'algorithme de CMV de Liu (2009) pour  $\alpha = 0,015$ . Leur composante en x est représenté sur la sous-figure (a) et celle en y sur la sous-figure (b).

Le mouvement décrit par l'équation (50) est imposé à chaque pixel de l'image HelioClim, puis en appliquant une interpolation linéaire pour générer l'image suivante sur la même grille d'origine. Une interpolation au plus proche voisin générerait des sauts brusques du déplacement, et une interpolation cubique modifie l'intensité globale de l'image ce qui rend caduque l'hypothèse de conservation de la luminosité. Enfin, pour faire face aux effets de bord, l'image choisie est répliquée de manière symétrique dans toutes les directions. Ces opérations assurent un mouvement strictement advectif des nuages dont le déplacement est connu et dans lesquelles les hypothèses du flot optique se trouvent dans leur domaine de validité ; Le mouvement artificiel sera ainsi utilisé comme un banc d'essai facilitant la caractérisation des performances des méthodes CMV.

L'algorithme de CMV proposé par Liu (2009) contient plusieurs paramètres d'entrée. Parmi ceux-ci, le paramètre fondamental à imposer est le facteur de régularisation  $\alpha$  compris entre 0 et 1, qui modélise la régularité spatiale du flot optique et pénalise les variations élevées du mouvement à détecter. Dans l'optique d'identifier une valeur de ce facteur de régularisation  $\alpha$  qui soit satisfaisante, l'algorithme de CMV est appliqué avec des valeurs de  $\alpha$  comprises entre 0,005 et 0,03. Le mouvement estimé au niveau du site n°1 est comparé au mouvement initialement imposé (n'importe quel point de l'image donnerait les mêmes résultats). La valeur de la RMSE en fonction de  $\alpha$  est représentée sur la figure 105, où la courbe bleue représente l'évolution de la RMSE pour le déplacement en x, et la courbe orange celle pour le déplacement en y. Dans la suite, la valeur  $\alpha = 0.015$  sera conservée à chaque application de cet algorithme de CMV sur les images satellites. L'estimation du mouvement issue de cette valeur de  $\alpha$  est représentée sur la figure 104.

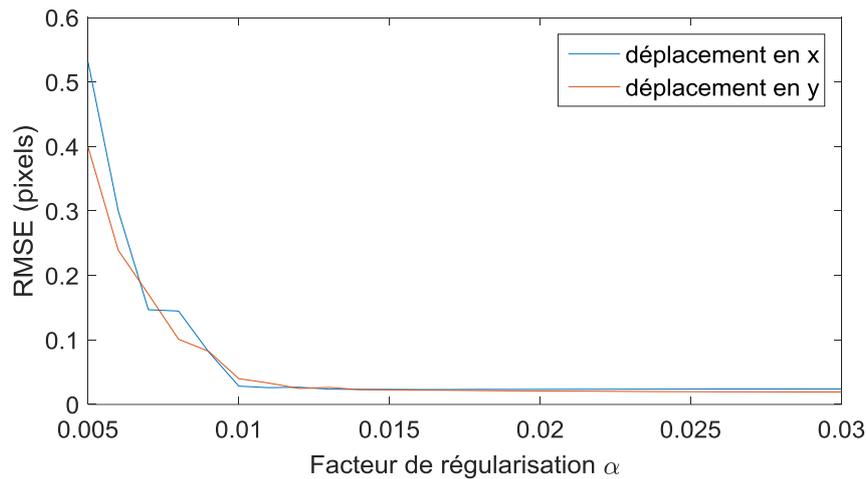


Figure 105 – Evolution de la RMSE calculée entre l'estimation du mouvement satellitaire en (x,y) par flot optique et le mouvement imposé initialement, en fonction du facteur de régularisation imposé en entrée de l'algorithme de Liu (2009).

À titre de comparaison, un algorithme d'appariement par corrélation a été appliqué. La taille de la zone de contexte est imposée à 5 pixels de large, et la taille de la zone de recherche évolue entre 7 et 15 pixels de large. Une légère amélioration de cet algorithme est également implémentée : lors de la recherche de la corrélation la plus grande entre le bloc de l'image initiale et le bloc de test, un barycentre est calculé sur le voisinage du maximum de corrélation afin d'améliorer la précision du vecteur déplacement estimé et ainsi obtenir des estimations sous-pixelliques. La figure 106 affiche la valeur de la RMSE calculée pour le mouvement synthétique choisi en fonction de la largeur de la zone de recherche. Ainsi, l'algorithme d'appariement par corrélation testé est moins performant que l'algorithme de flot optique de Liu (2009) avec  $\alpha \geq 0.007$  pour estimer le mouvement purement advectif proposé à l'équation (50). Enfin, pour la même paire d'images à traiter testée dix fois, l'algorithme d'appariement par corrélation met en moyenne 0.27 s à s'exécuter contre 0.17 s pour le flot optique de Liu (2009). L'algorithme d'appariement par corrélation n'est donc pas retenu pour l'estimation du mouvement des nuages à partir d'images satellites puisqu'il donne des résultats légèrement moins bons sur le mouvement synthétique testé et est moins rapide que le flot optique.

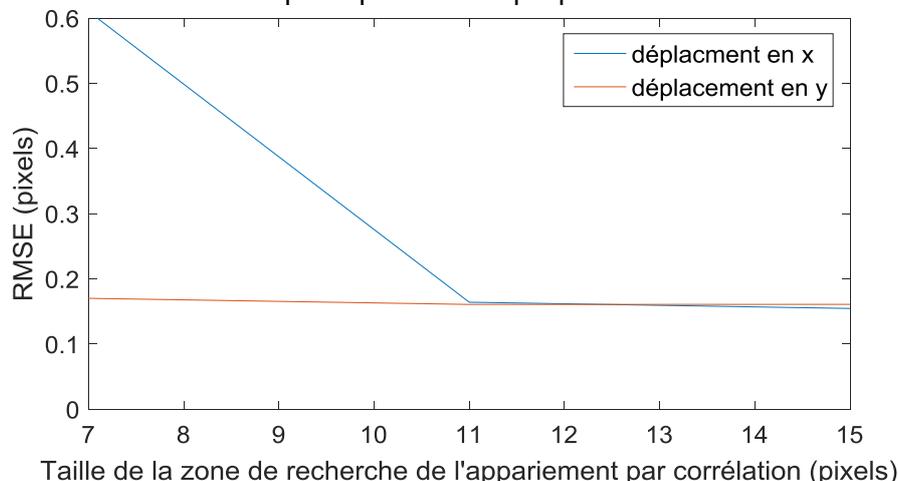


Figure 106 – Évolution de la RMSE calculée entre l'estimation du mouvement satellitaire en (x,y) par un algorithme d'appariement par corrélation et le mouvement imposé initialement, en fonction de la taille de la zone de recherche, et une zone de contexte de largeur 5 pixels.

### A1.2.2. Imposé aux images hémisphériques

Un nuage qui respecterait les hypothèses de mouvement purement advectif et de non-déformation n'apparaîtrait pas nécessairement entre deux images hémisphériques successives sans déformation. En effet, sur l'image hémisphérique, la distorsion optique génère une déformation géométrique lorsqu'un nuage d'épaisseur non négligeable se déplace horizontalement dans l'espace. De plus, les déformations des nuages semblent bien plus fréquentes avec la perception et la résolution des images hémisphériques qu'avec les images satellites où ces déformations sont moyennées par pixels de 3 km de côté. Ainsi, la méthode décrite au paragraphe précédent semble peu appropriée à l'évaluation des méthodes CMV sur les images hémisphériques.

La seconde méthode proposée dans ce paragraphe repose sur une évaluation pragmatique des CMV, en demandant à un opérateur expert d'apparier les 70 paires de points entre deux images hémisphériques séparées de 30 s, pour lesquels il a le plus grand degré de confiance dans l'association. La figure 107 illustre le résultat de l'appariement manuel proposé par l'expert entre deux images du 22/04/2015 à 12h15. Cet appariement est ensuite considéré comme le déplacement de référence.

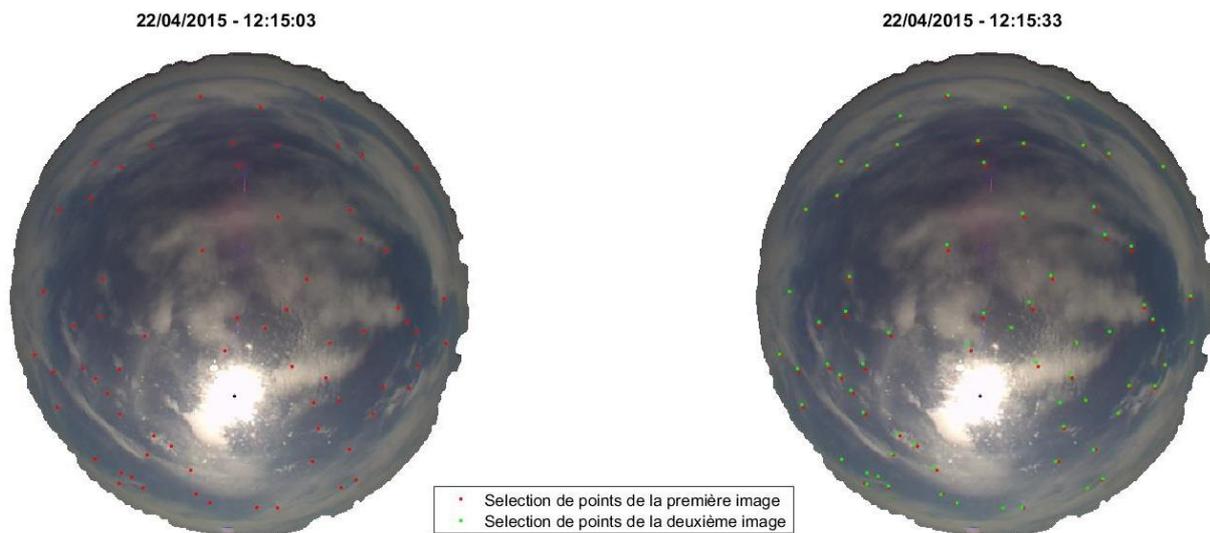


Figure 107 – Illustration de l'appariement de points sélectionnés par un opérateur expert entre une première image (en rouge) et une seconde image acquise 30 secondes plus tard (en vert).

Les algorithmes de flot optique et d'appariement par corrélation sont ensuite appliqués sur les images projetées dans un plan d'altitude arbitraire de 5 km, puis les vecteurs déplacements résultants sont reprojétés dans le repère de la caméra hémisphérique. Les vecteurs de déplacement ainsi estimés sont ensuite comparés à l'appariement manuel pour chaque valeur du facteur de régularisation (pour le flot optique) et de la taille de la zone de recherche (pour l'appariement par corrélation). La figure 108 (resp. 109) décrit l'évolution de la RMSE calculée selon les déplacements en x et y entre les CMV estimés par flot optique (resp. appariement par corrélation) et l'appariement manuel de l'expert selon le paramètre concerné. Les deux algorithmes convergent autour de la valeur de 1 pixel de RMSE, cependant les vitesses de calcul entre les deux sont très différentes : le

flot optique de Liu (2009) effectue l'estimation des CMV autour de 1,7 s quel que soit le facteur de régularisation choisi, alors que l'appariement par corrélation effectue l'estimation en 5,4 s pour une zone de recherche de taille 7 pixels jusqu'à 40 s pour une taille de recherche de 15 pixels. Dans le chapitre V, du fait de sa rapidité d'exécution et de sa précision, l'estimation des CMV à partir des images hémisphériques sera systématiquement effectuée à partir de l'algorithme de flot optique avec un facteur de régularisation de 1, assurant la plus faible RMSE sur les déplacements en x et en y.

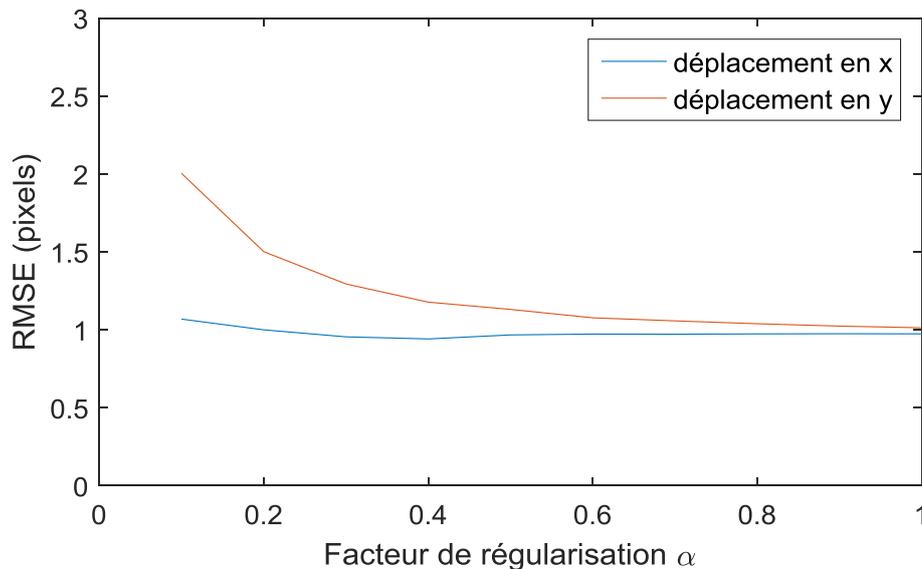


Figure 108 – Evolution de la RMSE calculée entre l'estimation du mouvement sur l'image hémisphérique en (x,y) par flot optique et le mouvement déterminé manuellement, en fonction du facteur de régularisation imposé en entrée de l'algorithme de Liu (2009).

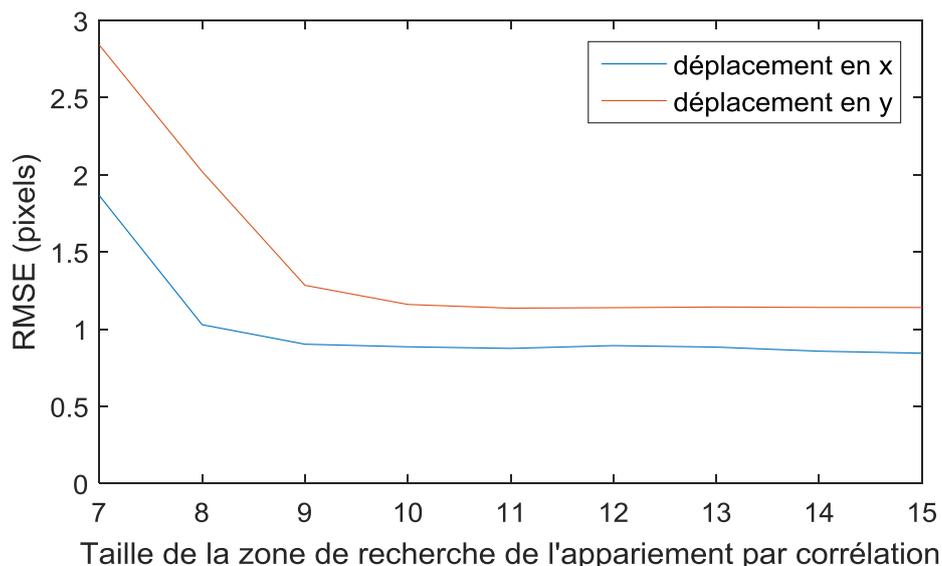


Figure 109 – Evolution de la RMSE calculée entre l'estimation du mouvement sur l'image hémisphérique en (x,y) par appariement par corrélation et le mouvement déterminé manuellement, en fonction de la taille de la zone de recherche, et une zone de contexte de largeur 5 pixels.



---

## Annexe A2 - Sur-échantillonnage temporel des images satellites

---

La limite de résolution temporelle des images satellites constitue un frein à l'utilisation concomitante de ces images avec les images hémisphériques. Dans cette thèse, un référentiel spatial commun est proposé afin d'effectuer la fusion de ces informations. Cependant, il est également nécessaire de travailler dans un référentiel temporel commun. À partir de cet objectif, plusieurs pistes peuvent être envisagées. Par exemple, il peut être suffisant de travailler au pas de temps 15 min et ainsi n'utiliser qu'un sous-ensemble des images hémisphériques disponibles. Cette question s'intègre dans un cadre plus large sur la manière d'effectuer la comparaison entre une estimation du rayonnement solaire et la mesure. Dans notre cas, il est souhaité d'obtenir des estimations et prévisions à haute résolution temporelle, d'où la recherche d'un échantillonnage à la minute de toutes les sources d'informations disponibles. Les mesures pyranométriques sont d'ores et déjà acquises au pas de temps 1 min et seule l'image hémisphérique la plus proche de la minute ronde est utilisée (c'est-à-dire une image sur six). Enfin, la difficulté de travailler dans un référentiel temporel commun à haute résolution réside dans l'étape de sur-échantillonnage temporel des images satellites, passant ainsi d'un pas de temps de 15 min à 1 min. Dans cette annexe, il a été choisi d'appliquer uniquement ce sur-échantillonnage temporel aux images HelioClim puisque leur estimation du rayonnement solaire intrinsèque autorise une comparaison avec les mesures *in situ*.

Appliquer une technique simple d'interpolation temporelle de manière directe en effectuant une moyenne pondérée de l'image présente avec l'image précédente pour en générer les images intermédiaires a deux défauts majeurs : d'abord cette méthode ne permet pas de prévoir les images intermédiaires à venir, et ne prend pas en considération le mouvement des nuages. Ainsi, cette méthode a tendance à générer des courbes moyennes d'estimation du rayonnement solaire extrêmement dépendantes de la méthode d'interpolation sélectionnée.

Cette annexe présente deux manières d'effectuer l'interpolation temporelle des images satellites avec une étape préliminaire d'estimation du mouvement des nuages entre deux images successives. L'algorithme de flot optique de Liu (2009) dont les paramètres d'implémentation sont discutés dans l'annexe A1 est utilisé afin d'extraire un champ de vecteurs déplacements des nuages (CMV) entre deux images successives. La première méthode présentée dans la suite est une méthode causale – qui ne dépend que des valeurs précédentes – qui permet de prévoir les images futures à partir de l'image présente et de l'image précédente. La seconde méthode au contraire utilise l'image présente et l'image

suivante – du « futur » – afin de générer des images intermédiaires de manière moins dépendante de l'estimation des CMV. Cette seconde méthode est donc utilisée pour générer des estimations *a posteriori* connaissant l'image suivante pour affiner le résultat des images intermédiaires.

## A2.1. Méthode de sur-échantillonnage causale

Dans le cas de la méthode causale, le problème revient à celui de la propagation des CMV déterminés entre les instants  $t - \Delta t$  et  $t$  à l'horizon  $h$  souhaité afin d'estimer l'image  $I_{t+h}$ . L'horizon peut par exemple varier de 1 à 15 min si l'objectif est de générer des images intermédiaires, ou peut dépasser 15 min si l'objectif est la prévision des images futures pour en déduire des prévisions de rayonnement. Une telle propagation a été envisagée de deux manières différentes. Une première méthode que l'on nommera eulérienne consiste à supposer un déplacement rectiligne et uniforme de chaque élément, c'est-à-dire que le vecteur déplacement calculé entre  $t - \Delta t$  et  $t$  est associé à la particule dont la vitesse restera inchangée tout au long de son parcours. Une deuxième méthode appelée lagrangienne consiste à considérer le champ de vecteurs comme constant au cours du temps, chaque particule suit ainsi le champ de vecteurs estimé. Ces deux méthodes sont simplificatrices de la complexité de l'évolution des masses nuageuses dans l'espace et le temps. L'approche eulérienne, bien que très simple dans l'attribution d'un vecteur vitesse constant, peut dans certains cas être préférable à l'approche lagrangienne. Cette dernière va en effet stopper la propagation des nuages si le déplacement estimé est nul sur une partie de l'image. Malgré cet artefact, uniquement la méthode lagrangienne sera utilisée et expliquée dans la suite de ce paragraphe.

Notons  $\delta p_{t-\Delta t, t}$  les déplacements des nuages en  $x$  et  $y$  entre l'image  $I_{t-\Delta t}$  et l'image  $I_t$ .  $\Delta t$  est le pas de temps initial de la séquence d'images satellites,  $dt$  est le pas de temps final du sur-échantillonnage. La formule itérative suivante délivre une estimation de l'image  $I_{t+h}$  à partir de l'image  $I_t$  et du champ de déplacement  $\delta p_{t-\Delta t, t}$  :

$$\forall h \in [dt; \Delta t], \quad \begin{cases} I_{t+h}(p_t) = I_t(p_{t+h}) \\ p_{t+h} = p_t - \frac{h}{\Delta t} \delta p_{t-\Delta t, t} \end{cases} \quad (51)$$

avec  $p_t$  les matrices de coordonnées de l'image  $I_t$ . Cette formulation suppose la persistance des matrices de déplacement  $\delta p_{t-\Delta t, t}$ , mais l'itération attribue le déplacement à différents pixels au cours du temps. Autrement dit, chaque particule suit les lignes du champ de vecteurs déplacements et se voit attribuer une vitesse qui varie au cours du temps. Cette méthode fournit des images sur n'importe quel horizon en se basant uniquement sur les deux dernières images disponibles, ce qui en fait une méthode adaptée à la prévision satellitaire du rayonnement solaire.

## A2.2. Méthode de sur-échantillonnage non causale

Sur-échantillonner temporellement les images satellites peut également s'envisager de manière non causale. Afin de générer les images intermédiaires, cette méthode propage de manière croisée à l'horizon souhaité  $h$  l'image à l'instant  $t$  et l'image à l'instant  $t + \Delta t$ , puis la moyenne pondérée de ces deux propagations est réalisée comme suit :

$$\forall h \in [dt; \Delta t], I_{t+h}(p) = (1 - \eta) \cdot I_t(p + \eta \cdot \delta p_{t,t+\Delta t}) + \eta \cdot I_{t+\Delta t}(p + (\eta - 1) \delta p_{t,t+\Delta t}) \quad (52)$$

avec  $\eta = 1 - \frac{dt}{\Delta t}$ , et  $\delta p_{t,t+\Delta t}$  le déplacement entre l'image  $I_t$  et  $I_{t+\Delta t}$ . L'illustration de cette méthode est disponible à la figure 110. Dans ce cas l'horizon  $h$  varie strictement entre le pas de temps du sur-échantillonnage et le pas de temps d'origine. La propagation effectuée à l'équation (52) peut être considérée comme une propagation eulérienne de l'image  $I_t$  vers le futur et de l'image  $I_{t+\Delta t}$  vers le passé.

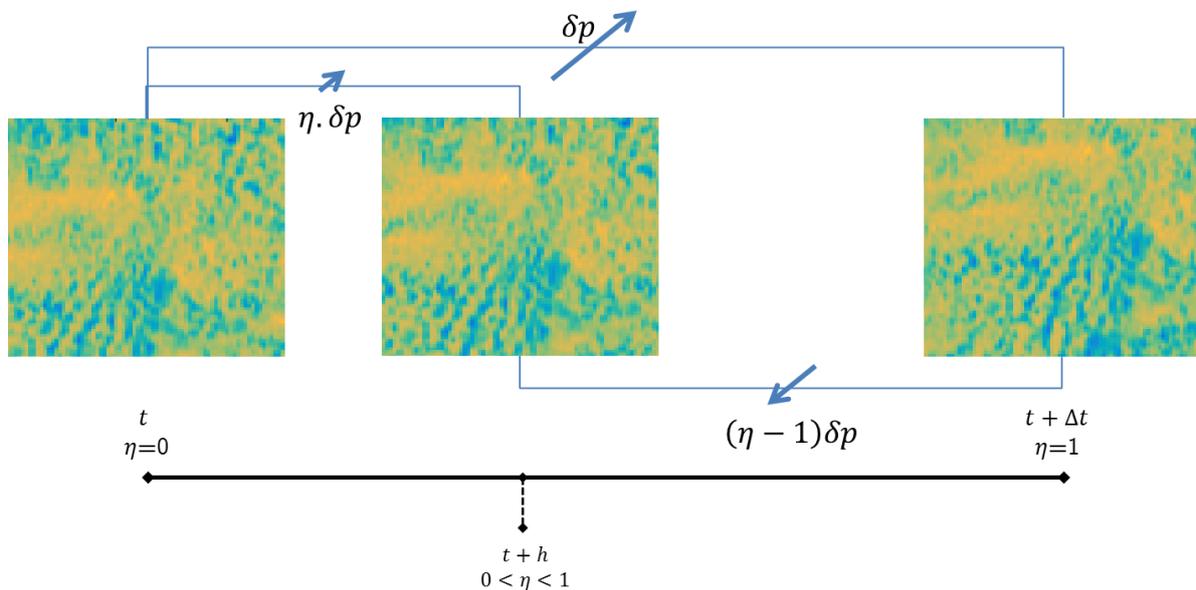


Figure 110 – Schéma explicatif de la méthode de sur-échantillonnage temporel des images satellites.

Ces deux méthodes – causale et non causale – ont été appliquées aux indices de ciel clair d'HelioClim sur le mois d'août 2015, et la comparaison avec la mesure pyranométrique donne des scores assez proches pour les deux méthodes d'environ 38% de RMSE. Cette erreur est en partie associée à la limite de résolution spatiale des images satellites qui empêchent une description précise de la localisation des nuages, de l'effet parallaxe et des erreurs inhérentes à la base de données HelioClim. En guise de comparaison, le suréchantillonnage au pas de temps 1 min sans utiliser les CMV, par simple interpolation linéaire entre deux points – donc non causal – donne une RMSE de 64 % sur le mois d'août 2015.



---

## Références

---

- Andersson, G., Donalek, P., Farmer, R., Hatziaargyriou, N., Kamwa, I., Kundur, P., Martins, N., Paserba, J., Pourbeik, P., Sanchez-Gasca, J., Schulz, R., Stankovic, A., Taylor, C., Vittal, V., 2005. Causes of the 2003 major grid blackouts in North America Europe, and recommended means to improve system dynamic performance. *IEEE Trans. Power Syst.* 20, 1922–1928. doi:10.1109/TPWRS.2005.857942
- Antonanzas, J., Osorio, N., Escobar, R., Urraca, R., Martinez-de-Pison, F.J., Antonanzas-Torres, F., 2016. Review of photovoltaic power forecasting. *Sol. Energy*. doi:10.1016/j.solener.2016.06.069
- Badescu, V., Gueymard, C.A., Cheval, S., Oprea, C., Baci, M., Dumitrescu, A., Iacobescu, F., Milos, I., Rada, C., 2013. Accuracy analysis for fifty-four clear-sky solar radiation models using routine hourly global irradiance measurements in Romania. *Renew. Energy* 55, 85–103. doi:10.1016/j.renene.2012.11.037
- Badosa, J., Wood, J., Blanc, P., Long, C.N., Vuilleumier, L., Demengel, D., Haefelin, M., 2014. Solar irradiances measured using SPN1 radiometers: Uncertainties and clues for development. *Atmos. Meas. Tech.* 7, 4267–4283. doi:10.5194/amt-7-4267-2014
- Bengulescu, M., 2017. Analyse temps-fréquence des données de rayonnement solaire reçu au sol. <http://www.theses.fr>.
- Bernecker, D., Riess, C., Angelopoulou, E., Hornegger, J., 2014. Continuous short-term irradiance forecasts using sky images. *Sol. Energy* 110, 303–315. doi:10.1016/j.solener.2014.09.005
- Bertin, C., Cros, S., Saint-Antonin, L., Schmutz, N., 2015. Prediction of optical communication link availability: real-time observation of cloud patterns using a ground-based thermal infrared camera, in: *Optics in Atmospheric Propagation and Adaptive Systems XVIII*. p. 96410A. doi:10.1117/12.2194920
- Beyer, H., Heilscher, G., Bofinger, S., 2004. A robust model for the MPP performance of different types of PV-modules applied for the performance check of grid connected systems. Eurosun. Freibg.
- Beyer, H., Martinez, J.P., Suri, M., Torres, J., Lorenz, E., Müller, S., Hoyer-Klick, C., Ineichen, P., 2009. Report on Benchmarking of Radiation Products. Report 70–74.
- Black, M.J., Anandan, P., 1996. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Comput. Vis. Image Underst.* 63, 75–104. doi:10.1006/cviu.1996.0006
- Blanc, P., Espinar, B., Geuder, N., Gueymard, C., Meyer, R., Pitz-Paal, R., Reinhardt, B., Renné, D., Sengupta, M., Wald, L., Wilbert, S., 2014. Direct normal irradiance related definitions and applications: The circumsolar issue. *Sol. Energy* 110, 561–577. doi:10.1016/j.solener.2014.10.001
- Blanc, P., Massip, P., Kazantzidis, A., Tzoumanikas, P., Kuhn, P., Wilbert, S., Schüler, D., Prah, C., 2017a. Short-term forecasting of high resolution local DNI maps with

- multiple fish-eye cameras in stereoscopic mode, in: AIP Conference Proceedings. doi:10.1063/1.4984512
- Blanc, P., Remund, J., Vallance, L., 2017b. 6 – Short-term solar power forecasting based on satellite images, in: Renewable Energy Forecasting. pp. 179–198. doi:10.1016/B978-0-08-100504-0.00006-8
- Blanc, P., Wald, L., 2012. The SG2 algorithm for a fast and accurate computation of the position of the Sun for multi-decadal time period. *Sol. Energy* 86, 3072–3083. doi:10.1016/j.solener.2012.07.018
- Bletterie, B., Pfajfar, T., 2007. Impact of Photovoltaic generation on voltage variations-how stochastic is PV. *CIREN 19th Int. Conf. Electr. Distrib.* 21–24.
- Blinn, J.F., 1982. A Generalization of Algebraic Surface Drawing. *ACM Trans. Graph.* 1, 235–256. doi:10.1145/357306.357310
- Boughorbel, S., Jarray, F., El-Anbari, M., 2017. Optimal classifier for imbalanced data using Matthews Correlation Coefficient metric. *PLoS One* 12. doi:10.1371/journal.pone.0177678
- Bristol, E.H., 1990. Swinging door trending: adaptive trend recording. *Proc. ISA Natl. Conf.* 45, 749–753. doi:10.1017/CBO9781107415324.004
- Buie, D., Monger, A.G., Dey, C.J., 2003. Sunshape distributions for terrestrial solar simulations. *Sol. Energy* 74, 113–122. doi:10.1016/S0038-092X(03)00125-7
- Calbó, J., Long, C.N., González, J.A., Augustine, J., McComiskey, A., 2017. The thin border between cloud and aerosol: Sensitivity of several ground based observation techniques. *Atmos. Res.* 196, 248–260. doi:10.1016/j.atmosres.2017.06.010
- Campbell, G., Purdom, J., Beyond, C.V., 1996. Asynchronous stereo height and motion estimation from multiple satellite images. *Spie* 2812, 95–110. doi:10.1117/12.254059
- Campbell, G.G., 1998. Asynchronous stereo height and motion analysis: applications.
- Campbell, G.G., Holmlund, K., 2004. Geometric cloud heights from Meteosat. *Int. J. Remote Sens.* 25, 4505–4519. doi:10.1080/01431160410001726076
- Cany, C., 2017. Interactions entre énergie nucléaire et énergies renouvelables variables dans la transition énergétique en France : adaptations du parc électrique vers plus de flexibilité. Université Paris-Saclay. <http://www.theses.fr>
- CEA, 2016. Mémento sur l'énergie.
- Chauvin, R., 2016. Évaluation de la ressource solaire pour la gestion optimisée de centrales CSP. <http://www.theses.fr>. Perpignan.
- Chauvin, R., Nou, J., Thil, S., Grieu, S., 2015. Modelling the clear-sky intensity distribution using a sky imager. *Sol. Energy* 119, 1–17. doi:10.1016/j.solener.2015.06.026
- Cheng, H.Y., 2016. Hybrid solar irradiance now-casting by fusing Kalman filter and regressor. *Renew. Energy* 91, 434–441. doi:10.1016/j.renene.2016.01.077
- Chicco, G., Cocina, V., Di Leo, P., Spertino, F., Massi Pavan, A., 2016. Error assessment of solar irradiance forecasts and AC power from energy conversion model in grid-Connected photovoltaic systems. *Energies* 9, 8. doi:10.3390/en9010008
- Chow, C.W., Urquhart, B., Lave, M., Dominguez, A., Kleissl, J., Shields, J., Washom, B., 2011. Intra-hour forecasting with a total sky imager at the UC San Diego solar energy testbed. *Sol. Energy* 85, 2881–2893. doi:10.1016/j.solener.2011.08.025

- Chu, Y., Pedro, H.T.C., Li, M., Coimbra, C.F.M., 2015. Real-time forecasting of solar irradiance ramps with smart image processing. *Sol. Energy* 114, 91–104. doi:10.1016/j.solener.2015.01.024
- Crispel, P., Roberts, G., 2018. All-sky photogrammetry techniques to georeference a cloud field. *Atmos. Meas. Tech.* 11, 593–609. doi:10.5194/amt-11-593-2018
- Cros, S., Liandrat, O., Sebastien, N., Schmutz, N., 2014a. Extracting cloud motion vectors from satellite images for solar power forecasting, in: *International Geoscience and Remote Sensing Symposium (IGARSS)*. pp. 4123–4126. doi:10.1109/IGARSS.2014.6947394
- Cros, S., Sébastien, N., Liandrat, O., Schmutz, N., 2014b. Cloud pattern prediction from geostationary meteorological satellite images for solar energy forecasting, in: *Spiedigitalibrary.Org*. p. 924202. doi:10.1117/12.2066853
- Cros, S., Turpin, M., Lallemand, C., Sébastien, N., Schmutz, N., 2015. Soleksat: A flexible solar irradiance forecasting tool using satellite images and geographic web-services, in: *31st European Photovoltaic Solar Energy Conference and Exhibition*. pp. 2128–2132.
- Cui, M., Zhang, J., Feng, C., Florita, A.R., Sun, Y., Hodge, B.M., 2017. Characterizing and analyzing ramping events in wind power, solar power, load, and netload. *Renew. Energy* 111, 227–244. doi:10.1016/j.renene.2017.04.005
- Cui, M., Zhang, J., Florita, A.R., Hodge, B.M., Ke, D., Sun, Y., 2016. An optimized swinging door algorithm for identifying wind ramping events. *IEEE Trans. Sustain. Energy* 7, 150–162. doi:10.1109/TSTE.2015.2477244
- Cui, M., Zhang, J., Florita, A.R., Hodge, B.M., Ke, D., Sun, Y., 2015. An optimized swinging door algorithm for wind power ramp event detection. *IEEE Power Energy Soc. Gen. Meet.* doi:10.1109/PESGM.2015.7286272
- Cutler, N., Kay, M., Jacka, K., Nielsen, T.S., 2007. Detecting, categorizing and forecasting large ramps in wind farm power output using meteorological observations and WPPT. *Wind Energy* 10, 453–470. doi:10.1002/we.235
- Da Silva Fonseca, J.G., Oozeki, T., Ohtake, H., Shimose, K.I., Takashima, T., Ogimoto, K., 2014. Forecasting regional photovoltaic power generation - A comparison of strategies to obtain one-day-ahead data, in: *Energy Procedia*. Elsevier, pp. 1337–1345. doi:10.1016/j.egypro.2014.10.124
- Dambreville, R., 2014. *Prévision du rayonnement solaire global par télédétection pour la gestion de la production d'énergie photovoltaïque*. Université Grenoble Alpes.
- Delta-Devices, T., 2011. *Sunshine Pyranometer The new SPN1 measures Global ( Total ) and Diffuse radiation and Sunshine Duration – in one instrument*.
- Derrien, M., Le Gléau, H., 2005. MSG/SEVIRI cloud mask and type from SAFNWC. *Int. J. Remote Sens.* 26, 4707–4732. doi:10.1080/01431160500166128
- Diagne, H.M., David, M., Lauret, P., Boland, J., 2012. Solar irradiation forecasting: state-of-the-art and proposition for future developments for small-scale insular grids. *World Renew. Energy Forum 2012* 1–8. doi:10.1016/j.rser.2013.06.042
- Diagne, M., David, M., Lauret, P., Boland, J., Schmutz, N., 2013. Review of solar irradiance forecasting methods and a proposition for small-scale insular grids. *Renew. Sustain. Energy Rev.* doi:10.1016/j.rser.2013.06.042

- Dobashi, Y., Nishita, T., Yamashita, H., Okita, T., 1999. Using metaballs to modeling and animate clouds from satellite images. *Vis. Comput.* 15, 471–482.  
doi:10.1007/s003710050193
- Du, P., Baldick, R., Tuohy, A., 2017. *Integration of Large-Scale Renewable Energy into Bulk Power Systems*, Springer. doi:10.1007/978-3-319-55581-2
- Ela, E., Diakov, V., Ibanez, E., Heaney, M., 2013. Impacts of Variability and Uncertainty in Solar Photovoltaic Generation at Multiple Timescales. *Natl. Renew. Energy Lab.*  
doi:NREL/TP-5500-58274
- Engerer, N.A., Mills, F.P., 2015. Validating nine clear sky radiation models in Australia. *Sol. Energy* 120, 9–24. doi:10.1016/j.solener.2015.06.044
- Escrig, H., Batlles, F.J., Alonso, J., Baena, F.M., Bosch, J.L., Salbidegoitia, I.B., Burgaleta, J.I., 2013. Cloud detection, classification and motion estimation using geostationary satellite imagery for cloud cover forecast. *Energy* 55, 853–859.  
doi:10.1016/j.energy.2013.01.054
- Espinar, B., Ramírez, L., Drews, A., Beyer, H.G., Zarzalejo, L.F., Polo, J., Martín, L., 2009. Analysis of different comparison parameters applied to solar radiation data from satellite and German radiometric stations. *Sol. Energy* 83, 118–125.  
doi:10.1016/j.solener.2008.07.009
- Evans, K.F., 1998. The Spherical Harmonics Discrete Ordinate Method for Three-Dimensional Atmospheric Radiative Transfer, *Journal of the Atmospheric Sciences*.  
doi:10.1175/1520-0469(1998)055<0429:TSHDOM>2.0.CO;2
- Ferreira, C., Gama, J., Matias, L., Botterud, A., Wang, J., 2011. A survey on wind power ramp forecasting. *Argonne Natl. Lab.* doi:10.2172/1008309
- Florita, A., Hodge, B.M., Orwig, K., 2013. Identifying wind and solar ramping events. *IEEE Green Technol. Conf.* 147–152. doi:10.1109/GreenTech.2013.30
- Frey, R.A., Ackerman, S.A., Liu, Y., Strabala, K.I., Zhang, H., Key, J.R., Wang, X., 2008. Cloud detection with MODIS. Part I: Improvements in the MODIS cloud mask for Collection 5. *J. Atmos. Ocean. Technol.* 25, 1057–1072.  
doi:10.1175/2008JTECHA1052.1
- Frías-Paredes, L., Mallor, F., Gastón-Romeo, M., León, T., 2017. Assessing energy forecasting inaccuracy by simultaneously considering temporal and absolute errors. *Energy Convers. Manag.* doi:10.1016/j.enconman.2017.03.056
- Frías-Paredes, L., Mallor, F., León, T., Gastón-Romeo, M., 2016. Introducing the Temporal Distortion Index to perform a bidimensional analysis of renewable energy forecast. *Energy* 94, 180–194. doi:10.1016/j.energy.2015.10.093
- Fu, C.L., Cheng, H.Y., 2013. Predicting solar irradiance with all-sky image features via regression. *Sol. Energy* 97, 537–550. doi:10.1016/j.solener.2013.09.016
- Gastón-Romeo, M., Leon, T., Mallor, F., Ramírez-Santigosa, L., 2011. A Morphological Clustering Method for daily solar radiation curves. *Sol. Energy* 85, 1824–1836.  
doi:10.1016/j.solener.2011.04.023
- Gauchet, C., Blanc, P., Espinar, B., Charbonnier, B., Demengel, D., 2012. Surface solar irradiance estimation with low-cost fish-eye camera. *COST WIRE Work. “Remote Sens. Meas. Renew. Energy”* 4 pp.
- Gaumet, J.L., Heinrich, J.C., Cluzeau, M., Pierrard, P., Prieur, J., 1998. Cloud-base

- height measurements with a single-pulse erbium-glass laser ceilometer. *J. Atmos. Ocean. Technol.* 15, 37–45.  
doi:10.1175/1520-0426(1998)015<0037:CBHMWA>2.0.CO;2
- GCOS, 2010. Implementation plan for the global observing system for climate in support of the UNFCCC. *World Meteorol. Organ.* 138, 180.
- Ghonima, M.S., Urquhart, B., Chow, C.W., Shields, J.E., Cazorla, A., Kleissl, J., 2012. A method for cloud detection and opacity classification based on ground based sky imagery. *Atmos. Meas. Tech.* 5, 2881–2892. doi:10.5194/amt-5-2881-2012
- Gueymard, C.A., 2017. Cloud and albedo enhancement impacts on solar irradiance using high-frequency measurements from thermopile and photodiode radiometers. Part 1: Impacts on global horizontal irradiance. *Sol. Energy* 153, 755–765.  
doi:10.1016/j.solener.2017.05.004
- Gueymard, C.A., 2012. Clear-sky irradiance predictions for solar resource mapping and large-scale applications: Improved validation methodology and detailed performance analysis of 18 broadband radiative models. *Sol. Energy* 86, 2145–2169.  
doi:10.1016/j.solener.2011.11.011
- Harrouni, S., 2008. Fractal Classification of Typical Meteorological Days from Global Solar Irradiance: Application to Five Sites of Different Climates, in: *Modeling Solar Radiation at the Earth's Surface*. pp. 29–55.
- Harrouni, S., Guessoum, A., Maafi, A., 2005. Classification of daily solar irradiation by fractional analysis of 10-min-means of solar irradiance. *Theor. Appl. Climatol.* 80, 27–36. doi:10.1007/s00704-004-0085-0
- Heinle, A., Macke, A., Srivastav, A., 2010. Automatic cloud classification of whole sky images. *Atmos. Meas. Tech.* 3, 557–567. doi:10.5194/amt-3-557-2010
- Hoff, T.E., Perez, R., 2012. Modeling PV fleet output variability. *Sol. Energy* 86, 2177–2189. doi:10.1016/j.solener.2011.11.005
- Hoff, T.E., Perez, R., 2010. Quantifying PV power Output Variability. *Sol. Energy* 84, 1782–1793. doi:10.1016/j.solener.2010.07.003
- Hoff, T.E., Perez, R., Kleissl, J., Renne, D., Stein, J., 2013. Reporting of irradiance modeling relative prediction errors. *Prog. Photovoltaics Res. Appl.* 21, 1514–1519.  
doi:10.1002/ppp.2225
- Horváth, Á., Davies, R., 2001. Simultaneous retrieval of cloud motion and height from polar-orbiter multiangle measurements. *Geophys. Res. Lett.* 28, 2915–2918.  
doi:10.1029/2001GL012951
- Huang, H., Yoo, S., Yu, D., Huang, D., Qin, H., 2012. Correlation and local feature based cloud motion estimation. *Proc. Twelfth Int. Work. Multimed. Data Min.* 1–9.  
doi:10.1145/2343862.2343863
- Ineichen, P., 2016. Validation of models that estimate the clear sky global and beam solar irradiance. *Sol. Energy* 132, 332–344. doi:10.1016/j.solener.2016.03.017
- Ineichen, P., 2008. Conversion function between the Linke turbidity and the atmospheric water vapor and aerosol content. *Sol. Energy* 82, 1095–1097.  
doi:10.1016/j.solener.2008.04.010
- Ineichen, P., 2006. Comparison of eight clear sky broadband models against 16 independent data banks. *Sol. Energy* 80, 468–478. doi:10.1016/j.solener.2005.04.018

- Inman, R.H., Pedro, H.T.C., Coimbra, C.F.M., 2013. Solar forecasting methods for renewable energy integration. *Prog. Energy Combust. Sci.* 39, 535–576.  
doi:10.1016/j.pecs.2013.06.002
- International Energy Agency, 2017a. Key world energy statistics.  
doi:10.1017/CBO9781107415324.004
- International Energy Agency, 2017b. World Energy Outlook 2017 763.  
doi:10.1787/weo-2017-en
- International Energy Agency, 2014. Solar Photovoltaic Energy, Technology Roadmap. International energy agency report. doi:10.1007/SpringerReference\_7300
- Jarvis, A., Reuter, H.I.I., Nelson, A., Guevara, E., 2008. Hole-filled seamless SRTM data V4. *Int. Cent. Trop. Agric.*
- Jeanjean, G., De Metz-noblat, B., 1997. Stabilité dynamique des réseaux électriques industriels. *Cah. Tech.* 185.
- Johnson, R.W., Hering, W.S., Shields, J.E., 1989. Automated visibility & cloud cover measurements with a solid-state imaging system.
- Kaiser, J.W., Peuch, V.-H., Benedetti, A., Boucher, O.J., R., E., Holzer-Popp, T., Morcrette, J.-J., Wooster, M.J., 2012. The pre-operational GMES atmospheric service in MACC-II and its potential usage of SENTINEL-3 observations. *ESA Spec. Publ. SP- 708*, held ESA-ESRIN 708, 4–7.
- Kamath, C., 2010. Understanding wind ramp events through analysis of historical data. 2010 IEEE PES Transm. Distrib. Conf. Expo. Smart Solut. a Chang. World.  
doi:10.1109/TDC.2010.5484508
- Kaur, A., Nonnenmacher, L., Pedro, H.T.C., Coimbra, C.F.M., 2016. Benefits of solar forecasting for energy imbalance markets. *Renew. Energy* 86, 819–830.  
doi:10.1016/j.renene.2015.09.011
- Kazantzidis, A., Tzoumanikas, P., Bais, A.F., Fotopoulos, S., Economou, G., 2012. Cloud detection and classification with the use of whole-sky ground-based images. *Atmos. Res.* 113, 80–88. doi:10.1016/j.atmosres.2012.05.005
- Kazantzidis, A., Tzoumanikas, P., Blanc, P., Massip, P., Wilbert, S., Ramirez-Santigosa, L., 2017. Short-term forecasting based on all-sky cameras, in: *Renewable Energy Forecasting: From Models to Applications*. Woodhead Publishing, pp. 153–178.  
doi:10.1016/B978-0-08-100504-0.00005-6
- Kleissl, J., 2013. Solar energy forecasting and resource assessment.
- Koren, I., Remer, L.A., Kaufman, Y.J., Rudich, Y., Martins, J.V., 2007. On the twilight zone between clouds and aerosols. *Geophys. Res. Lett.* 34.  
doi:10.1029/2007GL029253
- Kostylev, V., Pavlovski, A., 2011. Solar Power Forecasting Performance - Towards Industry Standards, in: *1st International Workshop on the Integration of Solar Power into Power Systems*. Aarhus, Denmark.
- Kudo, S., 2015. Characterising collective ramp events of distributed photovoltaic systems in Brisbane, Australia.
- Kuhn, P., Wilbert, S., Prah, C., Schüler, D., Haase, T., Hirsch, T., Wittmann, M., Ramirez, L., Zarzalejo, L., Meyer, A., Vuilleumier, L., Blanc, P., Pitz-Paal, R., 2017. Shadow camera system for the generation of solar irradiance maps. *Sol. Energy* 157,

- 157–170. doi:10.1016/j.solener.2017.05.074
- Kuhn, P., Wirtz, M., Killius, N., Wilbert, S., Bosch, J.L., Hanrieder, N., Nouri, B., Kleissl, J., Ramirez, L., Schroedter-Homscheidt, M., Heinemann, D., Kazantzidis, A., Blanc, P., Pitz-Paal, R., 2018. Benchmarking three low-cost, low-maintenance cloud height measurement systems and ECMWF cloud heights against a ceilometer. *Sol. Energy* 168, 140–152. doi:10.1016/j.solener.2018.02.050
- Kurtz, B., Mejia, F., Kleissl, J., 2017. A virtual sky imager testbed for solar energy forecasting. *Sol. Energy* 158, 753–759. doi:10.1016/J.SOLENER.2017.10.036
- Lave, M., Kleissl, J., 2013. Cloud speed impact on solar variability scaling - Application to the wavelet variability model. *Sol. Energy* 91, 11–21. doi:10.1016/j.solener.2013.01.023
- Lave, M., Kleissl, J., Stein, J., 2013. A wavelet-based variability model (WVM) for solar PV power plants. *Sustain. Energy, IEEE Trans.*
- Lefevre, M., Albuissou, M., Wald, L., 2002. Description of the software Heliosat-2 for the conversion of images acquired by Meteosat satellites in the visible band into maps of solar radiation available at.
- Lefèvre, M., Oumbe, A., Blanc, P., Espinar, B., Gschwind, B., Qu, Z., Wald, L., Schroedter-Homscheidt, M., Hoyer-Klick, C., Arola, A., Benedetti, A., Kaiser, J.W., Morcrette, J.J., 2013. McClear: A new model estimating downwelling solar radiation at ground level in clear-sky conditions. *Atmos. Meas. Tech.* 6, 2403–2418. doi:10.5194/amt-6-2403-2013
- Lew, D., Bird, L., Milligan, M., Speer, B., Wang, X., Carlini, E.M., Estanqueiro, A., Flynn, D., Gomez-lazaro, E., Menemenlis, N., Orths, A., Pineda, I., Smith, J.C., Söder, L., Sorensen, P., 2013. Wind and Solar Curtailment Preprint. *Int. Work. Large-Scale Integr. Wind Power Into Power Syst.* 1–7.
- Li, Q., Lu, W., Yang, J., 2011. A hybrid thresholding algorithm for cloud detection on ground-based color images. *J. Atmos. Ocean. Technol.* 28, 1286–1296. doi:10.1175/JTECH-D-11-00009.1
- Liu, C., 2009. Beyond pixels: exploring new representations and applications for motion analysis. Massachusetts Institute of Technology.
- Long, C.N., Dutton, E.G., 2010. BSRN Global Network recommended QC tests, V2. *J. Clim.* 25, 8542–8567. doi:10.1175/JCLI-D-11-00618.1
- Long, C.N., Sabburg, J.M., Calbó, J., Pagès, D., 2006. Retrieving cloud characteristics from ground-based daytime color all-sky images. *J. Atmos. Ocean. Technol.* 23, 633–652. doi:10.1175/JTECH1875.1
- Lonij, V.P.A., Brooks, A.E., Cronin, A.D., Leuthold, M., Koch, K., 2013. Intra-hour forecasts of solar power production using measurements from a network of irradiance sensors. *Sol. Energy* 97, 58–66. doi:10.1016/j.solener.2013.08.002
- Lorenz, E., Hammer, A., Heinemann, D., 2004. Short term forecasting of solar radiation based on satellite data, in: *EUROSUN2004, ISES Europe Solar Congress.* pp. 841–848.
- Lorenz, E., Hurka, J., Karampela, G., Heinemann, D., Beyer, H., Schneider, M., 2008. Qualified Forecast of ensemble power production by spatially dispersed grid-connected PV systems. *23rd Eur. Photovolt. Sol. Energy Conf.* 3285–3291.

- doi:10.4229/23rdEUPVSEC2008-5AO.8.6
- Lorenz, E., Remund, J., Müller, S.C., Traunmüller, W., Steinmaurer, G., Pozo, D., Antonio, J., Fanego, V.L., Ramirez, L., Romeo, M.G., Kurz, C., Pomares, L.M., Guerrero, C.G., 2009. Benchmarking of different approaches to forecast solar irradiance, in: 24th European Photovoltaic Solar Energy Conference. Hamburg, Germany, pp. 25–34. doi:10.4229/24thEUPVSEC2009-5BV.2.50
- Lorenz, E., Scheidsteiger, T., Hurka, J., Heinemann, D., Kurz, C., 2011. Regional PV power prediction for improved grid integration. *Prog. Photovoltaics Res. Appl.* 19, 757–771. doi:10.1002/pip.1033
- Lovejoy, S., Mandelbrot, B.B., 1985. Fractal properties of rain, and a fractal model. *Tellus A* 37 A, 209–232. doi:10.1111/j.1600-0870.1985.tb00423.x
- Lucas, B.D.B., Kanade, T., 1981. An iterative image registration technique with an application to stereo vision, in: International Joint Conference on Artificial Intelligence. pp. 674–679.
- Luoma, J., Kleissl, J., Murray, K., 2012. Optimal inverter sizing considering cloud enhancement. *Sol. Energy* 86, 421–429. doi:10.1016/J.SOLENER.2011.10.012
- Lynch, C., O'Mahony, M.J., Guinee, R.A., 2015. A novel 24 kalman filter bank estimator for solar irradiance prediction for PV power generation, in: 2015 IEEE 42nd Photovoltaic Specialist Conference, PVSC 2015. doi:10.1109/PVSC.2015.7356002
- Maafi, A., Harrouni, S., 2003. Preliminary results of the fractal classification of daily solar irradiances. *Sol. Energy* 75, 53–61. doi:10.1016/S0038-092X(03)00192-0
- Macrae, R., Dixon, S., 2010. Accurate Real-time Windowed Time Warping. *ISMIR* 423–428.
- Mahani, S.E., Gao, X., Sorooshian, S., Imam, B., 2000. Estimating cloud top height and spatial displacement from scan-synchronous GOES images using simplified IR-based stereoscopic analysis. *J. Geophys. Res. Atmos.* 105, 15597–15608. doi:10.1029/2000JD900064
- Marquez, R., Coimbra, C.F.M., 2013. Intra-hour DNI forecasting based on cloud tracking image analysis. *Sol. Energy* 91, 327–336. doi:10.1016/j.solener.2012.09.018
- Marquez, R., Coimbra, C.F.M., 2012. Proposed Metric for Evaluation of Solar Forecasting Models. *J. Sol. Energy Eng.* 135, 011016. doi:10.1115/1.4007496
- Marquez, R., Pedro, H.T.C., Coimbra, C.F.M., 2013. Hybrid solar forecasting method uses satellite imaging and ground telemetry as inputs to ANNs. *Sol. Energy* 92, 176–188. doi:10.1016/j.solener.2013.02.023
- Mathiesen, P., Kleissl, J., 2011. Evaluation of numerical weather prediction for intra-day solar forecasting in the continental United States. *Sol. Energy* 85, 967–977. doi:10.1016/j.solener.2011.02.013
- Matthews, B.W., 1975. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *BBA - Protein Struct.* 405, 442–451. doi:10.1016/0005-2795(75)90109-9
- Mayer, B., Kylling, A., 2005. Technical note: The libRadtran software package for radiative transfer calculations - description and examples of use. *Atmos. Chem. Phys.* 5, 1855–1877. doi:10.5194/acp-5-1855-2005
- Mejia, F., 2017. Remote sensing of clouds for solar forecasting applications.

- Mellit, A., Pavan, A.M., 2010. A 24-h forecast of solar irradiance using artificial neural network: Application for performance prediction of a grid-connected PV plant at Trieste, Italy. *Sol. Energy* 84, 807–821. doi:10.1016/j.solener.2010.02.006
- Menzel, W.P., 2001. Cloud Tracking with Satellite Imagery: From the Pioneering Work of Ted Fujita to the Present. *Bull. Am. Meteorol. Soc.* 82, 33–47. doi:10.1175/1520-0477(2001)082<0033:CTWSIF>2.3.CO;2
- Mills, A., 2010. Understanding Variability and Uncertainty of Photovoltaics for Integration with the Electric Power System. *Electr. J.*
- Mills, A., Wiser, R., 2010. Implications of Wide-Area Geographic Diversity for Short-Term Variability of Solar Power. *Energy* 1–10. doi:LBNL-3884E
- Ministère de l'Écologie, 2017. Chiffres clés du climat - France et monde. *Serv. l'observation des Stat.* 52.
- Morcrette, J., Barker, H., Cole, J., Iacono, M., Pincus, R., 2008. Impact of a new radiation package, McRad, in the ECMWF integrated forecasting system. *Mon. Weather Rev.* 136, 4773–4798.
- Müller, J., Fowler, G., Dammann, K., Rogers, C., Buhler, Y., Flewin, J., 2013. MSG Level 1.5 Image Data Format Description, EUM/MSG/ICD/105.
- Oberländer, D., Prah, C., Wilbert, S., Müller, S., Stanicki, B., Hanrieder, N., 2015. Cloud shadow maps from whole sky imagers and voxel carving, in: *International Conference Energy & Meteorology*. Boulder, Colorado (USA).
- Ohmura, A., Dutton, E.G., Forgan, B., Fröhlich, C., Gilgen, H., Hegner, H., Heimo, A., König-Langlo, G., McArthur, B., Müller, G., Philipona, R., Pinker, R., Whitlock, C.H., Dehne, K., Wild, M., 1998. Baseline Surface Radiation Network (BSRN/WCRP): New Precision Radiometry for Climate Research. *Bull. Am. Meteorol. Soc.* 79, 2115–2136. doi:10.2307/26214877
- Otsu, N., 1975. A threshold selection method from gray-level histograms. *Automatica*.
- Oumbe, A., Qu, Z., Blanc, P., Lefèvre, M., Wald, L., Cros, S., 2014. Decoupling the effects of clear atmosphere and clouds to simplify calculations of the broadband solar irradiance at ground level. *Geosci. Model Dev.* 7, 1661–1669. doi:10.5194/gmd-7-1661-2014
- Pelland, S., Remund, J., Kleissl, J., Oozeki, T., Brabandere, K. De, 2013. Photovoltaic and Solar Forecasting: State of the Art. *Int. Energy Agency Photovolt. Power Syst. Program. Rep.* IEA PVPS T14 1–40. doi:978-3-906042-13-8
- Peng, Z., Yu, D., Huang, D., Heiser, J., Yoo, S., Kalb, P., 2015. 3D cloud detection and tracking system for solar forecast using multiple sky imagers. *Sol. Energy* 118, 496–519. doi:10.1016/j.solener.2015.05.037
- Perez, R., David, M., Hoff, T.E., Jamaly, M., Kivalov, S., Kleissl, J., Lauret, P., Perez, M., 2016. Spatial and Temporal Variability of Solar Energy. *Found. Trends® Renew. Energy* 1, 1–44. doi:10.1561/27000000006
- Perez, R., Kivalov, S., Schlemmer, J., Hemker, K., Renné, D., Hoff, T., 2010. Validation of short and medium term operational solar radiation forecasts in the US. *Sol. Energy* 84, 2161–2172.
- Perez, R., Kivalov, S., Schlemmer, J., Hemker, K.J., Zelenka, A., 2010. Improving the performance of satellite-to-irradiance models using the satellite's infrared sensors.

- Proc. Am. Sol. Energy Soc. Annu. Conf.
- Perez, R., Lorenz, E., Pelland, S., Beauharnois, M., Van Knowe, G., Hemker, K., Heinemann, D., Remund, J., Müller, S.C., Traunmüller, W., Steinmayer, G., Pozo, D., Ruiz-Arias, J.A., Lara-Fanego, V., Ramirez-Santigosa, L., Gaston-Romero, M., Pomares, L.M., 2013. Comparison of numerical weather prediction solar irradiance forecasts in the US, Canada and Europe. *Sol. Energy* 94, 305–326.  
doi:10.1016/j.solener.2013.05.005
- Peruchena, C.F., Amores, A.F., 2017. Uncertainty in monthly GHI due to daily data gaps. *Sol. Energy* 157, 827–829. doi:10.1016/j.solener.2017.08.080
- Peuch, V., Rouil, L., Tarrason, L., Elbern, H., 2009. Towards European-scale Air Quality operational services for GMES Atmosphere. 9th EMS Annu. Meet. 9th Eur. Conf. Appl. Meteorol. Abstr. held Sept. 28-Oct. 2, 2009 Toulouse, Fr.
- Pfister, G., McKenzie, R.L., Liley, J.B., Thomas, A., Forgan, B.W., Long, C.N., 2003. Cloud Coverage Based on All-Sky Imaging and Its Impact on Surface Solar Irradiance. *J. Appl. Meteorol.* 42, 1421–1434.  
doi:10.1175/1520-0450(2003)042<1421:CCBOAI>2.0.CO;2
- Platnick, S., King, M.D., Ackerman, S.A., Menzel, W.P., Baum, B.A., Riédi, J.C., Frey, R.A., 2003. The MODIS cloud products: Algorithms and examples from terra. *IEEE Trans. Geosci. Remote Sens.* 41, 459–472. doi:10.1109/TGRS.2002.808301
- Qu, Z., Oumbe, A., Blanc, P., Espinar, B., Gesell, G., Gschwind, B., Klüser, L., Lefèvre, M., Saboret, L., Schroedter-Homscheidt, M., Wald, L., 2016. Fast radiative transfer parameterisation for assessing the surface solar irradiance: The Heliosat-4 method. *Energy Meteorol.* 26, 33–57.
- Quesada-Ruiz, S., Chu, Y., Tovar-Pescador, J., Pedro, H.T.C., Coimbra, C.F.M., 2014. Cloud-tracking methodology for intra-hour DNI forecasting. *Sol. Energy* 102, 267–275. doi:10.1016/j.solener.2014.01.030
- Ransome, S., Funtan, P., 2005. Why hourly averaged measurement data is insufficient to model PV system performance accurately, in: 20th European Photovoltaic Solar Energy Conference, Barcelona.
- Remund, J., Wald, L., Lefèvre, M., Ranchin, T., Page, J., 2003. Worldwide Linke Turbidity Information. *Ises* 2003 41.
- REN21, 2017. Renewables 2017: global status report, Renewable and Sustainable Energy Reviews. doi:10.1016/j.rser.2016.09.082
- Reno, M.J., Hansen, C.W., 2016. Identification of periods of clear sky irradiance in time series of GHI measurements. *Renew. Energy* 90, 520–531.  
doi:10.1016/j.renene.2015.12.031
- Rigollier, C., Lefèvre, M., Wald, L., 2004. The method Heliosat-2 for deriving shortwave solar radiation from satellite images. *Sol. Energy* 77, 159–169.  
doi:10.1016/j.solener.2004.04.017
- Roebeling, R., Feijt, A., Stammes, P., 2006. Cloud property retrievals for climate monitoring: Implications of differences between Spinning Enhanced Visible and Infrared Imager (SEVIRI) on METEOSAT-8 and Advanced Very High Resolution Radiometer (AVHRR) on NOAA-17. *J. Geophys. Res. Atmos.* 111, D20.
- Roesch, A., Wild, M., Ohmura, A., Dutton, E.G., Long, C.N., Zhang, T., 2011.

- Assessment of BSRN radiation records for the computation of monthly means. *Atmos. Meas. Tech.* 4, 339–354. doi:10.5194/amt-4-339-2011
- Roser, M., Geiger, A., 2009. Video-based raindrop detection for improved image registration, in: 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops 2009. IEEE, pp. 570–577. doi:10.1109/ICCVW.2009.5457650
- Roy, G., Hayman, S., Julian, W., 2001. Sky analysis from CCD images: cloud cover. *Light. Res. Technol.* 33, 211–221. doi:10.1177/136578280103300402
- Ruiz-Arias, J.A., Alsamamra, H., Tovar-Pescador, J., Pozo-Vázquez, D., 2010. Proposal of a regressive model for the hourly diffuse solar radiation under all sky conditions. *Energy Convers. Manag.* 51, 881–893. doi:10.1016/j.enconman.2009.11.024
- Saint-Drenan, Y.-M., Wald, L., Ranchin, T., Dubus, L., Troccoli, A., 2018. An approach for the estimation of the aggregated photovoltaic power generated in several European countries from meteorological data. *Adv. Sci. Res* 15, 51–62. doi:10.5194/asr-15-51-2018
- Sakoe, H., Chiba, S., 1978. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoust.* 26, 43–49.
- Sayeef, S., Heslop, S., Cornforth, D., Moore, T., Percy, S., Ward, J.K., Berry, A., Rowe, D., 2012. Solar intermittency: Australia's clean energy challenge 198.
- Scaramuzza, D., Martinelli, A., Siegwart, R., 2006. A toolbox for easily calibrating omnidirectional cameras, in: IEEE International Conference on Intelligent Robots and Systems. IEEE/RSJ, pp. 5695–5701. doi:10.1109/IROS.2006.282372
- Schroedter-Homscheidt, M., 2016. The Copernicus Atmosphere Monitoring Service (CAMS) Radiation Service in a nutshell, 22nd SolarPACES Conference 2016. doi:10.5194/amt-6-2403-2013.
- Seiz, G., 2003. Ground-and satellite-based multi-view photogrammetric determination of 3D cloud geometry.
- Seiz, G., Tjemkes, S., Watts, P., 2007. Multiview cloud-top height and wind retrieval with photogrammetric methods: Application to Meteosat-8 HRV observations. *J. Appl. Meteorol. Climatol.* 46, 1182–1195. doi:10.1175/JAM2532.1
- Sengupta, M., Habte, A., Kurtz, S., Dobos, A., Wilbert, S., Lorenz, E., Stoffel, T., Renné, D., Myers, D., Wilcox, S., Blanc, P., Perez, R., 2015. Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications, Technical report, NREL/TP-5D00-63112.
- Shields, J., Karr, M., Burden, A., Johnson, R., Mikuls, V., Streeter, J., Hodgkiss, W., 2009. Research toward Multi-site Characterization of Sky Obscuration by Clouds.
- Shields, J.E., Karr, M.E., Burden, A.R., Mikuls, V.W., Streeter, J.R., Johnson, R.W., Hodgkiss, W.S., 2010. Whole Sky Imager Characterization of Sky Obscuration by Clouds for the Starfire Optical Range 102.
- Stengel, M.S., Kniffka, A.K., Meirink, J.F.M., Lockhoff, M.L., Tan, J.T., Hollmann, R.H., 2014. CLAAS: The CM SAF cloud property data set using SEVIRI. *Atmos. Chem. Phys.* 14, 4297–4311. doi:10.5194/acp-14-4297-2014
- Stern, N., 2006. Stern Review Report on the Economics of Climate Change.
- Tapakis, R., Charalambides, A.G., 2013. Equipment and methodologies for cloud

- detection and classification: A review. *Sol. Energy* 95, 392–430.  
doi:10.1016/j.solener.2012.11.015
- The Paris Agreement, 2015. , unfccc.int.
- Tina, G.M., De Fiore, S., Ventura, C., 2012. Analysis of forecast errors for irradiance on the horizontal plane. *Energy Convers. Manag.* 64, 533–540.  
doi:10.1016/j.enconman.2012.05.031
- Vallance, L., Charbonnier, B., Paul, N., Dubost, S., Blanc, P., 2017. Towards a standardized procedure to assess solar forecast accuracy: A new ramp and time alignment metric. *Sol. Energy* 150, 408–422. doi:10.1016/j.solener.2017.04.064
- Van der Laan, M., Pollard, K., Bryan, J., 2003. A new partitioning around medoids algorithm. *J. Stat. Comput. Simul.* 73, 575–584. doi:10.1080/0094965031000136012
- Vernay, C., Blanc, P., Pitaval, S., 2013. Characterizing measurements campaigns for an innovative calibration approach of the global horizontal irradiation estimated by HelioClim-3. *Renew. Energy* 57, 339–347. doi:10.1016/j.renene.2013.01.049
- Voyant, C., 2011. Prédiction de séries temporelles de rayonnement solaire global et de production d'énergie photovoltaïque à partir de réseaux de neurones artificiels 6134, 257.
- Vuilleumier, L., Félix, C., Vignola, F., Blanc, P., Badosa, J., Kazantzidis, A., Calpini, B., 2017. Performance evaluation of radiation sensors for the solar energy sector. *Meteorol. Zeitschrift* 26, 485–505. doi:10.1127/metz/2017/0836
- Wald, L., 1999. Some terms of reference in data fusion. *IEEE Trans. Geosci. Remote Sens.* 37, 1190–1193. doi:10.1109/36.763269
- Wang, G., Kurtz, B., Kleissl, J., 2016. Cloud base height from sky imager and cloud speed sensor. *Sol. Energy* 131, 208–221. doi:10.1016/j.solener.2016.02.027
- Wood-Bradley, P., Zapata, J., Pye, J., 2012. Cloud tracking with optical flow for short-term solar forecasting. *50Th Conf. Aust. Sol. Energy Soc.* 2–7.
- Yamashita, M., 2004. Cloud cover estimation using multitemporal hemisphere imageries. *Proc. XXth Congr. Int. Soc. Photogramm. Remote Sens.* 818–821.  
doi:10.1175/1520-0450(1986)025<1280:CCEUBS>2.0.CO;2
- Yang, D., Kleissl, J., Gueymard, C.A., Pedro, H.T.C., Coimbra, C.F.M., 2018. History and trends in solar irradiance and PV power forecasting: A preliminary assessment and review using text mining. *Sol. Energy* 168, 60–101.  
doi:10.1016/j.solener.2017.11.023
- Yang, D., Quan, H., Disfani, V.R., Liu, L., 2017a. Reconciling solar forecasts: Geographical hierarchy. *Sol. Energy* 146, 276–286.  
doi:10.1016/J.SOLENER.2017.02.010
- Yang, D., Quan, H., Disfani, V.R., Rodríguez-Gallegos, C.D., 2017b. Reconciling solar forecasts: Temporal hierarchy. *Sol. Energy* 158, 332–346.  
doi:10.1016/j.solener.2017.09.055
- Yang, D., Walsh, W.M., Zibo, D., Jirutitijaroen, P., Reindl, T.G., 2013. Block matching algorithms: Their applications and limitations in solar irradiance forecasting. *Energy Procedia* 33, 335–342. doi:10.1016/j.egypro.2013.05.074
- Yang, J., Lu, W., Ma, Y., Yao, W., 2012. An automated cirrus cloud detection method for a ground-based cloud image. *J. Atmos. Ocean. Technol.* 29, 527–537.

- doi:10.1175/JTECH-D-11-00002.1
- Yang, J., Min, Q., Lu, W., Ma, Y., Yao, W., Lu, T., Du, J., Liu, G., 2016. A total sky cloud detection method using real clear sky background. *Atmos. Meas. Tech.* 9. doi:10.5194/amt-9-587-2016
- You, S., Tan, R., 2013. Adherent raindrops detection and removal in video, in: *IEEE Conference (Ed.), Computer Vision and Pattern Recognition (CVPR)*. pp. 2–3.
- Younes, S., Claywell, R., Muneer, T., 2005. Quality control of solar radiation data: Present status and proposed new approaches. *Energy* 30, 1533–1549. doi:10.1016/j.energy.2004.04.031
- Zadeh, L.A., 1975. *Fuzzy logic and approximate reasoning*. Springer 30.
- Zhang, J., Florita, A., Hodge, B.-M., Freedman, J., 2014. Ramp forecasting performance from improved short-term wind power forecasting, in: *Proceedings of the ASME Design Engineering Technical Conference*. doi:10.1115/DETC2014-34775
- Zhang, J., Hodge, B.-M., Florita, A., Lu, S., Hamann, H.F., Banunarayanan, V., 2013. Metrics for Evaluating the Accuracy of Solar Power Forecasting. *3rd Int. Work. Integr. Sol. Power into Power Syst.* 17436, 1–10.

## RÉSUMÉ

---

L'exploitation de l'énergie solaire pose des défis liés à la nature variable de la ressource concernée : le rayonnement solaire. Son comportement intermittent est un problème pour la gestion du réseau électrique et des centrales photovoltaïques. L'une des solutions largement envisagées est la prévision de la production photovoltaïque à différents horizons.

L'objectif de cette thèse est d'explorer de nouvelles voies pour améliorer les prévisions existantes du rayonnement solaire, pour des horizons allant de quelques minutes à quelques heures, en exploitant les synergies possibles entre les mesures pyranométriques, les images hémisphériques du ciel prises au sol et les images acquises par les satellites météorologiques géostationnaires. Ces deux types d'images ont des couvertures spatiales, des résolutions spatio-temporelles et des points de vue très différents.

L'approche proposée dans cette thèse exploite cette différence de points de vue afin d'affiner la géolocalisation des nuages en 3D par stéréoscopie, dont l'évolution des ombres peut alors être estimée et prévue. Un simulateur géométrique de la méthode a été développé et permet d'en identifier certains avantages et limitations. La géolocalisation des nuages appliquée à des images réelles a permis d'élaborer des schémas d'estimations et de prévisions performants du rayonnement solaire incident. Enfin, pour compléter l'analyse usuelle de ces performances, deux nouvelles métriques ont été proposées dans l'optique de quantifier deux notions essentielles : le suivi des rampes et l'alignement temporel de la prévision par rapport à la mesure.

## MOTS CLÉS

---

Prévision solaire, photovoltaïque, mesures pyranométriques, images hémisphériques, images satellites

## ABSTRACT

---

The exploitation of solar energy raises challenges related to the variable nature of the resources involved: the incident solar irradiance. Its intermittent behavior is an issue for grid and photovoltaic power plants management. One of the solutions that have been widely considered is the forecast of photovoltaic production at different time horizon.

The aim of this thesis is to explore new ways for improving the existing solar irradiance forecasts, for horizons ranging from the present moment to few hours, by exploiting possible synergies between pyranometric measurements, hemispherical images of the sky taken from the ground and images acquired by geostationary meteorological satellites. These two types of images have completely different spatial coverage, spatio-temporal resolutions and are taken from two different locations.

The proposed approach in this thesis exploits this difference in points of view in order to geolocate the clouds in 3D by stereoscopy, whose shadows dynamics can then be estimated and forecasted. A geometric simulator of the method has been developed and allows identifying certain advantages and limitations. The geolocation of clouds applied to real images made it possible to develop efficient estimates and forecasts of incident solar irradiance. Finally, to complete the usual analysis of these performances, two new metrics have been proposed in order to quantify two essential notions: the monitoring of the ramps and the temporal alignment of the forecast with the measurements.

## KEYWORDS

---

Solar forecasting, photovoltaic, pyranometric measurements, hemispherical images, satellite images