



HAL
open science

Développement d'une méthodologie pour l'évaluation de l'incidence des variabilités interannuelles et de plus long-terme de la ressource solaire sur l'analyse de risque financier d'un projet de centrale solaire photovoltaïque

Mélodie Trolliet

► To cite this version:

Mélodie Trolliet. Développement d'une méthodologie pour l'évaluation de l'incidence des variabilités interannuelles et de plus long-terme de la ressource solaire sur l'analyse de risque financier d'un projet de centrale solaire photovoltaïque. Energie électrique. Université Paris sciences et lettres, 2020. Français. NNT : 2020UPSLM040 . tel-03123362

HAL Id: tel-03123362

<https://pastel.hal.science/tel-03123362v1>

Submitted on 27 Jan 2021

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

**Développement d'une méthodologie pour l'évaluation de
l'incidence des variabilités interannuelles et de plus
long-terme de la ressource solaire sur l'analyse de
risque financier d'un projet de centrale solaire
photovoltaïque**

Soutenue par

Mélodie TROLLIET

Le 11 mai 2020

Ecole doctorale n° 621

**Ingénierie des Systèmes,
Matériaux, Mécanique,
Energétique**

Spécialité

**Energétique et Génie des
Procédés**

Composition du jury :

Laurent LINGUET Professeur des universités, IRD, Guyane	<i>Président</i>
Jean-Louis DUFRESNE Directeur de recherche, LMD, Paris	<i>Rapporteur</i>
Nathalie PHILIPPON Chargée de recherche, IGE, Grenoble	<i>Rapporteur</i>
Sylvie PAREY Chercheur senior, EDF R&D	<i>Examinatrice</i>
Yves-Marie SAINT-DRENAN Ingénieur de recherche, Mines ParisTech	<i>Examineur</i>
Philippe BLANC Directeur de recherche, Mines ParisTech	<i>Directeur de thèse</i>

Résumé en français

L'installation de grandes centrales solaires photovoltaïques est stratégique dans un contexte de transition énergétique vers des systèmes de production d'énergie décarbonés et renouvelables. L'important investissement financier que représente une telle installation étant concentré lors de la mise en place de la centrale, une analyse de risques financiers est réalisée. Il a été identifié, parmi toutes les sources d'incertitudes propres à cette analyse, un besoin de prise en compte des variations de long terme de l'éclairement. Cette thèse se propose de contribuer à la caractérisation de ces variations de long terme.

Une étude statistique et bibliographique a permis de montrer que les méthodes de caractérisation de l'éclairement sur la durée de la vie de la centrale sont limitées. En particulier, l'utilisation d'indicateurs statistiques (*e.g.* P90) a été discutée. L'hypothèse de stationnarité temporelle de l'éclairement, admise de manière implicite dans la quasi-totalité des études recensées, a été questionnée. Cela nous a menés à une étude approfondie des variations de long terme de l'éclairement. Celles-ci ont été caractérisées grâce à un outil de décomposition temps-fréquence développé au cours de cette thèse. Nous avons distingué trois classes de variabilité : la variabilité intra-annuelle, la variabilité annuelle à décennale, et la variabilité multi-décennale. Ces trois classes ont été analysées pour des bases de données de natures diverses : mesures de long terme du réseau GEBA, données satellitaires CLARA-A2, données de ré-analyse MERRA-2, et données issues du modèle climatique IPSL-CM6A-LR.

Pour l'ensemble des bases de données, une répartition latitudinale de la variabilité a pu être constatée, avec une faible influence de la variabilité inter-annuelle dans la zone extratropicale, et forte proche de l'équateur. Pour la première zone, influencée majoritairement par la variabilité intra-annuelle, l'utilisation de quatre ans de données historiques est suffisante pour prendre en compte de manière correcte l'ensemble des variations de l'éclairement. En revanche, pour les régions de la zone tropicale influencées par la variabilité annuelle à décennale, l'utilisation de 30 années de données historiques est recommandée. Une alternative à l'utilisation de ce large échantillon de données a été abordée, avec une possible modélisation des oscillations caractéristiques de ces échelles. Pour la variabilité multi-décennale, l'utilisation de plus de 30 années de données est préconisée. Une grande diversité des structures de variabilité en fonction de la base de données considérée a été observée. En particulier, la base de données GEBA détecte une plus forte influence de la variabilité inter-annuelle que les autres bases de données. Cela questionne sur la représentation des variations inter-annuelle dans les bases de données grillées. À ce stade des recherches, aucune base de données grillées ne s'est dégagée. Pour un site donné, il est ainsi recommandé de toutes les considérer, et d'en étudier l'ensemble les échelles de variations.

Résumé en anglais

In a context of energy transition from carbon energy to carbon-free energy, the installation of solar power plants is strategic. This kind of installation represents a great investment which is mostly made at the start of the installation. Therefore, a financial risk assessment is realised. We have identified, among all sources of uncertainty related to this analysis, a necessity to take the long term variations of the irradiance into account.

A statistical analysis and a literature review have shown that current methods used to characterize the irradiance on the lifetime of the installation are limited. In particular, the use of statistical indicators (e.g. P90) have been discussed. The hypothesis of temporal stationarity of the irradiance, which is assumed in most of the study, has been questioned. This led us to deepen the analysis of the long term variations of the irradiance. These variations have been characterized with the use of a time-frequency decomposition tool, that was developed during this PhD. This brought us to distinguish three classes of variability : the intra-annual variability, the annual to decadal variability, and the multi decadal variability. These three classes have been analysed for long term databases. The databases are of various kinds: long term measurements of the GEBA network, CLARA-A2 satellite data, MERRA-2 reanalysis data, and data from climate models IPSL-CM6A-LR.

For this set of databases, variability has shown a latitudinal repartition: the inter-annual variability has a weak influence on the extra tropical zone, and a strong influence close to the equator. For the extra tropical zone, which is mostly influenced by the intra-annual variability, the use of four years historical data will be sufficient in order to take into account all the variation of the irradiance in a correct way. On the other hand, for the tropical zone influenced by the annual to decadal variability, the use of 30 years historical data is recommended. An alternative to the use of this great amount of data has been addressed, through a possible modelisation of the characteristic oscillations of each scales. Regarding the multi decadal variability, the use of more than 30 years of data is encouraged. A large diversity of the variability structures has been observed, depending on the databases. In particular, the GEBA database shows a greater influence of the inter annual variability than the other databases. This result questions the representativity of the inter annual variations in the gridded databases. At this point of the research work, none of the gridded databases has stood out. For one location of interest, it is therefore recommended to consider them all, and to study the full range of variation scales.

Table des matières

Chapitre 1

Introduction	11
1 De la transition énergétique à la transition électrique	13
2 Le rôle des énergies renouvelables dans la transition électrique : le cas du solaire photovoltaïque	15
3 Introduction à l'éclairement solaire global	19
4 Méthodes de caractérisation du potentiel solaire	23
5 Objectifs de la thèse	25

Chapitre 2

Description des données	27
1 Introduction	29
2 Variables d'intérêts	29
3 Données de mesure <i>in-situ</i> La base de données Global Energy Balance Archive - GEBA	30
4 Données satellitales La base de données CLARA A2	40
5 Données de ré-analyses La base de données MERRA-2	45
6 Données de modèles climatiques La base de données IPSL-CM6A-LR	46
7 Exemple de séries temporelles d'éclairement pour l'ensemble des bases de données	49
8 Conclusion	51

Chapitre 3

Limitations des méthodes d'estimation du potentiel solaire	53
1 Introduction	55
2 Analyse d'un exemple d'indicateur : le P90	55
3 Génération de séries temporelles d'éclairement : analyse bibliographique	67
4 Conclusion	78

Chapitre 4

Etude de la stationnarité temporelle de l'éclairement	79
1 Introduction	81
2 Analyse bibliographique relative aux tendances de l'éclairement	81

3	Comparaison des résultats issus de la bibliographie à l'analyse de mesures locales en Chine	88
4	Conclusion	94
Chapitre 5		
	Description de l'outil de décomposition temps-fréquence	95
1	Introduction	97
2	Description de la méthode de décomposition temps-fréquence	97
3	Validation de la méthode de décomposition temps-fréquence	105
4	Conclusion	113
Chapitre 6		
	Décomposition temps-fréquence des indices de clartés mensuels de la sélection de stations GEBA.....	115
5	115
1	Introduction	117
2	Etude de la stationnarité temporelle des stations GEBA.....	117
3	Quels sont les modes de variabilité inter-annuels permettant d'expliquer le signal d'indice de clarté ?.....	121
4	L'influence de chaque élément sur le signal original est-elle stationnaire ?.....	123
5	Conclusion	126
Chapitre 7		
	Décomposition temps-fréquence des indices de clartés mensuels de différentes bases de données	127
1	Introduction	129
2	Rappel de la méthode d'analyse	129
3	Variabilité des indices de clarté issues de CLARA-A2	130
4	Variabilité des indices de clarté issues de MERRA-2	133
5	Variabilité des indices de clarté issues de IPSL-CM6A-LR	135
6	Discussion.....	142
7	Variabilité inter-annuelle de l'éclairement.....	152
8	Conclusion.....	155
Chapitre 8		
	Conclusions et perspectives	157

Chapitre 1

Introduction

Table des matières

Introduction	11
1 De la transition énergétique à la transition électrique	13
2 Le rôle des énergies renouvelables dans la transition électrique : le cas du solaire photovoltaïque	15
3 Introduction à l'éclairement solaire global	19
4 Méthodes de caractérisation du potentiel solaire	23
5 Objectifs de la thèse	25

La transition énergétique consiste à mettre progressivement en place un nouveau système énergétique mondial. Ce dernier se veut pérenne, respectueux de l’environnement, répondant aux impératifs climatiques et d’évolution démographique, et utilisant des énergies renouvelables. Compte tenu de la dépendance météorologique de la plupart de ces énergies renouvelables, la prise en compte de la variabilité climatique dans les enjeux énergétiques est primordiale. Ainsi, climat et énergie sont deux disciplines complémentaires pour penser ce sujet d’actualité qu’est la transition énergétique. Passant de transition énergétique à transition électrique (section 1), des énergies renouvelables à l’énergie solaire photovoltaïque (section 2), puis du potentiel solaire à l’éclairage solaire (sections 3 et 4), l’objectif de cette introduction est de balayer de manière synthétique ces différents thèmes, afin d’appréhender de façon précise les tenants et les aboutissants propres au travail de thèse effectué. Le point focal de ce dernier se trouve dans l’estimation et la caractérisation des variations inter-annuelles de l’éclairage, et sera explicité conjointement aux objectifs de ce travail de thèse en section 5.

1 De la transition énergétique à la transition électrique

Le concept de transition énergétique diffère fortement selon les pays et les acteurs impliqués. Cependant, un point commun à l’ensemble des concepts répertoriés est l’évolution de la consommation énergétique primaire des dernières décennies - présentée à titre illustratif au niveau mondial en figure 1.

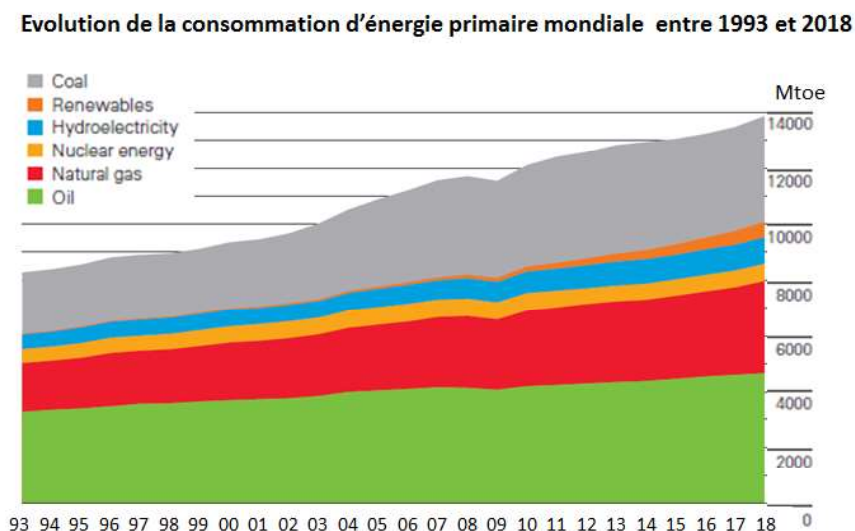


Figure 1 : Evolution de la consommation d’énergie primaire mondiale entre 1993 et 2018, exprimée en millions de tonnes équivalent pétrole (Mtoe, Million Tonnes of Oil Equivalent).

Source : (BP, 2019)

Il est clair, au vu de la figure 1, que la consommation d’énergie mondiale n’a cessé d’augmenter depuis les années 1990 : elle passe de 8 000 millions de tonnes équivalent pétrole (Million Tonnes of Oil Equivalent, Mtoe) en 1993 à près de 14 000 Mtoe pour l’année 2018. Cela correspond à une

1. De la transition énergétique à la transition électrique

augmentation de 6 000 Mtoe en 25 ans. Pour la dernière année 2018, une augmentation de la consommation d'énergie primaire mondiale de 2.9 % par rapport à l'année précédente est visible (BP, 2019). On peut noter sur la figure 1 que la quasi-totalité de la consommation d'énergie primaire est d'origine carbonée : la majorité de l'énergie consommée est d'origine pétrolière (4 800 Mtoe en 2018), de charbon (4 000 Mtoe en 2018) et de gaz naturel (3 100 Mtoe en 2018). Le reste de l'énergie primaire mondiale a été fournie en 2018 par de l'hydro électricité (1 000 Mtoe), de l'énergie nucléaire (50 Mtoe), et des énergies renouvelables (40 Mtoe).

La consommation importante d'énergies d'origine fossile depuis plusieurs dizaines d'années mène à deux conséquences majeures : la raréfaction des ressources, et la contribution à des changements environnementaux historiques. En effet, le rôle de la consommation d'énergie d'origine fossile dans le changement climatique - par leurs émissions de gaz à effet de serre - a été démontré de manière irréfutable (IPCC, 2013), avec pour conséquence des impacts de plus en plus documentés (*e.g.* BERRY *et al.*, 2010; CURTIS *et al.*, 2006; J. HANSEN *et al.*, 2016; HIRABAYASHI *et al.*, 2013; PARK *et al.*, 2018; SCHMIDT *et al.*, 2014).

La production d'électricité, un des pôles majeurs de la consommation d'énergie primaire mondiale, est à elle seule responsable de 40 % des émissions de gaz à effet de serre mondiales (IPCC, 2013). Cela s'explique aisément par l'origine carbonée de cette électricité : en 2017, l'électricité fut produite à près de 65 % par des énergies fossiles (REN21, 2019, p. 21). Sur ces 65 %, 40 % furent produites par des centrales à charbon, 20 % par des centrales à gaz, et 5 % par des centrales à fuel. Les 35 % de la production d'électricité restante se répartissent entre l'hydroélectricité (15 %), le nucléaire (10 %), et les énergies renouvelables (10 %). Avec l'émergence de l'économie numérique et la part croissante des véhicules électriques dans les moyens de transports, on peut s'attendre à ce que la part de la production électrique dans la consommation énergétique mondiale augmente dans le futur. **Avec l'omniprésence des énergies fossiles dans la production d'électricité, la forte émission de gaz à effet de serre associée à celle-ci, ainsi que la croissance extrêmement rapide de la consommation électrique mondiale (*e.g.* +3.7 % en 2018, (IEA, 2019)), il est clair que la transition électrique joue un rôle central dans la transition énergétique.**

Les stratégies de transition électrique varient en fonction des contraintes locales et des politiques nationales considérées. Il peut par exemple s'agir de décarboner l'électricité, de réduire les prélèvements de ressources naturelles, de décentraliser la politique énergétique propre à l'électricité, de définir de nouveaux modèles de production d'électricité pour pouvoir produire dans des zones reculées, assurer l'autonomie énergétique, etc. Quelle que soit l'ambition de chaque pays, le chemin de transition choisi est structuré par des cadres politiques et législatifs nationaux définissant des objectifs chiffrés.

Pour atteindre ces objectifs, un rôle croissant des énergies renouvelables dans le bouquet énergétique est nécessaire. Par exemple, l'Europe s'est engagée à augmenter jusque 32 % la part de la consommation d'énergie d'origine renouvelable dans la consommation d'énergie finale (*cf.* Décret n° 2016-1442 du 27 octobre 2016 relatif à la programmation pluriannuelle de l'énergie, 2016).

2 Le rôle des énergies renouvelables dans la transition électrique : le cas du solaire photovoltaïque

L'agence internationale pour les énergies renouvelables (*International Renewable Energy Agency*, IRENA) définit les énergies renouvelables comme « toute forme d'énergie produite à partir de sources renouvelables et de manière durable ». Afin d'affiner les termes de « renouvelables » et « durable », souvent sujets à des interprétations diverses, nous pouvons compléter cette définition par deux définitions : celle de Hansen et Percebois (J.-P. HANSEN & PERCEBOIS, 2017), qui définissent les énergies renouvelables comme des énergies « prélevées sur des flux naturels et non sur des stocks qui ne se reconstituent pas » ; et celle de l'Observatoire des Energies Renouvelables (Observ'ER), qui soulignent que **les énergies renouvelables sont des énergies qui « n'engendrent peu ou pas d'émissions polluantes, participent à la lutte contre l'effet de serre et les rejets de CO₂ dans l'atmosphère, facilitent la gestion raisonnée des ressources locales, génèrent des emplois »** (Observ'ER, 2019a). Cette dernière définition oblige à adopter un point de vue plus large que celui de la gestion de la ressource, prenant en compte de l'ensemble des impacts des énergies renouvelables, autant sur l'environnement que sur la société, tout au long de leur durée de vie.

L'IRENA liste les énergies renouvelables en six catégories : bioénergies, énergies géothermiques, hydroélectricité, énergies marines, énergies solaires, et énergies éoliennes (IRENA, 2018). L'énergie solaire photovoltaïque et l'énergie éolienne représentent la majorité de la puissance installée en énergie renouvelable dans le monde : en 2018, sur les 1246 GW_c de puissance installée en énergie renouvelable, 88 % étaient du solaire photovoltaïque (près de 500 GW_c) et de l'éolien (près de 600 GW_c) (REN21, 2019). Le secteur de l'énergie solaire photovoltaïque est particulièrement dynamique. D'après l'IEA (2019), l'énergie solaire sera la source principale d'électricité dès 2040. Actuellement, les trois pays les plus dynamiques, possédant la plus grande puissance installée, sont la Chine, les Etats-Unis et le Japon, avec respectivement 170 GW_c, 60 GW_c et 55 GW_c de puissance installée cumulée. Même pour des pays moins dynamiques que ces trois derniers, la puissance photovoltaïque installée raccordée est en croissance rapide, comme l'illustre la figure 2 pour le cas de la France. Cette dernière possédait en 2018 une puissance installée cumulée de 8.5 GW_c. **Ces raisons (omniprésence de l'énergie solaire photovoltaïque dans la puissance installée en énergie renouvelable dans le monde et fort dynamisme du secteur) illustrent la pertinence de cette thèse, qui s'intéresse uniquement à l'énergie solaire photovoltaïque.**

2. Le rôle des énergies renouvelables dans la transition électrique : le cas du solaire photovoltaïque

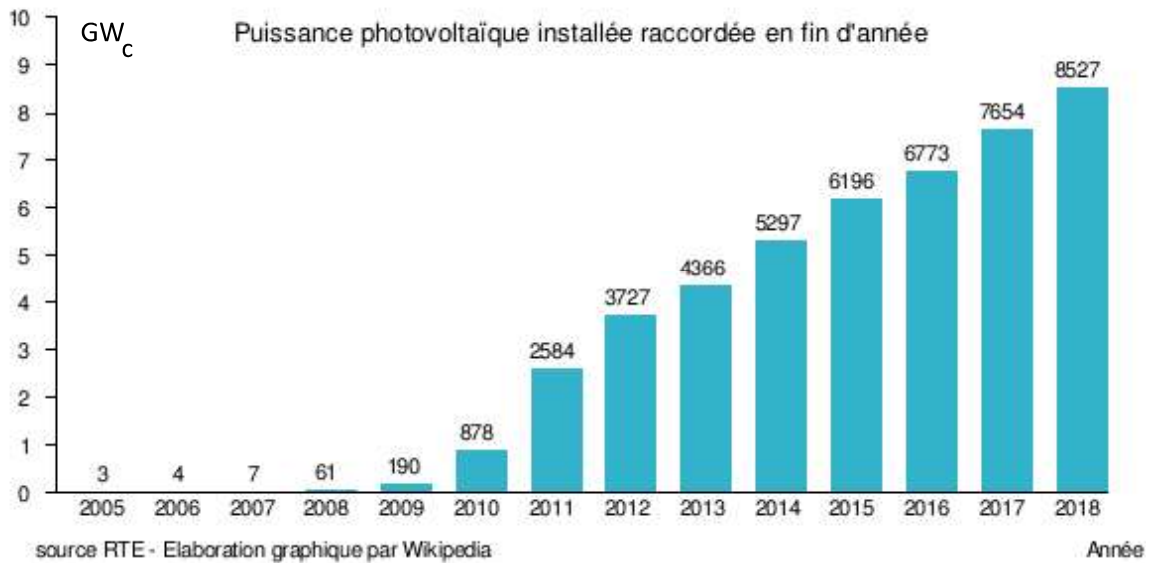


Figure 2: Evolution de la puissance photovoltaïque installée raccordée en France
Source : RTE – Wikipédia

On peut constater sur la figure 2 que la puissance du parc photovoltaïque français a été multipliée par 140 en 10 ans (fin 2008 - fin 2018), pour arriver à une capacité installée de plus de 8.5 GW fin 2018. Cela représente une progression de 12.7 % par rapport à l'année 2017. A titre indicatif, cela représente 2.1 % de la production électrique nationale pour l'année 2018 (RTE), soit 11.6 % de plus que pour l'année 2017.

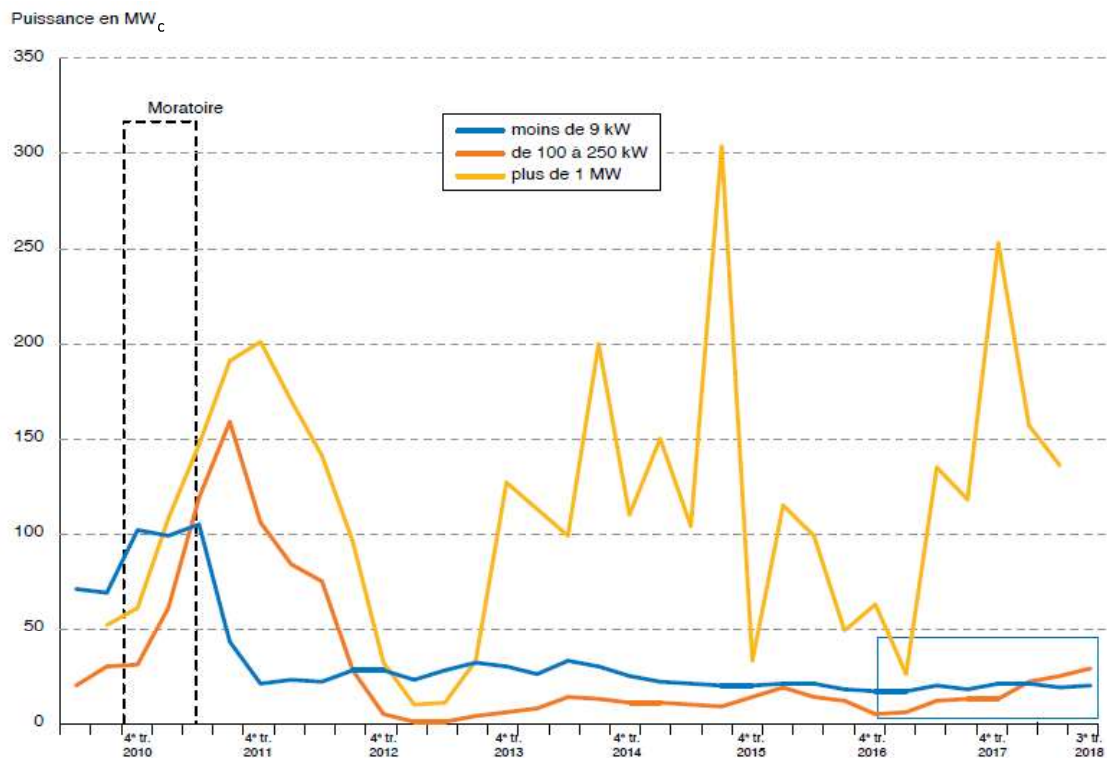


Figure 3 : Evolution des puissances raccordées par trimestre sur trois segments du marché.
Source : (Observ'ER, 2019b), d'après les chiffres du SDES et de l'observatoire de l'énergie solaire photovoltaïque.

On constate grâce à la figure 3, pour la France, que la majorité de la puissance installée depuis 2010 l'est sous forme de centrales solaires de plus de 1 MW_c (courbe jaune). On peut par exemple retrouver la mise en service du parc de Constantin, le plus grand parc photovoltaïque d'Europe, responsable d'un quart de l'accroissement de la puissance installée française en 2015. La quantité de centrales solaires installées sur le territoire français devrait continuer à augmenter dans le futur : entre autres, EDF a annoncé fin 2017 un programme d'investissement de 25 milliards d'euros pour la construction de 30 GW_c de centrales solaires photovoltaïques entre 2020 et 2035 (EurObserv'ER, 2019), quand ENGIE s'est engagée à tripler, voire quadrupler, la capacité photovoltaïque installée, passant ainsi de 550 MW_c en 2015 à près de 2,2 GW_c en 2021. Par ailleurs, la dynamique du secteur solaire photovoltaïque s'illustre à l'échelle internationale, avec par exemple la création de 3 605 000 emplois directs et indirects dans ce domaine pour l'année 2017-2018, dont 60 % sont situés en Chine, contre 3 % dans l'Union Européenne (REN21, 2019). Comme annoncé, la Chine connaît la dynamique du secteur photovoltaïque la plus forte à l'échelle mondiale. Elle produit près de 30 % de la production solaire photovoltaïque mondiale, avec une progression de +89 % en 2014, +53 % en 2015, +68 % en 2016, +74 % en 2017 (IEA, 2019). **L'essor du secteur photovoltaïque, en France comme à l'international, se fait donc principalement par l'installation de centrales photovoltaïques de grande taille.**

Cependant, malgré cette dynamique, différentes instances – comme en France le SER (Syndicat des Energies Renouvelables) - soulignent que le rythme actuel de développement des énergies renouvelables est insuffisant pour satisfaire les objectifs de puissance installée définis par les différentes lois de transition énergétique, tout comme ceux définis par l'Accord de Paris pour limiter le réchauffement climatique (*e.g.* Govillot *et al.*, 2019; Villavicencio & Finon, 2018). Pour atteindre ces objectifs, toute contribution visant à favoriser l'installation de puissance installée d'énergie renouvelable dans le monde est d'intérêt.

Différents freins à l'installation de puissance photovoltaïque ont pu être répertoriés dans le cadre de cette thèse. Ces freins ont été répertoriés grâce à la mise en place d'entretiens individuels d'une vingtaine d'acteurs du domaine de l'énergie solaire et à un travail de collaboration présenté en annexe I. On peut notamment citer parmi ces freins la question de la disponibilité des surfaces nécessaires à de telles installations, la question de l'anthropisation des sols ou encore le maintien de la biodiversité. Un autre frein important a été identifié : les réticences de certaines banques à financer l'installation de nouvelles centrales photovoltaïques dans un contexte très concurrentiel. Or, un des paramètres essentiels à l'installation de centrales solaires se trouve évidemment dans le financement propre à ces projets de grande envergure. Au regard du bilan financier réalisé sur l'ensemble des coûts d'une installation photovoltaïque (coût de l'installation, frais de fonctionnement fixes, coûts du combustible, coûts de maintenance propres à la production, etc.), il apparaît que le plus gros pôle de finance réside dans le coût du capital installé (DAMBRINE, 2019). **L'essentiel des dépenses est donc concentré au début du projet d'installation solaire. Ce fait impose d'avoir recours dans la majorité des cas à un acteur financier externe (*e.g.* banque, fonds d'investissement) afin d'assurer les fonds indispensables au lancement du projet.** Pour ordre de grandeur, le coût total moyen par kW_c d'une installation solaire en 2016 est aux alentours de 1 100 € (IRENA, 2018). Cela correspond à un investissement de 330 M€ pour la plus grande centrale française (300 MW_c) construite en 2015 à

2. Le rôle des énergies renouvelables dans la transition électrique : le cas du solaire photovoltaïque

Cestas (Gironde), et à un investissement de 1 650 M€ pour la plus grande centrale au monde (1500 MW_c), construite en Chine la même année.

Dans un grand nombre de cas, la construction d'une centrale photovoltaïque de grande taille se fait par le biais d'appels d'offres : des développeurs de projets sont mis en concurrence pour un site d'intérêt. La proposition conduisant au coût actualisé de l'électricité le moins cher, entre autres critères de sélection, remporte l'appel d'offres. Ce contexte concurrentiel pousse les développeurs de projet à une prise de risque maximale. **Ainsi, le recours à des acteurs financiers externes dès la première phase du projet, comme le contexte concurrentiel d'appels d'offres, imposent une estimation extrêmement précise et rigoureuse de la rentabilité future de la centrale afin de garantir le retour sur investissement.** Le projet solaire va être évalué sur sa capacité à générer suffisamment de revenus sur une durée typique de 20 à 30 ans, pour pouvoir rembourser progressivement l'emprunt bancaire, suivant des échéances annuelles ou mensuelles prédéfinies. Cette capacité de remboursement dépend de celle de la centrale à produire de l'énergie sur l'ensemble de sa durée de vie, mais aussi année par année, voire mois par mois, pour éviter des problèmes de trésorerie. La production d'énergie dépend elle-même très fortement, mais pas uniquement, de l'éclairement solaire disponible sur le site d'intérêt (PFEIFROTH *et al.*, 2018; STOFFEL, 2013). **Pour un site d'intérêt, l'estimation du potentiel solaire passera par une caractérisation de la quantité d'éclairement solaire incident au sol sur une durée de 20 à 30 ans, sur des échelles mensuelles à inter-annuelles, afin d'assurer une estimation correcte des revenus attendus** (MÜLLER *et al.*, 2009; SCHNITZER *et al.*, 2012; MÜLLER *et al.*, 2014; FERNÁNDEZ-PERUCHENA *et al.*, 2016; BONJEAN STANTON *et al.*, 2016; HULD, 2017; WILD *et al.*, 2015; ROMERA *et al.*, 2017; McMAHAN *et al.*, 2013a).

Il a été identifié au cours des entretiens individuels une incertitude sur la caractérisation de la ressource solaire disponible pour un site donné. Celle-ci concerne tout autant la quantité d'éclairement solaire disponible sur site et ses variations, que les incertitudes associées aux méthodes d'estimation de la ressource solaire. Ces lacunes peuvent introduire des biais dans l'analyse de risques réalisée en amont de l'installation de centrales solaires photovoltaïques, expliquant notamment une partie des réticences actuelles de certains banquiers à l'investissement. **Notre travail vise à contribuer à l'acquisition des connaissances nécessaires à l'augmentation du nombre de centrales solaires photovoltaïques, dans le but d'atteindre les objectifs climatiques et de transition énergétique fixés.**

Avant d'entrer plus en détail dans les méthodes de caractérisation du potentiel solaire, une introduction aux propriétés de l'éclairement solaire utiles dans le cadre de projets photovoltaïques nous a semblé nécessaire.

3 Introduction à l'éclairement solaire global

Pour rappel, l'éclairement solaire énergétique, abrégé en éclairement, est le flux énergétique solaire reçu par unité de surface (Wald, 2007). Il s'exprime en W/m^2 . On peut définir le rayonnement au sommet de l'atmosphère (*Top of Atmosphere*, TOA) comme l'éclairement annuel moyen arrivant à incidence normale sur la couche externe de l'atmosphère. Pour une distance de 1 unité astronomique ($1,5 \cdot 10^8$ km), on parle alors de la constante solaire, renommé depuis peu éclairement solaire total (MEFTAH, 2017). La valeur recommandée de l'éclairement solaire total est actuellement de $1361 W/m^2$ (International Astronomical Union General Assembly, 2015; KOPP & LEAN, 2011). En considérant la rotation de la Terre et l'angle avec lequel les rayons atteignent l'atmosphère selon l'endroit et la saison considérée, l'éclairement TOA sur un plan horizontal est estimé en moyenne sur 24h (intégrant la nuit) à un quart de l'éclairement solaire total, soit $340 W/m^2$. Celui-ci présente une distribution latitudinale, avec un maximum proche de l'équateur, diminuant progressivement à l'approche des pôles. Ces variations sont totalement prédictibles car elles dépendent principalement de paramètres astronomiques bien connus. Il existe en effet de nombreuses références dans la littérature détaillant avec précision le calcul de la position du soleil et les algorithmes utilisés couramment pour réaliser ce calcul (e.g. BLANC & WALD, 2012; BLANCO-MURIEL *et al.*, 2001; BRETAGNON & FRANCOU, 1988; GRENA, 2012; MICHALSKY, 1988; REDA & ANDREAS, 2004; RIGOLLIER *et al.*, 2000). À l'inverse, estimer la quantité d'éclairement reçue à la surface de la Terre à partir de l'éclairement au sommet de l'atmosphère peut s'avérer complexe. En effet, au cours de son trajet, la lumière va entrer en interaction avec les différents composants de l'atmosphère (e.g. nuages, aérosols, gaz), et être partiellement réfléchiée, absorbée ou transmise. Il en résulte trois composantes fondamentales de l'éclairement solaire couramment utilisées dans le domaine de l'énergie solaire, représentées schématiquement sur la figure 4.

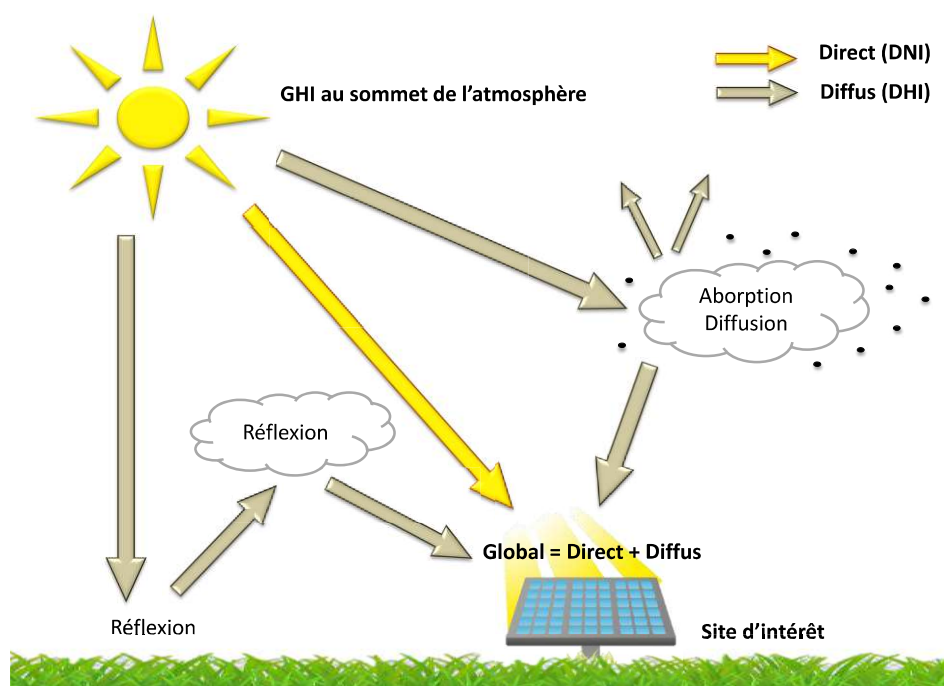


Figure 4 : Représentation schématique des trois composantes fondamentales de l'éclairement. Inspiré de Dambreville (2014).

Ces trois composantes sont :

- L'éclairement direct normal (Direct Normal Irradiance - DNI), correspondant à la part du rayonnement provenant d'un petit angle solide centrée sur la position du soleil, et reçue suivant un plan perpendiculaire à la direction de la position du soleil.
- L'éclairement diffus horizontal (Diffuse Horizontal Irradiance - DHI), correspondant à la partie de l'éclairement arrivant sur un plan horizontal de toutes les directions – ou angle solide – autres que celles considérées pour le rayonnement direct. Il est principalement composé de l'éclairement réfléchi par le sol et les composants atmosphériques, et de l'éclairement diffusé par les nuages, les aérosols, les gaz, etc.
- L'éclairement global horizontal (Global Horizontal Irradiance - GHI), correspondant à l'ensemble de l'éclairement reçu au sol. Ce dernier est donc équivalent à la somme de l'éclairement diffus et direct.

Dans le cadre de projets solaires, la ressource solaire correspond à la quantité de GHI disponible sur site. Ainsi, nous considérerons tout au long de cette thèse uniquement le GHI.

La quantité de ressource solaire est par essence variable dans le temps et dans l'espace. Il est évident pour tout un chacun que l'éclairement varie entre le jour et la nuit, entre l'été et l'hiver pour les moyennes latitudes, et que le Sud de la France est plus éclairé en moyenne que le Nord de l'Ecosse. Cette variabilité explique le rôle central de l'estimation de la ressource solaire dans l'estimation de la production énergétique, et ainsi dans la performance financière de tout projet solaire (McMAHAN *et al.*, 2013b). Autrement dit, à l'opposé des énergies fossiles dont la puissance peut être contrôlée, la puissance maximale fournie est directement tributaire de la quantité de ressource solaire disponible. A titre illustratif, la figure 5 présente les variations mensuelles de l'éclairement mesurée à Beijing, en Chine. Cette série temporelle appartient à la base de données GEBA (GILGEN & OHMURA, 1999) qui sera détaillée en section 3 du chapitre 2.

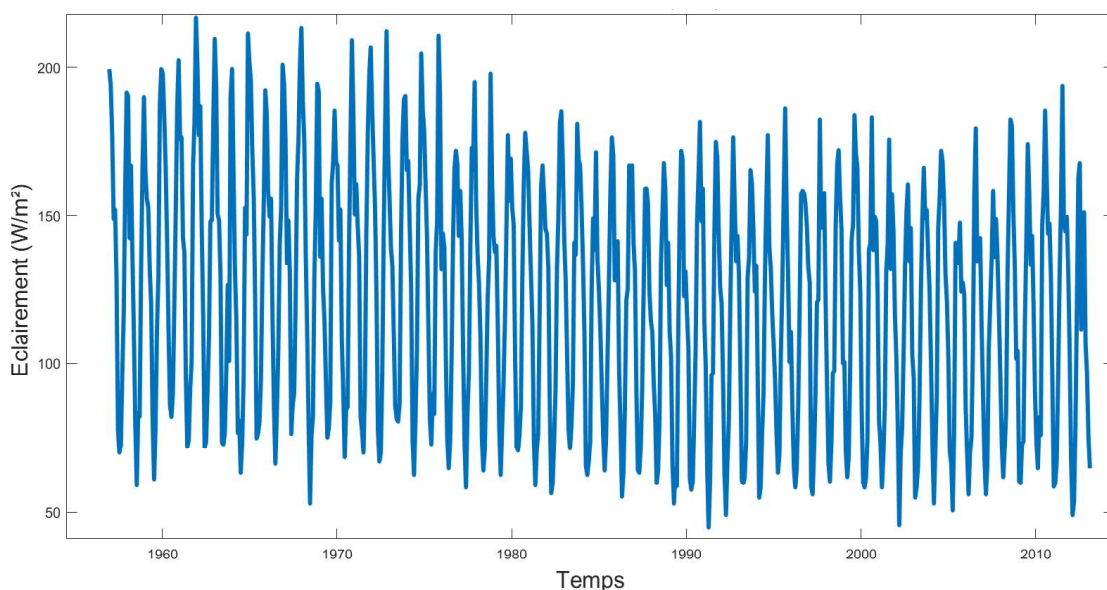


Figure 5 : Séries temporelles des moyennes mensuelles d'éclairement global sur plan horizontal (GHI) mesurées par la station Beijing (GEBA) – Chine.

La figure 5 illustre la variabilité temporelle de l'éclairement : variabilité suivant des échelles mensuelles à multi-annuelles. L'éclairement moyen de cette station est de 122 W/m^2 . On peut tout d'abord distinguer les cycles saisonniers, faisant varier l'éclairement entre 50 W/m^2 et 220 W/m^2 avec une période de 12 mois. L'éclairement peut donc plus que quadrupler en fonction du mois considéré. Le signal observé n'est pas parfaitement sinusoïdal : on peut par exemple constater pour les mois d'été des variations de l'intensité de l'éclairement différentes d'une année à l'autre. Enfin, des variations supérieures à l'année sont également perceptibles. Par exemple, il semblerait que l'éclairement de cette station ait en moyenne diminué entre les années 1960 et 1990 de plusieurs dizaines de W/m^2 . On peut également distinguer à partir des années 1980 des variations décennales quasi périodiques de la moyenne de l'éclairement, avec des minimums relatifs autour de 1990 et 2000, et des maximums relatifs autour des années 1985 et 1995. Cette courte analyse visuelle nous laisse deviner différentes échelles de variations temporelles de l'éclairement : intra-annuelle et inter-annuelle.

Les variations intra-annuelles comprennent notamment les cycles saisonniers et diurnes. Ces cycles sont notamment déterminés par la géométrie solaire. Cette dernière est extrêmement bien caractérisée, ce qui la rend facilement prévisible. On peut également compter parmi les variations intra-annuelles les variations rapides de la ressource solaire, dues par exemple aux conditions météorologiques. Celles-ci peuvent réduire de plus de moitié la production d'énergie d'un jour à l'autre. Ces variations, bien que complexes, sont observées, modélisées et prédites avec une précision toujours croissante.

Les variations inter-annuelles sont composées des cycles inter-annuels et des tendances de long terme de l'éclairement. Soulignons que les variations identifiées correspondent à des tendances lorsque les variations sont d'une échelle supérieure à la fenêtre temporelle d'analyse. Ainsi, il est impossible de dire si ce que l'on qualifie de tendance correspond à une réelle tendance de long terme, ou à une partie d'un cycle de plus long terme, qui apparaîtrait comme une tendance de par la fenêtre d'étude restreinte temporellement. L'ensemble de ces variations prévisibles sont connues avec une très faible précision, ce qui les rend aujourd'hui extrêmement complexes à prévoir. La variabilité inter-annuelle comprend également des phénomènes abrupts affectant l'éclairement, telles que les éruptions volcaniques (bien que ces dernières soient certainement expliquées par des lois géophysiques pour l'instant inconnues).

La bibliographie reflète les différences en termes de niveau de connaissance entre les deux échelles abordées (intra-annuelles et inter-annuelles) par une répartition extrêmement inégale dans la description de celles-ci. A titre illustratif, on peut citer KLEISSL *et al.* (2013), qui consacre un chapitre entier à la description des variations intra-annuelles, contre un seul paragraphe dédié aux variations inter-annuelles. Cependant, MÜLLER *et al.* (2014) soulignent que le rayonnement estimé à partir des 30 années de données historiques, sans prise en compte de la variabilité de long terme, pourrait aboutir à une sous-estimation ou à une surestimation du potentiel solaire futur. Cela se comprend aisément dans le cas théorique d'une tendance linéaire positive affectant l'ensemble d'une série de données de 60 valeurs : les 30 premières valeurs seront plus faibles que les 30 suivantes. Autrement dit, l'étude des 30 premières valeurs ne permettra pas de représenter de façon correcte les 30 valeurs suivantes. Pour le cas de stations allemandes connaissant une augmentation de l'éclairement moyen

de près de 10 W/m^2 entre 1980 et 2010, MÜLLER *et al.* (2014) montrent que l'absence de prise en compte de cette tendance dans l'estimation du potentiel solaire entraîne une sous-estimation de ce dernier pouvant aller jusque près de 10 %. **La connaissance de l'ensemble des échelles de variabilité du rayonnement solaire est donc indispensable pour une bonne estimation du potentiel solaire.**

La figure 6 présente les variations spatiales des moyennes de l'éclairement sur plusieurs années, calculées à partir des données de ré-analyse MERRA-2 fournies par la NASA (GELARO *et al.*, 2017) sur l'ensemble du globe entre 1980 et aujourd'hui. Cette base de données propose notamment des estimations de l'éclairement au sol, en plus de multiples paramètres météorologiques de surface. Les données sont grillées selon une grille de 0.5° en latitude (approx. 55 km à l'équateur) et 0.625° en longitude (approx. 71.5 km à l'équateur). Le jeu de données d'éclairement mensuel MERRA-2, utilisé pour calculer les moyennes multi-annuelles d'éclairement ici, est détaillé en section 5 du chapitre 2.

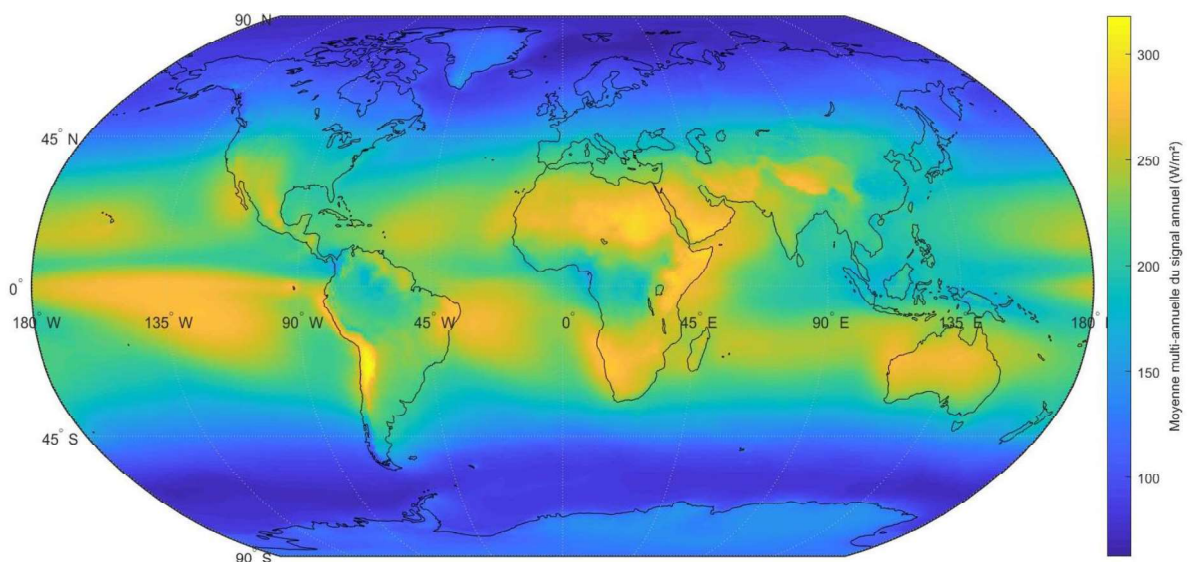


Figure 6 : Carte de la moyenne multi-annuelle de séries temporelles mensuelles d'éclairement fournies par la base de données MERRA 2

Il est clair au vu de cette figure que l'éclairement au sol varie fortement spatialement. On retrouve assez schématiquement une distribution latitudinale, avec un maximum proche de l'équateur autour de 300 W/m^2 , diminuant à l'approche des pôles pour arriver à un minimum inférieur à 100 W/m^2 . On observe cependant des écarts à cette distribution latitudinale. Ces variations évoluent de manière complexe dans l'espace. On voit par exemple une bande de faible éclairement proche de l'équateur, qui pourrait correspondre à la zone de convergence inter tropicale (ITCZ, Intertropical convergence zone). Cette zone est en effet connue pour être une zone de forte couverture nuageuse de par les courants ascendants élevant l'air chaud et humide de surface, créant par condensation des nuages de type cumulus. En plus de l'effet évident de la couverture nuageuse sur l'intensité de l'éclairement au sol, d'autres composants atmosphériques tels que les gaz ou les aérosols - provenant par exemple de l'activité humaine, des déserts, des feux de forêts ou encore des éruptions volcaniques - vont jouer un rôle important sur la ressource solaire. On constate également des zones

de très forte intensité de rayonnement, comme le long de la Cordillère des Andes, autour du Soudan en Afrique, en Australie, ou encore sur l'Atlantique et le Pacifique équatorial. Ces zones de fort éclairement peuvent également être reliées aux dynamiques atmosphérique et océanique, avec notamment les différentes zones de hautes et basses pressions atmosphérique et les gyres océaniques (*e.g.* Parding *et al.*, 2016; Simpson & Dickey, 1981). **Bien que la caractérisation explicite de la variabilité spatiale de l'éclairement sorte du contexte de cette thèse, étant donné les relations évidentes entre les structures spatiales de l'éclairement et les phénomènes atmosphériques à l'échelle synoptique, un rapprochement des résultats au regard de ces phénomènes physiques semble pertinent.**

En conclusion, ces deux figures illustratives des variabilités spatiales et temporelles de l'éclairement soulignent l'importance de la prise en compte de ces variations pour une bonne caractérisation du potentiel solaire. En particulier, il est évident que dans un contexte industriel, une analyse de l'éclairement à l'échelle d'un site particulier doit être réalisée pour prendre en compte toutes les spécificités du climat local. En ce qui concerne les variations temporelles, une prise en compte fiable et précise de l'ensemble des échelles de variations (intra-annuelle et inter-annuelle) semble nécessaire.

Au vu des différentes variations de l'éclairement, on comprend mieux la complexité de la caractérisation du potentiel solaire. Dans le cadre de projets solaires, différentes méthodes existantes de caractérisation peuvent être répertoriées.

4 Méthodes de caractérisation du potentiel solaire

Le potentiel solaire est principalement estimé à partir d'indicateurs statistiques sur la disponibilité de l'éclairement. Les méthodes de calcul des indicateurs sont assez diverses. Cependant, le plus souvent, **les indicateurs statistiques sont estimés à partir de l'analyse de séries temporelles historiques d'éclairement. Ces séries temporelles historiques sont typiquement constituées de 10 à 20 ans de données d'éclairement, agrégées à l'échelle annuelle** (McMAHAN *et al.*, 2013a). Ces données sont des données considérées comme une référence : elles sont issues des données mesurées sur le site d'intérêt et/ou estimées par satellites sur la grille correspondante. L'analyse statistique de ce jeu de données offre un ensemble d'indicateurs, tel que l'éclairement moyen ou médian du site, utilisé comme caractéristique typique du potentiel solaire. À cette analyse statistique de la ressource solaire s'ajoute une estimation des incertitudes liées aux indicateurs, qui peuvent être regroupées en deux catégories (Kariuki & Sato, 2018; McMAHAN *et al.*, 2013a; MOSER *et al.*, 2017) : les incertitudes liées à la qualité des données utilisées, et les incertitudes liées à la variabilité inter-annuelle de l'éclairement.

Premièrement, on peut compter les incertitudes liées à la qualité des données utilisées : e.g. incertitudes de mesures dépendant de l'instrumentation choisie, incertitudes des modèles de conversion d'images satellites en données de rayonnement, incertitudes propres à la méthode de

descente d'échelle de la maille au point d'intérêt, etc. Dans la pratique, un minimum de 10 ans de données est recommandé pour pouvoir caractériser statistiquement l'éclairement sur le site d'intérêt (GUEYMARD, 2008; VIGNOLA *et al.*, 2012). Idéalement, des données couvrant au moins la durée du prêt sont attendues par les banquiers, soit typiquement 20 ans. D'autres soutiennent que des données couvrant une période climatique serait un minimum pour garantir la représentativité de l'analyse, soit entre 30 et 50 ans de données (MILLER *et al.*, 2013). En parallèle, il est recommandé d'utiliser des données de long terme provenant de stations au sol de qualité. Cependant, compte tenu de la difficulté et du coût de l'acquisition de données *in-situ* de haute qualité, il est très rare d'avoir à sa disposition plus de 20 ans de données sur le site d'intérêt. Dans la quasi-totalité des cas, le site prédisposé à recevoir la centrale solaire ne dispose à l'origine d'aucune mesure *in-situ* de l'éclairement. Des données grillées sont alors utilisées pour estimer le potentiel du site (HABTE *et al.*, 2014; HOYER-KLICK *et al.*, 2009). Pour affiner la représentativité géographique des données grillées, une descente d'échelle de la maille au point est nécessaire, ajoutant nécessairement une incertitude aux données (FERNÁNDEZ-PERUCHENA *et al.*, 2015). Cette incertitude peut être limitée en choisissant des données grillées avec la meilleure résolution possible (MEYER, 2010). Par conséquent, les données grillées utilisées sont majoritairement des données satellitaires, offrant des données de qualité long terme à une résolution souvent inférieure à 0.1°. Pour réaliser la descente d'échelle de la maille au point, Fitch's Rathings (2018) conseille de prendre au moins une année de mesure sur le site d'intérêt, et d'appliquer une correction aux données de rayonnement estimées par satellites. Cette recommandation se retrouve largement dans la littérature (*e.g.* POLO *et al.*, 2011; Vernay *et al.*, 2012) et est appliquée dans la grande majorité des projets. Pour finir, une descente d'échelle plus fine peut être réalisée, grâce à l'analyse des effets d'ombrages des données satellitaires (RUIZ-ARIAS *et al.*, 2010). La série temporelle de données considérée va donc être soumise à des incertitudes propres au jeu de données utilisées. **Ces incertitudes sont relativement bien connues et communiquées par les fournisseurs de données** (McMAHAN *et al.*, 2013b; MOSER *et al.*, 2017).

La seconde source majeure d'incertitude se trouve dans la variabilité inter-annuelle de l'éclairement. Statistiquement, on pourrait définir la variabilité inter-annuelle de la ressource solaire comme la quantité traduisant l'amplitude de la dispersion des années particulières de rayonnement par rapport à la moyenne multi-annuelle (SENGUPTA *et al.*, 2015). La variabilité inter-annuelle est complexe à estimer pour deux principales raisons : premièrement car, comme nous l'avons vu précédemment, l'état des connaissances sur la caractérisation des variations de long terme de l'éclairement est limité ; ensuite car le jeu de données utilisé est nécessairement limité de par la disponibilité temporelle des données satellitaires considérées. Ainsi, toute variation de la ressource solaire supérieure à la longueur de la fenêtre d'analyse ne pourra pas être prise en compte, qu'elle soit due à un cycle de grande période, à une hypothétique tendance de long terme, ou à un événement particulier tel qu'une éruption volcanique.

Les méthodes de quantification de l'incertitude due à la variabilité inter-annuelle de la ressource solaire peuvent varier fortement. Cela qui peut entraîner de gros écarts dans les valeurs associées à cette incertitude. La méthode la plus simple repose sur le calcul de ou de l'écart type ou de la variance de l'échantillon de données annuelles, sous la forme du coefficient de variation (*e.g.* Wilcox & Gueymard, 2011). Ce dernier est défini comme le ratio entre l'écart type de l'éclairement annuel et la moyenne inter-annuelle de l'éclairement sur la période correspondante. Une alternative

trouvée fréquemment dans la littérature repose sur l'utilisation de probabilité d'excédences (P10 et P90). Cette alternative sera étudiée en détail en chapitre 3. Notons que ces approches basées sur l'analyse statistique de la série historique permettent d'estimer l'incertitude due à la variabilité inter-annuelle de l'éclairement, mais ne permettent pas de prendre en compte l'influence de ces variations sur le potentiel solaire futur. Or, comme introduit précédemment par l'étude de Müller *et al.* (2014), cette dernière peut être importante.

Le potentiel solaire d'un site est donc estimé en pratique par une analyse statistique d'une série temporelle, typiquement de 20 ans de données d'éclairement, à laquelle sont parfois associées les incertitudes liées à la qualité des données utilisées et à la variabilité inter-annuelle. L'incertitude liée à l'influence des variations inter-annuelles est aujourd'hui importante. Elle pourrait être réduite grâce à une meilleure compréhension des cycles et des tendances de long terme influençant l'éclairement.

5 Objectifs de la thèse

Le premier objectif de cette thèse est de **faire un état de l'art sur les méthodes d'estimation du potentiel solaire**. Dans un premier temps, nous avons exploré les analyses statistiques classiquement appliquées aux données historiques, dans le but de construire des indicateurs statistiques. Ces indicateurs statistiques ont pour vocation d'être utilisés comme support d'aide à la décision. Il sera mis en évidence qu'en estimant le potentiel solaire à partir d'une analyse statistique historique, il est implicitement supposé que les variations inter-annuelles de l'éclairement sont négligeables. Il est également supposé que la quantité d'éclairement passée est représentative du potentiel solaire futur. La question de la sensibilité des indicateurs statistiques au nombre de données disponibles sera soulevée. Dans un second temps, pour garantir la significativité statistique des indicateurs fournis, différentes méthodes de l'état de l'art visant la génération de données synthétiques plausibles ont été considérées. Une étude des hypothèses sous-jacentes à ces méthodes, généralement ni explicitées ni discutées, a été réalisée.

Une hypothèse sous-jacente, presque systématiquement postulée dans les méthodes étudiées, est l'hypothèse de stationnarité temporelle de l'éclairement (LOHMANN *et al.*, 2006; TOMSON *et al.*, 2008; BADESCU, 2008; WILCOX & GUEYMARD, 2010). Cependant, d'autres études montrent une non-stationnarité de l'éclairement, avec la présence de variations de long terme, qualifiées d'éclaircissement global et d'obscurcissement global (*e.g.* MÜLLER *et al.*, 2014; PFEIFROTH *et al.*, 2018; SANCHEZ-LORENZO *et al.*, 2015; WILD, 2009). Le second objectif de la thèse repose ainsi sur **l'étude approfondie de l'hypothèse de stationnarité temporelle de l'éclairement**. Pour cela, une analyse de la bibliographie disponible sur les tendances de long terme de l'éclairement a été réalisée. L'objectif de cette analyse a été de déterminer si l'information disponible dans la bibliographie est adaptée à l'aide à la décision dans un contexte de projet solaire photovoltaïque.

L'étude de la stationnarité temporelle de l'éclairement a mis en lumière l'intérêt de l'utilisation d'une méthode d'analyse des séries temporelles d'éclairement qui soit temporelle et multi-échelle. Cet outil permettrait de synthétiser l'information disponible dans la bibliographie, tout en caractérisant les modes de variabilité inter-annuels de l'éclairement. Ainsi, le troisième objectif de la thèse porte sur **l'identification et la caractérisation des variations de long terme de l'éclairement solaire**. Pour cela, une analyse des modes de variabilité propres à l'éclairement a été réalisée à l'échelle globale. Cette analyse nous a mené à distinguer trois classes de variabilité : la variabilité intra-annuelle, la variabilité annuelle à décennale, et la variabilité multi-décennale. Ces trois classes de variabilité ont été étudiées de manière distincte pour différentes bases de données. Grâce à la compréhension issue de ces résultats, il pourrait être possible, à terme, de réduire les incertitudes liées à l'estimation de la variabilité inter-annuelle dans les projets solaires.

Il est possible de résumer les trois objectifs de la thèse comme suit :

- Objectif 1** Faire le point sur les possibilités et les limites des méthodes de caractérisation du potentiel solaire disponibles dans la littérature.
- Objectif 2** Etudier de manière approfondie la stationnarité temporelle de l'éclairement.
- Objectif 3** Identifier et caractériser les variations de long terme de l'éclairement solaire.

Le document est structuré en huit chapitres. Le chapitre 1 correspond au présent chapitre introductif. Le chapitre 2 présente les différentes données qui seront exploitées au long de la thèse : les données *in-situ* GEBA, les données de ré-analyses MERRA-2 et les données de modèles climatiques IPSL-CM6. Le chapitre 3 s'attache à l'objectif 1, et explore les possibilités et les limites des méthodes de caractérisation du potentiel solaire de l'état de l'art. Le chapitre 4 aborde l'objectif 2, en se basant sur la bibliographie connexe au sujet, sur des mesures de stations et sur des tests statistiques de stationnarité temporelle. Le chapitre 5 présente la méthode de décomposition temps-fréquence choisie pour explorer en détail les variations de long terme de l'éclairement. Les chapitres 6 et 7 s'attachent tous deux à l'objectif 3. Le chapitre 6 analyse les données *in-situ* GEBA, présentées dans le chapitre 3, grâce à la méthode de décomposition temps-fréquence choisie. Cette étude est limitée par le nombre de données GEBA disponible et la couverture spatiale peu dense des stations considérées. Ainsi, le chapitre 7 propose de discuter et d'affiner les conclusions issues de l'analyse des stations GEBA grâce à différentes bases de données grillées. Ces dernières possèdent un long historique et/ou une couverture spatiale mondiale. En conclusion, le chapitre 8 propose un ensemble de conclusion pour notre étude, et ouvre des perspectives grâce à un exemple illustratif. Ce dernier prendra en compte d'un des modes de variabilité identifié pour une estimation de l'éclairement solaire.

Chapitre 2

Description des données

Table des matières

Description des données	27
1 Introduction	29
2 Variables d'intérêts	29
3 Données de mesure <i>in-situ</i> La base de données Global Energy Balance Archive - GEBA.....	30
4 Données satellitales La base de données CLARA A2	40
5 Données de ré-analyses La base de données MERRA-2	45
6 Données de modèles climatiques La base de données IPSL-CM6A-LR.....	46
7 Exemple de séries temporelles d'éclairement pour l'ensemble des bases de données	49
8 Conclusion.....	51

1 Introduction

Différentes bases de données d'éclairement vont être utilisées pour appuyer, illustrer ou étayer notre étude. En particulier, de nombreuses analyses vont être réalisées grâce à des données *in-situ* de long terme. Par la suite, pour d'assurer une couverture spatiale globale et une disponibilité temporelle plus importante des données étudiées, nous serons amenés à utiliser une base de données satellitaires, une base de données de ré-analyses et différentes réalisations d'un modèle climatique.

Après avoir défini de manière précise les variables d'intérêt pour notre étude (section 2), chaque section de ce chapitre s'attache à la description d'une base de données étudiée. Les données *in-situ* étant les plus largement utilisées dans notre étude, la section s'attachant à leur description sera plus détaillée.

2 Variables d'intérêts

Nous allons étudier dans le cadre de cette thèse deux variables : le rayonnement global horizontal (GHI), en moyenne mensuelle, et l'indice de clarté correspondant (KT). Le KT correspond au ratio entre la quantité de rayonnement reçu au sol et la quantité de rayonnement correspondante reçu au sommet de l'atmosphère (BLACK *et al.*, 1954). Ainsi, l'indice de clarté KT caractérise la transparence de l'atmosphère. Il varie classiquement entre 0.1 pour un ciel très couvert et 0.8 pour un ciel clair. En ôtant les variations TOA du GHI, le KT permet de s'affranchir d'une grande partie de l'influence du cycle saisonnier sur l'éclairement. Cela est intéressant car le cycle annuel est d'une telle amplitude qu'il peut dans, certains cas, rendre difficile à identifier et caractériser d'autres échelles de variations. De plus, le KT permet de s'affranchir de la dépendance latitudinale de ce cycle saisonnier, et donc de comparer les variations inter-annuelles observées entre différentes latitudes.

Un autre indice a été envisagé : l'indice de ciel clair (KC). Il correspond au ratio entre la quantité de rayonnement reçu au sol et le rayonnement qui aurait été reçu sol en l'absence de nuage. Le KC tiens donc compte de l'absorption et de la diffusion des gaz et des particules atmosphériques (C. A. GUEYMARD, 2012). Ainsi, en plus de s'affranchir de l'influence de la position du soleil, le KC s'affranchit de l'influence de l'épaisseur optique de l'atmosphère, rendant compte, notamment, de la vapeur d'eau, des effets de l'ozone et des aérosols. Cet indice de ciel clair varie typiquement entre 0.05 et 1. Les résultats des études réalisées à partir de ces deux indices étant très proches à l'échelle mensuelle, nous avons choisi de présenter ici uniquement l'indice construit à partir de la méthode la plus simple : **l'indice de clarté calculé à partir de l'éclairement TOA, lui-même calculé à partir de la géométrie Terre-Soleil.**

3 Données de mesure *in-situ*

La base de données Global Energy Balance Archive - GEBA

3.1 Présentation

Nous avons choisi d'utiliser la base de données GEBA (Global Energy Balance Archive, GILGEN & OHMURA, 1999). Cette base de données a été créée et est maintenue par l'ETH Zurich. Elle met à disposition un grand nombre de mesures météorologiques de surface concernant le flux d'énergie de surface à l'échelle mensuelle. Les données sont en libre accès via le site <http://www.geba.ethz.ch>. Les données d'éclairement solaire en moyennes mensuelles fournies sont toutes exprimées en W/m².

Cette base de données a été sélectionnée car **elle rassemble des mesures** :

- **couvrant des climats variés.** En effet, GEBA rassemble plus de 2 500 stations de mesure tout autour du globe, pour un total de près de 500 000 données d'irradiation en moyenne mensuelle. La répartition géographique de ces 2500 stations est présentée sur la figure 7.
- **de long terme**, typiquement multi-décennales. Bien que la période temporelle couverte par la base de données dépende de la station considérée, elle offre des séries temporelles de données de 10 ans minimum, pour aller jusqu'à des séries temporelles de 91 ans pour la station Potsdam.
- **de grande qualité.** Les données proviennent de différents réseaux et organismes de référence, tels que le World Radiation Data Centre (WRDC), de réseaux de météorologie nationaux ou encore de réseaux de recherche reconnus comme le BSRN – Baseline Surface Radiation Network – maintenu par le World Radiation Monitoring Center (WRMC), le réseau ARM – Atmospheric Radiation Measurement – maintenu par le US department of Energy et ou encore SurfRad - maintenu par la NASA (National Aeronautics and Space Administration).

De par leur qualité et leur disponibilité multi-décennale, ces données ont été utilisées dans de nombreuses études sur la détection de variations de long terme de l'éclairement global (e.g. CHIACCHIO & WILD, 2010; GILGEN & OHMURA, 1999; NORRIS & WILD, 2007; PHILIPONA *et al.*, 2009; SANCHEZ-LORENZO *et al.*, 2013; SANCHEZ-LORENZO *et al.*, 2015; WILD, 2009).

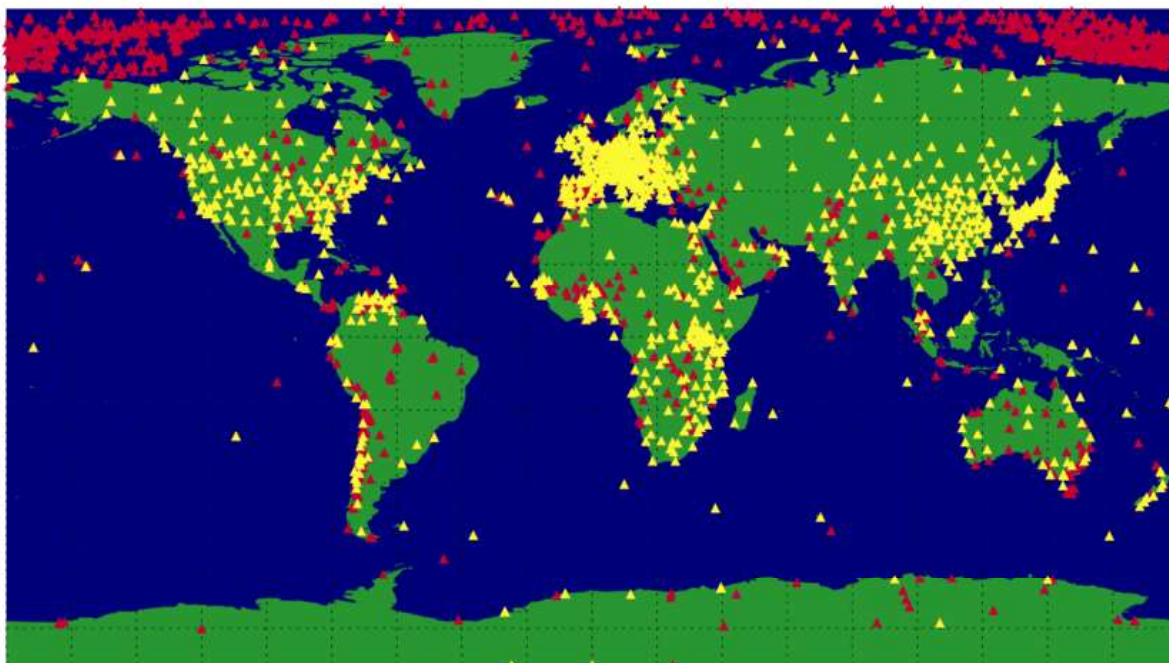


Figure 7 : Répartition géographique des 2500 stations GEBA. En jaune sont représentées les stations possédant au moins trois années de données. Figure extraite de (WILD et al., 2017).

3.2 Le contrôle de qualité des données GEBA

Malgré la grande qualité du réseau GEBA, différentes sources d’erreurs peuvent affecter la donnée mensuelle du rayonnement solaire pour les sites de mesure sélectionnés. Certaines erreurs sont communes aux mesures pyranométriques standards : dérives d’étalonnage des capteurs, pannes électriques ou électronique, phénomènes météorologiques comme le givre, la buée ou encore les salissures du dôme par les poussières ou les déjections d’oiseaux, etc. D’autres sources d’erreurs sont spécifiques à la mesure de long terme : variations des méthodes et des moyens de maintenance sur plusieurs décennies, changements d’instrumentation n’ayant pas les exactement les mêmes caractéristiques en dépit de leurs étalonnages, etc. L’estimation précise, la modélisation et la correction de ces erreurs sur la mesure sont complexes. Cependant, un travail conséquent a été réalisé par l’ETH Zurich pour qualifier la qualité de chaque série temporelle mise à disposition. Les données GEBA étant les plus utilisées dans cette thèse, **une attention particulière a été portée à comprendre la procédure de contrôle de qualité réalisé par l’ETH Zurich. Une sélection des stations disponibles a été réalisée, afin d’assurer l’adéquation entre les données utilisées et les objectifs de la thèse.**

3.2.1 Le contrôle de qualité réalisé par l'ETH Zurich

L'ETH Zurich propose une procédure de contrôle de qualité à l'ensemble des données contenues dans la base de données GEBA, décrit par Gilgen et Ohmura (1999). Nous nous intéresserons ici uniquement au contrôle de qualité des données d'éclairement global horizontal. **Le contrôle de qualité s'appuie sur quatre facteurs principaux :**

- **Le type d'instrument utilisé et le système d'acquisition de données associé.** La grande majorité des stations sont équipées de pyranomètres qui mesurent l'éclairement avec une incertitude relative estimée à 2 % (GILGEN & OHMURA, 1999). Il s'agit donc d'instruments de grande qualité, encourageant à considérer les données issues de la base GEBA comme *a priori* des données de qualité.
- **La maintenance de l'instrument et du système d'acquisition.** En effet, il est évident que la maintenance des instruments de mesure est cruciale pour la qualité des séries temporelles observées.
- **Le nombre et la taille des périodes de données manquantes.** En effet, la présence et la taille de données manquantes à l'échelle intra-journalière peuvent conduire à une complétion des données, induisant des incertitudes dans les données agrégées à la journée, au mois ou à l'année.
- **L'attention prise dans l'évaluation et le traitement des données lors de l'enregistrement initial des données.**

Plus précisément, le contrôle de qualité de l'ETH Zurich se déroule en cinq étapes, rassemblées dans le tableau 1. À chaque étape (colonne 1) est associé un critère (colonne 3). Ce critère est vérifié par des analyses automatisées sur les valeurs ou par des tests visuels (colonne 2). Les tests visuels sont réalisés par des scientifiques de l'équipe s'occupant de la station de mesure considérée. Ce dernier point est crucial dans le contrôle de qualité, et souligne la bonne prise en considération de l'expertise des scientifiques et du personnel de chaque station. En effet, le personnel d'une station est en général le seul à avoir accès aux enregistrements originaux, à l'historique des instruments et de la maintenance sur site, etc. Les colonnes 4 et 5 explicitent le code qualité (colonne 5) associé à l'adéquation ou non des données aux critères d'intérêts (colonne 4). La signification du code qualité est détaillée dans le tableau 2. Ce code est associé à chacune des données mensuelles. **Le code qualité de la série temporelle correspondante est égal au code qualité minimum de l'ensemble des mois contenus dans la série de données.**

Malgré ce contrôle de qualité, qui peut être qualifié de strict, l'existence de données erronées ne peut pas être exclue. Les erreurs résiduelles sont quantifiées à l'aide de comparaisons de long terme entre deux séries temporelles mensuelles mesurées sur la même station pour des périodes de temps coïncidente, ou entre deux stations voisines possédant un climat proche. **L'erreur relative pour l'éclairement global est estimée de l'ordre de 5 % pour les données mensuelles, et de 2 % pour les données annuelles par Gilgen & Ohmurha (1999).** Les auteurs ont également mis en évidence que les erreurs détectées par les différents contrôles de qualités étaient de nature aléatoire (*i.e.* équiréparties en intra-journalier, en intra- et inter-mensuel et en annuel) à l'échelle globale. Cela assure ainsi que **le contrôle de qualité n'induit pas de biais systématique sur les données.**

Pour notre étude, **nous n'avons sélectionné que les données répondant aux codes qualité 7 ou 8. Cela correspond à 689 stations**, dont la répartition géographique est présentée sur la figure 8. Pour chaque station, les moyennes multi-annuelles des sommes annuelles d'irradiation sont représentées en couleur.

On constate sur la figure 8 une hétérogénéité de la couverture géographique des stations considérées. Les moyennes latitudes de l'hémisphère Nord (30° N – 60° N) sont particulièrement bien représentées. D'autres zones ne sont pas couvertes, telles que les océans, l'Amérique du Sud ou encore les Pôles. La majorité des stations sont situées en Europe, aux États-Unis et en Asie, quand le continent Africain, l'Océanie et l'Amérique centrale ne présentent que quelques stations. On retrouve dans les moyennes d'irradiations annuelles une répartition latitudinale, avec un minimum proche des pôles, et qui croît graduellement à l'approche de l'équateur. Cette répartition latitudinale n'est clairement visible que sur l'hémisphère Nord, de par la meilleure couverture de cette zone. Bien que la couverture géographique des stations GEBA ne soit pas homogène, les stations couvrent une large zone géographique.

Comme annoncé précédemment, nous sommes intéressés dans le cadre de notre étude par les stations GEBA offrant des séries temporelles de grande qualité, avec une couverture la plus globale possible, mais également de long terme. Ainsi, nous avons réalisé une nouvelle sélection de stations, afin de retenir uniquement les stations disposant de plus de 30 années complètes de données mensuelles consécutives.

Étape	Type de tri	Critère	Conséquence	Code
ETH 1	Automatisé	Les données doivent être comprises entre 0 W/m ² et la valeur du rayonnement au sommet de l'atmosphère.	Si le critère est satisfait	2
			Si le critère n'est pas satisfait	5
ETH 2	Visuel	Les données doivent être comprises entre 0 W/m ² et une valeur théorique, issue d'une base de données de couverture nuageuse externe.	Si la donnée de couverture nuageuse existe pour le point et la date et que le critère est satisfait	8
			Si la donnée de couverture nuageuse considérée est une climatologie et que le critère est satisfait	7
			Si la donnée de couverture nuageuse existe pour le point et la date mais que le critère n'est pas satisfait	2
			Si la donnée de couverture nuageuse considérée est une climatologie et que le critère n'est pas satisfait	3
			Si aucune donnée de nuages n'est disponible pour le point ou la période d'intérêt	5
ETH 3	Automatisé et visuel	Pour les séries possédant plus de 15 années validées par ETH 1 et ETH 2, les écarts-types des séries temporelles annuelles sont comparés à un seuil défini en fonction du climat considéré.	Si la série temporelle validée par ETH 1 et ETH 2 possède moins de 15 ans	5
			Si la série est inférieure au seuil défini	8
			Si la série est inférieure ou supérieure au seuil, un contrôle visuel est appliqué	8 ou 2
ETH 4	Visuel	Comparaison visuelle entre les données annuelles et mensuelles GEBA de stations voisines, ou possédant un climat d'irradiation similaire.	Si les stations ont des résultats similaires	8
			Si les stations n'ont pas des résultats similaires	5
ETH 5	Visuel	Comparaison des moyennes climatologique d'irradiations de stations aux climats similaires.	Si les stations ont des résultats similaires	8
			Si les stations n'ont pas des résultats similaires	5

Tableau 1 : Résumé des étapes du contrôle de qualité appliqué aux séries temporelles du réseau GEBA. (inspiré de Gilgen & Ohmura (1999))

Code qualité	Etat de la donnée	Description
Blanc	Non déterminé	Les données n'ont pas encore été contrôlées
1	Erroné	Les données sont jugées valides par le contrôle de qualité automatique, mais sont rejetées après contrôle visuel d'un climatologue.
2	Erroné	Les données sont jugées erronées par le contrôle de qualité automatique et par le contrôle de qualité visuel réalisé par le climatologue.
3	Erroné	Les données sont jugées erronées lors d'une comparaison avec la couverture nuageuse d'une large zone. Il est donc possible qu'elles soient valides si l'irradiation du site d'intérêt est particulière.
4	Non signalé	Aucune indication n'est donnée par les auteurs quant à un code qualité 3.
5	Non déterminé	Les données ne peuvent être validées par le contrôle de qualité automatique ni visuel car la série temporelle ne satisfait pas les conditions requises à l'application du test considéré.
6	Valide	Les données sont jugées erronées par le contrôle de qualité automatique, mais sont réhabilitées par le contrôle visuel d'un climatologue.
7	Valide	Les données sont jugées valides par le contrôle de qualité automatique comme par celui réalisé par le climatologue. Une attention doit cependant être portée au fait que les comparaisons avec la couverture nuageuse ont été réalisées sur une zone large de par l'absence de données sur le point d'intérêt.
8	Valide	Les données sont jugées valides par le contrôle de qualité automatique comme par celui réalisé par le climatologue.

Tableau 2 : Code qualité fourni par l'ETH Zurich concernant les données GEBA (inspiré de Gilgen & Ohmurha (1999))

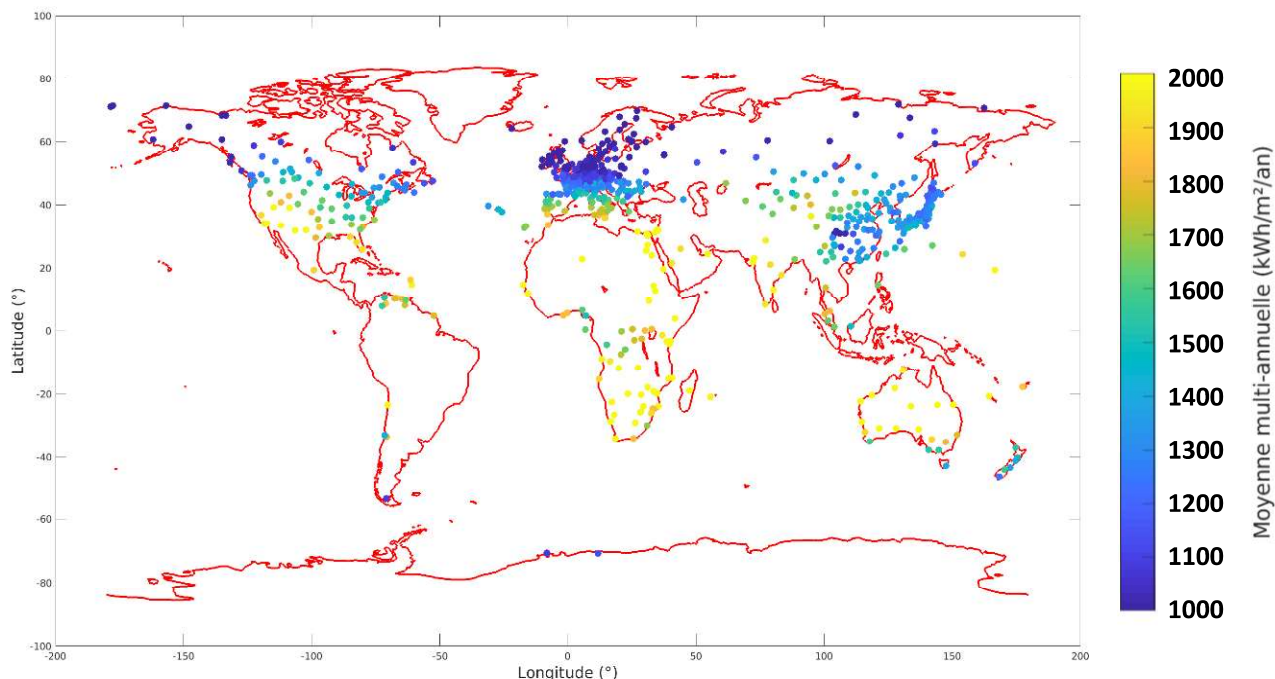


Figure 8 : Répartition géographique des 689 stations GEBA sélectionnées.
Couleur : moyenne multi-annuelle d'irradiation annuelle (kWh/m²).

3.2.2 Sélections de stations GEBA de long terme

Une première sélection de stations GEBA été réalisée dans le cadre de cette thèse. **Le but de cette sélection est de retenir, parmi les 689 stations GEBA de grande qualité, les stations possédant au moins une période de plus de 30 années de données mensuelles consécutives, absolument sans données manquantes. 63 stations GEBA satisfont ce critère.** La figure 9.a illustre la répartition géographique de ces stations. De même que précédemment, les irradiations annuelles moyennes, calculées à partir des données annuelles disponibles pour chaque station, sont représentées en couleur.

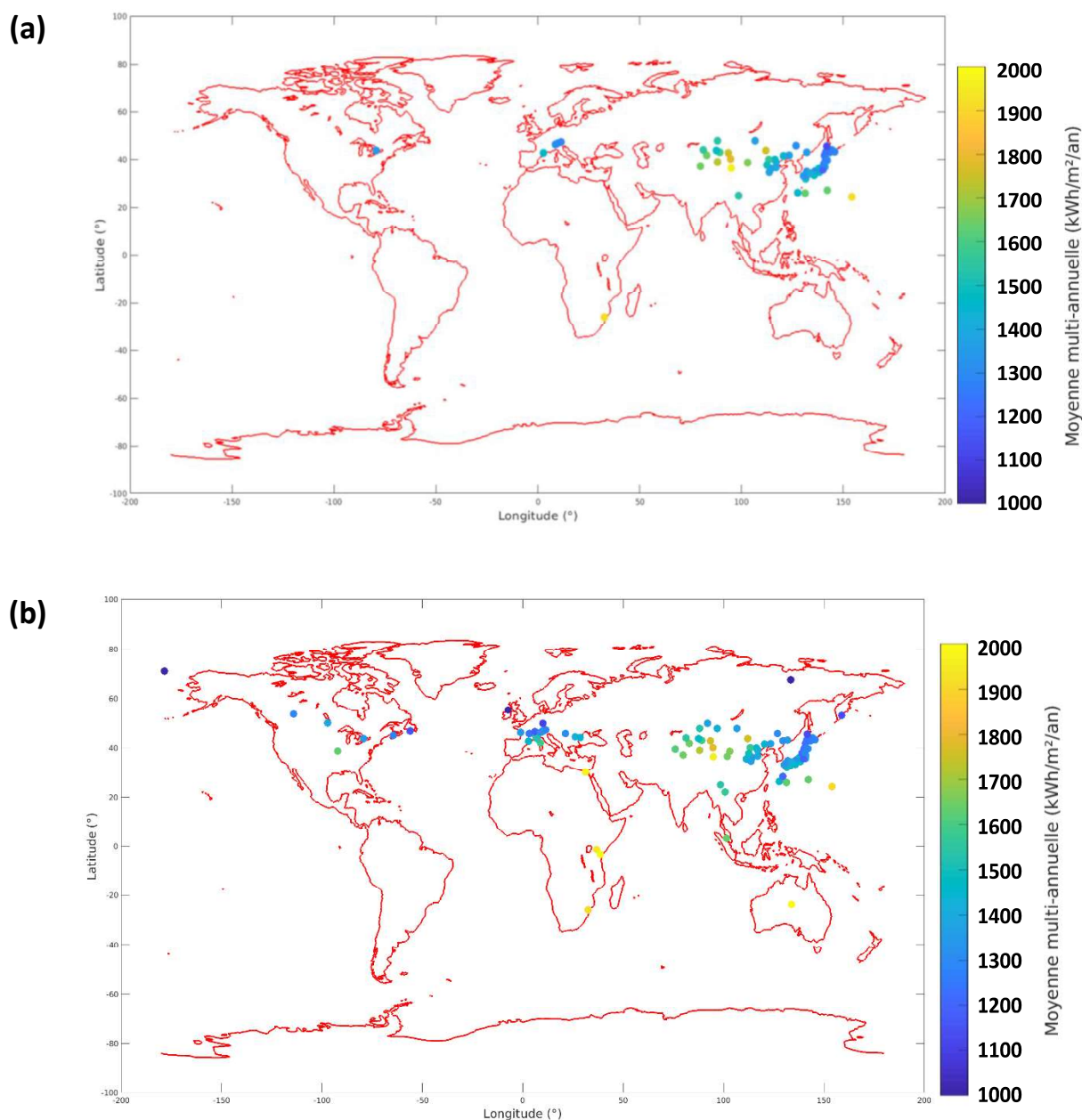


Figure 9 : (a) Répartition géographique des 63 stations GEBA sélectionnées, (b) Répartition géographique des 104 stations GEBA sélectionnées. Couleur : moyenne multi-annuelle d'irradiation annuelle (kWh/m^2).

La figure 9.a montre que cette sélection réduit drastiquement la représentativité de la couverture géographique des stations par rapport à la sélection initiale de 689 stations. Très peu de zones sont couvertes, avec la très grande majorité des stations situées en Asie, et en particulier au Japon (31 stations). L'Europe est uniquement couverte par quatre stations, quand l'Amérique et l'Afrique ne possèdent qu'une seule station de plus de 30 années complètes consécutives. A l'exception de la station africaine, l'ensemble des 63 stations se trouve entre 20 °N et 50 °N.

L'analyse de la diversité de représentations climatiques est complétée en annexe II par une analyse de la répartition des stations au regard de la classification des climats de Köppen & Geiger (1936) et actualisée par Peel *et al.* (2007). Il est évident que cette sélection de stations semble trop drastique et peu appropriée dans l'optique d'étudier la variabilité de long terme de différents climats. Cependant, **cette première sélection de stations GEBA reste d'intérêt : elle constitue un ensemble de stations de très grande qualité, pour lequel aucun artefact dû à une discontinuité de la série de données ne pourra être observé sur au moins trente années consécutives. Elle pourra ainsi servir de référence solide aux différentes analyses réalisées.**

Afin de garantir une meilleure couverture géographique à la sélection de stations GEBA, tout en conservant la qualité et l'aspect de long terme, nous avons procédé à une sélection un peu moins drastique de stations. Pour cette dernière, nous avons relâché la contrainte sur la présence de données mensuelles manquantes : **un maximum de deux données mensuelles manquantes consécutives est toléré. Le nouvel ensemble de stations satisfaisant ces conditions est constitué de 104 stations**, dont les localisations sont représentées sur la figure 9.b.

La figure 9.b montre à nouveau une hétérogénéité de la couverture géographique des stations considérées, mais en nette amélioration comparativement à l'ensemble précédant. Cette disparité reflète dans les grandes lignes la répartition globale de l'ensemble des stations GEBA présenté en figure 8 : certaines zones sont très bien couvertes, comme le Nord-Est de l'Asie, l'Europe centrale ou le Japon, quand d'autres zones ne sont pas couvertes, telles que les océans, l'Amérique du Sud, ou encore les Pôles. La majorité des stations sont situées aux moyennes latitudes de l'hémisphère Nord, quand le continent Africain et l'Océanie ne possèdent que quelques stations. Cependant, des différences sont notables : la Russie, le Nord de l'Europe et l'Amérique centrale ne sont par exemple pas représentés par cet ensemble de stations, contrairement à la sélection initiale de 689 stations. **Au regard de cette disparité, nous pouvons avancer que les conclusions résultantes de l'analyse de ces stations ne pourront pas être généralisées à une échelle globale.** Quoi qu'il en soit, ce sous-ensemble de long terme sélectionné de manière drastique recouvre une grande diversité de conditions climatiques, se traduisant en termes d'intensité d'irradiation comme de variabilité de celui-ci. On trouve par exemple des stations dans des zones désertiques, qui montrent une moyenne élevée, comme observé sur les stations désertiques d'Afrique et d'Océanie avec une irradiation annuelle de plus de 2200 kWh/m². À l'opposé, d'autres stations se trouvent dans des climats beaucoup plus variables, telles que les régions soumises aux processus de la mousson en Asie du Sud-Est. La répartition de ces stations au regard des zones climatiques de références de Köppen & Geiger (1936), actualisée par Peel et al (2007), montre également la représentation de différents climats par ce sous-ensemble de stations (annexe III).

Cette sélection de 104 stations est donc bien de qualité, de long terme, et de couverture globale. L'annexe IV rassemble les caractéristiques des 104 stations sélectionnées. Il est cependant important de garder à l'esprit qu'en autorisant deux données mensuelles manquantes consécutives, il est possible qu'un biais systématique s'introduise dans les analyses, si par exemple pour une station donnée deux mois d'été manquent de façon systématique. En ce sens, la sélection de 63 stations GEBA complètes permettra de valider les résultats issus de la sélection de 104 stations.

Enfin, il est à noter que le calcul de l'éclairement moyen annuel en W/m^2 , utilisé par la suite, est obtenu en divisant l'irradiation annuelle correspondante en Wh/m^2 par le facteur 8760 (24×365) pour une année non-bissextille. De la même manière, l'éclairement moyen mensuel est obtenu en divisant l'irradiation mensuelle correspondant par un facteur pouvant être 672, 696, 720, 744 selon que le nombre de jours dans le mois considéré est 28, 29, 30 ou 31.

3.3 Complétion de la sélection de 104 stations GEBA

Au cours de la thèse, une décomposition temps-fréquence des séries temporelles de la sélection de 104 stations GEBA a été réalisée. Nous verrons que la méthode de décomposition choisie, détaillée en chapitre 5, réagit très mal aux données manquantes. En ce sens, une complétion des données de la sélection GEBA a été réalisée.

Les séries de données sélectionnées ne comportant au plus que 2 mois manquants consécutifs sur des périodes de plus de 30 ans, nous avons choisi d'interpoler les données manquantes sur la base d'un indice mensuel défini comme le rapport des éclairement mensuels sur les moyennes multi-annuelles de ces derniers.

L'influence de cette complétion de données peut être aisément estimée à partir de la sélection de 63 stations GEBA complètes sur plus de 30 ans. Pour cela, l'expérimentation suivante est réalisée : (i) le nombre moyen de données manquantes de la série de 104 stations GEBA est calculé : il s'agit de trois données manquantes. (ii) Pour chaque station, trois données sont aléatoirement sélectionnées pour être retirées du jeu de données, créant ainsi artificiellement une série temporelle incomplète. Au maximum, deux données consécutives sur les trois sont autorisées. (iii) Les données manquantes sont complétées avec la méthode de complétion choisie. (iv) L'écart quadratique moyen (Root Mean Square Error - RMSE) entre l'échantillon initial et l'échantillon complété permet d'estimer la différence entre les deux signaux. (v) Les étapes allant de (ii) à (iv) sont reproduites 100 000 fois pour assurer une robustesse statistique suffisante aux résultats. (vi) La moyenne des RMSE est calculée pour l'ensemble des tirages et l'ensemble des stations.

Cette analyse a montré que l'écart quadratique moyen entre l'échantillon original et l'échantillon possédant trois données manquantes complétées est de moins de 1 %. Ainsi, malgré sa simplicité, la méthode de complétion de données ne semble pas introduire d'erreur notable dans la série complétée de données. Ce résultat est d'autant plus fiable que, sur l'ensemble des 104 stations GEBA, 101 possèdent moins de 10 données manquantes, les trois restantes étant composées de 15,

42 et 60 données manquantes. Dans ce dernier cas extrême de 60 données mensuelles manquantes, recensé pour la station Wrangel, le RMSE s'élève à 1.2 % si les données manquantes sont disposées de façon aléatoire, et à 6 % pour le cas théorique où les 60 données manquantes seraient successives deux à deux. Cette situation étant extrême, elle n'a pas remis en question la méthode de complétion choisie. **Une nouvelle sélection de données a donc été construite, composée de 104 stations GEBA offrant des séries temporelles d'irradiation mensuelles complètes sur au moins 30 années consécutives.**

3.4 Calcul de l'indice de clarté

Il existe aujourd'hui de nombreuses références dans la littérature détaillant avec précision le calcul de la position du soleil (BRETAGNON & FRANCOU, 1988), ainsi que les algorithmes utilisés couramment pour réaliser ce calcul (BLANC & WALD, 2012; BLANCO-MURIEL *et al.*, 2001; GRENA, 2008, 2012; MICHALSKY, 1988; REDA & ANDREAS, 2004; RIGOLLIER *et al.*, 2000). Nous ne détaillerons ni ces calculs, ni ces algorithmes ayant chacun un temps de calcul, une précision et une période de validité répondant à un besoin spécifique. Pour calculer le TOA utilisé dans le calcul de l'indice de clarté de la sélection de stations GEBA, nous utiliserons les données issues du modèle Solar Geometry 2 (SG2) (BLANC & WALD, 2012). Celui-ci propose des données de 1980 à 2030 avec une erreur proche de 0.002°, pour un temps de calcul adapté aux contraintes de cette thèse. **En divisant chacune des 104 séries temporelles GEBA d'éclairement mensuelles par la série temporelle d'éclairement TOA correspondante, 104 séries temporelles d'indice de clarté mensuel ont été calculées.**

3.5 Conclusion

Deux sélections, issues de la base de données GEBA, ont été présentées. La première est constituée de 63 stations fournissant des données mensuelles complètes sur plus de 30 ans. La seconde est constituée de 104 stations. Pour chaque station, une série temporelle de plus de 30 ans est disponible à l'échelle mensuelle, avec au maximum deux données manquantes consécutives. Cette dernière sélection de données constitue un échantillon de qualité, de long terme et de couverture globale. Pour satisfaire les contraintes futures d'une décomposition temps-fréquence, les séries temporelles de chaque station ont été complétées. Elles seront utilisées tout au long de la thèse : pour étudier la robustesse statistique des méthodes d'estimation du potentiel en chapitre 3, pour illustrer la diversité des comportements de long terme de l'éclairement en chapitre 4, pour étudier les variations inter-annuelles de l'éclairement à l'aide d'une décomposition temps-fréquence en chapitre 6.

4 Données satellitaires La base de données CLARA A2

L'éclairement calculé à partir d'images satellites est fréquemment utilisé pour estimer la ressource solaire d'un site (HABTE *et al.*, 2014; HOYER-KLICK *et al.*, 2009). Parmi l'ensemble des bases de données satellitaires de long terme et de couverture globale d'éclairement, cinq bases de données ont retenu notre attention. Les caractéristiques de ces bases de données sont résumées par le tableau 3.

Base de données	Début	Fin	Résolution spatiale	Référence associée
CERES EBAF-surface	2000	2018	1° x 1°	(KATO <i>et al.</i> , 2018),
ISCCP	1983	2017	1° x 1°	(ZHANG <i>et al.</i> , 2004)
GEWEX SRB	1983	2007	1° x 1°	(STACKHOUSE <i>et al.</i> , 2009)
Cloud_cci AVHRR-PM	1982	2016	0.5° x 0.5°	(STENGEL <i>et al.</i> , 2019)
CLARA-A2	1982	2015	0.25° x 0.25°	(KARLSSON <i>et al.</i> , 2017)

Tableau 3 : Résumé des caractéristiques de cinq bases de données satellitaires pertinentes pour notre étude.

Pour notre étude, nous nous sommes concentrés sur la base de données CLARA-A2. Celle-ci permet d'avoir *a priori* la meilleure résolution spatiale des cinq bases de données recensées, pour une disponibilité temporelle de plus de 30 ans.

4.1 Présentation

La base de données CLARA-A2 (*the CM SAF cCloud, Albedo and Radiation dataset from AVHRR data—edition 2*) est générée et distribuée par le CM-SAF (*EUMETSAT Climate Monitoring Satellite Application Facility*). Elle propose une actualisation du contenu de la base de données CLARA-A1 (KARLSSON *et al.*, 2013), grâce à des améliorations méthodologiques décrites en détail par Karlsson *et al.* (2017). **En particulier, une extension du jeu de données est réalisée, afin d'offrir, entre autres variables, des données d'éclairement global entre 1982 et 2015 sur l'ensemble du globe.** Les données de CLARA-A2 utilisées ici sont disponibles à l'échelle trihoraire et mensuelle via le site <http://navigator.eumetsat.int>. Les données d'éclairement sont estimées à partir des images acquises par les radiomètres-imageurs AVHRR (Advanced Very High Resolution Radiometer) embarqués sur les satellites météorologiques en orbites polaires NOAA et METOP. Plus précisément, elles sont dérivées à partir d'un abaque décrit par Mueller *et al.* (2009), utilisant des données auxiliaires de vapeur d'eau, d'albédo de surface et de contenu en aérosol. **Les données calculées sont fournies selon une grille régulière de 0.25° x 0.25° (approx. 30 km x 30 km à l'équateur).**

La précision des données mensuelles d'irradiation fournies par CLARA-A2 a été largement validée pour des périodes temporelles différentes et des climats différents, comme par exemple pour la Chine (Y. WANG et al., 2018), la Scandinavie (BABAR *et al.*, 2018), ou encore pour l'Europe (URRACA *et al.*, 2017). Sur l'ensemble des études, une déviation absolue moyenne (MAD) proche de 10 W/m^2 a été constatée pour des données mensuelles, pour un faible biais négatif de l'ordre de quelques W/m^2 . Cette base de données a été utilisée dans le cadre de nombreux articles sur les variations de long terme de l'éclairement (PFEIFROTH, BOJANOWSKI, *et al.*, 2018; PFEIFROTH, SANCHEZ-LORENZO, *et al.*, 2018).

Le calcul d'une valeur journalière d'éclairement est considéré comme valide s'il utilise au moins 20 données horaires valides. De même, pour le calcul d'une moyenne mensuelle utilisée ici, 20 données journalières valides sont requises.

4.2 Extraction d'un sous-jeu de données de CLARA-A2

La principale limitation de la base de données CLARA-A2 réside dans la forte présence de données manquantes dans les séries temporelles offertes. Comme souligné par Pfeiforth *et al.* (2018) et Babar *et al.* (2018), le principal facteur limitant la disponibilité des données est le faible nombre de radiomètres AVHRR (Advanced Very-High-Resolution Radiometer) : pendant la période entre 1982 et 1991, un seul instrument était en fonctionnement ; pendant la période 1992-2015, deux instruments AVHRR étaient en fonctionnement (TRENTMANN & KOTHE, 2016). Cela entraîne deux conséquences. Tout d'abord, les pannes ou les indisponibilités d'instruments sont fréquentes. Sur la période entre 1982 et 1991, la proportion de données manquantes peut aller jusque 50 %. Celles-ci sont souvent consécutives, ce qui rend difficile d'envisager une complétion des séries temporelles de qualité. Par conséquent, nous avons choisi de limiter la fenêtre temporelle de nos analyses préliminaires entre 1992 et 2015, soit 24 ans de données. Cependant, même sur la période de 1992 à 2015, de larges périodes de pannes coïncidentes des deux instruments peuvent être recensées. Deux périodes de pannes sont notables, entraînant ainsi deux périodes de données manquantes dans la base de données CLARA-A2 : de novembre 1994 à janvier 1995, et d'août à décembre 2000. Quelle que soit la série temporelle considérée, un minimum de 5 données manquantes consécutives sera donc observé entre 1992 et 2001.

La seconde conséquence du faible nombre de radiomètres AVHRR s'explique par le fait que les deux instruments disponibles sont situés sur deux satellites NOAA en orbite basse héliosynchrones. Ainsi, le nombre de passages des satellites en une journée est élevé autour des pôles, et faible sur la zone tropicale. Or, sachant que comme annoncé précédemment, une donnée journalière CLARA-A2 n'est considérée valide que si elle est calculée à partir d'au moins 20 données horaires, on comprend qu'un grand nombre de données manquantes vont être observées dans la zone tropicale.

La figure 10.a représente le nombre maximal de données manquantes consécutives sur la période 1992-2015, pour chaque pixel de la base de données CLARA-A2

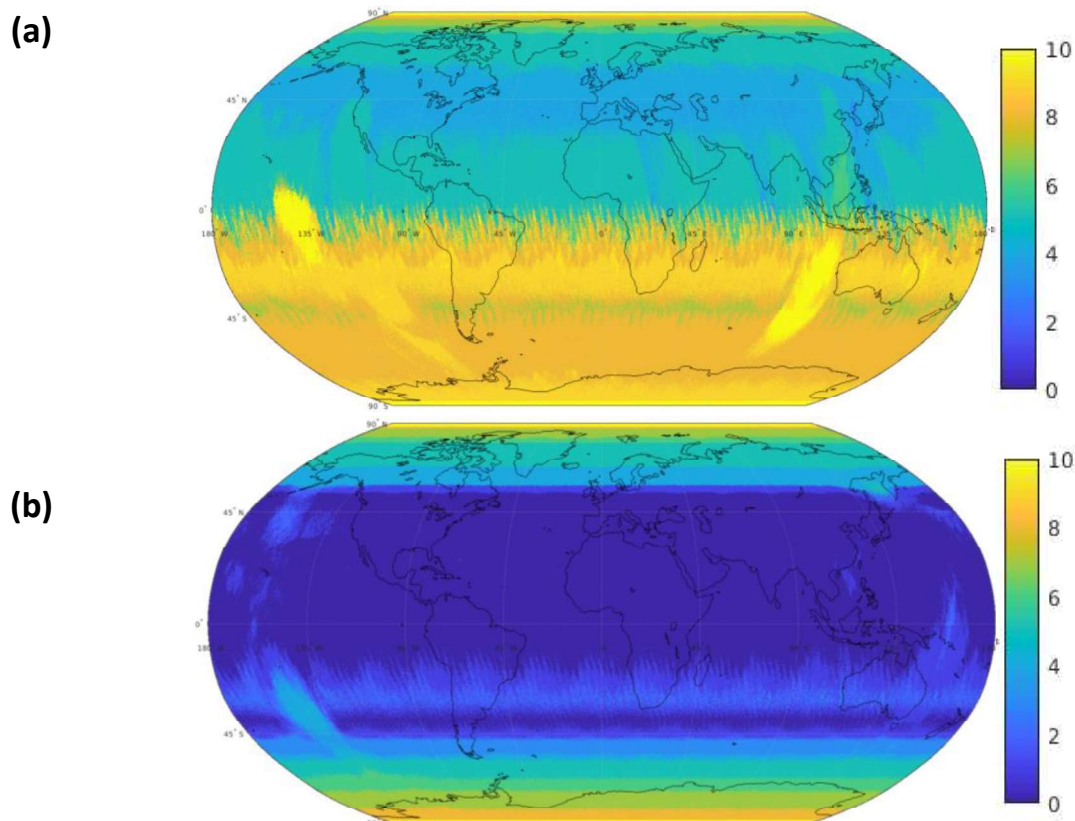


Figure 10 : Nombre maximum de données mensuelles consécutives manquantes pour chaque pixel de la base de données CLARA-A2. (a) entre 1992 et 2015, (b) entre 2001 et 2015

On observe sur la figure 10.a un grand nombre de données manquantes sur la période 1992-2015, allant jusque 10 mois consécutifs manquant au maximum. Au minimum, 5 données manquantes consécutives sont observées (bleu clair). Cette information est cohérente avec la période d'indisponibilité des deux radiomètres de 5 mois abordée précédemment. **Ce grand nombre de données manquantes nous a amené à limiter la fenêtre temporelle de l'ensemble de nos analyses à 15 années de données, entre 2001 et 2015.** La figure 10.b présente le nombre maximal de données consécutives manquantes pour chaque pixel, pour la période entre 2001 et 2015.

Pour cette dernière période, il est clair que l'essentiel du nombre de données consécutives manquantes se trouvent situés dans des latitudes supérieures à 55°. Cela s'explique par la qualité limitée des données fournies par l'algorithme dans les régions pour lequel l'albédo varie fortement dans le temps, dues aux périodes d'enneigement. En effet, l'algorithme est basé sur l'utilisation de cartes constantes d'albédo. Ainsi, les régions connaissant un couvert neigeux variable sont des régions très soumises aux données manquantes dans CLARA-A2.

Comme pour la base de données GEBA, les pixels présentant des séries temporelles possédant plus de deux données consécutives manquantes ont été exclus. On restreint donc globalement la zone d'étude entre 45° S et 45° N. **De plus, si plus de la moitié de la série temporelle est constituée de données manquantes - plus de 90 données manquantes pour la période entre 2001 et 2015 – ont également exclus.**

La figure 11 représente le nombre total de données manquantes dans les séries temporelles CLARA-A2. En jaune sont représentées les zones exclues de notre étude, c'est-à-dire les zones avec plus de deux données consécutives manquantes et/ou avec plus de 90 données manquantes.

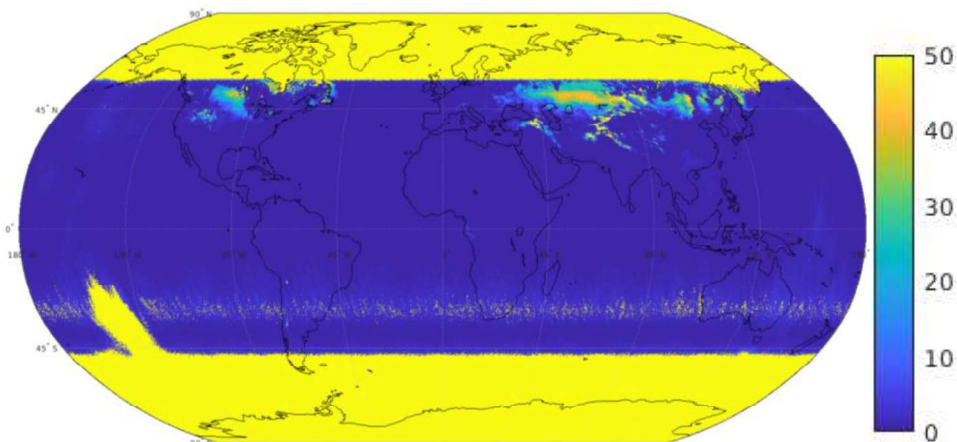


Figure 11 : Nombre de données mensuelles manquantes entre 2001 et 2015. En jaune sont représentées les zones exclues dans notre étude.

On constate que pour la grande majorité de la zone d'étude considérée, les séries temporelles possèdent moins de 10 données mensuelles d'éclairement manquantes, sur les 180 données que constitue la période comprise entre 2001 et 2015. Deux régions peuvent être identifiées avec un plus grand nombre de données manquantes : une zone au centre du Canada et une couvrant une partie importante de l'Eurasie : Sud de la Russie, Mongolie, Kazakhstan, Ukraine, une partie de la Turquie, etc. Ces deux régions possèdent entre 20 et 50 données manquantes. Cette information sera utile à considérer lors de l'analyse des résultats issus du sous-jeu de données CLARA-A2.

En définitive, le sous-jeu de données CLARA-A2 de 15 ans, sur la zone entre 45°S et 45°N, a été complété grâce à la même méthode d'interpolation que les données GEBA. Comme pour ces dernières, l'introduction de biais lors de la complétion des données se trouve limité par le faible nombre de données manquantes consécutives imposées. A titre illustratif, la figure 12 présente la moyenne multi-annuelle de l'éclairement pour chaque pixel de la base de données CLARA-A2

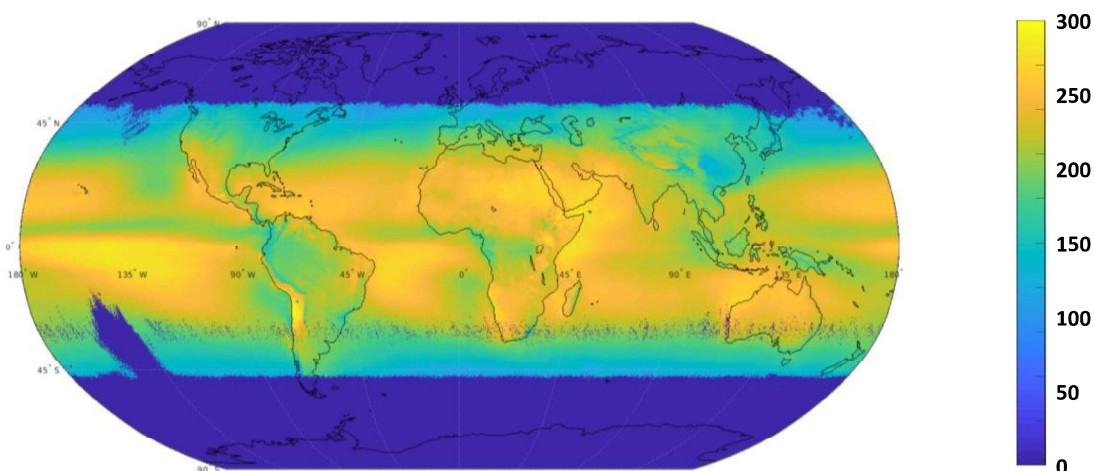


Figure 12 : Carte de la moyenne multi-annuelle de séries temporelles mensuelles d'éclairement fournies par la base de données CLARA-A2, en W/m^2 .

L'éclairement au sol varie fortement spatialement. On observe des structures caractéristiques de l'éclairement. Par exemple, un faible éclairement le long de l'équateur est observé. Cela est très probablement lié à l'ITCZ, comme supposé pour MERRA-2. Un fort éclairement est observé dans la zone tropicale, en particulier sur les bassins océaniques. Des structures de fortes intensités sont visibles en Afrique Saharienne, en Australie et au Sud de la Cordillère des Andes.

4.3 Calcul de l'indice de clarté

L'indice de clarté pour chaque point de la grille de la sélection de données CLARA-A2 a été calculé à partir des données d'éclairement TOA fournies par SG2. Plus précisément, pour une longitude donnée, des séries temporelles mensuelles de quatre années d'éclairement - trois années classiques et une année bissextile - ont été calculées pour l'ensemble des latitudes. Pour limiter le temps de calcul, les résultats calculés pour une longitude donnée ont été dupliqués sur l'ensemble de la grille. Les quatre années ont également été dupliquées pour constituer des séries temporelles de 23 années. Autrement dit, il est supposé que les variations de l'éclairement TOA supérieures à quatre ans sont négligeables. Cette dernière hypothèse peut être discutée au regard des cycles solaires de 11 ans aujourd'hui connus et documentés (FRÖHLICH & LEAN, 2004). L'erreur associée à cette dernière hypothèse est estimée à 0.3 % (FRÖHLICH & LEAN, 2004).

Les séries temporelles d'indice de clarté pour chaque point valide de la grille CLARA-A2 ont été calculées en faisant le ratio des données d'éclairement originales avec les données d'éclairement TOA ainsi estimées.

4.4 Conclusion

Les données CLARA-A2 ont pour avantage de proposer, théoriquement, des séries temporelles d'éclairement sur plus de 30 ans pour une couverture globale. Elles ont montré au cours de plusieurs études leurs pertinences dans l'analyse de variations de long terme de l'éclairement. Cependant, un grand nombre de données manquantes ont pu être répertoriées pour certaines périodes temporelles, comme pour certaines zones géographiques. Ces constatations nous ont amenés à construire un sous-jeu de données CLARA-A2, constitué de 15 années de données complétées. Ce sous-jeu de données ne sera analysé que sur les zones définies représentatives, soit les zones globalement comprises entre 45°S et 45 °N. Il sera utilisé de façon à étudier les variations inter-annuelles de l'éclairement à l'aide d'une décomposition temps-fréquence en chapitre 6.

5 Données de ré-analyses

La base de données MERRA-2

Les données de ré-analyses sont issues de modèles météorologiques, contraints de façon à ce que les sorties du modèle reproduisent le plus fidèlement possible ce qui a été observé. Elles ont l'avantage d'offrir une couverture globale, pour un historique de l'ordre de quelques dizaines d'années. En particulier, l'éclairement est calculé à partir d'un modèle de transfert radiatif utilisant un ensemble de variables d'état assimilés par le modèle, ne comprenant pas de mesures pyranométriques *in-situ*. Ainsi, la représentation de l'éclairement va dépendre de la représentation de l'ensemble des variables indirectement liés, utilisées pour son calcul. Les deux familles de ré-analyses les plus largement utilisées sont ERA et MERRA (Modern-Era Retrospective Analysis for Research and Applications), dont les deux dernières versions sont MERRA-2 (GELARO *et al.*, 2017) et ERA5 (HERSBACH, 2016). L'étude a ici été faite pour la base de données MERRA-2. Elle pourra dans le futur être reproduite pour la base de données ERA5.

5.1 Présentation

La base de données MERRA-2 (Modern-Era Retrospective Analysis for Research and Applications, version 2) est fournie par la NASA. Elle propose notamment des données d'éclairement au sol, en plus de multiples paramètres météorologiques, comme la vitesse et la direction du vent, la concentration en vapeur d'eau, etc. Elle possède naturellement beaucoup de caractéristiques communes avec sa version précédente, la base de données MERRA (RIENECKER *et al.*, 2011). Elle inclut cependant d'importantes mises à jour, détaillées par (GELARO *et al.*, 2017). En particulier, la prise en compte des aérosols a été nettement révisée. **MERRA-2 propose des données d'éclairement au sol selon une grille de 0.5° en latitude (approx. 55 km à l'équateur) et 0.625° en longitude (approx. 71.5 km à l'équateur). Les données mensuelles utilisées ici sont calculées à partir des moyennes horaires disponibles entre 1980 et aujourd'hui.**

Cette base de données étant relativement récente, peu de validations portant sur l'éclairement au sol ont été publiées au moment de la rédaction. Deux études notables ont pu être répertoriées : Bosilovich *et al.* (2015) et Krakauer & Cohan (2017). Bosilovich *et al.* (2015) indiquent que les performances de MERRA-2 varient en fonction de la région considérée. En moyenne, les auteurs soulignent que MERRA-2 a une tendance à la surestimation systématique de l'éclairement dans les tropiques, de l'ordre de 10 W/m², contre une sous-estimation de ce dernier dans les moyennes et hautes latitudes, également de l'ordre de 10 W/m². Krakauer & Cohan (2017) étudient quant à eux la variabilité inter-annuelle de MERRA-2 au regard de stations *in-situ*. Plus précisément, la variabilité inter-annuelle est étudiée à partir de données mensuelles mesurées pour 278 stations entre 1980 et 2016. Les auteurs concluent que l'ordre de grandeur fourni par MERRA-2 pour l'estimation de la variabilité inter-annuelle est correct, bien que globalement plus faible que celui des stations. Pour exemple, la figure 6 du chapitre 1 présente une carte de la moyenne multi-annuelle de séries temporelles mensuelles d'éclairement fournies par la base de données MERRA 2.

5.2 Calcul de l'indice de clarté

L'indice de clarté a été calculé de façon similaire pour MERRA-2 et CLARA-A2. Chaque série temporelle d'éclairement a été divisée par une série temporelle d'éclairement TOA issue de SG2.

5.3 Conclusion

La ré-analyse MERRA-2 propose des séries temporelles d'éclairement sur près de 40 ans avec une couverture globale. Bien que la qualité de cette base de données relativement récente reste à estimer de façon fine, elle semble adaptée à l'étude des variations de long terme de l'éclairement.

6 Données de modèles climatiques La base de données IPSL-CM6A-LR

Les données de modèles de circulation globale (global circulation model – GCM), ici appelés modèles climatiques, contiennent un historique de données d'éclairement de l'ordre de la centaine d'années, tout en conservant une couverture géographique globale. Les données d'éclairement sont calculées à partir de modèles numériques basés sur des équations physiques de transfert radiatif. Ces modèles sont couplés, entre autres, à des modèles de dynamique atmosphérique et océanique. Il existe actuellement plusieurs dizaines de modèles climatiques de référence, rassemblés dans le cadre des projets CMIP (Coupled Model Intercomparison Project) (MEEHL *et al.*, 2000). En effet, les projets CMIP ont pour objectif de coordonner à l'échelle mondiale la conception et la distribution des simulations des climats passés et futurs issus des modèles climatiques (EYRING *et al.*, 2016). Cela passe principalement par un exercice d'inter-comparaison d'un large ensemble de modèles, dont la dernière version est le projet CMIP6 (EYRING *et al.*, 2016).

Les données historiques fournies par CMIP6 permettent d'estimer la capacité des modèles à reproduire le climat, sa variabilité et les éventuelles tendances de long terme qui le caractérisent (IPCC, 2013). Elles sont forcées par des données d'observations telles que la variabilité du soleil, les aérosols volcaniques, ou encore les changements de gaz à effets de serre et d'aérosols causés par l'activité humaine (EYRING *et al.*, 2016). Nous nous sommes intéressés ici à l'étude de la variabilité et des éventuelles tendances de long terme de l'éclairement solaire, calculé par un modèle particulier : celui de l'Institut Pierre Simon Laplace (IPSL).

6.1 Présentation

La base de données IPSL-CM6A-LR est constituée de 31 réalisations décrivant l'éclairement global à la surface du globe pour des périodes allant de 1850 à 2015. Ces réalisations sont issues de la version du modèle de l'IPSL associée au projet CMIP6 : IPSL-CM6A. Cette version du modèle n'étant disponible que depuis fin 2018, elle n'a pas à ce jour fait l'objet d'une publication descriptive.

Cependant, l'ensemble des informations d'intérêt peuvent être trouvés sur <https://explore.es-doc.org/cmip6/models/ipsl/ipsl-cm6a-lr> et <http://forge.ipsl.jussieu.fr/igcmg/wiki/IPSL-CM6-articles>.

On retrouve sur ces pages que le modèle IPSL-CM6A a bien été conçu dans le but de représenter les processus majeurs impliqués dans le climat, d'évaluer la sensibilité climatique, et de contribuer à la compréhension de la variabilité climatique. Ainsi, dans le cadre de notre étude, cette base de données semble tout à fait indiquée. **Elle représente l'éclairement mensuel sur 79 niveaux verticaux, pour une résolution de 1.25° en latitude (approx. 140 km à l'équateur) et de 2.5° en longitude (approx. 280 km à l'équateur).**

Soulignons que l'exercice classique de comparaison entre mesures d'éclairement in-situ et données d'éclairement issu de modèle climatique peut être discuté. En effet, des outils de comparaisons classiques tels que le coefficient de corrélation ou le RMSE testent la coïncidence temporelle des deux séries temporelles. Or, les modèles climatiques ont pour objectifs de présenter une série temporelle plausible propre à une région donnée. Ainsi, il n'y a à priori aucune raison pour que la série plausible calculée soit coïncidente temporellement avec la série temporelle effectivement réalisée. Autrement dit, les sorties de modèles climatiques sont construites pour être utilisées de manière probabiliste : le modèle peut fournir un ensemble de réalisations plausibles de l'éclairement en un point, plutôt qu'une série temporelle suivant en tout instant la série temporelle observée. Ainsi l'ensemble de réalisations données par un modèle représente à l'échelle mensuelle, voire annuelle, l'ensemble des climats plausibles pour un point. La série temporelle mensuelle ou annuelle mesurée devrait donc, si l'éclairement est correctement représenté par le modèle, être comprise dans la gamme de possibles dégagés par l'ensemble des réalisations.

La base de données IPSL-CM6-LR étant relativement récente, peu de travaux ont été réalisés pour estimer la représentativité des données d'éclairement issues de ce modèle climatique, telle que définie précédemment. Il a été documenté pour la version précédente du modèle de l'IPSL utilisée dans CMIP5, une surestimation de l'éclairement au sol de l'ordre de 5 W/m^2 (e.g. Jerez *et al.*, 2015; LI, 2013; Wild, 2013 pour le cas de l'Europe), principalement causée par un déficit dans la représentation de la couverture nuageuse (BARTÓK *et al.*, 2017). Cependant, étant donné les changements substantiels faits dans le modèle entre les deux versions, en particulier sur la représentation de la couverture nuageuse, ces résultats ne peuvent pas être transposés simplement à cette nouvelle version du modèle.

De même que précédemment, la figure 13 présente la moyenne multi-annuelle de l'éclairement pour chaque pixel de la base de données IPSL-CM6A-LR. Cette carte illustrative est issue d'une réalisation particulière, sélectionnée parmi les 31 réalisations offertes par la base de données. La sélection de cette réalisation est détaillée au cours du chapitre 7.

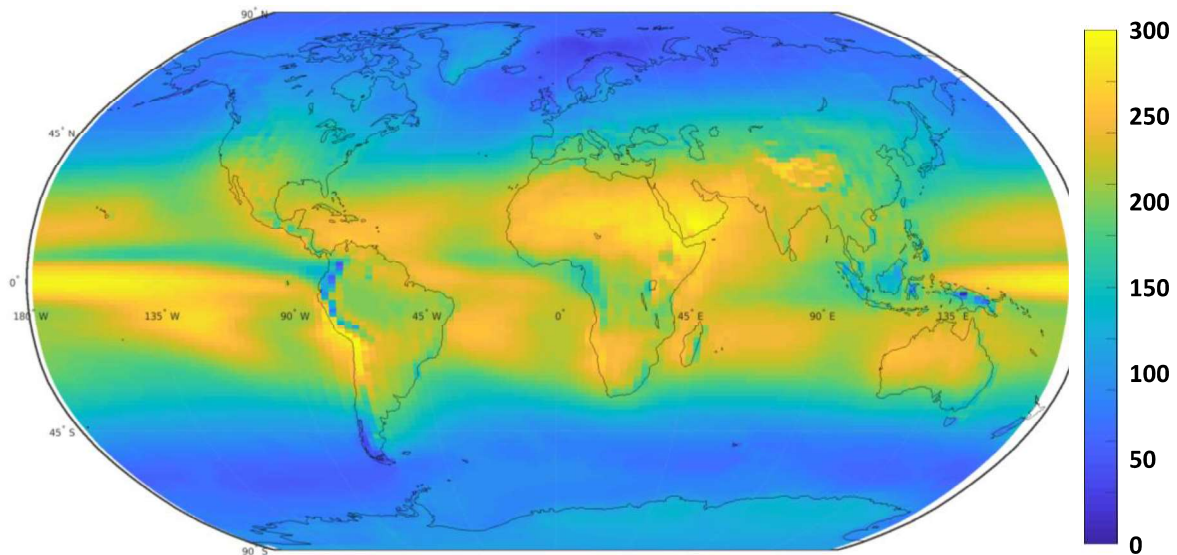


Figure 13 : Carte de la moyenne multi-annuelle de séries temporelles mensuelles d'éclairement fournies par la base de données IPSL-CM6A-LR.

L'éclairement au sol varie à nouveau spatialement. On retrouve assez schématiquement une distribution latitudinale du même ordre de grandeur que les bases de données précédentes : entre 0 et 300 W/m^2 . On retrouve la majorité des structures visibles sur CLARA-A2 et MERRA-2 : diminution de l'éclairement autour de l'ITCZ, fort éclairement dans la zone tropicale, structures de fortes intensités au Nord de l'Inde, en Australie et au Sud de la Cordillère des Andes.

6.2 Calcul de l'indice de clarté

Les données d'éclairement TOA sont fournies par l'IPSL sur l'ensemble de la grille. Chaque série temporelle d'éclairement, pour chaque pixel de chaque réalisation du modèle, a été divisée par la série temporelle d'éclairement TOA correspondante.

6.3 Conclusion

Nous utiliserons, dans le cadre de cette thèse, 31 réalisations historiques du modèle IPSL-CM6A-LR. Il est à noter que différentes réalisations sont également disponibles pour un horizon temporel allant de 2015 jusque 2100. Ces projections sont construites à partir de scénarios socio-économiques détaillant les évolutions possibles d'un grand nombre de paramètres. Bien qu'en perspective, la possibilité de prendre en compte la diversité des futurs possibles, autant en termes de climat, de démographie, de pollutions que d'habitudes de vie soit intéressante, le cadre de la thèse nous pousse à nous cantonner ici uniquement à l'utilisation de la partie historique des sorties de modèles climatiques. En cas de bonne représentation de l'éclairement et de ses variations par le modèle, utiliser ces réalisations pourrait, dans le futur, être une façon prometteuse de s'affranchir de la question de la représentativité du climat passé face à un futur par définition inédit.

7 Exemple de séries temporelles d'éclairement pour l'ensemble des bases de données

À titre illustratif, cette section propose une représentation des moyennes annuelles d'éclairement pour chacune des trois bases de données présentées. Les moyennes annuelles ne sont calculées que pour les années possédant 12 mois valides. Cela explique les éventuelles données manquantes dans les séries temporelles représentées, en particulier dans les bases de données GEBA ou CLARA-A2. La figure 14a présente les variations inter-annuelles de l'éclairement pour la station japonaise de Kochi pour la période 1975-2010, et la figure 14b pour la station japonaise Nara pour la période 1960-2010. Pour les bases de données grillées, le pixel correspondant à la station GEBA a été sélectionné. Aucune descente d'échelle de la maille au point n'a été réalisée.

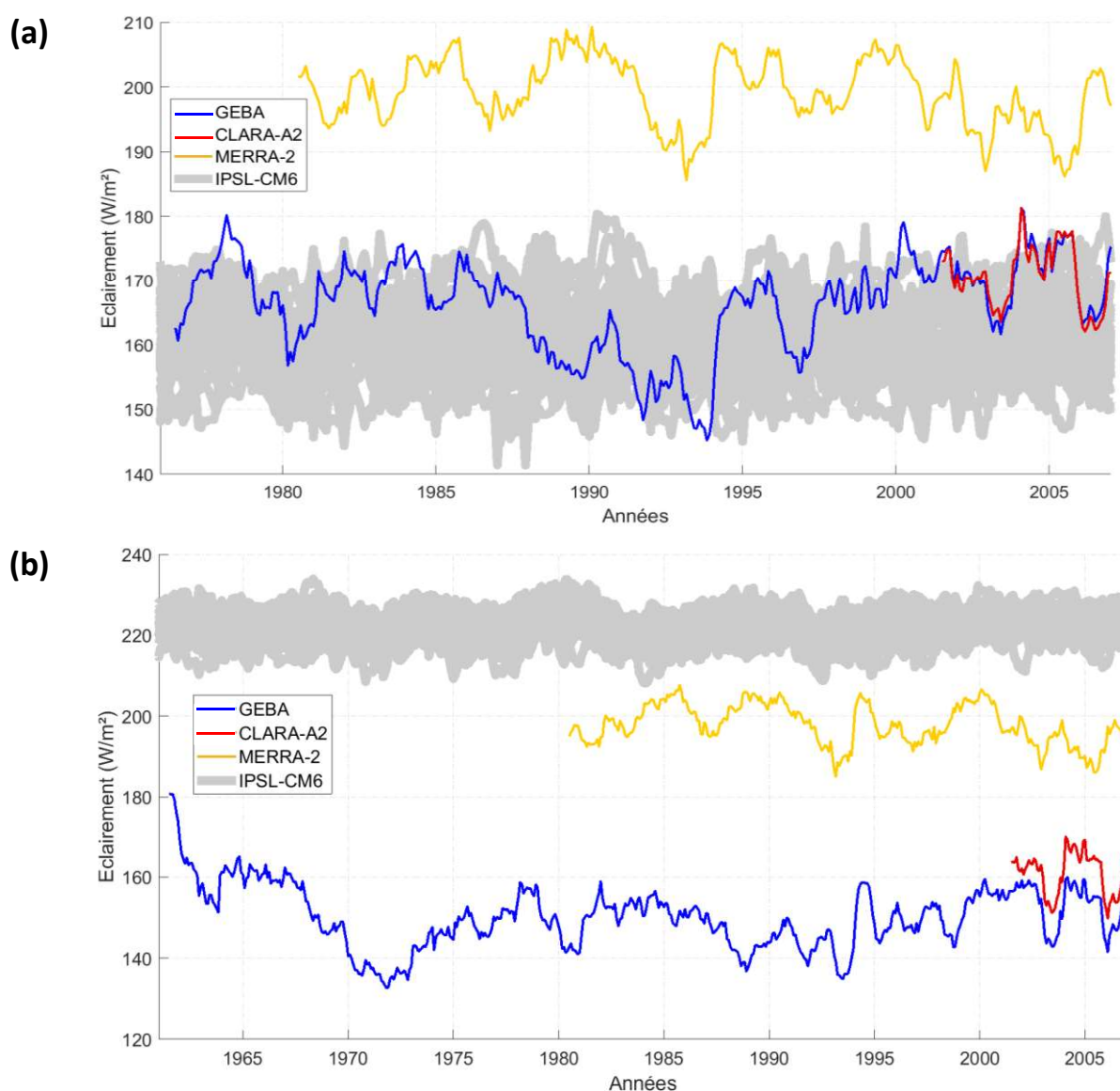


Figure 14 : Séries temporelles des moyennes annuelles glissantes d'éclairement pour les stations de Kochi (a) et de Nara (b). En couleurs sont représentées les bases de données considérées : la base de données GEBA (bleu), CLARA-A2 (rouge), MERRA-2 (jaune) et IPSL-CM6-LR (gris).

7. Exemple de séries temporelles d'éclairement pour l'ensemble des bases de données

Les deux stations présentées ont été sélectionnées car elles représentent deux cas illustratifs différents : un cas pour lequel les données issues des différentes bases de données sont en accord, et un cas où différents biais caractéristiques, parmi ceux qui ont été observés sur l'ensemble des stations, sont représentés. Différentes variations inter-annuelles peuvent être observées grâce aux mesures d'éclairement GEBA. Dans le premier cas, l'éclairement semble diminuer entre 1980 et 1995, avant de remonter jusque approximativement dans les années 2000, puis de se stabiliser. À l'inverse, l'éclairement pour la station Nara est relativement constant dans le temps après 1970. Entre 1960 et 1970, une forte diminution de l'intensité de l'éclairement est constatée, de l'ordre de 40 W/m^2 . Malgré la relative proximité géographique de ces deux stations, des comportements différents de l'éclairement annuel peuvent donc être constatés.

En ce qui concerne les données satellitales, les données CLARA-A2 ne couvrent qu'une partie de la série temporelle, n'étant disponibles qu'à partir de 2001. Grâce aux quelques années de comparaison disponible, il est possible d'avancer que celles-ci sont très proches des données mesurées pour les deux stations. Ainsi, au vu de la similarité des données *in-situ* et satellitales sur cette période, on comprend pourquoi ces dernières peuvent servir de données de référence pour le calcul du potentiel solaire. On peut néanmoins noter une tendance à la surestimation de CLARA-A2 pour la station Nara, de l'ordre de quelques W/m^2 . Ce biais, bien qu'observé sur quelques stations, est certainement dû aux particularités locales des stations considérées, étant donné que la bibliographie annonce plutôt un biais négatif de CLARA-A2 de quelques W/m^2 .

En ce qui concerne la base de données de ré-analyse, MERRA-2 a tendance à nettement surestimer le rayonnement pour les deux stations, de l'ordre de 15 à 30 %. Cette tendance se retrouve sur un grand nombre de stations, bien que des sous-estimations de l'éclairement puissent également être observées pour quelques stations.

Comme annoncé précédemment, les sorties de modèles climatiques sont considérées de façon ensembliste. Ainsi, les 31 sorties de modèles sont représentées conjointement, donnant une gamme des possibles dans laquelle la série temporelle mesurée devrait s'inscrire. C'est le cas pour la station Kochi, pour laquelle la série temporelle GEBA (bleu) est globalement comprise dans la zone grisée. Cependant, pour cette station, les modèles ne semblent pas reproduire la diminution d'éclairement mesurée entre 1980 et 1995. À l'opposé, pour la station Nara, l'ensemble des modèles est au-dessus de la courbe des données *in-situ*. Cela traduit une surestimation de l'ordre de 70 W/m^2 du rayonnement au sol calculé par les modèles. Cette surestimation est retrouvée pour une dizaine de stations GEBA. À l'inverse, il n'est pas impossible de trouver des stations pour lesquels les modèles climatiques sous-estiment le rayonnement observé. Ce résultat n'est pas étonnant, étant donnée notamment la taille du pixel considéré : la différence de résolution spatiale entre les modèles climatiques et les mesures ponctuelles peuvent largement expliquer ces écarts. En effet, le rayonnement dépend directement notablement de la couverture nuageuse, dont la dynamique se produit en grande partie à une échelle nettement inférieure à la maille du modèle.

Les observations faites pour ces deux stations ne peuvent absolument pas être généralisées. Il est évident qu'une analyse visuelle entre une station particulière et un pixel, dont la taille peut faire plus de 100 km, ne pourrait pas être représentative de la qualité d'une base de données. Ces exemples permettent néanmoins de se familiariser avec les bases de données introduites au cours de ce chapitre.

8 Conclusion

Quatre bases de données ont été détaillées dans ce chapitre : une base de données de mesures, une base de données satellitale, une ré-analyse et une base de données regroupant un ensemble de réalisations issues d'un modèle climatique. Les bases de données sélectionnées sont respectivement GEBA, CLARA-A2, MERRA-2 et IPSL-CM6A-LR. Les méthodes de calcul de l'indice de clarté ont été explicitées. La gestion des données manquantes dans GEBA et CLARA-A2 a également été détaillée, aboutissant à la sélection de sous-jeux de données pour chacune des deux bases de données.

Les caractéristiques de chaque base de données sont résumées succinctement dans le tableau 4.

Base de données		Début	Fin	Résolution temporelle	Résolution spatiale (km x km)	Disponible via
Mesures	GEBA	~1950	2005	Mois	Point	ETH Zurich
Satellites	CLARA-A2	1982 2001 ici	2015	30 min	5 x 5	EUMETSAT CM-SAF
Ré-analyses	MERRA-2	1980	Présent 2015 ici	Heure	55 x 71.5	NASA
Modèles climatiques	IPSL- CMIP6-LR	1850	2015	Jours	140 x 280	IPSL

Tableau 4 : Résumé des caractéristiques de chaque base de données introduite au cours du chapitre.

En conclusion, les méthodes de calcul de l'éclairement utilisées dans chaque base de données sont intrinsèquement différentes, ce qui en rend l'étude conjointe intéressante.

Chapitre 3

Limitations des méthodes d'estimation du potentiel solaire

Table des matières

Limitations des méthodes d'estimation du potentiel solaire	53
1 Introduction	55
2 Analyse d'un exemple d'indicateur : le P90.....	55
2.1 L'indicateur P90 : définition	55
2.2 Présentation des méthodes de calcul de l'indicateur P90	57
2.3 Evaluation des méthodes de calcul de l'indicateur P90.....	58
2.4 Discussion sur l'utilisation du P90	66
3 Génération de séries temporelles d'éclairement : analyse bibliographique	67
3.1 Méthodologie de l'analyse bibliographique.....	67
3.2 Méthode de Fernández-Peruchena <i>et al.</i> (2015).....	69
3.3 Méthode de Larañeta <i>et al.</i> (2018)	72
3.4 Résumé et conclusions de l'analyse bibliographique	Erreur ! Signet non défini.
4 Conclusion.....	78

1 Introduction

Notre étude s'insère dans un contexte d'analyse de risques financiers de centrales solaires photovoltaïques. L'estimation du potentiel solaire d'un site est un paramètre important à prendre en compte lors de cette analyse de risques. Cette prise en compte est usuellement faite par une analyse statistique de l'historique de l'éclairement du site choisi. Cette analyse découle sur un ensemble d'indicateurs statistiques, utilisés comme supports d'aide à la décision.

L'objet de ce chapitre est d'explorer les méthodes usuelles d'estimation du potentiel solaire. En premier lieu, nous étudierons l'utilisation du P90, un indicateur statistique très utilisé par les développeurs de projets photovoltaïques (section 2). Cette étude est construite autour d'expérimentations simples, numériques ou réalisées à partir de données de mesures de références. Ensuite, nous discuterons de la validité de méthodes de génération de séries temporelles d'éclairement utilisées pour l'estimation du potentiel solaire (section 3). En particulier, quatre hypothèses sous-jacentes, qui conditionnent la validité des résultats des méthodes, ont été identifiées : (i) la robustesse des résultats issus de calculs réalisés sur des petits échantillons, (ii) l'indépendance des données d'éclairement à l'échelle mensuelle ou annuelle, (iii) la stationnarité temporelle de l'éclairement, et enfin (iv) la capacité de représenter de façon robuste l'échantillon de données considéré par une distribution paramétrique.

2 Analyse d'un exemple d'indicateur : le P90

Parmi l'ensemble des indicateurs statistiques utilisés dans le cadre de l'analyse de risques de projet solaire, l'un d'eux a retenu notre attention : le P90.

2.1 L'indicateur P90 : définition

Le P90 correspond à la valeur associée à une probabilité d'excédence définie : il s'agit de la valeur annuelle d'éclairement pour laquelle 10 % des valeurs de la série temporelle totale sont inférieures. Autrement dit, pour une variable aléatoire x , la valeur du P90 est définie par :

$$P90 = X_{p90} \mid P(x < X_{p90}) = 0.10$$

Cet indicateur est très utilisé en pratique, de façon conjointe avec le P50, le P75, voire le P99 (FitchRatings, 2018; McMAHAN *et al.*, 2013a). Pour une distribution symétrique, le P50 (*i.e.* la médiane) correspond aussi à la moyenne des valeurs considérées. Il est couramment utilisé pour la mise en place de scénario de production défavorable – scénario « pire cas » - permettant de garantir, pour un certain niveau de confiance, la viabilité du projet photovoltaïque vis-à-vis de la ressource solaire.

La **Erreur ! Source du renvoi introuvable.** représente schématiquement le P50 et le P90 pour le cas d'une population suivant une loi normale de moyenne nulle et d'écart type égale à 0.2. Dans cet exemple, le P50 représente bien la moyenne des valeurs.

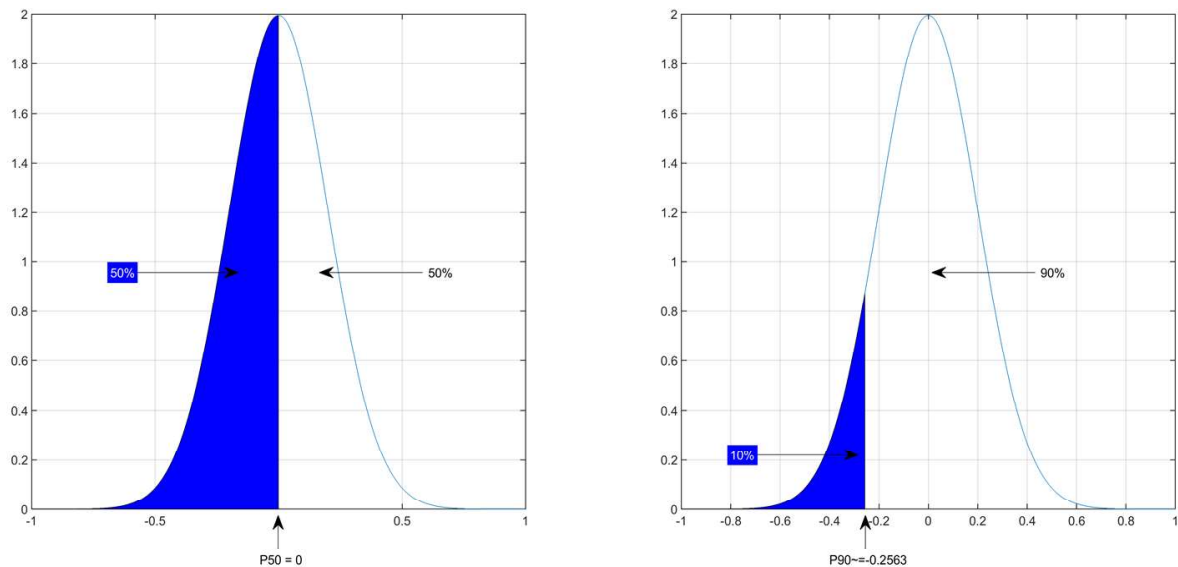


Figure 15 : Représentation du P50 (gauche) et du P90 (droite) pour une distribution gaussienne de moyenne nulle et d'écart type égal à 0.2.

Il est possible de trouver dans la bibliographie, sous l'appellation de P90, des indicateurs de probabilité d'excédence évalués à partir de données agrégées sur des périodes multi-annuelles différentes (CEBECAUER & SURI, 2015; ŞEN, 2017). Le P90 peut ainsi correspondre au dixième percentile soit de l'éclairement moyen annuel, soit de la moyenne de cet éclairement sur l'ensemble de la durée de vie de la centrale, soit des moyennes glissantes de plusieurs années. Il est clair que ces indicateurs P90 auront des significations très différentes les uns des autres, compte tenu de l'effet de lissage *a priori* engendré par l'agrégation multi-annuelle.

Dans notre cas, nous appellerons P90 l'indicateur calculé à partir d'éclairements moyens annuels. Cette définition est utilisée par la majorité des utilisateurs (McMAHAN *et al.*, 2013b). En effet, après lancement du projet, il est d'usage de comparer la ressource mesurée chaque année sur site au P90 estimé *a priori*. Cela permet de mettre en évidence au fil de l'eau la survenue d'anomalies de l'éclairement.

Il n'existe à l'heure actuelle aucun consensus ou standard industriel décrivant une méthode de calcul des probabilités d'excédence (McMAHAN *et al.*, 2013b). Ainsi, de nombreuses méthodes d'estimation du P90 – souvent à partir de l'estimation de quantiles - existent dans la bibliographie (*e.g.* DIELMAN *et al.*, 1994; McMAHAN *et al.*, 2013b; WEI *et al.*, 2015; ZIELINSKI, 2004). Trois principales familles de méthodes peuvent être dégagées : les méthodes empiriques, s'appuyant sur la construction et l'analyse de fonctions de distributions cumulatives empiriques ; les méthodes dites par ajustement de distributions paramétriques, s'appuyant sur les propriétés de distribution associées aux données originales ; et les méthodes d'estimation par noyau, s'appuyant sur l'utilisation de noyaux divers comme la distribution gaussienne ou celle d'Epanechnikov.

Il est à noter que la première famille de méthodes peut être vue comme un cas particulier de la troisième famille : il s'agit d'une estimation par noyau, avec comme noyau une fonction porte. En ce sens, bien que cela ne soit pas représentatif de l'ensemble des méthodes d'estimation par noyau, nous sommes contents de présenter, puis d'illustrer, les possibilités et les limites des deux premières familles de méthodes. Pour plus de détails, nous recommandons de consulter par exemple Cheng & Parzen (1997) et Sheather & Marron (2012).

2.2 Présentation des méthodes de calcul de l'indicateur P90

Nous nous attacherons à l'étude des sources de variation et d'incertitude associées au calcul des deux dernières familles de méthodes de calcul du P90. En ce sens, nous avons choisi de procéder en deux étapes. Dans un premier temps, nous détaillerons les méthodes empiriques et par ajustement de distributions paramétriques de calcul du P90. Dans un second temps, nous illustrerons par des expérimentations simples les limites et les possibilités du P90 calculé à partir de ces deux méthodes. Ces dernières seront numériques, ou utiliseront des données de mesure de référence.

2.2.1 Méthode empirique

Pour bien appréhender la méthode empirique, prenons l'exemple de jeux de données restreints. **Pour un échantillon de 10 valeurs, toute valeur comprise entre le premier et le deuxième minimum peut être choisie.** Comme nous nous intéresserons à un cas extrême, nous choisirons ici la valeur minimale de cet intervalle, c'est-à-dire la valeur minimale de la série. On voit bien que 90 % des valeurs se trouvent au-dessus de ce minimum. **Pour un échantillon de 20 valeurs, toute valeur comprise entre le deuxième et troisième minimum peut être choisie. On retiendra pour le P90 le second minimum. Ainsi, pour une série temporelle d'éclairement annuel de 20 ans, il suffira de calculer le second minimum pour estimer la valeur du P90.**

D'un point de vue numérique, il est possible de trouver des variantes à notre approche pour estimer le P90 à partir d'échantillons de petites tailles (ZIELINSKI, 2004). Le logiciel Matlab calcule par exemple le P90 pour un échantillon de 10 données en considérant la moyenne entre les deux plus petites valeurs, de 20 données en moyennant la seconde et la troisième plus petite valeur, etc. Pour l'estimation empirique du P90, nous privilégierons dans la suite la première méthode d'estimation.

Pour un échantillon de 30 valeurs, la majorité des méthodes calcule le P90 à partir de la fonction de distribution cumulative des données d'origines. Le P90 correspond alors à la valeur - ou à l'interpolation linéaire entre les deux valeurs les plus proches - associée à une densité de 0.1.

2.2.2 Méthode par ajustement de distributions paramétriques

Une seconde méthode courante d'estimation du P90 repose sur l'ajustement d'une distribution paramétrique à l'échantillon initial de données. Un exemple de description et d'application sur 13 sites de cette méthode est proposé par Fernandèz-Peruchena *et al.* (2016). Elle se décompose en quatre temps :

1. **La détermination du modèle de distribution associé au jeu de données.** Pour cela, les auteurs testent l'adéquation entre un ensemble de distributions paramétriques et les données historiques à l'aide de différents tests statistiques d'ajustement. Ils concluent sur le fait que l'éclairement annuel peut être représenté par une loi normale.
2. **L'estimation des paramètres propres à la distribution choisie.** Dans cette étude, ils déterminent les paramètres – moyenne et écart type – de la distribution gaussienne pour chaque station.
3. **Le calcul des percentiles d'intérêt.** Le P90 est notamment estimés théoriquement à partir des propriétés de la distribution normale, en utilisant la fonction d'erreur inverse.
4. **Le calcul de l'incertitude sur ce percentile.** L'incertitude est calculée théoriquement à partir des paramètres de la distribution normale.

Sous couvert de la validité de l'hypothèse d'une distribution gaussienne sous-jacente, cette méthode permet une estimation plus fine de l'indicateur P90 qu'avec une méthode empirique. De plus, elle permet de calculer l'incertitude associée à l'estimation du P90. Ce dernier point est d'un intérêt certain pour une bonne compréhension et une bonne utilisation de l'indicateur dans un contexte d'aide à la décision.

2.3 Evaluation des méthodes de calcul de l'indicateur P90

Malgré l'utilisation courante des méthodes présentées précédemment, celles-ci peuvent être discutées. Cette section a pour but de mettre en lumière les possibilités et les limites de l'utilisation actuelle de l'indicateur P90, autant en terme scientifique que d'interprétation.

Pour cela, un ensemble d'expérimentations statistiques a été réalisé. **Ces expérimentations n'ont pas pour but d'être réalistes : le but ici est de simplifier le problème, pour mettre en évidence la robustesse et la significativité statistique effective de l'analyse réalisée.** Si l'estimation de l'indicateur P90 n'est pas robuste dans un cas théorique simplifié, il semble discutable de le considérer robuste dans une réalité plus complexe à modéliser. Notons que la démarche proposée dans cette section est purement statistique, et ne dépend donc pas de la quantité considérée. Ainsi, les résultats issus d'expérimentations ayant recours à des données annuelles d'éclairement resteront tout à fait valables pour le P90 calculé à partir de valeurs de productibles.

2.3.1 Méthode empirique

Pour le cas d'un P90 calculé à partir d'un échantillon de 10 données, les limitations du calcul du P90 apparaissent rapidement. Comme annoncé, avec si peu de données historiques, le scénario « pire cas » est alors équivalent à l'année ayant connu l'éclairement le plus bas. Ce minimum pourrait tout à fait être une valeur aberrante, que ce soit pour des problèmes de mesure ou par la survenue d'un événement météorologique extrême, de type éruption volcanique par exemple.

Prenons le cas de l'estimation du P90 à partir de 20 valeurs pour la suite de cette section. Réalisons une expérimentation numérique simple, dans un cas simplifié à l'extrême : un site pour lequel l'éclairement mesuré sur une année est indépendant des années précédentes ou futures, où l'ensemble des valeurs mesurées peuvent être correctement représentées par une distribution gaussienne, et enfin où les données futures suivent la même loi de probabilité que les données historiques. **Autrement dit, nous considérons les valeurs indépendantes et identiquement distribuées (noté i.i.d. par la suite), donc stationnaires temporellement, auxquelles il peut être ajusté une estimation paramétrique gaussienne (moyenne et écart type).**

L'expérimentation proposée est composée de six étapes :

1. Un tirage aléatoire de 20 valeurs suivant une distribution gaussienne de moyenne nulle et d'écart type 1 est réalisé. Ce tirage représente les 20 valeurs d'éclairement annuel de l'échantillon historique. Les données considérées sont donc bien issues d'un tirage indépendant dans une distribution normale ;
2. Le P90 est calculé sur ces 20 valeurs : il s'agit du second minimum ;
3. Un second tirage aléatoire de 20 valeurs suivant une distribution gaussienne de mêmes paramètres que la première distribution est réalisé. Ce tirage représente les 20 années futures d'éclairement sur la durée de vie de la centrale. L'éclairement est donc bien supposé identiquement distribué, et donc stationnaire temporellement ;
4. On étudie la validité du P90 calculé. Pour cela, on compte le nombre de valeurs du second tirage inférieures au P90 calculé à partir de la première série.
5. L'expérience est reproduite ici 100 000 fois, pour assurer une robustesse statistique suffisante aux conclusions de l'étude.
6. Un histogramme du nombre d'années excédant le P90 est réalisé.

Les résultats de cette expérience numérique sont présentés en figure 16. **Si l'estimation du P90 est correcte, le nombre de valeurs issu du second tirage inférieures au P90 devrait être de deux.** Autrement dit, il est attendu que durant les 20 ans suivant le lancement du projet, deux années aient un éclairement annuel moyen inférieures au P90. Si moins de 2 valeurs sont inférieures au P90, alors le P90 a été sous-estimé ; si plus de 2 valeurs sont inférieures au P90, alors le P90 a été surestimé. Ce critère est largement utilisé lors de l'analyse de performance de centrales solaires photovoltaïques : chaque année, le productible mesuré sur le site est comparé au P90 estimé *a priori*, lors de l'analyse de risques financier du projet. À partir de deux années ayant connu une production inférieure au P90, les banques peuvent être amenées à demander des analyses complémentaires, voire des provisions financières, à l'opérateur de la centrale solaire photovoltaïque. Ainsi, cette analyse relève d'un aspect important de l'exploitation d'une centrale solaire : l'utilisation du P90 sous la forme de seuil lors d'un comptage.

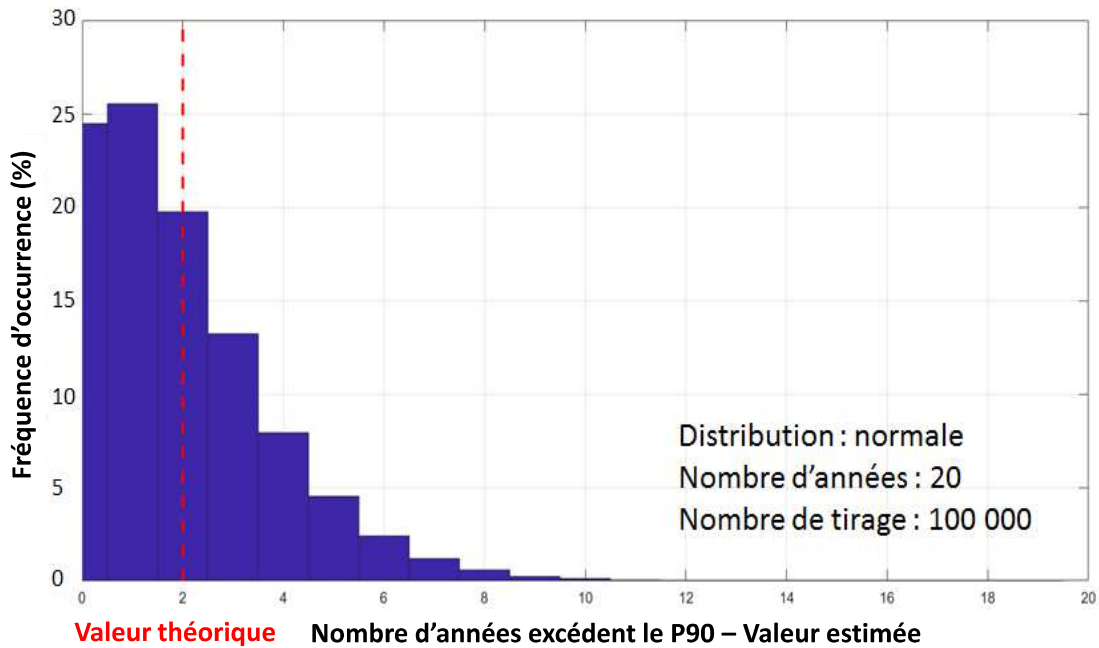


Figure 16 : Histogramme du nombre de valeurs du 2nd tirage inférieures au P90 calculé à partir du 1^{er} tirage.

Sous hypothèse gaussienne et i.i.d, on voit que sur les 100 000 tirages, uniquement 20 % des cas se trouvent satisfaire la compréhension usuelle du P90. En d'autres termes, pour plus de 80 % du temps, l'estimation du P90 n'est pas conforme à l'attente des utilisateurs, et le potentiel solaire sera mal représenté par l'analyse statistique de la série de données historique. Pour 50 % des cas, le P90 est sous-estimé. Par analogie, il y a plus de production que prévu à partir de l'analyse de l'échantillon historique. Dans une logique conservatrice de l'estimation du potentiel solaire visant à garantir la viabilité du projet, cela pourrait ne pas sembler inquiétant, voire être positif. Cependant, une sous-estimation du potentiel solaire pour 50 % des cas peut freiner l'investissement dans le domaine du solaire photovoltaïque. En effet, le développement de projets photovoltaïque peut alors apparaître comme une solution non viable sur un site. À l'opposé, dans 30 % des cas, le P90 est surestimé. Cela signifie que plus de 2 années de production seront inférieures au P90. À l'extrême, il n'est pas impossible d'avoir 8 ou 9 années sur les 20 années suivant la construction de la centrale qui soient inférieures au P90 annoncé. Cela peut bien évidemment éveiller les craintes des financeurs.

Ces résultats soulignent les limites à la fois de l'estimation empirique du P90 avec un nombre aussi restreint de données, à la fois et les limites dues à l'utilisation du P90 dans un processus de comptage. Ces limites peuvent être interprétées comme provenant autant du faible nombre de données dans le premier tirage, représentant l'échantillon de données historiques, que du faible nombre de données dans le second tirage, représentant l'échantillon de données futures. On peut alors s'interroger sur l'incertitude associée à l'estimation du P90 si le P90 était calculé à partir de milliers de données historiques. **Autrement dit, dans un cas idéal où le P90 est connu parfaitement, sans aucune incertitude, quelle serait alors l'incertitude associée au faible nombre de données futures ?**

Il est possible de répondre à cette question en retirant les étapes 1 et 2 de la méthodologie présentée précédemment. Autrement dit, au lieu d'estimer le P90 à partir du tirage aléatoire de 20

données, on considère une valeur théorique du P90. Les étapes 4 à 6 restent les mêmes : on compte le nombre de valeurs du second tirage inférieur à ce P90 théorique pour un grand nombre de cas. On s'affranchit ainsi des incertitudes sur les paramètres de la distribution, et donc de l'estimation du P90. L'histogramme des résultats du comptage est présenté en figure 17. En bleu ciel est représentée la probabilité théorique d'obtenir une valeur inférieure au P90 en fonction du nombre d'excédences.

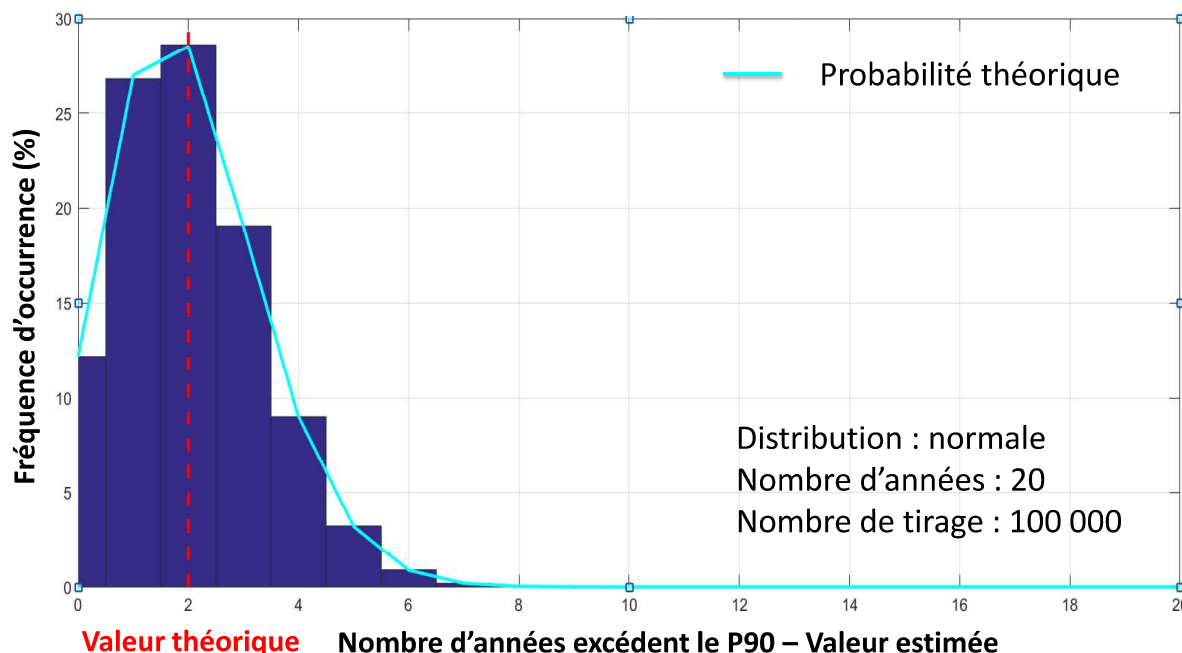


Figure 17 Histogramme du nombre de valeurs du second tirage inférieures au P90 théorique.

On constate des différences notables entre la figure 16 et la figure 17. En particulier, le nombre de cas satisfaisant la définition du P90 passe de 20 % à près de 30 %. Le nombre de cas pour lesquels le P90 est sous-estimé passe quant à lui de 50 % à 40 %. On voit donc que bien que l'estimation du P90 soit plus fiable dans ce second cas, elle reste loin de coïncider avec l'estimation attendue par les utilisateurs. **Ce point est central, car il montre que même dans un cas idéal, où le P90 est parfaitement connu, l'utilisation du P90 dans un processus de comptage serait tout de même soumise à l'incertitude propre au nombre d'années futures limité par la durée de vie de la centrale.**

Bien que ces résultats remettent en cause la pertinence de l'indicateur P90 avec si peu de données – historiques ou futures –, ils ne sont pas surprenants. En effet, ceux-ci sont cohérents avec les résultats théoriques issus d'une analyse mathématique du problème. Dans un cas i.i.d., une expérience de tirage aléatoire peut être décrite par une loi binomiale. Dans notre cas, on cherche à connaître pour 20 tirages, quelle est la probabilité que la valeur tirée x soit inférieure à la valeur idéale du P90 notée X_{p90} . En notant k le nombre d'excédences, soit le nombre de tirages pour lequel la valeur tirée est inférieure à celle du P90, on cherche :

$$P(k) = \binom{20}{k} P(x < X_{p90})^k (1 - P(x < X_{p90}))^{20-k}$$

Or, par définition, si la valeur X_{p90} est correcte, $P(x < X_{p90}) = 0.1$. D'où on peut écrire :

$$P(k) = \binom{20}{k} 0.1^k (0.9)^{20-k}$$

La courbe représentée sur la figure 17 en bleu ciel suit cette équation. On voit que cette courbe est très proche de l'histogramme des valeurs trouvées lors de l'expérimentation numérique. Pour le cas idéal, soit pour $k=2$, on retrouve bien 28,5 % de chance que le P90 corresponde à l'indicateur attendu. **Il est intéressant de souligner que ce résultat est indépendant de la loi de distribution sous-jacente.** En effet, que le tirage aléatoire de données soit effectué dans loi de distribution gaussienne ou une autre loi de probabilité, le phénomène de comptage sera toujours mathématiquement décrit par une loi binomiale, pourvu que l'on soit bien dans le cas i.i.d.

En conclusion, l'incertitude associée à l'estimation du P90 par une méthode empirique, et à l'utilisation de celui-ci dans un processus de comptage, est très grande. Même pour un cas idéal pour lequel le P90 est parfaitement connu, les incertitudes restent très grandes, de par la relativement faible durée de vie de la centrale. Ainsi, il a été conclu que l'incertitude propre au phénomène de comptage est indépendante de l'incertitude d'estimation du P90, tout comme de la distribution des données historiques.

À l'issu de cette expérimentation, deux questions restent à approfondir : l'incertitude propre à l'estimation du P90, et la détermination d'un ajustement de distribution paramétrique aux données d'éclairement. En effet, nous avons ici pris l'exemple d'un éclairage suivant une distribution gaussienne, mais il est possible d'imaginer d'autres distributions représentant les données d'éclairement.

2.3.2 Méthode par ajustement de distributions paramétriques

La distribution normale est régulièrement associée aux données d'éclairement. Cet ajustement de distribution paramétrique permet de calculer théoriquement la valeur et l'incertitude du P90, telle que cela l'a été illustré en section 2.2.2 grâce à la méthode de Fernandèz-Peruchena (2016). Ce calcul étant théorique, il n'est pas soumis à l'incertitude propre au faible nombre de données considéré. Ainsi, ajuster une distribution normale aux données historiques peut être vu comme une façon de contourner, au moins en partie, les limitations propres à l'utilisation d'un petit échantillon.

Cependant, il est possible de s'interroger sur l'incertitude associée à l'ajustement de distribution paramétrique à un échantillon de petite taille. En effet, si la distribution normale représente de façon significative les données historiques, il ne peut être *a priori* exclu que d'autres lois de distributions représentent de façon tout aussi significative ces mêmes données. **Autrement dit, est-il possible d'ajuster à une série temporelle plusieurs modèles de distribution théorique avec la même significativité ? Dans ce cas, est-ce que les différents modèles conduisent à de grandes différences dans l'estimation du P90 ?**

Pour répondre à cette question, nous avons choisi de réaliser à nouveau une expérimentation simple et illustrative. Nous utiliserons comme support les données d'éclairement annuelles de la sélection de 104 stations GEBA, présentées en section 2 du chapitre 2. Plus précisément, l'expérimentation proposée est composée de deux étapes.

Dans un premier temps, trois lois de probabilité classiques sont ajustées à chacune des 104 séries temporelles annuelles GEBA. Les lois de probabilité choisies sont la loi gaussienne, la loi de Weibull et la loi de Gumbel. La loi gaussienne est totalement définie par la donnée de la moyenne et de l'écart type de la série de données. Les lois de Weibull et de Gumbel font partie des lois d'extremums généralisées. Elles sont aussi définies par deux paramètres : un paramètre de forme et un paramètre d'échelle pour la distribution de Weibull, un paramètre d'échelle et un paramètre de position pour la distribution de Gumbel. Ces trois lois, décrites en annexe VI, sont couramment utilisées pour représenter les données d'éclairement (*e.g.* Beyer *et al.*, 2010; FERNÁNDEZ-PERUCHENA, GASTÓN, *et al.*, 2015; Hansen, 1999; Şen, 2017; Tina & Gagliano, 2011)

Dans un second temps, l'adéquation entre les données historiques et les différentes lois de probabilités classiques sont quantifiées à l'aide d'un test statistique de référence : le test de Kolmogorov-Smirnov (KS) (MARSAGLIA *et al.*, 2003). Ce test non paramétrique, décrit en annexe V, compare les fonctions de distribution cumulatives (CDF) de la série temporelle observée et de celle modélisée par l'ajustement de distribution paramétrique choisi. À partir de l'écart entre les CDF, il conclut quant à la significativité de l'ajustement aux données considérées, avec un seuil de confiance de 0.05. Ce test est couramment utilisé pour l'étude de l'éclairement dans le cadre de projet solaire (FERNÁNDEZ-PERUCHENA, BLANCO, *et al.*, 2015; FERNÁNDEZ-PERUCHENA, GASTÓN, *et al.*, 2015; LARRAÑETA *et al.*, 2018; ŞEN, 2017).

La figure 18 illustre les résultats du test statistique KS. Chaque distribution est représentée par une ligne, contenant les résultats pour les 104 stations GEBA. Sur la base de l'analyse d'au moins 30 ans de données, les couleurs indiquent si la station correspondante rejette l'hypothèse nulle associée au test statistique ou non, pour un niveau de confiance de 0.05. Autrement dit, la couleur jaune indique que la série de données d'éclairement ne peut pas être représentée par la loi en question. La couleur bleue indique qu'il est possible que la série soit représentée par la loi en question. Les valeurs des p-values sont cohérentes avec les valeurs indiquées, et sont données en annexe VI.

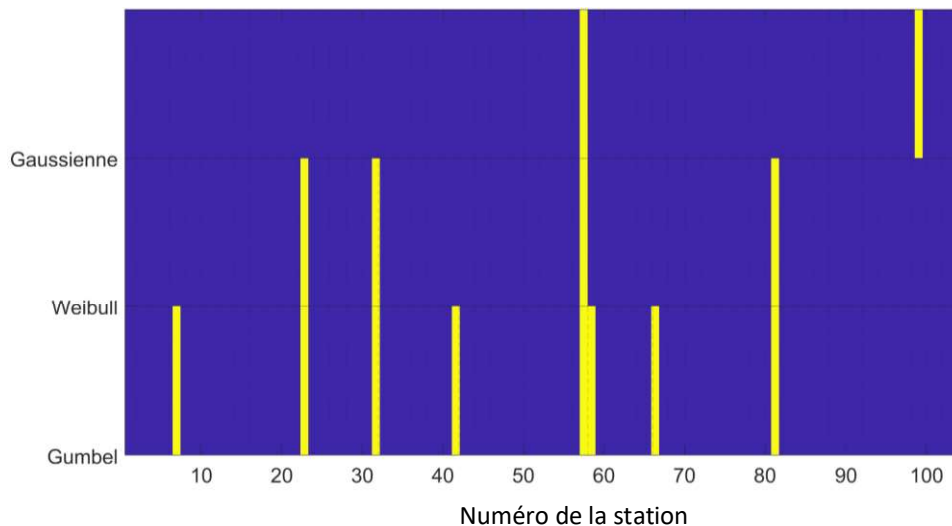


Figure 18 : Résultats du test de Kolmogorov-Smirnov pour trois distributions : gaussienne, de Weibull et de Gumbel. Le test est appliqué aux 104 séries temporelles annuelles GEBA. Les couleurs représentent le rejet (jaune) ou le non rejet (bleu) de l'hypothèse nulle.

Il semble évident au regard de la figure 18 que pour la plupart des stations, nous ne sommes pas en mesure de rejeter l'hypothèse nulle. Cela revient à dire qu'il est impossible de mettre en évidence une loi de distribution, parmi celles considérées, qui représente de façon plus significative que les autres les données historiques. Pour trois stations, nous ne sommes pas en mesure de rejeter l'hypothèse nulle pour la distribution gaussienne (bleu), mais nous sommes en mesure de la rejeter pour les deux autres distributions considérées (jaune). Pour ces trois stations, la distribution gaussienne est donc la plus adaptée pour représenter les données d'éclaircissement historiques. Une station japonaise, celle de Morioka (station n° 58), ne peut être représentée par aucune des trois distributions considérées. Trois stations japonaises et une station indonésienne pourraient être représentées avec la même robustesse par des lois gaussienne ou de Weibull. Enfin, une station allemande, Würzburg (station n° 100), pourrait être représentée de façon équivalente par une distribution de Weibull ou de Gumbel. Au total, malgré la diversité observée, il est impossible d'identifier une unique distribution paramétrique représentant significativement les données historiques pour 101 stations sur les 104 considérées. **Autrement dit, pour la majorité des stations, les trois lois de distribution considérées représentent avec la même robustesse les données historiques.**

S'il est impossible d'identifier une distribution représentant de manière plus robuste les données historiques, il est crucial d'étudier l'influence de la distribution ajustée sur le calcul du P90. Autrement dit, la valeur du P90 estimée à partir des paramètres de la distribution choisie va-t-elle dépendre fortement de la distribution elle-même ?

Pour estimer la sensibilité du P90 à la distribution paramétrique choisie, nous avons calculé les P90 à partir des CDF de chacune des trois distributions considérées. Pour chacune des 104 stations, les valeurs des trois P90 sont présentées sous forme de tableau en annexe VII. Pour chaque station, les écarts entre les trois P90 ont également été calculés. Ces écarts sont représentés sur la figure 19.

Pour ordre de grandeur, le tableau des P90 présente des valeurs allant de 72.9 W/m² pour la station irlandaise de Malin (n°49), à 179.1 W/m² pour la station australienne d’Alice Springs (n°4). On retrouve bien dans le P90 une indication quant aux climats auxquels les stations appartiennent : Malin est située dans un climat particulièrement nuageux, subissant saisonnièrement de longues périodes de nuit, alors que Alice Springs est située dans un climat désertique. Les moyennes des P90, toutes stations confondues, dépendent de la distribution à partir de laquelle elles sont calculées. Elles valent 111.5 W/m² pour la distribution gaussienne, 109.6 W/m² pour la distribution de Weibull et 109.0 W/m² pour la distribution de Gumbel. Ces moyennes illustrent le fait que les distributions de Gumbel et de Weibull offrent des résultats très proches en termes de P90 en moyenne.

Cela se retrouve sur la figure 19. La première ligne présente les écarts entre les distributions gaussienne et de Gumbel, la seconde ligne les écarts entre les distributions gaussienne et de Weibull, et la dernière ligne les écarts entre les distributions de Gumbel et de Weibull. En grisé sont représentées les stations pour lesquelles au moins une des deux distributions ne peut pas être ajustée aux données.

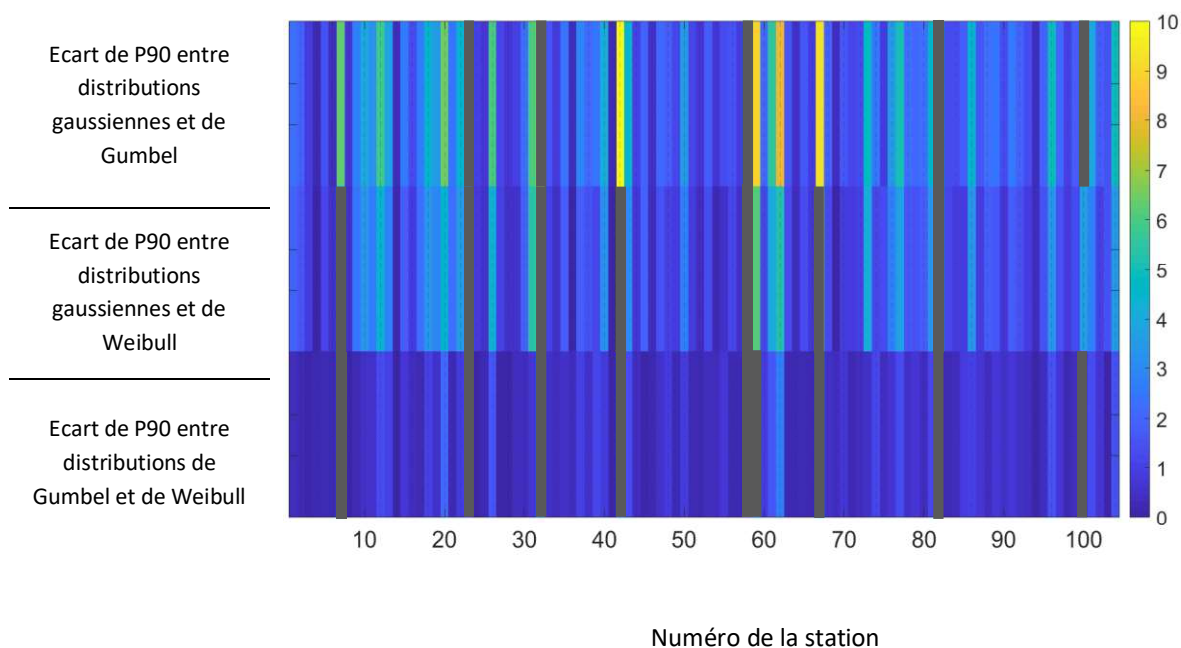


Figure 19 : Ecart entre les P90 calculés à partir des fonctions cumulatives de distribution de trois ajustements de distributions : de Gauss, de Weibull et de Gumbel. Les résultats sont calculés pour 104 stations GEBA, et exprimés en W/m².

On retrouve sur la figure 19 que les P90 estimés à partir des distributions de Gumbel et de Weibull sont très proches : la dernière ligne est beaucoup plus foncée que les deux autres. Au mieux, les P90 sont identiques, ce qui est le cas pour quatre stations. Au pire, ils sont légèrement différents, telle que c’est le cas pour la station japonaise Hikone (n°32). Pour cette dernière station, l’écart de P90 entre les deux distributions est maximal, et vaut 3.4 W/m². En moyenne, l’écart entre les P90 est de 0.7 W/m².

À l'inverse, les P90 calculées à partir des distributions de Gumbel et de Weibull sont très différents de ceux calculés par une distribution gaussienne, comme le montre les deux premières lignes de la figure 19. Les P90 issus de la distribution gaussienne diffèrent en moyenne de 2.6 W/m^2 par rapport à la distribution de Gumbel, et de 2.0 W/m^2 par rapport à la distribution de Weibull. Au maximum, l'écart entre les P90 des distributions de Gumbel et gaussienne est de 12.4 W/m^2 pour la station malaisienne de Kuala Lumpur. Il est de 10.0 W/m^2 pour les distributions de Weibull et gaussienne pour la même station. Ces écarts représentent respectivement 11 % et 9 % de la valeur moyenne multi-annuelle de l'éclairement. **Autrement dit, en fonction de la station et de la distribution ajustée aux données historique, le P90 peut varier jusqu'à de 10 % de sa valeur.**

En conclusion, pour les trois distributions paramétriques utilisées, il est difficile d'identifier une distribution qui représente de façon plus significative que les autres les données historiques. Cette difficulté est principalement due à l'incertitude associée à l'ajustement de distribution sur un échantillon de données de petite taille. Cela rend l'incertitude associée au calcul du P90 par ajustement de distributions paramétriques important pour certaines stations. Cette méthode de calcul ne permet donc pas de caractériser de façon fiable le potentiel solaire futur d'un site.

2.4 Discussion sur l'utilisation du P90

Il est clair au regard des analyses précédentes que l'utilisation du P90 est délicate. D'une part, son estimation – qu'elle soit faite par une méthode empirique ou par un ajustement de distribution paramétrique - est sujette à de grandes incertitudes. Nous avons vu que cette incertitude est en partie liée à la petite taille de l'échantillon de données historiques. D'autre part, son interprétation peut être biaisée compte tenu de la durée de vie d'une centrale photovoltaïque, qui comporte un nombre d'année statistiquement limité.

Pour contourner l'utilisation du P90 sur la ressource solaire, et plus généralement l'utilisation d'indicateurs statistiques pour estimer le potentiel solaire, certaines études proposent d'utiliser un ensemble de séries temporelles d'éclairement (FREEMAN et al., 2018; LARRAÑETA et al., 2018). En faisant ainsi l'information fournie gagne en robustesse statistique : on passe d'une prévision déterministe à une prévision stochastique, pour laquelle il est possible d'estimer de façon précise les incertitudes associées aux méthodes de construction de l'ensemble des séries temporelles. Dans cette idée, les méthodes de génération de multiples séries temporelles d'éclairement, à l'échelle mensuelle ou annuelle, ont été explorées.

3 Génération de séries temporelles d'éclairement : analyse bibliographique

La section suivante s'attache à l'étude d'une sélection de méthodes de génération de séries temporelles synthétiques. Chaque méthode sera présentée, puis la représentativité des données annuelles générées sera discutée. Enfin, les hypothèses sous-jacentes à la méthode seront explicitées et discutées. Pour des raisons de concision, nous avons choisi de ne pas décrire en détail l'ensemble des méthodes analysées. Nous analyserons en détail deux des méthodes rencontrées. Les résultats issus de cette analyse seront mis en perspectives avec les résultats issus de l'ensemble des analyses.

3.1 Méthodologie de l'analyse bibliographique

L'angle choisi pour cette étude bibliographique est celui de l'étude des hypothèses implicites ou explicites conditionnant les performances de chaque méthode. Ce choix est motivé par les conclusions établies jusqu'alors : le nombre de données peut jouer de façon très importante sur la robustesse des résultats issus d'analyses statistiques. Or, les méthodes de générations de données étudiées par la suite sont principalement statistiques. Il est donc probable que leur robustesse dépende également de la taille de l'échantillon. Ainsi, nous avons porté une attention particulière à la façon dont les auteurs considèrent cette hypothèse.

Dans la même idée, une attention a été portée sur trois autres hypothèses classiques, retrouvées largement dans la bibliographie : l'indépendance des données d'éclairement, sa stationnarité temporelle, et la capacité de représenter de façon robuste l'échantillon de données considéré par une distribution paramétrique.

Ces hypothèses peuvent être questionnées pour le cas de l'éclairement. Tout d'abord, un ensemble d'oscillations climatiques et de variabilités multi-annuelles, affectant la dynamique atmosphérique sur plusieurs années consécutives, peut être trouvé dans la bibliographie (KNUDSEN *et al.*, 2011; MARCHANT *et al.*, 2007; MO *et al.*, 1998; PAETH *et al.*, 2003). Or, l'éclairement dépend grandement de la couverture nuageuse, de la composition gazeuse et de la quantité d'aérosols dans l'atmosphère, tout comme de leurs possibles couplages. Celui-ci devrait donc être impacté par ces variations multi-annuelles. Ainsi, l'éclairement ne serait alors pas indépendant d'une année à l'autre.

Ensuite, l'hypothèse de stationnarité temporelle peut être questionnée par la présence éventuelle de changements abrupts ou de long terme de l'éclairement annuel. Dans ce cas-là, l'échantillon de données historiques ne serait alors plus représentatif de l'éclairement annuel futur du site d'intérêt. Il se trouve que des tendances de long terme dans les mesures d'éclairement sont recensées dans la bibliographie (*e.g.* LOHMANN *et al.*, 2006; MÜLLER *et al.*, 2014; WILD, 2009; Wild, 2012).

3. Génération de séries temporelles d'éclairement : analyse bibliographique

Enfin, les expérimentations réalisées dans le cadre de la discussion de l'indicateur P90 peuvent fournir des éléments questionnant la qualité de la représentation d'un échantillon de petite taille par une distribution paramétrique.

En conclusion, les quatre hypothèses étudiées peuvent être résumées comme il suit :

- (i) La robustesse des résultats issus de calculs réalisés sur des échantillons de petite taille.**
- (ii) L'indépendance des données d'éclairement à l'échelle mensuelle ou annuelle.**
- (iii) La stationnarité temporelle de l'éclairement. Celle-ci est souvent traduite par la correspondance entre la distribution associée aux données historiques et celles associée aux données futures d'éclairement.**
- (iv) La capacité de représenter de façon robuste l'échantillon de données considéré par une distribution paramétrique.**

Notons que les hypothèses (ii) et (iii) sont régulièrement considérées de façon conjointes : on parle alors d'hypothèse i.i.d. pour indépendant et identiquement distribuée.

Afin de répertorier la présence ou l'absence de ces quatre hypothèses dans chacune des méthodes étudiées, un tableau synthétique a été construit. Ce tableau sera implémenté pour chaque méthode. Le tableau complet est disponible en section **Erreur ! Source du renvoi introuvable..**

À titre illustratif, le tableau 5 rassemble les différentes conclusions de l'analyse de la pratique industrielle réalisée jusqu'à présent. Les symboles traduisent si l'hypothèse correspondante est postulée ou non dans la méthode considérée : X signifie que l'hypothèse n'est pas postulée, V qu'elle l'est. Les couleurs illustrent les efforts faits par les auteurs pour quantifier ou dépasser les limitations propres à chacune des hypothèses.

Hypothèse Méthodes	Robustesse d'un faible nombre de données	Indépendance des données	Stationnarité temporelle	Robustesse d'un ajustement de distribution
P90 empirique	V	X	V	X
P90 par ajustement de distribution	V	X	V	V

Tableau 5 : Analyse des hypothèses de différentes méthodes d'estimation du P90.

On retrouve sur le tableau 5 que dans les deux méthodes citées, l'influence d'un échantillon historique de petite taille sur le calcul statistique d'indicateurs tels que le P90 n'est pas considéré. De même, l'éclairement historique est considéré représentatif de l'éclairement futur, supposant implicitement la stationnarité temporelle de l'éclairement dans les deux cas. Pour le cas du calcul du P90 de façon empirique, un ajustement de distribution paramétrique aux données historiques est parfois utilisé. La couleur rouge souligne que les hypothèses sont la plupart du temps posées de façon implicite, ce qui limite grandement les efforts fait pour quantifier les écarts à ces hypothèses. L'indépendance des données annuelles ou mensuelles n'est pas supposée dans ces deux méthodes.

Pour démonstration, nous avons choisi de décrire et discuter de deux méthodes particulières : celle proposée par Fernandèz-Peruchena *et al.* (2015) et celle proposée par Larrañeta *et al.* (2018). Ces méthodes couplent des approches déterministes et stochastiques : elles ajoutent des fluctuations de court terme à des données d'éclairement générées par un modèle adapté. Les sections suivantes s'attacheront à la présentation puis à la discussion de chacune de ces deux méthodes, avant de synthétiser les résultats de l'ensemble des études considérées.

3.2 Méthode de Fernandèz-Peruchena *et al.* (2015)

3.2.1 Principe de la méthode

Fernandèz-Peruchena *et al.* (2015) proposent une méthode de génération d'années météorologiques possibles (Possible Meteorological Year, PMY). Les auteurs définissent une PMY comme une série annuelle d'éclairement statistiquement consistante avec les estimations de la variabilité mensuelle et annuelle locale. La variabilité d'un site est estimée à partir de l'écart type d'une série de données *in-situ* de long terme (15 années). Une grande attention est donc portée au fait que la série de données générées respecte la variabilité inter-annuelle de long terme du site.

Les données annuelles sont générées de la façon suivante : dans un premier temps, la moyenne et l'écart type multi-annuels de la série temporelle historique sont calculées. Par la suite, un nombre est tiré aléatoirement dans une distribution normale de moyenne nulle et d'écart type égale à l'écart type calculé précédemment. La PMY correspond à la somme de la moyenne multi-annuelle et du nombre aléatoire. Cela revient, sous une forme, à ajuster une distribution gaussienne à la série temporelle historique.

Pour estimer la qualité de la méthode, les auteurs comparent les données générées avec l'échantillon de données historiques. Il n'y a donc pas de distinction entre l'échantillon d'apprentissage, utilisé pour calibrer la méthode, et un échantillon de validation. La comparaison est réalisée pour 18 sites possédant des mesures de long terme, reconnues de grande qualité. Ces stations sont choisies pour être représentative de différents climats. En plus des indices de performance et de similitude, les (CDF) empiriques des données générées et des données mesurées sont comparées. Les CDF des données historiques et des données synthétiques générées sont remarquablement proches, comme le rapporte la figure 20. Les résultats de la génération de données mensuelles sont encore plus encourageants

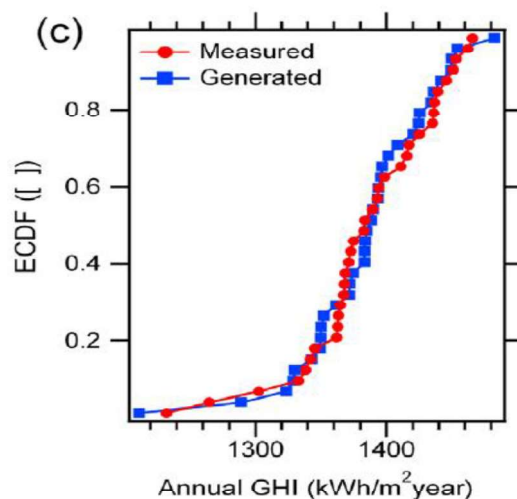


Figure 20 : Comparaison de CDF empiriques entre données d'éclairement in-situ mesurées à Eugene (USA) et données synthétiques générées. Figure extraite de Fernandèz-Peruchena et al. (2015)

Cette méthode semble donc tout à fait indiquée pour générer des données synthétiques dont la CDF est proche des données observées. Elle est donc adaptée si l'on s'intéresse à la quantité globale de la ressource disponible sur site durant la période d'analyse. **Bien sûr, ces résultats ne sont valables que si l'éclairement annuel respecte les hypothèses sous-jacentes à l'étude.** Parmi celles-ci on peut citer :

- **L'estimation statistique de la moyenne et de l'écart type à partir d'un échantillon de données restreint (maximum de 15 années) est robuste statistiquement.**
- **Les données annuelles d'éclairement peuvent être représentées de façon robuste grâce à une loi gaussienne.**
- **La variabilité du jeu de données historiques est représentative de la variabilité du jeu de données futures.** Autrement dit, le jeu de données futures peut être représenté par la même distribution paramétrique que le jeu de données historiques. Cela revient à dire que l'éclairement est supposé stationnaire temporellement.
- **Les valeurs annuelles d'éclairement ne présentent pas de corrélation entre elles.**

3.2.2 Analyse de la méthode au regard des hypothèses d'intérêt

Ces hypothèses peuvent être discutées au regard des hypothèses introduites en section 3.1 :

- (i) **Le manque de robustesse des résultats issus de l'estimation statistique de l'écart type à partir d'un échantillon de données restreint est largement documenté (e.g. AHN & FESSLER, 2003; BRETHERTON, 2014).** Celle-ci peut également être discutée en fonction **du climat considéré.** En effet, si le climat du site d'intérêt est relativement stable, l'écart type sera rapidement constant et son estimation sera alors robuste même avec un faible nombre d'années disponibles. À l'inverse, plus un climat sera variable, plus le nombre d'années à considérer pour avoir une bonne estimation de l'écart type sera important. Autrement dit, une adéquation entre la variabilité calculée à partir de l'écart type et les fluctuations météorologiques du site peut être observée (e.g. Ineichen (2014) en Europe,

Cebecauer et Suri (2015) en Afrique du Sud, Copper *et al.* (2017) en Australie). Par exemple, pour un même site, le coefficient de variations (correspondant pour rappel au ratio entre l'écart type et la moyenne multi-annuelle) peut être estimé à 1.3 % pour une série de données de 1 an, contre de 0.2 % pour 20 ans (CEBECAUER & SURI, 2015). Une estimation de l'incertitude associée au calcul de l'écart type serait donc souhaitable pour une bonne interprétation des résultats.

Il apparait donc, en plus de la sensibilité statistique du résultat à la taille de l'échantillon, une sensibilité aux fluctuations météorologiques locales. Malheureusement, il est impossible avec l'état actuel des connaissances, de connaître précisément à l'avance la variabilité de l'éclairement propre à un site particulier. Une connaissance fine de cette variabilité permettrait de déterminer le nombre idéal de données à considérer dans le calcul de l'écart type.

- (ii) En construisant les années de façon indépendante les unes des autres, les auteurs supposent que la corrélation entre les années successives est négligeable. **Cependant, l'existence de variations d'éclairement de long terme remet en cause cette hypothèse.** Ces variations peuvent être de l'ordre de quelques années, avec par exemple les impacts de l'oscillation Nord Atlantique (NAO, *North Atlantic Oscillation*) sur le couvert nuageux et donc sur le rayonnement (*e.g.* POZO-VÁZQUEZ *et al.*, 2004), ou encore l'impact d'El Niño (ENSO, *El Niño Southern Oscillation*) sur l'éclairement (*e.g.* DAVY & TROCCOLI, 2012; LI, 2013). Des variations de l'ordre de la dizaine d'années sont également répertoriées, avec par notamment les tendances de long terme décrites par Wild *et al.* (2009). Au vu de ces études, il semble probable que selon la zone géographique considérée, l'éclairement annuelle ne soit indépendant d'une année sur l'autre.
- (iii) L'hypothèse de stationnarité temporelle se retrouve sous deux formes dans l'étude de Fernandéz-Peruchena *et al.* (2015) : la correspondance imposée entre la variabilité des séries temporelles historiques et futures, et l'utilisation du jeu de données historiques pour valider la méthode de génération. Il ne sera donc pas envisageable que la série de données générées prennent en compte des événements hors de la variabilité des 20 dernières années. **Or, comme annoncé précédemment, il est possible que l'ensemble des variations du signal ne soit pas correctement représenté par un historique de 20 ans. Encore une fois, une connaissance fine de la variabilité du site étudié permettrait de déterminer si l'hypothèse de stationnarité est fondée ou non.** Cette hypothèse ne trouve pas d'appui d'un point de vue physique ou statistique dans l'article.
- (iv) Enfin, la méthode proposée suppose l'adéquation entre l'échantillon de données historiques et son ajustement par distribution gaussienne. Comme vu précédemment, cette hypothèse est discutable pour un petit échantillon de données, au regard de l'incertitude associée à l'ajustement d'une distribution particulière plutôt qu'une autre.

En résumé, l'ensemble des quatre hypothèses étudiées sont supposées dans l'étude, de façon implicite (robustesse d'un faible nombre de données et stationnarité temporelle) ou explicite (indépendance et robustesse de l'ajustement de distribution). Ces différentes informations sont résumées dans la troisième colonne du tableau 6.

Pour aller plus loin, Larrañeta *et al.* (2018) proposent une méthode de génération de données cherchant à s'affranchir de l'hypothèse de stationnarité temporelle supposé ici.

3.3 Méthode de Larrañeta *et al.* (2018)

3.3.1 Principe de la méthode

Larrañeta *et al.* (2018) proposent une méthode de génération de données d'éclairement direct en incidence normale (DNI). Cette méthode n'a pas été développée explicitement pour l'éclairement global sur plan horizontal. Cependant, il est tout à fait possible de s'en inspirer pour imaginer une méthode parente de génération de données synthétiques pour le GHI.

Dans un premier temps, les auteurs génèrent une série de données mensuelles de DNI. Pour cela, ils estiment pour chaque mois le profil type de ciel clair moyen ajusté au site d'intérêt. La CDF empirique des différences entre la valeur du profil type et la mesure effective de ciel clair est ensuite construite. Enfin, un tirage aléatoire dans une distribution uniforme est réalisé. La valeur tirée, comprise entre 0 et 1, est associée à une valeur de DNI par l'intermédiaire de la CDF empirique. La donnée synthétique d'éclairement direct correspond à la somme de ces deux composantes : la valeur issue du profil type, et la valeur statistique de la différence avec ce profil. Ainsi, un ensemble de valeurs de DNI pour chaque mois est généré.

Dans un second temps, les données mensuelles calendaires sont aléatoirement concaténées, puis agrégées pour former des données synthétiques annuelles. Cette concaténation permet d'obtenir un jeu de données annuelles d'éclairement possédant une variabilité inter-annuelle différente de celle de la série d'origine. Par exemple, en combinant plusieurs mois synthétiques correspondant à l'éclairement minimum dans la série d'origine, il est possible de créer une année avec un éclairement plus faible que toutes les années mesurées. Ainsi, les auteurs soutiennent qu'il est possible d'avoir une génération de données annuelle plus proche des données observées en générant dans un premier temps des données mensuelles.

Il est intéressant de noter que les auteurs partageant les conclusions de l'analyse réalisée en section 2.4 : ils proposent de générer un ensemble de séries temporelles d'éclairement direct plutôt qu'une série temporelle unique de données. Pour évaluer la qualité de la méthode, les auteurs comparent les CDF d'un an de données mensuelles d'une part, et de 50 années générées d'autre part. La figure 21 présente un exemple de CDF empirique pour les données mesurées et générées.

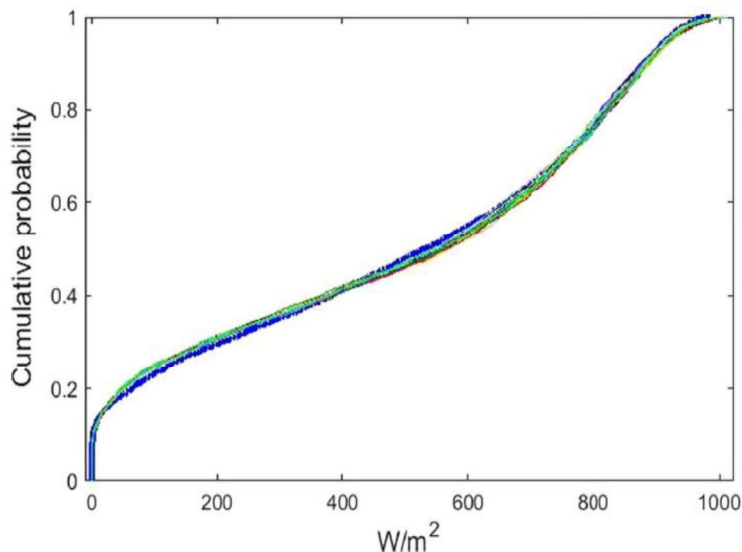


Figure 21 : Comparaison de CDF empiriques entre données d'éclairément in-situ mesurées à Séville (Espagne) en 2014 (ligne épaisse bleue) et 50 données synthétiques annuelles. Figure extraite de Larrañeta et al. (2018).

On observe que les courbes générées sont très proches des mesures du site, en bleu. **Ainsi, cette méthode semble appropriée pour étendre un jeu de données, sous couvert de la validité des hypothèses postulées dans l'étude, à savoir :**

- **L'indépendance des données à l'échelle mensuelle et annuelle.** En effet, en concaténant des mois tirés aléatoirement dans l'ensemble de mois générés correspondants, il est supposé implicitement que les données mensuelles ne sont pas corrélées. De même, en associant de manière aléatoire les années générées, il est supposé implicitement que les données annuelles ne sont pas non plus auto-corrélées.
- **Il est supposé dans l'étude que la validation de la méthode à partir d'une unique année permet de tirer des conclusions quant à la validité de la méthode présentée.**

3.3.2 Analyse critique de la méthode au regard des hypothèses d'intérêt

Cette méthode peut être discutées au regard des quatre hypothèses introduites en section 3.1 : la robustesse des résultats issus de calculs réalisés sur des échantillons de petite taille, l'hypothèse i.i.d et la qualité d'un ajustement de distribution à l'échantillon de données.

- (i) Les auteurs n'introduisent pas d'indicateurs ou d'estimateurs statistiques qui pourraient être influencés par la petite taille de l'échantillon historique. En ce sens, ils prennent en compte la taille de l'échantillon de manière satisfaisante. Soulignons néanmoins que la validation à partir d'une année de données mensuelle de DNI entraîne des incertitudes sur les performances de la méthode qu'il pourrait être utile de quantifier.

3. Génération de séries temporelles d'éclairement : analyse bibliographique

- (ii) Les auteurs ne considèrent pas de corrélation entre les mois ou les années individuelles dans plusieurs étapes : lors de l'estimation du profil type de ciel clair, lors du tirage aléatoire au sein de la CDF des différences entre profil type et valeur mesurée, ou encore lors de la concaténation aléatoire de mois d'éclairement synthétiques pour créer le jeu de données annuel d'intérêt.

De la même façon que pour la méthode de Fernandéz-Peruchena *et al.* (2015), l'existence d'oscillations climatiques et de cycles de long terme de l'éclairement peut remettre en question l'indépendance des données annuelles. En ce qui concerne l'indépendance des données mensuelles, elle peut également être discutée au regard d'oscillations climatiques telles que NAO par exemple. Pour ce cas, des blocages atmosphériques sur plusieurs semaines ou plusieurs mois peuvent être répertoriés (CHIACCHIO & WILD, 2010; HANNA *et al.*, 2015). Ainsi, la probabilité d'avoir plusieurs mois d'éclairement consécutifs faibles dans la région islandaise, ou forts autour des Açores, est importante. Dans ce contexte, il est difficile d'envisager un rayonnement indépendant à l'échelle mensuelle.

- (iii) **Les auteurs proposent une contribution à l'état de l'art en s'affranchissent de l'hypothèse de stationnarité temporelle de l'éclairement.** En effet, la série temporelle annuelle générée peut avoir une variabilité interannuelle différente de la série temporelle historique. Cependant, pour ce faire, ils concatènent de manière aléatoire différents mois synthétiques. **Premièrement, lors de cette concaténation, la création d'années d'éclairement extrêmes, particulièrement éclairées ou particulièrement peu éclairées, est fortuite et non maîtrisée.** À l'extrême, 10 mauvais mois peuvent être concaténés, donnant une année d'éclairement beaucoup plus faible que ce qui est physiquement possible pour le site considéré. Cette constatation interroge la représentativité physique des données générées. **De plus, la méthode de génération mensuelle (somme d'un profil moyen estimé à partir de l'historique et d'une différence entre ce profil et la mesure) peut être vue plus comme un bruitage du profil moyen historique que comme un réel affranchissement de l'hypothèse de stationnarité temporelle.**
- (iv) Les auteurs n'ont pas recours à un ajustement de distribution paramétrique. En ce sens, ils dépassent les limites soulevées par l'ajustement d'une distribution à un échantillon de petite taille.

En résumé, les auteurs proposent de générer un grand nombre de données et de séries temporelles pour contourner les limitations propres à la taille de l'échantillon initial. Ils n'ajustent pas de distribution paramétrique aux données historiques. La méthode proposée est innovante par rapport à l'état de l'art, car elle permettant d'avoir un écart type différent entre l'échantillon d'origine et celui généré. Cependant, la variabilité générée sera fortuite et peu représentative de la physique sous-jacente à l'éclairement. Enfin, les auteurs considèrent des données indépendantes à l'échelle mensuelle et annuelle, ce qui peut être remis en question en présence de variations de long terme de l'éclairement. Ces différentes infirmations sont résumées dans la quatrième colonne du tableau 6.

La section suivante traite de la généralisation de ces résultats par la réalisation d'une analyse similaire pour d'autres méthodes de l'état de l'art.

3.4 Résumé et conclusions de l'analyse bibliographique

Cette section présente succinctement les résultats issus de l'analyse de cinq études de l'état de l'art. Ces études ont été sélectionnées, car présentent des méthodologies différentes. Elles ont été analysées au regard des quatre hypothèses choisies. L'ensemble des résultats exposés par la suite sont regroupés dans le tableau 6tableau 6.

Cebecauer and Suri (2015) explicitent une partie des méthodes utilisées dans le logiciel de support d'aide à la décision d'investissement photovoltaïque Solargis (ŠÚRI *et al.*, 2011). En particulier, ils détaillent les méthodes de calcul des années météorologiques typiques (TMY). Une TMY est une série temporelle artificielle de douze mois construite pour être représentative des conditions moyennes (P50). Ils étendent le concept du TMY à d'autres niveaux de probabilités d'excédence, dont le P90. Ces probabilités d'excédences sont estimées de façon empirique à partir de la série temporelle annuelle historique. Les TMY sont usuellement construites en sélectionnant 12 mois parmi une série temporelle de long terme d'éclairement. La sélection se fait de façon à ce que les moyennes, ou les P90, de la TMY et de la série temporelle multi-annuelle soient égales. Les 12 mois sont ensuite concaténés pour fournir une valeur annuelle de TMY. Enfin, les TMY calculées sont aléatoirement concaténées pour construire une série temporelle d'éclairement sur la durée de vie de la centrale.

On retrouve dans cette méthode trois éléments abordés précédemment : le calcul de probabilité d'excédence de façon empirique (section 2.3.1) ; la correspondance imposée entre des paramètres statistiques de l'échantillon historique et futur (section 3.2.2), et la concaténation de mois et d'années générées (section 3.3.2). Comme vu précédemment, cette méthode sera donc soumise aux incertitudes associées : au calcul statistique à partir d'échantillons de petites tailles ; à l'absence de prise en compte de l'autocorrélation entre les données mensuelles et annuelles ; et à l'absence de prise en compte de possibles non stationnarité de l'éclairement. Pour cette méthode, aucun ajustement de distribution aux données historiques n'est réalisé.

Tomson *et al.* (2008) étudient quant à eux la dynamique de l'éclairement solaire à plusieurs échelles. En ce qui concerne l'éclairement inter-annuel, les auteurs postulent explicitement sa stationnarité temporelle dans plusieurs sections du document. À l'inverse, ils illustrent la présence de tendances linéaires de rayonnement pour une station particulière (Toravere, en Estonie). Pour cette station, ils travaillent à ôter les tendances linéaires observées – estimées à partir de 35 ans de données au plus. Ils remettent ainsi en cause implicitement la stationnarité temporelle. De la même façon, l'indépendance des données annuelles d'éclairement est explicitement postulée dans certaines sections, et questionnée sur des exemples particuliers.

L'information relative aux quatre hypothèses est complexe pour cet exemple. Nous avons conclu que ces deux hypothèses – indépendance des données annuelles et stationnarité temporelle – sont de manière générale postulées par les auteurs, bien que ceux-ci proposent une réflexion sur la validité de ces hypothèses grâce à un exemple d'application. En ce qui concerne la prise en compte de la taille de l'échantillon dans l'analyse des résultats, celle-ci n'est pas mentionnée : aucune donnée

3. Génération de séries temporelles d'éclairement : analyse bibliographique

concernant la significativité des tendances linéaire calculée à partir de 35 données au plus n'est fournie. Aucun ajustement de distribution paramétrique n'est réalisé.

Thevenard & Pelland (2013) proposent quant à eux une méthode de génération de série temporelle de productible sur 30 ans. Un effort notable réside dans la prise en compte de manière fine d'un ensemble d'incertitudes pouvant jouer sur le productible. Cela est notamment fait à l'aide du logiciel de simulation de technico-économique pour les énergies renouvelable – notamment solaire photovoltaïque – appelé *System Advisor Model (SAM)*, proposé par le NREL (National Renewable Energy Laboratory) (FREEMAN et al., 2018). SAM modélise cet ensemble d'incertitudes à partir de séries temporelles multiples. Pour générer 100 séries temporelles d'éclairement – *a minima* – les auteurs distinguent deux cas en fonction de la variabilité inter-annuelle du site considéré : si la variabilité inter-annuelle est faible, la série temporelle générée est une concaténation de 30 TMY ; si la variabilité inter-annuelle est grande, les auteurs génèrent des séries temporelles de dix années. Les dix années sont choisies à partir de la CDF de la série temporelle historique pour représenter au mieux la diversité des conditions d'éclairement que connaît le site. À partir des séries annuelles d'éclairement générées, le logiciel calcule l'incertitude associée à la ressource solaire en postulant que les erreurs suivent une distribution normale. Enfin, les données de productibles finales sont générées. Dans le processus de génération, un ajustement de distribution à un échantillon de petite taille est réalisé. L'indépendance des données annuelles de productibles est également postulée.

On retrouve plusieurs points soulevés précédemment dans la méthode de génération détaillée ci-dessus. L'indépendance à l'échelle mensuelle est postulée lors du calcul de TMY. L'indépendance à l'échelle annuelle est postulée lors du calcul du productible. L'influence du petit échantillon considéré dans le calcul de CDF n'est pas abordée. L'hypothèse de stationnarité est implicitement supposée par la construction des séries annuelles futures à partir de la CDF empirique de quelques données historiques. Cependant, des efforts sont faits pour s'affranchir de chacune des hypothèses. Par exemple, les hypothèses d'indépendance à l'échelle annuelle et de robustesse d'un ajustement de distribution ne sont pas postulées pour la génération de données d'éclairement. C'est le fait que le productible soit considéré comme indépendant d'une année à l'autre qui implique, sous une certaine forme, une indépendance annuelle de l'éclairement. De même, l'ajustement d'une distribution paramétrique est réalisé sur les données de productibles, et non d'éclairement. Dans la même idée, un effort est fait quant à l'hypothèse de stationnarité temporelle, en distinguant deux cas en fonction de la variabilité d'un site. Les auteurs réduisent les incertitudes associées au nombre de données pour le calcul en aval des indicateurs telles que les P90 sur le productible en utilisant des séries multiples d'éclairement.

Le même type d'analyse a été réalisé pour deux autres études : celle de Wilcox et Gueymard (2010), et celle de Abreu *et al.* (2018). La première est similaire à celle de Thevenard & Pelland (2013), dans le sens où elle propose de s'affranchir de l'hypothèse de stationnarité temporelle en distinguant les zones de faible et de forte variabilité inter-annuelle. Elle diffère car elle suppose la robustesse d'un ajustement de distribution. La seconde est similaire à celle Cebecauer and Suri (2015), dans le sens où les auteurs génèrent des séries temporelles à partir de TMY pour des stations *in-situ*. Les conclusions de nos analyses sur ces deux études sont donc également similaires.

Le tableau 6 rassemble les résultats exposés jusqu'ici. Pour aller plus loin, il serait possible d'étudier des méthodes de génération de séries temporelles synthétiques d'éclairement plus complexes, basées sur des méthodes de *machine learning* telles que celles décrites dans Badescu *et al.* (2014).

Méthodes	Hypothèse	Robustesse d'un faible nombre de données	Indépendance des données	Stationnarité temporelle	Robustesse d'un ajustement de distribution
P90 empirique		V	X	V	X
P90 par ajustement de distribution		V	X	V	V
Fernandèz-Peruchena <i>et al.</i> (2015)		V	V	V	V
Larrañeta <i>et al.</i> (2018)		X	V	V	X
Cebecauer and Suri (2015)		V	V	V	X
Tomson <i>et al.</i> (2008)		V	V	V	X
Thevenard & Pelland (2013)		V	V	V	X
Wilcox et Gueymard (2010)		V	X	V	V
Abreu <i>et al.</i> (2018)		V	V	V	X

Tableau 6 : Analyse des hypothèses d'intérêts postulées dans différentes études de l'état de l'art. (V, X) traduisent si l'hypothèse considérée est postulée ou non. Les couleurs représentent les efforts faits pour dépasser ou quantifier les incertitudes liées à ces hypothèses : rouge pour une hypothèse posée implicitement ou considéré comme vraie sans justification, orange pour un effort fait par les auteurs pour dépasser l'hypothèse, sans pour autant y parvenir, blanc pour une hypothèse non postulée.

Premièrement, soulignons que la robustesse des résultats calculés à partir d'un faible nombre de données est postulée dans l'ensemble des études considérées. Cela s'explique aisément par la faible disponibilité, en pratique, des données de mesures d'éclairement. Ainsi, travailler avec un échantillon de données de petite taille est une contrainte fondamentale, dans la bibliographie comme dans cette thèse. La quantification de l'incertitude associée à cette hypothèse devra par la suite être toujours explicitée.

Deux hypothèses se dégagent du tableau : l'hypothèse d'indépendance des données d'éclairement annuelles et celle de stationnarité temporelle de ces données. Bien qu'elles soient classiques dans le domaine de la caractérisation du potentiel solaire, les écarts à ces hypothèses sont rarement considérés dans les études, ou alors supposés négligeables face aux autres sources d'incertitudes. Cela nous a amenés à envisager une étude plus approfondie de ces hypothèses.

Nous nous sommes attachés dans la suite de ce document à étudier la validité de l'hypothèse de stationnarité temporelle. En effet, en présence de variations de long terme, l'hypothèse d'indépendance de l'éclairement à l'échelle annuelle serait discréditée de fait. Les méthodes de l'état de l'art ne pourront alors pas être appliquées sans réserve. À l'inverse, si l'hypothèse de stationnarité temporelle s'avère valide, l'hypothèse d'indépendance devra faire l'objet d'une étude approfondie. Si ces deux hypothèses sont valides, alors les méthodes de l'état de l'art décrites précédemment pourront être appliquées sans réserve.

4 Conclusion

Dans un premier temps, un ensemble d'expérimentations très simples ont remis en question l'utilisation de l'indicateur P90 dans l'analyse de risques financiers d'une centrale solaire photovoltaïque. En particulier, il a été constaté que même dans le cas d'un P90 connu de manière idéale avec des hypothèses en pratique restrictives de i.i.d., l'utilisation de celui-ci est associée à une très grande incertitude. Cette incertitude est imputable au court horizon temporel futur considéré, qui est limité par la durée de vie de la centrale sur quelques dizaines d'années. La difficulté de discréditer l'ajustement de trois distributions classiques aux données historiques a été montrée que l'influence de la distribution choisie sur la valeur du P90. Ces constatations nous ont mené à la conclusion suivante : il serait plus adapté de fournir à l'utilisateur un ensemble de séries temporelles futures plausibles pour un site, plutôt qu'un ou plusieurs indicateurs statistiques tels que le P90.

Dans cette idée, nous avons discuté en détail deux méthodes de l'état de l'art proposant la génération d'ensembles de données futures dans le cadre de la prévision solaire. L'étude de ces méthodes a été faite au regard de quatre hypothèses utilisées couramment dans l'état de l'art, bien que souvent formulées implicitement : la robustesse des résultats issus de calculs réalisés sur des échantillons de petite taille, l'indépendance des données d'éclairement, la capacité de représenter de façon robuste l'échantillon de données considéré par une distribution paramétrique et la stationnarité temporelle de l'éclairement. Nous avons pu discuter en détail la plausibilité de ces hypothèses pour le cas de l'éclairement. En conclusion, il est apparu que les hypothèses d'indépendance et de stationnarité temporelle étaient fréquentes dans les méthodes de l'état de l'art, souvent sans être explicitées, appuyées ou justifiées, et que les incertitudes induites par ces hypothèses sur la robustesse des résultats étaient rarement quantifiées.

La présence de possibles tendances de long terme ou de variations de long terme de l'éclairement n'étant pas *a priori* réfutable, nous avons choisi d'approfondir la question de la stationnarité temporelle de l'éclairement. Ce choix est soutenu par le fait que si la stationnarité temporelle est remise en cause, l'hypothèse d'indépendance annuelle des données le sera également.

Chapitre 4

Etude de la stationnarité temporelle de l'éclairage

Table des matières

1. Introduction	81
2. Analyse bibliographique relative aux tendances de l'éclairage.....	81
2.1 Comparaison des méthodologies utilisées dans différentes études relatives aux tendances d'éclairage	82
2.2 Comparaison des résultats issus d'études relatives aux tendances d'éclairage .	85
3. Comparaison des résultats issus de la bibliographie à l'analyse de mesures locales en Chine	88
3.1 Résumé du consensus bibliographique sur la Chine	88
3.2 Analyse visuelle des stations GEBA au regard de la bibliographie	89
3.3 Analyse statistique des stations GEBA au regard de la bibliographie	90
4. Conclusion.....	94

1 Introduction

L'hypothèse de stationnarité temporelle de l'éclairement est implicitement postulée dans les phases d'un montage de projet photovoltaïque, comme dans l'état de l'art (*e.g.* Fernandéz-Peruchena *et al.*, 2015; Larrañeta, Moreno-Tejera, Silva-Pérez, & Lillo-Bravo, 2015). Cette hypothèse étant dans la grande majorité des cas implicite, elle ne trouve pas de fondement scientifique, et n'est pas le sujet de discussions. À l'inverse, il existe dans la bibliographie des études portant sur l'existence d'oscillations climatiques de long terme et de tendances de l'éclairement (*e.g.* Müller *et al.*, 2014; Pfeifroth, Sanchez-Lorenzo, Manara, Trentmann, & Hollmann, 2018; Sanchez-Lorenzo *et al.*, 2015; Wild, 2009). La présence de telles variations de long terme de l'éclairement invalide l'hypothèse de stationnarité.

Une contradiction est donc visible à ce stade : certains travaux sont fondés sur une hypothèse de stationnarité temporelle de l'éclairement, quand d'autres montrent une non-stationnarité de la même quantité. Partant de ce constat, nous avons décidé d'approfondir l'analyse de la validité de cette hypothèse.

Pour cela, une analyse des méthodes utilisées dans la bibliographie portant sur les tendances de long terme de l'éclairement a été réalisée. **L'objectif de cette analyse a été de déterminer si l'information disponible dans la bibliographie est adaptée à l'aide à la décision dans un contexte de projet solaire photovoltaïque.** Si les méthodes utilisées offrent des résultats similaires à l'échelle du point, alors l'information fournie par la bibliographie permettra une prise en compte fine des variations de long terme de l'éclairement.

2 Analyse bibliographique relative aux tendances de l'éclairement

À première vue, les résultats issus de la bibliographie ont l'air de faire consensus. L'étude la plus connue traitant des variations de long terme de l'éclairement est la revue de Wild (2009). Cette méta-analyse recense les résultats de plus de 50 études portant sur les tendances de l'éclairement entre 1960 et 2000. **Deux périodes temporelles sont distinguées. Ces périodes sont caractérisées par des comportements différents en termes d'éclairement : une période de décroissance, de 1960 à 1990, connue sous l'appellation d'obscurcissement global (*global dimming*), et une période de croissance, de 1980 à 2000, connue sous l'appellation d'éclaircissement global (*global brightening*) (OHVRIL *et al.*, 2009).** Plus récemment, ces tendances ont également été confirmées à partir de mesures *in-situ* (CHIACCHIO & WILD, 2010; SANCHEZ-LORENZO *et al.*, 2015) et d'estimations satellitaires (Posselt *et al.*, 2014; Sanchez-Lorenzo *et al.*, 2013). Müller (2014) confirme que la confiance dans ces résultats est d'autant plus grande qu'ils se retrouvent dans un grand nombre d'études. Ces études portent sur l'éclairement de sites appartenant à différents climats, ce qui exclue l'influence d'un climat en particulier sur les tendances observées. Ils utilisent des instruments différents, limitant ainsi les biais de mesure possibles. Müller (2014) souligne également que

l'obscurcissement global et l'éclaircissement global se retrouvent également dans des études sur des quantités qui présentent des niveaux de corrélations notables avec l'éclairement, tels que la température journalière, l'albédo planétaire, l'évaporation (WILD, 2009), ou bien sûr dans la production photovoltaïque elle-même (B. MÜLLER et al., 2009).

Partant de ce constat, nous avons étudié les différentes méthodes utilisées. Si les méthodes sont variées et les résultats similaires, cela confortera une nouvelle fois la robustesse des résultats. Ainsi, la section suivante s'attache à la description des méthodes recensées. Par la suite, le consensus entre les résultats issus des différentes méthodes a été examiné. Enfin, la représentativité locale de celui-ci a été éprouvée.

2.1 Comparaison des méthodologies utilisées dans différentes études relatives aux tendances d'éclairement

Deux étapes communes à l'ensemble des études analysées peuvent être discutées : l'analyse visuelle des données d'éclairement, et leurs analyses statistiques. L'analyse visuelle comprend dans la majorité des cas une méthode de prétraitement des données. Ce prétraitement est composé d'un filtrage des données sur une fenêtre temporelle glissante. Il permet de mettre visuellement en évidence d'éventuelles tendances de l'éclairement. L'analyse statistique, quant à elle, est composée d'un ajustement linéaire des données. La représentativité de cet ajustement est généralement quantifiée par un test statistique.

Au cours de notre étude, une grande diversité de méthodologies a pu être constatée. En particulier, trois points ont attiré notre attention : la diversité des méthodes et des fenêtres temporelles glissantes employées pour le prétraitement des données, ainsi que la diversité des tests statistiques utilisés pour estimer la représentativité de l'ajustement linéaire. Les informations relatives à ces trois points et issues d'un échantillon de la bibliographie étudiée sont regroupées dans le tableau 7.

Les méthodes de filtrage des données utilisées pour l'analyse visuelle sont diverses. Selon l'étude considérée, on peut référencer l'utilisation d'ajustements linéaires, de moyennes glissantes, d'ajustement de polynômes du 4^{ème} ordre aux données, de filtres passe-bas gaussiens, ou encore de filtres de Lanczos. Ces différentes méthodes de filtrage mettent en évidence des variations de long terme de l'éclairement. Dans toutes les études citées, ces variations sont définies par les auteurs comme des tendances d'éclairement. Rappelons que, comme introduit au cours du chapitre 1, compte tenu de la finitude de la période historique d'analyse, il est bien souvent impossible de dire si le signal observé correspond à une réelle tendance de long terme, ou à une partie d'un cycle plus long terme. Ce dernier apparaîtrait comme une tendance sur la courte durée d'étude considérée. Ainsi, il faut comprendre tendance au sens statistique sur l'échelle de temps considérée. Nous préférons parler de variabilité de long terme pour notre part. De plus, l'analyse d'un unique signal filtré peut introduire des biais d'analyse. Par exemple, un filtre passe bas peut lisser une anomalie mensuelle ponctuelle de telle façon qu'elle apparaisse comme une tendance de long terme dans le signal filtré. Ainsi, une analyse du signal à plusieurs échelles est nécessaire pour éviter des biais d'analyses.

ARTICLE	ANALYSE VISUELLE		ANALYSE STATISTIQUE	
	METHODE DE FILTRAGE	FENETRE TEMPORELLE GLISSANTE	PERIODE	TEST STAT
Dutton (2006)	Moyenne glissante	5 ans	1975-2005	t-test
	Ajustement de polynôme du 3 ^{ème} ordre	/		
Lohmann (2006)	/	/	1984-2001	t-test
Norris & Wild (2007)	Filtres passe-bas de Lanczos	5 ans	1965-2004	Incertitude des moindres carrés
Sanchez-Lorenzo <i>et al.</i> (2013)	Filtres passe-bas gaussiens	13 ans	1985-2010	Mann-Kendall
Wild <i>et al.</i> (2015)	Ajustements linéaires	/	2006-2049	t-test
Sanchez-Lorenzo <i>et al.</i> (2015)	Filtres passe-bas gaussiens	21 ans	Différentes sous-périodes entre 1961-2007	Mann-Kendall
Manara <i>et al.</i> (2016)	Filtres passe-bas gaussiens	11 ans	Différentes sous-périodes entre 1955-1995	Mann-Kendall
Sanchez-Lorenzo <i>et al.</i> (2017)	Filtres passe-bas gaussiens	1 ans 3 ans	1983-2010 et 1994-2010	Mann-Kendall
Wild <i>et al.</i> (2017)	Moyennes glissantes	5 ans	/	/
	Filtres passe-bas gaussiens	21 ans		
	Ajustement de polynôme du 4 ^{ème} ordre	/		
Pfeifroth <i>et al.</i> (2018)	Filtres passe-bas gaussiens	1 an	Différentes sous-périodes entre 1983-2015	Incertitude des moindres carrés

Tableau 7 : Résumé des méthodes de prétraitement des données et des fenêtres temporelles glissantes utilisées pour la représentation des variations de long terme dans différents articles de la bibliographie.

2. Analyse bibliographique relative aux tendances de l'éclaircissement

Ensuite, le tableau 7 montre que d'une étude à l'autre, les méthodes de prétraitement des données utilisent des fenêtres temporelles glissantes différentes, allant de 36 mois à 21 ans. Cependant, la largeur de la fenêtre glissante choisie pour le prétraitement des données influe sur l'information dégagée par celui-ci : par exemple, l'information mise à jour par une moyenne glissante sur 5 ans sera très différente de celle dégagée par une moyenne glissante sur 10 ans. De plus, plus la fenêtre glissante est large, plus l'influence d'effets de bords - inhérents aux méthodes de détection de variations de long terme - sera grande. Ces effets de bords peuvent causer des problèmes d'interprétation, de par la présence d'artefacts numériques ou statistiques dans le signal considéré. Ils ne sont malheureusement pas explicités dans la majorité des études citées dans le tableau 7.

Enfin, une diversité de tests statistiques a été constaté. Certaines études utilisent des tests paramétriques et asymptotiques, tels que le test de Student (TEUKOLSKY et al., 1992). Cependant, au vu du faible nombre de données historiques disponibles, l'utilisation de tels tests peut être discutée. À l'inverse, certaines études ont recours à des tests non paramétriques, tels que le test de Man-Kendall (KENDALL, 1975; MANN, 1945). Enfin, la dernière méthode répertoriée s'appuie sur l'incertitude calculée lors de la méthode des moindres carré (STORCH et al., 2000). Cette méthode consiste à comparer les variances entre la série temporelle observée et la série temporelle linéaire qui lui est associée. Si les variances sont proches (à un seuil de 95 %), alors l'ajustement linéaire est considéré de robuste. Ces tests sont appliqués sur des périodes temporelles variées. Certains auteurs soulignent la sensibilité des résultats des tests statistiques à la période considérée (PFEIFROTH, SANCHEZ-LORENZO, et al., 2018; SANCHEZ-LORENZO et al., 2017; SANCHEZ-LORENZO et al., 2015).

En conclusion, une diversité dans les méthodes de détection des variations de long terme est observée. Comme annoncé, cette diversité peut être vue comme un atout : si l'ensemble des résultats calculés à partir des méthodes diverses offrent un consensus, il y a de fortes chances que ce dernier soit robuste. À l'inverse, si les résultats divergent, il sera difficile pour un monteur de projet photovoltaïque de dégager une information adaptée à ses besoins. Comment savoir si les différences observées entre les études sont dues à une différence de méthode de prétraitement des données, de largeur de fenêtre utilisée, d'effets de bords, ou bien à une réelle différence entre les bases de données ?

Il sera alors utile d'adopter une approche guidée par des considérations physiques et/ou mathématiques pour le choix de méthodes de prétraitement des données, de la largeur de la fenêtre glissante utilisée, ou des tests statistiques employés. Cela permettrait de grandement améliorer la représentativité, la comparabilité, et l'interprétation des résultats entre les études (e.g. Huang Hoang, Parey, & Dacunha-Castelle, 2009).

Ainsi, après avoir comparé les méthodologies utilisées dans différentes études, nous nous intéressons dans la prochaine section à la comparaison des résultats issus de ces études.

2.2 Comparaison des résultats issus d'études relatives aux tendances d'éclairement

L'analyse bibliographique réalisée montre que pour certaines zones géographiques, les tendances d'éclairement font consensus, quand pour d'autres zones, il est difficile de trouver des résultats similaires entre les études.

Par exemple, il est clair que l'Est de l'Asie, avec en particulier la Chine et l'Inde, est une zone très documentée pour laquelle les résultats d'études variées font consensus (*e.g.* revues de Hayasaka, 2016- Chine et Japon; revue de WILD, 2009 - Global). Un obscurcissement global en accord avec les tendances mondiales est observé entre 1950 et 1990. Il est suivi pendant quelques années par un éclaircissement global. À partir des années 1995, une anomalie par rapport aux tendances mondiales est observée : l'éclaircissement global n'est plus détecté.

À l'inverse, des zones telles que l'Europe présentent des résultats beaucoup plus divergents entre les études considérées. Une tendance positive de l'éclairement relativement uniforme sur l'Europe (avec néanmoins une tendance plus forte sur l'Europe de l'Est) annoncée par de nombreuses études (NORRIS & WILD, 2007; WILD, 2009; OHMURA, 2009; CHIACCHIO & WILD, 2010; PHILIPONA et al., 2009; SANCHEZ-LORENZO et al., 2013; SANCHEZ-LORENZO et al., 2015; SANCHEZ-LORENZO et al., 2017; PFEIFROTH, SANCHEZ-LORENZO, et al., 2018). Cependant, la caractérisation de cette tendance positive varie en fonction de l'étude considérée. Par exemple, il est possible de mettre en évidence une disparité géographique de la tendance en fonction des études considérées.

À titre illustratif, la figure 22 rassemble les conclusions de l'étude menée par Sanchez-Lorenzo *et al.* (2017), quand la figure 23 présente les conclusions de l'étude de Pfeifroth *et al.* (2018). Ces figures illustrent les tendances d'éclairement observées par trois bases de données satellitaires : SARAH-2, CLARA-2, et une base de données dérivée de ces deux dernières. L'analyse est réalisée pour des périodes différentes : 1983-2015, 1983-2010, 1992-2015, 1994-2010. Les auteurs s'attachent à l'observation de la même quantité : la tendance de l'éclairement au sol ramenée à la décennie, exprimée en $W/m^2/décennie$. Ainsi, malgré la diversité des périodes temporelles observées, les chiffres présentés sur ces deux figures sont comparables, puisque ramenés sur la même période de temps. Autrement dit, il est possible de comparer les résultats issus de la même base de données pour des périodes temporelles différentes (*cf.* figure 22), comme les résultats calculés sur la même période temporelle pour des bases de données différentes (*cf.* figure 23).

2. Analyse bibliographique relative aux tendances de l'éclairement

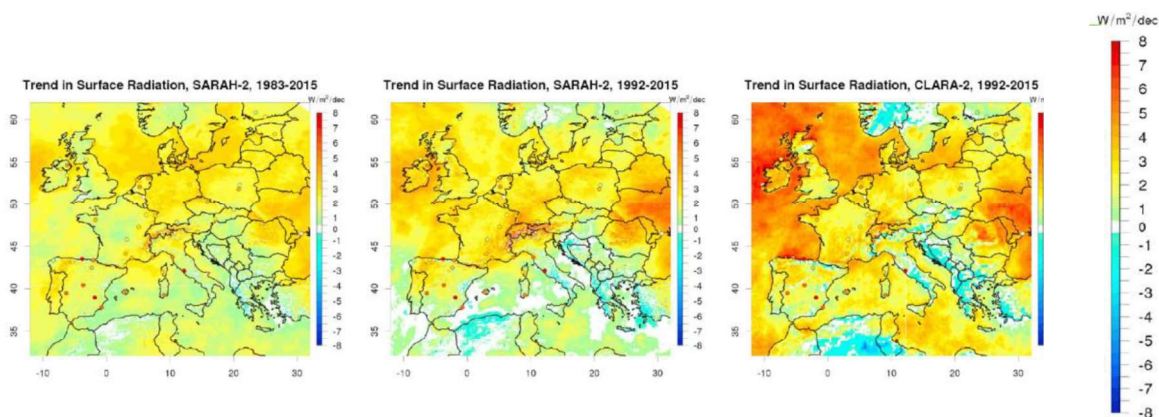


Figure 22 : Tendances d'éclairement sur l'Europe dérivée des données SARAH-2 et CLARA-2, 1983-2015 et 1992-2015. Figure extraite de Pfeifroth et al. (2018).

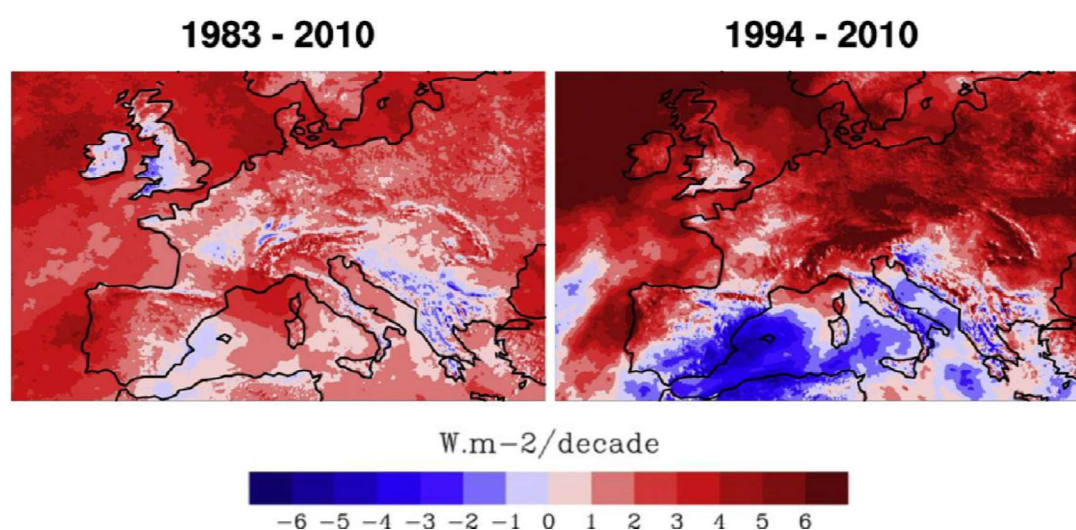


Figure 23 : Tendances d'éclairement sur l'Europe dérivée des données CM-SAF, 1983-2010 et 1994-2010. Figure extraite de Sanchez-Lorenzo et al. (2017)

On peut constater des différences notables entre les tendances reproduites par les trois bases de données satellitaires étudiées par les auteurs. On peut par exemple citer une augmentation de l'éclairement plus marquée à partir du milieu des années 1990 pour Sanchez-Lorenzo *et al.* (2017) (*cf.* figure 23), ce qui n'est pas forcément évident dans Pfeifroth *et al.* (2018) (*cf.* figure 22). Les tendances d'éclairement sur l'Ukraine sont représentées de la même manière pour l'ensemble des périodes et des bases de données, sauf pour la période 1994-2010 de Sanchez-Lorenzo *et al.* (2017). Les tendances observées sur l'Espagne et sur l'Angleterre varient fortement en fonction des bases de données, avec à l'extrême des tendances négatives de $4 \text{ W/m}^2/\text{décennie}$ entre 1994-2010 pour le Sud de l'Espagne, contre une absence de tendance notable dans les autres bases de données. De plus, de larges différences sont observables sur les zones maritimes et océaniques, avec des différences de l'ordre de $4 \text{ W/m}^2/\text{décennie}$ entre SARAH-2 et CLARA-2. Enfin, on peut souligner que pour ces deux études, l'ensemble des tendances semblent significatives. Comme vu dans le tableau 7, la significativité des tendances est estimée par deux méthodes différentes dans les études : le test statistique de Mann-Kendall, et la méthode des moindres carrés.

Cette divergence de résultats est extrêmement complexe à analyser. En effet, les différences de valeurs et de répartition géographique de ces valeurs pourraient notamment être dues à une différence dans :

- les méthodes algorithmiques de calcul des données d'éclairement à partir d'images satellites utilisées dans les trois bases de données.
- les données analysées : les uns s'attachent aux données d'éclairement quand les autres étudient des anomalies mensuelles de l'éclairement
- les largeurs des fenêtres temporelles considérées. Lors des représentations illustratives de séries temporelles dans les deux articles, Pfeifroth *et al.* (2018) utilisent un filtre gaussien de 12 mois, quand Sanchez-Lorenzo *et al.* (2017) utilisent un filtre gaussien de 36 mois.
- les périodes temporelles comparées. L'éclaircissement global et l'obscurcissement global sont fortement dépendants de la période considérée. Cela est illustré par la figure 23. Pour une même étude, donc faites par les mêmes auteurs, avec les mêmes méthodes et le même type de données, le résultat est extrêmement sensible à la période temporelle observée. Pour n'en citer qu'un exemple, on peut observer entre les deux représentations de la figure des différences de l'ordre de 5 W/m²/décennie sur l'ouest de la mer méditerranée.

Pointer parmi ces paramètres ceux qui sont à l'origine des différences observées n'est évident. Des analyses complémentaires à partir des données sources de chacune des études est nécessaire. Ce type d'analyse scientifique poussée n'est pas compatible avec les contraintes industrielles d'un développeur de projet photovoltaïque. **En cas de divergence de l'information disponible dans la bibliographie, la diversité des méthodes est donc un frein à l'utilisation de celle-ci pour une application industrielle. Une harmonisation des méthodes serait souhaitable.** Cependant, la diversité de méthodes est intéressante – et même souhaitable - en science, afin de mettre en lumière les limitations de chacune d'entre elles. **Une alternative possible pourrait être l'utilisation d'une méthode permettant de synthétiser l'ensemble de l'information bibliographique obtenue à l'aide de méthodologies variées.**

Cet exemple illustratif peut être étendu pour la majorité des régions mondiales. Nous avons néanmoins pu identifier quelques régions dont les résultats font consensus dans l'ensemble des études, appelées zones de consensus bibliographique dans la suite. Pour ces zones, il pourrait être possible d'utiliser l'information disponible dans la bibliographie pour estimer les variations de long terme de l'éclairement dans le passé. Cependant, il est possible que des structures spatiales fines puissent entraîner des différences entre les résultats décrit sur une zone et la mesure de l'éclairement en un point. Dans cette optique, les résultats de la bibliographie ont été confrontés à l'analyse de mesures locales pour une zone illustrative : la Chine.

3 Comparaison des résultats issus de la bibliographie à l'analyse de mesures locales en Chine

Pour notre étude, les mesures *in-situ* sont les stations GEBA décrites en chapitre 2. **Comme pour l'essentiel des études recensées, notre analyse se composera d'une analyse visuelle de quelques stations GEBA et d'une analyse statistique de ces mêmes stations.** Cette analyse a été réalisée pour l'ensemble des 104 stations GEBA décrite en section 3 du chapitre 2. Nous avons ici choisi de ne présenter que le cas illustratif de la région chinoise. Cette région a retenu notre attention car elle est une zone de consensus bibliographique, est très documentée, et est couverte par une dizaine de stations GEBA.

3.1 Résumé du consensus bibliographique sur la Chine

Une revue très complète portant sur les variations de long terme de l'éclairement en Chine est réalisée par Hayazaka (2016). Rassemblant près d'une centaine d'études, ce travail met en évidence le consensus suivant : sur la région chinoise, l'éclairement diminue entre les années 1950 et 1990, augmente entre les années 1990 et 1995, puis stabilise jusqu'à aujourd'hui. Il n'y a donc pas, pour cette région, d'éclaircissement global. Pour expliquer cela, l'hypothèse principale réside dans le lien entre les variations de la quantité d'aérosols dans l'atmosphère et les variations de l'éclairement. En particulier, le fort taux d'émission d'aérosols dans l'atmosphère par la Chine ces dernières décennies pourrait expliquer la stabilisation de l'éclairement après les années 1995 (HAYASAKA, 2016; MISHCHENKO *et al.*, 2007; NORRIS & WILD, 2007; RUCKSTUHL *et al.*, 2008; STANHILL & COHEN, 2001; STREETS *et al.*, 2006).

Le travail de revue de Hayazaka (2016) souligne néanmoins une disparité des tendances observées au sein de la région. Par exemple, l'anomalie à l'éclaircissement global est beaucoup plus importante dans le Nord de la Chine que dans le Sud. Autrement dit, l'éclairement continue d'augmenter dans les régions du Sud de la Chine, alors que celui-ci diminue fortement dans les régions du Nord. D'après les études répertoriées par Hayazaka (2016), cela s'explique aisément par le nombre beaucoup plus conséquent de villes, et donc d'émission d'aérosols, dans la région Nord de la Chine. L'auteur souligne qu'une grande diversité en termes de variabilité d'éclairement peut être observée en fonction de la proximité de la station de mesure à une ville. Cependant, la région étant bien documentée, il est possible de trouver de l'information pour des zones géographiques assez restreintes, de l'ordre de quelques degrés.

3.2 Analyse visuelle des stations GEBA au regard de la bibliographie

L'analyse visuelle est réalisée pour quatre stations GEBA du Nord de la Chine. Ces stations sont situées à des latitudes similaires. La figure 24 montre en étoiles de couleurs la localisation géographique de ces quatre stations GEBA. L'analyse visuelle est réalisée sur les moyennes glissantes annuelles de l'éclairement. Les moyennes annuelles sont calculées à partir de séries temporelles mensuelles complétées, décrites au cours du chapitre 2. À droite de la figure 24 se trouvent les quatre séries temporelles correspondantes.

L'information fournie par la bibliographie sur cette région restreinte est similaire à celle sur la région chinoise : l'éclairement diminue entre les années 1950 et 1990, puis se stabilise.

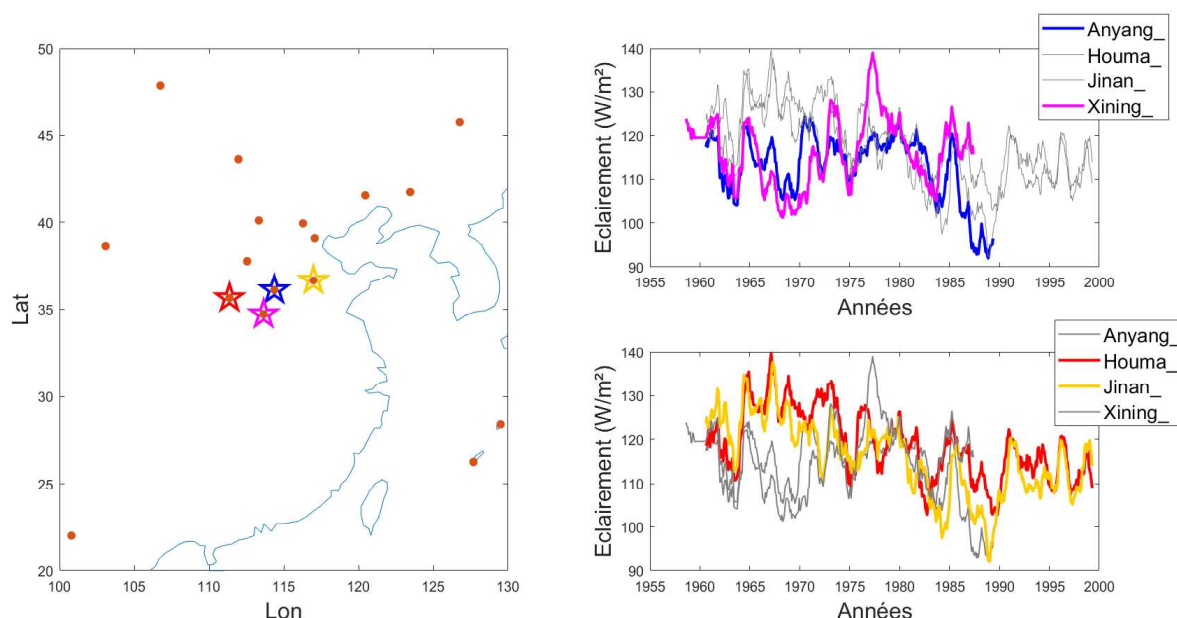


Figure 24 : Répartition géographique des stations GEBA sélectionnées pour l'étude sur la zone d'intérêt (droite). Les étoiles de couleurs mettent en évidence les stations dont les séries temporelles d'éclairement correspondantes sont représentées à gauche.

Une diversité de comportements entre les stations est observée : deux stations correspondent à des éclaircissements annuels relativement constants entre 1950 et 1990 (rose et bleu) tandis que deux stations correspondent à un éclaircissement nettement à la baisse entre 1965 et 1990 (jaune et rouge). Malgré cette diversité, un ensemble d'événements communs à l'ensemble des stations nous confortent sur la qualité des mesures observées. Par exemple, une diminution de l'éclairement est visible pour les quatre stations en 1963, en 1974, ou encore en 1982.

Plus précisément, les deux stations Houma et Jinan sont très proches. Elles suivent globalement les tendances décrites par la bibliographie à partir de 1965 (rose et bleu) : une diminution de l'éclairement jusque 1990, suivi d'une stabilisation de ce dernier. Cependant, la tendance à la diminution de l'éclairement entre 1950 et 1960 n'est pas visuellement perceptible. Pour la station

3. Comparaison des résultats issus de la bibliographie à l'analyse de mesures locales en Chine

Xining, le rayonnement est globalement constant sur l'ensemble de la série temporelle, avec une grande amplitude : on peut compter près de 40 W/m^2 entre le maximum et le minimum d'éclairement pour cette station. L'éclairement pour la station Anyang est également constant jusque 1980. Après cette date, une nette diminution de l'éclairement est perceptible, de plusieurs dizaines de W/m^2 .

Les différences en termes de tendance d'éclairement observées sur les quatre stations ne peuvent pas être expliquées par la proximité des stations à la ville de Pékin. En effet, les deux villes les plus proches géographiquement sont deux stations montrant des comportements différents : les stations Anyang (bleu) et Jinan (jaune).

En conclusion, nous pouvons avancer que même sur des zones géographiques restreintes, pour lesquelles la bibliographie fait consensus, différentes variations inter-annuelles peuvent être observées visuellement en fonction des stations considérées. Même si une tendance régionale claire est identifiée dans les études, des structures spatiales fines font que ces tendances peuvent ne pas se retrouver localement dans la région considérée. Ainsi, malgré la qualité des études portant sur la caractérisation des variations long terme de l'éclairement, une analyse à partir des séries temporelles propre au site choisi semble nécessaire. Une analyse courante visant à identifier les variations de long terme est l'application de tests statistiques de stationnarité.

3.3 Analyse statistique des stations GEBA au regard de la bibliographie

Pour rappel, un test statistique est une procédure de décision statistique permettant de valider ou de rejeter une hypothèse, appelée hypothèse nulle, en fonction d'un échantillon de données. Dans notre cas, l'hypothèse nulle H_0 est la stationnarité de l'éclairement. L'hypothèse nulle, supposée vraie dans un premier temps, est rejetée si la statistique propre au test considéré ne satisfait pas une condition préalablement requise. Ici, H_0 est ici rejetée si les données ne peuvent pas être considérées comme stationnaires. La condition préalablement requise est définie en fonction d'un seuil de significativité, classiquement considéré à 95 %. Soulignons enfin que ne pas rejeter l'hypothèse nulle ne peut être considéré comme une preuve de la véracité de celle-ci. Cela ne peut être interprété que comme une absence de preuve suffisante pour valider H_0 . Autrement dit, ce n'est pas parce que le test ne détecte pas de non-stationnarité que l'éclairement est nécessairement stationnaire. Des tests statistiques complémentaires doivent être réalisés pour démontrer la stationnarité temporelle de l'éclairement. Ainsi, il a été choisi de réaliser simultanément des tests de stationnarité et de non-stationnarité. Cette analyse conjointe augmente la significativité de nos résultats malgré le faible échantillon de données d'origine.

Une analyse bibliographique des tests statistiques existants a été réalisée. Le détail de cette analyse est disponible en annexe IX. En résumé, les tests paramétriques (*i.e.* supposant que les données considérées suivent une loi de distribution connue, en général la loi normale) ont été écartés. Comme annoncé précédemment, cela s'explique par le manque de robustesse de l'ajustement d'une distribution à un échantillon de 10 à 30 valeurs, démontré au cours du chapitre 3. Le faible nombre de données disponibles justifie également que les tests asymptotiques aient été écartés. Pour augmenter la robustesse des conclusions quant à la stationnarité temporelle de l'éclairement, nous avons choisi

d'utiliser conjointement différents tests statistiques. Les tests choisis correspondent à des tests non-paramétriques simples et de références, construits avec des méthodologies diverses. Les trois tests retenus sont : les tests de non-stationnarité de Mann-Kendall (MK) (KENDALL, 1975; MANN, 1945) et de KPSS (Kwiatkowski, Phillips, Schmidt, and Shin) (SHIN & SCHMIDT, 1992), et le test de stationnarité de Phillips-Perron (PP) (PHILLIPS & PERRON, 1988).

Le test KPSS décompose le signal en la somme d'une tendance linéaire, d'une marche aléatoire et d'un terme d'erreur stationnaire. En d'autres termes, il évalue par défaut la tendance linéaire associée aux données d'entrée, la retranche, puis cherche la présence d'une marche aléatoire dans la série temporelle. Pour notre étude, nous nous sommes intéressés à la valeur du coefficient de tendance calculé par le modèle. Si celui-ci est inférieur à 10^{-3} , il pourra être conclu que la série temporelle n'est pas soumise à une tendance linéaire significative, et pourra être considérée comme stationnaire. L'information sur la présence d'une marche aléatoire ne sera pas analysée ici.

Le test MK compare chaque point avec l'ensemble des points précédant celui-ci dans la série temporelle. Si la différence est positive (resp. négative), une somme cumulative est incrémentée positivement (resp. négativement). En fonction de la valeur finale de la somme cumulative, il est possible de conclure quant à la présence d'une tendance positive dans le signal original (resp. négative).

Le test PP correspond à une version non paramétrique du test de Dickey-Fuller. Il teste l'adéquation entre le signal original et trois modèles construits autour de quatre paramètres : le terme précédent dans la série temporelle, une constante, une dérive temporelle et un terme d'erreur. Ces trois tests sont également présentés plus en détail en annexe IX.

Afin d'augmenter le nombre de données utilisées dans l'estimation des statistiques de chaque test, et donc la significativité des résultats issus des tests statistiques présentés, les données mensuelles et non annuelles des 104 stations GEBA ont été considérées. Pour éviter que les variations intra-annuelles n'affectent les résultats des tests statistiques utilisés, il est d'usage de minimiser l'influence du cycle saisonnier sur le signal. Pour cela, deux principales approches existent : traiter de manière différentielle les saisons, ou ôter la partie saisonnière du signal. Étant plus intéressés par les variations de long terme de l'éclairement que par les tendances saisonnières, nous avons choisi la deuxième approche.

La famille de méthodes la plus classique pour ôter le cycle saisonnier du signal consiste à s'affranchir de l'influence de la position et de la déclinaison du Soleil en calculant un indice de clarté (KT). La méthode utilisée pour calculer le KT pour les stations GEBA est détaillée en section 2.5 du chapitre 2. Pour rappel, le KT correspond au ratio entre la quantité de rayonnement reçu au sol et la quantité correspondante de rayonnement reçu au sommet de l'atmosphère.

Les trois tests statistiques introduits précédemment (PP, KPSS et MK) ont été appliqués sur les séries temporelles mensuelles d'indice de clarté. La figure 25 illustre les résultats de ces tests pour

3. Comparaison des résultats issus de la bibliographie à l'analyse de mesures locales en Chine

les quatre stations chinoises. Les résultats pour l'ensemble des 104 stations GEBA sont présentés en annexe X. En bleu sont représentées les stations dont l'indice de clarté est non stationnaire, et en vert les stations dont l'indice de clarté est stationnaire. Les couleurs foncées montrent quand l'hypothèse nulle est rejetée : on peut affirmer la stationnarité ou la non-stationnarité de la série temporelle. Les couleurs claires illustrent quand l'hypothèse nulle ne peut pas être rejetée : la stationnarité ou non stationnarité est supposé, mais non démontrée. Pour rappel, les tests de KPSS et de MK ont comme hypothèse nulle la stationnarité de l'indice de clarté. Le test de PP a pour hypothèse nulle la non-stationnarité de l'indice de clarté. Pour que les résultats soient cohérents, il faut que l'ensemble des tests aient la même couleur.

Station Test statistique	Anyang (station n°6)	Houma (station n°35)	Jinan (station n°37)	Xining (station n°101)
MK				
KPSS				
PP				

Figure 25 : Résultat des tests statistiques de stationnarité de Phillips-Peiron (PP), de non-stationnarité Kwiatkowski-Phillips-Schmid-Shin (KPSS) et Mann-Kendall (MK), appliqués à l'indice de clarté des quatre stations chinoise sélectionnées. Les couleurs représentent la stationnarité (vert) ou la non-stationnarité (bleu). L'intensité de la couleur représente le rejet de H_0 (foncé), ou l'absence de rejet (clair).

La figure 25 met en évidence que le test de KPSS ne détecte pas de tendance linéaire dans les quatre séries temporelles considérées. Cela se retrouve pour l'ensemble des 104 stations GEBA. À l'inverse, le test de MK détecte une non stationnarité pour les quatre stations d'intérêt. Ceci n'est pas le cas pour l'ensemble des 104 stations. **Pour les quatre stations de la région chinoise, une incompatibilité de ces deux tests est donc visible : les séries temporelles sont déclarées simultanément stationnaires et non stationnaires.** Les résultats pour les tests de PP et de MK ne sont pas en accord pour l'une des quatre stations : la station Houma. Visuellement, cette station semblait soumise à une diminution de l'éclairement entre 1965 et 2000. De même, les variations observées visuellement sur la station Jinan ne sont pas détectés par le test de PP. Les tendances annoncées par la bibliographie sur cette zone ne se retrouvent donc pas dans l'analyse statistique des stations GEBA.

On peut s'interroger sur les causes des divergences de résultats entre les trois tests utilisés. Celles-ci peuvent notamment venir de différences dans :

- la définition de la stationnarité temporelle, qui est propre à chaque test.
- le nombre de données considéré, qui est propre à chaque station GEBA.
- la période temporelle considérée, qui elle aussi diffère entre les stations. En ce sens, il est possible que de ne pas avoir subdivisé la période temporelle d'analyse ait limité la détection de tendances par les tests statistiques. Par exemple, pour la station Jinan, des tendances n'étaient visibles qu'à partir de 1965. Si la période d'analyse était restreinte

entre 1965 et 1990, il est possible que le test de PP ait détecté une non stationnarité. On touche ici à une limite des tests statistiques : ceux-ci sont sensibles à la période considérée. Il serait souhaitable d'utiliser une méthode de détection des variations de long terme qui conserve l'information temporelle. Ainsi, on pourrait détecter d'éventuelles tendances advenant que sur une sous période de l'échantillon de données considéré.

- la puissance du test. En effet, la probabilité de rejeter l'hypothèse nulle alors que celle-ci est correcte, et inversement, menant à des faux positifs ou à des faux négatifs, est différente pour les trois tests. Elle est d'autant plus importante que le nombre de données est faible. En effet, avec un petit échantillon de données, le rapport signal sur bruit est difficile à estimer.

En conclusion, les résultats obtenus avec les trois tests choisis divergent pour les quatre stations considérées. Cela est également vrai à l'échelle globale. En effet, le même exercice que celui décrit ci-dessus a été réalisé pour l'ensemble des stations GEBA. Sur les 104 stations GEBA, seulement 8 stations sont consistantes entre les trois tests utilisés. Les mesures de ces stations peuvent être déclarées stationnaires. **Autrement dit, pour 92 % des stations, les séries temporelles d'éclairement ne peuvent être déclarées ni stationnaire, ni non stationnaire. La répartition spatiale des stations trouvées stationnaires ne montre pas de structure spatiale claire. Ainsi, n'étant pas pour le moment en mesure de comprendre la distribution de ces stations, il paraît peu robuste de poser a priori une hypothèse de stationnarité de façon systématique lors de l'étude du potentiel solaire.**

Pour aller plus loin, il serait possible d'approfondir l'analyse statistique réalisée ici. Par exemple, des méthodes de détection de tendances plus poussées peuvent être considérées (*e.g.* HOANG *et al.*, 2009; PAREY *et al.*, 2018). Cependant, les résultats issus de ces tests statistiques, aussi précis soient-ils, donnent peu d'informations quant à la caractérisation des variations inter-annuelles de l'éclairement. En effet, ils indiquent si l'éclairement peut être considéré comme stationnaire ou non, donc s'il existe des tendances supérieures à la période d'analyse. Or, il est possible que des variations du signal inférieures à la période d'analyse soient indispensables pour représenter l'ensemble du signal. De même, la caractérisation de variations de plus long terme que la période d'analyse peut être indispensable dans un contexte prospectif, comme abordé au cours du chapitre 3. **Ainsi, une connaissance fine de l'ensemble des modes de variabilité peut être très utile - qu'ils soient de quelques années, décennaux ou multi-décennaux - pour une bonne compréhension du signal d'éclairement. Cette connaissance des variations multi-échelles de l'éclairement ne peut être établie à l'aide de tests statistique de stationnarité.**

4 Conclusion

Une étude bibliographique relative aux tendances de long terme de l'éclairement a été réalisée. Cette étude a permis de mettre en évidence différentes méthodes utilisées dans un ensemble de travaux. Une diversité des résultats issus de ces travaux a également pu être constatée. Le manque d'homogénéité des méthodes entraîne une difficulté à identifier les paramètres à l'origine des différences observées. Il est alors apparu que l'information dégagée par la bibliographie ne permet pas de caractériser les variations de long terme de l'éclairement de façon suffisamment fine pour une application industrielle. Nous avons conclu quant à l'utilité d'une méthode permettant de synthétiser l'ensemble de l'information bibliographique obtenue à l'aide des méthodologies variées. En ce sens, une méthode d'analyse multi-échelle permettrait de couvrir l'ensemble des fenêtres temporelles glissantes recensées dans la bibliographie.

Pour quelques régions, un consensus entre les différentes études a pu être relevé. Cependant, nous avons montré que, sur ces zones, les mesures en un point particulier peuvent être différentes de l'information fournie par ce consensus. Cela a été illustré pour le cas de la région chinoise par une analyse visuelle et statistique de stations GEBA. À nouveau, les limitations de l'information disponible dans un contexte d'aide à la décision ont été mises en évidence. Les limites des méthodes statistiques de détection des tendances d'éclairement ont également été soulignées. Nous avons conclu quant à l'intérêt d'une méthode permettant de caractériser les variations de long terme de l'éclairement, que ces variations soient multi-annuelles, décennales, ou de plus long terme. L'intérêt de conserver l'information temporelle du signal a également été souligné. Celle peut notamment permettre d'identifier des variabilités qui ne sont pas nécessairement présentes sur l'ensemble de la période temporelle étudiée.

Cela nous amène à envisager une méthode d'analyse des séries temporelles d'éclairement qui soit temporelle et multi-échelle. Cet outil permettrait de synthétiser l'information disponible dans la bibliographie, tout en caractérisant les modes de variabilités inter-annuels de l'éclairement.

Chapitre 5

Description de l’outil de décomposition temps-fréquence

Table des matières

Description de l’outil de décomposition temps-fréquence choisi	95
1. Introduction	97
2. Caractérisation d’une méthode de décomposition temps-fréquence.....	97
2.1 Choix d’une méthode.....	97
2.2 Description de la méthode.....	98
3. Validation de la méthode de décomposition temps-fréquence	105
3.1 Validation de la méthode sur la base d’un signal synthétique	105
3.2 Validation de la méthode au regard de la bibliographie	109
4. Conclusion.....	113

1 Introduction

Le chapitre précédent a mis en évidence une réelle difficulté à conclure quant à la stationnarité temporelle de l'éclairement. Pour aller plus loin, l'intérêt de l'utilisation d'une méthode d'analyse temporelle et multi-échelle est apparu. Celle-ci permettrait en effet de :

- **synthétiser l'information disponible dans la bibliographie** (aspect multi-échelle).
- **détecter des changements subvenant uniquement sur une partie de la période temporelle analysée** (aspect temporel).
- **détecter la présence d'anomalies ponctuelles**, qui pourraient avoir une forte influence sur l'analyse des variations de long terme par une méthode de filtrage (aspect temporel).
- **étudier simultanément l'ensemble des modes de variabilité de l'éclairement**. Ainsi, une compréhension fine des variations de l'ordre de quelques années, de quelques décennies, ou plus, sera possible (aspect multi-échelle). Cette compréhension est une étape importante vers la prise en compte des variations inter-annuelles de l'éclairement dans l'estimation du potentiel solaire.

Ce chapitre décrit en détail (section 2) et valide (section 3) la méthode d'analyse temporelle et multi-échelle choisie : une décomposition temps-fréquence des séries temporelles mensuelles.

2 Description de la méthode de décomposition temps-fréquence

2.1 Attentes vis-à-vis de la méthode

L'évaluation des performances des méthodes de décomposition temps-fréquence n'étant pas l'objectif de ce travail de thèse, notre choix a été guidé par la recherche de transparence et de simplicité de la méthode considérée. Des études utilisant des outils de traitement du signal plus élaborées pourraient être envisagées dans le futur. Celles-ci permettraient d'affiner les résultats proposés par la suite, sans pour autant remettre en question l'analyse mise en œuvre. **Pour assurer transparence et simplicité, deux contraintes ont été ajoutées aux motivations présentées en introduction.**

Premièrement, les éléments issus de deux décompositions temps-fréquence doivent être comparables. Ainsi, les méthodes de décomposition adaptatives aux données, telles que la décomposition en modes empiriques (EMD, *empirical mode decomposition*) (BENGULESCU *et al.*, 2016), ont été exclues. En effet, ces méthodes identifient les échelles caractéristiques propres à chaque signal à partir de la série temporelle initiale. Les échelles estimées ne sont donc pas identiques entre deux décompositions. La méthode choisie ici devra permettre de définir les échelles de variations étudiées comme paramètres d'entrées.

Ensuite, la méthode choisie doit être quasi-orthogonale. Autrement dit, il doit être possible de reconstruire parfaitement le signal d'origine à partir des éléments décomposés. Ainsi, il est possible de traiter les signaux de sortie indépendamment ou de façon conjointe. Par exemple, les variations intra-annuelles seront dans un premier temps traitées selon plusieurs échelles, puis seront fréquemment assemblées.

La méthode de décomposition temps-fréquence choisie est construite à partir d'un banc de filtres. L'adéquation de la méthode aux contraintes et motivations présentées précédemment sera développée et illustrée tout au long de ce chapitre.

2.2 Description de la méthode

La méthode sélectionnée peut se résumer en trois étapes, reproduites neuf fois. Il en résulte neuf séries temporelles filtrées, plus un résidu. Ces dix séries temporelles issues de la décomposition seront par la suite nommées éléments. Les étapes de la décomposition sont illustrées schématiquement dans la figure 26, pour les données d'éclairage de la station chinoise Houma. Détaillons les étapes pour la construction des deux premiers éléments de sortie du filtre, illustrées sur les deux premières lignes de la figure 26.

(1) **Une convolution est réalisée entre un signal original, noté y_{or} , et un sinus cardinal.** Pour assurer un bon compromis entre l'information temporelle et fréquentielle, le sinus cardinal est tronqué, puis apodisé, en utilisant la fonction d'apodisation de Blackman (OPPENHEIM et al., 1999). Les paramètres du sinus cardinal apodisé varient en fonction du filtre considéré. Ils sont définis par un noyau de convolution.

Le noyau de convolution est défini de telle sorte que chaque élément en sortie du filtre corresponde à une plage de périodes centrée autour des périodes particulières de : 3 mois (cycle saisonnier), 6 mois (échelle semestrielle), 12 mois (cycle annuel), 24 mois (échelle biannuelle), 36 mois, 72 mois (6 ans), 120 mois (échelle décennale), 180 mois (15 ans), 500 mois (près de 40 ans). Ces éléments issus de la décomposition seront notés respectivement $y_3, y_6, y_{12}, \dots, y_{500}$. Ainsi, le sinus cardinal apodisé est défini de façon à ce que le signal filtré, noté y_{fil} , corresponde à l'ensemble des variations supérieures à l'échelle considérée : $y_{fil,3}$ correspond à l'ensemble variations supérieur à 3 mois, $y_{fil,6}$ à l'ensemble des variations supérieures à 6 mois pour y_6 , etc.

Pour la première série temporelle, y_{or} correspond à la série temporelle originale du signal, et $y_{fil,3}$ à la première série temporelle filtrée.

(2) **Le signal filtré $y_{fil,3}$ est soustrait au signal original y_{or} .** Ainsi, le premier élément y_3 est composé uniquement des plus courtes plages de périodes, ici centrées en 3 mois.

(3) **Le signal filtré constitue le prochain signal filtré.** Autrement dit, pour la décomposition suivante, $y_{or} = y_{fil}$.

On retrouve ces mêmes étapes pour le second élément :

(1) Une nouvelle convolution entre le signal résiduel $y_{fil,3}$ et un second sinus cardinal apodisé est réalisé. Cette convolution filtre les données de façon à n'avoir plus que les périodes supérieures au second élément du noyau de convolution : ici 6 mois. On obtient donc $y_{fil,6}$.

(2) $y_{fil,6}$ est alors soustrait au premier signal filtré $y_{fil,3}$. Autrement dit, on enlève l'ensemble des plages de périodes supérieures à 6 mois à un signal composé de l'ensemble des périodes supérieures à 3 mois. Il en résulte le second élément de la décomposition, y_6 , composé de l'ensemble des plages de périodes comprises entre 3 et 6 mois.

(3) À nouveau, $y_{or} = y_{fil,6}$.

Ces trois étapes sont reproduites jusqu'au dernier élément du noyau de convolution. À l'issue de la décomposition temps-fréquence, dix éléments sont donc constitués. Le dixième élément, y_{residu} , représente l'ensemble des périodes connaissant des variations supérieures à 500 mois. Cet élément, comme y_{500} , rassemble les variations de très long terme du signal. Ils ne seront exploités que pour les données issues des modèles climatiques : ces dernières possèdent un historique assez long pour pouvoir fournir une information de qualité à ces échelles.

2. Description de la méthode de décomposition temps-fréquence

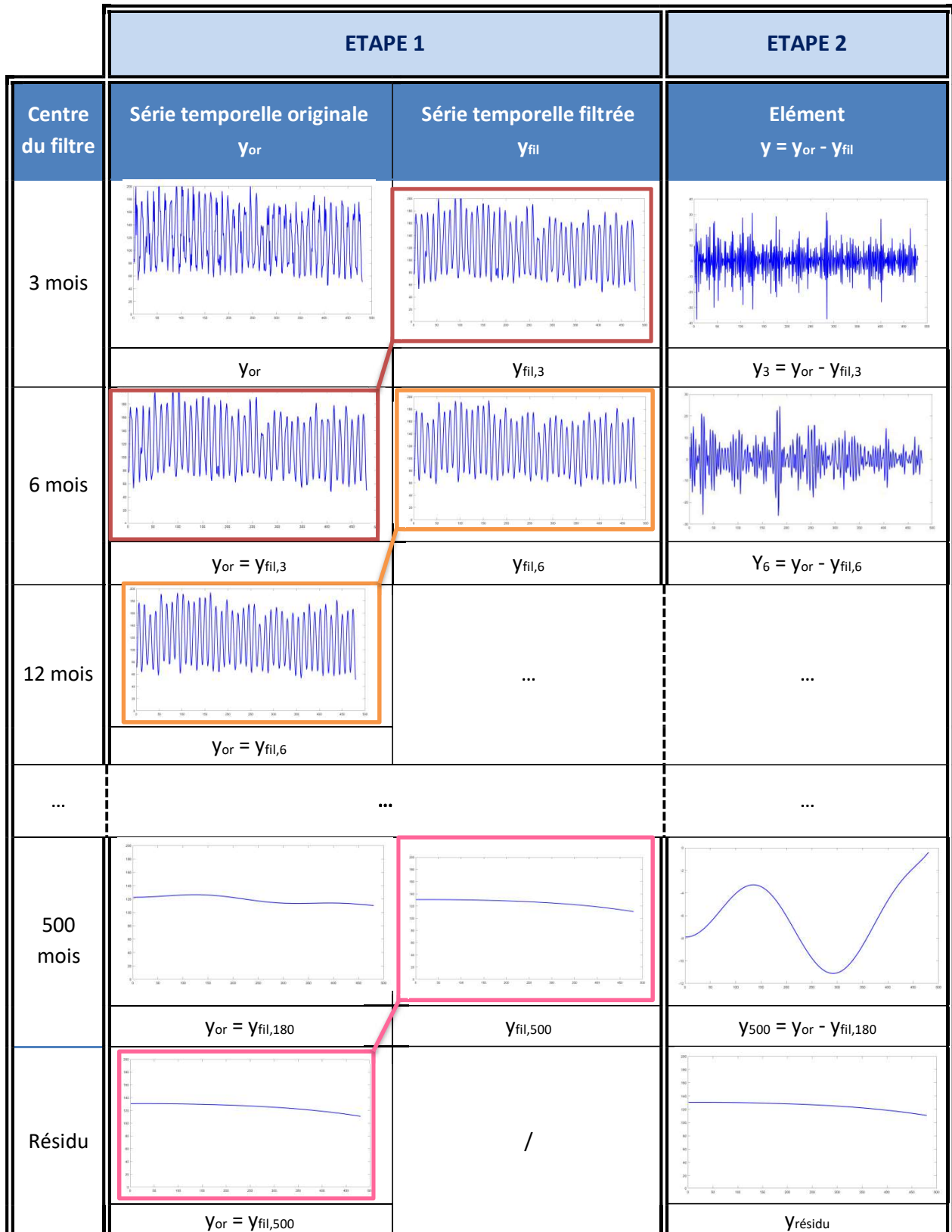


Figure 26 : Représentation schématique de la décomposition temps fréquence réalisée.

Les plages de périodes sélectionnées pour chaque élément sont illustrées par la figure 27. Cette dernière représente les modules de la transformée de Fourier de la réponse des différents filtres soumis à un même Dirac, aussi appelés fonctions de transfert de modulation.

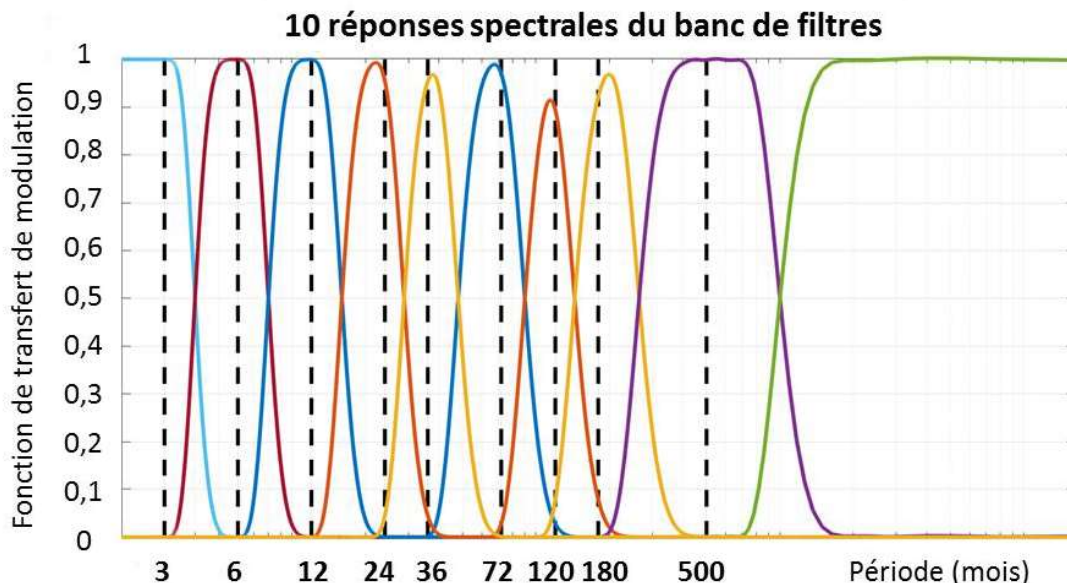


Figure 27 : Fonctions de transfert de modulation exprimées en périodes du banc de filtres utilisé.

On peut noter sur la figure 27 que le banc de filtre choisi offre des plages de périodes, et non une période particulière. Les neuf premiers éléments sont centrés sur les périodes citées précédemment. Le dixième élément, représenté en vert, représente l'ensemble des périodes supérieures à 500 mois.

Ces plages de périodes ne sont pas parfaitement rectangulaires. Comme annoncé, cela est dû à l'apodisation du sinus cardinal, assurant un compromis entre la précision temporelle et fréquentielle du banc de filtres choisi. Par conséquent, les pages de périodes se recoupent partiellement : par exemple, une partie du signal variant autour d'une période de 40 mois va se retrouver en partie dans l'élément centré sur 36 mois (jaune), et en partie dans l'élément centré sur 72 mois (bleu). Ainsi, si le signal connaît de fortes variations autour de 40 mois, les éléments centrés en 36 et 72 mois seront redondants. L'orthogonalité de la méthode de décomposition temps-fréquence sera alors remise en cause.

Cela souligne l'importance de choisir de façon précise le centre des plages de périodes : si les périodes caractéristiques de variations du signal se situent au centre de chaque plage de périodes, alors l'information redondante sera minimisée, et la décomposition sera quasi orthogonale, voire orthogonale. Dans notre cas, le centre des périodes a été sélectionné avec attention à partir de l'étude de Moron *et al.* (1998). Cette étude identifie des variations climatiques pour trois périodes caractéristiques : 72, 120 et 168 mois. Les auteurs mettent également en évidence des variations biannuelles. Ainsi, en centrant les plages de périodes sur ces valeurs, nous maximisons les chances que l'essentiel de l'information se trouve proche de la période centrale. Un mode de variabilité à 36 mois a été ajouté. Pour des raisons physiques évidentes, les variations intra-annuelles ont également été décomposées en variations saisonnières, biannuelles et annuelles.

D'autres limites de la décomposition temps-fréquence choisie sont à noter, à savoir :

- (i) la décomposition utilisant une convolution, celle-ci nécessite un traitement particulier relatif aux données manquantes. En effet, la présence de données manquantes va avoir des conséquences sur la décomposition du signal. Pour notre application simplement temporelle des filtres de convolution, les valeurs manquantes vont se retrouver étendues de part et d'autre des éléments, sur une portée liée au à la largeur des filtres correspondants. C'est pourquoi les données étudiées ont été complétées. Cette complétion est décrite au cours du chapitre 2.
- (ii) les données situées aux extrémités de la série temporelle seront altérées lors de la convolution par des effets de bord, liés la encore à la largeur des différents filtres appliqués en cascade. L'extension temporelle de la partie altérée sera d'autant plus importante que la largeur du sinus cardinal sera grande. À l'inverse, plus le sinus cardinal est apodisé, plus il est court, et plus le nombre d'élément valide sera important.

La figure 28 représente de manière schématique l'influence de la largeur du sinus cardinal sur l'extension temporelle des effets de bords.

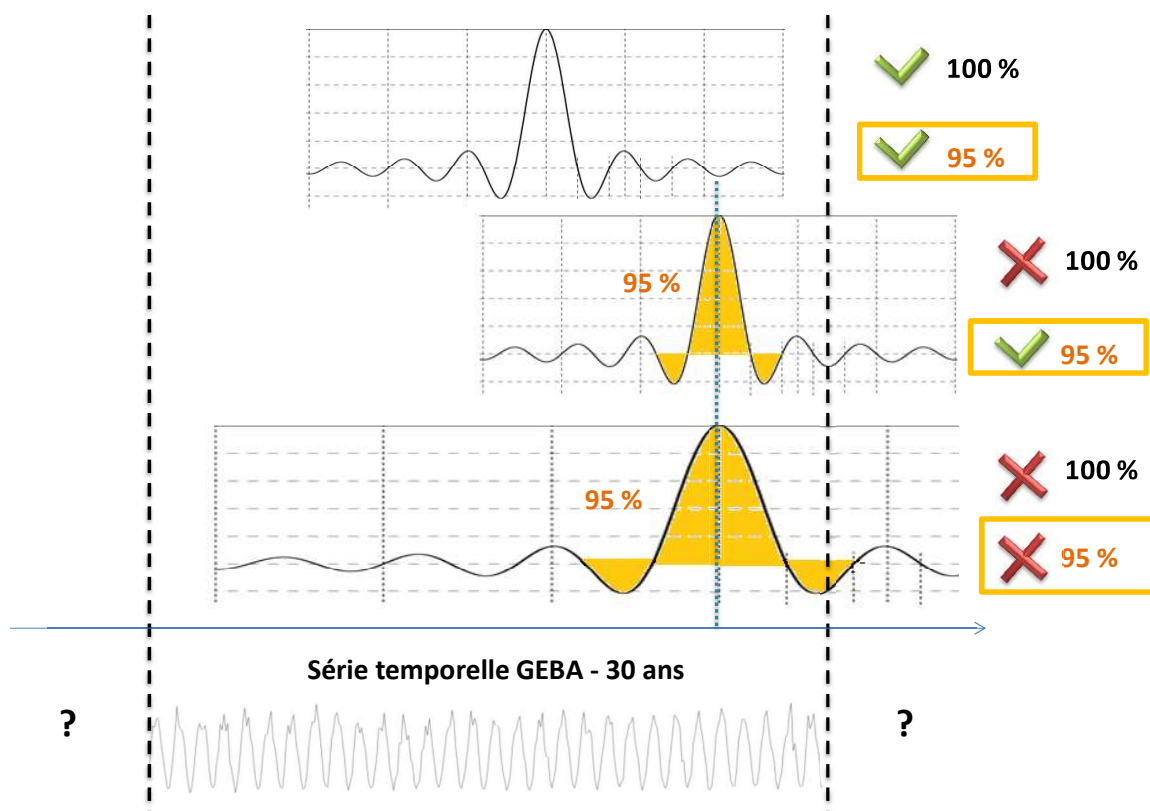


Figure 28 : Représentation schématique de l'influence des effets de bords sur le filtre considéré.

On voit bien sur la figure 28 que l'extension des effets de bords est dépendante de la largeur du sinus cardinal apodisé. Pour un sinus cardinal étroit, comme illustré en haut du schéma, peu de données seront affectées par les effets de bords. Pour un sinus cardinal large, comme c'est le cas en bas, un grand nombre de données seront influencées par les effets de bords. Ces données ne seront alors pas considérées comme valides. **La largeur du filtre, associée au faible nombre de données disponibles, entraîne une limitation intrinsèque du nombre de données valides pour notre analyse. Ainsi, les périodes supérieures à 72 mois auront un nombre de données valides très faible.** Or, les variations de long terme sont justement l'objet de cette thèse. Afin de limiter l'influence des effets de bords tout en conservant la quasi-totalité de l'information, nous avons choisi de réduire le rayon effectif du sinus cardinal. Ainsi, lors du filtrage, nous considérons valide les données filtrées par les premiers lobes du sinus cardinal apodisé. 95 % de l'information est comprise dans ces deux premiers lobes, comme illustré en orange sur la figure. On voit ainsi de façon schématique que les sinus cardinal apodisé au centre de la figure fournira plus de données valides dans ce second cas que dans le premier.

Pour les données affectées par des effets de bords, il est évidemment nécessaire de connaître précisément les données hors de la série temporelle originale utilisées lors du filtrage. Ces données vont déterminer l'ampleur de l'influence des effets de bord sur la donnée de sortie. Cela est également illustré sur la figure 28. Si les données à l'extérieur de la série temporelle sont de qualité, ou du moins représentative du climat considéré, l'information fournie par la décomposition pourra être considérée comme plausible. Dans ce but, une extension temporelle à droite et à gauche du jeu de données GEBA a été réalisée. Celle-ci s'appuie sur la série de données GEBA originale. Plus précisément, dans le cas où il existe des périodes de données valides avant ou après la période sélectionnée, une complétion des données similaires à celle effectuée pour les données manquantes a été appliquée (décrite en section 3.3 du chapitre 2). Bien évidemment, les intervalles de données manquantes pouvant être importants, la qualité de l'interpolation linéaire peut être discutable. Dans le cas où aucune donnée n'est disponible hors de la période sélectionnée, la série temporelle est reproduite à l'identique avant et après la période. Dans les deux cas, les données complétées ne pourront pas être considérées comme représentatives de l'éclairement du site à la période considérée. Elles n'ont pour but que de limiter l'influence des effets de bord en conservant dans les grandes lignes les variations climatologiques du site d'intérêt. Il sera indispensable de les distinguer clairement et proprement lors de l'analyse des décompositions.

La figure 29 représente le masque de données influencées par les effets de bords en fonction de l'élément considéré. Un exemple d'une série temporelle de 30 ans entre 1970 et 2000 a été choisi.

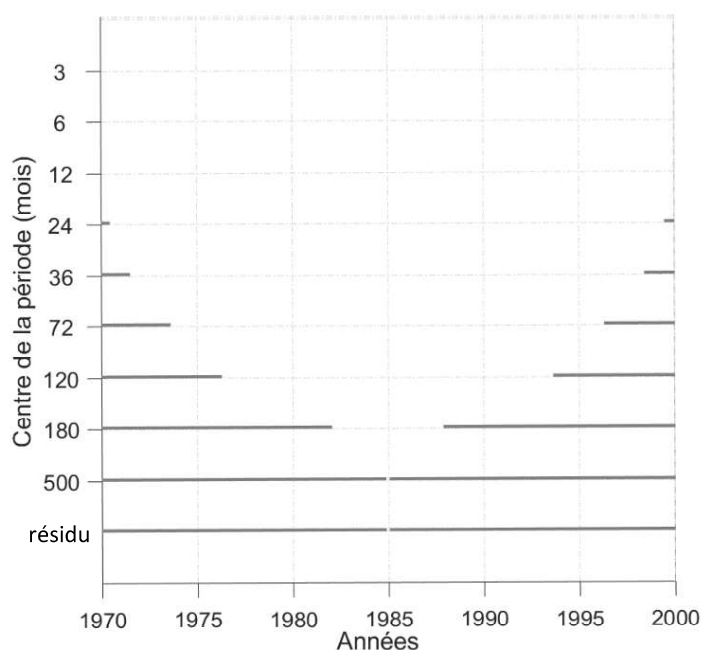


Figure 29 : Données affectées par des effets de bords en fonction de l'élément considéré, pour le cas illustratif d'une série temporelle entre 1970 et 2000.

On retrouve sur cette figure que le nombre de données influencées par les effets de bords est plus important pour les longues périodes que pour les courtes périodes. Les éléments y_{500} et $y_{\text{résidu}}$ n'ont aucune donnée valide pour une série temporelle de 30 ans. Ainsi, ces deux éléments n'auront aucune données valide pour la majorité des stations GEBA. L'élément y_{180} a également l'essentiel de sa série temporelle qui ne pourra être considéré comme valide. A l'inverse, les éléments y_3 , y_6 et y_{12} ne seront pas affectés par les effets de bords.

Cette figure nous encourage à distinguer trois classes de variabilités : la variabilité inter-annuelle, constituée des éléments y_3 , y_6 et y_{12} ; la variabilité inter-annuelle à décennale, composée des éléments y_{24} à y_{120} ; et la variabilité multi-décennale, composée des éléments y_{180} , y_{500} et $y_{\text{résidu}}$. Par la suite, ces trois classes de variabilité seront notées respectivement VIA, VAD et VMD. La première classe de variabilité sera considérée pour établir l'importance relative des variations intra-annuelles dans le signal total d'éclairage. La seconde classe de variabilité rassemble les éléments peu affectés par les effets de bords. Ces éléments pourront être correctement caractérisés grâce à la méthode de décomposition utilisée. La troisième classe de variabilité est fortement affectée par les effets de bords. Ainsi, des études complémentaires utilisant d'autres outils seront nécessaires pour caractériser de manière fine ces modes de variabilité. Cette classe sera néanmoins utilisée pour mettre en évidence l'importance ou non de la prise en compte des VMD pour une bonne compréhension du signal. Ainsi, notre outil permet bien d'étudier différentes échelles de variations de façon simultanée.

Soulignons que les délimitations temporelles des classes VIA, VAD et VMD ne sont pas exactes. En effet, comme vu précédemment, des recouvrements entre les échelles sont possibles. À l'extrême, des variations de périodes de 24 mois peuvent être trouvées dans la VIA, et de 180 mois dans la VAD.

3 Validation de la méthode de décomposition temps-fréquence

Pour garantir l'adéquation entre les caractéristiques de la méthode proposée et les attentes définies précédemment, différentes validations ont été réalisées. Dans un premier temps, les décompositions temps-fréquence de signaux synthétiques ont été analysées. Ces décompositions ont pour but de tester la détection par la méthode choisie de changements abrupts et de variations de long terme. Elles illustreront également l'intérêt d'une décomposition multi-échelle. Par la suite, une comparaison visuelle entre les résultats issus de la décomposition temps-fréquence et ceux issus de la bibliographie a été menée pour deux stations de référence. Cette comparaison permettra d'illustrer la capacité de la méthode choisie à synthétiser l'information disponible dans la bibliographie.

3.1 Expérimentation sur la base d'un signal synthétique

La décomposition de deux signaux synthétiques a été réalisée. Le premier signal choisi est composé d'une combinaison linéaire de sinus de périodes différentes : 6 mois, 12 mois, 42 mois et 115 mois. À cela on ajoute une tendance linéaire, et une modulation d'amplitude du sinus de période 115 mois. Ainsi, certaines périodes sont situées au centre des filtres associées (6 mois et 12 mois), quand d'autres sont intermédiaires à deux filtres (42 mois et 115 mois). On pourra donc étudier le comportement des éléments issus de la décomposition temps-fréquence dans un cas idéal et dans un cas de possibles recouvrements des filtres. Le second signal synthétique correspond à la même combinaison linéaire de sinus, à laquelle on ajoute un changement abrupt. Ce changement pourrait être dû, dans le cas de l'éclaircissement, à un changement d'instrument de mesure. C'est ici la détection de changement par la méthode multi-échelle qui est testée. Les deux signaux sont composés de 840 points, représentant artificiellement 840 mois, soit 70 ans de données d'éclaircissement. Ces deux signaux sont représentés sur la figure 30 et la figure 31.

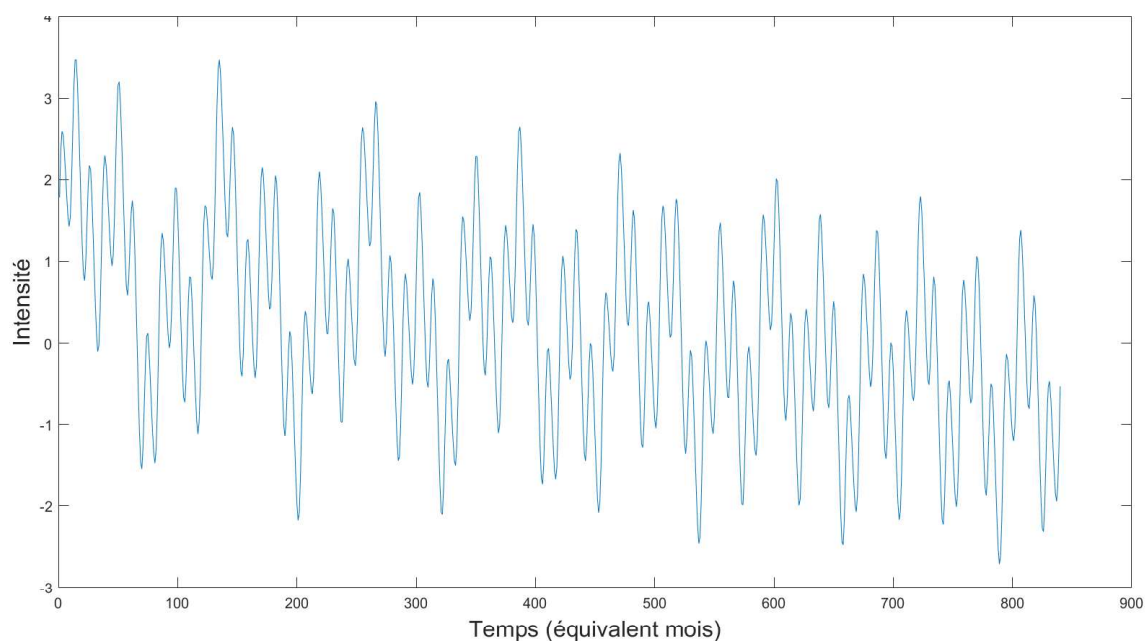


Figure 30 : Signal synthétique construit par une combinaison linéaire de sinus.

3. Validation de la méthode de décomposition temps-fréquence

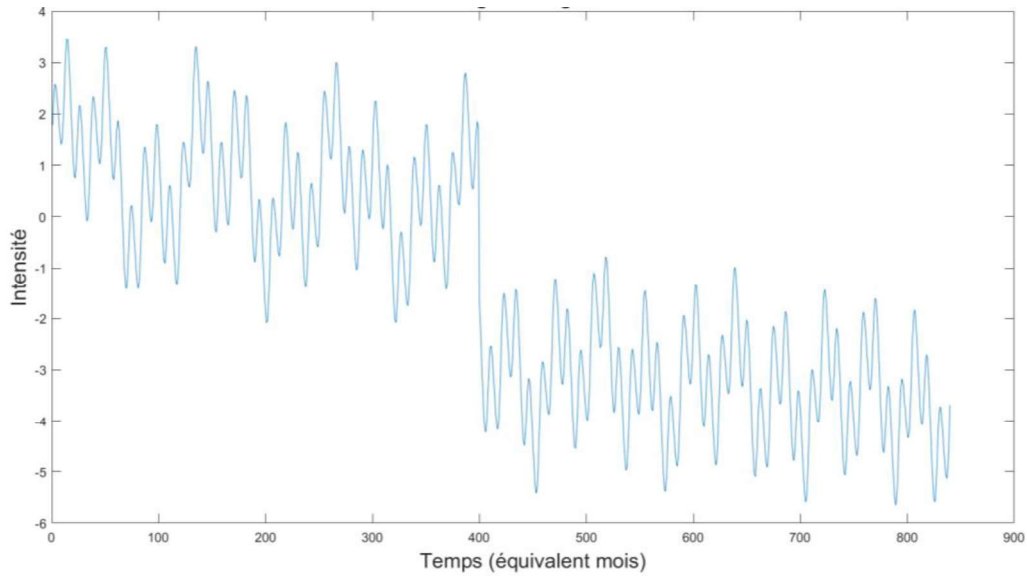


Figure 31 : Signal synthétique similaire à celui présenté figure 30, auquel s'ajoute un changement abrupt.

On retrouve sur la figure 30 et sur la figure 31 un ensemble de variations de périodes différentes, et une légère tendance à la décroissance du signal. Le changement abrupt du deuxième signal est visible le 400^{ème} mois.

La décomposition temps-fréquence du premier signal par le banc de filtres introduit auparavant est représentée sur la figure 32. À chaque élément issu de la décomposition est associé une couleur. Les données étant fortement influencées par les effets de bords sont colorées en gris pour chaque élément.

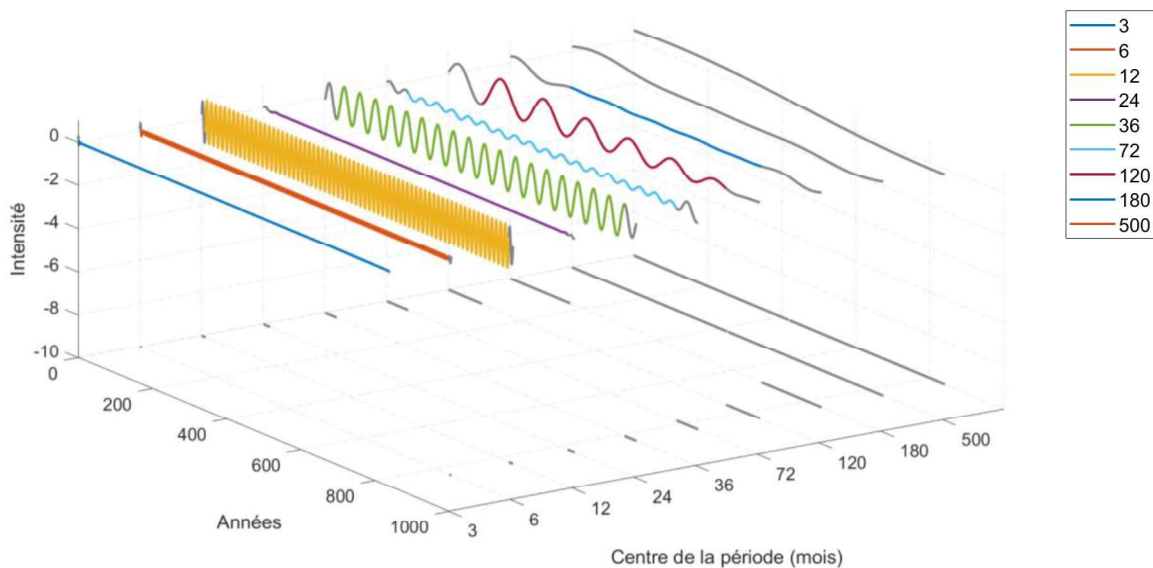


Figure 32 : Vue 3-D de la décomposition temps-fréquence du signal synthétique construit par une combinaison linéaire de sinus grâce au banc de filtre choisi. Les zones grisées sont affectées par les effets de bords.

On retrouve pour certains éléments une décomposition en accord avec les attentes vis-à-vis de la méthode. Par exemple, γ_3 et γ_{24} , ici respectivement en bleu foncé et en violet, sont identiquement nuls. Cela est cohérent avec bien l'absence de variation du signal pour cette période. De même, comme attendu, des variations sont visibles pour les éléments γ_6 et γ_{12} , en rouge et en jaune sur la figure. γ_{72}

À l'inverse, les éléments représentant les plages de périodes centrées sur 36 et 72 mois peuvent être analysés au regard de la figure 27. Comme annoncé précédemment, l'information contenue dans un signal ayant une période caractéristique de 42 mois sera répartie dans deux éléments par le banc de filtre : γ_{36} pour la plus grande part, et γ_{72} pour le reste. Cela se retrouve sur la figure 32, où une variation de période identique est observée pour γ_{36} et γ_{72} . Cette variation est d'une intensité forte pour γ_{36} , tracé en vert, et de plus faible intensité pour γ_{72} , tracée en bleu ciel.

Enfin, les variations de période de 115 mois dans le signal original se retrouvent dans l'élément centrée en 120 mois, γ_{120} , tracé en rouge. On retrouve également la modulation de l'intensité du signal avec le temps. Les éléments γ_{180} et γ_{500} sont quasiment nuls, sauf en bordure du signal. Ces zones correspondent aux zones de forte influence des effets de bord. Le dernier élément représentant le résidu du signal montre quant à lui une tendance linéaire très claire à la décroissance. Cette tendance est partiellement lissée par la représentation 3D. Cet élément, comme γ_{500} , sont complètement colorés en gris. On retrouve la difficulté d'étudier la variabilité de long terme du signal : même avec un équivalent de 70 années de données, les variations de long terme du signal ne sont pas précisément représentées. Ces plages de périodes ont été conservées car pourront être utiles pour des bases de données ayant un très long historique, telles que les modèles climatiques. Pour les autres bases de données, les trois derniers éléments, correspondant à la VMD, seront traités de manière groupée.

Pour compléter cette analyse, la figure 33 propose la décomposition temps-fréquence du second signal introduit précédemment, présenté en figure 31.

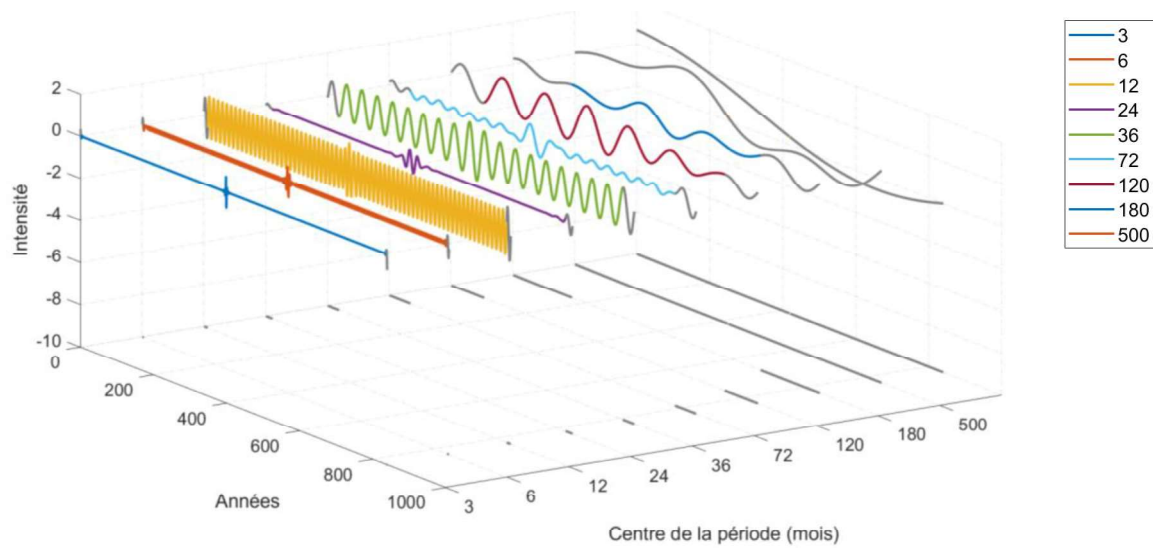


Figure 33 : Vue 3-D de la décomposition temps-fréquence du signal synthétique construit par une combinaison linéaire de sinus et soumis à un changement abrupt grâce au banc de filtre choisi. Les zones grisées sont affectées par les effets de bords.

On constate que pour les courtes périodes, peu de valeurs sont impactées par le changement abrupt. L'impact est très localisé autour de 400 mois, et bien visible sur les éléments y_3 , y_6 et y_{24} . Il est également distinguable pour y_{12} , y_{36} et y_{72} , qui connaissent des anomalies dans leurs variations sur une plus longue durée. À partir de y_{120} , en rouge, l'impact du changement abrupt est beaucoup moins localisable. Cependant, il reste important. En effet, on voit en comparaison avec la figure 32 que : l'atténuation de y_{120} est beaucoup moins visible, que les éléments y_{180} , y_{500} , et $y_{\text{résidu}}$ possèdent des variations importantes, qui n'existaient pas précédemment. Autrement dit, si l'on analyse uniquement l'élément $y_{\text{résidu}}$, il est possible d'analyser comme des variations de long terme des anomalies de la série temporelle. Cela souligne l'intérêt de la décomposition temps-fréquence choisie : en analysant l'ensemble des éléments, il est aisé d'identifier une anomalie de mesure, ainsi que les répercussions de telles anomalies entre les échelles.

En résumé, nous pouvons dire que le banc de filtres choisi est en mesure de représenter correctement les différents éléments du signal de façon simultanée. Cependant, les recouvrements des filtres constituant le banc de filtres peuvent être handicapants lors de l'analyse des décompositions. Les recouvrements sont minimisés par la correcte détermination des périodes centrales de chaque filtre. Cependant, de possibles artefacts dus aux recouvrements ne peuvent être écartés. Ceux-ci peuvent être approchés en étudiant l'orthogonalité de la décomposition. Celle-ci sera vérifiée de manière pragmatique lors de l'analyse de séries temporelles d'éclairement issues des différentes bases de données. Enfin, nous avons vu que l'outil proposé permet d'identifier rapidement un changement abrupt d'éclairement, qui pourrait entraîner un défaut d'interprétation des variations de long terme.

3.2 Validation de la méthode au regard de la bibliographie

La méthode choisie a pour but de synthétiser les informations disponibles dans la bibliographie. Ainsi, une comparaison visuelle entre les variations de l'éclairement fournies par la bibliographie et celles construites à partir des éléments de la décomposition temps-fréquence a été réalisée.

Le choix des sites de comparaison n'a pas été aisé. En effet, dans un grand nombre d'études, un composite de stations est utilisé pour mettre en évidence les variations de long terme (*e.g.* SANCHEZ-LORENZO *et al.*, 2015). La méthode de calcul de ces composites est rarement explicitée. Cependant, un traitement des données d'entrée différent entre la bibliographie et notre étude pourrait entraîner des biais dans les résultats issus de la comparaison visuelle. Ainsi, nous avons choisi de nous référer à des études étudiant de manière individuelle des stations appartenant au réseau GEBA. Pour ces stations, une représentation graphique des séries temporelles et de ses variations de long terme par les auteurs a été choisie. Les biais dus à l'implémentation des méthodes décrites par les auteurs ont ainsi été évités. **Au final, deux stations ont été sélectionnées : la station allemande de Postdam et la station espagnole de Murcia.**

Ces deux stations ont été utilisées pour illustrer les variations long terme de l'éclairement dans Wild *et al.* (2016) (Potsdam) et Sanchez-Lorenzo *et al.* (2013) (Murcia). Comme annoncé dans le tableau 7 du chapitre 4, Wild *et al.* (2016) représentent les variations de l'éclairement à l'échelle annuelle, ainsi que les moyennes glissantes sur cinq ans pour la première étude. La période d'analyse est 1937-2014. Ces variations sont présentées sur la figure 34. Pour la seconde étude, les variations de l'éclairement sont également représentées en moyennes annuelles, sur lesquelles un filtre passe bas gaussien de 13 ans est appliqué. La période d'analyse est 1970-2010. Les données correspondent à des anomalies annuelles calculées à partir de la moyenne d'éclairement entre 1991–2010. Ces variations sont présentées sur la figure 36.

Les deux stations choisies ne font pas partie de la sélection de 104 stations GEBA. Elles possèdent donc des trous de plus de deux mois consécutifs au sein de la série temporelle. La méthode de complétion de données décrite au chapitre 2 a été appliquée afin de compléter ces trous. 132 mois sont manquants dans la série temporelle de la station Postdam. Soulignons que ces 132 données manquantes correspondent en réalité à 11 années manquantes non consécutives. 57 années complètes sont disponibles. Pour la station Murcia, 17 données manquantes ont été répertoriées, affectant principalement deux années particulières.

Une décomposition temps-fréquence a été réalisée à l'aide du banc de filtres introduit précédemment. **Afin de fournir une information comparable à celle trouvée dans la bibliographie, nous avons construit trois séries temporelles à partir des éléments de la décomposition :**

- **Une somme de l'ensemble des plages de périodes supérieures à 12 mois.** On ôte ainsi les variations intra-annuelles du signal. **Il en résulte un équivalent de la moyenne glissante**

3. Validation de la méthode de décomposition temps-fréquence

de la série temporelle originale. Cette série temporelle pourra être comparée au signal annuel fournie par les deux études. Il sera ainsi possible d'identifier les différences entre les variations fournies par une moyenne annuelle et par une moyenne glissante.

- **Une somme des éléments supérieurs à 36 mois.** On ôte ainsi les variations inférieures à 6 ans. **Cela est comparable à la moyenne glissante sur 5 ans réalisée par Wild *et al.* (2016).**
- **Une somme des éléments supérieurs à 120 mois.** **Ce signal correspond à un filtrage des variations inférieures à 13 ans, représenté par Sanchez-Lorenzo *et al.* (2013).**

La figure 35 et la figure 37 rassemblent en rouge l'équivalent à la moyenne glissante annuelle, et en bleu l'équivalent des variations de long terme propre à chaque étude. A ces deux courbes s'ajoute en vert la moyenne annuelle de la série temporelle originale. Celle-ci est similaire à la moyenne annuelle représentée en noir dans les deux figures issues de la bibliographie. Il sera ainsi plus facile de comparer les figures issues de la bibliographie et les figures proposées par notre étude. En gris sont représentées les zones subissant une forte influence des effets de bords de la méthode.

En ce qui concerne la figure 35, comme attendu, des données des similarités claires peuvent être observées avec la courbe noire de la figure 34. Entre autres, des pics de fort éclaircissement peuvent être observées sur les deux figures en 1947, 1959 et 1982.

Les courbes rouges et vertes sont proches. On peut toutefois noter que l'équivalent de la moyenne glissante permet de détecter beaucoup plus de variabilité du signal.

Enfin, les variations de long terme de l'éclaircissement identifiées dans la bibliographie se retrouvent sur la figure 35. En particulier, les périodes d'obscurcissement global dès 1950 et jusqu'au milieu des années 1980, et d'éclaircissement global après le milieu des années 1980, sont clairement distinguables sur les deux représentations. L'éclaircissement global situé avant les années 1950 se retrouve également sur la figure 35, si l'on omet les données influencées par les effets de bords. Cependant, des différences notables sont visibles entre les courbes bleues des deux figures. Premièrement, les valeurs d'éclaircissement sont plus élevées sur la figure 35 que sur la figure 34. Cela peut aisément se justifier par le fait qu'une partie des variations du signal ne sont pas considérées sur la figure 35. Ensuite, des variations plus fines sont visibles sur cette dernière figure. En effet, deux modes de variabilité sont visibles sur la figure 35 : les variations de long terme très proches de celles visibles sur la figure 34, et des variations avec une période proche de 10 ans. **La méthode proposée permet donc de détecter non seulement les variations de long terme identifiées dans la bibliographie, mais aussi des modes de variabilité comprises entre l'année et la décennie. On a donc bien une détection simultanée de la VAD et de la VMD.**

En ce qui concerne la figure 37, des constatations similaires peuvent être faites. Les données annuelles sont très similaires avec celles observées sur la figure 36. La courbe rouge est proche de la courbe verte. Elle met toutefois en évidence des modes de variabilités qui ne sont pas détecté dans le cas de moyenne annuelles non glissantes. Enfin, les variations de long terme de l'éclaircissement sont très similaires entre les deux figures, que l'on omette ou non les effets de bords.

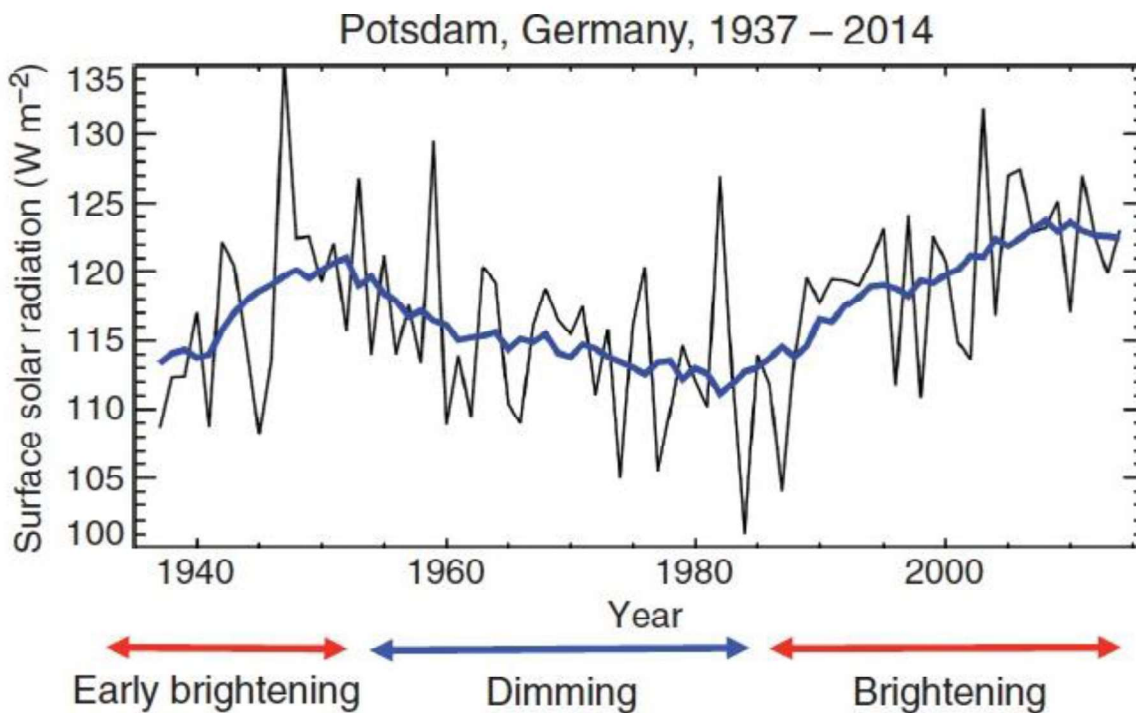


Figure 34 : Moyenne annuelle de l'éclairement (noir) et moyenne glissante sur cinq ans (bleu) pour la station GEBA Potsdam. Extraite de Wild et al. (2016).

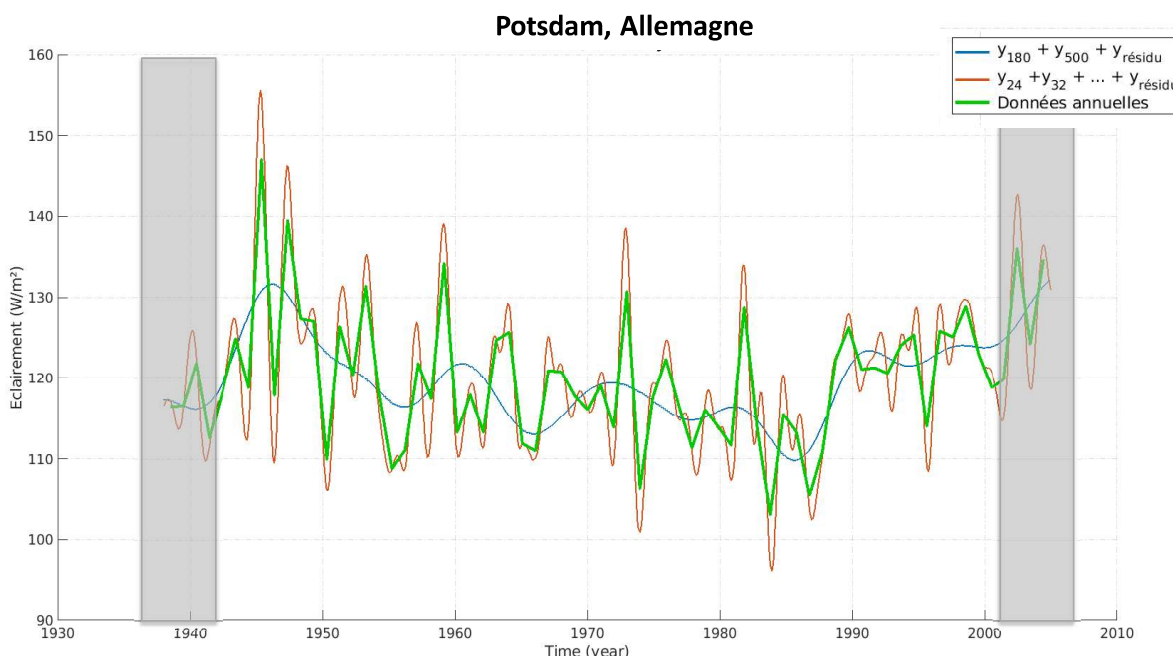


Figure 35 : Equivalent de moyenne annuelle, de moyenne annuelle glissante, et de moyenne glissante sur cinq ans de l'éclairement. Ces équivalents sont construits par une somme d'éléments de la décomposition temps-fréquence de la station Potsdam.

3. Validation de la méthode de décomposition temps-fréquence

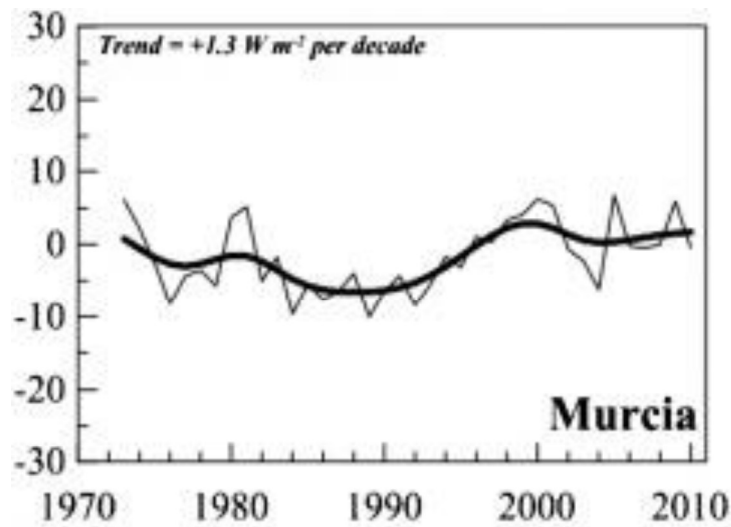


Figure 36 : Moyenne annuelle de l'éclairement (noir fin) et filtre passe bas gaussien de 13 ans (noir épais) pour la station GEBA Potsdam. Extraite de Sanchez-Lorenzo et al.

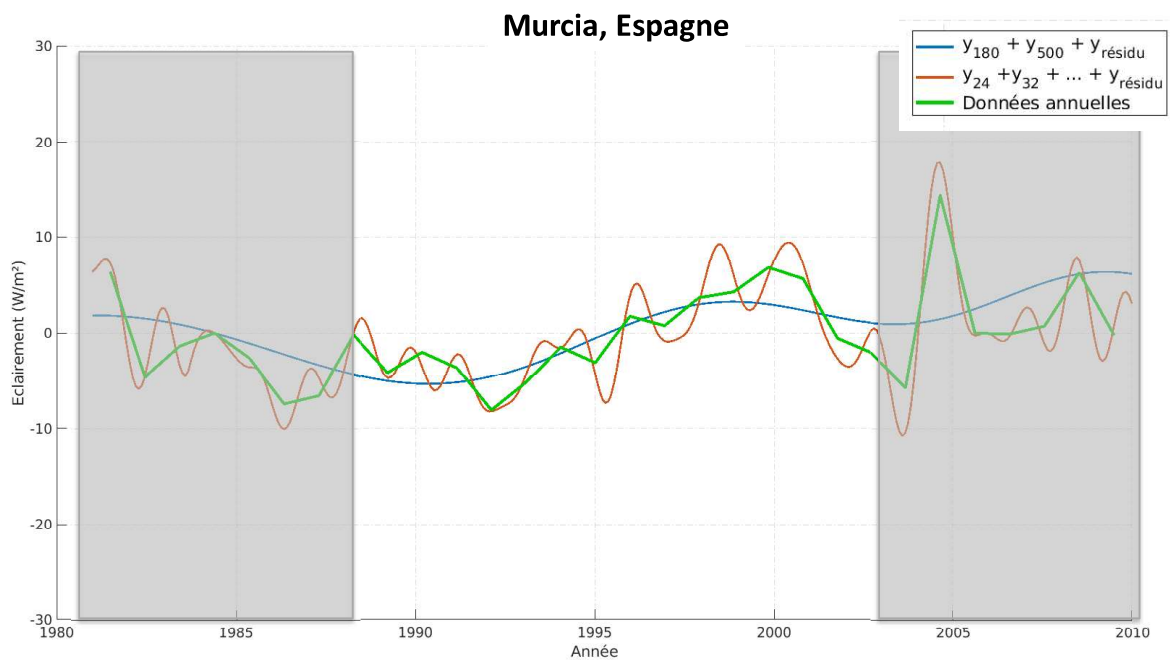


Figure 37 : Equivalent de moyenne annuelle, de moyenne annuelle glissante, et du filtrage par un filtre passe bas gaussien de 13 ans. Ces équivalents sont construits par une somme d'éléments de la décomposition temps-fréquence de la station Murcia.

En conclusion, ces deux exemples illustrent la capacité de la méthode de décomposition temps-fréquence à synthétiser les variations de long terme décrites dans la bibliographie. Ils ont également mis en évidence que l'information fournie par cette méthode est plus riche que celle proposée par la bibliographie.

4 Conclusion

Pour mettre en évidence et caractériser les variations de l'éclairement, il a été choisi d'utiliser une méthode de décomposition temps-fréquence. La méthode choisie est construite à partir d'un banc de filtres offrant des éléments centrés sur des périodes caractéristiques de l'éclairement. Cette méthode est multi-échelle, simple d'usage, et offre des modes de variabilités comparables entre décompositions. Les éléments issus de la décomposition peuvent être rassemblés pour reconstruire fidèlement le signal d'origine. Elle permet à la fois de synthétiser l'information disponible dans la bibliographie tout en offrant une information plus riche. Différentes validations de la méthode ont été réalisées. L'ensemble des résultats issus de ces validations nous conforte quant à la pertinence de la méthode choisie pour identifier et caractériser les variations de l'éclairement. Enfin, les effets de bords inhérents à la méthode nous ont amené à considérer trois classes de variabilité : la VIA, la VAD et la VMD. Ces trois classes structureront les analyses futures des résultats de la décomposition.

Chapitre 6

Décomposition temps-fréquence des indices de clartés mensuels de la sélection de stations GEBA

Table des matières

1	Introduction	117
2	Etude de la stationnarité temporelle des stations GEBA.....	117
3	Quels sont les modes de variabilité inter-annuels permettant d'expliquer le signal d'indice de clarté ?.....	121
4	L'influence de chaque élément sur le signal original est-elle stationnaire ?.....	123
5	Conclusion.....	126

1 Introduction

Afin de caractériser les modes de variabilité inter-annuels de l'éclairement, un outil de décomposition temps-fréquence a été développé. Cet outil a été identifié, caractérisé et validé. L'objet de ce chapitre est d'analyser les séries temporelles des stations GEBA à l'aide de cet outil.

Comme annoncé au cours du chapitre 2, les méthodes de génération de séries temporelles d'éclairement l'état de l'art possèdent de meilleures performances à l'échelle mensuelle. Ainsi, **la suite de notre analyse portera sur l'estimation des variabilités de séries temporelles mensuelles d'éclairement. Pour éviter les effets latitudinaux, nous allons à nouveau travailler sur l'indice de clarté.** Pour rappel, celui-ci correspond au rapport entre le signal d'éclairement incident au sol et celui, concomitant, au sommet de l'atmosphère. Les besoins de la communauté étant plus focalisée sur les variations inter-annuelles de l'éclairement, une comparaison entre les résultats obtenus sur l'indice de clarté mensuel et le GHI à l'échelle annuelle sera réalisée au cours du chapitre 7.

Trois classes de variabilité ont pu être distinguées au cours du chapitre 5 : la variabilité intra-annuelle (VIA), la variabilité annuelle à décennale (VAD), et la variabilité multi-annuelle (VMD). L'importance relative de chacune de ces trois classes de variabilité pour expliquer les séries temporelles d'indice de clarté a été étudiée. Cela a permis de distinguer d'une part les zones géographiques soumises à une forte variabilité inter-annuelle, et d'autre part les zones géographiques pour lesquelles il est possible de la négliger. Pour les stations soumises à une forte variabilité inter-annuelle, un effort de caractérisation de ces dernières a été réalisé.

2 Etude de la stationnarité temporelle des stations GEBA

Pour connaître l'influence relative de la variabilité inter-annuelle (VAD et VMD) par rapport à la variabilité intra-annuelle (VIA), une comparaison a été réalisée entre l'indice de clarté des stations GEBA d'une part, et la VIA des stations GEBA d'autre part. Cette comparaison revient à estimer les différences entre un signal ne possédant aucune variabilité inter-annuelle et le signal total. Si la différence est faible, alors la variabilité inter-annuelle sera considérée négligeable. Si la différence est grande, il sera alors nécessaire de prendre en compte les variabilités de long terme pour expliquer correctement le signal d'indice de clarté, et par voie de fait celui de l'éclairement.

Plus précisément, la comparaison peut se résumer en trois étapes. Premièrement, la série temporelle d'indice de clarté de chacune des 104 stations de la sélection GEBA a été décomposée par le banc de filtres présenté dans le chapitre précédent.

Ensuite, 104 séries temporelles ont été construites en sommant uniquement les trois premiers éléments pour chaque station : y_3 , y_6 et y_{12} . Cela revient à ôter l'ensemble des variations inter-annuelles du signal de l'indice de clarté. On a ainsi 104 séries temporelles représentatives de la VIA.

Troisièmement, la comparaison entre les variations du signal constituée des variations intra-annuelles et celles du signal original est réalisée pour chaque station. **Pour cela, le coefficient de détermination entre les deux séries temporelles a été calculé. Le coefficient de détermination correspond au carré du coefficient de corrélation. Il permet d'estimer le rapport de la variance expliquée par la régression sur la variance totale du signal.** Si celui-ci est grand, alors les deux signaux sont proches. L'utilisation du coefficient de détermination pour estimer les variations de long terme de l'éclairement se retrouve dans la méthode des moindres carrés (*cf.* chapitre 4 et Storch *et al.*, 2000). Ici, au lieu de comparer la série temporelle à un ajustement linéaire, nous la comparons à une estimation fine de la VIA. Autrement dit, au lieu de chercher une tendance dans les données, nous cherchons à identifier l'importance de la variabilité intra-annuelle. Ainsi, on s'affranchit de la contrainte de caractériser les variations inter-annuelles uniquement comme une tendance linéaire.

Soulignons que si la décomposition temps-fréquence est parfaitement orthogonale, la somme des coefficients de détermination calculés pour chaque élément est égale à 1. Ainsi, la variance qui n'est pas expliquée par la VIA correspond exactement à la variance expliquée par les variations inter-annuelles. Pour notre cas, il a été trouvé que la somme moyenne des coefficients de détermination pour les 104 stations GEBA est égale à 1.04. Nous considérerons par la suite les résultats issus de la décomposition comme quasi orthogonales. Pour faciliter l'analyse, et par soucis de comparaison entre les décompositions issues de différentes bases de données au chapitre suivant, nous avons néanmoins normalisé chaque coefficient de détermination, de façon à en ramener la somme à 1.

Il est à noter que l'indicateur choisi quantifie la similarité des variations du signal, et non la similarité des signaux en tant que tels. Ainsi, le biais systématique entre les deux signaux - dû au fait que l'un est centré en zéro par construction - n'influe pas sur les valeurs de coefficient de détermination. **De plus, rappelons que les 104 stations GEBA n'offrent hélas pas des séries temporelles coïncidentes temporellement. Ainsi, il est possible que les différences de coefficient de détermination observées entre stations soient dues à la variété des périodes temporelles plutôt qu'à des différences de variances en tant que telles. C'est pourquoi nous ne nous attacherons pas à une comparaison détaillée des différences entre les stations, mais plutôt à la répartition des variabilités entre les échelles intra-annuelles et inter-annuelles.**

La figure 38 présente la répartition spatiale des stations de la sélection GEBA pour le coefficient de détermination. Ce dernier est représenté en couleur, allant du bleu au jaune pur des valeurs allant de 0 à 1.

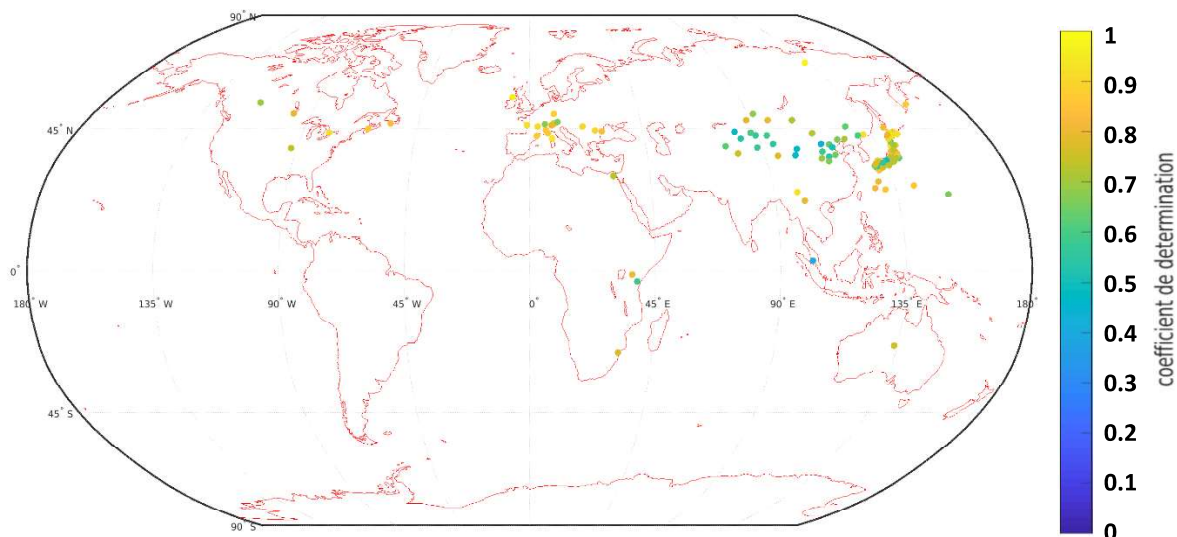


Figure 38 : Répartition géographique des coefficients de détermination entre les séries temporelles mensuelles d'indice de clarté et les séries temporelles représentatives de la VIA ($y_3 + y_6 + y_{12}$) (couleurs) des 104 stations GEBA sélectionnées.

Le coefficient de détermination moyen sur les 104 stations s'élève à 0.742. Cela revient à dire qu'en moyenne, 74.2 % des variations du signal sont expliquées par la VIA, ou encore que 25.8 % de la variance du signal ne peut être expliqué que par la variabilité inter-annuelle de l'indice de clarté.

Bien évidemment, ces chiffres varient en fonction de la station considérée. On peut constater sur cette carte de fortes différences dans le coefficient de détermination, avec une amplitude de 0.4 à presque 1. La station russe Vladivostok (station n°94) connaît le plus haut coefficient de détermination (0.975). Pour cette station, l'intérêt de prendre en compte la variabilité inter-annuelle du signal est donc très faible. À l'inverse, la station malaisienne Kuala Lumpur (station n°42) possède le coefficient de détermination le plus faible (0.402). Cette faible performance traduit la présence de variations de long terme, qui explique à près de 60 % la variance de la série temporelle de l'indice de clarté. Une représentation de la décomposition temps-fréquence de cette station est détaillée par la suite en section 4, figure 41.

En ce qui concerne la répartition géographique des coefficients de détermination, on peut distinguer des zones présentant des coefficients de détermination élevés, telles que l'Europe ou l'Amérique du Nord ; et des zones présentant des coefficients plus faibles, telles que l'Asie ou la zone équatoriale. Cependant, une certaine diversité des valeurs est distinguable au sein même de chacune de ces zones. Cela est illustré sur la figure 39. Cette dernière reprend les éléments de la figure 38 pour les zones de l'Europe et de l'Asie (restreinte à la Chine et au Japon).

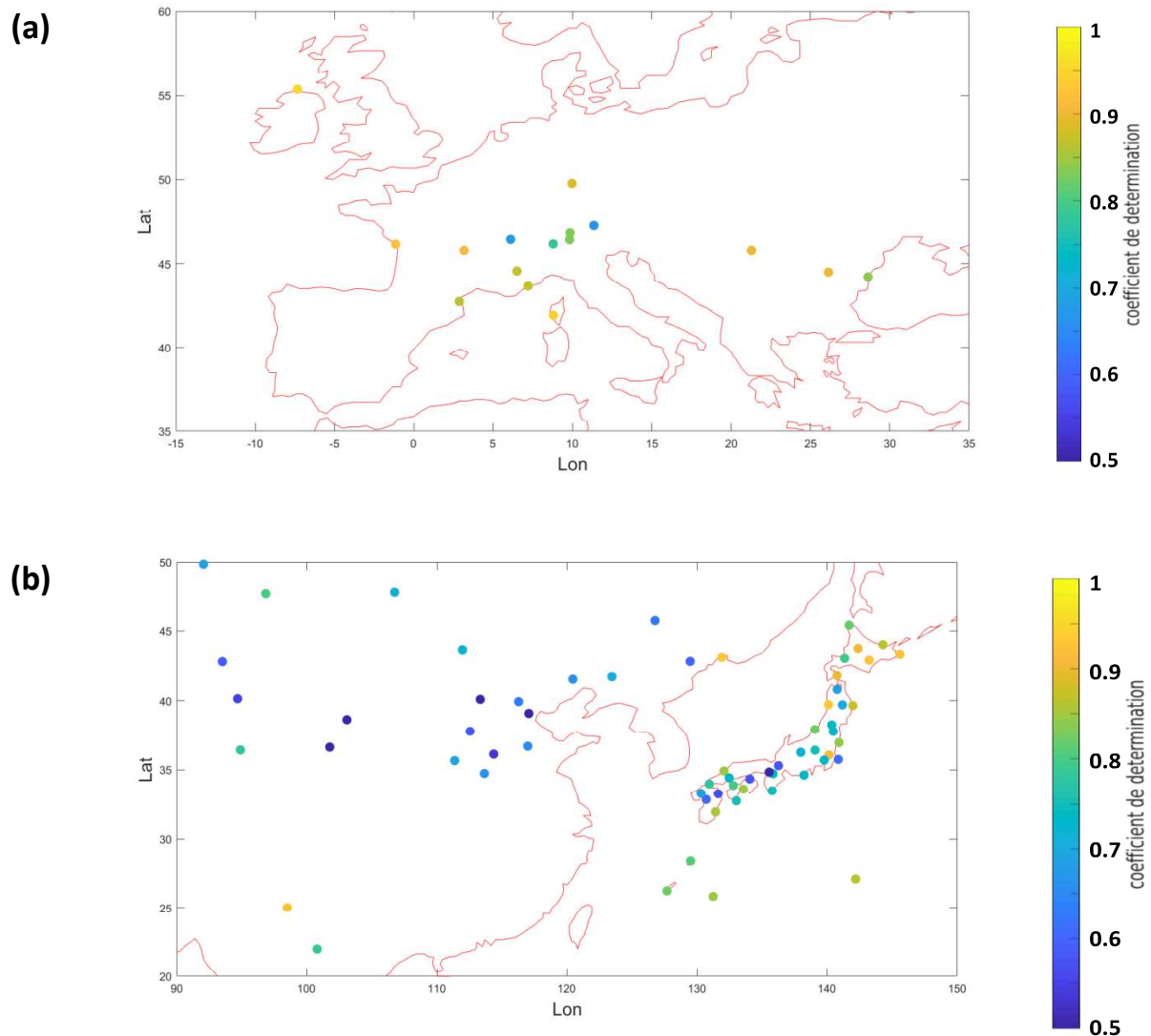


Figure 39 : Répartition géographique des coefficients de détermination entre les séries temporelles mensuelles d'indice de clarté et les séries temporelles représentatives de la VIA ($y_3 + y_6 + y_{12}$) (couleurs), pour la région Europe (a) et Asie (b).

La région européenne possède en moyenne des coefficients de détermination élevés, de 0.852. Cependant, une diversité des valeurs est distinguable, avec des coefficients allant de 0.961 pour la station irlandaise Malin (station n°49), à 0.664 pour la station autrichienne d'Innsbruck (station n°36). Autrement dit, si en moyenne les variations inter-annuelles expliquent au plus 15 % des variations de la série temporelle historique en Europe, il est possible de trouver localement des stations pour lesquelles cette part est plus importante, jusque 33 %, ou moins importante, jusque 4%. La part explicative des variations inter-annuelles peut donc varier pour une même région de près de 30%. Cela souligne l'importance d'une analyse à l'échelle du point.

Ces constatations sont également observables pour les cas de l'Asie. La moyenne des coefficients de détermination vaut pour cette région 0.729. Cependant, il est possible de trouver des stations ayant un coefficient de détermination de 0.430 (station n°10, Beijing), et donc une forte variabilité inter-annuelle. À l'inverse, pour d'autres stations, la VIA permet d'expliquer plus de 93 % du signal (station

n°59, Naha). Ainsi, la diversité des résultats même sur des zones à première vue similaires ne permet pas d'identifier de structure spatiale dans les coefficients de déterminations. Soulignons que des données possédant une couverture plus dense spatialement seraient d'intérêt pour confirmer et affiner ces constatations.

Il est clair que l'influence des variabilités inter-annuelles sur le signal varie fortement d'une station du réseau GEBA à l'autre. La VIA explique plus de 95 % du signal original pour uniquement deux stations (Malin et Vladivostok, respectivement station n°49 et station n°94). La majorité des stations ne peuvent donc être correctement représentées uniquement par la VIA. Ainsi, une caractérisation des modes de variabilité inter-annuels est intéressante. Pour cela, l'importance relative de la VAD et de la VAM dans le signal a été explorée. Comme vu au cours du chapitre 5, la VMD ne peut être étudiée de manière fine. En effet, le faible nombre de données disponibles dans GEBA associée aux effets de bords de la méthode ne permet pas une caractérisation fine de ces échelles. Elles seront donc étudiées conjointement. La VAD pourra à l'inverse être étudié élément par élément.

3 Quels sont les modes de variabilité inter-annuels permettant d'expliquer le signal d'indice de clarté ?

L'objectif est d'identifier les modes de variabilité les plus influents au sein de la variabilité inter-annuelle. Pour cela, nous avons calculé les coefficients de détermination entre chaque élément supérieur à l'année et la série temporelle historique. Les résultats issus de la décomposition temps-fréquence étant supposés quasi orthogonaux, chaque coefficient de détermination peut être interprété comme le pourcentage de variance du signal original expliqué grâce à l'élément considéré.

La *figure 40* présente en couleurs les valeurs des coefficients de détermination en fonction de la station et de l'élément y_i considéré. Afin de mettre en évidence d'éventuelles structures spatiales, les stations ont été triées en fonction de leur latitude. À gauche se trouvent les stations les plus au Sud, à droite les plus au Nord. Les lignes verticales permettent de repérer l'équateur et les deux tropiques situés à 30° N et 30° S. En ordonnée se trouve l'élément considéré. Les deux dernières lignes représentent la somme des coefficients de détermination des éléments appartenant respectivement à la VAD et à la VMD.

3. Quels sont les modes de variabilité inter-annuels permettant d'expliquer le signal d'indice de clarté ?

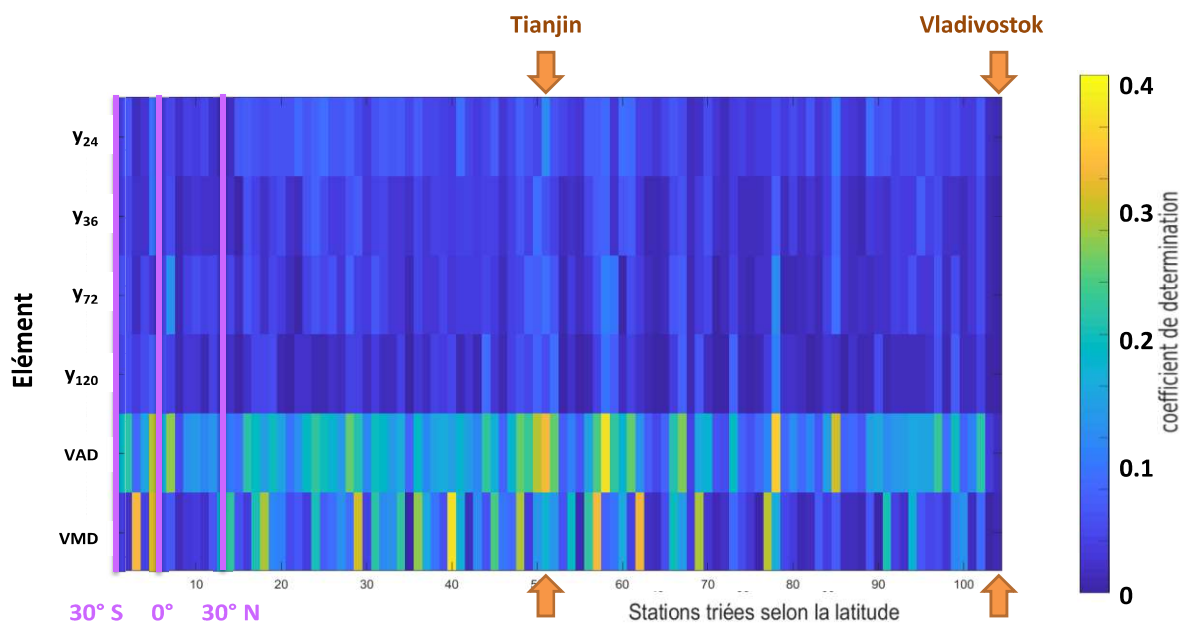


Figure 40 : Répartition de l'importance des variabilités inter-annuelles dans l'explication du signal total.

L'inhomogénéité de la couverture spatiale des 104 stations GEBA est évidente sur la figure 40, puisque la majorité des stations se trouvent au-dessus de 30 °N. Notons qu'il est extrêmement difficile de dégager d'éventuelles structures spatiales des valeurs présentées sur la figure 40. Cependant, une meilleure couverture géographique pourrait possiblement mettre en lumière des structurations spatiales invisibles sur ce jeu de données.

Il n'existe pas, au regard de cette figure, d'éléments de la VAD influençant la variabilité du signal de façon uniforme pour l'ensemble des stations. Pour un élément donné, les écarts maximaux entre les valeurs du coefficient de détermination sont très proches : elles varient entre 0.097 et 0.130. Le pourcentage des variations de la série temporelle historique expliqué par l'élément y_{24} est de 5.0 % en moyenne. Il varie selon la station considérée, entre 0.8 % pour la station russe de Vladivostok (station n°94, n°104 sur la figure 40), et 12 % pour la station chinoise de Tianjin (station n°85, n°51 sur la figure 40). Prendre en compte cette plage de périodes dans l'estimation de l'indice de clarté n'augmentera donc que peu la significativité du résultat pour la station de Vladivostok – ce qui est cohérent avec l'observation précédente soulignant que les variations de cette station sont majoritairement expliquées par les variations intra-annuelles – mais sera intéressante pour la station de Tianjin. Cette diversité de comportements entre les stations est également visible pour les autres éléments. En moyenne, les éléments y_{36} , y_{72} et y_{120} expliquent respectivement 3.6 %, 3.8 % et 2.9 % des variations du signal original. Cependant, la répartition des coefficients de détermination entre les stations n'est pas identique en fonction de l'élément considéré. Cette diversité rend difficile à identifier un élément particulier qu'il serait intéressant de considérer systématiquement pour bien prendre en compte les variations inter-annuelles du signal.

Ensuite, la VAD permet d'expliquer en moyenne 15.3 % des variabilités du signal original.

Bien évidemment, une grande diversité est à nouveau observée en fonction des stations considérées. La station possédant le plus fort coefficient de détermination est la station chinoise de Dunhuang (station n°22, n°58 sur la figure 40), avec un coefficient de détermination de 0.370. Autrement dit, cette station possède des variations expliquées à près de 40 % par la VAD. Cela met en évidence l'intérêt certain de travailler à la caractérisation de la VAD.

Enfin, la VMD permet d'expliquer en moyenne 10.4 % des variabilités du signal original.

Ainsi, en moyenne, il semble plus intéressant de caractériser finement la VAD que la VMD. Cela remet en cause la modélisation des variations inter-annuelles par une simple tendance d'éclairement. Bien évidemment, la réalité à l'échelle du point est différente. Pour certaines stations, telle que la station Xining (station n°100, n°40 sur la figure 40), la VMD est beaucoup plus influente que la VAD : elle explique 12.8 % du signal original, contre 8 % pour la VAD.

En conclusion, nous pouvons dire que la part explicative des éléments dans le signal original dépend fortement de la station considérée. L'analyse de la variabilité des séries temporelles en trois classes a mis en évidence l'intérêt certain de caractériser la VAD pour prendre en compte de manière correcte la variabilité inter-annuelle du signal. Pour certaines stations, la prise en compte de la VMD sera indispensable pour représenter les variations du signal. Cependant, étant donné le faible historique temporel des données utilisées, il nous est pour le moment impossible de caractériser correctement ces modes de variabilité. En ce sens, l'utilisation de bases de données de plus long terme serait d'intérêt. On peut notamment penser aux données de ré-analyse, proposant des données d'éclairement sur plus de 30 ans, ou encore aux modèles climatiques, offrant plus de 150 années de données historiques.

Pour finir, il est important de souligner que les analyses réalisées dans cette section s'affranchissent de la dimension temporelle de chaque élément. En effet, nous avons cherché à savoir dans quelle mesure chaque élément permet d'expliquer le signal global d'indice de clarté mensuel. Cependant, il n'est pas impossible que l'influence d'un élément ne soit pas constante dans le temps. À titre illustratif et exploratoire, la section suivante propose une analyse visuelle des variations temporelles de chaque élément pour deux stations GEBA.

4 L'influence de chaque élément sur le signal original est-elle stationnaire ?

Il est possible pour certaines stations de constater une modification dans le temps de l'amplitude d'un ou de plusieurs éléments. À titre illustratif, les figure 41 et figure 42 présentent deux exemples de décomposition temps-fréquence de séries temporelles d'indice de clarté issues des stations GEBA. Les deux stations considérées appartiennent à des climats différents. La station malaisienne Kuala Lumpur (station 42) possède la variabilité inter-annuelle la plus importante. Cette variabilité est répartie de façon relativement uniforme entre les différents modes de variabilité. La

4. L'influence de chaque élément sur le signal original est-elle stationnaire ?

station japonaise Sapporo (station n°76) possède une variabilité inter-annuelle principalement expliquée par la VMD. Elle a été préférée à la station Xining (station n°40) abordée précédemment, car elle possède un plus long historique temporel. En effet, Sapporo est la station possédant le plus grand nombre de données complètes de la sélection de station GEBA, soit 56 années complètes. L'axe des abscisses varie en fonction de la figure, car il est défini en fonction de la disponibilité temporelle de chaque station. En grisé sont représentées les données affectées par les effets de bords.

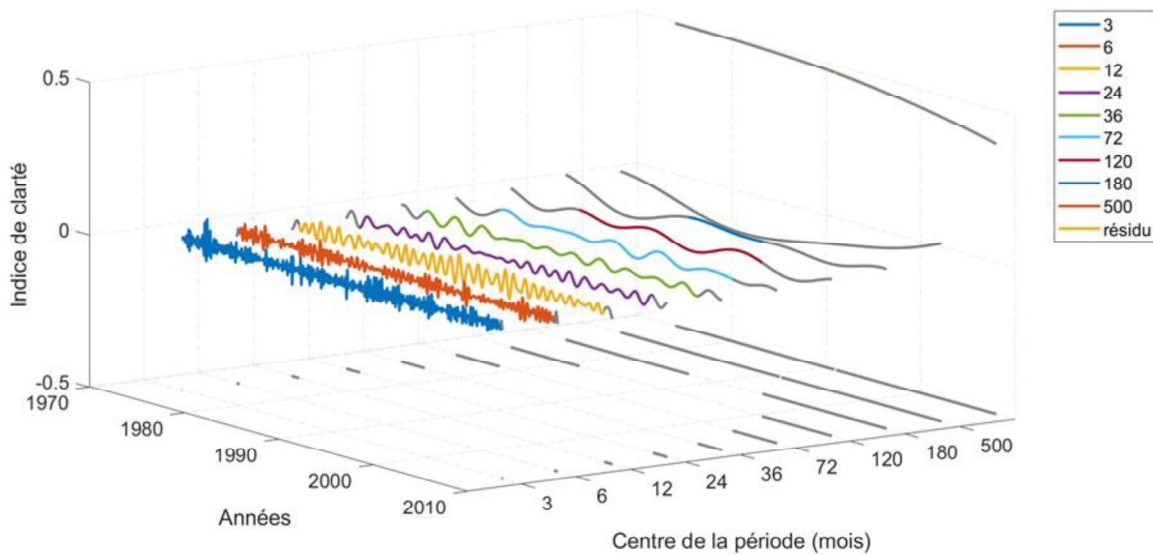


Figure 41 : Vue 3-D de la décomposition temps-fréquence de l'indice de clarté pour la station GEBA Kuala Lumpur. Les zones grisées sont affectées par les effets de bords.

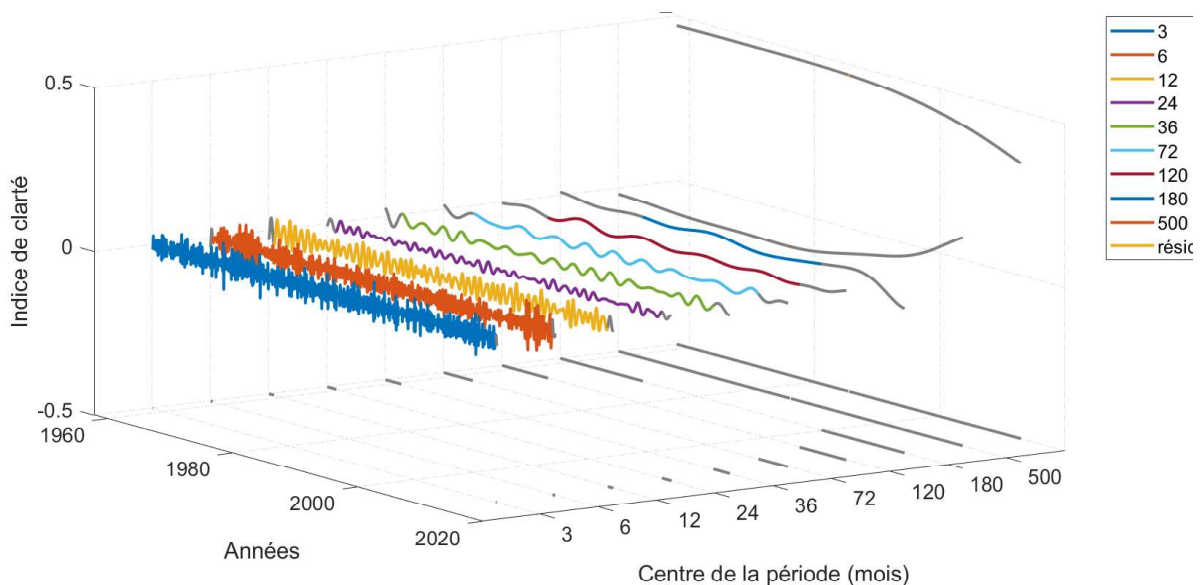


Figure 42 : Vue 3-D de la décomposition temps-fréquence de l'indice de clarté pour la station GEBA Sapporo. Les zones grisées sont affectées par les effets de bords.

CHAPITRE 6 : DECOMPOSITION TEMPS-FREQUENCE DES INDICES DE CLARTES MENSUELS DE LA SELECTION DE STATIONS GEBA

À l'échelle intra-annuelle, les deux éléments en bleu et en rouge connaissent des fortes variations d'amplitudes pour les deux stations. Pour y_3 , des pics d'intensité non périodiques sont observables. On peut par exemple noter pour Sapporo un pic d'intensité particulièrement important en 1977. Pour y_6 , des modulations de périodes chaotiques sont visibles sur les deux stations. Par exemple, des modulations d'amplitude peuvent être vues pour Kuala Lumpur. Ces événements affectant l'amplitude des variabilités ne montrent pas à première vue de structure qui pourrait permettre une explication physique.

En jaune sont représentées les variations annuelles. Pour les deux stations, des variations d'amplitude peuvent être observées. Ces variations ne peuvent pas être imputées à la course du Soleil, étant donné que nous étudions l'indice de clarté. Elles sont donc dues à des phénomènes atmosphériques qui semblent, pour les deux stations, cycliques. Une modulation de l'amplitude différente est toutefois visible en fonction de la station considérée. Par exemple, la station malaisienne connaît une modulation de l'amplitude entre 1980 et 2010, avec un maximum d'amplitude en 1990.

En ce qui concerne la VMD, on constate des amplitudes différentes entre les éléments y_{500} et $y_{\text{résidu}}$ selon la station considérée. La tendance à la baisse observée sur le résidu est plus intense pour la station de Sapporo. Les variations de y_{500} pour la station de Kuala Lumpur (diminution puis augmentation) ne se retrouvent pas pour la seconde station. Cette dernière connaît une forte augmentation des valeurs d'indice de clarté pour les quinze dernières années. Comme annoncé, ces variations ne peuvent être correctement analysées. En effet, elles pourraient être expliquées par un artefact dû aux effets de bords. De même, pour la station Kuala Lumpur, l'élément y_{180} ne possède que quelques points valides. La station Sapporo, qui possède 10 ans d'historique de plus que la station Kuala Lumpur, possède un nombre de données valide permettant une analyse prudente. Cela souligne l'intérêt d'avoir accès à un plus long historique de données pour la caractérisation de ces variabilités de très long terme. On touche à nouveau aux limitations du jeu de données GEBA sélectionné, à savoir le faible nombre de données disponible et la couverture géographique parcellaire des stations, malgré pourtant la qualité exceptionnelle, voire unique, du réseau GEBA sur ces deux critères.

Pour la VAD, des comportements différents sont à nouveau observés sur les deux stations. Pour la station Sapporo, l'élément y_{120} montre une amplitude faible, qui diminue avec le temps. À l'inverse, l'amplitude du signal pour la station Kuala Lumpur est très marquée, avec une légère augmentation d'amplitude avec le temps. Pour cette dernière, l'élément y_{72} montre des variations d'amplitude importante, avec une absence de signal entre 1981 et 1987. Cette anomalie ne semble pas être un artefact, car elle ne se retrouve pas sur les autres éléments pour des périodes temporelles coïncidentes. En effet, les diminutions d'intensité visibles sur les éléments y_{24} et y_{36} mois se trouvent respectivement entre 1987 et 1992 et entre 1982 et 2000. Des conclusions similaires peuvent être faites pour la station Sapporo. Pour y_{72} , un raccourcissement de la période de variations est observé à partir des années 1975 : d'une période régulière de 12 ans, on passe assez brusquement à une période proche de 10 ans. Le raccourcissement de la période coïncide avec une légère augmentation de l'amplitude du signal. Par la suite, aucun signal n'est détecté entre 2000 et 2005. Ce comportement singulier ne semble pas se retrouver sur d'autres stations appartenant au même climat, et pourrait donc être dû à un artefact de la méthode, ou encore à un problème de mesures tel qu'un changement

d'instrument. Cet élément illustre bien qu'en considérant uniquement la valeur de variance calculée à partir de l'ensemble de la période, de nombreux phénomènes peuvent être occultés. Pour γ_{36} , une modulation d'amplitude périodique de l'ordre de 20 ans est visible. Bien évidemment, il est difficile de conclure quant à la représentativité de cette modulation à partir de deux périodes observées de 20 ans. Enfin, pour l'élément γ_{24} , une grande modulation d'amplitude est également visible, avec aléatoirement deux périodes plus intenses que la moyenne.

Ces deux exemples illustrent la diversité de comportement des éléments issus de la décomposition temps-fréquence de l'indice de clarté. Ces derniers varient en fonction des stations, en fonction des éléments considérées, et en fonction du temps. Ainsi, en considérant uniquement la valeur de variance calculée à partir de l'ensemble de la période temporelle, de nombreux phénomènes peuvent être occultés. À nouveau, une analyse temporelle multi-échelle pour le site d'intérêt s'impose.

5 Conclusion

L'analyse des différentes classes de variabilité a été réalisée grâce à l'outil de décomposition temps-fréquence présentée en chapitre 5. En conclusion, une part non négligeable des variations du signal original ne peut pas être expliquée par la VIA. Autrement dit, en ne considérant pas les variabilités inter-annuelles du signal d'indice de clarté mensuelle, plus de 20 % de la variance totale des séries temporelles ne peut pas être expliquée en moyenne. L'influence des variabilités inter-annuelles sur le signal global varie fortement d'une station à l'autre, sans qu'une structure spatiale ne puisse être dégagée. Ces variabilités peuvent être négligeables pour un faible nombre de stations : la part des variations inter-annuelle dans les variations du signal est inférieure à 5 % pour seulement deux stations. Ainsi, pour la majorité des stations, les variabilités inter-annuelles ont été identifiées comme d'intérêt pour une correcte prise en compte des variations du signal.

Les variations de la série temporelle d'indice de clarté de certaines stations peuvent être grandement expliquées par la VAD. En ce qui concerne la VMD, celle-ci est moins influente en moyenne que la VAD. Elle reste cependant indispensable pour expliquer les variations inter-annuelles pour un nombre important de stations. Cependant, il nous est pour le moment impossible de caractériser correctement ces modes de variabilités, l'historique temporel des données utilisées étant restreint. Dans la même idée, la faible couverture géographique de la base de données GEBA utilisée ne nous a pas permis de dégager de structuration spatiale des résultats. Ainsi, pour avancer dans la caractérisation des modes de variabilités de l'indice de clarté, une analyse similaire utilisant des bases de données possédant des séries temporelles de très long terme sur une couverture globale dense spatialement serait intéressante.

Chapitre 7

Décomposition temps-fréquence des indices de clartés mensuels de différentes bases de données

Table des matières

1	Introduction	129
2	Rappel de la méthode d'analyse	129
3	Variabilité des indices de clarté issues de CLARA-A2	130
4	Variabilité des indices de clarté issues de MERRA-2	133
5	Variabilité des indices de clarté issues de IPSL-CM6A-LR	135
5.1	Sélection d'une réalisation du modèle IPSL-CM6A-LR.....	135
5.2	Variabilité des indices de clarté issues d'une réalisation de IPSL-CM6A-LR	140
6	Discussion.....	142
6.1	Analyse des similarités et différences entre les quatre bases de données.....	142
6.2	Discussion des structures de variabilité inter-annuelle identifiées.....	147
7	Variabilité inter-annuelle de l'éclairement.....	152
8	Conclusion.....	155

1 Introduction

L'analyse des différentes classes de variabilité a été réalisée pour les stations GEBA. Pour ces dernières, la prise en compte des variations inter-annuelles s'est avérée nécessaire pour représenter de façon correcte les variations du signal. Il a été mis en évidence que l'analyse spatiale des modes de variabilité observés est difficile, de par la répartition lacunaire des stations GEBA. De plus, il est apparu que bien que la disponibilité temporelle des stations GEBA soit exceptionnelle, le nombre de données reste trop faible pour caractériser de façon robuste les modes de variabilité multi-décennaux.

Dans l'objectif d'augmenter la disponibilité temporelle et spatiale des séries temporelles d'indice de clarté, et ainsi d'aller plus loin dans la caractérisation des modes de variabilité, des données grillées ont été utilisées. Les trois bases de données grillées choisies – CLARA-A2, MERRA-2 et IPSL-CM6A-LR – sont présentées en détail au cours du chapitre 2. La première base de données est issue des mesures satellitales. En effet, il est possible à partir d'images satellites de calculer l'éclairement sur une grille de résolution souvent inférieure à 0.1° , pour quelques dizaines d'années en moyenne. La seconde est un produit de ré-analyse. Elle couvre une période temporelle plus longue que pour les données satellitales, pour une résolution spatiale de l'ordre de 0.5° . Enfin, la troisième base de données est issue des modèles climatiques. Ces derniers offrent un historique de données d'éclairement, et ainsi d'indice de clarté, de 150 ans, tout en conservant une couverture géographique globale à une résolution de l'ordre de la centaine de kilomètres.

Dans un premier temps, l'étude sera réalisée à partir de l'indice de clarté mensuel calculé pour chaque base de données. La section 2 rappelle la méthode d'analyse utilisée. Les sections 3 à 5 proposent une analyse des variabilités estimées à partir des décompositions temps-fréquence de chaque base de données grillées. La section 6 analyse les différences et les similitudes des résultats fournis par l'ensemble des bases de données. Dans un second temps, les résultats de ces analyses seront confrontés à ceux obtenus à partir d'un indicateur couramment utilisé dans la pratique : le coefficient de variation calculé à partir des moyennes annuelles de GHI.

2 Rappel de la méthode d'analyse

La méthode d'analyse est similaire à celle mise en place pour l'étude de l'indice de clarté des stations GEBA. Pour rappel, nous avons traité de façon différenciée les variations intra annuelles (VIA), les variations annuelles à décennales (VAD), et les variations multi-décennales (VMD). Pour cela, une décomposition temps-fréquence des séries temporelles d'indice de clarté a été réalisée pour chaque point de grille des trois bases de données considérées. Les trois premiers éléments, centrées en 3 mois, 6 mois et 12 mois, ont été sommés. Les variations de la série temporelle résultante représente la VIA. Les coefficients de détermination entre cette série temporelle et la série de données historique ont été calculés. Si celui-ci est proche de 1, alors l'influence des variabilités inter-annuelles sur les variations du signal peut être négligée. L'analyse des valeurs de ces coefficients de détermination se trouvent sections 3 à 5.

Les coefficients de détermination entre chaque élément restant (pour rappel y_{24} , y_{36} , y_{72} , y_{120} , y_{180} , y_{500} et $y_{\text{résidu}}$) et la série temporelle historique ont été calculés. Chaque coefficient de détermination représente la fraction de la variance totale expliquée par la variance de l'élément considéré. Les coefficients de détermination entre les éléments de la VAD (y_{24} , y_{36} , y_{72} et y_{120}) et la série temporelle permettront de quantifier l'importance de la prise en compte de tel ou tel mode de variabilité pour un site donné. Leur analyse se trouvent dans les sections 3 à 5. Les éléments de la VMD (y_{180} , y_{500} et $y_{\text{résidu}}$) seront à nouveau traités de façon conjointe, puisque hormis pour IPSL-CM6A-LR, les bases de données ne contiennent que 40 ans de données au maximum. La discussion de la variabilité supérieure à 10 ans est faite au sein de la section 6.

Pour finir, rappelons que la décomposition temps-fréquence décrite dans le chapitre 5 n'est pas parfaitement orthogonale. Autrement dit, la somme des coefficients de détermination n'est pas parfaitement égale à 1. Nous avons vu que pour le cas de GEBA, la moyenne des sommes des coefficients de détermination était égale à 1.04. Pour les trois bases de données choisies, la somme des coefficients de détermination est en moyenne toujours inférieure à 1, avec un minimum pour la base de données MERRA-2 à 0.84. Cela signifie qu'il y a des corrélations significatives entre les éléments de la décomposition que l'on peut interpréter comme un recoupement d'information. Il est intéressant de noter qu'avec le même outil, la fraction d'informations redondante est différente entre les bases de données. Par ailleurs, elle varie de façon importante en fonction du point considéré. Pour faciliter les comparaisons des résultats issus des différentes bases de données, les coefficients de détermination ont été normalisés.

3 Variabilité des indices de clarté issues de CLARA-A2

La figure 43.a représente la répartition géographique des coefficients de détermination calculés entre les séries temporelles mensuelles et la VIA. Les valeurs de coefficient de détermination varient entre 0 et 1, et sont représentée par des couleurs allant du bleu au jaune. On retrouve également en bleu foncé les larges régions exclues de l'analyse pour des raisons de données manquantes. L'identification de ces régions est décrite en section 4.2 du chapitre 2.

La grande majorité des pixels présentent des coefficients de détermination supérieurs à 0.95. Autrement dit, les séries temporelles d'indice de clarté associées au pixel sont expliquées à plus de 95% par la VIA. Trois régions tropicales font exception, avec des coefficients de détermination plus faibles que 0.95 :

- Premièrement, une bande suivant l'équateur dans l'océan Pacifique connaît des coefficients de détermination très faibles, avec un minimum inférieur à 0.2. À l'ouest du Pacifique, cette bande s'étend entre la latitude 10° S et 5°N, pour des coefficients proches de 0.5.
- La deuxième région couvre l'ensemble de l'Indonésie, avec des coefficients de détermination allant de 0.5 à 0.9. En particulier, les coefficients de détermination proches

CHAPITRE 7 : DECOMPOSITION TEMPS-FREQUENCE DES INDICES DE CLARTES MENSUELS DE DIFFERENTES BASES DE DONNEES

de 0.5 de trouvent proche de la Malaisie, et sur la mer de Banda (située entre l'Indonésie, la Papouasie-Nouvelle-Guinée et les Philippines).

- Enfin, la troisième région est l'Afrique équatoriale. Plus précisément, deux zones de coefficients de détermination allant de 0.6 à 0.9 peuvent être identifiées : la zone du Congo et de l'Ouganda, et la zone atlantique à l'Est du Gabon.

On se réfèrera dans la suite à ces trois régions en parlant du Pacifique équatorial, de l'Indonésie et de l'Afrique équatoriale. Pour ces trois régions, une prise en compte de la variabilité inter-annuelle sera nécessaire. Enfin, on peut également souligner la présence de coefficients de détermination proches de 0.8 sur le Nord de l'Amérique du Sud (Colombie, Venezuela, Equateur et Brésil) et sur une partie de l'Océan Indien.

La figure 43.b représente la répartition géographique de la somme des coefficients de détermination des éléments de la VAD. Le détail de la répartition géographique des coefficients de détermination pour chaque élément est présenté en annexe XI, pour l'ensemble des bases de données étudiées.

Comme attendu, une complémentarité est observée entre les deux figures : les pixels très fortement expliqués par la VIA (en jaune sur la figure 43.a) ne sont pas expliqués par la VAD (en bleu foncé sur la figure 43.b). De même, certaines zones peu expliquées par la VIA sont fortement expliquées par la VAD. En ce qui concerne les trois régions tropicales citées précédemment :

- La région de l'Indonésie connaît des coefficients de détermination allant jusque 0.5 pour le Nord de la mer de Banda. Autrement dit, sur cette dernière région, près de 50% des variations du signal sont comprises entre l'année et la décennie. Ainsi, sans prise en compte de ces modes de variabilité, il est possible de passer à côté de près de 50% des variations du signal.
- Les variations de la bande équatoriale sont partiellement expliquées par la VAD. Cette dernière explique jusque 50% sur les 80% des variations manquantes. De plus, la zone Ouest du Pacifique est beaucoup moins étendue latitudinalement sur la figure 43.b que sur la figure 43.a. La bande équatoriale traversant le Pacifique est également moins étendue longitudinalement. Ainsi, pour ces régions, les variations multi-décennales seront nécessaires pour comprendre l'ensemble des variations du signal.
- Enfin, pour la région d'Afrique équatoriale, une partie des variations situées sur les terres sont expliquées par la VAD, alors que les variations situées à l'ouest du Gabon, sur l'océan Atlantique, ne sont pas du tout expliquées par cette classe de variabilité.

Des valeurs de coefficients de détermination entre 0.1 et 0.2 peuvent également être trouvées pour les régions du Nord de l'Amérique du Sud et l'océan Indien.

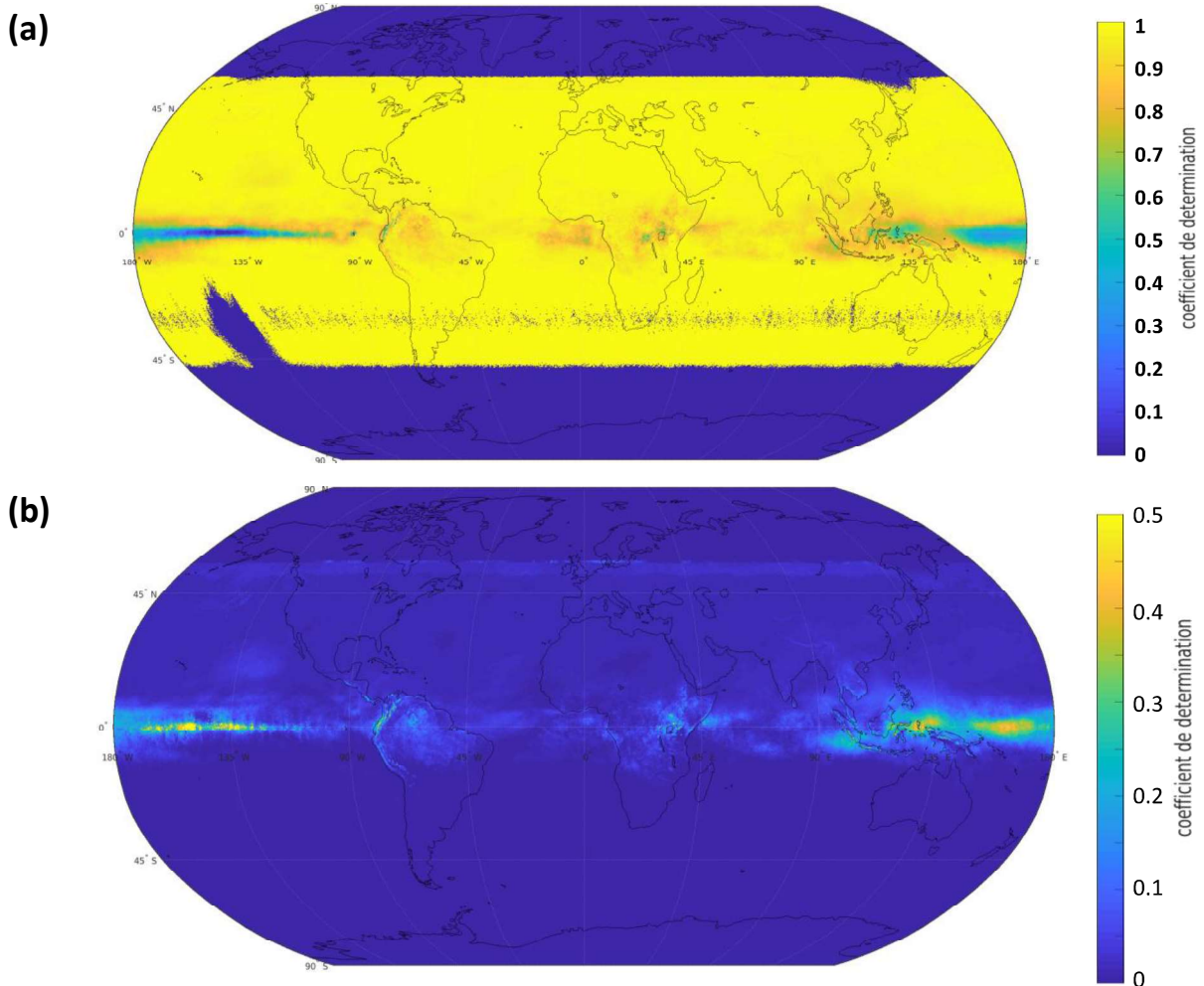


Figure 43 : (a) Répartition géographique des coefficients de détermination entre les séries temporelles mensuelles historiques et représentatives de la VIA. (b) Répartition géographique de la somme des coefficients de détermination entre les séries temporelles mensuelles historiques et représentatives de la VAD.
Base de données CLARA-A2.

En résumé, pour la zone extratropicale, la totalité des variations des pixels peuvent être représentées par la VIA. Pour la zone tropicale, la VAD permet d'expliquer une part non négligeable du signal total de l'indice de clarté, allant jusque 50% sur certaines régions. Ainsi, les modes de variabilité permettant d'expliquer le signal varient fortement en fonction de la région considérée, voir du pixel considéré. Cela se retrouve si l'on analyse la répartition des coefficients de détermination élément par élément. Par exemple, l'élément γ_{24} permet d'expliquer partiellement les variations du signal située en Malaisie (coefficients de détermination proches de 0.2), en moins grande mesure les variations au centre de l'Indonésie (coefficients de détermination autour de 0.1), et expliquent très peu les variations de la région d'Afrique équatoriale ou du centre du Pacifique équatorial (coefficients de détermination quasi nul). À l'inverse, l'élément γ_{120} permet d'expliquer en grande partie les variations du signal pour l'Ouest du Pacifique (coefficients de détermination supérieurs à 0.2), ainsi qu'une partie des variations de la bande équatoriale au centre du Pacifique.

4 Variabilité des indices de clarté issues de MERRA-2

La figure 44.a représente les mêmes quantités que la figure 43.a. D'après Krakauer & Cohan (2017), la variabilité inter-annuelle des bases de données CLARA-A2 et MERRA-2 sont assez similaires. Les résultats issus de ces bases de données devraient donc être assez proches, bien que les périodes temporelles couvertes par les deux bases de données soient différentes (2001-2015 1980-2015).

On constate à nouveau que la majorité des pixels possède des coefficients de détermination entre le signal original et la VIA supérieurs à 0.95. Les variations inter-annuelles seront donc négligeables.

À nouveau, une bande de faibles coefficients de détermination, proches de 0.5, est visible sur le Pacifique équatorial, ainsi que sur l'Indonésie. Cette bande est de largeur relativement constante entre 135° W et le centre de l'Indonésie à 120° E. Elle s'étend entre 10° S et 10° N. La zone d'Afrique équatorial connaît également des faibles coefficients de détermination, autour de 0.6 pour la région terrestre proche du Congo, et entre 0.8 et 0.9 pour la région atlantique à l'Ouest du Gabon. Cette région est plus étendue que précédemment : elle descend jusqu'en bas de l'Afrique.

De nombreuses autres structures caractérisées par des coefficients de détermination inférieurs à 0.95 peuvent être identifiées. Des coefficients de détermination proches de 0.5 peuvent être trouvés le long de la côte Ouest de l'Amérique du Sud. Des régions de coefficients de détermination moyens, autour de 0.6, se trouvent sur les régions océaniques à l'Ouest de la Malaisie et à l'Ouest de l'Amérique du Nord. Enfin, un grand nombre de structures caractérisées par des coefficients de détermination entre 0.8 et 0.9 sont décelables sur la Mongolie, l'Est de la cordillère des Andes ; ou encore sur le Nord de l'Amérique du Sud. Pour l'ensemble de ces zones, la prise en compte des variabilités inter-annuelles sera nécessaire pour une bonne prise en compte des variations du signal.

La figure 44.b représente la même quantité que la figure 43.b : elle illustre la part des variations du signal expliquée par les variations annuelles à décennales. Pour la région équatoriale, il est clair que la grande majorité des variations du Pacifique équatorial est expliquée par la VAD. En effet, pour l'ensemble de la région équatoriale identifiée précédemment, les valeurs de coefficients de détermination excèdent 0.5, à l'exception de l'Ouest de la Malaisie et de l'Afrique équatoriale. Pour ces zones, les variations multi-décennales seront d'intérêt.

Plusieurs des régions identifiées sur la figure 44.a (Mongolie, côte Ouest de l'Amérique du Sud, région pacifique à l'Ouest des USA, Nord de l'Amérique du Sud) ne sont expliquées que partiellement par la VAD. Pour l'ensemble de ces zones, la VMD sera donc à prendre en compte. Enfin, un ensemble de structures parcellaires, caractérisées par des coefficients de détermination entre 0.1 et 0.2, peut être observé : une bande parcourant les USA, l'Atlantique et l'Afrique de 40° N à 30° N, une structure longeant la côte Ouest de l'Afrique à partir de l'équateur, différentes zones dans l'océan Indien, etc.

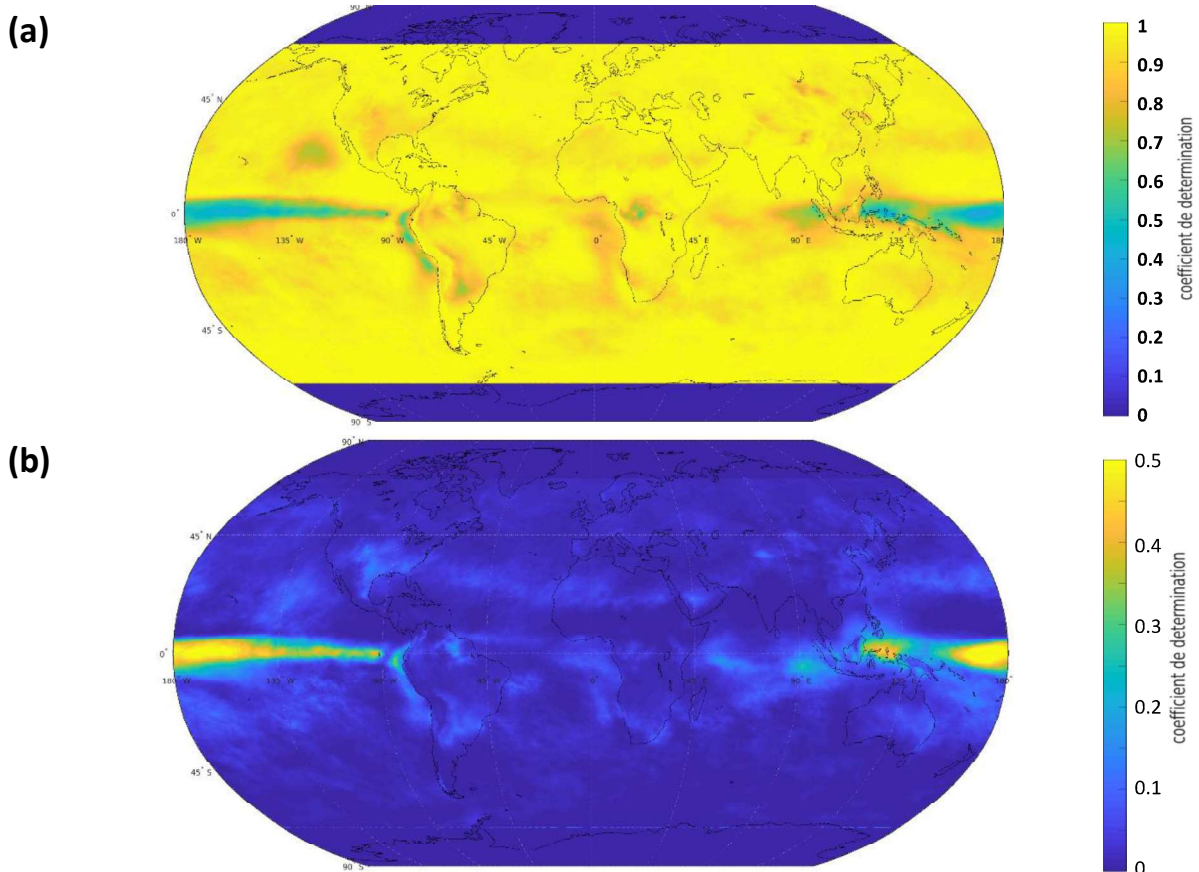


Figure 44 : (a) Répartition géographique des coefficients de détermination entre les séries temporelles mensuelles historiques et représentatives de la VIA. (b) Répartition géographique de la somme des coefficients de détermination entre les séries temporelles mensuelles historiques et représentatives de la VAD. Base de données MERRA-2.

En résumé, une grande diversité des coefficients de détermination peut être à nouveau observée en fonction du point considéré. La région extratropicale reste majoritairement expliquée par la VIA. Pour d'autres régions, prendre en compte la VAD permet d'expliquer plus de 50 % du signal. En ne considérant que les régions pour lesquelles la variabilité inter-annuelle est importante, nous pouvons avancer que la VAD permet d'expliquer une part importante des variations du signal.

Une diversité des structures est également observée entre les différents éléments de la décomposition temps-fréquence. L'élément γ_{24} rassemble une grande proportion des structures caractérisées par un coefficient de détermination entre 0.1 et 0.2 citées précédemment. De plus, il rassemble une grande part des variations du Pacifique équatorial (coefficient de détermination supérieur à 0.2). Dans la même idée, les variations à l'Est de la cordillère des Andes et en Afrique du Sud sont quasiment exclusivement expliquées par l'élément γ_{120} (coefficient de détermination proche de 0.1). Ainsi, pour cette zone, il sera suffisant de prendre en compte l'élément γ_{120} pour bien prendre en compte les variations inter-annuelles du signal. A l'inverse, pour le Pacifique équatorial, il sera nécessaire de prendre en compte l'ensemble des éléments γ_{24} , γ_{36} et γ_{72} . Autrement dit, même pour une région ayant des coefficients de détermination similaires sur la figure 44.b, la prise en compte de la VAD peut être différente en fonction de la région considérée. A ce stade, cela souligne l'importance d'une analyse de l'ensemble des plages de période à l'échelle de la région, voir du pixel considéré.

5 Variabilité des indices de clarté issues de IPSL-CM6A-LR

Comme discuté dans le chapitre 2, la base de données IPSL-CM6A-LR ne peut être considérée de la même manière que les autres bases de données étudiées. En effet, elle est composée de 31 réalisations de 150 ans d'éclairement, transformées en 31 réalisations d'indice de clarté. Pour rappel, ces 31 réalisations permettent de représenter la diversité des climats possibles pour une région donnée. On a ainsi un ensemble de séries temporelles plausibles pour la même période temporelle, définissant une gamme de possibles. La section suivante s'attache à la description de la méthode d'analyse choisie pour traiter cette information probabiliste au regard des attentes de notre étude.

5.1 Sélection d'une réalisation du modèle IPSL-CM6A-LR

Nous avons vu que si les modèles n'ont pas pour prétention de reproduire fidèlement les séries temporelles, il n'y a pas de raison apparente à ce qu'ils ne reproduisent pas fidèlement des quantités indépendantes du temps. Ainsi, l'utilisation de modèles climatiques pour estimer la variance de l'indice de clarté semble adaptée, puisqu'il s'agit d'une quantité indépendante du temps.

Pour les objectifs de notre étude, nous avons choisi de faire une démonstration de la pertinence de la méthode d'analyse proposée (décomposition temps-fréquence, calcul et analyse de coefficients de détermination) pour une seule réalisation. Si cette dernière est choisie avec soin, son analyse sera suffisante pour dégager les structures générales de variances propres au modèle climatique considéré. Bien sûr, dans le futur, il serait possible, et souhaitable, de reproduire la méthodologie proposée pour l'ensemble des 31 réalisations. À partir de ces résultats, il serait alors possible de calculer les densités de probabilités pour chaque quantité d'intérêt. L'information probabiliste dégagée pourra permettre de valider les résultats présentés ici, ainsi que de discuter plus finement de la représentation de la variance propre au modèle IPSL-CM6A-LR.

La première étape pour l'étude de la base de données IPSL-CM6A-LR est donc la sélection attentive d'une réalisation. La sélection d'une réalisation a été préférée à sa construction à partir des 31 réalisations disponibles a été écartée. En effet, en prenant par exemple la moyenne des 31 réalisations, une grande partie de la physique sous-jacente à chaque réalisation serait perdue. De la même manière, en construisant un composite de réalisations, les autocorrélations des séries temporelles de chaque réalisation seraient perdues.

Le cas idéal serait de pouvoir sélectionner la réalisation médiane des 31 réalisations, qui serait ensuite analysée au regard des réalisations maximales et minimales. Cependant, il n'existe pas de réalisation qui soit médiane, maximale, ou minimale pour l'ensemble des pixels. Ceci est illustré par la figure 45. Celle-ci représente les écarts relatifs entre la variance moyenne des 31 réalisations et la variance de chaque pixel pour deux réalisations. En jaune sont représentés les pixels pour lesquels la variance est supérieure à la moyenne, en bleu les pixels pour lesquels elle est inférieure. En vert se trouvent les pixels dont la variance est proche de la variance moyenne. Ainsi, si une réalisation médiane, maximale, ou minimale existait, les cartes seraient respectivement verte, bleue et jaune.

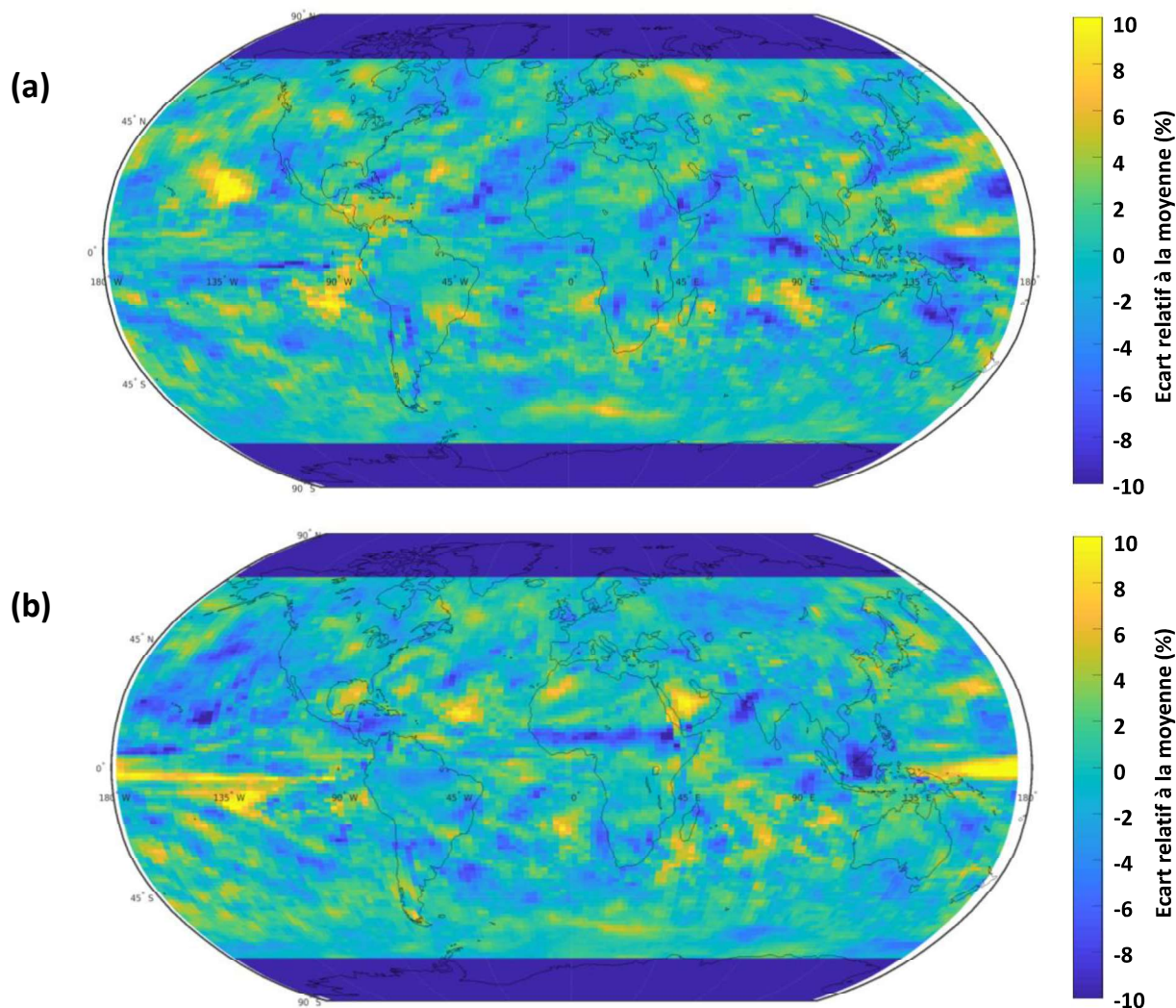


Figure 45 : Exemples illustratifs des écarts relatifs entre la valeur de la variance de deux réalisations du modèle IPSL-CM6A-LR, et la valeur moyenne de la variance entre les 31 réalisations.

Il est évident que pour une réalisation, les écarts à la variance moyenne varient fortement en fonction de la région considérée. Pour la figure 45.a, des écarts importants, de l'ordre de 10 %, peuvent être observés par exemple sur le Pacifique et l'Atlantique Nord, à l'Ouest de l'Amérique du Sud, ou encore au Sud de l'Afrique. D'autres régions, telle que la Brésil ou l'Afrique équatoriale, connaissent peu d'écart à la moyenne. Cela illustre qu'une réalisation peut être maximale sur une région, médiane sur une autre, et minimale sur une troisième région. Les structures observées sur la figure 45.a ne se retrouvent pas sur la figure 45.b, ou des écarts à la moyenne de 10 % sont observés dans le Pacifique tropical, sur l'Afrique tropicale ou sur le Moyen-Orient. **Il est donc difficile d'identifier une réalisation médiane ou extrême pour l'ensemble des pixels. Ainsi, le choix a été porté sur la sélection d'une réalisation particulière.**

Pour assurer la représentativité de la réalisation choisie, nous avons travaillé à identifier les régions pour lesquelles l'ensemble des réalisations sont en accords. Le choix d'une réalisation a été fait de façon à ce que l'écart à la moyenne des 31 réalisations sur ces régions soit faible. Pour cela, nous avons calculé, puis représenté, la dispersion des modèles pour l'écart relatif de la variance de

CHAPITRE 7 : DECOMPOSITION TEMPS-FREQUENCE DES INDICES DE CLARTES MENSUELS DE DIFFERENTES BASES DE DONNEES

l'indice de clarté. Autrement dit, pour chaque pixel, nous avons calculé la différence entre les écarts de variances relatifs maximaux et minimaux trouvées parmi les 31 réalisations. Cette quantité est représentée figure 46. Les régions pour lesquelles la dispersion entre les deux extrêmes est forte sont colorées en jaune ; les régions pour lesquelles les différentes réalisations sont en accord sont colorées en bleus.

Soulignons que la dispersion entre les réalisations peut être considérée comme une incertitude associée à la variance de chaque pixel. Si la dispersion est faible, les résultats issus des différentes réalisations sont en accord. La gamme de possible déterminée par le modèle est alors restreinte. Autrement dit, quel que soit le climat considéré, la valeur de la variance varie peu. Cela revient à dire que l'information relative à la variance est bien caractérisée statistiquement par le modèle. L'incertitude associée à la variance pour ce pixel est donc faible. À l'inverse, si la dispersion est grande, chaque réalisation offre une valeur de variance différente. L'incertitude associée à la valeur donnée du pixel considéré est donc grande.

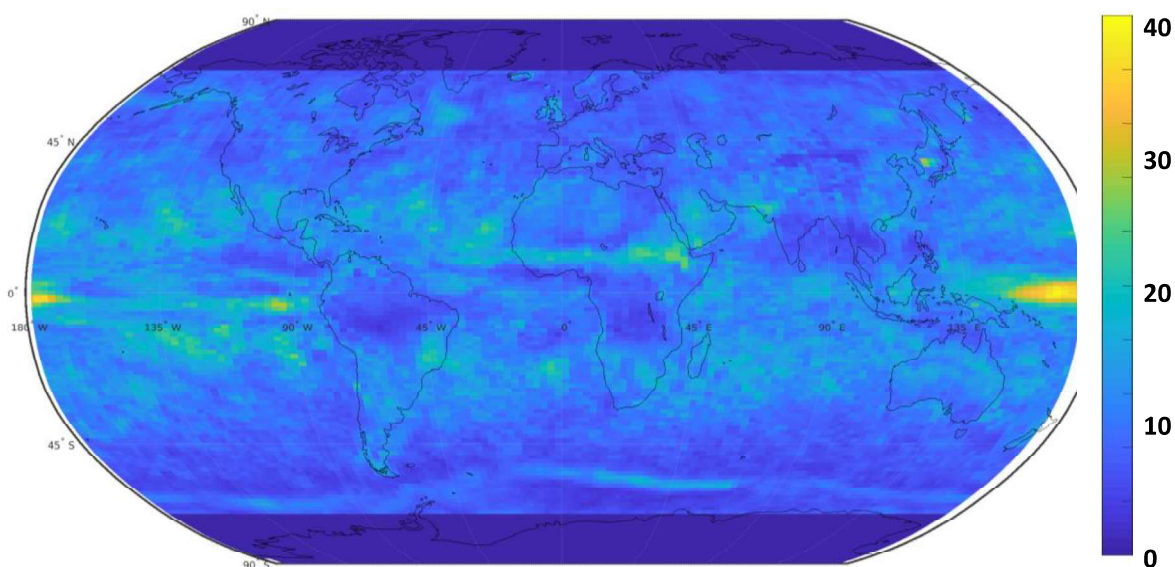


Figure 46 : Dispersion entre les deux réalisations extrêmes des 31 réalisations du modèle IPSL-CM6A-LR. La dispersion est définie par l'écart maximal entre les variances (écart relatif) pour chaque pixel.

La dispersion entre les réalisations du modèle n'est pas uniforme spatialement. En effet, si la majorité des pixels sont caractérisés par une différence maximale de moins de 20 %, certaines régions montrent une dispersion plus importante. En particulier, la dispersion entre les réalisations est très grande pour la zone du Pacifique équatorial situé à l'Est de l'Indonésie, proche de 40 %. Certaines zones à l'Est du Pacifique équatorial (autour de 135° W et 90° W) et de l'Afrique équatoriale (autour de 45° E) se trouvent également sujette à une grande dispersion. Pour ces zones, la variance de l'indice de clarté est représentée de façon très différente en fonction de la réalisation considérée : l'incertitude associée au calcul de la variance est grande.

La figure 47 illustre les écarts relatifs à la moyenne des 31 réalisations pour la réalisation sélectionnée. Pour rappel, on s'attend à ce que l'écart relatif soit faibles pour les zones où la dispersion est faible, et à ce que l'écart relatif ne soit pas extrêmes pour les zones où la dispersion est forte.

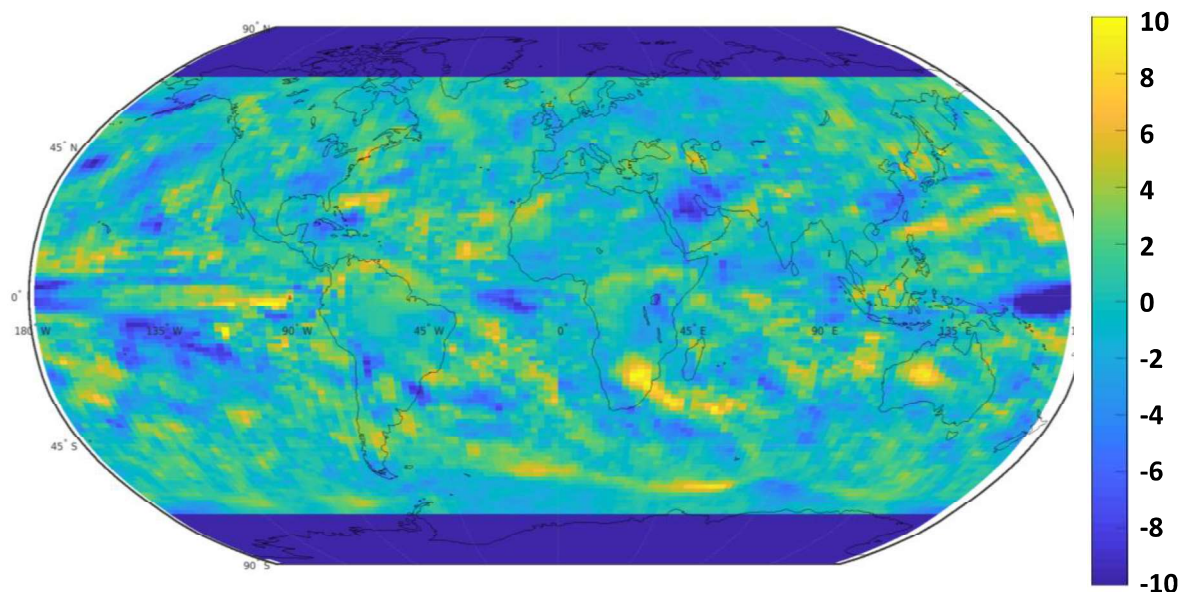


Figure 47 : Ecarts relatifs entre les valeurs de variance de la réalisation du modèle IPSL-CM6A-LR sélectionné, et la valeur de variance moyenne entre les 31 réalisations. L'écart relatif, calculé pour chaque pixel, est exprimé en pourcentage et représentée en couleurs.

Des valeurs importantes d'écarts relatifs (-10 %) sur la zone du Pacifique équatorial à l'Ouest de l'Indonésie peuvent être observées. Cet écart n'est toutefois pas extrême au regard de la dispersion entre les modèles sur cette zone, qui peut atteindre jusque 40 %. Les mêmes constatations peuvent être faites pour les deux zones situées à l'Est du Pacifique équatorial (autour de 135 °W et 90 °W) : l'écart relatif absolu est proche de 10 %, pour un écart maximal entre les différentes réalisations de l'ordre de 35 %. **À l'inverse, les régions du Sud de l'Afrique, du Moyen-Orient, tout comme la région équatoriale de l'océan Atlantique, connaissent un écart relatif important, pour une dispersion entre les réalisations faibles. L'information dégagée par la suite sur ces régions ne devra donc pas être généralisée, car dépend de la réalisation choisie.**

Pour faciliter l'analyse de cette réalisation par la suite, les informations précédentes ont été résumées visuellement sur la figure 48. La figure 48.a illustre en jaune les régions pour lesquelles la dispersion entre les 31 réalisations du modèle est supérieure à 20 %. En vert sont représentées les régions pour lesquelles l'écart relatif entre la réalisation choisie et la moyenne des réalisations est supérieure à 5 %. Dans l'idéal, les zones jaunes et vertes devraient se recouper. La figure 48.b représente les valeurs absolues de la variance pour chaque pixel de la réalisation choisie.

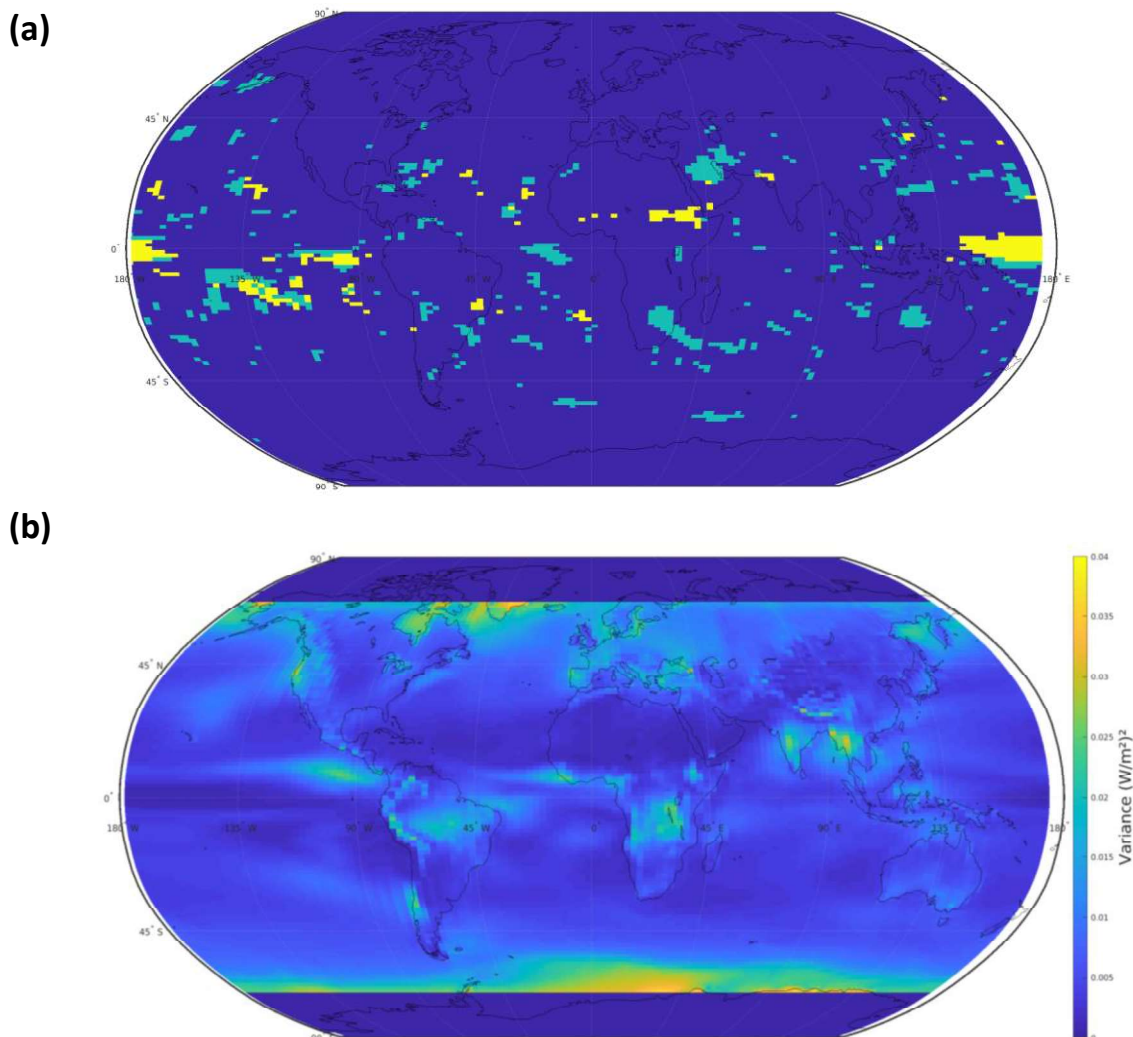


Figure 48 : (a) Représentation des régions à considérer avec attention : pixels dont la dispersion entre les écarts relatifs des différentes réalisations est supérieure à 20% (jaune), pixels dont l'écart relatif entre la réalisation sélectionnée et la moyenne des réalisations est supérieure à 5% (vert). (b) Répartition géographique des variances des séries temporelles d'indice de clarté pour la réalisation sélectionnée.

La figure 48.a résume les constatations faites précédemment en ce qui concerne les régions à considérer avec attention. En jaune, trois principales apparaissent : deux zones dans le Pacifique équatoriale (à l'Ouest de l'Indonésie, et à l'Est du bassin), ainsi qu'une zone à l'Est de l'Afrique équatoriale. Ces zones sont soumises à une forte incertitude. Certaines régions mises en évidence par la couleur verte se recoupent bien avec les structures jaunes. Cela est attendu, puisque la tolérance en termes d'écart relatif à la moyenne est plus grande pour ces zones lors de la sélection de la réalisation. Cependant, certaines structures colorées en vertes ne se recoupent pas avec des zones colorées en jaune. Il s'agit principalement, comme annoncé précédemment, de structures sur le Moyen-Orient, le Sud de l'Afrique l'Ouest, l'Australie et le centre de l'Atlantique équatorial. Les résultats des analyses effectuées sur ces zones seront donc propres à la réalisation choisie, et non caractéristiques du modèle considéré. Soulignons que ces zones ne sont pas nécessairement les zones de forte valeur de variances pour la réalisation considérée.

On constate sur la figure 48.b des structures géographiques dans la variance de l'indice de clarté. Par exemple, des zones de fortes variances peuvent être observées dans l'océan austral, sur le Groenland, à l'Est du Pacifique équatorial et de l'Atlantique équatorial, à l'Ouest de la Thaïlande, ou encore dans la région africaine du Congo. Ces zones ne correspondent ni à des zones de forte dispersion des modèles, ni à des zones pour lesquelles la réalisation considérée est proche de la moyenne. Nous pouvons donc en conclure que le modèle de l'IPSL détecte une forte variance sur ces zones. Ces structures de variabilité sont similaires à celles observées pour MERRA-2 (KRAKAUER & COHAN, 2017).

5.2 Variabilité des indices de clarté issues d'une réalisation de IPSL-CM6A-LR

La figure 49.a représente la même quantité que la figure 43.a et figure 44.a, c'est-à-dire la répartition géographique des coefficients de détermination entre les éléments de la VIA et la série temporelle historique disponibles entre 1850 et 2015. À nouveau, l'essentiel des pixels est caractérisé par des coefficients de détermination supérieurs à 0.95.

Une région sur l'Ouest du Pacifique équatorial montre des coefficients de détermination de l'ordre de 0.5. Différentes structures caractérisées par des coefficients de détermination entre 0.7 et 0.9 peuvent être identifiées : structures dans le Pacifique tropical, dans l'océan Indien, sur le Sud de l'Atlantique, ou encore sur l'Est de l'Australie.

La figure 49.b représente la même quantité que la figure 43.b et figure 44.b, à savoir la répartition géographique des coefficients de détermination entre les éléments de la VAD et la série temporelle historique.

Une forte complémentarité peut être observé entre la figure 49.a et la figure 49.b. En effet, la principale région identifiée sur la figure 49.a, sur l'Ouest du Pacifique équatorial, semble être totalement expliquée par les variations annuelles à décennales (coefficients de détermination de l'ordre de 0.5). Les structures de coefficients de détermination entre 0.7 et 0.9 identifiés sur la figure 49.a se retrouvent de façon extrêmement similaire sur la figure 49.b, avec des coefficients de détermination 0.1 et 0.3. Il semblerait donc que pour cette base de données, l'essentiel des variations inter-annuelles soient comprises entre l'année et la décennie.

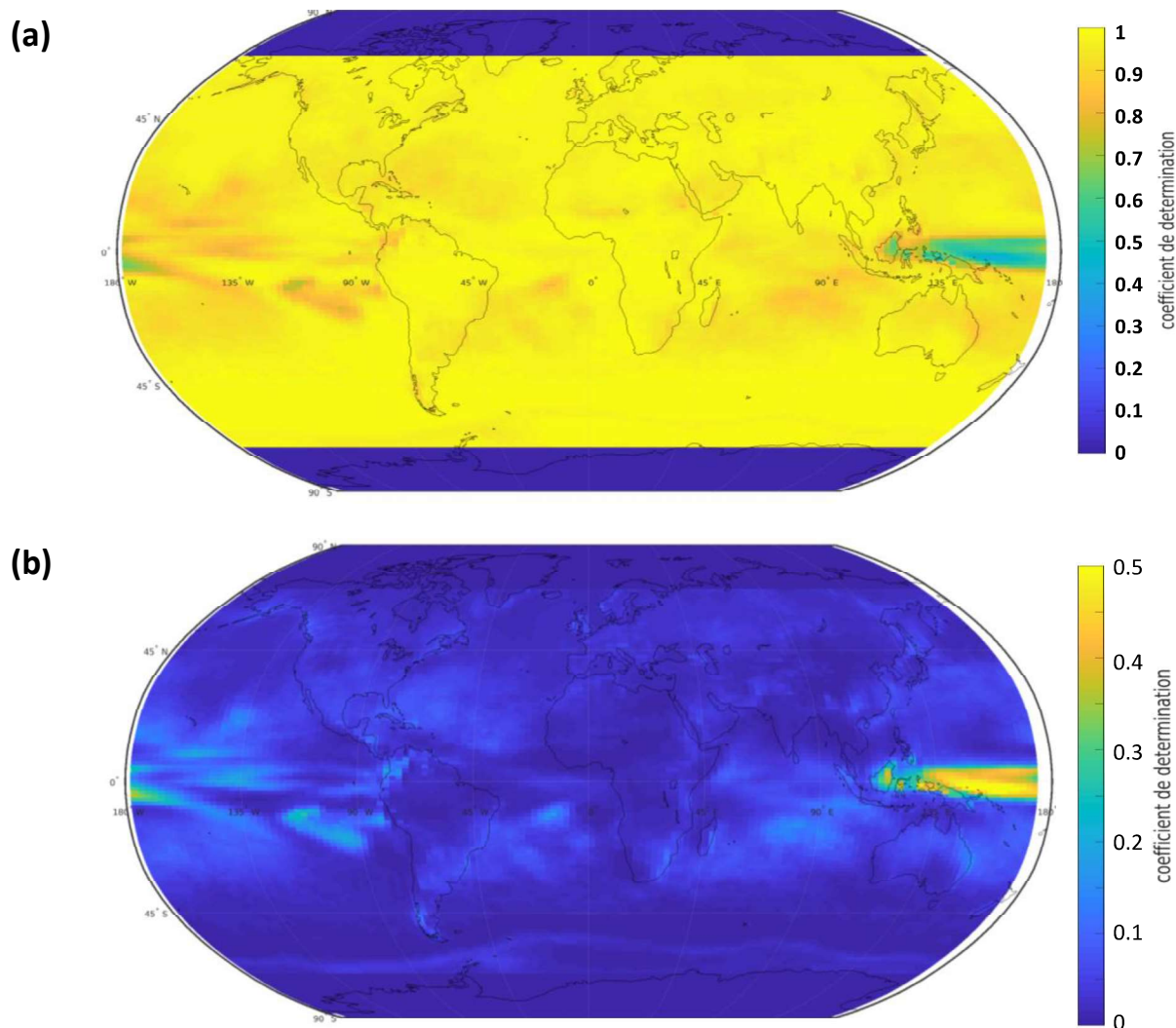


Figure 49 : (a) Répartition géographique des coefficients de détermination entre les séries temporelles mensuelles historiques et représentatives de la VIA. (b) Répartition géographique de la somme des coefficients de détermination entre les séries temporelles mensuelles historiques et représentatives de la VAD.
Base de données IPSL-CM6A-LR.

En étudiant les répartitions géographiques des coefficients de détermination pour chaque élément, il peut être conclu que l'essentiel des variations des séries temporelles d'indice de clarté sont expliquées par les éléments γ_{24} et γ_{36} . Ce dernier élément explique en grande partie les variations du Pacifique équatorial.

En conclusion, nous pouvons avancer que la prise en compte des variations annuelles à décennales est essentielle pour cette base de données : elles permettent d'expliquer la quasi-totalité des variations inter-annuelles. Au sein de cette classe de variabilité, les variations biannuelles et trisannuelles semblent les plus influentes.

6 Discussion

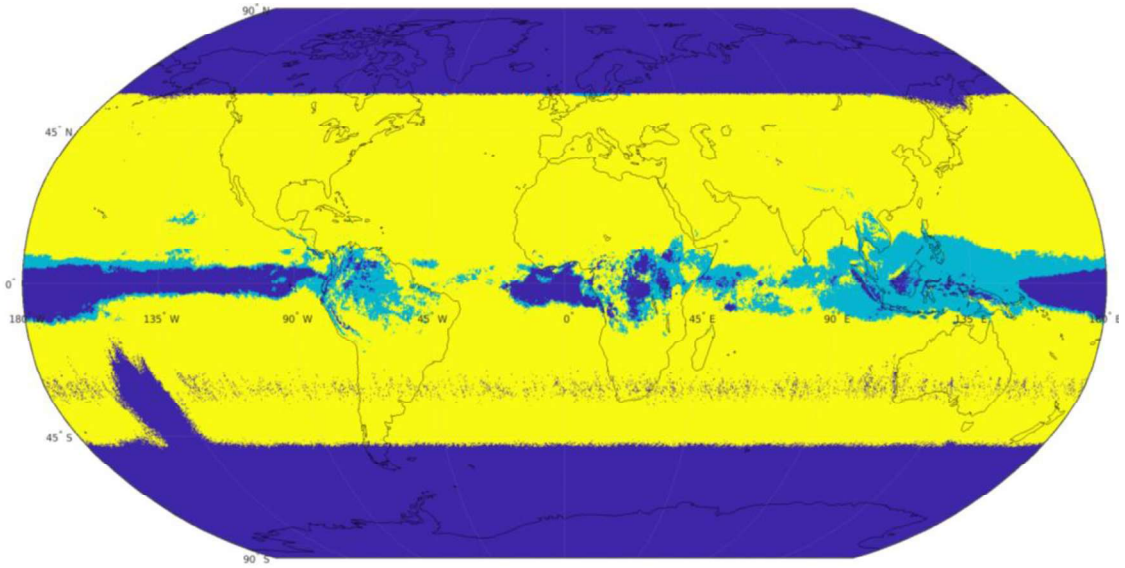
Des différences entre les résultats issus des trois bases de données étudiées peuvent être observées. La section suivante (section 6.1) analyse ces différences. Les résultats issus de la base de données GEBA seront également discutés (cf. chapitre 6). Ainsi, le but de cette section est de mettre en évidence d'éventuelles zones de consensus entre les quatre bases de données considérées au cours de la thèse. Les zones identifiées comme fortement influencées par la variabilité inter-annuelle seront ensuite discutées au regard d'oscillations climatiques connues (section 6.2). Si des similarités sont observées, il serait alors possible d'utiliser la connaissance disponible sur ces oscillations climatiques pour prendre en compte les variations inter-annuelles de l'éclairement dans un contexte prospectif. Les cas de la VAD (section 6.2.1) et de la VMD (section 6.2.2) seront distingués.

6.1 Analyse des similarités et différences entre les quatre bases de données étudiées

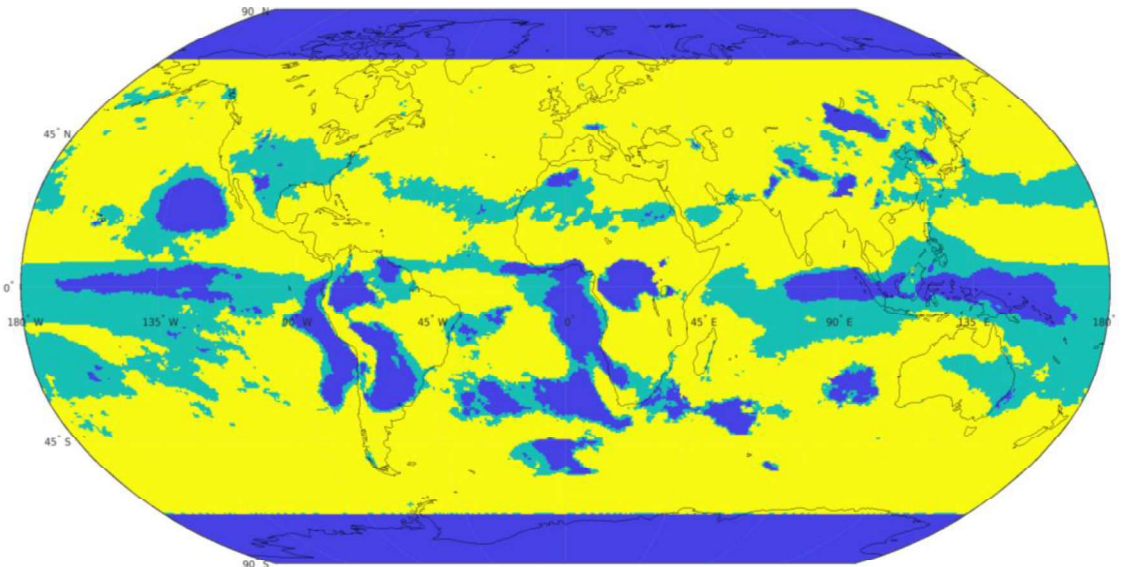
Afin de faciliter la comparaison entre les quatre bases de données considérées, nous avons choisi de représenter de façon visuelle les trois classes de variabilité abordées précédemment : la VIA, la VAD, et VMD. Pour cela, trois classes ont été constituées, puis cartographiées : les régions pour lesquelles la VIA permet d'expliquer plus de 95 % des variations du signal ; les régions pour lesquels la VIA et la VAD permettent d'expliquer plus de 95 % des variations du signal ; et enfin les régions pour lesquelles la VIA et la VAD ne permettent pas d'expliquer plus de 95 % du signal. Pour ces régions, il sera nécessaire de prendre en compte la VMD pour expliquer la même proportion de variation du signal. La première classe est représentée en jaune, la seconde en vert, et enfin, la troisième en bleu foncé sur la figure 50. Une attention devra être portée pour ne pas confondre cette dernière catégorie, illustrée en bleu foncé, des zones qui ne sont pas analysées, également en bleu foncé. La distinction reste toutefois aisée : les zones exclues de l'analyse sont quasi exclusivement des bandes latitudinales situées aux extrémités Nord et Sud des cartes. Elles sont décrites en détail dans le chapitre 2.

Cette classification est pertinente dans le cadre de notre étude. En effet, l'information issue de l'analyse des classes peut être traduite de la façon suivante : pour les régions colorées en jaune, l'essentiel des variations de l'indice de clarté sont dues à la VIA. Etant donné la large sélectivité des plages de périodes, décrite en chapitre 5, l'influence des variations supérieures à deux ans sera donc négligeable. Si on ajoute le fait qu'une observation de deux périodes est préférable pour limiter les incertitudes, l'utilisation quatre ans de données historique seront suffisantes pour prendre en compte de manière correcte l'ensemble des variations de l'indice de clarté. À l'inverse, pour les régions colorées en vert, une prise en compte de la VIA et de la VAD est nécessaire. À nouveau, de par la sélectivité des plages de périodes, les variations supérieures à 15 ans seront négligeables. 30 ans de données historiques seront ainsi requises pour une bonne estimation de l'indice de clarté sur ces régions. Nous verrons qu'une alternative à ce large échantillon de données est la prise en compte de la VAD par la modélisation des oscillations caractéristiques de ces échelles. Pour les zones bleu foncé, des études sur les modes de variabilité supérieurs à l'année devront être établies, pour estimer correctement les variations de l'indice de clarté dans un cadre prospectif.

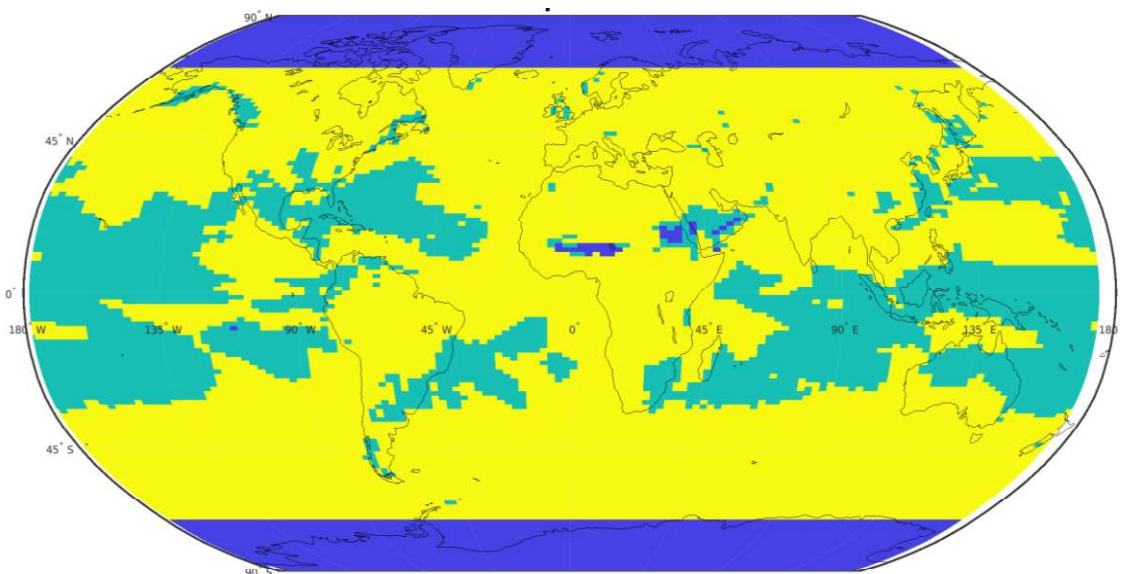
(a)



(b)



(c)



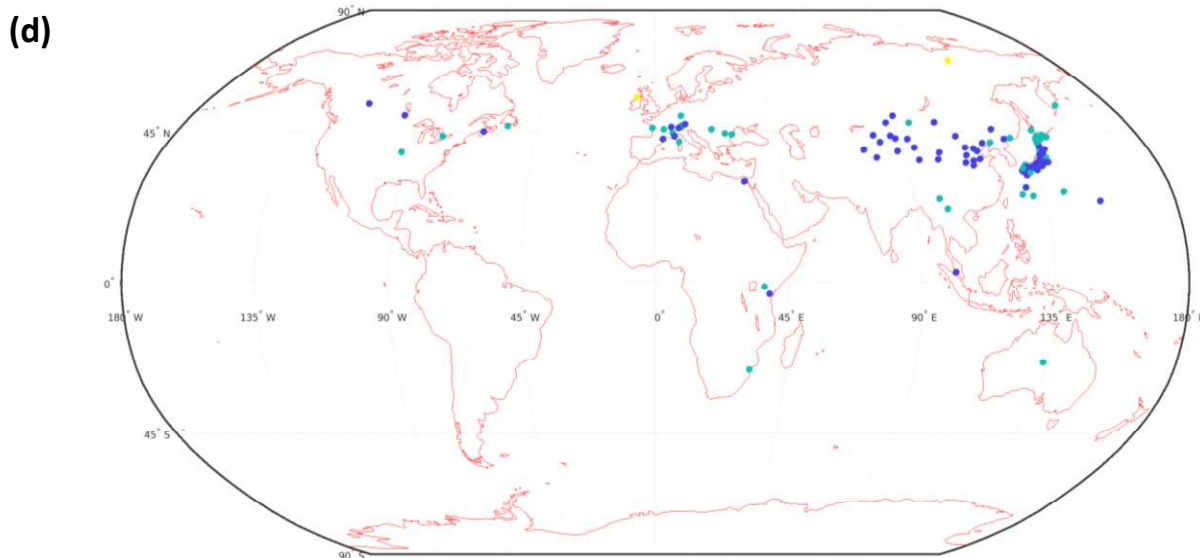


Figure 50 : Régions pour lesquelles les variations de la série temporelle d'indice de clarté est expliquée à plus de 95% par : la VIA (jaune), la VIA et la VMD (vert). En bleu sont représentées les zones pour lesquelles l'ensemble des variations inférieures à la décennie ne permettent pas d'expliquer 95% des variations du signal.
(a) : CLARA-A2, (b) : MERRA-2, (c) : IPSL-CM6A-LR, (d) : GEBA.

Même si elles ne sont pas sur des périodes de temps coïncidentes, il existe entre les trois bases de données grillées des structures similaires. Premièrement, la majorité du globe appartient à la première classe, et est colorée en jaune. Autrement dit, les VIA expliquent pour un grand nombre de pixel plus de 95% des variations du signal.

En ce qui concerne les variabilités inter-annuelles, différentes structures se retrouvent entre les trois bases de données. En particulier, la variabilité du Pacifique équatoriale est commune à l'ensemble des bases de données. Les échelles de variations expliquant cette variabilité ne sont néanmoins pas similaires : pour IPSL-CM6A-LR, elles sont principalement expliquées par la VAD ; pour CLARA-A2, elles sont expliquées majoritairement par la VMD, et pour MERRA-2, un ensemble des deux classes de variabilité est nécessaire pour expliquer ces variations.

L'ensemble des bases de données mettent également en évidence une structure de forte variabilité inter-annuelle l'Indonésie et sur l'océan Indien. Nous pouvons cependant souligner que l'étendue de la région affectée par les variations varie en fonction de la base de données : IPSL-CM6A-LR et MERRA-2 recensent des variations sur l'ensemble du bassin océanique, quand CLARA-A2 les situent sur une zone plus restreinte. Par ailleurs, pour cette région, IPSL-CM6A-LR et CLARA-A2 situent la majorité des variations inter-annuelles entre l'année et la décennie, quand MERRA-2 souligne l'importance des variations supérieures à la décennie sur une partie du bassin.

Pour finir sur les similarités entre les trois bases de données grillées, citons le Nord de l'Amérique du Sud. La variabilité détectée à cet endroit est à nouveau représentée différemment par

les trois bases de données : la zone détectée est beaucoup plus petite pour IPSL-CM6A-LR que pour les deux autres bases de données ; pour CLARA-A2, l'essentiel de la variabilité inter-annuelle est expliquée par la VAD, alors que pour MERRA-2, l'essentiel de la variabilité inter-annuelle est expliquée par la VMD. La zone détectée par cette dernière base de données est également beaucoup plus étendue, couvrant l'essentiel de l'Amérique du Sud, à l'exception du Brésil, de l'Argentine et du Chili.

A ces structures s'ajoutent un grand nombre de structures communes uniquement à deux bases de données grillées sur les trois étudiées. Par exemple, pour IPSL-CM6A-LR et MERRA-2, on peut observer une zone de variabilité dans le Sud-Est des Etats-Unis - expliquées par la VAD pour IPSL-CM6A-LR, et par la VMD pour MERRA-2- une zone de variabilité sur l'Est de l'Australie – expliquée par la VAD uniquement – et une zone sur l'Asie et le Japon – expliquée principalement par la VAD. On peut toutefois souligner que sur cette dernière zone, les régions de variabilités sont plus étendues pour MERRA-2 que pour IPSL-CM6A-LR : influence de la variabilité détectée sur le Nord de l'Inde, la Mongolie, l'intérieur de la Chine par exemple. Enfin, des zones similaires sont visibles pour quelques pixels en Europe, sur l'Afrique du Sud, le Nord-Est de l'Afrique et le Moyen-Orient. Une influence de la VMD est notable pour IPSL-CM6-LR pour ces deux dernières régions. Cependant, on se rappelle qu'il s'agit d'une des zones à traiter avec attention, étant donné la forte incertitude du modèle sur cette région. On privilégiera donc l'information fournie par MERRA-2, soit une influence de la VAD sur ces deux régions. Elles pourraient être reliées à une bande de variabilité identifiable entre 40 et 30° à travers tout l'Atlantique. Cette zone de variabilité recouvre également une grande partie des Etats-Unis d'Amérique.

Des similitudes peuvent également être trouvées entre CLARA-A2 et MERRA-2. On peut par exemple citer l'Est de l'Atlantique équatorial, pour lequel les deux bases de données détectent de la VMD. La région est cependant beaucoup plus étendue pour MERRA-2, longeant l'ensemble de la côte africaine à partir de l'équateur.

En résumé, nous pouvons avancer que beaucoup de similitudes dans les structures de la variabilité inter-annuelles sont communes à au moins deux des trois bases de données grillées considérées. Cependant, ces similitudes cachent une grande diversité des structures en fonction de la base de données considérée : diversité en termes d'extension spatiale, et diversité en termes d'échelles observée, avec une influence majoritaire de la VAD ou de la VMD selon les cas. Pour IPSL-CM6A-LR, la majorité de la variabilité inter-annuelle est explicable par la VAD. Pour MERRA-2, la variabilité inter-annuelle est très influente sur les variations du signal sur un grand nombre de régions, en comparaison avec les autres bases de données. Enfin, CLARA-A2 détectent une influence de la VAD et de la VMD uniquement pour la zone tropicale.

En ce qui concerne la base de données GEBA, deux stations uniquement ont des variations expliquées à plus de 95% pas la VIA : Malin, en Irlande, et Vladivostok, au Nord de la Russie. **Cela répartition n'est globalement pas en accord avec les structures identifiées précédemment grâce aux trois bases de données grillées.** En effet, l'essentiel des stations GEBA sont situées dans la zone extratropicale, zone pour laquelle l'essentiel des variations sont expliquées par la VIA d'après les bases de données grillées. Ainsi, de nombreuses régions dont les variations sont expliquées à plus de 95%

par la VIA ne sont pas retrouvées pour les stations GEBA : majorité de l'Europe, Amérique du Nord au-dessus de 45°N, Egypte, Asie centrale et Japon. Bien que MERRA-2 montre une importance de la variabilité inter-annuelle sur ces deux dernières zones, elle coïncide rarement spatialement avec la variabilité observée sur les stations GEBA.

Pour certaines stations, les variations inter-annuelles du signal sont majoritairement expliquées par la VAD. Ces stations sont réparties entre le Nord de l'Amérique, une partie de l'Europe, l'Australie, le Sud de la Chine, et une partie du Japon. Hormis pour l'Australie, ces résultats ne se retrouvent pas pour les bases de données grillées.

À l'inverse, les stations pour lesquelles l'essentiel de la variabilité inter-annuelle est majoritairement expliquée par le VMD couvrent l'essentiel de l'Asie, une partie du Japon, la Malaisie, l'Egypte et une partie de l'Europe. Ceci est à nouveau en contradiction avec les résultats précédents : pour la Malaisie et le Japon, les trois bases de données grillées trouvent que la variabilité inter-annuelle est majoritairement expliquée par la VAD, pour l'Egypte et l'Europe, elles ne détectent pas d'influence de la variabilité inter-annuelle.

En résumé, nous pouvons avancer que pour l'ensemble des trois bases de données grillées considérées, la majorité des pixels ont une variabilité expliquée à plus de 95% par la VIA. Pour ces régions, utiliser quatre ans de données historique serait suffisant pour prendre en compte de manière correcte l'ensemble des variations de l'indice de clarté. Cependant, cette information ne se retrouve pas pour les stations GEBA. Pour les quatre bases de données, nous avons montré que la VAD était essentielle pour prendre en compte plus de 95% des variations du signal sur certaines régions. Pour prendre en compte correctement cette classe de variabilité dans un contexte prospectif, un minimum de 30 années de données historique est recommandée. Pour avoir accès à un tel historique en tout point, nous avons vu au cours du chapitre 2 que le recours à des données satellitales ou de ré-analyses était d'usage. Cependant, la section précédente a mis en évidence une diversité de la représentation de la VAD entre une base de données satellitale, CLARA-A2, et une base de données de ré-analyse, MERRA-2. Nous avons également vu que la VAD détectée par ces deux bases de données coïncide rarement avec celle détectée par les données des stations GEBA. Ainsi, à ce stade, il est difficile de conclure sur la pertinence de l'utilisation de 30 années de ré-analyses ou de données satellitales pour une prise en compte correcte de la VAD. Enfin, hormis pour IPSL-CM6A-LR, une part non négligeable des pixels sont soumis à des variations supérieures à la décennie.

La section suivante discute au regard des oscillations physiques connues les structures de variabilités observées.

6.2 Discussion des structures de variabilité inter-annuelle identifiées

6.2.1 Variabilité annuelle à décennales (VAD)

Pour estimer la VAD, une alternative possible à l'utilisation de 30 années de données historiques est la modélisation des variations propres à cette classe de variabilité. En effet, si ces variations sont correctement caractérisées et modélisées, il serait possible de les prendre en compte plus aisément. Plus précisément, il serait possible d'ajouter à un échantillon de données in-situ de petite taille, de l'ordre de quelques années, la VAD associée : on ajoute ainsi la variabilité qui ne peut pas être observée dans l'échantillon historique, restreint par sa petite taille.

Un premier pas vers cette modélisation réside dans le rapprochement entre la VAD observée et les oscillations climatiques connues. Par exemple, en ce qui concerne la zone du Pacifique équatoriale, la variabilité observée par l'ensemble des bases de données pourrait être liée à deux oscillations climatiques : ENSO (El Niño Southern Oscillation) et sa contrepartie La Niña. Le phénomène ENSO est connu comme le phénomène de couplage océan-atmosphère le plus important du globe en termes de variabilité interannuelle. Bien que différents indices existent, ENSO est le plus souvent défini par une variation dans la température de surface de l'océan (HANLEY *et al.*, 2003). Cependant, au regard des télé-connexions entre océan et atmosphère, il est attendu que ces variations aient une influence sur la couverture nuageuse, et ainsi sur l'éclairement. Une représentation schématique des anomalies de températures et de couverture nuageuse pour un événement ENSO et La Niña sont présentées par la figure 51.

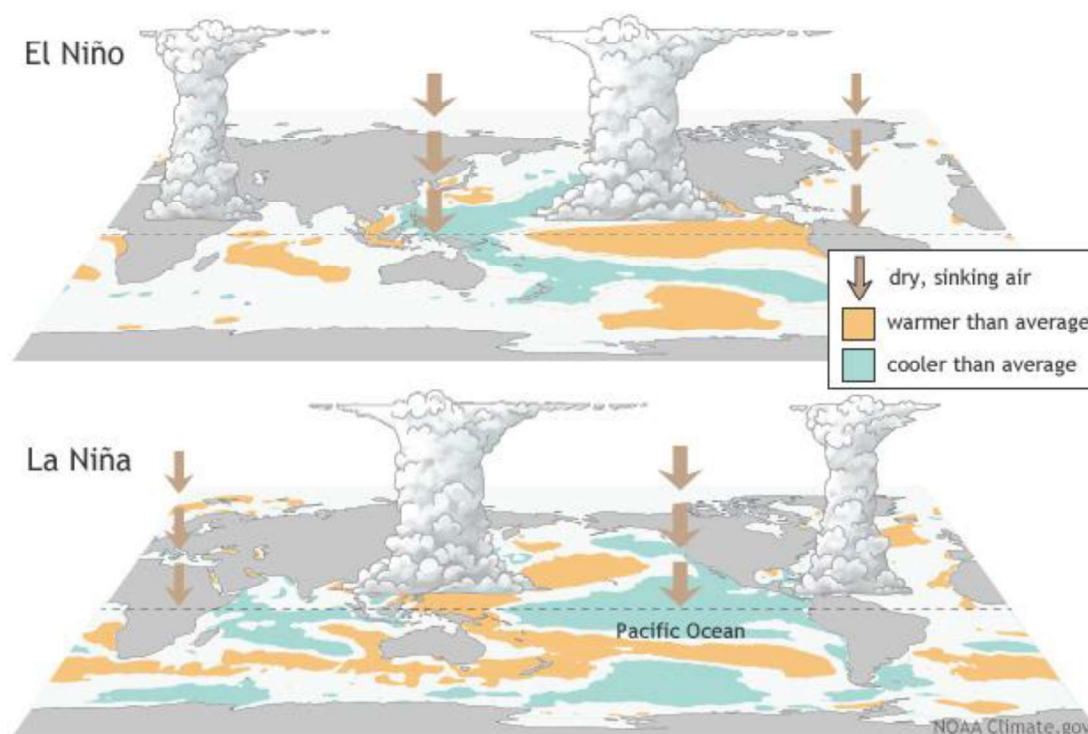


Figure 51 : Représentation schématique des anomalies de températures de surface (couleurs) et des modifications des zones de forte couverture nuageuse lors d'événements ENSO et La Niña.

Figure extraite de NOAA climate.gov

La principale variation de la couverture nuageuse lors des événements ENSO et La Niña est schématisée sur la figure 51 : on observe un déplacement longitudinal des zones de forte couverture nuageuse, avec une forte couverture nuageuse sur l’Afrique de l’Est et le centre du Pacifique pour un événement El Niño, contre une forte couverture sur l’Indonésie et l’Amérique centrale pour un événement La Niña.

Des similitudes peuvent être trouvées dans les structures visibles pour la température de surface et les structures observées pour l’indice de clarté dans les différentes bases de données grillées : variations sur le Pacifique équatorial, sur l’Indonésie, variations à l’Est de l’Australie, influence sur l’océan Indien, etc. Ces similitudes suggèrent un lien entre ENSO et les variations de l’indice de clarté. Ce lien est appuyé par la bibliographie pour différentes régions.

Premièrement un lien entre ENSO et la variabilité des pluies annuelle ou saisonnières a été recensé en Afrique du Sud (JANOWIAK, 1988; MAKARAU & JURY, 1997; NICHOLSON & GRIST, 2003; ROCHA & SIMMONDS, 1997), et en Afrique plus généralement (COOK, 2000; MISRA, 2003; PHILIPPON *et al.*, 2012). Les modifications des régimes de précipitations indiquent une modification de la répartition de la couverture nuageuse, pouvant avoir un impact sur le rayonnement, et ainsi l’indice de clarté. Ce lien est soutenu, par exemple, par la thèse de Peng Li (LI, 2015). Ce dernier montre une forte corrélation (0.6) entre ENSO et le GHI sur le Sud de l’Afrique, mais aussi sur l’Est de l’Océan Indien, contre une forte anti corrélation sur l’Angola, le Congo et la partie subéquatoriale de l’Océan Indien. Ainsi, il est possible qu’ENSO explique une partie des structures observées sur la figure 50.

Ensuite, des liens entre ENSO et le GHI en l’Australie ont également été recensés, notamment par Davy & Troccoli (2012). D’après ces auteurs, l’impact le plus important en termes de rayonnement se trouve en hiver, où la production d’énergie solaire peut varier jusqu’à 10 % entre un événement El Niño et La Niña, selon la zone géographique considérée. Cet ordre de grandeur est compatible avec nos observations, puisque pour MERRA-2 comme pour IPSL-CM6A-LR et GEBA, la part explicative des variations annuelles à décennale sur cette région est trouvée entre 10% et 15% (*cf.* figure 44 et figure 49). Bruce & Cooper (2017) lient également ENSO et la variabilité du rayonnement sur la côte australienne et le Queensland.

Soulignons que ENSO a lieu avec une fréquence d’occurrence comprise entre 3 et 10 ans (MORON *et al.*, 1998). Ainsi, cette oscillation ne peut expliquer que la VAD de l’indice de clarté, observée en vert sur la figure 50. Autrement dit, l’influence d’ENSO n’est compatible qu’avec les variations observées par la figure 50 pour une partie du Pacifique équatorial et de l’Indonésie, le Pacifique Sud, l’Australie et l’Est de l’Asie.

CHAPITRE 7 : DECOMPOSITION TEMPS-FREQUENCE DES INDICES DE CLARTES MENSUELS DE DIFFERENTES BASES DE DONNEES

En ce qui concerne les variations inter-annuelles observées sur l'océan Indien pour l'ensemble des bases de données, il est possible de les mettre en lien avec une troisième oscillation climatique : IOD (*Indian Oscillation Dipole*). Il s'agit d'une oscillation assez irrégulière provenant d'une interaction océan-atmosphère assez similaire à celle d'ENSO (MARCHANT *et al.*, 2007). Elle a d'ailleurs été liée 50% du temps à un événement ENSO ces 100 dernières années (VINAYACHANDRAN *et al.*, 2009). Sa fréquence d'occurrence est ainsi également comprise entre l'année et la décennie. IOD touche en première approximation le climat du bassin indien, de l'Afrique à l'Indonésie (SAJI *et al.*, 1999), jusqu'à l'Australie (UMMENHOFER *et al.*, 2009). Les variations associées à IOD en termes de température de l'océan et de la convection nuageuse sont représentées sur la figure 52.

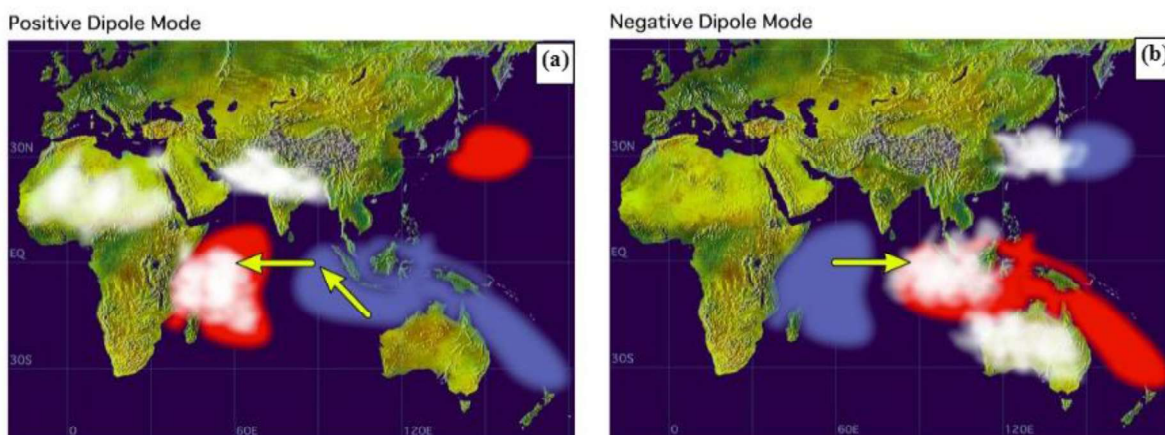


Figure 52 : Représentation schématique des variations de température de l'océan (rouge : anomalie positive, bleu : anomalie négative) et des variations de couverture nuageuses associées (blanc : activité accrue de la convection). Événement IOD. Source: the Japan Agency for Marine-Earth Science and Technology

On comprend au regard de la figure 52 que cette oscillation climatique puisse jouer sur l'éclairement l'océan Indien : la convection atmosphérique est amplifiée à l'Ouest du bassin lors de la phase positive du dipôle, et à l'Est du bassin lors de la phase négative de l'oscillation. Le lien entre IOD et couverture nuageuse est appuyé par la bibliographie (SAJI *et al.*, 1999), comme celui entre IOD et éclairement (LI, 2015).

Les variations observées sur ce bassin pourraient donc être liées à deux oscillations : IOD et ENSO. Avec cette nouvelle approche, nous pouvons également imaginer que les variations observées sur l'Australie, le Japon et l'Indonésie soient dues à des effets connexes de ENSO et de IOD : ces trois zones sont susceptibles de subir l'influence des deux oscillations (*cf.* figure 51 et figure 52). Les variations observées pour les trois bases de données sur l'Afrique équatoriale pourraient également trouver une source dans l'oscillation IOD. Soulignons cependant que la part des VMD mises en évidence par MERRA-2 et CLARA-A2 sur l'océan Indien comme sur l'Indonésie ne pourront pas être expliquées par IOD.

L'hypothèse d'une relation entre des oscillations climatiques (ENSO et IOD) et les variations de l'indice de clarté est donc plausible au regard de la bibliographie. Cependant, un travail conséquent d'attribution des variations de l'indice de clarté aux oscillations abordées reste à réaliser, pour valider ou invalider les liens proposés ici. Une première approche de ce travail est proposée par Krakauer & Cohan (2017). Les auteurs proposent des liens entre les données d'éclairement issues de MERRA-2 et différentes oscillations climatiques. Les auteurs identifient deux oscillations particulièrement influentes sur l'éclairement selon eux : PDO (*Pacific Decadal Oscillation*) et ENSO. La PDO sera discuté par la suite, car influe sur la VMD et non la VAD. D'après eux, ENSO pourrait expliquer près de 8 % des variations de l'éclairement pour les basses latitudes (entre 0° et 20°).

6.2.2 Variabilité multi-décennale (VMD)

En ce qui concerne la VMD, celle-ci peut recouvrir deux variabilités distinctes : les tendances de long terme, et les oscillations de périodes supérieures à la fenêtre d'analyse. Il est en effet possible de recenser un certain nombre d'oscillations climatiques de périodes supérieures à la décennie, pouvant avoir un impact sur l'éclairement. Par exemple, les variations multi-décennales observées dans le Pacifique équatorial et sur l'Indonésie par MERRA-2 et CLARA-A2 pourraient être liées à la PDO.

La PDO est une oscillation climatique touchant des zones géographiques très similaires à ENSO. Elle est également mesurée par une anomalie dans les températures de surface de l'océan. Des études mettent également en évidence des liens entre PDO et déplacement de la trajectoire des systèmes atmosphériques, et donc de la couverture nuageuse (QASMI et al., 2017; W. WANG et al., 2016). Cette oscillation cyclique connaît une période comprise entre 20 et 30 ans. Elle pourrait donc contribuer à expliquer la partie des variations observées sur le Pacifique équatorial qui ne peuvent pas être expliqués par les variations annuelles à décennale, ou encore la zone de variabilité observée sur le Sud-Est des Etats-Unis.

Enfin, soulignons que les variations inter-annuelles mises en évidence par les bases de données grillées sur l'océan Atlantique, les Etats-Unis et l'Amérique du Sud pourraient être liées à l'AMO (*Atlantic Multi-decadal Oscillation*) (ALEXANDER et al., 2014). Cette oscillation est assez proche de la PDO, mais a lieu dans l'océan Atlantique avec une période de l'ordre de 50 à 80 ans (KNUDSEN et al., 2011). Elle est définie à partir de la moyenne spatiale de la température de surface de l'océan dans l'Atlantique Nord. L'impact de l'AMO le plus documenté est celui sur les précipitation (Pour les Etats-Unis : Enfield, Mestas-Nuñez, et Trimble 2001; pour l'Afrique Sahélienne et la région Nord Est du Brésil : Knight, Folland, et Scaife 2006). L'AMO pourrait également être mise en lien avec la VMD observée par GEBA sur l'Europe. L'influence de cette oscillation sur la saison estivale de l'Europe a été exploré par Sutton and al.(2005). Ceux-ci mettent en évidence le rôle de l'AMO dans certaines vagues de chaleur européenne, dont l'été 2003. Ces résultats, confirmés par Cattiaux et al. (2011), suggèrent une relation entre AMO et schémas de circulations atmosphériques particuliers en Europe. L'influence de tels événements, et ainsi de l'AMO sur le rayonnement semble donc très probable.

Pour le cas de la VMD, le travail d'attribution des variations observées sur l'indice de clarté aux oscillations climatiques citées est plus complexe. En effet, les limitations du nombre de données disponible rendent difficile le travail de caractérisation de ces oscillations. Des études pourraient être faites grâce aux données IPSL-CM6A-LR, qui offre 150 ans de données historiques. Cependant, cette dernière base de données ne détecte que très peu de VMD.

On peut souligner qu'un manque de détection d'oscillations caractéristiques de la VMD a été documenté pour le cas général des modèles de CMIP5. Luo *et al.* (2018) soulignent que les modèles ont une propension à minimiser les tendances de température de l'océan. Cela pourrait être du, d'après les auteurs, à un manque de contraste entre les températures de surface des trois bassins océaniques. La quasi-totalité des oscillations citées précédemment étant définies par des variations de la température de l'océan, il est possible que cela puisse expliquer une déficience des modèles dans la représentation de la VMD. Qasmi *et al.* (2017) soulignent quant à eux que les modèles de CMIP5 sous-estiment les télé-connections entre l'océan et l'atmosphère, ainsi qu'une mauvaise prise en compte des rétroactions entre la température de surface et les nuages. Cela peut expliquer pourquoi les oscillations de long terme, principalement océaniques, ne sont pas décelables pour les variables atmosphériques telles que l'éclairement. L'AMO était également sous-estimée par les modèles CMIP5 (KNIGHT *et al.*, 2006; QASMI *et al.*, 2017). Comme annoncé au cours du chapitre 2, étant donné les modifications substantielles faites aux modèles entre les versions de CMIP5 et CMIP6, ces résultats pourraient être différents pour CMIP6. Ainsi, une étude plus approfondie de la représentation de la VMD dans les modèles de CMIP6 reste à réaliser, et en particulier du modèle IPSL-CM6A-LR étudié ici.

Des études complémentaires sont donc nécessaires pour caractériser la VMD, et ainsi pouvoir la prendre en compte de manière pertinente dans l'estimation du potentiel solaire. Soulignons cependant que notre étude a permis de mettre en évidence les zones pour lesquelles la VMD est particulièrement importante dans les variations du signal. Ainsi, la prochaine étape dans l'étude approfondie de la VMD pourrait être concentrée sur ces zones particulières.

Pour conclure, soulignons que la méthodologie développée au cours de cette thèse a été appliquée à l'indice de clarté mensuel. Il en a résulté une classification spatiale de la variabilité temporelle de ce dernier. Une telle classification est d'un intérêt certain pour envisager une prise en compte avisée de la variabilité inter-annuelle dans les méthodes d'estimation du potentiel solaire. En particulier, les méthodes de génération de données mensuelles citées au cours du chapitre 2 pourraient en bénéficier pour ajuster le nombre de données historiques utilisées. Il serait également possible d'utiliser l'information dégagée par cette classification et par les résultats de la décomposition temps-fréquence pour ajouter aux données mensuelles générées la variabilité manquante dans l'échantillon historique. Cependant, une classification similaire pour les données d'éclairement à l'échelle annuelle est également pertinente. En effet, nous avons vu que cette dernière quantité est celle couramment utilisée dans un contexte industriel (*cf.* chapitre 1 et 2). Ainsi, la section suivante étudie la similarité entre la variabilité inter-annuelle du GHI et les variabilités identifiées par notre étude. Si celles-ci sont proches, les résultats de notre étude pourront être également exploités dans un contexte industriel.

7 Variabilité inter-annuelle de l'éclairement

L'objectif de cette section est d'identifier dans quelle mesure les résultats obtenus à partir de séries temporelles d'indice de clarté mensuel sont transposables pour des séries temporelles d'éclairement annuel. Pour cela, nous avons étudié la similarité de nos résultats avec ceux issus d'un indicateur couramment utilisé dans la pratique industrielle : le coefficient de variations (CoV). Pour rappel, celui-ci correspond au ratio entre l'écart type et la moyenne de la série temporelle historique. Il est calculé à partir de données d'éclairement annuel. Il évalue l'importance des variations inter-annuelles par rapport à la moyenne du signal.

Nous avons estimé le CoV pour le cas de MERRA-2. Cette base de données étant la plus sujette à la variabilité inter-annuelle, son étude nous a semblé la plus intéressante. Nous nous sommes à nouveau appuyés sur les résultats de la décomposition temps-fréquence réalisée. L'équivalent d'un signal annuel a été construit en sommant les éléments issus de la décomposition supérieurs à 12 mois. Autrement dit, la VIA des signaux d'éclairement a été ôtée. Cette méthode se justifie par le fait que l'information propre aux variations du signal est plus riche pour ce signal que pour une simple moyenne annuelle d'éclairement (cf. chapitre 5).

Deux classes de variabilité ont ensuite été construites en fonction de la valeur du CoV : inférieure ou supérieure à 5 %. Les deux classes de CoV ont ensuite été comparées aux trois classes distinguées précédemment. Cependant, il ne sera pas possible de comparer rigoureusement les structures identifiées pour les deux quantités. En effet, nous n'avons malheureusement pas trouvé de relation directe entre le seuil appliqué sur le coefficient de détermination calculé précédemment et le seuil à appliquer sur le CoV. De plus, aucun seuil précis portant sur la valeur du CoV n'a été identifié dans la pratique des monteurs de projet. Ainsi, le seuil de 5 % a été choisi arbitrairement, dans le but de faciliter le discours. On représentera également la valeur absolue du CoV, afin de permettre d'ajuster le seuil si nécessaire.

Si les seuils de CoV et de coefficient de détermination étaient équivalents, alors les classes suivantes devraient être coïncidentes : CoV inférieur à 5 % et signal expliqué à 95 % par la VIA, CoV supérieur à 5 % et signal qui ne peut être expliqué à 95 % par la VIA. Ainsi, cette dernière classe de CoV devrait être coïncidente avec deux classes identifiées précédemment, regroupant l'influence de la VAD et de la VMD.

La figure 53 présente les deux classes de variabilité du CoV, ainsi que la valeur des coefficients de variations pour chaque pixel de la base de données MERRA-2.

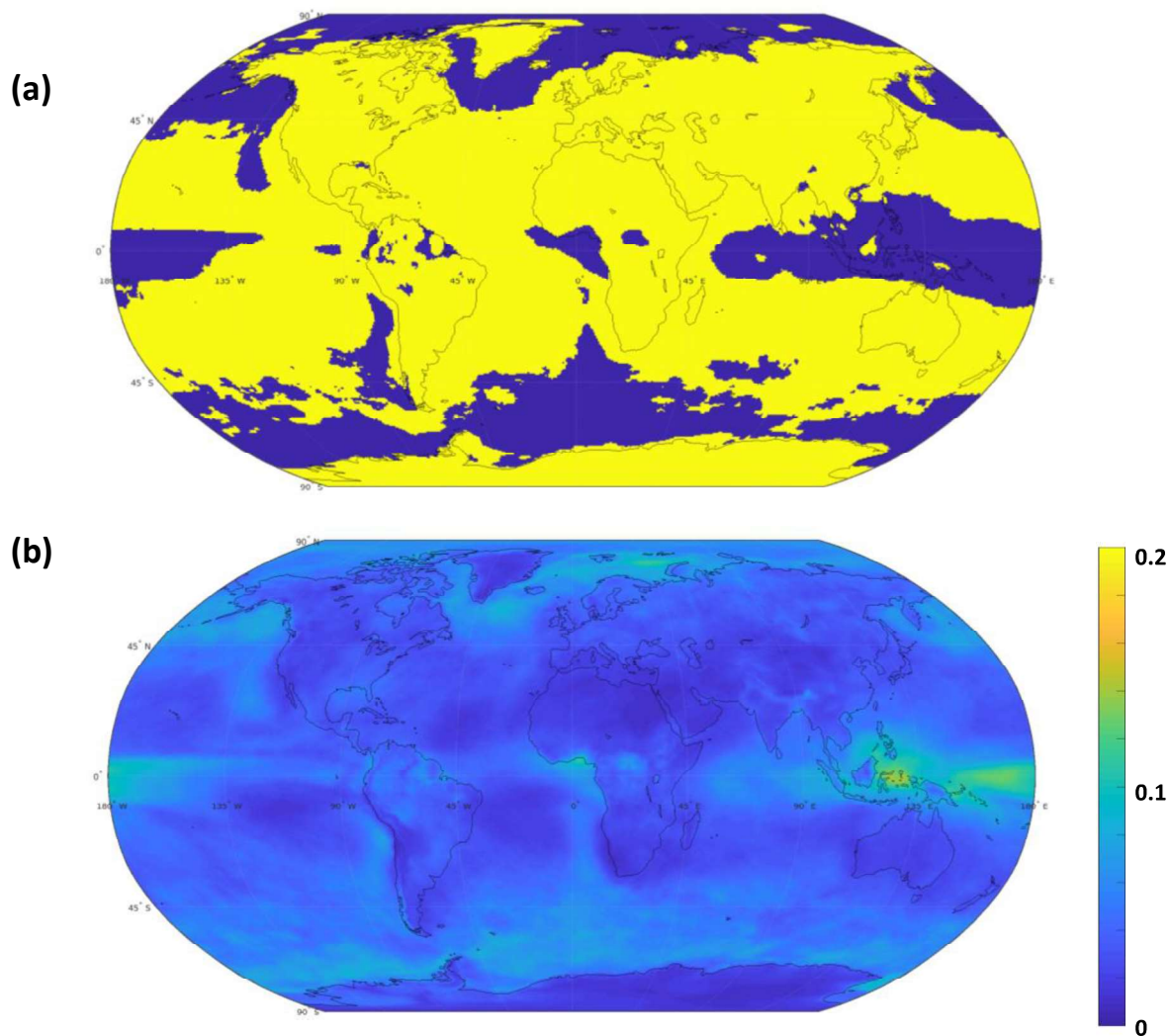


Figure 53 : (a) Régions pour lesquelles le coefficient de détermination est supérieur à 95 % (bleu), ou inférieur ou égal à 95 % (jaune). (b) Coefficient de variation des séries temporelle annuelle d'éclairement GEBA. Base de données MERRA-2.

Des structures similaires peuvent être observées entre les classes de variabilité issues de l'éclairement et de l'indice de clarté (cf. respectivement figure 50.c et figure 53.a). Des zones de fortes variabilités peuvent par exemple être observées dans la Pacifique équatorial, sur l'Indonésie et l'Océan Indien, sur l'Atlantique équatorial à l'Est du Gabon et dans les Terres, à l'Est de l'Amérique du Sud, ainsi que sur le Brésil.

À l'inverse, des différences peuvent également être répertoriées. Certaines zones identifiées comme variables précédemment ne se retrouvent pas ici. Par exemple, la bande parcourant l'Atlantique des Etats-Unis au Moyen Orient n'est pas visible ici. Il en est de même pour la variabilité observée sur l'Asie. Cela peut s'expliquer par le fait que si les variations du signal annuel sont très faibles, le coefficient de variation lui aussi faible, que les variations inter-annuelles soient responsables ou non de ces faibles variations. Dans notre étude, nous avons identifié la part des variations inter-annuelles expliquant les variations du signal quel que soit l'amplitude de ces dernières. Ainsi, notre étude est plus fine que celle réalisée à l'aide du coefficient de variation. Cependant, l'information

dégagée pourra être trop précise pour un industriel : pour estimer le potentiel solaire, il n'est pas nécessaire de prendre en compte des variations inter-annuelles qui soient d'amplitudes négligeables. Ainsi, étudier conjointement les valeurs absolues de la variance aux cartes présentées précédemment pourrait éviter l'application d'analyses complexes pour prendre en compte la variabilité inter-annuelle sur des pixels de variabilité faible.

Enfin, le CoV détecte des variations importantes sur le Nord et le Sud de l'ensemble des bassins océaniques. Ces zones n'étaient pas identifiées comme soumises à une forte variabilité inter-annuelle sur la figure 50.c. Il serait possible que sur ces zones, la variabilité intra-annuelle soit d'une telle importance qu'elle masquerait la forte variabilité inter-annuelle lors du calcul du coefficient de détermination. Une autre hypothèse réside dans le fait que proche des pôles, la moyenne inter-annuelle d'éclairement est nécessairement plus faible. Ainsi, les valeurs du CoV seront artificiellement amplifiées. Cette hypothèse est appuyée par les valeurs de moyennes multi-annuelles de MERRA-2, présentée en figure 6 du chapitre 1. Sur cette figure, on retrouve des zones de faibles moyennes pour les zones océaniques proches des pôles. Ces zones pourraient être dues à des artefacts de la méthode utilisée pour obtenir les données MERRA-2. Cela est soutenu par Hinkelman (2019), qui montre une sur-représentation de la couverture nuageuse dans les océans du Sud, et pour les régions océaniques des régions polaires de manière générale. Les régions identifiées par l'auteur comme biaisées coïncident avec les zones de fort coefficient de détermination observée sur la figure 53.b. Cette hypothèse est également appuyée par le fait qu'il n'a pas pu être répertorié au cours d'analyses complémentaires de variabilité particulièrement forte de la couverture nuageuse sur ces régions. Ainsi, ces zones de variabilité remettent plutôt en question la pertinence de l'utilisation du CoV pour des zones situées aux hautes latitudes que la méthode d'analyse appliquée au cours de cette thèse.

Cet exemple souligne certaines similarités entre notre étude et les résultats issus de l'analyse du coefficient de variation calculée à partir de moyennes annuelles. Les structures générales sont conservées. Cependant, notre étude permet de s'affranchir des limites propres au calcul du CoV. Elle offre également des résultats plus fins. Il a été mis en évidence l'intérêt de considérer les valeurs absolues de la variance en plus des cartes présentées précédemment.

8 Conclusion

L'analyse des modes de variabilité mis en évidence par la décomposition temps fréquence de trois bases de données grillées a été réalisée. Différentes structures de variabilités ont pu être observées pour les données satellitales, de ré-analyse, et de modèle climatique. Pour l'ensemble des bases de données grillées, 95% des variations du signal d'indice de clarté sont expliquées par la variabilité intra-annuelle pour la plus grande majorité de la zone extratropicale. Pour ces zones, les variations inter-annuelles peuvent être négligées. Nous avons conclu sur le fait que l'utilisation d'un échantillon de données historique de quatre ans était suffisante pour estimer correctement le potentiel solaire d'un site appartenant à une de ces zones. Cependant, cette information n'a pas été observée pour les stations GEBA, pour lesquelles seulement deux stations satisfont ce critère.

Pour l'ensemble des quatre bases de données, une part importante des variations inter-annuelles détectées ont pu être attribuées aux variations annuelles à décennales. Les régions pour lesquelles cette classe de variabilité a été jugé particulièrement importante ont été discutées au regard d'oscillations climatiques. Il a été suggéré qu'une étude approfondie pourrait permettre une caractérisation de ces modes de variabilité, dans le but de les incorporer à l'estimation du potentiel solaire.

Enfin, certaines régions du globe ont montré que la prise en compte des variations multi-décennale était nécessaire pour une estimation complète des variations du signal. L'étendu de ces régions est particulièrement réduite pour la base de données IPSL-CM6A-LR. Quelques oscillations climatiques pouvant avoir un effet sur les variations multi-décennales ont été abordées. Face au nombre de données important nécessaire à la bonne caractérisation de cette classe de variabilité, l'accent a été mis sur l'intérêt d'approfondir les connaissances des variations supérieures à la décennie.

La méthodologie proposée ici laisse entrevoir la possibilité d'une classification spatiale des zones pour lesquelles la VIA est suffisante, pour lesquelles la VAD est essentielle, et pour lesquelles l'ensemble des modes de variabilité est requis pour expliquer les variations du signal. Cette classification a été analysée au regard du coefficient de variation calculé sur les données annuelles d'éclairement.

Au final, au vu de la diversité des résultats entre les bases de données, tout comme entre les régions considérées, nous ne pouvons qu'une nouvelle fois recommander à ce stade de réaliser une étude approfondie des variations du signal pour le site d'intérêt, en analysant chaque série temporelle issue de la décomposition temps-fréquence.

Chapitre 8

Conclusions et perspectives

Ces travaux de thèse se sont insérés dans un contexte d'analyse de risques financiers réalisée en amont de l'installation de centrales solaires photovoltaïques. Il a été identifié, parmi toutes les sources d'incertitudes des analyses de risques financiers, un besoin de prise en compte des variations de long terme de l'éclairement. Pour contribuer à une meilleure prise en compte de ces variations, trois axes ont été dégagés, constituant les trois objectifs de cette thèse :

- Objectif 1** Faire le point sur les possibilités et les limites des méthodes de caractérisation du potentiel solaire disponibles dans la littérature.
- Objectif 2** Etudier de manière approfondie la stationnarité temporelle de l'éclairement.
- Objectif 3** Identifier et caractériser les variations de long terme de l'éclairement solaire.

L'ensemble des résultats dégagés par les études sont résumées par la suite.

La quantité d'information typiquement disponible pour caractériser la ressource d'un site n'est pas suffisante pour estimer le P90 de manière robuste.

Les analyses statistiques classiquement appliquées aux données historiques dans le but de construire des indicateurs statistiques ont été explorées. Leurs limitations ont été étudiées à partir de l'exemple de l'indicateur P90. La difficulté d'estimer de manière robuste le P90 à partir de séries temporelles constituées de 10 à 20 ans de données d'éclairement agrégées à l'échelle annuelle a été soulevée. La sensibilité des résultats au nombre de données disponibles a été mise en évidence. Autrement dit, avec un aussi faible échantillon de données historiques, le calcul du P90 est soumis à une grande incertitude. De plus, il a été mis en évidence que même pour un cas idéal, pour lequel le P90 est parfaitement connu, les incertitudes associées à cet indicateur restent très grandes en raison du nombre d'années limité de la durée de vie de la centrale. Ainsi, il a été conclu que les indicateurs statistiques tels que le P90 devraient être maniés avec précaution et, en particulier, que ses limitations devraient être bien comprises par les utilisateurs.

Ce résultat, issu d'analyses réalisées sur des données d'éclairements, reste entièrement valable dans le cas d'un P90 calculé sur le productible ou sur le revenu. Il paraît complexe d'envisager sur un court terme un changement d'indicateur en ce qui concerne le revenu. En effet, comme abordé précédemment, il s'agit de l'indicateur de référence utilisé par la communauté photovoltaïque comme par les banquiers. À l'opposé, il serait souhaitable d'éviter de caractériser la ressource solaire d'un site à l'aide d'un P90. Une alternative serait de proposer un ensemble de séries temporelles d'éclairement futur plausibles. À partir de cet ensemble de séries temporelles, un ensemble de séries temporelles de productible, puis de séries temporelles de revenus seraient générées. Le P90 serait ainsi uniquement calculé grâce à l'ensemble de séries temporelles de revenus générés, réduisant la cascade d'incertitudes associées à un calcul du P90 pour chaque quantité.

Cette pratique alternative est compatible avec de plus en plus de logiciels de support d'aide à la décision pour l'investissement photovoltaïque. Pour l'exemple du logiciel Solar Advisor Model (SAM) (Blair et al., 2014), couramment utilisé, les incertitudes associées à l'éclairement sont calculées à partir d'un ensemble de séries temporelles. Si l'utilisateur ne fournit qu'une série temporelle en entrée du logiciel, celui-ci va créer synthétiquement cinq séries temporelles à partir de la CDF de l'éclairement historique. Or, nous avons abordé la grande incertitude propre à la construction d'une CDF à partir de peu de données. Ainsi, en fournissant en entrée des séries temporelles plausibles et représentatives de la variabilité inter-annuelle du site, l'incertitude associée à l'éclairement sera plus précise et robuste qu'en ne fournissant qu'une série temporelle de données, dupliquée par la suite par le logiciel.

Cette constatation nous a mené à l'étude des méthodes de générations de séries temporelles plausibles d'éclairement disponibles dans la littérature.

Pour générer des données d'éclairement, la stationnarité temporelle est très fréquemment admise de manière implicite. Pourtant, cette hypothèse est largement discutable.

Une étude des hypothèses sous-jacentes aux méthodes de génération de données a été réalisée. Il a été mis en évidence que dans la grande majorité des cas, l'importance des variations inter-annuelles de l'éclairement est implicitement négligée. Cela revient à postuler l'hypothèse de stationnarité temporelle. Cette dernière étant implicite, les incertitudes associées à cette hypothèse sont rarement quantifiées.

À l'inverse, un ensemble d'études remettant en cause la stationnarité de l'éclairement ont été discutées. Ces études montrent qu'il n'est pas conseillé de postuler de manière systématique l'hypothèse de stationnarité temporelle de l'éclairement.

Cela nous a mené à une étude approfondie des travaux portant sur les variations de long terme de l'éclairement.

Les données disponibles dans les études sur les variations de long terme n'ont pas une résolution spatiale et temporelle suffisante pour répondre aux besoins industriels d'estimation de la ressource solaire.

La bibliographie disponible sur les variations de long terme de l'éclairement a été analysée. Deux cas ont été distingués : les régions pour lesquelles les résultats des études sont en désaccord, et celles pour lesquelles ils font consensus. Dans le premier cas, les divergences sont complexes à analyser pour un monteur de projet, de par la diversité des méthodes employées. Des études complémentaires s'imposent pour analyser ces divergences, ce qui est rarement compatible avec les contraintes temporelles d'une analyse industrielle. Dans le second cas, des différences peuvent être trouvées entre le consensus bibliographique et les mesures locales du réseau GEBA. La bibliographie ne semble donc pas assez précise pour prendre en compte les structures de variabilité fines, nécessaires pour l'étude d'un site particulier. Ainsi, l'information dégagée par la bibliographie n'est à nouveau pas adaptée pour une utilisation industrielle. L'analyse réalisée a souligné l'intérêt d'une méthode permettant de synthétiser l'information disponible dans la bibliographie.

Nous avons conclu quant à l'intérêt d'une meilleure caractérisation des cycles et des tendances de long terme influençant l'éclairement. Cette caractérisation permettrait de prendre en compte les variations de long terme dans les méthodes de génération de séries temporelles d'éclairement.

La caractérisation des variations de long terme par des tendances linéaires ne permet pas de représenter l'ensemble des variations inter-annuelles.

La première approche envisagée pour caractériser les variations de long terme fut une approche statistique. Ainsi, l'étude d'un ensemble de tests statistiques de stationnarité temporelle a été menée. Trois tests ont été sélectionnés, puis appliqués aux stations de mesure du réseau GEBA. Pour une majorité de stations, les résultats des tests ne sont pas en accord. Les causes possibles de ces divergences étant variées, cette approche a été mise de côté. Il a été souligné qu'en caractérisant les variations de long terme uniquement à l'aide de tendances linéaires sur une période temporelle restreinte, les variations inter-annuelles comprises dans des échelles inférieures à la période étudiée ne sont pas détectées.

Une seconde approche statistique, plus robuste, pourrait être appliquée dans le futur. Celle-ci, dont la méthodologie a été construite et testée au cours de la thèse, est inspirée des travaux portant sur la stationnarité temporelle de la température réalisés par *Huong et al.* (2009) et *Parey et al.* (2018). Elle repose sur des techniques de ré-échantillonnage de la série de données originale. Si la série de données originale et la série de données ré-échantillonnée montrent des résultats similaires, il est possible d'avancer que les données sont indépendantes et stationnaires. En effet, lors du processus de ré-échantillonnage, il est implicitement supposé l'indépendance des données, ainsi que l'absence de

mode de variabilité de long terme. Pour un lecteur intéressé, la méthode développée est succinctement résumée dans le paragraphe suivant.

La méthode proposée se compose de trois étapes distinctes : la mise en évidence d'éventuelles chaînes de corrélation au sein de la série temporelle mensuelle, le ré-échantillonnage de la série temporelle agrégée selon ces chaînes de corrélation, et la comparaison entre le signal ré-échantillonné et le signal original. L'identification rigoureuse et la conservation des chaînes de corrélation intra-annuelles est nécessaire pour une bonne interprétation des résultats. Sans cela, il serait impossible de déterminer si les éventuels écarts entre la série de données originale et la série ré-échantillonnée sont dus à des variations intra-annuelles ou inter-annuelles. Si des chaînes de corrélation sont identifiées, il suffit alors d'agréger la série temporelle initiale à l'échelle déterminée, avant de procéder au ré-échantillonnage et à la comparaison des deux séries temporelles. Si aucune chaîne de corrélation n'est identifiée, alors l'échelle annuelle pourra être choisie. Pour estimer les chaînes de corrélation, nous conseillons d'avoir recours à l'étude de la fonction de structure – ou variogrammes – des données considérées. En effet, celle-ci suppose uniquement la stationnarité par incréments de la série temporelle. Cette hypothèse est relativement faible face à celles postulées par d'autres méthodes. Par exemple, l'étude de l'autocorrélation des données suppose la stationnarité au second ordre de la série temporelle, c'est-à-dire que le signal est supposé stationnaire, tout comme son espérance. L'utilisation du modèle ARIMA suppose quant à elle la stationnarité des incréments.

Dans le cadre de cette thèse, il a été choisi d'utiliser des méthodes permettant d'identifier et de caractériser les variations inter-annuelles de l'éclairement quel que soit leurs échelles de variation. Ainsi, plutôt que d'utiliser des approches statistiques, nous nous sommes orientés vers une analyse temps-fréquence des séries temporelles.

Une décomposition temps-fréquence permet de synthétiser l'information fournie dans la bibliographie, tout en caractérisant les modes de variabilité inter-annuels de l'éclairement

Une méthode de décomposition temps-fréquence a été développée. Celle-ci permet de synthétiser l'information disponible dans la bibliographie tout en détectant la présence d'anomalies ponctuelles comme de changements subvenant uniquement sur une partie de la période temporelle analysée. Notre choix a également été guidé par la recherche de transparence et de simplicité de la méthode considérée. Il s'est porté sur un banc de filtres construit à partir de sinus cardinaux apodisés. Cette méthode a été soigneusement caractérisée et validée. Elle permet d'étudier simultanément l'ensemble des modes de variabilité de l'éclairement. Ainsi, une compréhension fine des variations de l'ordre de quelques années, de quelques décennies, ou plus, est possible.

Une nouvelle fois, le faible nombre de données historiques disponibles a conditionné les analyses possibles à l'aide de cet outil. Ainsi, trois classes de variabilité ont été distinguées : la variabilité intra-annuelle, la variabilité annuelle à décennale, et la variabilité multi-décennale. Ces trois classes de variabilité ont été étudiées de manière distincte pour différentes bases de données.

Des structures de variabilité communes à l'ensemble des bases de données ont été identifiées.

Les séries temporelles d'éclairement issues de quatre bases de données ont été décomposées. Différents modes de variabilité ont pu être identifiés pour la sélection de stations GEBA, les données satellitaires CLARA-A2, les données de ré-analyse MERRA-2 et les données du modèle climatique IPSL-CM6A-LR. Pour l'ensemble des bases de données, une répartition latitudinale de la variabilité a pu être constatée, avec une forte influence de la variabilité inter-annuelle proche de l'équateur, et faible dans la zone extratropicale. Ainsi, pour la majorité des zones extratropicales, l'essentiel des variations sont dues à la variabilité intra-annuelle. Pour ces zones, l'utilisation de quatre ans de données historiques sera suffisante pour prendre en compte de manière correcte l'ensemble des variations du signal. À l'inverse, pour des zones telles que le Pacifique équatorial, l'océan Indien, l'Indonésie, ou encore le Nord de l'Amérique du Sud, une prise en compte des variations inter-annuelles est nécessaire, d'après l'ensemble des bases de données.

Ces similitudes cachent une grande diversité des structures en fonction de la base de données considérée : diversité en termes d'extension spatiale, et diversité en termes d'influence relative des variations annuelles à décennales et des variations multi-décennales. Soulignons également que nous avons détecté une plus forte influence de la variabilité inter-annuelle avec la base de données GEBA qu'avec les autres bases de données. Cela questionne sur la représentation des variations inter-annuelle dans les bases de données grillées. Des analyses approfondies des modèles sous-jacents à chaque base de données étudiée seront intéressantes pour mieux comprendre cette représentation. Des études complémentaires à partir d'autres bases de données restent également à réaliser pour valider ce résultat.

Les variations inter-annuelles pour une même échelle diffèrent en fonction de la base de données utilisée.

Une classification spatiale de l'indice de clarté, puis de l'éclairement, a été réalisée. Pour certaines zones, il a été montré que la variabilité intra-annuelle est suffisante pour expliquer plus de 95 % des variations du signal. À l'inverse, pour d'autres zones, la prise en compte de la variabilité annuelle à décennale est nécessaire. Enfin, la troisième classe de variabilité a rassemblé les zones pour lesquels l'ensemble des modes de variabilité sont nécessaires pour une bonne représentation des variations du signal. Pour les quatre bases de données, nous avons montré que la variabilité annuelle à décennale est essentielle pour prendre en compte plus de 95% des variations du signal sur plusieurs régions, telle que l'Indonésie ou l'océan Indien. Pour prendre en compte correctement cette classe de variabilité dans un contexte prospectif, l'utilisation de 30 années de données historiques est recommandée. Cependant, une diversité des structures influencées par cette classe de variabilité a été observée entre les quatre bases de données. En particulier, la variabilité annuelle à décennale détectée par les trois bases de données grillées coïncide rarement avec celle détectée par les données des stations GEBA.

Pour certaines régions du globe, la prise en compte des variations multi-décennales est nécessaire pour une estimation complète des variations du signal. Pour ces régions, il est recommandé d'utiliser plus de 30 années de données historiques pour estimer le potentiel solaire. Une nouvelle fois, l'étendue des régions varie d'une base de données à l'autre. Pour la base de données GEBA, une forte proportion des stations montrent une forte variabilité multi-décennale. À l'inverse, pour la base de données IPSL-CM6A-LR, l'étendue de ces régions est très réduite, voir quasiment inexistante. En ce qui concerne les deux autres bases de données, des structures ont pu être identifiées dans le Pacifique équatorial, en Afrique centrale, et sur l'Est de l'Atlantique équatorial.

Étant donnée l'absence de consensus entre les quatre bases de données étudiées, il est difficile de conclure quant à l'importance relative de telle ou telle classe de variabilité pour une région donnée. À ce stade, il est recommandé de réaliser l'étude de l'ensemble des éléments issus de la décomposition temps-fréquence pour le site choisi. Pour aller plus loin, il serait utile de réaliser la même étude pour d'autres bases de données. En particulier, l'utilisation de bases de données satellitales possédant un historique complet plus long que CLARA-A2 serait utile. Par exemple, les bases de données ISCCP, GEWEX SRB ou Cloud_cci AVHRR-PM, citées et référencées dans le chapitre 2, pourraient être envisagées. La base de données de ré-analyse ERA-5 pourrait également être explorée. L'étude de différents modèles climatiques pourrait permettre d'identifier si l'absence de variabilité multi-décennale se retrouve dans l'ensemble des modèles climatiques, ou est propre aux modèles IPSL-CM6A-LR.

*L'étude des oscillations climatiques pourrait permettre
de modéliser la variabilité annuelle à décennale.*

Pour prendre en compte de manière correcte la variabilité annuelle à décennale, nous avons recommandé l'utilisation de 30 années de données. Une alternative à l'utilisation de ce large échantillon de données est la prise en compte de la variabilité annuelle à décennale par la modélisation des oscillations caractéristiques de ces échelles. Dans cette optique, un rapprochement visuel entre les régions influencées par cette classe de variabilité et les régions influencées par les oscillations climatiques connues a été réalisée. En particulier, les oscillations ENSO et IOD ont été discutées.

Pour aller plus loin dans cette approche, un travail d'analyse statistique pourrait être réalisé. Des indicateurs pourraient quantifier les similitudes entre les séries temporelles des indices climatiques et les éléments issus de la décomposition temps-fréquence, tels que par exemple le coefficient de corrélation ou le coefficient de détermination. Les indices climatiques quantifiant les oscillations étant majoritairement issus de mesures de la température de l'océan, l'intérêt d'un décalage temporel entre les deux séries pourra être exploré. Ainsi, l'inertie du système climatique, conduisant à un décalage entre les modifications de température de l'océan et la quantité d'éclairement d'un site parfois éloigné, pourra être pris en compte. À terme, si des similarités fortes sont observées, les indices climatiques pourraient être utilisés comme des indicateurs indirects de l'éclairement pour une plage de période données, et une région donnée.

La prise en compte de la variabilité multi-décennale permettrait d'améliorer la performance des méthodes d'estimation du potentiel solaire.

La difficulté de caractériser la variabilité multi-décennale est principalement due au nombre de données nécessaire important. Il a été question au cours de la thèse d'avoir recours aux données issues de modèles climatiques pour accéder à un historique de plus de cent ans d'éclairement. Il aurait même été possible, en combinant les différentes réalisations d'un modèle climatique, d'avoir accès à plus d'un millier d'années de données historiques. À première vue, l'utilisation de modèles climatiques nous a semblé d'autant plus pertinente qu'ils proposent également des données d'éclairement prospectives. Basées sur différents scénarios socio-économiques, les scénarios climatiques auraient ainsi pu permettre de prendre en compte les variations de long terme de l'éclairement dues à des évolutions sociétales, telles que par exemple les variations d'émissions d'aérosols anthropique dans les décennies à venir. Cependant, le fait que le modèle choisi ne montre pas de variabilité multi-décennale a écarté cette possibilité. Ce modèle ne semble donc pas permettre de contribuer en l'état à la caractérisation des variations multi-décennales de l'éclairement.

À ce stade, l'accent est donc mis sur l'intérêt d'approfondir les connaissances des variations supérieures à la décennie. Pour les régions soumises à cette classe de variabilité, l'intérêt de leur prise en compte est certain. Prenons l'exemple de la station de Nice. Cette station, appartenant au réseau GEBA, a été identifiée précédemment comme une station soumise à une forte variabilité multi-décennale. La série temporelle GEBA, disponible en 1982 et 2013, a été rallongée jusqu'en 2019 grâce aux données satellitales HelioClim-3 v5 (LEFEVRE et al., 2007). Cette opération peut se justifier par le fait qu'entre 2006 et 2013, période de recoupement des séries temporelles, le coefficient de corrélation entre les deux séries temporelles est supérieur à 0.999. La figure 54 présente la série temporelle mensuelle de l'indice de clarté. En rose sont également représentées les variations multi-décennales de cette série temporelle (éléments y_8 , y_9 et $y_{résidu}$).

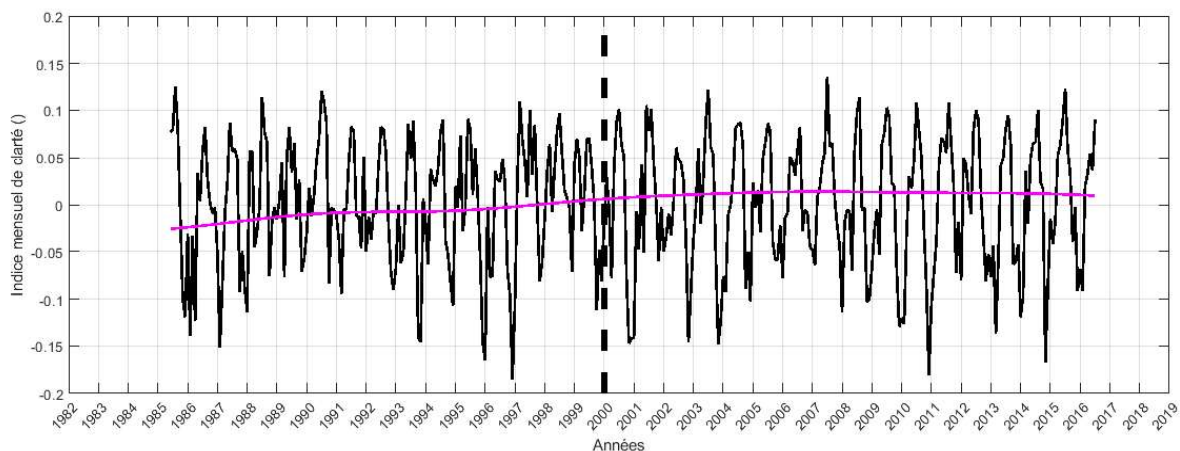


Figure 54 : Série temporelle d'éclairement mensuel pour la station de Nice. En rose sont représentées les variations multi-décennales de la série temporelle. Ces variations sont calculées à partir des trois plus longues plages de périodes issues de la décomposition temps-fréquence introduite chapitre 5.

Pour estimer le potentiel solaire de cette station, appliquons la méthode de Larrañeta et al. (2018) décrite au cours du chapitre 2. La période comprise entre 1982 et 1999 sera considérée comme l'échantillon historique de données. À partir de cet échantillon de près de 20 ans, le potentiel solaire du site sera calculé, représentant pour rappel les 20 années d'éclairement futur du site. Pour assurer une robustesse statistique à la méthode, 200 séries temporelles d'éclairement futur ont été calculées. Les données de la série temporelle originale entre 2000 et 2019 seront utilisées pour valider les résultats issus de la méthode de Larrañeta et al. (2018).

La figure 55.a présente la série temporelle mesurée, en rouge, et l'enveloppe des 200 réalisations futures, en bleu. Dans un cas idéal, la courbe rouge devrait se trouver entre les extremums des points bleus. La figure 55.b présente les mêmes quantités, avec toutefois en bleu les données générées, auxquelles les variations multi-décennales de la période considérée ont été ajoutées. Autrement dit, dans le cas hypothétique où les variations multi-décennales, telles que représentée en rose sur la figure 54, seraient parfaitement connus, il serait possible de les ajouter aux données générées par la méthode de Larrañeta et al. (2018). Dans ce cas, les prévisions seraient similaires à celles représentées sur la figure 55.b.

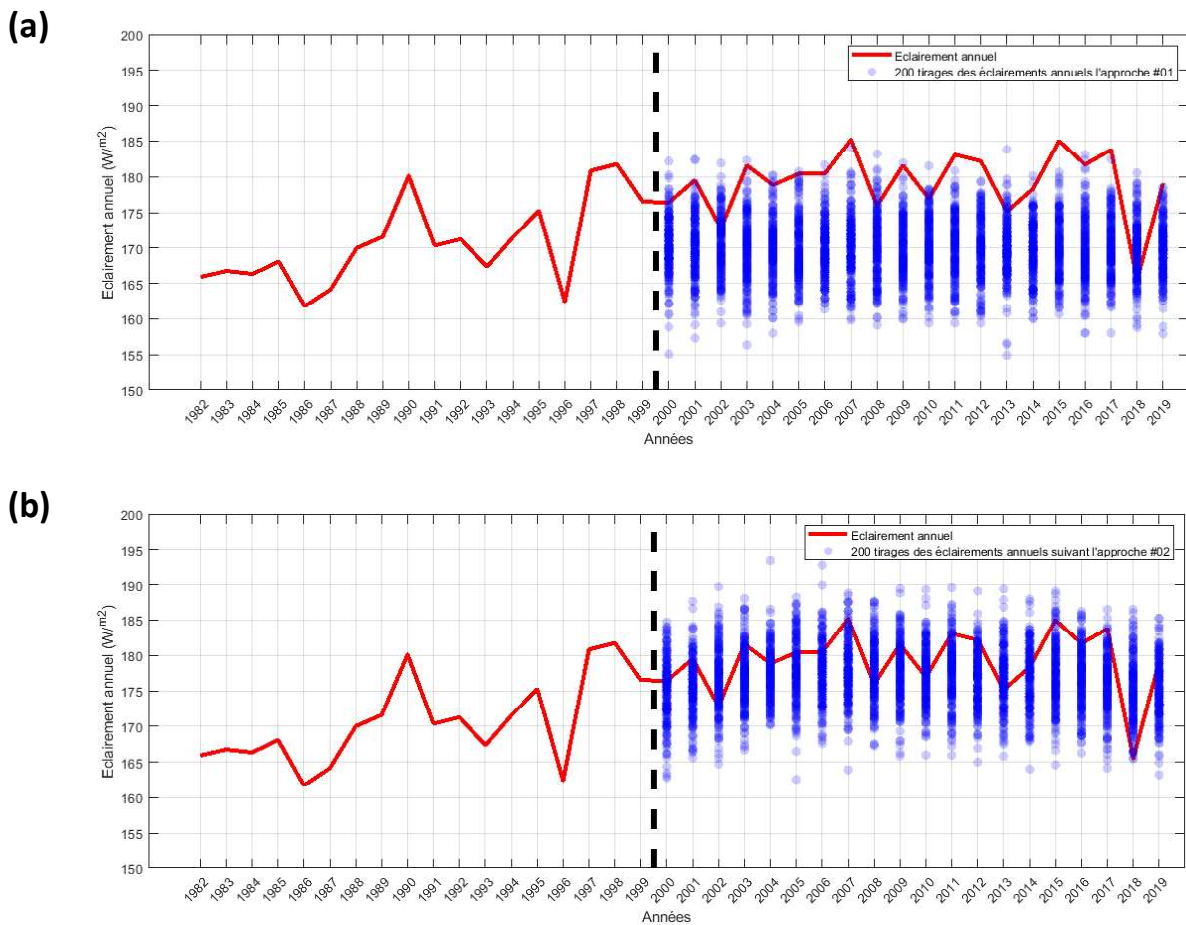


Figure 55 : Séries temporelle de l'éclairement annuel pour la station Nice (rouge). En bleu sont représentée les données générées sur la période 2000-2019. (a) Données générée par la méthode de Larrañeta et al. (2018). (b) Données générées par la même méthode, auxquelles on ajoute les variations multi-décennales.

L'intérêt de prendre en compte la variabilité multi-décennale est évidente au regard de cette figure. En effet, on voit bien sur la figure 55.a que l'éclairement moyen est systématiquement sous-estimé. À l'inverse, celui-ci est très bien modélisée si on ajoute aux données générées les variations supérieures à la décennie : la courbe rouge est comprise entre les points bleu pour la figure 55.b. Soulignons également l'intérêt d'avoir une génération de séries temporelles multiples : on ne compare pas une simple série temporelle générée, mais une gamme de possibles.

Grâce au logiciel Solar Advisor Model (SAM) cité précédemment, il est possible de calculer le productible associée à chacune des réalisations proposées. Ainsi, nous pouvons estimer l'influence de la prise en compte ou non de la variabilité multi-décennale sur l'estimation du productible futur, pour une centrale photovoltaïque installée à Nice. Le productible a été calculé à partir de la version 2020.2.29 du logiciel, pour une centrale de 1MW_c. L'inclinaison des modules est de 35° dans la direction Sud. Le facteur de charge moyen est estimé à 18.5 %, avec une dégradation des performances des modules de 0.6 % par an. La figure 56 présente les séries temporelles de productible, pour avec (b) et sans (a) prise en compte de la variabilité multi-décennale. En pointillé se trouve les valeurs du P90 calculée pour chaque année.

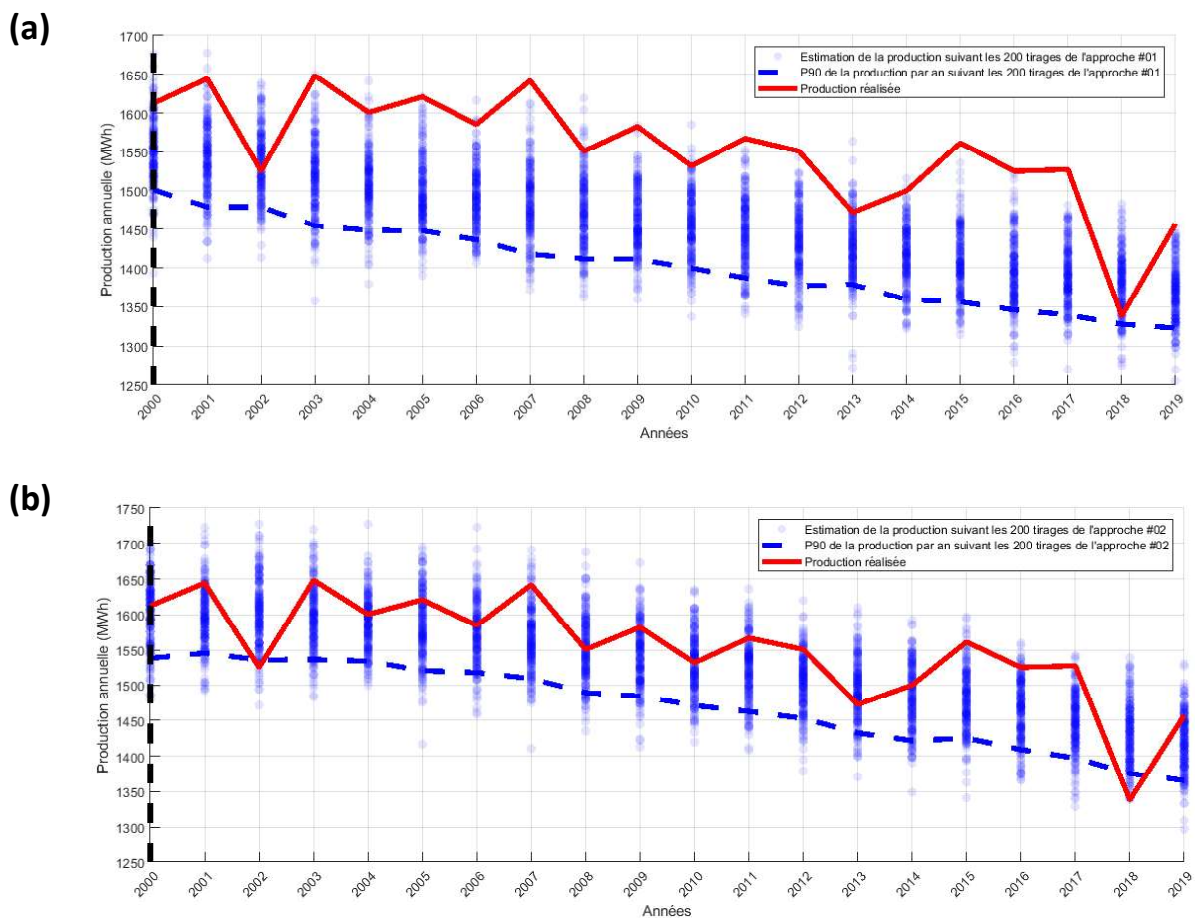


Figure 56 : Séries temporelle du productible annuel pour la station Nice (rouge). En bleu sont représentée les données générées sur la période 2000-2019. (a) Données générée par la méthode de Larrañeta et al. (2018). (b) Données générées par la même méthode, auxquelles on ajoute les variations multi-décennales.

Comme pour l'éclairément, il est clair que les performances de la deuxième méthode sont supérieures à la première méthode. Une décroissance du productible est visible pour les deux séries temporelles. Cela s'explique en grande partie par le vieillissement des panneaux fixé à 0.6 % par an. Cette tendance souligne l'intérêt d'avoir une représentation temporelle du productible.

Pour finir sur cet exemple illustratif, la figure 57 présente les fonctions de distributions cumulatives (CDF) du productible estimé par les deux méthodes. En bleu se trouve la première méthode, en rouge la seconde.

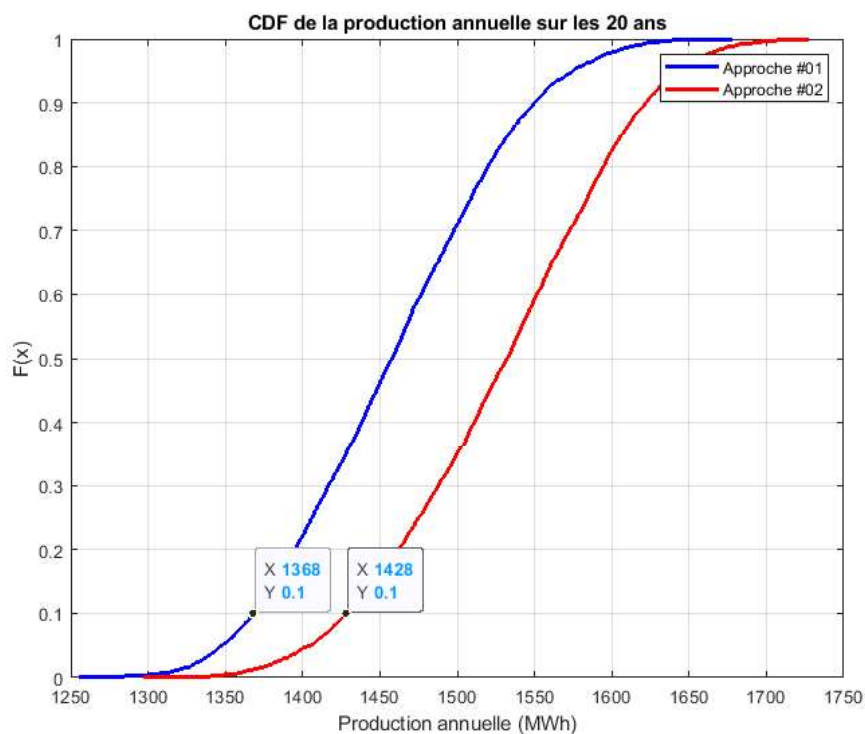


Figure 57 : Série temporelle d'éclairément mensuel pour la station de Nice. En rose sont représentées les variations multi-décennales de la série temporelle. Ces variations sont calculées à partir des trois plus longues plages de périodes issues de la décomposition temps-fréquence introduite chapitre 5.

L'écart entre les CDF est frappant. Pour le P90, un écart de 4.4 % est observée. Cet exemple illustratif montre l'intérêt de la prise en compte des modes de variabilité long terme pour une estimer de façon robuste le potentiel solaire.

Le champ d'application de la méthode développée est vaste : un fort potentiel d'analyses complémentaires a été identifié.

La méthode de décomposition temps-fréquence développée a été majoritairement utilisée dans le cadre d'analyse de variance de l'éclairement. En étudiant la variance de chaque élément, l'aspect temporel des résultats n'est pas considéré. Pourtant, nous avons souligné l'importance d'étudier pour le point d'intérêt la décomposition autant dans le domaine fréquentiel que temporel. Ainsi, une étude de la variance pourrait par exemple être d'intérêt.

Comme annoncé précédemment, cet outil pourrait également permettre d'identifier des coïncidences entre un élément particulier et un indice climatique. Ainsi, il pourrait être possible de recomposer un signal d'éclairement à partir d'un composite d'éléments. Par exemple, certains éléments pourraient être générés de façon aléatoire, quand d'autres seraient modélisés à partir des indices climatiques, ou encore à partir de la décomposition de bases de données de plus long terme. Cette idée se retrouve dans la littérature, avec notamment les études de Bengulescu *et al.* (2016, 2018), qui démontrent que certaines échelles caractéristiques issus d'une décomposition temps-fréquence (décomposition en modes empirique) sont de nature déterministe, quand d'autres sont stochastiques. Dans la même idée, la méthode d'analyse SSA (*Singular spectrum analysis*) (HASSANI, 2007), classe les éléments issus de la décomposition d'un signal (décomposition orthogonale aux valeurs propres) en trois catégories : tendance, modes oscillatoires et bruit.

Après avoir passé en revue les méthodes de l'état de l'art, cette thèse s'est donc attachée à l'étude des variations inter-annuelles de l'éclairement. En conclusion, l'intérêt de la prise en compte de la variabilité annuelle à décennale, ou multi-annuelle dépend de l'endroit considéré. Ainsi, l'identification de la classe de variabilité la plus influente, suivi d'une prise en compte propre à l'échelle identifiée, semble extrêmement pertinente. À terme, une classification des zones influencées par tel ou tel mode de variabilité pourrait permettre de faciliter les démarches. Un travail de compréhension des moteurs des modes de variabilité annuels à décennaux reste à réaliser, pour guider l'utilisateur dans la prise en compte de ces modes de variabilité. Enfin, un travail de caractérisation des modes de variabilité multi-décennaux sera également de grand intérêt.

Références

- ABREU, E. F., CANHOTO, P., PRIOR, V., & MELICIO, R. (2018). Solar resource assessment through long-term statistical analysis and typical data generation with different time resolutions using GHI measurements. *Renewable Energy*, *127*, 398–411.
- AHN, S., & FESSLER, J. (2003). *Standard errors of mean, variance, and standard deviation estimators*. The University of Michigan - EECS Department.
<https://web.eecs.umich.edu/~fessler/papers/files/tr/stderr.pdf>
- ALEXANDER, M. A., KILBOURNE, K. H., & NYE, J. A. (2014). Climate variability during warm and cold phases of the Atlantic Multidecadal Oscillation (AMO) 1871–2008. *Journal of Marine Systems*, *133*, 14–26. <https://doi.org/10.1016/j.jmarsys.2013.07.017>
- ALEXANDERSSON, H. (1986). A homogeneity test applied to precipitation data. *Journal of Climatology*, *6*(6), 661–675. <https://doi.org/10.1002/joc.3370060607>
- ALEXANDERSSON, H., & MOBERG, A. (1997). Homogenization of Swedish temperature data. Part I: Homogeneity test for linear trends. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *17*(1), 25–34. [https://doi.org/10.1002/\(SICI\)1097-0088\(199701\)17:1<25::AID-JOC103>3.0.CO;2-J](https://doi.org/10.1002/(SICI)1097-0088(199701)17:1<25::AID-JOC103>3.0.CO;2-J)
- ALIPPI, C., & ROVERI, M. (2006). An adaptive CUSUM-based test for signal change detection. *An Adaptive Cusum-Based Test for Signal Change Detection*, *4*.
<https://doi.org/10.1109/ISCAS.2006.1693942>
- ANDĚL, J., RUBIO, A., & INSUA, A. (1985). On periodic autoregression with unknown mean. *Aplikace Matematiky*, *30*(2), 126–139.
- BABAR, B., GRAVERSEN, R., & BOSTRÖM, T. (2018). Evaluating CM-SAF solar radiation CLARA-A1 and CLARA-A2 datasets in Scandinavia. *Solar Energy*, *170*, 76–85.
<https://doi.org/10.1016/j.solener.2018.05.009>
- BADESCU, V. (Ed.). (2014). *Modeling Solar Radiation at the Earth's Surface: Recent Advances ; [with Cd-Rom]*. Springer.
- BARTÓK, B., WILD, M., FOLINI, D., LÜTHI, D., KOTLARSKI, S., SCHÄR, C., VAUTARD, R., JEREZ, S., & IMECS, Z. (2017). Projected changes in surface solar radiation in CMIP5 global climate models and in EURO-CORDEX regional climate models for Europe. *Climate Dynamics*, *49*(7–8), 2665–2683. <https://doi.org/10.1007/s00382-016-3471-2>
- BENGULESCU, M., BLANC, P., & WALD, L. (2016). On the temporal variability of the surface solar radiation by means of spectral representations. *Advances in Science and Research*, *13*, 121–127. <https://doi.org/10.5194/asr-13-121-2016>
- BENGULESCU, M., BLANC, P., & WALD, L. (2018). *On the intrinsic timescales of temporal variability in measurements of the surface solar radiation*. /paper/On-the-intrinsic-timescales-of-temporal-variability-Bengulescu-Blanc/b6149f2b3bc91f23dc0e90f31bb711c77fcb4839

REFERENCES :

- BERRY, H. L., BOWEN, K., & KJELLSTROM, T. (2010). Climate change and mental health: A causal pathways framework. *International Journal of Public Health*, 55(2), 123–132.
- BEYER, H. G., FAUTER, M., SCHUMANN, K., SCHENK, H., & MEYER, R. (2010). Synthesis of DNI time series with sub-hourly time resolution. *16th SolarPACES Conference. Perpignan (France)*, 9.
- BLACK, J. N., BONYTHON, C. W., & PRESCOTT, J. A. (1954). Solar radiation and the duration of sunshine. *Quarterly Journal of the Royal Meteorological Society*, 80(344), 231–235. <https://doi.org/10.1002/qj.49708034411>
- BLAIR, N., DOBOS, A. P., FREEMAN, J., NEISES, T., WAGNER, M., FERGUSON, T., GILMAN, P., & JANZOU, S. (2014). *System Advisor Model, SAM 2014.1.14: General Description* (NREL/TP--6A20-61019, 1126294; p. NREL/TP--6A20-61019, 1126294). <https://doi.org/10.2172/1126294>
- 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. *Solar Energy*, 86(10), 3072–3083. <https://doi.org/10.1016/j.solener.2012.07.018>
- BLANCO-MURIEL, M., ALARCÓN-PADILLA, D. C., LÓPEZ-MORATALLA, T., & LARA-COIRA, M. (2001). Computing the solar vector. *Solar Energy*, 70(5), 431–441. [https://doi.org/10.1016/S0038-092X\(00\)00156-0](https://doi.org/10.1016/S0038-092X(00)00156-0)
- BONJEAN STANTON, M. C., DESSAI, S., & PAAVOLA, J. (2016). A systematic review of the impacts of climate variability and change on electricity systems in Europe. *Energy*, 109, 1148–1159. <https://doi.org/10.1016/j.energy.2016.05.015>
- BOSILOVICH, M. G., AKELLA, S., & COY, L. (2015). *MERRA-2: Initial Evaluation of the Climate* (No. 43; Technical Report Series on Global Modeling and Data Assimilation, p. 145). NASA. <https://gmao.gsfc.nasa.gov/pubs/docs/Bosilovich803.pdf>
- BP. (2019). *BP Statistical Review of World Energy*. <https://www.bp.com/content/dam/bp/business-sites/en/global/corporate/pdfs/energy-economics/statistical-review/bp-stats-review-2019-full-report.pdf>
- BRETAGNON, P., & FRANCOU, G. (1988). Planetary theories in rectangular and spherical variables—VSOP 87 solutions. *Astronomy and Astrophysics*, 202, 309–315.
- BREHERTON, C. S. (2014). *Lecture 3: Statistical sampling uncertainty*. <https://atmos.washington.edu/~breth/classes/AM582/lect/lect3-notes.pdf>
- BUIHAND, T. A. (1982). Some methods for testing the homogeneity of rainfall records. *Journal of Hydrology*, 58(1–2), 11–27. [https://doi.org/10.1016/0022-1694\(82\)90066-X](https://doi.org/10.1016/0022-1694(82)90066-X)
- CATTIAUX, J., VAUTARD, R., & YIOU, P. (2011). North-Atlantic SST amplified recent wintertime European land temperature extremes and trends. *Climate Dynamics*, 36(11–12), 2113–2128. <https://doi.org/10.1007/s00382-010-0869-0>
- CEBECAUER, T., & SURI, M. (2015). Typical Meteorological Year Data: SolarGIS Approach. *Energy Procedia*, 69, 1958–1969. <https://doi.org/10.1016/j.egypro.2015.03.195>
- CHENG, C., & PARZEN, E. (1997). Unified estimators of smooth quantile and quantile density functions. *Journal of Statistical Planning and Inference*, 59(2), 291–307. [https://doi.org/10.1016/S0378-3758\(96\)00110-3](https://doi.org/10.1016/S0378-3758(96)00110-3)

- CHIACCHIO, M., & WILD, M. (2010). Influence of NAO and clouds on long-term seasonal variations of surface solar radiation in Europe. *Journal of Geophysical Research: Atmospheres*, 115(D10). <https://doi.org/10.1029/2009JD012182>
- CONOVER, W. J. (1971). *Practical Nonparametric Statistics*. Wiley.
- COOK, K. H. (2000). The South Indian Convergence Zone and Interannual Rainfall Variability over Southern Africa. *Journal of Climate*, 13(21), 3789–3804. [https://doi.org/10.1175/1520-0442\(2000\)013<3789:TSICZA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2000)013<3789:TSICZA>2.0.CO;2)
- COOPER, J. K., & BRUCE, A. (2017). (13) *Interannual Variability of the Solar Resource across Australia*. https://www.researchgate.net/publication/322628186_Interannual_Variability_of_the_Solar_Resource_across_Australia
- CURTIS, L., REA, W., SMITH-WILLIS, P., FENYVES, E., & PAN, Y. (2006). Adverse health effects of outdoor air pollutants. *Environment International*, 32(6), 815–830.
- DAMBREVILLE, R. (2014). *Prévision du rayonnement solaire global par télédétection pour la gestion de la production d'énergie photovoltaïque* [Phdthesis, Université de Grenoble]. <https://tel.archives-ouvertes.fr/tel-01130251/document>
- DAMBRINE, F. (2019). Analyse micro-économique de l'intégration des EnR électriques intermittentes dans un système de production électrique. *Annales Des Mines, Responsabilités et Environnement*(93), 7–14.
- DAVY, R. J., & TROCCOLI, A. (2012). Interannual variability of solar energy generation in Australia. *Solar Energy*, 86(12), 3554–3560. <https://doi.org/10.1016/j.solener.2011.12.004>
- Décret n° 2016-1442 du 27 octobre 2016 relatif à la programmation pluriannuelle de l'énergie, 2016-1442 (2016).
- DICKEY, D. A., & FULLER, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366a), 427–431.
- DIELMAN, T., LOWRY, C., & PFAFFENBERGER, R. (1994). A comparison of quantile estimators. *Communications in Statistics - Simulation and Computation*, 23(2), 355–371. <https://doi.org/10.1080/03610919408813175>
- DUTTON, E. G., NELSON, D. W., STONE, R. S., LONGENECKER, D., CARBAUGH, G., HARRIS, J. M., & WENDELL, J. (2006). Decadal variations in surface solar irradiance as observed in a globally remote network. *Journal of Geophysical Research: Atmospheres*, 111(D19). <https://doi.org/10.1029/2005JD006901>
- ENFIELD, D. B., MESTAS-NUÑEZ, A. M., & TRIMBLE, P. J. (2001). The Atlantic Multidecadal Oscillation and its relation to rainfall and river flows in the continental U.S. *Geophysical Research Letters*, 28(10), 2077–2080. <https://doi.org/10.1029/2000GL012745>
- EurObserv'ER. (2019). *Photovoltaic barometer 2019 | EurObserv'ER*. <https://www.eurobserv-er.org/photovoltaic-barometer-2019/>
- EYRING, V., BONY, S., MEEHL, G. A., SENIOR, C. A., STEVENS, B., STOUFFER, R. J., & TAYLOR, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6)

REFERENCES :

- experimental design and organization. *Geoscientific Model Development*, 9(5), 1937–1958. <https://doi.org/10.5194/gmd-9-1937-2016>
- FERNÁNDEZ-PERUCHENA, C. M., BLANCO, M., GASTÓN, M., & BERNARDOS, A. (2015). Increasing the temporal resolution of direct normal solar irradiance series in different climatic zones. *Solar Energy*, 115, 255–263. <https://doi.org/10.1016/j.solener.2015.02.017>
- FERNÁNDEZ-PERUCHENA, C. M., GASTÓN, M., SÁNCHEZ, M., GARCIA-BARBERENA, J., BLANCO, M., & BERNARDOS, A. (2015). MUS: A multiscale stochastic model for generating plausible meteorological years designed for multiyear solar energy yield simulations. *Solar Energy*, 120, 244–256. <https://doi.org/10.1016/j.solener.2015.07.037>
- FERNÁNDEZ-PERUCHENA, C. M., RAMÍREZ, L., SILVA-PÉREZ, M. A., LARA, V., BERMEJO, D., GASTÓN, M., MORENO-TEJERA, S., PULGAR, J., LIRIA, J., MACÍAS, S., GONZALES, R., BERNARDOS, A., CASTILLO, N., BOLINAGA, B., VALENZUELA, R. X., & ZARZALEJO, L. F. (2016). A statistical characterization of the long-term solar resource: Towards risk assessment for solar power projects. *Solar Energy*, 123, 29–39. <https://doi.org/10.1016/j.solener.2015.10.051>
- FitchRatings. (2018). *Renewable Energy Project Rating Criteria, Global Infrastructure & Project Finance, Sector-Specific Criteria*. <https://www.fitchratings.com/site/re/10019874>
- FREEMAN, J. M., DIORIO, N. A., BLAIR, N. J., NEISES, T. W., WAGNER, M. J., GILMAN, P., & JANZOU, S. (2018). *System Advisor Model (SAM) General Description (Version 2017.9. 5)*. National Renewable Energy Lab.(NREL), Golden, CO (United States).
- FRÖHLICH, C., & LEAN, J. (2004). Solar radiative output and its variability: Evidence and mechanisms. *The Astronomy and Astrophysics Review*, 12(4), 273–320. <https://doi.org/10.1007/s00159-004-0024-1>
- GELARO, R., McCARTY, W., SUÁREZ, M. J., TODLING, R., MOLOD, A., TAKACS, L., TANGLES, C. A., DARMENOV, A., BOSILOVICH, M. G., REICHLE, R., WARGAN, K., COY, L., CULLATHER, R., DRAPER, C., AKELLA, S., BUCHARD, V., CONATY, A., da SILVA, A. M., GU, W., ... ZHAO, B. (2017). The modern-era retrospective analysis for research and applications, version 2 (MERRA-2). *Journal of Climate*, 30(14), 5419–5454. <https://doi.org/10.1175/JCLI-D-16-0758.1>
- GILBERT, R. O. (1987). *Statistical methods for environmental pollution monitoring*. John Wiley & Sons.
- GILGEN, H., & OHMURA, A. (1999). The Global Energy Balance Archive. *Bulletin of the American Meteorological Society*, 80(5), 831–850. [https://doi.org/10.1175/1520-0477\(1999\)080<0831:TGEBA>2.0.CO;2](https://doi.org/10.1175/1520-0477(1999)080<0831:TGEBA>2.0.CO;2)
- GILGEN, H., WILD, M., & OHMURA, A. (1998). Means and trends of shortwave irradiance at the surface estimated from global energy balance archive data. *Journal of Climate*, 11(8), 2042–2061.
- GOVILLOT, N., LAVERGNE, R., & VALÉRIAN, F. (2019). Évaluation macroéconomique de la transition électrique en France. *Annales Des Mines, Responsabilité et environnement*(93), 16–24.
- GRENA, R. (2008). An algorithm for the computation of the solar position. *Solar Energy*, 82(5), 462–470. <https://doi.org/10.1016/j.solener.2007.10.001>
- GRENA, R. (2012). Five new algorithms for the computation of sun position from 2010 to 2110. *Solar Energy*, 86(5), 1323–1337. <https://doi.org/10.1016/j.solener.2012.01.024>

- GUEYMARD, C. (2008). REST2: High-performance solar radiation model for cloudless-sky irradiance, illuminance, and photosynthetically active radiation – Validation with a benchmark dataset. *Solar Energy*, 82(3), 272–285. <https://doi.org/10.1016/j.solener.2007.04.008>
- 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. *Solar Energy*, 86(8), 2145–2169. <https://doi.org/10.1016/j.solener.2011.11.011>
- HABTE, A., SENGUPTA, M., ANDREAS, A., WILCOX, S., & STOFFEL, T. (2014). *Evaluating Solar Resource Data Obtained from Multiple Radiometers Deployed at the National Renewable Energy Laboratory: Preprint* (NREL/CP-5D00-62775). National Renewable Energy Lab. (NREL), Golden, CO (United States). <https://www.osti.gov/biblio/1158437>
- HANLEY, D. E., BOURASSA, M. A., O'BRIEN, J. J., SMITH, S. R., & SPADE, E. R. (2003). A Quantitative Evaluation of ENSO Indices. *Journal of Climate*, 16(8), 1249–1258. [https://doi.org/10.1175/1520-0442\(2003\)16<1249:AQEOEI>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)16<1249:AQEOEI>2.0.CO;2)
- HANNA, E., CROPPER, T. E., JONES, P. D., SCAIFE, A. A., & ALLAN, R. (2015). Recent seasonal asymmetric changes in the NAO (a marked summer decline and increased winter variability) and associated changes in the AO and Greenland Blocking Index. *International Journal of Climatology*, 35(9), 2540–2554. <https://doi.org/10.1002/joc.4157>
- HANSEN, J., SATO, M., HEARTY, P., RUEDY, R., KELLEY, M., MASSON-DELMOTTE, V., RUSSELL, G., TSELIODIS, G., CAO, J., RIGNOT, E., VELICOGNA, I., TORMEY, B., DONOVAN, B., KANDIANO, E., von SCHUCKMANN, K., KHARECHA, P., LEGRANDE, A. N., BAUER, M., & LO, K.-W. (2016). Ice melt, sea level rise and superstorms: Evidence from paleoclimate data, climate modeling, and modern observations that 2 °C global warming could be dangerous. *Atmospheric Chemistry and Physics*, 16(6), 3761–3812. <https://doi.org/10.5194/acp-16-3761-2016>
- HANSEN, J. W. (1999). Stochastic daily solar irradiance for biological modeling applications1Florida Agricultural Experiment Station, Journal Series No. R-06278.1. *Agricultural and Forest Meteorology*, 94(1), 53–63. [https://doi.org/10.1016/S0168-1923\(99\)00003-9](https://doi.org/10.1016/S0168-1923(99)00003-9)
- HANSEN, J.-P., & PERCEBOIS, J. (2017). *Transitions électriques: Ce que l'Europe et les marchés n'ont pas su vous dire* (Odile Jacob).
- HASSANI, H. (2007). *Singular Spectrum Analysis: Methodology and Comparison* [Text]. University Library of Munich, Germany. <https://EconPapers.repec.org/RePEc:pra:mprapa:4991>
- HAYASAKA, T. (2016). The Long-term Variation in Surface Shortwave Irradiance in China and Japan: A Review. *Journal of the Meteorological Society of Japan. Ser. II*, 94(5), 393–414. <https://doi.org/10.2151/jmsj.2016-024>
- HERSBACH, H. (2016). The ERA5 Atmospheric Reanalysis. *AGU Fall Meeting Abstracts*, 33. <http://adsabs.harvard.edu/abs/2016AGUFMNG33D..01H>
- HINKELMAN, L. M. (2019). The Global Radiative Energy Budget in MERRA and MERRA-2: Evaluation with Respect to CERES EBAF Data. *Journal of Climate*, 32(6), 1973–1994. <https://doi.org/10.1175/JCLI-D-18-0445.1>

REFERENCES :

- HIRABAYASHI, Y., MAHENDRAN, R., KOIRALA, S., KONOSHIMA, L., YAMAZAKI, D., WATANABE, S., KIM, H., & KANAE, S. (2013). Global flood risk under climate change. *Nature Climate Change*, 3(9), 816–821. <https://doi.org/10.1038/nclimate1911>
- HOANG, T. H., PAREY, S., & DACUNHA-CASTELLE, D. (2009). Multidimensional trends: The example of temperature. *The European Physical Journal Special Topics*, 174(1), 113–124. <https://doi.org/10.1140/epjst/e2009-01094-6>
- HOEFFDING, W. (1950). “Optimum” Nonparametric Tests. In N. I. Fisher & P. K. Sen (Eds.), *The Collected Works of Wassily Hoeffding* (pp. 227–236). Springer New York. https://doi.org/10.1007/978-1-4612-0865-5_11
- HOYER-KLICK, C., HUSTIG, F., SCHWANDT, M., & MEYER, R. (2009). Characteristic meteorological years from ground and satellite data. *Proc. of the 15th SolarPACES Symp. 2009*, 8.
- HULD, T. (2017). PVMAPS: Software tools and data for the estimation of solar radiation and photovoltaic module performance over large geographical areas. *Solar Energy*, 142, 171–181.
- IEA. (2019). *World energy statistics*. IEA. <https://www.iea.org/subscribe-to-data-services/world-energy-balances-and-statistics>
- International Astronomical Union General Assembly. (2015). *The IAU Strategic Plan 2010-2020: Astronomy for Development*. https://www.iau.org/static/resolutions/IAU2015_English.pdf
- IPCC. (2013). *AR5 Climate Change 2013: The Physical Science Basis*.
- IRENA. (2018). Renewable power generation costs in 2017. *International Renewable Energy Agency (IRENA): Abu Dhabi, UAE*, 160.
- JANOWIAK, J. E. (1988). An Investigation of Interannual Rainfall Variability in Africa. *Journal of Climate*, 1(3), 240–255. [https://doi.org/10.1175/1520-0442\(1988\)001<0240:AIOIRV>2.0.CO;2](https://doi.org/10.1175/1520-0442(1988)001<0240:AIOIRV>2.0.CO;2)
- JEREZ, S., THAIS, F., TOBIN, I., WILD, M., COLETTE, A., YIOU, P., & VAUTARD, R. (2015). The CLIMIX model: A tool to create and evaluate spatially-resolved scenarios of photovoltaic and wind power development. *Renewable and Sustainable Energy Reviews*, 42, 1–15. <https://doi.org/10.1016/j.rser.2014.09.041>
- KARIUKI, B. W., & SATO, T. (2018). Interannual and spatial variability of solar radiation energy potential in Kenya using Meteosat satellite. *Renewable Energy*, 116, 88–96.
- KARLSSON, K.-G., ANTTILA, K., TRENTMANN, J., STENGEL, M., FOKKE MEIRINK, J., DEVASTHALE, A., HANSCHMANN, T., KOTHE, S., JÄÄSKELÄINEN, E., SEDLAR, J., BENAS, N., van ZADELHOFF, G.-J., SCHLUNDT, C., STEIN, D., FINKENSIEPER, S., HÅKANSSON, N., & HOLLMANN, R. (2017). CLARA-A2: The second edition of the CM SAF cloud and radiation data record from 34 years of global AVHRR data. *Atmospheric Chemistry and Physics*, 17(9), 5809–5828. <https://doi.org/10.5194/acp-17-5809-2017>
- KARLSSON, K.-G., RIIHELÄ, A., MÜLLER, R., MEIRINK, J. F., SEDLAR, J., STENGEL, M., LOCKHOFF, M., TRENTMANN, J., KASPAR, F., HOLLMANN, R., & WOLTER, E. (2013). CLARA-A1: A cloud, albedo, and radiation dataset from 28 yr of global AVHRR data. *Atmospheric Chemistry and Physics*, 13(10), 5351–5367. <https://doi.org/10.5194/acp-13-5351-2013>

- KATO, S., ROSE, F. G., RUTAN, D. A., THORSEN, T. J., LOEB, N. G., DOELLING, D. R., HUANG, X., SMITH, W. L., SU, W., & HAM, S.-H. (2018). Surface Irradiances of Edition 4.0 Clouds and the Earth's Radiant Energy System (CERES) Energy Balanced and Filled (EBAF) Data Product. *Journal of Climate*, 31(11), 4501–4527. <https://doi.org/10.1175/JCLI-D-17-0523.1>
- KENDALL, M. (1975). Rank correlation methods (4th edn.) Charles Griffin. *San Francisco, CA*, 8.
- KLEISSL, J. (Ed.). (2013). *Solar energy forecasting and resource assessment* (First edition). Elsevier, AP, Academic Press in an imprint of Elsevier.
- KNIGHT, J. R., FOLLAND, C. K., & SCAIFE, A. A. (2006). Climate impacts of the Atlantic Multidecadal Oscillation. *Geophysical Research Letters*, 33(17). <https://doi.org/10.1029/2006GL026242>
- KNUDSEN, M. F., SEIDENKRANTZ, M.-S., JACOBSEN, B. H., & KUIJPERS, A. (2011). Tracking the Atlantic Multidecadal Oscillation through the last 8,000 years. *Nature Communications*, 2(1), 178. <https://doi.org/10.1038/ncomms1186>
- KOPP, G., & LEAN, J. L. (2011). A new, lower value of total solar irradiance: Evidence and climate significance: FRONTIER. *Geophysical Research Letters*, 38(1), n/a-n/a. <https://doi.org/10.1029/2010GL045777>
- KÖPPEN, W. P., & GEIGER, G. (1936). Das geographische System der Klimate In: W. Köppen and G. Geiger, Ed. *Handbuch Der Klimatologie (Handbuch Der Klimatologie, Vol. 1: C. Gebr, Borntraeger)*.
- KRAKAUER, N. Y., & COHAN, D. S. (2017). Interannual Variability and Seasonal Predictability of Wind and Solar Resources. *Resources*, 6(3), 29. <https://doi.org/10.3390/resources6030029>
- KUNDZEWICZ, Z. W. (2011). Nonstationarity in Water Resources - Central European Perspective1: Nonstationarity in Water Resources - Central European Perspective. *JAWRA Journal of the American Water Resources Association*, 47(3), 550–562. <https://doi.org/10.1111/j.1752-1688.2011.00549.x>
- KUNDZEWICZ, Z. W., & RADZIEJEWSKI, M. (2000). *Methodologies for trend detection*. 13.
- KWIATKOWSKI, D., PHILLIPS, P. C. B., SCHMIDT, P., & SHIN, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54(1–3), 159–178. [https://doi.org/10.1016/0304-4076\(92\)90104-Y](https://doi.org/10.1016/0304-4076(92)90104-Y)
- LARRAÑETA, M., MORENO-TEJERA, S., LILLO-BRAVO, I., & SILVA-PÉREZ, M. A. (2018). A methodology for the stochastic generation of hourly synthetic direct normal irradiation time series. *Theoretical and Applied Climatology*, 131(3–4), 1517–1527. <https://doi.org/10.1007/s00704-017-2188-4>
- LARRAÑETA, M., MORENO-TEJERA, S., SILVA-PÉREZ, M. A., & LILO-BRAVO, I. (2015). An improved model for the synthetic generation of high temporal resolution direct normal irradiation time series. *Solar Energy*, 122, 517–528. <https://doi.org/10.1016/j.solener.2015.09.030>
- LEFEVRE, M., WALD, L., & DIABATE, L. (2007). Using reduced data sets ISCCP-B2 from the Meteosat satellites to assess surface solar irradiance. *Solar Energy*, 81(2), 240–253.

REFERENCES :

- LI, P. (2015). *Temporal and Spatial Variability of Surface Solar Radiation over the South-West Indian Ocean and Reunion Island: Regional Climate Modeling* [LE2P - Laboratoire d'Énergétique, d'Électronique et Procédés - Faculté des Sciences et techniques]. <https://hal.univ-reunion.fr/LE2P/tel-01326960v1>
- LOHMANN, S., SCHILLINGS, C., MAYER, B., & MEYER, R. (2006). Long-term variability of solar direct and global radiation derived from ISCCP data and comparison with reanalysis data. *Solar Energy*, *80*(11), 1390–1401. <https://doi.org/10.1016/j.solener.2006.03.004>
- LUO, J.-J., WANG, G., & DOMMENGET, D. (2018). May common model biases reduce CMIP5's ability to simulate the recent Pacific La Niña-like cooling? *Climate Dynamics*, *50*(3–4), 1335–1351. <https://doi.org/10.1007/s00382-017-3688-8>
- MAKARAU, A., & JURY, M. R. (1997). Seasonal cycle of convective spells over southern Africa during austral summer. *International Journal of Climatology*, *17*(12), 1317–1332. [https://doi.org/10.1002/\(SICI\)1097-0088\(199710\)17:12<1317::AID-JOC197>3.0.CO;2-A](https://doi.org/10.1002/(SICI)1097-0088(199710)17:12<1317::AID-JOC197>3.0.CO;2-A)
- MANARA, V., BRUNETTI, M., CELOZZI, A., MAUGERI, M., SANCHEZ-LORENZO, A., & WILD, M. (2016). Detection of dimming/brightening in Italy from homogenized all-sky and clear-sky surface solar radiation records and underlying causes (1959–2013). *Atmospheric Chemistry and Physics*, *16*(17), 11145–11161. <https://doi.org/10.5194/acp-16-11145-2016>
- MANLY, B. F. J., & MACKENZIE, D. (2000). A cumulative sum type of method for environmental monitoring. *Environmetrics: The Official Journal of the International Environmetrics Society*, *11*(2), 151–166. [https://doi.org/10.1002/\(SICI\)1099-095X\(200003/04\)11:2<151::AID-ENV394>3.0.CO;2-B](https://doi.org/10.1002/(SICI)1099-095X(200003/04)11:2<151::AID-ENV394>3.0.CO;2-B)
- MANN, H. B. (1945). Nonparametric Tests Against Trend. *Econometrica: Journal of the Econometric Society*, *13*(3), 245–259. JSTOR. <https://doi.org/10.2307/1907187>
- MANN, H. B., & WHITNEY, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *The Annals of Mathematical Statistics*, *18*(1), 50–60. JSTOR.
- MARCHANT, R., MUMBI, C., BEHERA, S., & YAMAGATA, T. (2007). The Indian Ocean dipole – the unsung driver of climatic variability in East Africa. *African Journal of Ecology*, *45*(1), 4–16. <https://doi.org/10.1111/j.1365-2028.2006.00707.x>
- MARSAGLIA, G., TSANG, W. W., & WANG, J. (2003). Evaluating Kolmogorov's Distribution. *Journal of Statistical Software*, *8*(1), 1–4.
- MCGILCHRIST, C. A., & WOODYER, K. D. (1975). Note on a Distribution-free CUSUM Technique. *Technometrics*, *17*(3), 321–325. <https://doi.org/10.1080/00401706.1975.10489335>
- McMAHAN, A., GROVER, C., & VIGNOLA, F. (2013a). *Solar energy forecasting and resource assessment—Chapter 04—Evaluation of Resource Risk in Solar-Project Financing* (J. KLEISSL, Ed.; First edition). Elsevier, AP, Academic Press in an imprint of Elsevier.
- McMAHAN, A., GROVER, C., & VIGNOLA, F. (2013b). *Solar energy forecasting and resource assessment—Chapter 05—Bankable Solar-Radiation Datasets* (J. KLEISSL, Ed.; First edition). Elsevier, AP, Academic Press in an imprint of Elsevier.

- MEEHL, G. A., BOER, G. J., COVEY, C., LATIF, M., & STOUFFER, R. J. (2000). The Coupled Model Intercomparison Project (CMIP). *Bulletin of the American Meteorological Society*, 81(2), 313–318. JSTOR.
- MEFTAHA, M. (2017). *Nouveau regard sur quelques caractéristiques physiques du Soleil* [Phdthesis, Université Paris-Saclay]. <https://tel.archives-ouvertes.fr/tel-01593476>
- MEYER, R. (2010). Recommendations for Bankable Meteorological Site Assessments for Solar Thermal Power Plants. *Proc. of the SolarPACES Symp., Perpignan*, 8.
- MICHALSKY, J. J. (1988). The Astronomical Almanac's algorithm for approximate solar position (1950–2050). *Solar Energy*, 40(3), 227–235. [https://doi.org/10.1016/0038-092X\(88\)90045-X](https://doi.org/10.1016/0038-092X(88)90045-X)
- MILLER, S. D., HEIDINGER, A. K., & SENGUPTA, M. (2013). *Solar energy forecasting and resource assessment—Chapter 03—Physically Based Satellite Methods* (J. KLEISSL, Ed.; First edition). Elsevier, AP, Academic Press in an imprint of Elsevier.
- MISHCHENKO, M. I., GEOGDZHAYEV, I. V., ROSSOW, W. B., CAIRNS, B., CARLSON, B. E., LACIS, A. A., LIU, L., & TRAVIS, L. D. (2007). Long-Term Satellite Record Reveals Likely Recent Aerosol Trend. *Science*, 315(5818), 1543–1543. <https://doi.org/10.1126/science.1136709>
- MISRA, V. (2003). The Influence of Pacific SST variability on the precipitation over Southern Africa. *Journal of Climate*, 16(14), 2408–2418. <https://doi.org/10.1175/2785.1>
- MO, R., FYFE, J., & DEROME, J. (1998). Phase-locked and asymmetric correlations of the wintertime atmospheric patterns with the ENSO. *Atmosphere-Ocean*, 36(3), 213–239. <https://doi.org/10.1080/07055900.1998.9649612>
- MORON, V., VAUTARD, R., & GHIL, M. (1998). Trends, interdecadal and interannual oscillations in global sea-surface temperatures. *Climate Dynamics*, 14(7–8), 545–569. <https://doi.org/10.1007/s003820050241>
- MOSER, D., BUONO, M. D., JAHN, U., HERZ, M., RICHTER, M., & BRABANDERE, K. D. (2017). Identification of technical risks in the photovoltaic value chain and quantification of the economic impact. *Progress in Photovoltaics: Research and Applications*, 25(7), 592–604. <https://doi.org/10.1002/pip.2857>
- MÜLLER, B., HEYDENREICH, W., KIEFER, K., & REISE, C. (2009). More Insights from the Monitoring of Real World PV Power Plants—A Comparison of Measured to Predicted Performance of PV Systems. *24th European Photovoltaic Solar Energy Conference, 21-25 September 2009, Hamburg, Germany*, 5.
- MÜLLER, B., WILD, M., DRIESSE, A., & BEHRENS, K. (2014). Rethinking solar resource assessments in the context of global dimming and brightening. *Solar Energy*, 99, 272–282. <https://doi.org/10.1016/j.solener.2013.11.013>
- MÜLLER, R., MATSOUKAS, C., GRATZKI, A., BEHR, H. D., & HOLLMANN, R. (2009). The CM-SAF operational scheme for the satellite based retrieval of solar surface irradiance—A LUT based eigenvector hybrid approach. *Remote Sensing of Environment*, 113(5), 1012–1024. <https://doi.org/10.1016/j.rse.2009.01.012>

REFERENCES :

- NICHOLSON, S. E., & GRIST, J. P. (2003). The seasonal evolution of the atmospheric circulation over West Africa and equatorial Africa. *Journal of Climate*, 16(7), 1013–1030. [https://doi.org/10.1175/1520-0442\(2003\)016<1013:TSEOTA>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<1013:TSEOTA>2.0.CO;2)
- NORRIS, J. R., & WILD, M. (2007). Trends in aerosol radiative effects over Europe inferred from observed cloud cover, solar “dimming,” and solar “brightening”. *Journal of Geophysical Research: Atmospheres*, 112(D8). <https://doi.org/10.1029/2006JD007794>
- Observ'ER. (2019a). *5 familles énergies renouvelables*. http://www.energies-renouvelables.org/energies_renouvelables.asp
- Observ'ER. (2019b). *Le baromètre des énergies renouvelables électriques en France*. http://www.energies-renouvelables.org/observ-er/html/energie_renouvelable_france.asp
- OHMURA, A. (2009). Observed decadal variations in surface solar radiation and their causes. *Journal of Geophysical Research: Atmospheres*, 114(D10). <https://doi.org/10.1029/2008JD011290>
- OHVRIL, H., TERAL, H., NEIMAN, L., KANNEL, M., UUSTARE, M., TEE, M., RUSSAK, V., OKULOV, O., JÖEVEER, A., KALLIS, A., OHBRIL, T., TEREZ, E. I., TEREZ, G. A., GUSHCHIN, G. K., ABAKUMOVA, G. M., GORBARENKO, E. V., TSVETKOV, A. V., & LAULAINEN, N. (2009). Global dimming and brightening versus atmospheric column transparency, Europe, 1906–2007. *Journal of Geophysical Research: Atmospheres*, 114(D10). <https://doi.org/10.1029/2008JD010644>
- OPPENHEIM, A., SCHAFFER, R., & BUCK, J. (1999). *Discrete-Time Signal Processing*.
- PAETH, H., LATIF, M., & HENSE, A. (2003). Global SST influence on twentieth century NAO variability. *Climate Dynamics*, 21(1), 63–75. <https://doi.org/10.1007/s00382-003-0318-4>
- PARDING, K. M., LIEPERT, B. G., HINKELMAN, L. M., ACKERMAN, T. P., DAGESTAD, K.-F., & OLSETH, J. A. (2016). Influence of Synoptic Weather Patterns on Solar Irradiance Variability in Northern Europe. *Journal of Climate*, 29(11), 4229–4250. <https://doi.org/10.1175/JCLI-D-15-0476.1>
- PAREY, S., HOANG, T. T. H., & DACUNHA-CASTELLE, D. (2018). Future high-temperature extremes and stationarity. *Natural Hazards*. <https://doi.org/10.1007/s11069-018-3499-1>
- PARK, C.-E., JEONG, S.-J., JOSHI, M., OSBORN, T. J., HO, C.-H., PIAO, S., CHEN, D., LIU, J., YANG, H., PARK, H., KIM, B.-M., & FENG, S. (2018). Keeping global warming within 1.5 °C constrains emergence of aridification. *Nature Climate Change*, 8(1), 70–74. <https://doi.org/10.1038/s41558-017-0034-4>
- PEEL, M. C., FINLAYSON, B. L., & MCMAHON, T. A. (2007). Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences Discussions*, 4(2), 439–473.
- PETTITT, A. N. (1979). A Non-Parametric Approach to the Change-Point Problem. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 28(2), 126–135. <https://doi.org/10.2307/2346729>
- PFEIFROTH, U., BOJANOWSKI, J. S., CLERBAUX, N., MANARA, V., SANCHEZ-LORENZO, A., TRENTMANN, J., WALAWENDER, J. P., & HOLLMANN, R. (2018). Satellite-based trends of solar radiation and cloud parameters in Europe. *Advances in Science and Research*, 15, 31–37. <https://doi.org/10.5194/asr-15-31-2018>

- PFEIFROTH, U., SANCHEZ-LORENZO, A., MANARA, V., TRENTMANN, J., & HOLLMANN, R. (2018). Trends and Variability of Surface Solar Radiation in Europe Based On Surface- and Satellite-Based Data Records. *Journal of Geophysical Research: Atmospheres*, 123(3), 1735–1754. <https://doi.org/10.1002/2017JD027418>
- PHILIPONA, R., BEHRENS, K., & RUCKSTUHL, C. (2009). How declining aerosols and rising greenhouse gases forced rapid warming in Europe since the 1980s. *Geophysical Research Letters*, 36(2). <https://doi.org/10.1029/2008GL036350>
- PHILIPPON, N., ROUAULT, M., RICHARD, Y., & FAVRE, A. (2012). The influence of ENSO on winter rainfall in South Africa. *International Journal of Climatology*, 32(15), 2333–2347. <https://doi.org/10.1002/joc.3403>
- PHILLIPS, P. C. B., & PERRON, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2), 335–346. <https://doi.org/10.1093/biomet/75.2.335>
- POLO, J., ZARZALEJO, L. F., MARCHANTE, R., & NAVARRO, A. A. (2011). A simple approach to the synthetic generation of solar irradiance time series with high temporal resolution. *Solar Energy*, 85(5), 1164–1170. <https://doi.org/10.1016/j.solener.2011.03.011>
- POSSELT, R., MÜLLER, R., TRENTMANN, J., STOCKLI, R. C., & LINIGER, M. A. (2014). A surface radiation climatology across two Meteosat satellite generations. *Remote Sensing of Environment*, 142, 103–110. <https://doi.org/10.1016/j.rse.2013.11.007>
- POZO-VÁZQUEZ, D., TOVAR-PESCADOR, J., GÁMIZ-FORTIS, S. R., ESTEBAN-PARRA, M. J., & CASTRO-DÍEZ, Y. (2004). NAO and solar radiation variability in the European North Atlantic region: NAO AND SOLAR RADIATION. *Geophysical Research Letters*, 31(5), n/a-n/a. <https://doi.org/10.1029/2003GL018502>
- QASMI, S., CASSOU, C., & BOÉ, J. (2017). Teleconnection Between Atlantic Multidecadal Variability and European Temperature: Diversity and Evaluation of the Coupled Model Intercomparison Project Phase 5 Models. *Geophysical Research Letters*, 44(21), 11,140-11,149. <https://doi.org/10.1002/2017GL074886>
- REDA, I., & ANDREAS, A. (2004). Solar position algorithm for solar radiation applications. *Solar Energy*, 76(5), 577–589. <https://doi.org/10.1016/j.solener.2003.12.003>
- REEVES, J., CHEN, J., WANG, X. L., LUND, R., & LU, Q. Q. (2007). A Review and Comparison of Changepoint Detection Techniques for Climate Data. *Journal of Applied Meteorology and Climatology*, 46(6), 900–915. <https://doi.org/10.1175/JAM2493.1>
- REN21. (2019). *Renewables 2019 Global Status Report*. <https://www.ren21.net/gsr-2019>
- RIENECKER, M. M., SUAREZ, M. J., GELARO, R., TODLING, R., BACMEISTER, J., LIU, E., BOSILOVICH, M. G., SCHUBERT, S. D., TAKACS, L., KIM, G.-K., BLOOM, S., CHEN, J., COLLINS, D., CONATY, A., da SILVA, A. M., GU, W., JOINER, J., KOSTER, R. D., LUCCHESI, R., ... WOOLLEN, J. (2011). MERRA: NASA's Modern-Era Retrospective Analysis for Research and Applications. *Journal of Climate*, 24(14), 3624–3648. <https://doi.org/10.1175/JCLI-D-11-00015.1>
- RIGOLLIER, C., BAUER, O., & WALD, L. (2000). On the clear sky model of the ESRA — European Solar Radiation Atlas—With respect to the heliosat method. *Solar Energy*, 68(1), 33–48. [https://doi.org/10.1016/S0038-092X\(99\)00055-9](https://doi.org/10.1016/S0038-092X(99)00055-9)

REFERENCES :

- ROCHA, A., & SIMMONDS, I. (1997). Interannual variability of south-eastern African summer rainfall. Part 1: Relationships with air–sea interaction processes. *International Journal of Climatology*, 17(3), 235–265. [https://doi.org/10.1002/\(SICI\)1097-0088\(19970315\)17:3<235::AID-JOC123>3.0.CO;2-N](https://doi.org/10.1002/(SICI)1097-0088(19970315)17:3<235::AID-JOC123>3.0.CO;2-N)
- ROMERA, R., GAERTNER, M. Á., SÁNCHEZ, E., DOMÍNGUEZ, M., GONZÁLEZ-AMELEMÁN, J. J., & MIGLIETTA, M. M. (2017). Climate change projections of medicanes with a large multi-model ensemble of regional climate models. *Global and Planetary Change*, 151, 134–143.
- RUCKSTUHL, C., PHILIPONA, R., BEHRENS, K., COLLAUD COEN, M., DÜRR, B., HEIMO, A., MÄTZLER, C., NYEKI, S., OHMURA, A., VUILLEUMIER, L., WELLER, M., WEHRLI, C., & ZELENKA, A. (2008). Aerosol and cloud effects on solar brightening and the recent rapid warming. *Geophysical Research Letters*, 35(12), n/a-n/a. <https://doi.org/10.1029/2008GL034228>
- RUIZ-ARIAS, J. A., CEBECAUER, T., TOVAR-PESCADOR, J., & ŠÚRI, M. (2010). Spatial disaggregation of satellite-derived irradiance using a high-resolution digital elevation model. *Solar Energy*, 84(9), 1644–1657. <https://doi.org/10.1016/j.solener.2010.06.002>
- SAID, S. E., & DICKEY, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3), 599–607. <https://doi.org/10.1093/biomet/71.3.599>
- SAJI, N. H., GOSWAMI, B. N., VINAYACHANDRAN, P. N., & YAMAGATA, T. (1999). A dipole mode in the tropical Indian Ocean. *Nature*, 401(6751), 360–363. <https://doi.org/10.1038/43854>
- SANCHEZ-LORENZO, A., CALBÓ, J., & WILD, M. (2013). Global and diffuse solar radiation in Spain: Building a homogeneous dataset and assessing their trends. *Global and Planetary Change*, 100, 343–352. <https://doi.org/10.1016/j.gloplacha.2012.11.010>
- SANCHEZ-LORENZO, A., ENRIQUEZ-ALONZO, A., WILD, M., TRENTMANN, J., VICENTE-SERRANO, S. M., SANCHEZ-ROMERO, A., POSSELT, R., & HAKUBA, M. Z. (2017). Trends in downward surface solar radiation from satellites and ground observations over Europe during 1983–2010. *Remote Sensing of Environment*, 189, 108–117. <https://doi.org/10.1016/j.rse.2016.11.018>
- SANCHEZ-LORENZO, A., WILD, M., BRUNETTI, M., GUIJARRO, J. A., HAKUBA, M. Z., CALBÓ, J., MYSTAKIDIS, S., & BARTOK, B. (2015). Reassessment and update of long-term trends in downward surface shortwave radiation over Europe (1939–2012). *Journal of Geophysical Research: Atmospheres*, 120(18), 9555–9569. <https://doi.org/10.1002/2015JD023321>
- SCHMIDT, G. A., SHINDELL, D. T., & TSIGARIDIS, K. (2014). Reconciling warming trends. *Nature Geoscience*, 7(3), 158–160. <https://doi.org/10.1038/ngeo2105>
- SCHNITZER, M., THUMAN, C., JOHNSON, P., & TRUEPOWER, A. (2012). *The impact of solar uncertainty on project financeability: Mitigating energy risk through on-site monitoring*. 5.
- ŞEN, Z. (2017). Probabilistic innovative solar irradiation estimation. *International Journal of Energy Research*, 41(2), 229–239. <https://doi.org/10.1002/er.3610>
- SENGUPTA, M., HABTE, A., KURTZ, S., DOBOS, A., WILBERT, S., LORENZ, E., STOFFEL, T., RENNE, D., GUEYMARD, C. A., 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* (p. 236 p.) [Report]. National Renewable Energy Laboratory. <https://hal-mines-paristech.archives-ouvertes.fr/hal-01184753>

- SHEATHER, S. J., & MARRON, J. S. (2012). Kernel Quantile Estimators. *Journal of the American Statistical Association*, 85(410), 410–416.
- SHIN, Y., & SCHMIDT, P. (1992). The KPSS stationarity test as a unit root test. *Economics Letters*, 38(4), 387–392. [https://doi.org/10.1016/0165-1765\(92\)90023-R](https://doi.org/10.1016/0165-1765(92)90023-R)
- SIEGEL, S., & CASTELLAN, N. J. (1988). Measures of association and their tests of significance. *Nonparametric Statistics for the Behavioral Sciences*, 224-312.
- SIMPSON, J. J., & DICKEY, T. D. (1981). The Relationship between Downward Irradiance and Upper Ocean Structure. *Journal of Physical Oceanography*, 11(3), 309–323. [https://doi.org/10.1175/1520-0485\(1981\)011<0309:TRBDIA>2.0.CO;2](https://doi.org/10.1175/1520-0485(1981)011<0309:TRBDIA>2.0.CO;2)
- SNEYERS, R. (1975). *Sur l'analyse statistique des séries d'observations*. Secrétariat de l'Organisation Météorologique Mondiale.
- SONALI, P., & NAGESH KUMAR, D. (2013). Review of trend detection methods and their application to detect temperature changes in India. *Journal of Hydrology*, 476, 212–227. <https://doi.org/10.1016/j.jhydrol.2012.10.034>
- STACKHOUSE, P. W., GUPTA, S. K., COX, S. J., MIKOVITZ, J. C., & ZHANG, T. (2009). *The NASA/GEWEX Surface Radiation Budget Version 3.0: Results And Analysis*. 11, 13112.
- STANHILL, G., & COHEN, S. (2001). Global dimming: A review of the evidence for a widespread and significant reduction in global radiation with discussion of its probable causes and possible agricultural consequences. *Agricultural and Forest Meteorology*, 107(4), 255–278. [https://doi.org/10.1016/S0168-1923\(00\)00241-0](https://doi.org/10.1016/S0168-1923(00)00241-0)
- STENGEL, M., STAPELBERG, S., SUS, O., FINKENSIEPER, S., WÜRZLER, B., PHILIPP, D., HOLLMANN, R., POULSEN, C., CHRISTENSEN, M., & MCGARRAGH, G. (2019). Cloud_cci AVHRR-PM dataset version 3: 35 year climatology of global cloud and radiation properties. *Earth System Science Data*. <https://doi.org/https://doi.org/10.5194/essd-2019-104>
- STOFFEL, T. (2013). *Solar energy forecasting and resource assessment—Chapter 01—Terms & Definitions* (J. KLEISSL, Ed.; First edition). Elsevier, AP, Academic Press in an imprint of Elsevier.
- STORCH, H. von, ZWIERS, F. W., & LIVEZEY, R. E. (2000). Statistical analysis in climate research. *Nature*, 404(6778), 544.
- STREETS, D. G., WU, Y., & CHIN, M. (2006). Two-decadal aerosol trends as a likely explanation of the global dimming/brightening transition. *Geophysical Research Letters*, 33(15). <https://doi.org/10.1029/2006GL026471>
- ŠŮRI, M., CEBECAUER, T., & SKOCZEK, A. (2011). SolarGIS: Solar data and online applications for PV planning and performance assessment. *26th European Photovoltaics Solar Energy Conference*.
- SUTTON, R. T., & HODSON, D. L. R. (2005). Atlantic Ocean Forcing of North American and European Summer Climate. *Science*, 309(5731), 115–118. <https://doi.org/10.1126/science.1109496>

REFERENCES :

- TERRY, M. E. (1952). Some Rank Order Tests which are most Powerful Against Specific Parametric Alternatives. *The Annals of Mathematical Statistics*, 23(3), 346–366. <https://doi.org/10.1214/aoms/1177729381>
- TEUKOLSKY, S. A., FLANNERY, B. P., PRESS, W. H., & VETTERLING, W. T. (1992). Numerical recipes in C. *SMR*, 693(1), 59–70.
- THEVENARD, D., & PELLAND, S. (2013). Estimating the uncertainty in long-term photovoltaic yield predictions. *Solar Energy*, 91, 432–445. <https://doi.org/10.1016/j.solener.2011.05.006>
- TINA, G., & GAGLIANO, S. (2011). Probabilistic analysis of weather data for a hybrid solar/wind energy system. *International Journal of Energy Research*, 35(3), 221–232. <https://doi.org/10.1002/er.1686>
- TOMSON, T., RUSSAK, V., & KALLIS, A. (2008). Dynamic Behavior of Solar Radiation. In V. BADESCU (Ed.), *Modeling Solar Radiation at the Earth's Surface* (pp. 257–281). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-77455-6_10
- TRENTMANN, J., & KOTHE, S. (2016). *CM SAF Cloud, Albedo, Radiation dataset, AVHRR-based, Edition 2 (CLARA-A2) Surface Radiation Products Product User Manual (2.1)*. https://www.cmsaf.eu/SharedDocs/Literatur/document/2016/saf_cm_dwd_pum_gac_rad_2_1_pdf.pdf?__blob=publicationFile
- TROCCOLI, A., GOODESS, C., JONES, P., PENNY, L., DORLING, S., HARPHAM, C., DUBUS, L., PAREY, S., CLAUDEL, S., & KHONG, D.-H. (2018). Creating a proof-of-concept climate service to assess future renewable energy mixes in Europe: An overview of the C3S ECEM project. *Advances in Science and Research*, 15, 191–205.
- UMMENHOFER, C. C., SEN GUPTA, A., ENGLAND, M. H., & REASON, C. J. C. (2009). Contributions of Indian Ocean Sea Surface Temperatures to Enhanced East African Rainfall. *Journal of Climate*, 22(4), 993–1013. <https://doi.org/10.1175/2008JCLI2493.1>
- URRACA, R., GRACIA-AMILLO, A. M., KOUBLI, E., HULD, T., TRENTMANN, J., RIIHELÄ, A., LINDFORS, A. V., PALMER, D., GOTTSCHALG, R., & ANTONANZAS-TORRES, F. (2017). Extensive validation of CM SAF surface radiation products over Europe. *Remote Sensing of Environment*, 199, 171–186. <https://doi.org/10.1016/j.rse.2017.07.013>
- VAN DER WAERDEN, B. L. (1952). Order tests for the two-sample problem and their power. *Indagationes Mathematicae (Proceedings)*, 453–458.
- VERNAY, C., PTAVAL, S., & BLANC, P. (2012). Calibration of long-term global horizontal irradiation estimated by HelioClim-3 through short-term local measurement campaigns: Extending of the results to European and African sites. *World Renewable Energy Forum*, 3, Pages 1943-1950-ISBN 978-162276092-3. <https://hal.archives-ouvertes.fr/hal-00769606>
- VIGNOLA, F., GROVER, C., LEMON, N., & McMAHAN, A. (2012). Building a bankable solar radiation dataset. *Solar Energy*, 86(8), 2218–2229. <https://doi.org/10.1016/j.solener.2012.05.013>
- VILLAVICENCIO, M., & FINON, D. (2018). The social efficiency of electricity transition policies based on renewables. Which ways of improvement? *The Energy Journal*, (Article En Cours de Soumission).

- VINAYACHANDRAN, P. N., FRANCIS, P. A., & RAO, S. A. (2009). Indian Ocean dipole: Processes and impacts. *Current Trends in Science*, 569–589.
- WANG, W., MATTHES, K., OMRANI, N.-E., & LATIF, M. (2016). Decadal variability of tropical tropopause temperature and its relationship to the Pacific Decadal Oscillation. *Scientific Reports*, 6(1), 1–12. <https://doi.org/10.1038/srep29537>
- WANG, Y., TRENTMANN, J., YUAN, W., & WILD, M. (2018). Validation of CM SAF CLARA-A2 and SARAH-E Surface Solar Radiation Datasets over China. *Remote Sensing*, 10(12), 1977. <https://doi.org/10.3390/rs10121977>
- WEI, L., WANG, D., & HUTSON, A. D. (2015). An Investigation of Quantile Function Estimators Relative to Quantile Confidence Interval Coverage. *Communications in Statistics - Theory and Methods*, 44(10), 2107–2135. <https://doi.org/10.1080/03610926.2013.775304>
- WILCOX, S., & GUEYMARD, C. A. (2010). *Spatial and temporal variability of the solar resource in the United States*. 8.
- WILD, M. (2009). Global dimming and brightening: A review. *Journal of Geophysical Research: Atmospheres*, 114(D10). <https://doi.org/10.1029/2008JD011470>
- WILD, M. (2012). Enlightening Global Dimming and Brightening. *Bulletin of the American Meteorological Society*, 93(1), 27–37. <https://doi.org/10.1175/BAMS-D-11-00074.1>
- WILD, M. (2013). Long term trends of solar radiation. *Greenhouse Gases*, 27.
- WILD, M. (2016). Decadal changes in radiative fluxes at land and ocean surfaces and their relevance for global warming. *Wiley Interdisciplinary Reviews: Climate Change*, 7(1), 91–107. <https://doi.org/10.1002/wcc.372>
- WILD, M., FOLINI, D., HAKUBA, M. Z., SCHÄR, C., SENEVIRATNE, S. I., KATO, S., RUTAN, D., AMMANN, C., WOOD, E. F., & KÖNIG-LANGLO, G. (2015). The energy balance over land and oceans: An assessment based on direct observations and CMIP5 climate models. *Climate Dynamics*, 44(11–12), 3393–3429. <https://doi.org/10.1007/s00382-014-2430-z>
- WILD, M., FOLINI, D., HENSCHER, F., FISCHER, N., & MÜLLER, B. (2015). Projections of long-term changes in solar radiation based on CMIP5 climate models and their influence on energy yields of photovoltaic systems. *Solar Energy*, 116, 12–24. <https://doi.org/10.1016/j.solener.2015.03.039>
- WILD, M., OHMURA, A., SCHÄR, C., MÜLLER, G., FOLINI, D., SCHWARZ, M., HAKUBA, M. Z., & SANCHEZ-LORENZO, A. (2017). The Global Energy Balance Archive (GEBA) version 2017: A database for worldwide measured surface energy fluxes. *Earth System Science Data*, 9(2), 601–613. <https://doi.org/10.5194/essd-9-601-2017>
- ZHANG, Y., ROSSOW, W. B., LARCIS, A. A., OINAS, V., & MISHCHENKO, M. I. (2004). Calculation of radiative fluxes from the surface to top of atmosphere based on ISCCP and other global data sets: Refinements of the radiative transfer model and the input data. *Journal of Geophysical Research: Atmospheres*, 109(D19). <https://doi.org/10.1029/2003JD004457>
- ZIELINSKI, R. (2004). Optimal Quantile Estimators Small Sample Approach. *Polish Academy of Sciences. Institute of Mathematics*.

REFERENCES :

Table des figures

Figure 1 : Evolution de la consommation d'énergie primaire mondiale entre 1993 et 2018, exprimée en millions de tonnes équivalent pétrole (Mtoe, Million Tonnes of Oil Equivalent). Source : (BP, 2019)	13
Figure 2 : Evolution de la puissance photovoltaïque installée raccordée en France Source : RTE – Wikipédia ...	16
Figure 3 : Evolution des puissances raccordées par trimestre sur trois segments du marché. Source : (Observ'ER, 2019b), d'après les chiffres du SDES et de l'observatoire de l'énergie solaire photovoltaïque.....	16
Figure 4 : Représentation schématique des trois composantes fondamentales de l'éclairément. Inspiré de Dambreville (2014).	19
Figure 5 : Séries temporelles des moyennes mensuelles d'éclairément global sur plan horizontal (GHI) mesurées par la station Beijing (GEBA) – Chine.	20
Figure 6 : Carte de la moyenne multi-annuelle de séries temporelles mensuelles d'éclairément fournies par la base de données MERRA 2.....	22
Figure 7 : Répartition géographique des 2500 stations GEBA. En jaune sont représentées les stations possédant au moins trois années de données. Figure extraite de (WILD et al., 2017).	31
Figure 8 : Répartition géographique des 689 stations GEBA sélectionnées. Couleur : moyenne multi-annuelle d'irradiation annuelle (kWh/m ²).....	35
Figure 9 : (a) Répartition géographique des 63 stations GEBA sélectionnées, (b) Répartition géographique des 104 stations GEBA sélectionnées. Couleur : moyenne multi-annuelle d'irradiation annuelle (kWh/m ²).	36
Figure 10 : Nombre maximum de données mensuelles consécutives manquantes pour chaque pixel de la base de données CLARA-A2. (a) entre 1992 et 2015, (b) entre 2001 et 2015	42
Figure 11 : Nombre de données mensuelles manquantes entre 2001 et 2015. En jaune sont représentées les zones exclues dans notre étude.	43
Figure 12 : Carte de la moyenne multi-annuelle de séries temporelles mensuelles d'éclairément fournies par la base de données CLARA-A2, en W/m ²	43
Figure 13 : Carte de la moyenne multi-annuelle de séries temporelles mensuelles d'éclairément fournies par la base de données IPSL-CM6A-LR.	48
Figure 14 : Séries temporelles des moyennes annuelles glissantes d'éclairément pour les stations de Kochi (a) et de Nara (b). En couleurs sont représentées les bases de données considérées : la base de données GEBA (bleu), CLARA-A2 (rouge), MERRA-2 (jaune) et IPSL-CM6-LR (gris).	49
Figure 15 : Représentation du P50 (gauche) et du P90 (droite) pour une distribution gaussienne de moyenne nulle et d'écart type égal à 0.2.	56
Figure 16 : Histogramme du nombre de valeurs du 2 nd tirage inférieures au P90 calculé à partir du 1 ^{er} tirage. .	60

TABLE DES FIGURES :

Figure 17 : Histogramme du nombre de valeurs du second tirage inférieures au P90 théorique. 61

Figure 18 : Résultats du test de Kolmogorov-Smirnov pour trois distributions : gaussienne, de Weibull et de Gumbel. Le test est appliqué aux 104 séries temporelles annuelles GEBA. Les couleurs représentent le rejet (jaune) ou le non rejet (bleu) de l’hypothèse nulle..... 64

Figure 19 : Ecarts entre les P90 calculés à partir des fonctions cumulatives de distribution de trois ajustements de distributions : de Gauss, de Weibull et de Gumbel. Les résultats sont calculés pour 104 stations GEBA, et exprimés en W/m^2 65

Figure 20 : Comparaison de CDF empiriques entre données d’éclairement in-situ mesurées à Eugene (USA) et données synthétiques générées. Figure extraite de Fernandéz-Peruchena et al. (2015) 70

Figure 21 : Comparaison de CDF empiriques entre données d’éclairement in-situ mesurées à Séville (Espagne) en 2014 (ligne épaisse bleue) et 50 données synthétiques annuelles. Figure extraite de Larrañeta et al. (2018). 73

Figure 22 : Tendances d’éclairement sur l’Europe dérivée des données SARAH-2 et CLARA-2, 1983-2015 et 1992-2015. Figure extraite de Pfeifroth et al. (2018)..... 86

Figure 23 : Tendances d’éclairement sur l’Europe dérivée des données CM-SAF, 1983-2010 et 1994-2010. Figure extraite de Sanchez-Lorenzo et al. (2017) 86

Figure 24 : Répartition géographique des stations GEBA sélectionnées pour l’étude sur la zone d’intérêt (droite). Les étoiles de couleurs mettent en évidence les stations dont les séries temporelles d’éclairement correspondantes sont représentées à gauche. 89

Figure 25 : Résultat des tests statistiques de stationnarité de Phillips-Peiron (PP), de non-stationnarité Kwiatkowski-Phillips-Schmid-Shin (KPSS) et Mann-Kendall (MK), appliqués à l’indice de clarté des quatre stations chinoise sélectionnées. Les couleurs représentent la stationnarité (vert) ou la non-stationnarité (bleu). L’intensité de la couleur représente le rejet de H_0 (foncé), ou l’absence de rejet (clair). 92

Figure 26 : Représentation schématique de la décomposition temps fréquence réalisée. 100

Figure 27 : Fonctions de transfert de modulation exprimées en périodes du banc de filtres utilisé..... 101

Figure 28 : Représentation schématique de l’influence des effets de bords sur le filtre considéré. 102

Figure 29 : Données affectées par des effets de bords en fonction de l’élément considéré, pour le cas illustratif d’une série temporelle entre 1970 et 2000..... 104

Figure 30 : Signal synthétique construit par une combinaison linéaire de sinus..... 105

Figure 31 : Signal synthétique similaire à celui présenté figure 3, auquel s’ajoute un changement abrupt. 106

Figure 32 : Vue 3-D de la décomposition temps-fréquence du signal synthétique construit par une combinaison linéaire de sinus grâce au banc de filtre choisi. Les zones grisées sont affectées par les effets de bords. 106

Figure 33 : Vue 3-D de la décomposition temps-fréquence du signal synthétique construit par une combinaison linéaire de sinus et soumis à un changement abrupt grâce au banc de filtre choisi. Les zones grisées sont affectées par les effets de bords..... 108

Figure 34 : Moyenne annuelle de l'éclairement (noir) et moyenne glissante sur cinq ans (bleu) pour la station GEBA Potsdam. Extraite de Wild et al. (2016).	111
Figure 35 : Equivalent de moyenne annuelle, de moyenne annuelle glissante, et de moyenne glissante sur cinq ans de l'éclairement. Ces équivalents sont construits par une somme d'éléments de la décomposition temps-fréquence de la station Potsdam.	111
Figure 36 : Moyenne annuelle de l'éclairement (noir fin) et filtre passe bas gaussien de 13 ans (noir épais) pour la station GEBA Potsdam. Extraite de Sanchez-Lorenzo et al.	112
Figure 37 : Equivalent de moyenne annuelle, de moyenne annuelle glissante, et du filtrage par un filtre passe bas gaussien de 13 ans. Ces équivalents sont construits par une somme d'éléments de la décomposition temps-fréquence de la station Murcia.	112
Figure 38 : Répartition géographique des coefficients de détermination entre les séries temporelles mensuelles d'indice de clarté et les séries temporelles représentatives de la VIA ($y_3 + y_6 + y_{12}$) (couleurs) des 104 stations GEBA sélectionnées.	119
Figure 39 : Répartition géographique des coefficients de détermination entre les séries temporelles mensuelles d'indice de clarté et les séries temporelles représentatives de la VIA ($y_3 + y_6 + y_{12}$) (couleurs), pour la région Europe (a) et Asie (b).	120
Figure 40 : Répartition de l'importance des variabilités inter-annuelles dans l'explication du signal total.	122
Figure 41 : Vue 3-D de la décomposition temps-fréquence de l'indice de clarté pour la station GEBA Kuala Lumpur. Les zones grisées sont affectées par les effets de bords.	124
Figure 42 : Vue 3-D de la décomposition temps-fréquence de l'indice de clarté pour la station GEBA Sapporo. Les zones grisées sont affectées par les effets de bords.	124
Figure 43 : (a) Répartition géographique des coefficients de détermination entre les séries temporelles mensuelles historiques et représentatives de la VIA. (b) Répartition géographique de la somme des coefficients de détermination entre les séries temporelles mensuelles historiques et représentatives de la VAD. Base de données CLARA-A2.	132
Figure 44 : (a) Répartition géographique des coefficients de détermination entre les séries temporelles mensuelles historiques et représentatives de la VIA. (b) Répartition géographique de la somme des coefficients de détermination entre les séries temporelles mensuelles historiques et représentatives de la VAD. Base de données MERRA-2.	134
Figure 45 : Exemples illustratifs des écarts relatifs entre la valeur de la variance de deux réalisations du modèle IPSL-CM6A-LR, et la valeur moyenne de la variance entre les 31 réalisations.	136
Figure 46 : Dispersion entre les deux réalisations extrêmes des 31 réalisations du modèle IPSL-CM6A-LR. La dispersion est définie par l'écart maximal entre les variances (écart relatif) pour chaque pixel.	137
Figure 47 : Ecarts relatifs entre les valeurs de variance de la réalisation du modèle IPSL-CM6A-LR sélectionné, et la valeur de variance moyenne entre les 31 réalisations. L'écart relatif, calculé pour chaque pixel, est exprimé en pourcentage et représentée en couleurs.	138
Figure 48 : (a) Représentation des régions à considérer avec attention : pixels dont la dispersion entre les écarts relatifs des différentes réalisations est supérieure à 20% (jaune), pixels dont l'écart relatif entre la réalisation	

TABLE DES FIGURES :

sélectionnée et la moyenne des réalisations est supérieure à 5% (vert). (b) Répartition géographique des variances des séries temporelles d'indice de clarté pour la réalisation sélectionnée. 139

Figure 49 : (a) Répartition géographique des coefficients de détermination entre les séries temporelles mensuelles historiques et représentatives de la VIA. (b) Répartition géographique de la somme des coefficients de détermination entre les séries temporelles mensuelles historiques et représentatives de la VAD. Base de données IPSL-CM6A-LR. 141

Figure 50 : Régions pour lesquelles les variations de la série temporelle d'indice de clarté est expliquée à plus de 95% par : la VIA (jaune), la VIA et la VMD (vert). En bleu sont représentées les zones pour lesquelles l'ensemble des variations inférieures à la décennie ne permettent pas d'expliquer 95% des variations du signal. (a) : CLARA-A2, (b) : MERRA-2, (c) : IPSL-CM6A-LR, (d) : GEBA. 144

Figure 51 : Représentation schématique des anomalies de températures de surface (couleurs) et des modifications des zones de forte couverture nuageuse lors d'événements ENSO et La Niña. Figure extraite de NOAA climate.gov..... 147

Figure 52 : Représentation schématique des variations de température de l'océan (rouge : anomalie positive, bleu : anomalie négative) et des variations de couverture nuageuses associées (blanc : activité accrue de la convection). Evènement IOD. Source: the Japan Agency for Marine-Earth Science and Technology 149

Figure 53 : (a) Régions pour lesquelles le coefficient de détermination est supérieur à 95 % (bleu), ou inférieur ou égal à 95 % (jaune). (b) Coefficient de variation des séries temporelle annuelles d'éclairement GEBA. Base de données MERRA-2. 153

Tableaux

Tableau 1 : Résumé des étapes du contrôle de qualité appliqué aux séries temporelles du réseau GEBA. (inspiré de Gilgen & Ohmurha (1999))	34
Tableau 2 : Code qualité fourni par l'ETH Zurich concernant les données GEBA (inspiré de Gilgen & Ohmurha (1999))	35
Tableau 3 : Résumé des caractéristiques de cinq bases de données satellitaires pertinentes pour notre étude.....	40
Tableau 4 : Résumé des caractéristiques de chaque base de données introduite au cours du chapitre.	51
Tableau 5 : Analyse des hypothèses de différentes méthodes d'estimation du P90.....	68
Tableau 6 : Analyse des hypothèses d'intérêts postulées dans différentes études de l'état de l'art. (V, X) traduisent si l'hypothèse considérée est postulée ou non. Les couleurs représentent les efforts faits pour dépasser ou quantifier les incertitudes liées à ces hypothèses : rouge pour une hypothèse posée implicitement ou considéré comme vraie sans justification, orange pour un effort fait par les auteurs pour dépasser l'hypothèse, sans pour autant y parvenir, blanc pour une hypothèse non postulée.....	77
Tableau 7 : Résumé des méthodes de prétraitement des données et des fenêtres temporelles glissantes utilisées pour la représentation des variations de long terme dans différents articles de la bibliographie.....	83

Annexe I

Afin de garantir le bien fondé du travail de cette thèse, un travail conséquent d'entretiens individuels d'acteur du domaine solaire a été réalisé. En particulier, une collaboration avec l'équipe SoDa de la société Transvalor a été mise en place dans le cadre du projet européen ECEM (TROCCOLI et al., 2018). SoDa coordonne et organise la distribution de données de rayonnement solaire et d'outils météorologiques pour l'étude, la prospection, et la surveillance de systèmes solaires thermiques et photovoltaïques. Avec 70 souscriptions annuelles et 342 000 requêtes automatiques journalières uniquement pour l'accès à une base de données particulière disponible sur le site de SoDa, SoDa rassemble une large communauté d'utilisateurs. Cet échantillon d'utilisateurs nous a semblé suffisamment important pour mettre en évidence un corpus de questions représentatif des questionnements globaux de la communauté industrielle photovoltaïque.

Nous nous sommes particulièrement intéressés aux questions pour lesquelles l'équipe SoDa n'avait pas de réponse immédiate. Nous avons regroupé les différentes questions utilisateurs répertoriés comme faisant écho à des questions de recherches plus fondamentales en 3 catégories. Chaque catégorie touche à un thème distinct : la représentativité des données historiques pour la prévision future, les possibles tendances passées et futures de rayonnement, et la fréquence d'occurrence future de d'importantes anomalies de rayonnement. Une formulation adaptée, de manière à être générale, des questions répertoriées sont rassemblées dans le Tableau 8.

Catégorie	Question utilisateur
Représentativité de l'historique pour la prévision future	« Dans quelle mesure les estimations du potentiel solaire d'un site réalisées à partir de données « bancables » historiques seront-elles encore valables dans 10 ans ? Dans 20 ans ? Dans 50 ans ? »
	« Nous avons pour habitude de travailler avec des données de 10 ou 20 ans. Les banques seraient rassurées si nous utilisions des données de plus long terme dans les rapports bancaires. Est-il envisageable d'utiliser d'autres sources de données, telles que les données de modèles climatiques, comme données bancables ? »
	« Nous considérons que plus l'échantillon de données de rayonnement est long, l'analyse qui en découle sera fiable. En prenant en compte le changement climatique, serait-il préférable de prendre des périodes d'analyses plus courtes, disons entre 5 et 10 ans plutôt que entre 15 et 20 ans ? »
	« Avez-vous une idée de l'évolution de la ressource solaire dans le futur ? Y aura-t-il un impact du changement climatique dans un futur proche ? Et pour combien de temps ? »

<p>Existence et prise en compte de tendances de rayonnement</p>	<p>« A priori, les températures ont tendance à augmenter, le taux d'aérosols dans l'atmosphère augmente depuis quelques années... Comment est-ce que cela affecte la production solaire ? Dans quelle mesure ces impacts dépendent de la zone géographique considérée, du climat, de la saison, de l'horizon temporel ? »</p> <p>« J'ai l'habitude de lire des messages claires et précis quant à l'évolution futurs des températures. Pourquoi n'est-il pas possible d'avoir un message similaire pour le rayonnement ? »</p> <p>« Est-ce que la vapeur d'eau et les aérosols vont changer dans le futur ? Est-ce que cela va jouer sur le spectre du rayonnement solaire reçu au sol par nos panneaux ? »</p>
<p>Occurrence de fortes anomalies de rayonnement</p>	<p>« Dans quelle mesure une mauvaise production annuelle peut-elle se reproduire durant la durée de vie de mon projet ? Est-ce que cela peut se produire plusieurs années de suite ? A partir de quand dois-je remettre en cause la viabilité de mon projet ? »</p>

Tableau 8: Résumé des questions utilisateurs d'intérêt scientifique dans le cadre de l'estimation de la ressource solaire

Dans le cadre de cette étude, nous pouvons mettre en évidence trois questions d'intérêt scientifique sous-jacentes à ce *corpus*, faisant écho aux objectifs de la thèse :

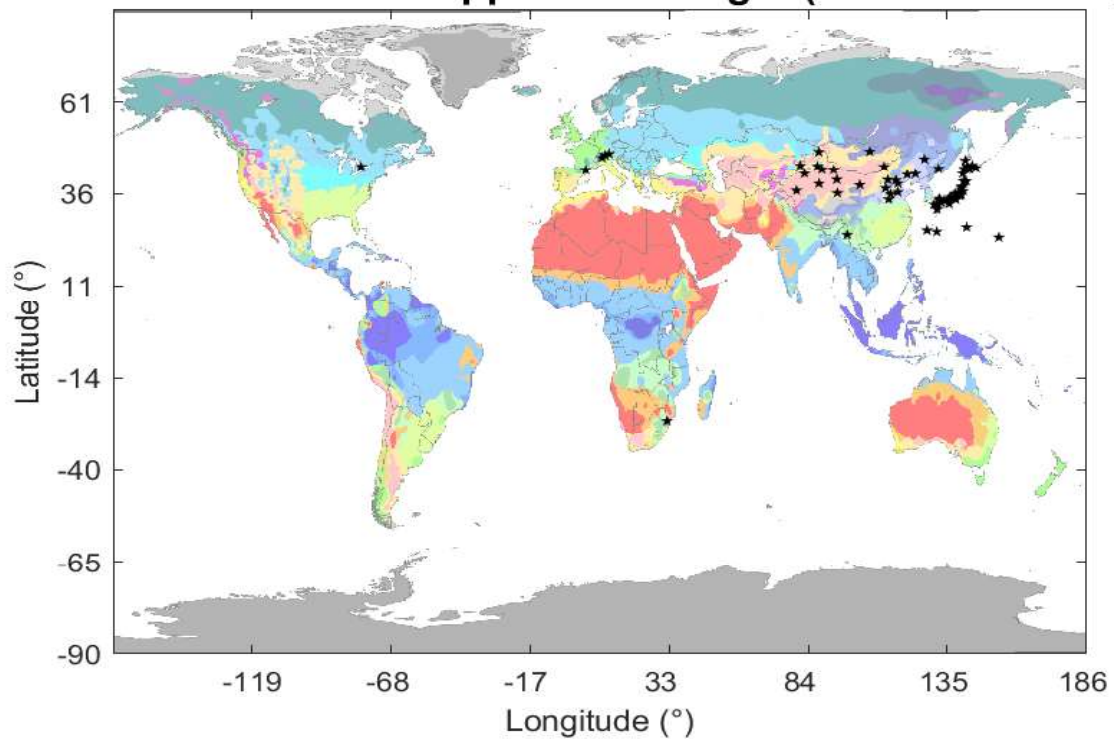
Dans le cadre d'un rapport bancaire, nous considérons entre 10 et 20 ans de données annuelles. Soit entre 10 et 20 données. Si peu de données peuvent-elles permettre de fournir des résultats statistiquement significatifs et robustes ? On retrouve l'objectif n°1 de la thèse.

L'analyse de données historiques comprises entre 10 et 20 ans peut-elle nous donner une information fiable quant à la ressource solaire future ? Autrement dit, cet échantillon couvre-t-il toutes les situations probables en termes de rayonnement ? Cet échantillon est-il suffisant pour représenter la variabilité interannuelle du site ? On retrouve l'objectif n°3 de la thèse.

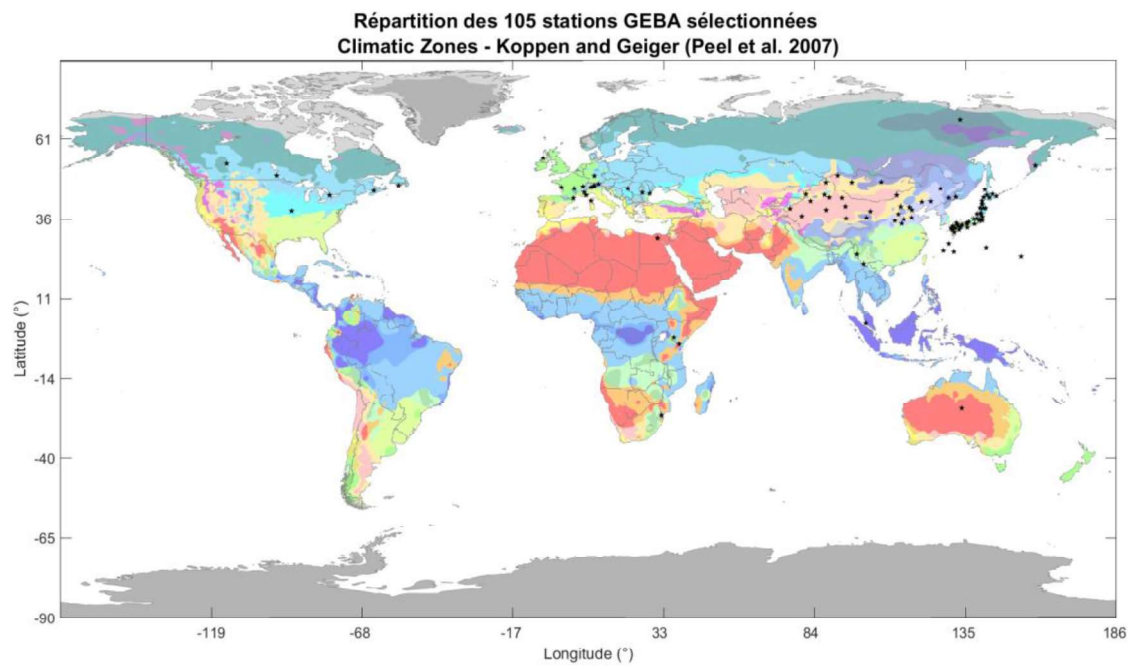
Le changement climatique a-t-il un impact significatif sur la quantité de rayonnement au sol qui devrait être pris en compte à l'échelle de la durée de vie de la centrale ? On retrouve, sous une forme plus large, l'objectif n°2 de la thèse.

Annexe II

**Répartition des 63 stations GEBA sélectionnées
Climatic Zones - Koppen and Geiger (Peel et al. 2007)**



Annexe III



Annexe IV

Station n°	Nom	Latitude (positif Nord)	Longitude (positif Est)	Année de début	Année de fin	Nombre de données mensuelles valides	Moyennes mensuelles (W/m ²)
St 1	Abashiri	44	144,3	1976	2007	390	107,4
St 2	Ajaccio	41,9	8,8	1978	2013	517	131,9
St 3	Akita	39,7	140,1	1975	2013	623	101,2
St 4	Alice	-23,8	133,9	1970	2006	526	188,2
St 5	Altay	47,7	88,1	1960	2000	492	129
St 6	Anyang	36,1	114,4	1961	1990	360	112,7
St 7	Aomori	40,8	140,8	1961	2007	568	103
St 8	Asahikawa	43,8	142,4	1976	2007	390	101,2
St 9	Bahtim	30,2	31,3	1969	2000	415	169,3
St 10	Beijing	39,9	116,3	1958	2013	680	122,2
St 11	Bucharest	44,5	26,1	1964	2009	554	117,4
St 12	Cairo	30,1	31,3	1969	2001	430	162,3
St 13	Chaoyang	41,6	120,5	1964	2000	452	121,6
St 14	Chichijima	27,1	142,2	1970	2013	530	136,2
St 15	Choshi	35,7	140,9	1973	2002	368	110,9
St 16	Clermont Ferrand	45,8	3,2	1978	2010	403	103,4
St 17	Columbia	38,8	-92,2	1951	1980	378	135,7
St 18	Constanta	44,2	28,6	1970	2004	476	120,2
St 19	Corvatsch	46,4	9,8	1981	2013	395	129,4
St 20	Datong	40,1	113,3	1961	2000	480	131,1
St 21	DavosDorf	46,8	9,9	1978	2007	360	111,4
St 22	Dunhuang	40,2	94,7	1966	2000	518	150,2
St 23	Edmonton	53,6	-114,1	1950	1987	494	107,2
St 24	Embrun	44,6	6,5	1980	2013	403	129,2

St 25	Erlianhaote	43,7	112	1958	2000	516	145,3
St 26	Fukushima	37,8	140,5	1961	2007	140,5	569
St 27	Golmud	36,4	94,9	1958	2000	94,9	163,3
St 28	Hakodate	41,8	140,8	1976	2007	140,8	102,5
St 29	Hamada	34,9	132,1	1961	2002	132,1	106,6
St 30	Hami	42,8	93,5	1961	2000	93,5	146,9
St 31	Harbin	45,8	126,8	1961	2013	126,8	109,4
St 32	Hikone	35,3	136,3	1961	2007	136,3	112,3
St 33	Hiroshima	34,4	132,5	1974	2007	132,5	115,9
St 34	Hotan	37,1	79,9	1958	2000	79,9	137,9
St 35	Houma	35,7	111,4	1961	2000	111,4	118,9
St 36	Innsbruck	47,3	11,4	1969	2005	11,4	103,1
St 37	Jinan	36,7	117	1961	2000	117	116
St 38	Jinghong	22	100,8	1961	2000	100,8	130,1
St 39	Kashgar	39,5	76	1958	2000	76	134,7
St 40	Kentville	45,1	-64,5	1961	1994	-64,5	106,8
St 41	Kochi	33,6	133,6	1976	2007	133,6	122,5
St 42	Kuala	3,1	101,6	1974	2008	101,6	138,7
St 43	Kumamoto	32,8	130,7	1961	2007	130,7	115
St 44	Kuqa	41,8	82,9	1958	1990	82,9	137
St 45	La	46,4	6,1	1981	2013	6,1	100,9
St 46	La	46,2	-1,2	1972	2013	-1,2	112,3
St 47	Locarno Monti	46,2	8,8	1972	2013	8,8	108,7
St 48	Maebashi	36,4	139,1	1972	2002	139,1	107,3
St 49	Malin	55,4	-7,3	1982	2013	-7,3	77,5
St 50	Maputo	-26	32,6	1955	2010	32,6	159,4
St 51	Matsumoto	36,3	138	1973	2006	138	122,3
St 52	Matsuyama	33,9	132,8	1966	2002	480	116,3

ANNEXE IV :

St 53	Minamidaitojima	25,8	131,2	1965	2007	555	136,6
St 54	Minamitorishima	24,3	154	1970	2013	530	160,1
St 55	Minqin	38,6	103,1	1961	2000	480	139,2
St 56	Miyako	39,7	142	1976	2007	552	111,8
St 57	Miyazaki	31,9	131,4	1977	2007	562	123,3
St 58	Morioka	39,7	141,2	1961	2007	570	106,4
St 59	Naha	26,2	127,7	1961	2013	633	122,7
St 60	Nairobi	-1,3	36,8	1957	1992	441	162,3
St 61	Nara	34,7	135,8	1961	2007	570	108,5
St 62	Naze	28,4	129,5	1961	2002	551	102,3
St 63	Nemuro	43,3	145,6	1976	2009	417	107,4
St 64	Nice	43,7	7,2	1982	2013	489	127,3
St 65	Niigata	37,9	139,1	1973	2002	366	102
St 66	Obihiro	42,9	143,2	1974	2007	413	107,2
St 67	Oita	33,2	131,6	1961	2007	569	111,9
St 68	Omaezaki	34,6	138,2	1973	2002	368	121,6
St 69	Onahama	37	140,9	1961	2007	570	115,8
St 70	Osaka	34,8	135,5	1973	2007	429	107,7
St 71	Perpignan	42,7	2,9	1981	2013	398	122,4
St 72	Petropavlovsk Kamchatski	53	158,8	1964	1997	541	95
St 73	Qira Ruoqiang	39	88,2	1958	2000	523	143,7
St 74	Saga	33,3	130,3	1961	2007	569	111,5
St 75	Saint Pierre	46,8	-56,2	1973	2013	492	93,2
St 76	Sapporo	43,1	141,3	1958	2013	676	103,2
St 77	Shenyang	41,7	123,5	1961	2013	635	114,4
St 78	Shimizu Ashizuri	32,7	133	1961	2008	573	123,7

St 79	Shimonoseki	34	130,9	1961	2007	565	111,9
St 80	Shionomisaki	33,5	135,8	1971	2008	574	122,1
St 81	Taiyuan	37,8	112,6	1961	2000	480	124,5
St 82	Takamatsu	34,3	134,1	1961	2007	570	117,3
St 83	Tateno	36,1	140,1	1958	1989	662	112
St 84	Tengchong	25	98,5	1961	2000	479	128,3
St 85	Tianjin	39,1	117,1	1959	2000	503	119,8
St 86	Timisoara	45,8	21,3	1964	2009	555	109,6
St 87	Tokyo	35,7	139,8	1973	2007	549	99,4
St 88	Toronto	43,7	-79,4	1955	1992	534	111,8
St 89	Turpan	42,9	89,2	1961	2000	491	132,6
St 90	Ulaangom	49,9	92,1	1964	1995	586	114,1
St 91	UlanBator	47,9	106,8	1964	2007	597	114,5
St 92	Ulyasutay	47,8	96,9	1967	2006	544	119,1
St 93	Urumqi	43,8	87,6	1959	1988	636	122,1
St 94	Verkhoyansk	67,6	133,4	1964	1995	368	83,1
St 95	Vladivostok	43,1	131,9	1964	1995	486	113,6
St 96	Voi	-3,4	38,6	1974	2003	439	173
St 97	Wakkanai	45,4	141,7	1961	2007	566	93,8
St 98	Winnipeg	49,9	-97,2	1950	1986	549	115,4
St 99	Wrangel	71	-178,5	1965	1994	318	86,9
St 100	Wuerzburg	49,8	10	1959	1988	595	92,4
St 101	Xining	36,6	101,8	1959	2000	503	136,3
St 102	Yamagata	38,3	140,4	1961	2007	564	107,3
St 103	Yanji	42,8	129,5	1960	2000	490	109,6
St 104	Yining	44	81,3	1961	2000	491	130,1
St 104	Zhengzhou	34,7	113,7	1961	2000	480	114,4

Annexe V

Test de Kolmogorov-Smirnov (MARSAGLIA et al., 2003)

Dans le cas pratique de l'ajustement d'une loi de probabilité paramétrique à une série temporelle d'éclairement, le test de Kolmogorov-Smirnov compare deux fonctions de distribution cumulatives (Cumulative Distribution Function, CDF) : l'une correspondant à la CDF des données initiales, la seconde correspondant à la CDF de l'ajustement paramétrique considéré. L'hypothèse nulle H_0 correspond au fait que la série de données d'éclairement pourrait être effectivement représentée par la distribution paramétrique ajustée. L'hypothèse alternative H_1 équivaut à dire qu'il n'est pas possible de représenter les données d'éclairement par la distribution d'intérêt. Plus précisément, la statistique du test correspond au maximum des différences absolues entre les deux CDF. Si la valeur de la statistique est supérieure ou égale à celle tabulée pour le seuil de confiance donné à 0.05, H_0 est rejetée.

Annexe VI

Loi normale

La loi normale est une loi de probabilité continue qui dépend de deux paramètres : son espérance (μ), et son écart type (σ). La densité de probabilité de la loi normale est donnée par :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

La figure 58 représente des exemples de fonctions de densités de probabilité de lois normales pour différents couples de paramètres (μ ; σ) : (0 ; 1) pour la courbe verte, (0 ; 0.5) pour la courbe violette, (0 ; 2) pour la courbe rouge, (-2 ; 5) pour la courbe jaune. On constate que plus l'écart type augmente, plus la fonction est étalée, ce qui fait de σ un paramètre d'échelle. La valeur de la moyenne μ correspond à l'abscisse du maxima de la fonction, ce qui fait de μ un paramètre de position. La courbe verte correspond à la loi normale centrée réduite, et est utilisée ici lors des tirages aléatoires.

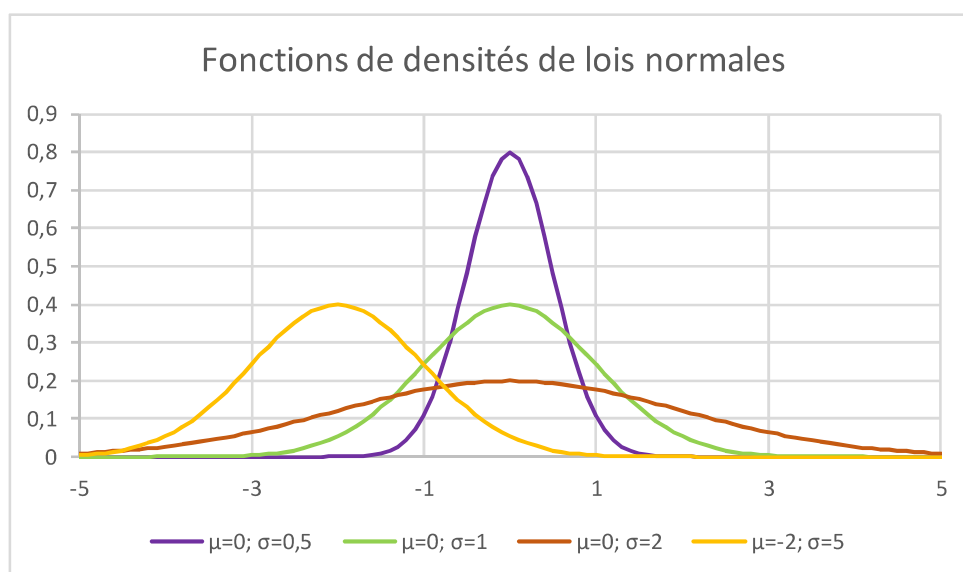


Figure 58 : Fonctions de densités de lois normales.

Loi de Weibull

La loi de Weibull correspond à un cas particulier de loi d'extremum généralisée au même titre que la loi de Gumbel. Ces deux lois ont donc pour objectif de représenter des phénomènes de valeurs extrêmes. Elles sont particulièrement utilisées dans le domaine de l'analyse de risques, où les scénarios « pire cas » sont d'intérêts.

La loi de Weibull est une loi de probabilité continue. Pour le cas de l'éclairement, valeur toujours positive, la loi de Weibull est définie par deux paramètres : le paramètre de forme (k) et le paramètre d'échelle (λ). Ces deux paramètres sont classiquement calculés de façon à maximiser la fonction de vraisemblance. Sa densité de probabilité est donnée par :

$$f(x; k, \lambda) = \frac{k}{\lambda} \left(\frac{x}{\lambda}\right)^{k-1} e^{-\left(\frac{x}{\lambda}\right)^k}$$

La figure 59 représente des fonctions de densités de probabilité de lois de Weibull pour différents couples de paramètres ($k ; \lambda$) : (2 ; 2) pour la courbe verte, (2 ; 0.5) pour la courbe violette, (3 ; 2) pour la courbe rouge. Comme attendu, la courbe s'aplatie quand le paramètre d'échelle λ augmente. k est un paramètre de forme, il va donc influencer la forme de la courbe : visuellement, on constate que plus k est grand, plus la courbe semble être symétrique.

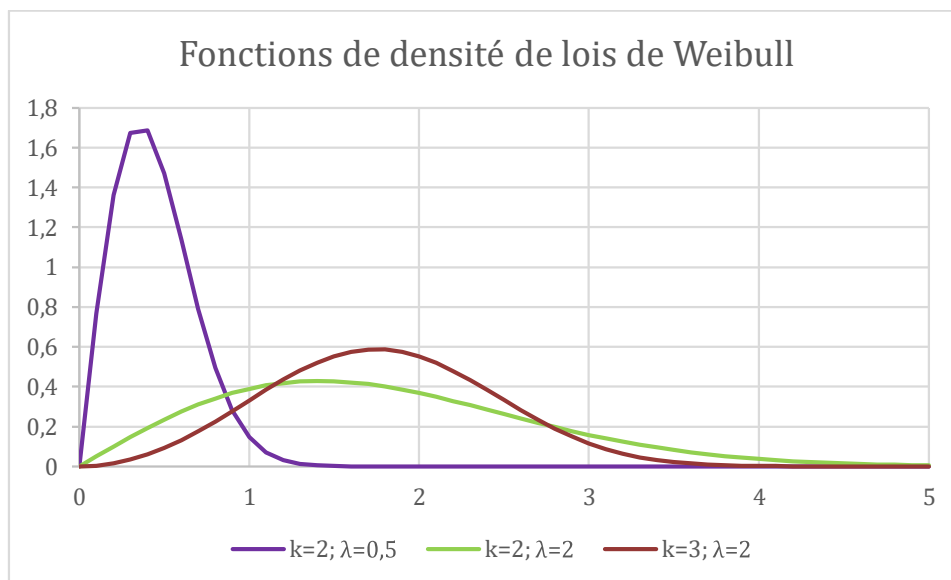


Figure 59 : Fonctions de densités de lois de Weibull.

Loi de Gumbel

Comme annoncé, la loi de Gumbel est une loi de probabilité continue appartenant à la famille des lois d'extremum généralisées. Elle dépend de deux paramètres : le paramètre de position (μ) et le paramètre d'échelle (β). Ces deux paramètres sont classiquement calculés de façon à maximiser la fonction de vraisemblance. Sa densité de probabilité est donnée par :

$$f(x; \mu, \beta) = \frac{z(x; \mu, \beta)e^{-z(x; \mu, \beta)}}{\beta}$$

$$\text{Avec } z(x; \mu, \beta) = e^{-\frac{x-\mu}{\beta}}$$

La figure 60 représente des fonctions de densités de probabilité de lois de Gumbel pour différents couples de paramètres (μ ; β) : (0 ; 2) pour la courbe verte, (0 ; 1) pour la courbe bleue, (5 ; 1) pour la courbe rouge. Comme attendu, μ étant un paramètre de position, il définit donc le maximum de la distribution, et β étant un paramètre d'échelle, sa valeur influe sur l'étalement de la courbe. On constate bien que plus β est grand et plus la courbe est aplatie.

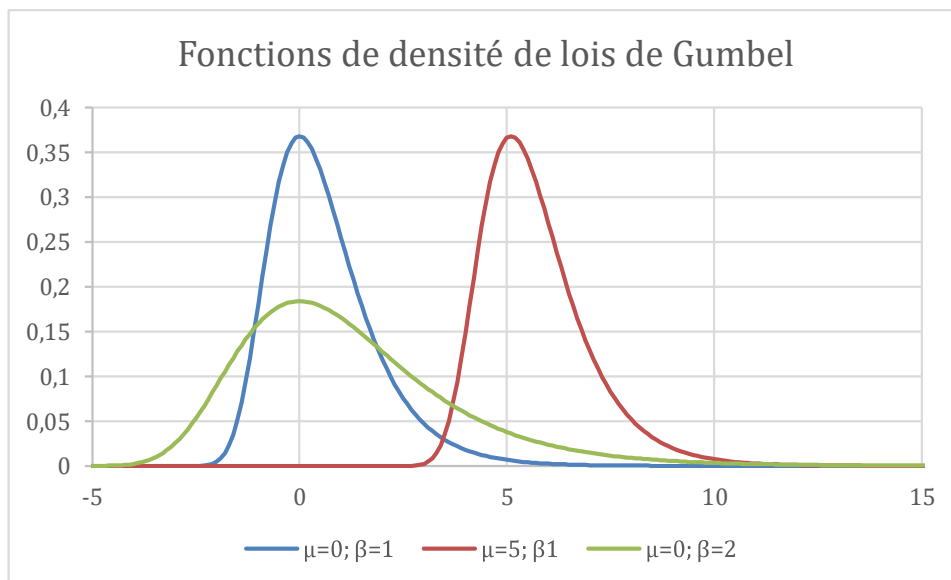


Figure 60 : Fonctions de densités de lois de Gumbel.

Annexe VII

P90 calculé à partir d'une loi	Normale (W/m ²)	Weibull (W/m ²)	Gumbel (W/m ²)		Normale (W/m ²)	Weibull (W/m ²)	Gumbel (W/m ²)
St 1	102,4	100,5	100,1	St 27	157,2	156,2	156,1
St 2	127,7	126,1	125,8	St 28	97,8	97,2	97,1
St 3	95,4	94,6	94,6	St 29	98,5	97,9	97,4
St 4	179,1	179,1	178,9	St 30	140,5	138,6	138,2
St 5	122,4	121,2	120,9	St 31	100,6	95,2	94,6
St 6	102,6	103,2	103,0	St 32	99,3	93,1	89,7
St 7	92,9	88,7	86,6	St 33	108,0	107,0	106,3
St 8	96,8	96,2	96,1	St 34	128,3	127,8	127,4
St 9	157,3	154,8	154,6	St 35	108,8	107,0	106,5
St 10	108,9	105,6	105,0	St 36	93,1	93,1	92,7
St 11	110,6	107,9	107,3	St 37	104,5	102,6	101,6
St 12	152,0	147,6	146,2	St 38	120,7	119,2	118,6
St 13	112,6	109,9	108,8	St 39	122,3	121,0	119,8
St 14	129,8	129,5	129,4	St 40	100,0	96,7	95,8
St 15	100,4	98,6	97,8	St 41	113,0	112,9	112,7
St 16	96,8	95,9	95,5	St 42	121,6	111,7	109,2
St 17	127,5	125,2	124,9	St 43	104,3	101,3	99,9
St 18	111,5	108,1	107,0	St 44	129,1	128,5	128,1
St 19	121,5	118,6	117,9	St 45	91,7	89,7	89,7
St 20	118,3	113,9	111,8	St 46	106,5	106,2	106,0
St 21	106,4	104,3	104,0	St 47	99,0	97,6	97,0
St 22	140,9	137,3	136,4	St 48	97,0	95,9	95,1
St 23	100,6	97,9	97,0	St 49	74,1	73,2	72,9
St 24	125,6	124,8	124,7	St 50	148,7	146,0	145,2
St 25	136,9	136,4	136,2	St 51	114,6	113,7	113,4
St 26	96,6	92,3	90,6	St 52	108,5	108,1	107,9

St 53	128,4	127,0	126,5	St 79	104,0	102,6	101,9
St 54	149,4	149,2	148,7	St 80	114,0	112,2	111,8
St 55	125,5	125,0	124,2	St 81	113,0	109,8	108,5
St 56	102,5	101,3	101,1	St 82	105,0	99,4	99,4
St 57	113,0	112,4	112,1	St 83	106,5	105,3	105,2
St 58	95,0	90,6	89,8	St 84	119,7	118,8	118,5
St 59	111,1	105,0	102,2	St 85	106,1	105,2	104,2
St 60	149,8	148,6	147,7	St 86	102,9	99,5	98,4
St 61	100,4	96,5	94,8	St 87	90,8	89,9	89,2
St 62	92,1	86,8	84,0	St 88	105,7	103,9	103,5
St 63	103,3	101,9	101,6	St 89	121,9	120,2	119,5
St 64	121,4	121,1	120,9	St 90	105,3	104,5	103,9
St 65	94,0	92,6	92,3	St 91	105,6	103,8	103,1
St 66	101,4	100,7	100,5	St 92	109,4	107,9	107,4
St 67	98,7	91,9	89,4	St 93	115,1	114,2	113,8
St 68	113,0	111,7	111,0	St 94	74,8	74,5	74,4
St 69	108,7	108,7	108,4	St 95	107,5	105,9	105,5
St 70	97,2	96,2	95,5	St 96	158,8	155,4	153,9
St 71	117,7	116,7	116,5	St 97	86,2	84,7	84,0
St 72	91,3	90,3	90,0	St 98	109,2	108,4	108,2
St 73	134,5	130,4	129,8	St 99	84,6	83,2	82,7
St 74	100,4	99,2	97,9	St 100	121,0	124,7	125,4
St 75	85,5	86,9	86,4	St 101	98,3	95,6	94,3
St 76	97,4	94,6	93,9	St 102	99,9	98,8	98,1
St 77	104,4	100,7	99,3	St 103	117,5	119,0	119,0
St 78	115,6	113,9	113,3	St 104	104,5	101,0	99,6

Annexe VIII

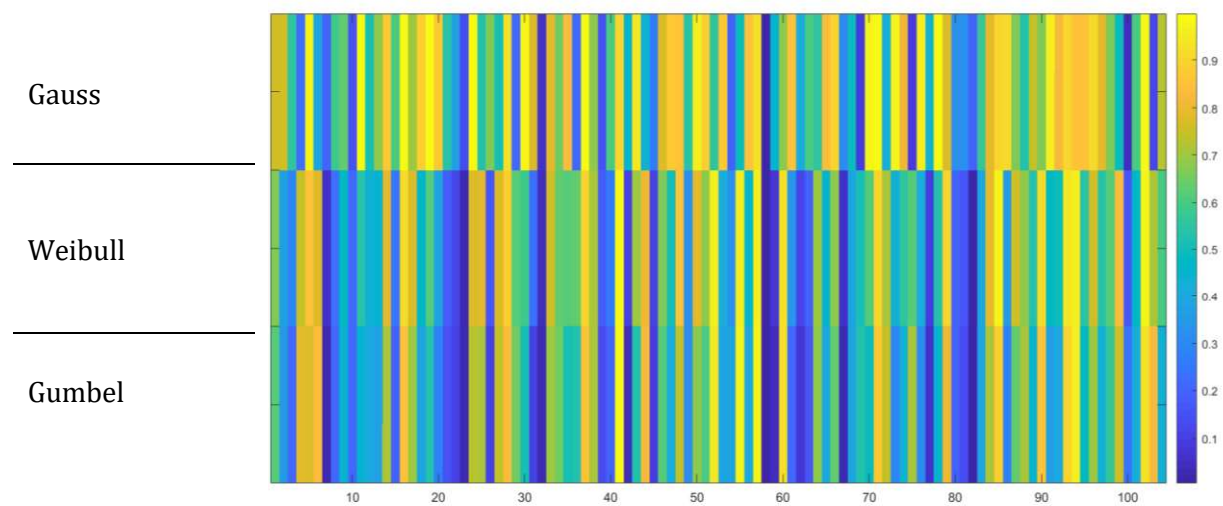


Figure 61 : p-value pour trois distributions d'intérêt : gaussienne, de Weibull et de Gumbel, appliquées aux 104 stations GEBA. Les couleurs représentent la valeur de la p-value.

Annexe IX

Depuis les années 1970, de nombreuses études ont été menées autour de méthodes statistiques de détection de changement, dans le contexte de séries temporelles de variables climatiques (BUISHAND, 1982; GILBERT, 1987; GILGEN et al., 1998; KUNDZEWICZ, 2011; KUNDZEWICZ & RADZIEJEWSKI, 2000; REEVES et al., 2007; SNEYERS, 1975; SONALI & NAGESH KUMAR, 2013). Usuellement, la stationnarité temporelle est définie par la correspondance entre les lois de probabilité des données passées et futures. Cette correspondance est qualifiée à l'aide de tests paramétriques, c'est-à-dire à l'aide de tests supposant que les données considérées suivent une loi de distribution connue, en général la loi normale. Cependant, nous avons bien vu au cours du chapitre 3 qu'il n'est pas aisé d'associer de façon robuste une loi de probabilité à l'échantillon considéré dans notre cas, car la série de données temporelle n'est composée que de 10 à 30 valeurs.

Une alternative possible réside dans l'utilisation de méthodes de transformation du jeu de données original, pour en faire un jeu de données qui suivent une loi normale, et ainsi pouvoir appliquer un test paramétrique sur ce dernier. À cause de la complexité de l'application de ces méthodes pour des échantillons aussi réduits que ceux considérés dans notre étude, cette alternative a été écartée. Les tests statistiques paramétriques n'ont donc pas été considérés dans le cadre de cette thèse.

Les tests choisis correspondent à des tests non-paramétriques simples et de références, construit avec des méthodologies diverses. Il sera considéré que si l'ensemble des tests convergent vers un résultat, celui-ci peut être considéré comme robuste. Plus précisément, nous pouvons détailler les tests sélectionnés en fonction de deux catégories :

- Les tests de non-stationnarité.
- Les tests de stationnarité.

Test statistiques de non-stationnarité

Parmi les tests de non-stationnarité basés sur la constance de paramètres les plus utilisés, nous pouvons citer les tests de détection d'une ou de plusieurs ruptures tels que les tests de Wilcoxon (Andél et al., 1985), le test de Mann-Whitney (MANN & WHITNEY, 1947), Terry-Hoeffding (HOEFFDING, 1950; TERRY, 1952), de Van der Waerden (VAN DER WAERDEN, 1952) et de Bell-Doksum (CONOVER, 1971), de Pettitt (PETTITT, 1979; SIEGEL & CASTELLAN, 1988; KUNDZEWICZ & RADZIEJEWSKI, 2000), ou encore le test de Kruskal-Wallis (SNEYERS, 1975), le test SNH (Standard Normal Homogeneity) (ALEXANDERSSON & Moberg, 1997), le test de Buishand (BUISHAND, 1982), le test du rapport de Von Neumann (ALEXANDERSSON, 1986), de CUSUM (MCGILCHRIST & WOODYER, 1975) ou encore de CUSUM adaptatif (MANLY & MACKENZIE, 2000; ALIPPI & ROVERI, 2006) ; les tests de détection de tendances linéaires, tel que le test de Mann-Kendall (MK) (KENDALL, 1975; MANN, 1945); ou les tests permettant de détecter aussi bien une tendance linéaire qu'une rupture tel que le test de KPSS (Kwiatkowski, Phillips, Schmidt, and Shin) (SHIN & SCHMIDT, 1992).

N'ayant aucune information ni sur le moment, ni sur le nombre des potentiels points de changement, nous avons écarté les tests de Wilcoxon, de Mann-Whitney, de Terry-Hoeffding, de Van der Waerden, de Bell-Doksum et de SNH. Le test de Kruskal-Wallis a été écarté car il suppose l'indépendance des échantillons considérés. Pour ce qui est du test de Buishand, bien qu'il puisse être utilisé sur des variables suivant des distributions quelconques, ses propriétés ont été particulièrement étudiées pour le cas d'une distribution normale. Nous ne l'avons donc pas considéré ici.

En conclusion, pour des raisons pratiques et opérationnelles, les tests de MK et de KPSS ont été sélectionnés. Ces deux tests adressent des questions spécifiques différentes. Ils peuvent donc être analysés conjointement, afin d'augmenter la robustesse des conclusions quant à la stationnarité temporelle de l'éclairement.

Test statistiques de non-stationnarité - KPSS

Pour le test de KPSS, ou test de Kwiatkowski-Phillips-Schmidt-Shin, le signal est supposé se décomposer en la somme d'une tendance linéaire, d'une marche aléatoire et d'un terme d'erreur stationnaire. Dans un premier temps, nous considérons le cas d'un signal sans tendance linéaire. Le signal est donc considéré comme la somme d'une marche aléatoire et d'un terme d'erreur stationnaire. L'hypothèse nulle H_0 correspond à la stationnarité de l'éclairement au sens d'une valeur nulle de la variance de la marche aléatoire. L'hypothèse alternative H_1 traduit la présence d'une racine unitaire, au sens d'une variance de la marche aléatoire positive. La statistique caractéristique de ce test K s'exprime comme :

$$\eta_{\mu} = \frac{1}{s^2 N^2} \sum_{t=1}^N S_t^2$$

Avec N la taille de l'échantillon, s^2 l'estimation de Newey-West pour la variance long terme et S_t la somme des écarts à la moyenne.

Le test consiste à comparer la valeur de la statistique avec la valeur de référence tabulée pour un seuil de confiance donné. La p-value est interpolée à partir d'une table trouvée dans Kwiatkowski et al. (1992).

Plus précisément, le test KPSS décompose le signal en la somme d'une tendance linéaire, d'une marche aléatoire et d'un terme d'erreur stationnaire. En d'autres termes, il évalue par défaut la tendance linéaire associée aux données d'entrée, la retranche, puis cherche la présence d'une marche aléatoire dans la série temporelle. Pour notre étude, nous nous sommes intéressés à la valeur du coefficient de tendance calculé par le modèle. Si celui-ci est inférieur à 10^{-3} , il pourra être conclu que la série temporelle ne connaît pas de tendance linéaire significative, et pourra être considérée comme stationnaire. L'information sur la présence d'une marche aléatoire ne sera pas analysée ici.

Test statistiques de non-stationnarité - MK

Le test de Mann-Kendall est un test non paramétrique permettant de détecter une tendance dans une série temporelle. La méthode consiste à relever le signe de la différence entre une valeur donnée et les $n-1$ autres valeurs de la série, et cela pour chaque point. Une somme cumulative S est ensuite effectuée, la valeur étant incrémentée de $+1$ si le signe relevé est positif, 0 en cas d'égalité des valeurs et -1 si le signe est négatif. Au final, une grande valeur positive (resp. négative) de S sera indicative de la présence d'une tendance positive (resp. négative). L'hypothèse nulle H_0 correspond à l'absence de tendance monotone dans la série temporelle, au sens $S = 0$. L'hypothèse alternative H_1 traduit la présence d'une tendance (au choix positive ou négative en fonction du signe de S) dans la série temporelle.

Condition : Les données ne doivent pas être auto-corrélées, ni collectés de façon saisonnière (alternance de tendances positives et négatives). Cette condition n'a pas été vérifiée dans notre cas, et pourrait être discutée comme illustré au cours du chapitre 3 pour le cas de l'éclairage. Cette condition peut expliquer une part des diversités observée dans les résultats issus des différents tests statistiques.

. La statistique caractéristique de ce test K s'exprime comme :

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \text{sign}(x_j - x_k)$$

Avec x_j la valeur de la série au temps j , n le nombre de valeurs de la série, et :

$$\text{sign}(x_j - x_i) = \begin{cases} 1 & \text{si } (x_j - x_i) > 0 \\ 0 & \text{si } (x_j - x_i) = 0 \\ -1 & \text{si } (x_j - x_i) < 0 \end{cases}$$

Le test de MK compare chaque point avec l'ensemble des points précédant celui-ci dans la série temporelle. Si la différence est positive (resp. négative), une somme cumulative est incrémentée positivement (resp. négativement). En fonction de la valeur finale de la somme cumulative, il est conclu quant à la présence d'une tendance positive dans le signal original (resp. négative). Ces deux tests sont présentés plus en détail en annexe IX.

Test statistiques de stationnarité

L'analyse bibliographique portant sur les tests de stationnarité a été plus succincte aux vues du fort consensus sur la qualité du test augmenté de Dickey Fuller (DICKEY & FULLER, 1979; SAID & DICKEY, 1984) et du test de Phillips-Perron (PP) (PHILLIPS & PERRON, 1988). Le premier étant un test paramétrique, nous n'avons considéré que le test de PP. Celui-ci correspond à une version non paramétrique du test de Dickey-Fuller. Il teste l'adéquation entre le signal original et trois modèles construits autour de quatre paramètres : le terme précédent dans la série temporelle, une constante, une dérive temporelle et un terme d'erreur. Ce modèle ainsi que la statistique associée sont détaillés en annexe X.

Test statistiques de stationnarité - PP

Ce test est une adaptation non paramétrique du test de Dickey Fuller. Les modèles sont identiques à ceux du test DF et pour chacun d'eux, il existe un calcul propre de deux statistiques. Nous n'avons pas jugé nécessaire de présenter ici l'ensemble des six modèles possibles. Plus de détail sont disponibles dans SAID & DICKEY (1984). L'hypothèse nulle H_0 réside dans la non-stationnarité de la série temporelle. L'hypothèse alternative H_1 correspond à la stationnarité de la série temporelle. Pour le cas testé ici d'un modèle sans constante ni dérive, les statistiques de ce test s'expriment comme :

$$Z(\hat{\varphi}) = T \hat{\varphi} - \frac{0.5 T^2 (\hat{\sigma}^2 - \hat{s}^2)}{\sum_{t=2}^T y_{t-1}^2}$$

$$Z(t_{\hat{\varphi}}) = \left(\frac{\hat{s}}{\hat{\sigma}}\right) t_{\hat{\varphi}} - \frac{0.5 T (\hat{\sigma}^2 - \hat{s}^2)}{(\hat{\sigma}^2 \sum_{t=2}^T y_{t-1}^2)^{1/2}}$$

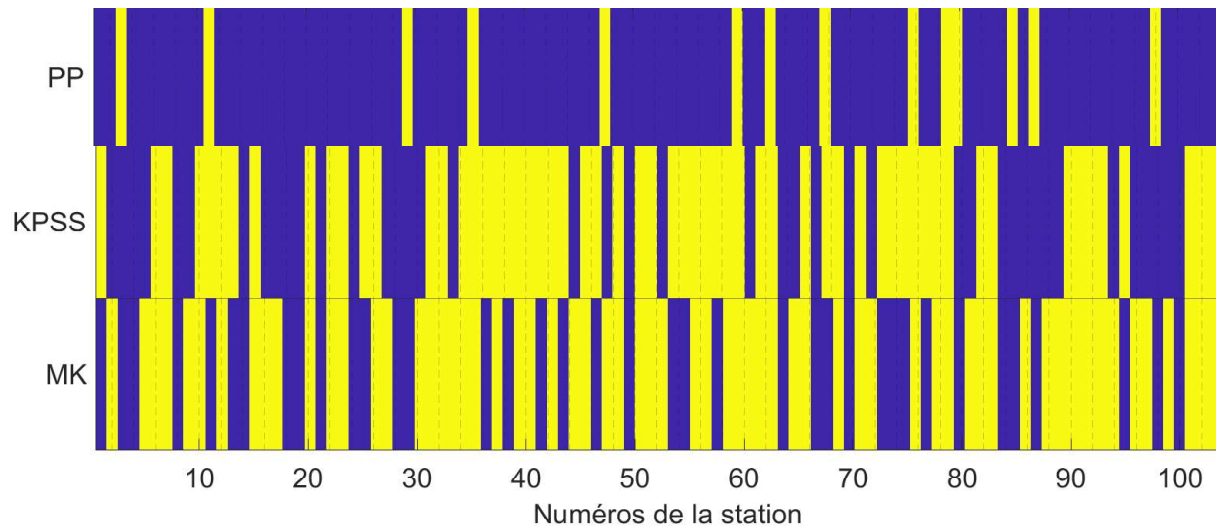
Avec $\hat{s}^2 = \frac{1}{T} \sum_{t=2}^T \hat{\varepsilon}_t^2$, $\hat{\sigma}^2 = \hat{s}^2 + \frac{2}{T} \sum_{l=1}^l w_{sl}$, ou $w_{sl} = 1 - \frac{s}{l+1}$ traduit le poids des auto-covariance. La valeur retenue pour l est, elle, donnée par $l = E T^{0.25}$.

Le test consiste à comparer la valeur de la statistique des valeurs tabulées pour un seuil donné.

Les trois tests statistiques introduits ont été implémentés sur la sélection de 104 stations GEBA, afin d'étudier la stationnarité temporelle de l'éclairement pour différents climats. La sélection de données GEBA décrites au cours du chapitre 2 remplit les critères de qualité donnés par Kundzewicz (2011), qui conseille de n'utiliser que des séries temporelles d'éclairement d'au moins 30 ans dans la détection de changements de variables climatiques. Ceci coïncide avec l'usage commun qui consiste à calculer les normales climatiques sur des périodes de 30 ans.

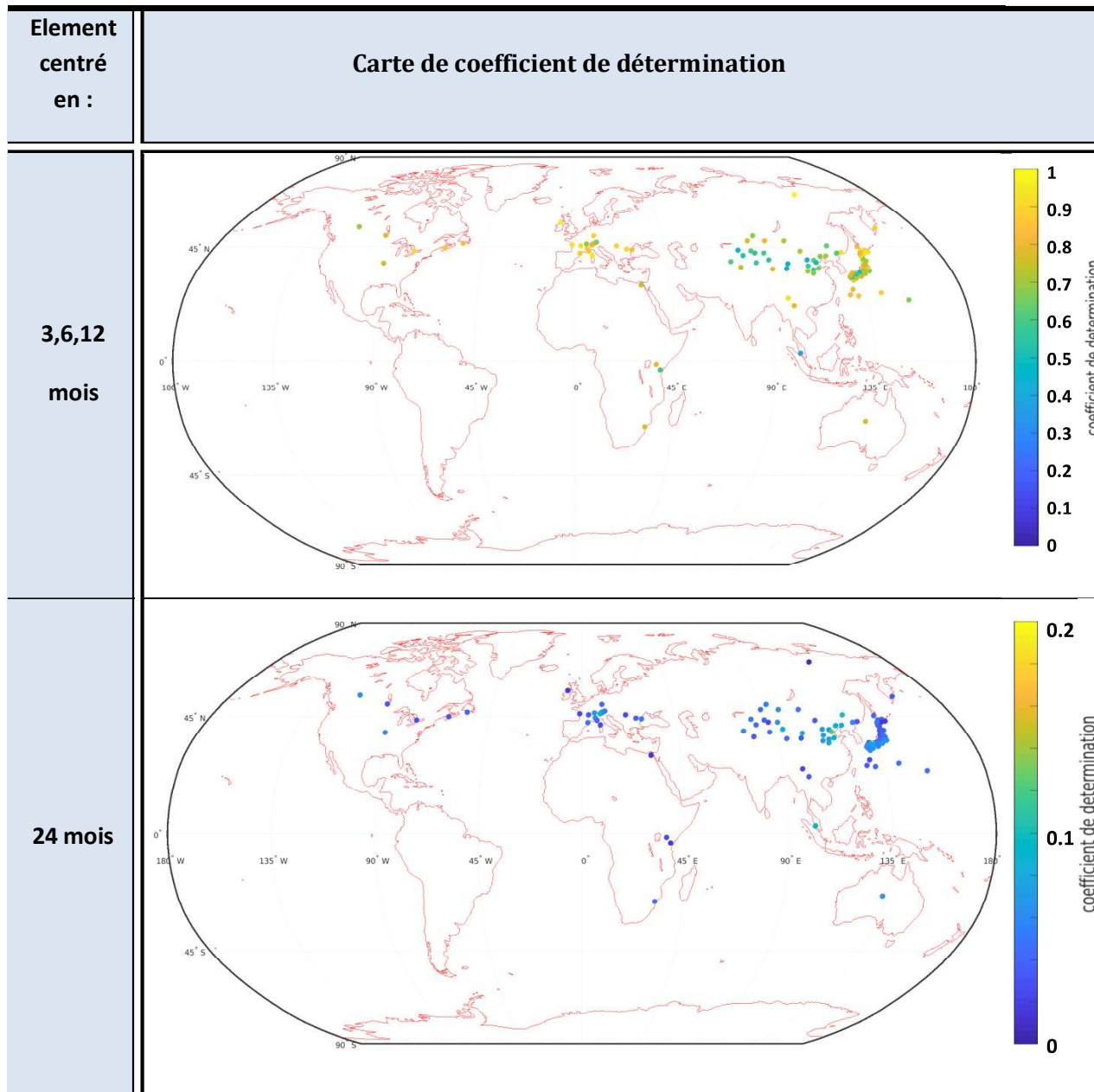
Annexe X

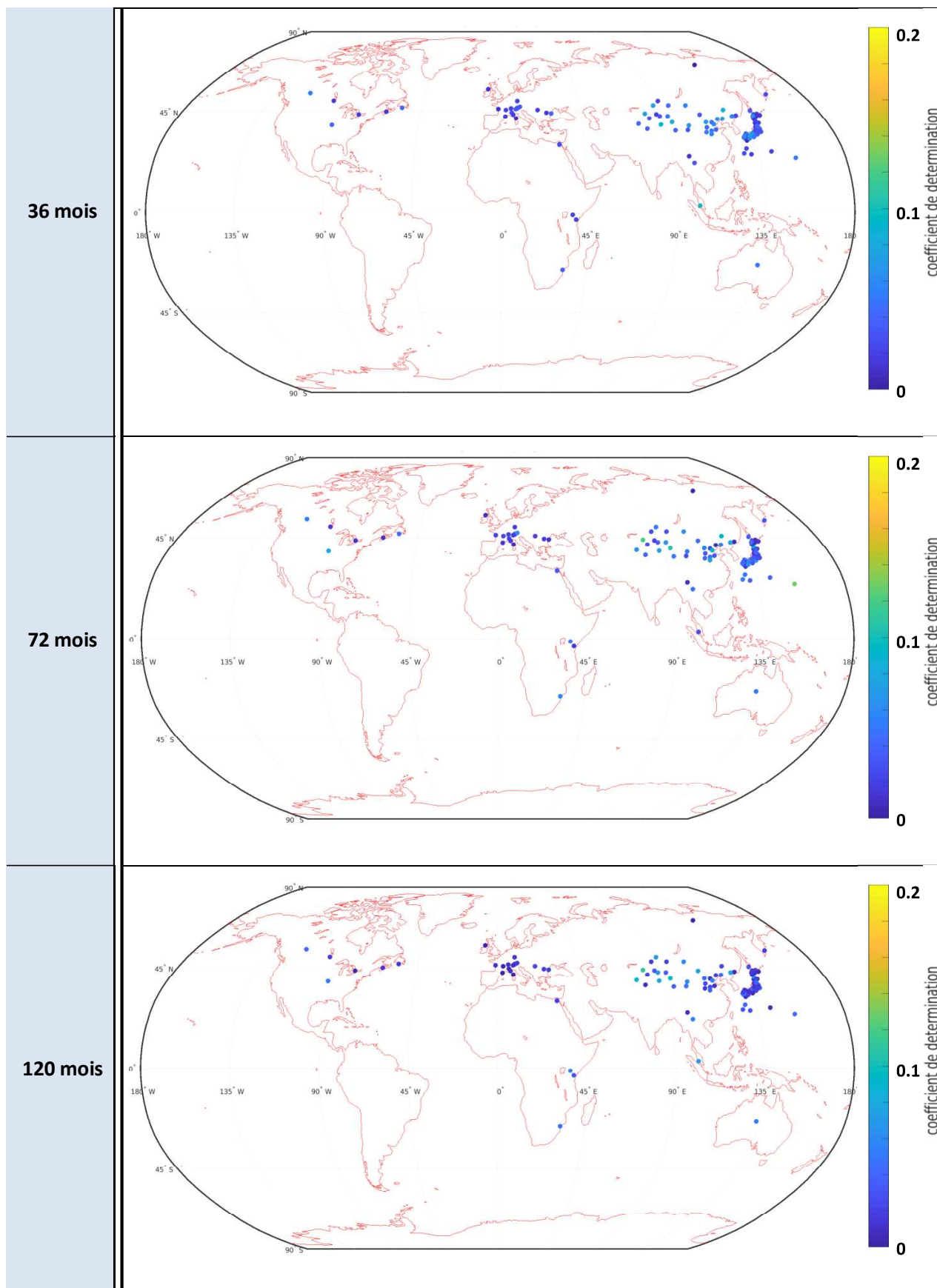
Résultats des tests statistiques pour les 104 stations GEBA

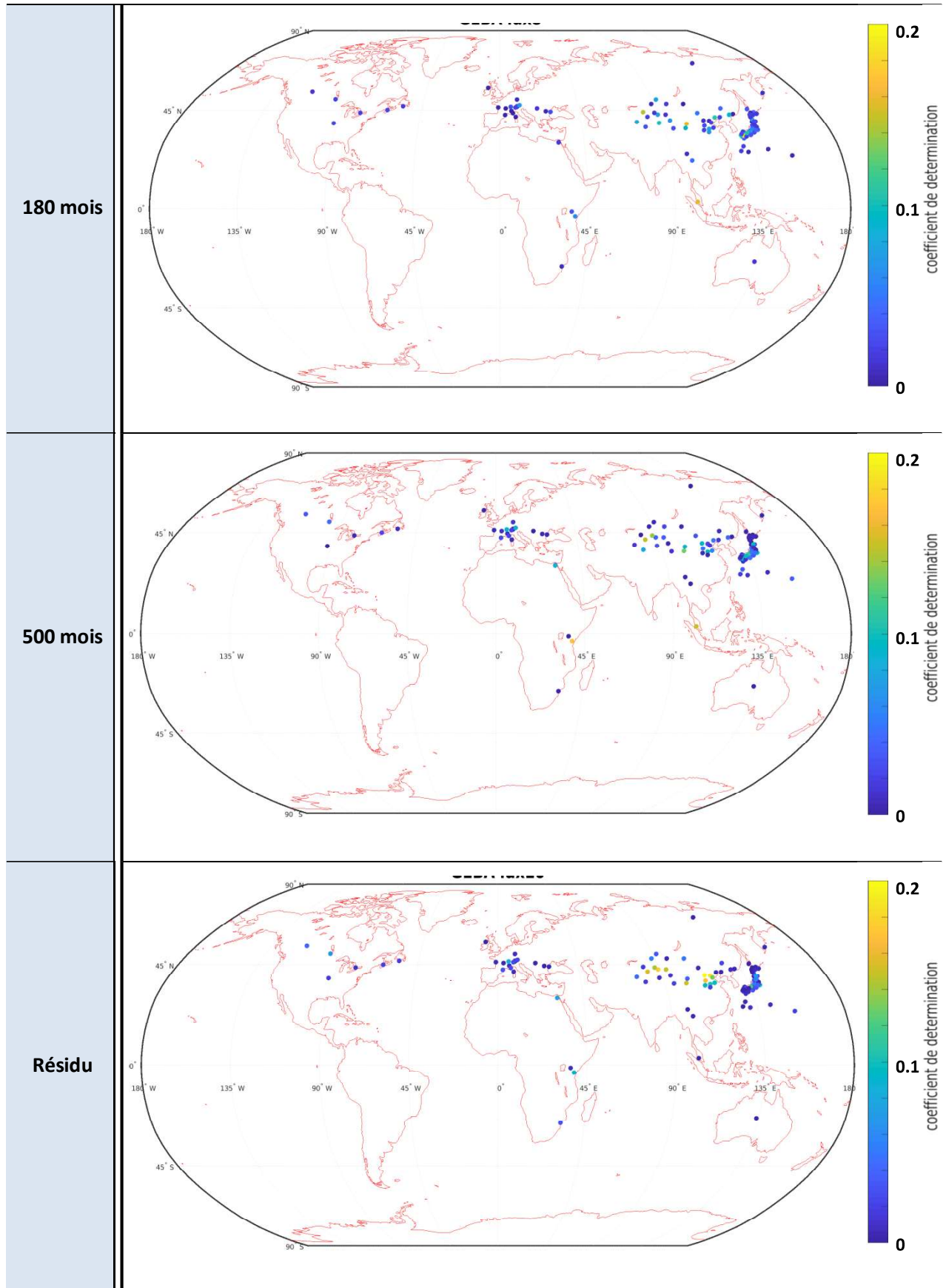


Annexe XI

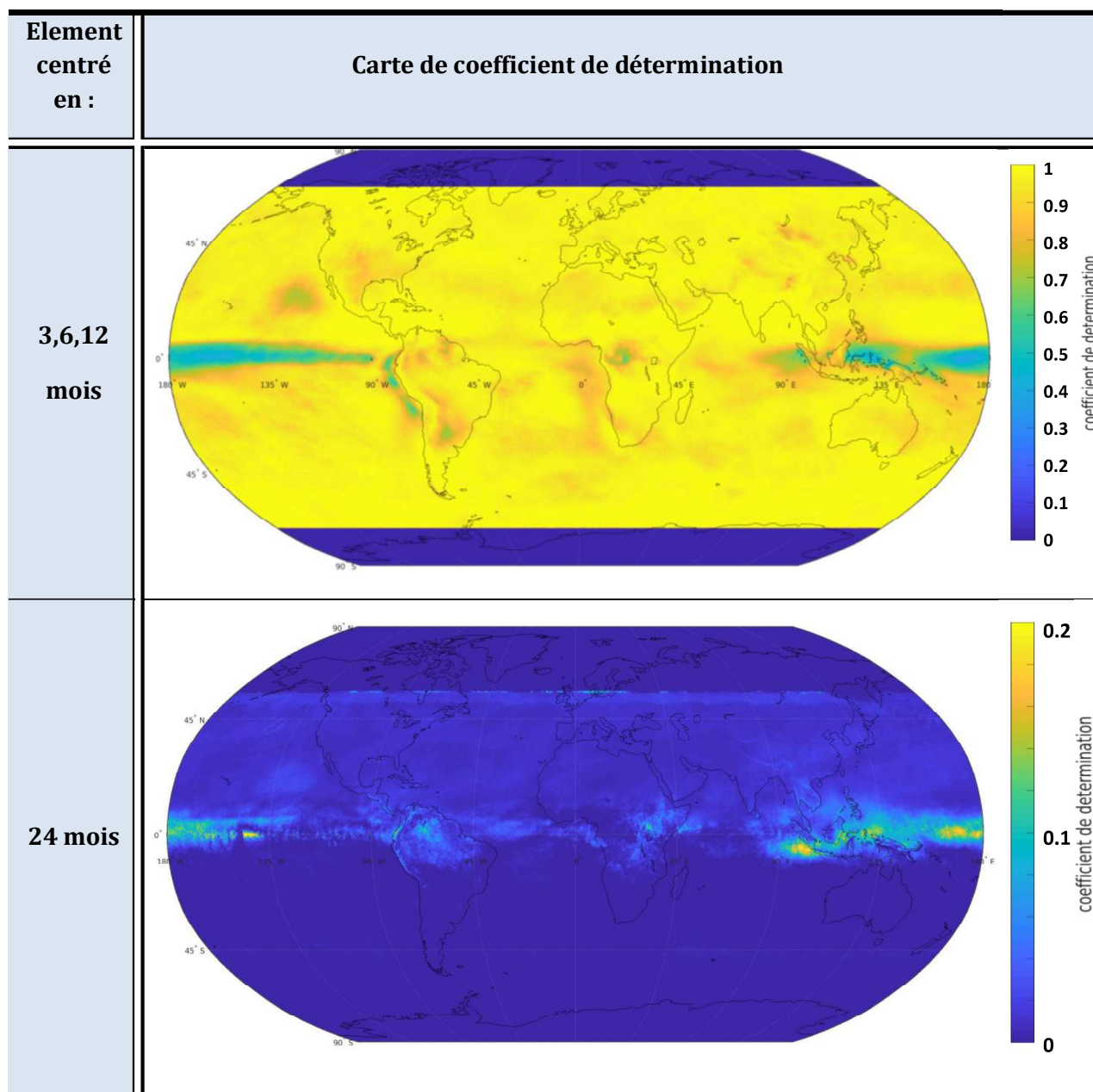
GEBA

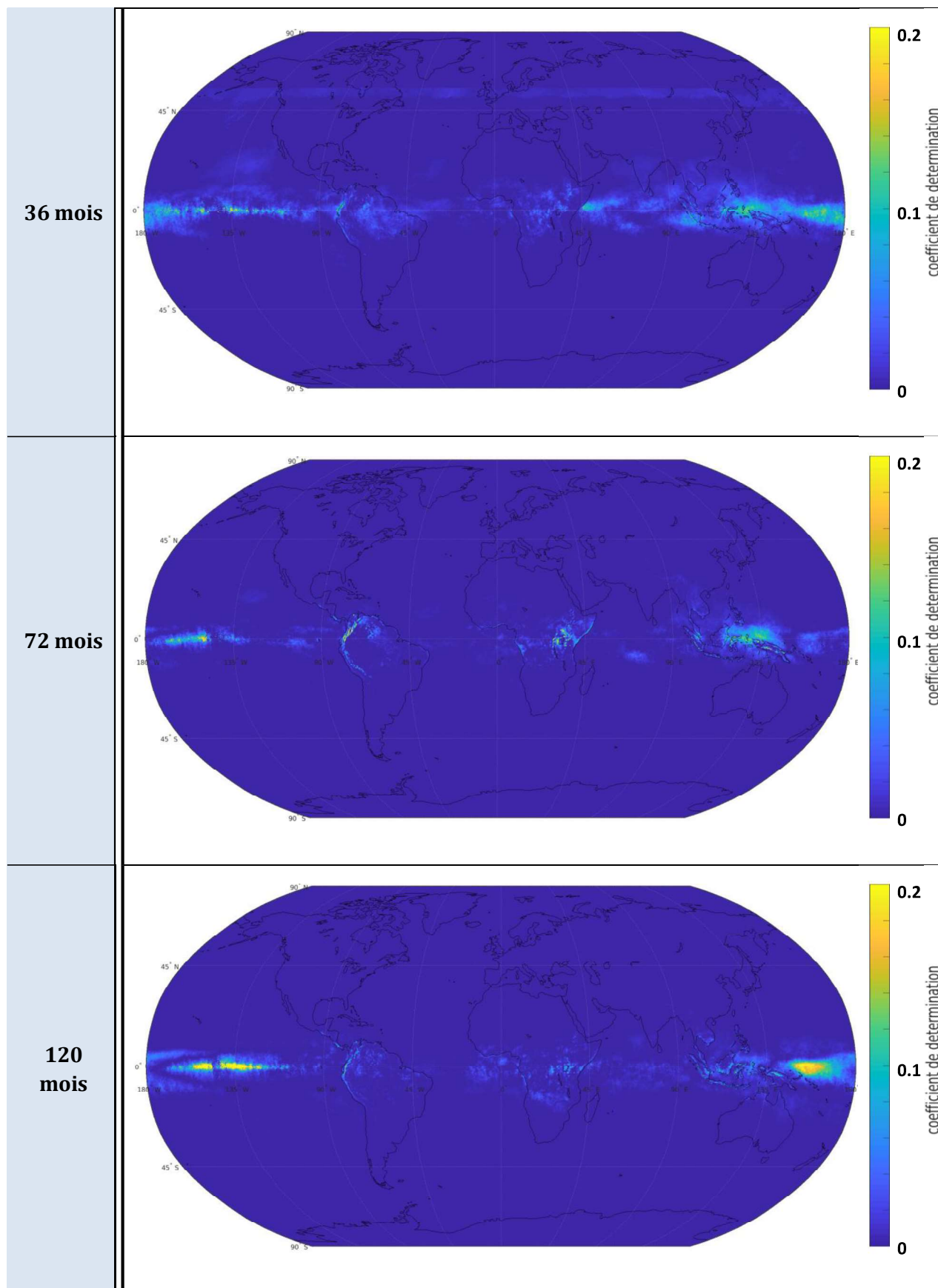


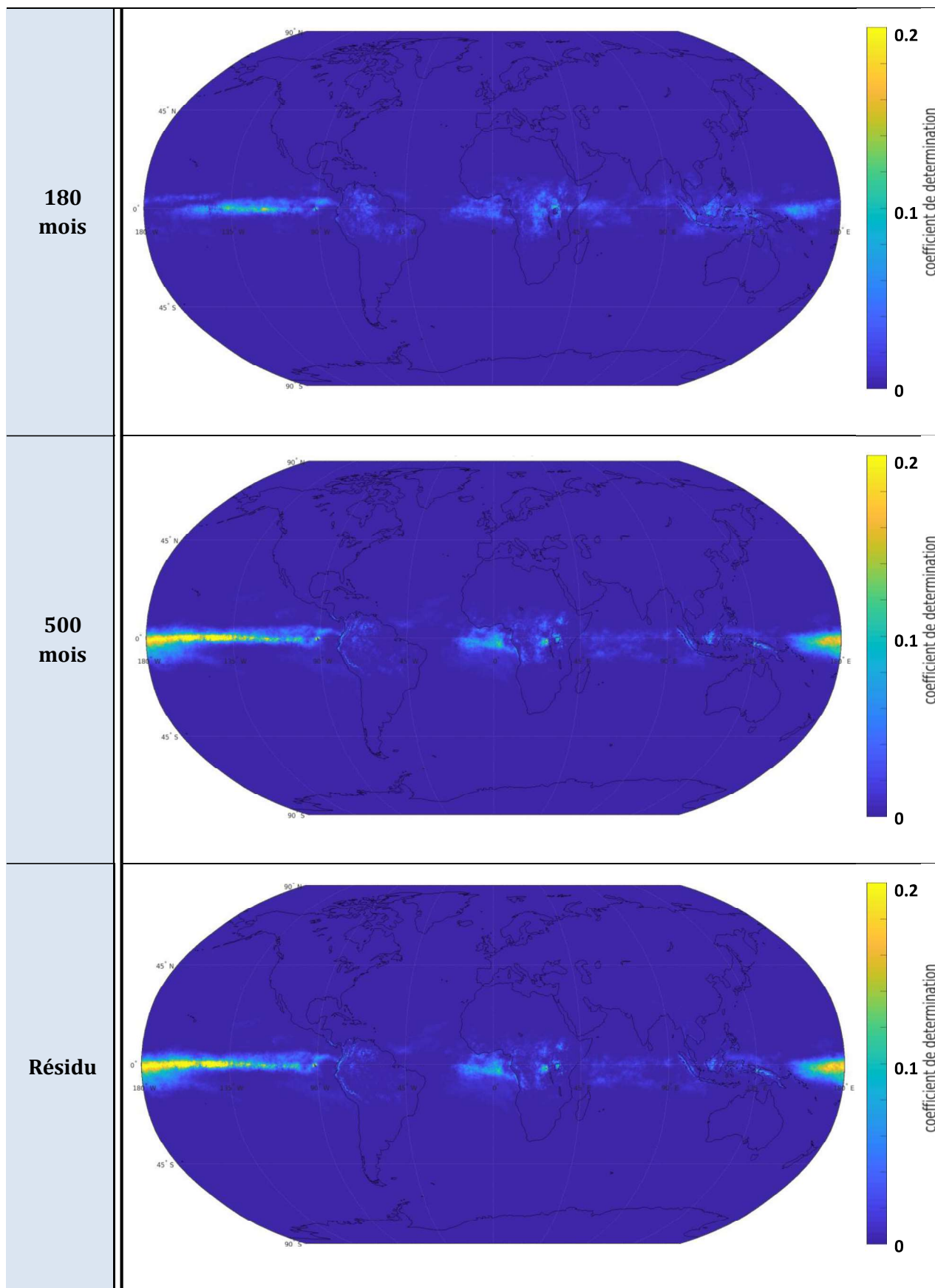




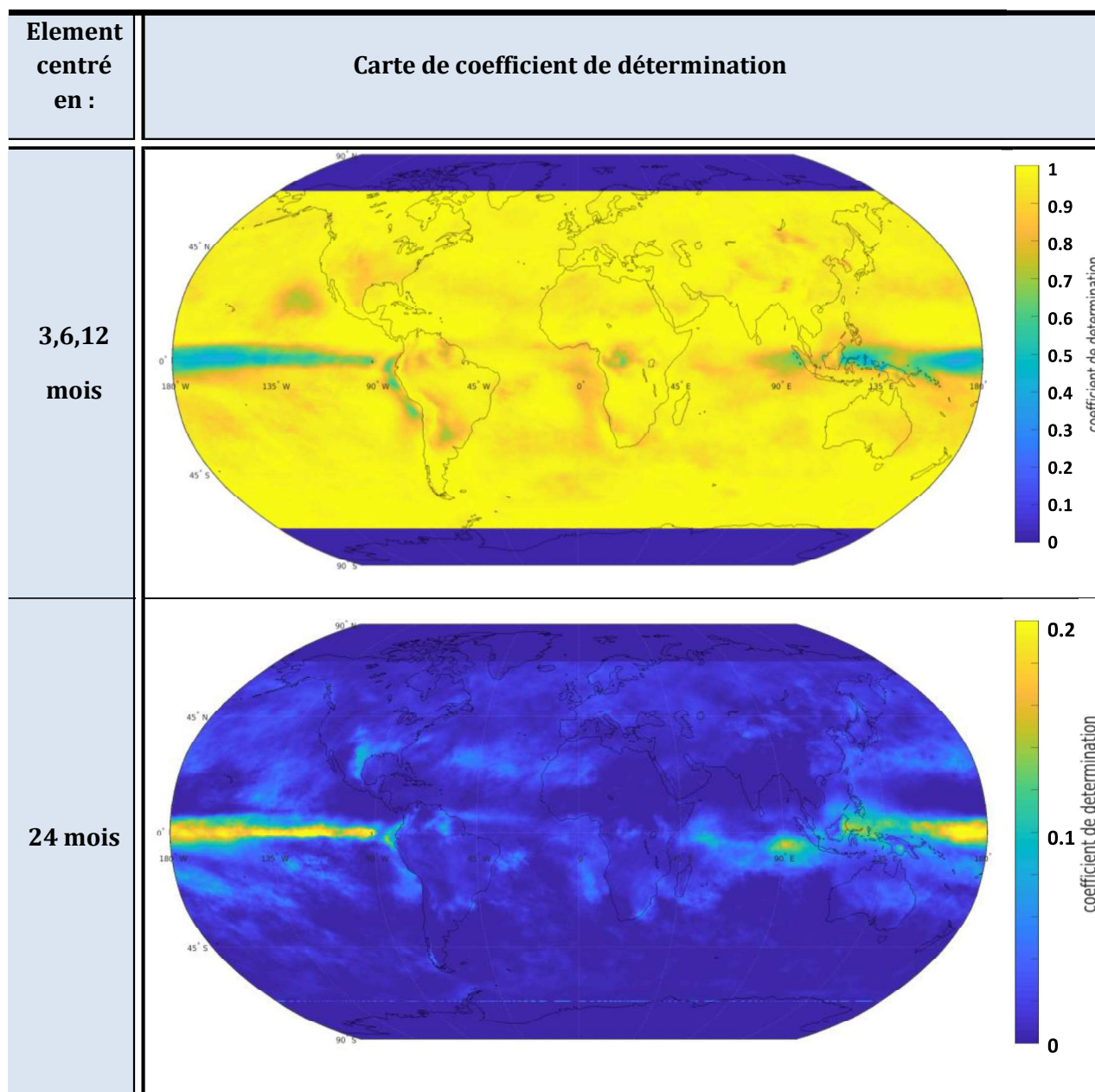
CLARA-A2

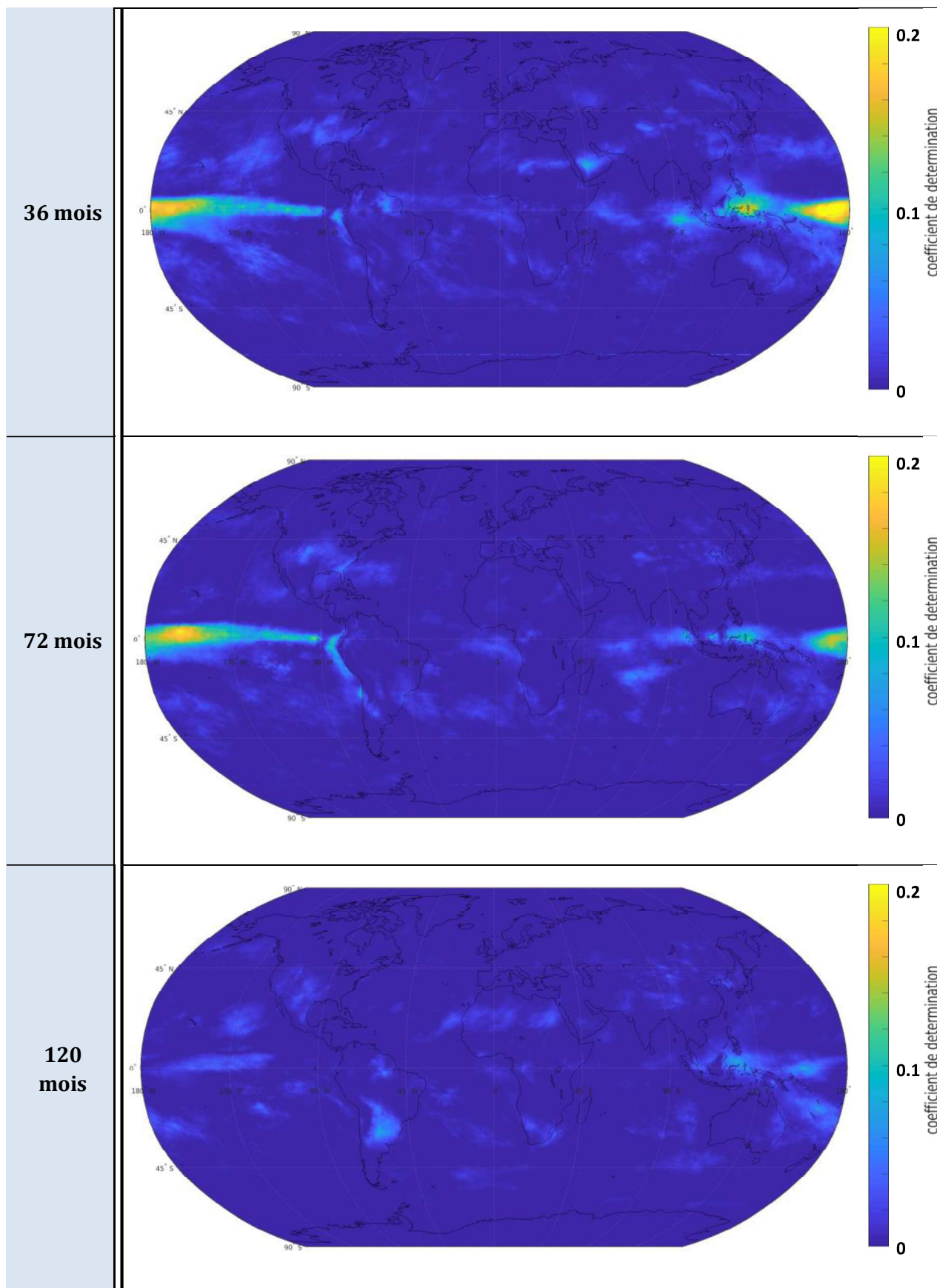


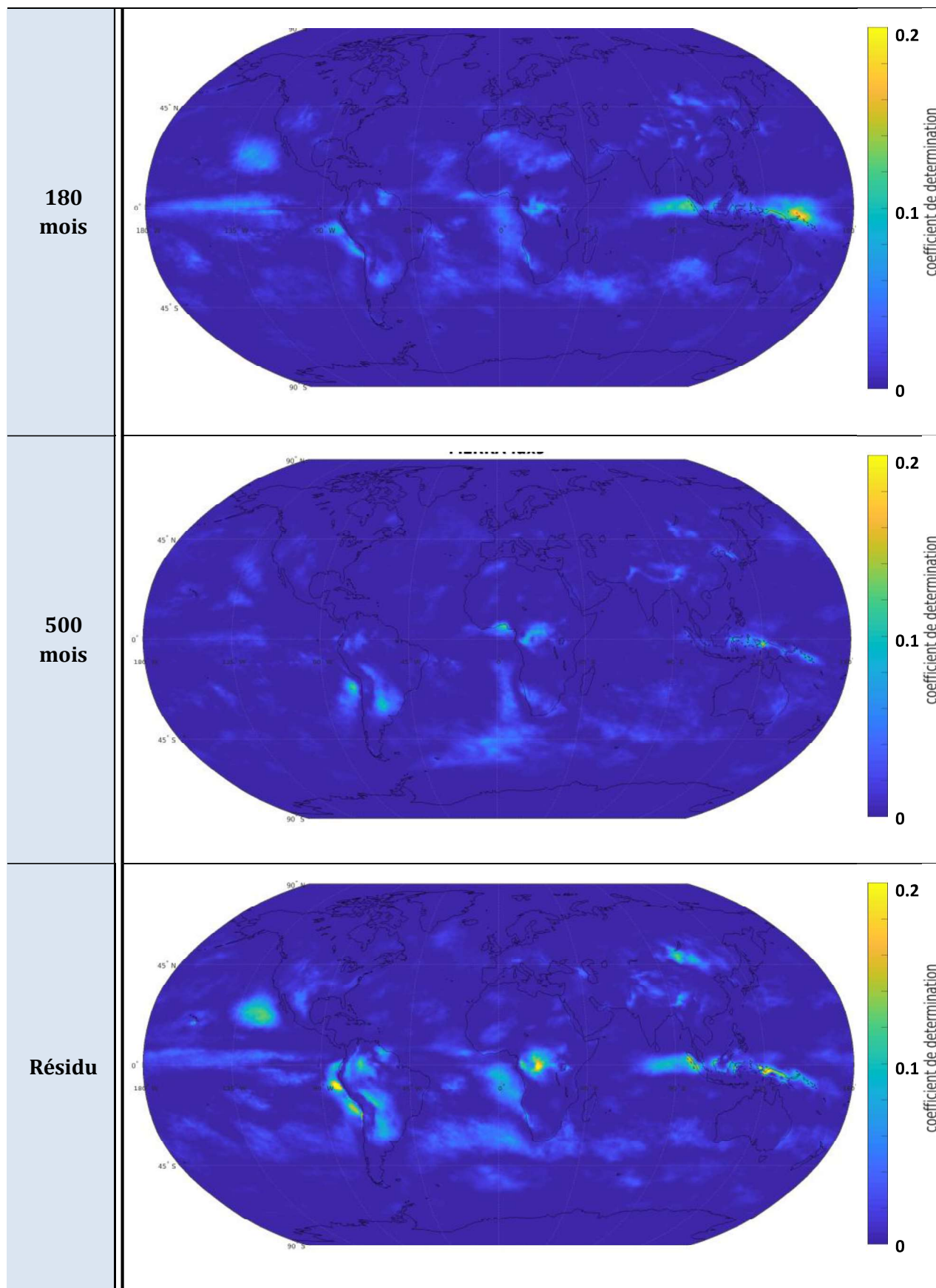




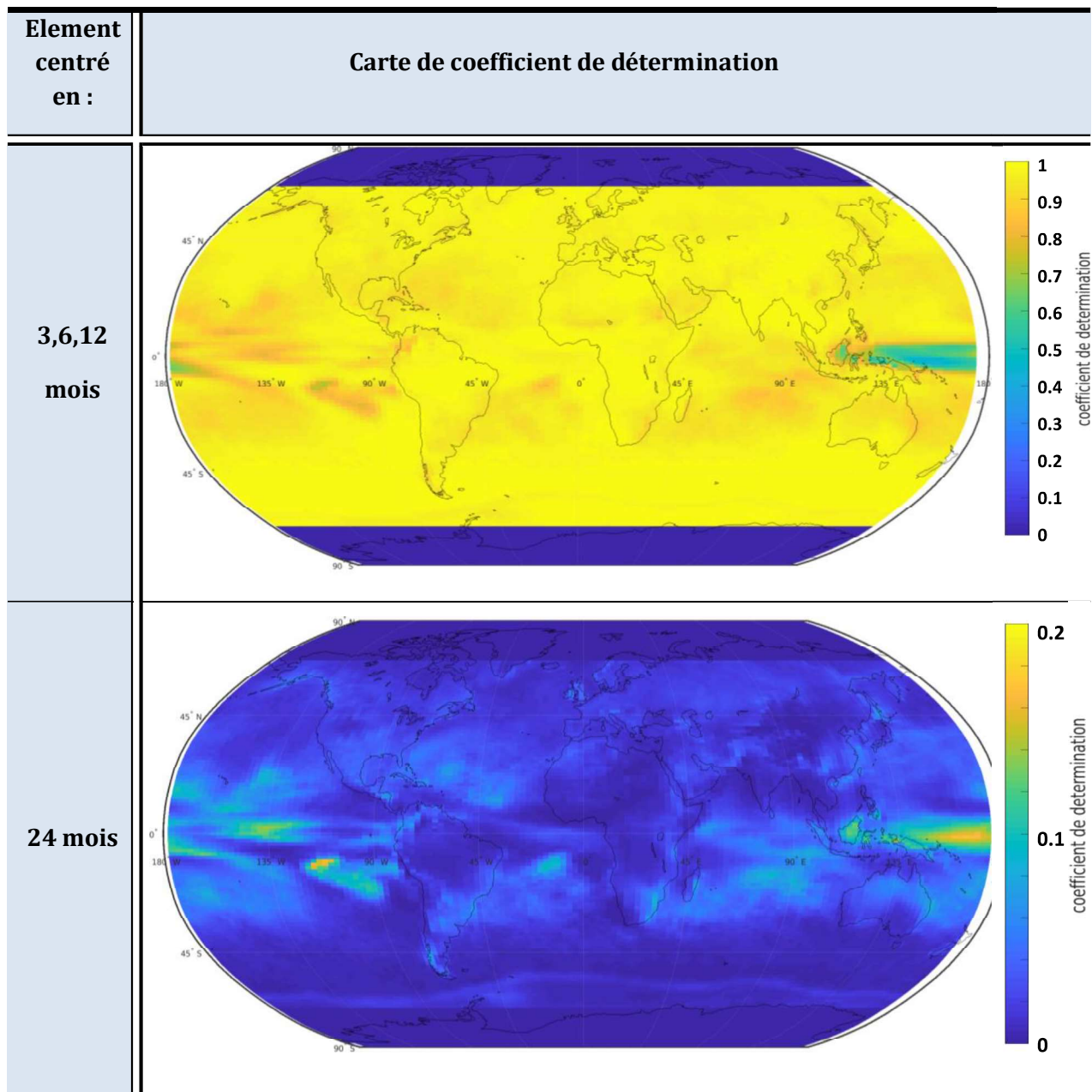
MERRA-2

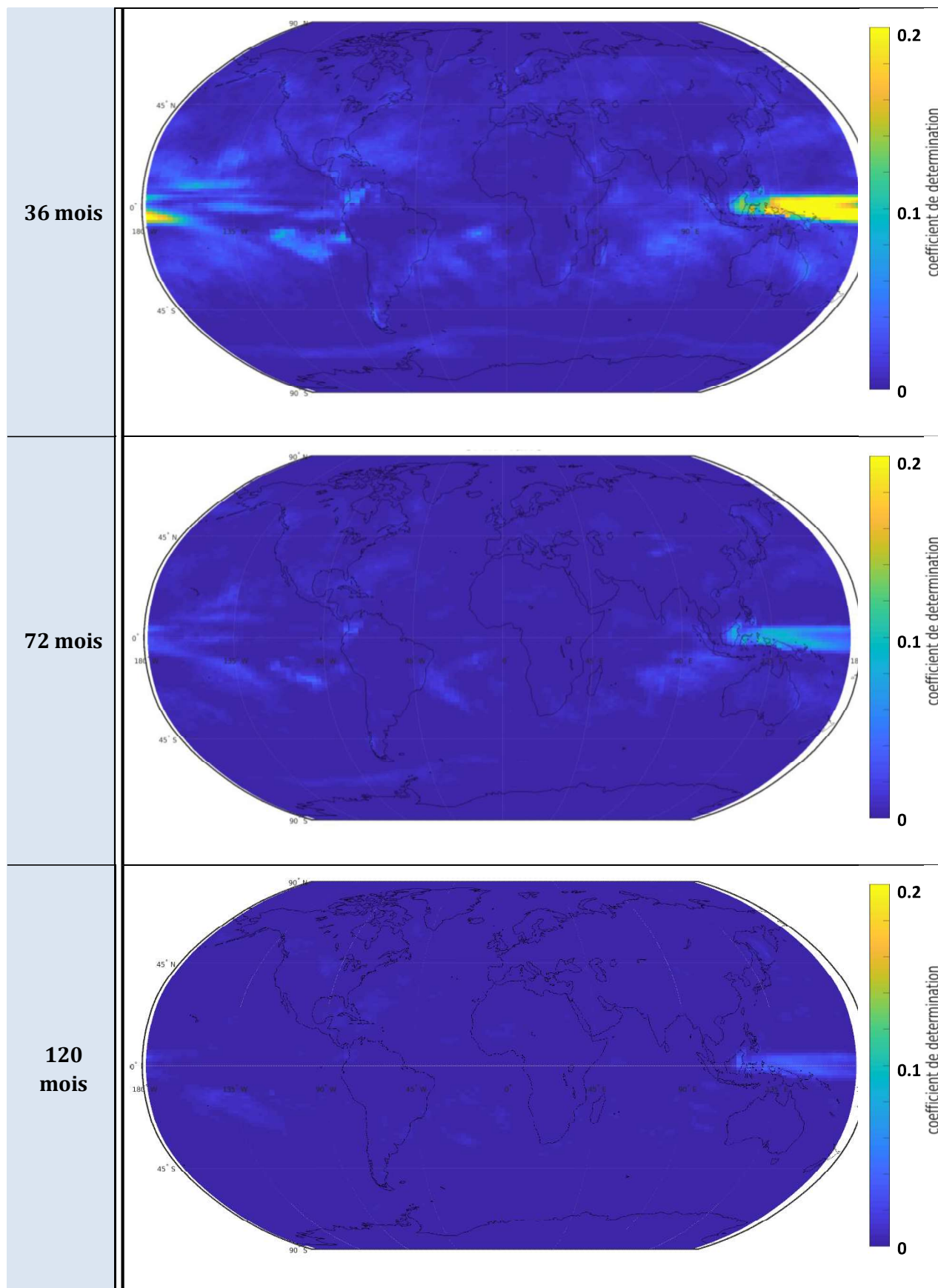


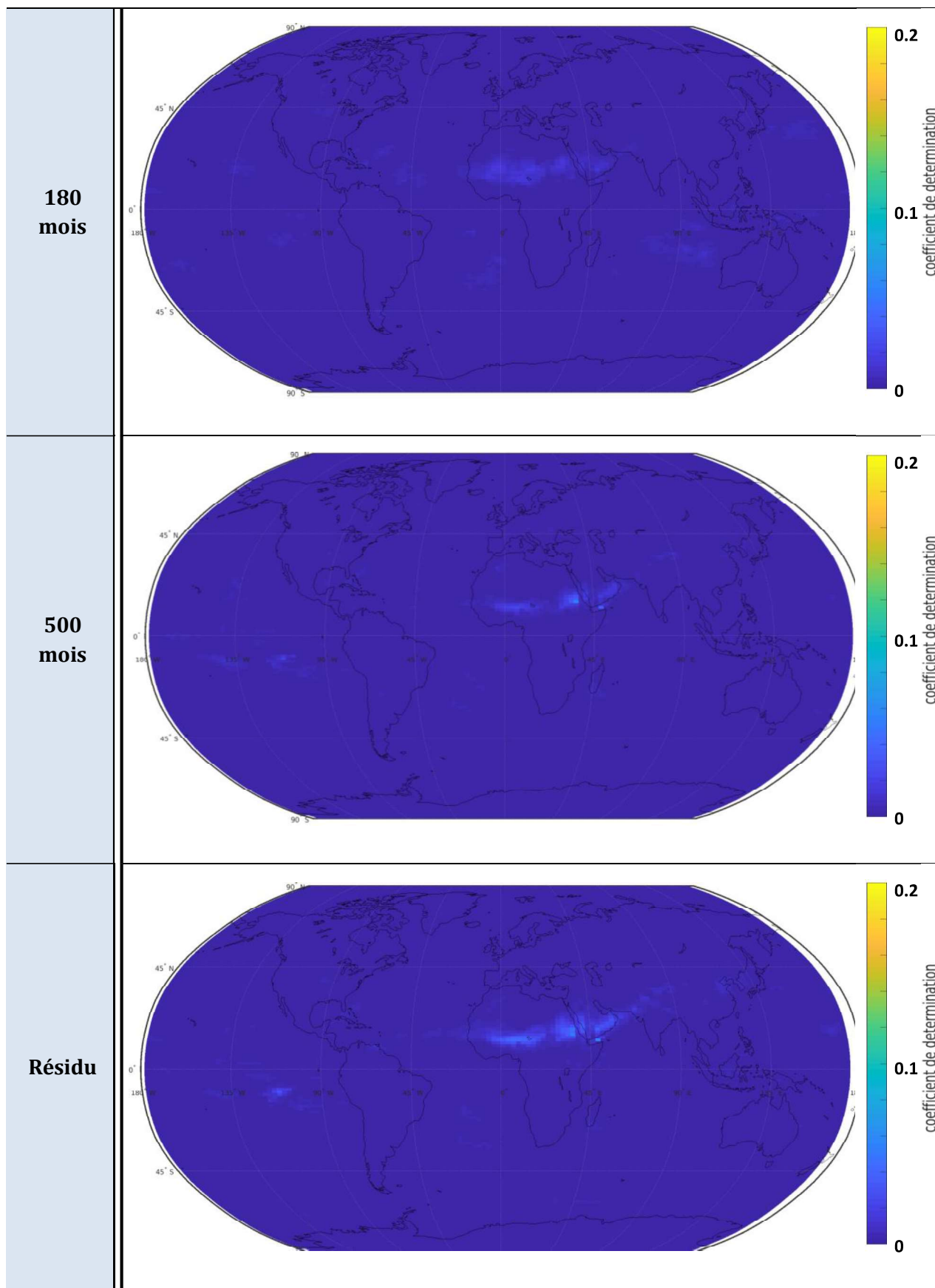




IPSL-CM6A-LR







RÉSUMÉ

Cette thèse se propose de contribuer à la caractérisation des variations de long terme de l'éclairement, dans un contexte d'analyse de risque financier de grandes centrales solaires photovoltaïques. L'utilisation d'indicateurs statistiques (e.g. P90) et de l'hypothèse de stationnarité temporelle de l'éclairement a été questionnée. Cela a mené à une caractérisation fine des variations de long terme de l'éclairement grâce à un outil de décomposition temps-fréquence développé au cours de cette thèse. Nous avons distingué trois classes de variabilité : la variabilité intra-annuelle, la variabilité annuelle à décennale, et la variabilité multi-décennale. Pour la première classe, l'utilisation de quatre ans de données historiques est suffisante pour prendre en compte de manière correcte l'ensemble des variations de l'éclairement. Pour la seconde classe, l'utilisation de 30 années de données historiques est recommandée. Pour la variabilité multi-décennale, l'utilisation de plus de 30 années de données est préconisée.

Les trois classes de variabilité ont été analysées pour des bases de données de natures diverses : mesures de long terme du réseau GEBA, données satellitaires CLARA-A2, données de ré-analyse MERRA-2, et données issues du modèle climatique IPSL-CM6A-LR. Une grande diversité des structures de variabilité en fonction de la base de données considérée a été observée.

MOTS CLÉS

Energie solaire, Rayonnement solaire reçu au sol, Variabilité temporelle, Variabilité climatique, Analyse temps-fréquence, Rayonnement solaire – Mesure, Rayonnement solaire – Satellite, Rayonnement solaire – Ré-analyse, Rayonnement solaire – Modèles climatiques

ABSTRACT

This PhD thesis proposes to contribute to the characterization of long-term variations in irradiance, in a context of financial risk analysis of large solar photovoltaic power plants. The use of statistical indicators (e.g. P90) and the hypothesis of temporal stationarity of the irradiance has been questioned. This led to a fine characterization of the long-term variations of the irradiance thanks to a time-frequency decomposition tool developed during this thesis. We distinguished three classes of variability: intra-annual variability, annual to decadal variability, and multi-decadal variability. For the first class, the use of four years of historical data is sufficient to correctly take into account all the variations in irradiance. For the second class, the use of 30 years of historical data is recommended. For multi-decadal variability, the use of more than 30 years of data is encouraged.

The three classes of variability have been analyzed for various types of databases: long-term measurements from the GEBA network, CLARA-A2 satellite data, MERRA-2 re-analysis data, and data from the IPSL-CM6A-LR climate model. A great diversity of variability structures according to the database considered was observed.

KEYWORDS

Solar energy, Irradiance, Temporal variability, Climate variability, Time-frequency analysis, Solar radiation - Measurement, Solar radiation - Satellite, Solar radiation - Re-analysis, Solar radiation - Climate models