Analyse temps-fréquence des données de rayonnement solaire reçu au sol

Marc Bengulescu

THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à MINES ParisTech

Time–frequency analysis of surface solar radiation data
Analyse temps–fréquence des données de rayonnement solaire
reçu au sol

École doctorale n°432

SCIENCES DES MÉTIERS DE L'INGÉNIEUR

Spécialité ENERGÉTIQUE ET PROCÉDÉS

COMPOSITION DU JURY :

M. Robert VAUTARD
Laboratoire des Sciences du Climat et de
l'Environnement – UMR 8212, Président du jury

M. Hans-Georg BEYER
Université des Îles Féroé, Rapporteur

M. Ted SOUBDHAN
Université des Antilles, Rapporteur

M. Milan PALUŠ
Académie Tchèque des Sciences, Examinateur

Mme Stéphanie DUBOST
Électricité de France R&D, Examinatrice

M. François G. SCHMITT
Laboratoire d'Océanologie et de Géosciences –
UMR 8187, Examinateur

M. Philippe BLANC
MINES ParisTech, Examinateur

M. Lucien WALD
MINES ParisTech, Examinateur

Soutenue par **Marc BENGULESCU**
le 12 juillet 2017

Dirigée par **Philippe BLANC**
Lucien WALD

MINES ParisTech

Centre Observation, Impacts, Energy

1 Rue Claude Daunesse, CS 10207

06904 Sophia Antipolis, France

This thesis was prepared as part of a PhD programme that the author has completed at Centre Observation, Impacts, Energy under a MINES ParisTech doctoral contract.

Le véritable voyage de découverte ne consiste pas à chercher de nouveaux paysages, mais à avoir de nouveaux yeux.

The real voyage of discovery consists not in seeking new landscapes, but in having new eyes.

Marcel Proust

Acknowledgements

I would like to express my sincere gratitude to all of those who made this thesis possible. I am indebted to my advisers Lucien Wald and Philippe Blanc for their continuous support of my PhD study and related research. I am also grateful to all my colleagues from Centre Observation, Impacts, Energy at MINES ParisTech in Sophia-Antipolis for the pleasurable moments that we have shared during the years that I have spent among them.

I would also like to thank all the members of my PhD examination committee for making my defence to be an enjoyable moment, and for their comments and suggestions.

Finally, I would also like to acknowledge my friends and family for their understanding and unending patience and for providing me with support and consideration when I most required them.

Résumé étendu du manuscrit

Chapitre 1 : Introduction

Cette thèse traite de l'éclairement solaire descendant reçu au sol, abrégé en SSI (*surface solar irradiance*), qui est la puissance du rayonnement provenant du Soleil reçue par unité de surface horizontale, sur l'ensemble du spectre. L'éclairement s'exprime en $\mathrm{W\,m^{-2}}$. Le SSI présente des variations importantes tant temporellement que spatialement. Certaines variations ont des échelles allant de la décennie au millénaire, voire plus. Hors atmosphère, ces variations peuvent avoir comme origine, la variabilité des émissions du Soleil, ou les modifications des paramètres orbitaux de la Terre. La révolution de la Terre autour du Soleil et l'obliquité de l'écliptique donnent lieu aux saisons astronomiques. La rotation de la Terre sur elle-même donne un cycle journalier. Les variations dans le temps et l'espace des constituants de l'atmosphère ayant un effet sur l'éclairement, comme par exemple les molécules d'air, les aérosols et les nuages, entraînent des variations du SSI à différentes échelles, de même que les effets dus aux ombrages, qu'ils soient d'origine orographique ou autre (végétation, bâtiments, etc.).

Cette thèse se concentre sur la variabilité temporelle du SSI et sur l'analyse de cette dernière. Plusieurs auteurs ont proposé des méthodes d'analyse du SSI, souvent limitées à un intervalle d'échelles temporelles, comme par exemple les changement décennaux [Trenberth et al., 2009 ; Abdel Wahab et al., 2010 ; Pachauri et al., 2014 ; Blanc et al., 2015], ou à très court terme [Yordanov et al., 2013 ; Lauret et al., 2016], ou encore à des échelles intermédiaires [Coskun et al., 2011 ; Medvigy and Beaulieu, 2012]. Est-il possible d'analyser la variabilité du SSI sur un large ensemble d'échelles temporelles à l'aide d'une seule méthode ? Il est préférable que cette méthode effectue une décomposition de la série temporelle du SSI en une série de composantes non corrélées présentant des échelles de temps distinctes. Il est également souhaitable qu'à chaque instant de la série temporelle, soient connues les composantes fréquentielles constituant le SSI. Les méthodes

dites *temps-fréquences* répondent bien à ces attentes [Cohen, 1989]. La non-linéarité et la non-stationnarité du SSI doivent être prises en compte [Zeng et al., 2013 ; Calif et al., 2016]. C'est ainsi que cette étude repose sur l'usage de la transformée de Hilbert-Huang [Huang et al., 1998], abrégée en HHT (*Hilbert-Huang transform*) qui présente notamment comme avantage d'être adaptative.

En effet, la HHT décompose une série temporelle en composantes déterminées entièrement par les propriétés locales (au sens temporel) de la série, sans présupposé d'un dictionnaire de fonctions élémentaires. Ces composantes sont appelées des modes intrinsèques de variation, ou fonctions modales intrinsèques, désignés ci-après par l'acronyme anglais IMF (*intrinsic mode functions*). Une représentation temps-fréquence de la série est construite à l'aide de ces IMFs. La HHT est et a été largement utilisée dans le domaine des géosciences [Huang and Wu, 2008 ; Huang et al., 2014 ; Tary et al., 2014], incluant le SSI [Duffy, 2004 ; Calif et al., 2013 ; Bengulescu et al., 2016a ; Bengulescu et al., 2016b ; Bengulescu et al., 2016c ; Calif et al., 2016 ; Bengulescu et al., 2017].

Quelle que soit la technique d'analyse utilisée, se pose toujours la question de la discrimination entre les composantes déterministes et les composantes de bruit définies comme les réalisations de processus stochastiques. Le moyen usuel pour la HHT est de supposer un modèle de bruit auquel sont comparés les IMFs [Huang and Wu, 2008 ; Franzke, 2009 ; Franzke, 2012]. Dans cette étude, une approche innovante a été adoptée d'après les travaux de Chen et al. [2013]. Elle ne suppose aucune connaissance *a priori* du bruit et teste l'hypothèse nulle de manière adaptative. L'usage combiné de la HHT et de cette approche permet une analyse fine du signal et la discrimination des IMFs déterministes du SSI et des IMFs présentant un aspect stochastique.

Les outils développés permettent d'extraire et d'analyser les échelles caractéristiques (les IMFs) présentes dans de longues séries temporelles de SSI. Cette étude montrera qu'il est possible d'identifier le type de climat par exemple, ou bien de relever que toutes les séries pour les stations étudiées présentent des composantes stochastiques de périodes inférieures à l'année et modulées par le cycle annuel par des interactions inter-échelles non-linéaires. Ces résultats ont été décrits par [Bengulescu et al., 2016a ; Bengulescu et al., 2016b ; Bengulescu et al., 2016c].

La variabilité du SSI est étudiée en analysant des séries temporelles du SSI mesurées par des pyranomètres de différentes stations de mesure dans le monde. Le choix s'est porté notamment sur des stations appartenant au réseau BSRN (*Baseline Surface Radiation Network*), dont les mesures sont de très bonne qualité [Ohmura et al., 1998]. Toutefois, il y a très peu de stations et les séries temporelles peuvent présenter des manques allant jusqu'à 13 % des mesures [Roesch et al., 2011]. En pratique, la connaissance du SSI est nécessaire partout dans le monde et des substituts à ces mesures ont été développés, comme les estimations du SSI faites à partir d'images issues de satellites météorologiques

géostationnaires [Lefèvre et al., 2014] ou par des modèles numériques de ré-analyse météorologique [Richardson and Andrews, 2014]. Un chapitre est dédié à la présentation de l'ensemble des mesures et estimations du SSI exploités dans cette thèse.

En appliquant les outils d'analyse à l'ensemble des données, une nouvelle approche a été proposée pour l'évaluation des capacités des estimations, par satellite et modèle numérique, à reproduire les mesures faites dans le réseau BSRN. Contrairement aux travaux de validation usuels, la comparaison est effectuée pour chaque échelle temporelle caractéristique. Il a notamment été trouvé que les modèles numériques offrent une qualité bien moindre que celle des estimations par satellite [Bengulescu et al., 2017], en parfait accord avec des études précédentes [Lohmann et al., 2006 ; Boilley and Wald, 2015] mais avec l'avantage qu'ici, les défauts et qualités sont identifiés échelle par échelle.

L'organisation du manuscrit est comme suit. Le chapitre 2 décrit les séries du SSI qui ont été analysées dans ce travail. Le chapitre 3 introduit la HHT de même qu'une brève revue d'autres méthodes temps-fréquence comme l'analyse en ondelettes et compare les avantages et inconvénients des unes et des autres pour le problème posé. La mise en œuvre de la HHT y est détaillée. Le chapitre 4 présente une analyse de différentes séries temporelles du SSI mesurées par différentes stations de mesure dans le monde. On y décrit les IMFs observés selon les stations, les similitudes et différences entre les IMFs des stations, et un clair contraste entre des IMFs ayant des origines déterministes et celle ayant une nature plus stochastique. Le chapitre 5 porte sur l'évaluation de la capacité des approches alternatives aux stations pyranométriques à reproduire la variabilité du SSI, et aborde l'influence des constituants de l'atmosphère dans la variabilité observée du SSI. Les conclusions de ce travail et les perspectives sont présentées dans le chapitre 6. Les références bibliographiques suivent puis l'annexe A qui reproduit les quatre articles publiés en accès libre, et l'annexe B donnant des graphiques supplémentaires relatifs au chapitre 5.

Chapitre 2 : Les jeux de données sur le rayonnement solaire

L'objet de ce chapitre est de décrire les différents jeux de données exploités dans cette thèse. Ces données jouent un rôle primordial dans cette thèse. Il y a deux types de jeux de données. Le premier est constitué de mesures du SSI faites par des pyranomètres dans des stations du réseau BSRN. Le deuxième type de jeu est constitué de mesures du SSI par des stations du WRDC (*World Radiation Data Center*) et d'estimations correspondantes effectuées au moyen d'images de satellite ou de modèles numériques de ré-analyse météorologique.

La thèse se limite au SSI journalier, c'est-à-dire à l'éclairement solaire global descendant reçu sur une surface horizontale (GHI, *Global Horizontal Irradiance*) en moyenne journalière.

Le réseau BSRN fournit des mesures de très bonne qualité du SSI pour plusieurs stations dans le monde [Ohmura et al., 1998]. Plusieurs stations offrent des séries temporelles plutôt complètes et couvrant au moins dix ans, tout en étant situées dans des climats différents. Ces caractéristiques en font des mesures du SSI idéales pour l'objectif de la thèse, notamment pour les chapitres 3 et 4.

Pour le premier jeu de données, dit BSRN, quatre stations de ce réseau ont été retenues : Boulder (Colorado, Etats-Unis), Carpentras (France), Payerne (Suisse) et Tateno (Japon). Leurs caractéristiques sont données dans le tableau 3.1 du Chapitre 2. Boulder (abrégé en BOU dans le texte) correspond à un climat Bsk dans la nomenclature de Köppen-Geiger, c'est-à-dire un climat frais de steppe de moyenne latitude. Le climat à Carpentras (CAR) est de type Csa sub-tropical méditerranéen. Celui de Payerne (PAY) est de type Cfb modéré de côte ouest maritime, et celui de Tateno (TAT) est de type Cfa sub-tropical humide de côte est.

Pour ces stations, dix ans de données ont été sélectionnées, de 2001 à 2010. Le contrôle de qualité recommandé par Roesch et al. [2011] a été appliqué aux mesures originales à 1 min de résolution temporelle. Seules les mesures ayant passé ce contrôle de manière satisfaisante ont été déclarées valides. Puis la moyenne journalière du SSI a été calculée pour les jours pour lesquels au moins 80 % des mesures ont été déclarées valides. Le SSI pour chaque jour manquant entouré de deux SSI journaliers valides a été estimé par interpolation linéaire des indices de clarté KT.

Pour le deuxième jeu, dit WRDC, les mesures de deux autres stations ont été exploitées : Wien Hohe Warte (VIE) en Autriche (48.25° N ; 16.35° E), et Kishinev (KIV) en Moldavie (47.00° N ; 28.82° E). Ces stations font partie du réseau GAW (*Global Atmosphere Watch*) et soumettent leurs mesures horaires au WRDC (*World Radiation Data Centre*) qui les distribue sous forme horaire et journalière après un contrôle de qualité [Tsvetkov et al., 1995]. Elles offrent des séries temporelles englobant la période de neuf ans du 1er février 2004 au 31 janvier 2013. Ces dates ont été choisies pour correspondre à d'autres séries temporelles discutées ci-après.

Pour mieux discerner les sources de variabilités, le jeu de données WRDC comprend les séries temporelles d'éclairement solaire journalier hors atmosphère et à la surface dans des conditions de ciel sans nuage — ciel clair — correspondantes aux lieux des stations VIE et KIV sur la même période de temps.

Les séries temporelles de l'éclairement hors atmosphère ont été générées au moyen de l'algorithme SG2 (*Solar Geometry 2*) [Blanc and Wald, 2012].

Les séries temporelles du SSI par ciel clair ont été générées à l'aide du modèle McClear [Lefèvre et al., 2013]. Le modèle McClear fait partie du service *Copernicus Atmosphere Monitoring Service* (CAMS) et prend en compte des estimations CAMS tri-horaires des propriétés des aérosols et des contenus de la colonne en ozone et vapeur d'eau [Hoyer-Klick et al., 2015].

A ce jeu de données WRDC ont été rajoutés trois autres séries temporelles de SSI journalier pour chacun des lieux des deux stations VIE et KIV. L'une des séries provient d'observations par satellite et les deux autres de modèles numériques de ré-analyses météorologiques.

Le SSI estimé par satellite vient de la base de données HelioClim-3 version 5 (HC3v5) [Blanc et al., 2011]. Cette dernière est obtenue par le traitement des images du satellite Meteosat par la méthode Heliosat-2 décrite par Rigollier et al. [2004] et modifiée par Qu et al. [2014].

Les deux modèles numériques de ré-analyses météorologiques sont les suivants :

- le produit ERA-Interim [ECMWF, 2009] du centre européen de météorologie (ECMWF, *European Center for Medium-range Weather Forecast*). Le SSI (variable « *surface solar radiation downwards* ») a été téléchargé depuis le système MARS (*Meteorological Archival and Retrieval System*) de ECMWF. La résolution spatiale est de 0.75° dans les deux directions et la résolution temporelle de 6 h. Les séries temporelles de SSI journalier pour VIE ou KIV sont obtenues par interpolation bilinéaire ;

- le produit MERRA-2 (*Modern-Era Retrospective analysis for Research and Applications*, Version 2) de l'agence spatiale NASA des Etats-Unis. Le SSI (variable *M2T1NXRAD*) est estimé suivant une résolution spatiale de 0.5° en latitude et de 0.65° en longitude et une résolution temporelle de 1 h.

Chapitre 3 : Une introduction à la transformée de Hilbert-Huang (HHT)

Ce chapitre est une introduction à la transformée de Hilbert-Huang. Comme vu précédemment, la méthode d'analyse de la variabilité temporelle du SSI doit être de type *temps-fréquence* [Cohen, 1989] afin d'effectuer une décomposition de la série temporelle en un ensemble de composantes non corrélées présentant des échelles de temps distinctes, en identifiant en chaque instant les fréquences instantanées correspondantes. Dans le cadre de la thèse, le choix de la méthode temps-fréquence s'est porté sur la transformée de Hilbert-Huang (HHT) [Huang et al., 1998].

La HHT est un algorithme qui décompose de manière adaptative une série temporelle en composantes déterminées entièrement par les propriétés locales de la série : la méthode ne

nécessite pas de dictionnaire *a priori* de fonctions élémentaires. Les composantes issues de la décomposition sont appelées des modes intrinsèques de variation, ou fonctions modales intrinsèques (IMF). Une représentation temps-fréquence de la série est construite à l'aide de ces IMFs. Bien qu'on puisse trouver de nombreuses publications, la HHT est un algorithme assez récent et mérite une introduction mathématique assez détaillée qui est le principal objet de ce chapitre.

A cette présentation est ajoutée une mise en perspective de la HHT avec deux méthodes d'analyse de la variabilité plus éprouvées et plus classiques : la transformée de Fourier discrète à court terme et les transformées en ondelettes. Il est ainsi montré que la HHT est tout à fait appropriée à l'étude.

La transformée de Fourier discrète à court terme

La transformée de Fourier discrète (DFT, *discrete Fourier transform*) est une transformée bien connue depuis les travaux de Cooley and Tukey [1965] proposant un moyen rapide de calcul numérique, la FFT (*fast Fourier transform*).

Telle quelle, la DFT ne permet pas une analyse temps-fréquence : Gabor [1946] a proposé une adaptation par fenêtre d'apodisation glissante donnant lieu à la transformée de Fourier à court terme (STFT, *short-term Fourier transform*) [Tary et al., 2014] :

$$X(\omega, \tau) = \int_{-\infty}^{\infty} x(t) w(t - \tau) e^{-j\omega t} \, \mathrm{d}t, \qquad \omega, \tau \in \mathbb{R}.$$

Bien que permettant une analyse temps-fréquence suivant un pavage rectangulaire constant, cette approche présente des limitations notamment liées au phénomène de « fuite spectrale » déjà présent dans la DFT ainsi qu'aux résolutions temporelle et spectrale de la STFT limitées par la taille de la fenêtre d'analyse définie *a priori* et non nécessairement adaptée au signal analysé.

La transformée en ondelette continue

La DFT ainsi que la STFT ne permettent pas d'analyser précisément des signaux non-stationnaires, notamment fréquemment observés en géosciences. La transformée en ondelette continue (CWT, *continuous wavelet transform*) a été introduite pour répondre à ce besoin par Grossmann et Morlet [1984]. Ces dernier définissent la CWT comme une décomposition du signal sur une base d'ondelettes, chaque ondelette étant obtenue par dilatation et translation de la même ondelette-mère.

$$\psi_{\lambda, \tau}(t) = \frac{1}{\sqrt{\lambda}} \, \psi\left(\frac{t - \tau}{\lambda}\right) \qquad \int_{-\infty}^{\infty} \psi(t) \, \mathrm{d}t = 0$$

La transformée en ondelette offre une représentation temps-fréquence, c'est-à-dire que l'on connaît à chaque instant les fréquences qui composent le signal à une incertitude près :

$$\mathcal{W}_x(\lambda, \tau) = \int_{-\infty}^{\infty} x(t)\, \psi_{\lambda,\tau}^*(t)\, \mathrm{d}t$$

Cela représente une amélioration par rapport à la STFT. Les signaux non-stationnaires peuvent être plus finement analysés. La linéarité de la transformée en ondelette reste cependant un problème pour l'analyse de signaux générés par des processus physiques non-linéaires, comme le SSI [Zeng et al., 2013]. Une grosse limitation dans l'usage de cette transformée reste le choix de l'ondelette-mère. Il n'y a pas une seule ondelette-mère et donc pas une seule transformée en ondelette d'un point de vue pratique. Un changement d'ondelette-mère peut mener à des résultats différents.

La transformée de Hilbert-Huang

Les deux précédentes transformées proposent une décomposition du signal sur une base fixe de fonctions de base, trigonométriques ou ondelettes, choisies *a priori*. Etant donné le caractère non-linéaire et non-stationnaire du SSI [Zeng et al., 2013], une analyse des données adaptative semble plus indiquée [Wu et al., 2011]. Idéalement, la méthode d'analyse temps-fréquence ne doit pas faire d'hypothèse quant aux propriétés du signal à analyser et doit s'adapter à ce dernier : la décomposition de la série en composantes non corrélées présentant des échelles de temps distinctes doit se faire naturellement, guidée par les variations tant globales que locales du SSI. Ceci assure que les composantes extraites ont une réalité physique [Wu et al., 2011]. La transformée de Hilbert-Huang, HHT répond à ces attentes [Huang et al., 1998].

La HHT comprend deux étapes distinctes décrites en détail dans ce chapitre. La première est appelée d*écomposition en modes empiriques*, et abrégée en EMD (*empirical mode decomposition*). Ces modes empiriques ou composantes (IMF) sont déterminés entièrement par les propriétés locales. La seconde étape est l'analyse spectrale de ces IMFs au moyen de l'analyse spectrale d'Hilbert.

La HHT repose sur le fait qu'une fréquence d'oscillation instantanée (locale) ne peut être définie que s'il y a oscillation, caractérisée par une symétrie locale du signal autour d'une moyenne nulle. Une IMF doit satisfaire deux conditions :

- sur l'ensemble de la série de données, le nombre d'extrema et le nombre de passages par zéro doivent différer d'au plus un ;
- en tout point de la série, la somme des enveloppes inférieure et supérieure, définies à l'aide des minima et maxima locaux, doit être nulle.

L'EMD est un processus itératif et extrait les IMFs au fur et à mesure, ainsi qu'un résidu qui ne peut pas être assimilé à une oscillation compte-tenu de la durée du signal

analysé. Ce résidu peut être interprété comme le résultat d'un filtre passe-bas appliqué à la série et donne la tendance [Moghtaderi et al., 2013]. Le signal original est la somme exacte des IMFs et du résidu.

Exemple d'analyse d'une série de SSI par la CWT et par la HHT

Cette section est une illustration de ce qui précède, et effectue pas à pas la HHT d'une série décennale de SSI journalier mesurée par la station du réseau BSRN à Carpentras (CAR). Elle présente également le résultat d'une analyse faite à l'aide de la CWT. La comparaison des deux analyses permet de mettre en évidence leurs traits communs et différences.
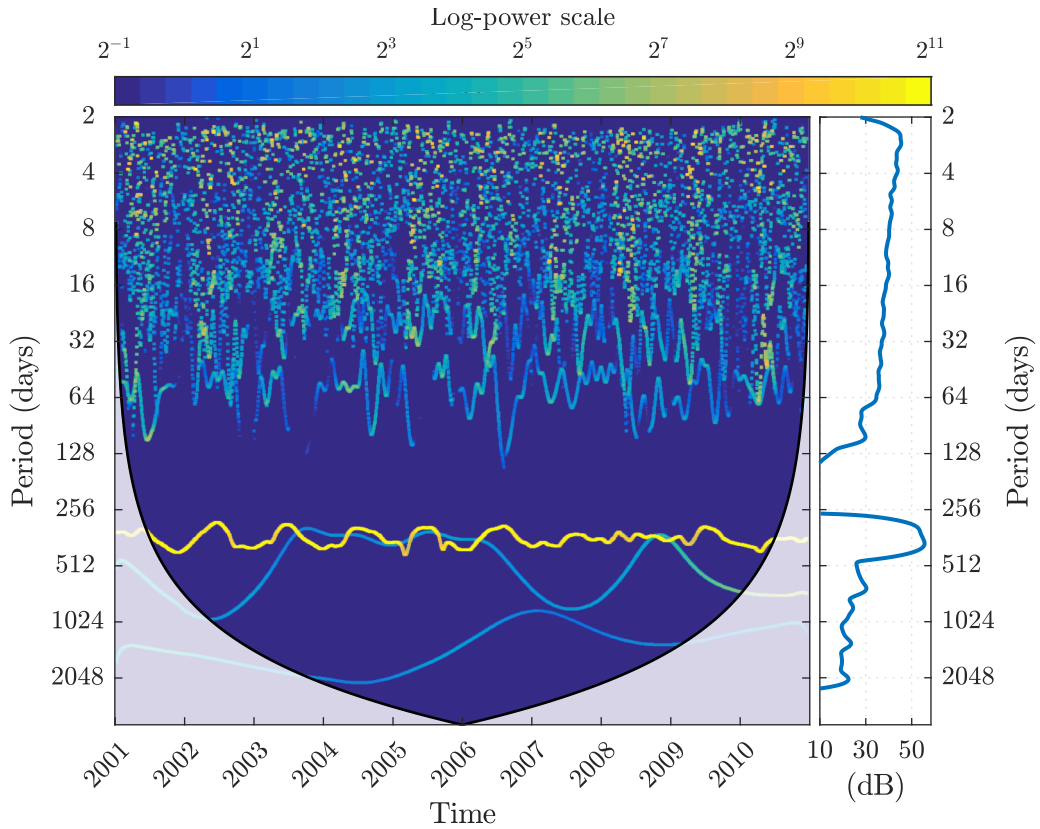
Figure 1 Spectre de Hilbert des dix ans de SSI à CAR. La couleur correspond à la puissance observée à chaque instant (abscisse) pour chaque échelle (ordonnée). Le panneau à droite est le spectre marginal de Hilbert et donne la puissance pour chaque échelle.

Ces résultats sont extraits de [Bengulescu et al., 2016a]. On trouve que les modes de variation du SSI sont bien mis en évidence par la CWT et la HHT. La HHT permet cependant une décomposition visiblement plus parcimonieuse et de mieux mettre en évidence certaines caractéristiques, comme la modulation annuelle. La HHT révèle que l'IMF la plus énergétique correspond à un mode de variabilité quasi-annuel. On observe également dans la représentation spectrale de Hilbert, un manque d'énergie entre 2-3 mois
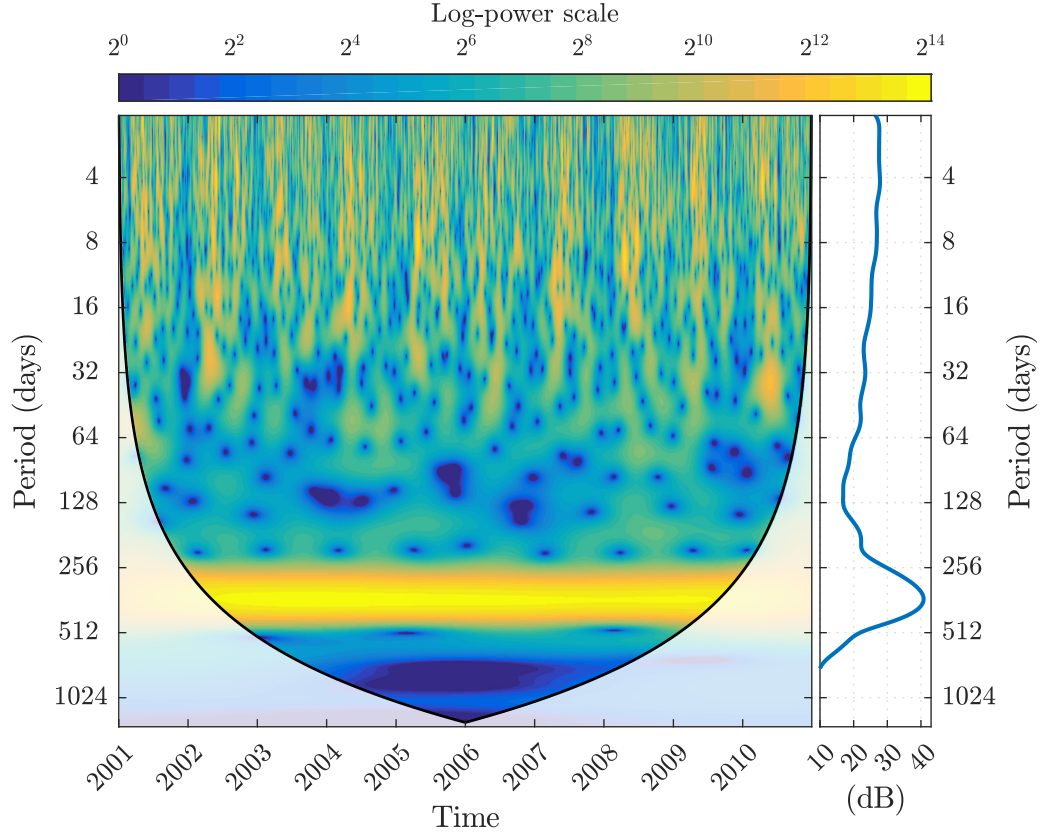
Figure 2 Spectre obtenu par ondelette des dix ans de SSI à CAR. La couleur correspond à la puissance observée à chaque instant (abscisse) pour chaque échelle (ordonnée). Le panneau à droite est le spectre marginal et donne la puissance pour chaque échelle.

et une année. Enfin, un autre aspect dans cette représentation est une sorte de plateau entre 2 jours et 2-3 mois, intervalle temporel dans lequel aucune échelle ne se détache.

Chapitre 4 : Variabilité caractéristique dans les mesures du SSI journalier

L'objet de ce chapitre est d'utiliser la HHT comme analyse temps-fréquence pour mettre en évidence et analyser les échelles temporelles caractéristiques de variabilité dans les séries temporelles de SSI journalier sur dix ans, du jeu de données BSRN comprenant les quatre stations BOU, CAR, PAY et TAT.

Dans une première partie, les distributions des périodes instantanées des différentes IMFs résultant de la HHT pour les quatre stations ont été extraites puis analysées.

La HHT ne révèle pas le même nombre d'IMFs pour chacune de ces stations avec 8 IMFs pour BOU et CAR, 9 pour PAY et 10 pour TAT. On note que les distributions des périodes instantanées des IMFs 1 à 5 sont très similaires pour les quatre stations avec des valeurs médianes suivant une suite dyadique : 3,5 jours, 7 jours, 14 jours, 28
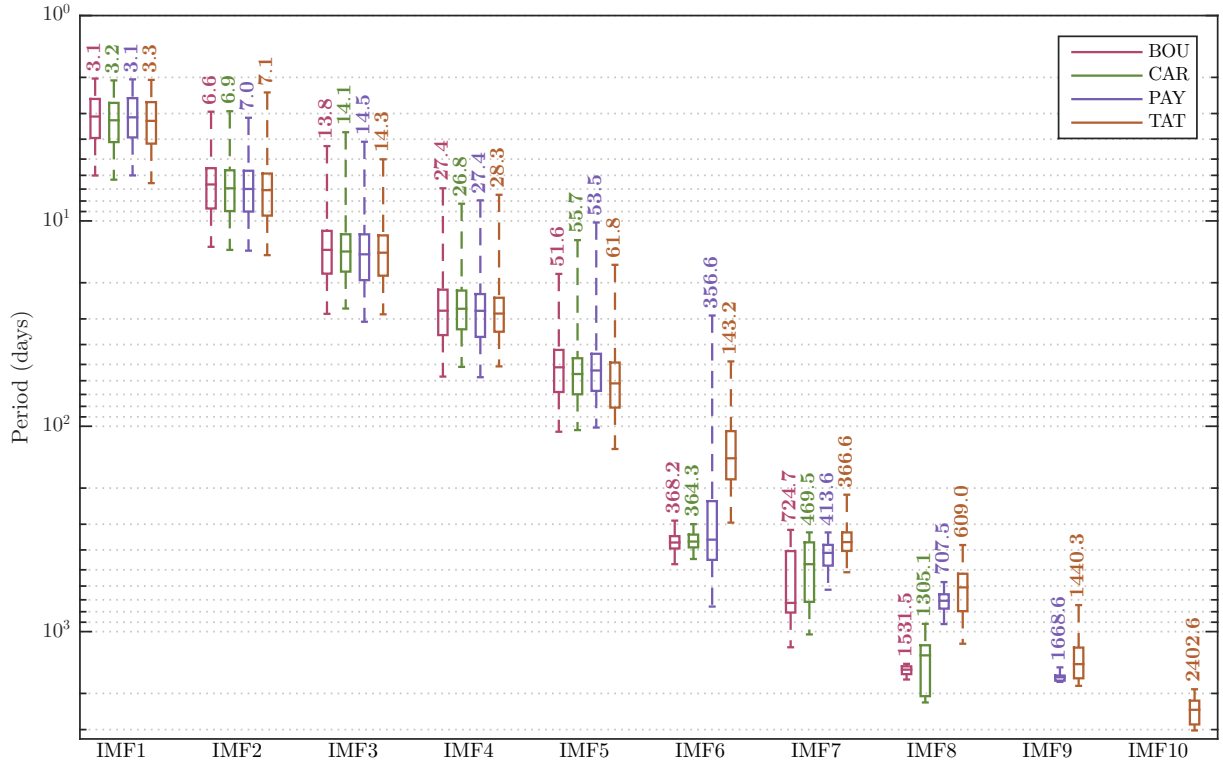
Figure 3 Boîtes à moustache des périodes instantanées pour chaque IMF de chaque station.

jours et 56 jours. Ces périodes dyadiques ont notamment été révélées par Lee et al. [2015] sur des séries temporelles de mesures par satellite du rayonnement solaire au sommet de l'atmosphère (TSI, *total solar irradiance*).

Par ailleurs, les IMFs 6 des stations BOU, CAR et PAY et l'IMF 7 de TAT présentent des valeurs médianes de périodes instantanées très proches du cycle annuel de révolution de la Terre autour du Soleil.

Au-delà de la période annuelle, on note que les IMFs présentent des différences notables dans leurs distributions de périodes instantanées avec toutefois des ressemblances entre d'une part les stations BOU et CAR, et d'autre part les stations PAY et TAT. Cette dernière est la seule présentant une dixième IMF avec une valeur médiane de périodes instantanées de 2402,6 jours.

La seule analyse des distributions des périodes instantanées des différentes IMFs n'étant pas suffisante, elle a été complétée par l'analyse des spectres de Hilbert correspondants, rendant compte des capacités d'analyse temps-fréquence de la HHT.

On note tout d'abord pour les quatre stations la présence d'un « plateau » de composantes de hautes fréquences avec des périodes entre 2 jours et environ 100 jours. On note que l'amplitude de ces composantes varie suivant un cycle saisonnier. Les deux stations BOU et CAR présentent un intervalle de temps compris entre 100 et 300 jours pour lequel il n'y a pas suffisamment de fluctuations intenses dans le SSI pour qu'une IMF puisse y
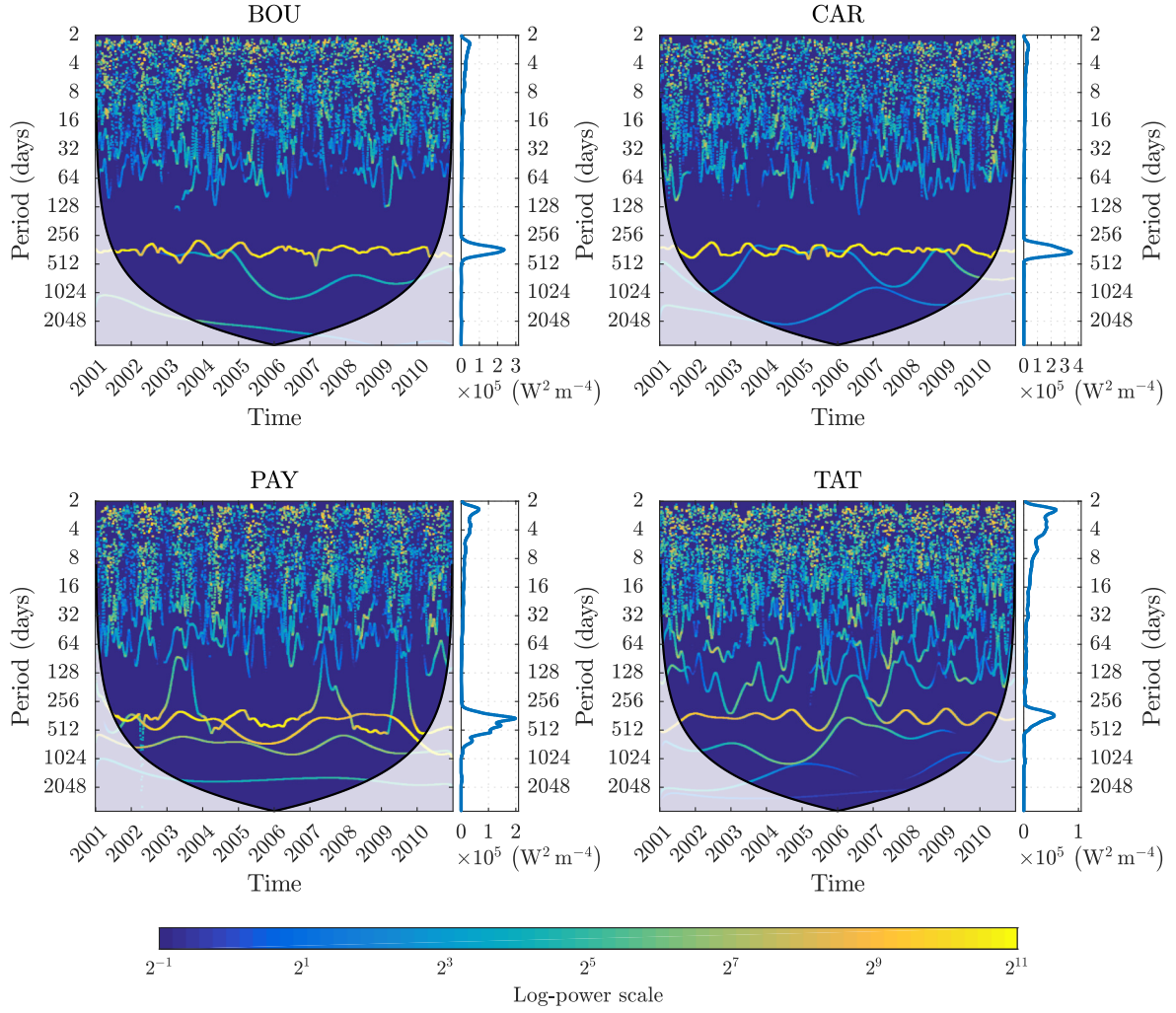
Figure 4 Spectre de Hilbert des dix ans de SSI à chaque station. La couleur correspond à la puissance observée à chaque instant (abscisse) pour chaque échelle (ordonnée). Le panneau à droite est le spectre marginal de Hilbert et donne la puissance pour chaque échelle.

être détectée. C'est aussi le cas pour PAY bien que la situation soit moins nette dans les années 2003, 2007 et 2009, où on observe un mixage d'IMFs associé à l'IMF 6. Quant à elle, TAT présente une IMF, la sixième, dans cet intervalle de temps.

Une certaine similarité des IMFs au-delà de la période annuelle avec des variabilités de l'activité solaire (QBOs, *solar quasi-biennal oscillations*) [Vecchio et al., 2012 , Kolotov et al., 2015] a été relevée.

Les similarités des échelles caractéristiques de variabilité entre les stations BOU et CAR d'une part, et PAY et TAT d'autre part, s'expliquent en partie par la nature de leurs climats respectifs, notamment l'occurrence de la couverture nuageuse : les stations BOU et CAR présentent plus couramment des situations de ciel clair tandis que le SSI à PAY et TAT est plus souvent perturbé par une couverture nuageuse variable et fragmentée.

Pour aller plus loin dans l'analyse de la variabilité du SSI journalier pour les quatre stations, la méthode proposée par Chen et al. [2013] pour discriminer les IMFs suivant leur nature déterministe ou stochastique a été utilisée. Cette méthode analyse, pour chaque IMF, la dépendance des fréquences moyennes pondérées (SWMF, *spectrum-weighted mean frequency*), au facteur de sous-échantillonnage, allant de 0.5 à 1, et permet de rejeter ou non l'hypothèse nulle que cette IMF est la réalisation d'un processus stochastique. Il a ainsi été montré que cette hypothèse nulle peut être rejetée seulement pour les IMFs correspondant au cycle annuel, c'est-à-dire les IMFs 6 pour BOU, CAR et PAY et l'IMF 7 pour TAT.

Chapitre 5 : La variabilité intrinsèque du SSI obtenu par des méthodes d'estimation alternatives aux mesures in-situ

Compte tenu de la faible densité mondiale de stations pyranométriques, des méthodes d'estimation du SSI alternatives ont été proposées. Les méthodes ayant recours aux satellites sont actuellement considérées comme effectivement complémentaires aux mesures in-situ [Lefevre et al., 2014]. Bien que reconnue comme étant moins précise que l'approche satellitaire [Lohmann et al., 2006 ; Boilley et Wald, 2015], une autre approche actuellement adoptée notamment pour des applications dans le domaine du photovoltaïque [Richardson et Andrews, 2014] consiste à recourir aux modèles numériques de ré-analyse météorologique.

L'objet de ce chapitre consiste à analyser comparativement les variabilités temporelles de séries de SSI journaliers du jeu de données WRDC décrit dans le chapitre 2, de deux stations (VIE et KIV), selon qu'elles proviennent de stations pyranométriques (GAW) ou de méthodes alternatives satellitaires (HC3v5) ou basées sur des modèles de ré-analyse météorologique (ERA-Interim et MERRA-2). Les séries temporelles correspondantes d'éclairement solaire hors atmosphère issues de SG2, et de SSI pour des condition de ciel sans nuage issues du modèle McClear ont fait l'objet de ces mêmes analyses comparatives.

La figure 5 représente pour la station VIE les spectres de Hilbert marginaux (i.e. intégrés temporellement) des différentes séries temporelles de SSI. Le spectre de l'éclairement hors atmosphère ne comprend qu'un mode largement dominant centré sur la période annuelle suivi d'un deuxième mode autour de 1400 jours. La série de SSI par ciel clair de McClear comprend le même type de mode annuel, épaulé par un deuxième mode plus faible dont le maximum correspond à une période de 800 jours. Elle comprend aussi des composantes de hautes fréquences avec des périodes de 2 jours à environ 300 jours. Les spectres marginaux des estimations et des mesures comprennent tous le même mode autour du cycle annuel, et présentent aussi tous une nette augmentation des composantes à la fois hautes et basses
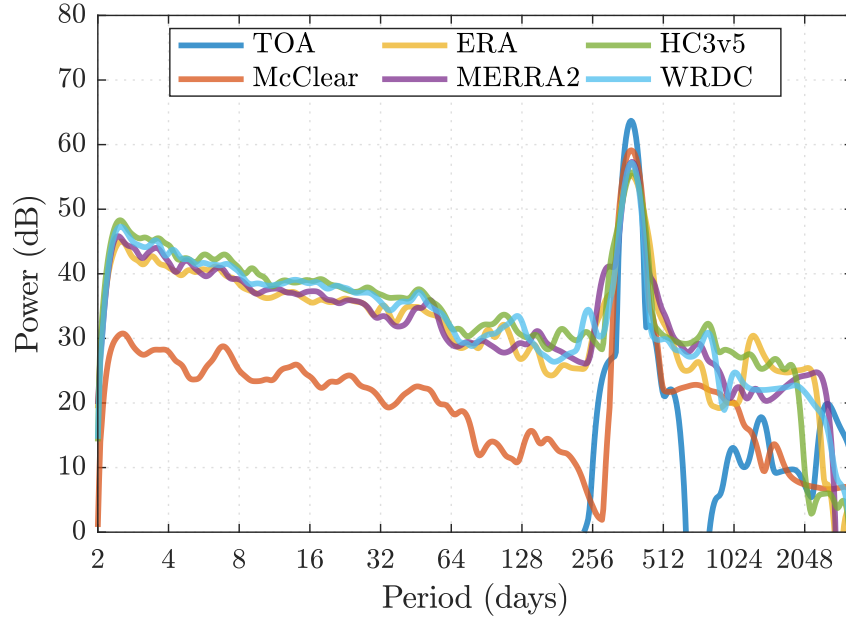
Figure 5 Spectres de Hilbert marginaux des différentes séries temporelles de SSI pour VIE.

fréquences par rapport à la fréquence annuelle. Bien que les spectres marginaux de Hilbert des séries issues des méthodes d'estimations soient semblables à celui de la série de mesures de SSI, une analyse comparative a été menée sur les séries elles-mêmes ainsi que pour les sept premières IMFs.

Les tableaux 1 et 2 montrent respectivement pour les deux stations VIE et KIV les coefficients de détermination et les racines carrées des erreurs quadratiques moyennes entre les IMFs révélées par la HHT appliquée aux différents jeux de données et celles des SSI de références issues des mesures in-situ. Ces comparaisons montrent très nettement que la méthode satellitaire HC3v5 surpasse très largement les deux méthodes issues des produits de ré-analyse météorologique ERA-Interim et MERRA-2 tant globalement sur les séries

Table 1 Coefficients de détermination et racines carrées des erreurs quadratiques moyennes entre les IMFs des SSI de références issues des mesures in-situ et celles révélées par la HHT pour les autres jeux de données. Pour VIE.

		IMF1	IMF2	IMF3	IMF4	IMF5	IMF7	Global
R^2								
	HC3v5	0.937	0.924	0.945	0.938	0.934	0.992	0.979
	ERA	0.639	0.636	0.675	0.755	0.703	0.987	0.921
	MERRA2	0.684	0.686	0.697	0.732	0.762	0.985	0.928
RMSE ($\mathrm{W\,m^{-2}}$)								
	HC3v5	9.1	6.2	4.2	3.6	2.8	5.7	14.1
	ERA	15.5	10.3	7.8	5.5	4.6	7.0	24.8
	MERRA2	16.1	9.7	7.5	5.6	4.3	8.4	26.4

Table 2 Coefficients de détermination et racines carrées des erreurs quadratiques moyennes entre les IMFs des SSI de références issues des mesures in-situ et celles révélées par la HHT pour les autres jeux de données. Pour KIV.

		IMF1	IMF2	IMF3	IMF4	IMF5	IMF7	Global
R^2								
	HC3v5	0.926	0.924	0.929	0.921	0.885	0.992	0.984
	ERA	0.611	0.677	0.692	0.753	0.670	0.984	0.931
	MERRA2	0.621	0.643	0.671	0.800	0.590	0.982	0.934
RMSE $(\mathrm{W\,m^{-2}})$								
	HC3v5	8.5	5.7	4.5	4.0	3.1	6.3	12.6
	ERA	13.9	8.7	6.5	5.2	4.5	8.4	24.5
	MERRA2	16.4	10.1	7.4	5.1	4.8	8.5	26.2

de SSI elles-mêmes que sur les différentes IMFs. L'écart de représentativité des IMFs se creuse à mesure que les fréquences des IMFs augmentent.

Chapitre 6 : Conclusion

Ce chapitre donne un bref résumé des résultats obtenus dans les chapitres précédents. Le travail de thèse a démontré la pertinence de l'usage de la HHT pour l'analyse des séries temporelles de SSI. L'analyse de séries de SSI obtenues dans des conditions climatiques différentes a mis en évidence des modes de variations (les IMFs) différents selon les climats. En ce qui concerne l'énergie solaire, cela a une portée pratique sur la sélection initiale de possibles sites d'implantation de centrales solaires.

Un autre résultat est que, quel que soit le climat parmi ceux explorés, le SSI présente des composantes stochastiques de périodes inférieures à l'année dont les amplitudes sont modulées par le cycle annuel par des interactions inter-échelles non-linéaires. Une technique a été développée pour extraire, décrire et caractériser ces composantes stochastiques. Dans leurs travaux sur la prévision du SSI, Ehnberg et Bollen [2005], Hoff et Perez [2010] ou Marquez et Coimbra [2013] soulignent l'importance d'une prévision de qualité de ces composantes. Le présent travail apporte une aide à une meilleure qualité de prévision par une caractérisation précise de ces composantes.

Toujours dans l'optique d'une prévision du SSI mais à plus long terme, de la journée à la saison typiquement, et puisqu'il a été montré que la description d'une série de SSI par ses IMFs est fonction du climat, il peut être envisagé de créer un catalogue de variations, comprenant donc différentes échelles de temps ; catalogue où on pourrait venir piocher une variabilité type par analogie avec une situation existante, et ainsi prédire un futur plus ou moins éloigné. Ce catalogue pourrait être bâti à partir de jeux de données de SSI dérivés de mesures par satellite, comme HelioClim-3. En ce qui concerne l'énergie solaire, cette

prévision des variations suivant de basses fréquences contribuerait à une meilleure prise de décision tant pour les investissements que les politiques énergétiques, et la prévision des variations de plus haute fréquence servirait au suivi de centrales solaires et de leur fonctionnement.

Cette thèse portait surtout sur l'évaluation d'une méthode pour l'analyse des séries temporelles du SSI. Plusieurs applications ont été esquissées, dont notamment, l'évaluation de la pertinence d'une série de SSI non-issue de mesures in-situ pour représenter les vraies variations du SSI. Ainsi, des conclusions peuvent-elles être tirées globalement mais également pour chaque échelle, avec pour portée pratique la mise à disposition de niveaux critères de performance et de comparaison permettant aux utilisateurs de données de ressource solaire de faire son choix.

On soulignera ici que l'approche d'analyse présentée peut s'appliquer à bien d'autres variables géophysiques et pas seulement au SSI. On notera également que si cette thèse n'a traité que la dimension temporelle, le formalisme peut être étendu à la dimension spatiale. Il sera alors nécessaire d'exploiter une autre technique d'EMD, comme la NA-MEMD (*noise-assisted multi-variate empirical mode decomposition*) ou la APIT-MEMD (*adaptative-projection intrinsically transformed multi-variate empirical mode decomposition*) afin d'obtenir une description du champ de SSI à l'aide du même nombre d'IMFs et les mêmes périodes afin de pouvoir procéder à des comparaisons et des analyses.

Table of contents

List of abbreviations

AM–FM   amplitude modulation – frequency modulation

APIT-MEMD   adaptive-projection intrinsically transformed multivariate empirical mode decomposition

BSRN   Baseline Surface Radiation Network

CAMS   Copernicus Atmosphere Monitoring Service

CEEMDAN   complete ensemble empirical mode decomposition with adaptive noise

COI   cone of influence

CWT   continuous wavelet transform

DFT   discrete Fourier transform

ECV   Essential Climate Variable

EEMD   ensemble empirical mode decomposition

EMD   empirical mode decomposition

ERA   European Reanalysis

FFT   fast Fourier transform

GCOS   Global Climate Observing System

GHI   global horizontal irradiance

HC3v5   HelioClim-3 version 5

HHT   Hilbert-Huang transform

HSA   Hilbert spectral analysis

ICEEMDAN   improved complete ensemble empirical mode decomposition with adaptive noise

IMF   intrinsic mode function

MERRA2   Modern-Era Retrospective analysis for Research and Applications, Version 2

NA-MEMD	noise-assisted multi-variate empirical mode decomposition
RMS	root-mean-square
RMSE	root-mean-square error
SSI	surface solar irradiance
STFT	short-time Fourier transform
TOA	top of the atmosphere
WMO	World Meteorological Organization
WRDC	World Radiation Data Centre

List of figures

List of tables

CHAPTER 1

Introduction

T HE Sun is of the utmost importance for the planet Earth. Not only did it play a central role in the formation of our solar system, but its mass also generates the gravity-well that enables the orbital revolution of our planet around the star's barycentre. But, most importantly, the Sun also represents the main power source for the Earth, outweighing by far any other contribution to energy budget of the planet. From a total solar radiative power of $4 \cdot 10^{26}$ W, similar in spectrum to that of a black body at approximately 5700 K, the electromagnetic radiation reaching the Earth is estimated to be of the order of magnitude of 10^{17} W [Beer et al., 2006]. The enormous theoretical potential of solar radiation and its implications for human energy needs truly stand out when one realizes that there is more solar energy impinging on the surface during one hour than the worldwide energy consumption for one whole year [Lewis and Crabtree, 2005].

This energy flux from the Sun that reaches our planet is the main driver behind the weather and climate systems on Earth [Trenberth et al., 2009]. It also has a strong influence on atmospheric photochemistry and on the biosphere [Madronich and Flocke, 1999]. Solar radiation also impacts human activities, such as photovoltaic power generation [Spertino et al., 2013], or agriculture [Gallagher and Biscoe, 1978]. As such, the surface solar irradiance has been recognized by the Global Climate Observing System (GCOS) as one of the Essential Climate Variables (ECVs) used for the characterization of the state of the global climate system and for long-term climate monitoring [GCOS, 2010].

Having recognised the importance of the solar radiation reaching the surface of the Earth, it is useful to introduce a formal definition of this physical quantity: surface solar irradiance (SSI) is the radiant flux density received from the Sun over the whole spectrum on a horizontal unit surface, at ground level; its unit is Watt per square meter ($\mathrm{W\,m^{-2}}$) [Wald, 2007]. The total SSI is the irradiance integrated over the entire spectrum covered by the solar radiation. The broadband SSI is the irradiance integrated over the part of

the spectrum where most of the total SSI is received the solar radiation, i.e. between 300 nm and 2800 nm. Depending on the domain of application, the SSI is also called solar flux or down-welling surface short-wave flux. For the scope of this study only the global SSI will be of interest. It is abbreviated in global horizontal irradiance (GHI). The GHI is measured by a thermopile pyranometer [WMO, 2014], which provides readings of the broadband SSI for a specific geographical location that are integrated both over time and over a solid angle of 2π sr [Noia et al., 1993].

The SSI is known to exhibit variations on a large dynamic range with respect to both time and geographical position. This is due to a wide array of factors. Some of these operate at long time-scales, from decades to millennia and beyond, and are related to the stellar variability of the Sun [Beer, 2000], or to changes in the orbital parameters of the Earth, e.g. obliquity of the ecliptic, eccentricity of the orbit, and axial precession [Beer et al., 2006]. Only changes in eccentricity effectively influence the amount of solar radiation reaching the top of the atmosphere, with the other two factors only affecting its relative spatial distribution [Beer et al., 2006]. During recent decades, a typical 11-year cycle of peak-to-peak variations of ∼0.1% has been observed in the solar flux incident upon the top of the atmosphere of the Earth [Lean, 2010]. The revolution of the Earth around the Sun and the obliquity of the ecliptic give rise to the yearly cycle of seasons. Another well-known factor is the rotation of the Earth around its own axis, that yields variations in the SSI with a period of one day. Atmospheric effects such as scattering and absorption, mainly due to O_2 and stratospheric O_3, water vapour, aerosols, and clouds [Madronich and Flocke, 1999] may exhibit even shorter time-scales. Orographic factors, such as shadowing effects due to relief, influence the SSI on a seasonal basis and down to very short time-scales.

In this context, better understanding the characteristic time-scales of variability of the solar radiation reaching the ground is the central topic of the present work. This theme is part of a farther reaching objective, namely to enhance knowledge of and to develop forecasting models for the spatio-temporal variability of the SSI, which can be ascribed to the research interests and scientific activities of MINES ParisTech.

The SSI is, like other ECVs, extensively used in science and policy circles for understanding climate evolution and guiding mitigation and adaptation efforts [Bojinski et al., 2014]. As such, long-term time-series of the SSI play a key role in science and policy consulting. They also strongly impact the engineering and financial sectors, providing support for, e.g., the selection of suitable sites for solar power plants or investment decisions, respectively [Schroedter-Homscheidt et al., 2006].

Therefore, to facilitate the task at hand, and without any loss of generality regarding the conclusions that may be drawn, the data under inquiry will be restricted to long-term time-series of daily means of the SSI. Throughout this work, only the temporal variability

will be of concern. On the one hand, proceeding in this manner has been motivated by the need to simplify computations and avoid dealing with large datasets of three-dimensional[1] SSI fields — an undertaking that, at this time, is proposed as a future research development. On the other hand, only having to account for one-dimensional variability allows for a greater emphasis on the data analysis and the discussion of the results, rather than getting tangled in the intricacy of the data itself.

In order to tackle the topic of the variability of the surface solar irradiance, the very concept of variability itself must first be properly addressed. Even though spatial variability is of course also of interest, as has been shown for example by Hoff and Perez, [2010], only the temporal dimension of SSI variability will be considered, in line with the scope of the study.

One of the most common ways of quantifying variability is through the use of simple descriptive statistics, such as the mean and the standard deviation [Hammer and Beyer, 2012]. Variability is then usually related to the time-scale at which analysis is carried out, with the fluctuations at the scale of interest being compared to the mean (or another low-pass filtered approximation) at the next larger scale. Thus, yearly sums of SSI are compared to their multiannual mean, daily sums to their monthly mean, minute values are compared to their corresponding hourly mean, etc. When, still, even finer (sub-second) scales are addressed, the second temporal derivative of irradiance may be used to quantify its instantaneous variability [Yordanov et al., 2013]. Another way to express SSI variability by using statistics is in terms of the probability density function of the irradiance, or by using a two-sample structure function, which is the root-mean-square (RMS) deviation of increments of irradiance, as described by Skartveit and Olseth, [1992]. Perez et al., [2016] propose a nominal variability metric defined as the standard deviation of the ramp rate of the clear sky index, i.e. the standard deviation of the fluctuations over a given period of the ratio of GHI to the GHI that would be observed if the sky were clear.

While adhering to the statistical approach and defining the RMS of a high-pass filtered version of the data as the *fluctuation factor* of the time-series, Otani et al., [1997] allude to employing the power spectral density as a way of describing variability. Hence, spectral methods provide a different approach of describing variability in that, by analysing the frequency contents of the data, a much clearer picture is rendered in terms of the scales, or the inverse of frequency, at which variations in the SSI occur. However, second-order measures, i.e. the power spectrum and the autocorrelation, provide no information about phase and are mostly suited to processing stationary time-series. More advanced spectral techniques, such as the wavelet transform, also yield phase information and can also be employed on non-stationary datasets. As such, Woyte et al., [2007] make use of wavelet

---

[1] Two spatial dimensions, corresponding to geographical latitude and longitude, and one temporal dimension.

periodograms to characterize fluctuations in solar irradiance, while Lave et al., [2013] use the wavelet transform to decompose a time-series of SSI for their proposed *Wavelet Variability Model*. Detrended fluctuation analysis is used by Pedron, [2012] to characterize persistence or long-range correlations in solar radiation datasets in the form of the Hurst exponent of a power law, which measures how fluctuations in the dataset scale with the size of a time-window of interest. The temporal variability of the Hurst exponent within a time-series is, in turn, an indication of the presence of non-stationary fluctuations, which point to multi-fractal behaviour. Calif et al., [2013] employ the power spectrum of a time-series in order to model these multi-fractal properties of the SSI by means of Hilbert spectral analysis, among other methods. This latter spectral method, first[2] applied to long-term SSI datasets by Duffy, [2004], is part of the so-called Hilbert-Huang transform [Huang et al., 1998] which plays a key role in this study, and will be detailed in due course.

Hitherto it has been shown that, generally, there is little consensus among the different uses of the very term (temporal) *variability*, with the implied meaning varying according to the field and scope with which it is employed. Hence, for the purposes of this study, the following definition of variability is adopted:

> **variability** — the fact or quality of being variable in some respect; tendency towards, capacity for, variation or change. [Oxford, 2015]

In spite of the large span of the characteristic scales of temporal variability of the SSI, most of the studies dealing with this physical quantity have focused primarily on a few selected time-scales of interest. As such, reports have either dealt with global averages and long-term trends [Trenberth et al., 2009; Abdel Wahab et al., 2010; Pachauri et al., 2014; Blanc et al., 2015], have only scrutinized the short-term, high-frequency variability [Yordanov et al., 2013; Lauret et al., 2016], or have focused exclusively on a few intermediate scales [Coskun et al., 2011; Medvigy and Beaulieu, 2012]. Although considerably differing in methods, the previously cited studies add valuable contributions to our knowledge of the SSI. But is it possible to analyse the variability of the SSI across multiple time-scales in a unitary way?

To do so, first a decomposition of the time-series into uncorrelated sub-constituents having distinct characteristic time-scales should be preferred. Analysis would then ensue in a like manner for each scale. The time-scales, or characteristic periods of a time-series can be identified with the inverse of the frequency at which the processes that generate the data occur. It then follows that methods portraying the changes of the spectral content of a time-series with respect to time are potentially good candidates, since they enable both the identification of the periodicities and of the dynamic evolution of the processes generating the data. A general class of useful signal processing techniques can thus be

---

[2] To the best knowledge of the author.

identified in the so-called time-frequency distributions, that depict the intensity (or energy) of a signal in the time and the frequency domains simultaneously [Cohen, 1989].

Another factor to be taken into account are the non-linear and non-stationary characteristics of the measured solar radiation data [Zeng et al., 2013; Calif et al., 2016]. Handling such data issued from the non-linear interaction of physical processes, often also found under the influence of non-stationary external forcings, calls for an adaptive data analysis approach [Wu et al., 2011].

In order to address the previous considerations, the study at hand will primarily make use of the Hilbert-Huang transform (HHT), an adaptive, data-driven analysis technique designed specifically for investigating non-linear and non-stationary data [Huang et al., 1998]. The HHT adaptively decomposes any dataset into basis functions that are derived solely from the local properties of the time-series. A time-frequency representation of the data can then be constructed from these basis functions. The HHT has seen extensive use in geophysical signal analysis and spectral estimation [Huang and Wu, 2008; Huang et al., 2014; Tary et al., 2014] and has also been employed on SSI datasets [Duffy, 2004; Calif et al., 2013; Bengulescu et al., 2016a; Bengulescu et al., 2016c; Bengulescu et al., 2016d; Calif et al., 2016; Bengulescu et al., 2017]. A similar technique has been independently proposed by Nagovitsyn, [1997] for the analysis of the non-linear, non-stationary, long range solar activity. The use of the HHT for the study of the temporal variability of the SSI is thus well-established. An introductory material detailing the inner workings of this data processing method is provided in a dedicated chapter.

Regardless, however, of the analysis techniques employed, when analysing data, there is always the need to discriminate between deterministic signals and what are assumed to be stochastic realizations of a noisy background [Rios et al., 2015]. The classical way to solve this when employing the HHT on geophysical signals, such as the SSI, is to presume some model for the background power spectrum, against which the identified features are then compared [Huang and Wu, 2008; Franzke, 2009; Franzke, 2012]. In contrast, the present study parts with the traditional approach, by adopting a novel, adaptive null-hypothesis that requires no *a priori* knowledge of the stochastic background processes, introduced by Chen et al., [2013]. Thus, the use of HHT combined with this innovative technique enables the discrimination of the deterministic signals from the quasi-stochastic ones independently of the class or the spectral characteristics of the random background fluctuations.

At this point it must be noted that one of the main outcomes of this study is a unitary approach, based on the HHT, for extracting and analysing the characteristic scales found in long-term time-series of the SSI. By means of this adaptive data analysis paradigm, new insights are gained into the variability of the SSI, allowing e.g. the identification of the type of climate based on the energetic intensity of certain time-scales — a key assessment for the initial site selection or the design and financing stages of a solar power plant

[Sengupta et al., 2015]. Another significant finding is that, irrespective of geographical location, the SSI exhibits high-frequency quasi-stochastic variability components which are amplitude-modulated by the deterministic annual cycle, through a non-linear cross-scale interaction. These outcomes are described in [Bengulescu et al., 2016a; Bengulescu et al., 2016c; Bengulescu et al., 2016d].

In order to analyse the temporal variability of the SSI, measurements of this physical quantity are needed, which are usually acquired by means of pyranometers at different ground-based stations across the globe. Indeed, pyranometric ground measurements of the SSI are of the utmost importance for this study, since this enterprise would not have been possible without quality ground SSI data, i.e. data having well-known, low uncertainty, few missing values, and long-term availability on the decennial scale. Nevertheless, it is found that even sources of high quality solar radiation measurements, such as the Baseline Surface Radiation Network (BSRN), a worldwide radiometric network providing accurate readings of the SSI at high temporal resolution [Ohmura et al., 1998], are spatially very sparse and may have data gaps of up to 13% [Roesch et al., 2011]. In practice, information about the SSI is often required at geographical locations different from any measuring station. But extending the representativeness of ground station measurements to surrounding areas cannot be applied to regions where the physical or climatological distance between stations is large [Zelenka et al., 1992]. Therefore, surrogate time-series are employed to supplement long-term solar resource assessment, such as satellite-based SSI estimates [Lefèvre et al., 2014] or radiation products provided by global atmospheric re-analyses from numerical weather models [Richardson and Andrews, 2014]. Improving the adaptation of these surrogate solar radiation datasets to a specific geographical site is an ongoing research effort [Polo et al., 2016]. Since long-term time-series of the SSI are the corner stone of the study, a dedicated chapter will detail all the datasets employed throughout the course of this study.

As such, another important outcome of this study can be summarized as a refinement of the existing methodology to assess the fitness for use of surrogate SSI products, through an improved in-depth comparison of their local time-scales of variability. While, hitherto, most studies have employed global statistical indicators [Espinar et al., 2009; Ineichen, 2014], the new approach allows the assessment of the fitness for use on a per time-scale basis. By using this approach, satellite-based methods are shown to outperform re-analyses products by a large margin in terms of estimating the measured SSI, across all time-scales [Bengulescu et al., 2017]. This finding is in agreement with previous studies that encourage the use of satellite-derived SSI over re-analyses products [Lohmann et al., 2006; Boilley and Wald, 2015].

Finally, the organization of the document can be summarized as follows. Chapter 2 describes the SSI datasets that lie at the foundation of this work. An introduction to the

adaptive data analysis approach of the Hilbert-Huang transform along with a brief review of the most known time-frequency analysis techniques are provided in chapter 3. Chapter 4 analyses the temporal variability of ground measurements of the SSI at different measuring stations around the world. The data-driven analysis of the temporal variability of the SSI is showcased. A statistical analysis of the intrinsic modes of variability is also carried out, with deterministic components being contrasted to the quasi-stochastic ones. Chapter 5 assesses the fitness for use of surrogate SSI data when pyranometric ground measurements are unavailable and also touches upon the role of atmospherically-induced variability in satellite estimates of solar radiation. The conclusions and outlook are presented in chapter 6. A bibliographical list concludes the study. The scientific outcomes of the study, in the form of four published research articles and as many communications, are listed in appendix A. Supplementary material for chapter 5 is offered in appendix B.

CHAPTER 2

Solar irradiance data

THIS chapter deals with the several datasets of measurements or estimates of the SSI that are investigated throughout this study. Since the analysis of the characteristic time-scales of variability of any physical quantity requires accurate measurements thereof, it follows that the data play a crucial role and are thus the cornerstone of the work at hand.

As previously mentioned in the introduction, this study will only deal with long-term time-series of the SSI, spanning approximately ten years. As such, the sampling rate of the data has been fixed at one day, i.e. the physical quantity of interest consists in daily means of the global solar irradiance that impinges on a unit horizontal surface at ground level. Various datasets are employed throughout this work, with a short summary thereof being provided below, preceding the more detailed description that is to follow in the dedicated sections.

First, the focus falls on pyranometric ground measurements stemming from the Baseline Surface Radiation Network (BSRN), a radiometric network providing high-quality readings of the SSI. The BSRN data are very accurate while having reduced uncertainty and good long-term availability, which makes them ideally suited for the intents and purposes of this study. SSI time-series of BSRN measurements are investigated in chapters 3 and 4, whence the temporal variability is inferred unencumbered by uncertainties or data unavailability. These datasets are described in section 2.1.

Nevertheless, in practice measuring stations offering such high-quality data are scarce, and therefore practitioners and researchers alike have to contend with other data sources. Such alternative data, sometimes of perhaps lower quality, can be e.g. issued from the radiometric networks of national meteorological agencies or from the World Radiation Data Centre (WRDC), a repository sponsored by the World Meteorological Organization (WMO) that ensures the availability of SSI data to the international scientific community.

Added to this, more often than not the geographical location of interest is situated too far away from any ground measurement station, so as to make intra- or extrapolation impractical. When this occurs, surrogate SSI data come into play, such as estimates derived from satellite imagery, or radiation products issued from numerical weather prediction models running in re-analysis mode. Measured SSI data from the WRDC repository, along with satellite and re-analysis estimates thereof, are analysed in chapter 5, which also deals with the topic of atmospherically-induced variability. As such, apart from these three datasets of SSI, chapter 5 also employs time-series of clear-sky and exo-atmospheric solar irradiance. All these datasets are detailed in section 2.2.

Lastly, since for all datasets the geographical locations for which the data were retrieved are situated at boreal latitudes, any henceforth reference to seasons and seasonal phenomena shall be understood as occurring in the northern hemisphere.

2.1   BSRN pyranometric data

The first set of data under scrutiny in this study consists in four ten-year time-series of daily means of SSI obtained from high-quality measurements performed at four different locations (figure 2.1 and table 2.1). The measurement stations are part of the BSRN, and provide accurate readings of the SSI at 1 min temporal resolution and with an uncertainty requirement at $5 \, \mathrm{W \, m^{-2}}$ [Ohmura et al., 1998]. For a particular BSRN station of interest for this study, situated in Payerne, Switzerland, Vuilleumier et al., [2014] list the combined uncertainties that affect SSI measurements as being in the range of 5.7–7.5 $\mathrm{W \, m^{-2}}$ for a small signal (50 $\mathrm{W \, m^{-2}}$) and in the range of 9.0–11.6 $\mathrm{W \, m^{-2}}$ for a large signal (1000 $\mathrm{W \, m^{-2}}$).

The four time-series for the period 2001–2010 contain each $10 \times 365 = 3650$ days each (leap days are omitted) and have been quality checked following the procedure of Roesch et al., [2011]. Next, daily means of SSI were then calculated from these raw time-series only if more than 80% of the data during daylight were valid. Lastly, any isolated missing daily means were completed by linear interpolation applied to the daily clearness index, K_T, which is the ratio between the daily mean of SSI and the daily mean of the total solar irradiance received on a horizontal surface at the top of atmosphere for the same geographical coordinates.

Four measuring stations were selected in order to capture various climatic conditions. Two are located in Europe (Carpentras, France; Payerne, Switzerland), one in Japan (Tateno), and one in North America (Boulder). Boulder (herefater abbreviated in BOU) experiences a mid-latitude steppe, cool type of climate (Köppen-Geiger: BSk), while at Carpentras (abbreviated in CAR) the climate is a humid subtropical, Mediterranean one (Köppen-Geiger: Csa). Both sites experience many sunny days during the year. As a rule
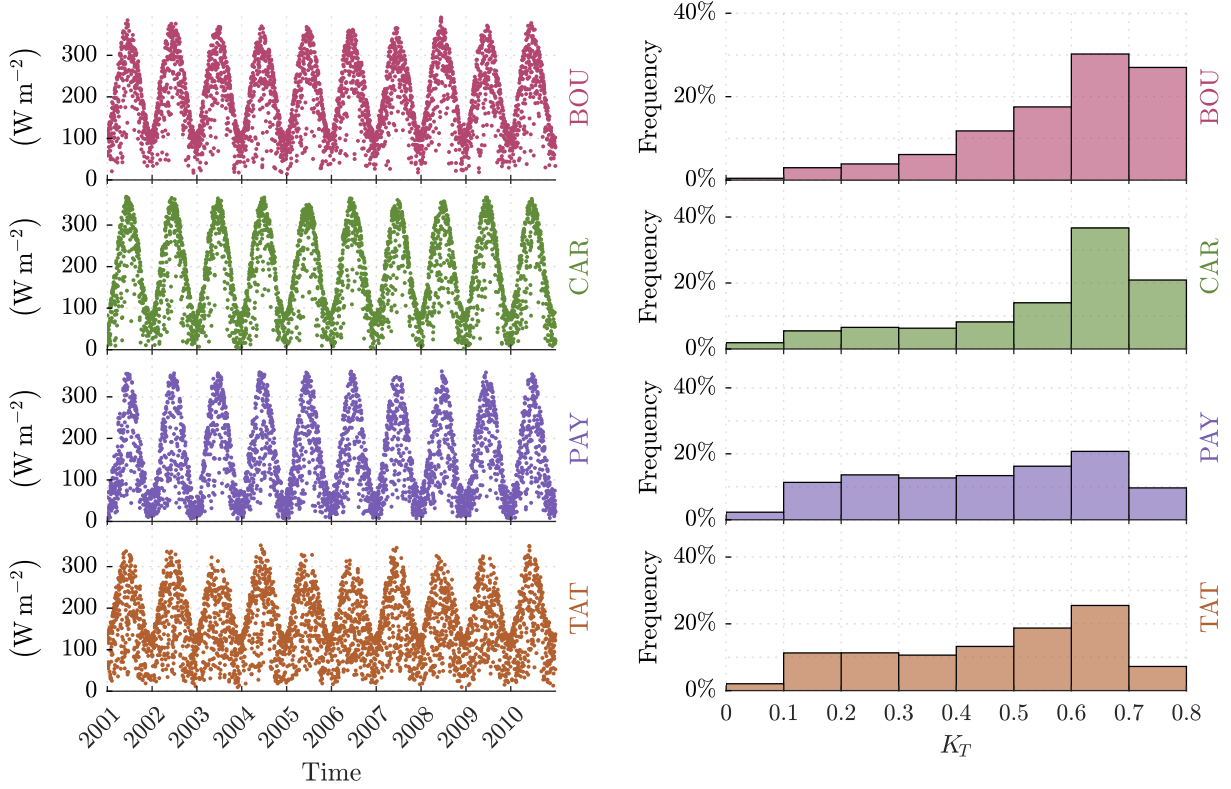
(a) The four decennial SSI time-series investigated in this study, spanning 2001 through 2010. Each point corresponds to a daily mean of SSI. Time markers on the abscissa indicate the start of the corresponding year.

(b) Histograms of the daily clearness index K_T over their decennial time-span. The relative frequency is expressed in percent.

Figure 2.1 The BSRN datasets employed throughout this study — (a) raw data and (b) clearness index histograms. From top to bottom: BOU, CAR, PAY, and TAT.

of thumb, K_T equal to 0.2–0.3 denotes cloudy, overcast conditions, while K_T around 0.7 indicates sunny conditions. Figure 2.1b exhibits the histograms of K_T for the four stations. One may observe the high frequencies of the greatest values of K_T for BOU and CAR. The median $\widetilde{K_T}$ is equal to 0.63 for both BOU and CAR, which means that half of the days exhibit K_T greater than 0.63. The climate in Payerne (PAY) is classified as marine west coast, mild (Köppen-Geiger: Cfb), and Tateno (TAT) has a humid subtropical, east coast climate (Köppen-Geiger: Cfa). Compared to BOU and CAR, PAY and TAT exhibit more uniform histograms, with less days with cloud-free conditions, and experience more overcast and broken clouds conditions. The median $\widetilde{K_T}$ is equal to 0.47 for PAY and 0.51 for TAT. Except for TAT, which is embedded in an urban setting, the stations are located in rural environments; the local topography for BOU and TAT is flat with grassy surfaces, while for CAR and PAY the area is hilly with cultivated surfaces [BSRN, 2015].

Table 2.1 Ground measurement stations listing.

Code	Location*		Latitude[†]	Longitude[†]	Climate[‡]
BOU	Boulder	(US)	40.0500	-105.0070	BSk
CAR	Carpentras	(FR)	44.0830	5.0590	Csa
PAY	Payerne	(CH)	46.8150	6.9440	Cfb
TAT	Tateno	(JP)	36.0581	140.1258	Cfa

* Country codes according to ISO 3166-1 alpha-2.
[†] Positive north for latitude and positive east for longitude, following ISO 19115.
[‡] Köppen-Geiger climate classification according to Kottek et al., [2006].

Data availability: The raw BSRN datasets employed in this study are made available by König-Langlo et al., [2015]. Archives containing the data can be found at http://doi.org/10/brqd.

2.2 WRDC pyranometric data and estimates of SSI

The data analysed in chapter 5 consist in multiple time-series of daily means of solar irradiance corresponding to two geographical locations in Europe: Vienna, Austria (48.25° N; 16.35° E), and Kishinev, Moldova (47.00° N; 28.82° E). The temporal coverage of the data is 9 years, from February 1st 2004 to January 31st 2013. The number of samples in each dataset is $9 \times 365 = 3285$ days (leap days are omitted). The dataset for Vienna (VIE henceforth) is plotted in figure 2.2a. A similar plot for Kishinev (hereafter KIV) is provided in figure 2.2b.

For the assessment of the fitness for use of surrogate SSI data in chapter 5, six datasets are employed for each geographical location:

1. modelled exo-atmospheric irradiance;
2. modelled clear-sky irradiance at ground level;
3. pyranometric ground measurements of the SSI;
4. estimates of the SSI based on Meteosat satellite imagery;
5. radiation products from the European Reanalysis (ERA)-Interim;
6. radiation products from the Modern-Era Retrospective analysis for Research and Applications, Version 2 (MERRA2).

The time-series of solar irradiance received by a horizontal surface at the top of the atmosphere (TOA), or exo-atmospheric irradiance, has been generated using the SG2 algorithm [Blanc and Wald, 2012]. This dataset is built using the constant value of 1367 $W\,m^{-2}$ for the total solar irradiance. The total solar irradiance is defined as the yearly average of the solar irradiance received on a plane always normal to the sun rays and
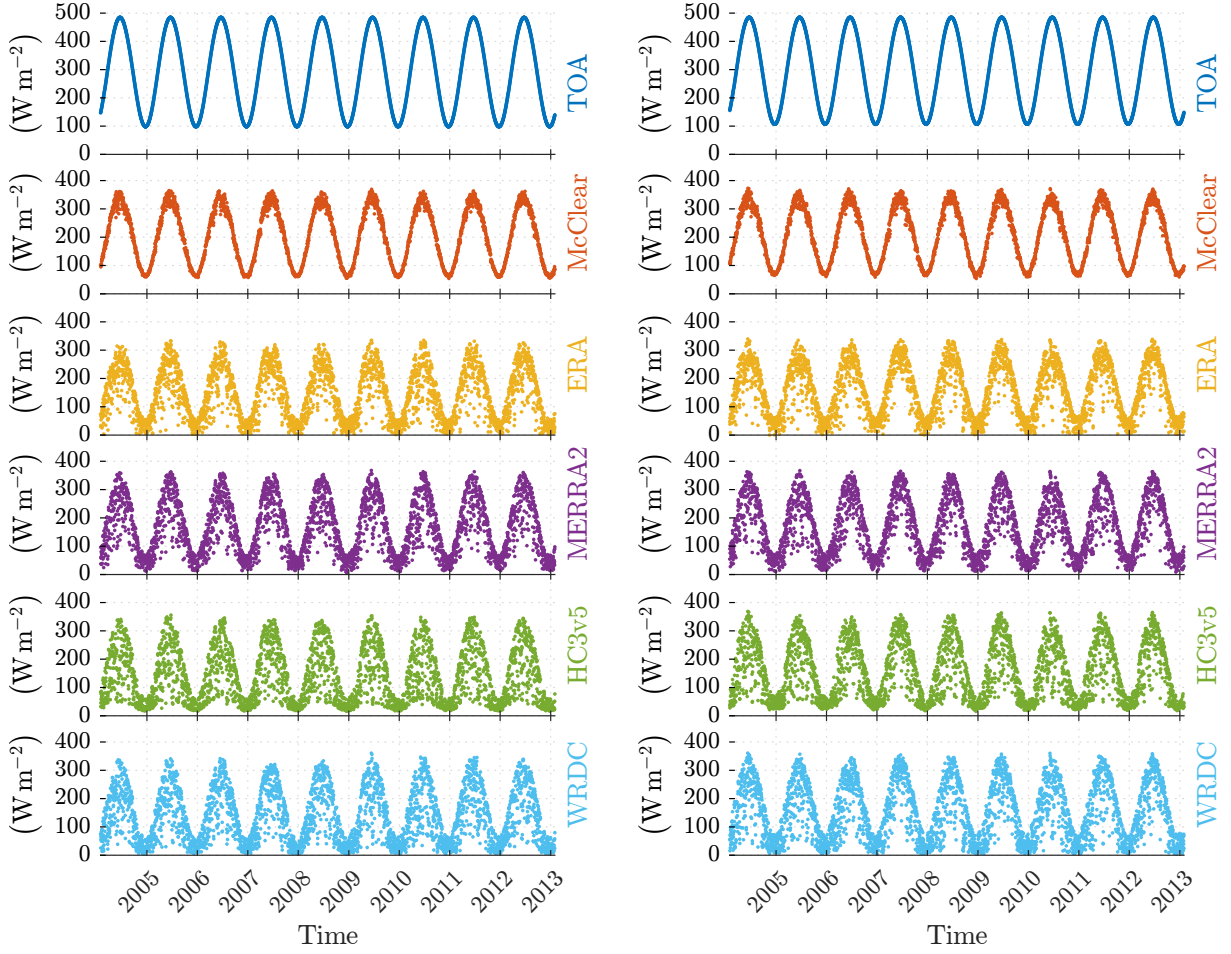
(a) The solar irradiance time-series for VIE

(b) The solar irradiance time-series for KIV

Figure 2.2 The solar irradiance time-series for (a) VIE, and (b) KIV investigated in chapter 5, spanning February 1st 2004 to January 31st 2013. From top to bottom: TOA, McClear, ERA, MERRA2, HC3v5, and WRDC. Each point corresponds to a daily mean of irradiance. Time markers on the abscissa indicate the start of the corresponding year.

located at the top of the atmosphere. Recently, the yearly mean of the total solar irradiance has been revised to 1361 W m^{-2} [Prša et al., 2016] or 1362 W m^{-2} [Meftah et al., 2014]. This change in amplitude is very minor and does not impact the validity of the analysis, however, since the total solar irradiance exhibits typical peak-to-peak variations of $\sim$0.1% on a 11-year cycle [Lean, 2010].

The dataset of downwelling surface solar irradiance, under clear-sky conditions (i.e. cloud-free), is generated using the McClear model [Lefèvre et al., 2013]. The McClear model is part of the Copernicus Atmosphere Monitoring Service (CAMS) and its inputs comprise 3-h estimates of the aerosol properties and total column contents of water vapour and ozone also provided by CAMS [Hoyer-Klick et al., 2015].

Estimates of the SSI derived from Meteosat satellite imagery by the Heliosat-2 method as described by Rigollier et al., [2004] and modified by Qu et al., [2014] are obtained from the HelioClim-3 version 5 (HC3v5) database [Blanc et al., 2011].

Pyranometric ground measurements of the daily SSI were obtained from the World Radiation Data Centre (WRDC) [Tsvetkov et al., 1995] for the two stations. No detailed information on data quality is provided except for a quality flag. Considering the high quality of maintenance of these stations, we consider that the data obey the good quality level set by the World Meteorological Organization [WMO, 2014] which specifies a 5% uncertainty expressed as the percentile 95 of the deviations (P95). If a normal law is assumed for the deviations, then there is a 0.3% chance that a deviation exceeds 1.5 times P95, i.e. 7.5% of the SSI. Given the global means of SSI of the WRDC datasets, both greater than 135 $\mathrm{W\,m^{-2}}$, the 7.5% or P99.7 uncertainty is 10.2 $\mathrm{W\,m^{-2}}$ for VIE and 11.3 $\mathrm{W\,m^{-2}}$ for KIV.

The ERA-Interim product "Surface Solar Radiation Downwards" [ECMWF, 2009], from 2004 to 2014, was retrieved using the `ecmwfapi` Python library on the Meteorological Archival and Retrieval System (MARS). The raw ERA-Interim data are a forecast of accumulated SSI expressed in $\mathrm{J\,m^{-2}}$, and has a spatial resolution of 0.75° × 0.75°. Both the H+12 forecast from 0 h UT re-analysis and from 12 h UT re-analysis were summed and then divided by the number of seconds in 24 h to obtain a daily SSI value in $\mathrm{W\,m^{-2}}$ for the 4 nearest points around the location. These values were then bi-linearly interpolated at the exact location.

The 1-h radiation diagnostics `M2T1NXRAD` from the MERRA2 [GMAO, 2015] have been extracted for the four nearest points and bi-linearly interpolated to generate the time-series at the exact location. The MERRA2 data are in $\mathrm{W\,m^{-2}}$ directly and have a spatial resolution of 0.5° × 0.65° in latitude and longitude. Values were then averaged over each day to obtain a daily SSI value in $\mathrm{W\,m^{-2}}$.

Data availability: The datasets described herein can be accessed as follows.

The datasets for both TOA and McClear, provided by CAMS, can be retrieved at: http://www.soda-pro.com/web-services/radiation/mcclear.

The ERA-Interim dataset can be accessed at: http://doi.org/10/bstv.

The MERRA2 radiation diagnostics `M2T1NXRAD` is available at: http://doi.org/10/bstw.

The WRDC datasets can be accessed at: http://wrdc.mgo.rssi.ru/wrdc_en.htm.

The HC3v5 dataset was downloaded from the SoDa Service, managed by Transvalor SA, at: http://www.soda-pro.com/web-services/radiation/helioclim-3-for-pay. Data are available to anyone for free for years 2004–2006 as a GEOSS Data-CORE (GEOSS Data Collection of Open Resources for Everyone) and for-pay for the most recent years with charge depending on requests and requester.

CHAPTER 3

An introduction to the Hilbert-Huang transform

As previously discussed in the introductory section, this chapter provides an overview of the new adaptive data analysis paradigm employed in this work. To recapitulate, the focus of this study lies on scrutinizing the characteristic scales of the temporal variability of the surface solar irradiance. But the time-scale is nothing more than the inverse of the frequency at which the processes that generate the data take place. Thus, a first prerequisite is that the analysis method of choice should be able to identify the frequency, or spectral content of a signal. This is necessary, but not sufficient, however, since the physical processes generating the data are not known beforehand and thus the assumption that their frequencies remain stable throughout time cannot be made, i.e. the time-series of SSI cannot been considered as stationary. Moreover, variability in frequency, or time-scale, would provide valuable information about the dynamics of the processes that drive the fluctuations of the SSI.

In this context, methods that depict the changes of the spectral content of a time-series with respect to time are established as potentially good candidates. A general class of useful signal processing techniques can thus be identified in the so-called time-frequency representations, that depict the intensity (or energy) of a signal in the time and the frequency domains simultaneously [Cohen, 1989]. Such techniques enable not only the identification of the periodicities, but also of the dynamic evolution of the physical mechanisms generating the data, and are commonly employed for geophysical signal processing [Tary et al., 2014].

Therefore, in this chapter, an attempt will be made to address some of the topics associated with various processing techniques of geophysical data, to highlight some of the particularities of different approaches, as well as to seek to identify the one that would

allow an as unobtrusive as possible scrutiny into the scales of temporal variability inherent to the SSI field.

In what follows, a survey of the effectiveness of several mathematical methods in revealing the variability of the SSI will be conducted. Three spectral methods that yield a time-frequency representation of the time-domain analysed signals will be of concern for this study. First, the (short-time) Fourier transform [Portnoff, 1980] will be discussed and used to introduce the concept of time-frequency analysis, the interest thereof being mostly historical. Next, following the evolution of time-frequency methods, the continuous wavelet transform [Grossmann and Morlet, 1984] will be presented. Lastly, a new paradigm in signal processing, the adaptive data-driven analysis spearheaded by the Hilbert-Huang transform [Huang et al., 1998], will be discussed. It will be shown that the Hilbert-Huang transform is a time-frequency analysis technique well suited for the purpose of the work at hand, namely the investigation of the characteristic scales of temporal variability of the SSI. This is not, by any account, an exhaustive collection of every possible procedure that can be used to identify the intrinsic scales of variability of the SSI, nor does it attempt to be. For a survey of spectral estimation methods employed in geophysical signal processing and interpretation the reader is kindly directed to the reviews of Ghil et al., [2002] and Tary et al., [2014].

3.1   A brief overview of time-frequency analysis

3.1.1   The (short-time) Fourier transform

The origins of the Fourier transform can be traced back to a paper submitted in 1807 to the *Académie des Sciences* in Paris by the French mathematician Joseph Fourier, concerning the study of heat transport in solid bodies [Grattan-Guinness, 1969]. The main insight behind the transform is that of decomposing a function into a trigonometric series, yielding its spectrum or frequency contents in the process. The transform, or more precisely a discrete version thereof, has become widely popular in a vast number of disciplines — including estimating the power spectra of geophysical data [Hinich and Clay, 1968] — after Cooley and Tukey, [1965] published their discovery[1] of a fast and efficient method of computing it, nowadays termed the fast Fourier transform (FFT).

Using complex exponential notation[2], the *Fourier transform* of a signal $x(t)$ can be written as:

$$X(f) = \mathcal{F}\left\{x\left(t\right)\right\}(f) = \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft}\,\mathrm{d}t, \qquad f \in \mathbb{R},\ j = \sqrt{-1}. \qquad (3.1)$$

---

[1] Although regarded as new at the time it was published, it was later found that the method, or variations of it, had been continuously re-invented along the time [Cooley et al., 1967; Heideman et al., 1984].

Conversely, the *inverse Fourier transform* of X is given by:

$$x(t) = \mathcal{F}^{-1}\left\{X\left(f\right)\right\}(t) = \int_{-\infty}^{\infty} X(f)e^{j2\pi ft}\,\mathrm{d}f, \qquad t \in \mathbb{R}. \tag{3.2}$$

As to what regards the terminology of the Fourier transform, if $x(t)$ is the original signal under scrutiny, then $X(f)$ is called the spectrum, or the frequency domain representation, of x. The variable t denotes time, measured in seconds, while f denotes frequency, measured in cycles per second, or Hertz. The inverse of the frequency is simply the period, or time-scale $T = f^{-1}$, measured in seconds. As such, the original signal $x(t)$ contains information about how events play out in time and its Fourier transform $X(f)$ describes how these events take place in the spectral realm, by revealing the frequency contents of $x(t)$. Thus, the Fourier transform in (3.1) and its inverse in (3.2) form a pair and provide a reversible mapping, allowing one to smoothly change the representation from one domain to the other.

The Fourier transform, as it has been introduced, deals only with continuous signals and requires an infinite number of terms to arrive at a result. However, the measurements of the SSI that this study attempts to analyse are stored in a digital format, hence in practice the discrete Fourier transform (DFT) should be employed. Thus, any henceforth mention of the Fourier transform shall be understood as referring to the DFT, computed by means of the FFT algorithm.

In this context, it is useful to introduce the definition[3] of the DFT, as per equation (3.3). The continuous signal $x(t)$ and the continuous spectrum $X(f)$, are replaced by the discrete signal $x[n]$ and the discrete spectrum $X[k]$, respectively, both of length N. The superposition operator, previously an integral, is replaced by the discrete sum and the frequency now only takes on discrete values — $f = k/N$.

$$X[k] = \sum_{n=0}^{N-1} x[n]e^{-j2\pi nk/N} \tag{3.3}$$

A distinct feature of the DFT is that it considers both the signal and the spectrum sequences as being periodic in N. As such, instead of considering the original signal as being composed of N samples in complete isolation, the DFT regards it as extending to infinity, with the same N samples repeating over and over. Although it is obvious that the first and last samples of the original signal are completely unrelated, from the point of view of the transform they are adjacent. When a signal is periodically extended,

---

[2] The imaginary unit notation j is adopted, following the electrical and control systems engineering convention, to avoid confusion with the electric current, whose notation is also i.

[3] See [Cochran et al., 1967].

however, incongruously with its fundamental period, discontinuities will appear at the boundaries of the observation, which are responsible for spectral contributions known as leakage. Nevertheless, this effect can be somewhat mitigated by means of a taper, or window function [Harris, 1978]. Thus, a drawback of using the DFT can already be identified. While not every signal is periodic in nature, by employing the Fourier transform for signal analysis one is, nevertheless, forced to adopt the circular "view of the world" that the DFT entertains.

Another imperfection of the DFT, termed the picket fence effect, also deserves some attention, as it is a source of uncertainty in interpreting the results obtained when using the Fourier transform for signal analysis [Bergland, 1969]. Recalling the definition from (3.3) and applying Euler's formula the DFT can be rewritten as:

$$X[k] = \sum_{n=0}^{N-1} x[n]e^{-j2\pi nk/N} = \sum_{n=0}^{N-1} x[n]\cos\left(\frac{2\pi nk}{N}\right) - j\sum_{n=0}^{N-1} x[n]\sin\left(\frac{2\pi nk}{N}\right) \qquad (3.4)$$

Therefore, according to (3.4), the discrete Fourier transform can be thought of as measuring the similarity between the signal of interest and a set of sine and cosine waves with different periods. These orthogonal sequences form a basis onto which the signal can be projected, called Fourier basis sequences. In the case of the DFT, however, the frequencies of the sine and cosine waves are confined to take on values from the discrete sequence $\{k/N \,|\, k \in \{0, 1, \ldots, N-1\}\}$. As such, from all the possible frequencies contained in the input signal, only those that exactly match one of the basis sequences will project onto the basis set as a single value, the rest being be represented as a non-zero linear combination of the whole set [Harris, 1978]. Hence, care must be taken when trying to determine whether these compound spectral components indeed have a physical meaning, or are rather just artefacts arising from the fact that the signal is being constrained by the mathematical apparatus used for analysis.

Added to this, the use of linear superposition of pure sinusoids itself introduces some ambiguity. This is because, whenever a wave deviates from a pure sinusoid, its Fourier transform will contains harmonics, i.e. extra frequency components that are required to reproduce the modified wave shape. These harmonics have no physical interpretation and may hinder an unambiguous reading of the Fourier spectrum for non-linear phenomena, i.e. phenomena associated with dynamical systems where "due to feedback or multiplicative effects between the components, the whole [system[4]] becomes something greater than the mere sum of its individual parts" [Boeing, 2016].

Another already prefigured issue of the DFT is also worth addressing, namely its global scope. Since each frequency component is represented as a linear superposition of waves

---

[4] Author's note

covering the whole span of the signal of interest, the spectrum is totally devoid of locality — absolutely no time-domain information is provided with respect to whether the spectral contents completely or partially overlap, or whether they are spaced apart, and if so, by how much.

Clearly, such a method that purely provides information about the frequency domains, while being completely oblivious to the time dimension, is not ideally suited for analysing the temporal variability of the SSI. An attempt to correct this was proposed in the classical paper of Gabor, [1946], under the name of "information diagrams" — nowadays called the short-time Fourier transform (STFT) — by adopting "the view that both time and frequency are legitimate references for describing a signal".

The main idea behind the STFT is to first pass the signal under scrutiny through an analysis window, and then take the Fourier transform. By sliding the window along the time axis, a time-frequency distribution is obtained. A definition of the STFT, adapted from [Tary et al., 2014], can be given as:

$$X(\omega, \tau) = \int_{-\infty}^{\infty} x(t)w(t - \tau)e^{-j\omega t}\,\mathrm{d}t, \qquad \omega, \tau \in \mathbb{R}. \tag{3.5}$$

In (3.5) the short-term spectrum, $X(\omega, \tau)$, is now a bivariate function representing the local behaviour of $x(t)$ at angular frequency $\omega = 2\pi f$, as viewed through the sliding window $w(t - \tau)$ [Portnoff, 1980]. In practice, the window, also called taper, is usually a Hann or Gaussian function [Tary et al., 2014], τ being the time delay to its center. For the remainder of the discussion, a Gaussian window function will be considered[5] — a detailed discussion focused on windows used for harmonic analysis can be found in [Harris, 1978]. When using the STFT on a computer, the continuous signal, $x(t)$, and window, $w(t)$, are replaced by their discrete counterparts $x[n]$ and $w[n]$ respectively, with the DFT also replacing the integral form of the Fourier transform.

It is interesting to introduce next the concept of time-frequency resolution, i.e. how fine a detail can be resolved by the STFT in time and in frequency. As recognized by Gabor, [1946], on the one hand, by widening the support of the window function the spectral leakage is reduced, but at the expense of diminished time resolution, as a longer window implies loosing locality in the time domain. On the other hand, shortening the taper enhances time resolution, but contributes to increased spectral smearing, thus loosing in terms of frequency localization. Figure 3.1 illustrates this concept, where the STFT is seen as superimposing a tiling on the time-frequency plane. In practice, the energy distribution of the signal in the time-frequency plane is rendered by means of the spectrogram, defined as the square modulus of the STFT from (3.5) [Rioul and Vetterli, 1991].

---

[5] When employing a Gaussian window, the STFT is also called the Gabor transform.
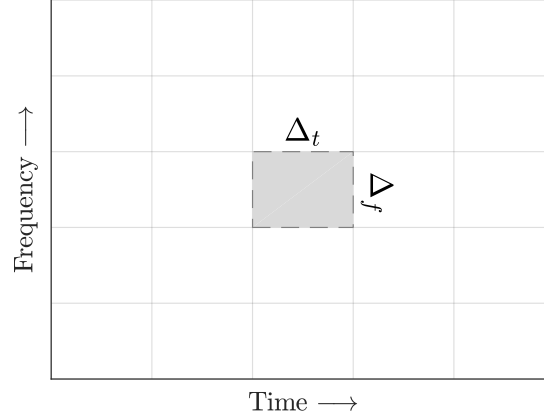
Figure 3.1 The tiling of the time-frequency plane by the short-time Fourier transform. The shaded rectangular area can be viewed as the region where most of the energy of one of the basis functions of the transform is concentrated.

The tiling can also be understood as how the analysed signals are sampled in time and frequency, as it depicts the approximate localisation of the basis functions of the transform [Herley et al., 1993]. Given a signal, $x(t)$, with unit norm, its time spread can be described by:

$$\Delta_t = \int_{-\infty}^{\infty} t^2 |x(t)|^2 \, \mathrm{d}t \tag{3.6}$$

By analogy, its bandwidth, based on its Fourier transform, $X(f)$, can be written as:

$$\Delta_f = \int_{-\infty}^{\infty} f^2 |X(f)|^2 \, \mathrm{d}f \tag{3.7}$$

Then, by virtue of the uncertainty principle, the area of a time-frequency tile can be regarded as approximately constant, as in Herley et al., [1993], i.e. time and frequency resolution cannot be arbitrarily refined, instead one must be traded for the other:

$$\Delta_t^2 \cdot \Delta_f^2 \geq \frac{\pi}{2} \tag{3.8}$$

Although it may seem that, by its ability to resolve events in time, the STFT provides for a better signal analysis technique, it is nevertheless plagued by the same problems as the underlying DFT on which it relies. Artefacts are induced whenever non-linear periodic components are present and spectral leakage still occurs if the support of the window does not match the wavelength of the spectral contents. And even if the STFT provides some information about the time domain, it is nevertheless still unable to temporally differentiate the signal contents below the scale of the window length, i.e. the signal is assumed to be stationary within the confines of the window. This last remark raises an interesting question: seeing that the window is fixed once and for all prior to the analysis, how can one ascertain that the chosen support of the window, in time or frequency, fits the signal under scrutiny, being, thus, appropriate for the study to be conducted?

3.1.2 The continuous wavelet transform

At this point, it would seem that the rigid window choice of the STFT may be something of a detriment in revealing all the information embedded within a signal, especially when it contains transient, non-stationary features. It appears as if a better approach would be to employ a somewhat more flexible windowing scheme, by continuously allowing the reshaping of the taper, trading frequency for time resolution based on the region of interest on the time-frequency plane.

This is precisely what Grossmann and Morlet, [1984] have proposed in one of the first articles describing the wavelet transform, namely to "analyse functions in terms of wavelets obtained by shifts [...] and dilations from a suitable basic wavelet" of fixed shape. The idea of decomposing a signal into "elementary wavelets" can be seen as a refinement of the original time-frequency representation introduced by Gabor, [1946]. Similarly to the Gabor transform, where the basis functions are Gaussian windows filled with sinusoids of different frequencies, the "basic wavelet" proposed for analysis by Goupillaud et al., [1984] is a also sine wave modulated by a Gaussian window. The wavelet decomposition retains the time-shifting approach, but, unlike the STFT, it does away with the shifts in frequency, instead replacing them with dilations, or contractions, of the basic wavelet. A side by side comparison of the basis functions employed by the Gabor transform and the wavelet transform is presented in figure 3.2.

The effect of these scaled basis functions on the tiling of the time-frequency plane, can be observed in figure 3.3. While the tiles of the time-frequency plane still obey the uncertainty principle from (3.8), the added flexibility with regard to the scaling of the basis functions enables the wavelet transform to achieve high frequency resolution when analysing slowly varying features and high temporal resolution when investigating rapid bursts and transient phenomena [Rioul and Vetterli, 1991]. It is worth mentioning that, in the context of the wavelet transform, the concept of frequency is intimately related to the time-scaling. Thus, low scales correspond to high-frequency — as the basis functions are compressed, the same number of oscillations occur faster in time. In contrast, high scales are associated with low frequency — owing to the dilation of the time-support, the oscillations take longer to complete.

To formalize this, consider a basic, or mother, wavelet, $\psi(t)$, from which, by scaling with some factor $\lambda \in \mathbb{R}$, $\lambda > 0$, and time-shifting by $\tau \in \mathbb{R}$, a family of wavelets can be defined:

$$\psi_{\lambda,\tau}(t) = \frac{1}{\sqrt{\lambda}} \psi \left(\frac{t-\tau}{\lambda} \right) \tag{3.9}$$

Then, the continuous wavelet transform (CWT) can be regarded as measuring the similarity between a signal of interest, $x(t)$ and the basis functions $\psi_{\lambda,\tau}(t)$, as in [Rioul
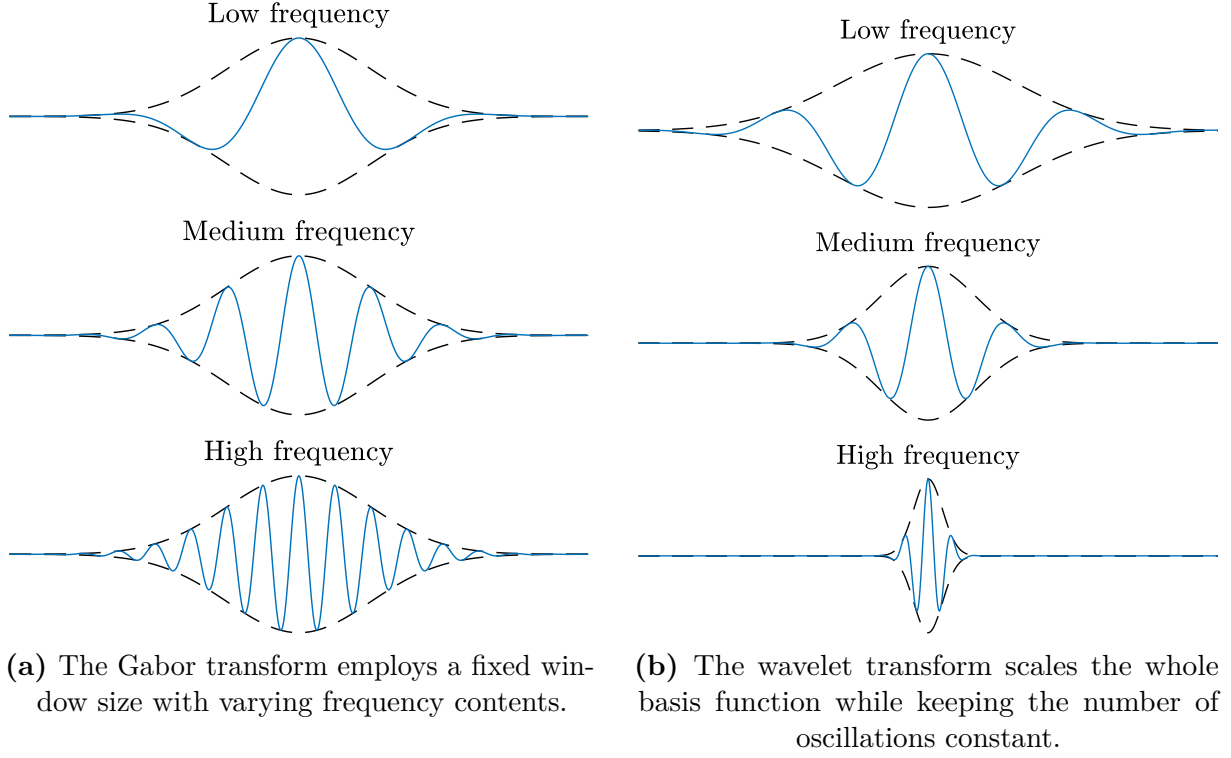
Low frequency

Low frequency

Medium frequency

Medium frequency

High frequency

High frequency

(a) The Gabor transform employs a fixed window size with varying frequency contents.

(b) The wavelet transform scales the whole basis function while keeping the number of oscillations constant.

Figure 3.2 The basis functions of the Gabor (a), and wavelet (b) transforms, adapted from [Goupillaud et al., 1984].

and Vetterli, 1991]:

$$\mathcal{W}_x(\lambda, \tau) = \int_{-\infty}^{\infty} x(t)\, \psi_{\lambda,\tau}^*(t)\, \mathrm{d}t \tag{3.10}$$

where * denotes complex conjugation.

In order to provide "an unfolding of the characteristics of a process in the scale-space plane" [Kumar and Foufoula-Georgiou, 1997], the *scalogram*[6] is employed to depict the distribution of the energy of the signal all over the time-scale (time-frequency) plane, i.e. the value of $|\mathcal{W}_x(\lambda, \tau)|^2$ is plotted for each λ and τ. Nevertheless this power spectrum has been found to be biased towards large scales, thus rectification of the coefficients through division by the scale is recommended in order to enable the comparison of the spectral peaks across scales [Liu et al., 2007].

The wavelet transform is seen to be linear, since it is an inner product between the signal and the wavelets [Farge, 1992]. For practical considerations, however, the CWT is best interpreted as a convolution operation, since this can be exploited to speed up the computation of the transform, by doing the convolution in the frequency domain, as a multiplication, via the FFT [Torrence and Compo, 1998]. However, the price to be paid

---

[6] In geophysical applications the scalogram is also known as the wavelet power spectrum [Torrence and Compo, 1998; Grinsted et al., 2004].
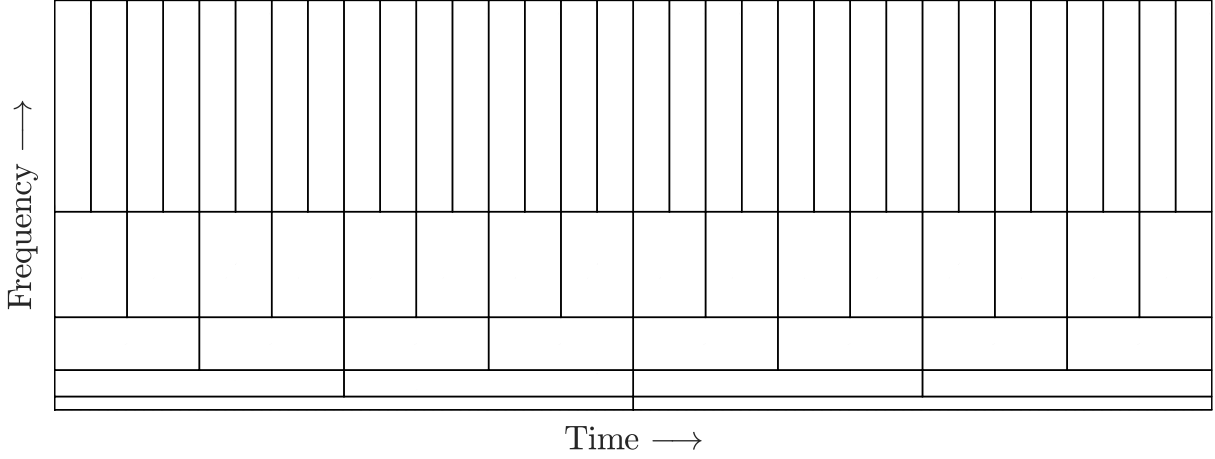
Figure 3.3 One possibility of tiling the time-frequency plane by means of the wavelet transform. Time resolution is better at higher frequencies, while frequency resolution is better at low frequency. All the rectangles are of equal area.

for this gain in time is the conformation to the circular world view of the DFT, inducing side effects with a range equivalent to the radius of the wavelet.

Coming back to the definition of the CWT from (3.10), the only condition imposed on a real or complex function $\psi(t)$ in order for it to be a wavelet is the so-called admissibility condition [Farge, 1992], which, when $\psi(t)$ is integrable, implies that it has zero mean:

$$\int_{-\infty}^{\infty} \psi(t)\, \mathrm{d}t = 0 \tag{3.11}$$

When looked at from a signal processing perspective, the condition (3.11) requires that the Fourier transform of the wavelet family $\psi_{\lambda,\tau}(t)$ establishes a bank of bandpass filters with constant ratio of width to center frequency, or constant-Q filters [Farge, 1992], providing for a logarithmic coverage of the frequency domain [Rioul and Vetterli, 1991], as pictured in figure 3.4.

At this point, by analogy with case of the Fourier transform, the following question must be raised: what are the implications of the linearity of the CWT when analysing non-linear phenomena? Can non-linear signals truly be represented by the new transform, or, as was the case of the STFT, will the confines of this linear technique also subject the results to the same mathematical artefacts without physical substrate?

Figure 3.5 attempts to provide an answer to this question, by illustrating side-by-side, the wavelet power spectrum of two signals, a pure sine wave on the left and a square wave on the right, both with a fundamental frequency of 6 Hz. For both time-frequency representations, each pixel denotes the amount of power through its color (see also binary logarithmic color-bar), at each time instant (abscissa) and at each frequency (ordinate).

The CWT in figure 3.5a renders a rather good description of the sinusoid, captured at an approximate frequency of 6 Hz. Most of the power is found to be concentrated near
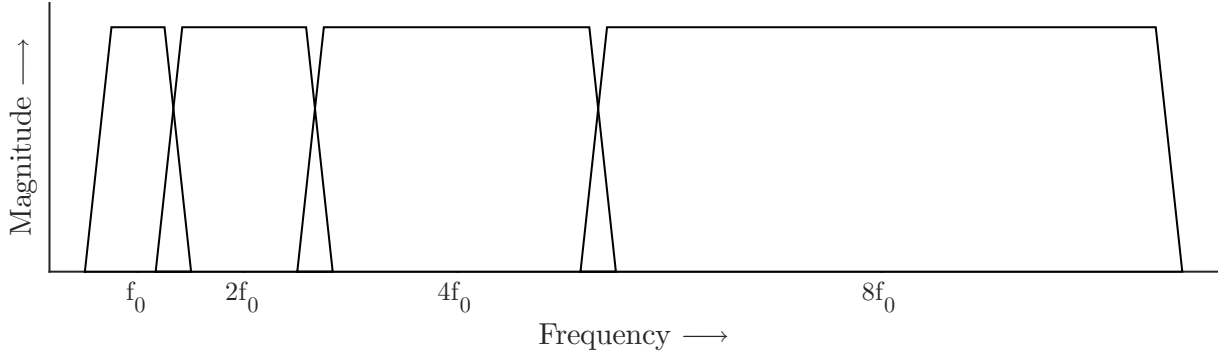
Figure 3.4 The wavelet transform can also be seen as logarithmically dividing the frequency plane, much like a bank of bandpass filters.

this frequency (light yellow hues), however spectral leakage smears the frequency bands up to 12 Hz (green and light blue tones). The CWT of the square wave in figure 3.5b contains many artefacts, due to the non-linear nature of the signal it tries to represent with a linear superposition of wavelets. This effect is most prominent at the time instants where the square waves switches between its two amplitude levels, as the CWT tries to represent these sharp discontinuities with the summation of wavelets of many different frequencies. Hence, another particularity of the wavelet transform is the fact that one must look in the high frequency region in order to identify discontinuities of components with low spectral order. Moreover, despite the fact that the square wave in figure 3.5b has a fundamental frequency of 6 Hz, its spectrum contains non-zero components at odd multiples of the fundamental, with diminishing power, from the 3$^{\text{rd}}$ up to the 9$^{\text{th}}$ harmonic, at 18 Hz, 30 Hz, 42 Hz and 49 Hz. As in figure 3.5a, spectral leakage smears power around the fundamental and its harmonics, a limitation rooted in the uncertainty principle from (3.8) — time resolution can be freely traded for frequency resolution, and vice-versa, but both cannot be arbitrarily increased at the same time.

Another noteworthy characteristic of figure 3.5 are the edge effects of the CWTs. Due to the wavelet not being completely localized in time [Grinsted et al., 2004], or, alternatively, due to the use of the FFT for computing CWT coefficients [Torrence and Compo, 1998], artefacts occur at the edges. Stated differently, because of the finite extent of the data being analysed the CWT coefficients get distorted by the discontinuity at the signal boundary, as the wavelet "spills over" the edge. The artefacts are more pronounced at low frequencies, owing to the fact that the wavelets have larger support at these scales, and gradually recede as frequency increases. To ensure that the influence of the edge artefacts is well represented, it is common practice to define the cone of influence (COI) as "the area in which the wavelet power caused by a discontinuity at the edge has dropped to e^{-2} of the value at the edge" [Grinsted et al., 2004]. In figure 3.5 the COI is marked by a white-out area; interpretation of this region is ambiguous at best.
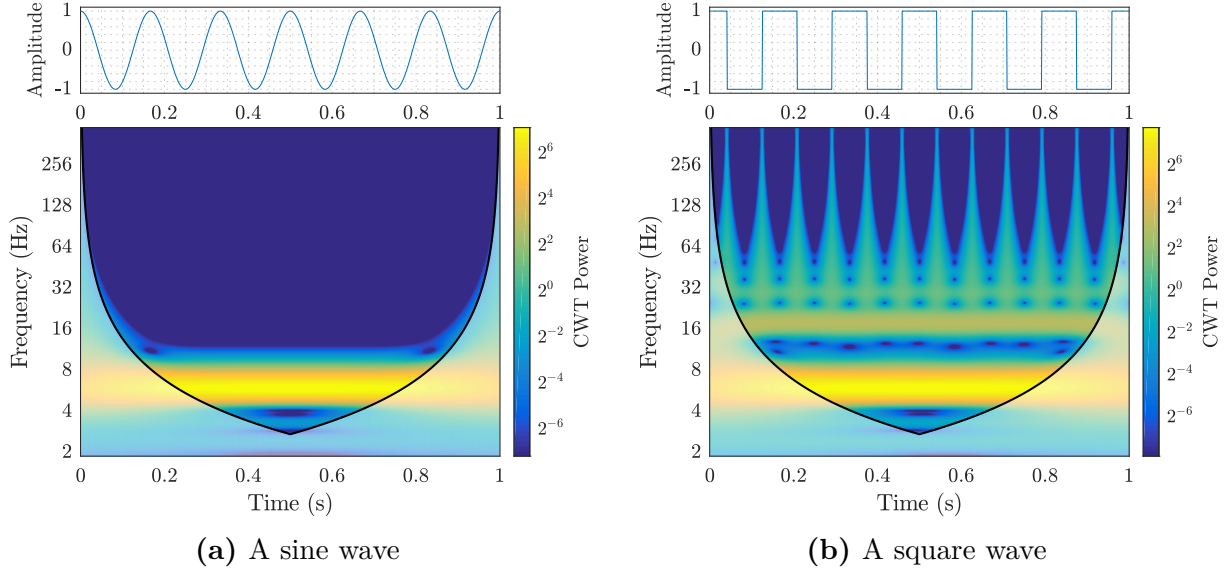
29

(a) A sine wave

(b) A square wave

Figure 3.5 Two signals with a fundamental frequency of 6 Hz (top), and the amplitudes of their CWT (bottom): a sine wave (a), and a square wave (b).

When employing wavelet analysis for geophysical applications, however, the choice of the wavelet is neither unique, nor arbitrary [Kumar and Foufoula-Georgiou, 1997]. Indeed, it would seem that a recurrent criticism of wavelet analysis is the arbitrary choice of the mother wavelet [Torrence and Compo, 1998]. When analysing a signal or process of interest, the choice among different categories and types of wavelet can seem a daunting task. Is one to prefer symmetric over asymmetric wavelets? Is orthogonality a factor? How smooth should the wavelet be? Are there any time or frequency domain constraints that play a role in the choice? These are all legitimate questions, with the selection of the wavelet ultimately depending on a combination of the above considerations; nevertheless "the appropriate choice should be made and justified on a case by case basis" [Kumar and Foufoula-Georgiou, 1997].

Numerous authors recommend the use of the complex-valued Morlet wavelet in conjunction with the CWT [Goupillaud et al., 1984; Grossmann and Morlet, 1984; Farge, 1992; Kumar and Foufoula-Georgiou, 1997; Torrence and Compo, 1998; Grinsted et al., 2004], which is defined as a plane wave of wave-number ω_0 modulated by a Gaussian envelope of unit width, as in (3.12).

$$\psi(t) = \pi^{-\frac{1}{4}} e^{-j\omega_0 t} e^{-\frac{t^2}{2}} \tag{3.12}$$

For the remainder of this study the CWT will imply the exclusive use of the complex Morlet function as a mother wavelet. Although it does not strictly verify the admissibility criterion from (3.11), taking $\omega_0 = 6$ in equation (3.12) yields an average that is close to zero within the computer round-off error [Farge, 1992]; this choice of parameters also provides

for a balanced trade-off between time and frequency resolution [Grinsted et al., 2004]. It can also be shown that the complex Morlet wavelet is analytic, i.e. it has no frequency components below zero. This is a desired feature, as it minimises the interference of positive and negative frequencies in the interpretation of results [Kumar and Foufoula-Georgiou, 1997].

In light of the previous discussion, and having found a shared framework which sets the stage for a comparison of the energy distribution onto the time-frequency plane of the short-time Fourier transform and the continuous wavelet transform, the advantages of the latter over the former have now become apparent. Owing to its changing time-frequency tiles, and empowered by the ability to reversibly trade frequency for time resolution, the CWT is found to be more revealing in rendering the characteristics of the analysed signal. The lead strength of the CWT is its ability to scrutinise non-stationary signals, being thus better suited to capture both periodic and transient features. Nevertheless, much like it was the case of the STFT, its linear nature might pose some difficulties in dealing with signals generated by non-linear processes. And, although the choice of an initial mother wavelet that would generate the basis of the decomposition is not fixed, this added flexibility is not a straightforward task, requiring some *a priori* insight into the features of the signal under analysis.

3.2 The Hilbert-Huang transform

The two previous sections have introduced the Fourier and the continuous wavelet transform, two data analysis techniques that yield a time-frequency energy distribution of the scrutinised signals. Although apparently different at a first glance, the inner workings of the two approaches are very similar, in the sense that they both try to decompose the data into a set of *a priori* chosen basis functions, by measuring the similarity between these functions and portions of the signal. The original information can then even be reconstructed as a linear superposition of the functions onto which it was projected, by weighing them with the previously determined coefficients of similarity. Yet, despite the fact that the transformation takes place within a well-established framework, the beforehand assumptions and constraints placed on the data by these methods introduce uncertainties in the interpretation of the results and in understanding the underlying physical phenomena [Wu et al., 2011]. Nevertheless, these two techniques have found widespread adoption in the earth sciences and geophysical data analysis fields [Labat, 2005]. However, when employing data analysis techniques the end goal is not finding out what mathematical properties the data may have, but rather revealing the physical traits embedded within [Huang and Wu, 2008].

Given the non-linear and non-stationary characteristics of the SSI [Zeng et al., 2013], handling such data issued from the non-linear interaction of physical processes, often also found under the influence of non-stationary external forcings calls for an adaptive data analysis approach [Wu et al., 2011]. The ideal analysis technique must make no *a priori* assumptions regarding the character of the data, i.e. neither linearity, nor stationarity should be presumed, since the nature of the processes that have generated the data is not usually known before the analysis is carried out. Adaptivity to the analysed data is also a desirable feature, i.e. letting the dataset itself decompose onto a set of basis functions determined by its local characteristic time scales, instead of a projection onto a predefined set of patterns. This ensures that the extracted components carry physical meaning, and that the influence of method-inherent mathematical artefacts on the rendered picture of temporal variability is kept to a minimum [Wu et al., 2011].

The aforementioned considerations are just some of the drivers behind the efforts of Huang et al., [1998] that have led to the development of an adaptive data analysis technique, with local scope, that enables the scrutiny of both non-stationary and non-linear signals — the Hilbert-Huang transform (HHT)[7]. The HHT is ideally suited for non-linear and non-stationary data and it adaptively decomposes any time-series into basis functions derived from the local properties of the data [Huang et al., 1998]. For the study at hand, the method is used to extract, depict and analyse the characteristic time-scales of variability of solar irradiance time-series. The data analysis method operates in the time domain and makes no *a priori* assumptions regarding the analysed dataset (stationarity or linearity). The method is also adaptive, letting the dataset decompose itself onto a finite number of locally derived data-driven basis functions [Wu et al., 2011], in contrast with the Fourier or wavelet transforms that impose a predefined set of trigonometric functions or wavelets for the decomposition. Further details on how the HHT compares to other spectral methods, the Fourier or wavelet transforms included, can be found in [Huang et al., 1998] and [Wu et al., 2011].

The transform consists of two separate steps. The first one, which is a key part of the method, called the empirical mode decomposition (EMD), adaptively decomposes the data into a sum of so-called intrinsic mode functions (IMFs), embedded oscillations that have a common time-scale. The second step consists in analysing the previously obtained IMFs by means of a Hilbert spectral analysis (HSA). In order to gain insight into the role played by the EMD and for a better understanding of the concept of IMF, a short review of Hilbert spectral analysis and of the concepts of instantaneous frequency and instantaneous amplitude are best introduced first.

---

[7] The term was coined by the National Aeronautics and Space Administration agency in the USA [Huang and Wu, 2008] and the acronym HHT is most probably a wordplay on the fast Fourier transform, i.e. HHT vs. FFT.

3.2.1   Hilbert spectral analysis

The use of Hilbert spectral analysis has been motivated by the need to better express non-stationarity and non-linearity [Huang et al., 1999; Huang and Wu, 2008; Huang et al., 2011]. As opposed to classical Fourier analysis, where frequency was defined for trigonometric basis functions spanning the length of the data with constant amplitude, the new approach extends the concepts to instantaneous amplitude and instantaneous frequency [Huang et al., 1998].

Thus, if $x(t)$ is a real-valued signal, its Hilbert transform can be written as:

$$\phi(t) = \mathcal{H}\left\{x\left(t\right)\right\}(t) = \frac{1}{\pi} P \int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} \, \mathrm{d}\tau \tag{3.13}$$

where P indicates the Cauchy principal value. In the frequency domain, the Hilbert transform of a real-valued signal can be expressed as:

$$\mathcal{F}\left\{\mathcal{H}\left\{x\left(t\right)\right\}\right\}(f) = -j\,\mathrm{sgn}\left(f\right) \cdot \mathcal{F}\left\{x\left(t\right)\right\}(f) \tag{3.14}$$

where $\mathrm{sgn}\left(\cdot\right)$ is the signum function.

The pair $\left\{x\left(t\right), \phi\left(t\right)\right\}$ can be used to construct the complex-valued analytic signal proposed by Gabor, [1946], as:

$$z(t) = \underbrace{x(t) + j\phi(t)}_{\text{Cartesian form}} = \underbrace{a(t)\, e^{j\theta(t)}}_{\text{polar form}} \tag{3.15}$$

or, alternatively, in the frequency domain:

$$
\begin{aligned}
\mathcal{F}\left\{z\left(t\right)\right\}(f) &= \mathcal{F}\left\{x\left(t\right)\right\}(f) + \mathrm{sgn}\left(f\right) \cdot \mathcal{F}\left\{x\left(t\right)\right\}(f) \\
&= \mathcal{F}\left\{x\left(t\right)\right\}(f) \cdot \left(1 + \mathrm{sgn}\left(f\right)\right) \\
&= \mathcal{F}\left\{x\left(t\right)\right\}(f) \cdot 2\,\mathrm{U}\left(f\right)
\end{aligned}
\tag{3.16}
$$

where $\mathrm{U}\left(\cdot\right)$ is the unit, or Heaviside, step function.

The polar form representation from (3.15) is then used to express the complex signal $z\left(t\right)$ in the amplitude modulation – frequency modulation (AM–FM) signal model, with the instantaneous amplitude given by:

$$a(t) = \sqrt{x^2(t) + \phi^2(t)} \tag{3.17}$$

and instantaneous phase derived as:

$$\theta(t) = \tan^{-1}\left(\frac{\phi(t)}{x(t)}\right) \tag{3.18}$$

The instantaneous frequency, i.e. the inverse of the local time-scale, is then just the first time derivative of the instantaneous phase:

$$f(t) = \frac{1}{2\pi} \frac{d\theta(t)}{dt} \tag{3.19}$$

The polar form in (3.15) provides insight into the local character of this representation as being the best local fit of $x(t)$ by a trigonometric function of time-varying amplitude and phase [Huang et al., 1998]. Using both the Cartesian and polar forms of the analytic signal, the original signal has the following representation within the AM–FM model:

$$x(t) = \mathcal{R}\left\{z\left(t\right)\right\} = \mathcal{R}\left\{a\left(t\right)e^{j\theta(t)}\right\} = \underbrace{a(t)}_{\text{AM}} \underbrace{\cos\theta(t)}_{\text{FM}} \tag{3.20}$$

As such, the role of the HSA is to decompose a signal of interest into two time-varying components, namely instantaneous amplitude and instantaneous frequency. The data can then be plotted as an amplitude (or energy, the square of amplitude) density distribution overlaid on the time-frequency space, called the Hilbert spectrum [Huang et al., 2011]:

$$H(f, t) = a\left(t\right)\cos\left(2\pi \int_0^t f\left(\tau\right)d\tau\right) \tag{3.21}$$

The time-integrated version of equation (3.21), the Hilbert marginal spectrum $h(f)$, is similar, but not identical to, the traditional Fourier spectrum, and it depicts "the accumulated amplitude (energy) over the entire data span [...] and offers a measure of the total amplitude (or energy) contribution from each frequency value" [Huang and Wu, 2008]:

$$h(f) = \int H(f, t)\,dt \tag{3.22}$$

However, in order to obtain meaningful instantaneous frequency values, the analysed data need to be the subject of some restrictions. A global domain condition is that the real part of its Fourier transform contain only positive frequencies [Huang et al., 1998]; the analytic signal already enforces this constraint [Gabor, 1946]. Yet, instantaneous frequency is a local feature of the data, and, since it implies that at any given time there is only one frequency value, the signal must also be mono-component [Huang et al., 1998; Huang and Wu, 2008]. As a direct consequence, the data also need to have zero average. To illustrate this on a simplified case, let $x(t)$ be the sum of a sine wave and a real constant:

$$x(t) = \sin(t) + \alpha, \qquad \alpha \in \mathbb{R} \tag{3.23}$$

Figure 3.6a shows the values of the function $x(t)$ for different values of the parameter α. The analytic signal, $z(t)$, is then derived according to (3.15). Figure 3.6b illustrates

(a) Original signal $x(t)$

(b) Instantaneous phase of $z(t)$

(c) Instantaneous frequency of $z(t)$
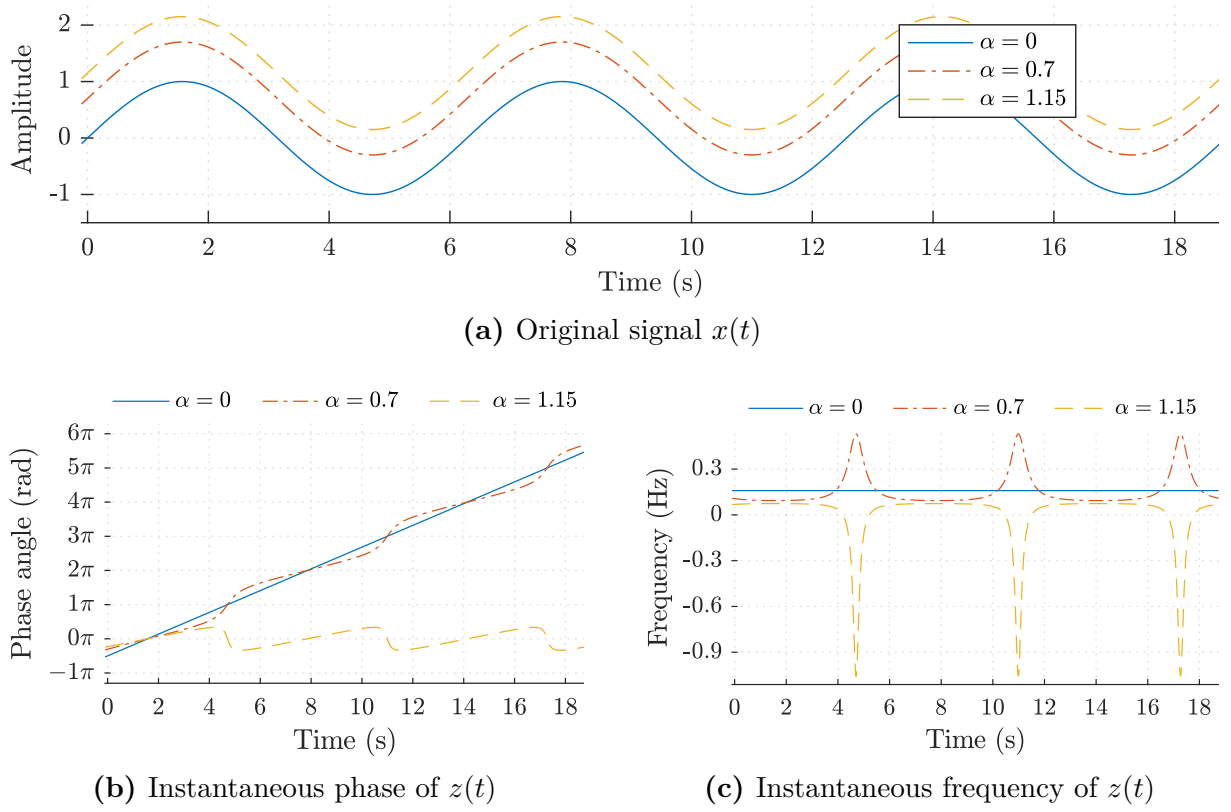
Figure 3.6 The signal from (3.23) for $\alpha \in \{0, 0.7, 1.15\}$ (a), and its instantaneous phase (b), and frequency (c). Adapted from [Huang et al., 1998].

the unwrapped instantaneous phase, $\theta(t)$, by using the definition from (3.18), while the instantaneous frequency, $f(t)$, computed according to (3.19), is pictured in figure 3.6c. It is interesting to notice that, even though the signals in figure 3.6a appear to have the same Fourier frequency contents, the only instantaneous frequency from figure 3.6c that has a meaningful physical interpretation corresponds to the signal with zero average. To make things worse, when $\alpha > 1$ the instantaneous frequency takes on negative values, even though the analytical signal has no negative spectral components [Gabor, 1946].

Thus, the recognition of the fact that instantaneous frequency can only be defined for signals that are locally symmetrical around their zero mean has led Huang et al., [1998] to introduce a new class of functions, known as intrinsic mode functions (IMFs)[8]. Defined based purely on its local properties, an IMF must satisfy two conditions:

- for the whole span of the data, the number of extrema and the number of zero crossings may differ at most by one;
- at any point of the data, the mean value of the envelopes defined by the local minima and the local maxima is zero.

---

[8] The term intrinsic mode function was adopted to represent the embedded oscillatory modes within the data [Huang et al., 1998].

While the adoption of IMFs still does not guarantee under all circumstances a perfect instantaneous frequency, it is, nevertheless, consistent with the physics of the scrutinised phenomena, even under the most adverse conditions [Huang et al., 1998].

3.2.2   The empirical mode decomposition

Although the previously introduced concepts of IMFs and instantaneous frequency are conceptually useful, real-world data are often complex, having multiple and often concurrent spectral components, and thus being unsuitable to direct Hilbert spectral analysis without first decomposing them into IMFs. One may incline to bypass this step with the risk of logically inconsistent results, such as having negative values for the instantaneous frequency of an analytical signal [e.g. figure 1(e)-(f) in Zappalà et al., 2016]. It was this precise need to extract the intrinsic oscillatory modes embedded in the data that has led Huang et al., [1998] to the development of the empirical mode decomposition (EMD).

The EMD is an adaptive method that decomposes data into a set of IMFs, operating directly in the temporal rather than in the frequency domain. The EMD uses an iterative process, akin to "sifting", to extract the embedded oscillatory modes, irrespective of the nature of the analysed signal, with the basis of the decomposition being *a posteriori* derived from the dataset itself [Huang et al., 1998; Huang and Wu, 2008].

The method can be described as follows: given a signal, $x(t)$, the local minima and the maxima are identified and then interpolated, e.g. via cubic splines[9], to form the upper, $U_x(t)$, and lower, $L_x(t)$, envelopes respectively. Let $\mathcal{M}\left\{\cdot\right\}$ be the operator which computes the local mean of its argument from the upper and lower envelopes:

$$\mathcal{M}\left\{x\left(t\right)\right\} = \frac{U_x\left(t\right) + L_x\left(t\right)}{2} \tag{3.24}$$

This local mean is then subtracted from the signal, yielding a proto-IMF:

$$h_1(t) = x(t) - \mathcal{M}\left\{x\left(t\right)\right\} \tag{3.25}$$

The first iteration of the sifting process is thus complete. By construction, this new signal, $h_1(t)$, should satisfy the definition of an IMF. In practice, however, this is often not the case, therefore the sifting continues iteratively, with the proto-IMF, $h_1(t)$, now taking the place of the original signal, $x(t)$, so that a deutero-IMF is obtained:

$$h_2(t) = h_1(t) - \mathcal{M}\left\{h_1\left(t\right)\right\} \tag{3.26}$$

---

[9] Cubic splines are the preferred interpolation method in the EMD [Rilling et al., 2003].

Further carrying out the procedure k times, the signal becomes

$$h_k(t) = h_{k-1}(t) - \mathcal{M}\{h_{k-1}(t)\} \tag{3.27}$$

and the iterations eventually stop when a local envelope symmetry condition is satisfied [Huang and Wu, 2008]. The sifting process then comes to an end and the first true IMF, designated as

$$c_1(t) = h_k(t) \tag{3.28}$$

will contain the highest frequency oscillation, or the finest scale, of the signal. Nevertheless, the residue

$$r_1(t) = x(t) - c_1(t) \tag{3.29}$$

may still have coarser features, thus the sifting procedure is repeated on the residue to extract IMFs of increasingly lower frequency

$$
\begin{aligned}
r_2(t) &= r_1(t) - c_2(t) \\
&\vdots \\
r_n(t) &= r_{n-1}(t) - c_n(t)
\end{aligned}
\tag{3.30}
$$

The decomposition stops when the residue becomes monotonic or has at most one extrema and no more IMFs can be extracted. A decomposition of the original signal into n elementary oscillatory functions and a residual trend is thus obtained:

$$x(t) = \sum_{i=1}^{n} c_i(t) + r_n(t) \tag{3.31}$$

The EMD process is formally described in pseudo-code in algorithm 1. Given the adaptive nature of the EMD, the IMFs $c_1, \ldots, c_n$ can be seen as the basis functions onto which the data are projected during decomposition and, by design, they have a well behaved Hilbert transform. Any two IMFs are locally orthogonal for all practical purposes, however, given the empirical nature of the method no theoretical guarantee can be provided. In practice, it is found that the relative difference between the variance of the input signal and the sum of variances of the IMFs (i.e. the spectral leakage) is typically less than 1%; only for extremely short data ranges does the leakage increase to 5%, comparable to that of a collection of pure trigonometric functions having the same data length [Huang et al., 1998]. The IMFs can also be shown to satisfy the envelope–carrier relationship, thus guaranteeing the existence of a unique true intrinsic amplitude function and of a unique phase function [Huang et al., 2013].

Algorithm 1 EMD

1: **function** EMD(x)

2:     $i \leftarrow 0$ ▷ Initialise IMF counter

3:     $r \leftarrow x$

4:     **while** $\neg$ IS_MONOTONIC(r) **do**   ▷ IMF extraction loop

5:         $i \leftarrow i + 1$ ▷ Increment IMF counter

6:         $h \leftarrow r$

7:         **while** $\neg$ IS_IMF(h) **do**   ▷ Sifting loop

8:             $[t_{min}, t_{max}, h_{min}, h_{max}] \leftarrow$ GET_EXTREMA(h) ▷ Identify extrema

9:             $L \leftarrow$ CUBIC_SPLINE(t_{min}, h_{min}) ▷ Construct lower envelope

10:            $U \leftarrow$ CUBIC_SPLINE(t_{max}, h_{max}) ▷ Construct upper envelope

11:            $h \leftarrow h - (L + U)/2$ ▷ Subtract mean of envelopes

12:        $c_i \leftarrow h$ ▷ Store IMF

13:        $r \leftarrow r - c_i$ ▷ Update residual

14:     **return** $c_1, \ldots, c_n, r$ ▷ Return IMFs and residue

---

After all the IMFs are extracted, what is left of the data, r, is called a trend or residue, which can no longer be considered as an oscillation at the span of the data. This trend can be interpreted as low-pass filtered version of the data [Moghtaderi et al., 2013].

By applying equations (3.19) and (3.20) to the IMFs $c_i(t), i \in \{1, \ldots, n\}$ from (3.31), the original data can now be written in terms of the AM–FM signal model as:

$$x(t) = \mathcal{R} \left\{ \sum_{i=1}^{n} a_i(t) \cdot e^{2\pi j \int_0^t f_i(\tau) \, \mathrm{d}\tau} \right\} + r_n(t) \tag{3.32}$$

which is essentially a sum of zero-local-mean oscillations having symmetric envelopes that are riding onto the EMD trend.

It is important to notice that, in the context of the EMD, the term *local* does not depend on the user, but is rather intrinsic to the data. This is best outlined by Rilling et al., [2003] when explaining the basics of the EMD:

"In fact, if we look at the evolution of a signal $x(t)$ between two consecutive extrema (say, two minima occurring at times t_- and t_+), we can heuristically define a (local) high-frequency part $\{d(t), t_- \leq t \leq t_+\}$, or local detail, which corresponds to the oscillation terminating at the two minima and passing through the maximum which necessarily exists in between them. For the picture to be complete, one still has to identify the corresponding (local) low-frequency part $m(t)$, or local trend, so that we have $x(t) = m(t) + d(t)$ for $t_- \leq t \leq t_+$. Assuming that this is done in some proper way for all the

oscillations composing the entire signal, the procedure can then be applied on the residual consisting of all local trends, and constitutive components of a signal can therefore be iteratively extracted."

Another topic that needs clarification at this stage is the so-called stoppage criterion [Huang and Wu, 2008] for the sifting operation, i.e. line 7 of algorithm 1. In the EMD, sifting has a two-fold purpose — to discard riding waves and to render the waves more symmetric. However, when iterated too many times the procedure can devoid the IMFs of physically meaningful amplitude variations [Huang et al., 1998]. Indeed, Wang et al., [2010] have proven that the IMF amplitude tends to become a constant when the number of sifting repetitions grows towards infinity.

When the EMD was first introduced, the original stopping condition proposed by Huang et al., [1998] was a Cauchy-like convergence test which required that the difference between two successive sifting results, SD_k, be smaller than some predetermined value, typically 0.2–0.3:

$$SD_k = \sum_{t=0}^{T} \left[\frac{|h_{k-1}(t) - h_k(t)|^2}{h_{k-1}^2(t)} \right] \tag{3.33}$$

Yet, the criterion defined by (3.33) was found to be rather unsatisfactory, for it does not guarantee that the function obtained is an IMF; it only requires that the squared difference between successive iterations be small [Huang and Wu, 2008]. Thus, a new stoppage criterion, termed S stoppage, was proposed by Huang et al., [1999], as follows: an IMF is found when the number of extrema differs from the numbers of zero-crossings by at most one, for S consecutive siftings. Huang et al., [2003] suggest values of $S \in \{4, \ldots, 8\}$ with 5 being most frequently used.

Furthermore, an alternative stoppage criterion was introduced by Rilling et al., [2003] in order to account for both the amplitude of the mean and the amplitude of the corresponding mode. Thus, if the mode amplitude is defined as the half-difference between the upper and lower envelopes

$$\mathcal{A}\{c_i(t)\} = \frac{U_{c_i}(t) - L_{c_i}(t)}{2} \tag{3.34}$$

an evaluation function is introduced as the ratio between the mean and the mode amplitude:

$$\delta(t) = \left| \frac{\mathcal{M}\{c_i(t)\}}{\mathcal{A}\{c_i(t)\}} \right| \tag{3.35}$$

The sifting procedure then goes on until $\delta(t) < \theta_1$ for some duration of the signal $(1 - \alpha)$, with the condition changing to $\delta(t) < \theta_2$ for the remaining part; the authors recommend adopting the following values for the parameters: $\alpha \approx 0.05$, $\theta_1 \approx 0.05$ and $\theta_2 \approx 10 \cdot \theta_1$.

Another stoppage criterion, is closely related to how the sifting loop controls the filter character of the EMD. On the one hand, an infinite number of sifting iterations

would asymptotically approach the result of the Fourier decomposition (i.e. constant amplitude envelopes) [Wang et al., 2010]. On the other hand, several studies performed on time-series of pure noise [Flandrin et al., 2004; Flandrin and Gonçalves, 2004; Wu and Huang, 2004] have shown the decomposition behaves like an adaptive wavelet-like dyadic filter bank, i.e. it has a constant-Q filter-bank character (see figure 3.4), if the number of sifting iterations is kept small, around 10, which also assures maximum component separation and minimum leakage [Wu and Huang, 2010]. Moreover, by adopting this fixed number of sifting stoppage criterion in the decomposition, it has been shown that the EMD has the same computational complexity as the FFT, although with somewhat larger constant factors [Wang et al., 2014]. Requiring only $\mathcal{O}\left(N \log\left(N\right)\right)$ operations and generating approximately $\log_2(N)$ IMFs, with N being the number of samples in the data, the method is thus suitable for use even in an operational setting. This stoppage criterion is currently the recommended one for practical applications [Wu and Huang, 2009; Wu and Huang, 2010].

One of the drawbacks of the original EMD is that it may introduce a phenomenon known as "mode mixing". This is the manifestation of oscillations with dissimilar time-scales in the same IMF, or the presence of oscillations with similar time-scales in different IMFs. A workaround was proposed by Wu and Huang, [2009] in the form of the ensemble empirical mode decomposition (EEMD). The idea was to run the decomposition over an ensemble of copies of the original signal to which white Gaussian noise has been added, with the final result obtained by averaging.

Although the EEMD improved the mode mixing problem, the different sums of signal and noise produced different numbers of modes, making the final averaging somewhat difficult. Added to this, the reconstructed signal still contained some residual noise, and thus was not identical to the original. To overcome this situation, Torres et al., [2011] have proposed another version of the EMD, the complete ensemble empirical mode decomposition with adaptive noise (CEEMDAN). This method also decomposes the white noise into modes, along with the signal, such that at each stage of the decomposition a particular noise is added and a unique residue is computed to obtain each mode. However, the modes of CEEMDAN still contain some residual noise and sometimes spurious modes appear in the early stages of the decomposition.

The next version of the method, the improved complete ensemble empirical mode decomposition with adaptive noise (ICEEMDAN), overcomes these issues by fixing the signal to noise ratio for all stages of the decomposition process [Colominas et al., 2014]. The approach allows for an exact decomposition of the data, i.e. the sum of all IMFs, including the trend, reconstructs the original time-series, and is more robust with respect to noise. The ICEEMDAN method is the one used in this study, and will be introduced next. The following notation is used for describing the algorithm. Let $\mathcal{E}_k\left\{\cdot\right\}$ be the

operator which produces the kth mode obtained by EMD, i.e. $\mathcal{E}_k\{x\} = c_k$, following the notation from algorithm 1. Also let $w_{(p)}$ be a set of p realizations of white Gaussian noise having zero mean and unit variance ($\mu_{w_i} = 0$, $\sigma_{w_i} = 1$, $\forall i = 1, \ldots, p$), and let $\langle \cdot \rangle$ denote the operation of averaging throughout these realizations. The ICEEMDAN can then be sketched according to algorithm 2, following Colominas et al., [2014].

As can be seen from the pseudo-code sketch, the ICEEMDAN first estimates the local mean by averaging throughout the ensemble of realization and then subtracts it from the original signal. By proceeding in this way, the amount of noise present in the modes is reduced. Furthermore, in order to reduce mode mixing, the method does not directly employ white noise but rather the ith IMF of the white noise realizations, $\mathcal{E}_i\{w_{(p)}\}$, to extract the ith mode of the signal. The parameter σ_w is the standard deviation of the white Gaussian noise and controls the desired signal to noise ratio between the added noise and the specific residue to which the noise is being added. The presently recommended value for this parameter is $\sigma_w = 0.2$ [Wu and Huang, 2009; Colominas et al., 2014], as detailed and numerically confirmed in [Colominas et al., 2012].

Algorithm 2 ICEEMDAN

1: **function** ICEEMDAN(x, p, σ_w)

2: $\quad i \leftarrow 1$ $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ ▷ Initialise IMF counter

3: $\quad r \leftarrow \left\langle \mathcal{M}\left\{ x + \sigma_w \dfrac{\sigma_x}{\sigma_{\mathcal{E}_1\{w_{(p)}\}}} \mathcal{E}_1\{w_{(p)}\} \right\} \right\rangle$ $\qquad$ ▷ Compute the first residual

4: $\quad c_i \leftarrow x - r$ $\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ ▷ Store the first IMF

5: $\quad$ **while** $\neg$ IS_MONOTONIC(r) **do** $\qquad\qquad\qquad$ ▷ IMF extraction loop

6: $\quad\quad i \leftarrow i + 1$ $\qquad\qquad\qquad\qquad\qquad\qquad$ ▷ Increment IMF counter

7: $\quad\quad h \leftarrow \left\langle \mathcal{M}\left\{ r + \sigma_w \sigma_r \mathcal{E}_i\{w_{(p)}\} \right\} \right\rangle$ $\qquad$ ▷ Compute the residual

8: $\quad\quad c_i \leftarrow r - h$ $\qquad\qquad\qquad\qquad\qquad\qquad$ ▷ Store IMF

9: $\quad\quad r \leftarrow h$ $\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ ▷ Update residual

10: $\quad$ **return** $c_{1,\ldots,n}$, r $\qquad\qquad\qquad\qquad$ ▷ Return IMFs and residue

---

3.3 Case study: CWT and HHT analysis of SSI data

With the introduction of the time-frequency analysis techniques now complete, the remaining chapters of this work will focus exclusively on the temporal variability of the SSI. However, to provide a guiding thread, the current section will present a step-by-step approach to the HHT analysis of the characteristic time-scales of variability, as embedded in a long-term time-series of SSI. Furthermore, the time-frequency portrait rendered by the HHT will be compared to the one obtained by the CWT, highlighting the commonly shared and contrasting features of the two representations. The data under scrutiny in this

section are a decennial time-series of daily means of SSI, recorded at the BSRN station in Carpentras, France (see CAR dataset in chapter 2). The aims and intents of this last section of chapter 3 are to ascertain what type of information about the temporal variability of the data the two spectral representations may convey.

3.3.1 The HHT in action

As previously mentioned, the first step of the HHT is to employ the EMD for decomposing the analysed time-series into intrinsic oscillatory components. To illustrate its workings, the eight IMFs of the CAR time-series are presented in figure 3.7 in the order they were obtained, from top to bottom. As EMD operates in time-domain, the IMFs have the same temporal support as the original data and, by construction, an average of zero and symmetrical upper and lower amplitude envelopes. It can be observed in figure 3.7 that as the decomposition progresses, the time-scale of the IMFs increases, i.e. the intrinsic oscillations are getting further spaced apart with increasing IMF number.

Another view of this is brought by figure 3.8, where the power spectral density (PSD) and a Fourier estimate of the mean period of each IMF are plotted. To aid the reader, the colours used to portray the individual IMF spectra are the same as for the time-domain representation from figure 3.7.

The spectral shapes of the IMF1,...,IMF5 are similar in form, i.e. bell curves, and their median periods roughly follow a dyadic scale, i.e. doubling with increasing IMF number as : 3.1 days $\rightarrow$ 7.4 days $\rightarrow$ 15.7 days $\rightarrow$ 32.0 days $\rightarrow$ 54.9 days. This doubling of the time-scale for these first five IMFs is the hallmark output of an efficient dyadic filter (see figure 3.4). Subsequently, it will be shown that this dyadic repartition is pertinent to identifying deterministic signals from random realizations of quasi-stochastic background processes. This finding is even more interesting, since the median periods have been estimated with a Fourier-based method, which measures period *globally* over the whole time range of the IMFs. By opposition, a measure of *local* period in the Hilbert sense is a much better estimate, since it has an uncertainty as low as a quarter wavelength of temporal resolution with respect to the average time-scale of the IMF [Huang et al., 2009].

Next, IMF6 from figure 3.7 can be readily identified by its median period of 364.9 days with the seasonal cycle. It also accounts for the most prominent visual feature in the original time-series (second panel from top (CAR) of figure 2.1a), its peaks and troughs denoting summer and winter, respectively. Another feature of this sixth IMF is its sharp peak of relatively high power, as opposed to the bell curve shape and the frequency spread of the previous five components. By contrast, no distinct visual features stand out in the PSD of IMF7 and IMF8, with median Fourier periods of 655.5 and 1370.5 days, respectively; nevertheless their time-domain representations reveal that they have very low amplitudes. Lastly, another point worth noticing with respect to figure 2.1a, is that
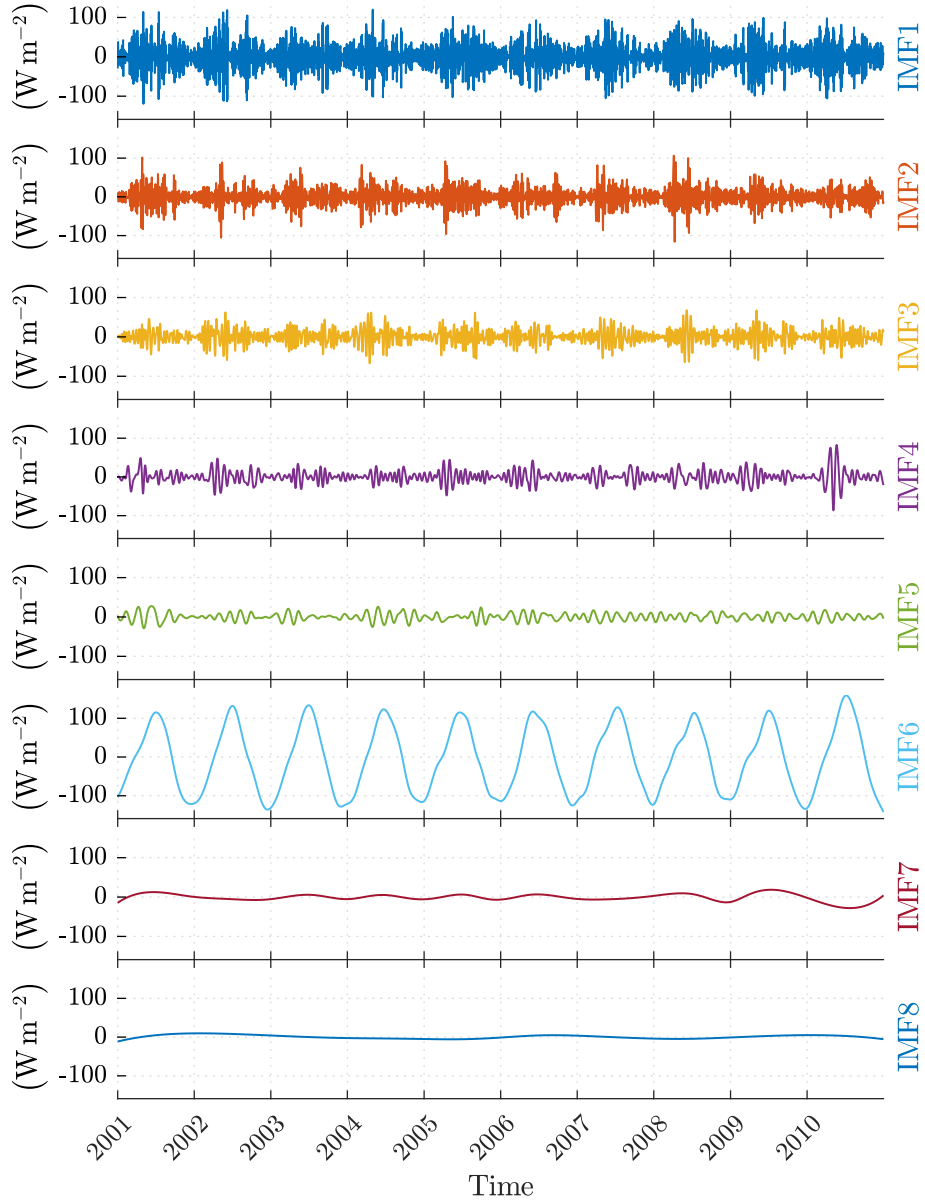
Figure 3.7 The eight IMFs obtained by decomposing the CAR time-series, from top to bottom IMF1,. . .,IMF8. The panels plot SSI (ordinate) versus time (abscissa). Time markers on the horizontal axes indicate January 1$^{\text{st}}$ of the corresponding year. The zero-centred oscillatory nature of the modes can be clearly seen. Also apparent is the local time-scale increase with mode number.

the first five modes of CAR seem to exhibit amplitude excursions in accordance with the phase and amplitude of IMF6, i.e. their amplitudes are greater during summer than during winter, in agreement with the yearly cycle.

After having extracted all the IMFs, all that is left of the original time-series is a residual, or EMD trend, that cannot be considered an oscillation at the time span of the original data (10 years). Stated otherwise, the trend can be thought of as an *offset* onto which all the zero mean embedded oscillatory components are riding — as per equation
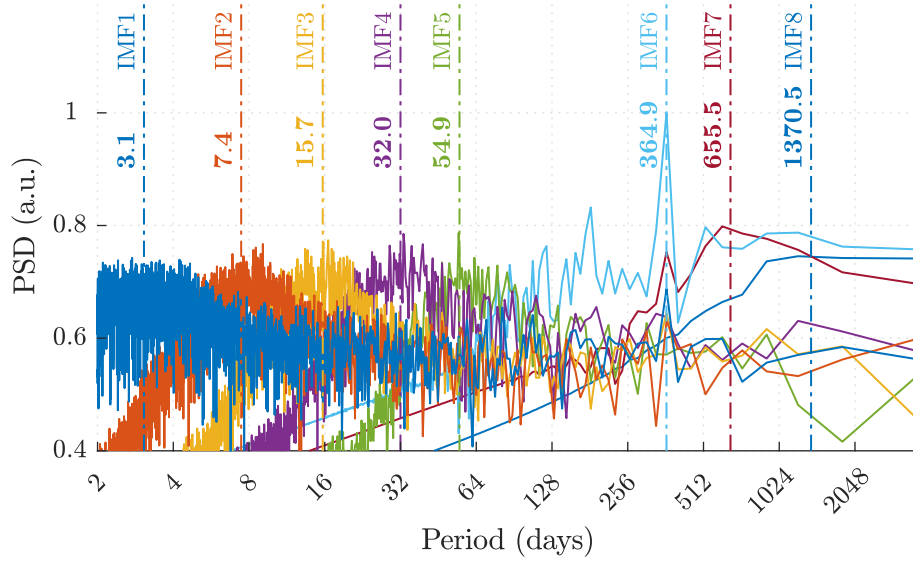
Figure 3.8 The power spectral density (PSD) of the eight IMFs for CAR (solid line) on a logarithmic scale normalized with respect to the power of the highest spectral peak. The period, or inverse frequency, runs on the abscissa in binary logarithm. The individual spectra are shown in the same colours as the IMFs from figure 3.7, from left to right: IMF1,. . .,IMF8. The Fourier estimates of the median periods, marked along the dash-dotted lines, are seen to increase with mode number. Notable features are the prominent spectral peak of IMF6 at ~ 365 days corresponding to the yearly cycle and the apparent dyadic repartition of the time-scales for IMF1,. . .,IMF5.

(3.32) — or as a low-pass filtered approximation of the original data. This can be clearly seen in figure 3.9, where the trend for CAR, plotted in blue, has a non-zero mean and is devoid of the oscillations of the IMFs depicted in figure 3.7. To graphically convey this, the yearly mean of the SSI, as described by Morel et al., [2010], is plotted in red alongside the EMD trend, in figure 3.9, whence it can be seen that the two curves follow each other closely, with the EMD trend taking on only slightly higher values.

Once the empirical mode decomposition is completed, the second and last step of the HHT consists in the Hilbert spectral analysis of the previously obtained IMFs. Each IMF and its Hilbert transform are used to construct a complex analytic signal, described by the AM–FM model. This decomposition into two time-varying parts corresponding respectively to instantaneous amplitude and instantaneous frequency is of interest for the purpose of this study. It enables the identification, in a time-varying sense, of how much power (i.e. the square of amplitude) occurs at which time-scale (i.e. the inverse of frequency).

Figure 3.10 provides a visual guide to this concept, by illustrating the AM–FM decomposition of IMF5 for the CAR time-series. The top panel (IMF5) of the figure reproduces the mode function, which is also the real part of the analytic signal from equation (3.15). The amplitude of the latter (AM), given in equation (3.17), which is the
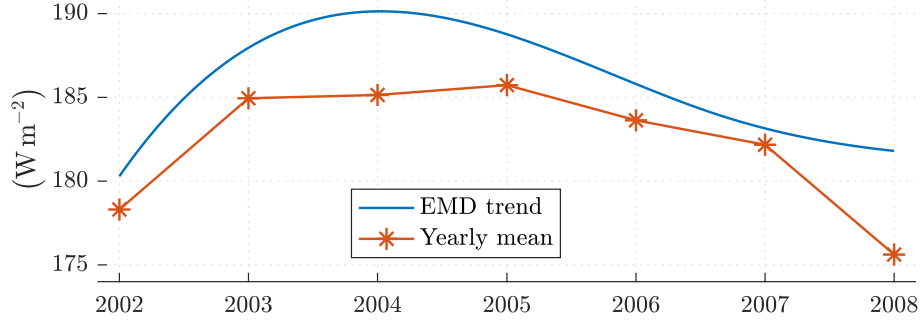
Figure 3.9 Comparison, for CAR, between the EMD trend and the yearly mean from Morel et al., [2010].

envelope of the original signal, is then extracted and plotted in the middle panel. Unlike the Fourier decomposition, the amplitude is not a constant, but rather a time-dependent function. Next, by removing the AM component from the signal through simple division, the frequency modulation component is obtained, i.e. the complex exponential in equation (3.15); the real part of this component is plotted in the bottom panel (FM). The FM is similar to a trigonometric function, but with a phase argument that unlike the Fourier transform is not a constant but a time-dependent function, as seen from equation (3.18). The local frequency (and its inverse, the local time-scale of the signal) is then just the first temporal derivative of this phase, as defined in equation (3.19). Owing to their time-varying character, the amplitude and frequency are usually encountered in the literature under the terms instantaneous amplitude, and instantaneous frequency, respectively.

The square of the instantaneous amplitude and the instantaneous frequency of the IMFs can then be used to represent the data as an energy density distribution overlaid on the time-frequency space, as per equation (3.21). This representation, called the Hilbert energy spectrum is defined[10] as "the energy density distribution in a time-frequency space divided into equal-sized bins of $\Delta_t \times \Delta_f$ with the value in each bin summed and designated as $a^2(t)$ at the proper time, t, and proper instantaneous frequency, f" [Huang et al., 2011].

An example of Hilbert spectral representation is given in figure 3.11 (left panel) where the CAR time-series is shown as an energy density distribution over-imposed on a time-frequency space as in equation (3.21). The abscissa is the same as for the original data, in the CAR panel from figure 2.1a, providing the means to temporally resolve the various features of the spectrum. The ordinate, having a logarithmic scale this time, indicates the local time-scale, or period, expressed in days. Lastly, colour encoding is used to logarithmically express the amount of power that the time-series contains at a certain point in time and at a certain period or scale. The colour bar at the top of the Hilbert energy spectrum provides a scale assigning adimensional numerical values to the colour

---

[10] For notation coherence, the symbol used in the original text to denote instantaneous frequency, ω, was replaced with f.
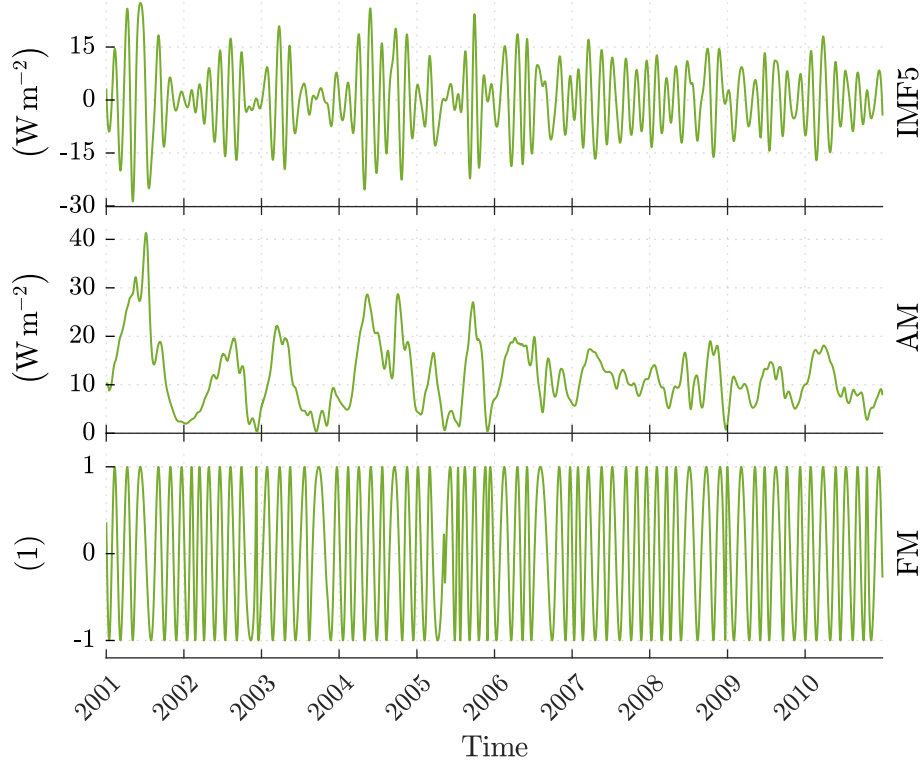
Figure 3.10 Hilbert spectral analysis of the fifth IMF of the CAR time-series. The intrinsic mode function (IMF5, top panel) is the product of its constituent slowly-varying amplitude modulation part (AM, middle panel) and of its rapidly-changing frequency-modulation component (FM, bottom panel). Time markers on the abscissa denote the beginning of the corresponding year.

encoding. As such, each pixel in the Hilbert spectrum is identified by three attributes – color, abscissa, and ordinate – through which it denotes the local power (color, log-scale) of the corresponding time-series, at a certain time (abscissa), and at a certain time-scale (ordinate, log-scale). For the sake of the readability, the spectrum is binned in time, scale, and colour space and has been smoothed. Hence, some aliasing may occur. Some features may be represented as continuous lines while others are rendered as point-like, especially where rapid frequency modulation takes place, such as in the high-frequency bands.

Interpretation of Hilbert spectral features at data boundaries must be done with care due to possible oscillations of the spline interpolants used in the EMD (see algorithm 1). This effect is similar to the COI in the wavelet transform [Torrence and Compo, 1998]. With the EMD edge effects are usually contained within a half-period of a component at data boundaries [Wu et al., 2011]. In figure 3.11 this region has been whitened out.

The plot in the right panel of figure 3.11 is the Hilbert marginal spectrum, or the time-integrated variant of the image at its left, indicating the amount of power contained in the time-series at each scale or period, expressed in dB. This time agnostic representation is comparable, but not identical to, the Fourier spectrum of the same time-series. It should be once again emphasized that the Hilbert marginal spectrum is obtained from *local*
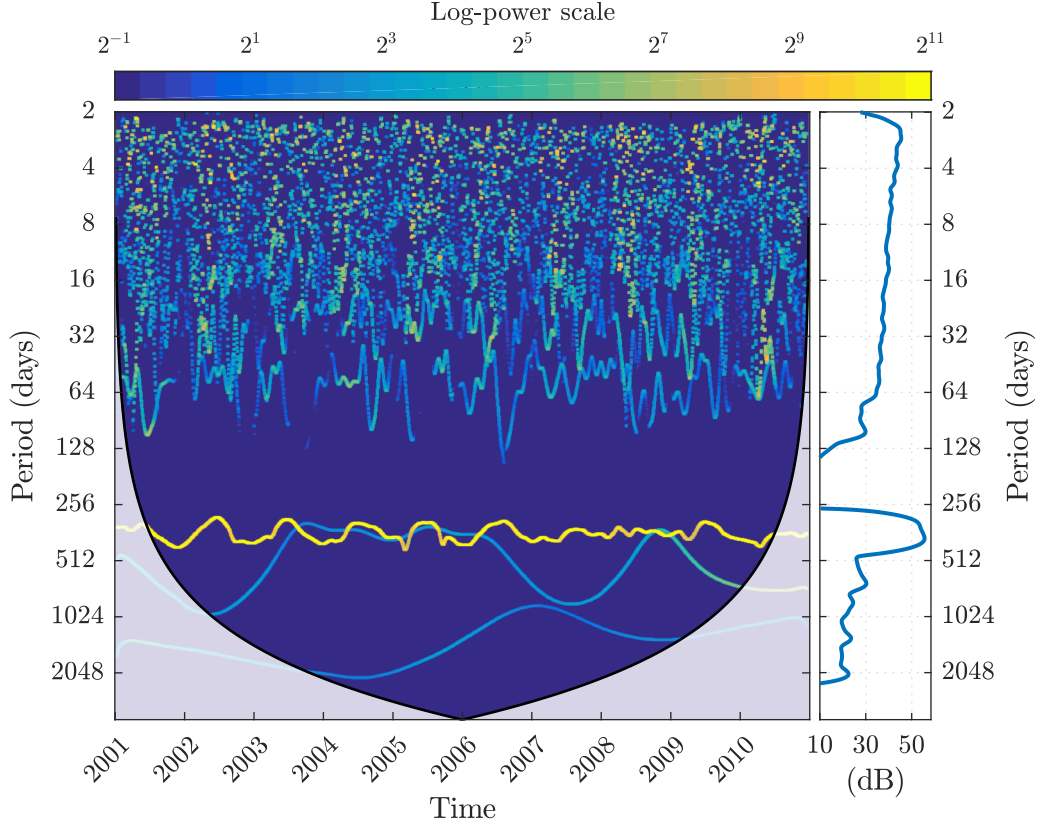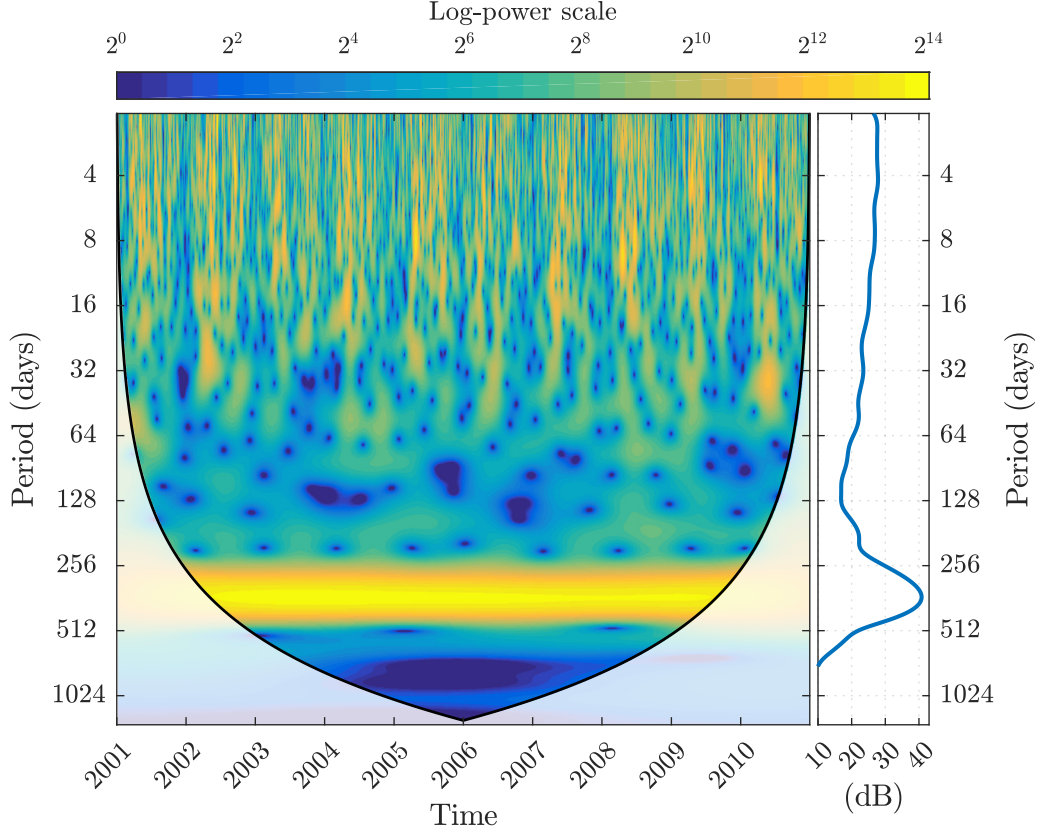
Figure 3.11 The Hilbert spectrum (left panel) of the ten year time-series of SSI under study, spanning 2001 through 2010. Pixel colour encodes power (logarithmic scale colour bar at the top) at each time (abscissa) and each scale (ordinate). Time markers on the abscissa denote the start of the corresponding year. The white-out area indicates the regions where edge effects become significant. The Hilbert marginal spectrum in the panel on the right indicates the amount of power at each scale, expressed in dB.

features of the data, its components having instantaneous amplitude and instantaneous frequency, as opposed to the *global* outlook of the Fourier spectrum whose constituents have constant amplitude and constant frequency throughout the whole domain.

The Hilbert spectrum depicted in figure 3.11 shows that the most energetic component is found oscillating around the one year period, in the form of a yellow trace, which clearly exhibits frequency modulation, i.e. with respect to 365 days, slight downward period shifts can be observed, followed by upward shifts. Also of notice is the fact that the domain of scale modulation of this yellow trace is somewhat restrained, as can be inferred from the support of this feature on the marginal spectrum where it accounts for the greatest power peak. Furthermore, two low power components can be identified as having periods between 4 years and 1 year, with two distinct spectral peaks showing on the marginal spectrum at about 1200 and 800 days, respectively. The interpretation of these peaks is ambiguous since most of the power of these components lies in the COI area. Another feature of the spectrum that stands out is the lack of power in the 2–3 months to one

47

year band, which is best viewed on the marginal spectrum. The EMD does not detect any spectral features in this region of the Hilbert spectrum and, owing to sparsity, assigns a negligible amount of power in this band. Another feature is the band of quasi-constant power in the high-frequency region, between 2 days (the Nyquist limit) and 2-3 months. The power of this plateau decreases with increasing scale. Modulation in amplitude of this band by the yearly cycle is readily apparent, and is most visible for the years 2005 and 2007 where the blue hues alternate with green-yellow ones.

3.3.2 The CWT in action

The continuous wavelet transform was also applied to the CAR time-series of SSI, producing the wavelet power spectrum depicted in figure 3.12. The axes of the wavelet power spectrum are analogous to the axes of the Hilbert energy spectrum in figure 3.11, except for the colour coding, which now spans a different range. The white-out regions that run along the edges of the figure indicate the COI that is excluded from the analysis. The global wavelet spectrum in the right panel is the time-integrated version, i.e. line-by-line sum, of the wavelet spectrum and indicates the amount of power at each scale, expressed in dB.

The wavelet power spectrum in figure 3.12 reveals that the most energetic component of the SSI time-series is confined between scales of approximately 256 and 512 days, as indicated by the thick orange power-band near the bottom of the plot. In CWT, spectral smearing and leakage occur because of the finite support of the analysing basic wavelet [Huang et al., 1998]. Power "spills over" to neighbouring scales and is assigned to this wider orange band on the plot instead of a distinct trace. This is also apparent in the support of this feature in the global wavelet spectrum. Is is also interesting to note, that because of the linear nature of the CWT, non-linearities of the yearly cycle are also identifiable in the green band at the scale of about 180 days (the second harmonic of the fundamental yearly cycle), which is dotted with blue patches that indicate low power occurring during this period. Furthermore, the grouping of yellow-orange "filaments" at scales of between 64 and 2 days could also be interpreted as high-order harmonics. However, the same feature may as well be indicative of amplitude modulation of these bands by the yearly cycle [similar to Liu et al., 2007, fig. 1], since this cycle can be inferred just from the alternation of blue-green and orange-yellow hues in this band. Power in this high-frequency band decreases with increasing scale, as can be inferred from the global wavelet spectrum, where a minima is also noticeable at around 128 days. This minima, centred in a larger depression, is a consequence of the low-powered regions occurring in the 150 to 64 days band in the left panel, denoted by dark blue patches, most notably between the second half of 2003 and 2005, the second half of 2005 and 2007, and from the second half of 2009 onward. These dark blue patches correspond to minimal levels of correlation between

Figure 3.12 The CAR wavelet power spectrum (left panel) of the ten year time-series of SSI under study, spanning 2001 through 2010. Pixel colour encodes power (logarithmic scale colour bar at the top) at each time (abscissa) and each scale (ordinate). Time markers on the abscissa denote the start of the corresponding year. The white-out area indicates the regions where edge effects become significant. The global wavelet spectrum in the right panel is the time-integrated version, i.e. line-by-line sum, of the wavelet spectrum and indicates the amount of power at each scale, expressed in dB.

the signal and the analysing wavelets at those scales and temporal localizations. A final remark must be made here by reiterating that owing to the overly redundant nature of the CWT, even after rectification, the global spectrum is best interpreted *qualitatively* and not *quantitatively.*

Conclusion   This chapter has provided a brief overview of two popular time-frequency signal processing methods: the STFT and the CWT, and has introduced the adaptive data-driven paradigm of the HHT. Through a case study, it has been shown that both the CWT and the HHT are able to capture the characteristic time-scales of variability of the data, i.e. the CAR time-series of the BSRN dataset. However, the HHT is slightly better suited for the task, in the sense that it is able to pick up more subtle features, such as the yearly period modulation, while also yielding a somewhat more sparse representation. This makes it a useful tool for the analysis of the temporal variability of the surface solar radiation. Nevertheless, both transforms provide a time-frequency-energy representation

of the CAR data, and reveal that the quasi-annual oscillations have the highest energetic contribution. A particular feature of the analysed data consists in a power depression in the spectrum in the 2–3 months to one year band. Lastly, a slightly slanted plateau between 2 days and 2-3 months can also be identified, whose power declines towards lower scales.

CHAPTER 4

The characteristic variability observed in measurements of the SSI

H AVING introduced the Hilbert-Huang transform as a useful time-frequency analysis technique for the purposes of the study at hand, the present chapter is concerned with employing the method for revealing and analysing the intrinsic temporal scales of the variability found in high quality ground measurements of the SSI. The data under scrutiny here consist of the decennial time-series of daily means of the global SSI received on a horizontal surface issued from different BSRN ground stations around the world, as described in section 2.1. First, the embedded oscillations of the data, roughly sorted by ranges of increasing time-scales, are extracted by the EMD. Next, HSA is applied to obtain an amplitude modulation – frequency modulation representation of the data. The time-varying nature of the characteristic time-scales of variability, along with the variations of the signal intensity, is thus revealed. The physical and statistical significance of these characteristic modes of variability is also investigated. To this end, a novel adaptive null-hypothesis is employed, in order to discriminate between the different features of the data, those that have a deterministic origin and those being realisations of various stochastic processes. Although the high-frequency components are apparently random in nature, nevertheless, a non-null, statistically significant rank correlation is evidenced between their amplitude envelopes and the yearly cycle. The eventual implications of these findings on the modelling and forecast of the SSI are lastly discussed.

4.1   HHT analysis of BSRN data

The EMD was used to decompose the four BSRN datasets (see figure 2.1a) into the constituent intrinsic mode functions. The IMFs obtained from the CAR time-series,

depicted in figure 3.7 in chapter 3, have already served as an illustrative example on the operation of the algorithm. The IMFs for the other datasets (not shown) are very similar and will be discussed in due time. It must be noted that, like CAR, the BOU time-series is decomposed into 8 IMFs, while the PAY data has 9 IMFs and 10 IMFs are obtained for TAT. Besides the IMFs, for each time-series the decomposition also yields a residual, or trend (also not shown, except for CAR — see figure 3.9). The EMD residual is an adaptively low-pass filtered version of the time-series and cannot be thought of as an oscillation at the 10-year time-span of the data. Since this work focuses mostly on the characteristic scales of temporal variability, these EMD trends along with their statistical significance and physical meaning do not fall within the scope of the study; for such discussion, see e.g. [Franzke, 2012].

By revisiting the Fourier spectra of the IMFs for CAR in figure 3.8 it can be observed that, owing to its median period of 364.9 days, IMF6 can be unambiguously associated with the yearly cycle, as dictated by the orbital parameters of the Earth–Sun system. IMF6 also accounts for the most prominent visual feature in the original data (figure 2.1a, CAR panel), with its maxima and minima denoting summer and winter, respectively. Further evidence is brought by the spectral shape of IMF6, distinguished by a sharp peak that has the largest power in figure 3.8. Also noteworthy is that IMF6 seems to modulate the previous five IMFs, i.e. the variability of the analytic amplitudes of the first five IMFs is coherent with the phase of IMF6, a phenomenon that is most visually distinguishable in figure 3.7 for the year 2007. Finally, the last two components, IMF7 and IMF8, having median periods of 655.5 days and 1370.5 days (figure 3.8), respectively, are seen to exhibit only deviations from zero of low amplitude in their temporal representation. Moreover, these fluctuations in amplitude are the greatest at the edges of the signal for both IMF7 and IMF8. Interpretation of these components should, thus, be done with care, since edge effects for the EMD are known to be usually contained within a half-period of a component at data boundaries [Wu et al., 2011], i.e. approximately one year for IMF7 and two years for IMF8.

However, it should be reiterated that this Fourier representation is based on a global approach and, as such, the frequency estimates do not take into account the local features of the data. Consequently, Hilbert spectral analysis was applied to the IMFs of the four BSRN time-series, yielding an AM–FM representation. The resulting Hilbert frequency components illustrate the spectral contents of each time-series in terms of its individual mode functions, as shown in figure 4.1, where the distribution of the local time-scale of each IMF is conveyed by means of box plots. The box plots of the instantaneous amplitudes of each IMF are given in figure 4.2. These box plot representations are nevertheless partial, as they only account for the distribution of time-scales and corresponding amplitudes
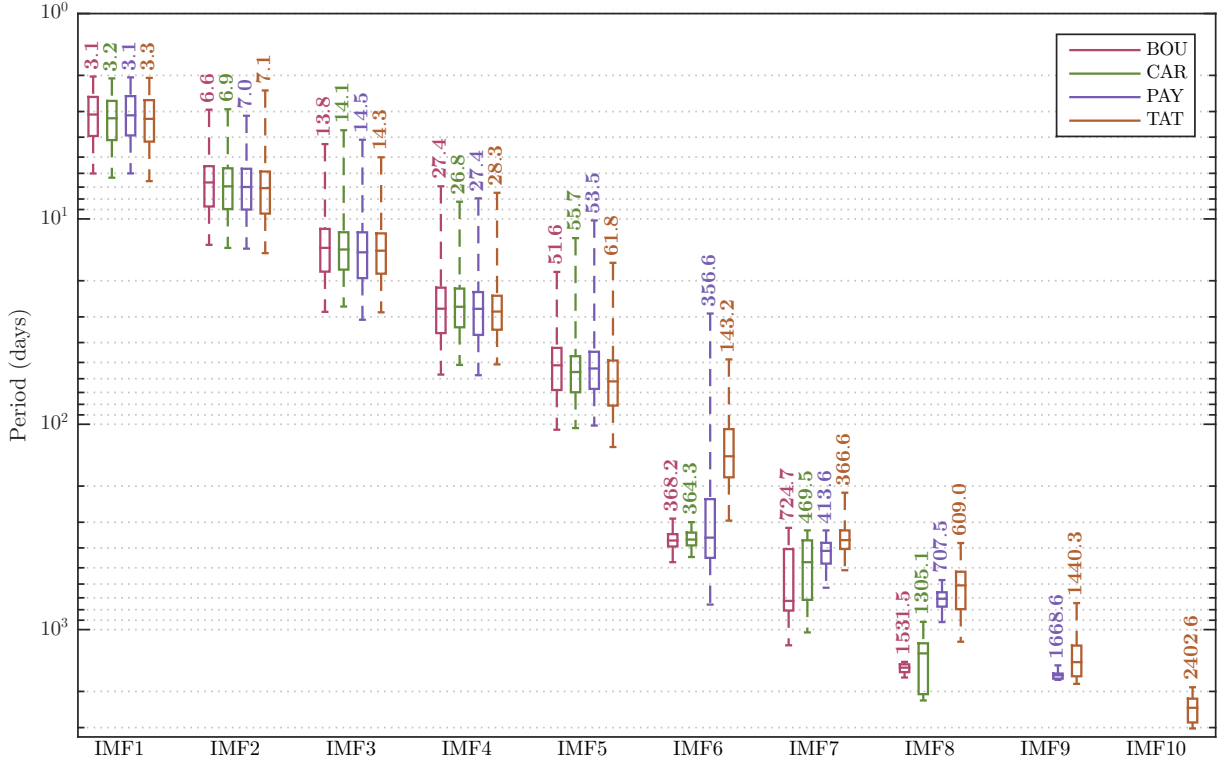
Figure 4.1 Box plot of the instantaneous time-scales of the IMFs for the four BSRN stations. The top and the bottom edges of the boxes represent the first ($Q1$) and, respectively, the third ($Q3$) quartiles. The bars inside boxes denote the second quartile ($Q2$), i.e. the median. The whisker length is at most 1.5 times the interquartile range, i.e. $1.5 \times (Q3 - Q1)$, hence the whiskers roughly correspond to ± 2.7 standard deviations, or equivalently $\backsim 99\%$ of the data, assuming normal distribution. The median for each box is expressed numerically above the lower whiskers. Outside the interval of the whiskers, instantaneous time-scales are omitted as outliers.

within and throughout the modes and does not take into account the temporal localization and variability of the events seen.

From the distribution of time-scales in figure 4.1 it can be observed that, for all time-series, IMF1,...,IMF5 have very similar median periods, that approximate the dyadic sequence: 3.5 days → 7 days → 14 days → 28 days → 56 days. Moreover, besides the notable similarity among the medians of these modes, for all the datasets both the interquartile ranges and the total ranges of these first five modes exhibit approximately the same variability. Added to this, IMF6 for BOU, CAR, and PAY, and IMF7 for TAT, whose median periods are respectively 368.2, 364.3, 356.6, and 366.6 days, can clearly be associated to the yearly cycle given by the revolution of the Earth around the Sun. This yearly component is very similar for BOU, CAR and to a lesser extent TAT, with an interquartile range that is concentrated around almost the same median value, the only minor difference being the slightly extended range for TAT of 200 days as opposed to 300 days for the other two. The PAY yearly mode differs from those of the other
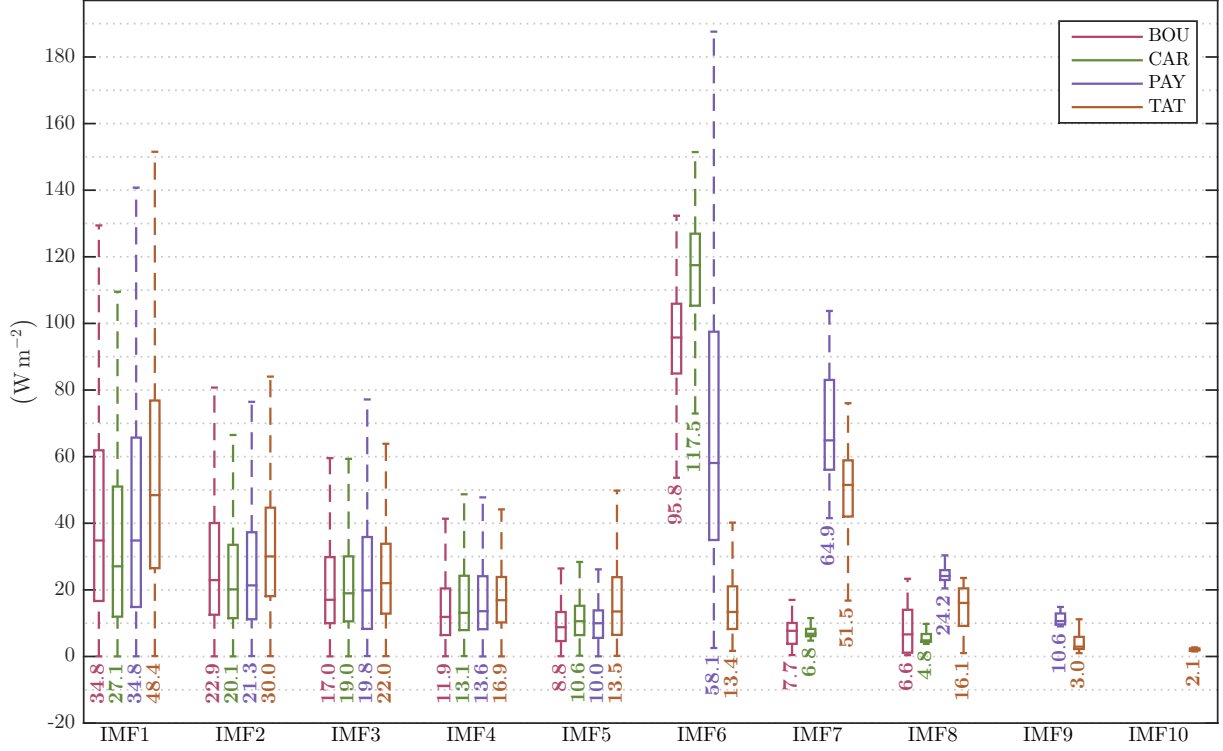
Figure 4.2 Box plot of the instantaneous amplitudes of the IMFs for the four stations. The bottom and the top edges of the boxes represent the first ($Q1$) and, respectively, the third ($Q3$) quartiles. The bars inside boxes denote the second quartile ($Q2$), i.e. the median. The whisker length is at most 1.5 times the interquartile range, i.e. $1.5 \times (Q3 - Q1)$, hence the whiskers roughly correspond to ± 2.7 standard deviations, or equivalently $\sim 99\%$ of the data, assuming normal distribution. The median for each box is expressed numerically below the lower whiskers. Outside the interval of the whiskers, instantaneous amplitudes are omitted as outliers.

stations, its interquartile range and foremost its range being much larger, the latter even overlapping the interquartile ranges of IMF5 and IMF4. This is a result of the mode mixing phenomenon described in section 3.2.2 that may arise with the EMD, i.e. the coexistence or mixing of different time-scales in the same IMF, mainly related to the intermittence of signal and to noise [Huang et al., 2003]. Nevertheless, the spectral part of IMF6 which overlaps IMF5 and IMF4 has very low power, as will be subsequently shown, thus this phenomenon does not influence the validity of the analysis. With this in mind, one notes that for BOU and CAR no spectral components are present in the 100 days to 300 days band. Furthermore, TAT is the only dataset that has a transitional mode of 143.2 days median period in between the first five IMFs common to all stations and the yearly cycle.

At this point, the Hilbert frequency distribution of the IMFs for CAR may be compared to the Fourier one from the PSD in figure 3.8. As already mentioned, the Hilbert estimates are based on *local* features of the data, and thus are more accurate than the Fourier ones when applied to non-stationary signals. This can be seen especially when comparing the

range of the first five high-frequency IMFs, which is upper bounded to about 100 days in figure 4.1, whereas in the PSD from figure 3.8 the spectra of the same components are seen to span the whole time-scale range. This also holds for IMF6, which has very narrow Hilbert period range, whose Fourier analogue is the sharp peak in the PSD of the same mode. Similar statements can be made for IMF7 and IMF8. To sum up, it is found that while the Hilbert period distributions of the modes have compact supports, the Fourier representations of the same components span the whole frequency range. Nevertheless, most of the power in the Fourier PSD is assigned to a frequency band that closely corresponds to the Hilbert range. Owing to the global nature of the Fourier transform, however, additional spectral coefficients are needed to provide a complete mathematical description of the data.

Resuming the discussion of the IMF time-scales from figure 4.1, it can be observed that the low-frequency, i.e. greater than one year, variability of the data, trend notwithstanding, is assigned into slightly overlapping (within the same time-series) IMFs that span the spectrum starting from the one year mark. For BOU and CAR time-series, there are only two modes extending beyond one year. First, IMF7 can be seen to span approximately the same range for both these stations, from about one year to slightly more than three years. For BOU however, the interquartile range and especially the median period is shifted towards higher periods, i.e. 724.7 days vs. 469.5 days for CAR. The last modes IMF8 of these stations are very different, with a very narrow range around the median of 1531.5 days for BOU, and a range of 900 days to over 2000 days and median of 1305.1 days for CAR. For the PAY data, the low-frequency components have narrower spectral support, with two IMFs (IMF7 and 8) that cover the band from 1 to 2.5 years and median periods of 413.6 and 707.5 days, and the IMF9 around 4.5 years ($\backsim$1668 days) with a very narrow range. It must also be noted that IMF7 for BOU, CAR and PAY have the same lower end support, and that the couple (IMF7, IMF8) of PAY taken together somehow emulates IMF7 for BOU and CAR. Lastly, TAT is the only dataset whose low-frequency variability is expressed by three components, IMF8,. . . ,IMF10, with mean periods of 609 days, 1440.3 days and 2402.6 days. While the first quartile of IMF9 coincides with the upper range of IMF8, the upper range of IMF9 is slightly below the lower range of IMF10, hence the last two modes do not overlap at all. By its range, IMF8 of TAT approximates IMF7 for BOU and CAR, but there is no proximity in terms of median or interquartile range. Similarly, IMF9 of TAT resembles IMF8 of CAR in terms of range, but their medians are not in close agreement and their interquartile ranges even less so.

With the scrutiny of the these low frequency components, the discussion of the time-scale distribution of the IMFs from figure 4.1 can now be concluded. However, as previously mentioned, this particular illustration, although instructive, is incomplete. First, the box plot representation does not take the instantaneous variations of frequency into account,

but renders global aggregates instead – much like the traditional Fourier methods, with the interquartile range spread in addition. This is done on purpose, with the intent of making it easier for the readership not accustomed to the HHT to create analogies with the more familiar methods (e.g. Fourier analysis, wavelets, etc.). Second and last, this particular representation is totally devoid of any information pertaining to the local amplitude, or power, or variance, of the data. With these consideration in mind, the Hilbert spectra of the data, a true time-frequency representation for non-linear and non-stationary data, will be discussed next.

Figure 4.3 depicts the Hilbert spectra of the four BSRN time-series: BOU, CAR, PAY and TAT. These spectra are similar to the one already introduced in figure 3.11, with the exception that the Hilbert marginal spectrum is no longer expressed in decibels.

The BOU Hilbert spectrum from the top left panel of figure 4.3 exhibits a high-frequency feature between 2 days and $\smile$ 100 days, which corresponds to the first five IMFs of the time-series. The instantaneous time-scales of these modes overlap (figure 4.1), hence the appearance on the Hilbert spectrum of a continuum instead of distinct bands. This spectral feature has relatively low power, that decreases with increasing period, as can be inferred from the sloped dent in the marginal Hilbert spectrum corresponding to this region. In the 2 days to 32 days band, amplitude modulation by the yearly cycle can be inferred from the periodic change in color, with yellow-green tones, occurring mostly during the high irradiance regime of summer, that turn blue during the winterly minima. Next, in the band between 100 and 300 days, a gap in the spectrum is apparent, as can also be inferred from the lack of support in this region for any of the BOU IMFs in figure 4.1. The yellow trace, corresponding to IMF6, exhibits frequency modulation around the one year period, seen as oscillations in the range of 300 to 450 days, which is also the support of this mode in the box plot of IMF time-scales. The colour of this IMF indicates that it has the highest power of all the components, as can also be inferred from the large peak on the marginal spectrum. The corresponding time-scale fluctuations are centred in 365 days, and frequency modulation is greatest during 2003 through 2005. From 2006 onwards, however, frequency modulation is less pronounced – perhaps capturing the low solar activity around the 2008 minimum in the eleven year cycle solar cycle [Hathaway, 2015]. The final two low-frequency, blue-green traces on the spectrum correspond to IMF7 and IMF8. For IMF7, mode mixing is apparent through the occasional sharing of the yearly time-scale band with IMF6, between mid-2003 and 2005. IMF7 has such low power that it fails to leave an imprint on the marginal spectrum and it seems to suddenly spring into existence during summer 2003, i.e. it has negligible amplitude during the first two and a half years. IMF8 starts out in light-green hues and slowly vanishes around 2007. This last mode for BOU has a median period of 1500 days (see figure 4.1), with most of its
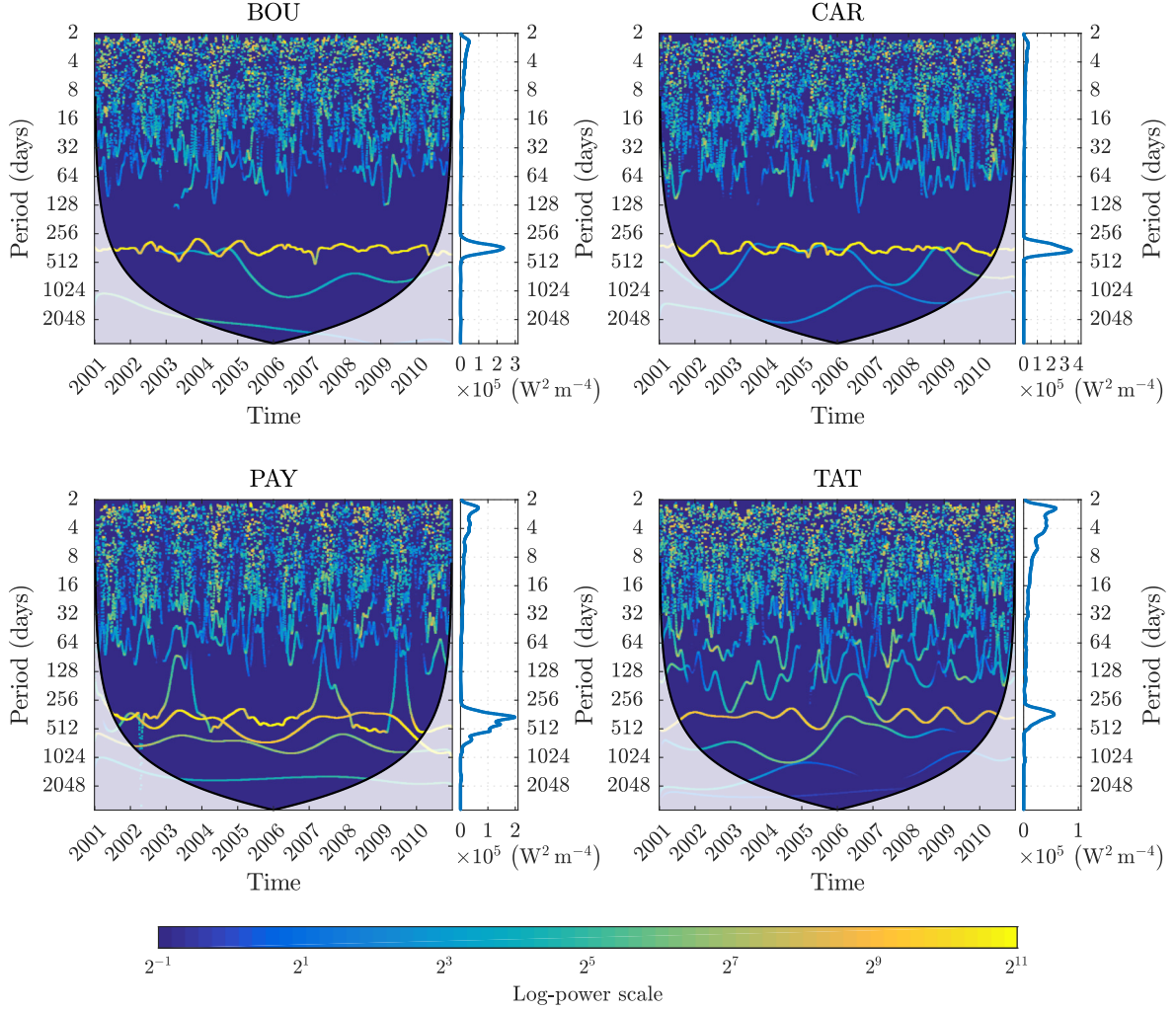
Figure 4.3 The Hilbert energy spectra of the 10-year BSRN time-series, spanning 2001 through 2010; clockwise from top left: BOU, CAR, TAT, PAY. Pixel colour encodes power (logarithmic scale colour bar at the bottom) at each time (abscissa) and each scale (ordinate). Time markers on the abscissa denote the start of the corresponding year. The white-out area indicates the regions where edge effects become significant. The Hilbert marginal spectra in the panels on the right indicate the amount of power at each scale.

power lying within edge effect territory; interpretation of this feature is thus ambiguous at best.

Referring to the CAR Hilbert spectrum, depicted in the upper right panel of figure 4.3, some features can be identified. A high-frequency plateau between 2 days and $\backsim$ 100 days is notable, corresponding to first five IMFs of the time-series. As can be seen in figure 4.1, the supports of these modes overlap, thus the appearance on the Hilbert spectrum of a contiguous plateau, as is the case for BOU. Also similar to BOU, the power of this feature is low, manifested by a slight indentation on the marginal Hilbert spectrum for this band. A cyclic shift in color can also be observed in this band, which points to an amplitude modulation mechanism linked to the seasonal cycle – darker blue tones during winter turn

57

yellow-green during summer. Similar to BOU, between 100 and 300 days a gap in the spectrum is again manifest, as hinted at by the time-scales of the CAR IMFs in figure 4.1, which do not cover this band. IMF6, in the form of the yellow trace oscillating around the one year period, in the range of 300 to 450 days, resembles the same mode for BOU. This component has the greatest power, denoted by the large peak on the marginal spectrum. It exhibits frequency modulation around 365 days, with shifts towards greater frequencies taking place predominantly during 2002 through 2006. Between 2007 and 2009, however, the frequency modulations are less pronounced – corresponding to a period when solar activity is at a minimum. The last two large time-scale, blue-green traces denote IMF7 and IMF8. The seventh exhibits some mode mixing with IMF6, between the end of 2003 and the first half of 2006, and shortly again before 2009. These two last components have such low power that they fail to leave a mark on the marginal spectrum.

In the Hilbert spectrum for the PAY data, shown in the lower left panel of figure 4.3, the contiguous high-frequency plateau between 2 days and ⌣ 100 days is also present, unsurprisingly, since the first five IMFs of the time-series closely resemble those for the CAR data. The power in this band, however, is slightly greater than for BOU or CAR, especially considering the predominance of yellow hues in the 2 − 4 days region that corresponds to the first and second IMFs. This is also apparent when looking at the marginal spectrum, where a distinct peak at this time-scale can be clearly made out. The amplitude modulation phenomenon in phase with the seasonal variations, previously identified in the BOU and CAR data, is even more pronounced in the PAY spectrum; once again dark blue tones that occur during low insolation in winter turn green and even yellow during the high irradiance regime of summer. Here too, between 100 and 300 days a gap in the spectrum is also apparent, with the notable exception of the mode mixing phenomena associated with IMF6 that occur during 2003, 2007 and 2009. As previously shown in figure 4.1, the lower range of the frequency distribution of the sixth mode overlaps the high frequency plateau, which is portrayed by the three jutting spikes from the yearly band into the sub-100 days plateau on the PAY spectrum. Since these protruding filaments have such low power that they leave no imprint on the marginal spectrum, their most probable origin can be attributed to some sort of numerical artefacts. The yearly variability of the data can be, unsurprisingly, identified with the sharp peak at roughly 365 days on the marginal spectrum. In terms of the Hilbert time-frequency-energy representation, however, the seasonal cycle cannot be attributed to one sole component. This is due to mode mixing as seen in figure 4.1 where the range of IMF7 completely overlaps that of IMF6. It could be argued that IMF6 should represent the "true" seasonal cycle, as indicated by the median of its frequency distribution, however for 2003 and 2007 the Hilbert spectrum reveals that whenever IMF6 extends its tendrils into the high-frequency plateau and drastically reduces its power, IMF7 seems to "pick up the slack" by reaching into the vacated one year band.

During 2008, and briefly during spring 2004, it is found that the two components trade places altogether, with IMF6 assuming lower values that IMF7 on the frequency scale. The last two low-frequency components, IMF8 and IMF9, can be seen here too to have relatively low power – both only just manage to make a very slight indentation on the marginal spectrum. For IMF9, the frequency spread in the Hilbert representation is in good agreement with its narrow support from figure 4.1.

The Hilbert spectrum for the TAT data, shown in the lower right panel of figure 4.3, also shows the contiguous high-frequency plateau between 2 days and $\backsim$ 100 days, owing to the first five IMFs of the data closely resembling those for BOU, CAR and PAY data. However, the power found in this band is much greater than for either BOU, CAR, or PAY, especially in the $2-4$ days region denoting the first two IMFs. For these time-scales of TAT, the marginal spectrum shows a peak that is even greater than the one associated with the yearly cycle. This feature is also evident in figure 4.2, where the upper whisker of the amplitude of IMF1 is seen to extend beyond 150 W m^{-2}, whereas it is less than 80 W m^{-2} for annual cycle depicted by IMF7. As a result, the slanting of the high-frequency plateau is clearly visible on the marginal spectrum, with a somewhat lesser slope that for the other stations. The amplitude modulation phenomenon in phase with the seasonal variations, also identified for the other datasets, is present here too, although to a lesser extent than for PAY. Unlike the previous datasets, no gap can is evident in the spectrum, owing to a sixth IMF that covers the region between 50 and 300 days, as shown in figure 4.1. The yearly variability of the time-series is denoted by the spectral peak at roughly 365 days on the marginal spectrum. In the Hilbert time-frequency representation the seasonal cycle is attributed to the dark yellow trace of IMF7, which can be seen in figure 4.1 to have a median time-scale of 366.6 days. Mode mixing in the yearly band is nevertheless apparent between the second half of 2005 through 2007, when IMF8 approaches the 365 days mark while IMF7 protrudes below 256 days. The last two low-frequency components, IMF9 and IMF10, have reduced power and fail to leave a mark on the marginal spectrum.

So far, all time-series have been shown to share a high-frequency constituent between 2 days and 100 days composed of five IMFs with mean periods following a dyadic sequence, and an IMF around 365 days that captures the yearly variability. For BOU, CAR and PAY, a low power region can be found in the 100 days to 300 days band. Beyond the one year time-scale, the low-frequency variability in the 1.5 years to 6 years band is captured by another two (BOU and CAR) or three (PAY and TAT) components. The TAT data is the only time-series that has an IMF in the low power band between the high-frequency feature and the yearly cycle (median period 143.2 days).

These features of the BSRN SSI datasets are now discussed in terms of their intrinsic temporal scales of variability, and of their physical statistical meaning.

4.2   The intrinsic time-scales of variability of the SSI

Firstly, the median periods of the IMFs composing the high frequency band are revisited. It has been shown in figure 4.1 that they follow a dyadic repartition, that approximate the series dyadic sequence: 3.5, 7, 14, 28, 56 days. Such a doubling in frequency in IMFs has been previously reported in astrophysical and geophysical signals. When investigating three independent datasets of satellite observations of the (exo-atmospheric) total solar irradiance (TSI), Lee et al., [2015] consistently find a similar dyadic scale progression of modes at 13.5, 27, and 54 days, statistically significant within the 95% level, that correspond to the 27-day solar rotation period and its (sub-)harmonics. Kolotkov et al., [2015] find intrinsic periodicities having an average of 25^{+7}_{-2} and 44^{+10}_{-5} days within five different solar proxy signals. The mean periods and the associated error bars in sub-/superscript, estimated at the half-level width of the corresponding probability histogram, were obtained by analysing the sunspot area for the whole Sun, and for the northern and southern solar hemispheres taken separately, the 10.7 cm radio flux intensity, and the helioseismic frequency shift. Compelling as it may seem, nevertheless, the imprint of a solar rotation signature on ground measurements of the SSI is highly unlikely, as it would imply the existence of hitherto unknown physical mechanisms in Earth's atmosphere [Thuiller, 2015]. The amplitudes of the IMFs of the TSI time-series and those of the IMFs in the SSI data differ at times by two orders of magnitude, e.g. compare figure 3.7 with figure 1 in [Lee et al., 2015]. If the solar rotation signature were to be seen in the IMFs of the SSI this would require the existence of amplifying processes. Stott et al., [2003] and Lockwood and Fröhlich, [2007] have studied the possibility of such a mechanism, and have concluded that, irrespective of the mechanisms invoked and of the amplification of the solar variability, for the past decades solar forcing is only a minor contributor and thus not able to account for most of the global warming observed in the second half of the twentieth century, which could be better explained by an increase in greenhouse gases. Further proof will be provided subsequently, this time from a signal theoretical point of view, in support of the view that it is unlikely that the solar rotation signature is captured in measurements of the SSI.

Secondly, in the 100 days to 300 days band, two of the stations, BOU and CAR, do not exhibit any variability. For PAY, the support of yearly IMF6 protrudes in this region, although its first quartile rests well below the 200 days mark. As shown in figure 4.3, the power of the portion of this IMF that extends into the high-frequency range is very small. Hence, while not totally devoid of spectral features, this band contains a negligible amount of power. A distinct mode is present at TAT in this band, whose median period of 143.2 days somehow seems to continue the dyadic sequence of the previous five modes. These findings are important for the modelling and forecasting of the SSI, as follows. On the one

hand, any model attempting to reconstruct or forecast the SSI for BOU and CAR should not contain any power in this band, or at least filter it out. For TAT, on the other hand, such models should ensure that the 100 days to 300 days region is not a spectral void. In section 4.5, evidence will be presented that the spectral band spanning from two days to 300 days seems to be composed mostly of random realisations of stochastic background processes, which can be modelled following, e.g. [Flandrin and Gonçalves, 2004; Rilling et al., 2005; Welter and Esquef, 2013; Kolotkov et al., 2016].

Thirdly, the median periods detected around the one year mark in all the datasets can be explained by the movement of revolution of the Earth around the Sun and the associated orbital parameters. The interpretation of these components is unambiguous, although for the PAY and, to a lesser extent, the TAT time-series, the yearly variability modes — IMF6 and IMF7, respectively — exhibit mode mixing, i.e. their total range overlaps some of the modes in the higher frequency band. Nevertheless, it will be subsequently be shown that it is indeed these components that account for variability at the one year time-scale in these datasets.

Lastly, the components indicative of low-frequency variability on time-scales greater than one year are discussed. The intrinsic time-scales found in these IMFs seem to match once more those pertaining to the so-called solar quasi-biennial oscillations, i.e. variations in the activity of the Sun exhibiting periodicities between 0.6 and 4 years [Bazilevskaya et al., 2015]. Again, Kolotkov et al., [2015] identify in the modes of the five solar proxies average periods of 395^{+46}_{-46}, 626^{+69}_{-113}, and 1423^{+196}_{-146} days respectively. Vecchio et al., [2012] also report solar quasi-biennial oscillations (QBOs) with time-scales from 1 year to 4.5 as being fundamental components of the variability of solar magnetic synoptic maps. Harrison, [2008] identifies a 1.68 year peak in the spectral domain by inferring cloud cover from measurements of the SSI, indicating that galactic cosmic rays, and not solar irradiance, may induce a cloud effect. Kirkby et al., [2011] and Kirkby et al., [2016] provide some tentative evidence for such an effect. Nevertheless, the same final precautions must be reiterated, similarly to the previous discussion concerning the high-frequency constituents, i.e. the similarity of the intrinsic periods for the different signals does not imply an unequivocal causal nexus.

4.3   The local climate imprint in the IMFs

It has been shown in section 2.1 that the four BSRN measuring stations experience different climates and exhibit differences in terms of the daily clearness index K_T. Figure 4.1 shows that the high frequency band composed of the first five IMFs is very much alike for all stations. This section investigates the possible relationship between local climate

and dissimilarities in terms of the repartition of the IMFs whose time-scales are larger that one year (IMFs 6 and higher).

It can be noted that the IMF6 for both BOU and CAR has a well-defined period (figure 4.1), with a median of respectively 368.2 and 364.3 days and very narrow interquartile range. In addition, for both stations, the IMF6 is the mode having the greatest amplitude and by far, compared to the other modes (figure 4.2). The IMFs 7 and 8 for CAR have less marked periods, i.e. the interquartile ranges are greater than for IMF6, and the amplitude of each IMF is very small. These observations may be related to the high frequency of cloud-free days seen in figure 2.1b because in absence of clouds, the variability of the daily mean of SSI is predominantly driven by the variability of the exo-atmospheric solar irradiance during the year.

On the contrary, PAY and TAT need four IMFs to account for the low frequency variability, i.e. one IMF more than BOU and CAR. IMF6 in PAY has a median period of 356.6 days, close to one year (figure 4.1) with a large interquartile range. The median amplitude of the IMF6 is approximately half of that of BOU or CAR (figure 4.2) and the amplitude exhibits large variations. The median amplitude of the IMF7 is similar to that of IMF6 while the period of the IMF7 is well marked with a narrow interquartile range. This may be related to the abundance of the presence of broken clouds at PAY and TAT compared to BOU or CAR that render the SSI signal highly intermittent. This intermittence of the signal could, in turn, explain the mode mixing observed in IMF6 [Huang et al., 2003].

Similar to PAY, TAT has a low median clearness index $\widetilde{K}_T^{\mathrm{TAT}} = 0.51$, which helps explain the presence of a sixth IMF (median period: 143.2 days) between the high-frequency components and the yearly IMF7 (median period: 366.6 days). Added to this, figure 4.3 shows that at TAT the magnitude of the spectral peak at three days is greater than the yearly one. In other words, there is much more power in the sub-year band at TAT than at PAY, or a lower signal-to-noise ratio of the yearly cycle. Hence, this high power of the stochastic components drives the EMD to assign an extra intrinsic mode for this region, as opposed to PAY, where the signal in this spectral band is assigned to the yearly IMF through mode mixing.

4.4 Discriminating deterministic signals from stochastic components in the IMFs

At this point, a question arises with regard to the physical and statistical significance of the identified spectral characteristics, namely how can one ascertain which features represent the expression of real, deterministic physical phenomena and which ones can be

attributed to random realizations of background processes. Such a method of discriminating between the "signal" and the "background" IMFs, proposed by Chen et al., [2013], will be described next.

4.4.1 Adaptive stochastic background null hypothesis

This subsection addresses the following questions: which confidence can be attributed to the information extracted by the EMD? More specifically, knowing that measurements may be contaminated by realizations of random processes, how can one ascertain that a certain IMF is the result of a real physical process as opposed to it possibly being a stochastic manifestations of the background?

In the past, several investigations have been carried out in order to identify the effects of the EMD when applied to time-series issued from various models, such as white, red, or fractional Gaussian noise, [Huang et al., 2003; Flandrin et al., 2004; Flandrin and Gonçalves, 2004; Wu and Huang, 2004; Flandrin et al., 2005; Rilling et al., 2005; Schlotthauer et al., 2009; Colominas et al., 2012]. As a result, it has been consistently shown that, irrespective of the assumed background model, the EMD acts as an efficient "wavelet-like" dyadic filter (see figure 3.4), decomposing the stochastic input into IMFs having the same spectral shape, but that are shifted in the frequency domain.

Nevertheless, the rejection of a null hypothesis based on an *a priori* assumed class of the background processes does not preclude the probability that the now statistically significant deemed signals originate from a stochastic process of a different kind. Furthermore, as the EMD is an adaptive, data-driven decomposition, it would be desirable to also employ a null hypothesis that shares the same characteristics, making no beforehand assumptions about the character of the background.

Following Flandrin, [2015] and Chen et al., [2013] this study will make use of the robust statistical properties of the EMD with respect to a wide class of background models, in order to adaptively contrast potential signals against the presumably stochastic background, as detailed hereafter. Owing to its dyadic filter character, the EMD decomposes stochastic inputs into IMFs having similar spectral shape, but that are translated to roughly the next lower octave in the spectral domain. When the sampling step is increased, i.e. the sampling frequency is reduced by fractionally re-sampling the input, these components cannot preserve their original locations in the spectral domain and will instead be shifted towards lower frequencies. Hence, significance testing of IMFs is done by verifying for each IMF if it remains unchanged in the time-frequency representation of the signal, on a normalized frequency scale, during fractional re-sampling.

A Hilbert marginal spectrum $h_i(f)$ is first constructed for each IMF $c_i(t)$ from its instantaneous amplitude $a_i(t)$ and instantaneous frequency $f_i(t)$, as per equations (3.21)

and (3.22). Next the spectrum-weighted mean frequency (SWMF) $\overline{f}_i$ of each IMF is computed [Chen et al., 2013]:

$$\overline{f}_i = \frac{\int h_i(f)\, f\, \mathrm{d}f}{\int h_i(f)\, \mathrm{d}f} \tag{4.1}$$

Then, the time-series is fractionally re-sampled by first up-sampling by an integer factor p, e.g. by inserting zeros, finite-impulse-response filtering the up-sampled data with a Kaiser-Bessel window normalized to account for the processing gain [Harris, 1978], then down-sampling the result by an integer factor q, such that $r = p/q$ is the desired final re-sampling rate. The original sampling rate Δt is thus made progressively larger, i.e. the time-spacing of the data points becomes:

$$\Delta t_r = \Delta t \cdot r, \quad r \in \{1.1, 1.2, \ldots, 1.9\} \tag{4.2}$$

For each sampling rate r, and for each IMF c_i, the SWMFs are then recomputed, obtaining a set $\overline{f}_{i,r}$. To enhance the visibility of the evolution of frequency as a function of the re-sampling rate, normalization is performed:

$$\widehat{f}_{i,r} = \frac{\overline{f}_{i,r}}{\overline{f}_{i,1}} \tag{4.3}$$

with $\overline{f}_{i,1}$ being the SWMFs of the modes of the data having the original sampling rate. Therefore, the normalized SWMFs for the IMFs of the original data will be unity, i.e. $\widehat{f}_{i,1} = 1, \forall i$.

Since the EMD is an efficient dyadic filter, frequency deviation from the unity line will occur for stochastic-like IMFs. It follows that when $\widehat{f}_{i,r} \simeq 1, \forall r$, the null hypothesis that mode i is the realization of stochastic processes can be rejected.

For an extensive set of examples of this adaptive statistical significance testing technique on simple test cases, the reader is kindly directed to the original report of Chen et al., [2013].

4.4.2 Signal detection through fractional re-sampling

The procedure outlined above was applied to the first eight IMFs of all the time-series and the results are presented in figure 4.4 (clockwise from top left: BOU, CAR, TAT, and PAY). First, each time-series was re-sampled with a fractional sampling rate up to a factor of two, i.e. the original uniform time-spacing of the data, Δt, was progressively made larger and larger, as described in equation (4.2): $\Delta t_l = \Delta t \cdot r$, where $r \in \{1.1, 1.2, \ldots, 1.9\}$ is the re-sampling rate and is running along the horizontal axis. Next, the HHT was used to
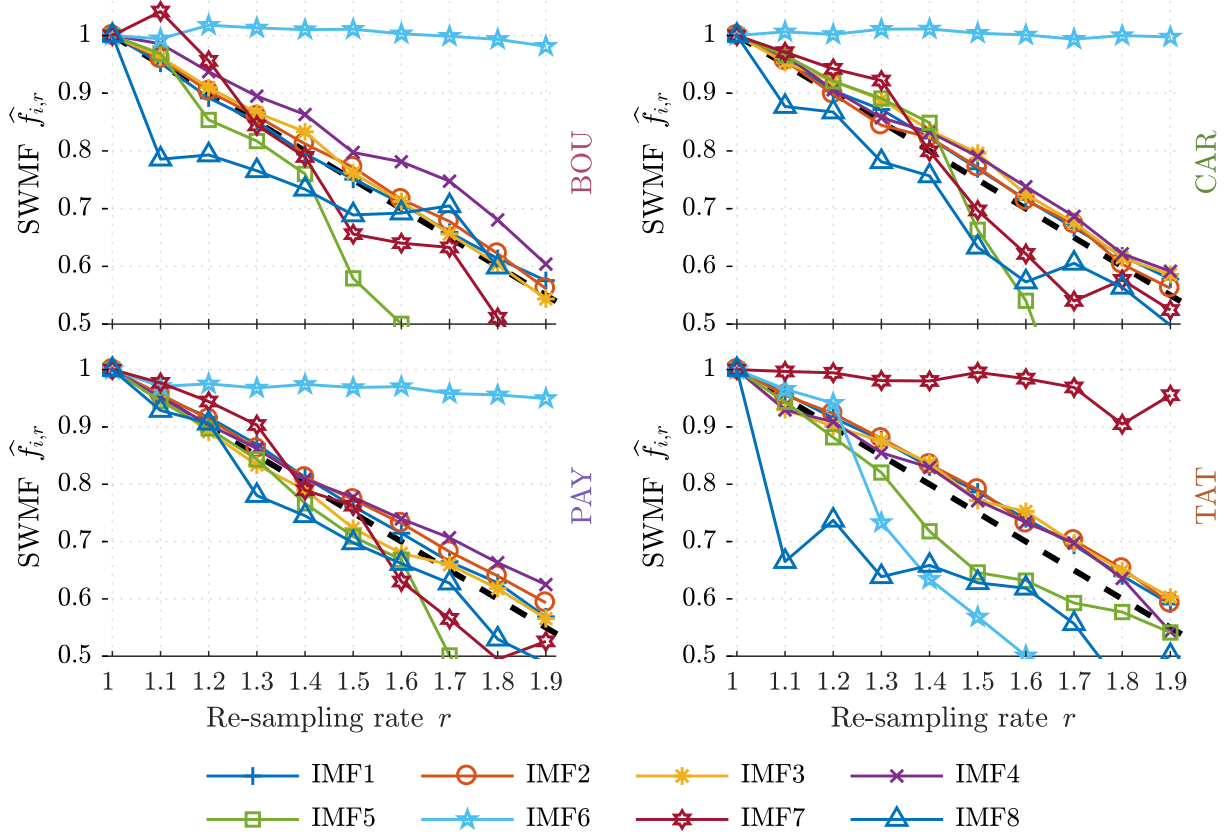
Figure 4.4 The drift of normalized SWMF $\widehat{f}_{i,r}$ (ordinate) of IMF c_i, with $i \in \{1 \ldots 8\}$, as a function of the re-sampling rate r (abscissa) for the four time-series. Clockwise from top left: BOU, CAR, TAT, and PAY. The black dashed diagonal depicts the behaviour of a purely stochastic time-series under an ideal dyadic filter. For all datasets, the only mode that maintains a quasi-constant frequency under fractional re-sampling is the IMF associated with the yearly cycle, i.e. IMF6 for BOU,CAR, and PAY and IMF7 for TAT. In all the other IMFs quasi-stochastic behaviour is apparent, through frequency down-shifting towards the next lower octave, approximately following the dashed line.

decompose the resulting time-series into IMFs and to compute their spectrum-weighted mean frequencies, following equation (4.1). In order to emphasize the effects of the fractional re-sampling on the spectral contents of the IMFs, these latter frequencies were then normalized by the SWMF of the original, non re-sampled data as per equation (4.3). For each dataset, this ratio is indicated on the y-axis, as $\widehat{f}_{i,r}$, with $i \in \{1 \ldots 8\}$ indicating the IMF number. It then becomes possible to follow the evolution of the normalized SWMF of each individual IMF as a function of the fractional re-sampling rate (figure 4.4). If the signal is purely stochastic, the IMFs of time-series undergo a translation towards lower frequencies under fractional re-sampling. Therefore, for those IMFs whose SWMFs are not down-shifted during re-sampling, the null hypothesis that they are purely stochastic can be rejected, i.e. they represent meaningful signals. Stated otherwise, an IMF c_i is deemed not to be stochastic in nature if its normalized SMWFs $\widehat{f}_{i,r}$ stay close to the unity line for all r. From figure 4.4 it can be observed that for all the stations,

the only component that maintains quasi-constant frequency under fractional re-sampling is the mode representative of the yearly variability, i.e. IMF6 for BOU, CAR and PAY, and IMF7 for TAT. All the other IMFs experience the previously described frequency down-shifting, hence for them the null hypothesis that they are purely random realizations cannot be rejected.

At this point, several precautionary notes are compulsory. First, the rule of inference used here is *modus tollens*, i.e. the results from figure 4.4 do not imply that the modes who experience down-shift in their SWMFs are purely stochastic in nature. It will be subsequently shown that, for the first five IMFs at least, this is indeed the case; although (quasi-)stochastic in nature, they are not completely devoid of information. Second, the result is mostly qualitative, since it is difficult to define a confidence interval owing to the adaptive nature of the null hypothesis that can account for different types of stochastic background processes. Third and last, the approach is best applied only to the high frequency modes, with respect to the data length and sampling, since by re-sampling spurious low-frequency oscillations may inadvertently be introduced [Chen et al., 2013]. This is further supported by the fact that as IMF number progresses, going toward larger time-scales, the region where the influence of edge effects becomes important is getting larger and larger, hence only adding uncertainty to the interpretation of the results. This is also the reason why this type of analysis was only carried out on the first eight IMFs of each dataset. As a corollary, unambiguous interpretations of QBO-like components seems to be out of reach with only 10 years of measurements.

4.5 Amplitude modulation through non-linear cross-scale coupling

This section investigates whether the first five IMFs can be modelled as being purely realizations of a stochastic background, or if they contain some form of information. To test this, the rank correlation between the yearly and sub-yearly IMFs and their envelopes, e.g the AM part in the middle panel of figure 3.10, has been computed for each SSI time-series. Kendall's rank correlation coefficient, τ, is a statistical measure of ordinal association describing how similar the orderings of the data are when ranked [Kendall, 1938]. It is employed here to establish whether each pair of the two variables, AMx and IMFy with $x, y \in \{1, \dots, 7\}$, may be regarded or not as independent. Non-zero values of τ denote dependence between the variables: $\tau = 1$ indicates perfect agreement between rankings, while $\tau = -1$ denotes perfect inverse agreement, i.e. one ranking is the reverse of the other.

The resulting rank correlation coefficients and the associated p-values, are presented in figure 4.5. For each panel, the columns denote the EMD modes (IMFx), and the rows their amplitude envelopes (AMy). The background colour of each cell (AMx, IMFy) indicates the rank correlation τ between IMFy and the AM part of IMFx within the same dataset. The legend of the colour encoding is found on the color bar at the bottom of the figure. The associated p-values, i.e. "the probability under a specified statistical model that a statistical summary of the data [...] would be equal to, or more extreme, than its observed value" [Wasserstein and Lazar, 2016], are presented numerically in each cell, for the sake of completeness and transparency. For BOU, CAR and PAY, IMF6 accounts for the yearly variability of the time-series, hence the rank correlation matrices are 6×6 in size. For TAT, the yearly mode is IMF7, thus in this case the rank correlation matrix has a size of 7×7. Two conclusions can be drawn from figure 4.5.

Values of τ significantly different from zero, shown in red, are recorded in the last column, for all stations. These demonstrate a modulation of the amplitude of the components having sub-year time-scales, i.e. AM1,...,AM5, respectively AM6 for TAT, by the yearly IMF, at a statistically significant level ($p \sim 0$). The effect is most pronounced for PAY, as inferred from the darker red shades (larger rank correlation coefficients).

For the BOU and CAR datasets the first row (AM1) exhibits blue and dark blue cells for IMF3,...,IMF5 at the statistically significant level. This indicates a negative rank correlation. Similar, but lighter, amplitude modulation is observed on the second row (AM2), but only by IMF4 and IMF5. For the PAY series, this negative rank correlation is greatly reduced for the first row (light blue tones) and is absent in the second row. For TAT no such correlation can be observed. At this point it is interesting to note, that in a similar way to the discussion from section 4.3, the different features of the datasets from figure 4.5 also enable a classification of the local climate experienced by the measuring stations.

It should be mentioned that the amplitude modulation of high-frequency stochastic components by lower frequency ones is also found in time-series of sunspot numbers [Chen et al., 2013] and in multiple solar proxies [Kolotkov et al., 2015]. The short term intrinsic periodicities in the solar proxies appear to be indicative of "randomly distributed dynamical processes in the solar atmosphere" that are closely related to the 11 year solar activity and therefore, unsurprisingly, the high-frequency modes are found to be modulated by this latter cycle [Kolotkov et al., 2016]. But this phenomenon is not limited to solar activity signals, and has also been identified in records of air temperature at surface [Paluš, 2014], and time-series of the sea level [Liu et al., 2007], and may indicate cross-scale non-linear couplings [Paluš, 2014; Huang et al., 2016].
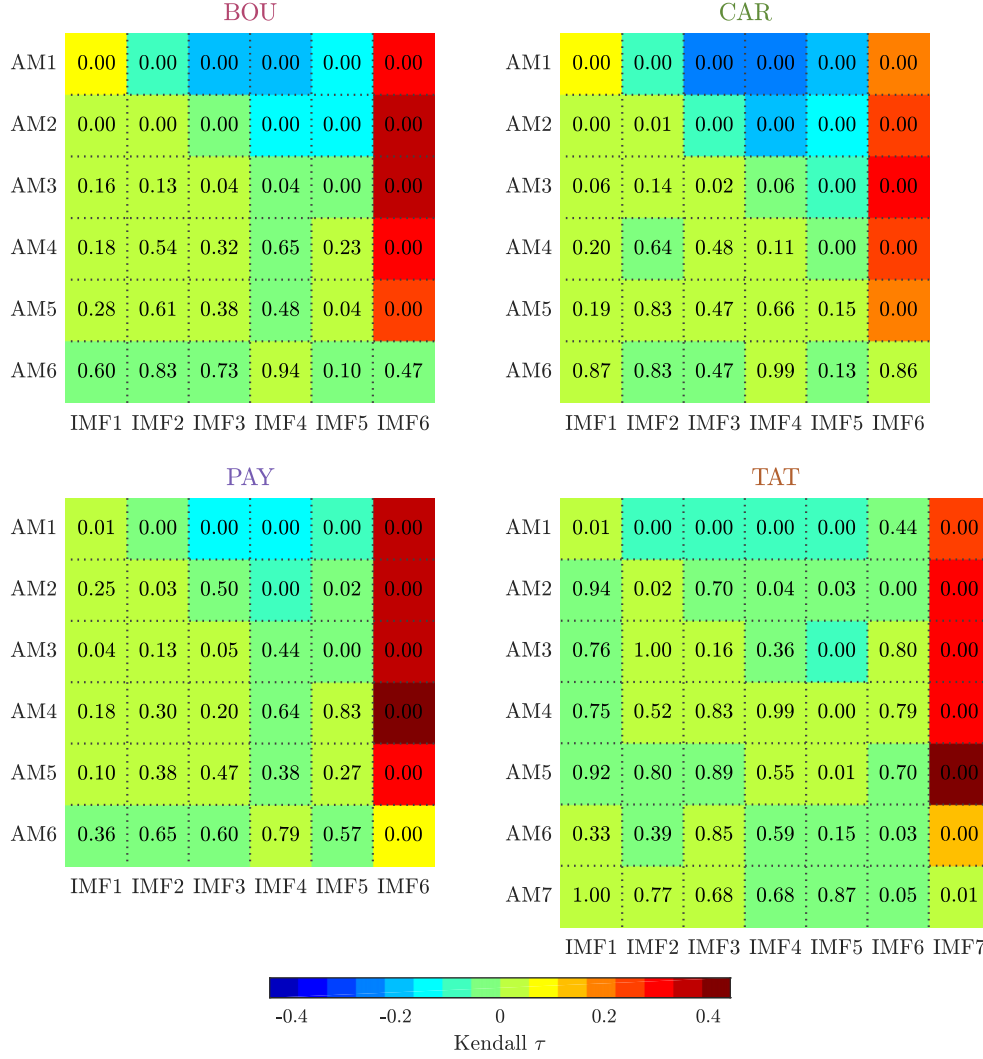
Figure 4.5 Rank correlations between IMFs and their AM components for BOU (top left), CAR (top right), PAY (bottom left) and TAT (bottom right). Kendall's rank correlation coefficient τ is colour-coded according to the colour bar on the bottom. IMFs run vertically, along the columns, and their AM components run horizontally, along the rows. The numeric values within the cells are the associated p-values.

Conclusion   To sum up, the HHT analysis of decennial time-series of daily means of measurements of the SSI, from distinct BSRN stations has revealed the following: the presence of a high-frequency (2-100 days) component consisting of quasi-stochastic IMFs that have been shown to be amplitude-modulated by the yearly cycle; a low power spectral band in the 100 days to 300 days region; a well-defined spectral peak at the one year mark accounting for the yearly variability; and multiple "QBO" components whose origin can be inconclusively attributed to quasi-stochastic random processes.

This separation of the (quasi-)periodic components of the signal from the apparently random realizations of a noisy background has been shown to significantly augment accuracy in modelling of time-series [Rios and de Mello, 2013]. These findings described in

this chapter can be thus directly used to improve models for estimating SSI from satellite images or forecasts of the SSI.

It has also been shown that the adaptive Hilbert-Huang Transform (HHT) is a versatile tool in analysing SSI datasets, exhibiting significant non-linearity and non-stationarity. First, it was employed to extract the intrinsic modes of variability of the SSI at distinct time-scales. It was proposed that a classification of the measuring stations according to climate and/or solar insolation conditions may be possible, based on the Hilbert spectral features of the data. Second, the HHT has been shown to be able to discriminate, without making any *a priori* assumptions about the nature of the data or of that of the random background, between the deterministic yearly cycle and the quasi-stochastic high-frequency components.

CHAPTER 5

The intrinsic variability of surrogate SSI estimates

IT is often the case that, in practice, information about the SSI is required at remote locations, far from any measuring station. However, extending the representativeness of ground station measurements to surrounding areas cannot be applied to regions where the physical and/or climatological distance between stations is large [Zelenka et al., 1992]. Using the nearest-neighbour approximation, Zelenka et al., [1992] have found that the relative root-mean-square error (RMSE) of the estimates is proportional to the square root of the distance between sites.

An alternative is to make use of satellite-based methods, whose outputs are a good supplement in long term solar resource assessment. Perez et al., [1997] have shown that when the distance from a station exceeds 50 km, in the case of daily aggregates, or 34 km for hourly data — Zelenka et al., [1999] estimate this as low as 20 km,— satellite-derived values of the SSI are more accurate than estimating using the nearby measuring point. Thus, given the scarcity and spatial sparsity of long-term ground measurements of the SSI satellite-derived estimates of surface solar radiation remain a good complement to ground station data [Lefèvre et al., 2014].

Yet another possibility for estimating the solar radiation at ground level is provided by global atmospheric re-analyses resulting from numerical weather models. The main benefits are the wide, even global, coverage and the spanning of multi-decennial time periods. However, some authors have found a large uncertainty relative to satellite-based irradiance estimates and advise against using data from re-analyses [Lohmann et al., 2006; Boilley and Wald, 2015]. Still, others find datasets of SSI from re-analyses to be suitable for photovoltaic applications [Richardson and Andrews, 2014].

Efforts to improve the adaptation of long-term (10+ years) re-analyses and satellite solar radiation datasets to a specific geographical site, by means of short-term (1 year at least) in-situ measurements, are also ongoing [Polo et al., 2016].

In this context, this chapter investigates and analyses the temporal variability in time-series of daily means of SSI for two geographical locations, at different time-scales, as found in the outputs of different models, satellite estimates, re-analyses, and ground measurements. To gain better insight into the causes of variability of the SSI, the down-welling solar short-wave irradiance is followed along its path through the atmosphere towards the surface. The solar irradiance at the top of the atmosphere (TOA) is easily modelled, disregarding the 11-year cycle of $\sim$0.1% amplitude variability, and is first analysed as a clean input signal, devoid of any atmospheric perturbations, in order to reveal the natural variability of the exo-atmospheric solar input. To account for the variability due to atmospheric effects such as scattering and absorption due to atmospheric gases, water vapour and aerosols, but excluding any influence of clouds, the outputs of a clear-sky model, i.e. a model providing SSI that should be observed if the sky were cloud-free, are then scrutinized. The role of clouds on variability is lastly inferred by analysing pyranometric ground measurements.

A further exploitation of this work deals with the use of the HHT to assess the fitness for use of satellite estimates and re-analyses data by comparison with the measured data. The novelty of this approach stems from the fact that, unlike previous studies where global statistical indicators are employed [Espinar et al., 2009], here the datasets are decomposed into their constituent characteristic time-scales before undergoing analysis. To this end, the adaptive, data-driven HHT is employed. In this way, the fitness for use of the modelled and estimated SSI can be assessed not only at a global, whole dataset level, but also on a per time-scale basis.

5.1   Analysis of the intrinsic time-scales of variability

The $12 = 2 \times 6$ datasets[1] described in section 2.2 for the sites Vienna (VIE) and Kishinev (KIV) have been decomposed by the EMD into 10 IMFs and a residual, as shown in figure 5.1. The modes IMF1,…,IMF10 and the residual (Res.) are plotted as YZ slices along the X axis, with time running on the Y axis, and amplitude on the Z axis. The zero-centred oscillatory nature of the IMFs can be clearly seen, as well as the local time-scale increase with mode number. The IMF1 has a time-scale of 3.0 days and exhibits a large variability in time. As the IMF rank increases, the time-scale increases and the

---

[1] Six time-series for each of the two sites: TOA, McClear, WRDC, HC3v5, ERA, and MERRA2.
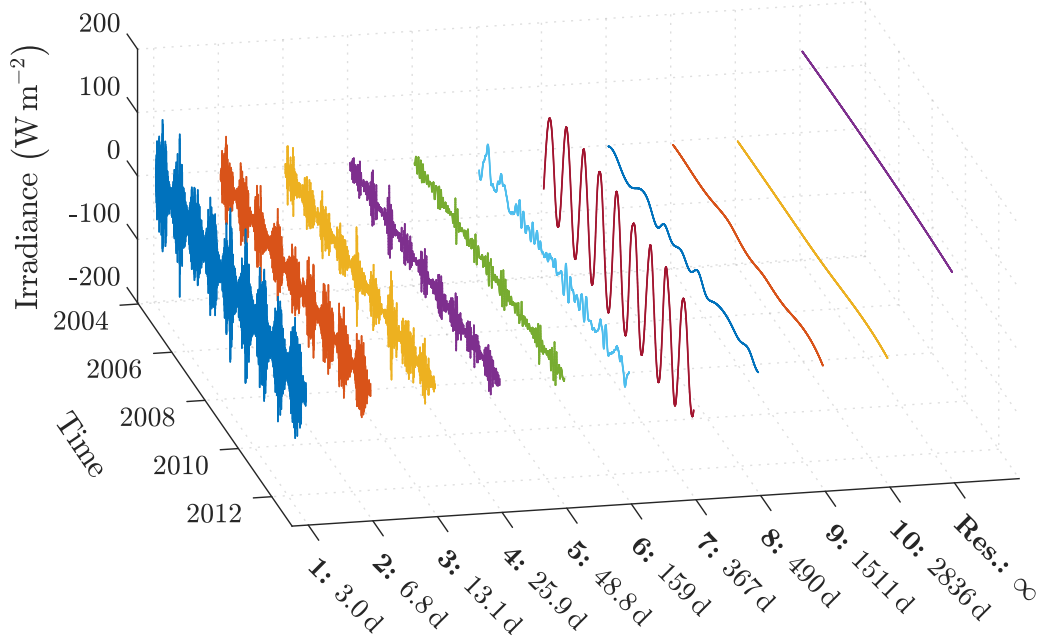
Figure 5.1 The IMFs obtained by decomposing the WRDC time-series for VIE, plotted as YZ slices, with the X-axis denoting their number (IMF1,...,IMF10) and time-scales in days. For reference, the residual (Res.) is also included. The amplitude, or irradiance, of each IMF is plotted on the Z-axis, and time runs on the Y-axis. Time markers denote January 1$^{\text{st}}$ of the corresponding year.

variability decreases. The exception is for IMF7 at 367 days which exhibits the greatest variability, as discussed later.

Similar plots were constructed for the rest of the datasets and are available as supplementary material in appendix B. To summarize the results, the mean characteristic scales of variability for all the IMFs of the VIE datasets have been compiled into table 5.1, while the corresponding mean amplitudes are given in table 5.2. Similar summaries for the KIV datasets are provided in table 5.3 for time-scales and table 5.4 for amplitude, respectively.

From tables 5.1 and 5.2 it can be seen that the most significant time-scale of variability present in the TOA time-series is around the 1 year mark, as evidenced by IMF6. The mean period of this component is 366 days and its mean amplitude of 186 W m^{-2} is two orders of magnitude greater than the other modes. From this, it can be inferred that this yearly mode, a result of the revolution of the Earth around the Sun, is the only significant scale of variability of the TOA. This result is unsurprising, since it can also be inferred from figure 2.2a top panel, where the TOA series does not exhibit any other variability apart from this mode (the 11-year variability is not modelled), i.e. it is almost a perfect sine wave with a period of one year. Adjacent to this sixth mode, IMF5 and IMF7 also display similar time-scales of 344 days and 366 days, however their mean amplitudes are 3.1 W m^{-2} and 5.4 W m^{-2} respectively, thus their origin is attributed to spectral leakage from the main mode. The first four IMFs have negligible amplitudes (0.1 W m^{-2}) and are

Table 5.1 Mean IMF time-scales in days for the VIE datasets

	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8	IMF9	IMF10
TOA	6.5	9.3	17.0	30.8	344	366	366	1431	3134	3242
McClear	3.0	6.9	13.2	25.3	55.3	366	417	803	2971	3432
WRDC	3.0	6.8	13.1	25.9	48.8	159	367	490	1511	2836
HC3v5	3.0	6.7	13.2	26.2	49.8	156	368	475	914	1645
ERA	3.1	6.8	13.1	25.8	50.8	180	371	576	1042	2838
MERRA2	3.0	6.7	12.8	26.0	52.1	217	368	568	1536	3351

Table 5.2 Mean IMF amplitudes in $\mathrm{W\,m^{-2}}$ for the VIE datasets

	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8	IMF9	IMF10
TOA	0.1	0.1	0.1	0.1	3.1	186	5.4	1.2	2.0	0.4
McClear	6.4	4.6	3.7	3.0	3.2	128	10	1.6	1.7	0.4
WRDC	44	24	19	16	12	16	96	8.3	4.2	2.8
HC3v5	47	26	20	17	13	14	91	12	6.5	2.1
ERA	33	20	16	13	10	19	89	6.9	5.0	2.2
MERRA2	38	21	16	13	10	30	94	4.7	4.5	3.1

probably the manifestation of residual noise. IMF8 has a mean time-scale of 1431 days and its low amplitude of $1.2\ \mathrm{W\,m^{-2}}$ is an indication that it might be a numerical artefact. IMF9 and IMF10 are also assumed to be non-physical, because their large periods of over 3100 days are indication that these components are entirely immersed in the time coverage where edge effect are non-negligible. As a reminder, edge effects are important at the half-period of a component at data boundaries, which for these two latter IMFs yields 1550 days forward from February 1st 2004 and 1550 backwards from January 31st 2013, spanning almost the entire data range.

For the McClear time-series of cloud-free SSI, issued from a clear-sky model taking only the transparency of the atmosphere into account and dependent mostly on aerosols, water vapour, and ozone, as well as for the rest of the VIE datasets, IMF1,...,IMF5 display remarkably similar features, such as monotonically decreasing amplitudes and time-scales that exhibit period doubling, roughly following the dyadic scale: 3 days $\rightarrow$ 6.8 days $\rightarrow$ 13.1 days $\rightarrow$ 26 days $\rightarrow$ 51 days. The amplitudes for McClear are 3 to 7 times less than in other time-series, and less than $6.5\ \mathrm{W\,m^{-2}}$. The break in the monotonic decrease of amplitude with scale for McClear — $3.2\ \mathrm{W\,m^{-2}}$ for IMF5 vs $3\ \mathrm{W\,m^{-2}}$ for IMF4 — is considered an artefact, since this monotonicity holds for the other datasets and also for KIV (see table 5.4). The presence of these first five dyadic scales in all the datasets, with the exception of TOA, allude to their possible origin as being cloud-free atmospheric

Table 5.3 Mean IMF time-scales in days for the KIV datasets

	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8	IMF9	IMF10
TOA	6.5	9.3	16.0	31.8	344	366	389	796	3157	3407
McClear	2.9	6.9	12.6	25.1	54.9	366	387	1054	2981	3813
WRDC	3.1	6.9	13.3	26.9	49.0	199	369	959	1552	1760
HC3v5	3.1	6.9	13.3	27.4	49.7	176	370	726	1639	2680
ERA	3.1	6.9	13.3	27.2	48.3	280	370	1015	1540	3367
MERRA2	3.0	6.9	13.4	26.8	48.3	298	374	746	1658	3533

Table 5.4 Mean IMF amplitudes in $\mathrm{W\,m^{-2}}$ for the KIV datasets

	IMF1	IMF2	IMF3	IMF4	IMF5	IMF6	IMF7	IMF8	IMF9	IMF10
TOA	0.1	0.1	0.1	0.1	3.3	182	5.6	1.4	2.6	0.5
McClear	7.7	5.1	3.8	3.5	3.1	130	6.6	1.6	0.9	0.5
WRDC	42	25	20	18	12	22	103	6.6	5.4	0.6
HC3v5	40	24	19	17	11	19	103	7.4	6.6	1.2
ERA	29	18	14	13	9.3	31	91	4.9	4.5	2.3
MERRA2	36	21	16	14	9.3	41	89	5.5	5.9	0.6

effects. This can also be observed in figure 2.2a, where, as opposed to the TOA, the raw McClear time-series is seen to exhibit slight high-frequency variability, and which tends to increase during the summer months. For McClear, there is little to no variability in the 2–3 months to 1 year band (see also figure 5.2, discussed later on). As for TOA, IMF6 of the McClear time-series is clearly associated with the yearly variability, by its mean period and amplitude of 366 days and 128 $\mathrm{W\,m^{-2}}$, respectively. Here too, this is the most energetic spectral component of the dataset. IMF7 of McClear has a median time-scale of 417 days and a median amplitude of 10 $\mathrm{W\,m^{-2}}$. IMF8,…,IMF10 have very low amplitudes of less than 1.8 $\mathrm{W\,m^{-2}}$, hence their origin cannot be unambiguously determined. This is especially the case with IMF9 and IMF10, which reside almost entirely in the edge effect region because of their large periods of about 3000 days and greater.

The WRDC dataset unsurprisingly shares many features with HC3v5, ERA and MERRA2 datasets, since the latter three are intended to be accurate estimates of the former. The rest of the results will be presented in a lumped form for these four time-series. For IMF1,…,IMF5, HC3v5 agrees better with WRDC than the re-analyses in terms of mean amplitudes and takes on only slightly higher values (1 to 3 $\mathrm{W\,m^{-2}}$ or less than 5% on average, table 5.2). This can also be seen in figure 2.2a, where the HC3v5 data exhibit a slight overestimation of the low SSI values for WRDC during winter. For the first five modes, both ERA and MERRA2 have mean amplitudes that are on average 17% less than

those of WRDC. This is also apparent in figure 2.2a, where fewer samples of less than $150\,\mathrm{W\,m^{-2}}$ are occurring between spring and autumn for the re-analyses time-series; this is mostly visible for the year 2007. For IMF6, HC3v5 closely follows WRDC, both in terms of time-scale: 156 days vs. 159 days, and amplitude: $14\,\mathrm{W\,m^{-2}}$ vs. $16\,\mathrm{W\,m^{-2}}$. For the same IMF6, both ERA and MERRA2 exhibit significantly greater time-scales: 180 days and 217 days, respectively, and amplitudes: $19\,\mathrm{W\,m^{-2}}$ and $30\,\mathrm{W\,m^{-2}}$. The first six modes are responsible for the high-frequency variability, which is also manifest in figure 2.2a, i.e. the estimates and ground measurements of the SSI are significantly more variable than the McClear time-series. For all these datasets the yearly variability cycle is captured by IMF7, with generally good agreement across datasets in terms of both time-scale (367 to 371 days) and amplitude (89 to $96\,\mathrm{W\,m^{-2}}$). The greatest modes, IMF8,...,IMF10, have amplitudes less than the uncertainty threshold, with the exception of IMF8 for HC3v5.

Generally similar results are also obtained for KIV, as summarized in tables 5.3 and 5.4, which only serve as a confirmation that the identified features are not site-specific. Apart minor differences in numerical values, the only notable exception in the KIV datasets are the significantly greater time-scales for IMF6 of 280 and 298 days respectively for ERA and MERRA2, as opposed to 199 and 176 days for WRDC and HC3v5, respectively.

5.2   Hilbert spectral analysis

The previous summary of the results, although informative, is static in the sense that only two features are used to characterize, in an approximative manner, each time-evolving IMF: the long-term average amplitude and time period. To make use of the full potential of the HHT, Hilbert spectra were also computed for all the datasets (not shown) and are provided in the supplementary material. From these Hilbert spectra the marginal, time-integrated, versions were computed and are presented in figure 5.2 for VIE and in figure 5.3 for KIV.

The TOA spectrum for VIE in figure 5.2 confirms the previous findings. For this dataset the only significant mode of variability is found at the one year mark, and has a power of about 63 dB. A slight peak of 17 dB is also present around 1400 days, corresponding to IMF8. The end region of the spectrum, from 2500 days onwards also contains some power, but the respective IMFs have been shown to be heavily affected by edge effects, thus the origin of this feature is ambiguous at best. For KIV the situation is very similar (figure 5.3), the TOA exhibits only a sharp yearly variability component (IMF6).

Apart from the Hilbert marginal spectra, it has been inferred from the mean time-scales and mean amplitudes of the decomposed data (tables 5.1, 5.2, 5.3, and 5.4) that the yearly mode of variability is the most prominent feature of all the datasets. This result is unsurprising since both stations are situated at mid-latitude and it can be explained
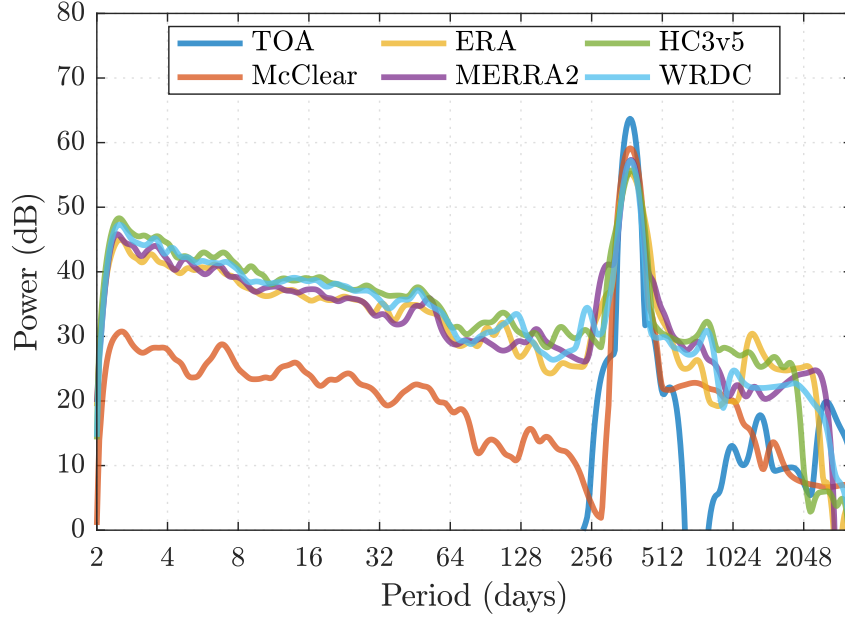
Figure 5.2 The Hilbert marginal spectra for the VIE datasets: TOA, McClear,WRDC, HC3v5, ERA, MERRA2. The abscissa indicates the time-scale on a base-2 logarithm, and the ordinate denotes power in dB.

in terms of orbital geometry through the yearly cycle of seasons; it can also be inferred by visually inspecting the raw data from figure 2.2 which shows large variability with periodicity of one year. For WRDC, HC3v5, ERA and MERRA2, there is good agreement with respect to this mode (IMF7) both in terms of amplitude and of time-scale. Apart from this yearly component, the TOA exhibits no other form of significant variability for either station, also in good agreement with its traces from figure 2.2, which register as undisturbed sinusoidal waveforms.

Compared to TOA, the McClear dataset for VIE is seen to introduce variability in the high-frequency regime, whose power decreases almost monotonically from 30 dB at 2 days, to about 2 dB at roughly 300 days (figure 5.2). Most of this variability occurs during summer, as observed in figure 2.2a (see also the full Hilbert spectrum of McClear in the supplementary material). The yearly variability component stands out again, this time with just less than 60 dB in power. From here, power decreases to a minor spectral shoulder of 22 dB at 800 days (IMF8), after which it fades out towards larger frequencies. For KIV, McClear introduces an almost monotonically decreasing band of high-frequency variability (figure 5.3). Unlike for VIE, here McClear does not drop abruptly up to one year, but exhibits a rebound of 17 dB around 150 days. This is interpreted as an artefact, induced by energetic oscillations in and near the left edge effect boundary in the full Hilbert spectrum (see supplementary material).

As previously shown, the high frequency variability (IMF1,...,IMF5) of HC3v5 matches more closely that of WRDC, while the re-analyses have slightly less power (up to 2–5 dB).
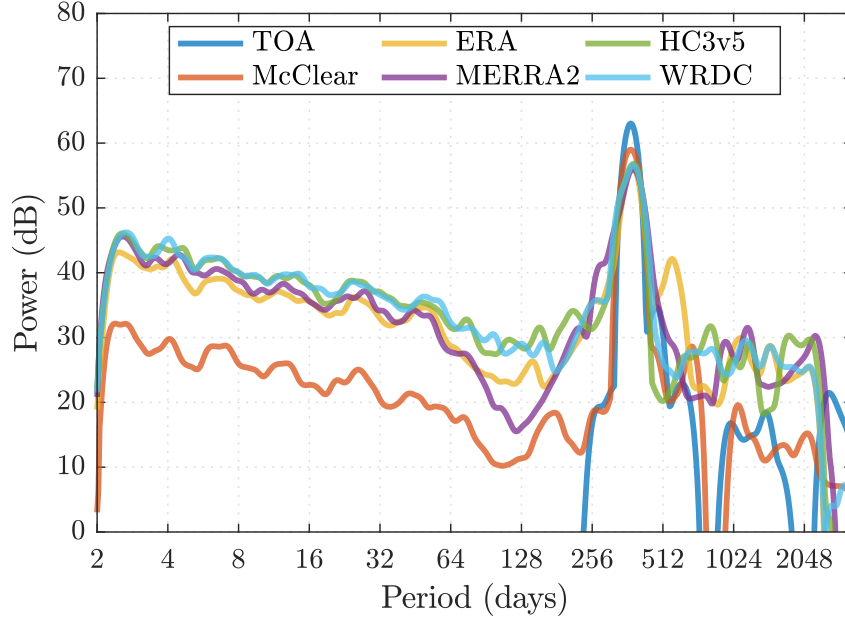
Figure 5.3 The Hilbert marginal spectra for the KIV datasets: TOA, McClear, WRDC, HC3v5, ERA, MERRA2. The abscissa indicates the time-scale on a base-2 logarithm, and the ordinate denotes power in dB.

The power of these features is 15 dB greater in the estimates and ground measurements of SSI than that found in the clear-sky regime. From 170 days to 256 days, WRDC and HC3v5 overpower the re-analyses. After 256 days, the power in the re-analyses overcomes that of WRDC and HC3v5, until approximately one year. This can also be seen from tables 5.1 and 5.2, where IMF6 for the re-analyses is seen to be greater than the other two time-series, both in terms of amplitude and of time-scale. Again, the yearly variability component is the largest spectral characteristic, with WRDC peaking at 57 dB. Here, the other dataset closely agree with the ground measurements as per the VIE results summary tables. After the 500 days mark the interpretation of the different spectral features is ambiguous, both because of their mean amplitudes failing to rise above the uncertainty level, and also because of the progressively large impact of the edge effects, especially towards the end of the spectrum.

The spectra for the KIV datasets in figure 5.3 are very similar to their VIE counterparts. From 2 days to 2 years, the HC3v5 spectrum follows the WRDC one more closely that the re-analyses. The latter two exhibit large downward excursions of 5 to 10 dB, from 70 days to roughly 150 days. For the yearly variability component, there is better agreement than for VIE between the estimates and the ground measured SSI data, all four of them peaking simultaneously at 57 dB. The large peak of 42 dB found in ERA at 530 days should be ignored, since it is caused by an energetic oscillation in the edge effect region (see supplementary material). As for the VIE datasets, after the 500 days mark the interpretation spectral features becomes ambiguous.

Taking both Hilbert spectra into account, the following characteristics can be outlined. High-frequency variability, from 2 days up to 2–3 months, is manifest in McClear through its first five IMFs. This feature is also present in the rest of the datasets, and with greater power when compared with McClear. Hence, this feature can be attributed to clear-sky (no cloud) atmospheric effects (scattering and absorption by ozone, water vapour, aerosols, etc.). Looking at the McClear graphs in figure 2.2, it becomes evident that this high-frequency variability manifests itself more strongly during the summer than during the winter months. In other words, the yearly cycle modulates the power of this high-frequency feature through a non-linear cross-scale amplitude-phase coupling. This feature is also apparent in the HC3v5, WRDC, ERA and MERRA2 datasets (see supplementary material) and is in agreement with the findings of the previous chapter (sections 4.4 and 4.5), where its stochastic nature was underlined. The time-scales for the individual modes composing this high-frequency feature agree well across these latter four stations. In terms of amplitudes, however, for VIE HC3v5 slightly overestimates the WRDC measurements while ERA and MERRA2 underestimate more severely, while for KIV all the SSI estimates yield lower values than the ground data, although less so for HC3v5. This somewhat anticipates the results of the next section and can also be inferred from tables 5.5 and 5.6, where for the first five IMFs, HC3v5 outperforms ERA and MERRA2 by a large margin. This result is a major contribution of the study, since Inman et al., [2013] have proposed that the accuracy of SSI forecasts crucially depends on the ability to forecast the stochastic component. As such, practitioners interested in modelling and forecasting the daily SSI will do better using satellite estimates that radiation products from re-analyses, at least for the two sites studied herein. This finding can be explained by the fact that re-analyses assimilate state variables such as temperature, moisture and wind, while the SSI is diagnostic. Stated differently, in re-analyses, radiation and cloud properties are derived from a model and, as such, they include the uncertainty of this model. Re-analyses often predict clear sky conditions while the actual conditions are cloudy [Boilley and Wald, 2015]. HC3v5 is based on Meteosat imagery (see section 2.2) and as such directly takes clouds into account.

Another significant point is that, for both VIE and KIV, the McClear datasets do not have a variability component in between this high-frequency feature and the yearly cycle. In other words, IMF6 for the McClear data represents the yearly cycle, unlike the ground measurements or the SSI estimates, where IMF6 is an intermediate component before the yearly component represented by IMF7. This has first been discussed as a "variability gap" by Bengulescu et al., [2016a] (see also section 3.3.1), when analysing a decennial dataset of daily means of SSI measured by BSRN ground station at Carpentras, France, that experiences clear-sky conditions for most of the year. Subsequently, Bengulescu et al., [2016c] (see also sections 4.2 and 4.3) have shown that this "variability gap" is also manifest

for a similar dataset of ground measurements taken at Boulder, Colorado, USA, a location that also experiences a high number of days with clear skies. Hence, this study confirms the fact that, indeed, a clear-sky atmosphere does not introduce any spectral features in between 2–3 months and 1 year. Since the ground data for both VIE and KIV feature such spectral components, it can be concluded that these two locations do not experience so many cloud-free days and/or that they experience a lot of broken clouds conditions. Here too, ERA and MERRA2 are outperformed by HC3v5, since the time-scale of this IMF6 in the re-analyses datasets is greatly different from the true time-scale found in ground measurements, and which is accurately reflected by the satellite estimates.

5.3 Per time-scale comparison of SSI estimates and ground measurements

Still another possibility of investigating the data is to make use of the adaptive, data-driven, time-domain filter character of the EMD. Looking at pairs of IMFs in the time domain, it is possible to construct 2D histograms of the satellite and re-analyses estimates of SSI compared to the concomitant ground measurements. This gives a good overview of the similarity, at each characteristic time-scale of variability, between satellite estimates of the SSI or re-analyses radiation products and the WRDC measurements, which serve as ground truth.

Figure 5.4a illustrates the 2D histogram for the second IMF of the HC3v5 and WRDC datasets for the VIE station. The colour of each pixel denotes the relative frequency of the WRDC–HC3v5 irradiance pairs, as encoded on the colour-bar on the right. In figure 5.4a, the pattern of dots has a positive slope which indicates a positive correlation between the two variables. A robust best-fit linear regression has been performed, with the resulting line shown in dash-dotted red. Plotted in solid black is the identity line; were the two datasets identical the scatters would fall exactly onto it. The equations for the two lines are indicated in the legend. The line describing the best-fit can be seen to deviate very little from the identity line, with a slope of 1.021 indicating that for IMF2 at VIE the satellite-derived SSI slightly over-estimates the SSI measured at ground level, and an intercept of –0.016, which is the expected mean irradiance value of HC3v5 when then WRDC irradiance is zero. Indeed, the linear regression model manages to explain 92.4% of the total variability, as indicated by the coefficient of determination ($R^2 = 0.924$). The points do not fall exactly onto the best-fit line and exhibit a small scattering, with a RMSE of 6.2 $W\,m^{-2}$. The characteristic time-scale of variability is also indicated on the plot, as 6.8 days.

(a) 2D histogram for IMF2 of HC3v5 and WRDC

(b) 2D histogram for IMF2 of ERA and WRDC
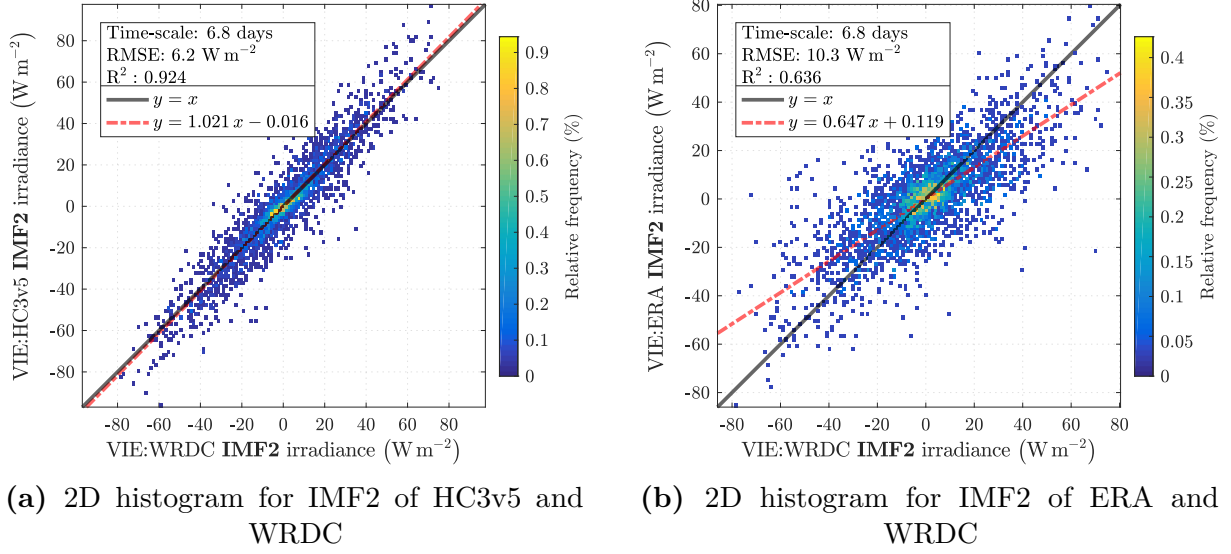
Figure 5.4 The 2D histogram for IMF2 for WRDC and (a) HC3v5, and (b) ERA, for VIE. Each pixel encodes relative frequency according to the colour-bar on the right. The solid black line denotes the identity line and the dash-dotted red line represents the best fit line. The linear regression equation is indicated in the legend. The time-scale, root-mean-square error and coefficient of determination are indicated in the panel above the legend.

Figure 5.4b presents a similar graph, but for ERA and WRDC datasets. Here the best-fit line deviates significantly from the identity line, with a slope of 0.647 and an intercept of 0.119. This is also reflected in the coefficient of determination of $R^2 = 0.636$. The scatter is also significantly larger, with a RMSE of 10.3 $\mathrm{W\,m^{-2}}$.

Similar plots to those in figure 5.4 have been computed for IMF1,...,IMF5, IMF7 and globally, for the whole time-series, for the ground measurements and for the satellite and re-analysis estimates. Graphs are available in the supplementary material. Results, in terms of root-mean-square error and coefficient of determination, are indicated in tables 5.5 and 5.6 for VIE and KIV, respectively. IMF6 has been excluded, because tables 5.1 and 5.3 show that the time-scale of this mode exhibits significant disparities between the re-analyses and the ground measurements, thus the comparison is meaningless. IMF8,...,IMF10 have also been excluded because the mean amplitudes are below the uncertainty level (see tables 5.2 and 5.4).

Table 5.5 shows that for VIE, on a per time-scale basis as well as globally, the closest estimate of ground measurements of the SSI is the HC3v5 dataset, both in terms of explained variance, and in terms of scatter. The lowest coefficient of determination for HC3v5 is 0.924 for the weekly variability (IMF2), meaning that figure 5.4a represents the worst case scenario for this particular dataset. The largest percentage of variance explained, 99.2%, is attained for the yearly variability (IMF7). Globally, the HC3v5 accounts for 97.9% of the observed variability (RMSE = 14.1 $\mathrm{W\,m^{-2}}$), outclassing ERA with 92.1% (RMSE = 24.8 $\mathrm{W\,m^{-2}}$) and MERRA2 with 92.8% (RMSE = 26.4 $\mathrm{W\,m^{-2}}$).

Table 5.5 Statistical indicators for correlations at different time-scales between SSI estimates and ground measurements for VIE

		IMF1	IMF2	IMF3	IMF4	IMF5	IMF7	Global
R^2								
	HC3v5	0.937	0.924	0.945	0.938	0.934	0.992	0.979
	ERA	0.639	0.636	0.675	0.755	0.703	0.987	0.921
	MERRA2	0.684	0.686	0.697	0.732	0.762	0.985	0.928
RMSE ($\mathrm{W\,m^{-2}}$)								
	HC3v5	9.1	6.2	4.2	3.6	2.8	5.7	14.1
	ERA	15.5	10.3	7.8	5.5	4.6	7.0	24.8
	MERRA2	16.1	9.7	7.5	5.6	4.3	8.4	26.4

Table 5.6 Statistical indicators for correlations at different time-scales between SSI estimates and ground measurements for KIV

		IMF1	IMF2	IMF3	IMF4	IMF5	IMF7	Global
R^2								
	HC3v5	0.926	0.924	0.929	0.921	0.885	0.992	0.984
	ERA	0.611	0.677	0.692	0.753	0.670	0.984	0.931
	MERRA2	0.621	0.643	0.671	0.800	0.590	0.982	0.934
RMSE ($\mathrm{W\,m^{-2}}$)								
	HC3v5	8.5	5.7	4.5	4.0	3.1	6.3	12.6
	ERA	13.9	8.7	6.5	5.2	4.5	8.4	24.5
	MERRA2	16.4	10.1	7.4	5.1	4.8	8.5	26.2

For IMF1,...,IMF5, MERRA2 outperforms ERA in terms of R^2 except for the monthly variability (IMF4), as also reflected in the range of this coefficient of [0.684; 0.762] for MERRA2 as opposed to a range of [0.636; 0.755] for ERA. The yearly variability of the ground measurements is better expressed by ERA than by MERRA2, both in terms of coefficient of determination and RMSE. Generally, all the datasets exhibit monotonically decreasing RMSE for the first five modes and very good agreement for the yearly variability ($R^2 > 0.985$).

Similar statements can be made about the results for KIV, presented in table 5.6. Here too, the HC3v5 dataset outperforms ERA and MERRA2 both in terms of coefficient of determination and RMSE, across all time-scales and also at the whole time-series level. The minimum R^2 for HC3v5 is 0.885 for IMF5, while the minimum values of R^2 for ERA end MERRA occur for IMF1 and are 0.611 and 0.621, respectively. As for VIE, all the

datasets exhibit monotonically decreasing RMSE for the first five modes and very good agreement for the yearly variability ($R^2 > 0.982$).

Conclusion   In this chapter the characteristic time-scales of variability found in long-term time-series of daily means of ground measurements or estimates of SSI have been investigated. The fitness for use of satellite estimates of the SSI and radiation products re-analyses as alternatives to pyranometric ground measurements has also studied. The novelty of this approach is the use of the adaptive, data-driven Hilbert-Huang Transform (HHT) to decompose the datasets into their distinct characteristic time-scales of variability before undergoing analysis.

It has been shown that the TOA only presents variability at the one year time-scale. The clear-sky atmosphere introduces stochastic high frequency variability, from 2 days to 2–3 months, which exhibits non-linear cross-scale phase-amplitude coupling with the yearly cycle. This feature is also present, and amplified, in ground measurements, satellite estimates and re-analysis products. The fact that the cloud-free atmosphere does not introduce variability from 2–3 months to one year, i.e. the "variability gap" alluded to in previous studies, has been confirmed. It has also been shown that, HC3v5 outperforms ERA and MERRA2 by a large margin in terms of estimating the measured SSI, not only at a global, whole dataset level, but also on a per time-scale basis, and especially with respect to the stochastic variability component. This has implications on the forecast and modelling of the SSI, where satellite estimates should be preferred instead of re-analysis products. This study, hence, refines the existing methodology to assess the fitness for use of surrogate SSI products, through an improved in-depth comparison of their local time-scales of variability.

CHAPTER 6

Conclusion and outlook

W ITH the end of the previous chapter, which marks the end of the original contributions brought forth, the study can now be briefly summarized as follows.

Following the introductory part, chapter 2 has described the SSI datasets that were at the foundation of this work. As such, the focus of analysis has consisted in long-term time-series of daily means of the SSI, spanning approximately ten years. It has been seen that the data have played a crucial role in this work, since the analysis of the characteristic time-scales of variability of the SSI demandes accurate measurements thereof. Nevertheless, in the absence of such high-quality data, as it often the case in practice, surrogate time-series, such as satellite estimates or radiation products from numerical weather re-analyses, may be employed. Consequently, such datasets have also made the object of scrutiny.

A brief overview of the adaptive data analysis paradigm of the Hilbert-Huang transform along with a short review of classical time-frequency analysis techniques have been offered in chapter 3. By means of a case study, it has been shown that both the CWT and the HHT are well suited for capturing the characteristic time-scales of variability of the SSI. It has been proposed, however, the HHT is slightly better suited for the task, in the sense is able to pick up more subtle features, such as the yearly period modulation, while also yielding a sparser representation of the data in the time-frequency plane. This makes it a useful tool for the analysis of the temporal variability of the surface solar radiation, as shown in [Bengulescu et al., 2016a].

Chapter 4 has dealt with the temporal variability of ground measurements of the SSI at different BSRN measuring stations around the world. The HHT data-driven analysis was showcased on revealing and analysing the temporal variability of high-quality measurements of the SSI. By means of this adaptive data analysis paradigm, new insights have been gained into the variability of the SSI, allowing e.g. the identification of the type of climate based on the energetic intensity of certain time-scales — a key assessment for

the initial site selection or the design and financing stages of a solar power plant [Sengupta et al., 2015].

Another significant finding is that, irrespective of geographical location, the SSI exhibits high-frequency quasi-stochastic variability components which are amplitude-modulated by the deterministic annual cycle, through a non-linear cross-scale interaction. These outcomes are documented in [Bengulescu et al., 2016c; Bengulescu et al., 2016d]. The statistical analysis of the intrinsic modes of variability for discriminating the deterministic components from the quasi-stochastic ones, needs not be limited to studies of the SSI, however, since the same methodology could also be employed on different geophysical signals, such as wind speed time-series, river discharge datasets, etc.

Furthermore, when modelling climate processes as dynamical systems with low-frequency oscillations and stochastic components, Chekroun et al., [2011] have shown that "even the 'approximately right' noise can help, rather than hinder". Here, a recipe was provided not only for extracting, but also for characterizing the quasi-stochastic constituents of long-term time-series of the SSI. Indications with respect to modelling these components have also been provided. With respect to SSI forecast models, it is exactly these components that are the focus of attention [Ehnberg and Bollen, 2005; Hoff and Perez, 2010; Marquez and Coimbra, 2013]. Inman et al., 2013 venture as far as stating that "the accuracy of the solar irradiance forecasting models depends almost exclusively on the ability to forecast the stochastic component". In this light, the recipe for identifying the stochastic components that has been put forth can be seen as one of the most significant contributions of this chapter. Here, some references can also be made to the existing literature where EMD-based forecasting schemes are described as first decomposing the time-series of interest into its constituent IMFs, followed by various forecasting schemes applied on a per-IMF basis, with the final forecast time-series result being the aggregation of the previous, per-time-scale predictors. Such approaches have been employed for forecasting the load factor of the electrical power grid [Zhu et al., 2007; Fan et al., 2013], the short-term price of electricity [Dong et al., 2011], the wind speed [Guo et al., 2012] and the month-ahead wind farm output power [Xiaolan and Hui, 2009], and various financial time-series, such as prices for oil futures contracts [Hong, 2011] or currency exchange rates [Fu, 2010]. Applying such forecasting schemes to time-series of SSI should be achievable without significant effort.

It has also been proposed that a classification of the measuring stations according to climate and/or solar insolation conditions may be possible, based on the Hilbert spectral features of the data. Thus, one future research pathway could consist in creating a catalogue of the variability of the solar resource, at different time-scales, on a global scale via satellite estimates of the SSI. Current meteorological re-analyses are too noisy in their estimates of the SSI to form the basis for such a catalogue [Boilley and Wald, 2015].

In terms of solar power production, the low-frequency variability data would aid with policy and investment decisions, while short-term variability would be of interest from a monitoring, operations and engineering perspective.

Lastly, chapter 5 has assessed the fitness for use of surrogate SSI data when pyranometric ground measurements are unavailable and has also briefly discussed the role of atmospherically-induced variability in satellite estimates of solar radiation. The novelty of the approach consists in the fact that the fitness for use of the modelled and estimated SSI has been assessed not only at a global, whole dataset level, but also on a per time-scale basis. A practical outcome of this analysis carried out in this chapter is that practitioners interested in modelling and forecasting the daily SSI will do better using satellite estimates that radiation products from re-analyses, at least for the two sites under consideration, as described in [Bengulescu et al., 2017]. Of course, the restrained geographical character of the analysis has to be acknowledged and, as a future exercise, it is proposed to extend it to many more geographical locations and possibly including several different satellite estimates and re-analyses radiation products, in order to determine whether the findings reported herein also hold for different regions and for different SSI surrogates.

Another limitation of the approach from chapter 5 needs to be pointed out. Before carrying out the analysis, EMD has been employed on each time-series and only modes with similar time-scales have been compared. That is, the mono-variate version of the EMD has been used, where mode alignment (identical time-scales for the IMFs across datasets) is not enforced. Nevertheless the non-alignment of modes is not to be considered a weakness of the approach. Because identical time-series will be decomposed into identical modes, by not enforcing similar time-scales across the modes of different datasets, changes in the time-scales of the modes also provide supplementary clues as to the fitness for use of the surrogate SSI datasets in lieu of ground measurements. However, mode alignment can be enforced by more advanced, multi-variate versions of the EMD. Two such techniques are the noise-assisted multi-variate empirical mode decomposition (NA-MEMD) introduced by ur Rehman et al., [2013] or the adaptive-projection intrinsically transformed multivariate empirical mode decomposition (APIT-MEMD) proposed by Hemakom et al., [2016]. The latter method is particularly of interest since it is also able to deal with power imbalances and inter-channel correlations found in multichannel data. With NA-MEMD or APIT-MEMD, all the datasets would be treated in a unitary manner as a single multi-variate signal, thus mode alignment would be enforced. This would also enable the use of of more advanced descriptors, such as multi-scale measures suitable for multi-variate datasets and inter-component measures, e.g. intrinsic correlation, intrinsic sample entropy or intrinsic phase synchrony [Looney et al., 2015]. However, the exercise is significantly more technical and is proposed as a future study.

Lastly, it must be pointed out that, although insightful, the work at hand represents but a minor personal contribution of the author to the vast knowledge of the SSI variability already accumulated, and that further research efforts are needed for the refinement, development and implementation of innovative techniques and approaches that will ultimately permit a better characterization of the temporal variability of the solar resource.

Bibliography

ABDEL WAHAB, M., EL-METWALLY, M., HASSAN, R., LEFÈVRE, M., OUMBE, A., and WALD, L. (2010). "Assessing surface solar irradiance and its long-term variations in the northern Africa desert climate using Meteosat images". In: *International Journal of Remote Sensing* vol. 31, no. 1, pp. 261–280 (cit. on p. 10). doi:10/ck533g.

BAZILEVSKAYA, G., BROOMHALL, A.-M., ELSWORTH, Y., and NAKARIAKOV, V. (2015). "A combined analysis of the observational aspects of the quasi-biennial oscillation in solar magnetic activity". In: *The Solar Activity Cycle*. Springer, pp. 359–386 (cit. on p. 61). doi:10/bcz3.

BEER, J. (2000). "Long-term indirect indices of solar variability". In: *Space Science Reviews* vol. 94, no. 1-2, pp. 53–66 (cit. on p. 8). doi:10/d8pzkn.

BEER, J., VONMOOS, M., and MUSCHELER, R. (2006). "Solar variability over the past several millennia". In: *Space Science Reviews* vol. 125, no. 1-4, pp. 67–79 (cit. on pp. 7, 8). doi:10/djw2kj.

BENGULESCU, M., BLANC, P., and WALD, L. (2015). "Assessing the temporal variability of the surface solar radiation with time-frequency-energy representations". In: *15th EMS annual meeting*. Vol. 12. Sofia, Bulgaria, pp. 2015–230 (cit. on p. 104). URL: http://meetingorganizer.copernicus.org/EMS2015/EMS2015-230.pdf.

BENGULESCU, M., BLANC, P., and WALD, L. (2016a). "On the temporal variability of the surface solar radiation by means of spectral representations". In: *Advances in Science and Research* vol. 13, pp. 121–127 (cit. on pp. 11, 12, 78, 83, 104). doi:10/bp6g.

BENGULESCU, M., BLANC, P., and WALD, L. (2016b). "Adaptive data analysis for characterizing the temporal variability of the solar resource". In: *European Geosciences Union General Assembly 2016*. Vol. 18. Geophysical Research Abstracts. European Geosciences Union. Vienna, Austria, EGU2016–14847 (cit. on p. 105). URL: http://meetingorganizer.copernicus.org/EGU2016/EGU2016-14847.pdf.

Bengulescu, M., Blanc, P., and Wald, L. (2016c). "On the intrinsic time-scales of temporal variability in measurements of the surface solar radiation". In: *Nonlinear Processes in Geophysics Discussions* vol. 2016. in review, pp. 1–35 (cit. on pp. 11, 12, 78, 84, 104). doi:10/brwj.

Bengulescu, M., Blanc, P., and Wald, L. (2016d). "Characterizing temporal variability in measurements of surface solar radiation and its dependence on climate". In: *Energy Procedia* vol. 97. European Geosciences Union General Assembly 2016, EGU Division Energy, Resources & the Environment (ERE), pp. 164–171 (cit. on pp. 11, 12, 84, 104). doi:10/bt46.

Bengulescu, M., Blanc, P., and Wald, L. (2016e). "Hilbert-Huang spectral analysis for characterizing the intrinsic time-scales of variability in decennial time-series of surface solar radiation". In: *European Geosciences Union General Assembly 2016*. Vol. 18. Geophysical Research Abstracts. European Geosciences Union. Vienna, Austria, EGU2016–14549 (cit. on p. 105). URL: http://meetingorganizer.copernicus.org/EGU2016/EGU2016-14549.pdf.

Bengulescu, M., Blanc, P., and Wald, L. (2016f). "Hilbert-Huang spectral analysis for the characterization of variability in satellite-derived time series of surface solar irradiance". In: *16th EMS Annual Meeting*. Vol. 13. European Meteorological Society. Trieste, Italy, pp. 2016–372 (cit. on p. 105). URL: http://meetingorganizer.copernicus.org/EMS2016/EMS2016-372.pdf.

Bengulescu, M., Blanc, P., Boilley, A., and Wald, L. (2017). "Do modelled or satellite-based estimates of surface solar irradiance accurately describe its temporal variability?" In: *Advances in Science and Research* vol. 14, pp. 35–48 (cit. on pp. 11, 12, 85, 104). doi:10/b37n.

Bergland, G. (1969). "A guided tour of the fast Fourier transform". In: *Spectrum, IEEE* vol. 6, no. 7, pp. 41–52 (cit. on p. 23). doi:10/fgv77x.

Blanc, P., Gschwind, B., Lefèvre, M., and Wald, L. (2011). "The HelioClim project: Surface solar irradiance data for climate applications". In: *Remote Sensing* vol. 3, no. 2, pp. 343–361 (cit. on p. 19). doi:10/chpk5b.

Blanc, P. and Wald, L. (2012). "The SG2 algorithm for a fast and accurate computation of the position of the Sun for multi-decadal time period". In: *Solar Energy* vol. 86, no. 10, pp. 3072–3083 (cit. on p. 17). doi:10/2h9.

Blanc, P., Coulaud, C., and Wald, L. (2015). "Yearly changes in surface solar radiation in New Caledonia". In: *Advances in Science and Research* vol. 12, no. 1, pp. 1–4 (cit. on p. 10). doi:10/brvs.

Boeing, G. (2016). "Visual Analysis of Nonlinear Dynamical Systems: Chaos, Fractals, Self-Similarity and the Limits of Prediction". In: *Systems* vol. 4, no. 4, p. 37 (cit. on p. 23). doi:10/b6rj.

BOILLEY, A. and WALD, L. (2015). "Comparison between meteorological re-analyses from ERA-Interim and MERRA and measurements of daily solar irradiation at surface". In: *Renewable Energy* vol. 75, pp. 135–143 (cit. on pp. 12, 70, 78, 84). doi:10/f259gv.

BOJINSKI, S., VERSTRAETE, M., PETERSON, T. C., RICHTER, C., SIMMONS, A., and ZEMP, M. (2014). "The concept of Essential Climate Variables in support of climate research, applications, and policy". In: *Bulletin of the American Meteorological Society* vol. 95, no. 9, pp. 1431–1443 (cit. on p. 8). doi:10/bcq2.

BSRN station listing (2015) (cit. on p. 16). URL: https://www.pangaea.de/ddi?request= bsrn/BSRNEvent&format=html&title=BSRN+Stations (accessed on December 9ᵗʰ 2015).

CALIF, R., SCHMITT, F. G., HUANG, Y., and SOUBDHAN, T. (2013). "Intermittency study of high frequency global solar radiation sequences under a tropical climate". In: *Solar Energy* vol. 98, pp. 349–365 (cit. on pp. 10, 11). doi:10/7xg.

CALIF, R., SCHMITT, F. G., and HUANG, Y. (2016). "Study of Local Correlations of the Simultaneous Wind Speed-Irradiance Measurements Using the Time Dependent Intrinsic Correlation Method". In: *Journal of Applied Nonlinear Dynamics* vol. 5, no. 4, pp. 373–390 (cit. on p. 11). doi:10/cdqz.

CHEKROUN, M. D., KONDRASHOV, D., and GHIL, M. (2011). "Predicting stochastic systems by noise sampling, and application to the El Niño-Southern Oscillation". In: *Proceedings of the National Academy of Sciences* vol. 108, no. 29, pp. 11766–11771 (cit. on p. 84). doi:10/bpt5kk.

CHEN, X., WANG, M., ZHANG, Y., FENG, Y., WU, Z., and HUANG, N. E. (2013). "Detecting Signals from Data with Noise: Theory and Applications". In: *Journal of the Atmospheric Sciences* vol. 70, no. 5, pp. 1489–1504 (cit. on pp. 11, 63, 64, 66, 67). doi:10/bb6d.

COCHRAN, W. T., COOLEY, J. W., FAVIN, D. L., HELMS, H. D., KAENEL, R. A., LANG, W. W., MALING JR, G. C., NELSON, D. E., RADER, C. M., and WELCH, P. D. (1967). "What is the fast Fourier transform?" In: *Proceedings of the IEEE* vol. 55, no. 10, pp. 1664–1674 (cit. on p. 22). doi:10/fhw8mm.

COHEN, L. (1989). "Time-frequency distributions-a review". In: *Proceedings of the IEEE* vol. 77, no. 7, pp. 941–981 (cit. on pp. 11, 20). doi:10/d2hp7f.

COLOMINAS, M. A., SCHLOTTHAUER, G., TORRES, M. E., and FLANDRIN, P. (2012). "Noise-assisted EMD methods in action". In: *Advances in Adaptive Data Analysis* vol. 4, no. 04 (cit. on pp. 41, 63). doi:10/9gp.

COLOMINAS, M. A., SCHLOTTHAUER, G., and TORRES, M. E. (2014). "Improved complete ensemble EMD: A suitable tool for biomedical signal processing". In: *Biomedical Signal Processing and Control* vol. 14, pp. 19–29 (cit. on pp. 40, 41). doi:10/7tp.

89

COOLEY, J. W. and TUKEY, J. W. (1965). "An algorithm for the machine calculation of complex Fourier series". In: *Mathematics of computation* vol. 19, no. 90, pp. 297–301 (cit. on p. 21). doi:10/btpkk5.

COOLEY, J. W., LEWIS, P. A., and WELCH, P. D. (1967). "Historical notes on the fast Fourier transform". In: *Proceedings of the IEEE* vol. 55, no. 10, pp. 1675–1677 (cit. on p. 21). doi:10/dgtzx6.

COSKUN, C., OKTAY, Z., and DINCER, I. (2011). "Estimation of monthly solar radiation distribution for solar energy system analysis". In: *Energy* vol. 36, no. 2, pp. 1319–1323 (cit. on p. 10). doi:10/cwbf42.

DONG, Y., WANG, J., JIANG, H., and WU, J. (2011). "Short-term electricity price forecast based on the improved hybrid model". In: *Energy Conversion and Management* vol. 52, no. 8, pp. 2987–2995 (cit. on p. 84). doi:10/df22qm.

DUFFY, D. G. (2004). "The application of Hilbert-Huang transforms to meteorological datasets". In: *Journal of Atmospheric and Oceanic Technology* vol. 21, no. 4, pp. 599–611 (cit. on pp. 10, 11). doi:10/b9td58.

EHNBERG, J. S. and BOLLEN, M. H. (2005). "Simulation of global solar radiation based on cloud observations". In: *Solar Energy* vol. 78, no. 2, pp. 157–162 (cit. on p. 84). doi:10/chvffx.

ESPINAR, B., RAMÍREZ, L., DREWS, A., BEYER, H. G., ZARZALEJO, L. F., POLO, J., and MARTÍN, L. (2009). "Analysis of different comparison parameters applied to solar radiation data from satellite and German radiometric stations". In: *Solar Energy* vol. 83, no. 1, pp. 118–125 (cit. on pp. 12, 71). doi:10/ftmgqt.

EUROPEAN CENTRE FOR MEDIUM-RANGE WEATHER FORECASTS (2009). *ERA-Interim Project*. Research Data Archive at the National Center for Atmospheric Research, Computational and Information Systems Laboratory (cit. on p. 19). doi:10/bstv. (accessed on October 27th 2016).

FAN, G.-F., QING, S., WANG, H., HONG, W.-C., and LI, H.-J. (2013). "Support Vector Regression Model Based on Empirical Mode Decomposition and Auto Regression for Electric Load Forecasting". In: *Energies* vol. 6, no. 4, pp. 1887–1901 (cit. on p. 84). doi:10/cdrp.

FARGE, M. (1992). "Wavelet transforms and their applications to turbulence". In: *Annual Review of Fluid Mechanics* vol. 24, no. 1, pp. 395–458 (cit. on pp. 27, 28, 30). doi:10/dzfvdn.

FLANDRIN, P., RILLING, G., and GONÇALVES, P. (2004). "Empirical mode decomposition as a filter bank". In: *Signal Processing Letters, IEEE* vol. 11, no. 2, pp. 112–114 (cit. on pp. 40, 63). doi:10/c9g7q3.

Flandrin, P. and Gonçalves, P. (2004). "Empirical mode decompositions as data-driven wavelet-like expansions". In: *International Journal of Wavelets, Multiresolution and Information Processing* vol. 2, no. 04, pp. 477–496 (cit. on pp. 40, 61, 63). doi:10/c9cswc.

Flandrin, P., Gonçalvès, P., and Rilling, G. (2005). "EMD equivalent filter banks, from interpretation to applications". In: *Hilbert-Huang Transform and Its Applications.* World Scientific Pub Co Pte Lt, pp. 57–74 (cit. on p. 63). doi:10/dxqw6v.

Flandrin, P. (2015). Personal communication (cit. on p. 63).

Franzke, C. (2009). "Multi-scale analysis of teleconnection indices: climate noise and nonlinear trend analysis". In: *Nonlinear Processes in Geophysics* vol. 16, no. 1, pp. 65–76 (cit. on p. 11). doi:10/fsgrzt.

Franzke, C. (2012). "Nonlinear trends, long-range dependence, and climate noise properties of surface temperature". In: *Journal of Climate* vol. 25, no. 12, pp. 4172–4183 (cit. on pp. 11, 52). doi:10/fx8xk3.

Fu, C. (2010). "Forecasting Exchange Rate with EMD-Based Support Vector Regression". In: *2010 International Conference on Management and Service Science*, pp. 1–4 (cit. on p. 84). doi:10/d2tm43.

Gabor, D. (1946). "Theory of communication. Part 1: The analysis of information". In: *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering* vol. 93, no. 26, pp. 429–441 (cit. on pp. 24, 26, 33–35). doi:10/2rz.

Gallagher, J. and Biscoe, P. (1978). "Radiation absorption, growth and yield of cereals". In: *The Journal of Agricultural Science* vol. 91, no. 01, pp. 47–60 (cit. on p. 7). doi:10/cp7hdk.

Ghil, M., Allen, M. R., Dettinger, M. D., Ide, K., Kondrashov, D., Mann, M. E., Robertson, A. W., Saunders, A., Tian, Y., Varadi, F., and Yiou, P. (2002). "Advanced spectral methods for climatic time series". In: *Reviews of Geophysics* vol. 40, no. 1, (3)1–(3)41 (cit. on p. 21). doi:10/d9fqvk.

Global Climate Observing System (2010). *Implementation Plan for the Global Observing System for Climate in Support of the UNFCCC (2010 Update).* Tech. rep. GCOS-138 (GOOS-184, GTOS-76, WMO-TD/No. 1523). Geneva, Switzerland: World Meteorological Organization (cit. on p. 7). URL: http://www.wmo.int/pages/prog/gcos/Publications/gcos-138.pdf (accessed on February 23rd 2017).

Global Modeling and Assimilation Office (2015). *MERRA-2 tavg1_2d_rad_Nx: 2D, 1-Hourly, Time-Averaged, Single-Level, Assimilation, Radiation Diagnostics v5.12.4.* Greenbelt, MD, USA, Goddard Earth Sciences Data and Information Services Center (GES DISC) (cit. on p. 19). doi:10/bstw. (accessed on October 27th 2016).

Goupillaud, P., Grossmann, A., and Morlet, J. (1984). "Cycle-octave and related transforms in seismic signal analysis". In: *Geoexploration* vol. 23, no. 1, pp. 85–102 (cit. on pp. 26, 27, 30). doi:10/fq6vcz.

GRATTAN-GUINNESS, I. (1969). "Joseph Fourier and the Revolution in Mathematical Physics". In: *IMA Journal of Applied Mathematics* vol. 5, no. 2, pp. 230–253 (cit. on p. 21). doi:10/bh9qhs.

GRINSTED, A., MOORE, J. C., and JEVREJEVA, S. (2004). "Application of the cross wavelet transform and wavelet coherence to geophysical time series". In: *Nonlinear Processes in Geophysics* vol. 11, no. 5/6, pp. 561–566 (cit. on pp. 27, 29–31). doi:10/d6dhzv.

GROSSMANN, A. and MORLET, J. (1984). "Decomposition of Hardy functions into square integrable wavelets of constant shape". In: *SIAM journal on mathematical analysis* vol. 15, no. 4, pp. 723–736 (cit. on pp. 21, 26, 30). doi:10/b9ddn3.

GUO, Z., ZHAO, W., LU, H., and WANG, J. (2012). "Multi-step forecasting for wind speed using a modified EMD-based artificial neural network model". In: *Renewable Energy* vol. 37, no. 1, pp. 241–249 (cit. on p. 84). doi:10/cj9d36.

HAMMER, A. and BEYER, H. (2012). "Solar Radiation, Spatial and Temporal Variability". English. In: *Encyclopedia of Sustainability Science and Technology.* Ed. by R. A. MEYERS. Springer New York, pp. 9744–9758 (cit. on p. 9). doi:10/7wd.

HARRIS, F. J. (1978). "On the use of windows for harmonic analysis with the discrete Fourier transform". In: *Proceedings of the IEEE* vol. 66, no. 1, pp. 51–83 (cit. on pp. 23, 24, 64). doi:10/bks35p.

HARRISON, R. G. (2008). "Discrimination between cosmic ray and solar irradiance effects on clouds, and evidence for geophysical modulation of cloud thickness". In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences.* Vol. 464. 2098. The Royal Society, pp. 2575–2590 (cit. on p. 61). doi:10/dqsd4x.

HATHAWAY, D. H. (2015). "The Solar Cycle". In: *Living Reviews in Solar Physics* vol. 12, no. 1, pp. 1–87 (cit. on p. 56). doi:10/bjxj.

HEIDEMAN, M., JOHNSON, D. H., and BURRUS, C. S. (1984). "Gauss and the history of the fast Fourier transform". In: *IEEE ASSP Magazine* vol. 1, no. 4, pp. 14–21 (cit. on p. 21). doi:10/fvnfs9.

HEMAKOM, A., GOVERDOVSKY, V., LOONEY, D., and MANDIC, D. P. (2016). "Adaptive-projection intrinsically transformed multivariate empirical mode decomposition in cooperative brain–computer interface applications". In: *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* vol. 374, no. 2065, p. 20150199 (cit. on p. 85). doi:10/bvjm.

HERLEY, C., KOVACEVIC, J., RAMCHANDRAN, K., and VETTERLI, M. (1993). "Tilings of the time-frequency plane: Construction of arbitrary orthogonal bases and fast tiling algorithms". In: *Signal Processing, IEEE Transactions on* vol. 41, no. 12, pp. 3341–3359 (cit. on p. 25). doi:10/b874vt.

Hinich, M. J. and Clay, C. S. (1968). "The application of the discrete Fourier transform in the estimation of power spectra, coherence, and bispectra of geophysical data". In: *Reviews of Geophysics* vol. 6, no. 3, pp. 347–363 (cit. on p. 21). doi:10/dx8gh4.

Hoff, T. E. and Perez, R. (2010). "Quantifying PV power output variability". In: *Solar Energy* vol. 84, no. 10, pp. 1782–1793 (cit. on pp. 9, 84). doi:10/dm7spw.

Hong, L. (2011). "Decomposition and Forecast for Financial Time Series with High-frequency Based on Empirical Mode Decomposition". In: *Energy Procedia* vol. 5, no. Supplement C. 2010 International Conference on Energy, Environment and Development - ICEED2010, pp. 1333–1340 (cit. on p. 84). doi:10/cjzqmh.

Hoyer-Klick, C., Lefèvre, M., Schroedter-Homscheidt, M., and Wald, L. (2015). *User's guide to the MACC-RAD services on solar energy radiation resources* (cit. on p. 18). doi:10.13140/RG.2.1.5016.7521.

Huang, J., Troccoli, A., and Coppin, P. (2014). "An analytical comparison of four approaches to modelling the daily variability of solar irradiance using meteorological records". In: *Renewable Energy* vol. 72, pp. 195–202 (cit. on p. 11). doi:10/f6jr24.

Huang, N. E., Shen, Z., Long, S. R., Wu, M. C., Shih, H. H., Zheng, Q., Yen, N.-C., Tung, C. C., and Liu, H. H. (1998). "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis". In: *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* vol. 454, no. 1971, pp. 903–995 (cit. on pp. 10, 11, 21, 32–37, 39, 48). doi:10/dr5ngd.

Huang, N. E., Shen, Z., and Long, S. R. (1999). "A new view of nonlinear water waves: The Hilbert Spectrum". In: *Annual Review of Fluid Mechanics* vol. 31, no. 1, pp. 417–457 (cit. on pp. 33, 39). doi:10/dj3v2z.

Huang, N. E., Wu, M.-L. C., Long, S. R., Shen, S. S., Qu, W., Gloersen, P., and Fan, K. L. (2003). "A confidence limit for the empirical mode decomposition and Hilbert spectral analysis". In: *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* vol. 459, no. 2037, pp. 2317–2345 (cit. on pp. 39, 54, 62, 63). doi:10/c8wsrv.

Huang, N. E. and Wu, Z. (2008). "A review on Hilbert-Huang transform: Method and its applications to geophysical studies". In: *Reviews of Geophysics* vol. 46, no. 2 (cit. on pp. 11, 31–34, 36, 37, 39). doi:10/fr2hk2.

Huang, N. E., Wu, Z., Long, S. R., Arnold, K. C., Chen, X., and Blank, K. (2009). "On instantaneous frequency". In: *Advances in Adaptive Data Analysis* vol. 1, no. 02, pp. 177–229 (cit. on p. 42). doi:10/dw9j3k.

Huang, N. E., Chen, X., Lo, M.-T., and Wu, Z. (2011). "On Hilbert spectral representation: a true time-frequency representation for nonlinear and nonstationary data".

In: *Advances in Adaptive Data Analysis* vol. 3, no. 01n02, pp. 63–93 (cit. on pp. 33, 34, 45). doi:10/b7m8xt.

Huang, N. E., Young, V., Lo, M., Wang, Y. H., Peng, C.-K., Chen, X., Wang, G., Deng, J., and Wu, Z. (2013). "The uniqueness of the instantaneous frequency based on intrinsic mode function". In: *Advances in Adaptive Data Analysis* vol. 5, no. 03, p. 1350011 (cit. on p. 37). doi:10/bx8p.

Huang, N. E., Hu, K., Yang, A. C., Chang, H.-C., Jia, D., Liang, W.-K., Yeh, J. R., Kao, C.-L., Juan, C.-H., Peng, C. K., et al. (2016). "On Holo-Hilbert spectral analysis: a full informational spectral representation for nonlinear and non-stationary data". In: *Phil. Trans. R. Soc. A* vol. 374, no. 2065, p. 20150206 (cit. on p. 67). doi:10/bhhc.

Ineichen, P. (2014). "Long term satellite global, beam and diffuse irradiance validation". In: *Energy Procedia* vol. 48. Proceedings of the 2nd International Conference on Solar Heating and Cooling for Buildings and Industry (SHC 2013), pp. 1586–1596 (cit. on p. 12). doi:10/b2j4.

Inman, R. H., Pedro, H. T., and Coimbra, C. F. (2013). "Solar forecasting methods for renewable energy integration". In: *Progress in Energy and Combustion Science* vol. 39, no. 6, pp. 535–576 (cit. on pp. 78, 84). doi:10/brhc.

Kendall, M. G. (1938). "A new measure of rank correlation". In: *Biometrika* vol. 30, no. 1-2, pp. 81–93 (cit. on p. 66). doi:10/fxdpmd.

Kirkby, J., Curtius, J., Almeida, J., Dunne, E., Duplissy, J., Ehrhart, S., Franchin, A., Gagné, S., Ickes, L., Kürten, A., et al. (2011). "Role of sulphuric acid, ammonia and galactic cosmic rays in atmospheric aerosol nucleation". In: *Nature* vol. 476, no. 7361, pp. 429–433 (cit. on p. 61). doi:10/cszcxm.

Kirkby, J., Duplissy, J., Sengupta, K., Frege, C., Gordon, H., Williamson, C., Heinritzi, M., Simon, M., Yan, C., Almeida, J., et al. (2016). "Ion-induced nucleation of pure biogenic particles". In: *Nature* vol. 533, no. 7604, pp. 521–526 (cit. on p. 61). doi:10/f3r2zr.

Kolotkov, D., Broomhall, A.-M., and Nakariakov, V. (2015). "Hilbert–Huang transform analysis of periodicities in the last two solar activity cycles". In: *Monthly Notices of the Royal Astronomical Society* vol. 451, no. 4, pp. 4360–4367 (cit. on pp. 60, 61, 67). doi:10/bcxn.

Kolotkov, D., Anfinogentov, S. A., and Nakariakov, V. M. (2016). "Empirical mode decomposition analysis of random processes in the solar atmosphere". In: *Astronomy and Astrophysics* vol. 592, A153 (cit. on pp. 61, 67). doi:10/br7c.

König-Langlo, G., Driemel, A., Raffel, B., and Sieger, R. (2015). *BSRN snapshot 2015-09, links to zip archives.* data set (cit. on p. 17). doi:10/brqd. (accessed on February 23rd 2017).

Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F. (2006). "World map of the Köppen-Geiger climate classification updated". In: *Meteorologische Zeitschrift* vol. 15, no. 3, pp. 259–263 (cit. on p. 17). doi:10/c37229.

Kumar, P. and Foufoula-Georgiou, E. (1997). "Wavelet analysis for geophysical applications". In: *Reviews of Geophysics* vol. 35, no. 4, pp. 385–412 (cit. on pp. 27, 30, 31). doi:10/cwrzz9.

Labat, D. (2005). "Recent advances in wavelet analyses: Part 1. A review of concepts". In: *Journal of Hydrology* vol. 314, no. 1-4, pp. 275–288 (cit. on p. 31). doi:10/bj6ttr.

Lauret, P., Perez, R., Aguiar, L. M., Tapachès, E., Diagne, H. M., and David, M. (2016). "Characterization of the intraday variability regime of solar irradiation of climatically distinct locations". In: *Solar Energy* vol. 125, pp. 99–110 (cit. on p. 10). doi:10/bcrf.

Lave, M., Kleissl, J., and Stein, J. S. (2013). "A wavelet-based variability model (WVM) for solar PV power plants". In: *Sustainable Energy, IEEE Transactions on* vol. 4, no. 2, pp. 501–509 (cit. on p. 10). doi:10/8kq.

Lean, J. L. (2010). "Cycles and trends in solar irradiance and climate". In: *Wiley Interdisciplinary Reviews: Climate Change* vol. 1, no. 1, pp. 111–122 (cit. on pp. 8, 18). doi:10/ffnr66.

Lee, J. N., Cahalan, R. F., and Wu, D. L. (2015). "The 27-day rotational variations in total solar irradiance observations: From SORCE/TIM, ACRIMSAT/ACRIM III, and SOHO/VIRGO". In: *Journal of Atmospheric and Solar-Terrestrial Physics* vol. 132, pp. 64–73 (cit. on p. 60). doi:10/6q4.

Lefèvre, M., Oumbe, A., Blanc, P., Espinar, B., Gschwind, B., Qu, Z., Wald, L., Schroedter-Homscheidt, M., Hoyer-Klick, C., Arola, A., Benedetti, A., Kaiser, J. W., and Morcrette, J.-J. (2013). "McClear: a new model estimating downwelling solar radiation at ground level in clear-sky conditions". In: *Atmospheric Measurement Techniques* vol. 6, no. 9, pp. 2403–2418 (cit. on p. 18). doi:10/f259gz.

Lefèvre, M., Blanc, P., Espinar, B., Gschwind, B., Menard, L., Ranchin, T., Wald, L., Saboret, L., Thomas, C., and Wey, E. (2014). "The HelioClim-1 database of daily solar radiation at Earth surface: an example of the benefits of GEOSS Data-CORE". In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* vol. 7, no. 5, pp. 1745–1753 (cit. on pp. 12, 70). doi:10/f259gx.

Lewis, N. S. and Crabtree, G. (2005). *Basic Research Needs for Solar Energy Utilization: Report on the Basic Energy Sciences Workshop on Solar Energy Utilization.* Tech. rep. U.S. Department of Energy (cit. on p. 7). URL: https://science.energy.gov/~/media/bes/pdf/reports/files/Basic_Research_Needs_for_Solar_Energy_Utilization_rpt.pdf (accessed on February 22nd 2017).

LIU, Y., SAN LIANG, X., and WEISBERG, R. H. (2007). "Rectification of the bias in the wavelet power spectrum". In: *Journal of Atmospheric and Oceanic Technology* vol. 24, no. 12, pp. 2093–2102 (cit. on pp. 27, 48, 67). doi:10/dv9zq4.

LOCKWOOD, M. and FRÖHLICH, C. (2007). "Recent oppositely directed trends in solar climate forcings and the global mean surface air temperature". In: *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences.* Vol. 463. 2086. The Royal Society, pp. 2447–2460 (cit. on p. 60). doi:10/bkg3th.

LOHMANN, S., SCHILLINGS, C., MAYER, B., and MEYER, R. (2006). "Long-term variability of solar direct and global radiation derived from ISCCP data and comparison with reanalysis data". In: *Solar Energy* vol. 80, no. 11, pp. 1390–1401 (cit. on pp. 12, 70). doi:10/fjn6mc.

LOONEY, D., HEMAKOM, A., and MANDIC, D. P. (2015). "Intrinsic multi-scale analysis: a multi-variate empirical mode decomposition framework". In: *Proc. R. Soc. A.* Vol. 471. 2173. The Royal Society, p. 20140709 (cit. on p. 85). doi:10/bvjn.

MADRONICH, S. and FLOCKE, S. (1999). "The Role of Solar Radiation in Atmospheric Chemistry". In: *Environmental Photochemistry.* Ed. by P. BOULE. Vol. 2 / 2L. The Handbook of Environmental Chemistry. Springer Berlin Heidelberg, pp. 1–26 (cit. on pp. 7, 8). doi:10/f259g2.

MARQUEZ, R. and COIMBRA, C. F. (2013). "Proposed metric for evaluation of solar forecasting models". In: *Journal of Solar Energy Engineering* vol. 135, no. 1, p. 011016 (cit. on p. 84). doi:10/brhb.

MEDVIGY, D. and BEAULIEU, C. (2012). "Trends in daily solar radiation and precipitation coefficients of variation since 1984". In: *Journal of Climate* vol. 25, no. 4, pp. 1330–1339 (cit. on p. 10). doi:10/dqkcvg.

MEFTAH, M., DEWITTE, S., IRBAH, A., CHEVALIER, A., CONSCIENCE, C., CROMME-LYNCK, D., JANSSEN, E., and MEKAOUI, S. (2014). "SOVAP/Picard, a spaceborne radiometer to measure the total solar irradiance". In: *Solar Physics* vol. 289, no. 5, pp. 1885–1899 (cit. on p. 18). doi:10/bvqs.

MOGHTADERI, A., FLANDRIN, P., and BORGNAT, P. (2013). "Trend filtering via empirical mode decompositions". In: *Computational Statistics & Data Analysis* vol. 58, pp. 114–126 (cit. on p. 38). doi:10/c4nqzz.

MOREL, J.-P., OLIVIÉR, J., and GUEYMARD, C. A. (2010). "Global Dimming and Brightening at Carpentras and Trappes (France)". In: *Eleventh Baseline Surface Radiation Network (BSRN) Scientific Review and Workshop.* Queenstown, New Zealand (cit. on pp. 44, 45). URL: http://www.wcrp-climate.org/documents/bsrn11rpt.pdf (accessed on February 23rd 2017).

NAGOVITSYN, Y. A. (1997). "A nonlinear mathematical model for the solar cyclicity and prospects for reconstructing the solar activity in the past". In: *Astronomy Letters*

vol. 23, pp. 742–748 (cit. on p. 11). URL: http://articles.adsabs.harvard.edu/cgi-bin/nph-iarticle_query?1997AstL...23..742N&data_type=PDF_HIGH&whole_paper=YES&type=PRINTER&filetype=.pdf (accessed on February 23rd 2017).

Noia, M., Ratto, C., and Festa, R. (1993). "Solar irradiance estimation from geostationary satellite data: I. Statistical models". In: *Solar Energy* vol. 51, no. 6, pp. 449–456 (cit. on p. 8). doi:10/b7nd2s.

Ohmura, A., Gilgen, H., Hegner, H., Müller, G., Wild, M., Dutton, E. G., Forgan, B., Fröhlich, C., Philipona, R., Heimo, A., König-Langlo, G., McArthur, B., Pinker, R., Whitlock, C. H., and Dehne, K. (1998). "Baseline Surface Radiation Network (BSRN/WCRP): New Precision Radiometry for Climate Research". In: *Bulletin of the American Meteorological Society* vol. 79, no. 10, pp. 2115–2136 (cit. on pp. 12, 15). doi:10/bvpw97.

Otani, K., Minowa, J., and Kurokawa, K. (1997). "Study on areal solar irradiance for analyzing areally-totalized PV systems". In: *Solar Energy Materials and Solar Cells* vol. 47, no. 1, pp. 281–288 (cit. on p. 9). doi:10/frvdrj.

Oxford English Dictionary (2015). Oxford University Press (cit. on p. 10). URL: http://www.oed.com (accessed on September 24th 2015).

Pachauri, R. K., Allen, M. R., Barros, V. R., Broome, J., Cramer, W., Christ, R., Church, J. A., Clarke, L., Dahe, Q., Dasgupta, P., Dubash, N. K., Edenhofer, O., Elgizouli, I., Field, C. B., Forster, P., Friedlingstein, P., Fuglestvedt, J., Gomez-Echeverri, L., Hallegatte, S., Hegerl, G., Howden, M., Jiang, K., Jimenez Cisneroz, B., Kattsov, V., Lee, H., Mach, K. J., Marotzke, J., Mastrandrea, M. D., Meyer, L., Minx, J., Mulugetta, Y., O'Brien, K., Oppenheimer, M., Pereira, J. J., Pichs-Madruga, R., Plattner, G.-K., Pörtner, H.-O., Power, S. B., Preston, B., Ravindranath, N. H., Reisinger, A., Riahi, K., Rusticucci, M., Scholes, R., Seyboth, K., Sokona, Y., Stavins, R., Stocker, T. F., Tschakert, P., van Vuuren, D., and van Ypserle, J.-P. (2014). *Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Ed. by R. Pachauri and L. Meyer. IPCC, Geneva, Switzerland (cit. on p. 10). URL: http://epic.awi.de/37530/1/IPCC_AR5_SYR_Final.pdf (accessed on February 23rd 2017).

Paluš, M. (2014). "Multiscale atmospheric dynamics: cross-frequency phase-amplitude coupling in the air temperature". In: *Physical Review Letters* vol. 112, no. 7, p. 078702 (cit. on p. 67). doi:10/bhpj.

Pedron, I. T. (2012). "Correlation and Persistence in Global Solar Radiation". In: *Solar Radiation.* Ed. by E. B. Babatunde. InTech (cit. on p. 10). doi:10/b6qk.

PEREZ, R., SEALS, R., and ZELENKA, A. (1997). "Comparing satellite remote sensing and ground network measurements for the production of site/time specific irradiance data". In: *Solar Energy* vol. 60, no. 2, pp. 89–96 (cit. on p. 70). doi:10/fn3t4w.

PEREZ, R., DAVID, M., HOFF, T. E., JAMALY, M., KIVALOV, S., KLEISSL, J., LAURET, P., and PEREZ, M. (2016). "Spatial and temporal variability of solar energy". In: *Foundations and Trends® in Renewable Energy* vol. 1, no. 1, pp. 1–44 (cit. on p. 9).

POLO, J., WILBERT, S., RUIZ-ARIAS, J. A., MEYER, R., GUEYMARD, C., SÚRI, M., MARTÍN, L., MIESLINGER, T., BLANC, P., GRANT, I., et al. (2016). "Preliminary survey on site-adaptation techniques for satellite-derived and reanalysis solar radiation datasets". In: *Solar Energy* vol. 132, pp. 25–37 (cit. on pp. 12, 71). doi:10/bs5k.

PORTNOFF, M. R. (1980). "Time-frequency representation of digital signals and systems based on short-time Fourier analysis". In: *Acoustics, Speech and Signal Processing, IEEE Transactions on* vol. 28, no. 1, pp. 55–69 (cit. on pp. 21, 24). doi:10/bmhmkr.

PRŠA, A., HARMANEC, P., TORRES, G., MAMAJEK, E., ASPLUND, M., CAPITAINE, N., CHRISTENSEN-DALSGAARD, J., DEPAGNE, É., HABERREITER, M., HEKKER, S., HILTON, J., KOPP, G., KOSTOV, V., KURTZ, D. W., LASKAR, J., MASON, B. D., MILONE, E. F., MONTGOMERY, M., RICHARDS, M., SCHMUTZ, W., SCHOU, J., and STEWART, S. G. (2016). "Nominal Values for Selected Solar and Planetary Quantities: IAU 2015 Resolution B3". In: *The Astronomical Journal* vol. 152, no. 2, p. 41 (cit. on p. 18). doi:10/bs5d.

QU, Z., GSCHWIND, B., LEFÈVRE, M., and WALD, L. (2014). "Improving HelioClim-3 estimates of surface solar irradiance using the McClear clear-sky model and recent advances in atmosphere composition". In: *Atmospheric Measurement Techniques* vol. 7, no. 11, pp. 3927–3933 (cit. on p. 19). doi:10/2gq.

UR REHMAN, N., PARK, C., HUANG, N. E., and MANDIC, D. P. (2013). "EMD via MEMD: multivariate noise-aided computation of standard EMD". In: *Advances in Adaptive Data Analysis* vol. 5, no. 02 (cit. on p. 85). doi:10/3t6.

RICHARDSON, D. B. and ANDREWS, R. W. (2014). "Validation of the MERRA dataset for solar PV applications". In: *2014 IEEE 40th Photovoltaic Specialist Conference (PVSC)*. IEEE, pp. 0809–0814 (cit. on pp. 12, 70). doi:10/bs5s.

RIGOLLIER, C., LEFÈVRE, M., and WALD, L. (2004). "The method Heliosat-2 for deriving shortwave solar radiation from satellite images". In: *Solar Energy* vol. 77, no. 2, pp. 159–169 (cit. on p. 19). doi:10/bfrn55.

RILLING, G., FLANDRIN, P., and GONÇALVES, P. (2003). "On empirical mode decomposition and its algorithms". In: *Proceedings of IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing NSIP-03*. Vol. 3. Grado (Italy), pp. 8–11 (cit. on pp. 36, 38, 39). URL: http://perso.ens-lyon.fr/patrick.flandrin/NSIP03.pdf (accessed on February 23rd 2017).

Rilling, G., Flandrin, P., and Gonçalves, P. (2005). "Empirical mode decomposition, fractional Gaussian noise and Hurst exponent estimation". In: *Proceedings (ICASSP '05) IEEE International Conference on Acoustics, Speech, and Signal Processing.* Vol. 4. Philadelphia (USA), pp. iv/489–iv/492 (cit. on pp. 61, 63). doi:10/fmnjq6.

Rios, R. A. and de Mello, R. F. (2013). "Improving time series modeling by decomposing and analyzing stochastic and deterministic influences". In: *Signal Processing* vol. 93, no. 11, pp. 3001–3013 (cit. on p. 68). doi:10/bdvd.

Rios, R. A., Parrott, L., Lange, H., and de Mello, R. F. (2015). "Estimating determinism rates to detect patterns in geospatial datasets". In: *Remote Sensing of Environment* vol. 156, pp. 11–20 (cit. on p. 11). doi:10/bdvf.

Rioul, O. and Vetterli, M. (1991). "Wavelets and signal processing". In: *IEEE signal processing magazine* vol. 8, no. 4, pp. 14–38 (cit. on pp. 24, 26, 28). doi:10/br22xh.

Roesch, A., Wild, M., Ohmura, A., Dutton, E. G., Long, C. N., and Zhang, T. (2011). "Assessment of BSRN radiation records for the computation of monthly means". In: *Atmospheric Measurement Techniques* vol. 4, no. 2, pp. 339–354 (cit. on pp. 12, 15). doi:10/dpqsc8.

Schlotthauer, G., Torres, M. E., Rufiner, H. L., and Flandrin, P. (2009). "EMD of Gaussian white noise: effects of signal length and sifting number on the statistical properties of intrinsic mode functions". In: *Advances in Adaptive Data Analysis* vol. 1, no. 04, pp. 517–527 (cit. on p. 63). doi:10/b9b7st.

Schroedter-Homscheidt, M., Delamare, C., Heilscher, G., Heinemann, D., Hoyer, C., Meyer, R., Toggweiler, P., Wald, L., and Zelenka, A. (2006). "The ESA-ENVISOLAR project: experience on the commercial use of Earth observation based solar surface irradiance measurements for energy business purposes". In: *Solar Energy Resources Management for Electricity Generation*, pp. 111–124 (cit. on p. 8). URL: http://www.envisolar.net/factsheets/ispra_paper_050204.pdf (accessed on February 23rd 2017).

Sengupta, M., Habte, A., Kurtz, S., Dobos, A., Wilbert, S., Lorenz, E., Stoffel, T., Renné, D., Gueymard, C. A., Myers, D., Wilcox, S., Blanc, P., and Perez, R. (2015). *Best Practices Handbook for the Collection and Use of Solar Resource Data for Solar Energy Applications.* Technical Report NREL/TP-5D00-63112. National Renewable Energy Laboratory (cit. on pp. 12, 84). doi:10/b2jp.

Skartveit, A. and Olseth, J. (1992). "The probability density and autocorrelation of short-term global and beam irradiance". In: *Solar Energy* vol. 49, no. 6, pp. 477–487 (cit. on p. 9). doi:10/dm6wdw.

Spertino, F., Di Leo, P., and Cocina, V. (2013). "Accurate measurements of solar irradiance for evaluation of photovoltaic power profiles". In: *PowerTech Grenoble 2013.* IEEE, pp. 1–5 (cit. on p. 7). doi:10/bz8f.

STOTT, P. A., JONES, G. S., and MITCHELL, J. F. (2003). "Do models underestimate the solar contribution to recent climate change?" In: *Journal of Climate* vol. 16, no. 24, pp. 4079–4093 (cit. on p. 60). doi:10/bzdgx4.

TARY, J. B., HERRERA, R. H., HAN, J., and BAAN, M. (2014). "Spectral estimation – What is new? What is next?" In: *Reviews of Geophysics* vol. 52, no. 4, pp. 723–749 (cit. on pp. 11, 20, 21, 24). doi:10/f259g3.

THUILLER, G. (2015). Personal communication (cit. on p. 60).

TORRENCE, C. and COMPO, G. P. (1998). "A practical guide to wavelet analysis". In: *Bulletin of the American Meteorological Society* vol. 79, no. 1, pp. 61–78 (cit. on pp. 27, 29, 30, 46). doi:10/bhbwhf.

TORRES, M. E., COLOMINAS, M. A., SCHLOTTHAUER, G., and FLANDRIN, P. (2011). "A complete ensemble empirical mode decomposition with adaptive noise". In: *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on.* IEEE, pp. 4144–4147 (cit. on p. 40). doi:10/b7sf6d.

TRENBERTH, K. E., FASULLO, J. T., and KIEHL, J. (2009). "Earth's global energy budget". In: *Bulletin of the American Meteorological Society* vol. 90, no. 3, pp. 311–323 (cit. on pp. 7, 10). doi:10/d3xccg.

TSVETKOV, A., WILCOX, S., RENNE, D., and PULSCAK, M. (1995). "International solar resource data at the World Radiation Data Center". In: *Proceedings of Solar '95 : The 1995 American Solar Energy Society Annual Conference, Minneapolis, Minnesota, July 15-20, 1995.* Ed. by R. CAMPBELL-HOWE and B. WILKINS-CROWDER. American Solar Energy Society, Boulder, Colorado, United States of America (cit. on p. 19). URL: https://www.osti.gov/scitech/biblio/109880 (accessed on February 23rd 2017).

VECCHIO, A., LAURENZA, M., MEDURI, D., CARBONE, V., and STORINI, M. (2012). "The dynamics of the solar magnetic field: polarity reversals, butterfly diagram, and quasi-biennial oscillations". In: *The Astrophysical Journal* vol. 749, no. 1, p. 27 (cit. on p. 61). doi:10/bcz5.

VUILLEUMIER, L., HAUSER, M., FÉLIX, C., VIGNOLA, F., BLANC, P., KAZANTZIDIS, A., and CALPINI, B. (2014). "Accuracy of ground surface broadband shortwave radiation monitoring". In: *Journal of Geophysical Research: Atmospheres* vol. 119, no. 24, pp. 13, 838–13, 860 (cit. on p. 15). doi:10/3nx.

WALD, L. (2007). "Solar Radiation Energy (Fundamentals and Theory)". In: *Solar Energy Conversion and Photoenergy Systems.* Ed. by J. BLANCO and S. MALATO. Encyclopedia of Life Support Systems (EOLSS). Developed under the Auspices of the UNESCO. Eolss Publishers, Paris, France (cit. on p. 7). URL: http://www.eolss.net (accessed on February 23rd 2017).

WANG, G., CHEN, X.-Y., QIAO, F.-L., WU, Z., and HUANG, N. E. (2010). "On intrinsic mode function". In: *Advances in Adaptive Data Analysis* vol. 2, no. 03, pp. 277–293 (cit. on pp. 39, 40). doi:10/b466fq.

WANG, Y.-H., YEH, C.-H., YOUNG, H.-W. V., HU, K., and LO, M.-T. (2014). "On the computational complexity of the empirical mode decomposition algorithm". In: *Physica A: Statistical Mechanics and its Applications* vol. 400, pp. 159–167 (cit. on p. 40). doi:10/3nz.

WASSERSTEIN, R. L. and LAZAR, N. A. (2016). "The ASA's Statement on p-Values: Context, Process, and Purpose". In: *The American Statistician* vol. 70, no. 2, pp. 129–133 (cit. on p. 67). doi:10/bc4d.

WELTER, G. S. and ESQUEF, P. A. A. (2013). "Multifractal analysis based on amplitude extrema of intrinsic mode functions". In: *Physical Review E* vol. 87, no. 3, p. 032916 (cit. on p. 61). doi:10/brmr.

WORLD METEOROLOGICAL ORGANIZATION (2014). *Guide to Meteorological Instruments and Methods of Observation (CIMO guide)*. Tech. rep. WMO-No. 8, Eighth Edition. Geneva, Switzerland: World Meteorological Organization (cit. on pp. 8, 19). URL: http://library.wmo.int/opac/doc_num.php?explnum_id=3121 (accessed on February 23rd 2017).

WOYTE, A., BELMANS, R., and NIJS, J. (2007). "Localized Spectral Analysis of Fluctuating Power Generation from Solar Energy Systems". In: *EURASIP Journal on Advances in Signal Processing* vol. 2007, no. 1, p. 080919 (cit. on p. 9). doi:10/bnvkvc.

WU, Z. and HUANG, N. E. (2004). "A study of the characteristics of white noise using the empirical mode decomposition method". In: *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* vol. 460, no. 2046, pp. 1597–1611 (cit. on pp. 40, 63). doi:10/d9p67j.

WU, Z. and HUANG, N. E. (2009). "Ensemble empirical mode decomposition: a noise-assisted data analysis method". In: *Advances in adaptive data analysis* vol. 1, no. 01, pp. 1–41 (cit. on pp. 40, 41). doi:10/fdnzfq.

WU, Z. and HUANG, N. E. (2010). "On the filtering properties of the empirical mode decomposition". In: *Advances in Adaptive Data Analysis* vol. 2, no. 04, pp. 397–414 (cit. on p. 40). doi:10/bscvxv.

WU, Z., HUANG, N. E., and CHEN, X. (2011). "Some considerations on physical analysis of data". In: *Advances in Adaptive Data Analysis* vol. 3, no. 01n02, pp. 95–113 (cit. on pp. 11, 31, 32, 46, 52). doi:10/cftghv.

XIAOLAN, W. and HUI, L. (2009). "One-Month Ahead Prediction of Wind Speed and Output Power Based on EMD and LSSVM". In: *2009 International Conference on Energy and Environment Technology*. Vol. 3, pp. 439–442 (cit. on p. 84). doi:10/fhf8pj.

Yordanov, G. H., Saetre, T. O., and Midtgard, O.-M. (2013). "100-millisecond resolution for accurate overirradiance measurements". In: *Photovoltaics, IEEE Journal of* vol. 3, no. 4, pp. 1354–1360 (cit. on pp. 9, 10). doi:10/7v8.

Zappalà, D. A., Barreiro, M., and Masoller, C. (2016). "Global Atmospheric Dynamics Investigated by Using Hilbert Frequency Analysis". In: *Entropy* vol. 18, no. 11, p. 408 (cit. on p. 36). doi:10/f9gtx5.

Zelenka, A., Czeplak, G., d'Agostino, V., Josefson, W., Maxwell, E., and Perez, R. (1992). "Techniques for supplementing solar radiation network data". In: *Report No. IEA-SHCP-9D-1*. Vol. 1. International Energy Agency (cit. on pp. 12, 70). URL: http://archive.iea-shc.org/publications/downloads/Task9_Vol1_Sept1992.pdf (accessed on February 23rd 2017).

Zelenka, A., Perez, R., Seals, R., and Renné, D. (1999). "Effective accuracy of satellite-derived hourly irradiances". In: *Theoretical and applied climatology* vol. 62, no. 3-4, pp. 199–207 (cit. on p. 70). doi:10/ct3pqh.

Zeng, Z., Yang, H., Zhao, R., and Meng, J. (2013). "Nonlinear characteristics of observed solar radiation data". In: *Solar Energy* vol. 87, pp. 204–218 (cit. on pp. 11, 32). doi:10/bcr8.

Zhu, Z., Sun, Y., and Li, H. (2007). "Hybrid of EMD and SVMs for Short-Term Load Forecasting". In: *2007 IEEE International Conference on Control and Automation*, pp. 1044–1047 (cit. on p. 84). doi:10/ck966h.

Appendices

APPENDIX A

Relevant scientific production

The scientific output of the doctoral programme, pertinent to the topic of this study, consists in four communications at international conferences and four original research articles published in open-access format. The following lists enumerate these scientific outcomes.

Research articles

- M. Bengulescu, P. Blanc, and L. Wald [2016a]. "On the temporal variability of the surface solar radiation by means of spectral representations". In: *Advances in Science and Research* vol. 13, pp. 121–127. doi:10/bp6g.
- M. Bengulescu, P. Blanc, and L. Wald [2016c]. "On the intrinsic time-scales of temporal variability in measurements of the surface solar radiation". In: *Nonlinear Processes in Geophysics Discussions* vol. 2016. in review, pp. 1–35. doi:10/brwj.
- M. Bengulescu, P. Blanc, and L. Wald [2016d]. "Characterizing temporal variability in measurements of surface solar radiation and its dependence on climate". In: *Energy Procedia* vol. 97. European Geosciences Union General Assembly 2016, EGU Division Energy, Resources & the Environment (ERE), pp. 164–171. doi:10/bt46.
- M. Bengulescu, P. Blanc, A. Boilley, and L. Wald [2017]. "Do modelled or satellite-based estimates of surface solar irradiance accurately describe its temporal variability?" In: *Advances in Science and Research* vol. 14, pp. 35–48. doi:10/b37n.

Communications

- M. Bengulescu, P. Blanc, and L. Wald [2015]. "Assessing the temporal variability of the surface solar radiation with time-frequency-energy representations". In: *15th EMS annual meeting.* Vol. 12. Sofia, Bulgaria, pp. 2015–230. URL: http://meetingorganizer.copernicus.org/EMS2015/EMS2015-230.pdf.

- M. Bengulescu, P. Blanc, and L. Wald [2016b]. "Adaptive data analysis for characterizing the temporal variability of the solar resource". In: *European Geosciences Union General Assembly 2016*. Vol. 18. Geophysical Research Abstracts. European Geosciences Union. Vienna, Austria, EGU2016–14847. URL: http://meetingorganizer.copernicus.org/EGU2016/EGU2016-14847.pdf.

- M. Bengulescu, P. Blanc, and L. Wald [2016e]. "Hilbert-Huang spectral analysis for characterizing the intrinsic time-scales of variability in decennial time-series of surface solar radiation". In: *European Geosciences Union General Assembly 2016*. Vol. 18. Geophysical Research Abstracts. European Geosciences Union. Vienna, Austria, EGU2016–14549. URL: http://meetingorganizer.copernicus.org/EGU2016/EGU2016-14549.pdf.

- M. Bengulescu, P. Blanc, and L. Wald [2016f]. "Hilbert-Huang spectral analysis for the characterization of variability in satellite-derived time series of surface solar irradiance". In: *16th EMS Annual Meeting*. Vol. 13. European Meteorological Society. Trieste, Italy, pp. 2016–372. URL: http://meetingorganizer.copernicus.org/EMS2016/EMS2016-372.pdf.

APPENDIX B

Supplement for chapter 5

List of figures in supplement

S1 2Dhist KIV WRDC-ERA IMF1

S2 2Dhist KIV WRDC-ERA IMF2
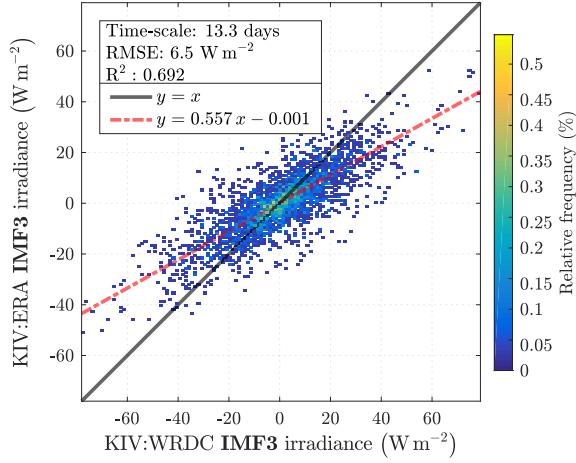
**S3** 2Dhist KIV WRDC-ERA IMF3
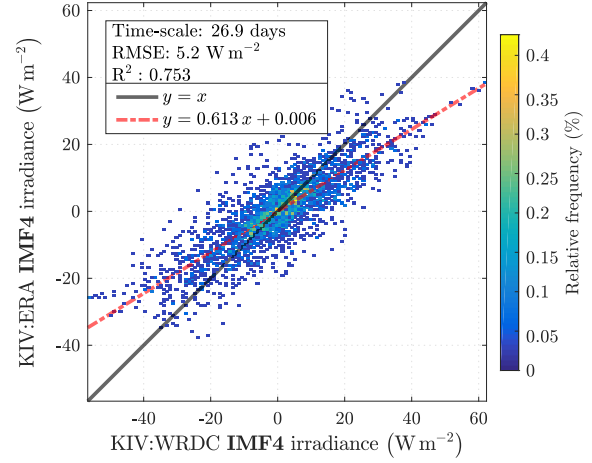

S4 2Dhist KIV WRDC-ERA IMF4
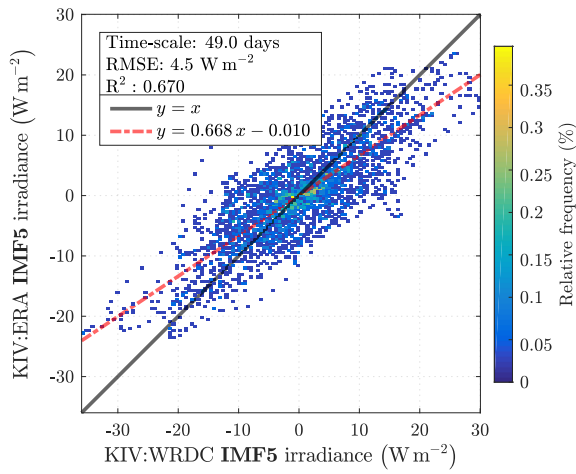
S5 2Dhist KIV WRDC-ERA IMF5
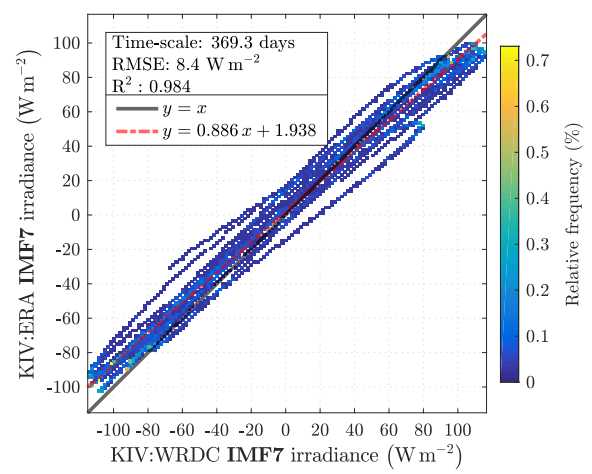
S6 2Dhist KIV WRDC-ERA IMF7

109
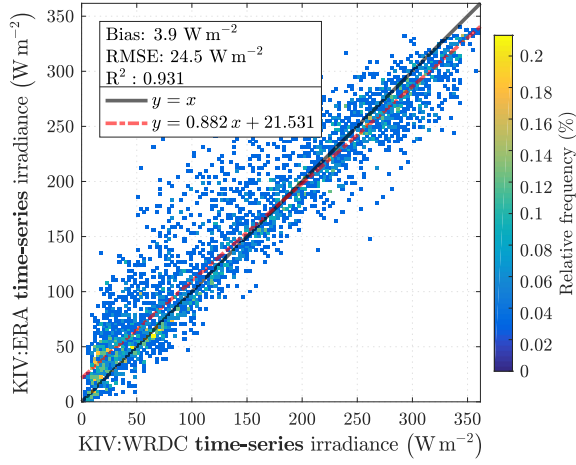
S7 2Dhist KIV WRDC-ERA time-series
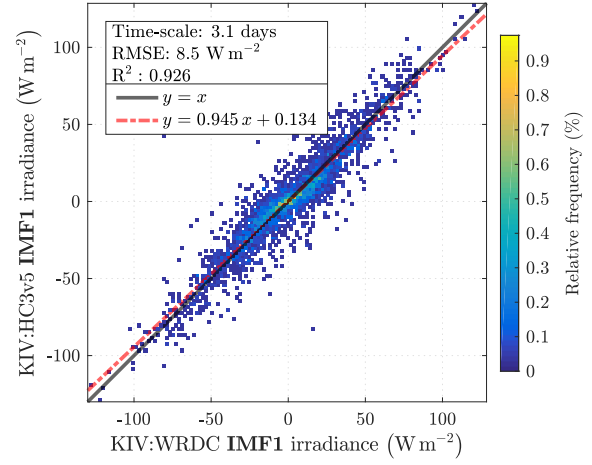
S8 2Dhist KIV WRDC-HC3v5 IMF1
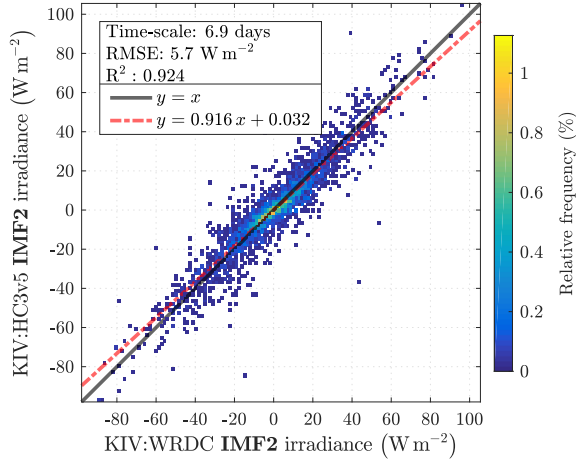
S9 2Dhist KIV WRDC-HC3v5 IMF2
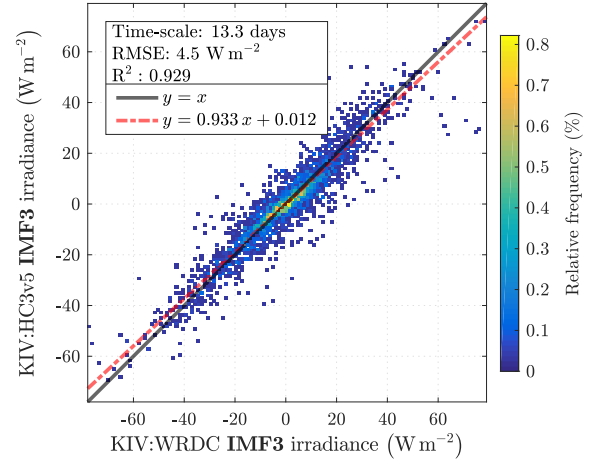
S10 2Dhist KIV WRDC-HC3v5 IMF3
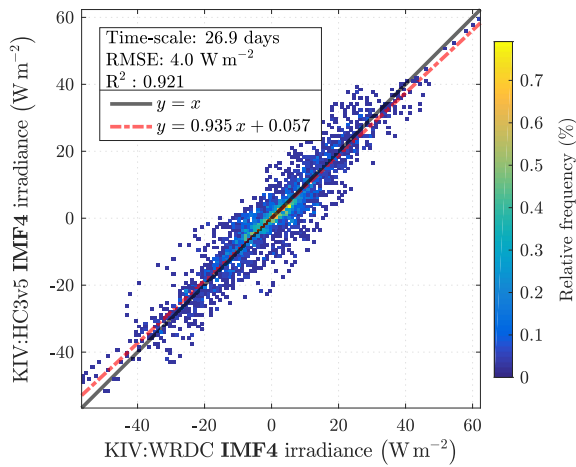
S11 2Dhist KIV WRDC-HC3v5 IMF4
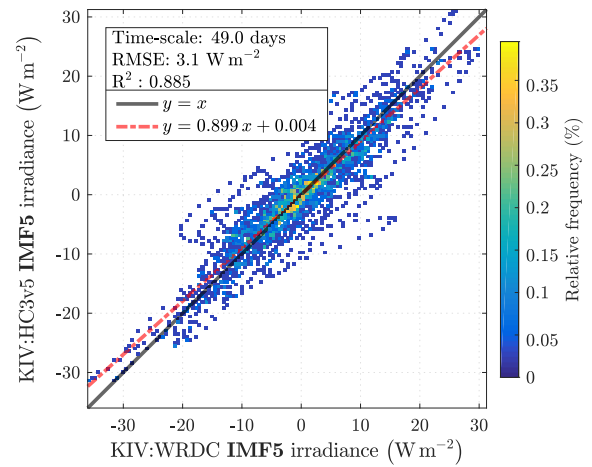
S12 2Dhist KIV WRDC-HC3v5 IMF5
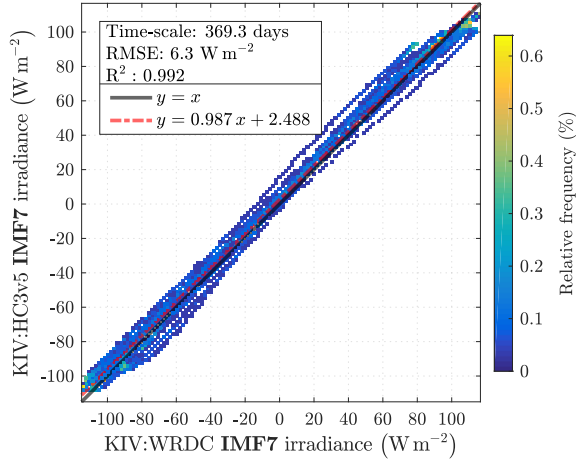
S13 2Dhist KIV WRDC-HC3v5 IMF7
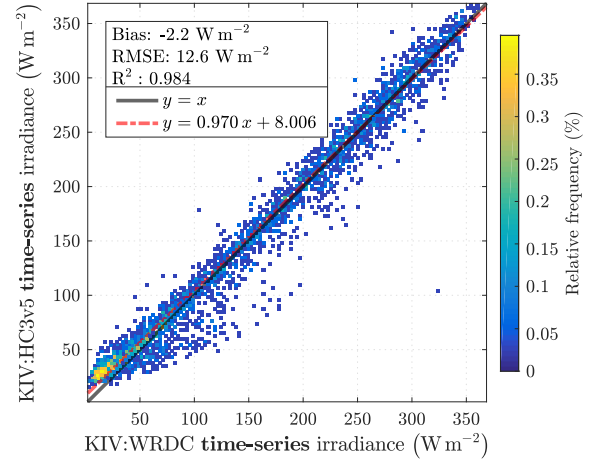
S14 2Dhist KIV WRDC-HC3v5 time-series
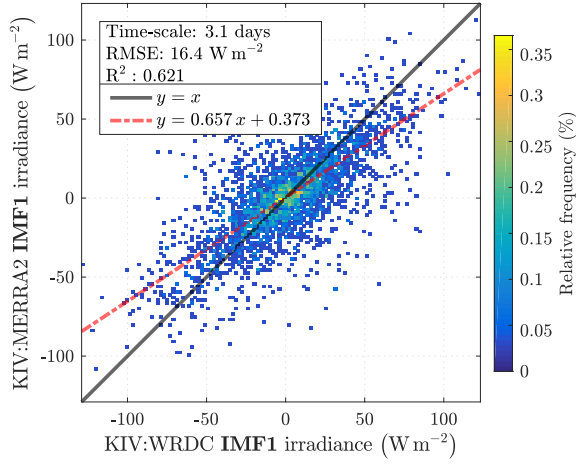
S15 2Dhist KIV WRDC-MERRA2 IMF1
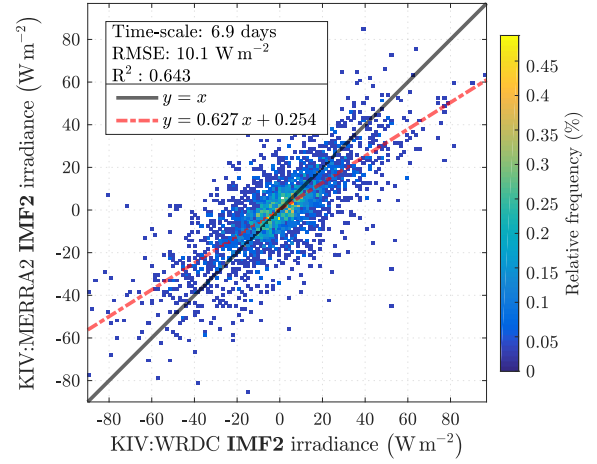
S16 2Dhist KIV WRDC-MERRA2 IMF2
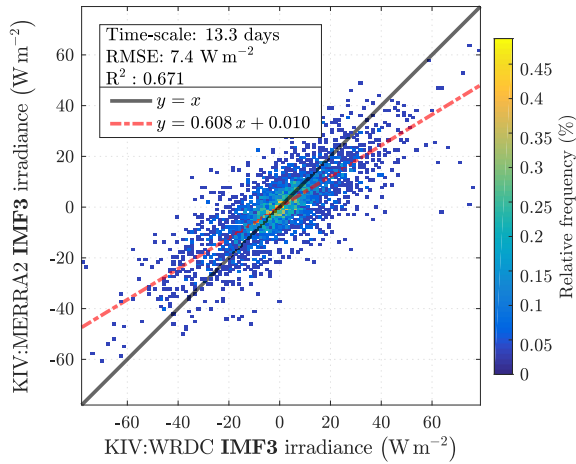

S17 2Dhist KIV WRDC-MERRA2 IMF3
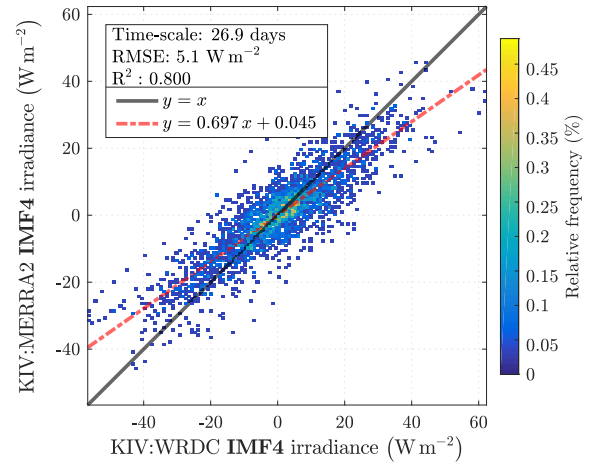
S18 2Dhist KIV WRDC-MERRA2 IMF4
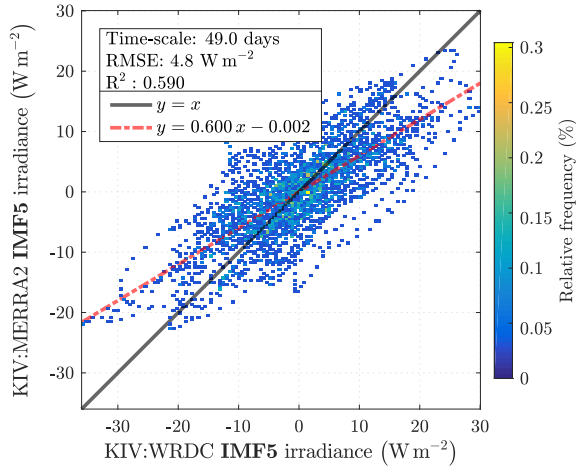
S19 2Dhist KIV WRDC-MERRA2 IMF5
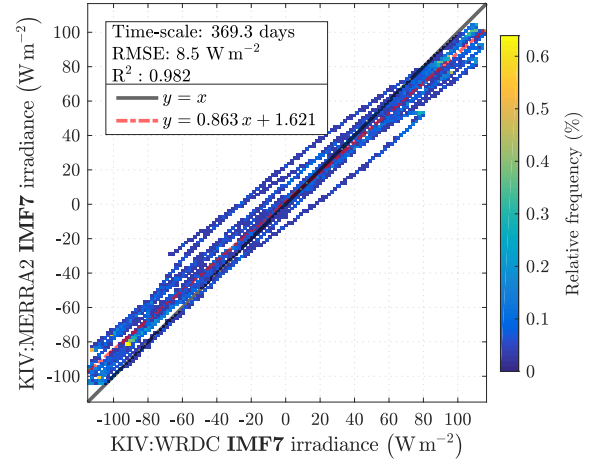
S20 2Dhist KIV WRDC-MERRA2 IMF7
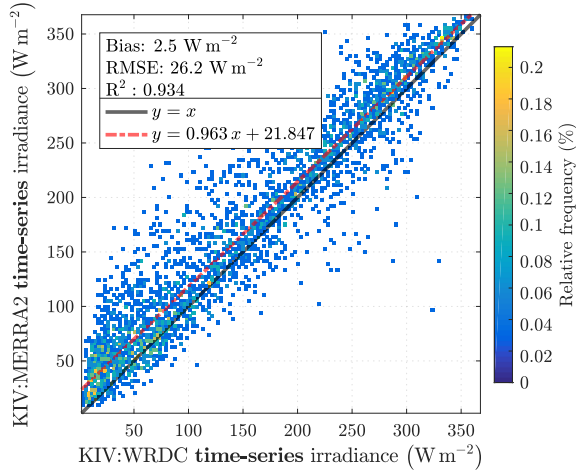
S21 2Dhist KIV WRDC-MERRA2 time-series
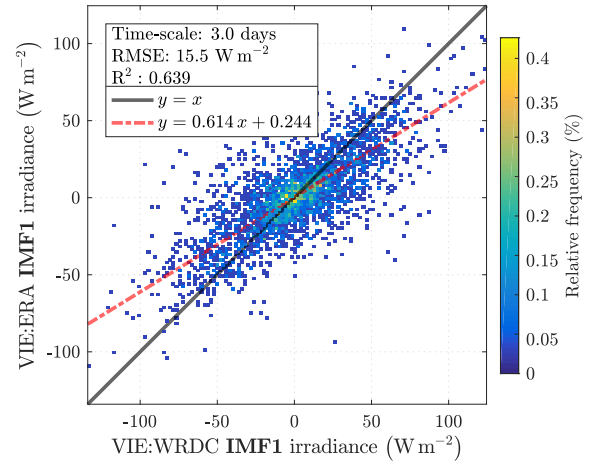
S22 2Dhist VIE WRDC-ERA IMF1
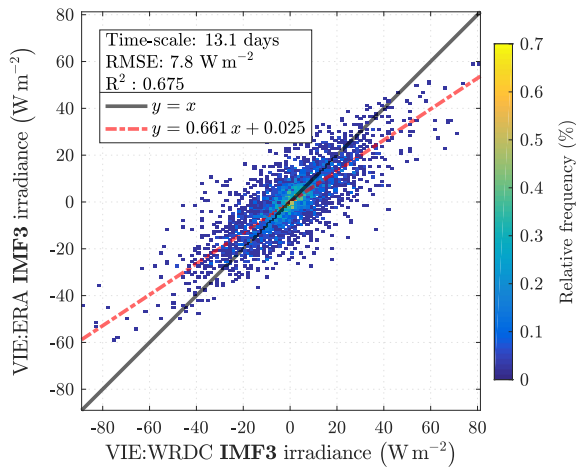
S23 2Dhist VIE WRDC-ERA IMF3
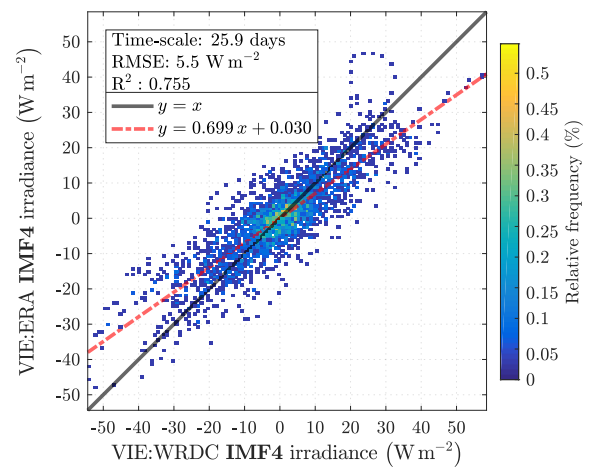
S24 2Dhist VIE WRDC-ERA IMF4
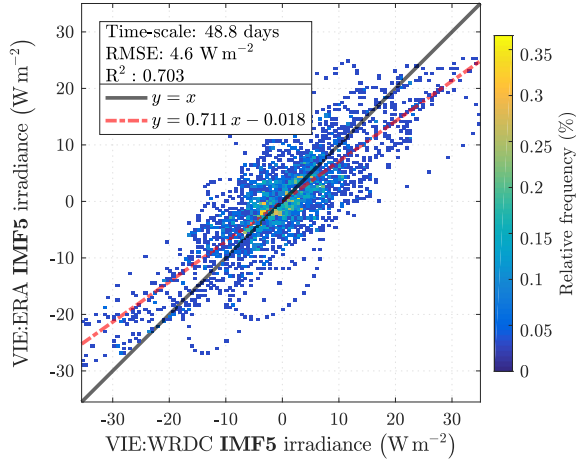
S25 2Dhist VIE WRDC-ERA IMF5
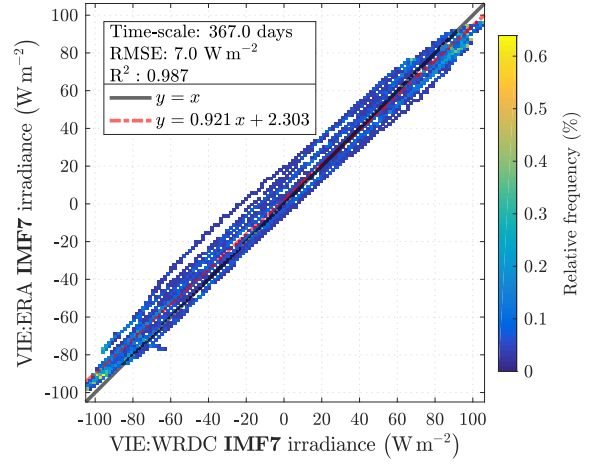
S26 2Dhist VIE WRDC-ERA IMF7
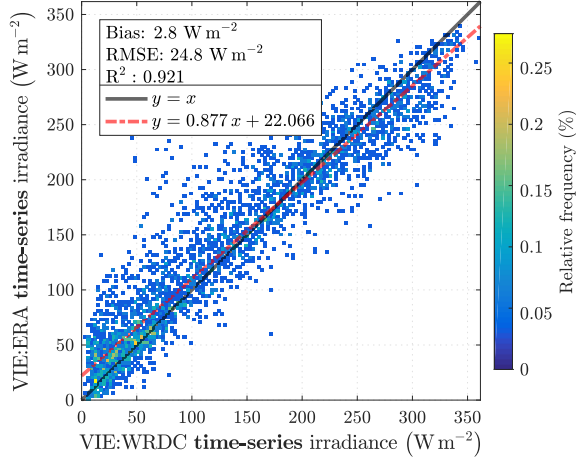
S27 2Dhist VIE WRDC-ERA time-series
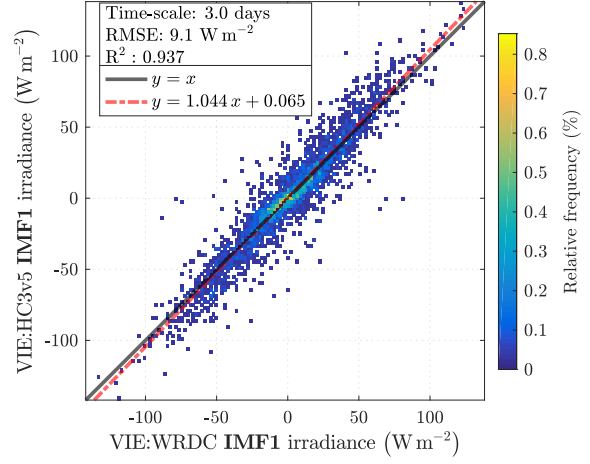
S28 2Dhist VIE WRDC-HC3v5 IMF1
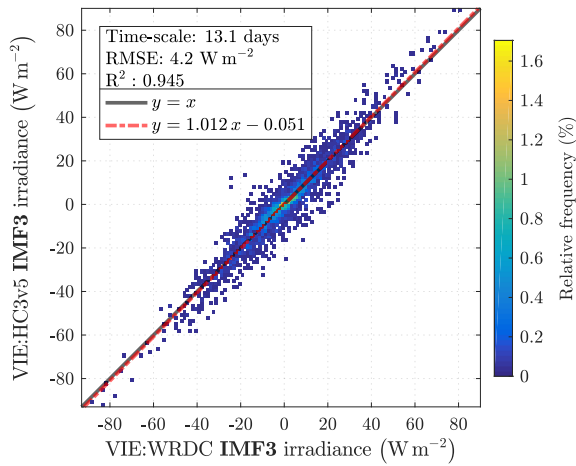
S29 2Dhist VIE WRDC-HC3v5 IMF3
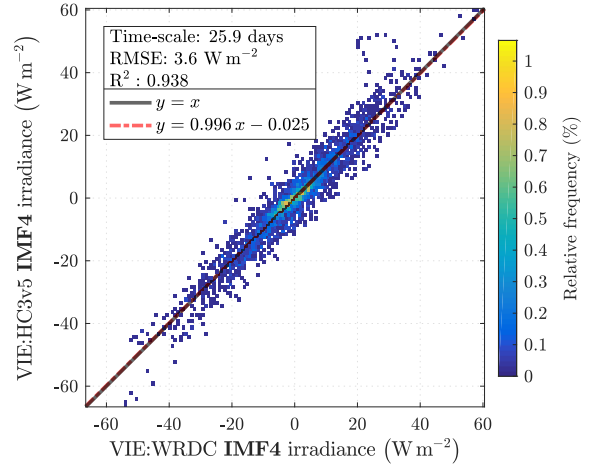
S30 2Dhist VIE WRDC-HC3v5 IMF4
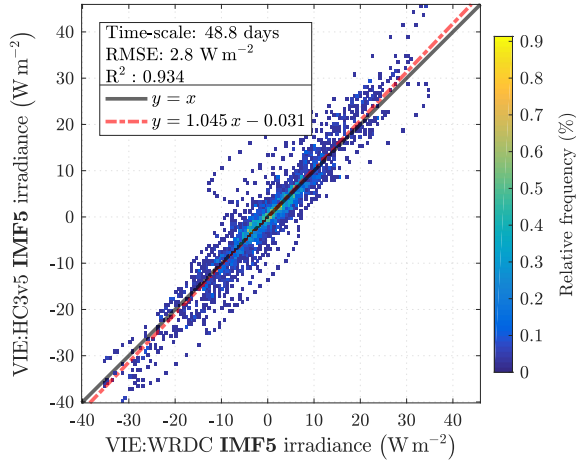
S31 2Dhist VIE WRDC-HC3v5 IMF5
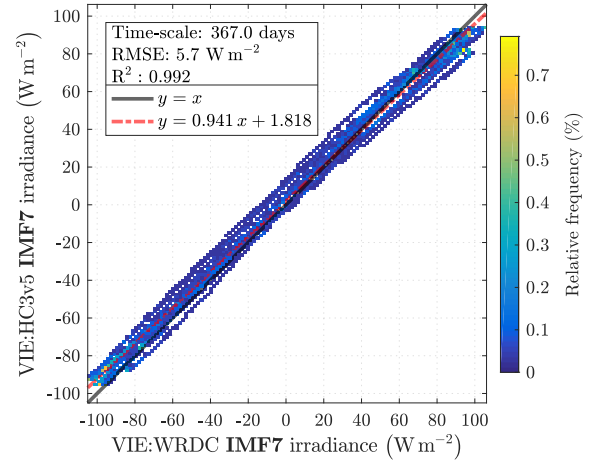
S32 2Dhist VIE WRDC-HC3v5 IMF7
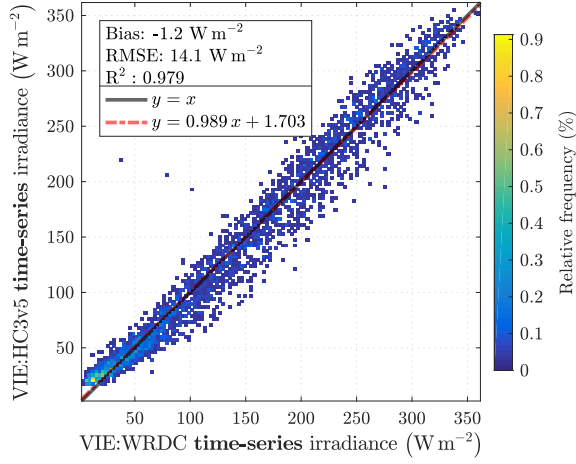
S33 2Dhist VIE WRDC-HC3v5 time-series
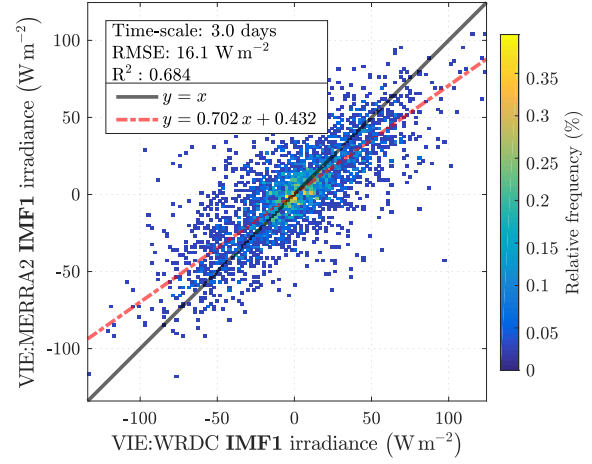
S34 2Dhist VIE WRDC-MERRA2 IMF1
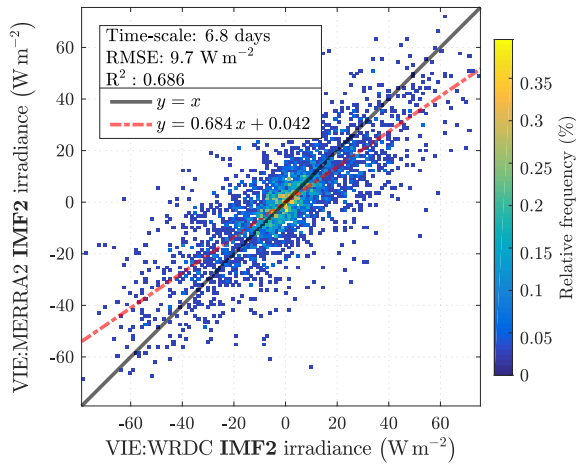
S35 2Dhist VIE WRDC-MERRA2 IMF2
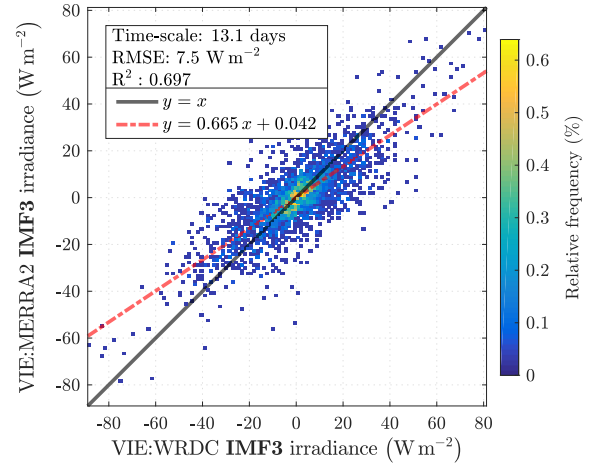
S36 2Dhist VIE WRDC-MERRA2 IMF3
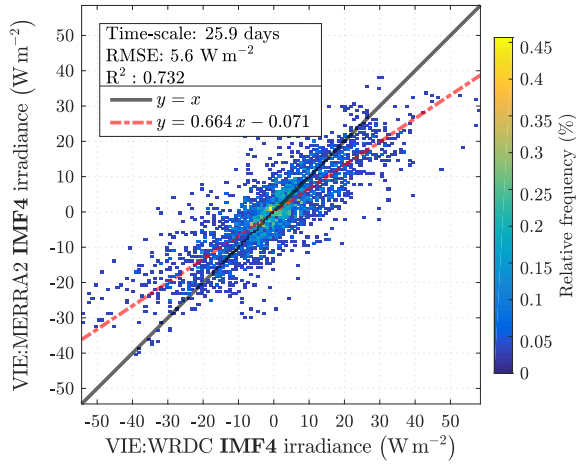
S37 2Dhist VIE WRDC-MERRA2 IMF4
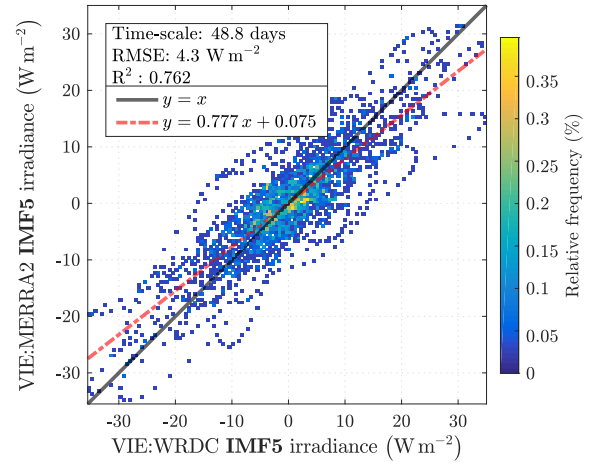
S38 2Dhist VIE WRDC-MERRA2 IMF5
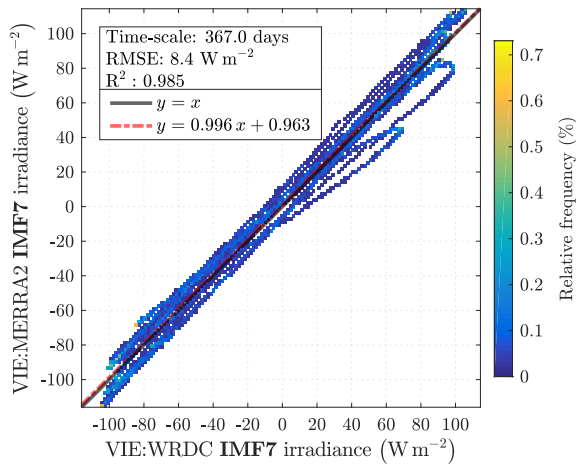
S39 2Dhist VIE WRDC-MERRA2 IMF7
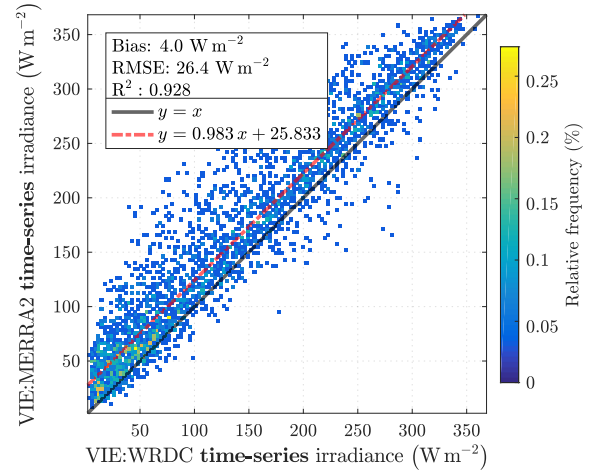
S40 2Dhist VIE WRDC-MERRA2 time-series
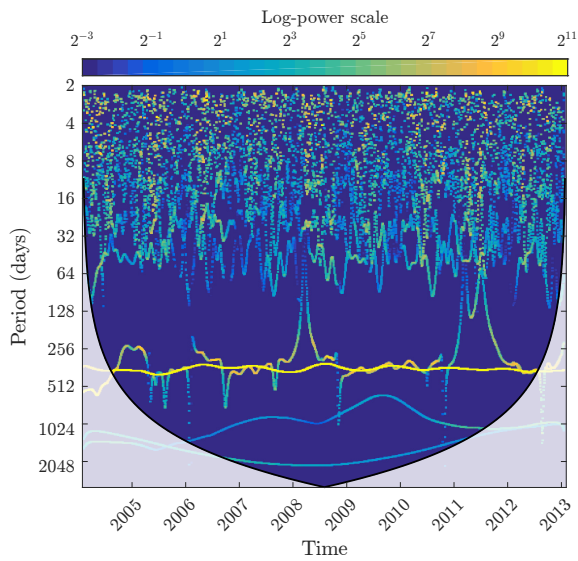
S41 HSA KIV ERA
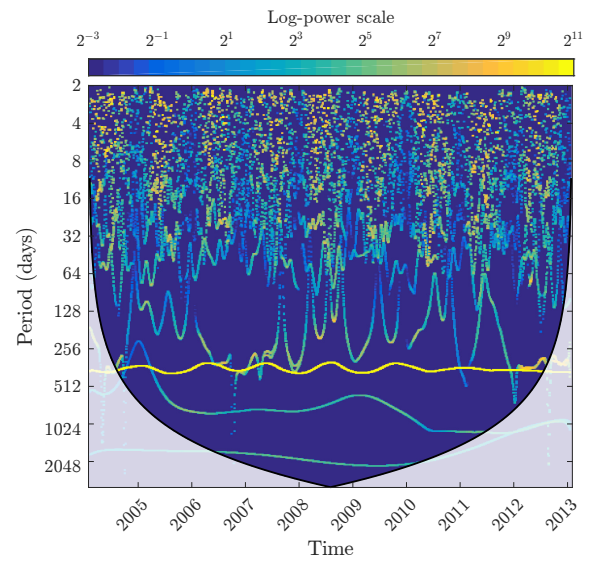
S42 HSA KIV HC3v5
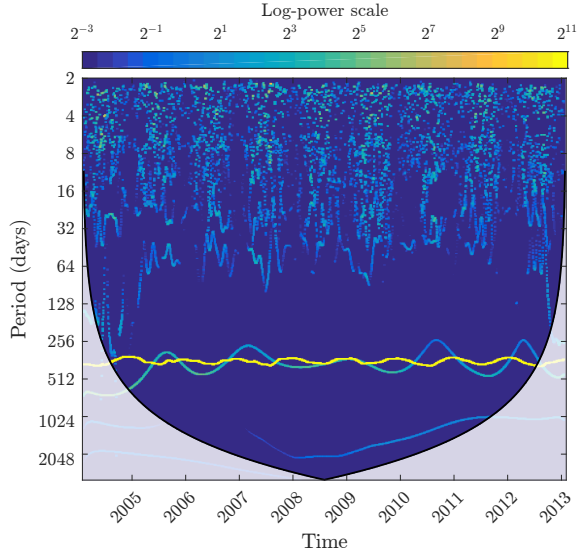
S43 HSA KIV McClear
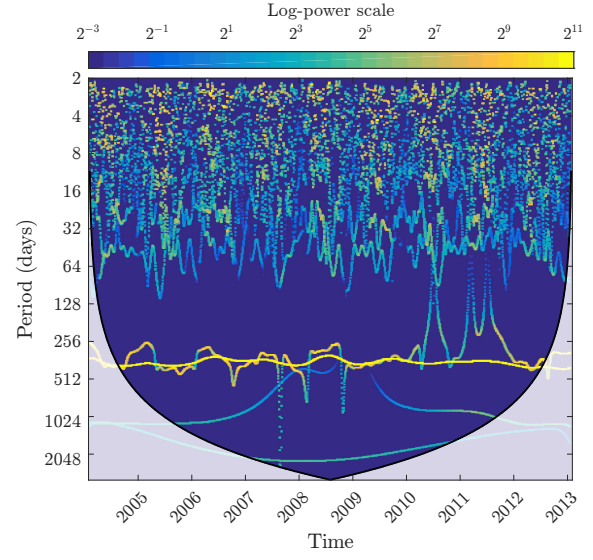
S44 HSA KIV MERRA2
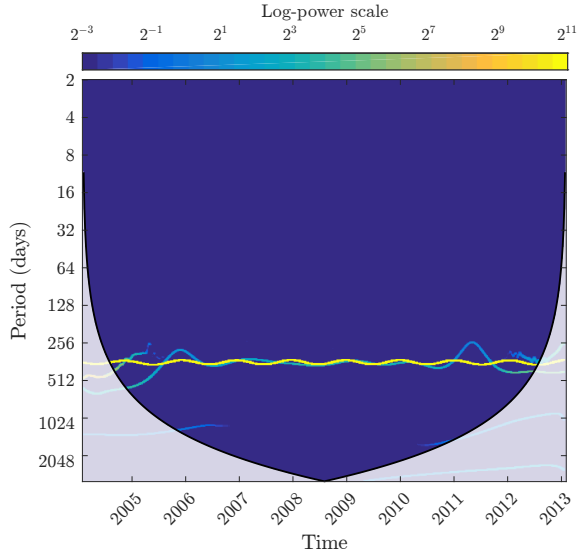
S45 HSA KIV TOA
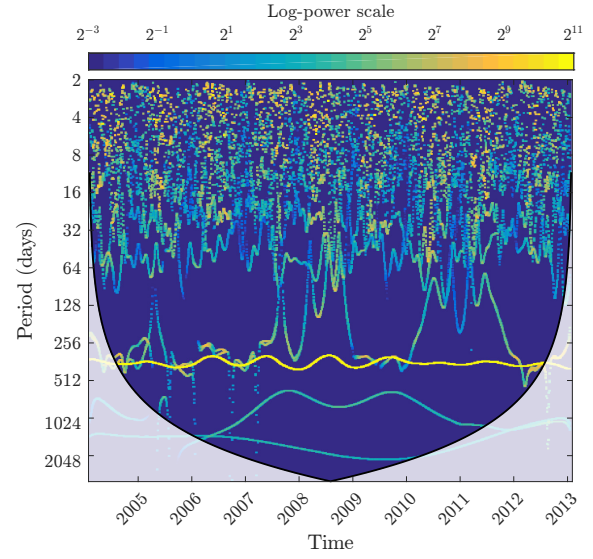
S46 HSA KIV WRDC
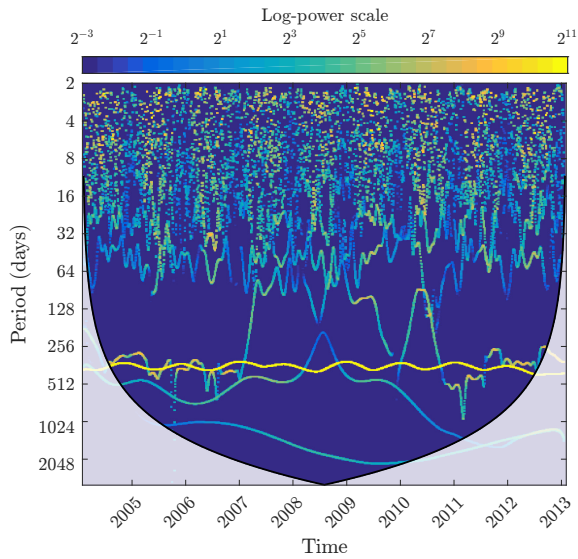
S47 HSA VIE ERA
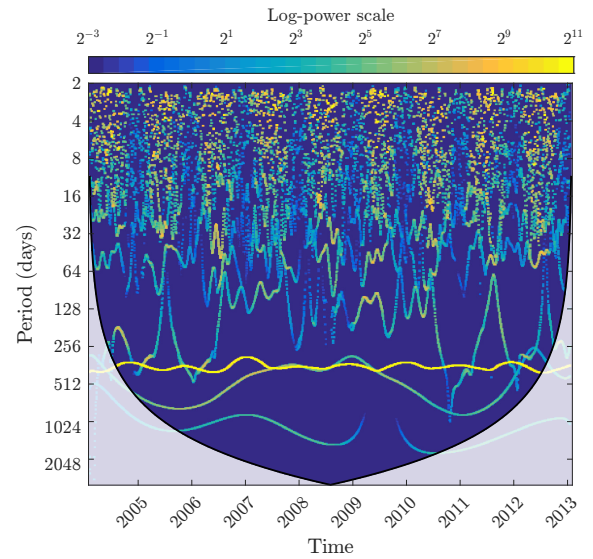
S48 HSA VIE HC3v5
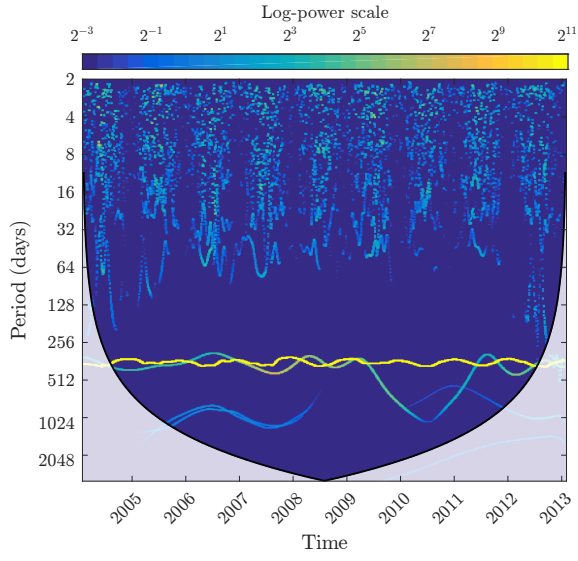
S49 HSA VIE McClear
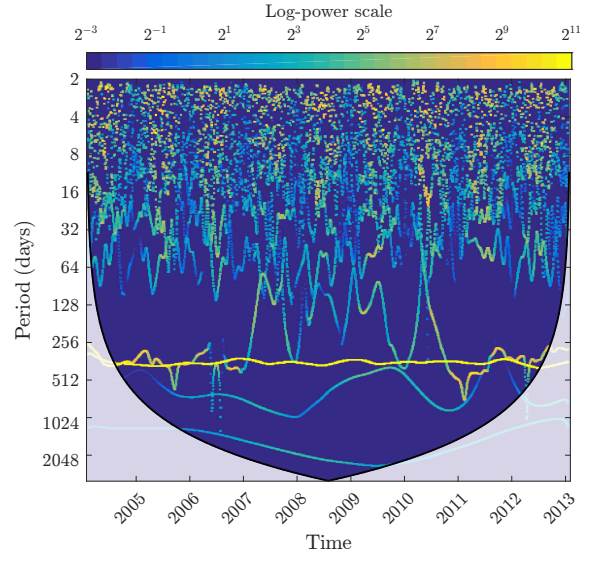
S50 HSA VIE MERRA2
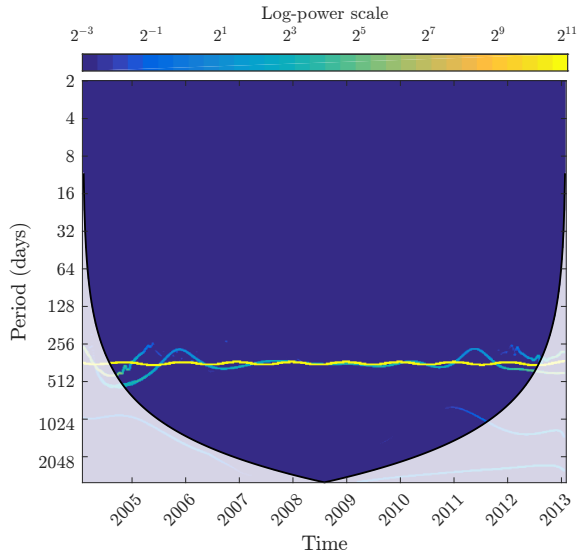
S51 HSA VIE TOA
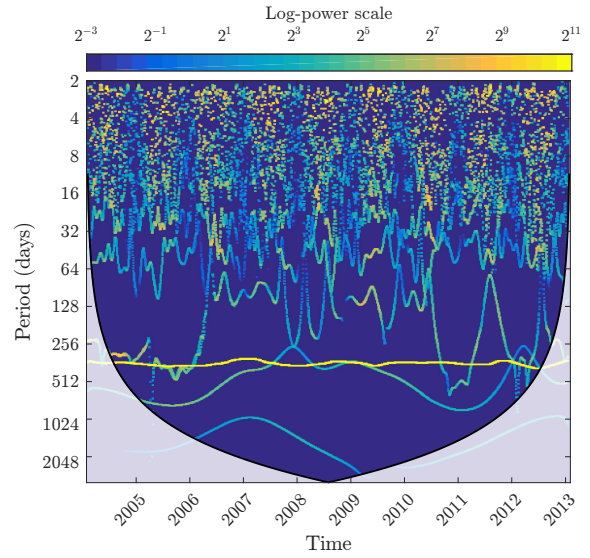
S52 HSA VIE WRDC
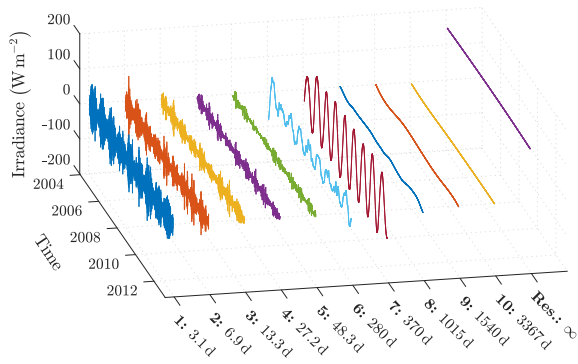
S53 IMFs KIV ERA
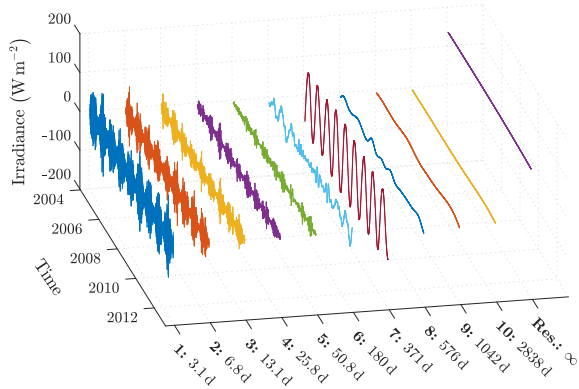

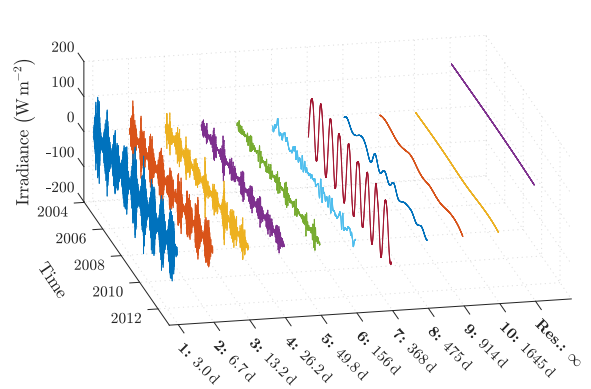**S54** IMFs KIV HC3v5

117
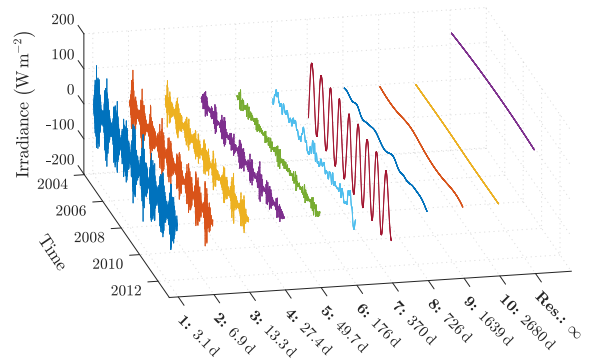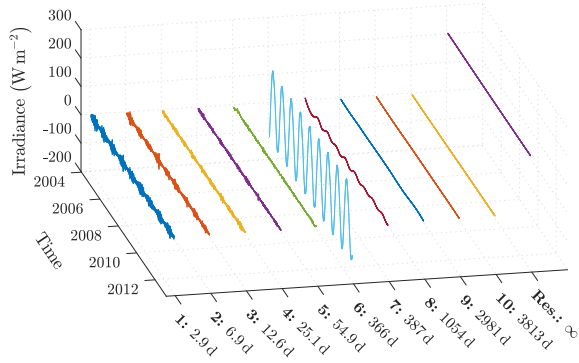
S55 IMFs KIV McClear
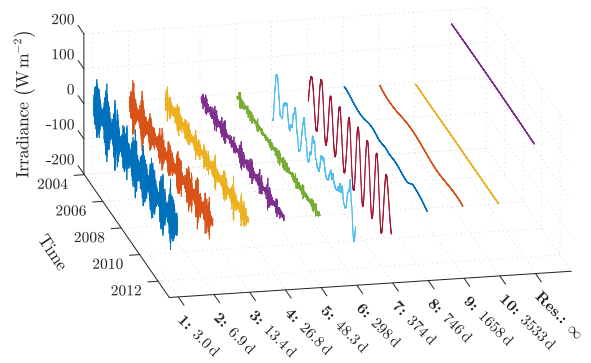
S56 IMFs KIV MERRA2
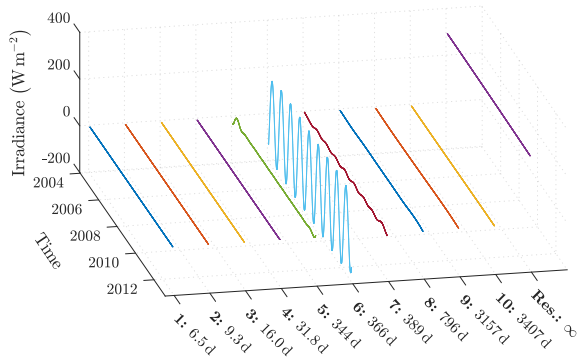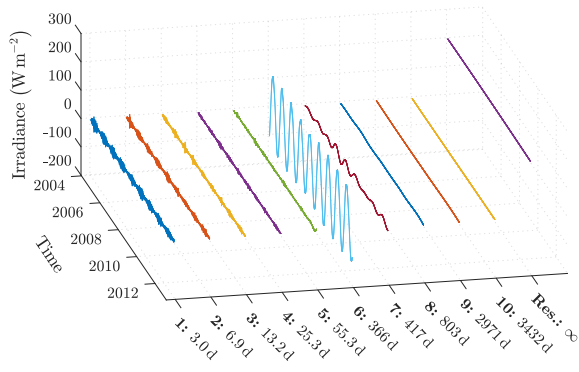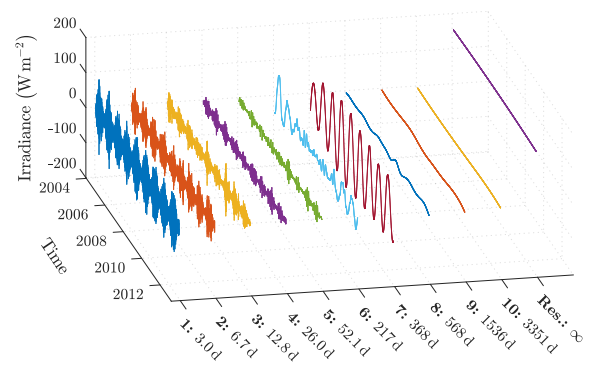
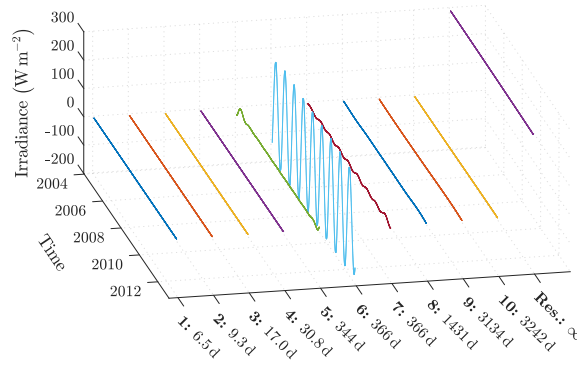

S57 IMFs KIV TOA
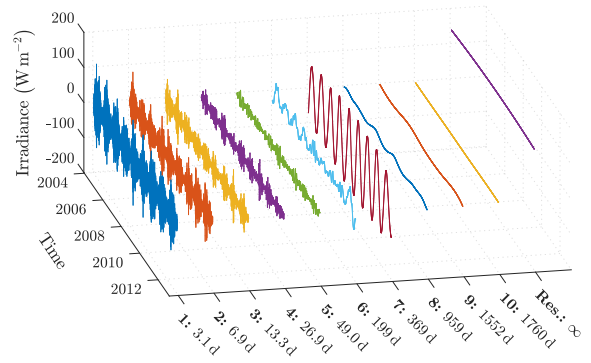
S58 IMFs KIV WRDC

S59 IMFs VIE ERA

S60 IMFs VIE HC3v5

S61 IMFs VIE McClear

S62 IMFs VIE MERRA2

S63 IMFs VIE TOA

Résumé

Cette thèse traite de la variabilité temporelle intrinsèque de l'éclairement solaire reçu au sol. Les échelles caractéristiques de variabilité sont mises en évidence par l'analyse de longues séries temporelles de moyennes journalières de l'éclairement, pour différents endroits du monde, issues de mesures pyranométriques au sol, d'estimations satellitaires ou de réanalyses météorologiques .

Compte-tenu de la nature non linéaire et non stationnaire des données, la transformée adaptative de Hilbert-Huang est utilisée comme outil d'analyse pour tenir compte de la diversité de ces échelles temporelles. On montre ainsi la nature variable des échelles caractéristiques et de leur intensité, ainsi que leur dépendance vis-à-vis du climat.

L'application d'une technique adaptative de ré-échantillonnage fractionnaire montre la juxtaposition d'une composante déterministe et d'une stochastique. Pour tous les jeux de données, le cycle annuel déterministe représente la plus grande partie de la variabilité. Toutes les séries temporelle contiennent une composante de variabilité stochastique à haute fréquence, qui est modulée en amplitude par le cycle annuel.

L'approche permet également d'évaluer, échelle par échelle, les performances des estimations satellitaires ou issues de ré-analyses par comparaison avec des mesures pyranométriques au sol. Une étude de cas confirme que les estimations satellitaires surpassent les ré-analyses à toutes les échelles temporelles.

Mots Clés

Énergie solaire, rayonnement solaire reçu au sol, variabilité temporelle, télédétection, analyse temps–fréquence.

Abstract

The center of focus for this PhD thesis is the intrinsic temporal variability of the surface solar irradiance (SSI). The characteristic time-scales of variability are revealed by analysing long-term time-series of daily means of SSI, such as ground measurements, satellite estimates, or radiation products from global atmospheric re-analyses, for different geographical locations around the world.

To account for the wide range of the time-scales of variability, and given the non-linear and non-stationary nature of the data, the adaptive, data-driven Hilbert-Huang Transform is employed as an analysis tool. The time-varying nature of the characteristic time-scales of variability, along with variations in intensity, are thus revealed.

An adaptive fractional re-sampling technique is used to discriminate between the deterministic and the stochastic variability constituents. For all datasets, the deterministic yearly cycle is found to account for the largest part of variability. Furthermore, all time-series are found to contain a high-frequency stochastic variability component, that exhibit cross-scale amplitude modulation by the yearly cycle.

A refinement to existing methods for assessing the fitness for use of surrogate SSI products in lieu of ground measurements is also proposed. A case study confirms that satellite estimates outperform re-analyses across all time-scales.

Keywords

Solar energy, surface solar radiation, temporal variability, remote sensing, time–frequency analysis.