



HAL
open science

Intermittency of Global Solar Radiation over Reunion island: Daily Mapping Prediction Model and Multifractal Parameters

Li Qi

► **To cite this version:**

Li Qi. Intermittency of Global Solar Radiation over Reunion island : Daily Mapping Prediction Model and Multifractal Parameters. Physics [physics]. Université de la Réunion, 2018. English. NNT : 2018LARE0016 . tel-01943744

HAL Id: tel-01943744

<https://theses.hal.science/tel-01943744>

Submitted on 4 Dec 2018

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

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



**THÈSE DE DOCTORAT DE
L'UNIVERSITÉ DE LA RÉUNION**

spécialité

Physique Energétique

École doctorale de Science Technologie et Santé de L'Université de La
Réunion

Présentée par

QI LI

Pour obtenir le grade de
DOCTEUR de l'UNIVERSITÉ DE LA RÉUNION

**Intermittency of Global Solar Radiation over Reunion
Island: Daily Mapping Prediction Model and Multifractal
Parameters**

Dirigée par **Pr Miloud Bessafi**

Soutenu le 17 Juillet, 2018

devant le jury composé de:

Pr. Miloud BESSAFI

Pr. Richard Emilion

Dr. Muhammad Zaid Dauhoo

Dr. Dominique Gay

Université de La Réunion

Université d'Orléans

Université de Maurice

Université de La Réunion

Supervisor

Reviewer

Reviewer

Member

Acknowledgements

I would like to express my special appreciation and thanks to my supervisor, Professor Miloud Bessafi, who has always encouraged me to do this research and has developed myself as a researcher in the best possible way. I really remembered some sentences he had always told me before: “This is the science and congratulations, you are making your own way on science”. You have cared so much about my work, and you have always responded to my questions and queries so promptly. I have been extremely lucky to be a PhD student of yours.

I would also like to thank my committee members, Professor Richard Emilion, Doctor Muhammad Zaid Dauhoo and Doctor Dominique Gay for serving as my committee members even at hardship. I also want to thank you for letting my defending be an enjoyable moment, and for your brilliant comments and suggestions, thanks to you.

I would also like to thank every member in LE2P laboratory. Thanks to its director, Professor Chabriat, to Pratricks, Kelly, Mathieu and other colleagues in the lab. Everyone has been so nice and has given me big help for my research.

I would also like to thank Professor Huang Yongxiang from the University of Xiamen. He taught me how to understand the method HHT, and gave me suggestions at the beginning of my research.

I would also like to thank my family, my parents. Words could not express how grateful I am. Special thanks to my wife Peng LI, who is also as my college. Without you, this thesis would never have been finished. Your love and support without any complaint and regrets enabled me to complete my PhD study: I love you.

Abstract

Due to the heterogeneous and rapidly-changing cloudiness, tropical islands, such as Reunion Island in the South-west Indian Ocean (SWIO), have significant solar resource that is highly variable from day-to-day. In this study, we propose a new approach for deterministic prediction of daily surface solar radiation(SSR) maps based on four linear regression models: multiple linear regression (MLR), principal component regression (PCR), partial least squares regression (PLSR), and stepwise regression (SR), that we have applied on the SARAH-E@5km satellite data (CM SAF) for the period during 2007-2016. To improve the accuracy of prediction, the multifractal parameters (H , C_1 and α) are proposed to include as new predictors in the predictive model. These parameters are obtained from the analysis of SSR intermittency based on arbitrary order Hilbert spectral analysis. This analysis is the extension of Hilbert Huang Transform (HHT) and it is used to estimate the generalized scaling exponent $\xi(q)$. It is the combination of the Empirical Mode Decomposition and Hilbert spectral analysis (EMD+HSA). In a first step, the multifractal analysis is applied onto one-second SSR measurements from a SPN1 pyranometer in Moufia in 2016. The mean sub-daily, daily and seasonal daily multifractal patterns are derived, and the scaling exponent $\zeta(q)$ is analyzed. In a second step, the intermittency study is conducted on one-minute SSR measurements from a SPN1 network with 11 stations in 2014. The spatial patterns for all the stations with the multifractal parameters H , C_1 and α are shown. The variability of singularity spectrum width is considered to study the spatial intermittency at the daily and seasonal scale. Based on this intermittency analysis from measurements at several stations, the universal multifractal parameters (H , C_1 and α) could be taken as new predictors for indicating the multifractal properties of SSR.

Keywords: global solar radiation; mapping prediction; intermittency; multifractality; scaling exponent; arbitrary order Hilbert spectral analyses; linear regression model

Résumé

Les îles tropicales sont soumises à un ennuagement hétérogène et changeant rapidement. Par ailleurs, elles ont une ressource solaire importante mais significativement variable d'un jour à l'autre. Dans le sud-ouest de l'océan indien (SWIO), La Réunion fait partie de ces îles tropicales ayant un potentiel solaire colossal mais fortement intermittent. Dans cette étude, nous proposons une nouvelle approche de prévision déterministe des cartes journalières rayonnement solaire (SSR), basée sur quatre modèles de régression linéaire : une régression linéaire multiple (MLR), une régression en composantes principales (PCR), une régression des moindres carrés (PLSR) et une régression pas à pas (stepwise--SR). Ces quatre régressions sont appliquées sur les données satellites SARA-H-E (CM SAF) à 5km de résolution entre 2007 et 2016, en vue d'en effectuer la prévision. Pour obtenir de meilleures performances, nous proposons d'inclure les paramètres multi-fractale (H , C_1 et α) comme nouveaux paramètres prédictifs. Ceux-ci sont obtenus à partir de l'analyse de l'intermittence du SSR basée sur la méthode d'analyse d'ordre spectral arbitraire de Hilbert. Cette analyse qui est une extension de la transformation d'Hilbert Huang (HHT) est utilisée afin d'estimer l'exposant d'échelle $\xi(q)$. On effectue la combinaison d'une décomposition en mode empirique et de l'analyse spectrale de Hilbert (EMD + HSA). Dans une première étape, l'analyse multi-fractale est appliquée sur une mesure du SSR d'une seconde échelle à partir d'un pyranomètre SPN1 à Moufia en 2016. La moyenne infra journalière, journalière et saisonnière de la structure multi-fractale a été dérivée, et la loi d'échelle d'exposants $\zeta(q)$ a été analysée. Dans une seconde partie, l'analyse de l'intermittence est effectuée sur les mesures du SSR, d'une période d'une minute, à partir le réseau de SPN1 cotenant 11 stations en 2014. Les modèles spatiaux pour toutes les stations avec les paramètres multi-fractales H , C_1 et α sont mis en évidence. La variabilité de la largeur du spectre de singularité est considérée pour étudier l'intermittence spatiale et la multifractalité dans l'échelle quotidienne et l'échelle saisonnière. Sur la base de ces analyses d'intermittence faites sur les mesures de plusieurs stations, les paramètres multi-fractaux universels (H , C_1 et α) pourraient être choisis comme de nouveaux prédictifs afin d'indiquer les propriétés multi-fractales du SSR.

Mots clés: rayonnement solaire global; prédiction de cartographie; intermittence; multifractalité; exposant d'échelle; Analyses spectrales de Hilbert d'ordre arbitraire; modèle de régression linéaire

Contents

| | |
|---|------------|
| 1 Introduction..... | 1 |
| 1.1 Background..... | 2 |
| 1.1.1 Climate change and sustainable development | 2 |
| 1.1.2 Renewable energy sources: Solar radiation | 4 |
| 1.1.3 Solar radiation prediction challenge..... | 7 |
| 1.1.4 Surface solar radiation observation and studies..... | 9 |
| 1.1.5 Reunion Island: Toward a sustainable island..... | 11 |
| 1.1.6 Reunion Island: Synoptic and local cloud covers context..... | 14 |
| 1.2 Objectives and strategy of the thesis | 15 |
| 1.2.1 Statistical model: Solar radiation regression prediction..... | 16 |
| 1.2.2 Intermittency: nonstationary fluctuations | 17 |
| 1.2.3 A spatio-temporal prediction model..... | 18 |
| 1.3 Outline of the thesis..... | 19 |
| 2 Data and Methodology | 21 |
| 2.1 Data..... | 22 |
| 2.1.1 Available solar radiation observations | 22 |
| 2.1.2 Study dataset..... | 25 |
| 2.2 Methodology | 31 |
| 2.2.1 Linear mapping regression model..... | 31 |
| 2.2.2 Pre-processing and decomposition method..... | 37 |
| 2.2.3 Fractal and Multifractal | 52 |
| 2.2.4 Intermittency: Multifractal analysis..... | 56 |
| 3 Daily mapping prediction..... | 65 |
| 3.1 Data processing..... | 66 |
| 3.1.1 Missing data: Filling technique | 66 |
| 3.1.2 Daily scale: Detrended and deseasonalized data | 68 |
| 3.1.3 Dataset normalization | 69 |
| 3.2 Linear models | 70 |
| 3.2.1 Independent/dependent variables | 70 |
| 3.2.2 PCA decomposition | 70 |
| 3.3 Results | 74 |
| 3.3.1 Test with SR, MLR, PCR and PLSR models..... | 74 |
| 3.3.2 Score evaluation: MAE, MSE, and RMSE | 81 |
| 4 Intermittency: multifractal framework and results | 83 |
| 4.1 Intermittency and multifractality analysis on one SPN1 station | 84 |
| 4.1.1 Mean sub-daily and daily multifractal pattern..... | 84 |
| 4.1.2 Seasonal variability of daily multifractal pattern..... | 105 |
| 4.2 Intermittency and multifractality analysis of SPN1 network | 108 |
| 4.2.1 Spatial pattern of intermittency within the daily scale | 108 |
| 4.2.2 Seasonal Spatial pattern of intermittency within the daily scale | 113 |

| | |
|--|------------|
| 5 Conclusion and Perspective | 125 |
| 5.1 Daily mapping prediction: A new reduced dimension approach | 126 |
| 5.2 Intermittency: A first attempt of spatial and temporal characterization for Reunion Island..... | 126 |
| 5.3 Consistency of log-stable parameter to be used as predictors in the daily mapping prediction model..... | 129 |
| 5.4 Future work plan..... | 129 |
| References | 131 |
| Appendix A --- List of Acronyms | 144 |
| Appendix B --- Conferences Attended | 146 |
| Appendix C --- List of Publications..... | 146 |

Chapter 1

Introduction

1.1 Background

1.1.1 Climate change and sustainable development

- Anthropogenic forcing: Climate scenarios

The climate system has been strongly affected by human's activities in recent years and due to anthropogenic emissions of Green-House gases (GHG) that are now higher than at any point in history. Climate changes have had widespread impacts on human and natural systems in recent years. GHG trap heat and make the planet warmer. The atmosphere and the ocean have warmed, the amounts of snow and ice have diminished, and the sea level has risen. Human activities are responsible for almost all of the increase in GHG in the atmosphere over the last 150 years. Anthropogenic GHG emissions have increased since the pre-industrial era, driven largely by economic and population growth, and are now higher than ever, leading to atmospheric concentrations of carbon dioxide, methane and nitrous oxide that are unprecedented in at least the last 800,000 years (IPCC, 2014). The effects of increase GHG, together with those of other anthropogenic drivers, have been detected throughout the climate system and the increase in GHG emissions is extremely likely to have been the dominant cause of the observed warming since the mid-20th century.

The GHG emissions associated with the provision of energy services are a major cause of climate change. Most of the observed increase in global average temperature since the mid-20th-century is due to the observed increase in anthropogenic GHG concentrations. The global average temperature has increased by 0.76 °C between 1850 to 1899 and 2001 to 2005, and the warming trend has increased significantly over the last 50 years. Climate change is one of the great challenges in the 21st-century. The news from NASA point out that: "October 2017 was the second warmest October in 137 years of modern record-keeping, according to a monthly analysis of global temperatures by scientists at NASA's Goddard Institute for Space Studies (GISS) in New York" (<https://climate.nasa.gov/news/>). Changes in many extreme weather and climate events have been observed since 1950 (IPCC, 2014). Some of these changes have been linked to human influences, including a decrease in cold temperature extremes, an increase in warm temperature extremes, an increase in extreme high sea levels and an increase in the number of heavy precipitation events in a number of regions. Climate changes have resulted in the changes of nature and human systems all over the world in recent decades, and these effects from the climate change feedback also increase the vulnerability of natural and human systems, especially for small islands.

- Mitigation and adaptation: Island vulnerability

Islands are heterogeneous in geomorphology, culture, ecosystems, populations, and hence also in their vulnerability to climate change. Vulnerabilities and adaptation needs are as diverse as the variety of islands between regions and even within nation states, often with little climate adaptation occurring in peripheral islands. The small islands face the risks resulting from current and future climate changes, including sea level rise, sea surface temperature increases, changing rainfall patterns, tropical and extratropical cyclones, and increasing air temperature (Nurse et al., 2014). Given the inherent physical characteristics of small islands, it is confirmed that there are high level of vulnerability of small islands to multiple stressors, both climate and non-climate. The high diversity of small islands in both physical and human attributes and their response to climate-related drivers means that climate change impacts, vulnerability, and adaptation will be variable from one island region to another and between countries in the same region. In the latest IPCC's report (IPCC, 2014), it is pointed out that "island vulnerability is often a function of four key stressors: physical, socioeconomic, socio-ecological and climate-induced. Islands faced with multiple stressors can therefore be assumed to be more at risk from climate impacts". GHG emissions from most small islands are negligible in relation to global emissions, yet small islands will most probably be highly impacted by climate change (Srinivasan, 2010). The combination of island geography and economic types informs the extent to which adaptation and mitigation actions might interact (Nunn, 2009). The geography and location of islands affect their sensitivity to hydro-meteorological and related hazards such as cyclones, floods, droughts, invasive alien species, vector-borne disease, and landslides. The capacity of island residents to cope with those hazards is often related to income levels, resources endowment, technology and knowledge. Rather, long-term climate impacts depend on the type of island and the adaptation strategy adopted. It is necessary and urgent to conduct mitigation and adaptation to the risks posed by global climate change in small islands.

- Mitigation and adaptation: Renewable energy

The most severe impacts of climate change may still be avoided if efforts are made to transform current energy systems. There are multiple means for lowering GHG emissions from the energy system, while still providing desired energy services (Dincer, 2000). Renewable energy sources have a large potential to displace emissions of GHG from the combustion of fossil fuels and thereby to mitigate climate change. Climate changes, energy demands and increasing government attention are all driving the development of renewable energy. If implemented properly, renewable energy sources can contribute to social and economic

development, to energy access, to a secure and sustainable energy supply, and to a reduction of negative impacts of energy provision on the environment and human health. Renewable energy sources (solar, wind, biomass, tides, geothermal heat....) play a role on providing energy services in a sustainable manner and, in particular, in mitigating climate change (Omer, 2008).

1.1.2 Renewable energy sources: Solar radiation

- Clouds trigger space and time scale surface solar radiation

Renewable energy is energy that is collected from renewable resources, which are naturally replenished on a human timescale, such as sunlight, wind, rain, tides, waves and geothermal heat (Ellabban et al., 2014). It often provides energy in four important areas: electricity generation, air and water heating/cooling, transportation and a variety of commercial and industrial uses (REN21, 2010). Among those different kinds of renewable energy resources, solar energy is regarded as the most promising renewable energy resources. Solar radiation incident at the Earth's surface is the principal energy source for life on the planet and largely determines the climatic conditions of our habitats. It has also major practical implications, including solar energy technologies and agricultural productivity (Wild, 2009). According to the recent earth energy budget estimates by (Wild et al., 2013; Figure 1.1), 22% of solar radiation reaching the top of the Earth's atmosphere is reflected back to space by clouds and 23% is absorbed by the clouds, gases and aerosols in the atmosphere. Only 54% of solar radiation reaches the Earth's surface, which will be referred as the surface solar radiation (SSR) in this thesis. SSR depends on the geographic location, orientation of the surface, time of the day, time of the year and atmospheric components (Boes, 1981). In their paper, Chiacchio and Vitolo (2012) indicate that clouds are the primary factor determining the amount of solar radiation received at a given point on the Earth's surface. Clouds modulate solar irradiance in all time and space scales. Low, thick clouds primarily reflect solar radiation and cool the surface of the Earth. High, thin clouds primarily transmit incoming solar radiation (Graham, 1999). Figure 1.2 shows the global distribution of cloud cover for January and July afternoon, which give a general information about cloud's distribution and the effects on SSR over the year. Those monthly data contain an entrancing record of the annual cycles of the weather: in these plots, we can see the development of summer convection across the U.S. Plains, the evolution of the India and China monsoons, the onset of "The Wet" in Australia, the annual cycle of the Intertropical Convergence Zone (ITCZ) and many other seasonal changes in the cloud record.

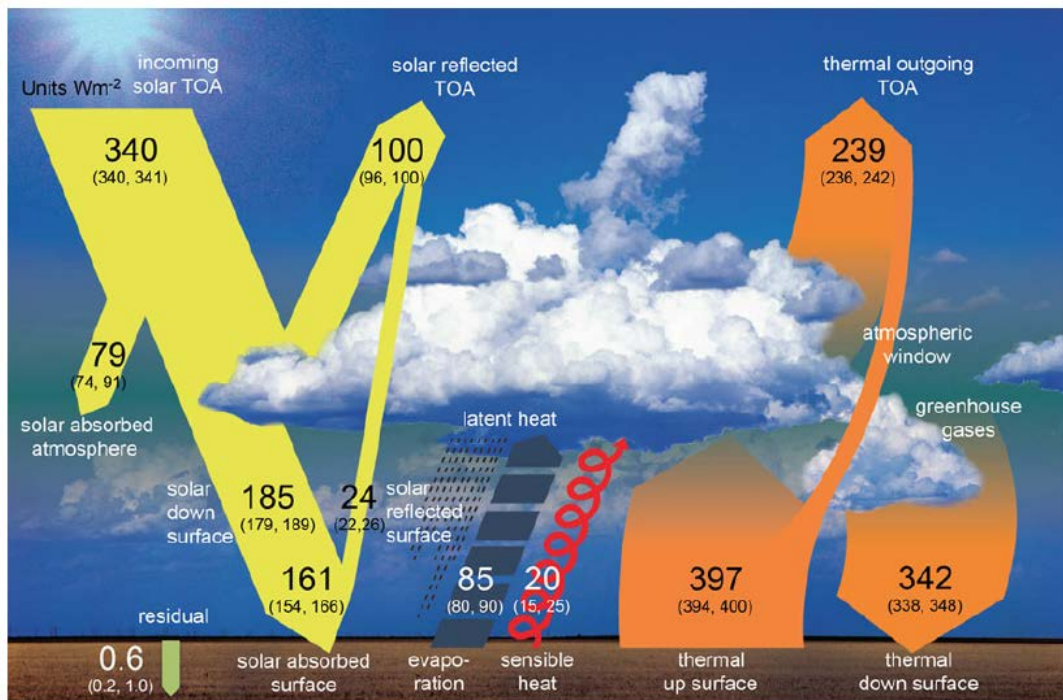


Figure 1.1 Schematic diagram of the global mean energy balance of the Earth. The numbers (W/m^2) indicate best estimates for the magnitudes of the globally averaged energy balance components together with their uncertainty ranges, representing present day climate conditions at the beginning of the twenty first century (from Wild et al., 2013).

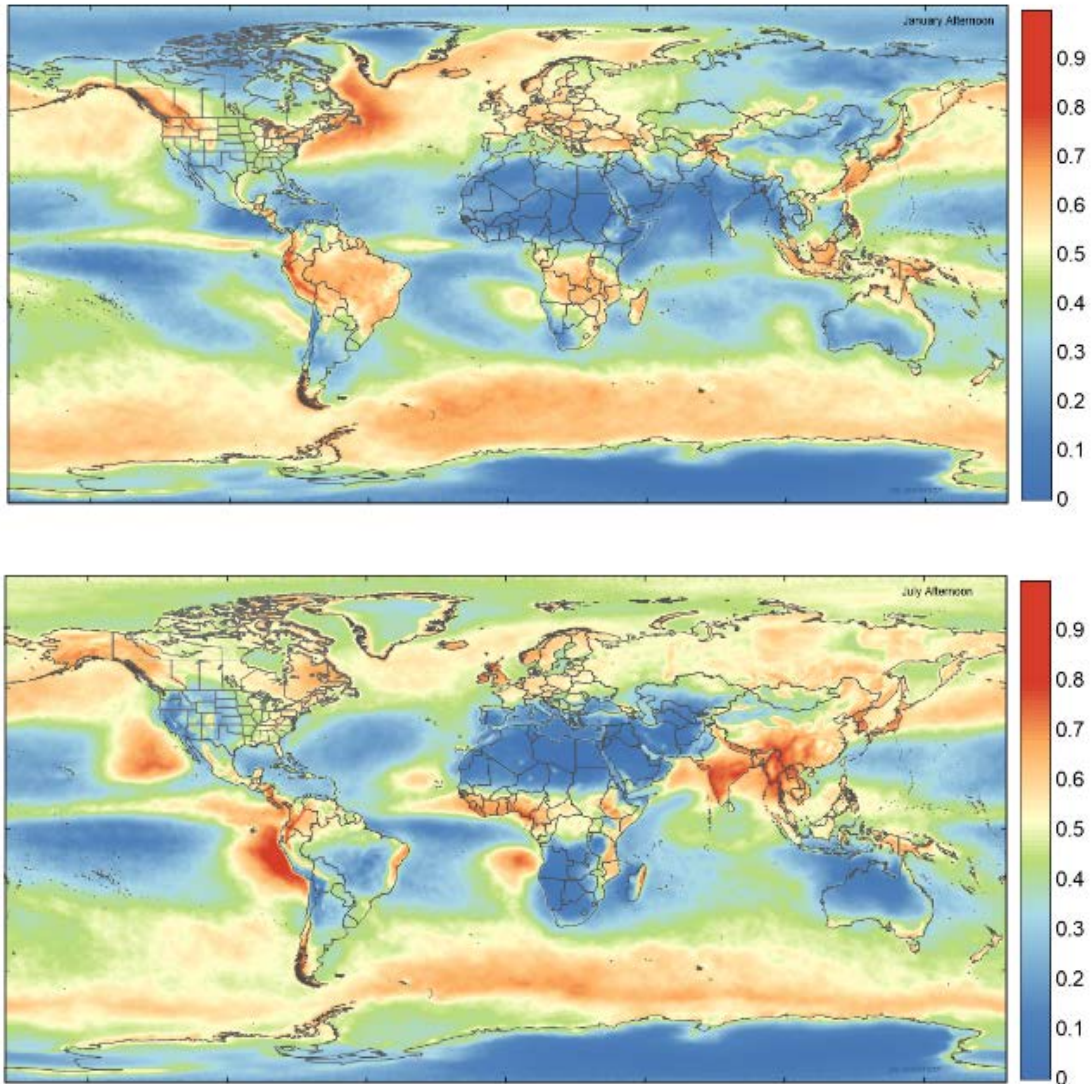


Figure 1.2 Global cloud cover for January and July afternoon as cloudiness derived from observations from the Aqua satellite (NASA, 2002-2015).

- Beyond seasonal cycle: Solar energy demands

In the current global energy system, fossil fuels still take the dominated place, even if the use of renewable energy has been increasing rapidly in recent years. Renewable energy costs are still higher than existing energy prices, but in various settings renewable energy is already competitive. Solar energy is one of the most promising resource and its technical potential is the highest among the renewable sources. Solar energy can be used directly for heating and lighting homes and other buildings, for generating electricity, for hot water heating, solar cooling, and for a variety of commercial and industrial uses (Garci, 2003). In recent decades,

the solar energy has been used for electricity generation based on photovoltaic (PV) technology. At this time, the demand of solar energy is increasing more rapidly than before.

- Deterministic versus stochastic variability: Intermittency

As pointed out previously, solar radiation received at the ground is triggered by clouds cover which fluctuate on different space and time scales. In addition, cloud cover pattern's complexity increases at small spatial and time scales like for example clouds related to turbulent forcing due to dynamic and thermodynamic instability in boundary layers. The variability and unpredictability of solar energy makes it belong to intermittent energy sources. In complex dynamical systems, intermittency is the irregular alternation of phases of apparently periodic and chaotic dynamics or different forms of chaotic dynamics (Edward, 2002). Different methods used to measure this phenomenon might give different definitions of intermittency. In the area of solar radiation, Tarroja et al. (2013) concern the concept of intermittency as the same meaning as fluctuation. They define the concept of the intermittency as the change in the magnitude of the total irradiation on a surface over a given time interval, which agrees with the concept of variability. On the contrary, Davis et al. (1997) imported the notion of intermittency from turbulence. They analyzed radiation data artificially generated from cloud data. The amplitude of the fluctuations or intermittency of solar irradiation could drastically increase in a few seconds or minutes in one place within one hour whereas other places would experience low fluctuations. This can take place at different time during the day. This is mainly due to intermittent cloud covering triggered by orographic turbulence, lee waves, and prevailing trade winds.

1.1.3 Solar radiation prediction challenge

The solar irradiance is variable and intermittent. It is not a deterministic variable due to the meteorological conditions. Cloud presence is the most important factors for solar radiation because the attenuation process by clouds is highly stochastic and it is difficult to predict how it will affect solar radiation. Estimating the parameters that would express the characteristics of solar energy in the prediction process is the challenge of this work. The intermittency study is the important basis for the prediction of solar radiation. There are a lot of limitations for the development and the utilization of this intermittent solar energy. A good knowledge of global solar radiation variations and accurate solar forecasts are needed. According to the input data, or time scale, SSR forecasting can be categorized into different types:

- Statistical and stochastic model

Statistical models based on solar radiation ground-based measurements, which are applied for forecast horizon from 5 minutes up to 6 hours. These statistical models are based on linear method (Glasbey and Allcroft, 2007; Huang et al., 2003; Voyant et al., 2014) and non-linear methods which are based on artificial intelligence techniques (Mellit et al., 2005), such as Artificial Neural Network (ANN) (Monjoly, 2017). Kutty et al. (2015) developed a regression model to predict the monthly global solar irradiance in Malaysia. Nalina et al. (2014) gives a description of regression models for forecast solar radiation at short term and the multivariate regression is carried out for to achieve higher accuracy. ANN provides a better way to predict solar radiation using more meteorological and geographical parameters as input, under the assumption that there exists a non-linear relationship between solar radiation and those parameters in the ANN model (Sivamadhavi, 2012). Mohandes et al. (1998), Reddy and Ranjan (2003), and Fadare et al. (2010) used the ANN model to predict the global solar radiation and justified the application of this non-linear model. ANN is regarded as one of the best tools for non-linear prediction (Monjoly et al., 2017; Paoli et al., 2010).

Kaplanis and Kaplani (2007) suggested a cosine wave correlation for daily global solar radiation and later they developed a stochastic model that uses measured solar radiation values in early mornings and predicts the radiations for the rest of the day followed by the approach in 2007 (Kaplanis and Kaplani, 2010). Hocaoglu (2011) also proposes a novel stochastic method for solar radiation modeling which uses the hidden Markov models.

- Dynamical model

Instead of using ground-based and satellite data, atmospheric models have been implemented in the solar radiation's prediction study a lot in recent years. Though the accuracy of those dynamical model is significantly less than the current satellite-based techniques (Lohmann et al., 2006; Kennedy et al., 2011; Bojanowski et al., 2014), they provide an alternative approach for solar radiation prediction. The Numerical Weather Prediction (NWP) models (Ruiz-Arias et al., 2011b; Ruiz-Arias et al., 2013a; Diagne et al. 2014) and Regional Climate Models (RCMs) are two main types of dynamical models. Using dynamical models to study solar radiation has its own advantages compared satellite methods, as they can perform longer period simulation with the whole atmospheric system to make a comprehensive analysis.

- Time scale of interest

Monjoly et al. (2017) made a hourly forecasting of global solar radiation using hybrid

method, in which they categorize forecasting approach of the solar radiation to different types based on the time scale of the input data: 1) dynamical models, which are applied fully by (Heinemann et al., 1990; Lorenz et al., 2012; Perez et al., 2011). These modelling processes have robustness for time scale higher than 6h forecast horizon. 2) Short-term radiation forecasting models (forecast horizon from few minutes to hours) based on clouds information using sky cameras and satellite images (Bosh et al., 2005; Marquez and Coimbra, 2013; Coimbra and Pedro, 2013). 3) Statistical models based on solar radiation ground-based measurements, which are applied for forecast horizon from 5 minutes up to 6 hours.

1.1.4 Surface solar radiation observation and studies

Solar radiation incident at the Earth's surface is the principal energy source for life on the planet and largely determines the climatic conditions of our habitats (Wild et al., 2009). Surface solar radiation as input parameter is critical for the mapping prediction in the photovoltaic industry (Hussain et al., 1999; Davy and Troccol, 2012; Monforti et al., 2014). There are different ways to predict the surface solar radiation typed by the dataset:

- Ground-based historical data

Solar radiation measurements directly from the ground stations in long-term are regarded as the most accurate source. However, with the difficulty of maintaining the equipment and the funding, the available direct measurements of solar radiation from radiometers are usually not sufficient. Figure 1.3 shows the geographical distribution of surface radiative fluxes observation sites from the Global Energy Balance Archive (GEBA, Gilgen et al., 1998; Ohmura et al., 1989; Wild et al. 2017), and the database of the Baseline Surface Radiation Network (BSRN, Ohmura et al., 1998). When the ground stations of solar radiation are sparse, there are some methods to obtain solar radiation estimates for locations where it is not directly measured. One simple way is to use spatial interpolation methods (Bechini et al., 2000; Ertekin and Evrendilek, 2007). However, the density of solar radiation measurements (Wild et al., 2013) is often not sufficient for reliable interpolation.

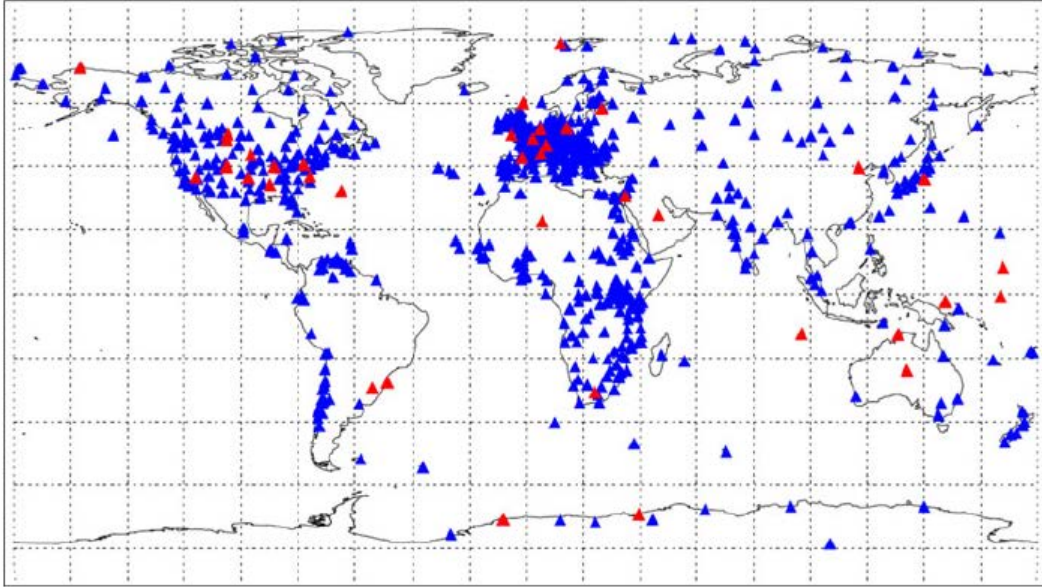


Figure 1.3 Geographical distribution of surface radiative fluxes observation sites from GEBA (760 sites in blue) and from BSRN (42 sites in red). From Wild et al., 2013.

- Satellite historical database

Satellite-based techniques have become the most common method for solar resource prediction at regional, continental or global scales (Brisson et al., 1999; Laszlo et al., 2008; Posselt et al., 2012a; Gueymard and Myers, 2008b) in recent decades. The available satellite dataset for solar radiation over South-West Indian Ocean (SWIO) is from the European Organization for the Exploitation of Meteorological Satellites (EUMETSAT), which established within its Satellite Application Facility (SAF) network a dedicated center, the SAF on Climate Monitoring (CM SAF). CMSAF is the dataset used in this study, satellite observations from geostationary weather satellites for solar radiation also include the American Geostationary Operational Environmental Satellites (GOES), the Indian National Satellite (INSAT), the Chinese Fengyun and the Japanese Geostationary Meteorological Satellite (GMS) (<http://www.cmsaf.eu>).

- Background on solar radiation studies

The best database would be the long-term measured data at the site of the proposed solar system. Solar radiation can be estimated through the empirical models using other available meteorological observations when there are not enough measurements (Ulgen and Hepbasli, 2009; Yohanna et al., 2011; Li et al., 2011; Karakoti et al., 2012). Sivamadhavi (2012) made the studies based on empirical models: Veeran and Kumar (1993) developed an Angstrom type

regression model correlating global radiation and sunshine hours to estimate the daily SSR at two locations in India. Ahmad and Ulfat (2004) employed the regression technique and proposed an Angstrom type empirical equation of first and second orders to for determine the daily global radiation at Karachi, Pakistan. Chandal et al. (2005) correlated the daily global radiation with temperature, latitude and altitude for various places in Egypt, Kuwait, Lebanon, Sudan and Saudi Arabia. Sabziparvar (2008) modified three existing radiation models to estimate daily global radiation in various cities in central arid desert of Iran by including parameters such as altitude, Sun-Earth distance and number of dusty days. A number of researchers previously evaluated the ability of high-resolution ($\sim 0.5^\circ$) regional climate models (RCMs) to estimate the mean state and variability of the SSR over the Arctic (Wyser et al., 2008), North America (Markovic et al. 2008), Europe (Kothe et al. 2011; Chiacchio et al. 2015) and West Africa (Kothe and Ahrens 2010). Zamora et al. (2005) and Lara-Fanego et al. (2012) tested mesoscale models' ability for short-term forecasts of global horizontal irradiance with high resolution at a few kilometers at some locations in the United States and Spain.

1.1.5 Reunion Island: Toward a sustainable island

- Demography and energy demands: Through last decades

Reunion Island is a French overseas territory and it is the only European region in the southern hemisphere. It lies at 20.8° South and 55.5° East, in the South-West Indian Ocean (SWIO) and is 800km off the east coast of Madagascar. It has a population of 842,767 in 2016 with a total area of 2512 km².

Introduction



Figure 1.4 SWIO area. The star indicates the position of Reunion Island. From Google Maps.

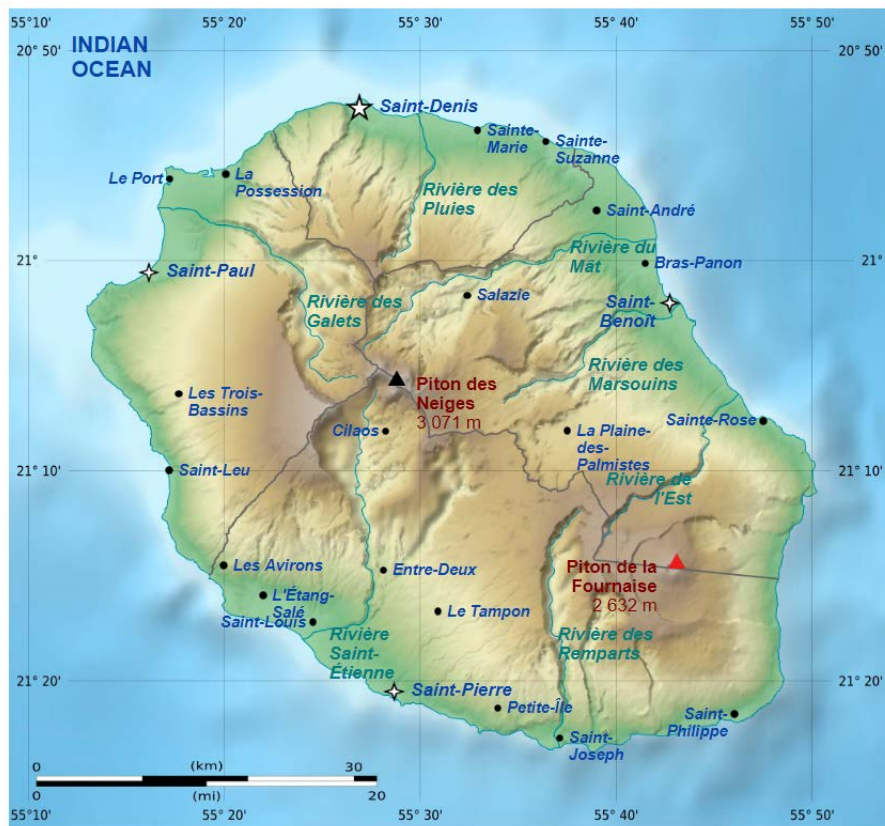


Figure 1.5 Map of Reunion Island. From Google Maps.

- Renewable energy: Solar resource among others

Even if Reunion Island is a French territory, it is far away from the metropole and any European territories or continent territories. So, it is difficult to make use of their energy production network, which make Reunion Island's development deeply relating on its own geography characters and energy production. Based on Reunion Island's location, renewable energy is regarded as a sustainable alternative for providing energy service to achieve its development. And solar energy is the most promising renewable energy among wind energy, ocean energy, hydro energy, biomass energy which are indirectly solar energy forms (Li, 2015). The solar technology research could apply to the abundant solar energy resource in Reunion Island for improving the energy production.

- Potential of solar energy

Figure 1.6 shows Global Horizontal Irradiance (hereafter GHI) (kWh/m^2) distribution and average annual temperatures on Reunion Island [Météo France, 2011]. Annual sunshine of Reunion Island is in the range of 1400-2500 h and can reach the value of 2900 h at the altitude lower than 400m. The monthly daily radiation reaches more than 6.5 kWh/m^2 during the austral summer season in some parts of the coastal region (altitude $< 300 \text{ m}$). Daily insolation is characterized by a strong evolution due to orographic cloud formations on the mountains. The two main applications of solar energy in Reunion are solar water heating and photovoltaic electricity (PV). And Reunion is actively engaged in the development and promotion of solar thermal and PV energy through various action plans and programs in the context of sustainable development, such as GERRI, PRERURE and so on, because of its solar energy potential.

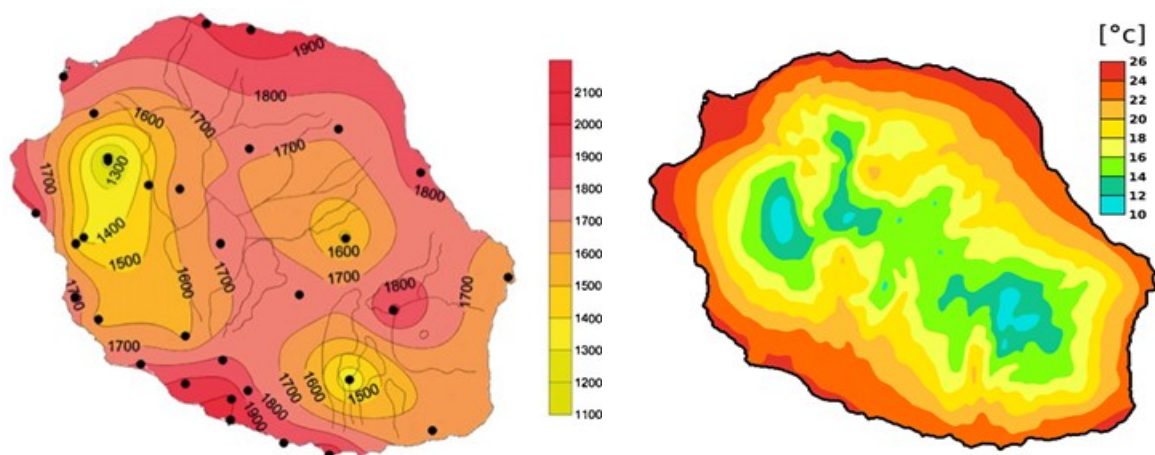


Figure 1.6 a) Annual average of Global Horizontal Irradiance (kWh/m^2) distribution. b) Annual average temperatures on Reunion Island [Météo France. 2011].

1.1.6 Reunion Island: Synoptic and local cloud covers context

- Tropical weather: Trade wind versus sea/land breeze

Reunion Island has a tropical maritime climate marked by two seasons (austral summer: November to April; austral winter: May to October) depending on the behavior of the southern Hadley cell and the Walker circulation (Baldy et al., 1996). Winter seasons are characterized by large-scale atmospheric subsidence and broken low-level cloudiness, while in summer, clouds are found both at low and high altitudes (Badosa et al., 2013). Morel et al. (2014) studied the precipitation over the island and indicated that: the prevailing winds named trade winds blow from the east and south-east to the west coast and the east coast has more rain than the west: average annual rainfall amounts present a marked west-east gradient, reaching cumulated values larger than 10-12m in the elevated sectors facing the dominant moisture fluxes associated with the trade winds. During the daytime, slope heating and differential sea-land heating result in anabatic wind and sea breeze that combine to create an inland flow forcing air up the slope and cloud formation along. The return flow is created by the combined blocking effect of the orography and the inversion layer above, brings clouds aloft towards the coast. The night scenario with land breeze and katabatic wind blowing down slope and clearing the island of orographic clouds (Badosa et al., 2015).

- Complex terrain: Leeward/windward/overcast clouds

Reunion is a mountainous island with very complex topography. It is famous for the two volcanos: the 'Piton des Neiges' (3070m) and the 'Piton de la Fournaise' (2632m, Figure 1.5). And there are three major central depressed areas: the Cirques of Mafate in the northwest, Salazie in the northeast and Cilaos in the southwest. Due to this complex terrain, Reunion island has its own typical synoptic characters as (Badosa et al., 2015) stated: on the windward side of the island, the thermal wind blows in a direction opposite to the trade wind, creating a convergence zone. The convergence zone is located few kilometers from the coast. On the windward side, trade winds slow down and orographic clouds are very frequent. The leeward side is not affected directly by the trade winds; thermal onshore winds prevail leading to topographic clouds. On the coasts parallel to the synoptic wind direction, trade winds are accelerated due to the Venturi effect and clouds tend to be blown away.

- Marine boundary layer: Turbulence processes

The marine boundary layer (MBL) structure is important for the marine low cloud processes and the exchange of heat, momentum, and moisture between the ocean and the low atmosphere.

Clouds, precipitation and aerosol all play important related role on the MBL with turbulent energy. Therefore, the MBL also relates with the SSR. The turbulence in MBL trigger dynamic and thermodynamic processes like cloud formation and movement. As well known that turbulent motions occur over a wide range and long-term time scale. The variation in length and time scales is an important characteristic of turbulent flows. In the numerical and theoretical analysis of turbulent flows, there is always difficulty to describe that characteristic. The scaling laws are regarded as an important tool for describing turbulence, as the turbulence is based on order of magnitude estimates which follow from logical applications of scaling laws and dimensional analysis. The nature of the scaling behavior for different datasets is determined using q th-order structure functions analysis. The nonlinearity and concavity of the scaling exponent functions $\zeta(q)$ reveal the intermittent and the multifractal properties of datasets, which result from the complex interaction of the turbulent atmospheric and the energy converter systems. A scaling regime or power law presents the correlation of the form $f^{-\beta}$ over a broad range of time scales. The exponent spectral β is close to the exact 5/3 Kolmogorov value in the Fourier analysis, and 1/3 Kolmogorov value for multifractal analysis (Kolmogorov, 1941; Frisch, 1995; Seuront and Schmitt, 2005).

1.2 Objectives and strategy of the thesis

Because of the heterogeneous and rapidly changing cloudiness, tropical islands, such as Reunion Island in the SWIO, have significant solar resource that is highly variable from day-to-day. In this study, the deterministic daily SSR mapping prediction is proposed as a new approach. Depending on the results and the quality of the prediction, the intermittency analysis with the arbitrary order Hilbert spectral analyses method are applied to get the multifractal parameters as new predictors in the new prediction model. It could achieve better mapping performance. This is the first step of SSR daily mapping. Nevertheless, there is inherent limitation with this deterministic multivariate approach in regard to the SSR intermittency component. Obviously, it is expected that the stochastic part of cloud cover pattern form day-to-day must play a crucial role in SSR daily fluctuations. Thus, the daily stochastic fluctuation of SSR must be implemented in the deterministic native mapping model. This could be achieved using additional tools for non-stationary time series (Lyapunov exponents, state-space reconstruction, multifractal analysis ...). In this study, multifractal analysis is performed to quantify the intermittency of SSR over Reunion Island. The log-stable model is used to evaluate daily variations of three parameters (H, C_1 and α) which characterize the SSR

fluctuations observed at La Reunion. These parameters as a potential set of predictors in the mapping model prediction take account the daily stochastic fluctuation of SSR.

1.2.1 Statistical model: Solar radiation regression prediction

- Mapping approach: One-day ahead prediction

There are many studies on SSR prediction model which are based on one site time series prediction (univariate or multivariate model). The goal of the present work is to build a spatio-temporal multivariate model for mapping prediction. Each pixel of Reunion map could be predicted. So our new concept for one-day ahead SSR prediction is to combine spatial and temporal correlation for building a set of daily maps using SSR daily satellite data (CM SAF@5km). The regression model, including multiple linear regression (MLR), principal component regression (PCR), partial least squares regression (PLSR), stepwise regression (SR). Most common ones are tested at the first step in the mapping prediction.

- Dimensionality reduction

A regression model needs a set of predictors and predictands as we know. In this study, the satellite data used to do prediction is from SARAHE (CM SAF @5km), which covers Reunion Island with 360 pixels. In this sense, there is a large set of predictor and predictands. Building a prediction model with satellite data is not easy when the number of variables increases. It would cost a large CPU time to perform the regression process. Finding a way to reduce the size of the dimension set of the variables could be the first step in the analysis. There are many methods of dimension reduction. Some of these are Principal Component Analysis (PCA), Canonical Correlation Analysis (CCA), Kernel PCA, and k-Nearest Neighbours (k-NN) (Altma, 1992). PCA is commonly used in climate research as a tool to analyse meteorological fields with high spatio-temporal dimensionality. PCA aims at finding a new set of variables that capture most of the observed variance from the data through a linear combination of the original variables (Kutzbach, 1967; Wilks, 1995; Storch, 1993; Storch and Zwiers, 1999; Jolliffe, 2002).

- Stationary/nonstationary fluctuations: predictor set

As we use the linear regression model to conduct the mapping prediction, one important process in the linear model is to add new predictors that take into account the variability of the fluctuations. Because solar radiation belongs to nonstationary fluctuations, there are many options to set the predictors. The most common are the variance, skewness, kurtosis which are

based on second, third and fourth moment of the fluctuations respectively, in time series. Nevertheless, this moment based methods are not relevant for nonstationary time series. Thus, if the mapping prediction with linear regression model does not give accurate enough SSR results, other parameters that take into account the nonstationary processes must be introduced. To achieve this goal, the intermittency analysis of SSR needs to be performed.

1.2.2 Intermittency: nonstationary fluctuations

- Power law scale: Mono/multifractal processes

Verifying the existence of the fractal behavior in time series is the way to evaluate the non-stationarity of those time series. The power law scale is a common approach to characterize nonstationary fluctuations. The scaling exponent $\zeta(q)$ is used to characterize the scaling behaviour or measure the distance between monofractal and multifractal processes. The scaling exponent $\zeta(q)$ is linear for monoscaling processes and non-linear for multifractal processes. Several methods exist for estimating the scaling exponents $\zeta(q)$: structure function analysis, wavelet-based methods, detrended fluctuation analysis or multifractal detrended fluctuation analysis and arbitrary order Hilbert spectral analysis (Huang et al., 2008; Calif et al., 2013). In this study, the arbitrary order Hilbert spectral analysis which are the extension of Hilbert Huang Transform (HHT) (Huang et al., 1998, 1999) is used to estimate the scaling exponents $\zeta(q)$.

- Multifractal analysis: Log-stable model

As indicated before, the other parameters should be introduced into the linear regression model in order to take into account the nonstationary processes. Multifractal analysis could be helpful to give a way of building new predictors. In order to model intermittency or do multifractal analysis, many statistical functions have been proposed to fit the scaling exponents $\zeta(q)$. The log-stable model or universal multifractal proposed by Schertzer and Lovejoy (1987) and Kida (1991) is a multifractal model which can give some information about stochastic fluctuations. In that model, the scaling exponent is given by:

$$\zeta(q) = qH - \frac{C_1}{(\alpha-1)} (q^\alpha - q)$$

where $H = \zeta(1)$ the Hurst parameter defines the degree of smoothness or roughness of the field. The parameter C_1 is the fractal co-dimension of the set giving the dominant contribution to the mean ($q = 1$) and bounded between 0 and d (d the dimension space). It measures the inhomogeneity mean or the mean intermittency characterizing the sparseness of the field: the larger C_1 , the more the mean field is inhomogeneous. The multifractal Lévy parameter α is bounded between 0 and 2, where $\alpha = 0$ corresponds to the monofractal case and $\alpha = 2$

corresponds to the multifractal lognormal case. The parameter α measures the degree of multifractality. The concavity of $\zeta(q)$ is a characteristic of the intermittency (Frisch, 1995; Schertzer et al., 1997; Vulpiani and Livi, 2004).

- Daily multifractal parameters as predictors

The aim of this thesis is to study the daily mapping prediction of SSR over Reunion Island, and therefore daily predictors are needed. The CM SAF @5km satellite hourly data is used in this study. Therefore, pre-processing the data in daily scale as the first step can extract the infra-daily information for building daily multifractal parameters.

1.2.3 A spatio-temporal prediction model

- Toward a first step of a new mapping prediction

Based on the multifractal analysis, the multifractal parameters could be obtained with the log-stable model for the SSR. Through the mapping prediction work by the linear regression model and the intermittency study for the surface solar radiation, the multifractal parameters are taken into account of building a new spatio-temporal prediction model. In the first step, the multifractal log-stable parameters as the new predictors is used for the model, then the linear regression is preformed to do mapping prediction. Thus, the new linear model could integrate the multifractal parameters, producing more accurate prediction results.

- Synopsis of the model of one-day ahead prediction

Through the statement above, the model of one-day ahead prediction for SSR can be summarised as follows: 1) the arbitrary order Hilbert spectral analyses are applied to study the intermittency and multifractality of global solar radiation over Reunion Island. The method Empirical Mode Decomposition (EMD) (Huang et al., 1998, 1999) is used to decompose the input data. Through EMD, we can get the result of Intrinsic Mode Functions (IMFs). Each IMF has its own instantaneous frequency and amplitude. In order to analyze the arbitrary-order moments, a representation of the joint probability density function (PDF) of the solar radiation is estimated (Huang et al., 2008). The Hilbert spectrum is also calculated and the time scaling $\zeta(q)$ is fitted by a least squared method. After the intermittency or multifractal analysis, three multifractal parameters: H , C_1 and α are derived for every day and taken to study the multifractal process of the global solar radiation, which could be used as the new predictors; 2) The prediction is conducted using the linear regression model in which the new multifractal predictors are included. SR, MLR, PCR and PLSR which are common linear regression models

are applied to compare the prediction results; 3) The mapping is then done again with the prediction results.

- Fourth moment order versus intermittency parameters

As mentioned before, the variance, skewness, and kurtosis are based on two, three, and four moments of the fluctuations respectively in time series sets. Skewness characterizes only the shape of the distribution of the values, and the Kurtosis measures the relative peak or flatness of a distribution. When the multifractal parameters are considered, moments of order greater than four are required. The power law is extracted through the q th-order spatial structure function which is from the arbitrary order Hilbert spectral analysis. The scaling exponents $\zeta(q)$ (arbitrary-order statistical moment $q \geq 0$) expresses the multifractality, and the $\zeta(1)-1$ is the generalized Hurst exponent. As the statement above, the process is stationary if the spectral exponent (scaling exponent $\zeta(q)$ for Hilbert analysis) is less than 1 in Fourier analysis. So here, the generalized Hurst exponent cannot be used for stationarity.

1.3 Outline of the thesis

Chapter 2 presents the data and methodology used in this thesis. All the available solar radiation observations, which include EUMETSAT satellite data and ground-based data are listed and the dataset used in our study is also given. The linear regression model, including MLR, PCR, SR, PLSR, the data pre-processing and decomposition methods, and the intermittency study method-arbitrary order Hilbert Spectral Analysis are the main methods used in the thesis and introduced in detail in that chapter.

Chapter 3 shows the daily mapping prediction results with four different linear regression models: MLR, PCR, SR and PLSR. Initially, the SARAHE@5km satellite data (CM SAF) is pre-processed, then linear regression models are applied to do prediction. Finally, statistical analysis using MAE, MSE and RMSE is conducted in order to check the prediction results.

Chapter 4 studies the intermittency of the surface solar radiation and get the multifractal parameters. The analysis results are present in two parts. At first, the multifractal analysis is applied on one SPN1 station using second scale dataset. The mean sub-daily and daily and seasonal daily multifractal pattern are obtained, and the scaling exponent $\zeta(q)$ is analyzed. In the second part, the intermittency study is conducted on the SPN1 network using minute scale dataset. The spatial pattern for all the stations with the multifractal parameters H , C_1 and α are presented.

Chapter 5 gives the conclusion and outlines a new daily mapping prediction model with the multifractal parameters obtained from the intermittency study. As this study is based on the

linear regression prediction with multifractal parameters, the non-linear prediction models, such as ANN can be another way to fulfil the objective, which could be regarded as future development of the present work.

Chapter 2

Data and Methodology

2.1 Data

The most accurate source of solar radiation data is from direct measurements of solar radiation at ground stations, provided that the devices are well maintained and regularly calibrated. Various methods have been developed in order to obtain estimates of SSR for locations where it is not directly measured. One way is to apply interpolation from nearby stations. However, this method depends on the density of solar radiation measurements. Satellite observations also give an alternative way to derive solar radiation for over 30 years (Pinker and Laszlo, 1992; Gautier et al., 1980; Brisson et al., 1999; Lefèvre et al., 2007; Geiger et al., 2008; Laszlo et al., 2008). Due to the specific location of Reunion Island, the observational networks from which we can get data are rather limited. The available solar radiation observations, which include satellite and ground-based measurements and the dataset used in this study are presented separately in this section.

2.1.1 Available solar radiation observations

2.1.1.1 Satellite : EUMETSAT

- METEOSAT East: Instruments on board & space/time resolution

The European Organization for the Exploitation of Meteorological Satellites (EUMETSAT) is an intergovernmental organization and was founded in 1986. Its purpose is to supply weather and climate-related satellite data, images and products -24 hours a day, 365 days a year - to the National Meteorological Services of the Member and Cooperating States in Europe and other users worldwide. EUMETSAT established within its Satellite Application Facility (SAF) network a dedicated center, the SAF on Climate Monitoring (CM SAF, <http://www.cmsaf.eu>). CM SAF provides surface solar radiation climate data record based on observations from the eastern METEOSAT satellites situated over the Indian Ocean (these satellites which are officially named METEOSAT Indian Ocean Data Coverage (IODC), which are referred to as “METEOSAT East”). METEOSAT East covers [70S-70N, 10W-130E] and provides monthly, daily, and instantaneous data products. The resulting database consists of hourly values of global horizontal and direct normal irradiance at the Earth surface with a spatial resolution similar to the native pixel resolution of the satellite images.

- CMSAF products

The CM SAF's project objective is to provide homogeneous datasets relevant for climate monitoring, without the need of near real-time delivery of solar radiation estimates. CM SAF products are categorised into several groups and types. On one hand, there is the group of

routinely and near-real time produced data records in support to climate monitoring, called “Operational Products” (Environmental data record, EDRs and Interim climat data record, ICDRs). On the other hand CM SAF offers retroactively produced climate data records based on carefully intersensor calibrated radiances, called “Climate Data Records” (Thematic climat data record, TCDRs and Fundamental climat data record, FCDRs).

Operational Products include cloud products, surface radiation products, radiation fluxes at the top of atmosphere and water vapour and temperature products. Climate Data Records contain climate data records with DOI, cloud products, surface radiation products, radiation fluxes at the top of atmosphere, water vapour and temperature products and miscellaneous.

- Surface Radiation products

A 18-year long (1999-2016) surface solar radiation climate data records based on observations from the MVIRI instruments onboard the Meteosat First Generation has been generated by CM SAF: Surface Solar Radiation Data records – Heliosat (SARAH) - East. Each surface radiation parameter including the SIS is gridded onto a regular lat-lon grid with a resolution of $0.05 \times 0.05^\circ$ which is approximately 5km x 5km. The available time resolution ranges are hourly, daily mean and monthly mean values. All surface solar radiation data records are introduced in Table 2.1 with associated acronyms and units.

Table 2.1 SARAH-East data records

| Acronym | Product title | Unit |
|----------------|---------------------------------------|------------------|
| SIS | Surface Indoming Shortwave Irradiance | W/m ² |
| SID | Surface Indoming Direct Irradiance | W/m ² |
| DNI | Direct Normal Irradiance at surface | W/m ² |

2.1.1.2 Ground-based: pyranometer

- Technical overview

There is a relative dense world-monitoring network using relatively inexpensive but reliable pyrhelimeter and pyranometer instruments for broadband solar radiation measurements from the surface, with measurements starting in the 1880s (Roosen et al., 1973; Stothers, 1996). A pyranometer is a radiometer used to measure both the solar global horizontal irradiance (GHI) and diffuse horizontal irradiance (DHI) arriving at a particular location on the Earth’s surface.

A pyranometer needs to be properly calibrated from time to time in order to give the most possible accurate measurement results. The frequency at which a pyranometer needs to be calibrated depends on the type of pyranometer and environmental conditions that it is under use. A typical pyranometer does not require any power to operate. However, recent technical development includes use of electronics in pyranometers, which do require (low) external power.

- Measurement outputs

Global solar radiation is typically measured by a pyranometer. The global solar radiation (S_{Global}) detected by a pyranometer on the ground is composed of two components, the diffuse radiation (S_{Dif}) and the direct radiation (S_{Dir}), given by

$$S_{Global} = S_{Dif} + S_{Dir} = \int_{\lambda_1}^{\lambda_2} S_0(\lambda) T_{Dif}(\lambda, \mu_0) d\lambda + \mu_0 \int_{\lambda_1}^{\lambda_2} S_0(\lambda) T_{Dir}(\lambda, \mu_0) d\lambda \quad (2.1)$$

where $S_0(\lambda)$ is extraterrestrial solar irradiance at the wavelength λ , λ_1 and λ_2 are lower and upper spectral limits of the pyranometer, T_{Dif} and T_{Dir} are the spectral diffuse and direct transmittances, and μ_0 is the solar zenith angle cosine. The pyranometer usually has a spectral response between 0.3 and 3 μm .

- Calibration process and quality control

One should always double check with the manufacturer or recommendations of the World Meteorological Organization (WMO) to determine the most appropriate calibration frequency for a pyranometer under use in a particular application. For example, International Electrotechnical Commission (IEC, 2017) 61724-1:2017 standard for solar photovoltaic (PV) systems performance monitoring specifies that a pyranometer must be calibrated once every year for monitoring in utility scale PV systems while once every two years for monitoring large commercial PV systems.

According to the WMO's Guide to Meteorological Instruments and Methods of Observation (CIMO Guide), there are a variety of methods for calibrating pyranometers using the sun or laboratory sources. The following six methods are described in the WMO's CIMO Guide (WMO 2008):

1. By comparison with a standard pyrliometer for the direct solar irradiance and a calibrated shaded pyranometer for the diffuse sky irradiance.
2. By comparison with a standard pyrliometer using the sun as a source, with a removable shading disc for the pyranometer.

3. With a standard pyrheliometer using the sun as a source and two pyranometers to be calibrated alternately measuring global and diffuse irradiance.
4. By comparison with a standard pyrheliometer using the sun as source, under other natural conditions of exposure (for example, a uniform cloudy sky and direct solar irradiance not statistically different from zero).
5. In the laboratory, on an optical bench with an artificial source, either normal incidence or at some specified azimuth and elevation, by comparison with a similar pyranometer previously calibrated outdoors.
6. In the laboratory, with the aid of an integrating chamber simulating diffuse sky radiation, by comparison with a similar type of pyranometer previously calibrated outdoors.

It is mentioned in the WMO's CIMO Guide that the above-listed methods 1, 2, 3, and 4 are commonly used. However, it is essential that, except for method 2, either the zero irradiance signals for all instruments are known or pairs of identical model pyranometers in identical configurations are used. Ignoring these offsets and differences can bias the results significantly. Method 3 is considered to give very good results without the need for a calibrated pyranometer.

2.1.2 Study dataset

2.1.2.1 Satellite database: CMSAF SARAHE

- Spatial domain and temporal coverage

In this study, the SARAHE satellite dataset (Amillo et al., 2014; Huld et al., 2016) from CM SAF is used to do daily mapping prediction. The surface incoming shortwave radiation (SIS) from SARAHE is taken as the input data, which covers the time period 1999 to 2016, with a spatial resolution of 0.05x0.05 degrees. The data records cover monthly mean, daily mean and hourly instantaneous values, which are provided as netCDF (Network Common Data Format) files. The total area is METEOSAT East (70S-70N, 10W-130E). In this study, the selected domain is (21.55S-20.70S, 55.05E-56E) which covers Reunion island and a small part of the surrounding ocean.

- Solar radiation product: SIS

Surface solar radiation product SIS from CM SAF (SARAHE) is used as input data to do one-day ahead prediction with linear regression model. Ten years' product (2007-2016) covering Reunion Island and Indian Ocean (21.55S-20.70S, 55.05E-56E) are applied to four different

prediction models. The training dataset is from 2007 to 2011 and the test dataset is from 2012 to 2016.

- Missing data

In the SARA-E (1999-2016) satellite dataset, there are only 16 days have no records, and those days have then been regarded as missing data. When using these data, the missing data are taken off from the calculation.

2.1.2.2 Ground-based SPN1 pyranometer: LE2P database

- Historical database and time resolution

For the intermittency analysis, the data used is from the LE2P lab in the University of Reunion. Since 2010, LE²P has established the RCI_GS solar radiation network to provide global horizontal irradiance (GHI) and the diffuse horizontal irradiance (DHI) at 1-minute intervals using SPN1, 2-in-1 pyranometer from Delta-T devices. The RCI_GS network consists of 12 ground-based stations (with a reference unit dating back to 2008). Each station includes meteorological capabilities by WXT520 (Heintz et al., 2015).

- Located stations

Figure 2.1 provides a map of the network, the blue star indicates existing stations and the yellow star future stations. Figure 2.2 shows the SPN1 and the WXT520 separately and Figure 2.3 shows the typical station with SPN1 sunshine pyranometer.

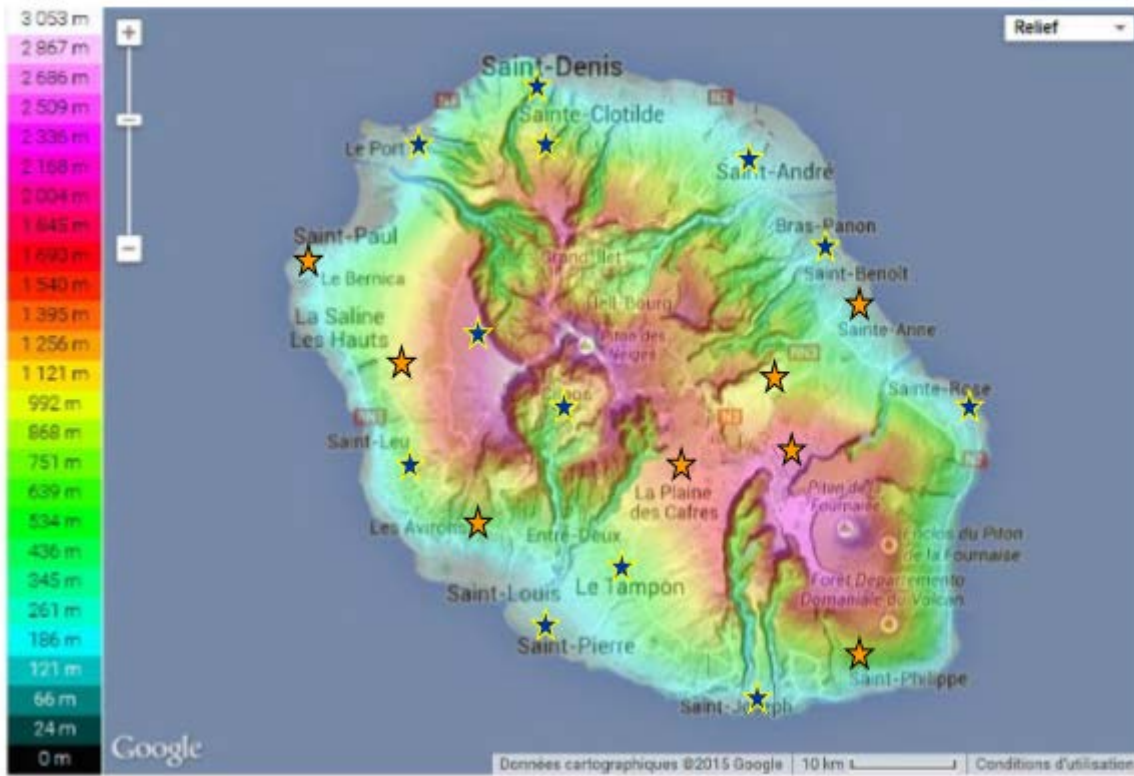


Figure 2.1 LE2P network of stations in Reunion Island. The blue star indicates existing stations and the yellow star future stations.

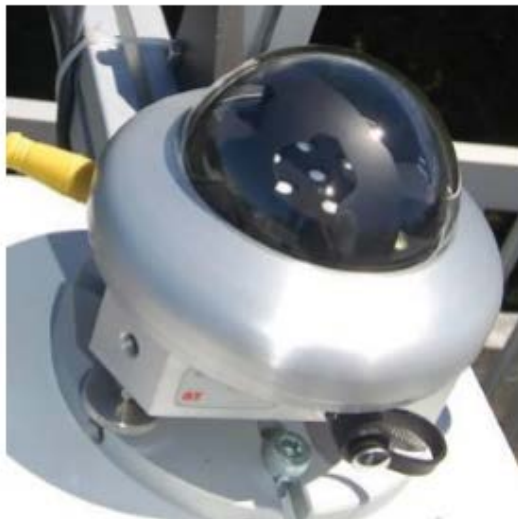


Figure 2.2 SPN1 (left) and the WXT520 (right), from lab LE2P.

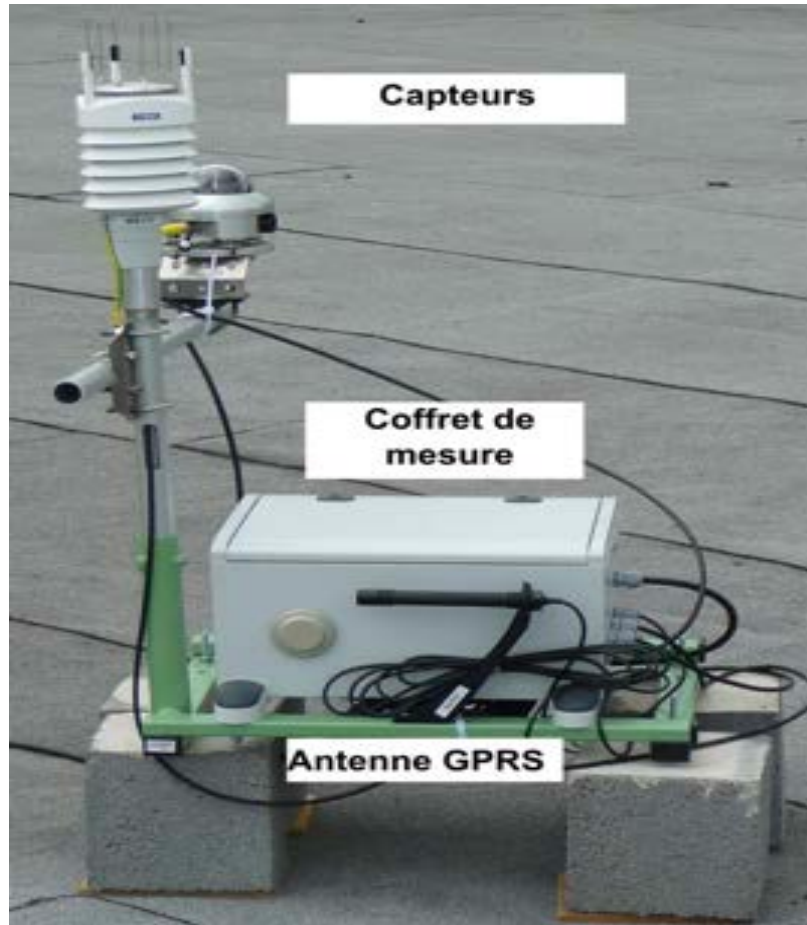


Figure 2.3 Typical station with a SPN1 sunshine pyranometer, from lab LE2P.

SPN1 is fitted with an array of seven miniature thermopile sensors and a specific computer-designed shading pattern which allows to measure the global and diffuse components of incident solar radiation at the same time. This pyranometer has a spectral response set between 400 - 2700 nm with a solar radiation measurement range of 0 to > 2000 W/m² and an overall accuracy of ± 10 W/m². It is a compact, light-weight, easy-to-use and easy-to-maintain instrument with no moving parts for tracking and no shade rings. Data from SPN1 may be collected through analogy voltage outputs or digital RS232 signal and the instrument is equipped with an internal heater. Its response time is 0.1s and zero off-set response < 3 W/m². The pyranometer's resolution is 0.6 W/m² with non-stability < 1%. The directional response of SPN1 is ± 20 W/m² and the spectral sensitivity is $\pm 10\%$ (0.4-2.7 μm). Its temperature response is $\pm 1\%$. The achievable uncertainty for SPN1 is 5% (Heintz et al., 2015).

Table 2.2 SPN1 parameters

| | |
|------------------------|-------------------------|
| Response Time | 0.1s |
| Zero off-set response | < 3W/m |
| Resolution | 0.6 W/m |
| Non-stability | < 1% |
| Directional response | ± 20 W/m |
| Spectral sensitivity | $\pm 10\%$ (0.4-2.7 mm) |
| Temperature response | $\pm 1\%$ |
| Achievable uncertainty | 5% |

WXT520 is also a compact and light-weight instrument that measures temperature, humidity, barometric pressure, precipitation, and wind speed and direction. The measurement range and accuracy are respectively $[-52 \text{ } +60] \pm 0.3^\circ\text{C}$ for temperature, $[0 \text{ } 100] \pm 3\%$ for humidity, $[600 \text{ } 1100] \pm 1$ hPa for pressure, cumulative accumulation with 0.01 mm for precipitation, $[0 \text{ } 60] \pm 0.3\text{m/s}$ for wind speed, $[0 \text{ } 360] \pm 3^\circ$ for wind direction.

Table 2.3 WXT520 parameters

| | |
|---------------------------|---------------------------------|
| Operating temperature | -52 ... + 60°C |
| Storage temperature | -60 ... + 70°C |
| Operation voltage | 5 ... 32 VDC |
| Typical power consumption | 3 mA at 12 VDC |
| Heating voltage | 5 ... 32 VDC. / 5 ... 30 VACRMS |

- Missing data

In this study, we have chosen to work with station data for 2014, because that year is common to all the stations except that located at Le Port. Table 2.4 lists the data information of the 12 stations in 2014, with the observations' days and missing days.

Table 2.4 Statistics of the station data in 2014

| | Stations | Period | Observation Days | Missing Days | Percentage of missing data |
|----|------------------------|------------------|-------------------------|---------------------|-----------------------------------|
| 1 | MOUFIA UR | 2008.12~2017.6 | 365 | 30 | 8.2% |
| 2 | MOUFIA BOIS DE NEFLES | 2012.10~present | 365 | 31 | 8.5% |
| 3 | LA POSSESSION | 2012.10~ present | 365 | 5 | 1.5% |
| 4 | SAINT LEU | 2012.10~ present | 365 | 9 | 2.5% |
| 5 | SAINT PIERRE LA VALLEE | 2012.10~ present | 365 | 7 | 1.9% |
| 6 | SAINT JOSEPH | 2013.9~2015.6 | 365 | 33 | 9.0% |
| 7 | SAINTE ROSE | 2013.9~2016.5 | 365 | 5 | 1.5% |
| 8 | SAINT ANDRE | 2012.10~ present | 365 | 32 | 8.8% |
| 9 | CILAOS PISCINE | 2013.6~2016.11 | 365 | 5 | 1.5% |
| 10 | BRAS PANON | 2010.11~2014.9 | 260 | 25 | 9.6% |
| 11 | TAMPON UR | 2012.12~2014.10 | 293 | 8 | 2.7% |
| 12 | Le Port | 2010.07~2012.12 | 0 | 0 | 0 |

When using these ground-based measurements to do intermittency analysis, the missing data and extreme (outliers) data as presented in Figure 2.4 for 2014 are not considered. The extreme and missing data are taken off from the whole data and the normal data for each station are taken to do calculation and analysis.

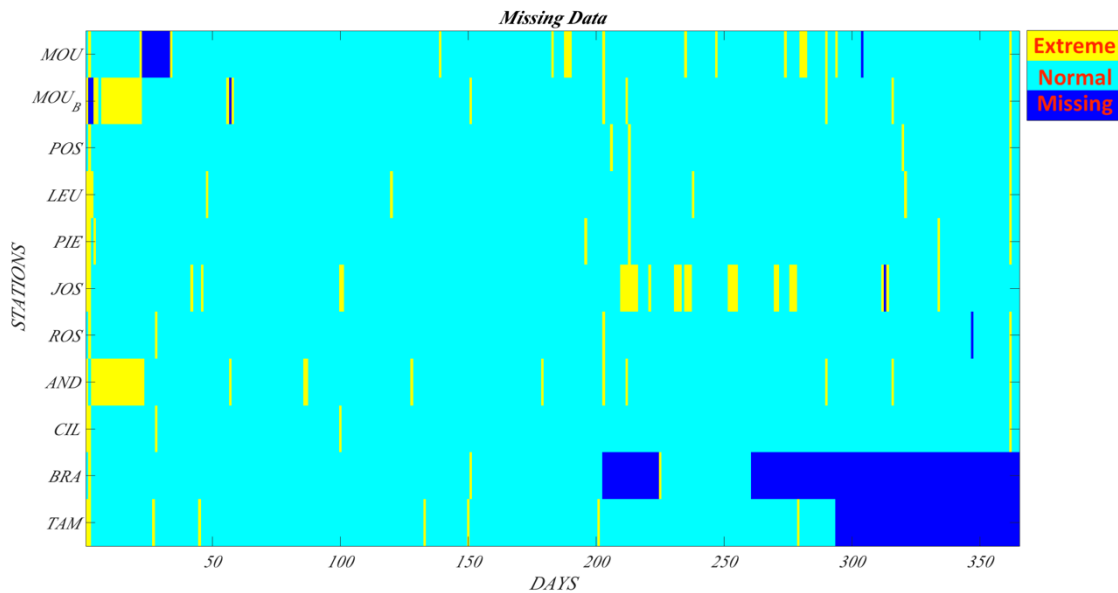


Figure 2.4 The missing data and extreme data

2.2 Methodology

2.2.1 Linear mapping regression model

2.2.1.1 Linear regression model

- Dependent/independent variables

There are many traditional models for forecasting: exponential smoothing, regression, time series, and composite model forecasts. Regression analysis is a statistical method to analyze quantitative data in order to estimate model parameters and make forecasts. In this thesis, the linear regression model is applied to predict the one-day ahead SSR.

The linear regression model is a statistical procedure that allows a researcher to study the linear relationship that relates two or more variables. This linear relationship summarizes the amount of change in one variable that is associated with changes in another variable or other variables. The prediction with the regression model provides an estimate of the future level of the independent variable. The prediction of the independent variable is important to an accurate forecast of the dependent variable. The independent variable may be regarded as causing

changes in the dependent variable, or the changes in the independent variable may occur prior to changes in the dependent variable. Normally, the independent variable is labelled the X variable, and the dependent variable as Y . The relationship between X and Y is shown on the Figure 2.5, with the independent variable X along the horizontal axis, and the dependent variable Y along the vertical axis. The objective of the regression model is to determine the straight line relationship that connects X and Y that best fits the data set. The straight line connecting any two variables X and Y is given by:

$$Y = a + bX, \quad (2.2)$$

where a is the Y intercept, and b is the slope of the line.

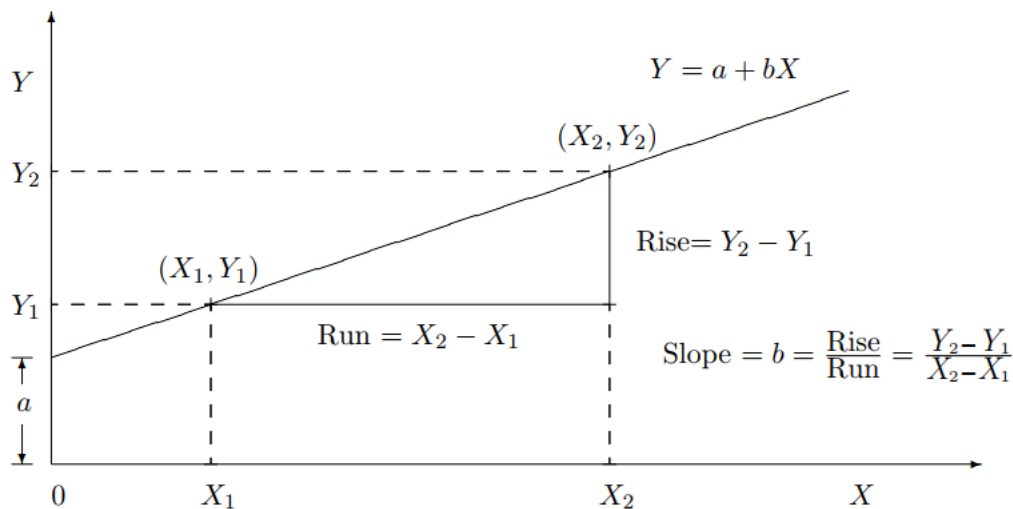


Figure 2.5 Diagrammatic representation of a straight line

- Linear fitting and parameters estimation

The regression model is to find a slope and intercept so that the straight line with that slope and intercept fits the points in the scatter diagram as closely as possible. The prediction procedure is to obtain an estimate of the line that best fits the points using the least square criterion. This criterion involves minimizing the sums of the squares of the errors of prediction. It is this minimization that produces the line that best fits the observed points.

When there are N paired data point (x_i, y_i) , a linear regression can be applied to approximate their relationship as follows:

$$\hat{Y} = a + bX. \quad (2.3)$$

The errors produced by this linear approximation can be estimated as:

$$Q = \sum_{i=1}^N \varepsilon_i^2 = \sum_{i=1}^N (\hat{y} - y_i)^2 . \quad (2.4)$$

The least square linear fit chooses coefficients a and b to produce a minimum value of the error Q . Thus, the following equations are solved:

$$\frac{\partial Q}{\partial a} = 0; \frac{\partial Q}{\partial b} = 0 . \quad (2.5)$$

By solving these equations, we can get the linear regression coefficients a and b . Then we get the linear regression line.

- Goodness of fit

In accuracy of the fit or the quality of the linear regression can be obtained by analyzing the variance. The root mean square error (RMSE), mean bias error (MAE), the mean square error (MSE) and R^2 value are four indices used to measure the quality of the fit. The MAE and the RMSE are given by:

$$MAE = \frac{1}{N} \sum_{i=1}^n |Xm_i - Xo_i| \quad (2.6)$$

and

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^n (Xm_i - Xo_i)^2} , \quad (2.7)$$

where N indicate the number of data, Xm and Xo are the modelled and observed daily respectively.

- Multiple Linear, Principal Component, Stepwise, Partial Least Square regression

In this study, the Multiple Linear Regression (MLR), Principal Component Regression (PCR), Stepwise Regression (SR), Partial Least Square Regression (PLSR) are used to perform prediction mapping. We described these briefly.

1) Multiple Linear Regression (MLR)

In the simple linear regression a single predictor variable X is used to model the response variable Y . However, in many applications, there is more than one factor that influences the response. MLR models thus describe how a single response variable Y depends linearly on a number of predictor variables. A MLR model with k predictor variables X_1, X_2, \dots, X_k and a response Y can be written as:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \epsilon, \quad (2.8)$$

where $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ are coefficients and ϵ is the residual term of the model.

The MLR could be thought as an extension of simple linear regression.

2) Principal Component Regression (PCR)

PCR is a technique for analyzing multiple regression data that suffer from multicollinearity. In PCR, a principal component analysis (PCA) is conducted on the design matrix and then the first k principal components are used to do the regression.

3) Stepwise Regression (SR)

Stepwise regression is an automated tool used in the exploratory stages of model building in order to identify a useful subset of predictors. In each step, a variable is considered for addition to or subtraction from the set of explanatory variables based on some prespecified criterion. The main approaches include: 1) forward selection, which involves starting with no variable in the model, testing the addition of each variable using a chosen model fit criterion, adding the variable whose inclusion gives the most statistically significant improvement of the fit and repeating this process until none improves the model to a statistically significant extent. 2) backward elimination, which involves starting with all candidate variables, testing the deletion of each variable using a chosen model fit criterion, deleting the variable whose loss gives the most statistically insignificant deterioration of the model fit, and repeating this process until no further variables can be deleted without a statistically significant loss of fit. 3) bidirectional elimination, a combination of the above, testing at each step for variables to be included or excluded.

4) Partial Least Square Regression (PLSR)

PLSR is a powerful and frequently applied technique in multivariate statistical process control when the process variables are highly correlated. It is a technique that reduces the predictors to a smaller set of uncorrelated components and performs least squares regression on these components, instead of on the original data. PLSR is especially useful when the predictors are highly collinear, or when there are more predictors than observations and ordinary least-squares regression either produces coefficients with high standard errors or fails completely. PLSR does not assume that the predictors are fixed, unlike MLR. This means that the predictors can be measured with error, making PLSR more robust to measurement uncertainty.

2.2.1.2 Dimensionality reduction: Principal Component Analysis

- Definition and properties: Short overview

Dimensionality reduction or dimension reduction is the process of reducing the number of random variables under consideration by obtaining a set of principal variables (Roweis et al.,

2000). Dimensionality reduction can also be seen as the process of deriving a set of degrees of freedom which can be used to reproduce most of the variability of a data set. Principal components analysis (PCA, Jolliffe, 1986) is a classical method that provides a sequence of best linear approximations to a given high-dimensional observation. It is one of the most popular techniques for dimensionality reduction. PCA is appropriate when you have obtained measures on a number of observed variables and wish to develop a smaller number of artificial variables (called principal components) that will account for most of the variance in the observed variables. The principal components may then be used as predictor or criterion variables in subsequent analyses.

PCA was invented in 1901 by Karl Pearson (1901) as an analogue of the principal axis theorem in mechanics. It was later independently developed and named by Harold Hotelling in the 1930s (Hotelling, 1933). Depending on the field of application, it is also named the discrete Karhunen–Loève transform (KLT) in signal processing, the Hotelling transform in multivariate quality control, proper orthogonal decomposition (POD) in mechanical engineering, singular value decomposition (SVD) of the matrix X (Golub and Van Loan, 1983), eigenvalue decomposition (EVD) of the covariance matrix XTX in linear algebra, factor analysis (for a discussion of the differences between PCA and factor analysis (Jolliffe, 1986), Eckart-Young theorem (Harman, 1960), or Schmidt-Mirsky theorem in psychometrics, empirical orthogonal functions (EOF) in meteorological science, empirical Eigen function decomposition (Sirovich, 1987), empirical component analysis (Lorenz, 1956), quasiharmonic modes (Brooks et al., 1988), spectral decomposition in noise and vibration and empirical modal analysis in structural dynamics.

- Principal component and Empirical Orthogonal Functions

EOF analysis is a decomposition of a signal or dataset in terms of orthogonal basis functions which are determined from the data. It is similar to performing a PCA on the data, except that the EOF method finds both time series and spatial patterns. Empirical Orthogonal Function technique aims at finding a new set of variables that capture most of the observed variance from the data through a linear combination of the original variables. EOF have been introduced in atmospheric science since the early 50's (Obukhov, 1947, 1960; Fukuoka 1951; Lorenz, 1956; Craddock, 1973). The EOF terminology is due to Lorenz (1956) who applied it in a forecasting project at the Massachusetts Institute of Technology. Since then, EOFs have become popular analysis tools in climate research. EOF is commonly used in climate research as a tool to analyze meteorological fields with high spatio-temporal dimensionality. The

leading EOF modes will typically describe large scale dynamical features in the field, and reconstruction of the field using a truncated subset of EOF can filter out small scale features or noise. Furthermore, EOF truncation may be useful for further statistical analysis by reducing the dimensionality of the data.

Here we present a brief description of EOF method, the covariance matrix is defined by Hannachi (2004):

$$\Sigma = \frac{1}{n-1} X'^T X' , \quad (2.9)$$

where X is data matrix which contains the covariance between any pair of grid points. The aim of EOF is to find the linear combination of all the variables, i.e. grid points, that explains maximum variance.

$$var(X'a) = \frac{1}{n-1} \|X'a\|^2 = \frac{1}{n-1} (X'a)^T (X'a) = a^T \Sigma a . \quad (2.10)$$

To make the problem bounded we normally require the vector a to be unitary. Hence the problem readily yields:

$$\max_a a^T \Sigma a , \quad s. t. a^T a = 1. \quad (2.11)$$

The solution is a simple eigenvalue problem (EVP):

$$\Sigma a = \lambda a. \quad (2.12)$$

By definition the covariance matrix Σ is symmetrical and therefore diagonalizable. The k^{th} EOF is simply the k^{th} eigenvector a_k of Σ , and the corresponding eigenvectors, have been sorted in decreasing order. The covariance matrix is also positive semi definite, hence all its eigenvalues are positive. The eigenvalue λ_k corresponding to the k^{th} EOF gives a measure of the explained variance by a_k , $k = 1, \dots p$. It is common to write the explained variance in percentage as

$$\frac{100\lambda_k}{\sum_{k=1}^p \lambda_k} \% . \quad (2.13)$$

The projection of the anomaly field X' onto the k^{th} EOF a_k , i.e. $c_k = X'a_k$ is the k^{th} principal component (PC).

$$C_k(t) = \sum_{s=1}^p x'(t, s) a_k(s). \quad (2.14)$$

- Reduction criteria

Normally we take all the PCs together which could give 95% explained variance in the EOF (Hannachi, 2004).

2.2.2 Pre-processing and decomposition method

2.2.2.1 Stationary/nonstationary time series

- Definition

As stated in the introduction, the data need to be pre-processed. So firstly it is necessary to do time series analysis. Time series analysis is about the study of data collected through time. Generally, a stationary process is one whose statistical properties do not change over time. A strictly stationary stochastic process is one where given t_1, \dots, t_l , the joint statistical distribution of X_{t_1}, \dots, X_{t_l} is the same as the joint statistical distribution of $X_{t_1+\tau}, \dots, X_{t_l+\tau}$ for all l and τ . This means that all moments of all degrees (expectations, variances, third order and higher) of the process are the same. It also means that the joint distribution of (X_t, X_s) is the same as $(X_{t+\tau}, X_{s+\tau})$ and hence cannot depend on s or t but only on the distance between s and t . Since this definition of strict stationarity is too strict, usually the weaker definition of stationarity is used, which means that the mean and the variance of a stochastic process do not depend on t and the autocovariance between X_t and $X_{t+\tau}$ only can depend on the lag τ (τ is an integer, the quantities also need to be finite). A process is only stationary or non-stationary. If the mean or the variance of the time series changes over time, it is non-stationary.

- Statistical properties

In statistics, the probability models in the part of time series analysis is the familiarity of the concepts like mean and covariance and the whole stochastic process paraphernalia. However, there is another way to study time series, which is the spectrum. The frequency or spectral approach determines how much energy is contained within a time series as a function of frequency (the spectrum).

- Spectral analysis: Fourier and Hilbert transforms

In the spectral analysis domain, the Fourier transform (FT) is the dominating tool which is better for stationary or linear data. The Hilbert transform is commonly used to study instantaneous frequency for non-stationary data.

----- Fourier Transform (FT)

The FT decomposes a signal to complex exponential functions of different frequencies, defining by the following two equations:

$$X(f) = \int_{-\infty}^{\infty} x(t) \cdot e^{-2j\pi f t} dt \dots \dots \quad (2.15)$$

$$x(t) = \int_{-\infty}^{\infty} X(f) \cdot e^{2j\pi ft} df \dots \dots \quad (2.16)$$

where t is the time, f is the frequency, and x denotes the signal in time domain, X denotes the signal in frequency domain. Equation (2.15) is called Fourier transform of $x(t)$, and equation (2.16) is the inverse Fourier transform of $X(f)$.

----- Hilbert Transform

The Hilbert transform of the function signal $g(t)$ is defined as:

$$H_g(t) = \frac{1}{\pi} (VP) \int_{-\pi}^{\pi} \frac{g(\tau)}{t-\tau} d\tau . \quad (2.17)$$

Because of the possible singularity at $\tau = t$, the integral is to be considered as a Cauchy principal value (VP). $H_g(t)$ is a convolution, $\frac{1}{\pi t} * g(t)$.

The spectrum of $H_g(t)$ is related to that of $g(t)$. From the convolution equation,

$$\mathcal{F}(H(t)) = \mathcal{F}\left(\frac{1}{\pi t}\right) \mathcal{F}(g(t)) , \quad (2.18)$$

where F is the FT. With a real signal $g(t)$ one can associate a complex function with the real part equal to $g(t)$ and the imaginary part equal to $H(g(t))$, $h(t) = g(t) - iH(g(t))$. In statistical signal analysis, this complex function $h(t)$ is known as the analytic signal. Analytic signals are important since they possess unique phase $\phi(t)$ which leads to the definition of the instantaneous frequency. If $h(t)$ is represented as $a(t) \cdot \exp\{i\phi(t)\}$, then the quantity $\frac{d\phi}{dt}$ is instantaneous frequency of the signal $g(t)$ at time t (Flandrin 1992,1999).

2.2.2.2 Time-frequency analysis

- Definition

The traditional data analysis methods, such as the Fourier transform, use pre-determined basis, which provide an effective tool to process linear and stationary data. However, there are still some limitations in applying these methods to analyze nonlinear and nonstationary data. Time-frequency analysis has been developed to overcome the limitations of the traditional techniques by representing a signal with a joint function of both time and frequency. Time-frequency analysis comprises those techniques that study a signal in both the time and frequency domains simultaneously.

Wavelet analysis opened a new path for time-frequency analysis. A significant breakthrough of wavelet analysis is the use of multi-scales to characterize signals. Another important approach in the time-frequency analysis which made more substantial progress is Empirical Mode Decomposition (EMD). The EMD method provides an effective tool to

decompose a signal into a collection of Intrinsic Mode Functions (IMF) that allow well-behaved Hilbert transforms for computation of physically meaningful time-frequency representation.

- First example: Wavelet Decomposition

In recent decades, wavelet transforms have become a widely used statistical methods (Daubechies, 1990; Grinsted et al., 2004; Du et al., 2006). The wavelet transform is similar to the FT (or much more to the windowed FT) with a completely different merit function. The main difference is that FT decomposes the signal into sines and cosines; in contrast the wavelet transform uses functions that are localized in both the real space and Fourier space. Generally, the wavelet transform can be expressed by the following equation:

$$F(a, b) = \int_{-\infty}^{\infty} f(x)\psi_{(a,b)}^*(x) dx , \quad (2.19)$$

where the * denotes complex conjugate symbol and ψ is some function.

As it is seen, the Wavelet transform is in fact an infinite set of various transforms, depending on the merit function used for its computation. That is why wavelet transform can be applied in different types. Normally wavelet transform is divided based on the wavelet orthogonality. The orthogonal wavelets is used for discrete wavelet transform development and non-orthogonal wavelets *for continuous* wavelet transform development. The discrete wavelet transform (DWT) returns a data vector of the same length as the input is. The continuous wavelet transform (CWT) in contrary returns an array one dimension larger than the input data.

DWT is an implementation of the wavelet transform using a discrete set of the wavelet scales and translations obeying some defined rules (<http://klapetek.cz/wdwt.html>). In other words, this transform decomposes the signal into mutually orthogonal set of wavelets. This is the main difference with CWT, or its implementation for the discrete time series sometimes called discrete-time continuous wavelet transform (DT-CWT). CWT is an implementation of the wavelet transform using arbitrary scales and almost arbitrary wavelets. The wavelets used are not orthogonal and the data obtained by this transform are highly correlated. For the discrete time series we can use this transform as well, with the limitation that the smallest wavelet translations must be equal to the data sampling. This is sometimes called Discrete Time Continuous Wavelet Transform (DT-CWT) and it is the most used way of computing CWT in real applications. In principle the CWT works by using directly the definition of the wavelet transform, i.e. we are computing a convolution of the signal with the scaled wavelet. For each scale we obtain by this way an array of the same length N as the signal has. By

using M arbitrarily chosen scales we obtain a field $N \times M$ that represents the time-frequency plane directly. The algorithm used for this computation can be based on a direct convolution or on a convolution by means of multiplication in Fourier space (this is sometimes called Fast Wavelet Transform). The choice of the wavelet that is used for time-frequency decomposition is the most important thing. By this choice we can influence the time and frequency resolution of the result. We cannot change the main features of WT by this way (low frequencies have good frequency and bad time resolution; high frequencies have good time and bad frequency resolution), but we can somehow increase the total frequency of total time resolution. This is directly proportional to the width of the used wavelet in real and Fourier space. If we use the Morlet wavelet for example (real part – damped cosine function) we can expect high frequency resolution as such a wavelet is very well localized in frequencies. In contrary, using Derivative of Gaussian (DOG) wavelet will result in good time localization, but poor one in frequencies.

- Second example: Empirical Mode Decomposition (EMD)

The Empirical Mode Decomposition (EMD) is based on the point that most of the signal is multi-component, which means that there exist different scales simultaneously (Cohen, 1995; Huang et al., 1998, 1999). The EMD is an iterative process, which decomposes real signals f into simpler signals (modes),

$$f(t) = \sum_{j=1}^M \psi_j(t). \quad (2.20)$$

Each “monocomponent” signal ψ_j should be representable in the form

$$\psi(t) = r(t) \cos \theta(t), \quad (2.21)$$

where the amplitude and phase are both physically and mathematically meaningful. Once a suitable polar parametrization is determined, it is possible to analyze the function f by processing these individual components.

In general, EMD may be applied either to sampled data or to functions of real variables by first identifying the appropriate time scales that reveals the physical characteristics of the studied system, decompose the function into modes ψ intrinsic to the function at the determined scales, and then apply the Hilbert transform to each of the intrinsic components.

According to Huang and collaborators (Robert et al., 2006), the EMD method is motivated “from the simple assumption that any data consists of different simple intrinsic mode oscillations.” There are three assumptions for EMD method: (1) the signal has at least two extrema: one maximum and one minimum; (2) the characteristic time scale is defined by the time lapse between the extrema; and (3) if the data were totally devoid of extrema but contained

only inflection points, then it can be differentiated once or more times to reveal the extrema. Final results can be obtained by integration(s) of the components.

Norden Huang et al. (1998, 1999) introduced that a function ψ of a real variable t is defined to be an IMF, if it satisfies two characteristic properties: (i) the difference between the number of local extrema and the number of zero-crossings must be zero or one; (ii) the running mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero. The EMD method is used to identify the intrinsic oscillatory modes by their characteristic time scales in the data empirically, and then decompose the data accordingly. Norden Huang et al. (1998, 1999) introduced the EMD algorithm, termed as “sifting process”.

The first step of sifting process is to define all the local extrema maxima and minima points by using the envelopes. The envelopes are constructed by a cubic spline. Once all the local extrema maxima points are identified, a cubic spline connects all the points and the upper envelope $u(t)$ is then constructed. Repeat the process to produce the lower envelope $l(t)$ for the local minima. Figure (2.6) and (2.7) show extrema maxima, extrema minima and upper and lower envelopes. All the data $X(t)$ are covered by the upper and lower envelopes, and $X(t)$ are between the envelopes. The mean of these two envelopes is defined as

$$m_1(t) = \frac{u(t)+l(t)}{2}. \quad (2.22)$$

The first component is then estimated by

$$h_1(t) = X(t) - m_1(t). \quad (2.23)$$

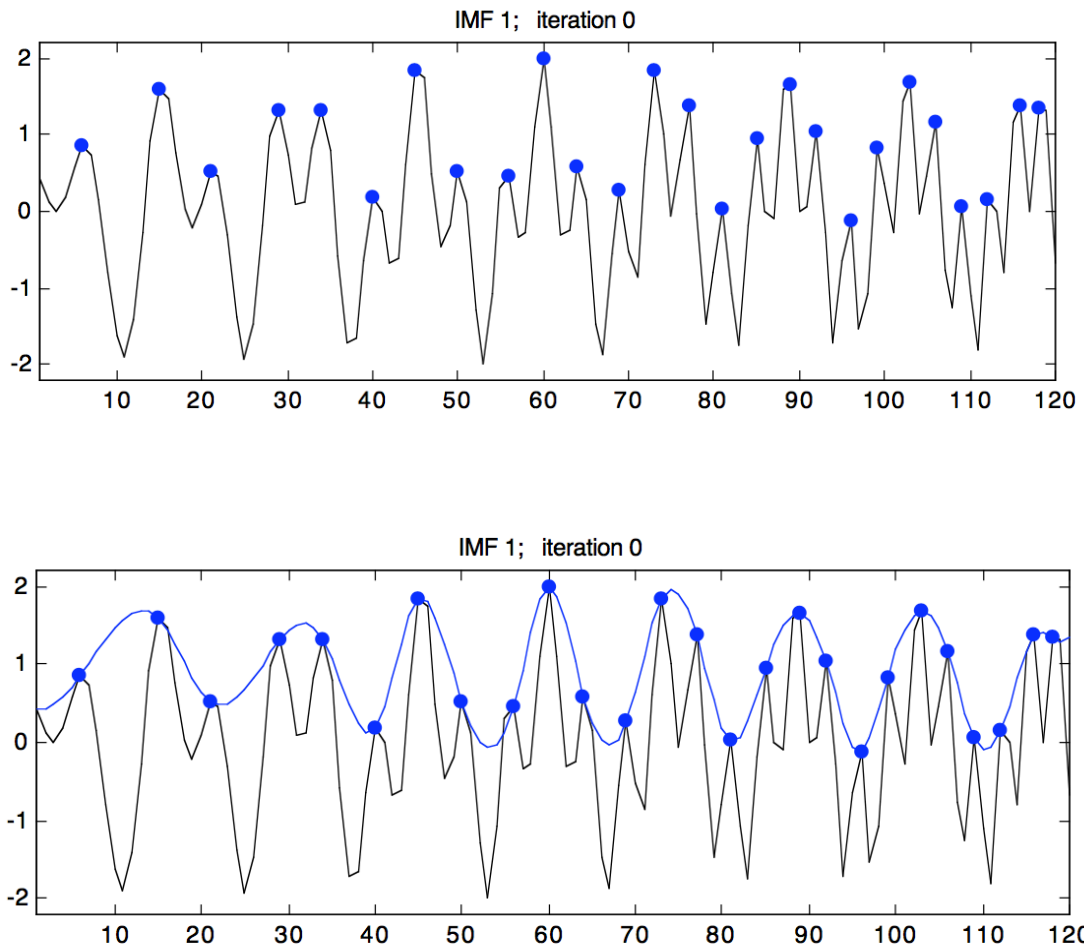


Figure 2.6 How to find the extrema maxima and upper envelope

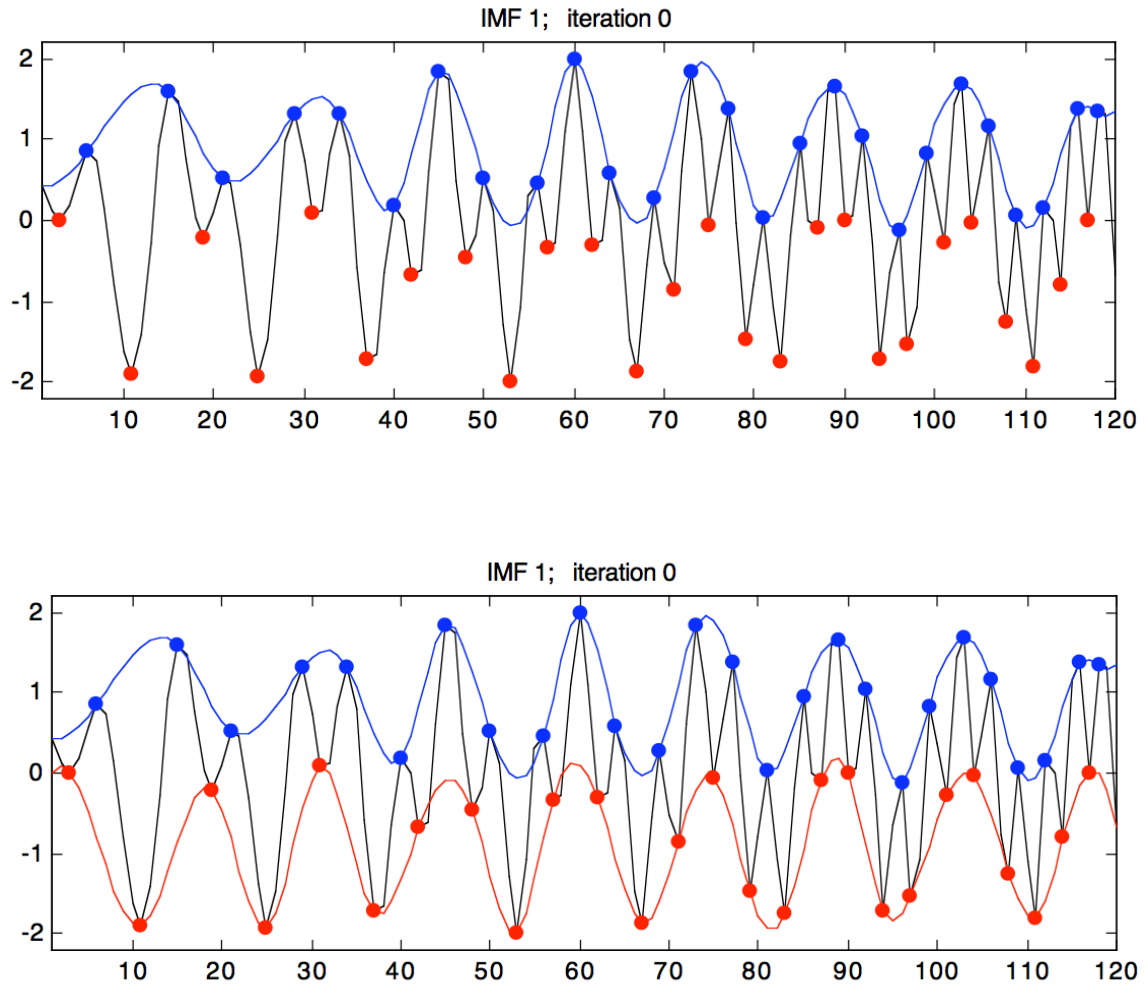


Figure 2.7 How to find the extrema minima and lower envelope

Ideally, $h_1(t)$ should be an IMF, but overshoots and undershoots can also generate new extrema, and shift or exaggerate the existing ones. $h_1(t)$ described above doesn't seem to satisfy all the requirements of an IMF. An illustration of the first sifting process for a real time series is shown in Figure 2.8. It identifies the original time series $X(t)$ is shown as thin solid black line; all the local extrema points, and construct the upper envelop $e_{max}(t)$, and the lower envelop $e_{min}(t)$ in blue and red line; calculate the running average $m_1(t)$ in pink line, and get the local detail $h_1(t)$ after 1st sifting. In the second sifting process, $h_1(t)$ is treated as new time series, then

$$h_{11}(t) = h_1(t) - m_{11}(t). \quad (2.24)$$

$$h_{12}(t) = h_{11}(t) - m_{12}(t). \quad (2.25)$$

Figure 2.8 Illustration of the sifting processes of EMD algorithm: the original data, in thin black line, with the upper and lower envelopes in blue and red lines, and the local mean in pink line.

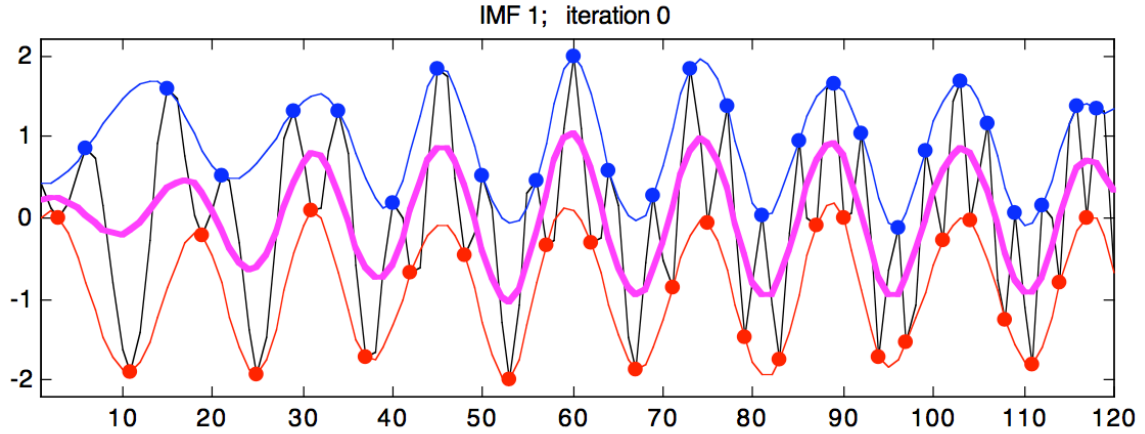


Figure 2.8 Illustration of the sifting processes of EMD algorithm

After the second sifting, if there are still local maxima below the zero line, repeat the sifting process j times, until all the local maxima are positive, all the local minima are negative, many waves are still symmetric, and until $h_{1k}(t)$ is an IMF.

$$h_{1k}(t) = h_{1(k-1)}(t) - m_{1k}(t). \quad (2.26)$$

The first IMF component from the data is designated as

$$C_1(t) = h_{1k}(t). \quad (2.27)$$

$C_1(t)$ should contain the finest scale or the shortest period component of the data. We can separate C_1 from the rest of the data, then, the residual $r_1(t)$

$$r_1(t) = X(t) - C_1(t). \quad (2.28)$$

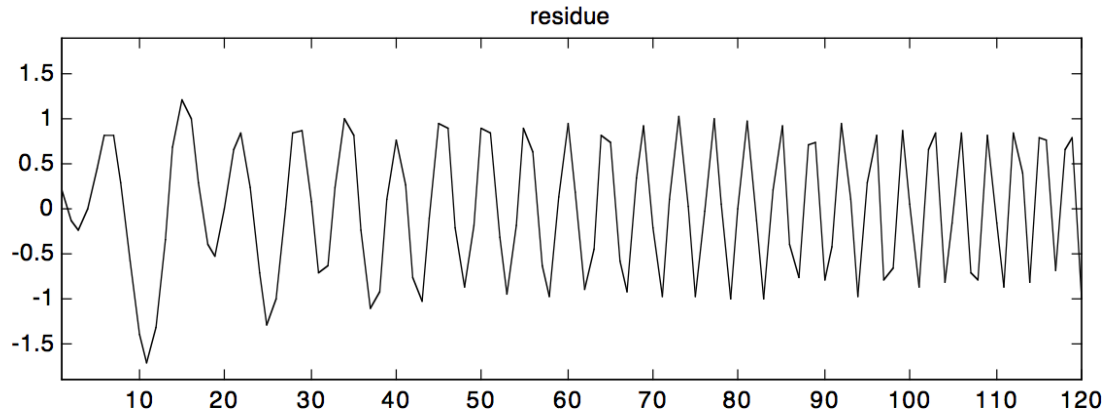
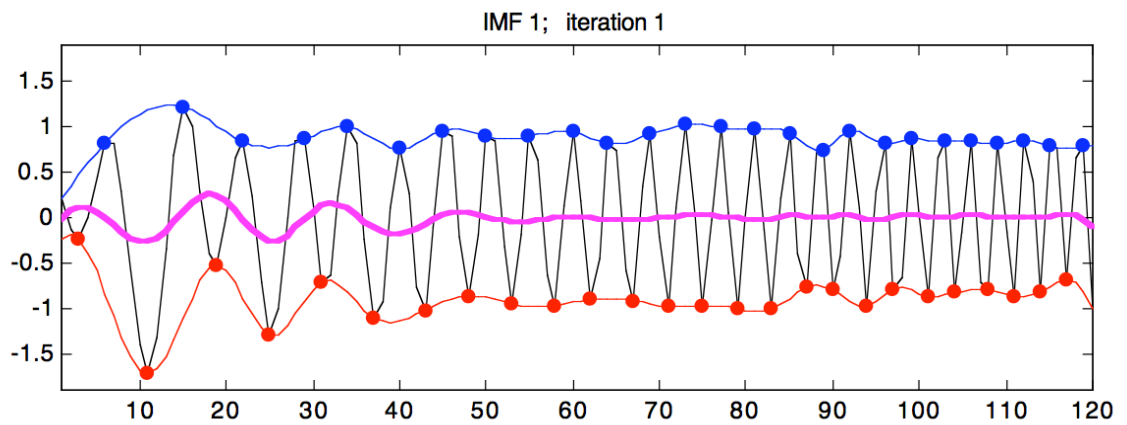
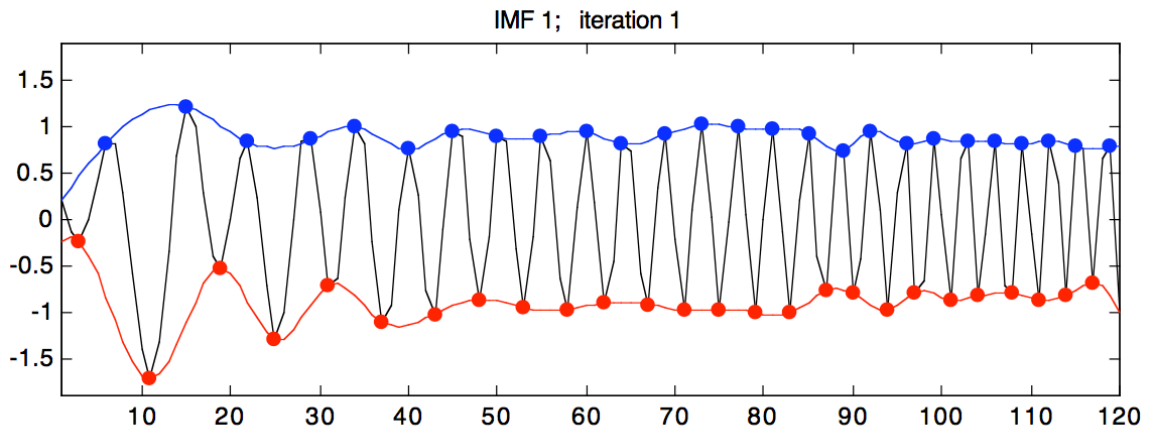
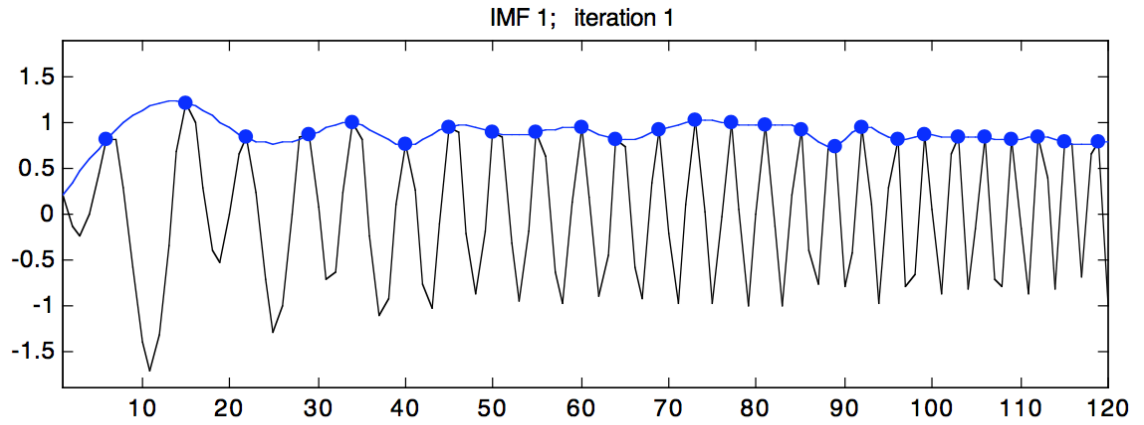
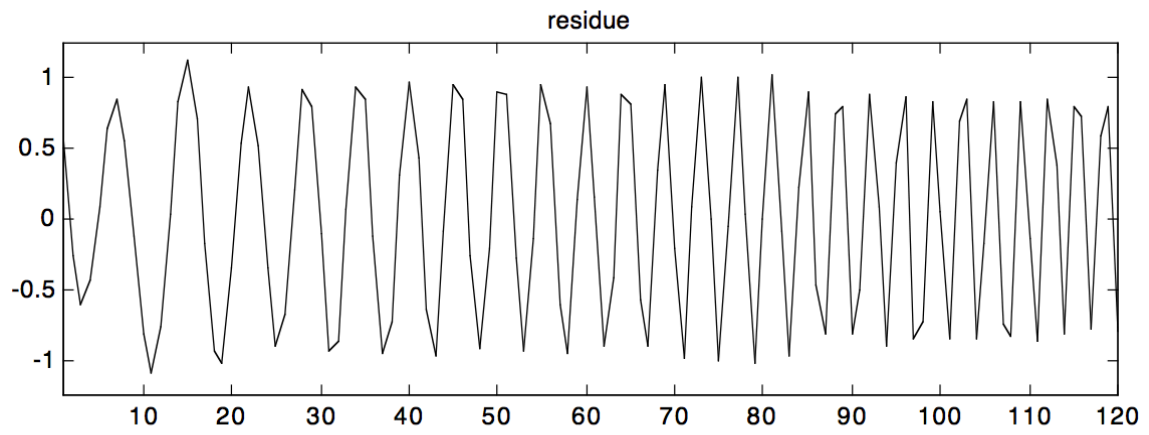
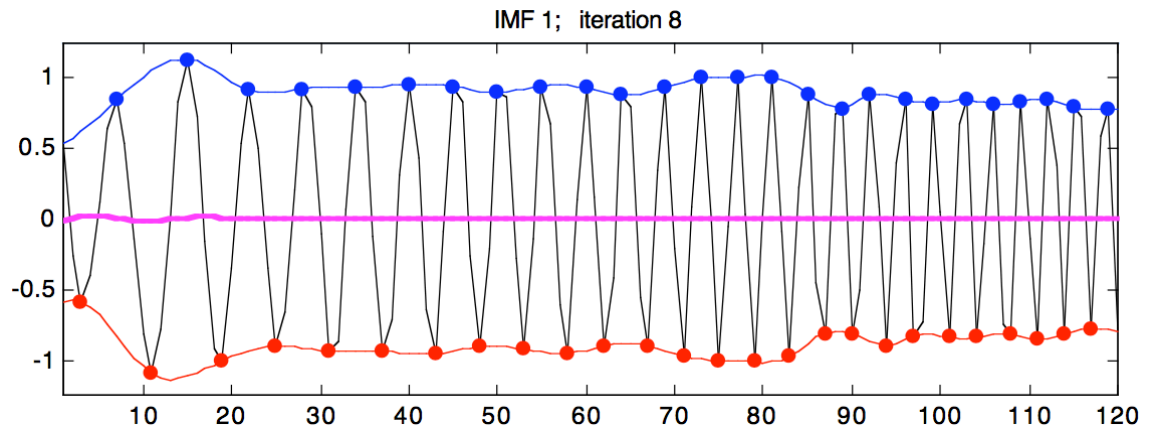
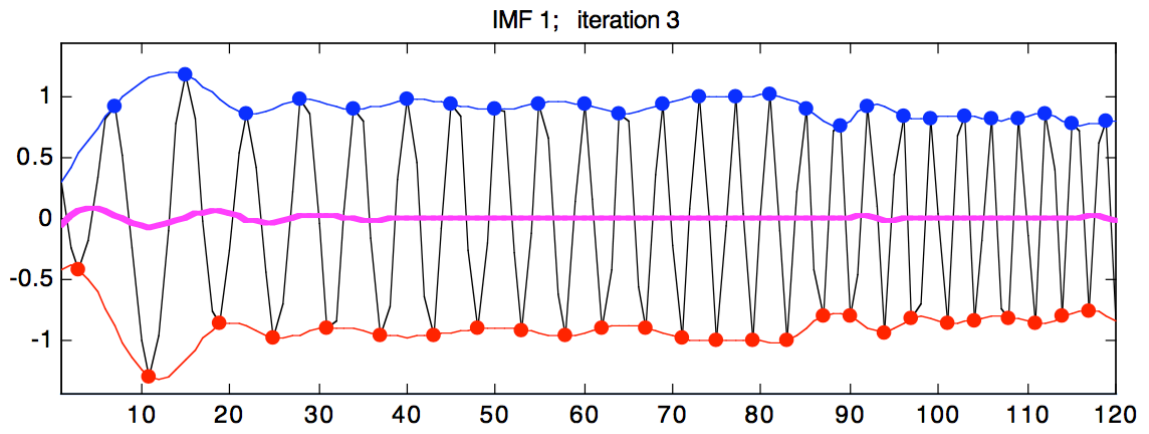


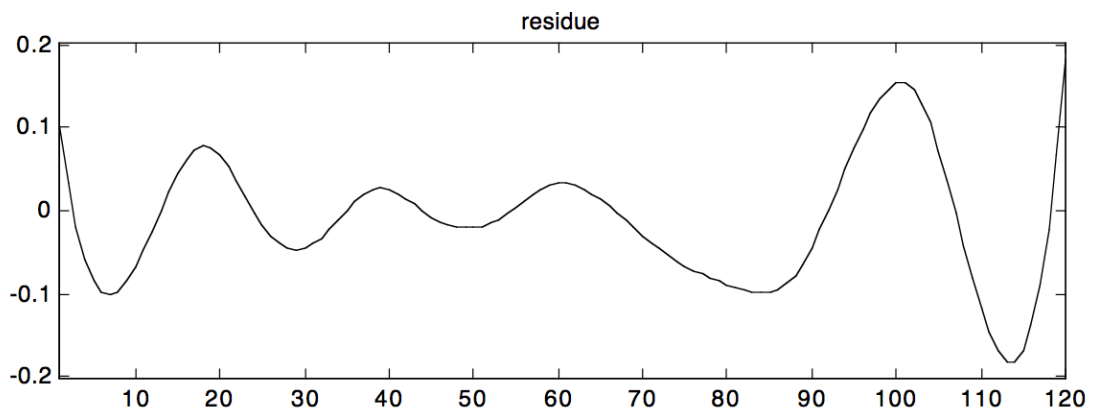
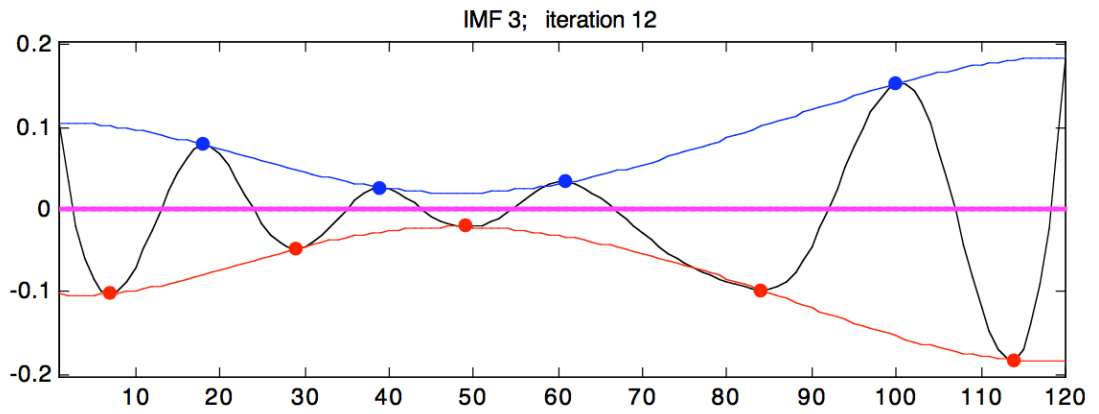
Figure 2.9 Illustration of the residue r_1

This residue $r_1(t)$, still contains information of longer period components, Iterate on h_1 if it violates the assumptions and restrictions. We repeat this procedure n times until $r_n(t)$ becomes a monotonic function or at most has one local extreme point. The result is THEN

$$r_2 = r_1 - C_2, \dots, r_n = r_{n-1} - C_n. \quad (2.29)$$







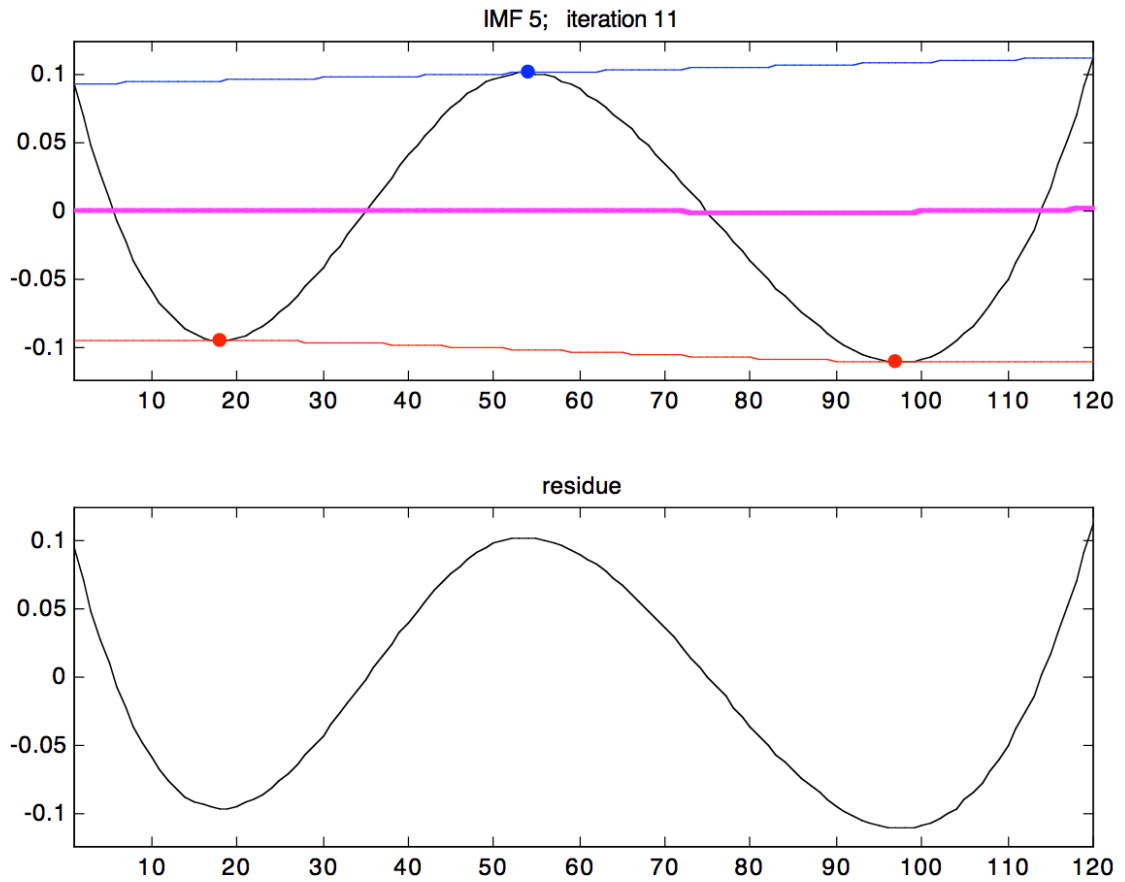


Figure 2.10 Some results of sifting process and some residues in different iteration.

The sifting process is completed when the residual r_n is so small that it is less than the predetermined value of substantial consequence, or when the residual r_n becomes a monotonic function from which no IMF can be extracted. The corresponding flow chart of this sifting process is shown in Figure 2.11.

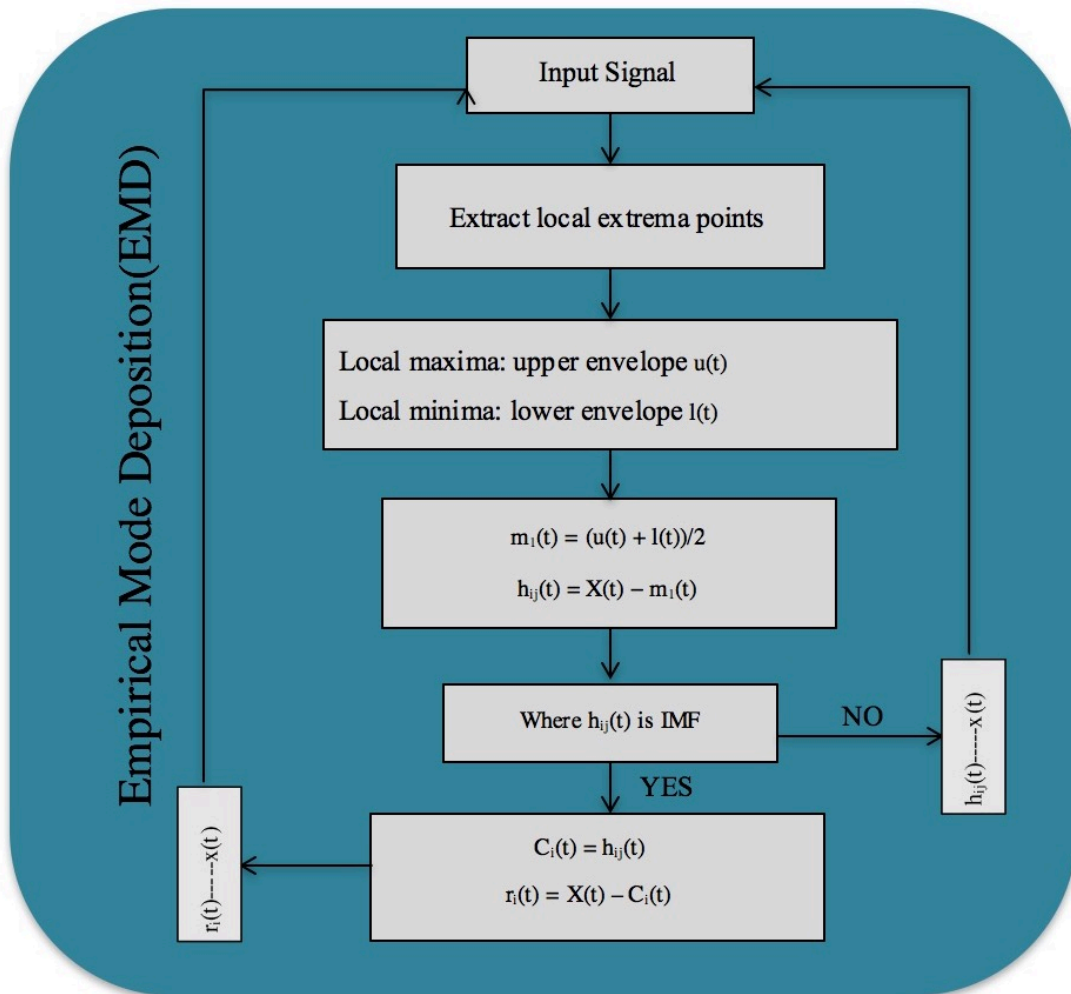


Figure 2.11 The flowchart describing sifting process for EMD algorithm.

Finally, we can get n IMFs with residual $r_n(t)$. IMF is the residual of mean-spline and input signal at each stage. An IMF represents a simple oscillatory mode as a counterpart to the simple harmonic function. Instead of constant amplitude and frequency, as in a simple harmonic component, the IMF can have a variable amplitude and frequency as functions of time. Figure 2.12 shows the calculation of IMFs. Note that IMFs are recorded at the bottom of the table.

| | IMF1 | IMF2 | IMF3 | IMF n |
|-----|--------------------------------|--------------------------------|--------------------------------|--------------------------------|
| | $X(t)$ | $X(t) - c_1 = r_1$ | $r_1 - c_2 = r_2$ | $r_{n-2} - c_{n-1} = r_{n-1}$ |
| 0 | $X(t) - m_1 = h_1$ | $r_1 - m_2 = h_2$ | $r_2 - m_3 = h_3$ | $r_{n-1} - m_n = h_n$ |
| 1 | $h_1 - m_{11} = h_{11}$ | $h_2 - m_{21} = h_{21}$ | $h_3 - m_{31} = h_{31}$ | $h_n - m_{n1} = h_{n1}$ |
| 2 | $h_{11} - m_{12} = h_{12}$ | $h_{21} - m_{22} = h_{22}$ | $h_{31} - m_{32} = h_{32}$ | $h_{n1} - m_{n2} = h_{n2}$ |
| 3 | $h_{12} - m_{13} = h_{13}$ | $h_{22} - m_{23} = h_{23}$ | $h_{32} - m_{33} = h_{33}$ | $h_{n2} - m_{n3} = h_{n3}$ |
| .. | | | | |
| k | $h_{1(k-1)} - m_{1k} = h_{1k}$ | $h_{2(k-1)} - m_{2k} = h_{2k}$ | $h_{3(k-1)} - m_{3k} = h_{3k}$ | $h_{n(k-1)} - m_{nk} = h_{nk}$ |
| IMF | $h_{1k} = c_1$ | $h_{2k} = c_2$ | $h_{3k} = c_3$ | $h_{nk} = c_n$ |

Figure 2.12 The calculation of IMFs

To guarantee that the IMF modes retain enough physical sense for both amplitude and frequency modulations, the sifting process can be stopped by any of the following predetermined criteria: either when the component c_n , or the residue r_n , becomes so small that it is less than the predetermined value of substantial consequence, or when the residue r_n , becomes a monotonic function from which no more IMF can be extracted. A stopping criterion has to be introduced to stop the sifting process. Different stoppage criteria can be used in the sifting process such as Cauchy SD convergence, S-method, etc. Their detailed description can be found elsewhere (Huang et al, 2003; Huang and Attoh-Okine, 2005; Huang and Shen, 2005; Huang et al., 1998, 1999; Rilling et al., 2003). In this study, the job of the stopping criteria is accomplished by limiting the size of the standard deviation (SD). SD is computed from the two consecutive successive sifting processes as

$$SD = \sum_0^r \left[\frac{|d_{i-1}(t) - d_i(t)|^2}{d_{i-1}^2(t)} \right] \tag{2.30}$$

A typical value proposed by (Huang et al., 1998) based on their experience is 0.2~0.3, which two Fourier spectra computed by shifting only five out of 1024 points from the same data, can have an equivalent SD of 0.2~0.3 calculated point-by-point. Therefore, a SD value of 0.2~0.3 for the sifting procedure is a very rigorous limitation for the difference between siftings. If a calculated SD is smaller than a given value 0.2~0.3, then the sifting stops and gives an IMF.

Another widely used criterion is based on 3 thresholds α , θ_1 and θ_2 , which are designed to guarantee globally small fluctuations in the mean while taking into account locally of large excursions (Rilling et al., 2003). In this study, the first criterion is conducted on the IMF.

Different versions of (Hilbert Huang transform) HHT algorithms have been implemented using various interpolation algorithms in the sifting process to evaluate the envelopes, such as Cubic spline interpolation (Huang et al., 1998; Rilling et al., 2003), Akima cubic spline interpolation (Radić, Pasarić and Šinik, 2004), etc. Various boundary conditions may also be considered during interpolation such as tied to zero, mirroring nearest extrema boundary conditions (Rilling et al., 2003), tied to signal, etc. In order to reduce the end point effects due to miscalculated boundary conditions during interpolation, several envelope estimation algorithms are used in practice including those based on neural networks (Pradhan and Lee, 2010a,b,c)), etc. Improved algorithms such as Bi-variate EMD (Rilling et al., 2007), Complex EMD (Tanaka and Mandic, 2007) and Multivariate EMD (Rehman and Mandic, 2010) for multi-channel and multi-dimensional signal analysis, Ensemble EMD for a higher accuracy in noisy data (Wu and Huang, 2006), are also available.

EMD algorithm described above can be considered to be well adapted for nonlinear and nonstationary data. More detail about the EMD algorithm can be found in (Huang et al., 1998, 1999; Rilling et al., 2003; Flandrin et al., 2004; Flandrin and Gonçalvès, 2004; Huang, 2005).

2.2.3 Fractal and Multifractal

2.2.3.1 Fractal

- Brief overview of fractal application domain

Time series generated by complex systems exhibit fluctuations on a wide range of time scales and/or broad distributions of the values. In both equilibrium and non-equilibrium situations, the natural fluctuations are often found to follow a scaling relation over several orders of magnitude. Such scaling laws allow for a characterization of the data and the generating complex system by fractal or multifractal scaling exponents, which can serve as characteristic fingerprints of the systems in comparisons with other systems and with models. Fractal scaling behavior has been observed in many time series from experimental physics, geophysics, medicine, physiology, and even social sciences.

A fractal is a rough or fragmented geometrical object that can be subdivided in parts, each of which is a reduced-size copy of the whole. Fractals are generally self-similar and independent of scale. Although the underlying causes of the observed fractal scaling are often

not known in detail, the fractal or multifractal characterization can be used for generating surrogate (test) data, modelling the time series, and deriving predictions regarding extreme events or future behavior (Kantelhardt, 2008). The main application, however, is still the characterization of different states or phases of the complex system based on the observed scaling behavior (Kantelhardt, 2008). The first scientist who applied fractal analysis to natural time series is (Mandelbrot, 1982) who included early approaches by Hurst regarding hydrological systems. In the last decade, fractal and multifractal scaling behavior has been reported in many natural time series generated by complex systems, including geophysics time series (recordings of temperature, precipitation, ozone levels, wind speed, seismic events, vegetational patterns, and climate dynamics), astrophysical time series (X-ray light sources and sunspot numbers), medical and physiological time series (recordings of heartbeat, respiration, blood pressure, blood flow, nerve spike intervals, human gait, glucose levels, and gene expression data), technical time series (internet traffic, highway traffic, and neutronic power from a reactor), social time series (finance and economy, language characteristics, fatalities in conflicts), and so on (Kantelhardt, 2008).

- Fractal and dimension

If one finds that a complex system is characterized by fractal or multifractal dynamics with particular scaling exponents, this will help in obtaining predictions on the future behaviour of the system and on its reaction to external perturbations or changes in the boundary condition. In addition, one could test and iteratively improve models of the system until they reproduce the observed scaling behaviour.

In a strict sense, most time series are one dimensional, since the values of the considered observable are measured in homogeneous time intervals. Hence, unless there are missing values, the fractal dimension of the support is $D(0) = 1$. However, there are rare cases where most of the values of a time series are very small or even zero, causing a dimension $D(0) < 1$ of the support. In these cases, one has to be very careful in selecting appropriate analysis techniques. Even if the fractal dimension of support is one, the information dimension $D(1)$ and the correlation dimension $D(2)$ can be studied. $D(2)$ is in fact explicitly related to all exponents studied in monofractal time series analysis.

- Hurst exponent: predictability of time series

When the time axis and the axis of the measured values $x(t)$ are not equivalent, a rescaling of time t by a factor a may require rescaling of the series values $x(t)$ by a different factor a^H in order to obtain a statistically similar picture. In this case the scaling relation

$$x(t) \rightarrow a^H x(at) . \quad (2.31)$$

Holds of an arbitrary factor a , describing the data as self-affine. The *Hurst* exponent H characterizes the type of self-affinity. For example, Brownian motion is characterized by $H = 0.5$. For a monofractal process, $D = 2 - H$.

The Hurst exponent is a useful statistical method for inferring the properties of a time series without making assumptions about stationarity. The Hurst exponent measures a relative tendency of a time series either regress strongly to the mean or to cluster in a direction. It is related to fractal dimension which gives measure for roughness of the time series. The Hurst exponent provides a measure for long-term memory of time series and very useful in forecasting (Stan et al., 2013).

2.2.3.2 Multifractal

- Generalized Hurst exponent

Many records do not exhibit a simple monofractal scaling behaviour, which can be accounted for by a single scaling exponent. There might exist crossover time scales separating regimes with different scaling exponents. In other cases, the scaling behaviour is more complicated, and different scaling exponents are required for different parts of the series. In even more complicated cases, such different scaling behaviour can be observed for many interwoven fractal subsets of the time series. In this case a multitude of scaling exponents is required for a full description of the scaling behaviour in the same range of time scales, and a multifractal analysis must be applied. A multifractal is a set of intertwined fractals. Self-similarity of multifractals is scale dependent. A deviation from a strict self-similarity is also called intermittency.

Two general types of multifractality in time series can be distinguished: (1) Multifractality due to a broad probability distribution (density function) for the values of the time series, e. g. a Levy distribution. In this case the multifractality cannot be removed by shuffling the series. (2) Multifractality due to different long-term correlations of the small and large fluctuations. In this case the probability density function of the values can be a regular distribution with finite moments, e. g., a Gaussian distribution. The corresponding shuffled series will exhibit

non-multifractal scaling, since all long-range correlations are destroyed by the shuffling procedure. Randomly shuffling the order of the values in the time series is the easiest way of generating surrogate data; however, there are more advanced alternatives. If both kinds of multifractality are present, the shuffled series will show weaker multifractality than the original series (Movahed et al., 2006). A multifractal analysis of time series will also reveal higher order correlations. Multifractal scaling can be observed if, e. g., three or four-point correlations scale differently from the standard two-point correlations studied by classical autocorrelation analysis. In addition, multifractal scaling is observed if the scaling behaviour of small and large fluctuations is different. For example, extreme events might be more or less correlated than typical events.

Multifractal processes could be seen as an extension of monofractal processes introduced by Mandelbrot and characterized by a single exponent such as Hurst parameter H or Fractal dimension D ($D = 2 - H$). The Hurst exponent is not sufficient for describing the dynamics of a multifractal process, needing a scaling exponent function $\zeta(q)$ or $K(q)$. The multifractal concepts were introduced with the multiplicative cascade model, to study the energy dissipation in the context of fully developed turbulence in the 1980's (Grassberger and Procaccia, 1983; Benzi et al., 1984; Parisi and Frisch, 1985). Let x be a nonstationary signal with stationary increments. For a scaling process, its increments $\Delta x = x(t + T) - x(t)$ possess scaling statistics of the form:

$$\langle |\Delta X|^q \rangle \sim T^{\zeta(q)}, \quad (2.32)$$

where T is the time increment and $\zeta(q)$ is the scale invariant exponent function, which is non-linear and concave. The estimation of the $(q, \zeta(q))$ curve for integer and non-integer moments, provides a full description of the stochastic process $x(t)$ at all intensities and all scales. This function characterizes the nature of its scaling behavior: $\zeta(q)$ is nonlinear and concave for a multifractal process and linear for a monofractal process. Furthermore, the concavity is an indication of intermittency. The parameter $H = \zeta(1)$ is the Hurst parameter. Monofractal processes correspond to a linear function $\zeta(q) = qH$, where Brownian motion is described by $H = 1/2$, fractional Brownian motion is described by $0 < H < 1$ and homogeneous non-intermittent turbulence is described by $H = 1/3$. In order to model intermittency, many statistical functions have been proposed to fit $\zeta(q)$ since the introduction of multifractals in the turbulence field.

- Log-stable model

In this study, we focus on the log-stable model or universal multifractal proposed by Schertzer and Lovejoy (1987) and Kida (1991):

$$\zeta(q) = qH - \frac{C_1}{(\alpha-1)} (q^\alpha - q) , \quad (2.33)$$

where $H = \zeta(1)$ the Hurst parameter defines the degree of smoothness or roughness of the field. The parameter C_1 is the fractal co-dimension of the set giving the dominant contribution to the mean ($q = 1$) and bounded between 0 and d (d the dimension space, here $d = 1$). It measures the inhomogeneity mean or the mean intermittency characterizing the sparseness of the field: the larger C_1 , the more the mean field is inhomogeneous. The multifractal Lévy parameter α is bounded between 0 and 2, where $\alpha = 0$ corresponds to the monofractal case and $\alpha = 2$ corresponds to the multifractal lognormal case. The parameter α measures the degree of multifractality, i.e, how fast the inhomogeneity increases with the order of the moments (Seuront et al., 1996 a, b).

The computation of multifractal indices α , and C_1 are obtained from Eq. (2.33) (Schertzer et al., 1997). For that, we analyze the following function

$$R(q) = q\zeta'(0) - \zeta(q) = \frac{C_1}{(\alpha-1)} q^\alpha . \quad (2.34)$$

Thus, the function $R(q)$ versus q will have a slope α and C_1 can be estimated by the intercept.

- Singularity spectrum

Another way to characterize a multifractal process is the singularity spectrum $M(\gamma)$. The singularity spectrum is a function used in multifractal analysis to describe the fractal dimension of a subset of points of a function belonging to a group of points that have the same Hurst exponent. Intuitively, the singularity spectrum gives a value for how fractal a set of points are in a function. Related to the scaling exponents $\zeta(q)$ presented above, $M(\gamma)$ could be present:

$$\gamma = d \zeta(q)/dq. \quad (2.35)$$

$$M(\gamma) = \gamma q - \zeta(q) + 1 . \quad (2.36)$$

Based on this, it is a monofractal process, $\gamma = H$ and $M(\gamma) = 1$.

2.2.4 Intermittency: Multifractal analysis

The objective of this thesis is to study the intermittency of GHI over Reunion Island and the multifractal properties of GHI time series have been highlighted in the introduction. Hence, this section will present the main methods of multifractal analysis used in this thesis.

2.2.4.1 Hilbert Spectral Analysis (HSA)

- Hilbert marginal spectrum

Hilbert spectral analysis (HSA) is associated with EMD. This spectral analysis is known to be a Hilbert transform applied to each IMF component extracted from the original signal by the EMD method. Thus the original signal $X(t)$ is written as a sum of IMF modes $C_i(t)$ and a residual $r_n(t)$ which is the mean trend or a constant. Figure 2.13 shows an example of original data and IMFs from EMD decomposition.

$$X(t) = \sum_{i=1}^N C_i(t) + r_n(t) \quad (2.37)$$

Even for data with zero mean, the final residue can still be different from zero. For data with a trend, then the final residue should be that trend. The process is indeed like sifting: to separate the finest local mode from the data first based only on the characteristic time scale. The sifting process, however, has two effects: (a) to eliminate riding waves; and (b) to smooth uneven amplitudes. Therefore, the sifting process should be repeated enough times. However, if too many times sifting are performed, the amplitude of the IMF modes will become constant, and the nonlinear wave profiles is then distorted, which means the modes lose their physical meaning (Huang et al., 1998, 1999).

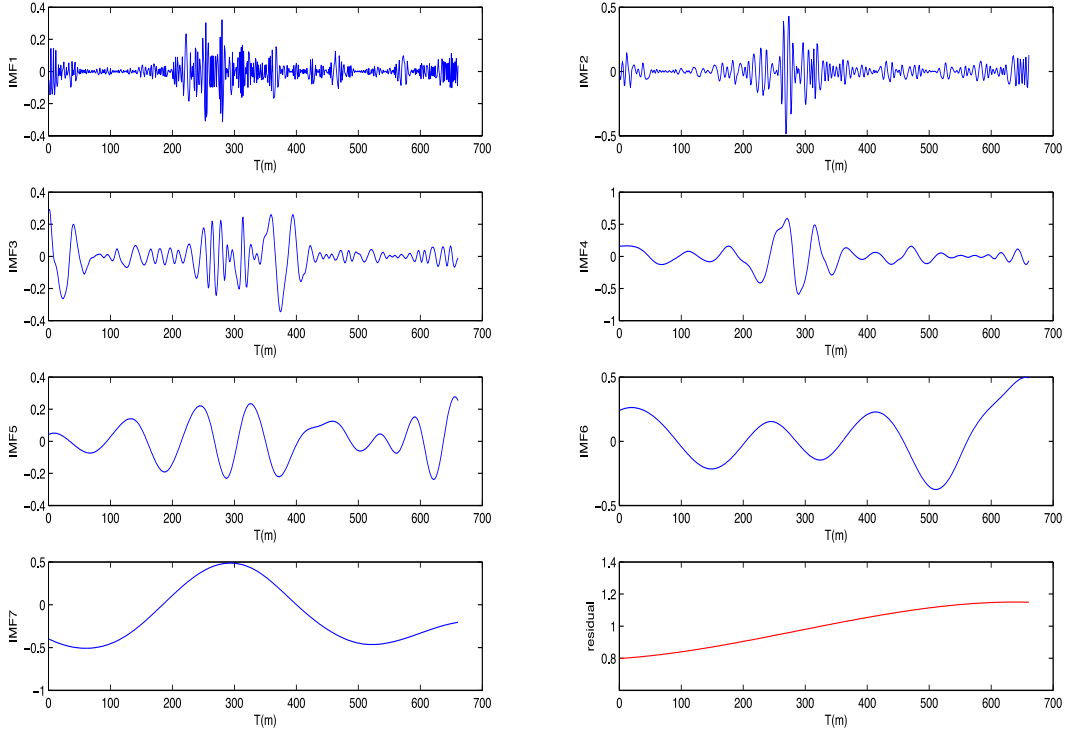


Figure 2.13 An example IMFs from EMD decomposition of global solar radiation data.

Giving $C_i(t)$ to be the IMF mode, HSA acts as a time frequency analysis to extract the energy-frequency information from the data (Huang et al. 1998, 1999; Cohen, 1995; Long et al., 1995). The Hilbert transform applied to one mode $C_i(t)$ is written as

$$\tilde{C} = \frac{1}{\pi} P \int_0^{\infty} \frac{C_i(t')}{t-t'} dt', \quad (2.38)$$

where P indicates the Cauchy principle value (Cohen, 1995; Long et al., 1995). With this definition, C and \tilde{C} form the complex conjugate pair, so we can construct an analytic signal z as

$$z = C + j\tilde{C} = A(t)e^{j\theta(t)}, \quad (2.39)$$

where C , \tilde{C} is real and imaginary part of a signal respectively. Theoretically, there are infinitely many ways of defining the imaginary part, but the Hilbert transform provides a unique way of defining the imaginary part so that the result is an analytic function. A brief tutorial on the Hilbert transform with the emphasis on its physical interpretation can be found in Bendat and Piersol (1986). The corresponding amplitude and phase functions are defined as,

$$Am = |z| = \sqrt{C^2 + \tilde{C}^2}. \quad (2.40)$$

$$\theta(t) = \arg(z) = \arctan \tilde{C} / C \quad (2.41)$$

The instantaneous frequency ω is determined from the phase $\theta(t)$ in equation. The notion of the instantaneous energy or the instantaneous envelope of the signal is well accepted. On the other hand, the notion of the instantaneous frequency has been highly controversial. Existing opinions range from editing it out of existence (Shekel, 1953) to accepting it but only for special “monocomponent” signals (Boashash, 1992; Cohen, 1995). In principle, there are some limitations for monocomponent on the data. At any given time, there is only one frequency value. Therefore, it can only represent one component, hence “monocomponent”.

$$\omega = d\theta / dt \quad (2.42)$$

Unfortunately, no clear definition of the “monocomponent” signal was given to judge whether a function is or is not “monocomponent”. For lack of a precise definition, “narrow band” was adopted as a limitation on the data for the instantaneous frequency to make sense (Schwartz et al., 1966). But the bandwidth limitation on the Hilbert transform to give a meaningful instantaneous frequency has never been firmly established. In order to obtain meaningful instantaneous frequency, restrictive conditions have to be imposed on the data as discussed by (Gabor, 1946; Bedrosian, 1963; Boashash, 1992): for any function to have a meaningful instantaneous frequency, the real part of its Fourier transform has to have only positive frequency. This restriction can be proven mathematically as shown in (Titchmarsh, 1948), but it is still global. (Huang et al., 1998, 1999) modify the restriction condition from a global one to a local one, and based on its local properties. They designated as intrinsic mode function (IMF) that we calculated above for defining the instantaneous frequency everywhere.

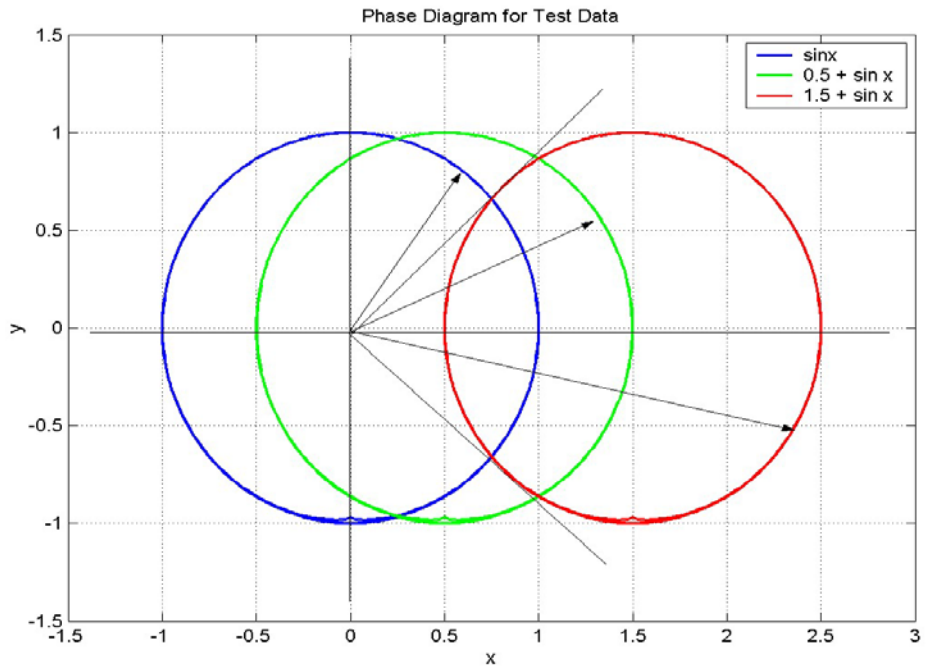
They consider some simple examples to illustrate these restrictions physically, by examining the function,

$$x(t) = \sin t \quad (2.43)$$

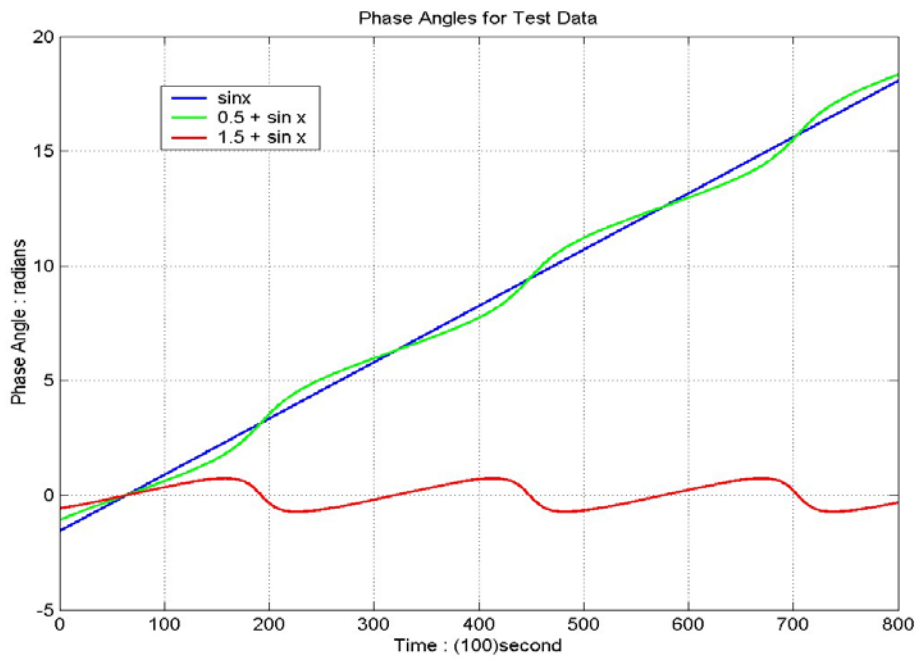
$$x(t) = a + \sin t \quad (2.44)$$

Its Hilbert transform is simply $\cos t$. The phase plot of x - y is a simple circle of unit radius as in figure 2.14 a). The phase function is a straight line as shown in figure 2.14 b) and the instantaneous frequency, shown in figure 2.14 c), is a constant as expected.

a)



b)



c)

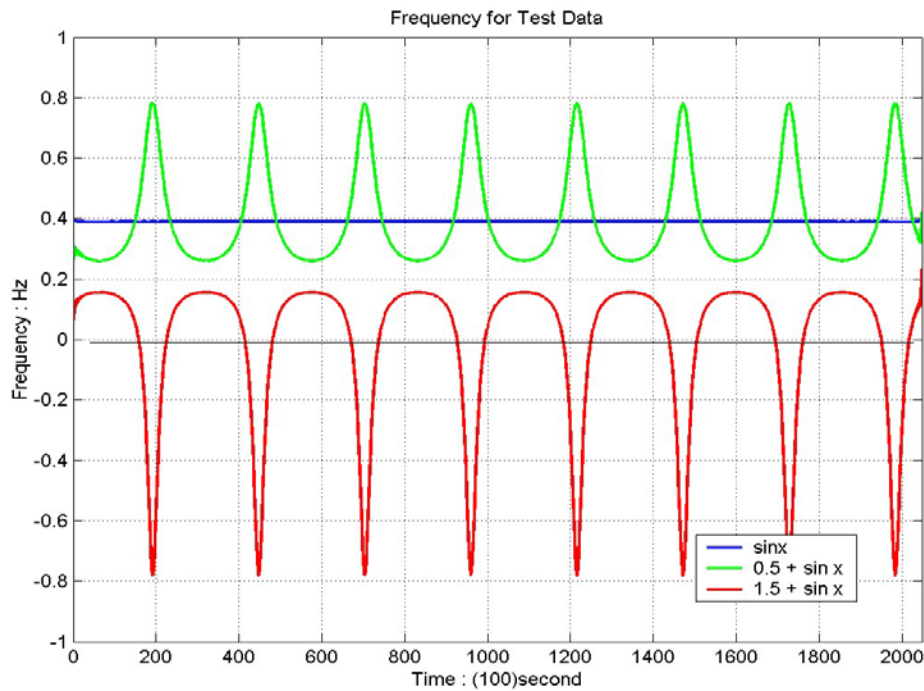


Figure 2.14 a) The phase plane for the model functions of $x(t) = a + \sin t$: blue cycle for $a = 0$; (b) green cycle for $a < 1$; (c) red cycle for $a > 1$; Figure 2.14 b) The unwrapped phase function of the model functions; Figure 2.14 c) The instantaneous frequency computed according to equation.

The phase plot of x - y is still a simple circle independent of the value of a , but the center of the circle will be displaced by the amount of a as illustrated in figure 2.14 a). If $a < 1$, the center is still within the circle. Under this condition, the function has already violated a restriction, for its Fourier spectrum has a DC (Direct Current) term; nevertheless, the mean zero-crossing frequency is still the same as in the case for $a = 0$, but the phase function and the instantaneous frequency will be very different as shown in figures 2.14 b, c. If $a > 1$, the center is outside the circle; thus, the function no longer satisfies the required conditions. Then both the phase function and the instantaneous frequency will assume negative values as shown in figures 2.14 b, c, which are meaningless. These simple examples illustrate physically that, for a simple signal such as a sine function, the instantaneous frequency can be defined only if we restrict the function to be symmetric locally with respect to the zero mean level.

For general data, any riding waves would be equivalent to the case of $a > 1$ locally; any asymmetric wave form will be equivalent to the case of $a < 1$, but non-zero, locally.

Neglecting the residual, the original time series is rewritten as

$$X(t) = \text{Re} \sum_{i=1}^N A_i(t) e^{i\theta_i(t)}, \quad (2.45)$$

where A_i and θ_i are the amplitude and phase time series of mode i and Re means real part (Huang, et al., 1998, 1999).

For each mode, the Hilbert spectrum is defined as the square amplitude $H(\omega, t) = A^2(\omega, t)$. $H(\omega, t)$ gives a local representation of energy in the time frequency domain. Then the Hilbert marginal spectrum $h(\omega)$ is written as

$$h(\omega) = \int_0^{\infty} H(\omega, t) dt. \quad (2.46)$$

The Hilbert marginal spectrum is similar to the Fourier spectrum, as it corresponds to the energy associated to the frequency (Huang et al., 1998, 1999). We do not give the validation and calibration detail of the Hilbert-Huang transform here. For details of the validation and calibration, we suggest to read the publication of Huang et al. (1998, 1999).

The Hilbert marginal spectrum is defined as a marginal integration of the Hilbert spectrum $H(\omega, t)$ over t . They define a joint probability density function ($p(\omega, A)$) of the instantaneous frequency ω and the amplitude A (Long et al., 1995; Huang et al., 2008, 2009a) for all the IMF modes. Thus the corresponding Hilbert marginal spectrum is rewritten as the marginal integral of the joint pdf $p(\omega, A)$ over A^2 .

$$h(\omega) = \int_0^{\infty} p(\omega, A) A^2 dA. \quad (2.47)$$

This expression concerns the second-order statistical moment.

- Joint probability density function

The joint probability density function (joint PDF) is a function used to characterize the probability distribution of a continuous random vector. It is a multivariate generalization of the probability density function (PDF). The generalization works as follows: the integral of the density of a continuous variable over an interval is equal to the probability that the variable will belong to that interval; the multiple integral of the joint density of a continuous random vector over a given set is equal to the probability that the random vector will belong to that set.

2.2.4.2 Arbitrary order Hilbert Spectral Analysis

In order to deal with scaling intermittent multifractal time series, as the extension of HHT, arbitrary-order Hilbert spectral analysis was proposed by (Huang et al., 2008, 2009) to characterize such scale invariant property of signals.

- Intrinsic Mode function and HSA

The arbitrary-order Hilbert spectral analysis is the combination of the Empirical Mode Decomposition (EMD) and Hilbert spectral analysis (HSA). The EMD method provides an effective tool to decompose a signal into a collection of Intrinsic Mode Functions (IMF) that allow well-behaved Hilbert transforms for computation of physically meaningful time-frequency representation, which are already stated before in this chapter. The HSA is performed to each obtained IMF component $I_m(t)$ extracted by the EMD method. The energy in a time-frequency space is estimated from the Hilbert spectrum, $H(\omega, t) = A^2(\omega, t)$. The Hilbert spectrum $h(\omega)$ being defined as:

$$h(\omega) = \int_0^N H(\omega, t) dt , \quad (2.48)$$

where N the total data length. The Hilbert spectrum $H(\omega, t)$ gives a measure of the amplitude from each frequency and time, while the marginal spectrum $h(\omega)$ gives a measure of the total amplitude from each frequency.

- Generalized Hilbert marginal spectrum

The second-order statistical moment was expressed by equation 2.45. A generalization of this definition is considered to arbitrary-order statistical moment $q \geq 0$ (Huang et al., 2008, 2009):

$$L_q(\omega) = \int_0^\infty p(\omega, A) A^q dA . \quad (2.49)$$

- Generalized Hurst scaling exponent

Hence, in the Hilbert space, the scale invariance is written as

$$L_q(\omega) \sim \omega^{-\xi(q)} , \quad (2.50)$$

where $\xi(q)$ is the corresponding generalized scaling exponent in the Hilbert space. This scaling exponent function is linked to scaling exponent function $\zeta(q)$ of structure functions analysis by the expression (Huang et al., 2008, 2009):

$$\zeta(q) = \xi(q) - 1. \quad (2.51)$$

This method provides a way to characterize the scale invariance in an amplitude-frequency space (Huang et al., 2008, 2010a, 2009a) and estimate the scaling exponent that characterizes intermittency process.

We do not give the validation and calibration detail of the arbitrary-order Hilbert spectral analysis here neither. For details of the validation and calibration, we suggest to read the publication of (Huang et al., 2008, 2009).

The study dataset, the prediction method and the multifractal analysis method are presented in this chapter. They provide the basic information for the results analysis in the following section: chapter 3 and 4.

Chapter 3

Daily mapping prediction

3.1 Data processing

3.1.1 Missing data: Filling technique

This chapter presents the daily mapping prediction results with the CM SAF data. The SARAHE satellite data (CM SAF) is used to perform the one-day ahead prediction for surface solar radiation (SSR) over Reunion Island. As I indicated in Chapter 2, the spatial resolution of this dataset is $0.05^\circ \times 0.05^\circ$ and the temporal coverage is from 2007 to 2016. The first five years (2007-2011) are used as training data, and the remaining five years (2012-2016) are used as test data in the prediction model. The domain used for the prediction is $55.05^\circ\text{E} \sim 56^\circ\text{E}$, $-21.55^\circ\text{S} \sim -20.70^\circ\text{S}$, which covers Reunion Island, as shown in the Figure 3.1.

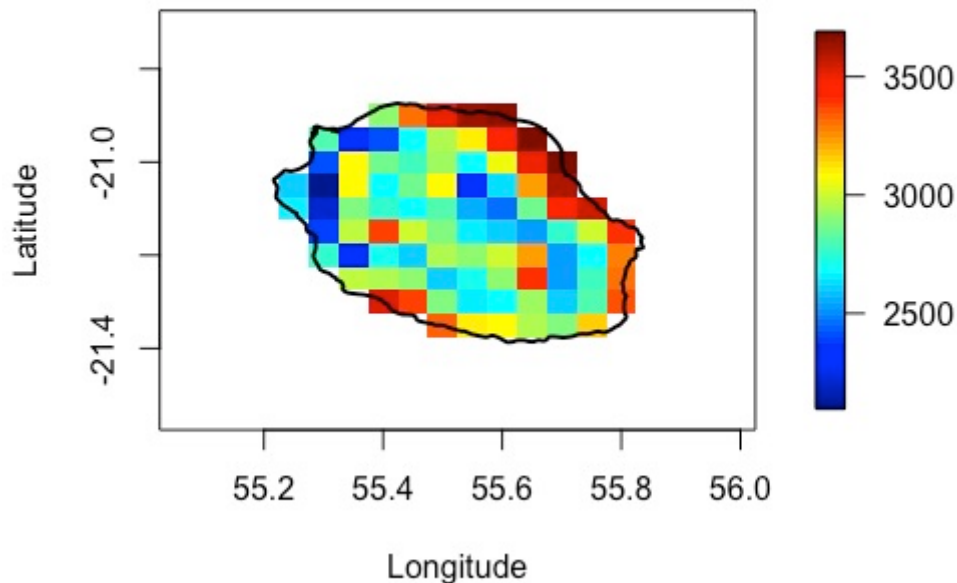


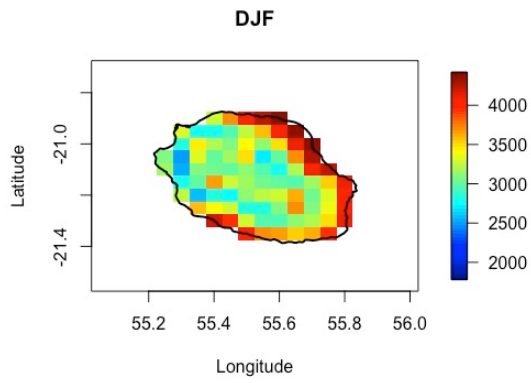
Figure 3.1 Domain of the CM SAF dataset for the prediction model.

Because those 10-years data are from satellite and there are not much missing data, the days for which we don't have record have been removed.

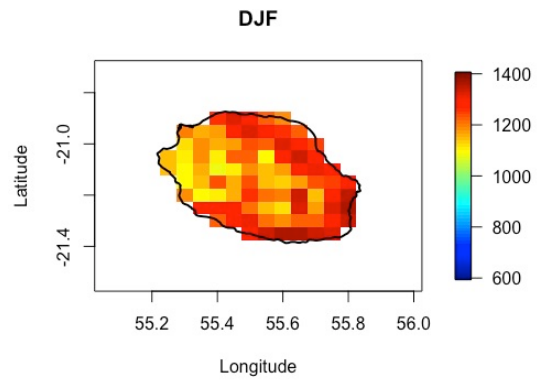
Figure 3.2 presents the seasonal mean and standard deviation of SSR in summer, winter and inter-season from 2007 to 2016 (10 years) over Reunion Island. From Figures 3.2 a), c), e) and h), the obvious seasonal variability can be observed. The austral summer season (DJF: December-January-February) obtains more SSR than austral winter season (JJA: June-July-August) over the whole island. During the inter-season (MAM: March-April-May and SON: September-October-November), there is more SSR over the northeast coast than over the

inland area, because in this period, strong thermal and wind inversion occur (Badosa et al., 2013). Turbulent flows and clouds activities over the island are thus exclusively vertical in the boundary layer.

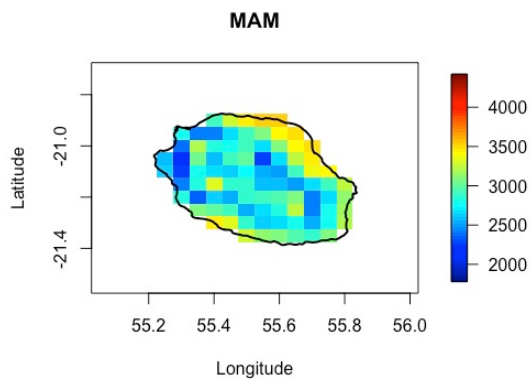
a)



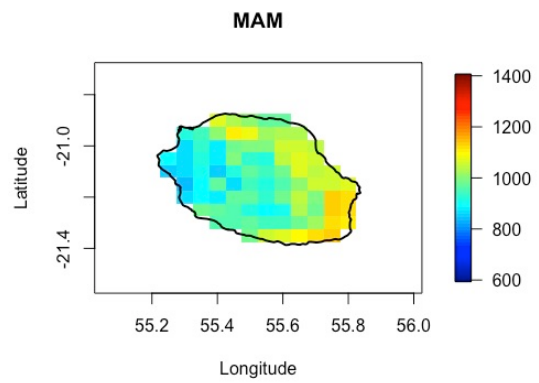
b)



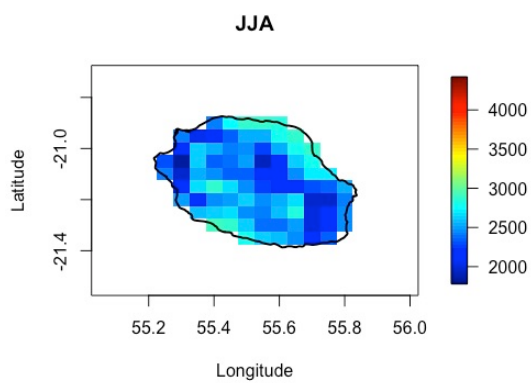
c)



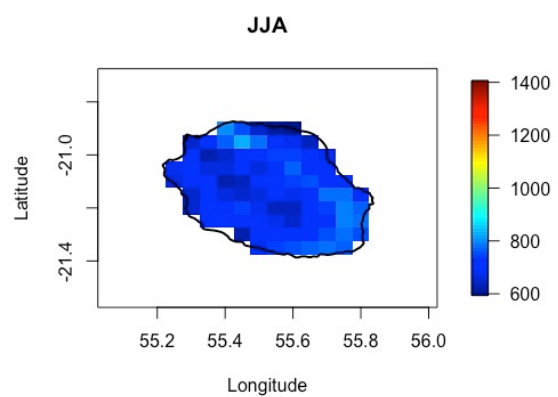
d)



e)



f)



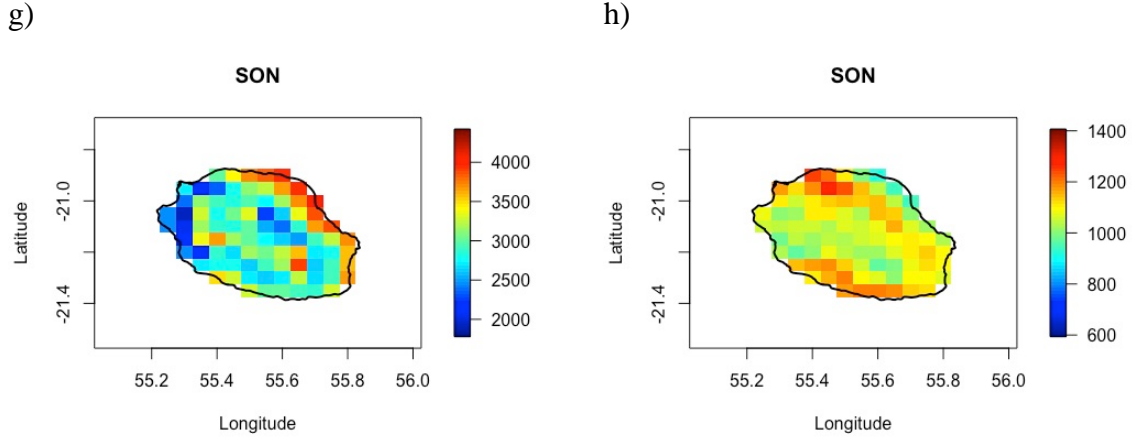


Figure 3.2 The seasonal mean (a; c; e; g) and standard deviation (b; d; f; h) of SSR in the summer, winter, and inter-season from 2007 to 2016 over Reunion (W/m^2).

3.1.2 Daily scale: Detrended and deseasonalized data

The total amount of solar irradiation received at one place on the Earth's surface and at a given latitude varies daily and seasonally. In order to overcome these two-time scale variabilities in each day, a filtering process has been applied to the GHI dataset for de-trending and de-seasonalizing. The adjusted seasonal index method which is a simple and convenient pre-processing method (Brockwell and Davis, 1986), was used.

Let $G(m, d)$ be GHI solar radiation observed at time m ($m \in [07:00 - 18:00]$) and day d ($d \in [1/1/2007 - 31/12/2016]$). The steps to compute the deseasonalized GHI time series using adjusted seasonal index method is as follow:

Step 1- Compute $\overline{G(d)}$ which is the GHI daily mean:

$$\overline{G(d)} = \frac{1}{M} \sum_{m=1}^M G(m, d). \quad (3.1)$$

Step 2- Compute the modified GHI time series $G_m(m, d)$ by dividing the original time series $G(m, d)$ by its daily mean:

$$G_m(m, d) = \frac{G(m, d)}{\overline{G(d)}}, \quad (3.2)$$

where $\overline{G(d)}$ is the GHI daily mean computed in step 1.

Step 3- Compute $\overline{G(m)}$ which is the GHI climatologically daily mean of the modified GHI time series $G_m(m, d)$:

$$\overline{G(m)} = \frac{1}{D} \sum_{d=1}^D G_m(m, d). \quad (3.3)$$

Step 4- Compute AF the adjusted factor of $\overline{G(m)}$:

$$AF = \frac{\sum_{m=1}^M \overline{G(m)}}{M}. \quad (3.4)$$

Step 5- Compute the deseasonalized adjusted GHI time series $G_{da}(m, d)$ by dividing modified GHI time series $G_m(m, d)$ (computed in step 2) by the GHI climatologically daily mean $\overline{G(m)}$ (computed in step 3) corrected by the adjusted factor AF (computed in step 4):

$$G_{da}(m, d) = AF \frac{G_m(m, d)}{\overline{G(m)}}. \quad (3.5)$$

This adjusted seasonal index method has been used in this study, however, the results obtained present the low correlation coefficient (0.2~0.3) between two days. Based on this consideration, it is not practical to apply on our satellite dataset (CM SAF). Thus, only the normalization has been done for pre-processing.

3.1.3 Dataset normalization

We perform normalization over the input data at first for pre-processing. Normalization is the process of restructuring a relational database in accordance with a series of so-called normal forms in order to reduce data redundancy and improve data integrity, which was first proposed by Codd (1970). If there is a dataset with various values on large differences between them, then it is better to obtain a more compact dataset through normalization. Normalization reduces dispersion of data collected in order to obtain better results (Neelamegam, 2016). Before performing the prediction, the training and test dataset are normalized. The Figures 3.3 and 3.4 give an example of the original and normalized GHI time series separately.

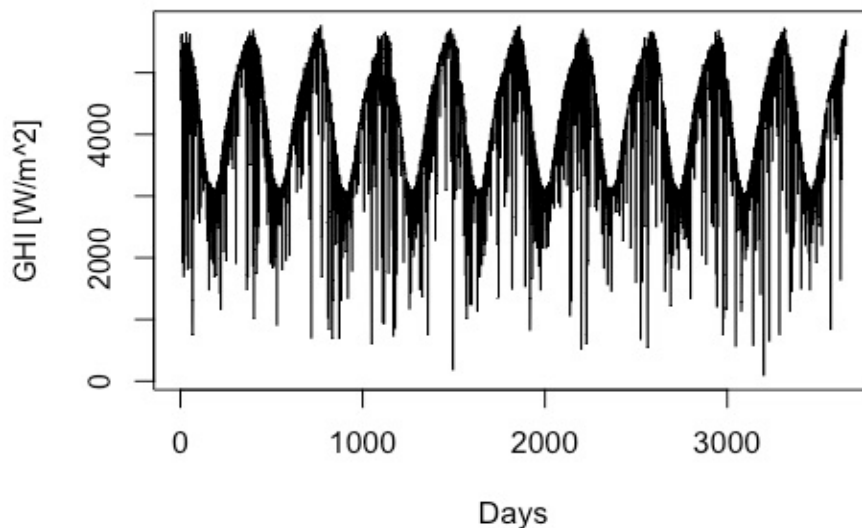


Figure 3.3 The original GHI time series with the trend for the year 2007~2016.

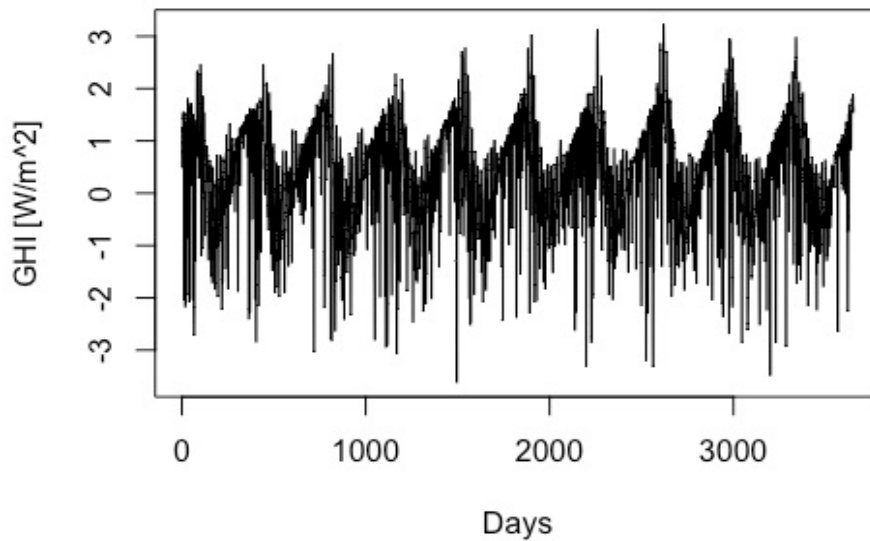


Figure 3.4 The normalized daily time series 2007~2016 in one grid point of Reunion.

3.2 Linear models

3.2.1 Independent/dependent variables

The aim of this section is to present the dependent variables and independent variables which will be used in the each linear prediction model. There are four linear regression models used in this study as mentioned in Chapter 2: SR, MLR, PCR and PLSR. 1) SR: the first three days' daily GHI in the domain area (all the grid points) are the independent variable in this linear model, and the dependent variable is the fourth day's daily GHI in the domain area (all the grid point); 2) MLR: the independent variables are the first three days' daily GHI at each grid point of the domain, and the dependent variable is the fourth day's daily GHI at each grid point of the domain; 3) PCR: each PC getting from the PCA analysis of GHI data is the independent variable and the predicted PC is the dependent variable; 4) PLSR: each PC getting from the PCA analysis of GHI data is the independent variable and the predicted PC is the dependent variable.

3.2.2 PCA decomposition

As explained in Chapter 2, applying linear regression models to do prediction requires to choose the independent and dependent variables as predict parameters and predicted value. The climate dataset, such as solar radiation data, are usually multi-dimensional: time, latitude, longitude..... All these dimensions stand all the variables. To interpret the data in a more meaningful form, it is therefore necessary to reduce the number of variables, to a set of

interpretable linear combinations of the data. PCA is the most common method for decomposition, which reduces the data size and simplifies the calculation. Each linear combination corresponds to a principal component with EOF.

- Hypothesis on EOF

PCA analysis could largely reduce the dimensionality for the heavy dataset, like multi-years GHI. When the PCA is applied in the linear prediction model, the PCR model is often described as follows:

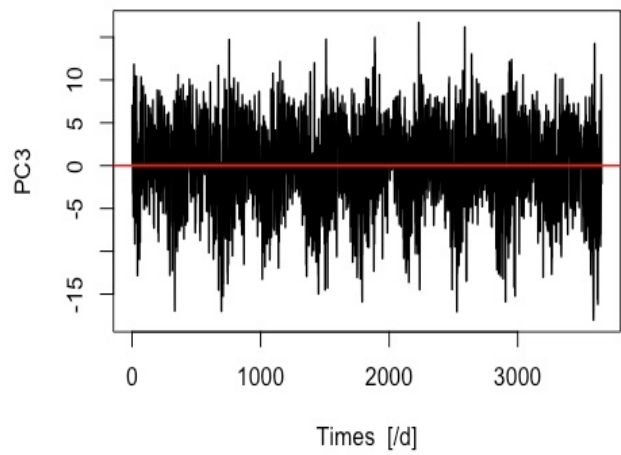
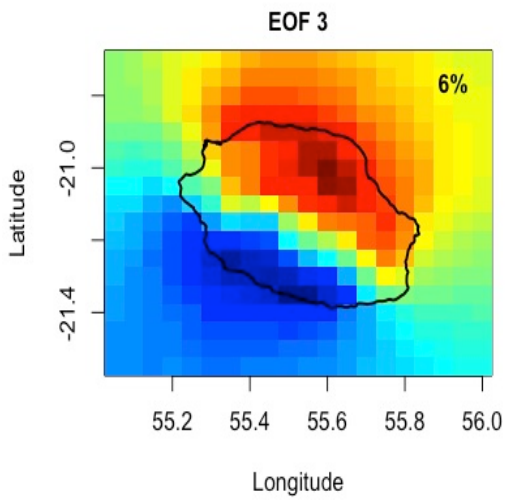
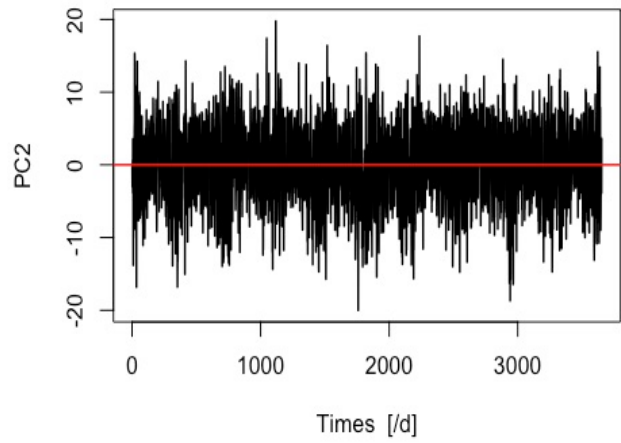
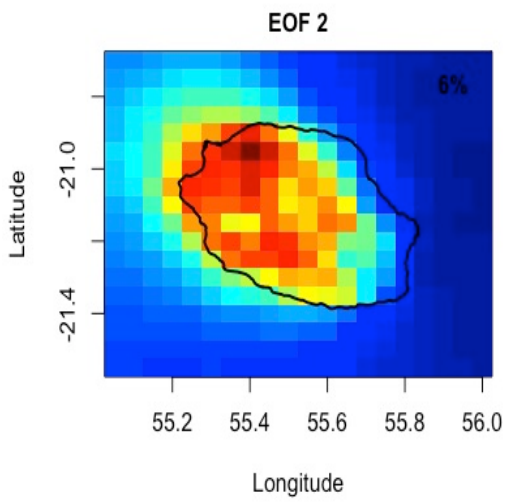
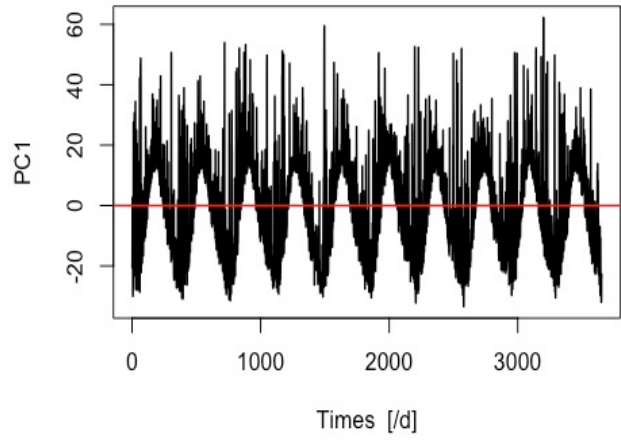
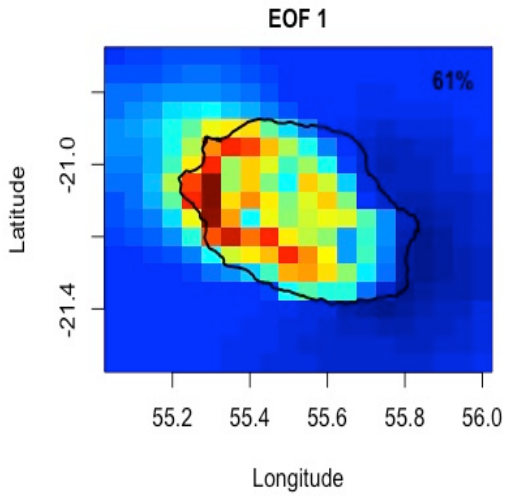
$$Y = \gamma_0 1 + g\gamma_1 + Z\gamma_2 + \varepsilon, \quad (3.6)$$

where γ_0 is the intercept, and Z is the matrix with each column as one of a few top PCs constructed by PCA from GHI daily dataset, or more generally, g is normalized daily GHI, and ε is the error term. The null hypothesis for the EOF is all regression coefficients are zero ($H_0: \gamma_1 = 0$). The basic idea of PCR is that we use the first few PCs to replace the original large GHI variables in model.

- PC as independent/dependent variables in the linear model

PCA is applied to the GHI data in 2007~2016 in order to obtain the PCs. The PCs which could explain 95% variance are used as the independent variables and the according predicted PCs are the dependent variable as output in the linear model. The first 5 PCs with their EOFs modes for the SSR data in 2007~2016 from CM SAF (SARAH-E@5km) are presented in Figure 3.5. EOF provides us with both the spatial and temporal patterns of the dominant modes of variability. The explained variance (%) of each PC is given in Figure 3.6, and 43 PCs together can explain 95% variance. In Figure 3.5, it is observed that the leading 3 modes of EOF explains 61%, 6% and 6% variance separately, and the first PC gives more contribution for SSR over Reunion Island. And these three PCs give obviously different spatial distribution of SSR over Reunion Island: first EOF mode shows more (less) SSR over north-west (south-west) part; Second EOF mode shows more (less) SSR over north (south) part; Third EOF mode shows more (less) SSR over north-east (south-west) part.

ing



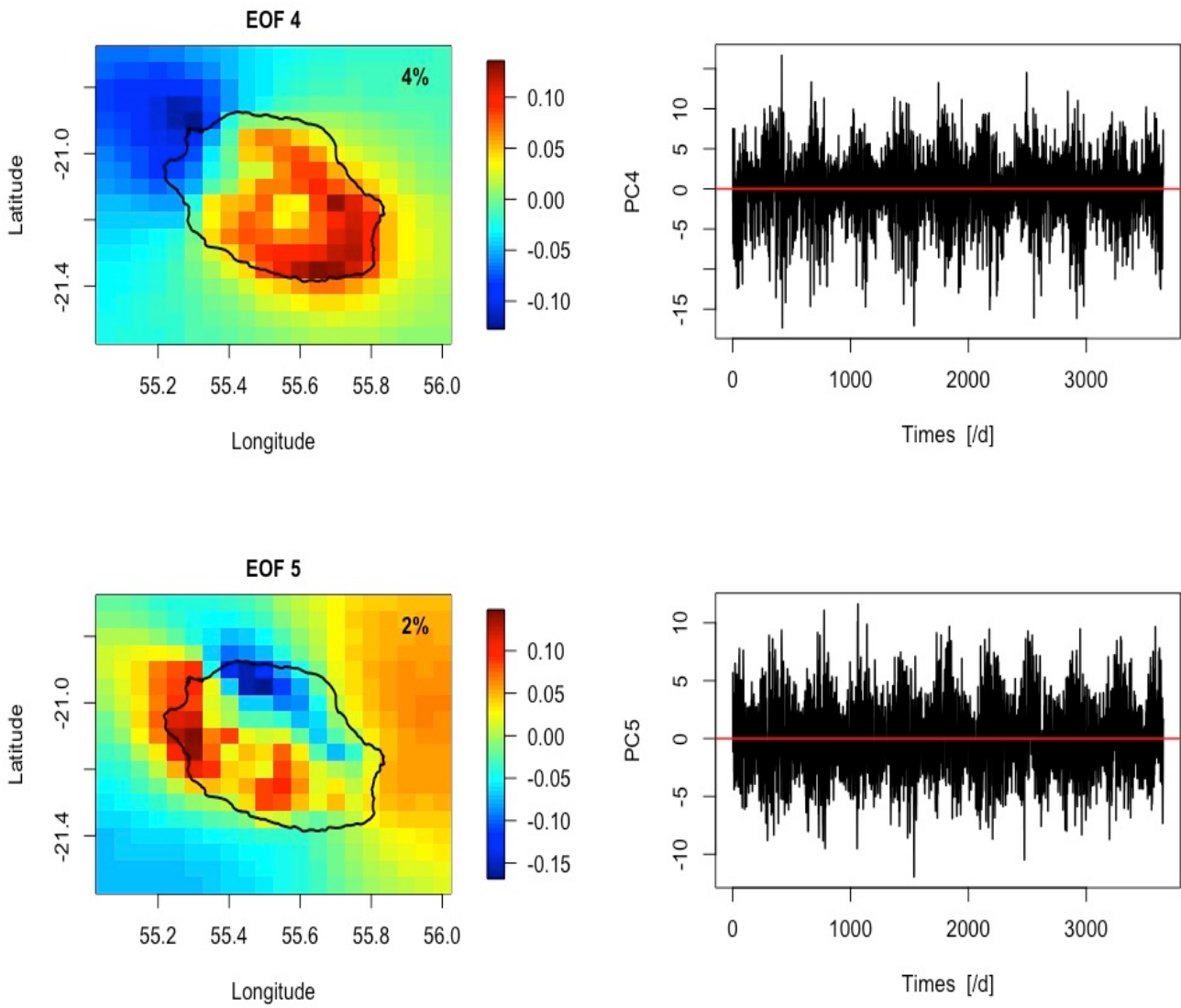


Figure 3.5 The leading five EOF modes (left panel) for GHI for its PCs (right panel) in 2007~2016.

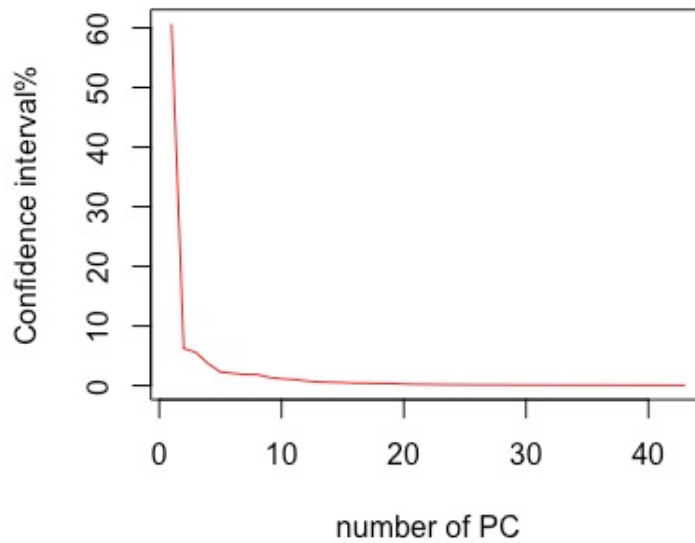


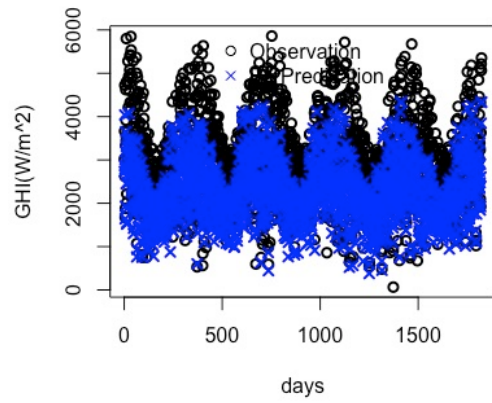
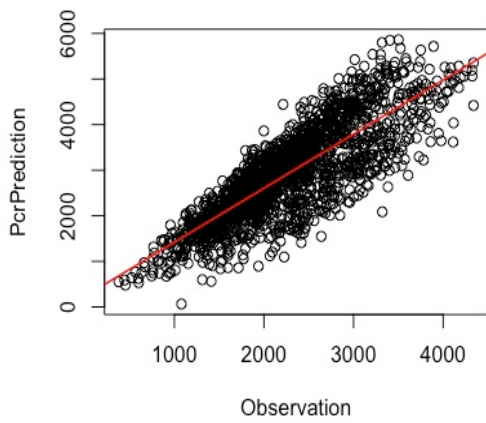
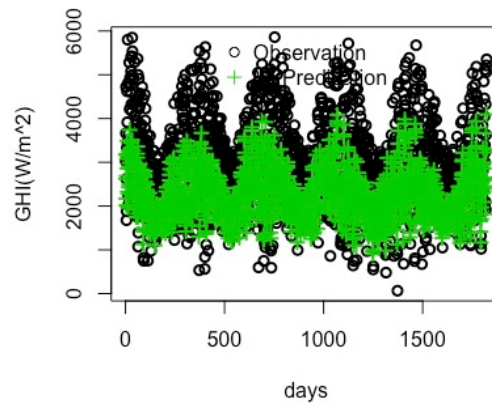
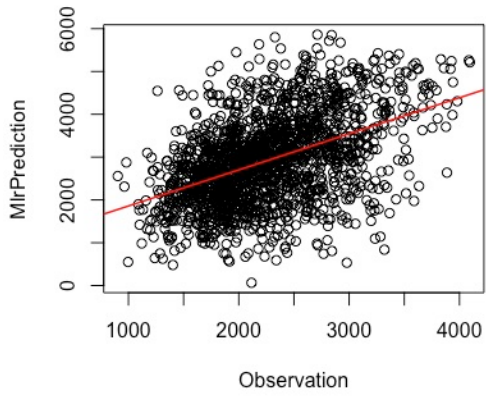
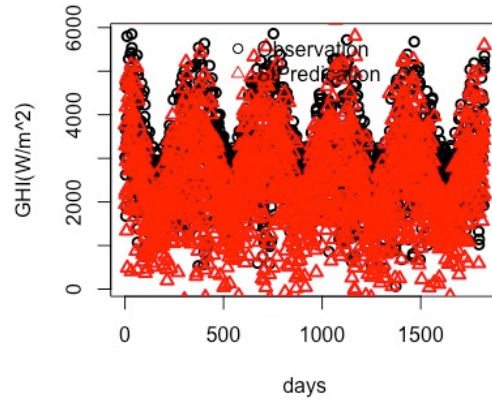
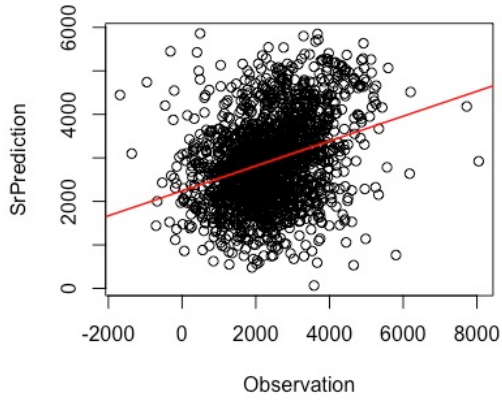
Figure 3.6 The explained variance (%) of each PC.

3.3 Results

3.3.1 Test with SR, MLR, PCR and PLSR models

Four linear regression models: SR, MLR, PCR and PLSR are used in this study to do prediction of SSR in 2007~2016 over Reunion Island. 2007-2011, these five years used as the training data and other five years from 2012~2016 as test data. Figure 3.7 presents the goodness of fitting in a plot of prediction against observation (left panel) and comparing the daily prediction with observation (right panel) at one grid point over Reunion for example for these four linear regression models 2012~2016. From this one grid point example, the four models give different prediction comparing to the observations. For this only one grid point, it can be seen at first that PCR prediction model fits against the observations better. Even though the PCR predicted less SSR in the summer seasons than observation, it gives a quite accurate value in the winter seasons as compared to SR (MLR/PLSR) which gives relatively poor prediction for both summer and winter. Thus, PCR gives better prediction results than the other three models.

Data mapping prediction



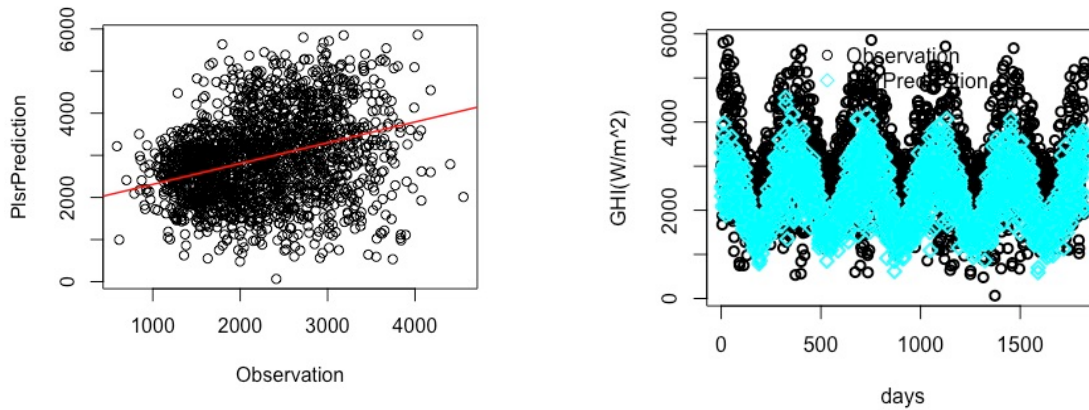


Figure 3.7 The goodness of fitting in a plot of prediction against observation (left panel) and comparing the daily prediction with observation (right panel) at one grid point in 2012~2016 over Reunion (W/m^2).

The re-mapping is conducted to the predicted data for spatial-comparison over the whole island. The 3rd July and the day on 31 December 2016 are taken as two example to show the mapping results (Figure 3.8 and Figure 3.9). These two days stand for the winter and summer season, and could feedback the character of seasonal variability. The top mapping in Figure 3.8 and Figure 3.9 is the observation and the other four mappings in the bottom are prediction mapping results from four linear regression models separately: SR, MLR, PCR and PLSR.

In Figure 3.8, SR, MLR, PCR and PLSR prediction all could present the obvious less SSR over the volcano area (Piton de la Fournaise) and the Cirque de Salazie comparing to the observation on 3 July 2016. The four models show similar spatial variability of SSR over the other areas over the island, and it seems that PCR model give closer SSR distribution (more SSR) over southern coastal line as the observation.

Data mapping prediction

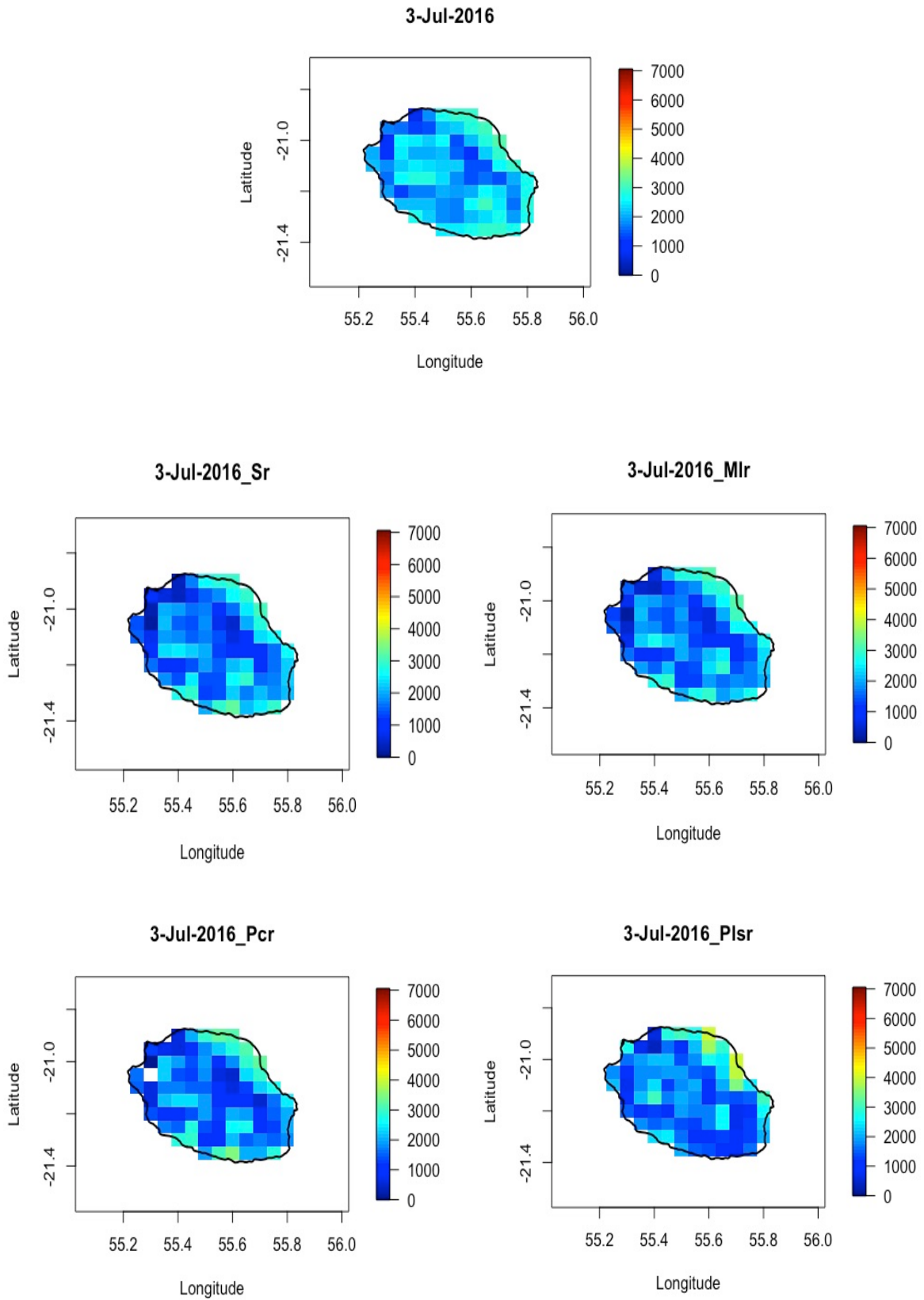


Figure 3.8 The SSR prediction mapping with four different linear regression methods against the original mapping (W/m^2) on 3 July 2016.

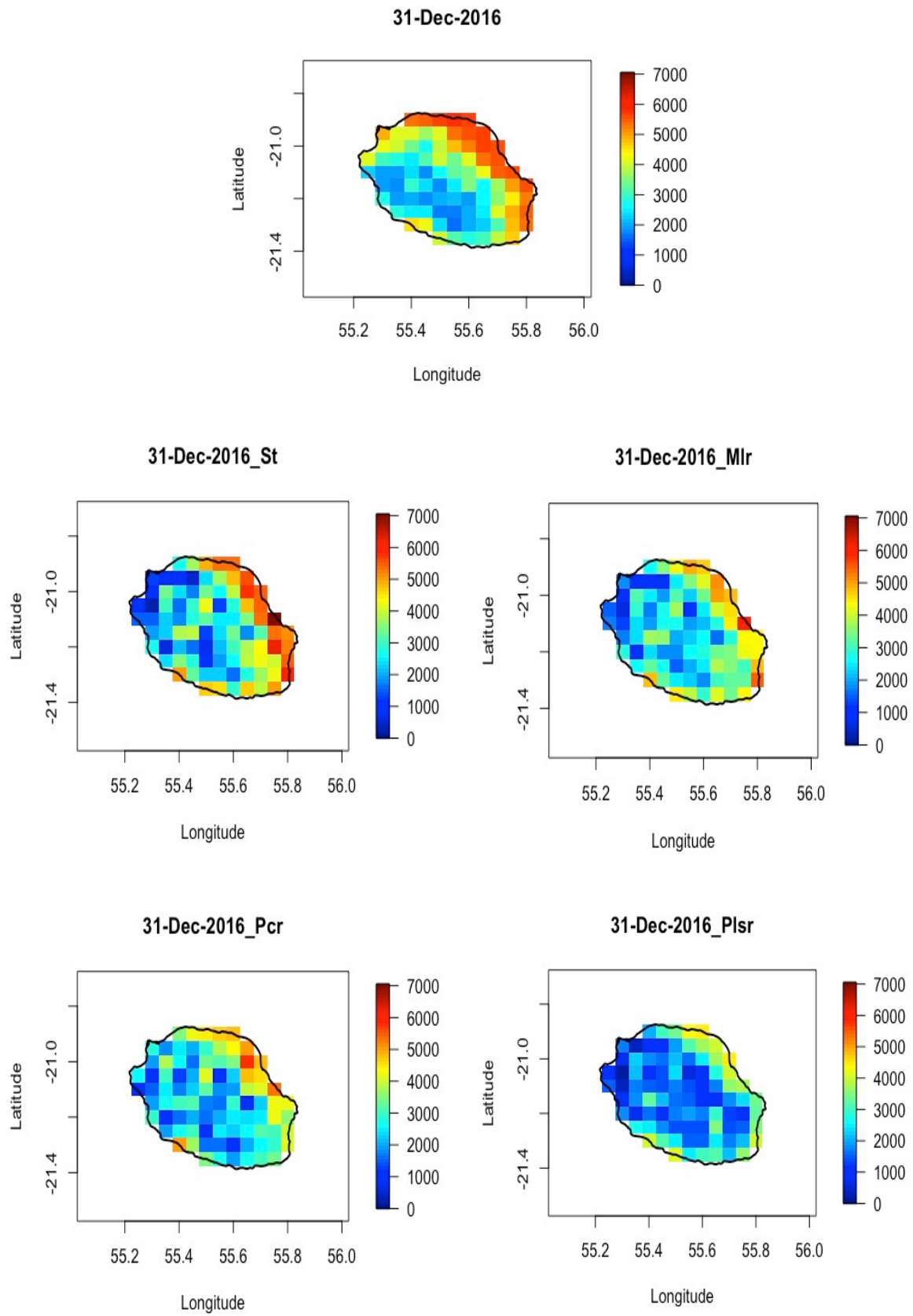


Figure 3.9 The SSR prediction mapping with four different linear regression methods against the original mapping (W/m^2) on 31 December 2016.

Figure 3.9(a) shows the observation of SSR on 31st December 2016. Figure 3.9(b) and (c) show the prediction due to SR, MLR, PCR and PLSR models respectively for 31st December 2016. All the four prediction models present more (less) daily SSR over north-east (three Cirques and the Volcano) area compared to the observation.

Figure 3.10 shows the multi-annual mean of daily SSR mapping prediction with SR, MLR, PCR and PLSR models in 2012~2016 comparing to the observation. Four prediction models all could present more (less) daily SSR over north-east (three Cirques and the Volcano) area as the observation in the five years. PCR seems to give less difference comparing to the observation than the other models.

Data mapping prediction

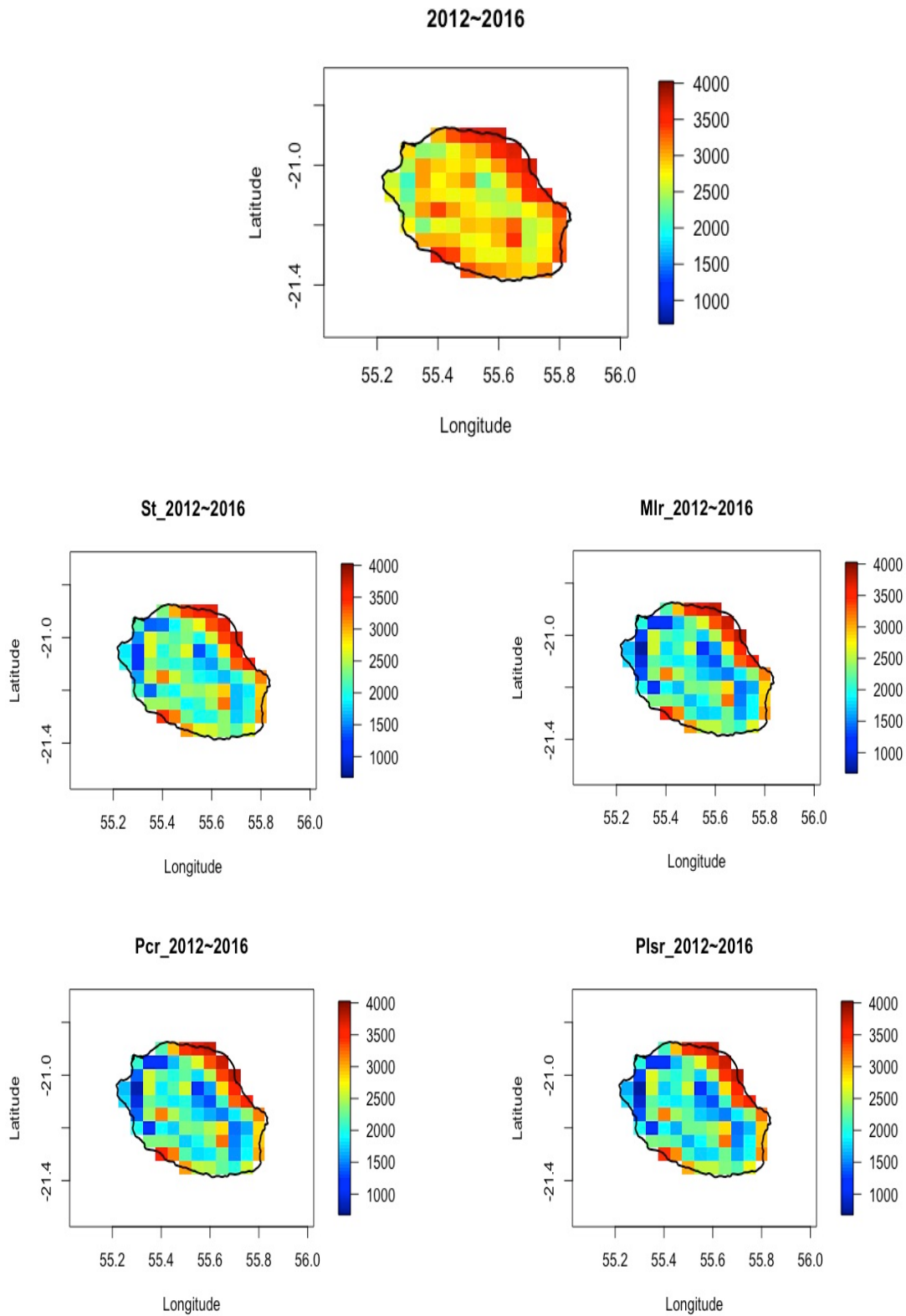


Figure 3.10 The multi-annual mean of daily SSR mapping prediction with SR, MLR, PCR and PLSR models in 2012~2016 compared to the observation.

3.3.2 Score evaluation: MAE, MSE, and RMSE

Table 3.1 Statistics of linear regression models

| | MAE (W/m ²) | MSE | RMSE (W/m ²) |
|------------|-------------------------|----------------|--------------------------|
| SR | 1.142e+3 | 2.15e+6 | 1.47e+3 |
| MLR | 1.00e+3 | 1.58e+6 | 1.26e+3 |
| PCR | 7.97e+2 | 9.24e+5 | 9.61e+2 |
| PLSR | 1.11e+3 | 1.93e+6 | 1.39e+3 |

Based on the mapping prediction results with four linear regression models (SR, MLR, PCR and PLSR), the statistical analysis using MAE, MSE and RMSE is conducted. Table 3.1 lists the calculation for the four models. Table 3.1 shows that PCR model has smallest MAE, MSE and RMSE as compared to the other three models. PCR model seems better for SSR mapping prediction over Reunion Island. Although PCR model gives better prediction results, it can be seen that (MAE, MSE and RMSE are quite big) the method is not quite accurate. Then, it is necessary to find a way to improve the accuracy of this prediction method.

Chapter 4

Intermittency: multifractal framework and results

As stated before, the arbitrary order Hilbert spectral analyses are applied in this thesis to study the intermittency and multifractality of global solar radiation over Reunion Island. The generalized scaling exponents $\xi(q)$ is estimated through the arbitrary order Hilbert spectral analyses which is the combination of the Empirical Mode Decomposition and Hilbert spectral analysis (EMD-HSA). Three parameters: Hurst exponent (H), the fractal co-dimension (C_1), and Lévy parameter (α) are taken to study the multifractal process of the global solar radiation. To achieve this multifractal analysis, we have used one highest available sampling rate of one-year GHI records (one observation per second) which is located at Saint-Denis (Moufia). In addition, we have also used twelve other stations with lower sampling rate of GHI records (one observation per minute) and time period greater than one year.

4.1 Intermittency and multifractality analysis on one SPN1 station

4.1.1 Mean sub-daily and daily multifractal pattern

As Moufia measurements are performed by a SPN1 which records the GHI per second from May to December in 2016, it is interesting to study the intermittency and multifractality of GHI sampling at this high frequency. This SPN1 is the A1211 pyranometer (Delta T Device) which is located on the S1 science building of University of Reunion. The datalogger was DLF1 (CR1000 Campbell Scientific, reference E3636). It gives records in a second rate without average. Firstly, we focus on the mean sub-daily and daily multifractal pattern of GHI. It is normal that the GHI during the day time (7h-18h) is much higher than at night. So, it is more interesting to just focus on the GHI variability within the day time for checking the annual and seasonal cycle. The arbitrary order Hilbert spectral analysis (EMD+HSA) is applied to GHI time series for obtaining the generalized scaling exponent and log-stable model parameters are analyzed. Then the multifractal processes could be found in the sub-daily and daily fluctuations.

Figure 4.1 gives the studied dataset in this part: daily GHI from May to December in 2016 sampling in second, which shows the daily variability of GHI for all the days in 2016. And Figure 4.2 presents one day's global solar radiation from 7h to 18h on 1-May-2016 and the corresponding normalized time sequence as an example. These two plots show the short time scales of this daily global solar radiation and exhibit the fluctuations stochastically distributed in time (seconds). The normalized time sequence is used as original data to the intermittency and multifractality analysis. This section gives mean sub-daily and daily multifractal pattern of global solar radiation at one station with the second dataset.

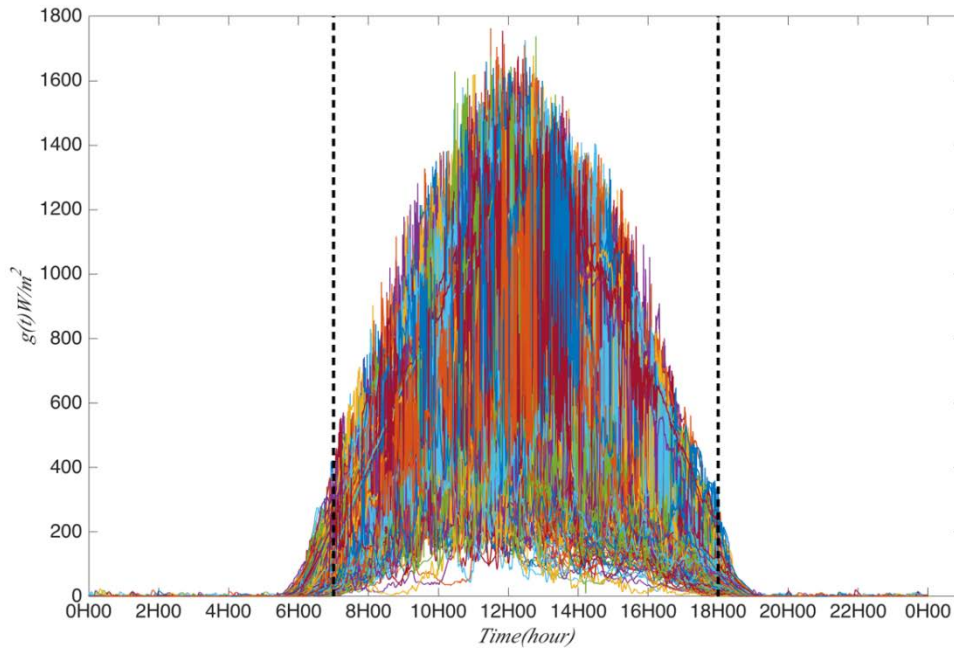
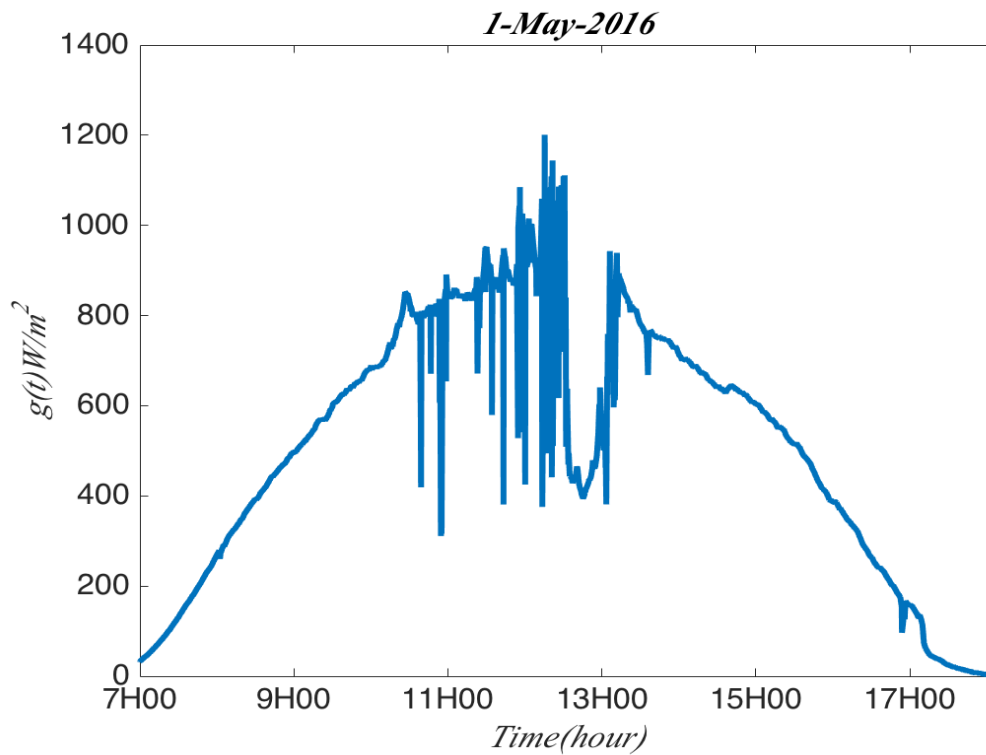


Figure 4.1 The daily global solar radiation GHI from May to December in 2016. 7h to 18h is the time period considered in this part.



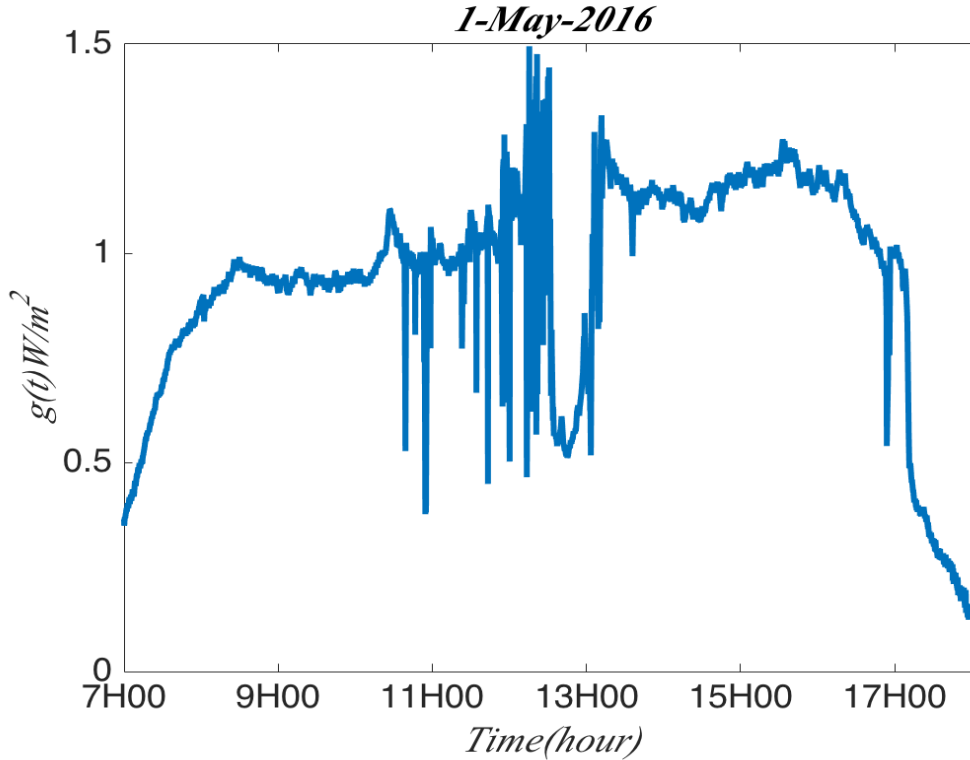


Figure 4.2 An example of typical daily global solar radiation sequence $g(t)$ (panel up) and the corresponding normalized solar radiation sequence $g_{\text{norm}}(t)$ (panel down) from 7h to 18h on 1-May-2016.

4.1.1.1 Fourier spectrum analysis: Power law and Kolmogorov spectrum

In the field of solar energy, the variability of a stochastic process as GHI is often characterized by a second order statistic such as the Fourier spectrum $E(f)$ (Lave and Kleissl, 2010). The power spectral density separates and measures the amount of variability occurring in different frequency bands. Scale invariance can be detected by computing $E(f)$. Through this Fourier spectrum analysis here, the objective is to detect if there is a law scale in the GHI time series.

For a scale invariant process, the following power law is obtained over a range of frequencies f :

$$E(f) \sim f^{-\beta}, \quad (4.1)$$

where β is the spectral exponent which contains information about the degree of stationarity or non-stationarity inside a time series (Marshak et al., 1994; Telesca et al., 2003):

- $\beta < 1$, the process is stationary,
- $\beta > 1$, the process is nonstationary,
- $1 < \beta < 3$, the process is nonstationary with increments stationary.

In Figure 4.3, the Fourier spectrum $E(f)$ of the normalized global solar radiation $gnorm(t)$ (represented in Figure 4.2: used for 01-May-2016) is given in log-log scale for four different days (01-May-2016, 03-August-2016, 01-October-2016 and 02-December-2016). These four days present the summer season (December), summer-to-winter interseason (May), winter season (August), and winter-to-summer season (October). All the spectrums for these four days display a power law behavior (the dash black line) close to the Kolmogorov spectrum ($\beta=-5/3$) for frequencies $1.07 \times 10^{-4} \leq f \leq 0.07$ Hz, corresponding to time scales $14 \leq T \leq 9346s$ (approximately 2.5h). The Fourier spectrum also displays a power law behavior close to the Kolmogorov spectrum (red line).

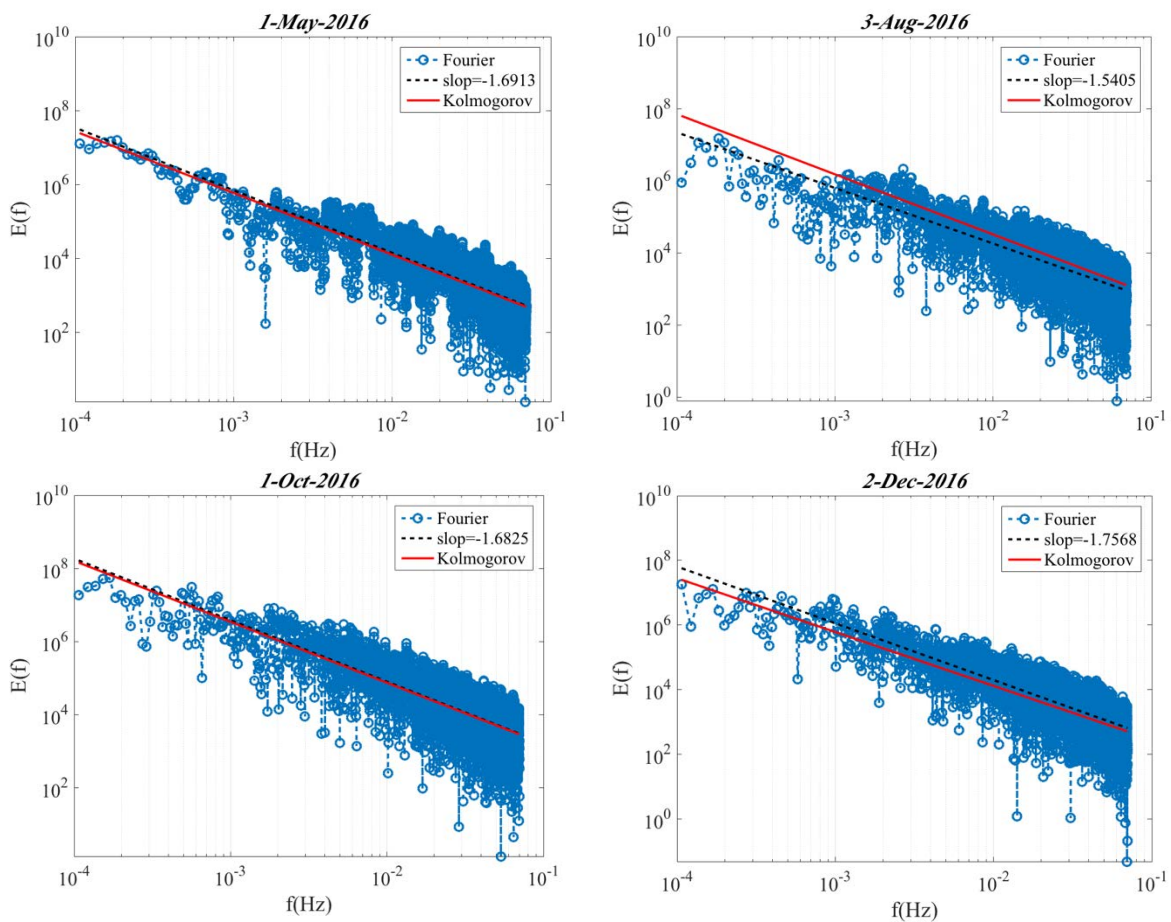


Figure 4.3 The power density spectrum $E(f)$ of the normalized time sequence on 4 different days (01-May-2016, 03-August-2016, 01-October-2016 and 02-December-2016) displaying a power law behavior (the dash black line). Fourier spectrum also displaying a power law behavior close to the Kolmogorov spectrum (red line).

Figure 4.4 illustrates the histogram of the spectral exponent β estimated separately for the whole year 2016. The idea is to emphasize that even if the spectral exponent is highly variable throughout the year as shown with 4 selected days of selected season, the distribution is

bounded and is centered around a mean value -1.68 . The spectral exponent ranges between 1.3 and 2.75. The Fourier spectrum displays a power law behavior with $1.46 < \beta < 1.86$ for 60% of the time sequences and $\beta > 2$ for 19% of sequences. It means that the sub-daily fluctuations are all the time nonstationary fluctuations ($\beta > 1$). Figure 4.5 presents the boxplot of the spectral exponent β estimated from Fourier spectra for the year 2016, which shows that the spectral exponent is centered around 1.7. Ten deciles split the spectral exponent β into 10 groups and the ninetieth percentile catches the value of the spectral exponent β is around 2.3 which separate the extremes.

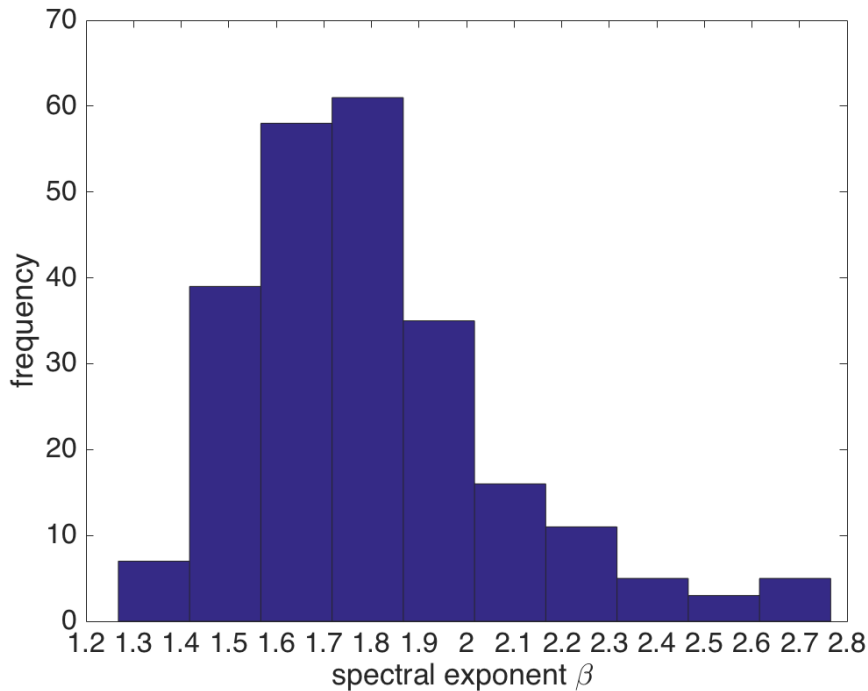


Figure 4.4 Histogram of the spectral exponent β estimated from Fourier spectra for the year 2016.

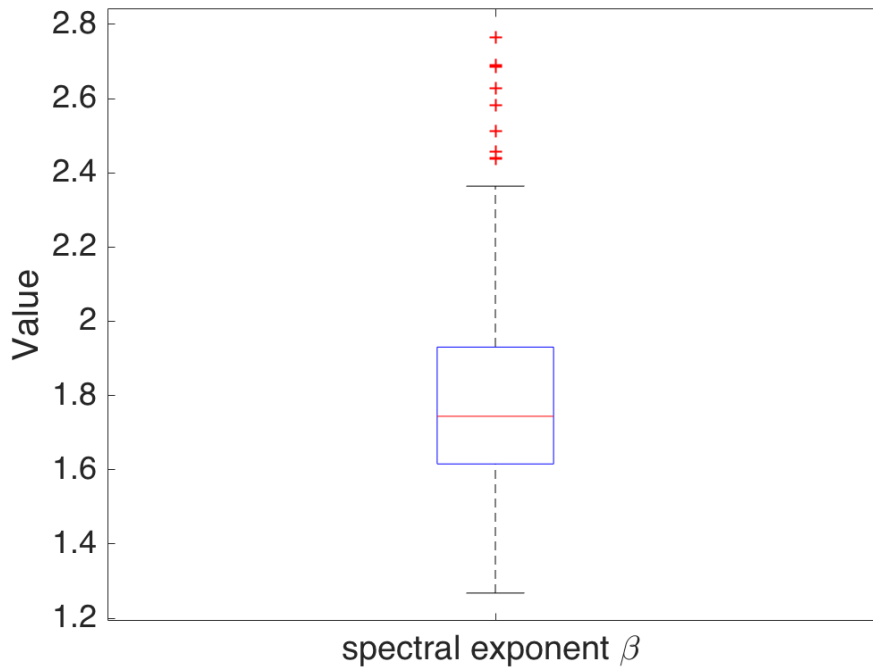


Figure 4.5 The boxplot of the spectral exponent β estimated from Fourier spectra for the year 2016.

As the analysis before, the spectral exponent β also presents a power law scale when comparing with other well-known nonstationary case as Kolmogorov shown for homogeneous turbulence ($\beta=5/3 \sim 1.66$). However, we still need to go further for checking what kind of non-stationarity we have in these sub-daily fluctuations with multifractal analysis.

4.1.1.2 Generalized scaling exponent $\xi(q)$

The arbitrary order Hilbert spectral analysis (EMD+HSA) is one of the well-known multifractal analysis techniques (structure function analysis; Multifractal detrended fluctuation analysis; wavelet analysis) and is applied in this study to normalized global solar radiation fluctuations $gnorm(t)$ for estimating the scaling exponent $\xi(q)$. The generalized exponent is obtained to see whether the sub-daily fluctuations are monofractal or multifractal processes.

First of all, EMD technique is applied to pre-process global solar radiation sequence and have a time-frequency decomposition in the IMF modes. In order to determine the energy-time frequency representation from the normalized global solar radiation fluctuations $gnorm(t)$, HSA is performed on each IMF component which is extracted from the EMD method. The instantaneous frequency and the amplitude of all the IMF modes are used to compute the joint probability density function (PDF) by the HSA. The marginal Hilbert spectrum could be defined with these frequency and amplitude:

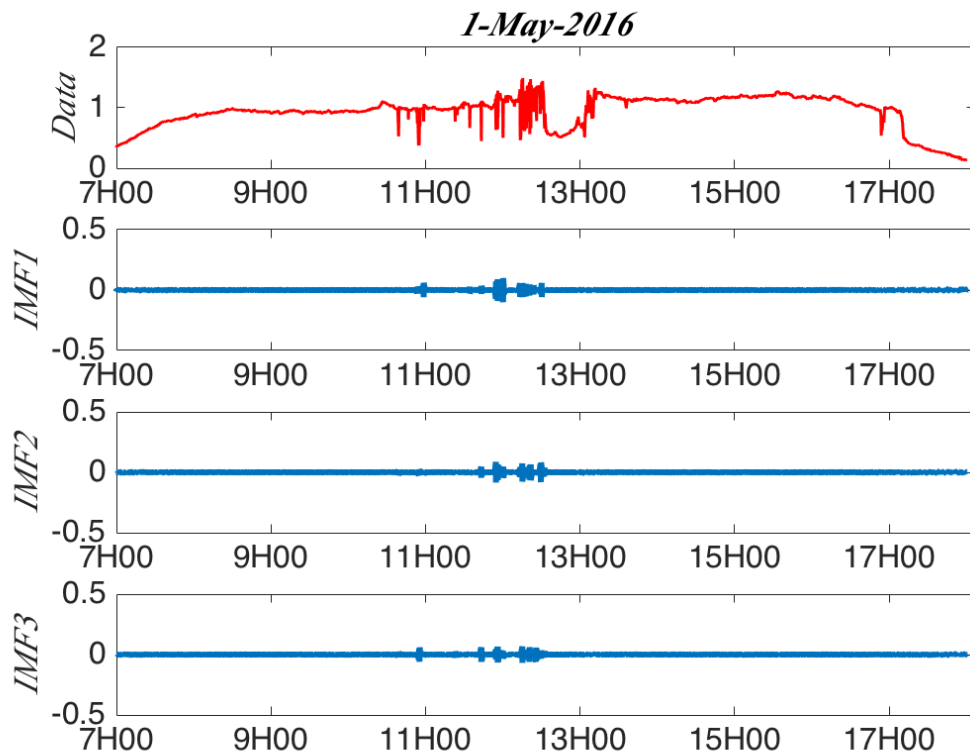
$$L_q(\omega) = \int_0^N p(\omega, A) A^q dA. \quad (4.2)$$

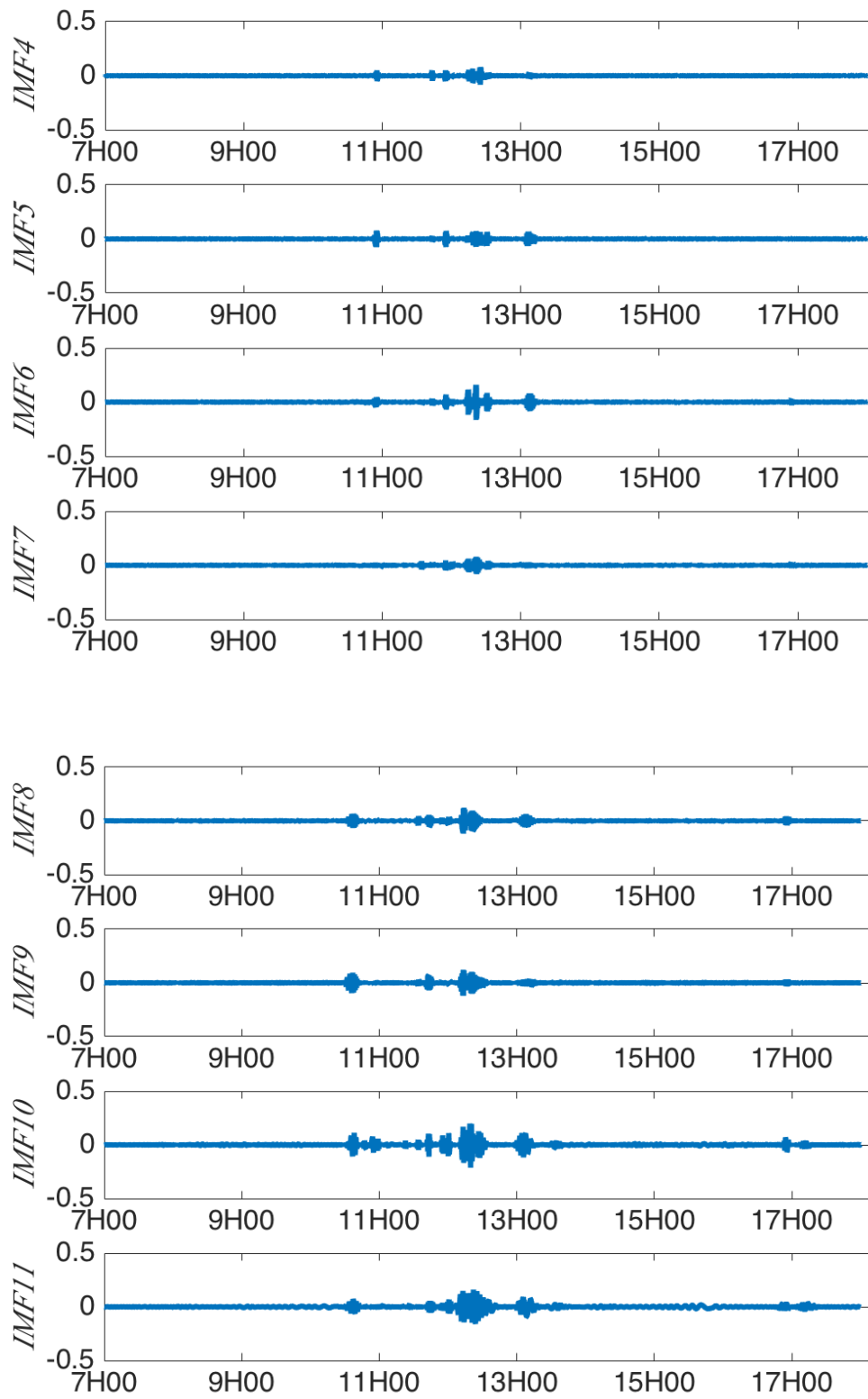
If we consider the arbitrary-order statistical moment $q \geq 0$, the scaling exponent $\zeta(q)$ can be obtained by:

$$L_q(\omega) \sim \omega^{-\xi(q)}. \quad (4.3)$$

$$\zeta(q) = \xi(q) - 1. \quad (4.4)$$

As an example, Figure 4.6 displays the EMD decomposition process into IMF modes and residual. The decomposition of global solar radiation into multiple time scales using EMD shows a decreasing of frequency scales ω_m with the mode index m . As mentioned previously, the EMD decomposition is a time-frequency decomposition which decomposes the time series into a set of high to low frequency nonstationary signal. Low frequency would reveal some daily cycle and its harmonic (e.g. residual and IMF 17). Figure 4.6 gives some idea on how intermittency triggers the GHI fluctuations.





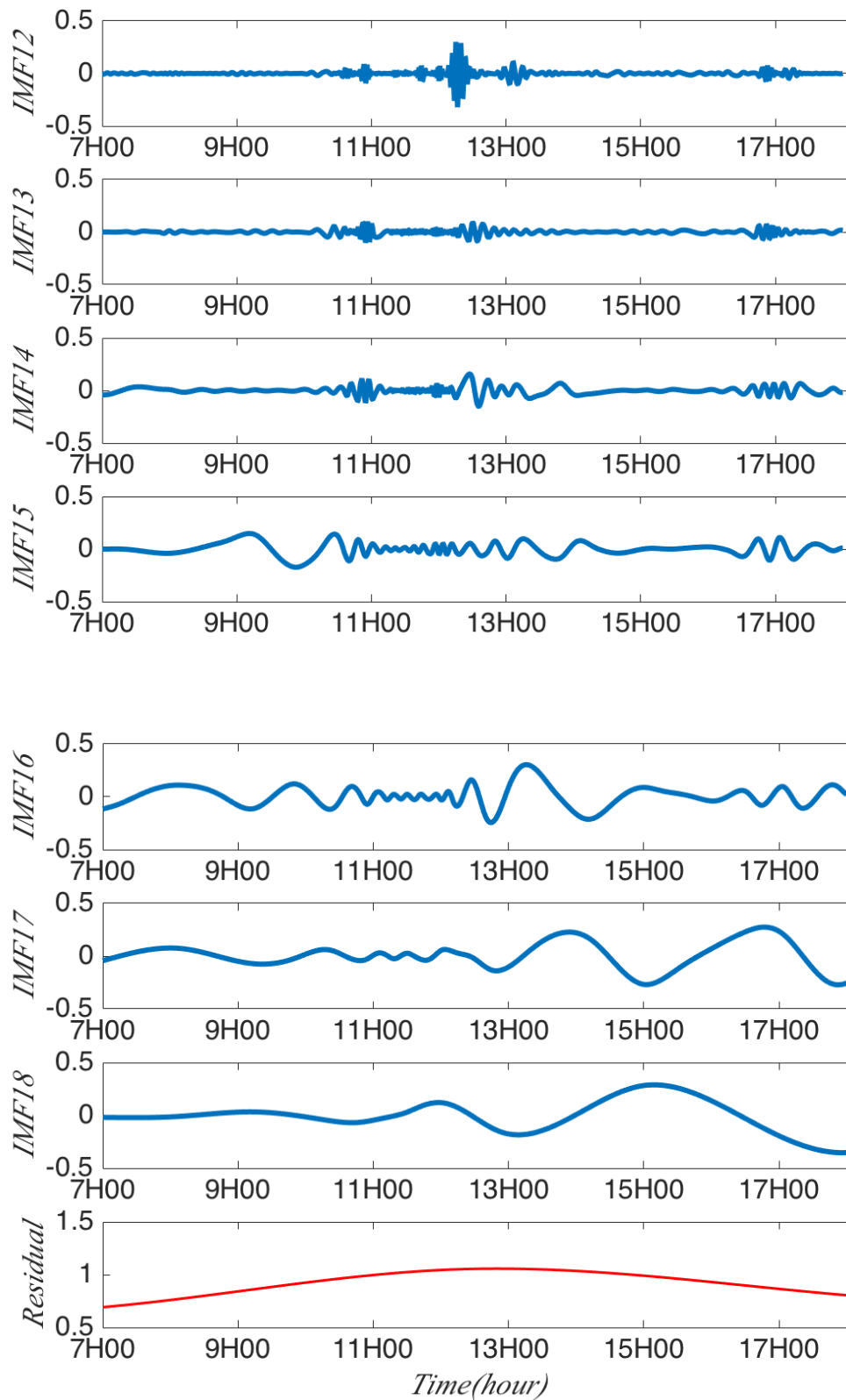


Figure 4.6 Decomposition of the signal $gnorm(t)$ on 1-May-2016 into 18 IMF modes with one residual.

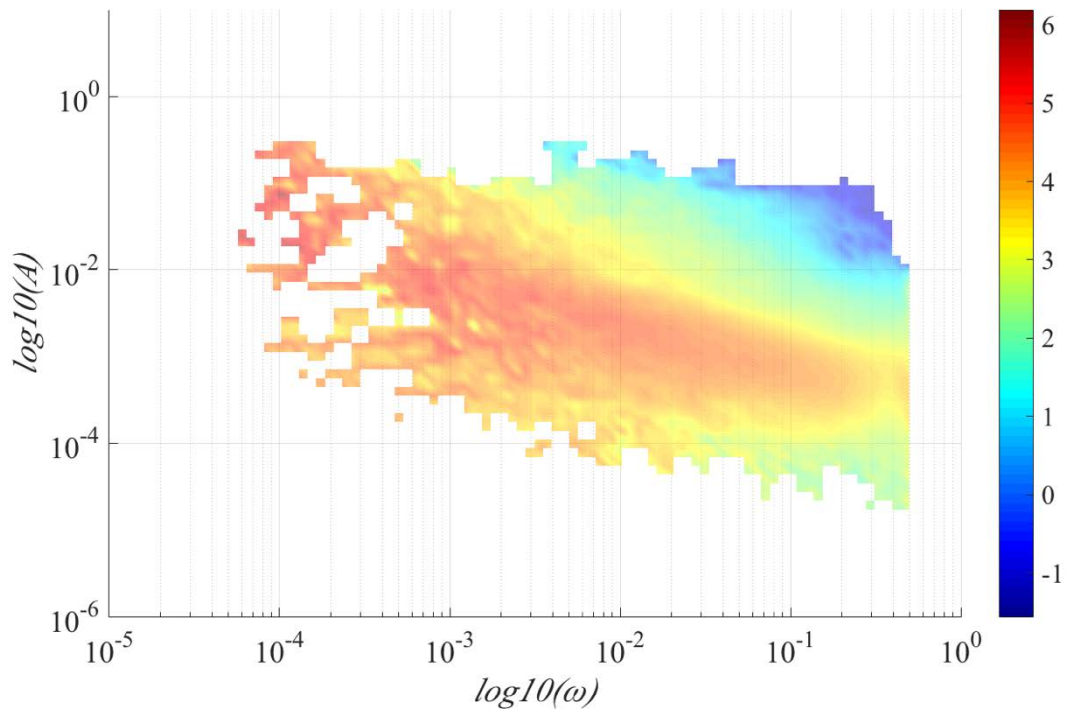


Figure 4.7 The probability density function, PDF $p(\omega, A)$ (in log scale) of the GHI recorded on 1-May-2016 in an amplitude-frequency space.

Figure 4.7 presents the PDF, $p(\omega, A)$ of the GHI record on 1 May 2016, which is from all the IMF modes. This figure is the first 2D amplitude-frequency representation of the PDF of GHI fluctuations, and it can be seen graphically that the amplitudes decrease with increasing frequencies with a scaling trend.

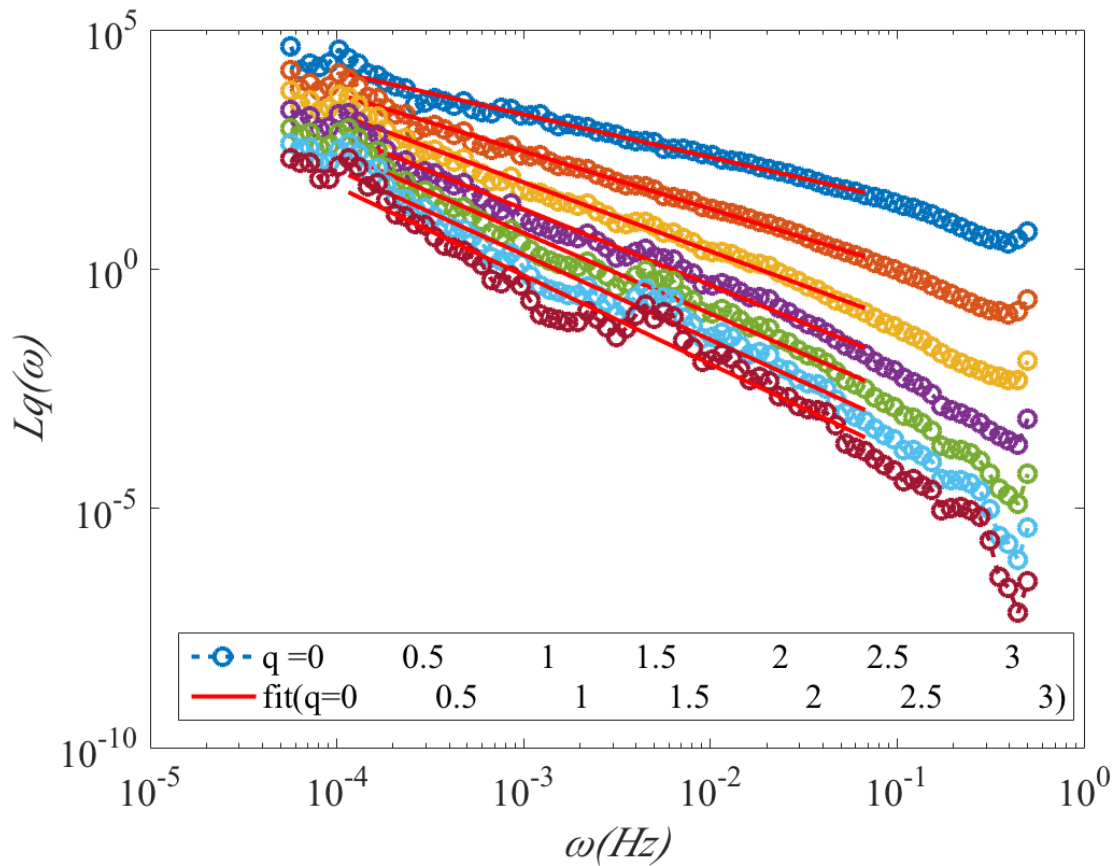


Figure 4.8 Arbitrary order Hilbert marginal spectrum $L_q(\omega)$ displaying a scaling in log-log plot in different orders of moments ($q=0, 0.5, 1, 1.5, 2, 2.5, 3$). For $0 \leq q \leq 3$ with 0.5 increment and for frequencies $1.07 \times 10^{-4} \leq f \leq 0.07$ Hz, corresponding to time scales $14 \text{ s} \leq T \leq 9346 \text{ s}$ on 1-May-2016. Power law behavior is observed on the inertial range.

For checking if there is a scaling law in the inertial range, the Arbitrary order Hilbert marginal spectrum is displaying a scaling in log-log plot in Figure 4.8. The plots in different orders of moment ($q=0, 0.5, 1, 1.5, 2, 2.5, 3$) are fitted by a least square, which indicate the scaling relationships. For frequencies $1.07 \times 10^{-4} \leq f \leq 0.07$ Hz (corresponding to time scales $14 \text{ s} \leq T \leq 9346 \text{ s}$), $q=0, 0.5, 1, 1.5, 2, 2.5, 3$ with 0.5 increment. The power law behavior is observed on the inertial range in the marginal spectrum. Each arbitrary-order statistical moment q reveals a law scale in the inertial range with a scaling exponent $\zeta(q)$ which corresponds to the slope of the straight line in the log-log plane.

The Hilbert marginal spectrum is also compared to Fourier spectrum in Figure 4.3 on 1-May-2016. Figure 4.9 shows that the Hilbert marginal spectrum can reproduce well the power law as Fourier spectrum which is the reference for spectral analysis. The Hilbert marginal spectrum could catch the same spectral feature as Fourier technique. Hilbert marginal spectrum could capture the power law behavior (marked by the vertical black lines) on the range of

frequency $1.07 \times 10^{-4} \leq f \leq 0.07$ Hz as it is done by the Fourier spectrum. This demonstrates the ability of Hilbert marginal spectrum to catch the power law inside time series.

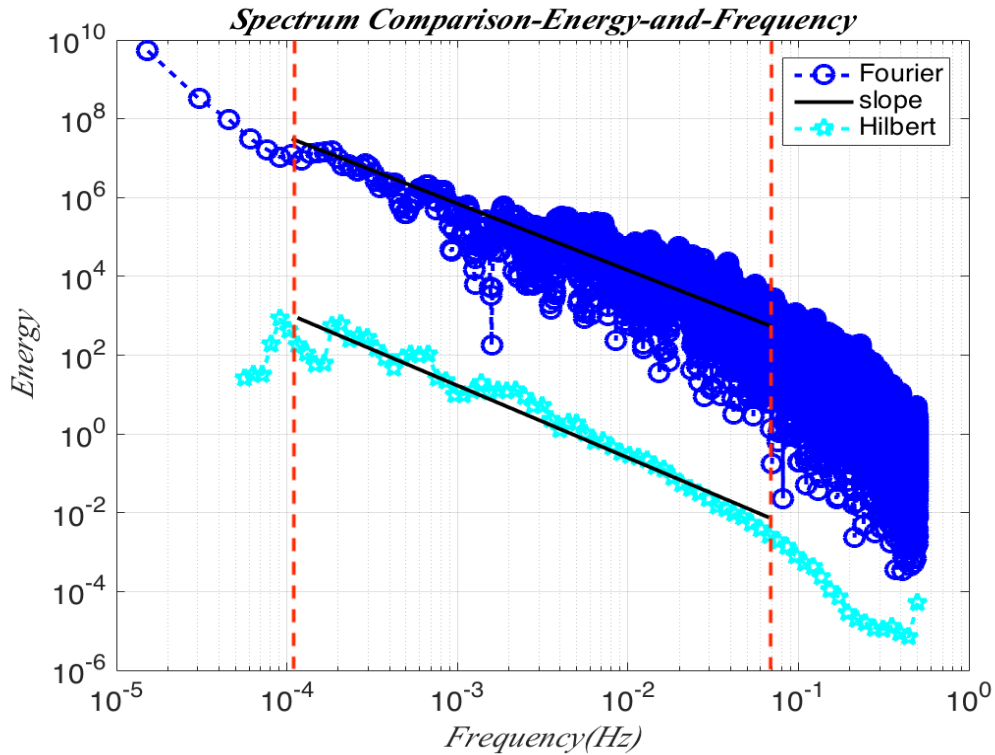


Figure 4.9 Comparison of the Hilbert marginal spectrum and Fourier spectrum for the global solar radiation at Station Moufia. A power law behavior is observed in the range of frequency $1.07 \times 10^{-4} \leq f \leq 0.07$ Hz, corresponding to time scales $14 \leq T \leq 9346$ s: this range is marked by the vertical black lines.

Figure 4.10 shows the scaling exponents $\xi(q)$ of GHI at station Moufia on 01-May-2016. This curve is calculated by the HHT and it is concave, which indicates the multifractal properties of the solar radiation. For comparison, a reference line $qH+1$ with $H = 1/3$ (dashed red line) is shown in the figure, which corresponds to monofractal processes. The scaling exponents $\zeta(q)$ in this figure is with error-bar plot, which used 95% confidence of the error at each q for the scaling exponents $\zeta(q)$. This error-bar plot gives the mean of the scaling exponents $\zeta(q)$ in the middle point with the positive and negative variance in the up and down.

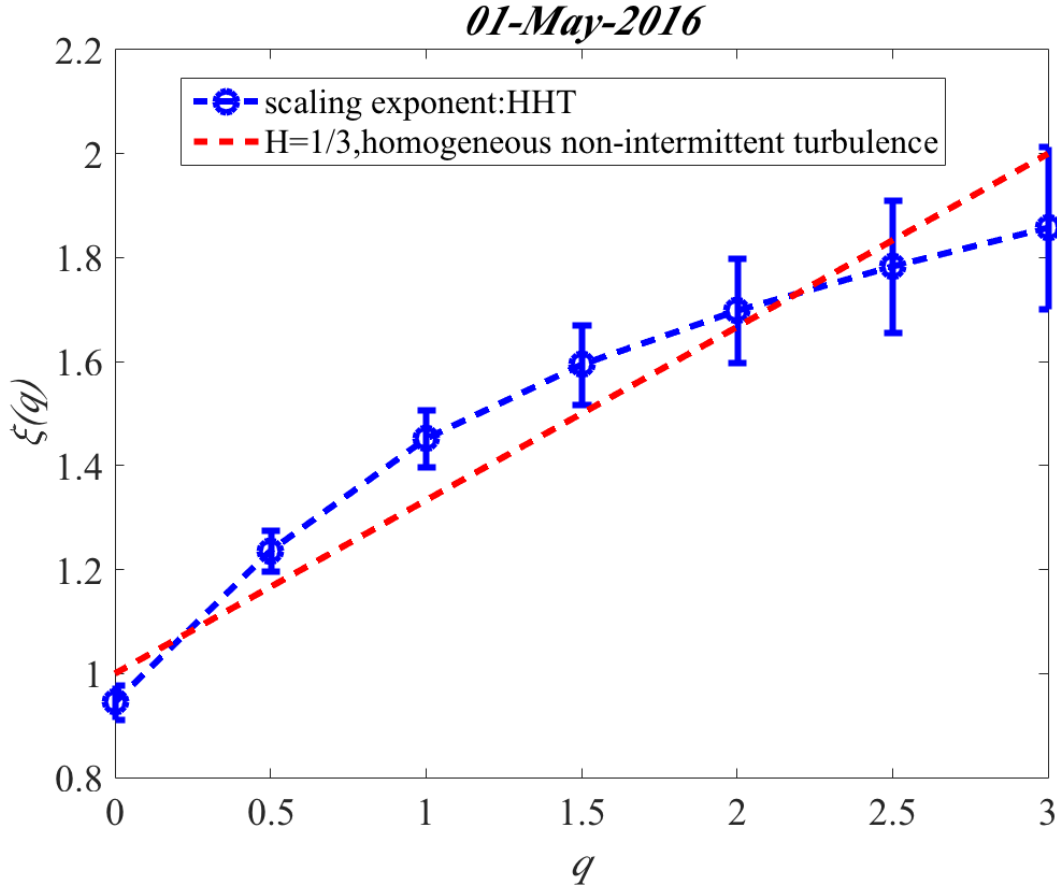


Figure 4.10 The scaling exponents calculated by the HHT compared to a non-intermittent linear model $qH+1$ with $H = 1/3$ (dashed red line) the Hurst exponent.

4.1.1.3 Log-stable model parameters H , C_1 and α

The generalized scaling exponent $\zeta(q)$ is obtained through the arbitrary order Hilbert spectral analysis in the last section. The multifractal process is detected for this sub-daily GHI fluctuation at station Moufia. Then this section is to characterize the concavity of the scaling exponent $\zeta(q)$ for analyzing the intermittency of GHI. The log-stable model is a well-known method on fitting the scaling exponent $\zeta(q)$ curve for getting the parameters describing the intermittency or multifractality. And also, these parameters: H , C_1 and α could be used after as predictor in linear forecast model. The following is the log-stable model proposed by Schertzer and Lovejoy (1987) and Kida(1991):

$$\zeta(q) = qH - \frac{C_1}{(\alpha-1)} (q^\alpha - q). \quad (4.5)$$

It is well-known that intermittency can be represented with a parametric model used in turbulence theory (Lovejoy and Schertzer, 2006; Lovejoy et al., 2001). As presented in the equations 2.33 and 2.34 in Chapter 2, the three parameters have physical meaning. Firstly, H -the Hurst parameter ($H = \zeta(1)$) defines the degree of smoothness or roughness of the field;

secondly, C_I measures the inhomogeneity mean or the mean intermittency characterizing the sparseness of the field and thirdly, which is bounded between 0 and d (d the dimension space, here $d = 1$); the multifractal Lévy parameter α measures the degree of multifractality, like how fast the inhomogeneity increases with the order of the moments. α is bounded between 0 and 2, where $\alpha = 0$ corresponds to the monofractal case and $\alpha = 2$ corresponds to the multifractal lognormal case.

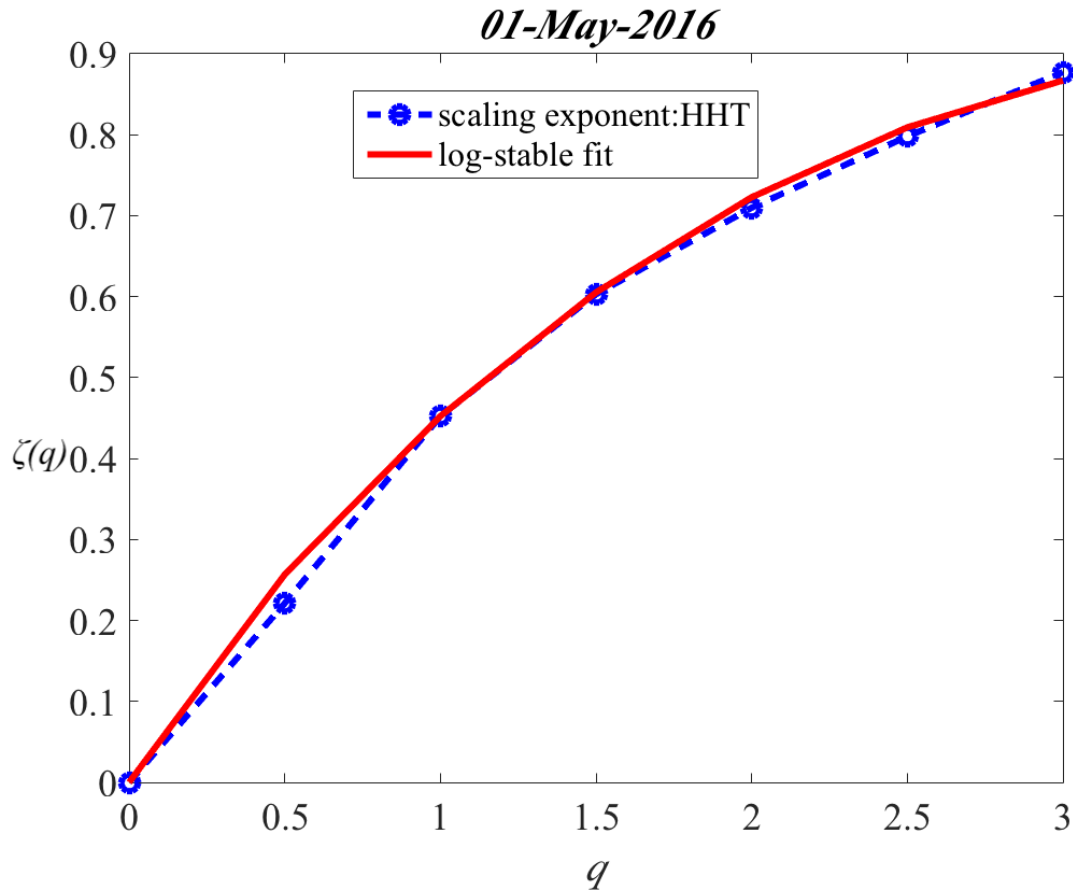


Figure 4.11 The scaling exponents (HHT) with log-stable fitting (red line) on 01-May-2016: $H=0.44$, $C_I=0.10$, $\alpha=1.66$.

Figure 4.11 shows the log-stable fitting of GHI fluctuations for 01-May-2016. Through the log-stable model, $H = 0.44$, $C_I = 0.10$, and $\alpha = 1.66$ are estimated. All these analyses are based on only one day for sub-daily GHI multifractal pattern. Then, we extend the study to all days of the year 2016 record for the daily GHI multifractal analysis, showing that the GHI fluctuations are highly variable from day-to day and thus the set of triplet (H, C_I, α) may be sensitive to this variability.

In order to analyze the distribution of the daily intermittency process, the classification method is applied to the daily GHI. Different class could present their intermittency

characterization. Many studies proposed their classification methods on solar radiation data. Muselli et al. (2000) applied a classification methodology with parameters defined from hourly clearness index profiles. Maafi and Harrouni (2003) and Harrouni et al. (2005) use fractal dimension and daily clearness index k_t as classification parameters. Three classes obtained by some specific thresholds of these parameters correspond to clear sky, partially clouded and overcast sky, respectively. Soubdhan et al. (2009) classified daily distributions of the clearness index kt by estimating a finite mixture of Dirichlet distributions without assuming any parametric hypothesis on the daily distributions in order to qualify the fluctuating nature of solar radiation under tropical climate. Bessafi et al. (2013) show interest and disadvantages of two approaches for classifying curves. The first is based on a vector representation of curves, the second offers the D'Urso and Vichi distance incorporating the mathematical properties of curves and based on the first and second finite derivatives. These two approaches are applied to the classification of sources of solar energy. In this study, the method of Bessafi et al. (2013) is chosen to do clustering GHI and analyze the intermittency in different classes. The GHI fluctuations at all scales and all classes are studied in this work. This classification would be a first step to check the variability of intermittency for each type of weather and thus the range value of (H, C_1, α) .

Here we just use the method of Bessafi et al. (2013) directly. This method uses the direct fraction noted as K_b to obtain the different classes.

$$K_b = 1 - K_d = 1 - \frac{f_{diff}}{f_{glo}} = \frac{f_{dir}}{f_{glo}}, \quad (4.6)$$

where f_{diff} means the diffuse part of the global solar radiation, and f_{glo} means the global solar radiation. Thus, when direct fraction note K_b is closed to 1, it means it is clear sky day and the surface could receive more solar energy. Otherwise, the K_b is closed to 0, which means it is cloudy day. And three data mining methodologies are applied in the classification: PCA, Ward and K-means clustering methods.

Figure 4.12 is the individual factor map for the GHI in 2016 sampling at SPN1 Moufia got from the PCA and the Figure 4.13 is the corresponding dendrogram (clustering tree) of hierarchical ascending classification performed on GHI records. The ascending hierarchical clustering creates a nested sequence of partitions of the patterns from a dissimilarity matrix, and proceeds by series of fusions of the n objects into groups (Gong and Richman, 1995). It produces a series of partitions of the data, P_n, P_{n-1}, \dots, P_1 . In this study, P_n consists of 238 single object clusters, and P_1 consists of a single group containing the 238 days. At each stage,

the ascending hierarchical clustering regroups the two clusters that are closest according to a Euclidean distance metrics. Five classes with GHI and K_b are obtained which are presented in Figure 4.14.

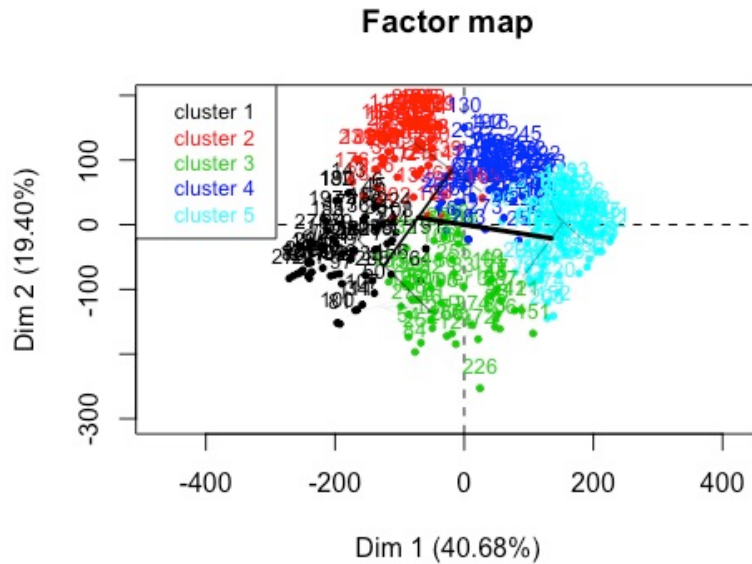


Figure 4.12 The individual factor map for the GHI in 2016 sampling at SPN1 Moufia.

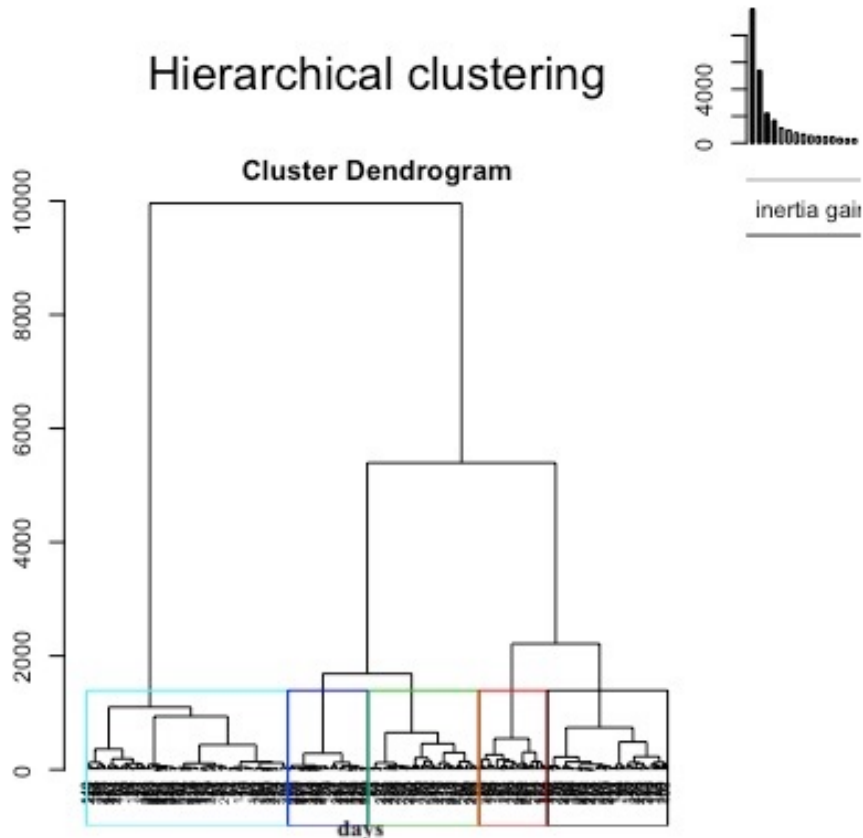


Figure 4.13 Dendrogram (clustering tree) of hierarchical ascending classification performed on GHI records on SPN1 Moufia station sampling at second in 2016.

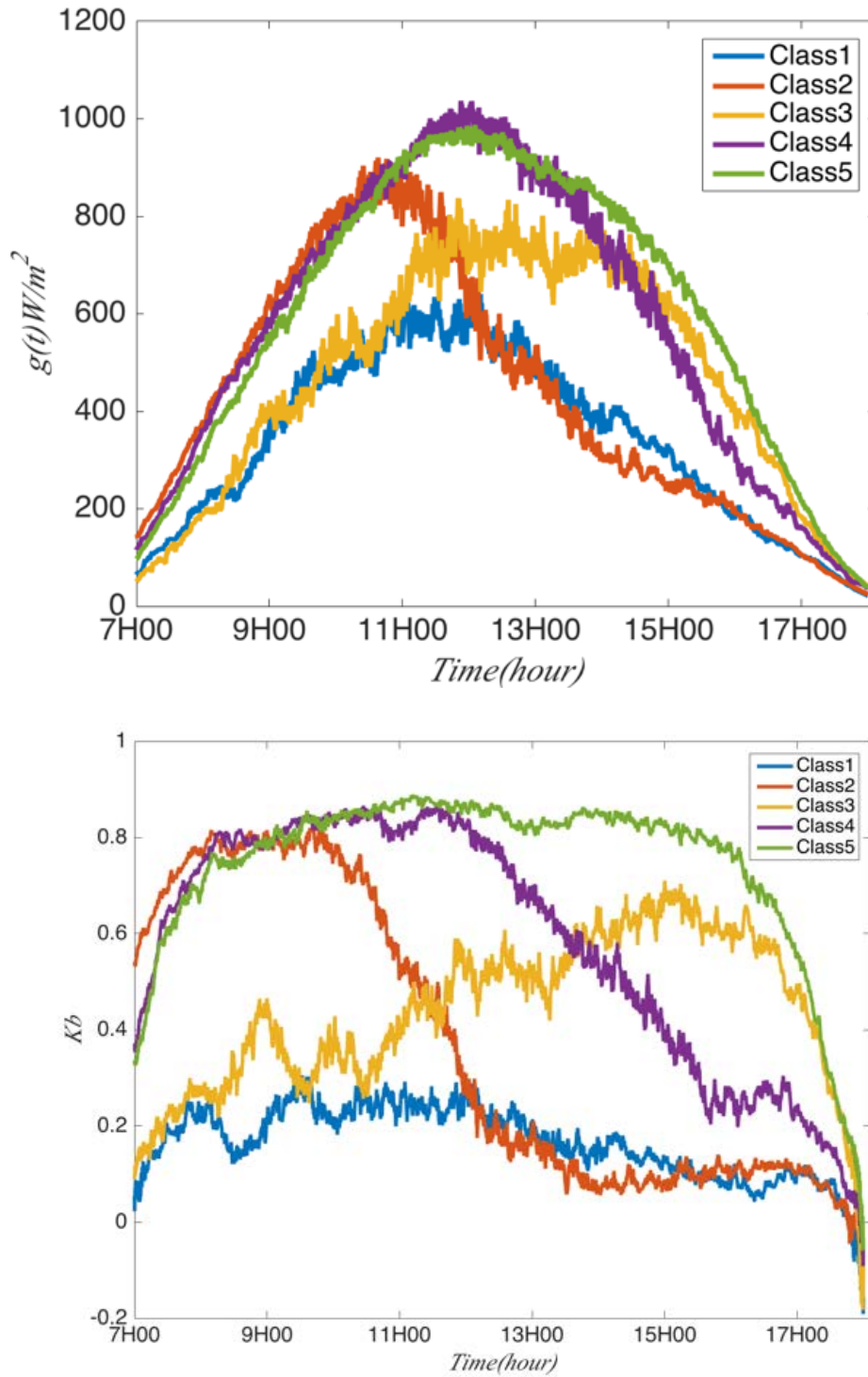


Figure 4.14 The classification for the GHI (up panel) and the K_b (down panel).

Class 1 corresponds to a very low level of sunshine all day. This class presents dominant local phenomena which include, on one hand, the weak trade winds accompanied by a flow of

moisture leading to significant effects of orographic clouds, and, on the other hand, the land breeze phenomenon induced by thermal contrasts (Badosa et al., 2013).

Class 2 has a sunny beginning until mid-morning around 09:00 - 09:30 AM and a cloudy afternoon. Diffuse radiation is dominant in the afternoon while the direct component is more important in the morning.

Class 3 corresponds to the days with a huge variability with the variable weather.

The performance of Class 4 is similar to that of Class 2, but with a stronger sunny regime during all morning until early afternoon. DHI predominates later in Class 4 than in Class 2.

Class 5 corresponds to a regime of good weather throughout the day. Intermittent clouds passing over the station do not have a systematic character since direct radiation dominates in this class.

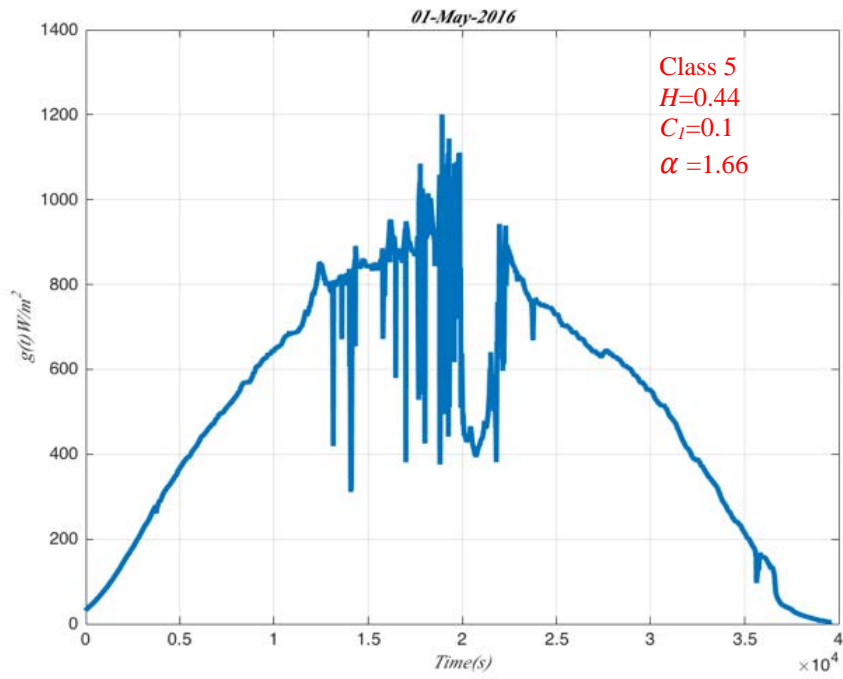
The multifractal parameters (H , C_1 , α) then could be classified into these five classes as in Table 4.1. The mean of the multifractal parameters (H , C_1 , α) in five classes are presented in the table. The mean value of H for class 1 (0.44) is larger than other classes and it is more closer to 0.5, indicating that GHI fluctuations in five classes are anti-persistent in average and the class 1 is more closer to random process. The mean values of C_1 are all under 0.1 for the five classes and the class 2 has larger C_1 than other classes indicating that the days in class 2 are more intermittent. $\alpha \in [1, 2]$ for class 1 and 2 indicate that those days are with multifractals with Levy generators and unbounded singularities.

Table 4.1 The mean of the multifractal parameters (H , C_1 , α) in five classes

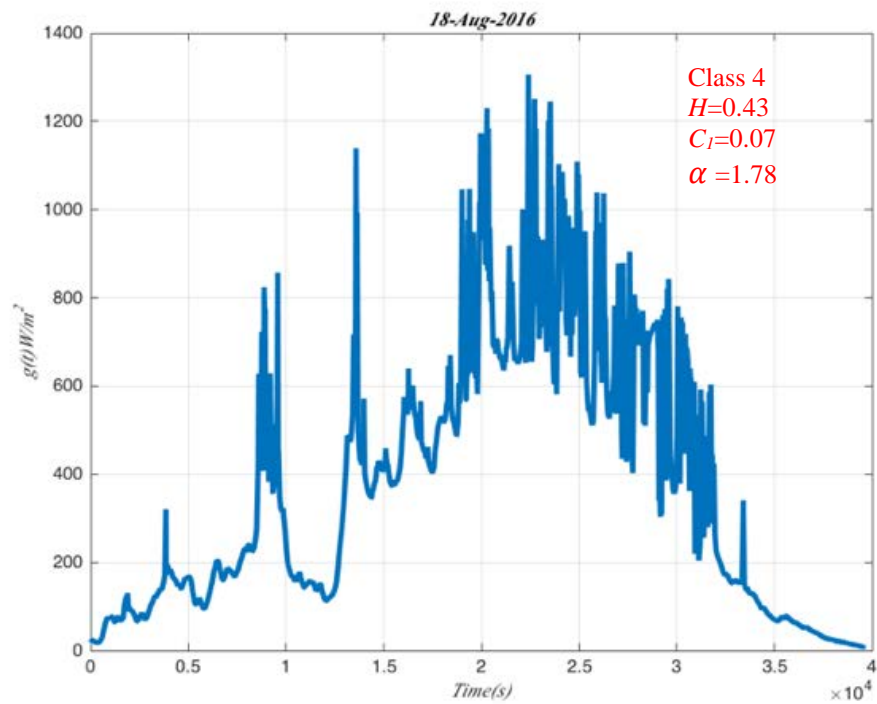
| | Class 1 | Class 2 | Class 3 | Class 4 | Class 5 |
|----------|---------|---------|---------|---------|---------|
| H | 0.44 | 0.37 | 0.34 | 0.37 | 0.35 |
| C_1 | 0.08 | 0.09 | 0.03 | 0.06 | 0.07 |
| α | 1.80 | 1.66 | 2.4 | 2.1 | 2.1 |

Therefore, Figure 4.15 illustrate the effect of three parameters: the H , C_1 , and α got from the log-stable model with the scaling exponent separately, to present the intermittency and multifractal properties of GHI. According to the classification, the days in Figure 4.15a) is sunny day (01-May-2016); the days in Figure 4.15b) and 4.15d) are sunny in the morning and cloudy in the afternoon (18-August-2016 and 09-May-2016); the day in Figure 4.15e) is cloudy day (18-December-2016), the day in Figure 4.15c) is cloudy in the morning and sunny in the afternoon (03-November-2016).

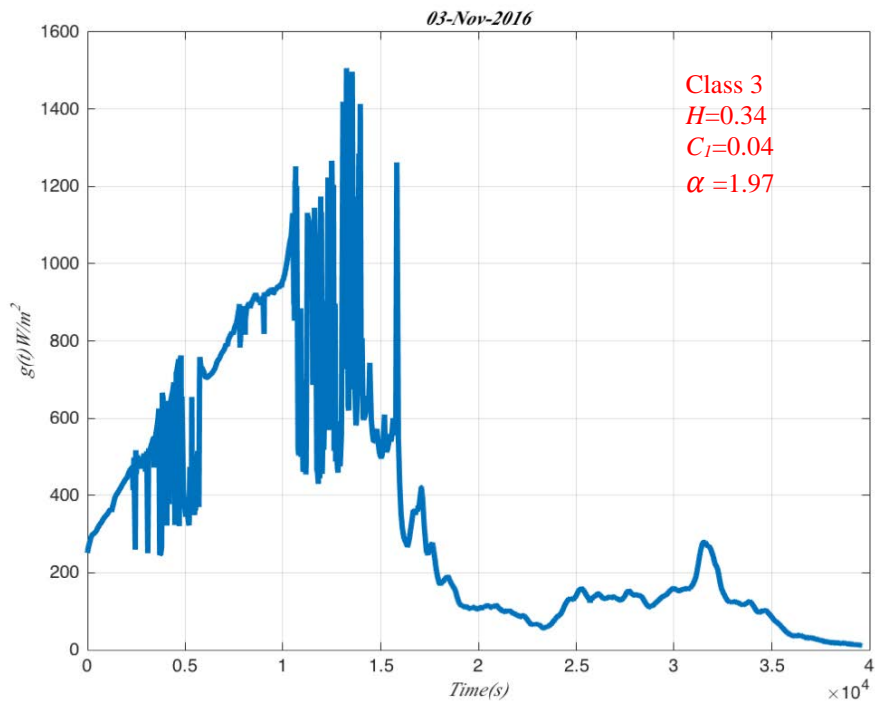
a)



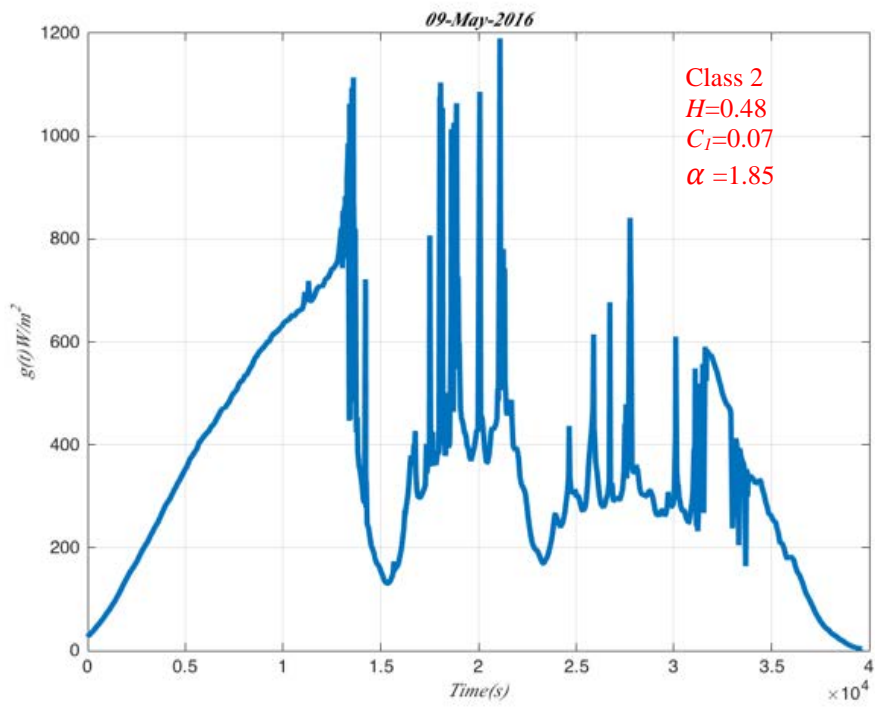
b)



c)



d)



e)

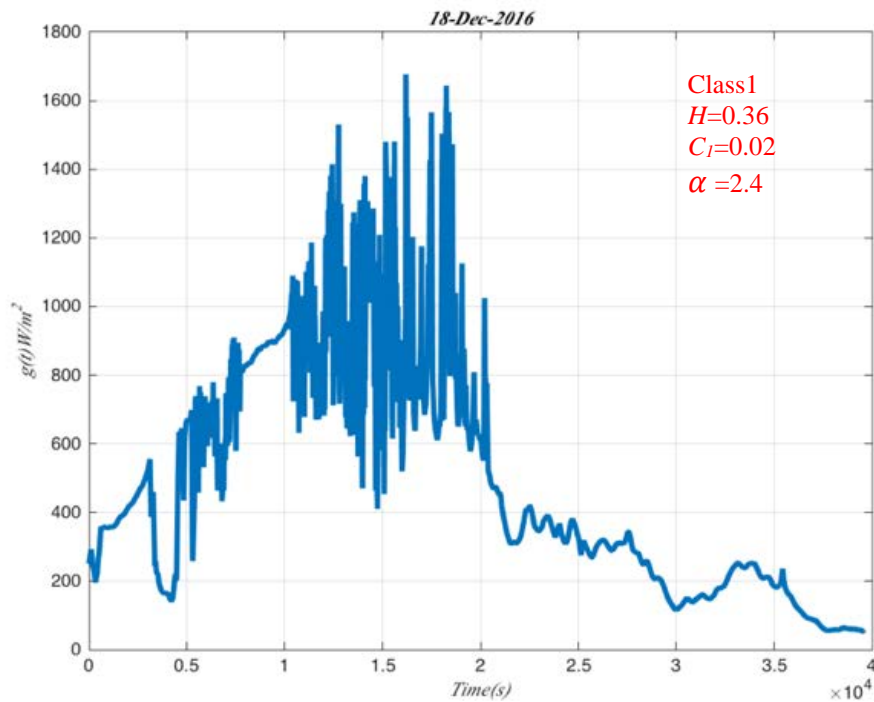


Figure 4.15 A few days of daily GHI sequences illustrating the varying effect of multifractal parameters, H , C_1 , and α .

The distribution for these three parameters in 2016 (238 days: May to December) is presented in Figure 4.16, which could be the statistics for the GHI in 2016. The values of H , C_1 , and α are grouped by frequency and intensity. The error-bar is also plotted in this figure to show the mean variation of these three parameters H , C_1 , and α . We make the distribution of these three parameters H , C_1 , and α comparing to the Beta and Gamma distribution. It has been found that the Hurst exponent H and parameter C_1 closer to Beta distribution, and the parameter α closer to Gamma distribution. Husak et al. (2006) used the gamma distribution to represent monthly rainfall in Africa for drought monitoring applications, which proved that the gamma distribution is suitable for roughly 98% of the locations over all months. Kadilar and Çakmakyapan (2016) proposed a new probability distribution to model the wind speed data and the Poisson-Gamma distribution modelled the data sets better. The beta-P distribution is examined for modelling hydrologic events by Murshed et al. (2018). Gamma distribution and Beta distribution are found in the rainfall, wind power and flood field. Here, it is excited to find that the universal parameters H , C_1 , and α for GHI give the Gamma distribution and Beta distribution in the solar energy field.

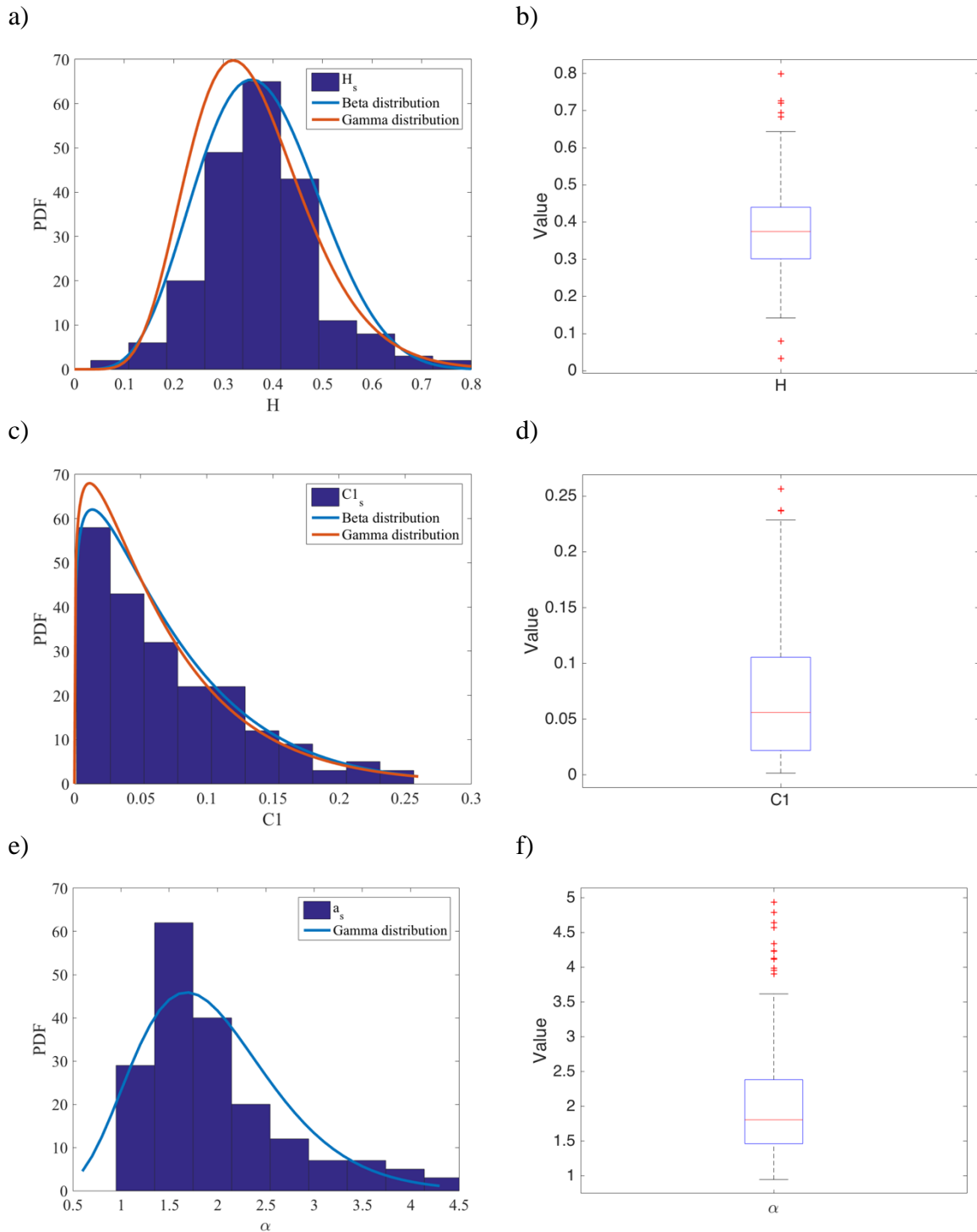


Figure 4.16 Distribution of parameter H , $C1$, and α for the 238 days of GHI in 2016 for Moufia station and the box-bar for each parameter.

4.1.2 Seasonal variability of daily multifractal pattern

After checking the sub-daily and daily GHI multifractal pattern, it would be interesting to see the GHI variability throughout the year and especially through the seasonal cycle (month by month). This section is to show the multifractal fluctuation of the GHI through the day month by month.

4.1.2.1 Summer, winter and interseason characteristics

As mentioned before, Reunion Island is a tropical island on the south hemisphere and months from November to April are in austral summer (the warm and humid season), May to October are in austral winter (the fresher and drier season). Normally the austral winter comes in July and August, and the austral summer reaches its peak in January and February. Because second data is only available from May to December in 2016 at station Moufia, the analysis on seasonal variability of daily multifractal pattern presented depends on these dataset (month by month) only. Figure 4.17 shows the distribution of monthly GHI from May to December in 2016.

In Figure 4.17, the lowest monthly value of global solar radiation appears in June and the highest in December (missing the data from January to April sampling on second).

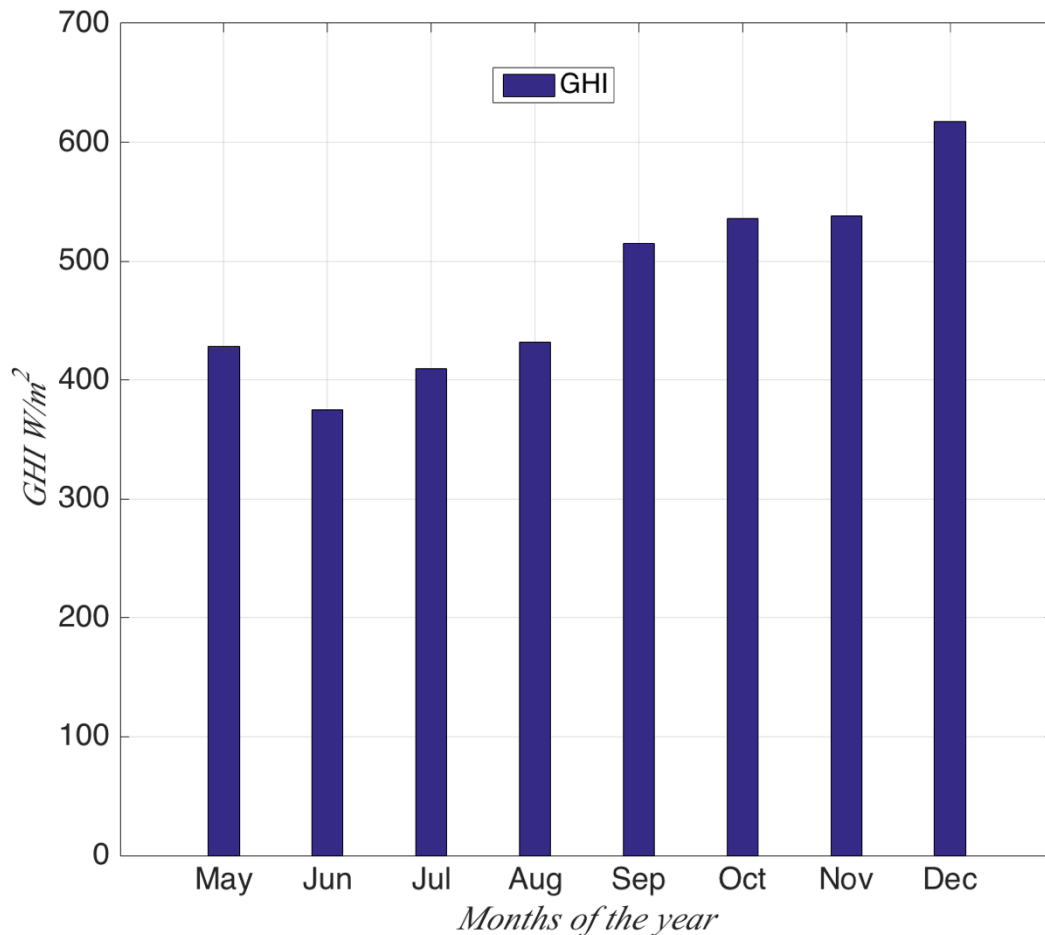
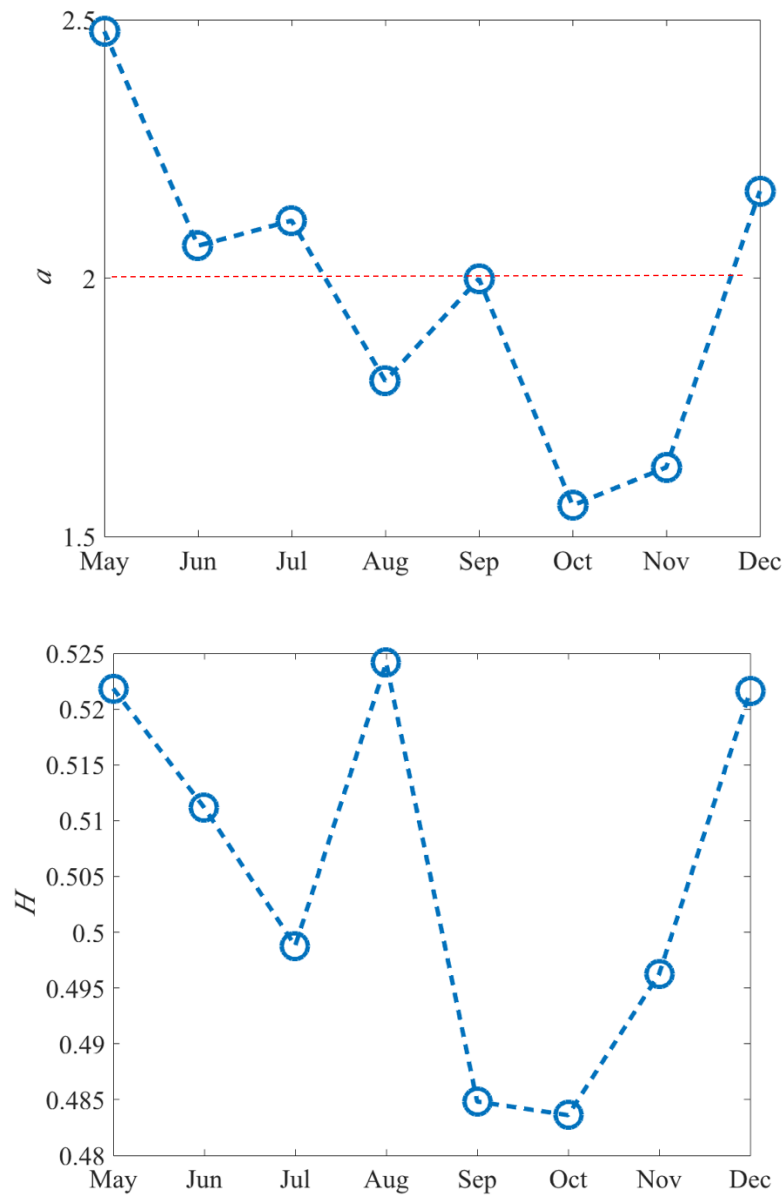


Figure 4.17 Monthly mean global solar radiation from May to December during 2016 (dataset sampling on second).

4.1.2.2 Month by month focus

Figure 4.18 shows the change of the universal multifractal parameter H , C_1 , and α in each month from May to December in 2016. Firstly, the monthly change for these three parameters indicates the seasonal variability of the multifractal pattern for GHI. The monthly mean of H in July, September, October and November are lower than 0.5 (anti-persistent), other months are bigger than 0.5 (persistent). $\alpha \in [1, 2]$ during August to November indicates that those months GHI are multifractal with Levy generators and unbounded singularities. The monthly C_1 are quite small in May and December which are inter-season months, normally they are more intermittent, but here they are more homogeneous for GHI than other months.



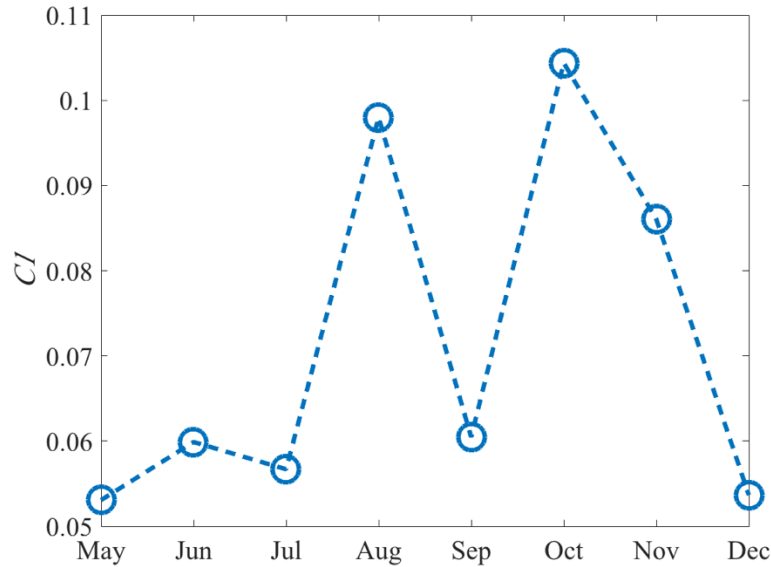


Figure 4.18 Universal multifractal parameter H , C_1 , and α in each month from May to December in 2016.

4.2 Intermittency and multifractality analysis of SPN1 network

The first part of this chapter focuses on GHI multifractality on one SPN1 station. This part we turn to the spatial pattern of GHI intermittency over the island using GHI from 11 stations in 2014. The seasonal spatial pattern of intermittency in daily scale is also studied over Reunion Island.

4.2.1 Spatial pattern of intermittency within the daily scale

4.2.1.1 Regionalization of log-stable parameters H , C_1 and α

Based on the analysis of the universal multifractal parameter H , C_1 , and α for the global solar radiation's intermittency, this part focuses on the regionalization of daily solar radiation's intermittency through the parameters H , C_1 and α .

Table 4.2 Universal multifractal parameter H , C_1 , and α value in each station in 2014

| | Stations | H | C_1 | α |
|----|------------------------|------|-------|----------|
| 1 | MOUFIA UR | 0.34 | 0.18 | 1.02 |
| 2 | MOUFIA BOIS DE NEFLES | 0.27 | 0.12 | 0.92 |
| 3 | LA POSSESSION | 0.27 | 0.11 | 1.07 |
| 4 | SAINT LEU | 0.27 | 0.11 | 1.10 |
| 5 | SAINT PIERRE LA VALLEE | 0.32 | 0.17 | 1.00 |
| 6 | SAINT JOSEPH | 0.31 | 0.13 | 1.05 |
| 7 | SAINTE ROSE | 0.34 | 0.15 | 1.03 |
| 8 | SAINT ANDRE | 0.19 | 0.07 | 1.65 |
| 9 | CILAOS PISCINE | 0.41 | 0.27 | 0.71 |
| 10 | BRAS PANON | 0.29 | 0.13 | 0.98 |
| 11 | TAMPON UR | 0.34 | 0.20 | 0.77 |

Table 4.2 presents the universal multifractal parameter H , C_1 , and α value in each station in 2014. The properties of the intermittency could be found from this table. Firstly, when we check the value of H , it could be found that all the H is less than 0.5, which indicates that the GHI fluctuations from day to day are anti-persistent in average. The value of H at Cilaos station is close to 0.5, indicating that the GHI fluctuations there are close to Brownian and random process.

The parameter C_1 is the fractal co-dimension measuring the mean intermittency: the larger C_1 , the more the signal is intermittent. Furthermore, $0 < C_1 < d$ with d the dimension space (here $d = 1$) (Calif et al., 2013). All the C_1 in the table 4.2 are around 0.15 and the Cilaos station gives 0.27, which means GHI over Cilaos is more heterogeneous than other place.

Lovejoy and Schertzer (2006) state that α values define five qualitative cases: 1) $\alpha = 2$ defines multifractals with Gaussian generators; 2) $\alpha \in [1, 2]$ defines multifractals with Levy generators and unbounded singularities; 3) $\alpha = 1$ defines multifractals with Cauchy generators;

4) $\alpha \in (0,1)$ defines multifractals with Levy generators and bounded singularities; 5) $\alpha = 0$ defines monofractal processes. In Table 4.2, we find that the stations at Saint Pierre, Moufia, La Possession, Saint Joseph, Saint Rose and Bras Panon exhibit α values close to 1 which means that the GHI fluctuation there is following a Cauchy process. The station Saint Andre, Saint Leu, Cilaos and Le Tampon are closed to Levy process with unbounded or bounded singularities.

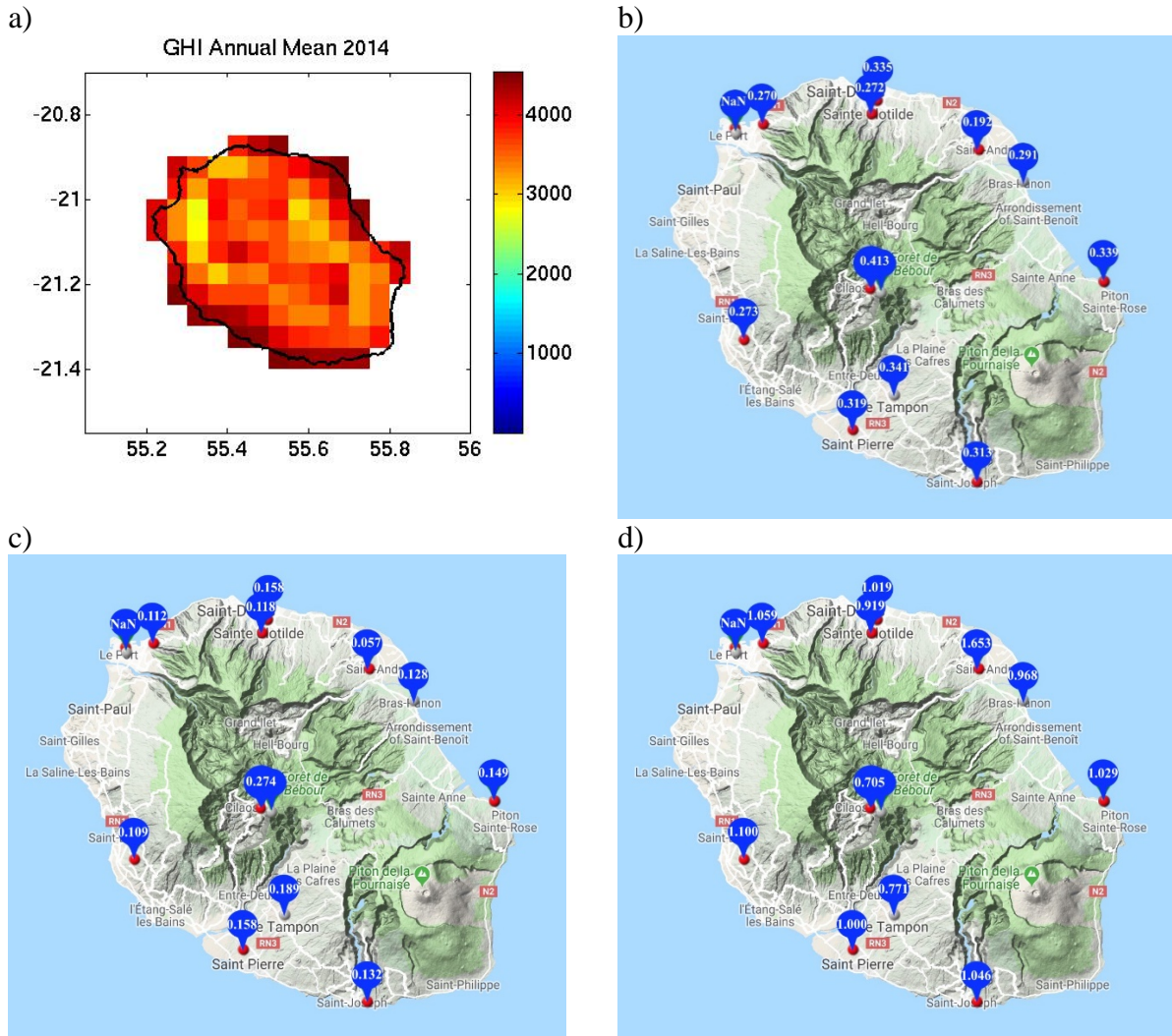


Figure 4.19 a) the annual mean of global solar radiation in 2014 over Reunion Island (W/m^2) calculated from CM SAF (SARAH-E@5km); b) the value of parameter H for 11 stations in 2014; c) the value of parameter C_1 for 11 stations in 2014; d) the value of parameter α for 11 stations in 2014.

It could be observed that 11 stations have different values of parameter H , C_1 , and α value in 2014, which firstly indicates that the daily global solar radiation's intermittency varies over space. There are two stations located over Moufia, but one (Moufia UR) has higher altitude than the other one (Moufia Bois de Nefles). These two stations have large differences for parameter H , C_1 , and α value, which could be explained by the fact that there are more clouds

cover in the mountain than closest to sea, so more multifractal or intermittency in the station Moufia Bois de Nefles normally. The station La Possession and Saint Leu which are located in the west of the island have close H , C_I , and α value in the Table 4.2. Western coast of Reunion Island usually obtain more surface solar radiation than the inland area as shown in Figure 4.19a): the maximum value of surface solar radiation annual mean in 2014 is 4842 W/m^2 over the coastal area. The southern stations which includes Siant PierreI La Vallee, Saint Joseph show the close H , C_I , and α value. The station in the Cilaos cirque (Cilaos PiscineI) gives largest (smallest) H value (α value), so it has roughness field characters and shows less degree of multifractal. The station Saint Andre has smallest H value (0.19) and largest α value (1.65), indicating the strong intermittency in this location. It is corresponding to the climate that the eastern part of the Reunion Island has more cloud weather and more precipitation than other places. Because the station Saint Andre is exposed to easterly trade winds that accelerate along the coast line and the daytime sea breeze force the moist sea air up on the slopes, frequently resulting in cloud formation (Badosa et al., 2013).

4.2.1.2 Singularity spectrum width by region

Another way to characterize a multifractal process is the singularity spectrum related to the scaling exponents $\zeta(q)$. The width of the singularity spectrum measures the multifractality level of the signal. The bigger amount of multifractality is present in the signal, the width of singularity spectrum becomes wider. The width of the multifractal spectrum reflect the temporal variation of the Hurst exponent (Ihlen, 2012).

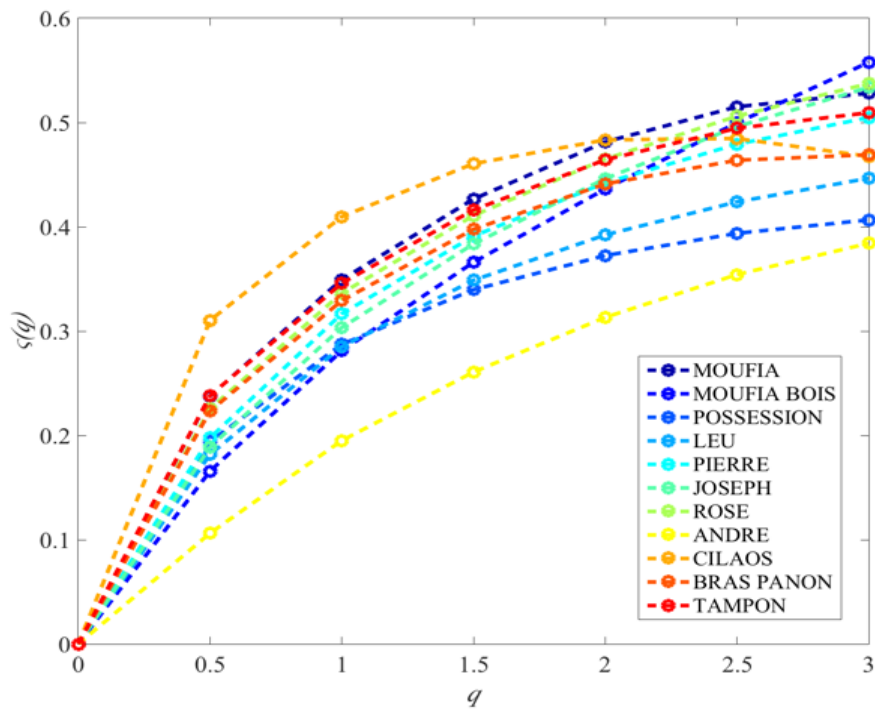


Figure 4.20 The scaling exponents $\zeta(q)$ calculated by arbitrary order Hilbert spectral analysis (EMD+HSA) for each station in 2014.

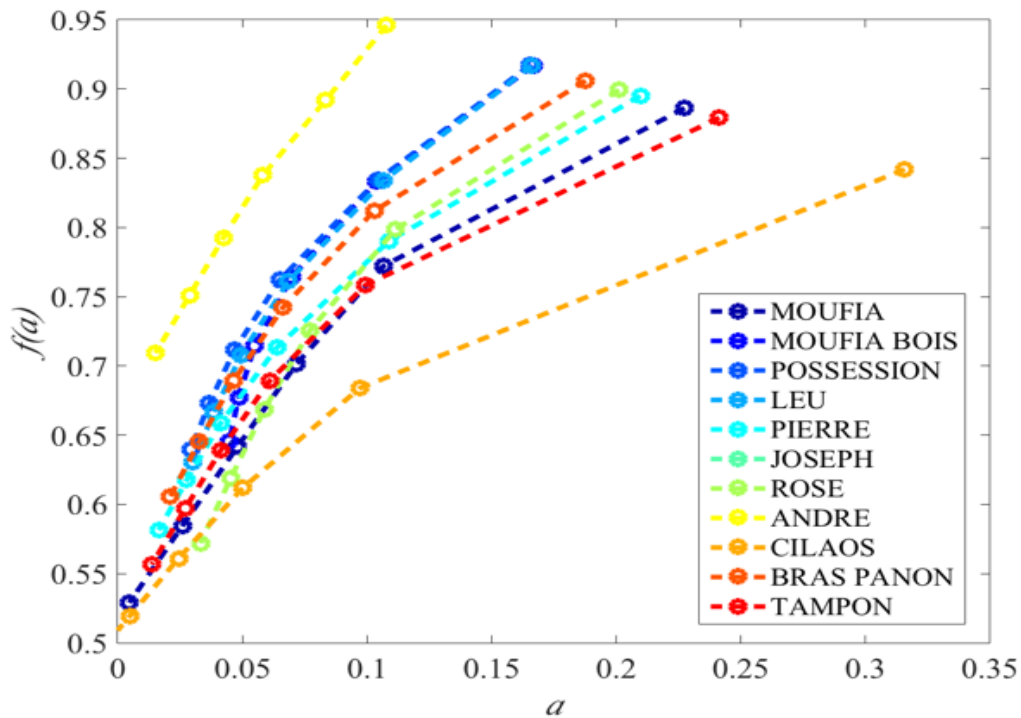


Figure 4.21 The singularity spectrum corresponding to the scaling exponents $\zeta(q)$ in Figure 4.22 for each station in 2014.

Table 4.3 The width of the singularity spectrum for each station in 2014

| Station | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 |
|---------|------|------|------|------|------|------|------|------|------|------|------|
| WSP | 0.22 | 0.12 | 0.14 | 0.14 | 0.19 | 0.17 | 0.17 | 0.09 | 0.33 | 0.17 | 0.23 |

***WSP: width of the singularity spectrum**

Figure 4.20 and 4.21 present the scaling exponents $\zeta(q)$ and the singularity spectrum separately for the 11 stations in 2014. For moments order $0 \leq q \leq 3$, all the scaling exponents functions observed in Figure 4.20 are nonlinear and concave and the corresponding singularity spectrum also show these concave characters, which highlights the multifractal and intermittent properties for the global solar radiation in the different regions over Reunion Island. Through the calculation of the width of the singularity spectrum for each station in 2014 in Table 4.3. The station 9 (Cilaos) has the largest width of the singularity spectrum (0.33) and the station 8 (St Andre) has the smallest width of the singularity spectrum (0.09), which indicates the strongest and weakest multifractality separately in this station network. It also reveals that station 3 (La Possession) and 4 (St Leu) present similar multifractality for they have the same width of the singularity spectrum (0.14). This is reasonable as La Possession and St Leu locate on the western side of the island, receiving more stable GHI relatively.

4.2.2 Seasonal Spatial pattern of intermittency within the daily scale

4.2.2.1 Consistency of regionalization of log-stable parameters H , C_1 and α through the year

Based on the 11 stations' spatial pattern analysis of intermittency by the universal multifractal parameter H , C_1 , and α , the intermittency of surface solar radiation for the seasonal scale over Reunion Island could also be presented spatially. This section is to study whether the intermittency indicated by the universal multifractal parameters could be consistent throughout the year.

Table 4.4, 4.5 and 4.6 present the value of parameter H , α , and C_1 at each station in each month of 2014 separately. Firstly it is observed that austral summer season (November-April) for each station has different value range (parameter H , C_1 , and α) with austral winter season (May-October). For parameter H , each station in winter has smaller value than in summer in general. As stated before, the smaller H , the smother for the solar radiation signal, which defines the degree of the roughness of the signal.

The Hurst exponent could classify time series into three types: 1) a Brownian time series, also known as a random walk. The Hurst exponent is close to 0.5 is indicative of a Brownian time series; 2) an anti-persistent time series, also known as a mean-reverting series, an increase will most likely be followed by a decrease or vice-versa. A Hurst exponent value between 0 and 0.5 is indicative of anti-persistent behavior and the closer the value is to 0, the stronger is the tendency for the time series to revert to its long-term means value; 3) a persistent time series. In a persistent time series an increase in values will most likely be followed by an increase in the short term and a decrease in values will most likely be followed by another decrease in the short term. A Hurst exponent value between 0.5 and 1.0 indicates persistent behavior; the larger the H value the stronger the trend.

Firstly, for the Hurst exponent H in the Table 4.4 at 11 stations in different month, all the value H are less than 0.5. The station Moufia (0.48), Saint Joseph (0.47) and Cilaos (0.47) in December, January and February give H value close to 0.5, indicating that the process path is similar to that of a Brownian motion. All the other stations in each month throughout the year are belong to the anti-persistent process.

As we stated in 4.2.1, there are five classes for the parameter α : 1) $\alpha = 2$ defines multifractals with Gaussian generators; 2) $\alpha \in [1, 2]$ defines multifractals with Levy generators and unbounded singularities; 3) $\alpha = 1$ defines multifractals with Cauchy generators; 4) $\alpha \in (0, 1)$ defines multifractals with Levy generators and bounded singularities; 5) $\alpha = 0$ defines monofractal processes. There are some stations present α value greater than 2 in May to October (austral winter) in red, which are also found in the paper of Calif et al. (2013). The others are $0 < \alpha < 2$ in each month. Almost all of the stations are multifractal with Levy bounded and unbounded singularities in Table 4.5.

Table 4.4 H value of each station in each month of 2014

| H | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|----------------------------|-------------|-------------|------|------|------|------|------|------|------|------|------|-------------|
| 1 Moutfia | 0.37 | 0.45 | 0.36 | 0.28 | 0.25 | 0.34 | 0.22 | 0.30 | 0.31 | 0.26 | 0.41 | 0.48 |
| 2 MOUFIA BOIS DE NEFLES | 0.34 | 0.34 | 0.30 | 0.26 | 0.24 | 0.27 | 0.15 | 0.21 | 0.15 | 0.25 | 0.36 | 0.39 |
| 3 LA POSSESSION | 0.34 | 0.26 | 0.29 | 0.26 | 0.21 | 0.22 | 0.25 | 0.27 | 0.22 | 0.24 | 0.28 | 0.40 |
| 4 SAINT LEU | 0.30 | 0.28 | 0.27 | 0.25 | 0.20 | 0.31 | 0.22 | 0.22 | 0.30 | 0.31 | 0.27 | 0.34 |
| 5 SAINT PIERRE | 0.42 | 0.32 | 0.28 | 0.32 | 0.31 | 0.31 | 0.18 | 0.30 | 0.26 | 0.33 | 0.38 | 0.42 |
| 6 SAINT JOSEPH | 0.47 | 0.47 | 0.34 | 0.29 | 0.29 | 0.23 | 0.15 | 0.24 | 0.23 | 0.40 | 0.31 | 0.35 |
| 7 SAINTE ROSE | 0.45 | 0.43 | 0.37 | 0.26 | 0.29 | 0.29 | 0.27 | 0.32 | 0.26 | 0.30 | 0.32 | 0.52 |
| 8 SAINT ANDRE | 0.30 | 0.31 | 0.13 | 0.15 | 0.16 | 0.16 | 0.14 | 0.19 | 0.06 | 0.16 | 0.21 | 0.33 |
| 9 CILAOIS PISCINE | 0.47 | 0.43 | 0.37 | 0.34 | 0.33 | 0.49 | 0.36 | 0.44 | 0.44 | 0.50 | 0.37 | 0.42 |
| 10 BRAS PANON | 0.41 | 0.40 | 0.23 | 0.24 | 0.25 | 0.33 | 0.23 | 0.33 | 0.21 | NaN | NaN | NaN |
| 11 TAMPON UR | 0.41 | 0.39 | 0.36 | 0.29 | 0.24 | 0.29 | 0.29 | 0.33 | 0.48 | 0.35 | NaN | NaN |

Table 4.5 α value of each station in each month of 2014

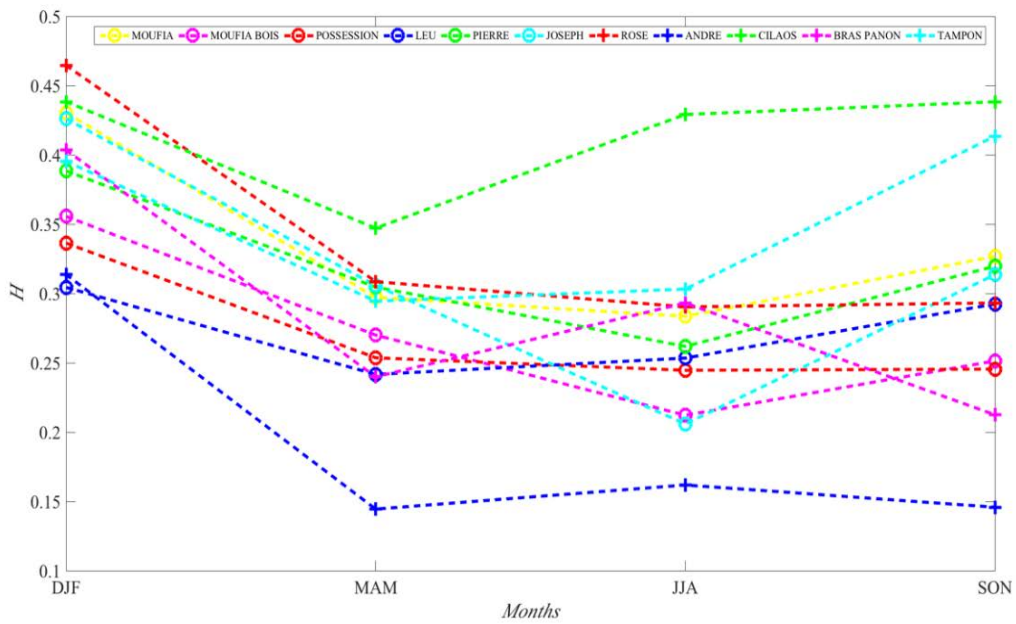
| α | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|----------------------------|------|------|------|------|-------------|-------------|-------------|------|-------------|-------------|------|------|
| 1 Moutfia | 0.65 | 1.26 | 0.58 | 0.98 | 0.67 | 1.01 | 1.59 | 1.70 | 0.53 | 0.70 | 0.72 | 1.83 |
| 2 MOUFLA BOIS DE NEFLES | 0.19 | 1.10 | 0.51 | 0.60 | 1.63 | 1.14 | 2.01 | 1.27 | 0.83 | 0.80 | 0.46 | 0.48 |
| 3 LA POSSESSION | 0.85 | 1.29 | 0.49 | 1.18 | 1.72 | 1.48 | 1.45 | 0.95 | 0.96 | 0.49 | 1.09 | 0.76 |
| 4 SAINT LEU | 0.73 | 1.51 | 0.84 | 0.86 | 2.09 | 1.21 | 1.66 | 1.12 | 0.80 | 0.22 | 1.48 | 0.68 |
| 5 SAINT PIERRE | 0.80 | 1.57 | 0.59 | 1.23 | 1.61 | 1.18 | 2.00 | 0.87 | 0.35 | 0.84 | 0.45 | 0.52 |
| 6 SAINT JOSEPH | 0.88 | 0.87 | 0.62 | 1.11 | 0.78 | 1.75 | 2.18 | 0.76 | 0.73 | 0.61 | 0.97 | 1.29 |
| 7 SAINTE ROSE | 0.77 | 1.26 | 0.58 | 0.98 | 0.67 | 1.01 | 1.59 | 1.70 | 0.53 | 0.70 | 0.72 | 1.83 |
| 8 SAINT ANDRE | 0.54 | 1.33 | 1.68 | 1.21 | 2.27 | 2.92 | 1.68 | 1.43 | 2.48 | 2.10 | 0.80 | 1.41 |
| 9 CILAOIS PISCINE | 0.74 | 0.54 | 0.23 | 0.48 | 0.89 | 0.92 | 0.49 | 0.28 | 0.62 | 0.68 | 1.02 | 1.56 |
| 10 BRAS PANON | 0.83 | 1.23 | 0.62 | 0.61 | 1.16 | 1.29 | 0.77 | 1.33 | 0.88 | NaN | NaN | NaN |
| 11 TAMPON UR | 0.67 | 0.73 | 0.44 | 0.69 | 1.45 | 0.88 | 0.69 | 0.37 | 0.69 | 1.12 | NaN | NaN |

Table 4.6 C_I value of each station in each month of 2014

| C_I | Jan | Feb | Mar | Apr | May | Jun | Jul | Aug | Sep | Oct | Nov | Dec |
|-------------------------|------|------|------|------|------|------|------|------|------|------|------|------|
| 1 Moufia | 0.28 | 0.22 | 0.18 | 0.13 | 0.12 | 0.12 | 0.07 | 0.07 | 0.22 | 0.12 | 0.22 | 0.14 |
| 2 MOUFIA BOIS DE NEFLES | 0.16 | 0.11 | 0.19 | 0.16 | 0.05 | 0.12 | 0.05 | 0.06 | 0.06 | 0.11 | 0.18 | 0.18 |
| 3 LA POSSESSION | 0.14 | 0.06 | 0.19 | 0.14 | 0.04 | 0.07 | 0.08 | 0.12 | 0.11 | 0.13 | 0.10 | 0.18 |
| 4 SAINT LEU | 0.10 | 0.03 | 0.13 | 0.09 | 0.04 | 0.15 | 0.07 | 0.13 | 0.24 | 0.09 | 0.04 | 0.18 |
| 5 SAINT PIERRE | 0.24 | 0.10 | 0.16 | 0.12 | 0.07 | 0.18 | 0.04 | 0.21 | 0.17 | 0.16 | 0.24 | 0.21 |
| 6 SAINT JOSEPH | 0.28 | 0.22 | 0.23 | 0.13 | 0.13 | 0.06 | 0.01 | 0.09 | 0.07 | 0.23 | 0.07 | 0.07 |
| 7 SAINTE ROSE | 0.18 | 0.22 | 0.18 | 0.13 | 0.12 | 0.12 | 0.07 | 0.07 | 0.22 | 0.12 | 0.22 | 0.14 |
| 8 SAINT ANDRE | 0.13 | 0.13 | 0.01 | 0.05 | 0.01 | 0.02 | 0.04 | 0.06 | 0.01 | 0.04 | 0.07 | 0.11 |
| 9 CLAOS PISCINE | 0.24 | 0.29 | 0.26 | 0.27 | 0.23 | 0.37 | 0.33 | 0.30 | 0.37 | 0.35 | 0.19 | 0.10 |
| 10 BRAS PANON | 0.25 | 0.19 | 0.08 | 0.12 | 0.03 | 0.13 | 0.14 | 0.07 | 0.14 | NaN | NaN | NaN |
| 11 TAMPON UR | 0.21 | 0.19 | 0.26 | 0.14 | 0.06 | 0.18 | 0.19 | 0.18 | 0.29 | 0.19 | NaN | NaN |

Table 4.6 presents the C_1 value of each station in each month of 2014. All the C_1 belong to $[0, 1]$. And each station seems have bigger C_1 value in austral summer (November to April) than austral winter (May to October), which indicates that GHI is more intermittent in austral summer in the different region of the island.

Figure 4.22 presents the universal multifractal parameter H , C_1 , and α on the austral summer (DJF), austral winter (JJA) and inter-seasons (MAM and SON) at 11 stations for observing the seasonal intermittency of the global solar radiation. All the stations have the larger H value in December-January-February than other seasons, corresponding to the summer season (the deep convective clouds and low-altitude clouds often prevail during in this season) over Reunion Island which are more intermittency. The trade wind is predominant during June-July-August over Reunion Island, so the larger α is observed in most station in this period which marking the degree of multifractal properties



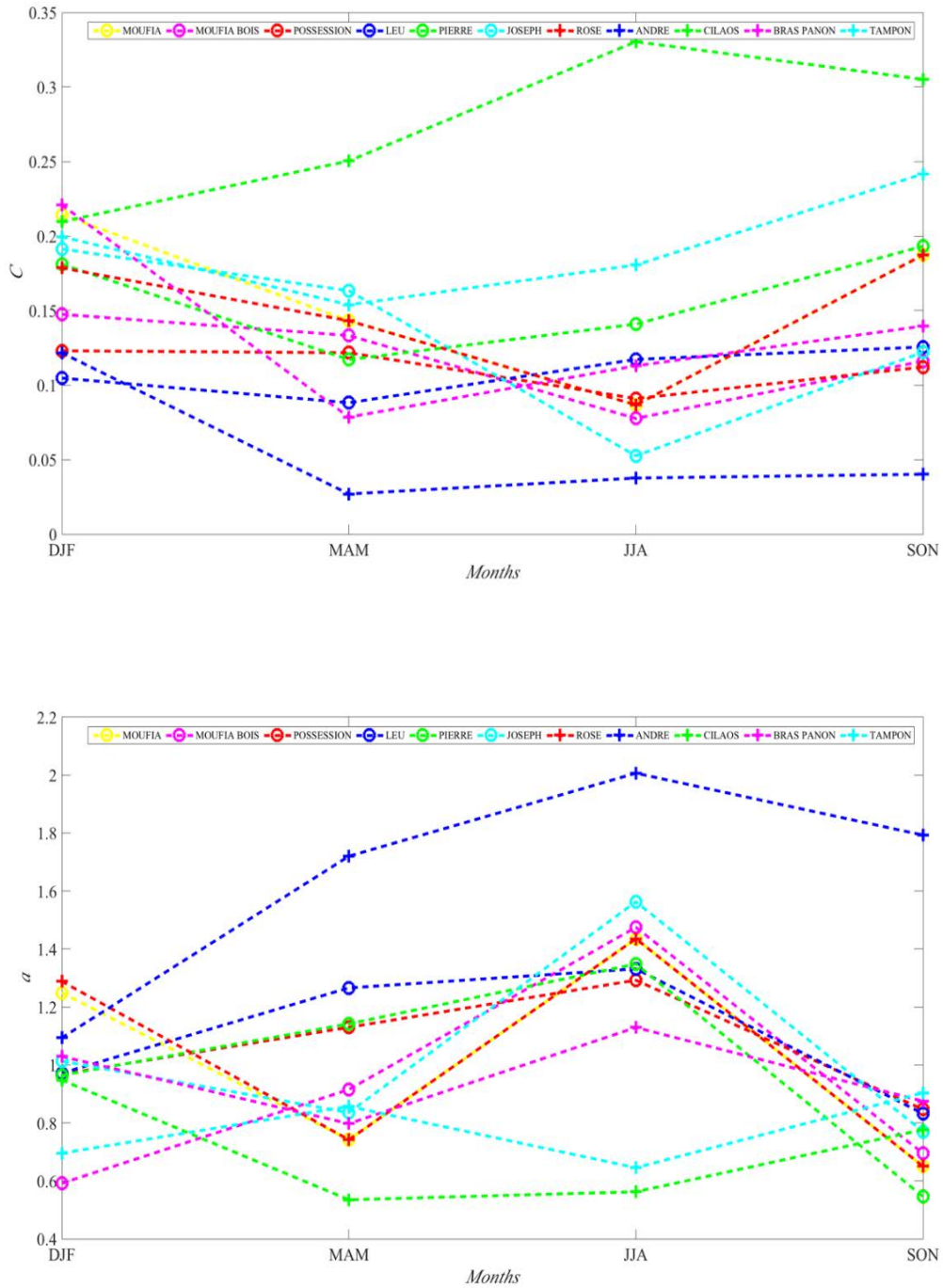


Figure 4.22 The universal multifractal parameter H , C_l , and α on the austral summer (DJF), austral winter (JJA) and inter-seasons (MAM and SON) at 11 stations.

4.2.2.2 Variability of Singularity spectrum width in summer, winter and interseason period

Same as the last section, the singularity spectrum corresponding to the scaling exponents $\zeta(q)$ on the austral summer (DJF), austral winter (JJA) and inter-seasons (MAM and SON) at

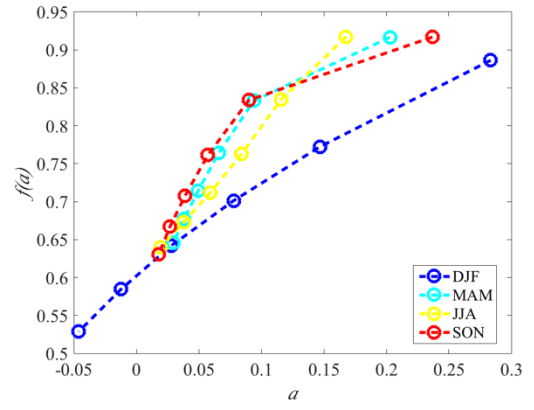
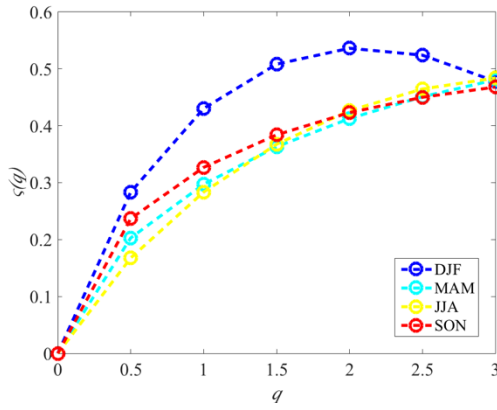
11 stations are studied in Figure 4.23 for seasonal intermittency. Firstly we could find that the singularity spectrum in each season for all stations are nonlinear and concave, indicating the multifractal and intermittent properties for the global solar radiation.

Secondly, each station presents different singularity spectrum pattern in summer, winter and inter-seasons, which could be explained as the spatial seasonal variability. Thirdly, summer season (December-January-February) presents relatively large range difference than other seasons. It also highlights the stronger intermittency in summer. The station La Possession and St Leu which located in the west coast of Reunion Island present close singularity spectrum pattern in different seasons. And the station St Pierre and St Joseph which in the south of Reunion Island also show the similar singularity spectrum pattern. These also indicate the seasonal spatial variability of surface solar radiation.

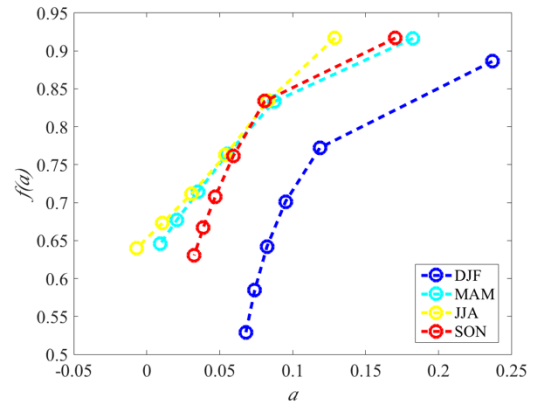
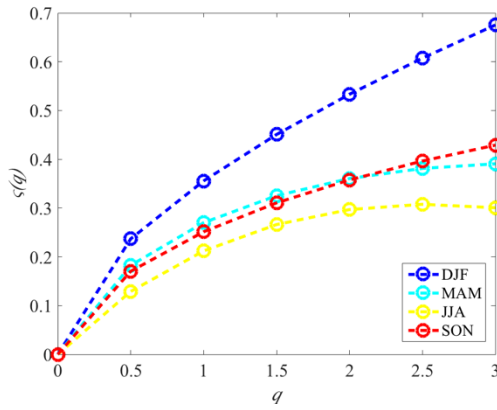
Thirdly, through calculating the width of the singularity spectrum on the austral summer (DJF), austral winter (JJA) and inter-seasons (MAM and SON) at 11 stations in Table 4.7, the station Cilaos has the largest width of the singularity spectrum (0.37) in the austral winter (June-July-August) and inter-season (September-October-November), and the station St Andre has the smallest width of the singularity spectrum (0.09) in the austral winter (June-July-August) and inter-season (September-October-November), which indicate the strongest and weakest seasonal multifractality separately in this station network. And these results are corresponding to the mean spatial variability throughout the year.

In summary, the spatial pattern of the intermittency and multifractality over SPN1 network in the seasonal scale and throughout the year could be presented and observed by studying the parameters H , C_1 and α and the singularity spectrum.

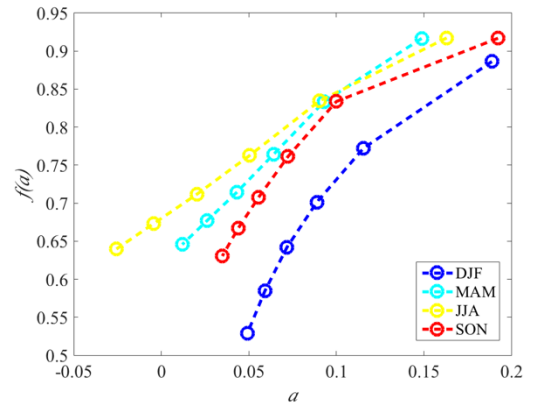
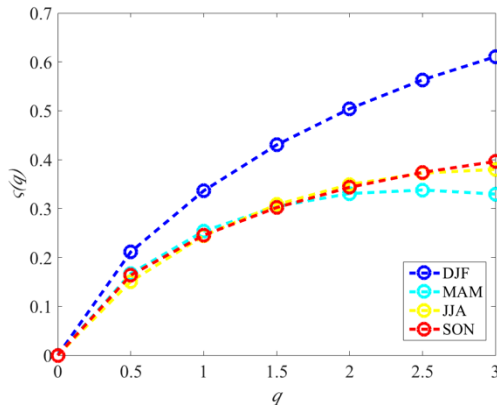
1
MOUFIA
UR



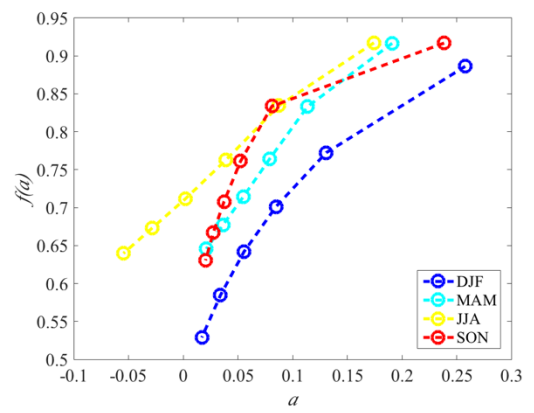
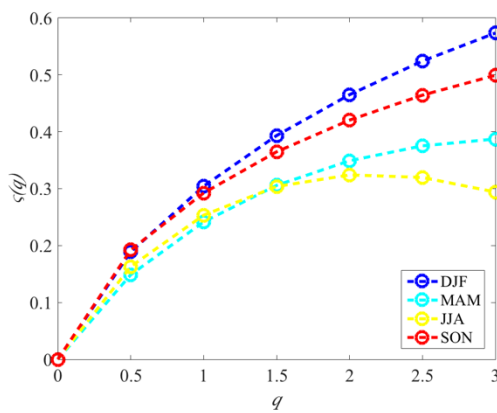
2
MOUFIA
BOIS DE
NEFLES



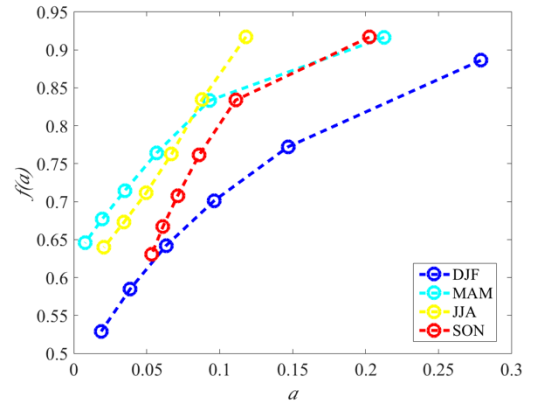
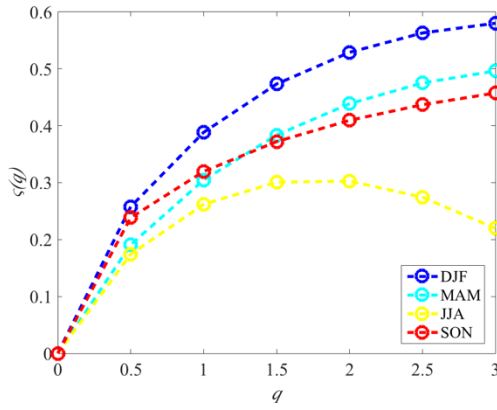
3
LA
POSSES
SION



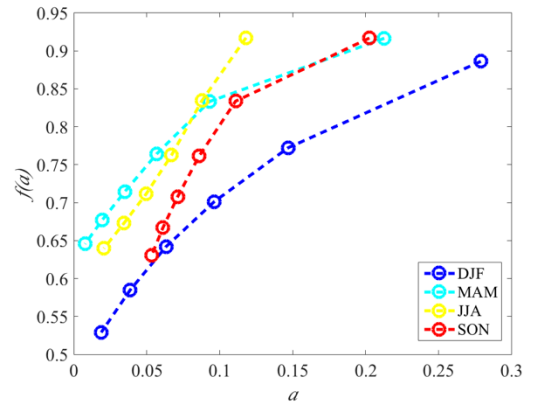
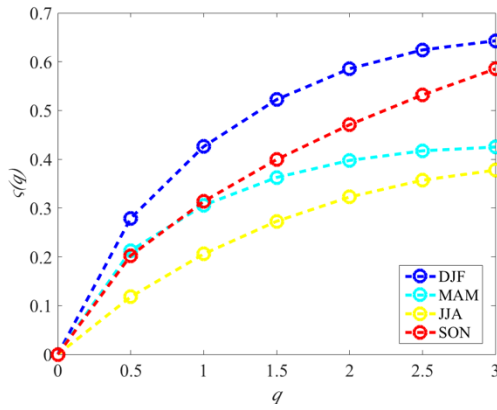
4
SAINT
LEU



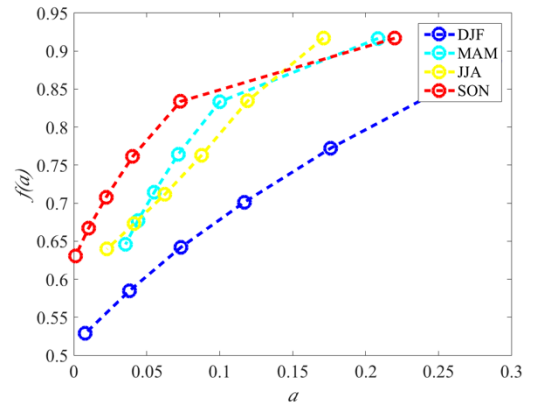
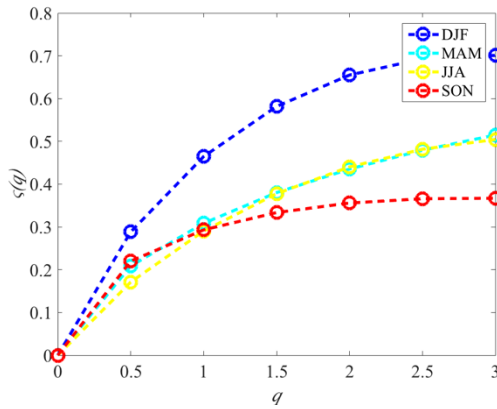
5
SAINT
PIERRE



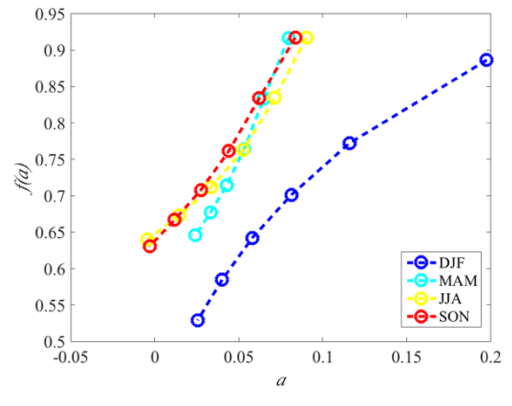
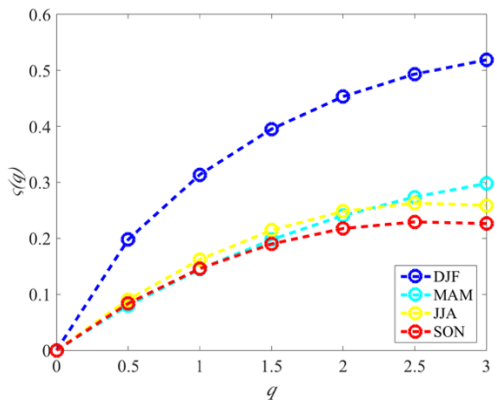
6
SAINT
JOSEPH



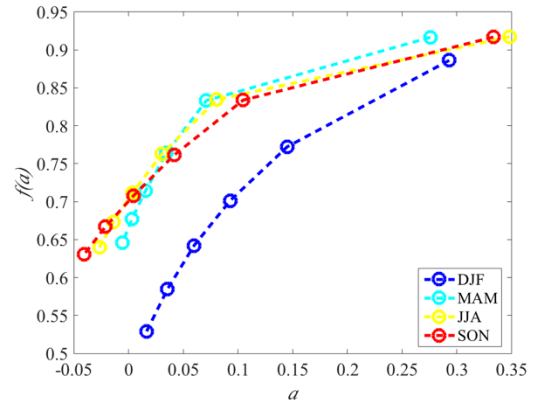
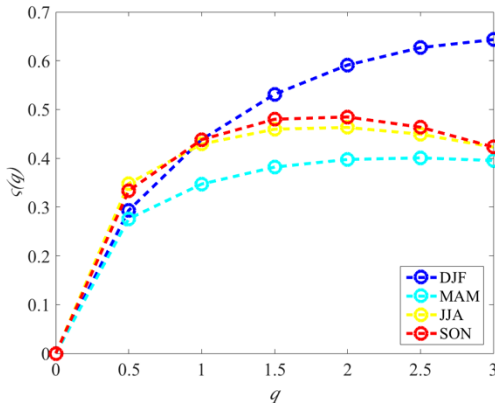
7
SAINTE
ROSE



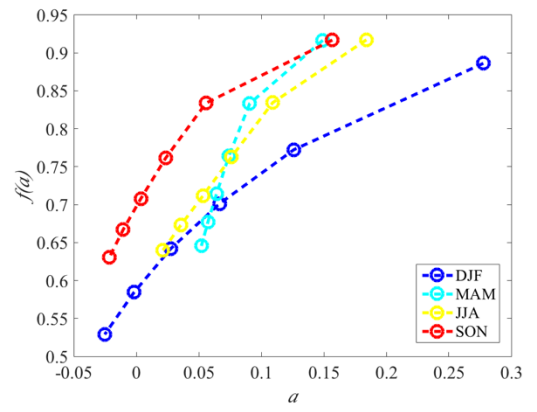
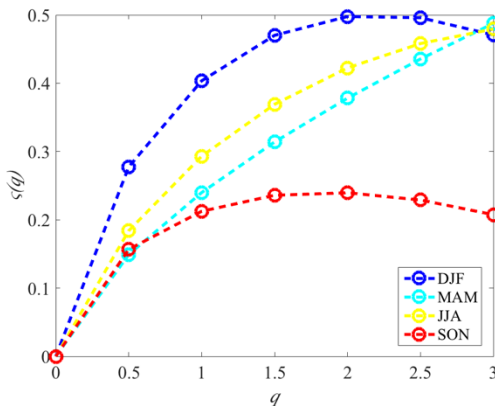
8
SAINT
ANDRE



9
CILAOS
PISCINE



10
BRAS
PANON



11
TAMPON
UR

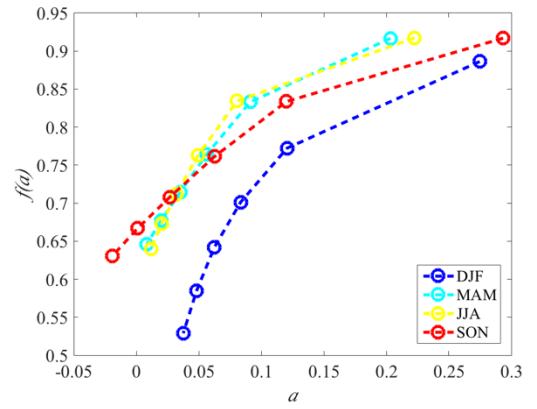
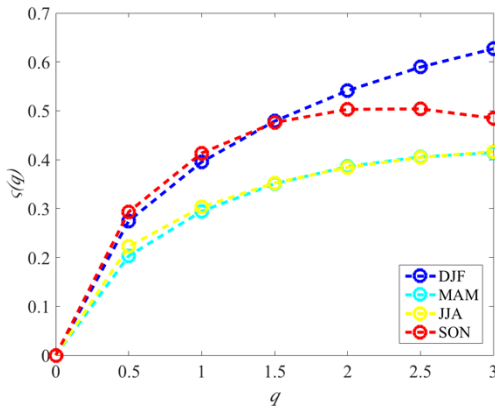


Figure 4.23 The singularity spectrum corresponding to the scaling exponents $\zeta(q)$ on the austral summer (DJF), austral winter (JJA) and inter-seasons (MAM and SON) at 11 stations.

Table 4.7 The width of the singularity spectrum on the austral summer (DJF), austral winter (JJA) and inter-seasons (MAM and SON) at 11 stations

| Seasons Stations | DJF | MAM | JJA | SON |
|--------------------------|------|------|-------------|-------------|
| 1. MOUFIA UR | 0.33 | 0.17 | 0.15 | 0.23 |
| 2. MOUFIA BOIS DE NEFLES | 0.17 | 0.17 | 0.14 | 0.14 |
| 3. LA POSSESSION | 0.16 | 0.18 | 0.14 | 0.14 |
| 4. SAINT LEU | 0.14 | 0.14 | 0.19 | 0.16 |
| 5. SAINT PIERRE | 0.24 | 0.17 | 0.23 | 0.22 |
| 6. SAINT JOSEPH | 0.26 | 0.20 | 0.10 | 0.15 |
| 7. SAINTE ROSE | 0.28 | 0.17 | 0.15 | 0.22 |
| 8. SAINT ANDRE | 0.17 | 0.06 | 0.09 | 0.09 |
| 9. CILAOS PISCINE | 0.28 | 0.28 | 0.37 | 0.37 |
| 10. BRAS PANON | 0.30 | 0.10 | 0.16 | 0.18 |
| 11 TAMPON UR | 0.24 | 0.19 | 0.21 | 0.31 |

Chapter 5

Conclusion and Perspective

The objective of this thesis is to build a daily mapping prediction model through the intermittency analysis of global solar radiation with multifractal parameters over Reunion Island. In a first study, the deterministic daily surface solar radiation mapping prediction is proposed as a new approach firstly in this study. Four different linear regression models are applied on the SARAH-E@5km satellite data (CM SAF) during 2007-2016 for prediction. Depending on the results and the quality of the prediction, the intermittency analysis with the arbitrary order Hilbert spectral analyses are applied to get the multifractal parameters (H , C_1 and α) as new predictors for achieving more accuracy mapping performance by the new prediction model. The main conclusion and the future work plan are presented in this chapter.

5.1 Daily mapping prediction: A new reduced dimension approach

A spatio-temporal multivariate model as a new approach is applied for mapping prediction using SARAH-E@5km satellite data (CM SAF) during 2007-2016 over Reunion Island. Four linear regression models: SR, MLR, PCR and PLSR are implemented and compared in the prediction process. The first five years 2007-2011 are used as the training data and other five years from 2012~2016 as test data. PCA as a dimension reduction method is used in PCR and PLSR models, which could give PCs (43 PCs could explain 95% variance) as input in the prediction model and reduce the dimension of the dataset. Through the goodness of fitting and re-mapping results against observation, and also the statistical analysis, PCR model has smallest MAE, MSE and RMSE comparing to other three models. PCR model is better for GHI mapping prediction over Reunion Island, even with quite big error.

5.2 Intermittency: A first attempt of spatial and temporal characterization for Reunion Island

The prediction results based on the linear regression model is not so satisfactory, which push us to study the intermittency and multifractal characters of GHI for improving the model's quality. The arbitrary order Hilbert spectral analyses which is the combination of the Empirical Mode Decomposition and Hilbert spectral analysis (EMD-HSA) are applied to study the intermittency and multifractality of GHI over Reunion Island. The generalized scaling exponents $\xi(q)$ is estimated through the arbitrary order Hilbert spectral analyses and three parameters: Hurst exponent (H), the fractal co-dimension (C_1), and Lévy parameter (α) are taken to study the multifractal process of the GHI. To achieve this multifractal analysis, we have used one available highest sampling rate of one-year GHI records (one observation per second) which is located at Saint-Denis (Moufia). In addition, we have used also twelve other

Conclusion and Perspective

stations with lower sampling rate of GHI records (one minute per second) and time period greater than one year.

Firstly, the GHI during the day time (7h-18h) sampling on second at one station Moufia from May to December in 2016 are applied to obtain the generalized scaling exponent and log-stable model parameters. The multifractal processes could be found in the sub-daily and daily fluctuations. A power law behavior with a spectral exponent $\beta=1.68$ close to the Kolmogorov spectrum is detected through Fourier spectrum analysis in this GHI time series, which also indicates that the sub-daily fluctuations of GHI are nonstationary. The scaling exponent $\zeta(q)$ is then estimated by the arbitrary order Hilbert spectral analysis and the multifractal properties is detected. The log-stable model parameters H , C_1 and α characterize the concavity of the scaling exponent $\zeta(q)$ for analyzing the intermittency of GHI. H -the Hurst parameter defines the degree of smoothness or roughness of the field; secondly, C_1 measures the inhomogeneity mean or the mean intermittency characterizing the sparseness of the field and thirdly, and the multifractal Lévy parameter α measures the degree of multifractality. The classification method is applied to the daily GHI for analyzing the distribution of the daily intermittency process and five classes with GHI and K_b are obtained. The multifractal parameters (H , C_1 , α) then are classified into these five classes which present the variability of intermittency for each type of weather and thus the range value of (H , C_1 , α). It has been found that the Hurst exponent H and parameter C_1 are more close to Beta distribution, and the parameter α is more close to Gamma distribution.

Secondly, the GHI variability throughout the day month by month also presents a multifractal fluctuation in the SPN1 network with 11 GHI stations record in 2014. The lowest monthly value of GHI appears in June and the highest in December (missing the data from January to April sampling on second) through the distribution month by month. The monthly change for the universal multifractal parameter H , C_1 , and α indicates the seasonal variability of the GHI multifractal pattern. All the H is less than 0.5, which indicates that the GHI fluctuations from day to day are anti-persistent in average. The value of H at Cilaos station is close to 0.5, indicating that the GHI fluctuations there are close to Brownian and random process. All the C_1 are around 0.15 and the Cilaos station gives 0.27, which means the field of Cilaos is more heterogeneous than other place. The stations at Saint Pierre, Moufia, La Possession, Saint Joseph, Saint Rose and Bras Panon exhibit α value close to 1 which means that the GHI fluctuation there is following a Cauchy process. The station Saint Andre, Saint Leu, Cilaos and Le Tampon are closed to Levy process with unbounded or bounded

Conclusion and Perspective

singularities. The singularity spectrum related to the scaling exponents $\zeta(q)$ is also present to characterize the multifractal process of the GHI time series in the station network. The width of the singularity spectrum measures the multifractality level of the signal. The bigger amount of multifractality is present in the signal, the width of singularity spectrum becomes wider. The width of the multifractal spectrum reflect the temporal variation of the Hurst exponent. The station 9 (Cilaos) has the largest width of the singularity spectrum (0.33) and the station 8 (St Andre) has the smallest width of the singularity spectrum (0.09), which indicate the strongest and weakest multifractality separately in this station network. It also reveals that the station 3 (La Possession) and 4 (St Leu) present the similar multifractality for they have the same width of the singularity spectrum (0.14). This is reasonable that La Possession and St Leu locate in the west of the island, receiving more stable GHI relatively.

The intermittency indicated by the universal multifractal parameters is consistent throughout the year by the analysis the seasonal spatial pattern of log-stable parameters H , C_1 and α . Firstly, the Hurst exponent H are all less than 0.5 for the 11 stations. The station Moufia (0.48), Saint Joseph (0.47) and Cilaos (0.47) in December, January and February give H value close to 0.5, indicating a Brownian process. All the other stations in each month throughout the year are belong to the anti-persistent process. Almost all of the stations are multifractal with Levy bounded and unbounded singularities. All the C_1 belong to $[0, 1]$. And each station seems have bigger C_1 value in austral summer (November to April) than austral winter (May to October), which indicates that GHI is more intermittent in austral summer in the different region of the island. All the stations have the larger H value in December-January-February than other seasons, corresponding to the summer season (the deep convective clouds and low-altitude clouds often prevail during in this season) over Reunion Island which are more intermittency. The trade wind is predominant during June-July-August over Reunion Island, so the larger α is observed in most station in this period which marking the degree of multifractal properties. When studying the singularity spectrum corresponding to the scaling exponents $\zeta(q)$ on the austral summer (DJF), austral winter (JJA) and inter-seasons (MAM and SON) at 11 stations, firstly we could find that the singularity spectrum in each season for all stations are nonlinear and concave, indicating the multifractal and intermittent properties of GHI. Secondly, each station presents different singularity spectrum pattern in summer, winter and inter-seasons, which could be explained as the spatial seasonal variability. Thirdly, summer season (December-January-February) presents relatively large range difference than other seasons. It also highlights the stronger intermittency in summer. The station La Possession and St Leu which located in the west coast of Reunion Island present close singularity spectrum

pattern in different seasons. And the station St Pierre and St Joseph which in the south of Reunion Island also show the similar singularity spectrum pattern. These also indicate the seasonal spatial variability of surface solar radiation. The station Cilaos has the largest width of the singularity spectrum (0.37) in the austral winter (June-July-August) and inter-season (September-October-November), and the station St Andre has the smallest width of the singularity spectrum (0.09) in the austral winter (June-July-August) and inter-season (September-October-November), which indicate the strongest and weakest seasonal multifractality separately in this station network. In summary, the spatial pattern of the intermittency and multifractality over SPN1 network in the seasonal scale and throughout the year could be presented and observed by studying the parameters H , C_1 and α and the singularity spectrum.

5.3 Consistency of log-stable parameter to be used as predictors in the daily mapping prediction model

Based on the intermittency study of GHI over Reunion Island, the parameters H , C_1 and α could be used as predictors in the daily mapping prediction model to characterize the multifractality of GHI. It could be the work in the next step.

5.4 Future work plan

- Merge intermittency parameters in linear mapping prediction

Adding the parameters H , C_1 and α as predictors in the linear mapping prediction model to improve the prediction results could be one of the future work. On the basis of the PCR model for daily mapping prediction of GHI, the parameters H , C_1 and α could be the independent variable after the PCA and the predicted H , C_1 and α are the dependent variable as output for example.

- Daily log-stable parameters prediction model

The non-linear prediction models, such as ANN could be another way to achieve the GHI prediction quality with the daily log-stable parameters H , C_1 and α . Through the prediction network with these parameters, the intermittency and multifractality of the GHI time series could be more considered and present. This could be another future work after this thesis.

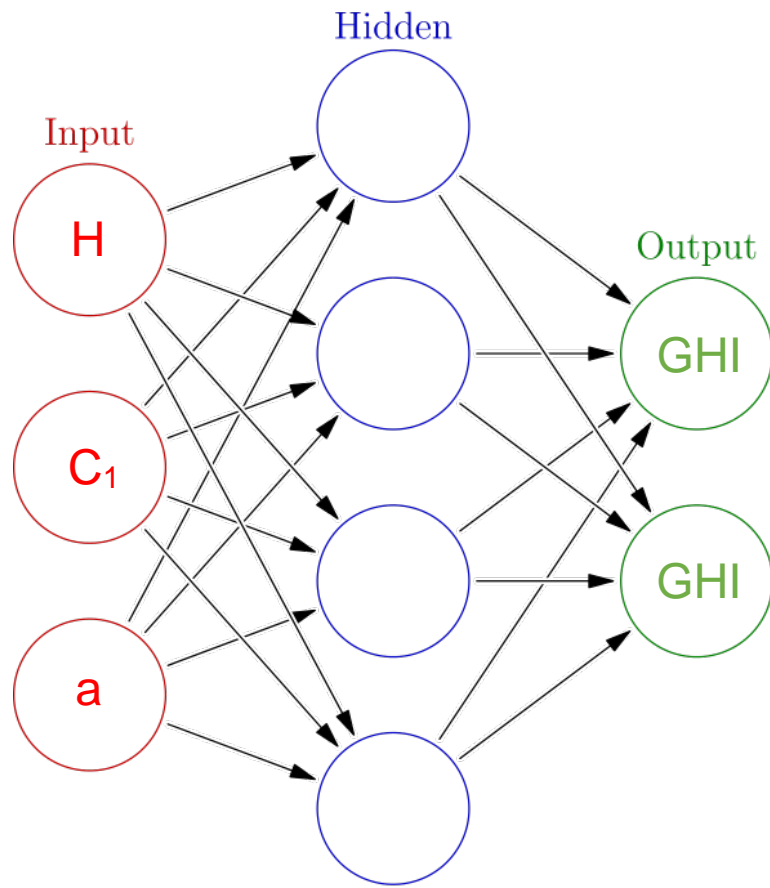


Figure 5.1 Daily log-stable parameters prediction model

References

- Ahmad F, Ulfat I (2004) Empirical models for the correlation of monthly average daily global solar radiation with hours of sunshine on a horizontal surface at Karachi, Pakistan. *Turkish Journal of Physics*, 28: 301-307.
- Altman N.S (1992) An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46 (3): 175-185.
- Amillo A.G, Huld T and Müller R (2014) A New Database of Global and Direct Solar Radiation Using the Eastern Meteosat Satellite, *Models and Validation Remote Sensing*, 6(9): 8165-8189.
- Badosa J, Haeffelin M, Chepfer H (2013) Scales of spatial and temporal variation of solar irradiance on Reunion tropical island. *Solar Energy* 88: 42-56.
- Badosa J, Haeffelin M, Kalecinski N, Bonnardot F and others (2015) Reliability of day-ahead solar irradiance forecasts on Reunion Island depending on synoptic wind and humidity conditions. *Solar Energy* 115: 306-321.
- Baldy S, Ancellet G, Bessafi M and others (1996) Field observations of the vertical distribution of tropospheric ozone at the island Reunion (southern tropics). *Journal of Geophysical Research*, 101: 23835-23849.
- Bedrosian E (1963) A product theorem for Hilbert transform. *Proc. IEEE* 51: 868-869.
- Bechini L, Ducco G, Donatelli M, Stein A (2000) Modelling, interpolation and stochastic simulation in space and time of global solar radiation. *Agriculture, Ecosystems & Environment*, 81(1): 29-42.
- Bendat J.S, Piersol A.G (1986) *Random data: analysis and measurement procedures*, New York: Wiley.
- Benzi R, Paladin G, Parisi G and Vulpiani A (1984) On the multifractal nature of fully developed turbulence and chaotic systems. *J. Phys*, 17: 3521-3531.
- Bessafi M, De Carvahlo F.A.T, Chartron P, Delsaut M and others (2013) Classification des journées en fonction des radiations solaires sur l'île de la Réunion. 45e journées de la Statistique, Toulouse, France.
- Boashash B (1992) Estimating and interpreting the instantaneous frequency of a signal- Part I. Fundamentals. *Proceedings of the IEEE*, 80: 520-538.
- Boes E.C (1981) Fundamentals of solar radiation. In Kreider JF and Kreith F, editors, *Solar Energy Handbook*, Chapter 2. McGraw-Hill, New York.
- Bojanowski J.S (2014) Quantifying solar radiation at the earth surface with meteorological and satellite data. Ph.D. Thesis. University of Twente: Netherlands.
- Bosh A.S, Person M.J and others (2015) The state of Pluto's atmosphere in 2012-2013. *Icarus*, 246: 237-246.
- Böttcher F, Barth S, Peinke J (2007) Small and large scale fluctuations in atmospheric wind speeds. *Stoch. Env. Res. Risk A*. 21: 299-308.
- Brisson A, Borgne P.L, Marsouin A (1999) Development of algorithms for surface solar irradiance retrieval at O&SI SAF low and mid latitudes. *Météo-France/SCEM/CMS O&SI SAF Low and Mid Latitudes*.

- Brockwell P.J and Davis R.A (1986) *Time Series: Theory and Methods*, Springer Series in Statistics.
- Brockwell P.J, Davis R.A (2002) *Introduction to time series and forecasting*. Springer.
- Brooks C.L, Karplus M, Pettit B.M (1988). *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*, Volume 71 (New York: J. Wiley).
- Calif R, Schmitt F.G, Huang, Y.X, Soubdhan T (2013) Intermittency study of high frequency global solar radiation sequences under a tropical climate. *Solar Energy*, 98: 349-365.
- Chatfield C (1975) *The Analysis of Time Series: Theory and Practice*. Chapman and Hall.
- Chaudary J.B.H, Ortega J, Smith G.H (2012) Spectral Analysis of Storm Waves using the SLEX algorithm and the Hilbert-Huang Transform. *Ocean Research*.
- Chiacchio M, Vitolo R (2012) Effect of cloud and atmospheric circulation patterns on the observed surface solar radiation in Europe. *J Geophys Res* 117: D18207.
- Chiacchio M, Solmon F and others (2015) Evaluation of the radiation budget with a regional climate model over Europe and inspection of dimming and brightening. *Journal of Geophysical Research: Atmospheres*, 120(5): 1951-1971.
- Codd E.F (1970) A Relational Model of Data for Large Shared Data Banks. *Communications of the ACM*, 13 (6): 377-387.
- Cohen L (1995) *Time-frequency analysis*. Prentice Hall PTR Englewood Cliffs, NJ.
- Coimbra C.F.M, Pedro H.T.C (2013) Stochastic-Learning Methods. *Solar Energy Forecasting and Resource Assessment*, 383-406.
- Craddock J.M (1973) Problems and prospects for eigenvector analysis in meteorology. *The statistician*, 22: 133-145.
- Cyril V, Christophe P.L, Marc M and Marie L.N (2013) Multi-horizon solar radiation forecasting for Mediterranean locations using time series models. *Renewable and Sustainable Energy Reviews*, 28: 44-52.
- Cyril V, Pierrick H and others (2014) Time series modeling and large scale global solar radiation forecasting from geostationary satellites data. *Solar Energy*, 102: 131-142.
- Daubechies I (1990) The wavelet transform, time-frequency localization and signal analysis. *IEEE Transactions on Informat* 36(5): 961-1005.
- Davis A.B, Marshk A, Cahalan R.F, Wiscombe W.J (1997) Interactions: solar and laser beams in stratus clouds, fractals & multifractals in climate & remote-sensing studies. *Fractals* 5, Issue supp02.
- Davy R.J, Troccoli A (2012) Interannual variability of solar energy generation in Australia. *Solar energy* 86: 3554-3560.
- Diggle P.J (1990) *Time Series: A Biostatistical Introduction*, Oxford University Press.
- Dincer I (2000) Renewable Energy and Sustainable Development: A Crucial Review. *Renewable and Sustainable Energy Reviews*, 4: 157-175.
- Diagne M.H, David M and others (2014) Post-processing of Solar Irradiance Forecasts from WRF Model at Reunion Island. *Energy Procedia*, Elsevier, 57: 1364 -1373.

- Du P, Kibbe W.A, Lin S.M (2006) Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics*, 22: 2059-2065.
- Duffie J.A, Beckman W.A (1980) *Solar Engineering of Thermal Processes*.
- Edward O (2002) *Chaos in dynamical systems*, Cambridge University Press, 323.
- Ellabban O, Abu-Rub H, Blaabjerg F (2014) Renewable energy resources: Current status, future prospects and their enabling technology. *Renewable and Sustainable Energy Reviews*, 39: 748-764 [749].
- Ertekin C, Evrendilek F (2007) Spatio-temporal modeling of global solar radiation dynamics as a function of sunshine duration for Turkey. *Agricultural and Forest Meteorology*, 145: 36-47.
- Fadare D.A (2010) The application of artificial neural networks to mapping of wind speed profile for energy application in Nigeria. *Applied Energy* 87(3): 934–942.
- Flandrin P (1992) Wavelet analysis and synthesis of fractional Brownian motion. *IEEE Transactions on Information Theory*, 38(2).
- Flandrin P (1999) *Wavelet analysis and its applications*. Cademic Press San Diego.
- Flandrin P, Gonçalves P (2004) Empirical Mode Decompositions as Data-Driven Wavelet-Like Expansions. *Int. J. Wavelets, Multires.* 2(4): 477-496.
- Flandrin P, Rilling G, Gonçalves P (2004) Empirical mode decomposition as a filter bank. *IEEE Sig. Proc. Lett.*, 11(2): 112-114.
- Frisch U (1995) *Turbulence: The Legacy of AN Kolmogorov*. Cambridge University Press.
- Fukuoka A (1951) The Central Meteorological Observatory, A study on 10-day forecast (A synthetic report). *The Geophysical Magazine*, Tokyo, XXII: 177-218.
- Garci L (2003) Renewable energy applications in desalination: state of the art. *Solar Energy* 75: 381-393.
- Gautier C, Diak G, Masse S and others (1980) A Simple Physical Model to Estimate Incident Solar Radiation at the Surface from GOES Satellite Data. *Journal of Applied meteorology*, 19: 1005-1012.
- Geiger B, Dominique C and others (2008) Land Surface Albedo Derived on a Daily Basis From Meteosat Second Generation Observations, *Geoscience and Remote Sensing*, *IEEE Transactions on*, 11: 3841-3856.
- Gilgen M, Hübner P, Höfelein C and others (1998) PCR-based detection of verotoxin-producing *Escherichia coli* (VTEC) in ground beef Détection par PCR de *Escherichia coli* produisant des vérotoxines dans la viande hachée. *Research in Microbiology*, 149(2): 145-154.
- Glasbey C.A, Allcroft D.J, Paulo M.J (2007) A latent Gaussian model for multivariate consumption data. *Food Quality and Preference*, 18: 508-516.
- Gabor D (1946) *Theory of communication*. *Proc. IEE* 93: 429-457.
- Golub G.H, Van Loan C.F (1983) *Matrix computations*. The Johns Hopkins University Press.

- Gong X.F, Richman M.B (1995) On the application of cluster analysis to growing season precipitation data in North America east of the Rockies. *JCLI*, 8: 897-931.
- Grassberger P, Procaccia I (1983) Measuring the strangeness of strange attractors. *Physica D*, 9(1-2): 189-208.
- Grinsted A, Moor J.C (2004) Application of the cross wavelet transform and wavelet coherence to geophysical time series. *Nonlinear Processes in Geophysics*, European Geosciences Union (EGU), 11(5/6): 561-566.
- Gordon R (2009) Predicting solar radiation at high resolutions: A comparison of time series forecasts. *Solar Energy*, 83(3): 342-349.
- Graham S (1999) Clouds and radiation. The earth's climate system constantly adjusts. Retrieved from <http://earth.observatory.nasa.gov/Features/Clouds>.
- Gueymard C.A, Myers D.R (2008b) Solar radiation measurement: progress in radiometry for improved modeling. V. Badescu (Ed.), *Modeling Solar Radiation at the Earth's Surface*, Springer.
- Hannachi A (2004) A Primer for EOF Analysis of Climate Data. Department of Meteorology, University of Reading.
- Harrouni S, Guessoum A, Maafi A (2005) Classification of daily solar irradiation by fractional analysis of 10-min-means of solar irradiance, *Theo. Appl. Climat.* 80: 27-36.
- Harman H.H (1960) *Modern factor analysis*. Chicago: University of Chicago Press.
- Heintz A, Cazal J.D and others (2015) Technical design and characterization of a ground based solar metrology network on Reunion Island. *Third Southern African Solar Energy Conference*, 578-591.
- Heinemann D, Luther J (1990) Control of renewable energy systems using Bayesian forecasting techniques and stochastic dynamic programming. *Clean and Safe Energy Forever*, A volume in International Solar Energy Society Proceedings Series, 407-411.
- Hocaoğlu F.O (2011) Stochastic approach for daily solar radiation modeling. *Solar Energy*, 85(2): 278-287.
- Hotelling H (1933) Analysis of a Complex of Statistical Variables Into Principal Components, *Journal of Educational Psychology*, 24: 417-441 & 498-520.
- Huang N. E, Shen Z, Long S. R and others (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. *Proc. R. Soc. London, Ser. A*, 454: 903-995.
- Huang N. E, Shen Z, Long S.R and others (1999) A new view of nonlinear water waves: The Hilbert Spectrum. *Annu. Rev. Fluid Mech.*, 31(1): 417-457.
- Huang N. E, Wu M.L, Long S.R and others (2003a) A confidence limit for the empirical mode decomposition and Hilbert spectral analysis. *Proc. R. Soc. London, Ser. A*, 459(2037): 2317-2345.
- Huang N.E, Wu M.L, Qu W, Long S.R and Shen S.S.P (2003b) Applications of Hilbert-Huang transform to non-stationary financial time series analysis. *Appl. Stoch. Model Bus.*, 19(3): 245-268.

- Huang N.E, Attoh-Okine N.O (2005) *The Hilbert-Huang Transform in Engineering*. CRC Press.
- Huang N. E (2005) *Hilbert-Huang Transform and Its Applications*. World Scientific. Chap1. Introduction to the Hilbert-Huang transform and its related mathematical problems, 1-26.
- Huang N.E, Shen S.S.P (2005) *Hilbert-Huang Transform and Its Applications*. Interdisciplinary Mathematical Sciences, Volume 5.
- Huang Y.X, Schmitt F.G, Lu Z and Liu Y (2008) An amplitude-frequency study of turbulent scaling intermittency using Hilbert spectral analysis. *Europhys. Lett.* 84.
- Huang Y.X, Schmitt F.G, Lu Z and Liu Y (2009a) Analyse de l'invariance d'échelle de séries temporelles par la décomposition modale empirique et l'analyse spectrale de Hilbert. *Traitement du Signal*, 25: 481-492.
- Huang Y.X, Schmitt F.G, Lu Z, and Liu Y (2009b) Analysis of Daily River Flow Fluctuations Using Empirical Mode Decomposition and Arbitrary Order Hilbert Spectral Analysis. *J. Hydrol.*, 373: 103-111.
- Huang Y.X, Schmitt F.G, Lu Z, and Liu Y (2009c) Autocorrelation function of velocity increments in fully developed turbulence. *Europhys. Lett.* 86.
- Huang Y.X, Schmitt F.G, Lu Z and Liu Y (2010a) Arbitrary order Hilbert spectral analysis for time series possessing scaling statistics: definition and validation. *Phys. Rev. E* (submitted).
- Huang Y.X, Schmitt F.G, Lu Z and Liu Y (2010b) Second order structure function in fully developed turbulence. *Phys. Rev. E* (submitted).
- Huld T, Salis E, Pozza A and others (2016) Photovoltaic energy rating data sets for Europe. *Solar Energy*, 133: 349-362.
- Husak G.J, Michaelsen J, Chris F (2006) Use of the gamma distribution to represent monthly rainfall in Africa for drought monitoring applications. *Int.J.Climatol* 27(7): 935-944.
- Hussain M, Rahman L, Rahman M.M (1999) Estimation of solar radiation for use in crop modeling. *Agric. Forest Meteorol* 91: 293-300.
- Ihlen E.A.F (2012) Introduction to Multifractal Detrended Fluctuation Analysis in Matlab. *Front Physiol*, 3: 141.
- IPCC, *Climate Change 2014 Synthesis Report*, 2014.
- Jia B, Xie Z, Dai A, Shi C, Chen F (2013) Evaluation of satellite and reanalysis products of downward surface solar radiation over East Asia: spatial and seasonal variations. *J Geophys Res Atmos* 118. <http://dx.doi.org/10.1002/jgrd.50353>.
- Jolliffe I.T (2002) *Principal Component Analysis*, Second Edition. Springer.
- Jolliffe I.T (1986) *Principal Component Analysis and Factor Analysis*, Springer, 115-128.
- Kadilar G.O, Çakmakyapan S (2016) The Lindley family of distributions: Properties and applications. *Hacettepe University Bulletin of Natural Sciences and Engineering Series B: Mathematics and Statistics*.
- Kantelhardt J.W, Zschiegner S.A, Koscielny-Bund E and others (2002) Multifractal detrended fluctuation analysis of nonstationary time series. *Physica A: Statistical Mechanics and its Applications*, 87-114.

- Kantelhardt J.W (2008) Fractal and Multifractal Time Series. Springer's Encyclopedia of Complexity and System Science.
- Kaplanis S, Kaplani E (2007) A model to predict expected mean and stochastic hourly global solar radiation $I(h;n_j)$ values. *Renewable Energy*, 32(8): 1414-1425.
- Kaplanis S, Kaplani E (2010) Stochastic prediction of hourly global solar radiation for Patra, Greece. *Applied Energy*, 87(12): 3748-3758.
- Karakoti I, Das P.K, Singh S.K (2012) Predicting monthly mean daily diffuse radiation for India. *Applied Energy*, 91: 412-425.
- Kennedy A, Dong X, Xi B and others (2011) A comparison of MERRA and NARR reanalyses with the DOE ARM SGP data. *J Clim*, 24: 4541-4557.
- Kendall M (1976) *Time Series*, Charles Griffin.
- Kida S (1991) Log stable distribution and intermittency of turbulence. *J. Phys. Soc. Jpn.* 60: 5-8.
- Kolmogorov A.N (1941) On the degeneration of isotropic turbulence in an incompressible viscous fluid. *Russian mathematicians in the 20th century*, 332-336.
- Kothe S, Ahrens B (2010) On the Radiation Budget in Regional Climate Simulations for West Africa. *J Geophys Res* 115: D23120.
- Kothe S, Dobler A, Beck A and Ahrens B (2011) The radiation budget in a regional climate model. *Clim Dyn*, 36: 1023-1036.
- Kutty H.A, Masral M.H, and Rajendran P (2015) Regression model to predict global solar irradiance in Malaysia. *International Journal of Photoenergy*, Volume 2015.
- Kutzbach J.E (1967) Empirical Eigenvectors of Sea-Level Pressure, Surface Temperature and Precipitation Complexes over North America. *Journal of Applied Meteorology*, 6: 791-802.
- Lara-Fanego V, Ruiz-Arias J.A, Pozo-Vázquez D, and others (2012) Evaluation of the WRF model solar irradiance forecasts in Andalusia (southern Spain). *Solar Energy*, 86: 2200-2217.
- Laszlo I, Ciren P, Liu H and others (2008) Remote sensing of aerosol and radiation from geostationary satellites. *Adv. Space Res* 41: 1882-1893.
- Lave M, Kleissl J (2010) Solar variability of four sites across the state of Colorado. *Renewable Energy*, 35(12): 2867-2873.
- Lefèvre M, Diabaté L, Wald L (2007) Using reduced data sets ISCCP-B2 from the Meteosat satellites to assess surface solar irradiance. *Solar Energy*, 81: 240-253.
- Li H, Ma W.B, Wang X.L, Lian Y.W (2011) Estimating monthly average daily diffuse solar radiation with multiple predictors: a case study. *Renewable Energy*, 36: 1944-1948.
- Li P (2015) Temporal and Spatial Variability of Surface Solar Radiation over the South-West Indian Ocean and Reunion Island: Regional Climate Modeling. *Université de la Réunion*,
- Lohmann S, Schillings C, Mayer B, Meyer R (2006) Long-term variability of solar direct and global radiation derived from ISCCP data and comparison with reanalysis data. *Solar Energy*, 80: 1390-1401.

- Long S.R, Huang N.E, Tung C.C, Wu M.L and others (1995) The Hilbert techniques: an alternate approach for non-steady time series analysis. *IEEE Geoscience and Remote Sensing Soc. Lett.*, 3: 6-11.
- Lorenz E.N, Heinemann D (2012) Prediction of solar irradiance and photovoltaic power. *Reference Module in Earth Systems and Environmental Sciences, Comprehensive Renewable Energy*, 1: 239-292.
- Lorenz E.N (1970) Climate change as a mathematical problem. *Journal of Applied Meteorology*, 9: 325-329.
- Lorenz E.N (1956) Empirical orthogonal functions and statistical weather prediction. Technical report, Statistical Forecast Project Report 1.
- Lovejoy S, Schertzer D, Stanway J.D (2001) Direct evidence of multifractal atmospheric cascades from planetary scales down to 1 km. *Phys. Rev. Lett.* 86: 5200-5203.
- Lovejoy S, Schertzer D (2006) Multifractals, cloud radiances and rain. *J. Hydrol.* 322: 59-88.
- Maafi A, Harrouni S (2003) Preliminary results of the fractal classification of daily solar irradiances, *Solar Energy*, 75 (1): 53-61.
- Mandelbrot B.B (1982) *The Fractal Geometry of Nature*.
- Markovic M, Jones C.G, Vaillancourt P.A, Paquin D and others (2008) An evaluation of the surface radiation budget over North America for a suite of regional climate models against surface station observations. *Clim Dyn* 31: 779-794.
- Marquez R, Coimbra C.F.M (2013) Intra-hour DNI forecasting based on cloud tracking image analysis. *Solar Energy*, 91: 327-336.
- Marshak A, Davis A, Cahalan A and Wiscombe W (1994) Bounded cascade models as nonstationary multifractals. *Physical Review E* 49: 55.
- Mauas P.J.D, Flamenco E, Buccino A.P (2008) Solar forcing of the stream flow of a continental scale south American river. *Phys. Rev. Lett.* 101, 168501.
- Mellit A, Benghane M, Hadj Arab A and Guessoum A (2005) A simplified model for generating sequences of global solar radiation data for isolated sites: Using artificial neural network and a library of Markov transition matrices approach. *Solar Energy*, 79(5): 469-482.
- Mohandes M, Rehman S and Halawani T.O (1998) Estimation of global solar radiation using artificial neural networks. *Renewable Energy*, 14(1-4): 179-184.
- Monforti F, Huld T, Bódis K, and others (2014) Assessing complementarity of wind and solar resources for energy production in Italy: A Monte Carlo approach. *Renewable Energy*, 63:576-586.
- Monjoly S, Andre M, Calif R and Soubdhan T (2017) Hourly forecasting of global solar radiation based on multiscale decomposition methods: A hybrid approach. *Energy*, 119:288-298.
- Morel B, Pohl B, Richard Y and others (2014) Regionalizing rainfall at very high resolution over La Réunion Island using a regional climate model. *Mon Wea Rev.* 142: 2665-2686.

- Movahed M.S, Jafari G.R, Ghasemi F and others (2006) Multifractal detrended fluctuation analysis of sunspot time series. *J. Stat. Mech.* 0602:P02003 10.1088:1742-5468.
- Muselli M, Poggi P, Notton G, Louche A (2000) Classification of typical meteorological days from global irradiation records and comparison between two Mediterranean coastal sites in Corsica Island. *Energy Conversion and Management*, 41:1043-1063.
- Murshed M.S, Seo Y.A, Park J.S, Lee Y (2018) Use of beta-P distribution for modelling hydrologic events. *Communications for Statistical Applications and Methods* 25: 15-27.
- Nalina U, Prema V, Smitha K and Urna Rao K (2014) Multivariate regression for prediction of solar irradiance. *Data Science & Engineering (ICDSE)*, 2014 International Conference.
- Nunn N (2009) The Importance of History for Economic Development, *Annual Review of Economics*, 1:65-92.
- Nurse L.A, McLean R.F and others, Small islands. In: *Climate Change 2014: Impacts, Adaptation, and Vulnerability*, 2014.
- Obukhov A.M (1960) The statistically orthogonal expansion of empirical functions. *Bull. Acad. Sci. USSR Geophys. Ser. (English Transl.)*, 288-291.
- Obukhov A.M (1947) Statistically homogeneous fields on a sphere. *Usp. Mat. Navk.*, 2:196-198.
- Oehler J.F, Lénat J.F, Labazuy P (2008) Growth and collapse of the Reunion Island volcanoes. *Bull Volcanol*, 70: 717-742.
- Ohmura A, Dutton E.G, Forgan B and others (1998) Baseline Surface Radiation Network (BSRN/WCRP): New Precision Radiometry for Climate Research. *Bulletin of the American Meteorological Society*. Vol. 79.
- Ohmura A, Gilgen J, Wild M (1989) Global Energy Balance Archive, GEBA.
- Omer A.M (2008) Energy, environment and sustainable development. *Renewable and Sustainable Energy Reviews*, 9: 2265-2300.
- Paoli C, Voyant C, Muselli M, Nivet M.L (2010) Forecasting of preprocessed daily solar radiation time series using neural networks. *Sol. Energy*, 84:2146–2160.
- Parisi G, Frisch U (1985) On the singularity structure of fully developed turbulence in turbulence and Predictability in Geophysical Fluid Dynamics. *Proceedings of the International School of Physics “Enrico Fermi”*, 84-87.
- Pearson K (1901) On lines and planes of closest fit to systems of points in space, *Philosophical Magazine*, Series 6, 2(11): 559-572.
- Peng li (2015) Temporal and Spatial Variability of Surface Solar Radiation over the South-West Indian Ocean and Reunion Island: Regional Climate Modeling. *Solar and Stellar Astrophysics*. Université de la Réunion.
- Perez T, Fossen H.T (2011) Practical aspects of frequency-domain identification of dynamic models of marine structures from hydrodynamic data. *Ocean Engineering*, 38(2-3):426-435.
- Pinker R.T, Laszlo I (1992) Modeling Surface Solar Irradiance for Satellite Applications on a Global Scale. *Journal of Applied Meteorology*, 31: 194-211.

- Posselt R, Mueller R, Trentmann J and others (2012a) Extension of the CM SAF surface radiation climatology beyond Meteosat first generation satellites. In: 12th EMS Annual Meeting & 9th European Conference on Applied Climatology (ECAC), Poland.
- Posselt R, Müller R, Stöckli R, Trentmann J (2011) Spatial and temporal homogeneity of solar surface irradiance across satellite generations. *Remote Sensing* 3(5): 1029-1046.
- Pradhan B, Lee S (2010a) Delineation of landslide hazard areas using frequency ratio, logistic regression and artificial neural network model at Penang Island, Malaysia. *Environmental Earth Sciences* 60: 1037-1054.
- Pradha B, Lee S (2010b) Regional landslide susceptibility analysis using backpropagation neural network model at Cameron Highland, Malaysia. *Landslides*, 7:13-30.
- Pradhan B, Lee S (2010c) Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modeling. *Environmental Modelling and Software*, 25: 747-759.
- Premalatha N, A Valan Arasu A (2016) Prediction of solar radiation for solar systems by using ANN models with different back propagation algorithms. *J. appl. res. Technol*, 14 (3).
- Radić V, Pasarić Z, Šinik N (2004) Analysis of Zagreb climatological data series using empirically decomposed intrinsic mode functions. *Geofizika*, 21(1).
- Reddy K.S, Ranjan M (2003) Solar resource estimation using artificial neural networks and comparison with other correlation models. *Energy Conversion and Management*, 44(15): 2519-2530.
- Rehman N, Mandic D. P (2010) Empirical mode decomposition for trivariate signals. *IEEE T. Signal Process.*
- REN21 (2010). *Renewables 2010 Global Status Report*, 15.
- Reno M.J, Hansen C.W, Stein J.S (2012) *Global Horizontal Irradiance Clear Sky Models: Implementation and Analysis*. Sandia report, SAND, 2012-2389.
- Rilling G, Flandrin P, Gonçalvès P (2003) On empirical mode decomposition and its algorithms. *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing*.
- Rilling G, Flandrin P, Goncalves P, Lilly J.M (2007) Bivariate empirical model decomposition. *IEEE Signal Process. Lett.* 14:936-939.
- Robert T, Dervins D, Fabre F and others (2006) Mrc1 and Srs2 are major actors in the regulation of spontaneous crossover. *EMBO J* 25(12):2837-46.
- Roosen, R. G, Angione R.J and Klemeke C.H (1973) Worldwide variations in atmospheric transmission: 1. Baseline results from Smithsonian observations, *Bull. Am. Meteorol. Soc.*, 54: 307-316.
- Roweis S.T, Saul L.K (2000) Nonlinear Dimensionality Reduction by Locally Linear Embedding. *Science*, 290(5500).
- Ruiz-Arias J.A, Dudhia J, Santos-Alamillos F.J, Pozo-Vázquez D (2013a) Surface clear-sky shortwave radiative closure intercomparisons in the weather research and forecasting model. *J Geophys Res Atmos*, 118 (17): 9901-9913.

- Ruiz-Arias J.A, Pozo-Vázquez D, Lara-Fanego V, Santos-Alamillos F.J, Tovar-Pescador J (2011b) A high-resolution topographic correction method for clear-sky solar irradiance derived with a numerical weather prediction model. *J Appl Meteorol Clim*, 50(12): 2460-2472.
- Sabziparvar A.A (2008) A simple formula for estimating global solar radiation in central arid deserts of Iran. *Renewable Energy*, 33(5): 1002-1010.
- Seuront L, Schmitt F, Schertze D and others (1996a) Multifractal intermittency of Eulerian and Lagrangian turbulence of ocean temperature and plankton fields. *Nonlinear Processes in Geophysics*, European Geosciences Union (EGU), 3 (4): 236-246.
- Seuront L, Schmitt F, Lagadeuc Y and others (1996b) Multifractal analysis of phytoplankton biomass and temperature in the ocean. *Geophysical*, 23: 3591-3594.
- Schertzer D, Lovejoy S (1987) Physical modeling and analysis of rain and clouds by anisotropic scaling multiplicative processes. *J. Geophys. Res.* 92 (D8): 9693-9714.
- Schertzer D, Lovejoy S, Schmitt F, Chigirinskaya Y and Marsan D (1997) Multifractal cascade dynamics and turbulent intermittency. *Fractals* 5 (3): 427-471.
- Schmitt F.G, Schertzer D, Lovejoy S (1999) Multifractal analysis of foreign exchange data. *Appl. Stoch. Models Data Anal.* 15 (1): 29-53.
- Schwartz M, Bennett W.R, Stein S (1966) *Communications systems and techniques*. New York: McGraw-Hill.
- Shekel J (1953) 'Instantaneous' frequency. *Proc. IRE* 41: 548.
- Sirovich L, Maxey M, Tarman H (1987) Analysis of turbulent thermal convection, *Proc. 6th Symposium on Turbulent Shear Flow*.
- Sirovich L (1987) Turbulence and the dynamics of coherent structures, Pt. I: Coherent Structures. *Quar. Appl. Math.*, Vol XLV, No. 3, 561-571.
- Sirovich L (1987) Turbulence and the dynamics of coherent structures, Pt. II: Symmetries and transformations, *Quar. Appl. Math.*, Vol. XLV, No. 3, 573-582.
- Sirovich L (1987) Turbulence and the dynamics of coherent structures, Pt. III: Dynamics and scaling, *Quar. Appl. Math.*, XLV(3): 583-590.
- Sivamadhavi V, Samuel Selvaraj R (2012) Prediction of monthly mean daily global solar radiation using Artificial Neural Network. *Journal of Earth System Science*, 121 (6): 1501-1510.
- Soubdhan T, Emilion R, Calif R (2009) Classification of daily solar radiation distributions using a mixture of dirichlet distributions, *Solar Energy*, 83 (7) :1056-1063.
- Stan C, Cristescu M.T, Buimaga-Iarinca L, Cristescu C.P (2013) Investigation on series of length of coding and non-coding DNA sequences of bacteria using multifractal detrended cross-correlation analysis. *J Theor Biol*, 321:54-62.
- Storch H.V, Zwiers F.W (1999) *Statistical analysis in climatology*. Cambridge University Press.
- Storch H.V, Zorita E, and Cubasch U (1993) Downscaling of Global Climate Change Estimates to Regional Scales: An Application to Iberian Rainfall in winter time. *Journal of Climate*. 6: 1161-1171.

- Stothers R. B. (1996), Major optical thickness perturbations to the stratosphere from volcanic eruptions: Pyrhelometric period, 1881-1960, *J. Geophys. Res.*, 101: 3901-3920.
- Sreenivasan K, Antonia, R (1997) The phenomenology of small-scale turbulence. *Annu. Rev. Fluid Mech.* 29: 435-472.
- Srinivasan U.T (2010) Economics of climate change: risk and responsibility by world region. *Climate Policy*, 10(3): 298-316.
- Tanaka T, Mandic D. P (2006) Complex empirical mode decomposition. *IEEE Signal Process. Lett.* 14: 101-104.
- Tarroja B, Mueller F, Samuelsen S (2013) Solar power variability and spatial diversification: implications from an electric grid load balancing perspective. *Int. J. Energy Res.* 37: 1002-1016.
- Telesca L, Colangelo G, Lapenna V and others. (2003) Monofractal and multifractal characterization of geoelectrical signals measured in southern Italy. *Chaos, Solitons & Fractals*, 18(2) 385-399.
- Titchmarsh E.C (1948) Introduction to the theory of Fourier integrals. Oxford University Press.
- Ulgen K, Hepbasli A (2009) Diffuse solar radiation estimation models for Turkey's big cities. *Energy Conversion and Management*, 50: 149-156.
- Veeran P.K, Kumar S (1993) Analysis of monthly average daily global radiation and monthly average sunshine duration at two tropical locations. *Renewable Energy*, 3(8): 935-939.
- Voyant C, Haurant P, Muselli M and others (2014) Time series modeling and large scale global solar radiation forecasting from geostationary satellites data. *Solar Energy*, Elsevier.
- Vulpiani A, Livi R (2004) *The Kolmogorov Legacy in Physics*. Springer, Berlin, Heidelberg, New York.
- Wild M, Truessel B, Ohmura A and others (2009) Global dimming and brightening: an update beyond 2000. *J Geophys Res Atmos* , 114: D00D13.
- Wild M, Folini D, Schär C and others (2013) The global energy balance from a surface perspective. *Clim Dyn*, 40: 3107-3134.
- Wild M, Ohmura A and others (2017) The Global Energy Balance Archive (GEBA) version 2017: a database for worldwide measured surface energy fluxes. *Earth Syst. Sci. Data*, 9: 601-613.
- Wilks P.A.D, English M.J (1995) A system for rapid identification of respiratory abnormalities using a neural network. *Medical Engineering & Physics*, 17(7): 551-555.
- Wu M, Hu C (2006) Empirical mode decomposition and synchro gram approach to cardiorespiratory synchronization. *Phys. Rev. E* 73:051917.
- Wyser K, Jones C, Du P, and others (2008) An evaluation of Arctic cloud and radiation processes during the SHEBA year: simulation results from eight Arctic regional climate models. *Clim Dyn*, 30(2-3): 203-223.
- Yohanna J.K, Itodo I.N, Umogbi V.I (2011) A model for determining the global solar radiation for Makurdi, Nigeria. *Renewable Energy*, 36: 1989-1992.

Zamora R.J, Dutton E.G, Trainer M, McKeen S.A, Wilczak J.M, Hou Y.T (2005) The accuracy of solar irradiance calculations used in mesoscale numerical weather prediction. *Mon Weather Rev*, 133: 783-792.

Appendix A --- List of Acronyms

| | |
|----------|--|
| AF | Adjusted Factor |
| ANN | Artificial Neural Network |
| BSRN | Baseline Surface Radiation Network |
| C_1 | Fractal Co-dimension |
| CCA | Canonical Correlation Analysis |
| CIMO | Commission Instruments and Methods of Observation |
| CM SAF | Satellite Application Facility on Climate Monitoring |
| CWT | Continuous wavelet transform |
| DC | Direct Current |
| DJF | Dec-Jan-Feb |
| DNI | Direct Normal Irradiance |
| DHI | Diffuse Horizontal Irradiance |
| DOG | Derivative of Gaussian |
| DWT | Discrete Wavelet Transform |
| DT-CWT | Discrete-time Continuous Wavelet Transform |
| EMD | Empirical Mode Decomposition |
| EDR | Environmental Data Record |
| EUMETSAT | Exploitation of Meteorological Satellites |
| EVD | Eigenvalue Decomposition |
| EVP | Eigenvalue Problem |
| FCRD | Fundamental Climate Data Record |
| FT | Fourier Transform |
| GEBA | Global Energy Balance Archive |
| GOES | Geostationary Operational Environmental Satellites |
| GHG | Green-House Gases |
| GHI | Global Horizontal Irradiance |
| GISS | Goddard Institute for Space Studies |
| GMS | Geostationary Meteorological Satellite |
| H | Hurst exponent |
| HHT | Hilbert Huang Transform |
| HSA | Hilbert Spectral Analysis |
| JJA | Jun-Jul-Aug |
| MAM | Mar-Apr-May |
| MAE | Mean Absolute Error |

| | |
|--------|---|
| MBL | Marine Boundary Layer |
| MLR | Multiple Linear Regression |
| MSE | Mean Square Error |
| MVIRI | Meteosat Visible and Infrared Imager |
| netCDF | Network Common Data Format |
| NWP | Numerical Weather Prediction |
| ICDR | Interim Climate Data Record |
| IEC | International Engineering Conference |
| IMFs | Intrinsic Mode Functions |
| INSAT | Indian National Satellite |
| IODC | Indian Ocean Data Coverage |
| IPCC | Intergovernmental Panel on Climate Change |
| ITCZ | Intertropical Convergence Zone |
| k-NN | k-Nearest Neighbours |
| KLT | Karhunen–Loève Transform |
| PCA | Principal Component Analysis |
| PCR | Principal Component Regression |
| PDF | Probability Density Function |
| PLSR | Partial Least Squares Regression |
| POD | Proper Orthogonal Decomposition |
| PV | Photovoltaic |
| RCMs | Regional Climate Models |
| RMSE | Root Mean Square Error |
| SAF | Satellite Application Facility |
| SARAH | Surface Solar Radiation Data Records Heliosat |
| SD | Standard Deviation |
| SID | Surface Indoming Direct Irradiance |
| SIS | Surface Indoming Shortwave Irradiance |
| SON | Sep-Oct-Nov |
| SR | Stepwise Regression |
| SSR | Surface Solar Radiation |
| SVD | Singular Value Decomposition |
| TCRD | Thematic Climate Data Record |
| WMO | World Meteorological Organization |

Appendix B --- Conferences Attended

SASEC2015 Third Southern African Solar Energy Conference 11 – 13 May 2015 Kruger National Park, South Africa

Appendix C --- List of Publications

Li Q, Bessafi M, Delage O, Chabriat JP, Li P (2015) Intermittency study of global solar radiation on Reunion Island using Hilbert-Huang Transform. Third Southern African Solar Energy Conference, 11-13 May, Kruger National Park, South Africa.

INTERMITTENCY STUDY OF GLOBAL SOLAR RADIATION ON REUNION ISLAND USING HILBERT-HUANG TRANSFORM

LI Qi, BESSAFI Miloud, DELAGE Olivier, CHABRIAT Jean-Pierre, LI Peng

Laboratory of Energy, Electronic and Process (LE²P),
 University of Reunion Island,
 15 Avenue Rene Cassin,
 CS 92003, 97 744 Saint-Denis Cedex 9,
 Reunion Island, France

E-mail: richie.qili@gmail.com

ABSTRACT

The characterization of the solar radiation variability is a fundamental step before prediction and is crucial to transform an intermittent source of energy into a stable one. As solar irradiance result from a nonlinear and non-stationary process, we use a multifractal approach based on the Hilbert-Huang Transform (HHT) consisting of an empirical mode decomposition (EMD) followed by a spectral analysis. In this paper, we will briefly introduce the HHT data analysis method. Such a recent adaptive data analysis method has been applied to Reunion Island global solar radiation time series of measurements with a sampling rate of 1/60 Hz over six years. Through the EMD, daily global solar radiation data were decomposed into several intrinsic mode functions (IMF). For each IMF, we estimate the amplitude, instantaneous frequency and Hilbert spectrum for the original data. From the comparison of Hilbert Spectrum and Fourier Spectrum, we find the calculated solar radiation power spectrum follows a power law behavior close to the Kolmogorov law. The method described in this paper provides an amplitude frequency representation of the global solar radiation sequences resulting in a probability density function and a scaling coefficient. The multifractal approach allows to extract parameters connected to the multifractal properties of the global solar radiation.

INTRODUCTION

With the rapid development of the global economy and society, requirements for renewable energy increase remarkably. Solar energy is considered as one of the most promising alternative energy resources. However, the spatial and temporal variability of solar radiation over various time scales is difficult for computing, controlling and balancing. Solar radiation shows the characteristics, nonlinear, nonstationary and intermittency in various process [1].

Intermittency is the irregular alternation of phases of apparently periodic and chaotic dynamics or different forms of chaotic dynamics (crisis-induced intermittency) [2 3]. In the area of solar radiation, Tarroja et al. give the concept of intermittency the same meaning that to the concept of fluctuation, defining the severity of the intermittency as the change in the magnitude of the total irradiation on a surface over a given time interval [4]. Thus, this meaning is agreed with the concept of variability. On the contrary, Davis et al. import the notion of intermittency from turbulence, though they analyze radiation data artificially generated from cloud data [5]. This one is the sense given to the concept of intermittency in the present work, that is, intermittency refers to the changes of the variability of daily solar radiation according to the scale considered. In this sense, intermittency and multifractality are synonym.

Intermittency and multiscaling properties have been found in many fields, such as rainfall [11], finance [9 10], turbulence [6-8 15], and geophysical fields [12 13 28].

Multiscaling intermittency is often characterized using a structure function of order $q > 0$ as the statistical moment of the fluctuations $\Delta X(\tau) = |X(t + \tau) - X(t)|$ [6 15]

$$\langle (\Delta X(\tau))^q \rangle \sim C_q \tau^{\zeta(q)} \quad (1)$$

where C_q is a constant and $\zeta(q)$ is a scale invariant moment function; it is also a cumulant generating function, which is nonlinear and concave and fully characterizes the scale-invariant properties of intermittency.

For this global solar radiation study, we consider the scaling exponent $\zeta(q)$ that characterizes the scaling behavior or measures the distance between a monofractal and multifractal

LETTRE D'ENGAGEMENT DE NON-PLAGIAT

Je, soussigné(e) **QILI** en ma qualité de doctorant(e) de l'Université de La Réunion, déclare être conscient(e) que le plagiat est un acte délictueux passible de sanctions disciplinaires. Aussi, dans le respect de la propriété intellectuelle et du droit d'auteur, je m'engage à systématiquement citer mes sources, quelle qu'en soit la forme (textes, images, audiovisuel, internet), dans le cadre de la rédaction de ma thèse et de toute autre production scientifique, sachant que l'établissement est susceptible de soumettre le texte de ma thèse à un logiciel anti-plagiat.

Fait à Saint-Denis le : 03/06/2018

Signature :

 Qili

Extrait du Règlement intérieur de l'Université de La Réunion
(validé par le Conseil d'Administration en date du 11 décembre 2014)

Article 9. Protection de la propriété intellectuelle – Faux et usage de faux, contrefaçon, plagiat

L'utilisation des ressources informatiques de l'Université implique le respect de ses droits de propriété intellectuelle ainsi que ceux de ses partenaires et plus généralement, de tous tiers titulaires de ces droits.

En conséquence, chaque utilisateur doit :

- utiliser les logiciels dans les conditions de licences souscrites ;
- ne pas reproduire, copier, diffuser, modifier ou utiliser des logiciels, bases de données, pages Web, textes, images, photographies ou autres créations protégées par le droit d'auteur ou un droit privatif, sans avoir obtenu préalablement l'autorisation des titulaires de ces droits.

La contrefaçon et le faux

Conformément aux dispositions du code de la propriété intellectuelle, toute représentation ou reproduction intégrale ou partielle d'une œuvre de l'esprit faite sans le consentement de son auteur est illicite et constitue un délit pénal.

L'article 444-1 du code pénal dispose : « Constitue un faux toute altération frauduleuse de la vérité, de nature à causer un préjudice et accomplie par quelque moyen que ce soit, dans un écrit ou tout autre support d'expression de la pensée qui a pour objet ou qui peut avoir pour effet d'établir la preuve d'un droit ou d'un fait ayant des conséquences juridiques ».

L'article L335_3 du code de la propriété intellectuelle précise que : « Est également un délit de contrefaçon toute reproduction, représentation ou diffusion, par quelque moyen que ce soit, d'une œuvre de l'esprit en violation des droits de l'auteur, tels qu'ils sont définis et réglementés par la loi. Est également un délit de contrefaçon la violation de l'un des droits de l'auteur d'un logiciel (...) ».

Le plagiat est constitué par la copie, totale ou partielle d'un travail réalisé par autrui, lorsque la source empruntée n'est pas citée, quel que soit le moyen utilisé. Le plagiat constitue une violation du droit d'auteur (au sens des articles L 335-2 et L 335-3 du code de la propriété intellectuelle). Il peut être assimilé à un délit de contrefaçon. C'est aussi une faute disciplinaire, susceptible d'entraîner une sanction.

Les sources et les références utilisées dans le cadre des travaux (préparations, devoirs, mémoires, thèses, rapports de stage...) doivent être clairement citées. Des citations intégrales peuvent figurer dans les documents rendus, si elles sont assorties de leur référence (nom d'auteur, publication, date, éditeur...) et identifiées comme telles par des guillemets ou des italiques.

Les délits de contrefaçon, de plagiat et d'usage de faux peuvent donner lieu à une sanction disciplinaire indépendante de la mise en œuvre de poursuites pénales.