博士論文

An stochastic approach for the multiscale analysis and reproduction of rainfall data at high spatial resolutions

(確率手法を利用した高い空間分解能の
降雨情報の多重スケール解析と再生)

ゴメス　ガルシア　マルティン

THE UNIVERSITY OF TOKYO

A stochastic approach for the multiscale analysis and reproduction of rainfall data at high spatial resolutions

Author

Martín Gómez García Alvéstegui

M.Eng., The University of Tokyo, 2011

B.Sc., Universidad Privada Boliviana, 2006

Dissertation Committee:

Professor Toshio KOIKE, Chairman

Professor Taikan OKI

Professor Kei YOSHIMURA

Professor Akiyuki KAWASAKI

Professor Keisuke HANAKI

Professor Yukari TAKAYABU

September 2016

A stochastic approach for the multiscale analysis and reproduction of rainfall data at high spatial resolutions

Author

Martín Gómez García Alvéstegui

Student ID: 37-137173

Thesis submitted in partial fulfillment
of the requirements for the degree of

Doctor of Philosophy

Department of Civil Engineering
Graduate School of Engineering

Approved by:

Prof. Toshio KOIKE
Committee Chairman

Prof. Taikan OKI

Prof. Kei YOSHIMURA

Prof. Akiyuki KAWASAKI

Prof. Keisuke HANAKI

Prof. Yukari TAKAYABU

September 2016

Publications

1. GÓMEZ GARCÍA, M., T. KOIKE, 2012. Multiscaling analysis of satellite-derived precipitation estimates linked to in-situ measurements. Annual Journal of Hydraulic Engineering, JSCE. 68 (4), I_169 – I_174.
2. GÓMEZ GARCÍA, M., T. KOIKE, 2016. Parameter-estimation methods for symmetric stable distributions: application to small samples of spatial fluctuations of rainfall. Spa Sta. 17, 50 – 70.

International Conferences

3. GÓMEZ GARCÍA, M., T. KOIKE, Inter-Scale Statistical Analysis of Fine-Resolution Rainfall Datasets over the Japanese Islands, EGU General Assembly, 12-17 April 2015.

ABSTRACT

Modeling the water cycle has become a more comprehensive procedure due to the developments in computer capability and the scientific understanding of the involved processes. However, because hydrological models and numerical weather prediction models are mostly mathematical solutions of nonlinear equations, the simulations are sensitive to spatial and temporal variations of the input data, being rainfall intensity one that causes considerable impact (Bell and Moore, 2000; Beven, 2001; Arnaud et al, 2002, 2011; Smith et al., 2004; Schuurmans and Bierkens, 2007; Younger et al., 2009; Liu et al., 2012; Song et al., 2015; Weijian et al., 2015). Despite the scientific contribution of physically-based models, there still remains the problem of spatial distribution of field measurements and the consequent sub-grid parameterization of land properties and atmospheric forcing data at small scales (high resolutions). Therefore, the use of this kind of models becomes futile in regions without the resources to afford the installation and operation of weather radars or dense meteorological stations.

Satellite-based products are an alternative source of estimations of rainfall intensity. The validation of these products is mostly done by comparing them with rain gauge measurements and, if available, with weather radars at various temporal and spatial scales. As a result, systematic errors, which generally show regional and seasonal trends, were identified by many researchers. The number of validation studies is probably immensurable considering the many available satellite-based products, and that only a fair number of rain gauge stations are needed to conduct such study (Ebert et al., 2007; Tian et al., 2007; Hossain and Huffman, Tian et al., 2009; Pombo et al., 2015; Rudlosky et al., 2016; and Maggioni et al., 2016). Motivated by how important it is to have accurate rainfall fields at high-resolutions, even in poorly gauged regions, this research aimed to conduct a stochastic analysis of the spatial structure of rainfall before establishing a strategy for the combination of the rain gauge measurements and satellite products. In this sense, Chapter 1 introduces the methodology and establishes the specific objectives of the research.

Rainfall intensity portrays distinctive spatial features which include intermittency, abrupt discontinuities, parallel bands and high-intensity clusters surrounded by consecutive larger areas of each time lower-intensity. One way of studying this extreme variability is to perform a multiscale stochastic characterization of the rainfall process. The hypothesis is that certain statistical aspects have a scale-invariant behavior. A particular kind of scale-invariance is called self-similarity or simple-scaling, which implies that a single parameter can describe the inter-scale relationship of the probability distributions. If a process is qualified as self-similar in distribution, it means that the type of distribution does not change in spite of scaling transformations. Early stochastic models representing multiscale variability of rainfall were based on empirical evidence (radar-based estimations of intensity) that self-similarity was present in certain statistical properties of the rainfall process.

Kumar and Foufoula-Georgiou (1993a, 1993b) proposed to conduct the multiscale analysis by decomposing the rainfall process into large- and small- scale components, where the small-scale components represent the details (fluctuations of intensity) that differentiate a process at high-resolution from the same process at lower resolutions. For this purpose, Chapter 2 makes a description of the filters used for the decomposition, which have been largely used in computer sciences for image processing and edge detection. A benefit of using these filters is that the algorithm can be easily inverted, which means that rainfall at high resolutions can be reproduced from low-resolution rainfall and information about the fluctuations.

Rainfall intensity fields are normally discretized into two-dimensional grids for representation purposes. The filters described in Chapter 2 are used in Chapter 3 with the purpose of characterizing the components of the rainfall field and prove the presence of self-similarity.

Specifically, the two–dimensional Discrete Wavelet Transform extracts directional wavelet coefficients using orthonormal bases constructed from wavelet functions. The probability distributions of the wavelet coefficients depict a large amount of values around zero (as a result of adjacent boxes with similar intensity) and positive and negative outliers representing the abrupt jumps of intensity. Because of these two conditions, distributions with non-Gaussian behavior (thick tails), like symmetric α-stable distributions, have proved to be a better fit (Perica and Foufoula-Georgiou, 1996; Ebtehaj and Foufoula-Georgiou, 2011). Perica and Foufoula-Georgiou (1996), however, demonstrated the presence of self-similarity in the distributions of standardized wavelet coefficients. Standardization was defined as the wavelet coefficients divided by their corresponding low-pass coefficients (local means). The standardization of wavelet coefficients, which flattened the tails making the distributions quasi-Gaussian, allowed to qualify the fluctuation processes as self-similar.

In this study, we revise the adjustment of the distributions of wavelet coefficients to symmetric α-stable distributions, which are described by two parameters: the characteristic exponent α, and the scale parameter γ. The characteristic exponent α determines the frequency of extreme values, and hence, it is an indicator of the degree of variability. We handle the hypothesis that this parameter can be an indicator of the degree of spatial variability of rainfall fields. Two conditions are necessary to demonstrate the existence of self-similarity when considering symmetric α-stable distributions. First, the characteristic exponent α should be the same at all scales, meaning that the type of distribution is invariant under these transformations. Second, the scale parameter γ should have exponential growth with a constant rate (power-law behavior) as the scale becomes larger (lower resolutions).

The estimation of the parameters of α-stable distributions is hampered by the absence of closed-formed expressions for the probability density function, which means that it can only be expressed as an infinite series or using improper integrals. Consequently, common methods of estimation (e.g., method of moments or maximum likelihood) cannot be applied

and alternative techniques have been developed. These techniques usually require large samples with values of frequency evenly located along the real line. Unfortunately, samples of wavelet coefficients usually show a large amount of near-zero values at small scales, and (the samples) are greatly reduced at large-scales as a result of the scaling transformations. These two unfavorable situations happen to affect the existing estimation techniques, as they either fail to give a result or generate very different values of α between scales.

A thorough study of the properties of symmetric stable distributions lead to the theoretical development of a new technique of estimation, presented in Chapter 3. Nolan (2013) defined the "amplitude" of multivariate α-stable distributions in such way that for the univariate case the amplitude becomes a distribution of absolute values. An expression for the Fractional Absolute Moments (FAM) were presented previously by several authors (e.g., Nikias and Shao, 1995). This expression relates the fractional absolute moments to the parameters α and γ. During this study, an expression that represents the evaluation of the FAM in the cumulative distribution function of the "amplitude" (of a univariate stable distribution) was derived. This expression happened to be a function of the parameter α and independent of the scale parameter γ. Therefore, the parameter α could be estimated using the empirical FAM evaluated at the empirical cumulative distribution function. Then, the estimation of the parameter γ can be done using the FAM expression and the estimated parameter α. This technique of estimation and the existing ones were compared through a Monte Carlo simulation. The results showed that the new technique can be even more accurate than the existing ones with reduced times of computation. The performance when applied to samples of wavelet coefficients is detailed.

This research examined the possibility of retaining the degree of variability expressed by the characteristic exponent α, by demonstrating that self-similarity can be revealed in the marginal distributions of wavelet coefficients without the need to standardize them. For this purpose, wavelet coefficients were adjusted to symmetric α-stable distributions with the new technique described above. With the purpose of encompassing different types of rainfall, five analysis sites in Japan were identified for the multiscale spatial analysis of fluctuations: Kanto (KAN), South Tohoku (STO), Hokkaido (HOK), Kyushu (KYU), and Shikoku (SHI). Each domain covers a 256 km × 256 km surface. The period of analysis is between the 2006 and 2009.

During the analyzed period, Radar-AMeDAS data was produced with a 1-km resolution. The multiscale two-dimensional discrete wavelet transform was applied to these datasets. Three sets of directional wavelet coefficients are extracted at each scale representing the fluctuations of intensity that are more prominent in the longitudinal, latitudinal and diagonal direction, respectively.

The empirical probability distribution of each set of wavelet coefficients can be adjusted to a symmetric α-stable distribution. Common techniques of estimation failed to estimate the

parameters at small-scales because of the large number of near-zero fluctuations, which are generated as a result of vast areas with the same value of rainfall intensity.

Throughout the analyzed scale range, which is between 1-km to 64-km resolution, two conditions need to be fulfilled in order to qualify the rainfall fluctuation processes as self-similar. First, the estimated values of the characteristic exponent α need to be almost invariant. Second, the scale parameter γ must have a power-law behavior. The second condition can be verified by adjusting the values of γ to a log-linear curve with logarithmic slope H, which is often called the self-similarity index (Samorodnitsky and Taqqu, 1994; Embrechts and Maejima, 2000). Only adjusted curves that showed a coefficient of determination, R^2, greater than 0.75 in all three directions were accepted. He results of this analysis is presented in Chapter 3.

The multiscale two-dimensional discrete wavelet transform can be inverted with relative simplicity, allowing to construct an algorithm that is able to generate high-resolution rainfall with a spatial distribution consistent with that of the true event. The input of this algorithm is (i) the rainfall-intensity measurement of only one rain gauge, (ii) a large-scale gridded dataset containing the location of high- and low- intensity areas, and (iii) the parameters of the probability distribution of wavelet coefficients. The structure of the algorithm has two main processes: estimation of the mean rainfall intensity and disaggregation. Chapter 4 is a complete description of such algorithm and presents the validation of its structure by computing fidelity metrics.

ACKNOWLEDGEMENT

First and foremost, I would like to express my deepest gratitude to my supervisor Prof. Toshio KOIKE with whom I had the privilege of holding many fruitful discussions about the present study. His deep insight and interest in building my knowledge through constant support and timely comments kept me motivated to expand the perspective with which I perceived the research activities. His constant encouragement and guidance will always be remembered and appreciated.

I would also like to thank the review committee members, Prof. Taikan OKI, Prof. Kei YOSHIMURA, Prof. Akiyuki KAWASAKI, Prof. Keisuke HANAKI, and Prof. Yukari TAKAYABU for their comments and contribution to this study. In the same manner, my gratitude is extended to all the professorship members, whose lectures and seminars I was able to attend during the Civil Engineering Department Graduate Program.

I want to recognize the help and support of all the REEL members and colleagues who always proved a willingly support and kind friendship along the three years in which I had the pleasure of being part of this laboratory. I would like to especially thank for the kind comments and fruitful discussions I had with Prof. Takeyoshi Chibana, Prof. Yukiko Hirabayashi, Dr. Petra Koudelova, Dr. Katsunori Tamagawa, Dr. Yohei Sawada, Dr. Rie Seto, Dr. Ralph Acierto, Dr. Huiyi Bao, Dr. Maheswor Schresta, Mr. Peter Lawford, Mr. Neluwala Panduka, Mr. Toride and Mr. Asif Naseer. I would also like to thank the support from Ms. Goda, Ms. Ikeda, Ms. Kanoh and Ms. Fujita.

I feel honored and thankful for being awarded with the Japanese Government (MEXT) Postgraduate Scholarship which granted the financial assistance to cover my monthly expenses, allowing me to focus in the graduate program.

I want to express my gratitude to the source of infinite support and words of encouragement, my parents and siblings, who somehow through distance found the way to always make me feel supported.

Finally, I would like to dedicate this document to my wife, who always showed me support, understanding and love.

Contents

List of Tables

List of Figures

1. INTRODUCTION

1.1 MOTIVATION

Precipitation is regarded as a key driving force of the water cycle, and for that reason, not only it is a controlling factor of the hydrological response of catchments, but also plays an important role in the prediction of weather at all scales of atmospheric motion. Because hydrological models and numerical weather prediction models are mostly mathematical solutions of nonlinear equations, the simulations are most sensitive to spatial and temporal variations of the input (e.g., rainfall, soil moisture), causing significant uncertainties in the results (Bell and Moore, 2000; Beven, 2001; Arnaud et al, 2002, 2011; Smith et al., 2004; Younger et al., 2009; Liu et al., 2012; Song et al., 2015; Weijian et al., 2015). A "realistic" representation of rainfall is such that at high resolutions is capable of depicting accurately the characteristic small-scale sudden variations of intensity. Considering that the output of these models is used by an increasingly number of scientists from a wide-range of fields, as well as decision-makers and local stakeholders, developing "realistic" precipitation fields, therefore, has become a major concern among geophysicists and hydrologists.

Historical precipitation observations are limited to in-situ measurements where the rain gauges can be monitored on a regular basis. These measurements are considered the true value of intensity over a point location and even valid for the nearby surroundings. Consequently, dense networks of rain gauge stations are necessary to adequately portray the spatial variability of precipitation fields. Dense networks are usually available in populated areas of developed countries. Conversely, the deployment of stations in developing regions or in vast unpopulated regions, some of which are rather important for the simulation of weather conditions (e.g., deserts and rainforests), is unfortunately sparse. Continuous two-dimensional fields can be obtained using different methods of spatial interpolation (e.g., Haberlandt, 2007; Ruelland et al., 2008), however, "realistic" representations will depend on the existence of a large number of closely-located stations.

Nowadays, satellite-based products are an alternative source of rainfall intensity as a result of algorithms that combine the soundings of different kind of sensors. Even though it would be possible to estimate rainfall with only one kind of sensor, the search for more coverage, accuracy and resolution encouraged the development of multi-sensor algorithms. The validation of satellite-based products is mostly done by comparing the estimates with rain gauge measurements and, if available, with weather radars at various temporal and spatial scales. As a result, systematic errors, which generally show regional and seasonal trends, are identified. The number of validation studies is probably immensurable considering the many available satellite-based products, and that only a fair number of rain gauge stations are needed to conduct such study. Being able to identify the bias of satellite products lead to the development of bias-correction schemes aimed to produce the most "realistic" representations of rainfall fields. These schemes, which try to take advantage of the fact that

rain gauge measurements provide "true-value" measurements while satellite products provide wide coverage, stand merely as local solutions since they are usually based on specific local characteristics like quantity and relative location of the rain gauge stations. Examples of the assessment of satellite products can be found in the works by Ebert et al. (2007), Tian et al. (2007, 2009, 2013), Hossain and Huffman (2008), Sapiano and Arkin (2009), Tian et al (2009), Anagnostou et al. (2010), Gebregiorgis and Hossain (2012, 2013), Pombo et al. (2015), Rudlosky et al. (2016), and Maggioni et al. (2016). Under this framework of analysis/comparison it becomes difficult to apply the same scheme in a different location. Tang and Hossain (2010) showed an interesting analysis of how the bias assessment can be transferred from a gauged region to an ungauged region, however, it relies on a large number of observations in the gauged region and on the ungauged region having the same precipitation regime as the one of the gauged region.

Motivated by how important it is to have accurate rainfall fields at high-resolutions even in poorly gauged regions, this research tried to tackle the issue with a different approach. Ebert et al. (2007) recognized that the reproduction of precipitation fields at hydrologically relevant scales should first involve thorough analyses of rainfall patterns. With that consideration, before combining the merits of rain gauge measurements (true value of intensity) and satellite products (wide-range coverage and relative location of storms), it was considered imperative to first study the spatial structure of the different rainfall patterns.

1.2 MULTISCALE STOCHASTIC ANALYSIS OF RAINFALL

One way to study the spatial structure of rainfall patterns is to search for a mathematical dependence from one scale of representation to another. Patterns with such behavior are usually found in fractal objects. Mandelbrot (1982) stated that the degree of variability and/or fragmentation of fractals tends to be independent of scale, and proposed to use this property to describe the variability exhibited by numerous natural formations.

Empirical evidence of Mandelbrot's conjectures were presented in a study of the area-perimeter relation of clouds and rain fields, which revealed how fractal mechanisms could be potentially used to characterize the formation processes (Lovejoy, 1982). Encouraged by this exciting realization and the technological advance of ground and space-borne remote sensors in the subsequent 30 years, geophysicists devised a few multiscale-analysis techniques that could help unravel the hidden structures and erratic behavior of rainfall. The results exposed a kind of symmetry, often called scale-invariance, in which certain aspects of the process remained invariant despite the transformations of scale. Consequently, various stochastic approaches were proposed relying mainly on the scale-invariance of the exponents/generators, and accomplishing exceptional representations of the characteristic variability of rainfall (Lovejoy and Mandelbrot 1985; Lovejoy and Schertzer, 1986; Schertzer and Lovejoy, 1987; Gupta and Waymire, 1990; Lovejoy and Schertzer, 1990; Kumar and Foufoula-Georgiou, 1993a, 1993b; Tessier et al., 1993; Over and Gupta, 1996; Veneziano et al., 1996; Menabde et al., 1997; Deidda et al., 2006; Lovejoy and Schertzer,

2006; Lovejoy et al, 2008; Ebtejah and Foufoula-Georgiou, 2011; Lovejoy and Schertzer, 2013).

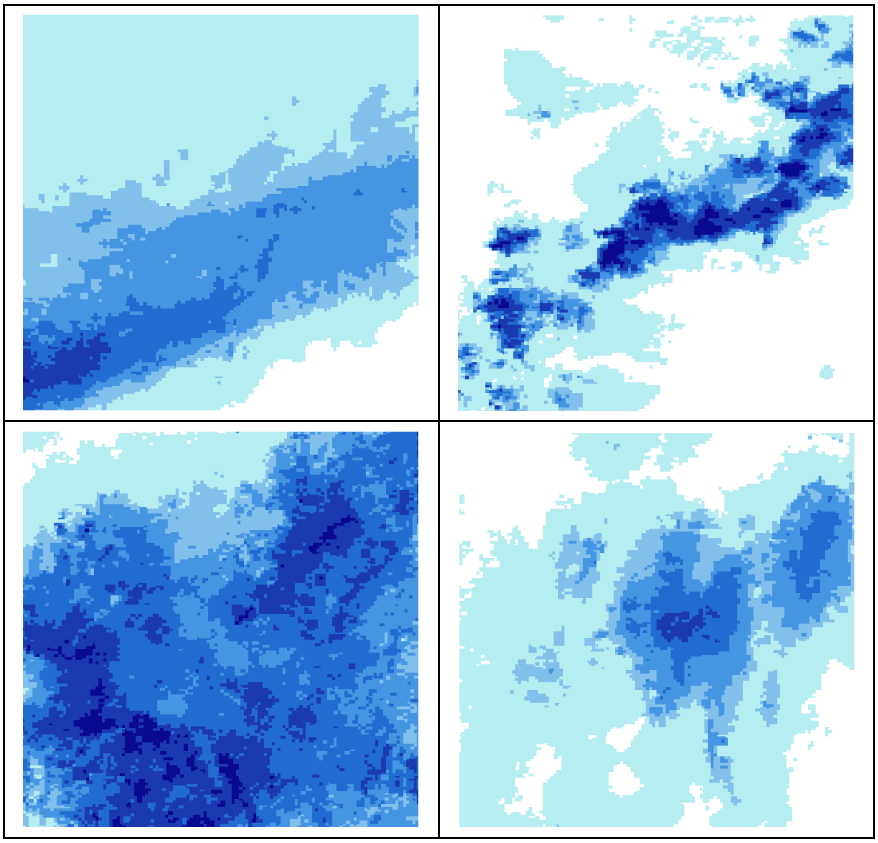


Figure 1.1 Spatial features/patterns of rainfall

Rainfall intensity extracted from the Radar-AMeDAS datasets over the Kanto region (See Section 3.4.1 for a complete description of the datasets).

Remotely-sensed estimations of rain-rate fields consistently portray distinctive spatial features that encompass intermittency, abrupt discontinuities, parallel bands and high-intensity clusters surrounded by consecutive larger areas of each time lower-intensity (Figure 1.1). The term intermittency refers to the probability of no-rain areas existing inside rain areas and becoming observable as the scale of representation becomes smaller.

Simple scaling or self-similarity is established if a single exponent can characterize the inter-scale relationship of the statistical properties of a given process. Models based on simple scaling, which simulate rainfall intensity as scaling sums of random increments (linear/additive structure), could not entirely reproduce the characteristic spatial features and extreme variability. Instead, it was empirically recognized that the statistical moments scale differently (i.e., there cannot be only one exponent). Another argument for ruling out trivial, self-similar processes is the fact that it seems incongruous to use a linear model to simulate a natural formation produced by non-linear physical processes (e.g., advection). In the light of these considerations, researchers discarded self-similar models and developed a few stochastic models exploring the properties of multifractals and random multiplicative cascades, which generalize the single scaling exponent to a continuous spectrum of

exponents. A more detailed description of the basic structure of multifractal models can be found in the works by Schertzer and Lovejoy (1987), Lovejoy and Schertzer (1990), and Gupta and Waymire (1990, 1993).

Alternatively, rather than focusing on the actual rain-rate, the characterization of the extreme variability of rainfall was accomplished by studying the scale-to-scale dependence of the probability distributions of local fluctuations. For this purpose, orthogonal wavelet decomposition has been utilized to capture the sudden directional fluctuations of rain rate fields (Kumar and Foufoula-Georgiou, 1993a, 1993b; Perica and Foufoula-Georgiou 1996a; Ebtehaj and Foufoula-Georgiou, 2011). The multiscale application of this decomposition, successfully revealed the presence of simple scaling in the marginal distributions of wavelet coefficients. The advantage of using this kind of filtering scheme is the possibility of inverting the decomposition with a relative computational simplicity in order to recover the original field.

Subsequent studies have put emphasis on relating the aforementioned stochastic models to atmospheric processes involved in the formation and development of rainfall (Over and Gupta, 1994; Perica and Foufoula-Georgiou, 1996b; Harris and Foufoula-Georgiou, 2001; Nykanen and Harris, 2003; Nykanen, 2008; Parodi et al., 2011). One hypothesis is that the frequency of extreme values generated by the models should be linked to the degree of variability of the different types of precipitation systems. For example, convective precipitation with rapidly changing intensity noticeably may exhibit more extreme fluctuations than stratiform precipitation, which usually portrays extensive rainfall with steady intensity.

1.3 PROBLEM STATEMENT

Radar-based estimations of rainfall intensity are often given as a two-dimensional arrange of cells. The probability distributions of fluctuations regularly depict a large amount of almost-zero values around the center (as a result of adjacent cells with similar intensity) and positive and negative outliers representing the abrupt jumps of intensity. Commonly-used symmetric, finite-moment distributions failed to capture these particular characteristics, in part, because the frequency of the outliers in the probability distribution of fluctuations makes the tails appear thicker than the tails of Gaussian distributions.

As it was mentioned previously, one of the objectives of this kind of stochastic analysis is to find scale-to-scale dependence of certain aspects of the rainfall field (self-similarity). In this case, it is necessary to show self-similarity in the distribution of fluctuations. Perica and Foufoula-Georgiou (1996a) identified that α-stable distributions (also known as Lévy distributions) can be adequately adjusted to rainfall fluctuations extracted from different types of rainfall. However, the adjusted distributions did not convincingly reveal self-similarity unless the wavelet coefficients were standardized. In that study, "standardized rainfall fluctuations" were defined as the wavelet coefficients (obtained with the oriented

band-pass filters) divided by their corresponding low-pass coefficients (local means). The standardization of wavelets flattened the tails making the distributions adjustable to Gaussian distributions regardless of the type of precipitation (degree of variability).

The characteristic exponent α of α-stable distributions determines the frequency of extreme values, and hence, it could be an optimal indicator of the degree of variability. Despite the fact that the findings in the study of Perica and Foufoula-Georgiou (1996a) are remarkable by showing self-similarity in the Gaussian distributions of the standardized fluctuations of rainfall, it is difficult not to contemplate the possibility of characterizing the degree of variability with α-stable distributions. If it were possible to demonstrate the presence of self-similarity in the α-stable distributions of rainfall fluctuations, it would open the possibility of relating the characteristic exponent α to the underlying physics that induce the formation and variability of rainfall. Moreover, accomplishing this challenge would allow to build a stochastic model that utilizes the characteristic exponent α for appropriately representing small-scale extreme variability.

1.4 OBJECTIVES

The overall objective of this research is to develop a stochastic methodology that allows to reproduce small-scale extreme variability of rainfall by combining the kind of rainfall data that is worldwide accessible. Linking in-situ measurements and satellite-based products to a multi-scale analysis of rainfall patterns (degree of variability) implies that the same framework might be fit for universal application if certain validations are undertaken beforehand. Therefore, this research is structured so as to fulfill the following specific objectives:

- Conduct a theoretical study of α-stable distributions, and parameter-estimation methods of symmetric α-stable distributions.
- Identify the method that is better suited for estimating the stable parameters given the particular difficulties of the empirical distributions of rainfall fluctuations.
- Apply the selected parameter-estimation method to samples extracted from rainfall data of ground-based weather radars over the Japanese Islands trying to contemplate different types of precipitation systems.
- Document and analyze the range of the stable parameters and evaluate the possibility of linking them to indicators of atmospheric instability, cloud formation and rainfall types
- Build and algorithm that combines a minimum amount of rain gauge stations and wide-coverage satellite products in order to reproduce accurate rainfall fields at high resolutions with the capacity of controlling extreme variability using the stable parameters of the distribution of fluctuations.

- Validate the performance of the algorithm by evaluating its applicability to different types of rainfall, its capacity for reproducing extremes using common statistical metrics.

2. DECOMPOSITION OF RAINFALL FIELDS

2.1 BACKGROUND

In the fields of computer vision and image processing, filtering is defined as the convolution of a mask with an image. Generally, the masks are either smoothing filters or frequency domain filters. Smoothing filters, also known as low-pass filters, are intended to reduce noise or remove the sudden changes of intensity producing smoother transitions. Frequency domain filters, also known as high-pass filters or derivative filters, are often used for sharpening images or for edge detection. Edge detection filters/operators are mostly classified in two categories, gradient and Laplacian. Figure 2.1 shows the effects of the convolution of an image with a smoothing mask (low-pass filter) and a sharpening mask.

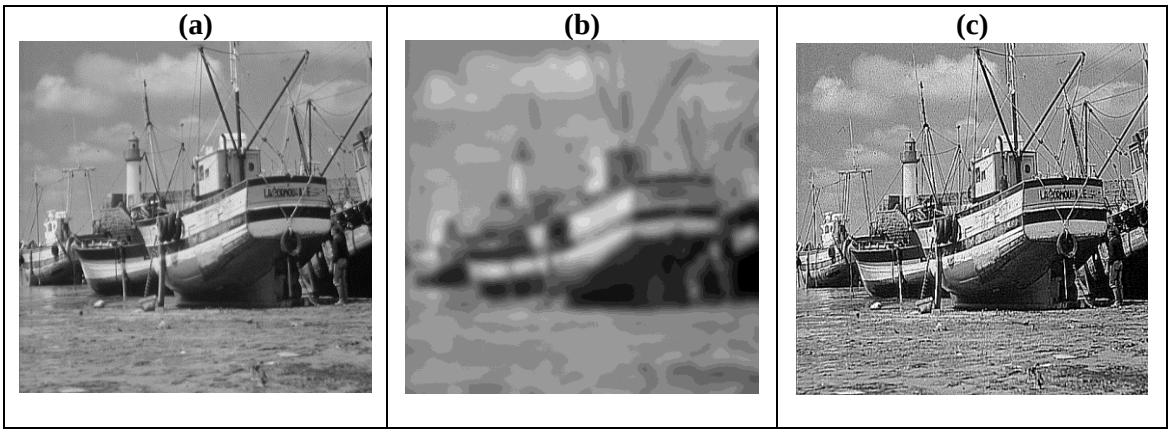


Figure 2.1 Effects on images of the convolution with masks

(a) This image is entitled "boat" and is part of the open source image library of the University of Southern California Signal and Image Processing Institute (USC-SIPI). (b) Convolution of the original image with a blurring mask (low-pass filter). (c) Convolution of the original image with a sharpening mask.

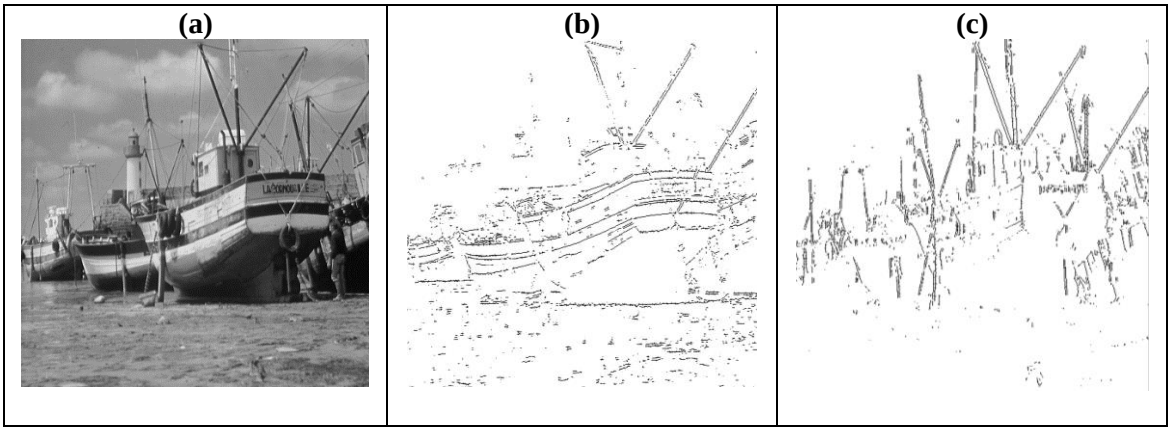


Figure 2.2 Edge detection with the Sobel operator.

(a) Original image. (b) Convolution of the original image with the horizontal Sobel operator. (c) Convolution of the original image with a vertical Sobel operator.

In this study, we are particularly interested in edge detection operators which are a collection of masks aimed to allocate the discontinuities and sharp changes of image intensity. Gradient operators look for extreme values in the first-order derivative space (e.g., Roberts, Prewitt, and Sobel operators). Laplacian operators look for zero-crossings in the second-order derivative space (e.g., Marrs-Hildreth operator). The Prewitt and Sobel operators are very similar 3×3 kernels that can be used to identify horizontal and vertical edges separately just by rotating the kernel, as shown in Figure 2.2.

A third way to detect edges in images is using two-dimensional discrete wavelets, which have been widely used in image processing for developing statistical prior models and image compression algorithms. The main advantages of this kind of filter are its multiscale applicability and the relative simplicity with which it can be inverted in order to get the original image.

2.2 THE ORTHONORMAL WAVELET TRANSFORM

In the field of signal processing Fourier transforms and wavelet transforms have been designed with the purpose of measuring the time-frequency variations of spectral components. By extending the concept of wavelet transforms into two-dimensions, they have been utilized to allocate discontinuities and study the statistical properties of the spatial patterns of digital images.

Since discretized representations of the continuous rainfall fields are similar to digital images, Kumar and Foufoula-Georgiou (1993a, 1993b) proposed to decompose rainfall intensity fields into large- and small-scale components utilizing wavelet transforms. The large-scale components hold information related to the mean behavior of the process while the small-scale components, named rainfall "local fluctuations" by Perica and Foufoula-Georgiou (1996a), represent the abrupt changes of intensity (sharp edges). The multiscale decomposition employed an orthonormal wavelet transform. The following is an explanation of the wavelet transform in a two-dimensional multiscale framework. For basic details about wavelets the reader is referred to the works of Mallat (1989, 2009), Daubechies (1992), and Kumar and Foufoula-Georgiou (1993a).

2.2.1 Multiscale approximation and wavelet transforms

Let $f(x) \in \mathrm{L}^2(\mathbb{R}^2)$ be a one-dimensional function, m the scale index, and $\mathrm{L}^2(\mathbb{R}^2)$ a vector space of measurable, square-integrable one-dimensional functions. The multiscale approximation of f at any scale m can be accomplished using an orthogonal projection onto a sequence of subspaces $\{V_m^1\}_{m\in\mathbb{Z}}$ of $\mathrm{L}^2(\mathbb{R}^2)$. The projections are undertaken using a scaling function ϕ. Denoting $\phi_{m,n}(x) = a_0^{m/2}\phi(a_0^m x - nb_0)$ as the integer translates of ϕ at scale m, the family $\{\phi_{m,n}\}_{n\in\mathbb{Z}}$ is an orthonormal basis of V_m^1 if $a_0 = 2$ and $b_0 = 1$, where n are all the possible integer translates that can be "covered" at scale m (Daubechies, 1992). Therefore, an orthonormal basis of V_m^1 is

$$\{\phi_{m,n}(x) = 2^{m/2}\phi(2^m x - n)\}_{n\in\mathbb{Z}}. \quad (2.1)$$

The inner products $a_m(n) = \{\langle f(x), \phi_{mn}\rangle\}_{n\in\mathbb{Z}}$ are the approximations (low-pass filtering) of $f(x)$ at intervals 2^m.

The approximation of $f(x)$ in V_{m-1}^1 has all the required information to obtain the approximation of $f(x)$ in the subspace V_m^1, which approximates $f(x)$ at a lower resolution. From this assertion, $V_m^1 \subset V_{m-1}^1$. In transforming into a lower resolution, the detail or difference of information between the approximations of $f(x)$ in the subspaces V_m^1 and V_{m-1}^1 is lost. However, this detail-information can be recovered utilizing wavelet transforms. Let W_m^1 be the detail subspace which is the orthogonal complement of V_m^1 in V_{m-1}^1:

$$V_{m-1}^1 = V_m^1 \oplus W_m^1. \quad (2.2)$$

Based on $\phi(x)$, it is possible to construct a wavelet transform, $\psi(x)$, that supports integer translates and dyadic dilates:

$$\psi_{m,n}(x) = 2^{m/2}\psi(2^m x - n). \quad (2.3)$$

Then, the family of integer translates, $\{\psi_{m,n}\}_{n\in\mathbb{Z}}$, form an orthonormal basis of W_m^1, and the family of integer translates and dyadic dilations, $\{\psi_{m,n}\}_{m,n\in\mathbb{Z}}$, form an orthonormal basis of $L^2(\mathbb{R}^2)$. Dilating or contracting the wavelet transform allows to identify the behavior of the function f in different frequency bands, making it a band-pass filter.

The orthogonal projection of the function $f(x)$ onto W_m^1 is obtained with a partial expansion in its wavelet basis

$$P_{W_m^1} f = \sum_{n=-\infty}^{+\infty} d_m(n)\, \psi_{m,n}, \quad (2.4)$$

where the wavelet coefficients $d_m(n)$ are the inner products $d_m(n) = \{\langle f(x), \psi_{mn}\rangle\}_{n\in\mathbb{Z}}$. The wavelet coefficients $d_m(n)$ contain the necessary details to reconstruct the approximation $a_{m-1}(n)$ from the approximation of $f(x)$ in V_m^1, i.e., $a_m(n)$.

The expansion of the function f in a wavelet orthogonal basis is thus an aggregation of the details at all scales

$$f = \sum_{m=-\infty}^{+\infty} P_{W_m^1} f = \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} d_m(n)\psi_{m,n}. \quad (2.5)$$

2.2.2 Two-dimensional multiscale approximation and two-dimensional wavelet transform

Let $f(x_1, x_2) \in L^2(\mathbb{R}^2)$ be a two-dimensional field, and $L^2(\mathbb{R}^2)$ a vector space of measurable, square-integrable two-dimensional fields. The multiscale approximation of f at

any scale m can be accomplished using an orthogonal projection onto a sequence of subspaces $\{V_m^2\}_{m\in\mathbb{Z}}$ of $L^2(\mathbb{R}^2)$.

A separable two-dimensional subspace is composed of the tensor product

$$V_m^2 = V_m^1 \otimes V_m^1. \tag{2.6}$$

Then, an orthonormal basis in V_m^2 can be constructed by

$$\left\{\Phi_{m,n_1,n_2} = \phi_{m,n_1}(x_1)\phi_{m,n_2}(x_2) = 2^m\phi(2^m x_1 - n_1)\phi(2^m x_2 - n_2)\right\}_{n_1,n_2\in\mathbb{Z}}, \tag{2.7}$$

where Φ_{m,n_1,n_2} is a two dimensional low-pass filter, and n_1, n_2 are the coordinates of a two-dimensional square grid that translates covering the field f. The inner products $A_m(n_1,n_2) = \left\{\langle f(x_1,x_2), \Phi_{m,n_1,n_2}\rangle\right\}_{n_1,n_2\in\mathbb{Z}}$ are the approximations (low-pass filtering) of $f(x,x_2)$ at scale m.

In order to construct a separable wavelet orthonormal basis, let W_m^2 be the detail subspace equal to the orthogonal complement of V_m^2 in V_{m-1}^2:

$$V_{m-1}^2 = V_m^2 \oplus W_m^2. \tag{2.8}$$

Expanding Eq. 2.7 using the equality in Eq. 2.6 yields $V_{m-1}^1 \otimes V_{m-1}^1 = (V_m^1 \otimes V_m^1) \oplus W_m^2$. Considering the decomposition of V_{m-1}^1 as in Eq. 2.2 and the distributivity of the tensor product over direct sums, it can be proven that

$$W_m^2 = (V_m^1 \otimes W_m^1) \oplus (W_m^1 \otimes V_m^1) \oplus (W_m^1 \otimes W_m^1). \tag{2.9}$$

Because $\left\{\phi_{m,n}\right\}_{n\in\mathbb{Z}}$ and $\left\{\psi_{m,n}\right\}_{n\in\mathbb{Z}}$ are respectively orthonormal bases of V_m^1 and W_m^1, an orthonormal basis of W_m^2 is given by the wavelet family

$$\left\{\Psi_{m,n_1,n_2}^1, \Psi_{m,n_1,n_2}^2, \Psi_{m,n_1,n_2}^3\right\}_{n_1,n_2\in\mathbb{Z}}, \tag{2.10}$$

where

$$\Psi_{m,n_1,n_2}^i = 2^m\Psi^i(2^m x_1 - n_1, 2^m x_2 - n_2),$$

$$\Psi^1(x_1,x_2) = \phi(x_1)\psi(x_2),$$

$$\Psi^2(x_1,x_2) = \psi(x_1)\phi(x_2),$$

$$\Psi^3(x_1,x_2) = \psi(x_1)\psi(x_2),$$

and $1 \le i \le 3$.

Similarly to the one-dimensional case, $L^2(\mathbb{R}^2)$ can be decomposed as the orthogonal sum of all the detail spaces $\{W_m^2\}_{m\in\mathbb{Z}}$, therefore, an orthonormal basis of $L^2(\mathbb{R}^2)$ is given by

$$\left\{\Psi^1_{m,n_1,n_2}, \Psi^2_{m,n_1,n_2}, \Psi^3_{m,n_1,n_2}\right\}_{m,n_1,n_2\in\mathbb{Z}}. \tag{2.11}$$

Consequently, and considering a two-dimensional field $f(x_1,x_2)$ with $1 \le x_1 \le L_1$ and $1 \le x_2 \le L_2$, the two dimensional field f can be expanded in an orthogonal wavelet basis as an aggregation of all the two-dimensional details at all scales:

$$f(x_1,x_2) = \frac{1}{\sqrt{L_1 L_2}} \sum_{i=1,2,3} \sum_{m=-\infty}^{\infty} \sum_{n_1=0}^{L_1-1} \sum_{n_2=0}^{L_1-1} D_m^i(n_1,n_2)\, \Psi^i_{m,n_1,n_2}(x_1,x_2). \tag{2.12}$$

The wavelet coefficients in Eq. 2.12 are the inner products

$$D_m^i(n_1,n_2) = \left\{\langle f(x_1,x_2), \Psi^i_{m,n_1,n_2}\rangle\right\}_{n_1,n_2\in\mathbb{Z}}. \tag{2.13}$$

Since the scaling function $\Phi(x_1,x_2)$ and the three wavelets $\Psi^1(x_1,x_2)$, $\Psi^2(x_1,x_2)$, and $\Psi^3(x_1,x_2)$ are defined as separable products of the functions ϕ and ψ, the orthonormal wavelet decomposition can be understood as a decomposition of the field $f(x_1,x_2)$ in a set of independent spatially oriented sub-bands.

By using the two-dimensional wavelet transform, the detail information between the approximation A_{m-1} and the discrete approximation A_m is divided into three sets of wavelet coefficients: D_m^1 (high horizontal correlation), D_m^2 (high vertical correlation), and D_m^3 (high horizontal and vertical correlation).

2.2.3 Haar wavelet transform

The multiscale decomposition framework and the pair of functions ϕ and ψ are designed so as to allow the discrete approximations to be non-redundant at all scales due to the linear independence of the translates, and to span completely the space in which the field is defined.

Given a scaling function (Figure 2.1a) of the form

$$\begin{aligned} &\phi(x) = 1 \quad 0 \le x < 1 \\ &\phi(x) = 0 \quad \text{otherwise,} \end{aligned} \tag{2.14}$$

the corresponding wavelet transform, known as the Haar wavelet (Figure 2.1b), has the form

$$\begin{aligned} &\psi(x) = 1 \quad\quad 0 \le x < 1/2 \\ &\psi(x) = -1 \quad 1/2 \le x < 1 \\ &\psi(x) = 0 \quad\quad \text{otherwise.} \end{aligned} \tag{2.15}$$

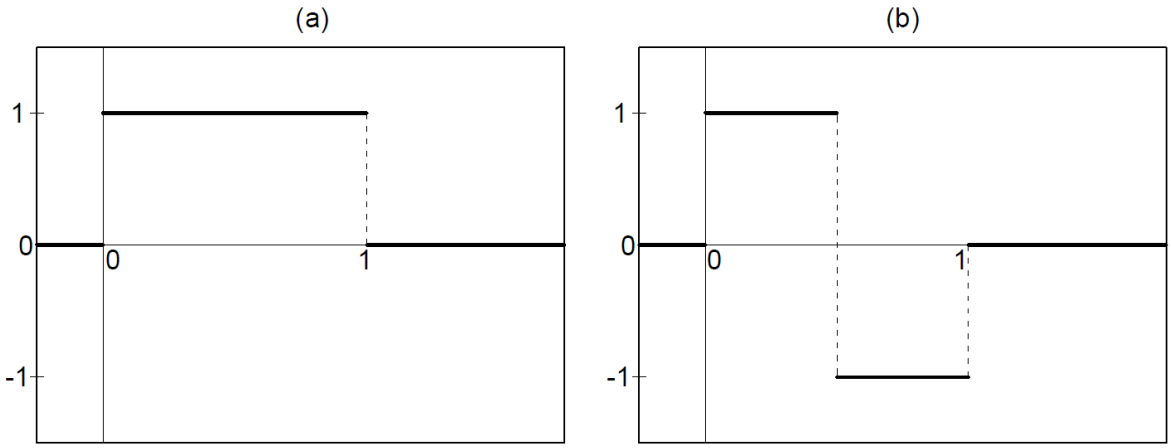


Figure 2.1. Scaling function and the corresponding Haar wavelet.

Note that the integration of ϕ is equal to 1 and the integration of ψ is equal to zero, while the norm of both functions in $L^2(\mathbb{R}^2)$ is equal to 1. The simplicity of this pair of functions and the interpretation that can be given to the sets of wavelet coefficients in a two-dimensional framework has made them useful in several image-processing and image-compression applications.

2.3 DISCRETE WAVELET TRANSFORM AND ITS APPLICATION TO RAINFALL FIELDS

Even though rainfall is a process constantly changing in time and continuous in the three spatial dimensions, in this study we analyze rainfall intensity as an idealized two-dimensional field sampled at surface-level and aggregated at fixed time steps. The representation of this two-dimensional field is done by using a two-dimensional vector (grid) comprising values of mean intensity.

Let X be the two-dimensional rainfall field containing a finite number of samples $L_1 \times L_2$. The highest resolution of representation, which corresponds to the scale of sampling, is denoted with the scale index $m = 0$. Traversing X through all the integer translates of the two-dimensional low-pass filter Φ_m (orthonormal basis of V_m^2), built with the scaling function in Eq. 2.14, yields the discrete approximation $\bar{X}_m$ of the rainfall field. The discrete approximation of X at scale $m = 0$ is $\bar{X}_0$, i.e., $\bar{X}_0 \equiv X$.

The discrete approximation is thus a decrease in resolution or a spatial averaging of the rainfall field, in which the set of values $\bar{X}_m$ can be considered "local means" of the rainfall field at the scale m. The detail information lost during the averaging can be recovered using the integer translates of the three oriented sub-band filters Ψ_m^1, Ψ_m^2, and Ψ_m^3 (orthonormal basis of W_m^2), yielding the three sets of wavelet coefficients $X'_{m,1}$, $X'_{m,2}$, and $X'_{m,3}$. Due to the definition of the wavelet transform in Eq. 2.15, the resulting wavelet coefficients can be interpreted as local variations of the rainfall field with respect to the local mean, hereafter denominated "local rainfall fluctuations" as in Kumar and Foufoula-Georgiou (1996a).

Large values of wavelet coefficients contained in $X'_{m,1}$ represent local fluctuations at scale m that are more sensitive in the east-west direction. Similarly, large values of wavelet coefficients contained in $X'_{m,2}$ represent local fluctuations at scale m that are more sensitive in the north-south direction, and large values of wavelet coefficients contained in $X'_{m,3}$ represent local fluctuations at scale m that are more sensitive in both directions (diagonal).

This discrete wavelet transform applied to rainfall-intensity representations is a multidimensional decomposition into "local means" and "local fluctuations" by using the dyadic dilates of the low-pass filter and the dyadic dilates of the three oriented sub-band filters. In the case of the Haar wavelet, the transformation into a lower resolution between any two adjacent scales $m-1$ and m is given by

$$\bar{X}_m(n'_1, n'_2) = 1/4\{\bar{X}_{m-1}(n_1, n_2) + \bar{X}_{m-1}(n_1+1, n_2) + \bar{X}_{m-1}(n_1, n_2+1) + \bar{X}_{m-1}(n_1+1, n_2+1)\}. \quad (2.16)$$

In Eq. 2.16, the discrete approximation $\bar{X}_{m-1}$ at scale $m-1$ with dimensions $L_1/2^{m-1} \times L_2/2^{m-1}$ is rescaled into the process $\bar{X}_m$ at scale m with dimensions $L_1/2^m \times L_2/2^m$. The location indexes n_1 and n_2 can take only odd integer values between 1 and $L_2/2^{m-1}-1$ at scale $m-1$. The location indexes n'_1 and n'_2 can take integer values between 1 and $L/2^m$ at scale m.

Similar expressions can be derived for computing the three sets of directional wavelet fluctuations at scale m as a function of the discrete approximation of the rainfall field at scale $m-1$:

$$X'_{m,1}(n'_1, n'_2) = 1/4\{[\bar{X}_{m-1}(n_1, n_2) + \bar{X}_{m-1}(n_1+1, n_2)] - [\bar{X}_{m-1}(n_1, n_2+1) + \bar{X}_{m-1}(n_1+1, n_2+1)]\},$$

$$X'_{m,2}(n'_1, n'_2) = 1/4\{[\bar{X}_{m-1}(n_1, n_2) + \bar{X}_{m-1}(n_1, n_2+1)] - [\bar{X}_{m-1}(n_1+1, n_2) + \bar{X}_{m-1}(n_1+1, n_2+1)]\},$$

$$X'_{m,3}(n'_1, n'_2) = 1/4\{[\bar{X}_{m-1}(n_1, n_2) + \bar{X}_{m-1}(n_1+1, n_2+1)] - [\bar{X}_{m-1}(n_1+1, n_2) + \bar{X}_{m-1}(n_1, n_2+1)]\}. \quad (2.17)$$

Note that the sets of directional wavelet fluctuations $X'_{m,i}$ represent the local variations of intensity at scale $m-1$. The combination of the expressions in Eq. 2.16 and Eq. 2.17 allows to construct the inverse discrete wavelet transform in order to obtain discrete approximations at higher resolutions.

A diagram of the iterative multiscale application of the discrete wavelet transform for computing local means and local fluctuations is shown in Figure 2.2.

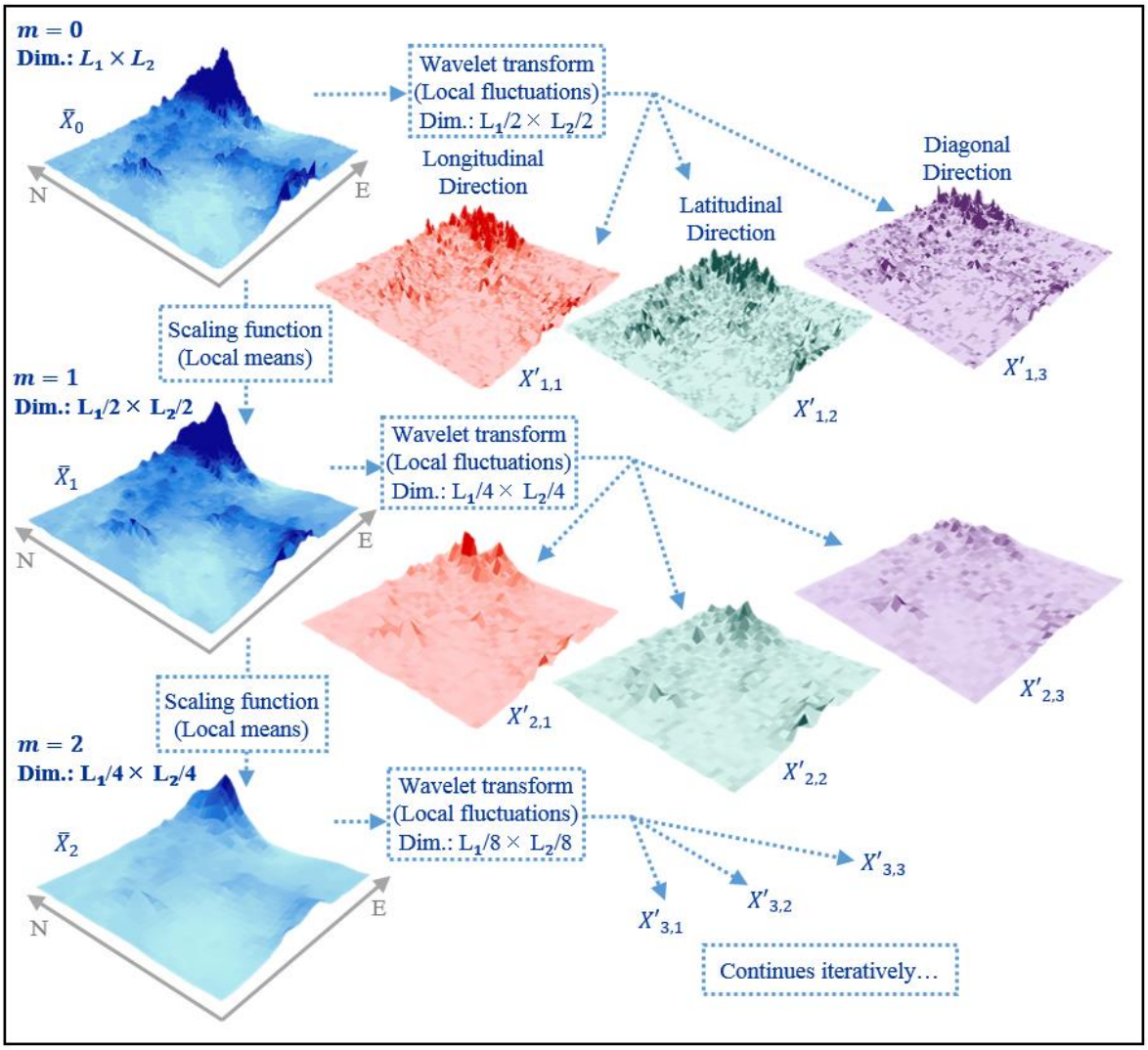


Figure 2.2 Two-dimensional discrete wavelet transform

2.4 UNDECIMATED DISCRETE WAVELET TRANSFORM

In Eq. 2.16 and Eq. 2.17, because the location indexes n_1 and n_2 can take only odd integer values for the sampling of wavelet coefficients, the discrete wavelet transform is space-variant. This decimated method of sampling implies that if the location indexes take even integer numbers, the statistical properties of the sample might be different, causing aliasing in each of the resulting sets.

The undecimated discrete wavelet transform, also called the stationary wavelet transform (Nason and Silvermann, 1995), is a redundant, over-complete representation of the original dataset. The projection of the original process onto the subspaces V_m^2 and W_m^2 in order to obtain the discrete approximation of the process and the directional sets of detail-information at scale m can be accomplished by using shifted versions of the bases presented in Section 2.2.

The undecimated discrete wavelet transform is a multiscale application of Eq. 2.18 and Eq. 2.19. Imposing periodic boundary conditions, the location indexes n_1 and n_2 can take odd and even integer values, and therefore, "fill the gaps" caused by the discrete wavelet transform. Consequently, the dimensions of the discrete approximations $\bar{X}_m$ and directional sets of wavelet coefficients at resolutions $X'_{m,i}$ at lower resolutions are exactly the same as the dimensions of $\bar{X}_m$, i.e., $L_1 \times L_2$. Having over-complete information can benefit the statistical parameterization of the directional sets, which is usually hindered by the limited information of the small samples generated with the discrete wavelet transform (Ebtehaj and Foufoula-Georgiou, 2011). A discussion about the use of decimated or undecimated samplings and the consequences of imposing periodic boundary conditions in the case of rainfall data is presented in the next Chapter.

$$\bar{X}_m(n'_1, n'_2) = 1/4\{\bar{X}_{m-1}(n_1, n_2) + \bar{X}_{m-1}(n_1 + 2^{m-1}, n_2) + \bar{X}_{m-1}(n_1, n_2 + 2^{m-1}) + \bar{X}_{m-1}(n_1 + 2^{m-1}, n_2 + 2^{m-1})\}. \quad (2.18)$$

$$X'_{m,1}(n'_1, n'_2) = 1/4\{[\bar{X}_{m-1}(n_1, n_2) + \bar{X}_{m-1}(n_1 + 2^{m-1}, n_2)] - [\bar{X}_{m-1}(n_1, n_2 + 2^{m-1}) + \bar{X}_{m-1}(n_1 + 2^{m-1}, n_2 + 2^{m-1})]\},$$

$$X'_{m,2}(n'_1, n'_2) = 1/4\{[\bar{X}_{m-1}(n_1, n_2) + \bar{X}_{m-1}(n_1, n_2 + 2^{m-1})] - [\bar{X}_{m-1}(n_1 + 2^{m-1}, n_2) + \bar{X}_{m-1}(n_1 + 2^{m-1}, n_2 + 2^{m-1})]\},$$

$$X'_{m,3}(n'_1, n'_2) = 1/4\{[\bar{X}_{m-1}(n_1, n_2) + \bar{X}_{m-1}(n_1 + 2^{m-1}, n_2 + 2^{m-1})] - [\bar{X}_{m-1}(n_1 + 2^{m-1}, n_2) + \bar{X}_{m-1}(n_1, n_2 + 2^{m-1})]\}. \quad (2.19)$$

3. SELF-SIMILARITY IN LOCAL FLUCTUATIONS OF RAINFALL

3.1 BACKGROUND

As it was introduced in Chapter 1, the stochastic analysis of the spatial structure of rainfall with the purpose of characterizing the characteristic irregular patterns has been undertaken for the past 30 years. Consequently the proposed stochastic models, aimed to reproduce the extreme spatial patterns of rainfall at high-resolutions, are mostly based on the scale invariance of the exponents/generators of some statistical aspect of the rainfall field (Lovejoy and Mandelbrot 1985; Lovejoy and Schertzer, 1986; Schertzer and Lovejoy, 1987; Gupta and Waymire, 1990; Lovejoy and Schertzer, 1990; Kumar and Foufoula-Georgiou, 1993a, 1993b; Tessier et al., 1993; Veneziano et al., 1996; Menabde et al., 1997; Deidda et al., 2006; Lovejoy and Schertzer, 2006; Lovejoy et al, 2008; Ebtejah and Foufoula-Georgiou, 2011; Lovejoy and Schertzer, 2013).

In this study, the scale-to-scale dependence of the probability distributions of rainfall local fluctuations is analyzed. Figure 3.1 shows an example of the histogram of the values of wavelet coefficients that is generated from a rainfall field at a specific orientation at a specific scale m. The histogram shows a large mass of values around the center as a consequence of large areas with similar values of intensity, which naturally do not generate significant local fluctuations. It also shows a large number of positive and negative outliers produced by the sudden jumps of intensity that are characteristic of the clustered and band-type spatial structure of rainfall.

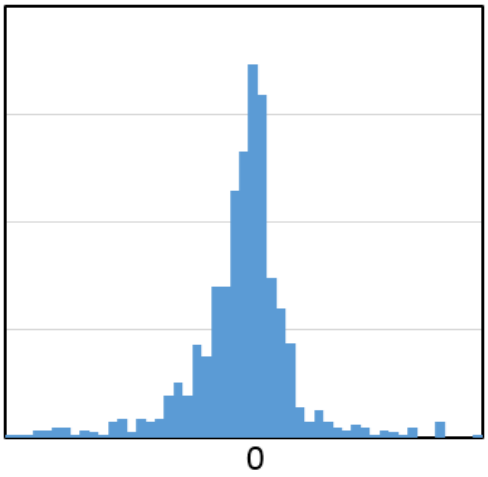


Figure 3.1 Example of the typical histogram of wavelet coefficients from rainfall fields

Kumar and Foufoula-Georgiou (1993b) proposed to adjust the empirical distributions to α-stable distributions. The characteristics of the probability density function of α-stable distributions agrees with the aforementioned features of the empirical distribution. Particularly, the tail of α-stable distributions decays at a lower rate than the tails of Gaussian

distributions, which is a condition that produces heavier tails and consequently a higher frequency of extreme values. Perica and Foufoula-Georgiou (1996a) were effectively able to fit the empirical distributions of wavelet coefficients to α-stable distributions. However, the scale-to-scale dependence of the stable parameters did not show a clear scale-invariance behavior. In that same study, "standardized rainfall fluctuations" were defined as the wavelet coefficients divided by their corresponding local means. The empirical distribution of the standardized rainfall fluctuations proved to be Gaussian and a power-law behavior of the standard deviations across scales was revealed. Moreover, the exponent characterizing the power-law behavior was found to be somewhat dependent on the convective instability of the pre-storm environment measured by the convective available potential energy (CAPE). Several subsequent studies were based on this remarkable finding about the gaussianity and scale-invariance of the distributions of standardized rainfall fluctuations. Examples include a rainfall disaggregation model (Perica and Foufoula-Georgiou, 1996b), the coupling of a mesoscale dynamical model with the previously-mentioned statistical disaggregation model (Zhang and Foufoula-Georgiou, 1997), stochastic assessment of the space-time dependence of rainfall processes (Venugopal et al., 1999a; Joseph et al., 2000), disaggregation of a numerical weather prediction model using a stochastic space-time disaggregation model (Venugopal et al., 1999b), analysis of the multiscale variability of observed microwave radiances and outputs of a cloud model (Harris and Foufoula-Georgiou, 2001), and the impact of the small-scale variability of rainfall on the atmospheric variables of a land-atmospheric scheme coupled with a mesoscale model (Nykanen and Foufoula-Georgiou, 2001).

Another interesting multiscale stochastic analysis of rainfall fluctuations was presented by Ebtehaj and Foufoula-Georgiou (2011). In that study, Gaussian Scale Mixtures are introduced as a model that can simultaneously, but separately, control the scale-to scale dependence of the probability distributions and the heavy-tail features.

In this study, we initially revise the adjustment of the distributions of wavelet coefficients to α-stable distributions. Because the characteristic exponent α defines the thickness of the tails (frequency of extreme values), the hypothesis is that this parameter can be an indicator of the degree of spatial variability of rainfall fields. For this purpose, the performance of a few parameter-estimation methods are evaluated with samples of wavelet coefficients. It was found that (i) at low-resolutions the size of the samples is small, and (ii) at high-resolutions an unusual large number of values equal to zero are comprised in the samples of wavelet coefficients since the source datasets usually store equal values of intensity in large areas that should at least have small fluctuations. These two conditions impact the estimation of stable parameters hampering the multiscale analysis of scale-to-scale dependence of the distributions.

Based on a theoretical review of α-stable distributions, this study proposes a new parameter-estimation method that diminishes the effect of the natural adverse conditions of the samples of wavelet coefficients, improving the multiscale analysis of scale-to-scale dependence of

the distributions. This chapter also shows the results of applying the proposed method to samples extracted from rainfall intensity datasets available in Japan, and how it was revealed the presence of self-similarity on the distributions of rainfall local fluctuations.

3.2 α-STABLE DISTRIBUTIONS AND ESTIMATION OF PARAMETERS

The generalized central limit theorem states that stable laws define the family of distributions that are the only possible limit of a normalized sum of independent, identically distributed random variables (rvs). Stability under addition and other mathematical properties allow stable distributions to model empirical data displaying skewness and heavy tails, which noticeably cannot be described by traditional Gaussian distributions. Stable distributions have been shown to demonstrate usefulness in many applied areas, such as physics, astronomy, telecommunications, acoustics, remote sensing, computer science, finance, and biology (Zolotarev, 1986; Uchaikin and Zolotarev, 1999).

Conventional methods used to adjust data to probability distributions (e.g., maximum likelihood, least squares, and sample moments) cannot be applied to stable distributions because the probability density function (pdf) and the cumulative distribution function (cdf) lack closed-form expressions. The estimation of parameters using quantile-based relative measures (McCulloch, 1986) and the regression of the empirical characteristic function (Koutrouvelis, 1980) are the two estimation methods regarded as most efficient in terms of accuracy and time consumption. Maximum likelihood estimation using non-closed-form expressions of the pdf is also considered reasonably accurate, although its computational effort was found to be unjustified. These different techniques tend to estimate different parameters when dealing with small samples that naturally fail to provide detailed information about the empirical distribution. Because parameter-estimation techniques aim to match different characteristics of the sample, assessing goodness of fit has to be performed using alternative procedures based on qualitative observations rather than quantitative evaluations (Nolan, 1999). Borak et al. (2011) sustained that there are no standard tests for assessing goodness of fit of stable distributions despite the development of new procedures, which are often conditioned to the availability of large samples (e.g., the stability test by Brcich et al., 2005; the minimization of the difference between the empirical and the estimated characteristic function by Matsui and Takemura, 2008). In this study, the estimation methods are firstly evaluated with Monte Carlo simulations, and then, by comparing how much the multiscale analysis of the scale-to-scale dependence of the distributions is improved.

Because the pdf of wavelet coefficients is bound to be symmetrical this analysis covered parameter-estimation methods restricted to symmetric stable distributions, which are defined by two parameters only: one that describes the heaviness of the tails and the other that is a parameter of scale. Aside from the McCulloch's (1986) quantile-type approximation and the regression-type method, four other estimation techniques were considered for evaluation. These additional techniques were postulated developing the concept and mathematical

properties of the absolute value of a rv following a stable distribution. Let X be a centered radially symmetric stable random vector, for which Nolan (2013) defined the amplitude as a univariate rv, R, that is equal to |X|. If X is a univariate rv X, the pdf of R is the result of adding the negative side of the pdf of X to its positive side. In the case of small samples, interpreting R as a distribution of absolute values happened to be beneficial to the estimation of parameters because insufficient data in one tail was complemented with data from the other. This realization improved the performance of simple techniques of estimation, i.e., those that rely on distribution quantiles. Additionally, it was found that the cdf evaluated at the fractional moments of R is a good estimator of the characteristic exponent of stable distributions.

3.2.1 Description of α-stable distributions

The Generalized Central Limit Theorem (Definition 1.1.4 in Samorodnitsky and Taqqu, 1994; Theorem 1 Section 2.5 in Nikias and Shao, 1995; Section 2.5 in Uchaikin and Zolotarev, 1999) states that a rv X is said to be stable if, for $N > 1$, there is a positive norming constant c_N and a real number d_N such that $X_1 + \cdots + X_N \stackrel{d}{=} c_N X + d_N$. In this expression, $X_1, \cdots, X_N$ are independent copies of X, the norming constant c_N is of the form $N^{1/\alpha}$, and $\stackrel{d}{=}$ stands for equality in the distribution (i.e., both sides follow the same probability law). Variable X is said to be stable in the strict sense if the previous expression holds for $d_N = 0$. Parameter $0 < \alpha \leq 2$, called the characteristic exponent of stable distributions, determines the rate of decrease of the tails. When $\alpha = 2$, the stable distribution becomes a Gaussian distribution. A rv that follows a stable distribution is usually expressed in terms of its characteristic function (cf)

$$\varphi(t) = \mathrm{E}[\exp(itx)] = \begin{cases} \exp\{-\gamma^{\alpha}|t|^{\alpha}[1 - i\beta\tan(\pi\alpha/2)\mathrm{sgn}(t)] + i\delta t\} & \alpha \neq 1 \\ \exp\{-\gamma|t|[1 + i\beta(2/\pi)\mathrm{sgn}(t)\ln|t|] + i\delta t\} & \alpha = 1, \end{cases} \quad (3.1)$$

where sgn (·) is the signum function, $-1 \leq \beta \leq 1$ is the skewness parameter, $\gamma > 0$ is the scale parameter, and $\delta \in \mathbb{R}$ is the location parameter (Definition 1.1.6 in Samorodnitsky and Taqqu, 1994). The parameterization of the cf of stable distributions might vary from one author to another. To simplify some numerical manipulations and, at the same time, to make the cf jointly continuous for all four parameters, Nolan (1999) proposed the parameterization

$$\varphi(t) = \mathrm{E}[\exp(itx)] = \begin{cases} \exp\{-\gamma^{\alpha}|t|^{\alpha}[1 + i\beta\tan(\pi\alpha/2)\mathrm{sgn}(t)((\gamma|t|)^{1-\alpha} - 1)] + i\delta_0 t\} & \alpha \neq 1 \\ \exp\{-\gamma|t|[1 + i\beta(2/\pi)\mathrm{sgn}(t)(\ln|t| + \ln\gamma)] + i\delta_0 t\} & \alpha = 1. \end{cases} \quad (3.2)$$

In both parameterizations, α, β, and γ have the same meaning, whereas the location parameter is related by $\delta = \delta 0 - \gamma\,\beta\tan(\pi\alpha/2)$ for $\alpha \neq 1$ and by $\delta = \delta_0 - \gamma\beta(2/\pi)\ln(\gamma)$ for $\alpha = 1$. Parameterization as in Eq. 3.2 allows the continuity of α and β in all of their parameter ranges, giving them clearer significance in characterizing the heaviness of tails and skewness, respectively. Henceforth, unless otherwise noted, the rv denoted by X ~ S(α, β, γ , δ) has parameterization as in Eq. 3.1.

With the exception of the Gaussian $\sim N(\mu,\sigma)$ or $\sim S(2,0,2^{-1/2}\sigma,\mu)$, the Cauchy$\sim S(1,0,\gamma,\delta)$, and the Lévy $\sim S(1/2,1,\gamma,\delta)$ distributions, closed-form expressions for stable pdfs and cdfs do not exist. However, the pdf $f_X(x)$ can be expressed as the Fourier transform of its cf:

$$f_X(x) = \mathcal{F}(\varphi)(x) = \int_{-\infty}^{\infty} \varphi(t)\exp(-2\pi i x t)\,dt,$$

where symbol $\mathcal{F}$ denotes the Fourier transform. The pdf and the distribution function are considered continuous in the real line even though they lack closed-form expressions. Conversely, because f_X meets suitable conditions, employing the Fourier inversion theorem the cf can be obtained as the inverse Fourier transform of f_X, i.e., $\varphi(t) = \mathcal{F}^{-1}(f_X)(t)$.

Nolan (1997) proposed several numerical algorithms based on direct integration that compute the pdf and the cdf of stable distributions. These algorithms cover the entire range of the parameters and give accurate results except for very extreme values ($x \to \infty$) and particular boundary cases of the parameters. Matsui and Takemura (2006) reviewed the numerical difficulties in extreme cases and presented, as a solution, alternative expressions for symmetric distributions based on asymptotic expansions. Fast Fourier transforms (FFT) can also be used to compute the pdf with accuracy that depends on the sample size (Mittnik et al., 1999a; Menn and Rachev, 2006). For large samples, FFT methods can be computationally more efficient than direct integration; however, the accuracy is practically the same.

One important characteristic of stable distributions is the inverse-power decay of the tails that makes them thicker than the tails of Gaussian distributions. A consequence of this condition is the nonexistence of variance. In fact, the p th-order fractional absolute moments $\mathrm{E}|X|^p$ are infinite for all $p \geq \alpha$, and $\mathrm{E}|X|^p < \infty$ for $0 \leq p < \alpha$ (Property 1.2.16 in Samorodnitsky and Taqqu, 1994; Theorem 2 Section 2.5 Nikias and Shao, 1995). Only for the Gaussian distribution is $\mathrm{E}|X|^p < \infty$ for all $p \geq 0$.

The pdf of stable distributions is unimodal, and the location of the mode is a function of the skewness parameter β (Fofack and Nolan, 1999). For a fixed value of scale γ and characteristic exponent $\alpha \geq 1$, the mode is equal to the location parameter δ when $\beta = 0$, smaller than δ when $\beta > 0$, and greater than δ when $\beta < 0$. The opposite occurs for $\alpha < 1$; i.e., the mode is shifted toward the positive side when $\beta > 0$, and vice versa. Another particularity of distributions with the same characteristic exponent $\alpha < 1$ is that they are stochastically ordered in the interval $-1 \leq \beta \leq 1$, which is not the case for $\alpha \geq 1$ (Property 1.2.14 of Samorodnitsky and Taqqu, 1994). The skewness parameter β also affects the asymptotic behavior of the tails.

For a stable rv $X\sim S(\alpha,\beta,\gamma,\delta)$, the standardized form $Y = (X-\delta)/\gamma$ has a distribution of the form $Y\sim S(\alpha,\beta,1,0)$.

3.2.2 Symmetric α-stable distributions

The particular case of $\beta = 0$ results in symmetric stable distributions, often abbreviated as $S\alpha S$. Gaussian distributions are similar to $S\alpha S$ distributions; i.e., they both show a bell-shaped smooth pdf and symmetry with respect to the location parameter δ. However, there are three principal differences: (i) $S\alpha S$ distributions are more leptokurtic, showing higher density for values around δ. (ii) The pdf of Gaussian distributions is higher than that of $S\alpha S$ distributions in an intermediate range, as a consequence of the slenderness around δ. (iii) The rate of decay of $S\alpha S$ distributions (inverse-power) makes the tails heavier than those of Gaussian distributions whose decay is exponential. For $\alpha = \{2.0; 1.5; 1.0; 0.5\}$, Fig. 3.2 illustrates the pdf of standard $S\alpha S(\alpha, 0,1,0)$ distributions.

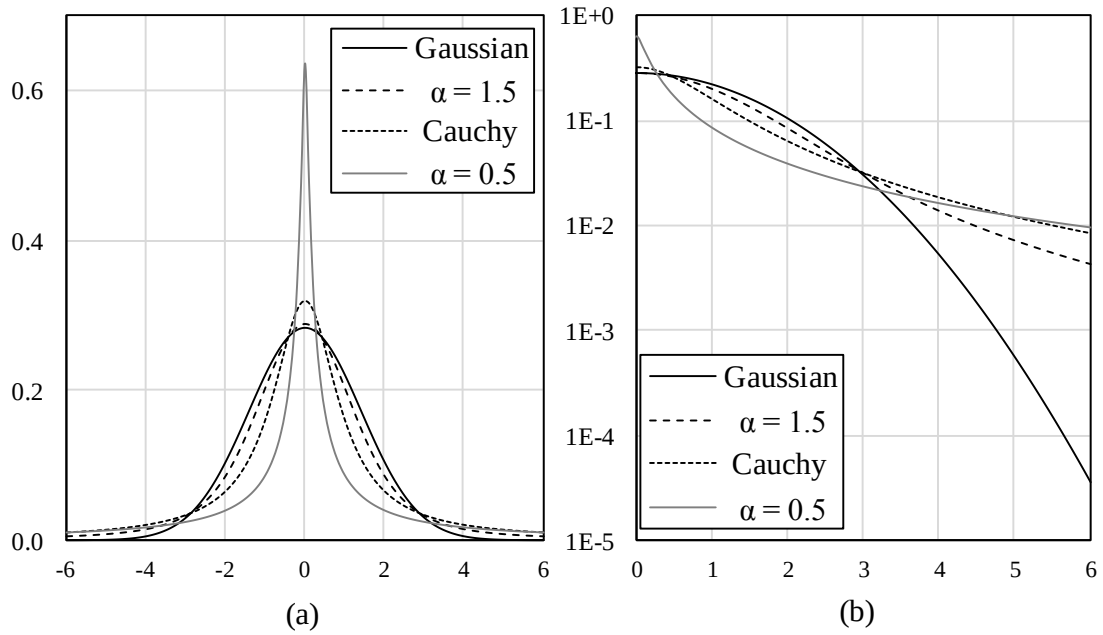


Figure 3.2 Pdfs for $S\alpha S$ distributions

The pdfs are computed with direct-integral algorithms (Nolan, 1997). Vertical axis corresponds to (a) f_X and (b) $\log(f_X)$ for clearer appreciation of the tails.

If location parameter δ (equivalent to the mode in $S\alpha S$ distributions) is equal to zero, the cf of the rv $X \sim S\alpha S(\alpha, 0, \gamma, 0)$ becomes

$$\varphi(t) = \exp(-\gamma^\alpha |t|^\alpha). \tag{3.3}$$

The pth-order fractional absolute moments for $X \sim S(\alpha, 0, \gamma, 0)$ are given by

$$\mathrm{E}|X|^p = C_{p,\alpha}\gamma^p, \tag{3.4}$$

where

$$C_{p,\alpha} = \frac{2^{p+1}\Gamma[(p+1)/2]\Gamma(-p/\alpha)}{\alpha\sqrt{\pi}\Gamma(-p/2)},$$

$\Gamma(\cdot)$ is the gamma function and $0 < p < \alpha$. Eq. 3.4 has been demonstrated in different ways (Nikias and Shao, 1995), although it can be simplified from the general expression for $-1 \leq \beta \leq 1$ (Property 1.2.17 in Samorodnitsky and Taqqu, 1994). Moreover, for $\alpha > 1$, $-1 \leq \beta \leq 1$, $\delta \in \mathbb{R}$, and $1 < p < \alpha$, expressions for δ-centered fractional absolute moments $\mathrm{E}|X - \delta|^p$ were derived using derivatives of cfs by Matsui and Pawlaz (2014). Even though those expressions are not in a closed form, they can be simplified to obtain Eq. 3.4 if $\alpha > 1$, $\beta = 0$, and $\delta = 0$.

Symmetric stable rvs are all considered conditionally Gaussian (Proposition 1.3.1 in Samorodnitsky and Taqqu, 1994). For $0 < \alpha < 2$, a normal rv $Z \sim N(0, \sqrt{2}\gamma)$, and a stable rv $B \sim S(\alpha/2, 1, (\cos\pi\alpha/4)^{2/\alpha}, 0)$, the rv $X \sim S\alpha S(\alpha, 0, \gamma, 0)$ with cf as in Eq. 3.2 can be expressed as

$$X \overset{d}{=} B^{1/2} Z \tag{3.5}$$

with B and Z independent. Note that B is totally skewed to the positive side (i.e., the pdf is equal to 0 for all $x \leq 0$) because $\alpha/2 < 1$and $\beta = 1$. If a standard normal rv $Z \sim N(0,1)$ is considered, the scale parameter of B must then be changed to $B \sim S(\alpha/2, 1, 2\gamma^2(\cos\pi\alpha/4)^{2/\alpha}, 0)$ such that the cf of X remains as in Eq. 3.2.

3.2.3 Amplitude of univariate α-stable distributions

Nolan (2013) defined the amplitude of a d-dimensional centered radially symmetric random vector $\mathbf{X}$ that follows a stable distribution as a univariate rv $R = |\mathbf{X}| = \sqrt{X_1^2 + \cdots + X_d^2}$. The same study proposed a few expressions for the pdf and the cdf of R when $d \geq 2$, deriving from the multivariate expression of Eq. 3.5. The inference for $d = 1$ is that univariate rv $X \sim S\alpha S(\alpha, 0, \gamma, 0)$ with pdf $f_X(x)$ has an amplitude R such that the pdf $f_R(x) = f_X(x) + f_X(-x)$ for $x \geq 0$, and $f_R(x) = 0$ for $x < 0$. If $\delta = 0$, $f_X(x) = f_X(-x)$. Consequently, $f_R(x)$ is two times $f_X(x)$ for $x \geq 0$, as shown in Fig. 3.3a. Under this circumstance, $P(X \geq x) = P(-X \leq x)$; thus, the cdf of the amplitude is $F_R(x) = F_X(x) - F_X(-x) = 2F_X(x) - 1$ for $x \geq 0$, as shown in Figure 3.3b. It is important to mention that R cannot be considered stable, but the consequences of its definition can be beneficial to the estimation of the parameters of symmetric stable distributions.

Related to the objective of this study is the analysis of the fractional moments of R. Because the univariate case is considered, expressions for $\mathrm{E}|X|^p$ do not differ from that presented in the previous section. Rearranging Eq. 3.4 gives the expression

$$\mathrm{E}|X|^p = \mathrm{E}(R^p) = C_p \Gamma(1 - p/\alpha)\gamma^p, \tag{3.6}$$

where C_p is a constant dependent only on p:

$$C_p = \frac{2^p \Gamma\left(\frac{p+1}{2}\right)}{\sqrt{\pi}\Gamma(1-p/2)}.$$

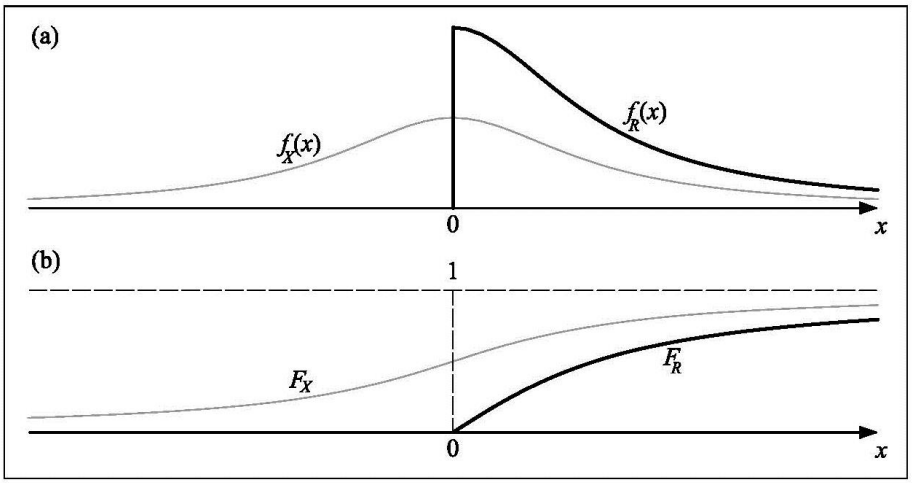


Figure 3.3 Illustration of the pdf and the cdf of the amplitude *R*

The rv R corresponds to a rv $X \sim S\alpha S(1.1,0,1,0)$, where $f_R(x) = 2f_X(x)$ and $F_R(x) = 2F_X(x) - 1$ for $x \geq 0$.

It is worth noting that the two stable parameters that define the distribution of X appear in Eq. 3.6. Analysis of probability $P\{[\mathrm{E}(R^p)]^{1/p} \geq R\} = F_R\{[\mathrm{E}(R^p)]^{1/p}\}$ revealed that it is independent of scale parameter γ:

$$F_R\{[\mathrm{E}(R^p)]^{1/p}\} = 2/\pi \int_0^\infty s^{-1}\sin(s)\exp\left\{-\frac{s^\alpha}{[C_p\Gamma(1-p/\alpha)]^{\alpha/p}}\right\} ds \qquad (3.7)$$

Proof

Let a and $a + h$ (with $h > 0$) be continuity points of $F_X(x)$. The inversion theorem of cfs states that

$$F_X(a+h) - F_X(a) = \lim_{T\to\infty}(1/2\pi)\int_{-T}^{T}(it)^{-1}\left(1 - e^{-ith}\right)e^{-ita}\varphi(t)\mathrm{d}t. \qquad (3.7.1)$$

Replacing $a = 0$ results in $F_X(a+h) - F_X(a) = F_X(h) - F_X(0)$. The right-hand side of Eq. 3.7.1 can be rearranged into

$$F_X(a+h) - F_X(a) = \lim_{T\to\infty}\left[-(i/2\pi)\int_{-T}^{T} t^{-1}\varphi(t)dt + (i/2\pi)\int_{-T}^{T} t^{-1}e^{-ith}\,\varphi(t)\mathrm{d}t\right]. \qquad (3.7.2)$$

For $a < c < b$, the Cauchy principal value (PV) of $\int_a^b \omega(t)\mathrm{d}t$ is denoted by

$$PV\int_a^b \omega(t)\mathrm{d}t = \lim_{\varepsilon\to 0+}\left[\int_a^{c-\varepsilon}\omega(t)dt + \int_{c+\varepsilon}^{b}\omega(t)\mathrm{d}t\right],$$

where $\int_a^c \omega(t)\mathrm{d}t$ and $\int_c^b \omega(t)\mathrm{d}t$ are divergent integrals. If the interval of integration is $[a = -T, b = T]$, $\omega(t)$ is an odd function with a pole at c, and $c = 0$, the limit of

$\int_{-T}^{T} \omega(t)\mathrm{d}t$ exists in the sense of the Cauchy PV and it is equal to 0, even for $T \to \infty$ (Chapter 7 of Ponnusamy, 2012).

Because the function $t^{-1}\varphi(t)$ is odd with a pole at 0, Eq. 3.7.2 should be written as

$$F_X(h) - F_X(0) = \lim_{T\to\infty}\left[-(i/2\pi)PV\int_{-T}^{T} t^{-1}\varphi(t)\mathrm{d}t + (i/2\pi)\int_{-T}^{T} t^{-1}e^{-ith}\,\varphi(t)\mathrm{d}t\right].$$

The limit of the first integral in the sense of the Cauchy PV is equal to 0. The second integral can be solved by applying the change of variable $s = th$, yielding

$$F_X(h) - F_X(0) = \lim_{T\to\infty}(i/2\pi)\int_{-T/h}^{T/h} s^{-1}e^{-is}\varphi(s/h)\,\mathrm{d}s,$$

and applying Euler's formula $e^{-ix} = \cos x - i\sin x$,

$$= \lim_{T\to\infty}\left[(1/2\pi)\int_{-T/h}^{T/h} s^{-1}\sin(s)\varphi(s/h)\mathrm{d}s + (i/2\pi)\int_{-T/h}^{T/h} s^{-1}\cos(s)\varphi(s/h)\,\mathrm{d}s\right].$$

Because the function $s^{-1}\cos(s)\varphi(s/h)$ is odd with a pole at 0, the previous expression should be written as

$$= \lim_{T\to\infty}\left[(1/2\pi)\int_{-T/h}^{T/h} s^{-1}\sin(s)\varphi(s/h)\mathrm{d}s + (i/2\pi)PV\int_{-T/h}^{T/h} s^{-1}\cos(s)\varphi(s/h)\,\mathrm{d}s\right].$$

The limit of the second integral in the sense of the Cauchy PV is equal to 0. The first integral contains an even function $\chi(t)$, such that $\int_{-T}^{0}\chi(t)\mathrm{d}t = \int_{0}^{T}\chi(t)\mathrm{d}t$, and therefore $\int_{-T}^{T}\chi(t)\mathrm{d}t = 2\int_{0}^{T}\chi(t)\mathrm{d}t$. Changing the interval of integration and replacing it with the cf $\varphi(t) = \exp(-\gamma^{\alpha}|t|^{\alpha})$ results in

$$F_X(h) - F_X(0) = (1/\pi)\int_{0}^{\infty} s^{-1}\sin(s)\,e^{-\gamma^{\alpha}[s/h]^{\alpha}}\,\mathrm{d}s.$$

Considering that $F_R(h) = 2[F_X(h) - F_X(0)]$,

$$F_R(\mathrm{h}) = 2/\pi\int_{0}^{\infty} s^{-1}\sin(s)\,e^{-\gamma^{\alpha}[s/h]^{\alpha}}\,ds. \qquad (3.7.3)$$

It is easy to see that values of $h = [\mathrm{E}(R^p)]^{1/p}$ eliminate the scale parameter γ in Eq. 3.7.3. Equation 3.7 can be obtained by replacing Eq. 3.6 in 3.7.3. □

Using the Cauchy PV is not necessarily the most elegant solution because it deals with improper integrals. However, it should be noted: (i) the interval of integration defined in the inversion theorem is symmetric with respect to $t = 0$; and (ii) the integrals contain odd functions that are integrable at $\lim_{T\to\infty} t = \pm T$ and that have the same rate of approximation to infinity at both sides of $t = 0$. These conditions make it possible to assume that the limits of the integrals in the sense of the Cauchy PV exist and are equal to 0.

3.2.4 Simulation of α-stable rvs

An efficient algorithm for simulating sequences of stable rvs has been presented by Chambers et al. (1976), hereafter denoted as the CMS algorithm. A detailed proof as well as some notes about the change of parameterization is presented by Weron (1996). Eqs. 3.8 and 3.9 correspond to the parameterization given by Weron (1996). For a rv U uniformly distributed on $(-\pi/2, \pi/2)$ and an independent exponential rv W with mean 1, the algorithm for simulating the rv $Y \sim S(\alpha, \beta, 1, 0)$ is:

$$Y = (1+\zeta^2)^{1/2\alpha} \frac{\sin\{\alpha(U+\xi)\}}{\{\cos(U)\}^{1/\alpha}} \left[\frac{\cos\{U-\alpha(U+\xi)\}}{W}\right]^{\frac{1-\alpha}{\alpha}} \quad \text{for } \alpha \neq 1, \text{ and} \tag{3.8}$$

$$Y = \frac{1}{\xi}\left\{\left(\frac{\pi}{2}+\beta U\right)\tan U - \beta \ln\left(\frac{\frac{\pi}{2} W \cos U}{\frac{\pi}{2}+\beta U}\right)\right\} \quad \text{for } \alpha = 1, \tag{3.9}$$

where $\zeta = -\beta\tan(\pi\alpha/2)$, $\xi = \pi/2$ for $\alpha = 1$, and $\xi = (1/\alpha)\arctan(-\zeta)$ for $\alpha \neq 1$. Because Eqs. (8) and (9) are given for a standard stable rv, the simulation of $X \sim S(\alpha, \beta, \gamma, \delta)$ as a function of Y is $X = \gamma Y + \delta$ for $\alpha \neq 1$, and $X = \gamma Y + (2/\pi)\beta\gamma \ln\gamma + \delta$ for $\alpha = 1$. If $X \sim S(\alpha, \beta, \gamma, \delta)$ has parameterization (2), then the simulation of X is given by $X = \gamma[Y - \beta\tan(\pi\alpha/2)] + \delta$ for $\alpha \neq 1$, and $X = \gamma Y + \delta$ for $\alpha = 1$.

The CMS algorithm is regarded as the most accurate and it is currently implemented in the latest statistical computing environments (e.g., Veillette, 2012; Liang and Chen, 2013).

3.2.5 Estimation of stable parameters

Techniques commonly used to estimate the parameters of probability distributions are not suitable for stable distributions because they rely on pdfs with closed-form expressions and on the existence of finite moments. Alternatively, the fitting of samples has been done through numerical methods using quantile relative measures (Fama and Roll, 1971; McCulloch, 1986; Dominicy and Veredas, 2013), regression of the empirical cf (Arad, 1980; Koutrouvelis, 1980; Kogon and Williams, 1998; Besbeas and Morgan, 2008), and approximate maximum-likelihood estimation (Mittnik et al., 1999b; Nolan, 2001). The evaluation and comparison of the different techniques is often undertaken using statistical metrics computed from Monte Carlo simulations. The results have shown acceptable performances of all these methods, especially when large samples were considered. Even though the maximum-likelihood method performs with slightly better accuracy, quantile-type and regression-type methods are usually preferred because of the lesser computational effort (Besbeas and Morgan, 2008; Weron, 2011). In this section, we review the basic descriptions of two existing methods and detail the procedures of a few others based on the concept of the amplitude of $S\alpha S$ distributions.

3.2.5.1 Regression of the empirical cf (RCF)

Consider a random sample $x_1, x_2, \ldots x_n$ from a population with a stable distribution $S(\alpha, \beta, \gamma, \delta)$. The empirical cf, $\hat{\varphi}(t)$, is defined as the expected value of $\exp(it_k x_j)$, i.e., $\hat{\varphi}(t) = n^{-1} \sum_{j=1}^{n} \exp(it_k x_j)$. This expression implies that the empirical cf is bounded by unity, making all its moments finite. Because of these conditions, and by the law of large numbers, $\hat{\varphi}(t)$ is considered a consistent estimator of $\varphi(t)$. The selection of the real-valued set of evaluation points, $\{t_k, k \in \mathbb{Z}\}$, often differs from one author to another. Besbeas and Morgan (2008) presented a literature review and compared the performance of the different ways in which t_k was defined.

The RCF method, introduced by Koutrouvelis (1980), is often regarded as the estimation method that best balances complexity of programming and computation time. Manipulations of the cf, as shown in Eq. 3.1, give

$$\log(-\log|\varphi(t)|^2) = \log(2\gamma^\alpha) + \alpha \log|t|, \tag{3.10}$$

and, considerations of principal values apart, the real $\Re[\varphi(t)]$ and imaginary $\Im[\varphi(t)]$ parts of the cf can lead to the expression:

$$\arctan\left(\frac{\Im[\varphi(t)]}{\Re[\varphi(t)]}\right) = \delta t + \beta \gamma^\alpha \tan(\pi\alpha/2) \operatorname{sgn}(t) |t|^\alpha. \tag{3.11}$$

Expression 3.10 depends only on parameters α and γ, suggesting that both estimators $\hat{\alpha}$ and $\hat{\gamma}$ could be found by computing the regression of $y = \log(-\log|\varphi_n(t)|^2)$ on $w = \log|t|$ in the model: $y_k = \mu_0 + \alpha w_k + \epsilon_k$, where $\mu_0 = \log(2\gamma^\alpha)$ and ϵ_k denotes an error term. If it is known a priori that the random sample is from a population with a $S\alpha S(\alpha, 0, \gamma, 0)$ distribution, Eq. 3.11 is unnecessary. Moreover, note that an expression equivalent to Eq. 3.10 could be found by rearranging the cf in Eq. 3.3. Nevertheless, the method for estimating all four parameters is explained next for completeness.

Once $\hat{\alpha}$ and $\hat{\gamma}$ are estimated and set equal to α and γ, Eq. 3.11 may be used to obtain estimators $\hat{\beta}$ and $\hat{\delta}$ through a second regression. The procedure, explained by Koutrouvelis (1980), involves an initial estimate of γ, which is computed using Fama and Roll's (1971) formula, and an initial estimate of δ, which is found by computing the 25% truncated mean. After the computation of estimators $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}$, and $\hat{\delta}$, these values are fixed as the initial condition of the subsequent iteration. The procedure is repeated iteratively until some criterion of convergence is satisfied. The evaluation points t_k are uniformly spaced, where k is a positive integer between 9 and 134, depending on the initial value of α and the sample size.

Kogon and Williams (1998) resolved the discontinuity of the cf near $\alpha = 1$ by deriving the linear regression model from Eq. 3.2, which is the continuous expression of the cf. Additional improvements of the model proposed by Kogon and Williams (1998) include the

elimination of the iteration process and the use of only 10 uniformly spaced evaluation points. Furthermore, the study proposed the use of McCulloch's (1986) expressions for obtaining the initial estimators. The model developed by Kogon and Williams (1998) is slightly faster than the RCF method, which renders an improved performance for distributions with parameters near $\alpha = 1$ and $\beta \neq 0$. However, the RCF method remains more accurate for other values of α (Weron, 2011; Borak et al., 2011).

In this study, a MATLAB toolbox (Alpha–Stable distributions by Veillette, 2012), which uses McCulloch's (1986) expressions for the initial estimates of the parameters, was used to generate the results of the RCF method presented in later sections.

3.2.5.2 Quantile-based relative measures (QRM)

Improving an earlier work on the estimation of parameters via quantiles conducted by Fama and Roll (1971), McCulloch (1986) proposed estimators for all four parameters, which eliminated the previous asymptotic bias and covered a broader range of the characteristic exponent ($\alpha \geq 0.6$). Let us denote by x_r the r-th population quantile such that the probability of the rv being less than or equal to x_r is r. McCulloch defined

$$\upsilon_\alpha = (x_{0.95} - x_{0.05})/(x_{0.75} - x_{0.25}), \text{ and } \upsilon_\beta = (x_{0.95} + x_{0.05} - 2x_{0.50})/(x_{0.95} - x_{0.05}), \quad (3.12)$$

where υ_α is a relative measure of the heaviness of the tails with respect to the body of the distribution, and υ_β is a relative measure of the skewness. If $\hat{\upsilon}_\alpha$ and $\hat{\upsilon}_\beta$ are the corresponding sample values, then $\hat{\upsilon}_\alpha$ and $\hat{\upsilon}_\beta$ are considered consistent estimators of υ_α and υ_β, respectively. As these expressions were found to be independent of γ and δ, the estimators $\hat{\alpha}$ and $\hat{\beta}$ can be obtained by linearly interpolating the values of $\hat{\upsilon}_\alpha$ and $\hat{\upsilon}_\beta$ using lookup tables computed for different values of α and β.

Because of the proportionality between γ and the difference between any two quantiles for fixed values of α and β, the following relative measure of scale was defined:

$$\upsilon_\gamma = (x_{0.75} - x_{0.25})/\gamma. \quad (3.13)$$

If $\hat{\upsilon}_\gamma$ is the corresponding sample value, the consistent estimator $\hat{\gamma}$ can also be linearly interpolated using lookup tables. A similar expression for the location parameter was also presented. However, the discontinuity of the cf for distributions near $\alpha = 1$ and $\beta \neq 0$ produces singularities that make interpolation futile. The solution to overcome this pitfall, which also assures the continuity of the cf, involves a change of variable using the median of the shifted data. All the aforementioned lookup tables are presented in McCulloch (1986).

3.2.5.3 Modified quantile-based relative measures (MQM)

In the case of zero-mode $S\alpha S$ distributions, x_r is equal to $-x_{1-r}$. Let $|x|_{r^*}$ be the r^*-th population quantile of a univariate rv $R = |X|$ with $X \sim S\alpha S(\alpha, 0, \gamma, 0)$, where $r^* =$

$2(r-0.5)$ for $r \geq 0.5$. Derived directly from Eq. 3.12, the relative measure v_α can be expressed as

$$v_\alpha = |x|_{0.90}/|x|_{0.50}. \tag{3.14}$$

Similarly, derived from Eq. 3.13, the relative measure of scale results in

$$v_\gamma = 2|x|_{0.50}/\gamma. \tag{3.15}$$

The MQM method estimates $\hat{\alpha}$ and $\hat{\gamma}$ using Eqs. 3.14 and 3.15 and the tabulated data computed by McCulloch (1986). As only $S\alpha S$ distributions are considered, only the columns corresponding to $\beta = 0$ in those tables are relevant to this method. In order to compare this method with the QRM method, the estimation of parameters with the QRM method was done by fixing the parameter $\beta = 0$.

3.2.5.4 Fractional absolute moments (FAM) method

Equations 3.6 and 3.7 relate α and γ with the pth fractional absolute moment of the amplitude R. This condition suggests the possibility of computing the estimators $\hat{\alpha}$ and $\hat{\gamma}$ from a random sample using only a predefined value of p.

Given a random sample $x_1, x_2, \ldots, x_n$ from a population with distribution $S\alpha S(\alpha, 0, \gamma, 0)$, the empirical pth fractional absolute moment is $m_p = n^{-1}\sum_{j=1}^{n}|x_j|^p$, which may be considered an unbiased estimator of $\mathrm{E}(R^p)$. The absolute values of the random sample are then arranged in ascending order for the purpose of computing values of the empirical cdf, $q(j) = (j-0.5)/n$ for $j = 1, 2, \ldots, n$. The empirical cdf at ${m_p}^{1/p}$ can be calculated by applying linear interpolation between two corresponding adjacent values of $q(j)$. We denote this value by $\hat{q}_p$, which may be considered a consistent estimator of the cdf evaluated at $[\mathrm{E}(R^p)]^{1/p}$.

Because of the monotonically decreasing relationship observed in Eq. 3.7 between $F_R\{[\mathrm{E}(R^p)]^{1/p}\}$ and α in the interval $p < \alpha \leq 2$ (see Fig. 3.4a), it is possible to numerically interpolate estimator $\hat{\alpha}$ for a given value of $\hat{q}_p$. Because Eq. 3.7 is valid only if $\alpha > p$, the use of small random samples might cause the following issue. If a sample comes from a population with $S\alpha S$ distribution, $0 < p < \alpha$, and p is assigned a value almost equal to the parameter α of the distribution, the empirical fractional absolute moment m_p should be a large number tending to infinity in order to be a good estimator of $\mathrm{E}(R^p)$. However, infinite empirical moments can only be obtained from samples with an infinite number of drawings, which is practically impossible. Considering an alternative case in which p is assigned a value close to α, such that $\mathrm{E}(R^p)$ is still a large number (but not infinity), it would require a sample with a large number of evenly distributed outliers to represent adequately the heavy tails, which is a condition that cannot be fulfilled by small or even moderately large samples.

For this reason, the value of p should be such that $p/\alpha \leq \tau'_{max} < 1$, where τ'_{max} is an upper bound that assures the computation of representative values of m_p.

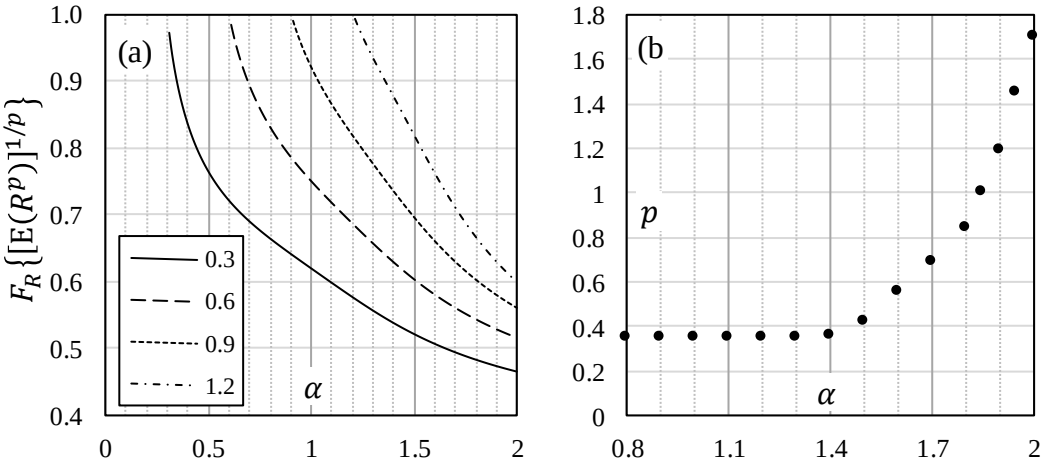


Figure 3.4 α vs. $F_R\{[\mathrm{E}(R^p)]^{1/p}\}$ and p_α for different values of α

(a) Curves obtained by evaluating Eq. 3.7 with values of $p = \{0.3,0.6,0.9,1.2\}$. When the fractional absolute moments tend to infinity, as α approaches p, note that the values of $F_R\{[\mathrm{E}(R^p)]^{1/p}\}$ tend toward 1. (b) Average results of multiple simulation runs to determine the behavior of the estimator $\hat{\alpha}$ for different values of p. Black points correspond to p_α.

Another restriction needs to be established following careful observation of Eq. 3.7, which contains an oscillatory integral with asymptotic decay at infinity. The value of p should be such that $0 < \tau_{min} \leq p/\alpha \leq \tau_{max} < 1$. If p/α is outside the interval $[\tau_{min}, \tau_{max}]$, the asymptotic decay might become slower, producing large errors in the numerical approximation of the integral. The value of τ_{max} may not be equal to τ'_{max} because τ'_{max} depends on the characteristics of the sample. On the other hand, the values of τ_{max} and τ_{min} depend on the tolerances established in the integral-approximation algorithms of the utilized computing environment.

Choosing a value of p is complicated because the value of α cannot be known a priori nor is it known whether it approximates p. Simulation runs were undertaken to determine the accuracy of the estimation of $\hat{\alpha}$ using Eq. 3.7 and several values of p. Figure 3.4b shows the results of the simulations, which represent the average behavior of 500 replications and sample sizes $n = \{50,100,150,200\}$. A value of p that returned the most accurate estimate of $\hat{\alpha}$ was identified, which is denoted as p_α. This analysis indicates that the estimation of parameter α of $S\alpha S$ distributions using fractional absolute moments is statistically more accurate if the p_αth-order absolute moment is used. In fact, the performance in estimating $\hat{\alpha}$ and $\hat{\gamma}$ with Eqs. 3.6 and 3.7 was found to match the simulation results of the RCF method while being approximately five times faster.

Table 3.1 Values of p_α for different values of the characteristic exponent α

α	2.00	1.95	1.90	1.85	1.80	1.70	1.60	1.50	1.40	≤1.30
p_α	1.70	1.45	1.19	1.00	0.84	0.69	0.56	0.42	0.36	0.35

The FAM method is structured as follows. First, a value of p_α is interpolated in Table 3.1 using an initial estimation of $\hat{\alpha}$ obtained with the truncated sample mean (TSM) method (next section). Second, by replacing p_α in Eq. 3.7, an estimate of $\hat{\alpha}$ is obtained. Finally, the estimator $\hat{\gamma}$ is computed by making $\alpha = \hat{\alpha}$ in Eq. 3.6 and solving for γ. The range in which the FAM method can be applied depends on the adjustment of the tolerances of the integral-approximation algorithm of the utilized computing environment. These adjustments may also affect the time of computation. The observations and results presented in this study correspond to the default tolerances set for the "integral" algorithm of MATLAB R2014a. Acceptable results of simulations of this method were obtained for $0.8 \leq \alpha \leq 2.0$.

3.2.5.5 Truncated sample mean (TSM) method

During the development of the FAM method, it was identified that small samples cannot generate representative empirical fractional absolute moments. The TSM method is proposed to overcome this shortcoming, given the assumptions below.

If a given sample is truncated at the quantile $|x|_{r^*}$, all the drawings less than or equal to $|x|_{r^*}$ are considered to contain sufficient information with which to estimate the parameters of the distribution. The expected value of the amplitude R truncated at the quantile $|x|_{r^*}$ can be expressed as $\mathrm{E}(R|R \leq |x|_{r^*})$; for simplicity, hereafter, this is denoted as just $\mathrm{E}(R)_{\mathrm{T}p^*}$. For different values of α, this expected value can be estimated by means of numerical integration of the pdf using Nolan's (1997) algorithms. Similarly, the probability of $\mathrm{E}(R)_{\mathrm{T}p^*} \geq R$, denoted as $F_R\left(\mathrm{E}(R)_{\mathrm{T}p^*}\right)$, can also be computed using Nolan's (1997) algorithms. The value of $F_R\left(\mathrm{E}(R)_{\mathrm{T}p^*}\right)$ represents the cdf evaluated at the expected value of the truncated pdf. Table 2 shows the two described parameters computed for $p^* = 0.99$ and $\gamma = 1$. Even though the expected value of the amplitude R exists for $\alpha > 1$ only, values of $\mathrm{E}(R)_{\mathrm{T}p^*}$ and $F_R\left(\mathrm{E}(R)_{\mathrm{T}p^*}\right)$ can be computed for values of $\alpha \leq 1$ by truncating the distribution.

The TSM method estimates the parameters of $S\alpha S$ distributions as follows. Given a random sample $x_1, x_2, \ldots x_n$ from a population with distribution $S\alpha S(\alpha, 0, \gamma, 0)$, $\overline{|x|}_{\mathrm{T}0.99}$ denotes the mean of the absolute values of all drawings that are less than or equal to the 0.99th quantile. $\overline{|x|}_{\mathrm{T}0.99}$ is then considered an unbiased estimator of $\mathrm{E}(R)_{\mathrm{T}0.99}$. The absolute values of the random sample are then arranged in ascending order with the purpose of computing values of the empirical cdf, $q(j) = (j - 0.5)/n$ for $j = 1, 2, .., n$. The empirical cdf at $\overline{|x|}_{\mathrm{T}0.99}$ can be calculated by applying linear interpolation between two corresponding adjacent values of $q(j)$. This value is denoted $\hat{q}_{T0.99}$ and it may be considered a consistent estimator of the distribution function evaluated at the expected value of the truncated pdf $F_R(\mathrm{E}(R)_{\mathrm{T}0.99})$. Using Table 3.2, the estimator $\hat{\alpha}$ and a value of $\mathrm{E}|X|_{T0.99}^{\gamma=1}$ can be obtained

by linearly interpolating with $\hat{q}_{T0.99}$. Then, the estimator $\hat{\gamma}$ can be computed using the expression shown in Eq. 3.16, because $\mathrm{E}(R)_{\mathrm{T0.99}}$ is the product of γ and $\mathrm{E}|X|_{\mathrm{T0.99}}^{\gamma=1}$.

$$\hat{\gamma} = \overline{|x|}_{\mathrm{T0.99}} / \mathrm{E}|X|_{\mathrm{T0.99}}^{\gamma=1}. \tag{3.16}$$

Different values of the p^*-th quantiles were tested for these analysis; however, the value of $p^* = 0.99$ returned the most satisfying results.

Table 3.2 Expected value of the truncated pdf and its corresponding value in the distribution function

α	$\mathrm{E}\|X\|_{\mathrm{T0.99}}^{\gamma=1}$	$F_{\|X\|}(\mathrm{E}\|X\|_{\mathrm{T0.99}})$	α	$\mathrm{E}\|X\|_{\mathrm{T0.99}}^{\gamma=1}$	$F_{\|X\|}(\mathrm{E}\|X\|_{\mathrm{T0.99}})$
2.00	1.0985	0.5627	1.20	1.8533	0.7210
1.95	1.1108	0.5668	1.15	1.9967	0.7340
1.90	1.1254	0.5716	1.10	2.1726	0.7468
1.85	1.1426	0.5773	1.05	2.3921	0.7595
1.80	1.1630	0.5841	1.00	2.6710	0.7719
1.75	1.1866	0.5917	0.95	3.0329	0.7844
1.70	1.2138	0.6003	0.90	3.5145	0.7969
1.65	1.2448	0.6097	0.85	4.1747	0.8096
1.60	1.2801	0.6199	0.80	5.1120	0.8224
1.55	1.3204	0.6309	0.75	6.5002	0.8355
1.50	1.3666	0.6426	0.70	8.6647	0.8488
1.45	1.4196	0.6549	0.65	12.2602	0.8623
1.40	1.4809	0.6678	0.60	18.7248	0.8759
1.35	1.5522	0.6810	0.55	31.5797	0.8894
1.30	1.6359	0.6943	0.50	60.7098	0.9027
1.25	1.7349	0.7077			

3.2.5.6 Logarithmic moments (LMO) method

Other techniques for estimating the parameters of $S\alpha S$ distributions employ fractional lower-order moments and LMO (Section 5.7 of Nikias and Shao, 1995). A Monte Carlo simulation showed that a method based on fractional (absolute and signed) lower-order moments and the method of Kogon and Williams (1998) estimated parameter α with a similar performance, while the LMO estimators returned poorer statistics (Dance and Kuruoğlu, 1999). Since Weron (2011) demonstrated that the RCF method is slightly more accurate than the method of Kogon and Williams (1998), we assume that the RCF method is also better than the method of moments proposed by Dance and Kuruoğlu (1999) for $\beta = 0$ and $\alpha > 1$. However, in order to have an idea of the range in which parameter α can be estimated using different techniques, the LMO method is detailed in this section and it will be included in the simulation undertaken in Section 3.2.6.

Even though the rv R has finite pth fractional lower-order moments for $p < \alpha$ only, its logarithm has moments of all orders (Section 5.8 in Uchaikin and Zolotarev, 1999). Based

on this definition, the parameters α and γ can be estimated using the first two logarithmic moments of the rv R (Nikias and Shao, 1995), as shown in Eqs. 3.17 and 3.18.

$$\mathrm{E}(\ln R) = \gamma_{Euler}\left(\frac{1}{\alpha} - 1\right) + \ln\gamma. \tag{3.17}$$

$$\mathrm{Var}(\ln R) = \frac{\pi^2}{6}\left(\frac{1}{\alpha^2} + \frac{1}{2}\right). \tag{3.18}$$

The application of the LMO method to real data is hindered in some cases where the samples have values equal to zero, which makes it impossible to compute the mean and variance of the logarithms.

3.2.6 Simulation of the parameter-estimation methods

This section presents the results of a simulation using the parameter-estimation methods previously detailed. The objective is to evaluate the level of accuracy obtained using the various techniques.

The procedure is as follows. A vector of stable random variates is generated with the CMS algorithm. The vector is subdivided into $K = 500$ samples of size n. The total number of simulation runs, K, determines the quantity of estimator points used for the statistical evaluation of the methods. This simulation was performed for several sample sizes that ranged from $n = 50$ to 20,000. As we are interested in the evaluation of the performance of small samples, the results for values of n equal to 50, 100, and 200 are presented below. However, the results for $n = 1000$ are also presented to show the performance of large samples. The MQM, FAM, TSM, and LMO methods use the absolute values of the random variates. The simulation was implemented for input values of $\alpha = \{0.8, 0.9, \ldots, 1.9, 2.0\}$, and constant input values for the other three parameters fixed at $\beta = 0$, $\gamma = 1$, and $\delta = 0$.

The estimator point obtained from the k-th simulated sample with input true parameter θ^* is denoted $\hat{\theta}_k$ $(k = 1, 2, \ldots, K)$. The results of this analysis are then expressed in terms of commonly used descriptive statistics, which include the mean

$$\bar{\theta} = K^{-1}\sum_{k=1}^{K}\hat{\theta}_k, \tag{3.19}$$

root mean square error

$$\mathrm{RMSE} = \left[K^{-1}\sum_{k=1}^{K}\left(\hat{\theta}_k - \theta^*\right)^2\right]^{1/2}, \tag{3.20}$$

and standard error of the mean

$$\mathrm{SEM} = K^{-1/2}\left[K^{-1}\sum_{k=1}^{K}\left(\hat{\theta}_k - \bar{\theta}\right)^2\right]^{1/2}. \tag{3.21}$$

It was observed that the mean value for all methods rapidly converged and stabilized after approximately $k = 250$ simulation runs. However, $K = 500$ was maintained in this analysis such that the results could be compared with those of previous evaluations of fitting models, e.g., the studies of Akgiray and Lamoureux (1989) and Besbeas and Morgan (2008).

Table 3.3 Simulation results for estimator $\hat{\alpha}$

α	Statistics[(a)]	RCF	QRM	MQM	FAM	TSM	LMO
		n = 50					
1.90	Mean	1.864	1.847	1.825	1.833	1.839	1.773
	RMSE	0.169	0.226	0.225	0.197	0.190	0.332
	SEM	74.0	98.1	94.7	83.0	80.5	137.3
1.70	Mean	1.682	1.689	1.704	1.665	1.718	1.732
	RMSE	0.234	0.295	0.251	0.263	0.221	0.313
	SEM	104.2	131.7	112.2	116.8	98.4	139.2
1.50	Mean	1.502	1.503	1.542	1.505	1.554	1.605
	RMSE	0.257	0.345	0.265	0.291	0.270	0.377
	SEM	114.8	154.4	117.2	130.2	118.3	162.1
1.30	Mean	1.277	1.360	1.336	1.315	1.390	1.425
	RMSE	0.262	0.352	0.249	0.302	0.276	0.363
	SEM	116.6	154.9	110.2	134.7	117.0	152.4
1.10	Mean	1.039	1.151	1.118	1.097	1.199	1.210
	RMSE	0.245	0.327	0.218	0.271	0.269	0.322
	SEM	106.1	144.6	97.1	121.4	111.8	135.4
0.90	Mean	0.835	0.994	0.930	0.939	1.040	0.970
	RMSE	0.210	0.299	0.188	0.254	0.274	0.221
	SEM	89.2	126.8	83.0	112.2	105.3	93.9
		n = 100					
1.90	Mean	1.888	1.889	1.838	1.853	1.832	1.816
	RMSE	0.117	0.163	0.188	0.157	0.167	0.269
	SEM	52.2	72.8	79.2	67.0	68.4	114.2
1.70	Mean	1.700	1.705	1.719	1.704	1.673	1.728
	RMSE	0.166	0.243	0.204	0.194	0.202	0.286
	SEM	74.2	108.4	90.6	86.6	89.3	127.1
1.50	Mean	1.491	1.500	1.521	1.509	1.472	1.558
	RMSE	0.174	0.264	0.205	0.213	0.204	0.307
	SEM	77.8	118.2	91.4	95.3	90.5	135.1
1.30	Mean	1.289	1.292	1.316	1.302	1.259	1.409
	RMSE	0.168	0.235	0.178	0.204	0.220	0.304
	SEM	75.1	105.1	79.1	91.1	96.7	126.8
1.10	Mean	1.065	1.109	1.110	1.099	1.069	1.140
	RMSE	0.165	0.211	0.154	0.192	0.213	0.209
	SEM	72.1	94.4	68.6	85.8	94.3	91.9
0.90	Mean	0.876	0.951	0.925	0.924	0.898	0.934
	RMSE	0.118	0.186	0.132	0.184	0.184	0.139
	SEM	51.9	79.9	57.9	81.4	82.4	60.3

[(a)] The values of SEM are multiplied by a factor of 10^4

The overall analysis of the results, presented in Table 3.3, suggests that parameter α can be estimated with any of the methods and that similar satisfactory outcomes could be expected when dealing with large samples. For small samples, the main observations are outlined as follows:

- The MQM method uses the Eqs. 3.14 and 3.15, which are modifications of Eqs. 3.12 and 3.13 of the QRM method. The hypothesis was that these modifications could improve the accuracy of the estimation in small samples. However, the improvement

can be seen in the values of RMSE and SEM for $\alpha^* < 1.7$, while the mean of the estimators is slightly better for a few cases only.

Table 3.3 (Continuation) Simulation results for estimator $\hat{\alpha}$

α	Statistics[a]	RCF	QRM	MQM	FAM	TSM	LMO
		n = 200					
1.90	Mean	1.902	1.926	1.877	1.886	1.876	1.816
	RMSE	0.077	0.124	0.141	0.110	0.123	0.237
	SEM	34.4	54.3	62.4	48.6	53.9	99.2
1.70	Mean	1.704	1.730	1.730	1.708	1.686	1.706
	RMSE	0.113	0.204	0.163	0.136	0.144	0.246
	SEM	50.4	90.3	71.7	60.9	63.9	110.0
1.50	Mean	1.499	1.487	1.524	1.511	1.486	1.562
	RMSE	0.119	0.189	0.148	0.151	0.139	0.252
	SEM	53.1	84.1	65.5	67.4	62.0	109.3
1.30	Mean	1.293	1.263	1.310	1.296	1.278	1.348
	RMSE	0.106	0.154	0.118	0.136	0.139	0.197
	SEM	47.4	66.9	52.5	61.0	61.2	85.5
1.10	Mean	1.092	1.098	1.113	1.107	1.084	1.130
	RMSE	0.108	0.144	0.109	0.144	0.154	0.142
	SEM	48.3	64.6	48.5	64.3	68.7	61.9
0.90	Mean	0.896	0.939	0.916	0.927	0.899	0.922
	RMSE	0.078	0.140	0.095	0.140	0.150	0.094
	SEM	34.9	60.0	41.9	61.6	67.2	40.9
		n = 1000					
1.90	Mean	1.901	1.978	1.900	1.900	1.900	1.882
	RMSE	0.038	0.092	0.087	0.071	0.073	0.131
	SEM	17.2	21.8	38.8	31.7	32.8	58.2
1.70	Mean	1.702	1.741	1.709	1.701	1.699	1.717
	RMSE	0.051	0.120	0.085	0.072	0.064	0.147
	SEM	22.7	50.6	38.0	32.0	28.6	65.4
1.50	Mean	1.497	1.459	1.501	1.497	1.492	1.508
	RMSE	0.051	0.092	0.061	0.068	0.059	0.110
	SEM	22.8	36.9	27.1	30.3	26.0	49.3
1.30	Mean	1.296	1.251	1.302	1.294	1.289	1.295
	RMSE	0.048	0.085	0.054	0.069	0.064	0.082
	SEM	21.4	30.9	24.3	30.9	28.1	36.8
1.10	Mean	1.098	1.086	1.100	1.100	1.091	1.104
	RMSE	0.041	0.068	0.044	0.065	0.072	0.055
	SEM	18.1	29.6	19.8	29.2	31.8	24.4
0.90	Mean	0.900	0.923	0.905	0.901	0.898	0.903
	RMSE	0.037	0.065	0.042	0.072	0.073	0.040
	SEM	16.7	27.3	18.6	32.3	32.4	17.8

[a] The values of SEM are multiplied by a factor of 10^4

- With some exceptions, the LMO method often shows the poorest results. This observation does not mean that the descriptive statistics are always unacceptable. For example, although the results of all methods return similar values of RMSE and SEM for $n = 200$ and $\alpha^* = 1.1$, the LMO method shows only a slightly worse mean estimator.
- The RCF method shows exceptional statistics for values of α^* approaching 2.0; however, the results tend to be less accurate for $\alpha^* < 1.5$. The discontinuity of the characteristic function at $\alpha = 1$ could be the cause of this deterioration.

- For values of $\alpha^* < 1.6$, the performance of the FAM method shows, in most cases, the most accurate mean estimate of α. Its statistics are slightly surpassed by the results of the other methods for values of α approaching 2.
- The TSM method, along with the LMO method, showed the worst mean estimators for sample sizes of $n = 50$ and $n = 100$. However, its values of RMSE and SEM are usually better than the average of all methods.

For the estimators $\hat{\gamma}$, the general observation from the results shows outstanding performance by the MQM method, while the RCF, FAM, TSM and LMO methods show fairly acceptable results. The method that returns the poorest statistics is the QRM method.

3.3 CONDITIONS FOR SELF-SIMILARITY IN α-STABLE DISTRIBUTIONS

A process $X(\cdot)$ is called self-similar in distribution if for every $a > 0$, there exists some unique $H > 0$ such that $X(a\,\cdot) \stackrel{d}{=} a^H X(\cdot)$ (Chapter 7 in Samorodnitsky and Taqqu, 1994; Embrechts et al., 1997). The scaling exponent H is called the self-similarity index. This expression implies that if the process $X(\cdot)$ goes through a transformation of scale, the distribution of the transformed process will be equal to the distribution of the original process multiplied by a^H.

The discrete wavelet transform uses two-dimensional dyadic dilates of the scaling function (low-pass filter) in order to obtain discrete approximations at multiple resolutions. The finite number of samples comprised in $\bar{X}_m$ is one fourth the of the samples comprised in $\bar{X}_{m-1}$. The same conclusion can be drawn for the number of samples in the sets of wavelet coefficients $X'_{m,i}$ (for $i = 1,2,3$) at each scale. Considering that the highest resolution of representation is $m = 0$ (scale of sampling) and that discrete representations at lower resolutions correspond to positive integer values of m, it is not difficult to imply that the coefficient a^H should be equal to $4^{(m-1)H}$ in a wavelet transform framework. Even though the undecimated discrete wavelet transform generates samples with the same dimensions as the original data, the redundant information comprises similar copies of the samples generated with the discrete wavelet transform. Thus, it is not wrong to assume $a^H = 4^{(m-1)H}$ for undecimated samplings as well.

It can be deduced from the previous definition, for symmetric α-stable distributions, the process $X \sim S\alpha S(\alpha, 0, \gamma, 0)$ is self-similar in distribution if the transformed process has distribution $S\alpha S(\alpha, 0,\ a^H\gamma, 0)$, or $S\alpha S\big(\alpha, 0,\ 4^{(m-1)H}\gamma, 0\big)$. Therefore, two conditions are necessary to demonstrate the existence of self-similarity when considering $S\alpha S$ distributions. First, the characteristic exponent α should be the same, meaning that the frequency of extreme values is invariant under these transformations. Second, the scale parameter γ should have exponential growth with rate H (power-law behavior) as the resolution becomes

lower. The self-similarity index *H* can be found by making a logarithmic regression of the estimated scale parameters at all scales.

3.4 RESULTS

This Section presents the results of the procedures necessary to determine the presence of self-similarity in the distributions of local fluctuations of rainfall, represented by the sets of wavelet coefficients at different scales. The datasets used for this analysis are the rainfall intensity composites denominated Radar-AMeDAS, which are available across the Japanese Islands. The extraordinary characteristics of the datasets and the possibility to stochastically analyze the spatial patterns of different precipitation systems are conditions that probably could not be matched in other locations of the world.

3.4.1 Area of study and characteristics of the rainfall intensity datasets

Japan, consisting of four main islands and thousands of smaller surrounding ones, forms an elongated arc covering a range of latitude of about 22 degrees. The Sea of Japan separates the Japanese archipelago from the Pacific coast of the Asian continent. The four main islands (Honshu, Hokkaido, Kyushu and Shikoku), shown in Figure 3.5, have vast mountainous regions that are about three-fourths of Japan's total land surface. The particular geographical features, combined with the ocean currents, happen to have a direct influence on the weather of the diverse climatic zones that range from subarctic in the north to subtropical in the south. Furthermore, the climatic conditions and the annual precipitation patterns can also be quite different between the Pacific side and the Sea of Japan side.

Mainly three types of precipitation systems take place annually over Japan and the surrounding ocean waters: cold fronts, the baiu front, and tropical cyclones. Heavy precipitation is also generated by local convective updraft over the plains during summer (e.g., Kanto region) and by orographic lift; however, the extent is much smaller than the ones of the three aforementioned precipitation systems.

Extratropical cyclones, picking up moisture as they pass over the Sea of Japan, cause cold frontal precipitations during spring, autumn and early winter. If a well-developed cyclone reaches the coasts of the Sea of Japan, bands of high-intensity rainfall or snowfall precipitate over the Main Island (Honshu). However, if the cyclone is not strong enough, it might precipitate only over the Hokuriku region, or move upward the mountain range and precipitate over the northwest areas of the Chubu mountains without approaching the plains of the Kanto region (Yoshida and Asuma, 2004; Kusaka and Kitahata, 2009).

The baiu front is part of the East Asian Summer Monsoon. This phenomenon, which is a long quasi-stationary precipitation belt, is the result of a high-pressure mass of cold air above the Sea of Okhotsk encountering a high-pressure mass of moist warm air developed over the Pacific Ocean (Ohba et al., 2014). Southwesterly flow along the western rim of the North

Pacific subtropical high introduces warm moist air into the front. Consequently, damp weather and prolonged periods of high-intensity rainfall are characteristic of the baiu front season.

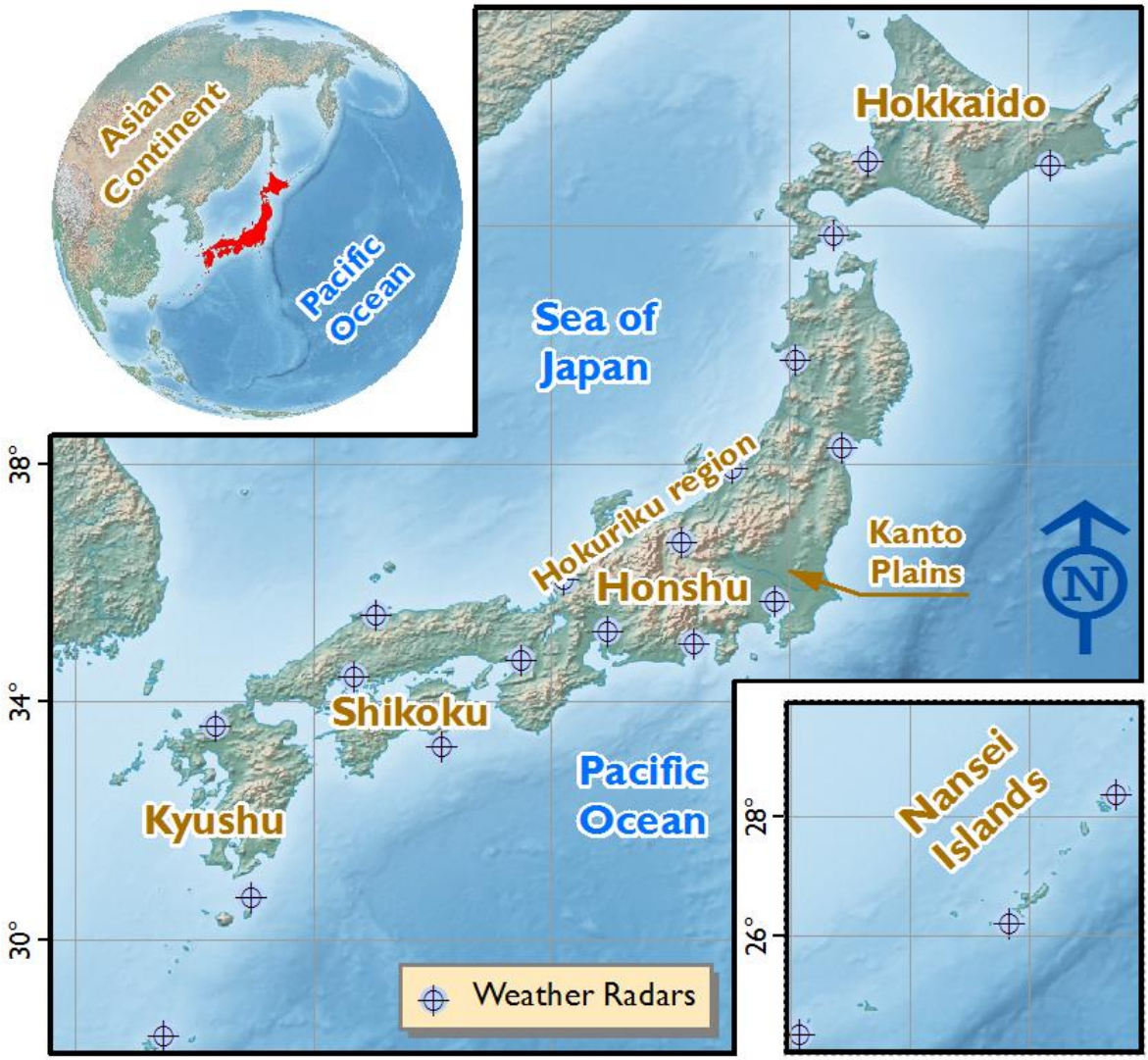


Figure 3.5 Geographic location of the main islands of Japan and the weather radars of the JMA

Additionally, Japan is regularly struck by tropical cyclones, which originate from large masses of low-pressure air in the North Pacific Ocean. There is no specific season for the occurrence of tropical cyclones since the trajectory and intensity are quite different from each other. However, when a tropical cyclone approaches Japan, it often heads towards the Nansei islands, Kyushu or Shikoku. The classification, which is not the same for all regions of the world, depends on the intensity defined by the wind velocity in the storm. The Japan Meteorological Agency (JMA) classifies the most severe tropical cyclone in the North Western Pacific region as a typhoon. Classifications of less intense tropical cyclones include different degrees of tropical depression and tropical storm. Besides the strong winds, tropical cyclones often cause heavy rainfall. The interaction of an existing midlatitude front and a tropical cyclone may cause the transition of the storm into an extratropical front (Klein et al.,

2000). Kamahori (2012) reviewed previous studies about the relation between rainfall and tropical cyclones, showing that these storms cause between 5 and 7 percent of the seasonal rainfall in the northwest Pacific region. Moreover, the same research concluded that rainfall rate increases with storm intensity and that it precipitates concentrically to the center within a 5-degree radius. Figure 3.6 portrays schematic trajectories and location of these three precipitation systems.

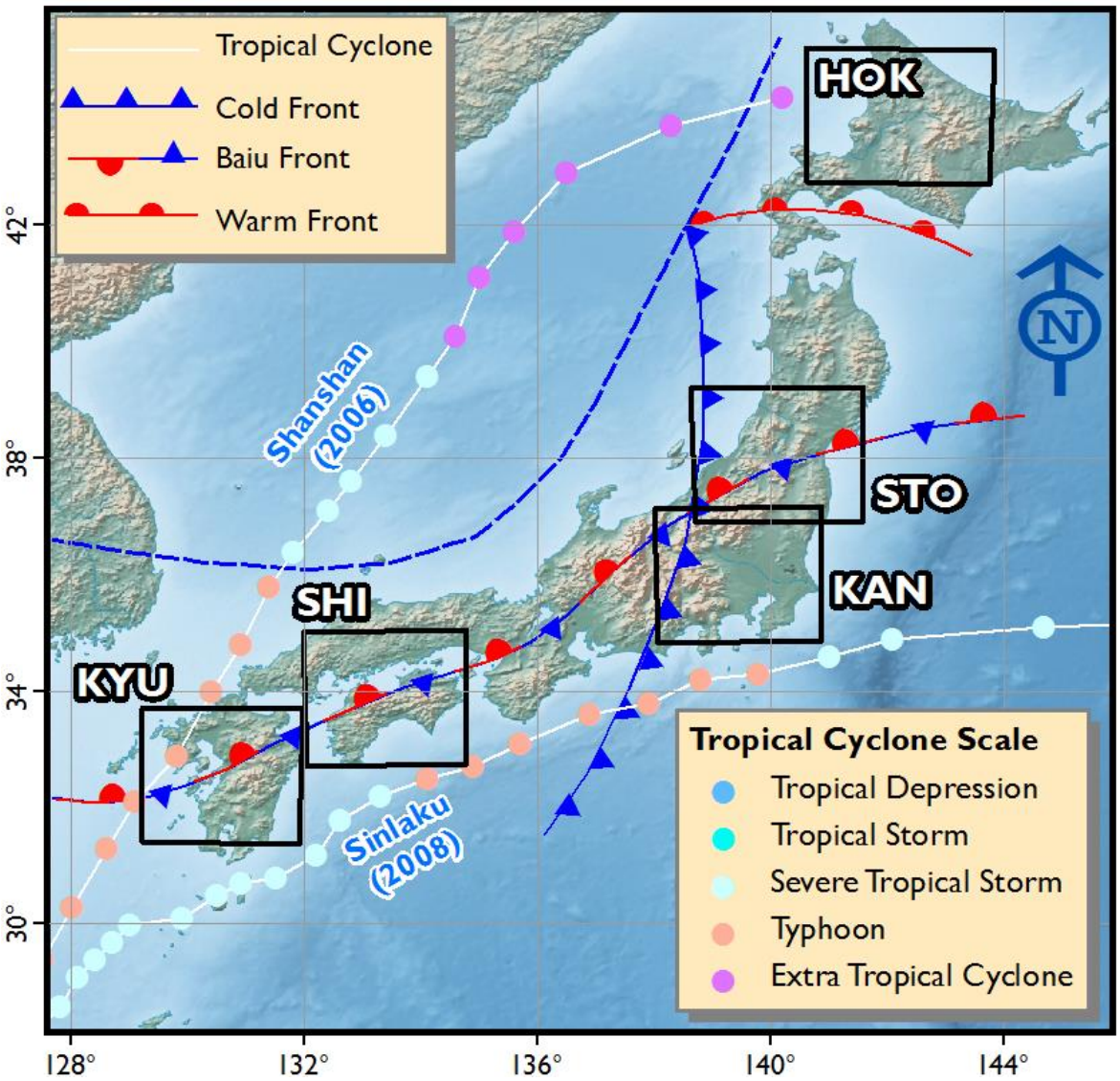


Figure 3.6 Examples of precipitation systems over Japan and the surrounding waters.

The Baiu front is considered quasi-stationary, however, it moves slowly northward as the warm air of the Pacific Ocean becomes predominant, marking the beginning of the hot summer season. The blue dashed line is an example of the trajectory that follow the extratropical cyclones generating cold fronts. Two tropical cyclones, which stroke Japan causing damage and torrential rains, are shown in this figure: typhoon Shanshan in September 2006, and typhoon Sinlaku in September 2008.

The structure of precipitation systems is often a combination of stratiform and convective components (Houze, 1993). Stratiform precipitation is generated by stably stratified clouds,

e.g., nimbostratus. The formation of this type of clouds is usually a consequence of the slow ascent of unsaturated air ahead of warm fronts. Stratiform precipitation tends to be widespread, to have a steady intensity, and to last for several hours. On the other hand, convective clouds (e.g., cumulus congestus and cumulonimbus) are formed by atmospheric instability, which leads to the strong updraft of warm moist air through either thermal convection or frontal lifting. High intensity, short duration and limited extent are typical features of convective rainfall. Thus, the different types of rainfall events can be classified as stratiform, localized convective cells generated by strong local updrafts (this classification encompasses single-cell storms, multicellular storms, supercells and squall lines), and convective cells embedded in large convective or stratiform rain zones. Having accurate, high-resolution observations of intensity for these different types of events is undoubtedly essential for the assessment of the link between multiscale statistical features and the degree of variability.

With the purpose of encompassing the different types of rainfall, five analysis sites in Japan were identified for the multiscale analysis of fluctuations conducted in this study: Kanto (KAN), South Tohoku (STO), Hokkaido (HOK), Kyushu (KYU), and Shikoku (SHI). Each site, shown in Figure 3.6, covers a squared area with a 256-km side.

The JMA is the institution responsible for monitoring several weather-observation networks across the Japanese Islands. The weather information is collected via geostationary satellites (Himawari series), ground-based radars, radiosonde, wind profilers, and a network of 1300 surface-observation stations commonly known as AMeDAS (Automated Meteorological Data Acquisition System). The stations, which form one of the densest networks in the world, are laid out at average intervals of approximately 17 km throughout Japan. The surface data measured and monitored by AMeDAS encompasses rainfall intensity, snow depth, wind, surface pressure, temperature, humidity, visibility and sunshine duration. The JMA also produces and distributes the so-called Radar-AMeDAS datasets, covering the Japanese islands and surrounding ocean waters. These datasets are a composite of rainfall/snowfall intensity estimated using the soundings of 20 C-Band Doppler weather radars and calibrated with the measurements of the AMeDAS network. The location of the radars is shown in Figure 3.5. There have been improvements in the spatial and temporal resolution of Radar-AMeDAS since its release in 1988. The spatial resolution changed from 5 km to 2.5 km in 2001, and then to 1 km in 2006; the temporal resolution changed from hourly to every thirty minutes in 2003. These two-dimensional, gridded datasets comprise values of rainfall rate in millimeters per hour (mm/hr) rounded to positive integers. Additionally, the real value 0.4 is used for drizzle and zero is used to represent no-rain. The value of 0.4 was set as threshold because rainfall with lower intensities and echoes from non-weather targets are not easily distinguished in the radar soundings.

3.4.2 Criteria for selection of the events

In order to clarify the use of certain words in the subsequent sections, let us make the following definitions:

- "grid": two-dimensional, equally-spaced arrange of elements.
- "cell": a single element of a grid.
- "rainfall event" or simply "event": a grid of rainfall intensity data accumulated over a specified period of time.
- "rain area": for a single event, the area of a single cell in km^2 multiplied by the number of cells in which the comprised value of rainfall intensity is greater than zero.
- "no-rain area": for a single event, the area of a single cell in km^2 multiplied by the number of cells in which the comprised value of rainfall intensity is equal to zero.

The analysis was done to events that happened between February 2006 and December 2009. Since the area of the analysis sites is 256×256 km2 and the Radar-AMeDAS data has a 1-km resolution in the period of analysis, the rainfall intensity data for each event is stored as a grid of 256×256 cells. Even though the data is available every thirty minutes between 2006 and 2009, the events considered in this study are hourly accumulations. The accumulation period was set in order to match the hourly measurements of the rain gauges of the AMeDAS network. Hereafter, if an event is referred to as KAN_2006.05.20.15, for example, it means that the data corresponds to the intensity in mm/hr of the rainfall precipitated over the Kanto site on May 20th 2006 between 14:01 and 15:00 hours.

The samples of events considered for analysis were selected establishing two conditions:

- The rain area mean intensity has to be greater than or equal to 1 mm/hr.
- The rain area has to be greater than or equal to 25% of the total surface of the analysis site. In other words, the number of cells that belong to the rain area has to be greater than or equal to $0.25 \times 256 \times 256$.

The first condition excludes those events in which drizzle is dominant. The analysis of spatial fluctuations is practically futile in the case of drizzle because the grid cells comprise mostly values of 0.4, showing no apparent variability. The second condition excludes small rain areas which may not depict scale-to-scale dependence across all the considered scales. A more comprehensive explanation of the reason for excluding events with small rain areas is given later. Table 3.4 shows the sample size or number of events that satisfy these two conditions in each analysis site.

Table 3.4 Characteristics of the analysis sites and sample size (number of events)

Region	Code	UTM coordinates of the Southwest corner		UTM Zone	Nº of events
Kanto	KAN	232500E	3858000N	54N	3210
South Tohoku	STO	295000E	4085500N	54N	3824
Hokkaido	HOK	468500E	4731000N	54N	1900
Kyushu	KYU	520000E	3476000N	52N	3895
Shikoku	SHI	223000E	3623000N	53N	3536

3.4.3 Conditional sampling based on rain or no-rain areas

Figure 3.7 shows Radar-AMeDAS data for the event KAN_2008.09.22.00 and its approximation at scale $m = 4$, $\bar{X}_4$, corresponding to a 16-km resolution. Comparing the no-rain areas in both representations, it can be appreciated that the rain-area becomes larger as the resolution becomes lower, which is partly the definition of intermittency of rainfall. Regarding this observation, an additional consideration for the selection of the sample is established.

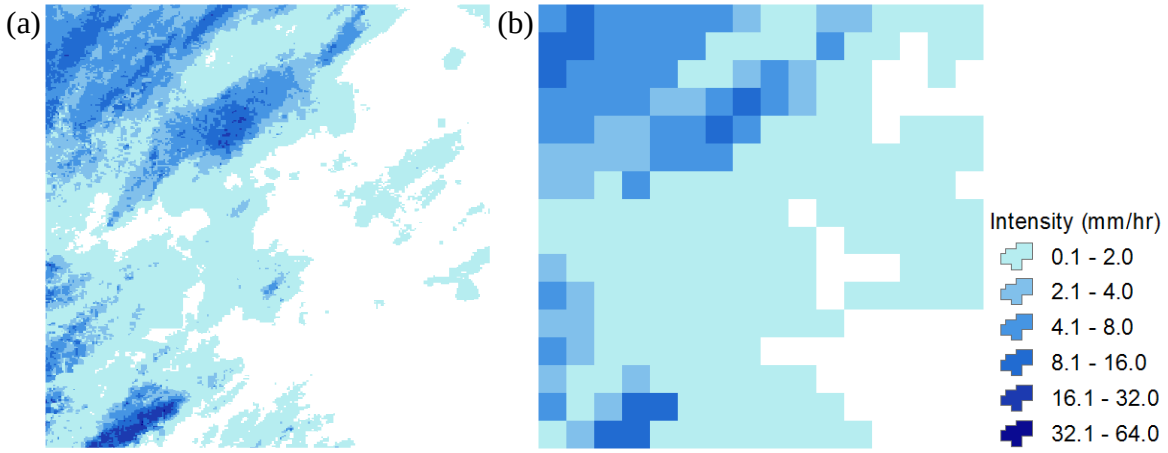


Figure 3.7 Radar-AMeDAS data for the event KAN_2008.09.22.00

(a) Rainfall intensity at scale $m = 0$ correspondent to 1-km resolution, $\bar{X}_0$, and (b) approximation at scale $m = 4$ correspondent to 16-km resolution, $\bar{X}_4$.

In Eq. 2.16, each value contained in the set $\bar{X}_m(n'_1, n'_2)$ is a function of the four values $\bar{X}_{m-1}(n_1, n_2)$, $\bar{X}_{m-1}(n_1, n_2 + 1)$, $\bar{X}_{m-1}(n_1 + 1, n_2)$ and $\bar{X}_{m-1}(n_1 + 1, n_2 + 1)$, which form a 2×2 sub-grid of cells that belong to the approximation of the rainfall field at scale $m - 1$ (See Figure 3.8). If these four values are equal to zero, the resulting value of $\bar{X}_m(n'_1, n'_2)$ is also equal to zero. In the same manner, by analyzing Eq. 2.17 one can arrive to the same conclusion, i.e., zero-values will be generated in the sets of wavelet coefficients $X'_{m,i}$ in the locations correspondent to no-rain areas. With the purpose of avoiding the inclusion of these non-relevant zero-values, the following condition is established: "the values generated in Eq. 2.17 are considered part of the sample of wavelet coefficients only if at least one of the four values $\bar{X}_{m-1}(n_1, n_2)$, $\bar{X}_{m-1}(n_1, n_2 + 1)$, $\bar{X}_{m-1}(n_1 + 1, n_2)$ and

$\bar{X}_{m-1}(n_1+1, n_2+1)$ is greater than zero. The following examples are given according to the schematic in Figure 3.8:

- Because $\bar{X}_{m-1}(n_1, n_2+1)$ is greater than zero, the correspondent value of $X'_{m,i}(n'_1, n'_2)$ is included in the sample.
- Because $\bar{X}_{m-1}(n_1-2, n_2-1)$, $\bar{X}_{m-1}(n_1-2, n_2-2)$ and $\bar{X}_{m-1}(n_1-1, n_2-2)$ are greater than zero, the correspondent value of $X'_{m,i}(n'_1-1, n'_2-1)$ is included in the sample.
- Because $\bar{X}_{m-1}(n_1, n_2-2)$, $\bar{X}_{m-1}(n_1, n_2-1)$, $\bar{X}_{m-1}(n_1+1, n_2-2)$ and $\bar{X}_{m-1}(n_1+1, n_2-1)$ are all equal to zero, the correspondent value of $X'_{m,i}(n'_1, n'_2-1)$ is not included in the sample.

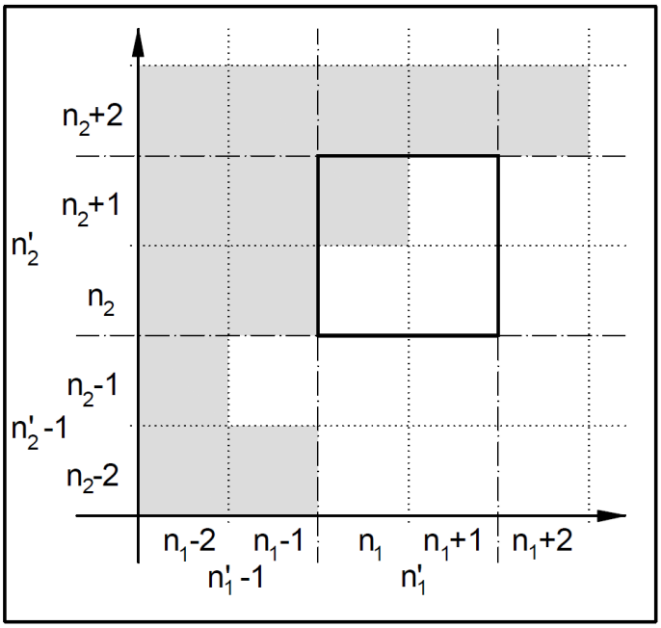


Figure 3.8 Schematic of the discrete approximation of the rainfall field.

The cells separated by dash-dot lines correspond to the approximation at scale m with location indexes (n'_1, n'_2). The cells separated by dotted lines correspond to the approximation at scale $m-1$ with location indexes (n_1, n_2). The cells with grey background comprise values of intensity greater than 0 while the cells with white background comprise values of intensity equal to zero.

3.4.4 The effect of zero-values on the estimation methods of stable parameters

In spite of the rule established in the previous section for filtering non-relevant zeros generated in no-rain areas, the samples of wavelet coefficients are still prone to have a large number of zero-values at high resolutions. Because Radar-AMeDAS datasets only comprise integer values, at high resolutions it is likely to find large areas (a large number of adjacent cells) with the same value of intensity. Hence, the sets of wavelet coefficients, which are meant to capture local variations of intensity, may contain an unusual number of zero-values.

Most of the parameter-estimation methods presented in Section 3.2.5 have a poor performance when the samples have a large number of zeros. A brief explanation of the effect of this situation is given for each method:

- **QRM method**: In Eq. 3.12, if $x_{0.75}$ and $x_{0.25}$ are equal to zero due to a large number of zero-values in the sample, v_α becomes undefined. For this reason, neither $\hat{\alpha}$ nor $\hat{\gamma}$ can be determined.
- **MQM method**: Likewise, in Eq. 3.14, if $|x|_{0.50}$ is equal to zero due to the large number of zero-values, v_α becomes undefined. For this reason, neither $\hat{\alpha}$ nor $\hat{\gamma}$ can be determined.
- **RCF method**: The large number of zero values produces an ill-conditioned linear system of equations in the regression of the cfs (Eq. 3.10 and Eq. 3.11), which is commonly known as a rank-deficient problem. This implies that one or more rows in the coefficient matrix are nearly linear combinations of some or the remaining rows. Consequently, the system has more than one possible solution. As the regression is done iteratively until some level of convergence is reached, the different solutions that are obtained in each iteration preclude the estimators to converge. Therefore, in some cases this method does not return reasonable values.
- **FAM method**: A large number of zero values causes the empirical pth fractional absolute moment, m_p, to be close to zero. Then, the evaluation of ${m_p}^{1/p}$ at the empirical cdf yields a value approaching 1.0. This situation, combined with the boundary conditions at which the oscillatory integral of Eq. 3.7 is evaluated, causes the numerical interpolation of the estimator $\hat{\alpha}$ to diverge with each iteration. Consequently, in some cases this method does not return reasonable values.
- **TSM method**: The QRM and MQM methods compute relative measures of the distribution as a function of fixed quantiles. For example, in Eq. 3.12, computing $x_{0.75}$ gives the value of the variable that is greater than or equal to 95% of the elements in the distribution. However, this value is not able to inform about how extreme the remaining 5% of the distribution is. Differently from those methods, the TSM method uses the mean of the absolute values of the truncated distribution, therefore, not only representing a relative measure of the distribution, but also having a weighted impact of all the values of the distribution. This feature allowed to estimate values of $\hat{\alpha}$ and $\hat{\gamma}$ when the sample has a large number of zero-values. Nevertheless, the estimator $\hat{\alpha}$ in some cases is considerably smaller than the estimators at lower resolutions.
- **LMO method**: Since the logarithm of zero is undefined, this method cannot be utilized for samples comprising zero-values.

From the six methods, only the TSM method is able to return estimates of the stable parameters of samples with a large amount of zero-values, with no exceptions. For a good

portion of the selected rainfall events, the issues of the five other methods listed above are evident when trying to estimate the stable parameters of the sets of wavelet coefficients $X'_{1,i}$, i.e., the variations of intensity at the highest resolution. Depending on the method and on the event, the inability to estimate the stable parameters of the sets of wavelet coefficients $X'_{2,i}$, and even $X'_{3,i}$, was also observed.

3.4.5 Procedure and methods for the computation of results

In Section 3.2.1, it was mentioned that the lack of closed-form expressions of the pdf and cdf hinders not only the estimation of the parameters, but also resolving which of the existing estimation methods performs better than the others. Even though there are tests based on qualitative comparisons (e.g., q-q plots and p-p plots) and some new procedures aimed to assess goodness of fit of α-stable distributions (e.g., Brcich et al., 2005 and Matsui and Takemura, 2008), they usually rely on the existence of large samples and do not contemplate samples with an unusual number of zeros. Therefore, choosing the most suited parameter-estimation method should consider the characteristics of the samples and the purpose of such characterization.

As an initial step, the characteristics of the samples that signify setbacks in the estimation of their stable-parameters are clearly identified:

4. At high resolutions, in most events, the samples of wavelet coefficients comprise a large number of zero-values. Apart from the TSM method, in the previous section it was determined that a large amount of zero-values hinders the estimation of the parameters.
5. At low resolutions, the estimation of the parameters of the distribution might be precluded by the limited information that fails to represent the "shape" of the distribution.

Since the main objective of this Chapter is to show the presence of self-similarity in the distributions of rainfall local fluctuations, the methods of estimation are going to be evaluated by how well they overcome the two setbacks enumerated above and how well they contribute to the achievement of said objective.

Considering the conditional sampling explained in Section 3.4.3 that rules out zero-values coming from no-rain areas, wavelet coefficients from rain areas can be sampled applying on of three variations of the discrete wavelet transform: (1) the Discrete Wavelet Transform (DWT), (2) The Undecimated Discrete Wavelet Transform (UDWT), and (3) a combination of both. Each variation generates samples with different sizes. The details are listed below.

1. **DWT**: The multiscale sampling of wavelet coefficients is carried out according to Eq. 2.16 and Eq. 2.17. Let $L_1 \times L_2$ be the number of samples in the original

datasets. Then, the sample size of the sets $\overline{\boldsymbol{X}}_m$ and $X'_{m,i}$ are reduced to $(L_1 \times L_2)/4^m$.

2. **UDWT**: The multiscale sampling of wavelet coefficients is carried out according to Eq. 2.18 and Eq. 2.19. Since periodic boundary conditions are imposed (Nason and Silverman, 1995), the number of samples in the sets $\overline{\boldsymbol{X}}_m$ and $X'_{m,i}$ should be equal to $L_1 \times L_2$ at all scales. This kind of imposition is useful when analyzing continuous signals or digital images with repeating patterns. However, periodic boundary conditions should not be applied in rainfall datasets. The probability of the existence of no-rain areas and the different patterns that can be found at opposite ends of the data grids might generate additional information that should not be considered in the characterization of wavelet coefficients. Thus, the UDWT sampling in this study considers finite boundaries, reducing sample size of the sets $\overline{\boldsymbol{X}}_m$ and $X'_{m,i}$ to $(L_1 - 2^m + 1) \times (L_2 - 2^m + 1)$.

3. **A combination of both**: Compared to the DWT method, the UDWT comes at the cost of a larger number of arithmetic operations, which implies a larger storage of digital data and a longer time of computation. For this reason, it was considered to sample the wavelet coefficients by obtaining the discrete approximation of the rainfall field at all possible scales of representation using the DWT method (Eq. 2.16), and then, obtaining the wavelet coefficients at each scale using the UDWT method (Eq. 2.19). The sample size of the sets $X'_{m,i}$ are reduced to $(2^{1-m}L_1 - 1) \times (2^{1-m}L_2 - 1)$.

Table 3.5 Maximum sample sizes of the directional sets of wavelet coefficients at all possible scales of representation

		Sample size		
Scale m	**Resolution**	**DWT**	**UDWT**	**Combination**
0	1-km	--	--	--
1	2-km	16384	65025	65025
2	4-km	4096	64009	16129
3	8-km	1024	62001	3969
4	16-km	256	58081	961
5	32-km	64	50625	225
6	64-km	16	37249	49
7	128-km	4	16641	9
8	256-km	1	1	1

Table 3.5 shows the maximum sample size that can be obtained at each scale with the different sampling methods. Recall that the sample sizes can be reduced even more if the event has no-rain areas. Note that because the dimensions of the original data at each site is $L_1 = 256 \times L_2 = 256$, the lowest resolution of representation is at scale $m = 8$. If periodic boundary conditions are considered in the UDWT sampling method, the variations of information in the samples cause adverse, significant differences in the estimation of the

stable parameters, especially when the no-rain areas are large. Curiously, even though the samples generated by the UDWT and by the combined method are quite different in size, the estimation of the parameters does not show significant variations, even at scales $m = 6$ and $m = 7$. Consequently, the results presented in this study are obtained using the **combined method** of sampling.

At scale $m = 1$ the samples comprise a large number of zero-values, hampering the performance of the parameter-estimation methods. Therefore, the scale range in which all methods can be compared is $m_{\min} = 2 \leq m \leq m_{\max} = 7$. For a set of directional wavelet coefficients $X'_{m,i}$, let $\alpha_{m,i}$ and $\gamma_{m,i}$ be respectively the characteristic exponent and the scale parameter of the symmetric α-stable distribution at scale m. Recalling the definition of self-similarity in distribution given in Section 3.3, if the relations

$$\alpha_{m_{\min},i} = \alpha_{m_{\min+1},i} = \cdots = \alpha_{m_{\max},i}, \text{ and} \tag{3.22}$$

$$\gamma_{m,i} = \gamma_{m_{\min},i} 4^{(m-1)H_i} \tag{3.23}$$

hold independently and simultaneously in the 3 directions ($i = \{1,2,3\}$) , the multidimensional process of local fluctuations of that rainfall event is said to be self-similar in distribution in the scale range $m_{\min} \leq m \leq m_{\max}$.

Because some bias is expected in the estimation of the parameters, apart from the LMO method, which cannot be applied at any scale, the rest of the methods are going to be evaluated by the degree of variation between scales of the estimated characteristic exponent of the α-stable distributions. Thus, when samples of rainfall fluctuations are considered, the method that is most suitable for the estimation of stable-parameters is going to be the one that can hold expression 3.22 with the least variation between scales. For that purpose, let us define

$$\sigma_i = \left[(m_{\max} - m_{\min} + 1)^{-1} \sum_{m_{\min}}^{m_{\max}} \left(\hat{\alpha}_{m,i} - \bar{\alpha}_i\right)^2\right]^{1/2} \text{ and}$$

$$\bar{\alpha}_i = (m_{\max} - m_{\min} + 1)^{-1} \sum_{m_{\min}}^{m_{\max}} \hat{\alpha}_{m,\mathrm{i}}, \tag{3.24}$$

where $\bar{\alpha}_i$ is the arithmetic mean of the estimators $\hat{\alpha}_{m,i}$, and σ_i is a measure of the invariance of the estimators $\hat{\alpha}_{m,i}$ between scales. As an example, Figure 3.9 shows a comparison of the results for the rainfall event KAN_2009.11.11.16. The correspondent values of σ_i are presented in Table 3.6., which denote that the TSM method shows an overall better performance.

Table 3.6 Values of σ_i for the event KAN_2009.11.11.16 in the scale range $2 \leq m \leq 6$

i	RCF	QRM	TSM	FAM	MQM
1	0.178	0.050	0.013	0.059	0.060
2	0.209	0.035	0.059	0.084	0.084
3	0.121	0.075	0.030	0.095	0.032

Table 3.7 shows the mean values of σ_i correspondent to all the analyzed rainfall events in the Kanto analysis site. At scale $m = 7$, which has a sample size equals to 9, the estimators of $\hat{\alpha}_{m,i}$ approximate 2.0, which corresponds to the Gaussian distribution. Although with less frequency, the estimators at scale $m = 6$ have similar outcome. Such results can be an indication that the rainfall fluctuations of the event can be qualified as self-similar only in the scale ranges $2 \leq m \leq 5$ or $2 \leq m \leq 6$. However, if the analyzed event is part of a storm developing at a larger atmospheric scale, the results depicted in Figure 3.9 make evident that the information of extreme variations that can be obtained at scales $m = 6$ or $m = 7$ is quite limited for an analysis site of 256 km $\times$ 256 km. In which case, it would be difficult to rule out the possibility that self-similarity holds in a wider scale range. The TSM method was found to yield the smallest mean values of σ_i, followed by the MQM method. The RCF method and the FAM method, regarded as the most accurate methods according to the Monte Carlo simulation in Section 3.2.6, yielded the worst results. The reason for the failure of these two methods is the deficient estimation at high-resolutions, in which the amount of zero-values is not large enough to prevent the methods from working, but enough to produce unexpected results.

Table 3.7 Mean values of σ_i for the Kanto analysis site

i	RCF	QRM	TSM	FAM	MQM
	$2 \leq m \leq 5$				
1	0.081	0.050	0.032	0.071	0.040
2	0.083	0.050	0.032	0.071	0.042
3	0.078	0.054	0.033	0.090	0.043
	$2 \leq m \leq 6$				
1	0.097	0.071	0.052	0.077	0.058
2	0.098	0.072	0.053	0.076	0.060
3	0.090	0.073	0.050	0.089	0.057

Based on the previous analysis of the inter-scale variation of the estimators of the characteristic exponent, the TSM method proves to be the method that is most suitable for the estimation of stable-parameters when samples of rainfall fluctuations are considered. Moreover, note that the TSM method is the only method that is able to return estimates of the stable-parameters at scale $m = 1$. Therefore, the results presented in this study (next section) are computed with the TSM method in the scale range $1 \leq m \leq 7$.

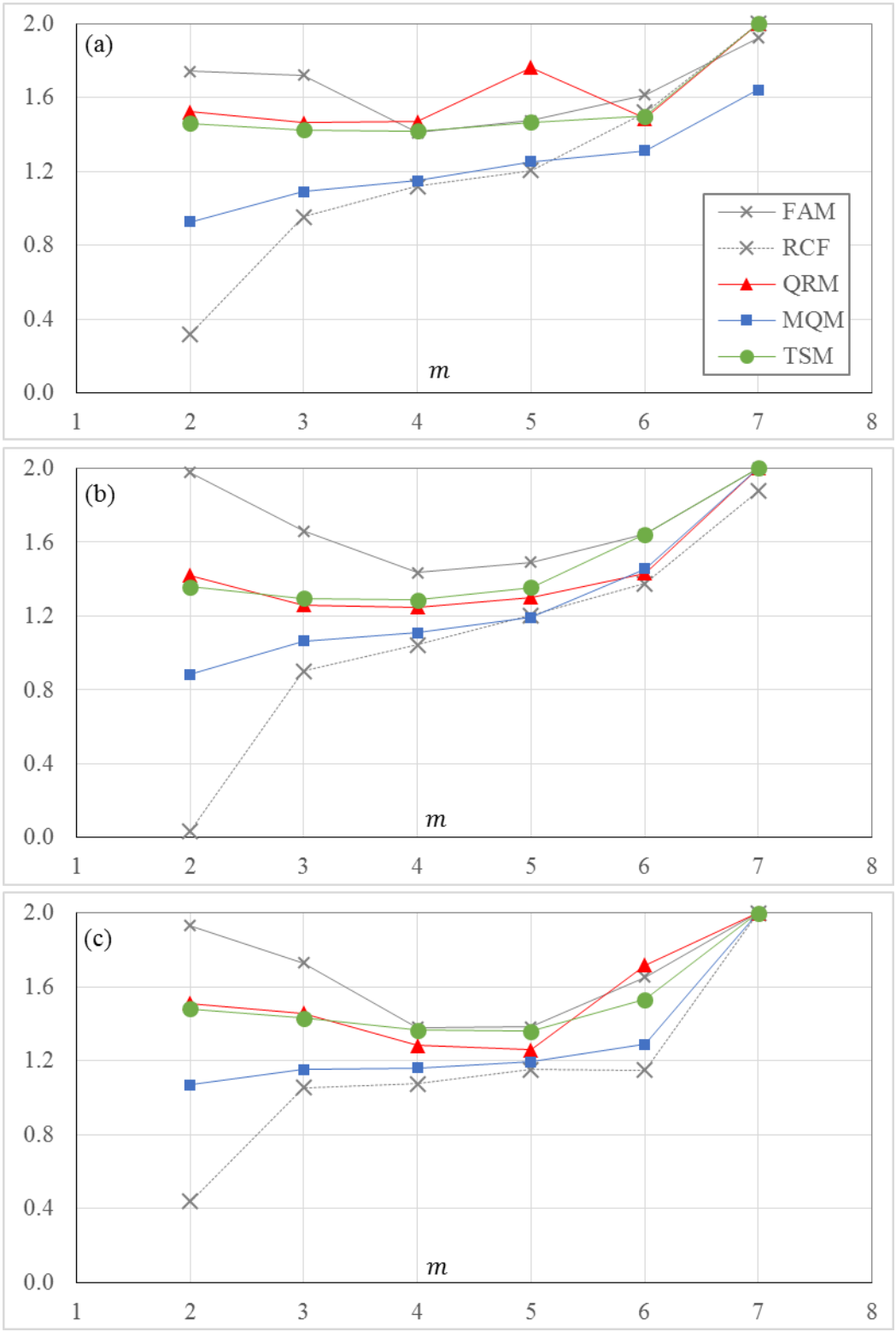


Figure 3.9 Estimators $\hat{\alpha}_{m,i}$ for the event KAN_2009.11.11.16 with different estimation-methods

The results are divided in three graphs representing the estimators of the characteristic exponent of the α-stable distributions of the wavelet coefficients (a) in the horizontal direction ($i = 1$), (b) in the vertical direction ($i = 2$), and (c) in the diagonal direction ($i = 3$).

With regard to the scale parameter, applying base-4-logarithm to both sides of Eq. 3.23 yields

$$\log_4(\gamma_{m,i}) = \log_4(\gamma_{m_{\min},i}) + (m-1)H_i.$$

This linear expression implies that $\gamma_{i,m_{\min}}$ and H_i can be determined by means of regression. Let ρ_i^2 be the coefficient of determination of the regression. The results are accepted only if $\rho^2{}_i$ is greater than 0.75 in all three directions. If this condition is not fulfilled, the regression is applied to the estimators in a narrower scale range, i.e., $m_{\min} \leq m \leq m_{\max} - 1$.

As an example, Figure 3.10 shows the logarithmic regression of the estimated scale parameters for the event KAN_2009.11.11.16.

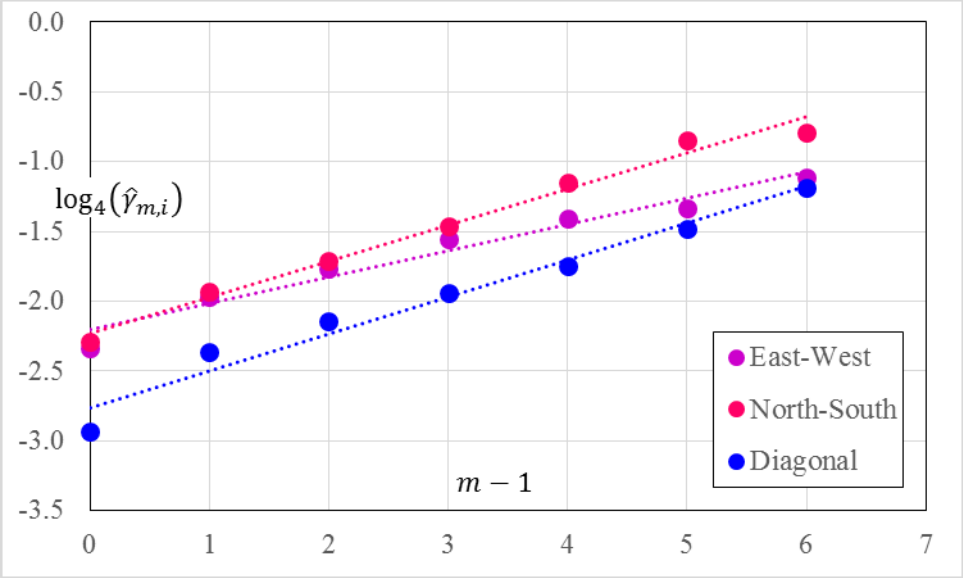


Figure 3.10 logarithmic regression of the estimated scale parameters for the event KAN_2009.11.11.16

The colored circles represent the values of $\log_4(\hat{\gamma}_{m,i})$ correspondent to each $m-1$ in the three directions. The self-similarity index H_i is the slope of the dotted lines, which represent the best fit of the logarithmic regression. Considering that the sets of wavelet coefficients $X'_{1,i}$ represent the variations of intensity at scale $m = 0$, the point in which the dotted lines cross the vertical axis correspond to $\gamma_{1,i}$. The coefficients of determination are $\rho^2{}_{i=1,2,3} = \{0.98, 0.99, 0.98\}$.

3.4.6 Presentation of the Results

The procedure detailed in the previous section was applied for computing the estimators $\alpha_{m,i}$ and $\gamma_{m,i}$ of the probability distribution of rainfall fluctuations (wavelet coefficients) contained in the sets $X'_{m,i}$. The events selected for analysis in each site was previously detailed in Table 3.4.

3.4.6.1 Estimation of the characteristic exponent

The TSM method proved to overcome the setbacks produced by the nature of the samples better than the other parameter-estimation methods. Even though this method is the only that, with no exceptions, returns estimators $\hat{\alpha}_{1,i}$, it does not mean that the large amount of zeros does not impact the estimated values. In some cases, like the one shown in Figure 3.9, the values of $\hat{\alpha}_{1,i}$ are similar to the ones obtained at lower resolutions. However, in other cases the values of $\hat{\alpha}_{1,i}$ are much smaller than the ones at lower resolutions, which may affect the computation of the mean $\bar{\alpha}_i$ (Figure 3.11). The events in which this adversity is observed happen to have a large number of contiguous cells comprising the same value of low intensity and vast no-rain areas. Figure 3.12 shows the Radar-AMeDAS data correspondent to the event KAN_2009.02.25.09 in which no-rain areas occupy 27% of the total surface, cells comprising 0.4 mm/hr occupy 42% of the total surface, and cells comprising 1 mm/hr occupy 14% of the total surface. In order to avoid this bias, the computation of the mean $\bar{\alpha}_i$ considers only the estimators $\hat{\alpha}_{m,i}$ in the scale range $2 \leq m \leq 5$ for all events.

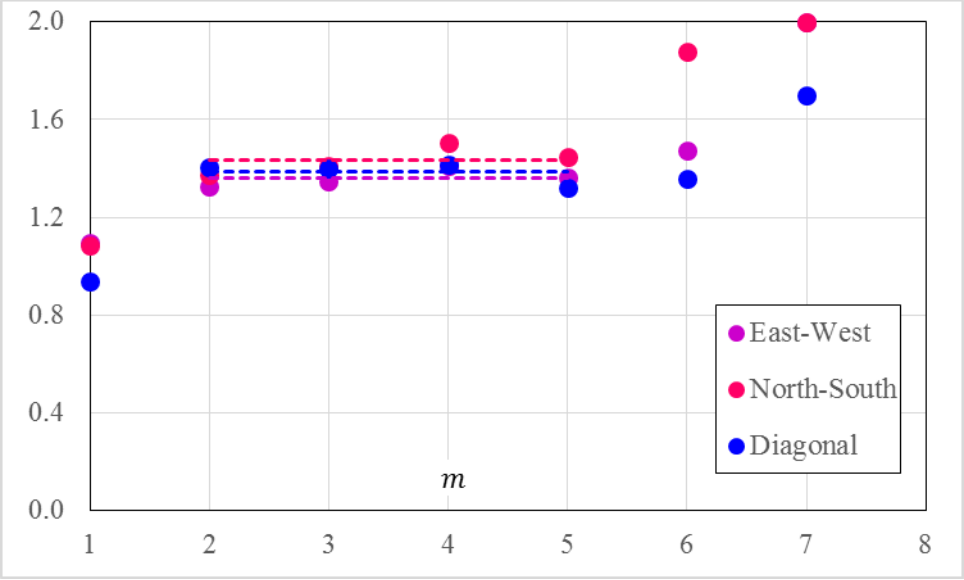


Figure 3.11 Estimators $\hat{\alpha}_{m,i}$ for $1 \leq m \leq 7$ and $i = \{1, 2, 3\}$ corresponding to the event KAN_2009.02.25.09

The dashed lines represent the values of $\bar{\alpha}_i$ in the scale range $2 \leq m \leq 5$. Note that the values corresponding to $\hat{\alpha}_{1,i}$ are noticeably smaller than at lower resolutions. If these values are included in the computation of $\bar{\alpha}_i$, it creates a biased mean value.

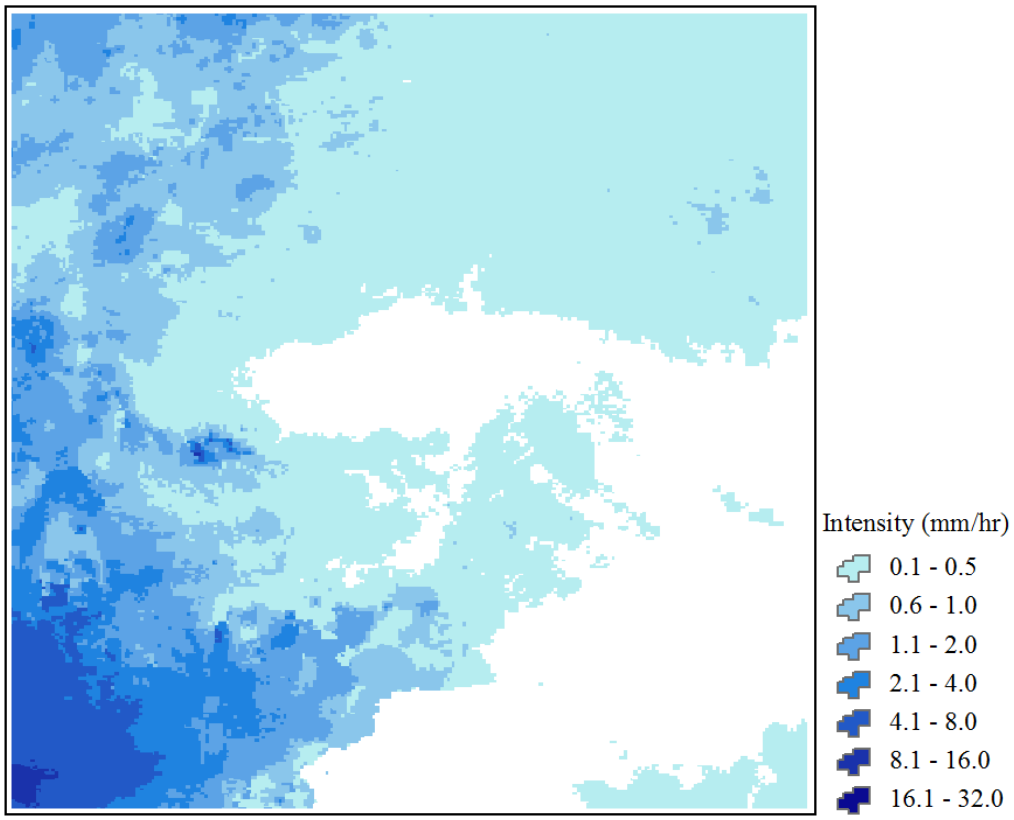


Figure 3.12 Radar-AMeDAS data for the event KAN_2009.02.25.09

The image depicts vast no-rain areas and a large number of contiguous cells with the same value of intensity.

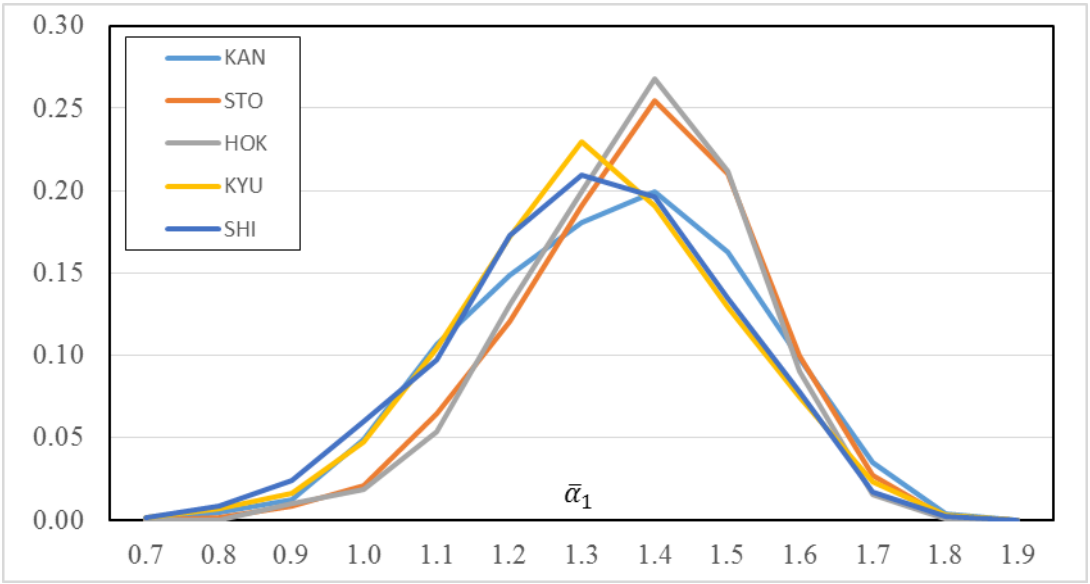


Figure 3.13 Normalized frequencies of $\bar{\alpha}_1$ (horizontal direction)

The values encompass all selected events in each of the analysis sites.

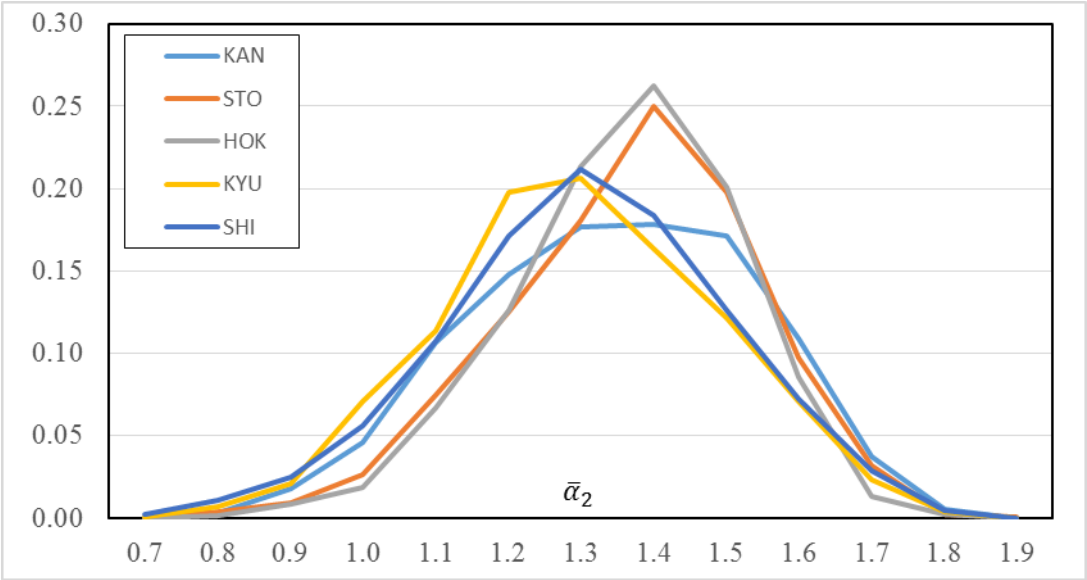


Figure 3.14 Normalized frequencies of $\bar{\alpha}_2$ (vertical direction).

The values encompass all selected events in each of the analysis sites.

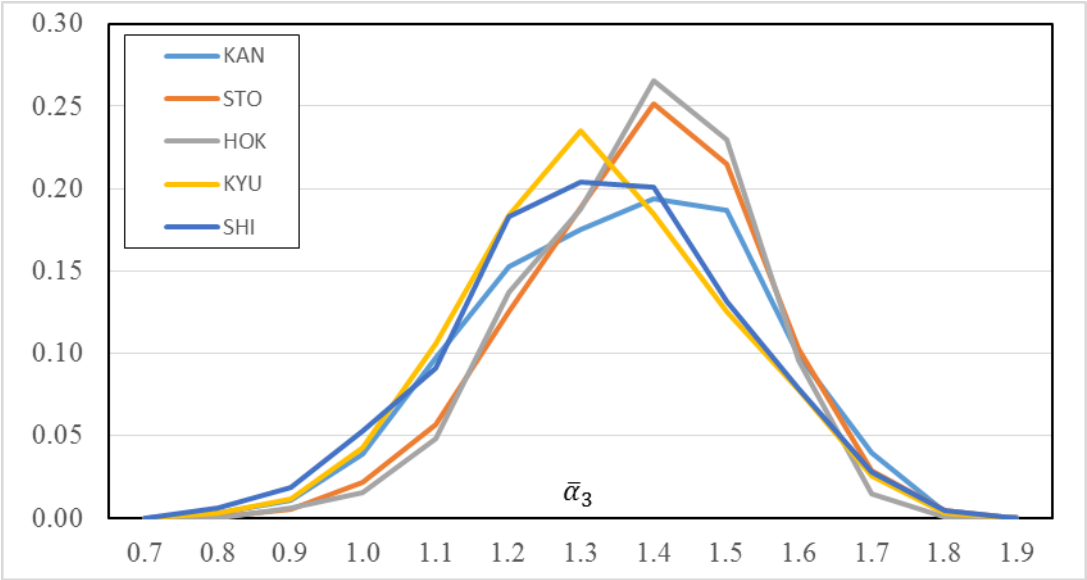


Figure 3.15 Normalized frequencies of $\bar{\alpha}_3$ (diagonal direction).

The values encompass all selected events in each of the analysis sites.

Figures 3.13, 3.14 and 3.15 show the normalized frequencies of $\bar{\alpha}_i$ encompassing the results obtained for all the selected events. Recall that the number of events in each analysis site is 3210 (KANto), 3824 (South TOhoku), 1900 (HOKkaido), 3895 (KYUshu), and 3536 (SHIkoku). These Figures are intended to show the parameter-range of $\bar{\alpha}_i$. The following notes can be extracted from the analysis of the frequencies:

- The highest frequencies of the South Tohoku and Hokkaido sites, which correspond to events with $\bar{\alpha}_i$ approximately equal to 1.4, are just above 0.25. When compared to the other three sites, the curves are leptokurtic depicting lower frequencies of events with $0.9 \leq \bar{\alpha}_i \leq 1.2$. The possible explanation for these characteristics can be the

fact that in the South Tohoku and Hokkaido sites the occurrence of rainfall with steady intensity and gradual changes is more frequent than rainfall with sudden variations. These regions are mostly hit by cold fronts and warm fronts.

- The frequencies of the Kanto site show a pick that coincides with the peak of the South Tohoku and Hokkaido sites. However, the curves show a symmetric distribution of frequencies in the parameter-range. Even though the Kanto region is frequently struck by cold fronts, there are occasional tropical cyclones causing heavy rainfall just before their dissipation stage and local convective activity over the planes and by the Pacific side of the mountain ranges.
- The highest frequencies of the Kyushu and Shikoku sites, which correspond to events with $\bar{\alpha}_i$ approximately equal to 1.3, are just above 0.20. When compared to the other three sites, the curves depict higher frequencies of events with $0.9 \leq \bar{\alpha}_i \leq 1.3$. The possible explanation for these characteristics can be the fact that in the Kyushu and Shikoku sites the occurrence of rainfall with sudden variations is more frequent than rainfall with steady intensity. These regions are often hit by strong tropical cyclones (e.g., typhoons).

Figures 3.16 to 3.21 are events with similar values of $\bar{\alpha}_i$ in the three directions. These events are shown with the purpose of visually appreciating the degree of spatial variation of intensity for different values of $\bar{\alpha}_i$ within its paremeter-range. As it was expected, the figures showing events with values of $\bar{\alpha}_i$ approaching 1.0 depict intermittency and changes of intensity within short-distances with more frequency than the events with larger values of $\bar{\alpha}_i$.

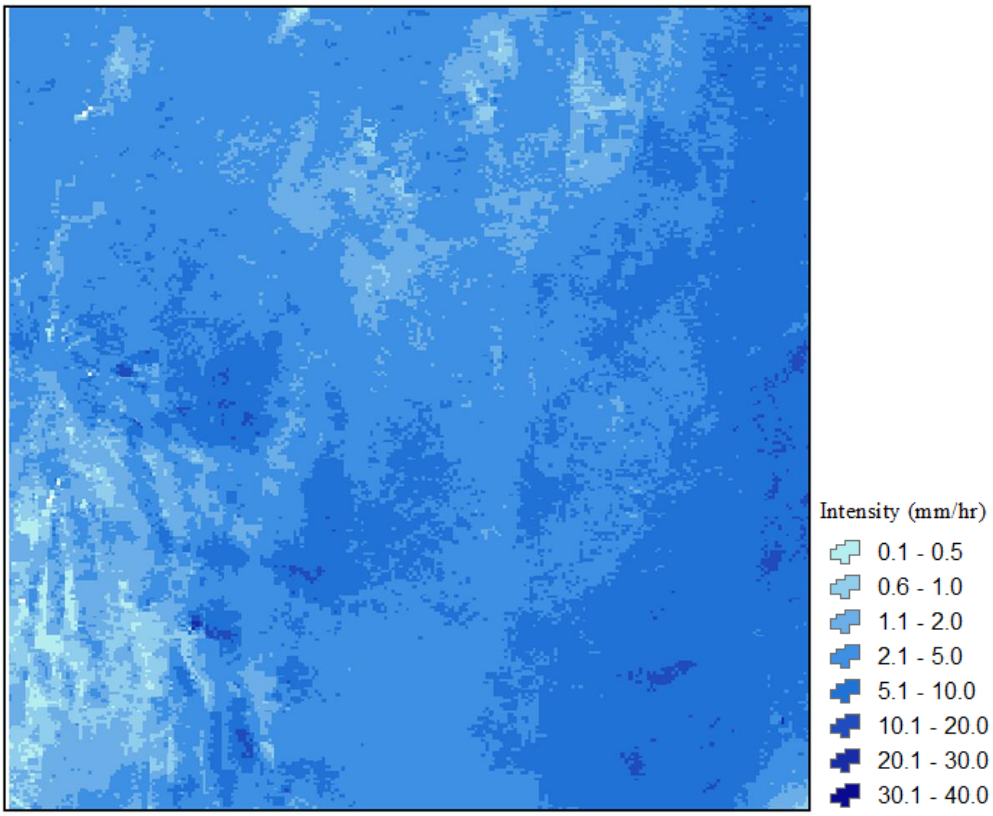


Figure 3.16 Radar-AMeDAS data for the event KAN_2006.10.06.08 with $\bar{\alpha}_{i=1,2,3} = \{1.75, 1.71, 1.74\}$

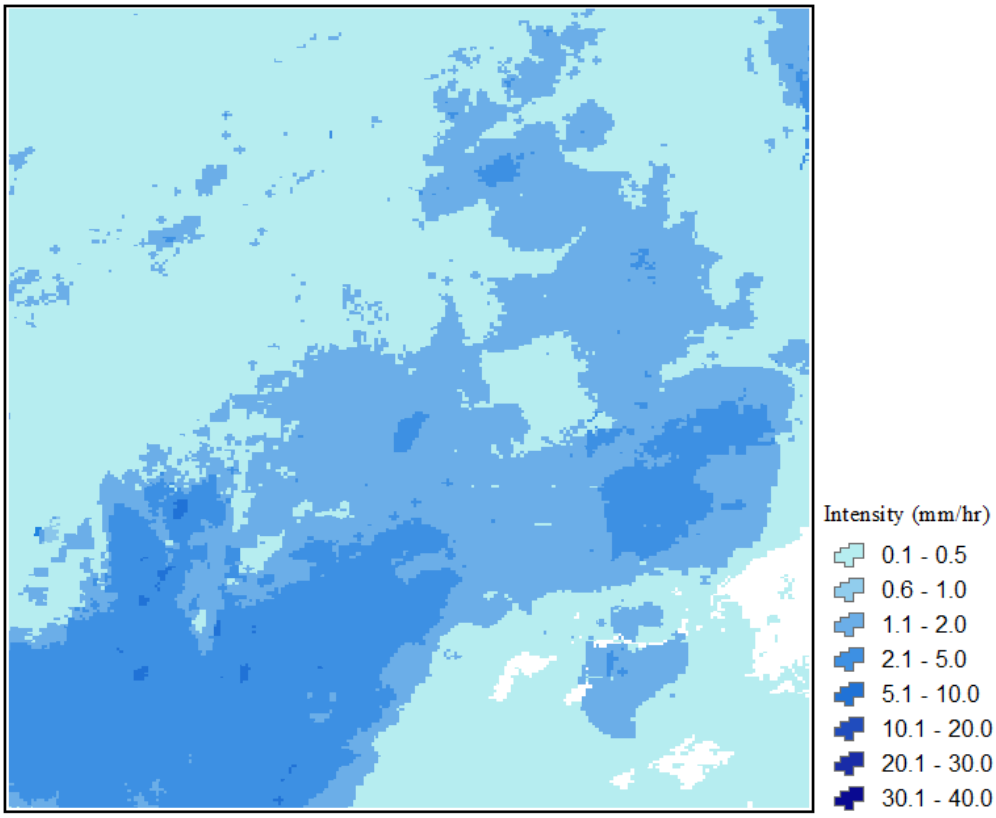


Figure 3.17 Radar-AMeDAS data for the event KAN_2009.01.23.04 with $\bar{\alpha}_{i=1,2,3} = \{1.60, 1.61, 1.63\}$

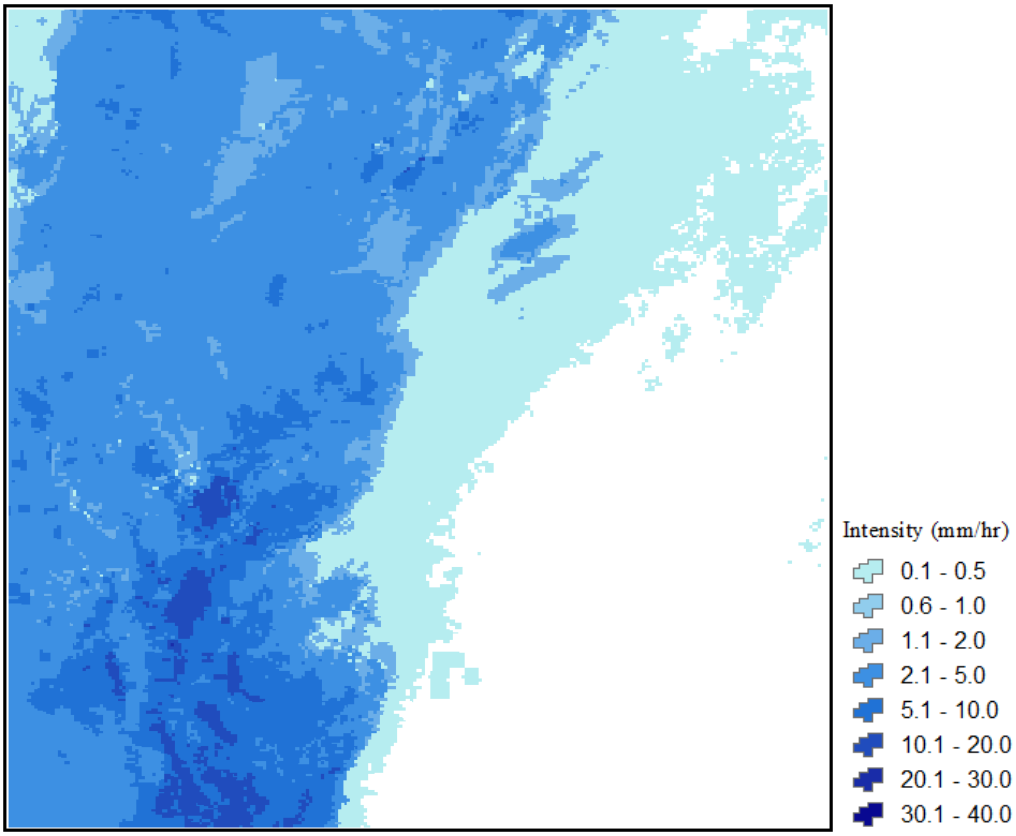


Figure 3.18 Radar-AMeDAS data for the event KAN_2008.12.05.15 with $\bar{\alpha}_{i=1,2,3} = \{1.51, 1.50, 1.47\}$

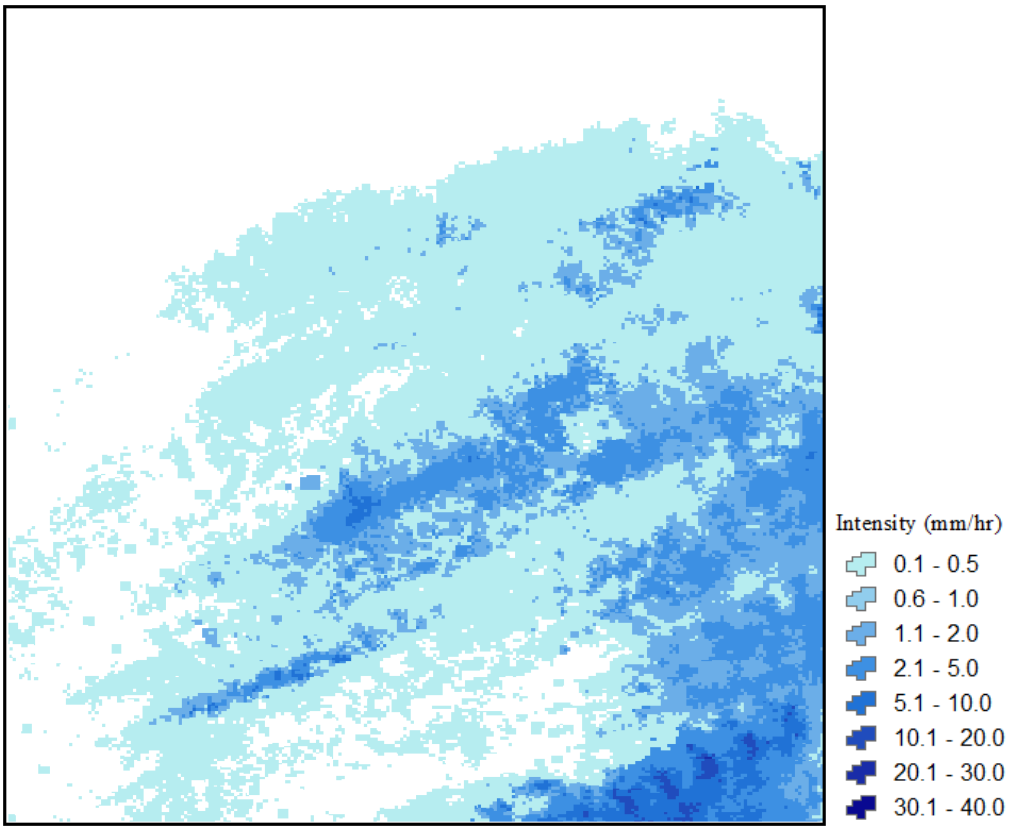


Figure 3.19 Radar-AMeDAS data for the event KAN_2006.10.29.03 with $\bar{\alpha}_{i=1,2,3} = \{1.41, 1.40, 1.40\}$

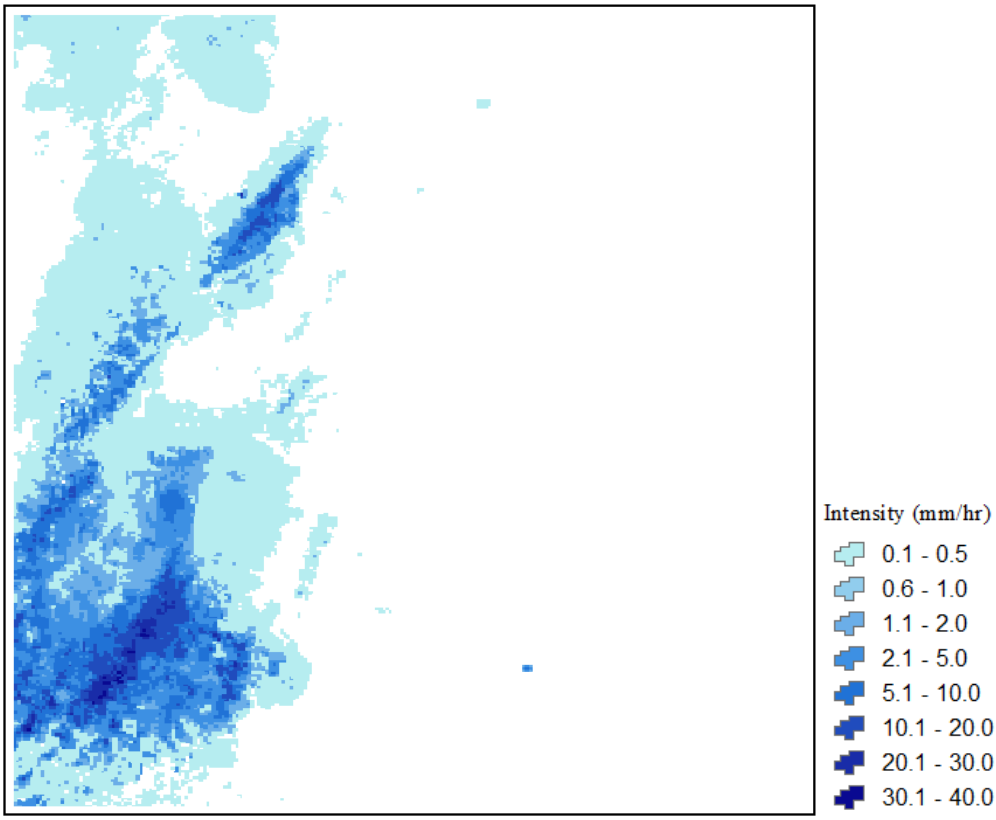


Figure 3.20 Radar-AMeDAS data for the event KAN_2007.09.10.06 with $\bar{\alpha}_{i=1,2,3} = \{1.21, 1.19, 1.20\}$

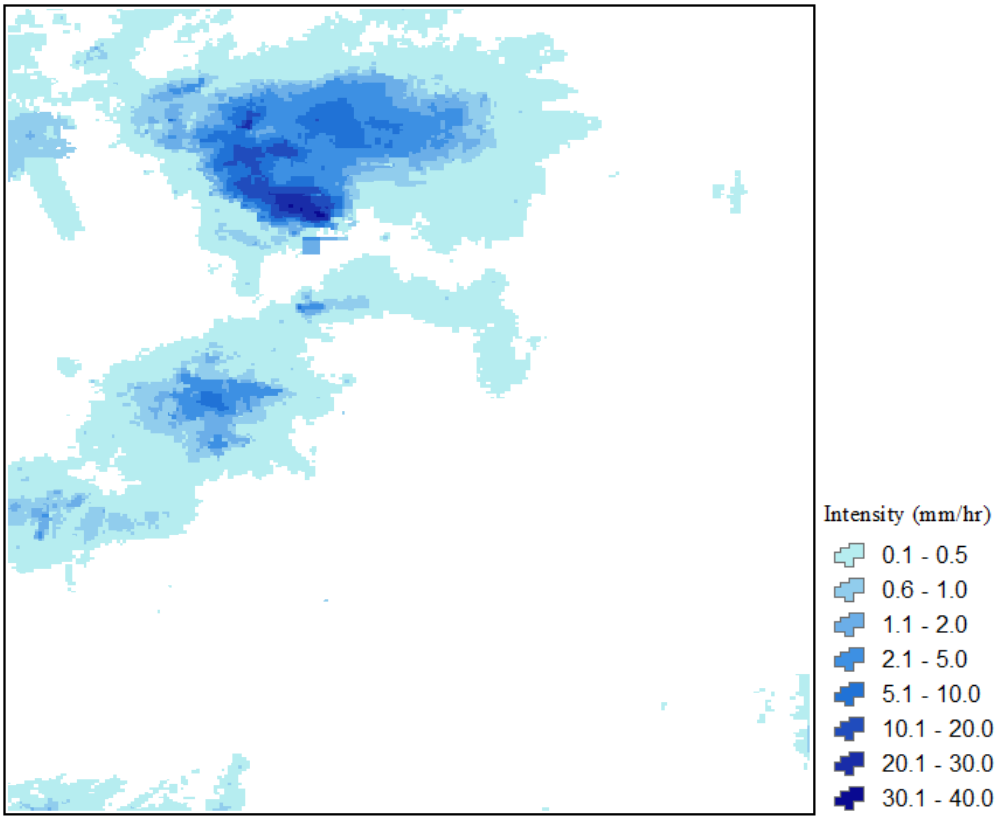


Figure 3.21 Radar-AMeDAS data for the event KAN_2009.08.07.07 with $\bar{\alpha}_{i=1,2,3} = \{1.00, 1.01, 1.04\}$

3.4.6.2 Estimation of the scale parameter

In the TSM method, the computation of the estimators $\hat{\gamma}_{m,i}$ is dependent on the computation of the estimators $\hat{\alpha}_{m,i}$, which seemingly could imply that bad estimators of the characteristic exponent generate bad estimators of the scale parameter. Even though the values $\bar{\alpha}_i$ correspond to the arithmetic mean of the estimators $\hat{\alpha}_{m,i}$ in the scale range $2 \leq m \leq 5$, the logarithmic regression of the estimators $\hat{\gamma}_{m,i}$ was carried out in the scale range $1 \leq m \leq 7$ with the purpose of observing how different would the resulting self-similarity index H_i be if the estimators $\hat{\gamma}_{m,i}$ of the lowest resolutions were also considered.

Table 3.8 shows the percentage of events that showed power-law behavior of $\gamma_{1,i}$ within different scale-ranges. In the previous section it was established that the estimation of H_i would be accepted only if the logarithmic regression of $\hat{\gamma}_{m,i}$ showed a coefficient of determination ${\rho^2}_i$ greater than or equal to 0.75 in the three directions. If this condition was not fulfilled, the regression was applied at narrower scale-ranges. Figures 3.22 to 3.25 show examples of events from the different columns in Table 3.8. The Figures also show a graph in which dashed lines represent the best-fit computed by the regression. Tables 3.9 to 3.12 show the values of $\gamma_{1,i}$, H_i and ${\rho^2}_i$ correspondent to those events. The parameter $\gamma_{1,i}$ corresponds to the point where the dashed lines cross the vertical axis in the graphs and the parameter H_i corresponds to the slope of the dashed lines. Additionally, Figure 3.26 shows an example of an event that falls in the category correspondent to the fifth column of Table 3.8.

Table 3.8 Percentage of events in which power law behavior of the scale parameter was observed within different scale ranges

	Range of scales					
	$2 \leq m \leq 7$	$2 \leq m \leq 6$	$2 \leq m \leq 5$	$2 \leq m \leq 4$		
	Range of resolutions					
Site	1-km to 64-km	1-km to 32-km	1-km to 16-km	1-km to 8-km	[(a)]	**Total**
KAN	81	6	3	2	8	100
STO	84	6	2	2	6	100
HOK	85	6	3	2	4	100
KYU	84	7	3	2	4	100
SHI	84	5	2	2	7	100

[(a)] Percentage of events showing self-similarity between 1-km resolution and 4-km resolution, or failing to have the values of ${\rho^2}_i$ greater than or equal to 0.75

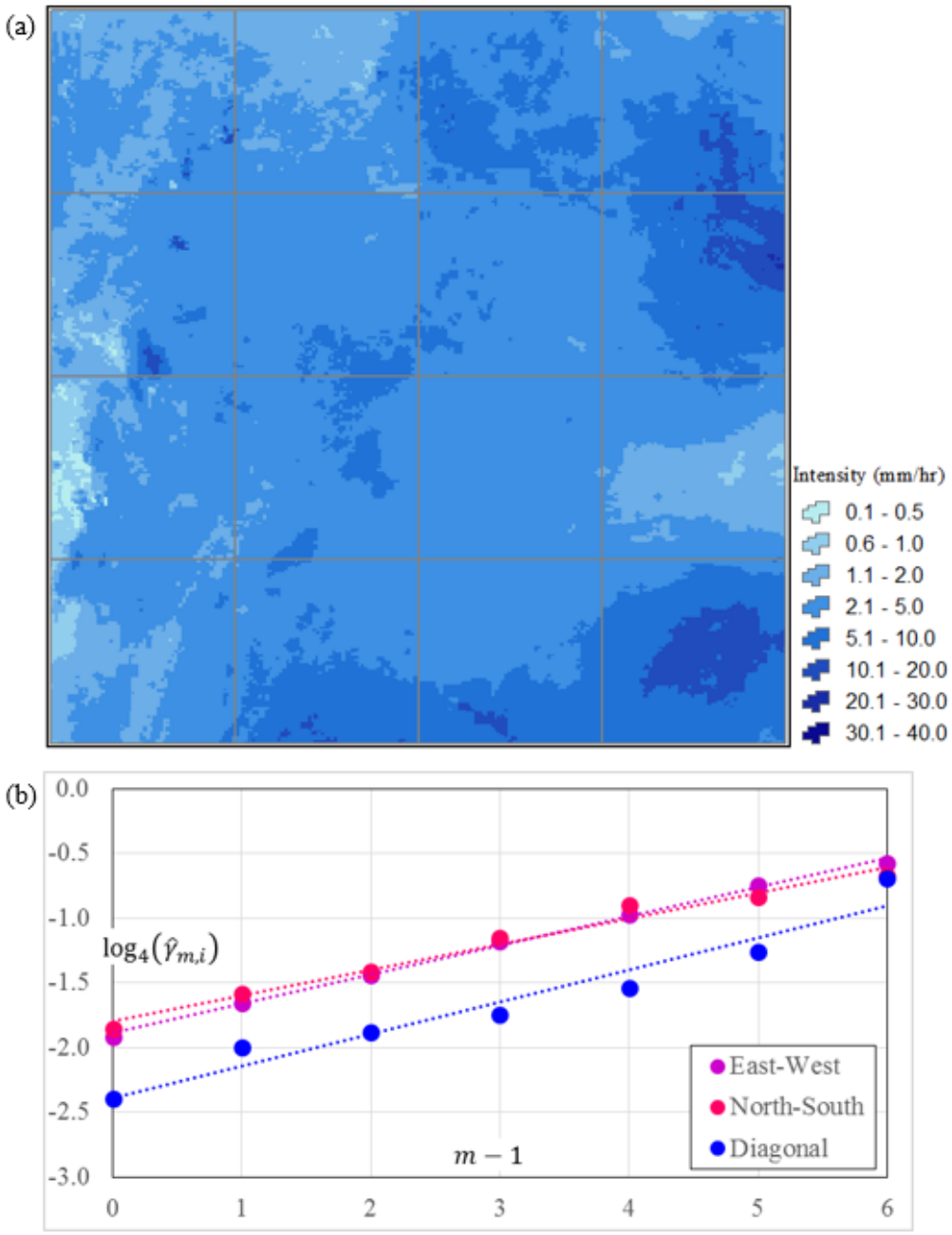


Figure 3.22 Examples of power-law behavior (1)

(a) Radar-AMeDAS data for the event KAN_2009.10.26.11. The grid at 64-km resolution, correspondent to scale $m = 7$, is shown for reference. (b) Graph showing the logarithmic regression of $\gamma_{m,i}$ on $m - 1$ showing acceptable correlation in the scale range $1 \leq m \leq 7$.

Table 3.9 Results of the logarithmic regression of $\gamma_{m,i}$ on $m - 1$ for the event KAN_2009.10.26.11

	i		
	1	2	3
$\gamma_{1,i}$	0.053	0.063	0.026
H_i	0.226	0.198	0.247
${\rho^2}_i$	0.998	0.991	0.970

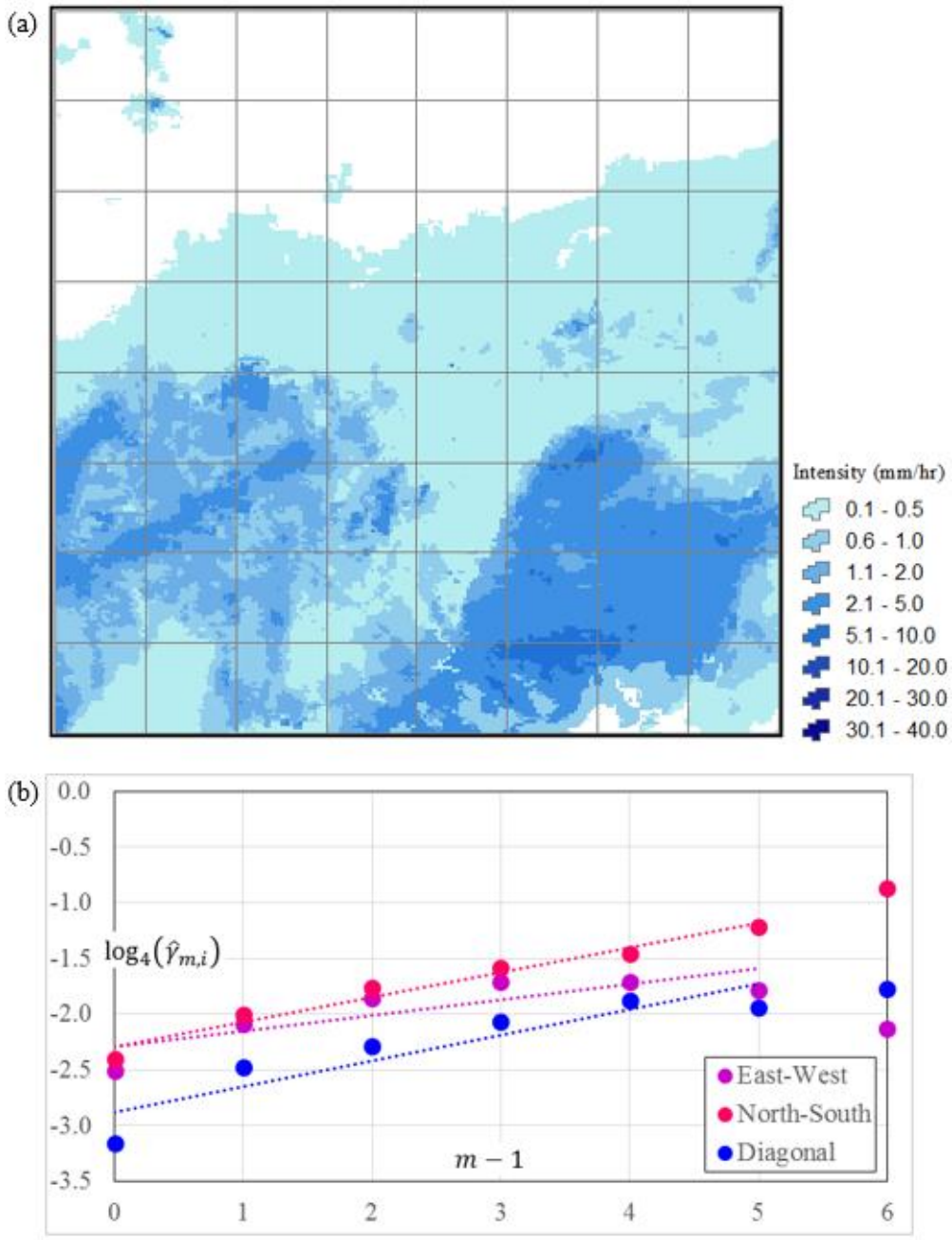


Figure 3.23 Examples of power-law behavior (2)

(a) Radar-AMeDAS data for the event KAN_2009.05.05.15. The grid at 32-km resolution, correspondent to scale $m = 6$, is shown for reference. (b) Graph showing the logarithmic regression of $\gamma_{m,i}$ on $m - 1$ showing acceptable correlation in the scale range $1 \leq m \leq 6$.

Table 3.10 Results of the logarithmic regression of $\gamma_{m,i}$ on $m - 1$ for the event KAN_2009.05.05.15

	i		
	1	2	3
$\gamma_{1,i}$	0.034	0.031	0.013
H_i	0.140	0.221	0.230
$\rho^2{}_i$	0.848	0.984	0.909

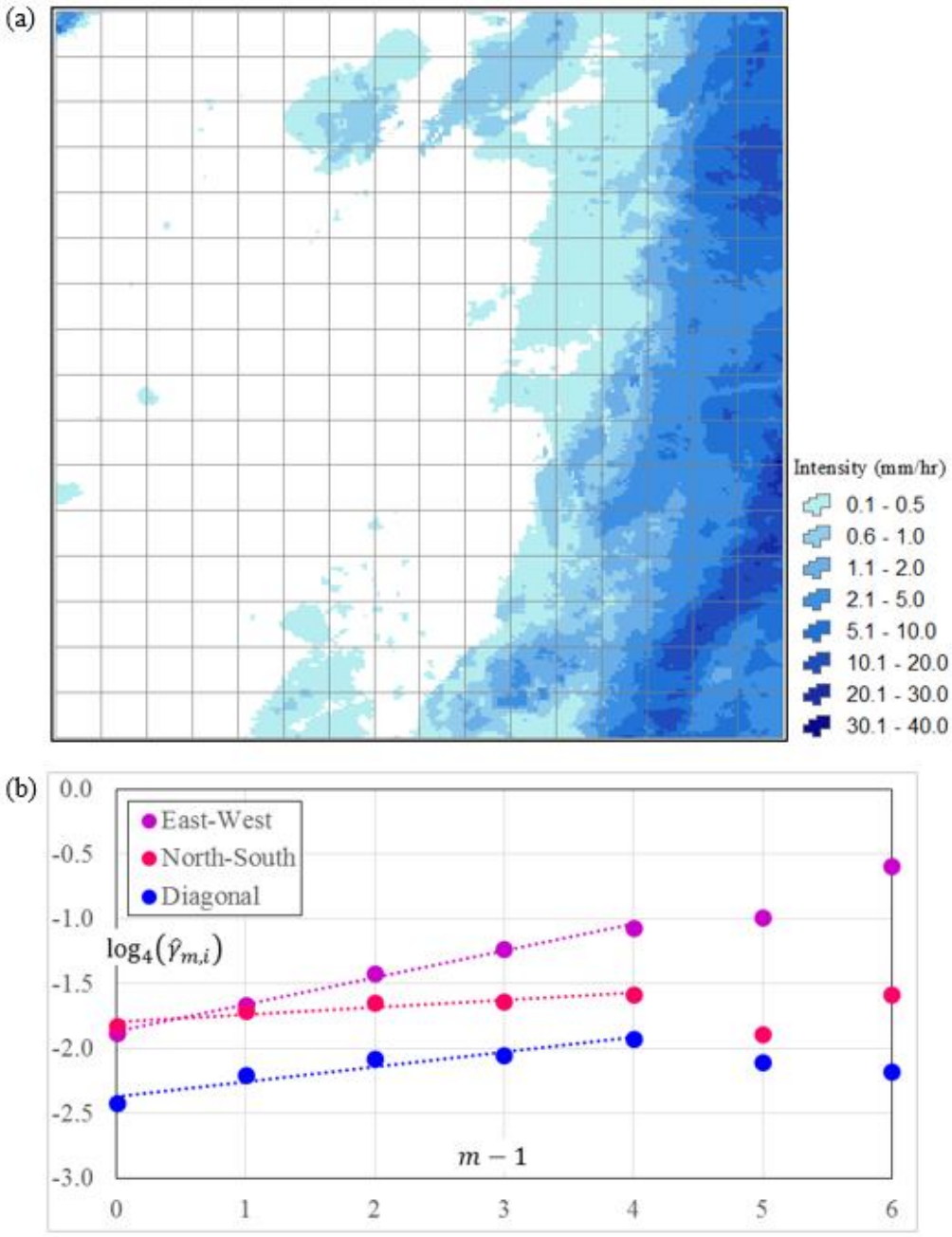


Figure 3.24 Examples of power-law behavior (3)

(a) Radar-AMeDAS data for the event KAN_2009.12.05.21. The grid at 16-km resolution, correspondent to scale $m = 5$, is shown for reference. (b) Graph showing the logarithmic regression of $\gamma_{m,i}$ on $m - 1$ showing acceptable correlation in the scale range $1 \leq m \leq 5$.

Table 3.11 Results of the logarithmic regression of $\gamma_{m,i}$ on $m - 1$ for the event KAN_2009.12.05.21

	i		
	1	2	3
$\gamma_{1,i}$	0.057	0.077	0.032
H_i	0.206	0.056	0.114
${\rho^2}_i$	0.997	0.947	0.963

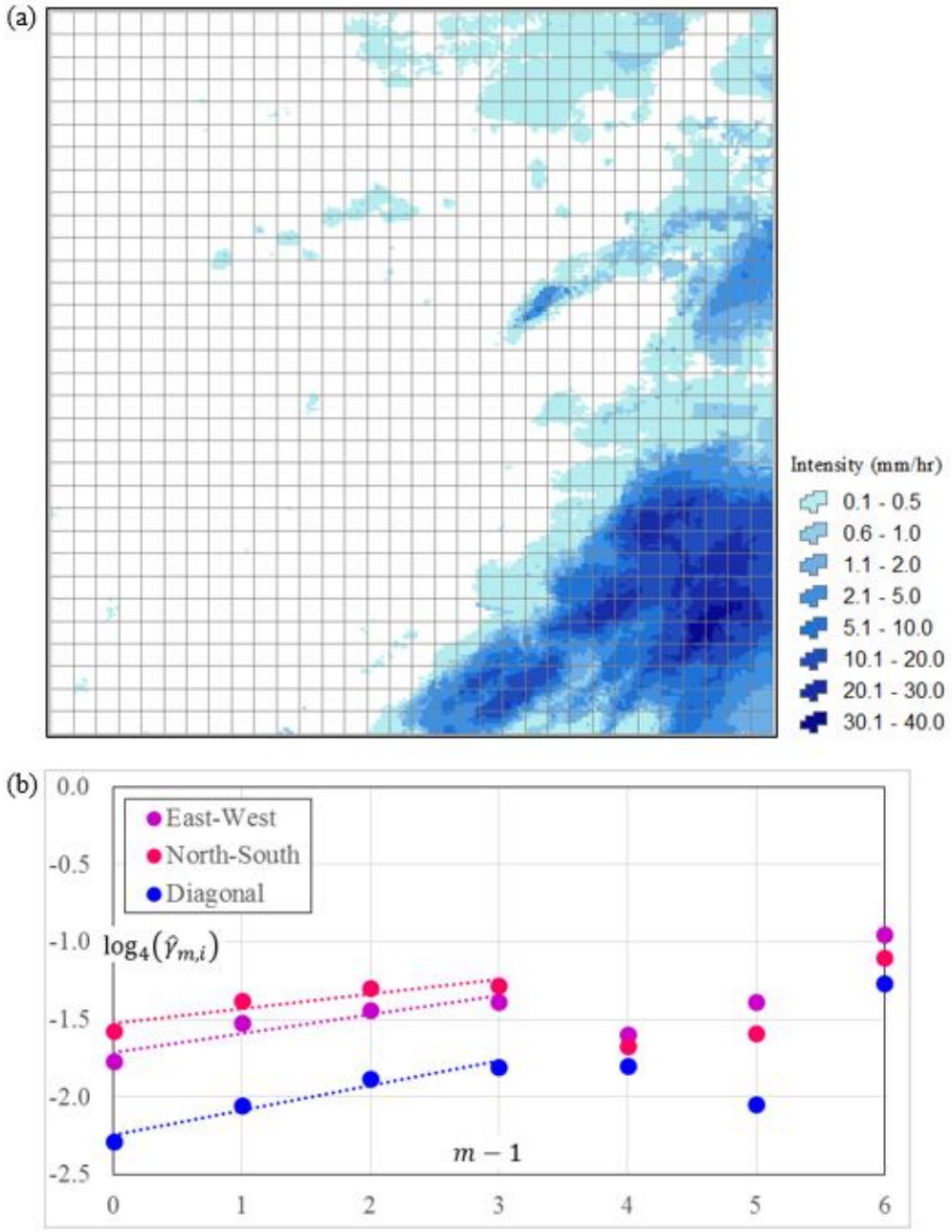


Figure 3.25 Examples of power-law behavior (4)

(a) Radar-AMeDAS data for the event KAN_2009.08.10.17. The grid at 8-km resolution, correspondent to scale $m = 4$, is shown for reference. (b) Graph showing the logarithmic regression of $\gamma_{m,i}$ on $m - 1$ showing acceptable correlation in the scale range $1 \leq m \leq 4$.

Table 3.12 Results of the logarithmic regression of $\gamma_{m,i}$ on $m - 1$ for the event KAN_2009.08.10.17

	i		
	1	2	3
$\gamma_{1,i}$	0.079	0.106	0.035
H_i	0.123	0.096	0.162
$\rho^2{}_i$	0.941	0.920	0.978

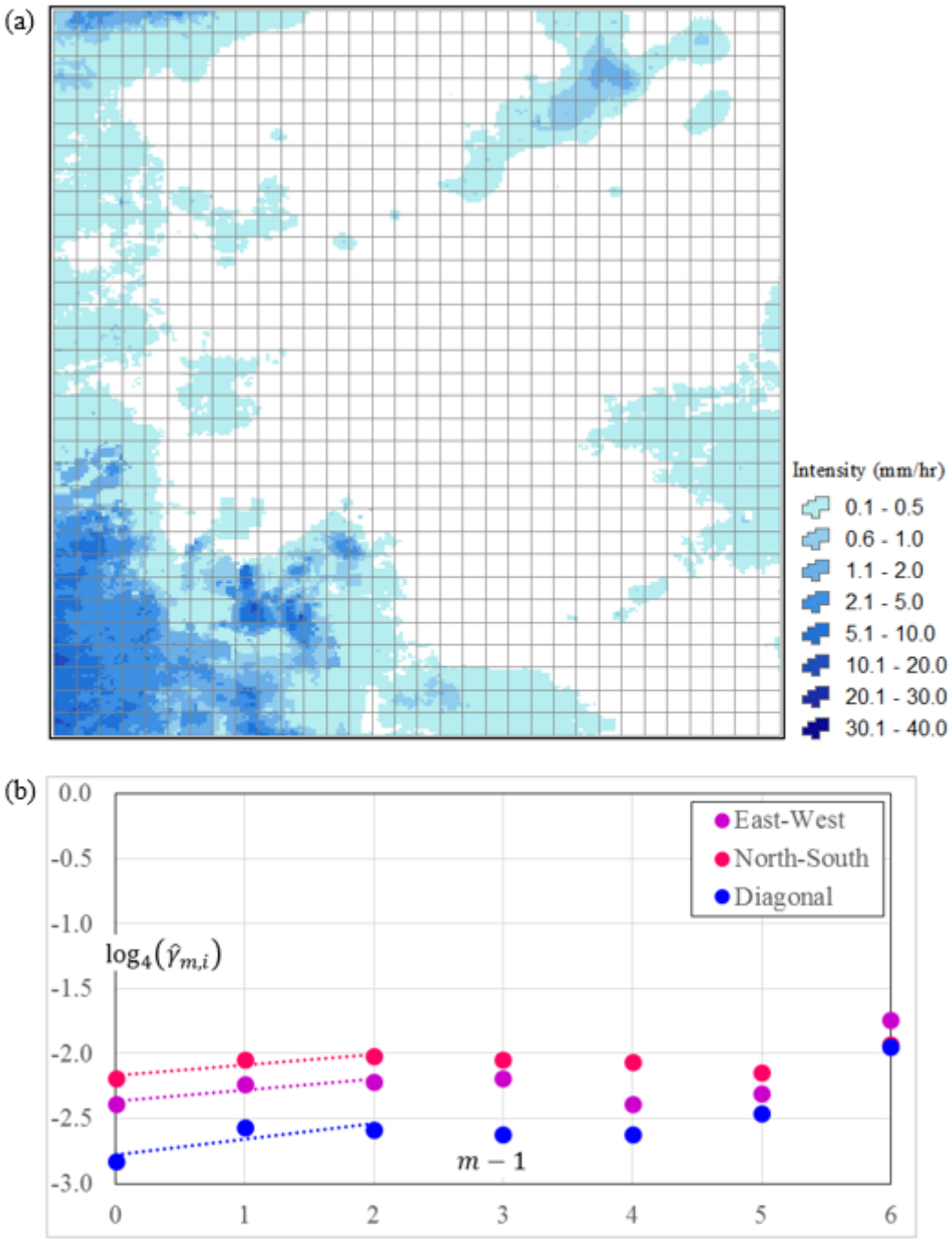


Figure 3.26 Examples of power-law behavior (5)

(a) Radar-AMeDAS data for the event KAN_2009.05.17.20. The grid at 8-km resolution, correspondent to scale $m = 4$, is shown for reference. (b) Graph showing the logarithmic regression of $\gamma_{m,i}$ on $m - 1$ showing acceptable correlation in the scale range $1 \leq m \leq 3$.

In order to have an idea of the parameter-range of the parameters estimated via logarithmic regression, Figure 3.27 shows the normalized frequencies of $\gamma_{1,i}$ and Figure 3.28 shows the normalized frequencies of H_i for all selected events in each of the analysis sites.

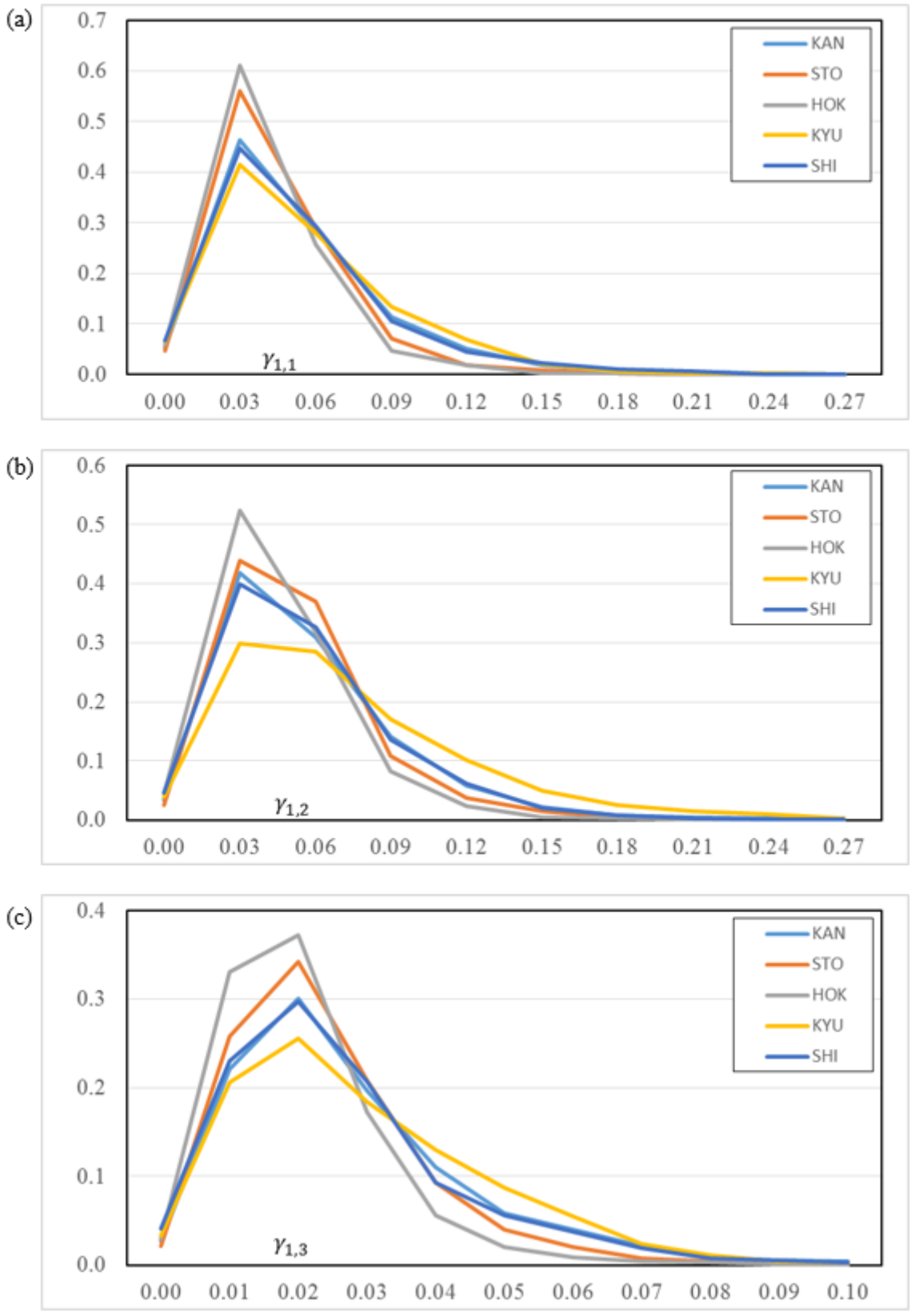


Figure 3.27 Normalized frequencies of $\gamma_{1,1}$, $\gamma_{1,2}$, and $\gamma_{1,3}$

The graphs include the results of the all selected events in each of the analysis sites for the (a) horizontal direction, (b) the vertical direction, and (c) the diagonal direction.

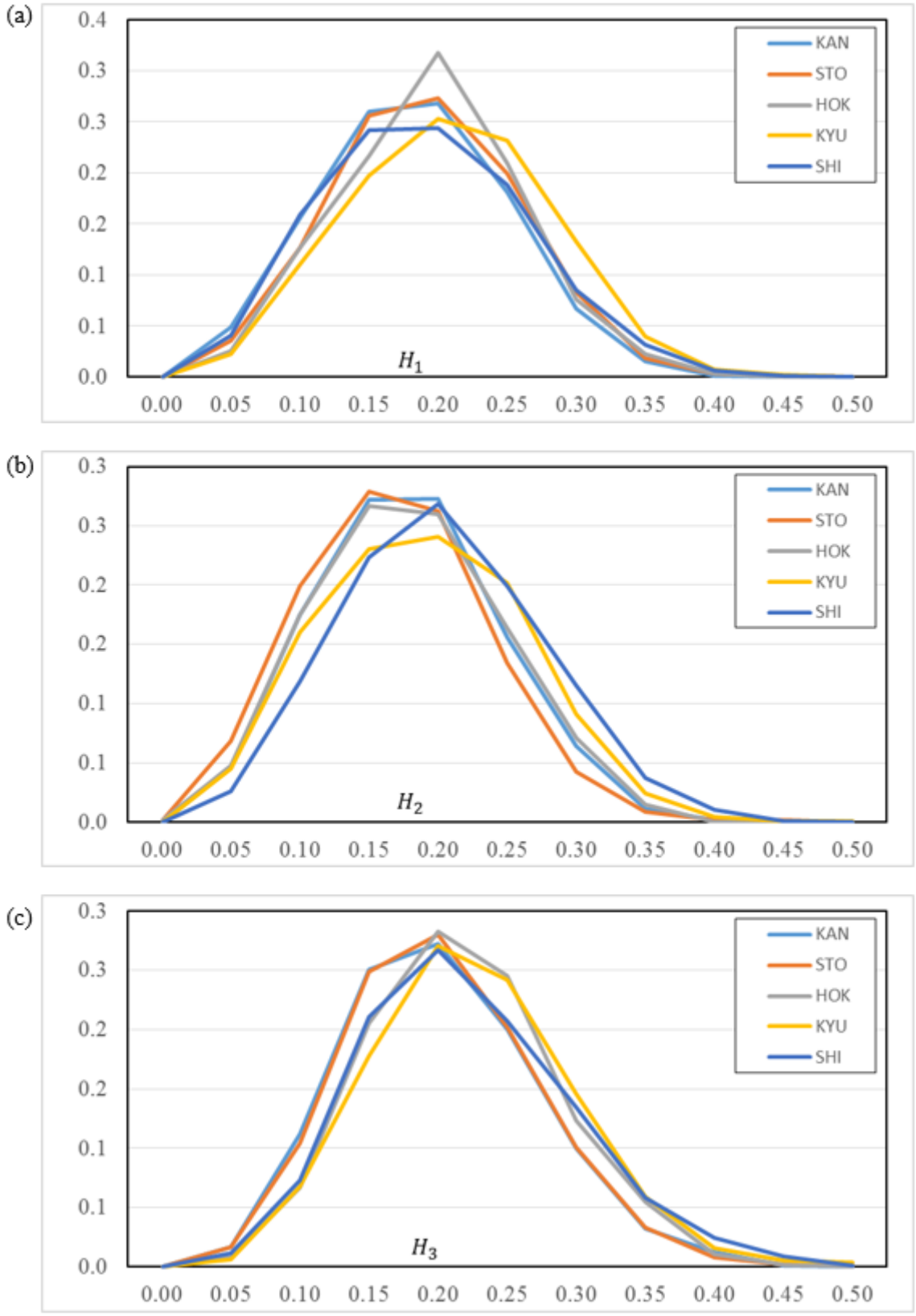


Figure 3.28 Normalized frequencies of H_1, H_2, and H_3

The graphs include the results of the all selected events in each of the analysis sites for the (a) horizontal direction, (b) the vertical direction, and (c) the diagonal direction.

Figure 3.27 denotes that the majority of the events have values of $\gamma_{1,i}$ between 0.00 and 0.09 in the east-west and north-south directions, and between 0.00 and 0.05 in the diagonal direction. The analysis sites that have more annual accumulated rainfall (Kyushu and Shikoku) seem to have a slightly higher frequency of extreme values.

On the other hand, Figure 3.28 shows that the frequencies are well-distributed in the parameter-range $0.00 \leq H_i \leq 0.50$ in all three directions for all analysis sites.

3.4.6.3 Inter-directional dependence of the parameters

The purpose of this section is to make a statistical description of the estimators of the parameters of α-stable distributions of rainfall fluctuations and the correspondent self-similarity indexes computed for the five analysis sites.

Despite the fact there is no a priori reason for the values of $\bar{\alpha}_i$ to be equal in all three directions, in a large percentage of the events similar values of $\bar{\alpha}_1$, $\bar{\alpha}_2$ and $\bar{\alpha}_3$ were found. For all the analysis sites, Figures 3.29 to 3.33 show the scatterplots of $\bar{\alpha}_1$ - $\bar{\alpha}_2$, and $\bar{\alpha}_1$ - $\bar{\alpha}_3$. In these Figures are included only those events in which the estimators $\hat{\gamma}_{m,i}$ proved to have power-law behavior in the scale-ranges $1 \leq m \leq 4$, $1 \leq m \leq 5$, $1 \leq m \leq 6$, and $1 \leq m \leq 7$ (corresponding to the events in the percentages shown in the first four columns of Table 3.8). Each graph presents the equation of a red dotted line that indicates the best fit for the pairs of values. Additionally, the graphs depict an error term, ε, marked by red dashed lines corresponding to the 5th and 95th percentiles of the differences of the pairs of values. Generally, the pairs of values in all analysis sites show good correlation to the best-fit lines. Therefore, it can be concluded that the characteristic exponent tends to be equal in all three directions, although, a slight difference of approximately ± 0.10 between $\bar{\alpha}_1$ and $\bar{\alpha}_2$ and between $\bar{\alpha}_1$ and $\bar{\alpha}_3$ is expected regardless of the type of rainfall.

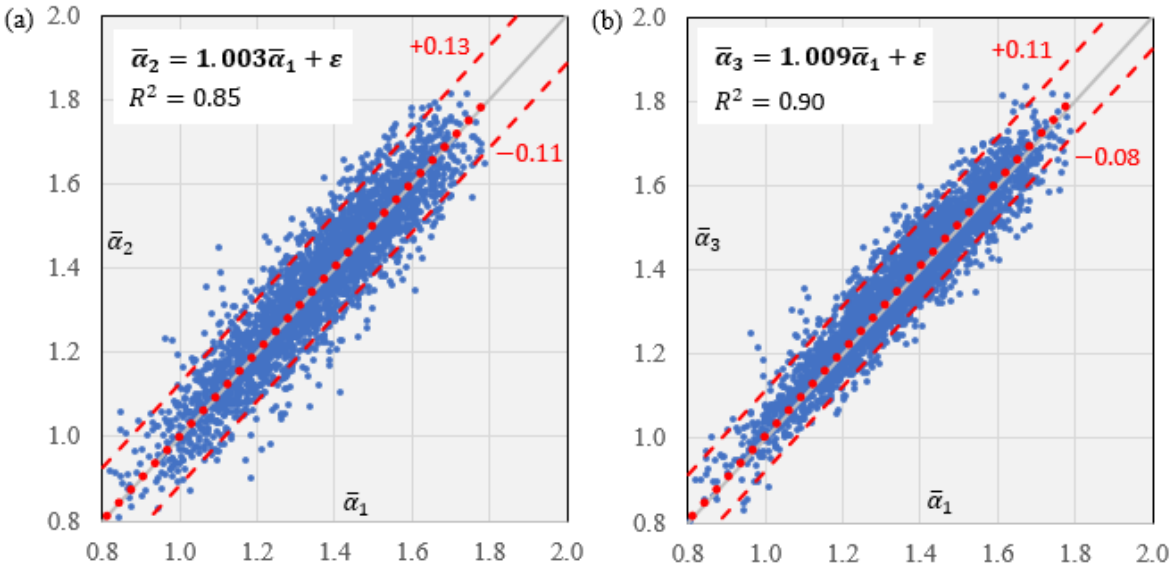


Figure 3.29 Scatterplots of (a) $\bar{\alpha}_1$ - $\bar{\alpha}_2$, and (b) $\bar{\alpha}_1$ - $\bar{\alpha}_3$ for the Kanto site.

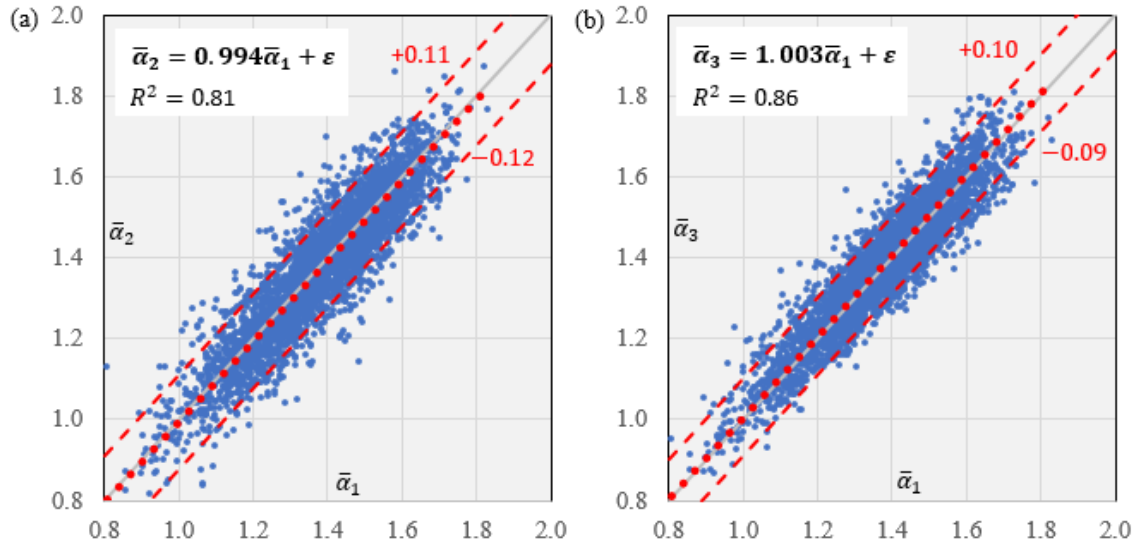


Figure 3.30 Scatterplots of (a) $\bar{\alpha}_1$ - $\bar{\alpha}_2$, and (b) $\bar{\alpha}_1$ - $\bar{\alpha}_3$ for the South Tohoku site.

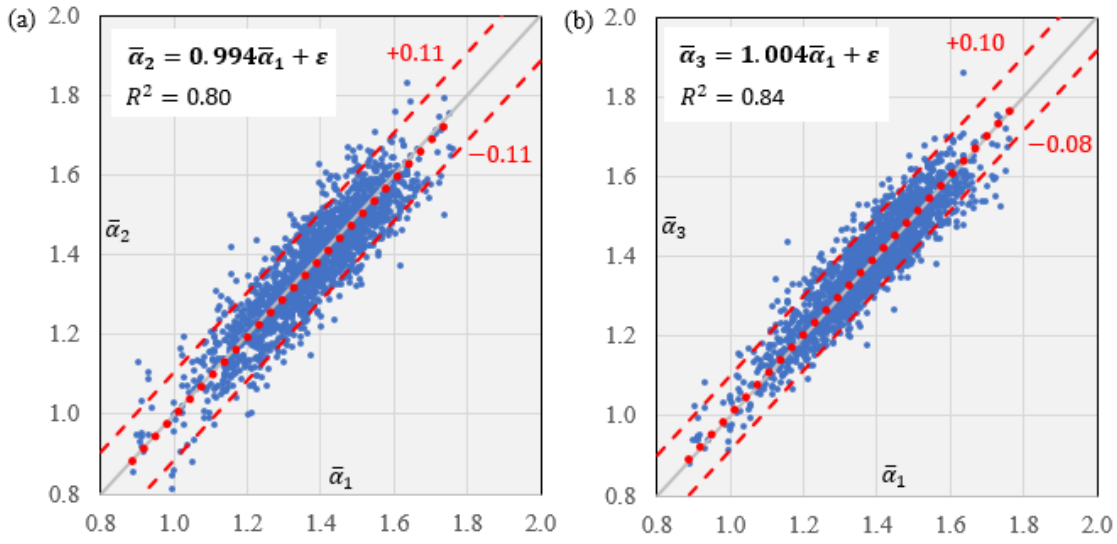


Figure 3.31 Scatterplots of (a) $\bar{\alpha}_1$ - $\bar{\alpha}_2$, and (b) $\bar{\alpha}_1$ - $\bar{\alpha}_3$ for the Hokkaido site.

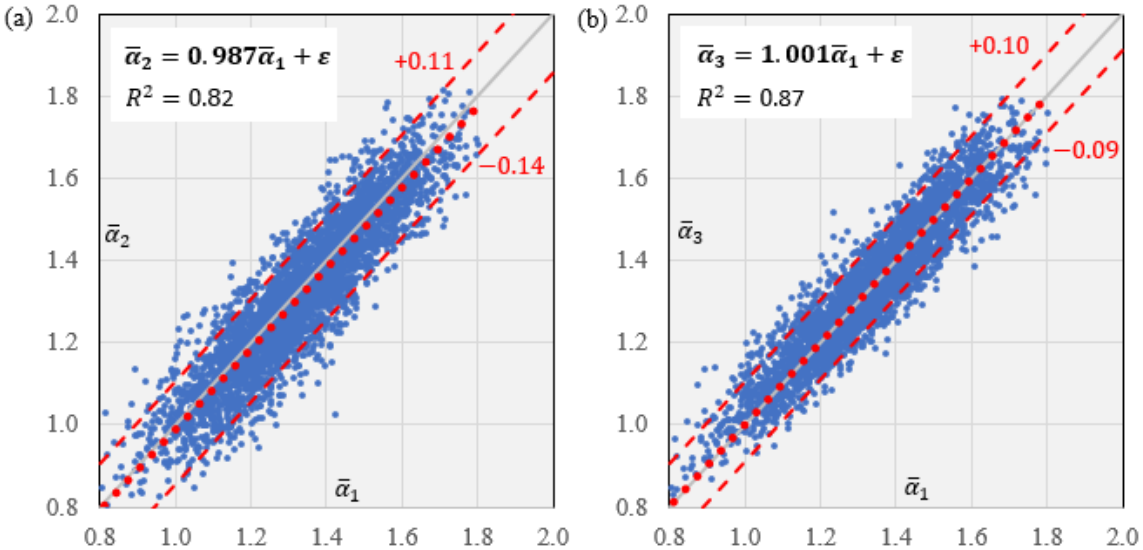


Figure 3.32 Scatterplots of (a) $\bar{\alpha}_1$ - $\bar{\alpha}_2$, and (b) $\bar{\alpha}_1$ - $\bar{\alpha}_3$ for the Kyushu site.

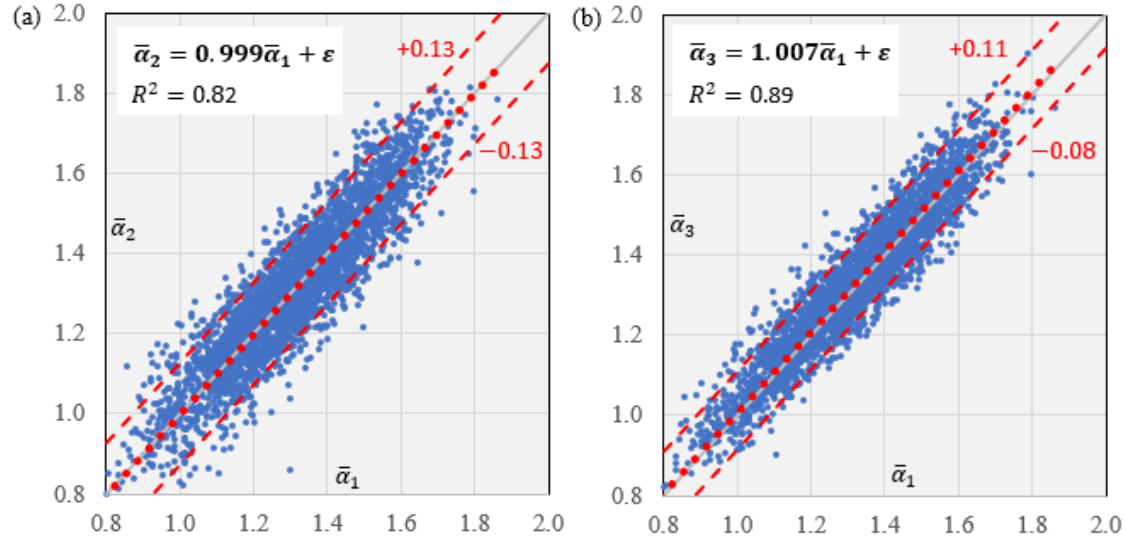


Figure 3.33 Scatterplots of (a) $\bar{\alpha}_1$ - $\bar{\alpha}_2$, and (b) $\bar{\alpha}_1$ - $\bar{\alpha}_3$ for the Shikoku site.

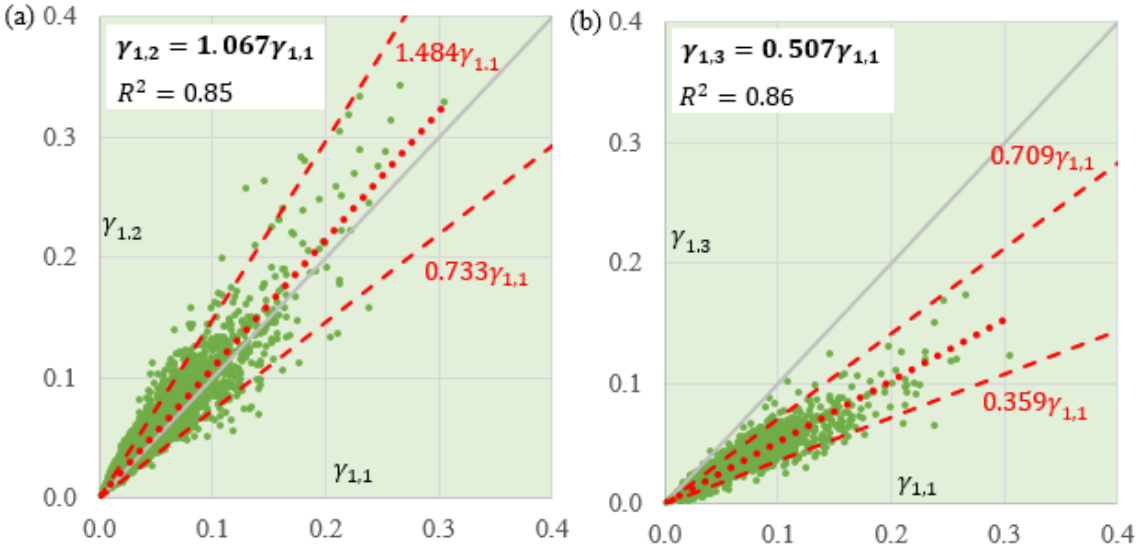


Figure 3.34 Scatterplots of (a) $\gamma_{1,1}$ - $\gamma_{1,2}$, and (b) $\gamma_{1,1}$ - $\gamma_{1,3}$ for the Kanto site.

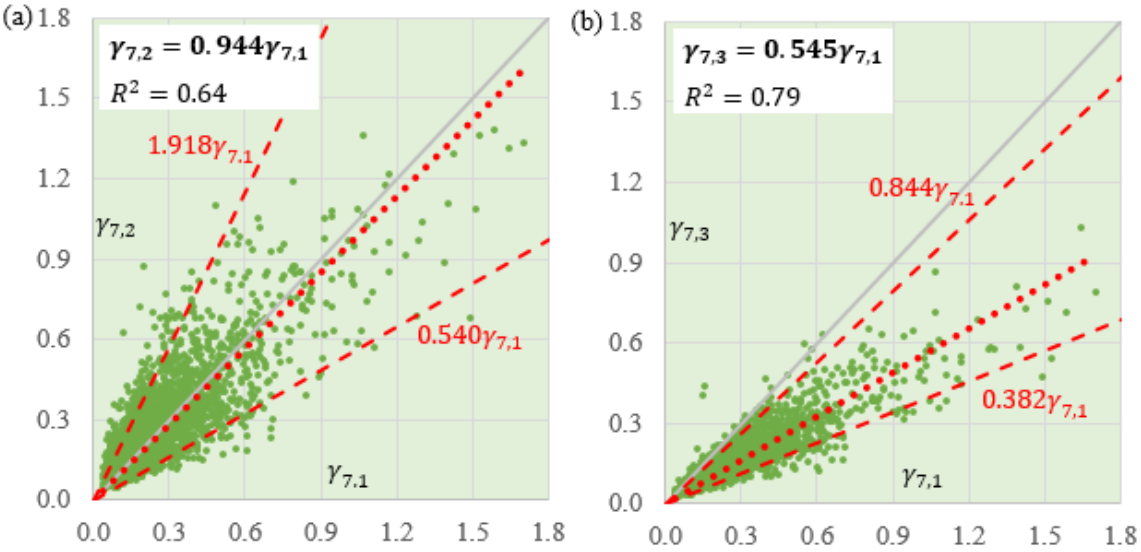


Figure 3.35 Scatterplots of (a) $\gamma_{7,1}$ - $\gamma_{7,2}$, and (b) $\gamma_{7,1}$ - $\gamma_{6,3}$ for the Kanto site.

Figures 3.34, 3.36, 3.38, 3.40 and 3.42 show the scatterplots of $\gamma_{1,1}$ - $\gamma_{1,2}$ and $\gamma_{1,1}$ - $\gamma_{1,3}$ for all the analysis sites. In these Figures are included only those events in which the estimators $\hat{\gamma}_{m,i}$ proved to have power-law behavior in the scale-ranges $1 \leq m \leq 4$, $1 \leq m \leq 5$, $1 \leq m \leq 6$, and $1 \leq m \leq 7$ (corresponding to the events in the percentages shown in the first four columns of Table 3.8). Similarly, Figures 3.35, 3.37, 3.39, 3.41 and 3.43 show the scatterplots of $\gamma_{7,1}$ - $\gamma_{7,2}$ and $\gamma_{7,1}$ - $\gamma_{7,3}$ for all the analysis sites. In these Figures are included only those events in which the estimators $\hat{\gamma}_{m,i}$ proved to have power-law behavior in the scale-range $1 \leq m \leq 7$ (corresponding to the events in the percentages shown in the first column of Table 3.8). Each graph presents the equation of a red dotted line that indicates the best fit for the pairs of values. Additionally, the graphs depict two red dashed lines corresponding to the 5th and 95th percentiles of the ratios of the pairs of values.

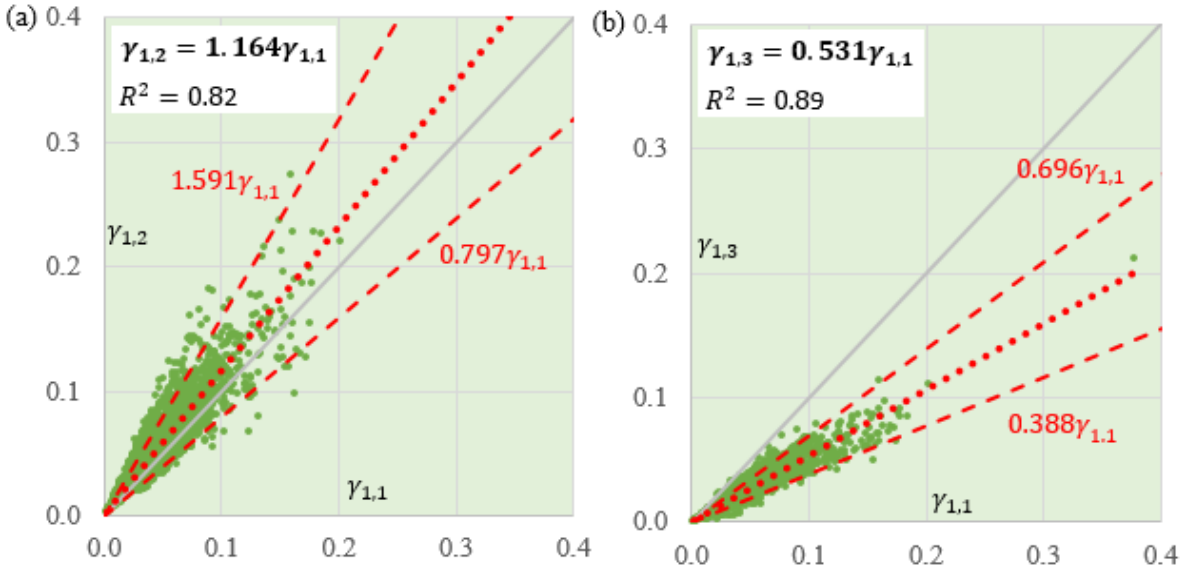


Figure 3.36 Scatterplots of (a) $\gamma_{1,1}$ - $\gamma_{1,2}$, and (b) $\gamma_{1,1}$ - $\gamma_{1,3}$ for the South Tohoku site.

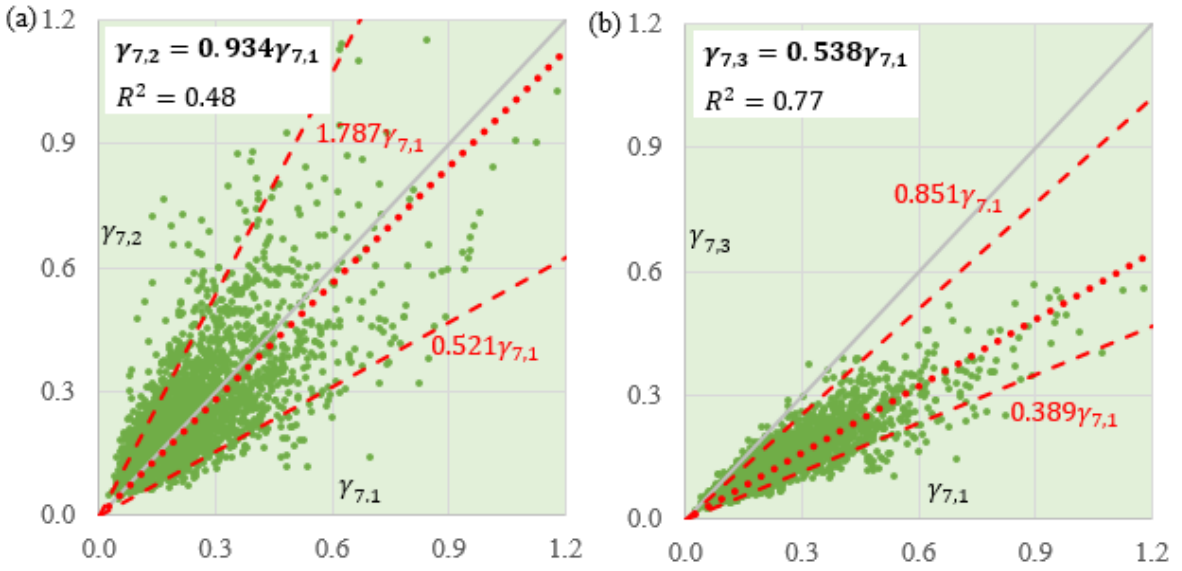


Figure 3.37 Scatterplots of (a) $\gamma_{7,1}$ - $\gamma_{7,2}$, and (b) $\gamma_{7,1}$ - $\gamma_{7,3}$ for the South Tohoku site.

In all the five analysis sites, the pairs of values of $\gamma_{1,1}$ - $\gamma_{1,2}$, and $\gamma_{1,1}$ - $\gamma_{1,3}$ show good correlation to the best-fit lines. However, due to a larger dispersion, the correlation of the pairs of values $\gamma_{7,1}$ - $\gamma_{7,2}$ and $\gamma_{7,1}$ - $\gamma_{7,3}$ is noticeably lower. It can be concluded that the scale parameter tends to have a mean trend with respect to other directions, although, the value of this trend happens to differ from one analysis site to another. This condition suggests that the parameter $\gamma_{m,i}$ should be somewhat dependent to the spatial structure of rainfall intensity. A difference between $\gamma_{m,1}$ and $\gamma_{m,2}$ and between $\gamma_{m,1}$ and $\gamma_{m,3}$ is expected, which happens to be larger as the magnitude of the parameters become larger and as the resolution becomes lower.

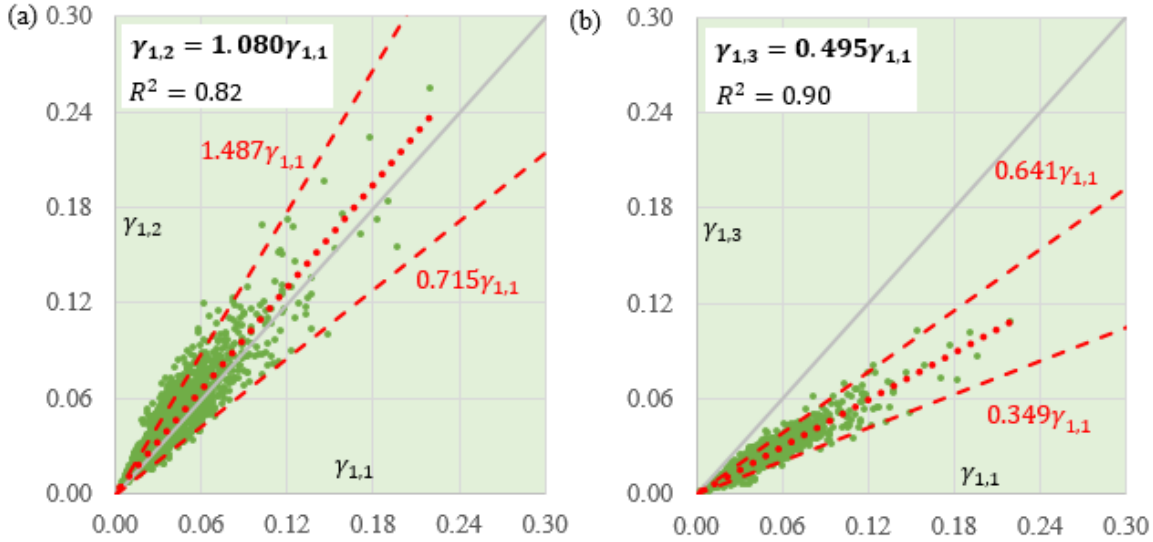


Figure 3.38 Scatterplots of (a) $\gamma_{1,1}$ - $\gamma_{1,2}$, and (b) $\gamma_{1,1}$ - $\gamma_{1,3}$ for the Hokkaido site.

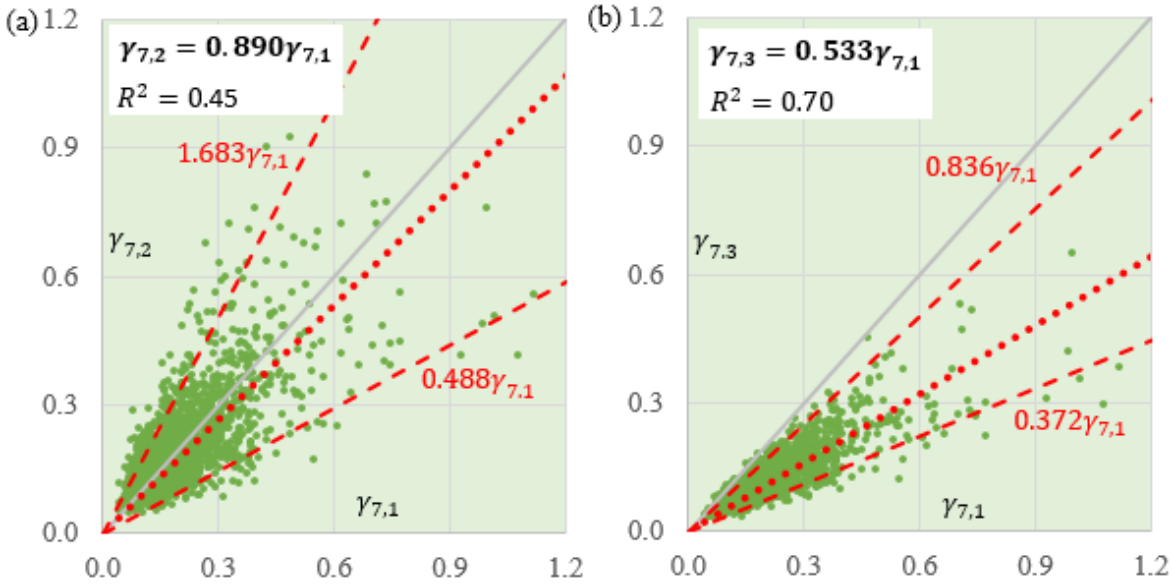


Figure 3.39 Scatterplots of (a) $\gamma_{7,1}$ - $\gamma_{7,2}$, and (b) $\gamma_{7,1}$ - $\gamma_{7,3}$ for the Hokkaido site.

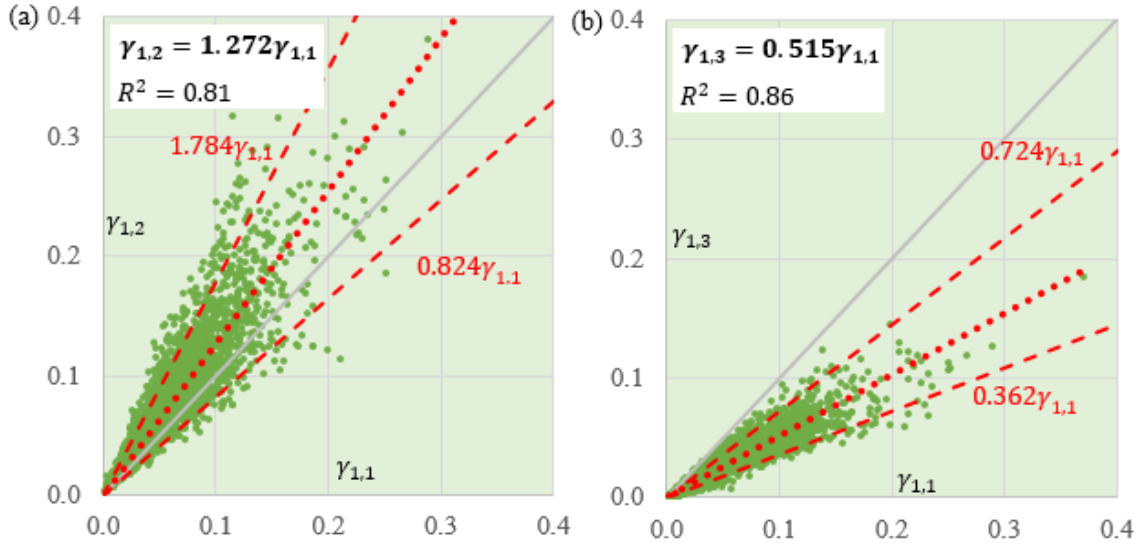


Figure 3.40 Scatterplots of (a) $\gamma_{1,1}$ - $\gamma_{1,2}$, and (b) $\gamma_{1,1}$ - $\gamma_{1,3}$ for the Kyushu site.

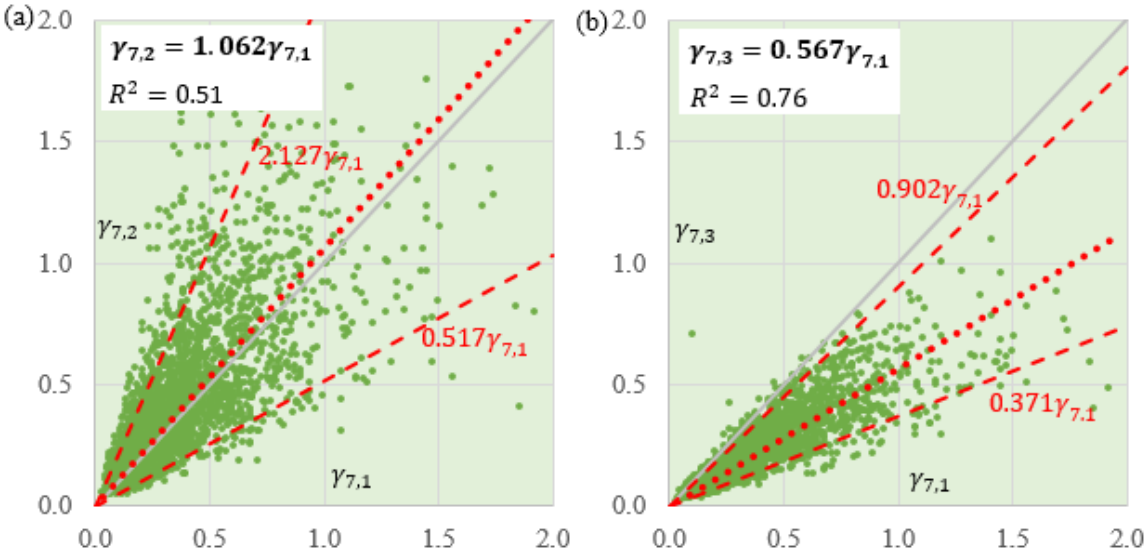


Figure 3.41 Scatterplots of (a) $\gamma_{7,1}$ - $\gamma_{7,2}$, and (b) $\gamma_{7,1}$ - $\gamma_{7,3}$ for the Kyushu site.

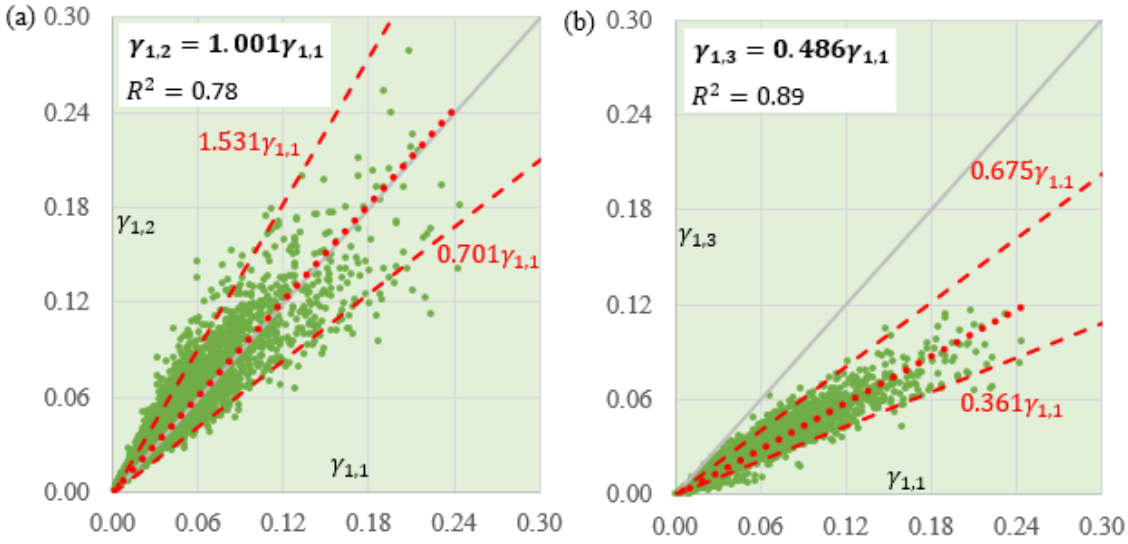


Figure 3.42 Scatterplots of (a) $\gamma_{1,1}$ - $\gamma_{1,2}$, and (b) $\gamma_{1,1}$ - $\gamma_{1,3}$ for the Shikoku site.

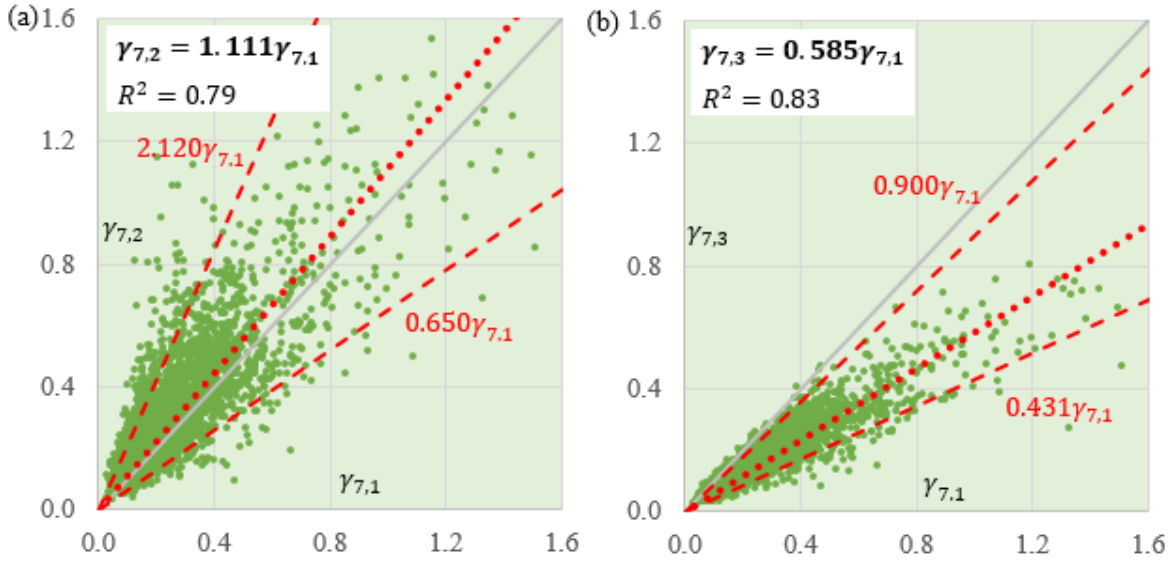


Figure 3.43 Scatterplots of (a) $\gamma_{7,1}$ - $\gamma_{7,2}$, and (b) $\gamma_{7,1}$ - $\gamma_{7,3}$ for the Shikoku site.

Figures 3.44 to 3.48 show the scatterplots of H_1 - H_2 and H_1 - H_3 for all the analysis sites. In these Figures are included only those events in which the estimators $\hat{\gamma}_{m,i}$ proved to have power-law behavior in the scale-ranges $1 \leq m \leq 4$, $1 \leq m \leq 5$, $1 \leq m \leq 6$, and $1 \leq m \leq 7$ (corresponding to the events in the percentages shown in the first four columns of Table 3.8). Each graph presents the equation of a blue dotted line that indicates the best fit for the pairs of values. Additionally, the graphs depict an error term, ε, marked by blue dashed lines corresponding to the 5th and 95th percentiles of the differences of the pairs of values. Generally, the best-fit lines show bad correlations due to the large differences between the pairs of values. It is difficult to affirm that the self-similarity index has some kind of inter-directional dependency. Nevertheless, the values of the 5th and 95th percentiles given in blue numbers in each graph might be helpful to establish a threshold if estimation is needed.

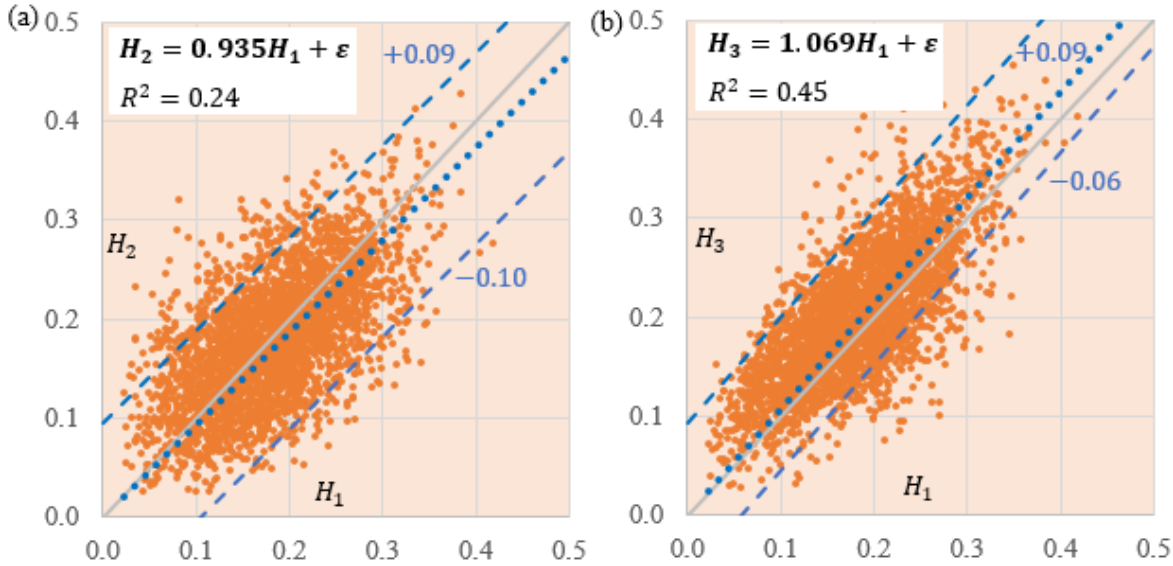


Figure 3.44 Scatterplots of (a) H_1 - H_2, and (b) H_1 - H_3 for the Kanto site.

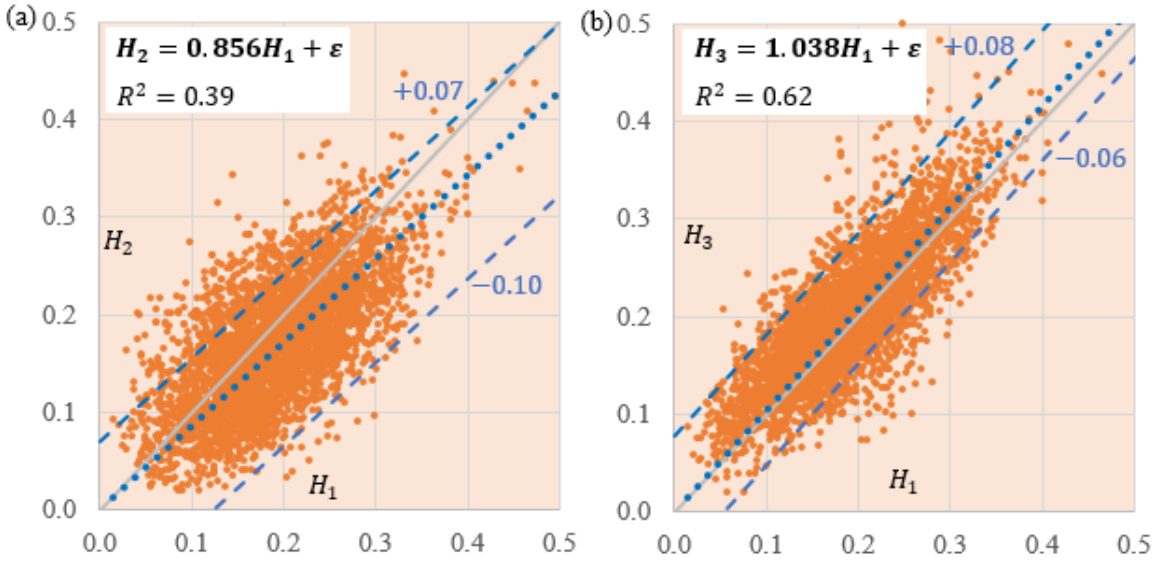


Figure 3.45 Scatterplots of (a) H_1 - H_2, and (b) H_1 - H_3 for the South Tohoku site.

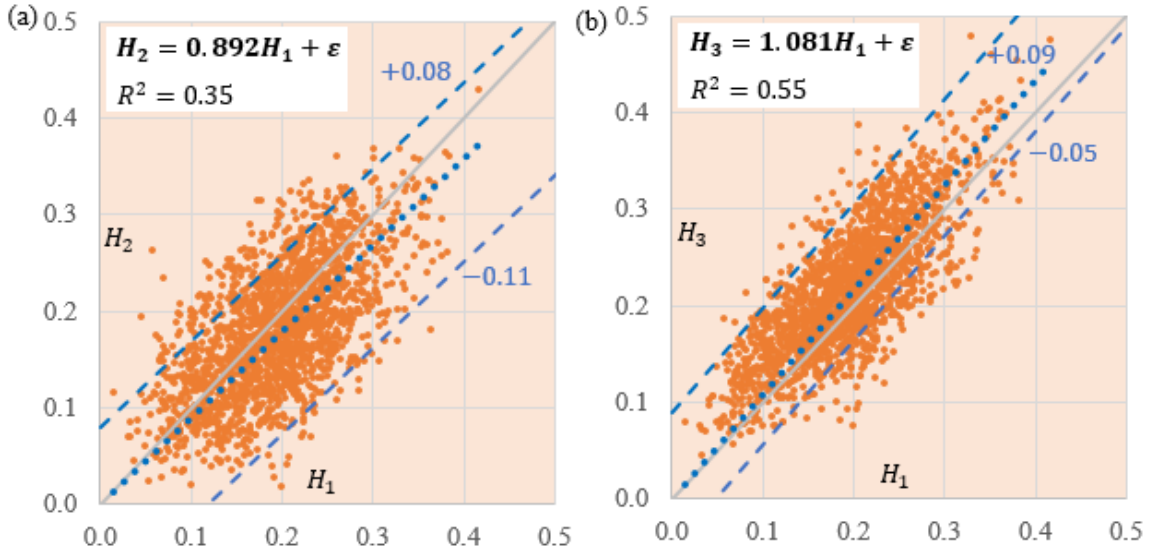


Figure 3.46 Scatterplots of (a) H_1 - H_2, and (b) H_1 - H_3 for the Hokkaido site.

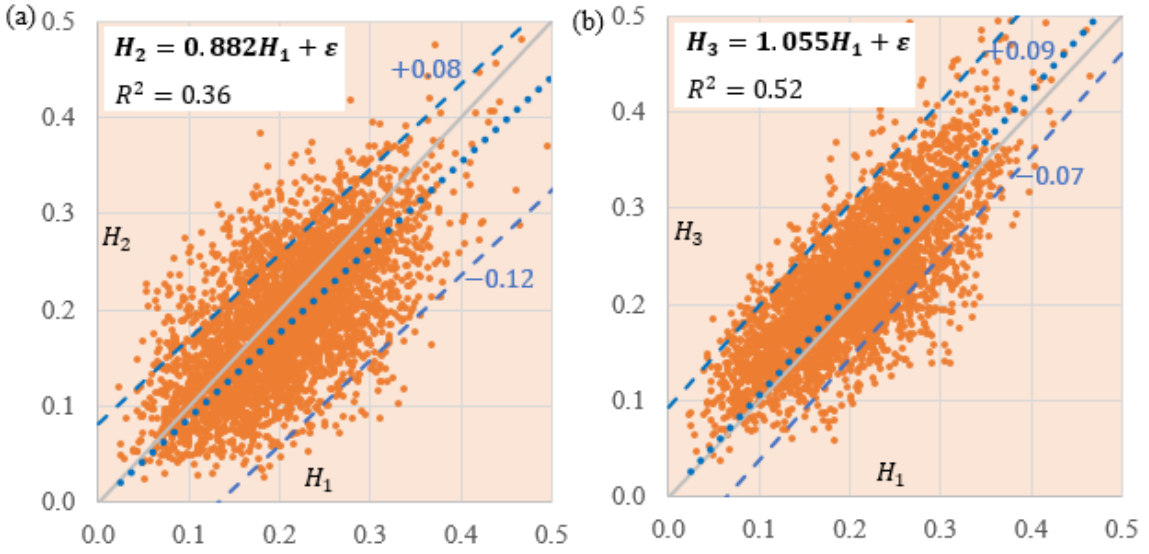


Figure 3.47 Scatterplots of (a) H_1 - H_2, and (b) H_1 - H_3 for the Kyushu site.

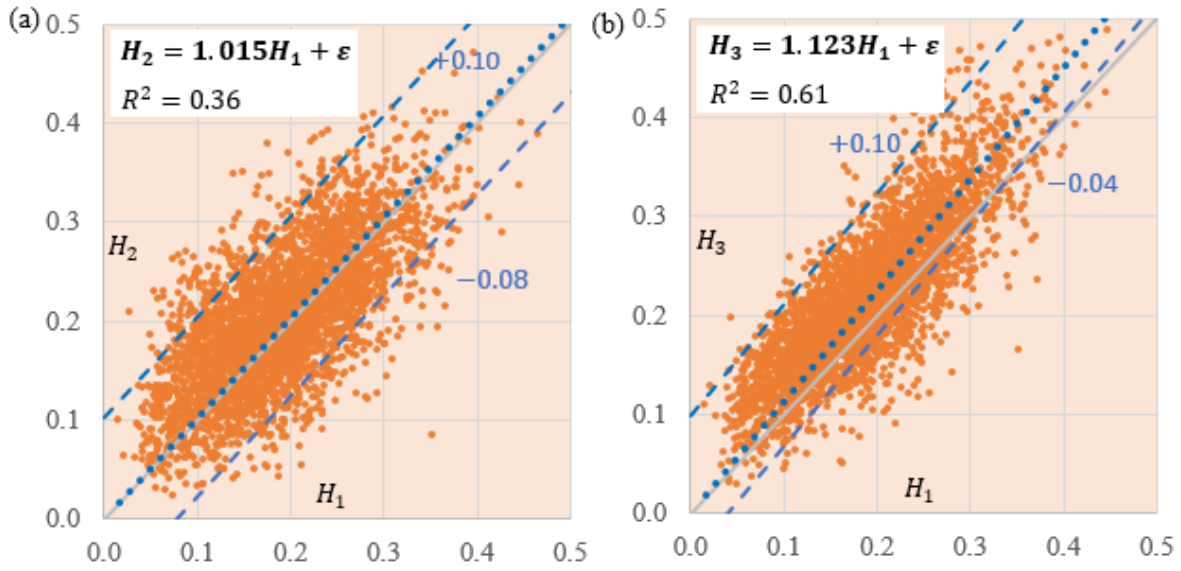


Figure 3.48 Scatterplots of (a) H_1 - H_2, and (b) H_1 - H_3 for the Shikoku site.

3.4.7 Summary of the results

The two-dimensional discrete wavelet transform using a decimated sampling method generates three sets of directional fluctuations (wavelets coefficients) that contain insufficient information for adequately estimating the stable parameters of the marginal distributions. On the other hand, the undecimated method of sampling generates samples of equal size at all scales creating over-complete, redundant samples that cause large computational effort. A combined method of sampling was used in this study that samples wavelet coefficients in an undecimated manner from the discrete approximation of the original field at each scale. The sampling considers wavelet coefficients (rainfall fluctuations) that correspond to the rain area only (conditional sampling).

In five analysis sites located in Japan with dimensions $256\,\text{km} \times 256\,\text{km}$, multiscale sampling was carried out with the purpose of stochastically characterizing the fluctuation processes and the possible scale-to-scale dependence of their statistical properties. The empirical pdf of the wavelet coefficients portray a large mass of values around the center and heavy tails. The latter feature is a direct consequence of the high frequency of extreme variations. These characteristics can be modeled by symmetric α-stable distributions, which are function of a characteristic exponent (α) and a scale parameter (γ). Because the characteristic exponent defines the rate of decay of the tails of the pdf, it has the potential to characterize the degree of variability of rainfall fields.

The properties of the symmetric α-stable distributions were revised with the purpose of designing a method of estimation that is able to overcome the difficulties that come naturally in the samples of rainfall fluctuations. Two methods were proposed: the FAM method and TSM method. Through Montecarlo simulations, the FAM method showed an accuracy that is comparable to the RCF method while employing less computational time. This and the other existing methods normally present problems when trying to estimate the parameters of

samples that contain large number of zeros at high resolutions. The TSM method, on the other hand, proved to have the capacity to overcome this issue.

The scale-to-scale dependence of the estimated parameters were analyzed. Independently at each direction, the estimators of the characteristic exponent α showed an almost invariant behavior between the 1-km and the 16-km resolutions (reaching the 32-km resolution in several cases). At the 32-km and 64-km resolutions, the estimated characteristic exponent α tends to have larger values than the ones estimated at higher resolutions. This situation can be caused by the insufficient information of the small samples at these scales or simply because self-similarity does not hold up to these scales. The former claim might be more probable since the samples come from events that develop at larger atmospheric scales. The scale parameter showed a power-law behavior in the scale range between 1-km and 64-km for a large portion of the considered events. These two conditions suggest that **the processes of rainfall fluctuations can be qualified as self-similar in distribution**. This finding is one of the most remarkable results of this research since it implies that the probability distribution of fluctuations at any scale of representation can be estimated if the distribution at a specific scale is known. Additionally, the fact that the process used to decompose the rainfall field into multiscale fluctuation fields is reversible allows to devise a model for downscaling rainfall fields with low resolutions in order to obtain rainfall fields at high resolutions (Chapter 4).

3.4.8 Link between the statistical model and the storm environment

Even though the samples of wavelet coefficients are obtained using independent orthonormal bases, the parameters of the distributions showed some degree of dependence with the parameters of the distributions at other directions, as shown in Section 3.4.6. This finding may reduce the number of unknown variables, which is important for the construction of stochastic disaggregation models. This comparison, additionally, helped drawing the following conclusions:

- Since the characteristic exponent $\bar{\alpha}_i$ appears to be quite similar in the three directions, the degree of variability (characterized by this parameter) does not seem to depend on factors that impose directionality in the storm (e.g., orography, wind speed or wind direction). It should be highlighted that this observation holds for all types of rainfall.
- The self-similarity index H_i characterizes the rate with which the magnitude of the fluctuations increase from one scale to larger scales. The results shown in Figures 3.44 to 3.48 depict a considerable difference between the values obtained at different directions suggesting that this parameter may be affected by factors that impose directionality in the storm.
- The scale parameter $\gamma_{1,i}$ gives an idea of the mean value of the fluctuations since it is linearly related to the mean of the absolute value of the fluctuations, as

demonstrated in sections 3.2.5.4 and 3.2.5.5. The values of these parameters appear to be somewhat related to the mean intensity of the conditional rainfall. This assertion, though, may not always be true. Naturally, extreme values of intensity in an event with a high degree of variability would cause large values of fluctuations. However, if the degree of spatial variability is low, large values of intensity don't necessary cause large values of fluctuations. The results shown in Figures 3.34 to 3.43 depict a trend between the scale parameters of different directions. This trend, however, is different from one analysis site to another. This condition suggests that the magnitude of the scale parameter is influenced by factors that impose directionality in the storm, that at the same time are caused by local conditions, i.e., orography.

From the observations made above, it can be concluded that the characteristic exponent and the self-similarity index can be dependent on the degree of variability of the storm. The following analysis is aimed to support or discard this claim.

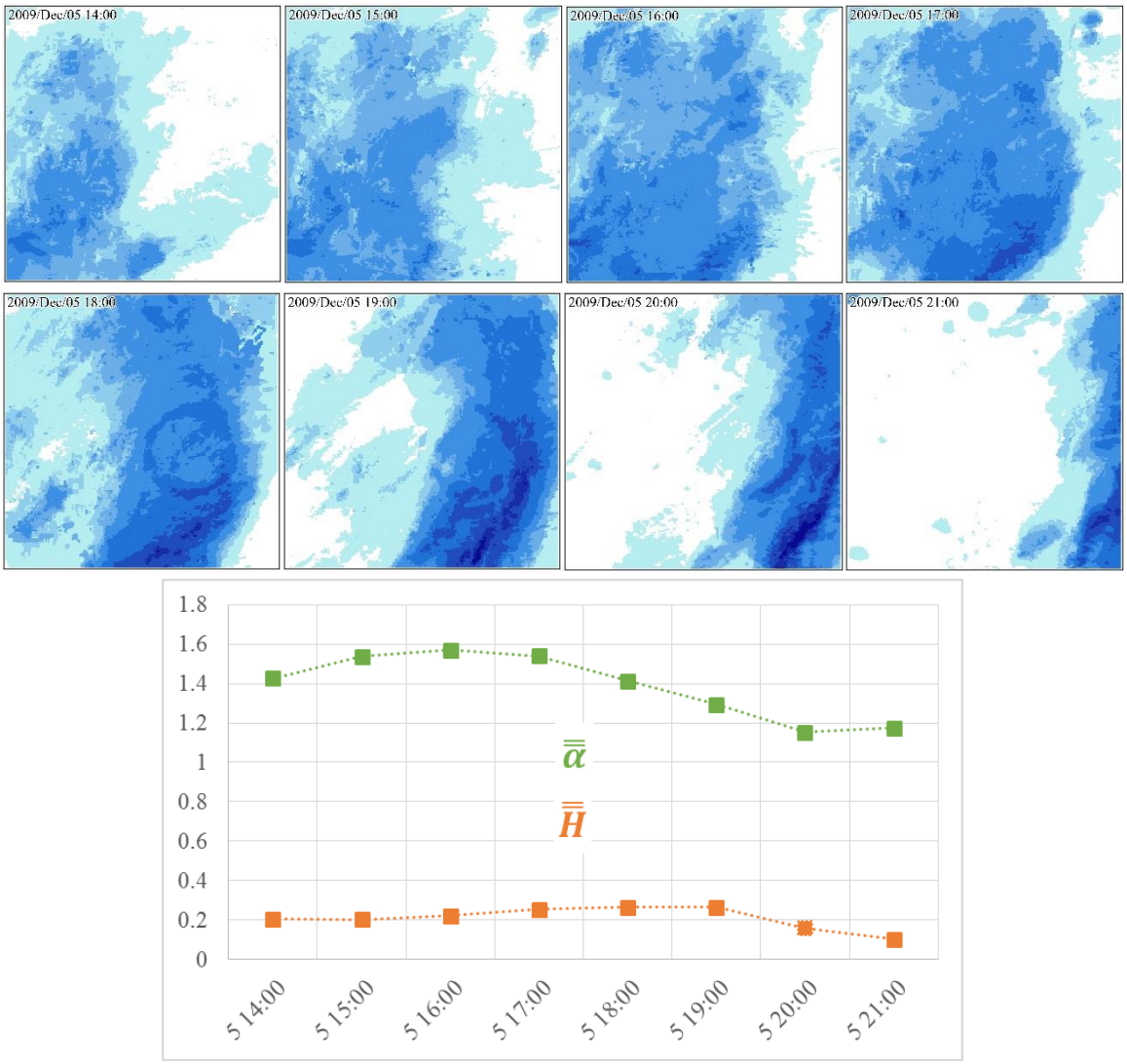
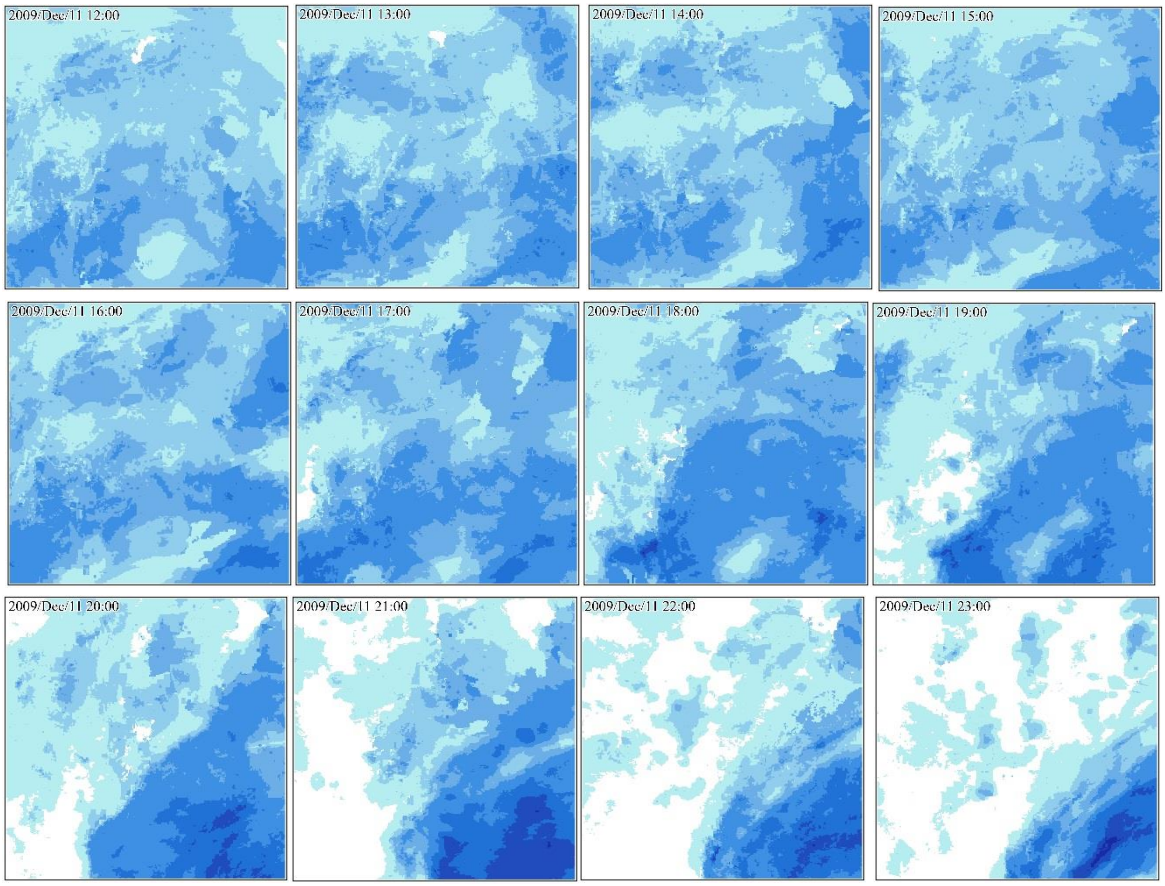


Figure 3.49 Time evolution of parameters $\bar{\bar{\alpha}}$ and $\bar{\bar{H}}$ (1)

The consecutive events correspond to a storm that happened on the Kanto analysis site on December 5th, 2009.

Let $\bar{\bar{\alpha}}$ and $\bar{\bar{H}}$ be the arithmetic mean of the values correspondent to the three directions for the characteristic exponent and the self-similarity index, respectively. Figures 3.49, 3.50 and 3.51 show the time-evolution of $\bar{\bar{\alpha}}$ and $\bar{\bar{H}}$ for three storms that happened over the Kanto analysis site.



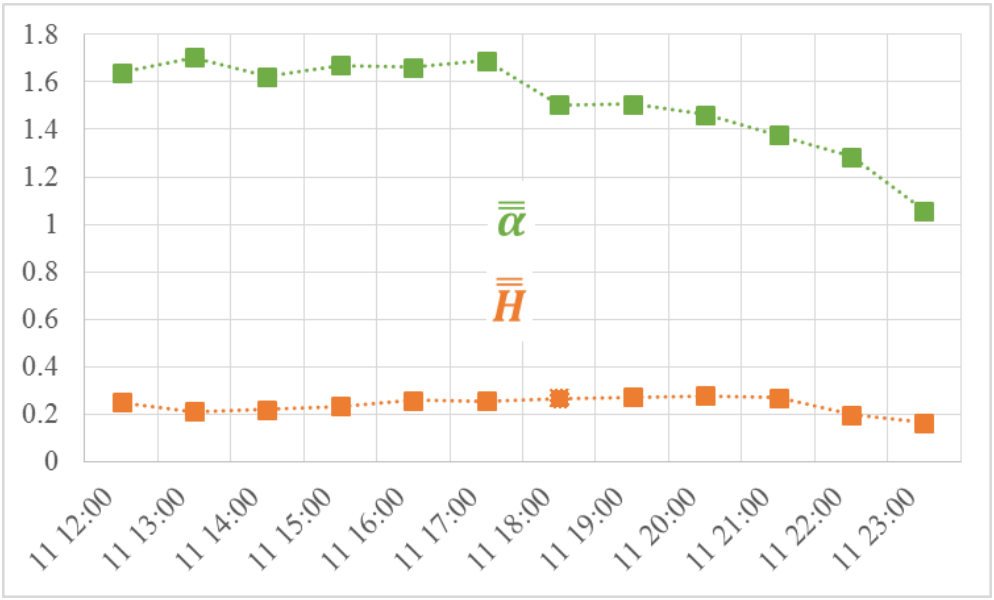


Figure 3.50 Time evolution of parameters $\bar{\bar{\alpha}}$ and $\bar{\bar{H}}$ (2)

The consecutive events correspond to a storm that happened on the Kanto analysis site on December 11th, 2009.

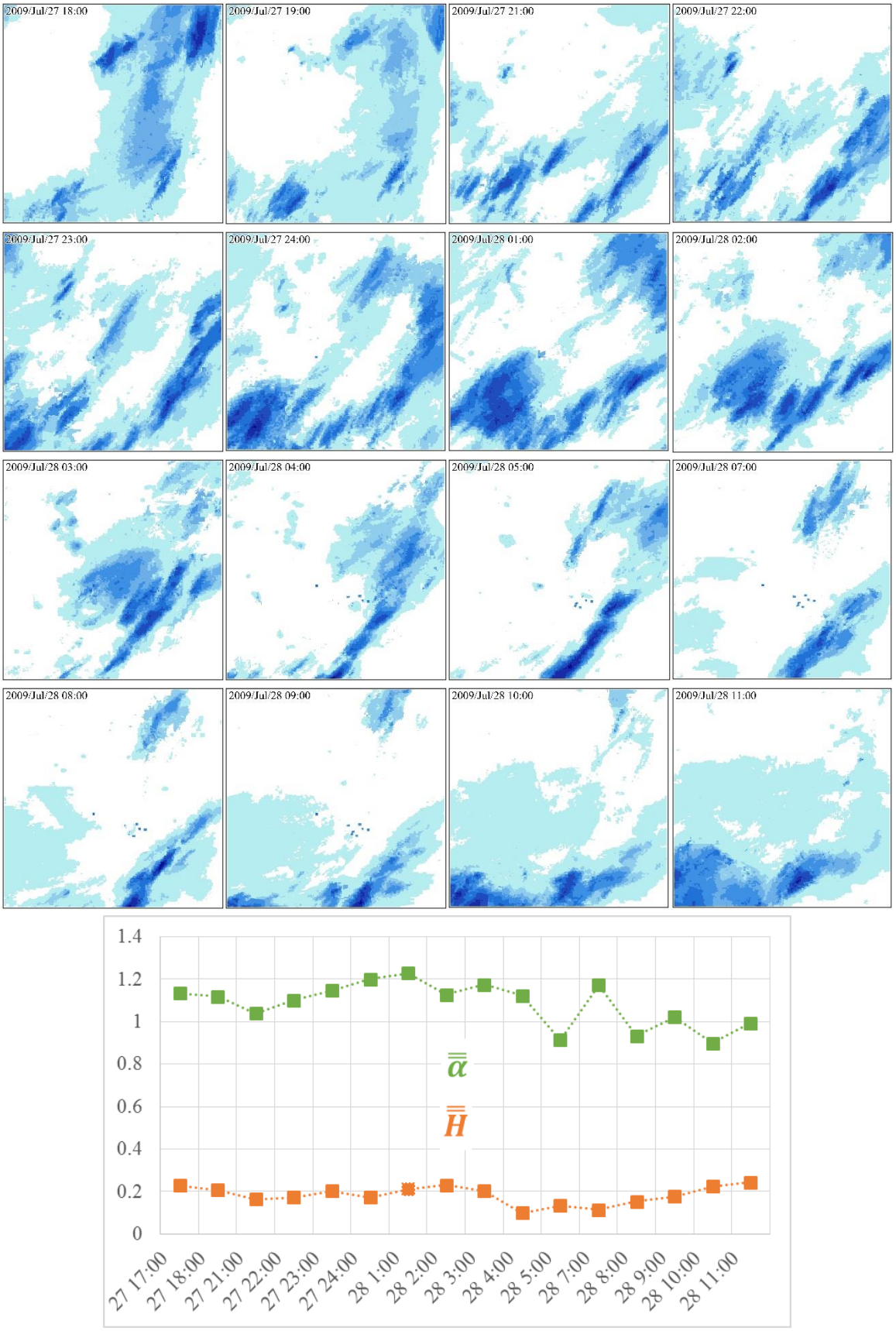


Figure 3.51 Time evolution of parameters $\bar{\bar{\alpha}}$ and $\bar{\bar{H}}$ (3)

The consecutive events correspond to a storm that happened on the Kanto analysis site on July 27th and July 28th, 2009.

The observation of the time-evolution of parameter $\bar{\bar{H}}$ does not show a direct relationship between the value of this parameter and the appearance of sudden changes of intensity (clustering of rainfall intensity or formation of bands). One reason for this condition can be the fact that the values in the three directions can be quite different. Therefore, even though it would seem logical that this parameter is related to the degree of variability, its value is also influenced by factors that impose directionality in the storm, as mentioned before.

The storms shown in Figures 3.49 and 3.50 depict initially events with mild spatial changes of intensity, which yield values of $\bar{\bar{\alpha}}$ greater than 1.4. However, the values of $\bar{\bar{\alpha}}$ tend to decrease below 1.4 when clusters or bands of high intensity form, producing sudden changes of intensity in short distances (extreme values of fluctuations). The storm shown in Figure 3.51 reveals the formation and rapid dissipation of multiple rainfall intensity clusters inside the analysis site. In this situation, the time-evolution of parameter $\bar{\bar{\alpha}}$ depicts values always below 1.4. It can be concluded that the higher the intensity of the appearing clusters (bands), the smaller the value of $\bar{\bar{\alpha}}$. **The hypothesis that the characteristic exponent α is related to the degree of variability is supported by this analysis**.

The estimation of the characteristic exponent α of the α-stable distribution of multiscale fluctuations for a given rainfall event, therefore, should be done by making some physical quantification of the degree of variability. As it was pointed out in Section 3.4.1, the two types of rainfall are generated by either stratiform clouds or clouds formed by convective activity. Stratiform rainfall is a consequence of the slow ascent of unsaturated air, it is widespread, shows steady intensity (mild spatial fluctuations) and has long duration. The trait that differentiates convective formation of clouds from the formation of stratiform clouds is **atmospheric instability**. As a consequence, convective rainfall has short duration, limited extent, and may have much higher intensity than stratiform rainfall. Intuitively, it can be said that values of α greater than 1.3 or 1.4 correspond to stratiform rainfall, while lower values of α are caused by convective activity. This conclusion leads to believe that measurements of the instability of the environment during the formation and precipitation of the rainfall event could be linked to the magnitude of parameter α.

Physical models (e.g., general circulation models, regional climate models) allow to handle atmospheric variables like temperature and relative humidity. Knowing these variables at several pressure levels allows to estimate parameters that give an idea of the atmospheric instability that causes the ascent of moist air, leading eventually to convective rainfall. Convective Available Potential Energy (CAPE) is a measure of atmospheric instability (stability) based on the potential energy that is transformed into kinetic energy and eventually develops into cumulus clouds. This energy is computed by integrating the difference of virtual temperature of the idealized rising air parcel and the surrounding atmosphere between the level of free convection and the level of neutral buoyancy. Larger values of CAPE would imply a more extreme rainfall event in terms of sudden changes of intensity both in spatial and temporal scales. Consequently, rainfall intensity (rainfall extremes) can be roughly related in a log-linear manner to CAPE (Lepore et al., 2015).

However, large values of CAPE are not necessarily always caused by convective activity. The ascent of a moist air parcel may be inhibited by a stable boundary layer, in which case another parameter, called Convective Inhibition (CIN), describes the energy needed by the rising parcel to overcome the boundary layer, and eventually reach the level of free convection. Alfieri et al. (2008) proposed a way to filter the large values of CAPE using a threshold of CIN in order to differentiate convective and stratiform rainfall. In that study, convective events are identified if CAPE ≥ 400 kJ/kg and CIN≥ -5 kJ/kg. Thus, future work related to the stochastic multiscale behavior of rainfall fluctuations using α-stable distributions could analyze the possible relationship between the characteristic exponent α and the classification-method by Alfieri et al. (2008) using CAPE and CIN. Being able to estimate this key parameter from physical models may be quite beneficial for the stochastic disaggregation of rainfall field using schemes based on discrete wavelet transforms.

Another way to differentiate stratiform and convective rainfall can be accomplished by analyzing the measurements from the multiple sensors on board of the Tropical Rainfall Measuring Mission (TRMM) satellite. This satellite carries the Precipitation Radar (PR), which provides detailed vertical distribution precipitation-sized particles inside systems, the TRMM Microwave Imager (TMI), which provides information on the vertical integrated ice and water path, the Visible and Infrared Scanner (VIRS), which provides information on cloud-top temperature and reflectance, and the Lightning Imaging Sensor (LIS), which estimates lightning-flash rates. Liu et al. (2008) devised an event-based methodology to define Precipitation Features (PFs) by grouping contiguous pixels with the PR 2A25 near-surface rain greater than 0 mm/hr, and classifying according to fixed thresholds set for different products (versions, levels of processing) of the other aforementioned sensors. This analysis was applied to TRMM measurements, which started in 1998, and the results are stored as a database administrated by the University of Utah. The identification of PFs through this methodology allows to define area coverage, volumetric rain rate, fraction of raining area, fraction of convective rain area, fraction of convective rain volume, and maximum height of the feature. This kind of information lead to the analysis of rain types inside different precipitation systems at different atmospheric scales located in different longitudes and latitudes. Examples include the studies of Liu and Zipser (2009), Jiang and Zipser (2010), Liu (2011), Thatcher et al. (2012), Yokoyama and Takayabu (2012), and Zhou et al. (2013), among others. In this manner, being able to identify the type of precipitation system and the consequent type of rainfall (type of convective formations or stratiform rainfall) from satellite observations could be used to estimate the characteristic exponent α if a fair relationship between this parameter and characteristics of the PFs (e.g., stratiform rain ratio, SRR) can be found.

4. STATISTICAL SPATIAL DISAGGREGATION OF RAINFALL

4.1 BACKGROUND

Different kinds of numerical schemes have been proposed for the disaggregation of low-resolution datasets into realistic high-resolution rainfall fields, which should be able to preserve the large-scale components of the storm while at the same time depict the small-scale details and distinctive spatial structure of rainfall. Examples include models based on inverse wavelet transforms (Perica and Foufoula-Georgiou, 1996b), multifractal processes (Menabde et al., 1997, 1999; Deidda, 2000), autoregressive processes (Guillot and Lebel, 1999; Bouchaud et al., 2000; Rebora et al., 2006), fractal interpolations (Tao and Barros, 2010; Nogueira and Barros, 2016), exponential Langevin-type models (Sapozhnikov and Foufoula-Georgiou, 2007), dictionary-learning for sparse representation using high-resolution rainfall patches (Ebtehaj et al., 2012) and variational downscaling via ℓ_1-norm regularization (Foufoula-Georgiou et al., 2014). The development of such schemes is triggered by the need to reduce the uncertainty in hydro-meteorological modeling that is typically caused by low resolution data and/or small scale variations of the rainfall fields, as reported in the studies of Smith et al. (2004), Schuurmans and Bierkens (2007), Younger et al. (2009), Arnaud et al. (2011), Liu et al. (2012), Song et al. (2015), and Weijian et al. (2015), among others.

Ferraris et al. (2003) compared three types of stochastic models based on multifractal processes, autoregressive processes, and point processes. The aim was to compare the ability of each of those models to reproduce various statistical properties of the rainfall fields. The results of that study showed that even though all methods yielded similar acceptable statistical results, they all had problems to reproduce specific spatial characteristics at high-resolutions. Since then, the developing of stochastic downscaling schemes has focused on enhancing the ability to reproduce the particular small-scale variability of rainfall while maintaining a small number of parameters (Nogueira and Barros et al., 2016, present a review on the developments of the last 15 years in statistical downscaling).

Another challenging issue that comes when trying to reproduce rainfall fields is intermittency. Because it is known that the presence of a large number of zeroes influences the scaling behavior of the parameters/generators (Harris et al., 1996; Verrier et al., 2010), most studies showing the performance of downscaling schemes focused on events with limited non-zero areas. In this study, however, we have demonstrated that the analysis of the fractality of rainfall fluctuations is not affected by the positive probability of the rainfall field being zero as it focuses on variations of intensity instead of rainfall intensity itself. Also, even though the presence of a large number of zeroes is a problem when trying to characterize the scaling behavior (Ebtehaj and Foufoula-Georgiou, 2011), the TSM method introduced in this study proved to successfully overcome this issue.

It was highlighted by Tao and Barros (2010) that one of the aims of statistical downscaling of rainfall fields should be devising a scheme without case-specific constraints and/or calibration requirements. Nevertheless, there has been an on-going discussion on whether the parameters of these schemes can be universal or dependent on physical quantities of the storm environment (Over and Gupta, 1994; Schertzer and Lovejoy, 2007). Undoubtedly, the marked differences in the results of the analyses of different types of rainfall with different stochastic models encourages the avid interest of linking the stochastic parameters to underlying atmospheric properties like atmospheric stability, wind speed and orography (Perica and Foufoula-Georgiou, 1996b; Nogueira et al., 2013; Nykanen, 2008). In this chapter, we intend to test the applicability of using the characteristic exponent of α-stable distributions to characterize the degree of variability of rainfall fluctuations. The approach is to construct an algorithm, which disaggregates low-resolution rainfall data, based on the inverse operation that was used to extract rainfall fluctuations at multiple scales. In this study, the high-resolution rainfall fields were decomposed into low-resolution fields (via a low-pass filter) and overcomplete fluctuation-fields (via a high-pass filter), which is the two-dimensional multiscale Discrete Wavelet Transform (DWT).

In a two-dimensional framework, the rainfall fluctuations, expressed by the wavelet coefficients, are separated in three directional sets that capture the variations of the field in the east-west direction, the north-west direction and simultaneously in both directions (diagonal direction). At a certain scale m, the combination of the low-resolution field $\bar{X}_m$ and the three directional sets of wavelet coefficients $X'_{m,i}$ can be combined to reconstruct the field at scale $m-1$. This operation is known as the Inverse Discrete Wavelet Transform (IDWT), as shown in Figure 4.1 (Mallat, 2009). The reconstruction can also be expressed as the sum of four Kronecker products:

$$\bar{X}_{m-1} = \bar{X}_m \otimes \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} + X'_{m,1} \otimes \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} + X'_{m,2} \otimes \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix} + X'_{m,3} \otimes \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \tag{4.1}$$

In Eq. 4.1, the elements in the two-dimensional vectors $\bar{X}_m$ and $X'_{m,\mathrm{i}}$ are increased by a factor of $2 \times 2 = 4$, and then, added to yield the field $\bar{X}_m$ at a higher resolution.

The algorithm of spatial disaggregation of rainfall detailed in this chapter uses iteratively the IDWT for obtaining rainfall fields at high resolutions. Even though the exact values of the directional sets $X'_{m,i}$ are not known, in the previous chapter the statistical characterization of these sets showed that symmetric α-stable distributions are a good fit. A multiscale analysis of the distributions determined that the characteristic exponents (α_i) show a somewhat steadiness within a defined scale-range. Additionally, the same analysis revealed that in each of the three directions the scale parameter ($\gamma_{m,i}$) obeys a power-law behavior defined by a single parameter known as the self-similarity index (H_i). This kind of scale-to-scale dependence, known as self-similarity, implies that if the stable-parameters

$\{\alpha_i, \gamma_{1,i}, H_i\}_{i=1,2,3}$ are known, the distribution of local fluctuations can be determined at all scales for which the power-law behavior holds. The following sections detail not only the structure of the disaggregation algorithm, but also the way in which the values of the distribution of wavelet coefficients are spatially distributed in order to reproduce the characteristic spatial features of the rainfall field.

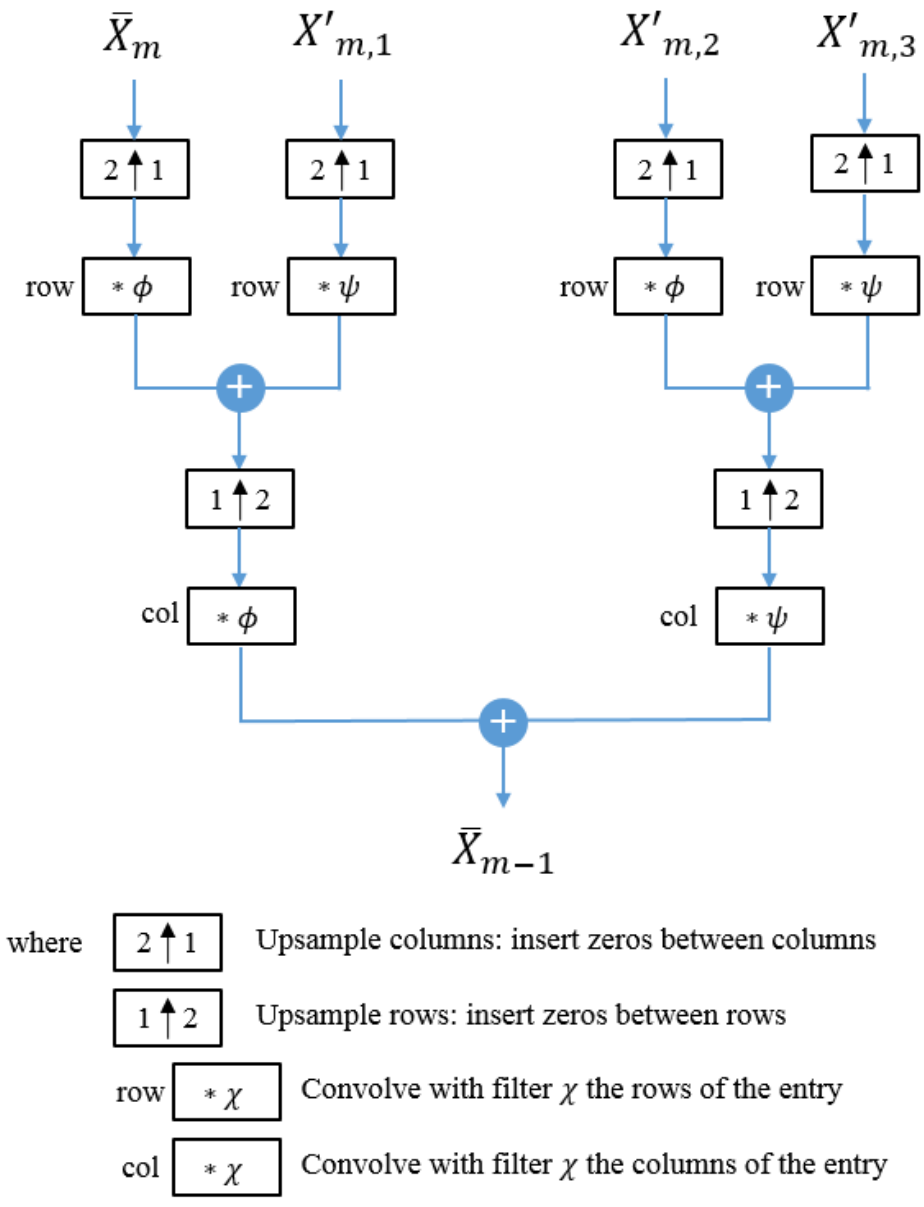


Figure 4.1 Filter bank for the IDWT

Reconstruction of $\bar{X}_{m-1}$ by inserting zeros between the rows and columns of $\bar{X}_m$ and $X'_{m,i}$ and convolving the outputs with the one-dimensional scaling function ϕ and the wavelet transform ψ.

4.2 MAIN STRUCTURE OF THE ALGORITHM

The disaggregation algorithm has as main inputs a low-resolution rainfall-intensity (LR) field, with a resolution correspondent to the scale m_{in}, and the stable parameters

$\{\alpha_i, \gamma_{1,i}, H_i\}_{i=1,2,3}$. The flow chart of the main structure of the algorithm is presented in Figure 4.2.

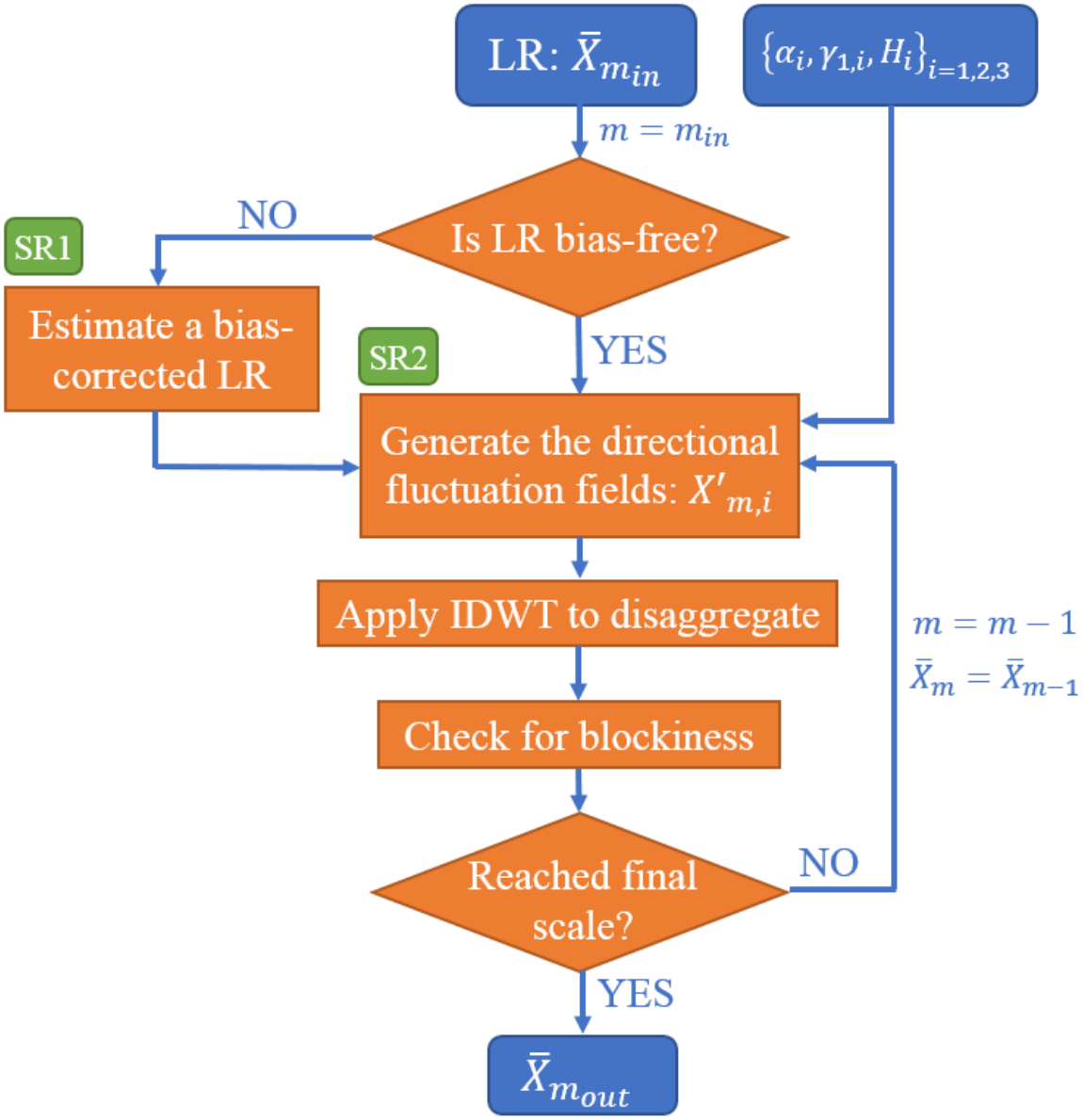


Figure 4.2 Main structure of the disaggregation algorithm

The algorithm checks initially if the LR field is biased or not. For this purpose, a subroutine (SR1) is used to estimate the mean rainfall intensity of the analysis site at the lowest resolution of representation. A second subroutine (SR2) is used to spatially distribute the rainfall fluctuations $X'_{m,i}$ at each scale before applying the IDWT. SR1 and SR2 are separately detailed in the next sections.

Because the typical resolution used in hydrological models is $1\ \text{km} \times 1\ \text{km}$, the final disaggregation scale $m = 0$ is the correspondent to that resolution. This assumption will also allow to verify the algorithm by comparing the algorithm's output to Radar-AMeDAS data.

4.2.1 Subroutine SR1

Since the LR field is usually an estimated product of rainfall intensity, it may be associated to some degree of bias. Consequently, it is necessary to devise a methodology for correction of that bias. Two situations are contemplated:

1. A previous bias-correction scheme for the LR field is not available.
2. A statistical analysis of the bias is available for LR field's historical data.

It is necessary to clarify that the bias correction scheme is not applied to each cell of the original LR field at scale m_{in}. Instead, the mean rainfall intensity $\bar{\bar{R}}$ of the analysis site is estimated/corrected. Note that if the dimensions of the analysis site are $L \times L$,

$$\bar{\bar{R}} = (4^m/L^2) \sum_{n_1=0}^{L/2^m} \sum_{n_2=0}^{L/2^m} \bar{X}_m(n_1, n_2) \tag{4.2}$$

for all possible scales m of representation. Therefore, $\bar{\bar{R}}$ should be equal at all scales.

4.2.1.1 Case 1: A previous bias-correction scheme is not available

Ebert et al. (2007) presented an analysis of accuracy (probability of detection and false alarm ratio) and bias assessment of several rainfall datasets derived from satellite-borne sensors and numerical weather prediction models. In that study, the rainfall products were compared to local weather radars and dense rain gauge networks in the U.S.A, North-Western Europe and Australia. Numerous studies compare satellite-derived products to local rain gauge networks and the mathematical solutions for spatially distributing the bias are quite diverse, some of which rely on geographic and orographic characteristics. Nevertheless, this section intends to assume that no information about the bias of the LR field is available, and proposes to estimate $\bar{\bar{R}}$ as a function of one rain gauge measurement, its relative position in the storm and the degree of variability expressed by the characteristic exponent of the distribution of rainfall fluctuations.

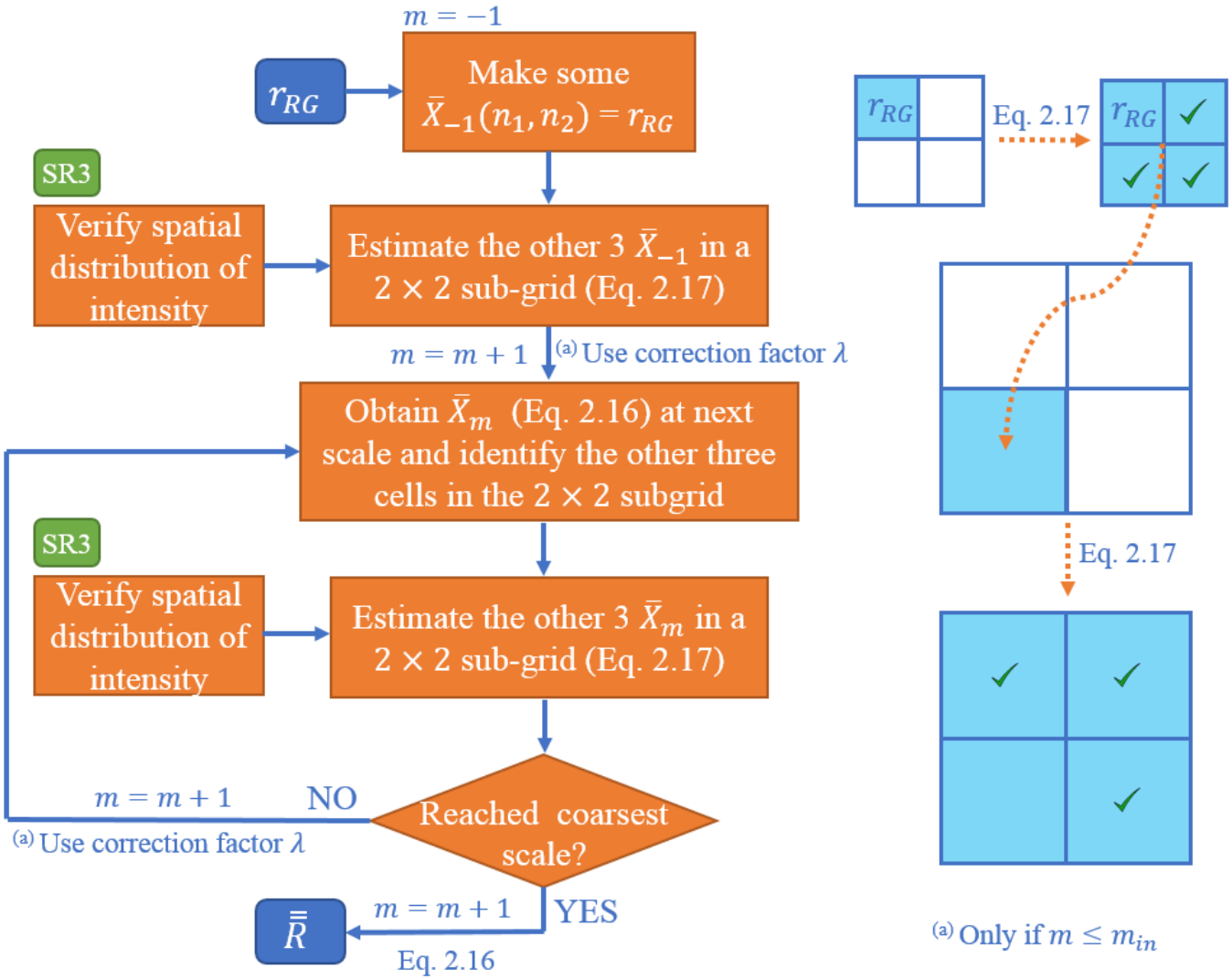


Figure 4.3 Flowchart for the estimation of $\bar{\bar{R}}$ (SR1 case 1)

This flowchart uses as input the rainfall intensity measurement (r_{RG}) of one rain gauge.

SR1 case 1 starts by assuming that the rainfall intensity measurement, r_{RG}, of one rain gauge station located inside the analysis site is equal to the intensity of the cell containing it at a scale $m = -1$, which corresponds to a 0.5-km resolution. This assumption is supported by the fact that the spatial variation of intensity within a 0.5 km × 0.5 km area (cell) may be negligible. Utilizing the expressions in Eq. 2.17, if the values of $X'_{0,1}$, $X'_{0,2}$, and $X'_{0,3}$ are known, and if one of the four values of $\bar{X}_{-1}$ that form a 2×2 sub-grid at scale $m = -1$ is assumed to be equal to r_{RG}, then, the other three values of $\bar{X}_{-1}$ can be obtained. Subsequently, the four values of intensity $\bar{X}_{-1}$ are averaged as in Eq. 2.16 to obtain the rainfall intensity $\bar{X}_0$ at scale $m = 0$. The cell containing the computed value of $\bar{X}_0$ is part of a 2×2 sub-grid at scale $m = 0$, and the unknown values of the other three cells can be estimated if the values of $X'_{1,1}$, $X'_{1,2}$, and $X'_{1,3}$ are known (Eq. 2.17). This procedure can be repeated iteratively until reaching the largest scale, which for the size of the analysis sites is the scale $m = 8$. The flowchart of SR1 for the estimation of $\bar{\bar{R}}$ is shown in Figure 4.3. An additional subroutine (SR3) is needed for verifying the correct spatial distribution of the four estimated values of the cells in the 2×2 subgrids.

A rain gauge station of the AMeDAS network was selected in each of the analysis sites. Then, between 2006 and 2009, all the events in which the rain gauge stations measured an hourly-accumulated intensity larger than a predetermined threshold were selected for carrying out the estimation of $\bar{\bar{R}}$ according to the process detailed above. The name of the rain gauge station, the value of threshold and the number of selected events are shown in Table 4.1. By comparing the estimated values of $\bar{\bar{R}}$ and the actual values aggregated from the Radar-AMeDAS data, it was found that in a considerable number of events the values of $\bar{\bar{R}}$ were overestimated. For this reason, for all scales $m < m_{min}$ it was considered necessary to apply a correction factor (λ) that was found to be function of the relative position of the rain gauge in terms of intensity (s_{LR}), the magnitude of the rain gauge measurement (r_{RG}), and the mean of the three characteristic exponents of the marginal α-stable distributions of directional rainfall fluctuations, simply computed as $\bar{\bar{\alpha}} = (1/3)[\alpha_1 + \alpha_2 + \alpha_3]$. The correction factor λ has to be applied applied while computing the average of the 2×2 subgrid as shown in Figure 4.3.

Table 4.1 Number of events selected for estimating $\bar{\bar{R}}$

Site	Rain Gauge Station	Threshold (mm/hr)	Sample size
Kanto	Maebashi	4	347
South Tohoku	Niigata	4	379
Hokkaido	Sapporo	4	201
Kyushu	Miyazaki	8	552
Shikoku	Kochi	7	357

The values of λ can be obtained from a look-up table presented in Table 4.2.

Table 4.2 Values of λ as a function of s_{LR}, r_{RG}, and $\bar{\bar{\alpha}}$

$s_{LR} \leq 0.27$			
	$r_{RG} < 5$	$5 \leq r_{RG} < 15$	$r_{RG} \geq 15$
$\bar{\bar{\alpha}} < 1.22$	0.88	0.88	0.88
$1.22 \leq \bar{\bar{\alpha}} < 1.50$	1.00	0.74	0.68
$1.50 \leq \bar{\bar{\alpha}} < 1.68$	0.87	0.87	0.76
$\bar{\bar{\alpha}} \geq 1.68$	1.00	1.00	1.00
$0.27 < s_{LR} \leq 0.60$			
	$r_{RG} < 12$	$12 \leq r_{RG} < 15$	$r_{RG} \geq 15$
$\bar{\bar{\alpha}} < 1.30$	0.84	0.84	0.82
$1.30 \leq \bar{\bar{\alpha}} < 1.55$	0.95	0.95	0.74
$1.55 \leq \bar{\bar{\alpha}} < 1.68$	0.92	0.92	0.80
$\bar{\bar{\alpha}} \geq 1.68$	0.95	0.88	0.88
$0.60 < s_{LR} \leq 0.90$			
	$r_{RG} < 15$	$15 \leq r_{RG} < 21$	$r_{RG} \geq 21$
$\bar{\bar{\alpha}} < 1.00$	0.62	0.62	0.62
$1.00 \leq \bar{\bar{\alpha}} < 1.20$	1.00	1.00	0.80
$1.20 \leq \bar{\bar{\alpha}} < 1.40$	0.88	0.88	0.88
$1.40 \leq \bar{\bar{\alpha}} < 1.68$	0.89	0.89	0.89
$\bar{\bar{\alpha}} \geq 1.68$	0.94	0.94	0.94
$s_{LR} > 0.90$			
	$r_{RG} < 15$	$15 \leq r_{RG} < 30$	$r_{RG} \geq 30$
$\bar{\bar{\alpha}} < 1.20$	0.82	0.56	0.56
$1.20 \leq \bar{\bar{\alpha}} < 1.40$	0.84	0.84	0.84
$1.40 \leq \bar{\bar{\alpha}} < 1.68$	0.94	0.94	0.80
$\bar{\bar{\alpha}} \geq 1.68$	0.96	0.88	0.88

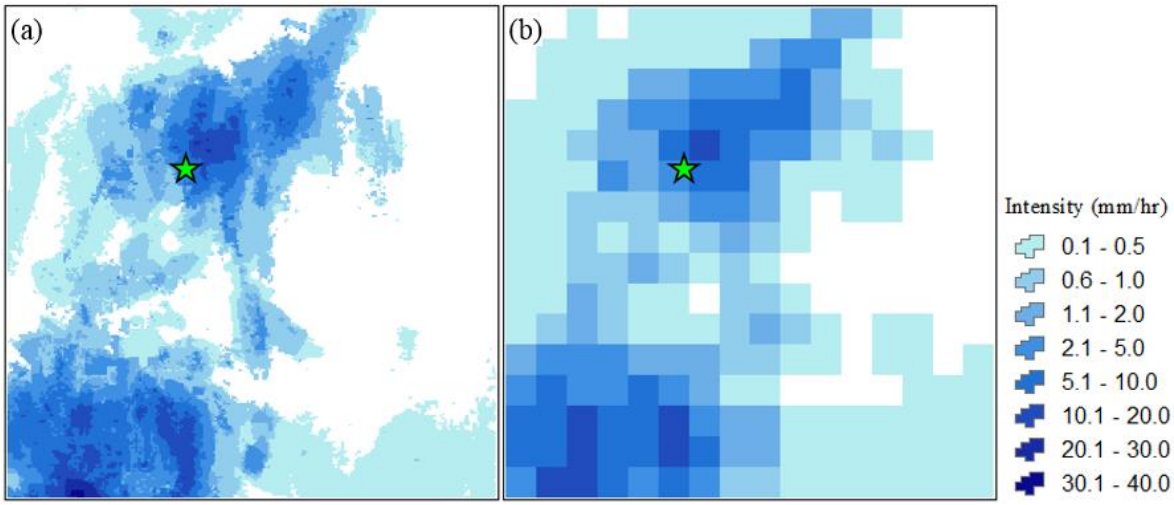


Figure 4.4 Radar-AMeDAS data for the event KAN_2009.07.24.03

(a) Original scale. (b) LR field computed from the original data after a transformation (aggregation) into scale $m = 4$. The green star represents the location of the Maebashi station. The largest value of the LR field is equal to 19.7 mm/hr, and the value of the cell in which the Maebashi station is located is equal to 3.2, yielding a value of $s_{LR} = 0.162$.

For each event, the parameter s_{LR} can be computed by dividing the rainfall intensity of the cell of the LR field that contains the rain gauge station by the largest value of intensity comprised in the LR field data. For example, consider the Radar-AMeDAS dataset of a particular event aggregated to the scale $m = 4$ (16-km resolution) to be the LR field. Figures

4.4 and 4.5 present respectively the values of s_{LR} for the KAN_2009.07.24.03 and KAN_2009.03.06.13 events considering the relative location of the Maebashi rain gauge station.

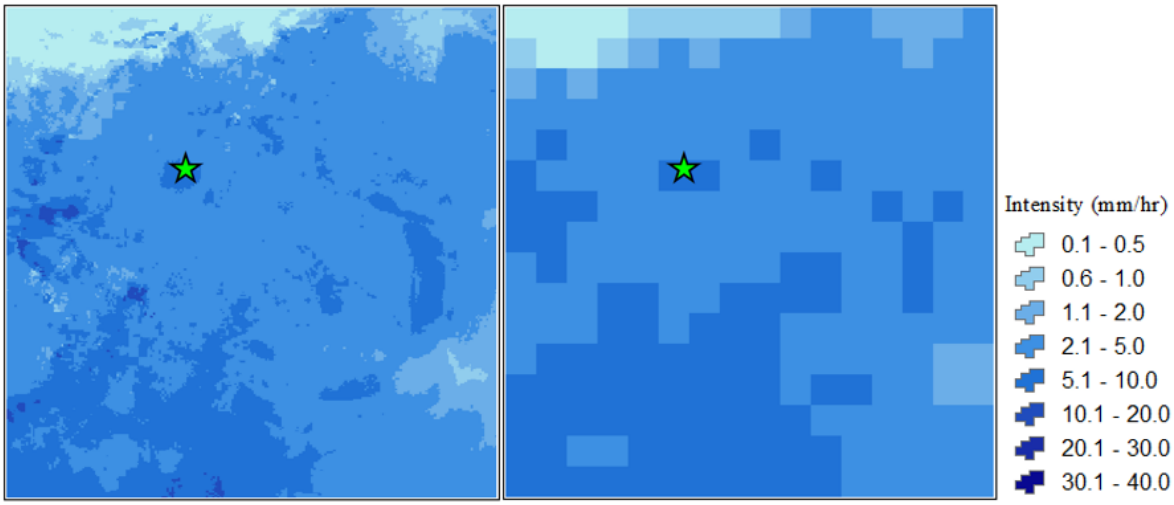


Figure 4.5 Radar-AMeDAS data for the event KAN_2009.03.06.13

(a) Original scale. (b) LR field computed from the original data after a transformation (aggregation) into scale $m = 4$. The green star represents the location of the Maebashi station. The largest value of the LR field is equal to 9.3 mm/hr, and the value of the cell in which the Maebashi station is located is equal to 6.9, yielding a value of $s_{LR} = 0.742$.

4.2.1.2 Case 2: A previous bias-correction scheme is available

The estimation of $\bar{\bar{R}}$ might be more accurate if a previous stochastic analysis of the LR field data was undertaken. In this study, we consider the use of rainfall estimates from the Tropical Rainfall Measuring Mission (TRMM) satellite. The TRMM project is a joint project between the National Aeronautics and Space Administration (NASA) and the Japan Aerospace Exploration Agency (JAXA), which provides different products through a combination of different satellites. The TRMM 3B42 real time product (RT), for example, uses the microwave rainfall estimates from the TRMM Microwave Imager (TMI) adjusted by the cloud vertical structure obtained from the Precipitation Radar (PR), as well as geosynchronous infrared data. Another product, the TRMM 3B42 research product (RP), includes a monthly rain gauge analysis which is used to calibrate the real time product. The TRMM 3B42 RP version 7.0 is used throughout this study, which are 3-hourly estimates of rainfall intensity with a $0.25° \times 0.25°$ resolution and a latitude coverage of 50°S - 50°N. For more details about the sensors and the algorithm of these rainfall products, refer to Huffman (2007, 2010).

The mean rainfall intensity in each analysis site was computed for the TRMM product and the Radar-AMeDAS datasets. Let us denote this values as $\bar{\bar{R}}_{\mathrm{TR}}$ and $\bar{\bar{R}}_{\mathrm{RA}}$, respectively. In order to make a fair comparison of both values, a clarification of the Radar-AMeDAS data is needed. For example, for a value of $\bar{\bar{R}}_{\mathrm{TR}}$ computed from the TRMM product

correspondent to 2009 December 2nd 3:00 GTM, the value of $\bar{\bar{R}}_{\mathrm{RA}}$ is computed from the Radar-AMeDAS data correspondent to 2009 December 2nd 3:30 GTM, which comprises hourly accumulated rainfall intensity between 2:30 and 3:30 GTM.

The values of $\bar{\bar{R}}_{\mathrm{TR}}$ and $\bar{\bar{R}}_{\mathrm{RA}}$ were compared with the purpose of observing the tendencies for overestimation or underestimation. The bias of $\bar{\bar{R}}_{\mathrm{TR}}$ with respect to $\bar{\bar{R}}_{\mathrm{RA}}$ is defined as

$$b_{TR} = \bar{\bar{R}}_{\mathrm{TR}} - \bar{\bar{R}}_{\mathrm{RA}}, \tag{4.3}$$

where positive values of b_{TR} imply overestimation of the TRMM product and negative values imply underestimation. The determination of bias was carried out in the five analysis sites between 2006 and 2009 for all those events in which the conditional mean rainfall intensity (mean rainfall intensity of the rain area), computed from the Radar-AMeDAS data, was larger than 1 mm/hr. The results were separated in intervals of the values of $\bar{\bar{R}}_{\mathrm{TR}}$ (Figure 4.6). The average value of b_{TR} in each interval is denoted by $\bar{\bar{b}}_{TR}$. For example, for the interval $0.3 \leq \bar{\bar{R}}_{\mathrm{TR}} < 0.5$, the value of $\bar{\bar{b}}_{TR}$ computed for the ensemble is equal to -0.35. Despite the fact that this number implies a tendency for underestimation, there is a number of events that returned positive values of b_{TR} in the interval $0.3 \leq \bar{\bar{R}}_{\mathrm{TR}} < 0.5$. Therefore, in addition to the values of $\bar{\bar{b}}_{TR}$, Figure 4.6 shows the probability of underestimation ($P(b_{TR} < 0)$) for each interval of $\bar{\bar{R}}_{\mathrm{TR}}$.

Generally, it was observed that regardless of the type of rainfall or location, the average tendency of the TRMM product is to underestimate if $\bar{\bar{R}}_{\mathrm{TR}} < 1.8$; and if $\bar{\bar{R}}_{\mathrm{TR}} \geq 1.8$, the average tendency of the TRMM product is to overestimate.

With the purpose of establishing. Therefore, denoting the standard deviation by

SR1 case 2 starts by estimating $\bar{\bar{R}}$ using the SR1 case 1 procedure without applying the correction factor λ. Let us denote this value by $\bar{\bar{R}}_{C1}$. Other parameters needed in this subroutine, which were already established in the bias assessment of TRMM products detailed above, are the average bias for each interval of $\bar{\bar{R}}_{\mathrm{TR}}$ (denoted b_{TR}), and the standard deviation for each interval of $\bar{\bar{R}}_{\mathrm{TR}}$ (denoted σ_{TR}). Then, the estimation of $\bar{\bar{R}}$ is done according to the flowchart presented in Figure 4.7. The value of $\bar{\bar{R}}_{\mathrm{TR}}$ returns a value of $P(b_{TR} < 0)$, a value of b_{TR} and value of σ_{TR}. Depending on the value of $P(b_{TR} < 0)$, $\bar{\bar{R}}_{C1}$ is compared with the bias-corrected value of $\bar{\bar{R}}_{\mathrm{TR}}$ plus/minus the standard deviation if applicable. For better comprehension, we give the following examples:

- If $P(b_{TR} < 0) \geq 0.7$, the minimum value that $\bar{\bar{R}}$ can take is the largest value between $\bar{\bar{R}}_{\mathrm{C1}}$ and the bias-corrected $\bar{\bar{R}}_{\mathrm{TR}}$, and the maximum value that $\bar{\bar{R}}$ can take is the smallest between $\bar{\bar{R}}_{\mathrm{C1}}$ and the bias-corrected $\bar{\bar{R}}_{\mathrm{TR}}$ plus a tolerance value equal to σ_{TR}.

The value of σ_{TR} is not subtracted from the bias-corrected $\bar{\bar{R}}_{\mathrm{TR}}$ for establishing the minimum because of the large probability of underestimation.

- If $P(b_{TR} < 0) < 0.3$, the minimum value that $\bar{\bar{R}}$ can take is the largest value between $\bar{\bar{R}}_{\mathrm{C1}}$ and the bias-corrected $\bar{\bar{R}}_{\mathrm{TR}}$ minus a tolerance value equal to σ_{TR}, and the maximum value that $\bar{\bar{R}}$ can take is the smallest between $\bar{\bar{R}}_{\mathrm{C1}}$ and the bias-corrected $\bar{\bar{R}}_{\mathrm{TR}}$. The value of σ_{TR} is not added to the bias-corrected $\bar{\bar{R}}_{\mathrm{TR}}$ for establishing the maximum because of the large probability of overestimation.

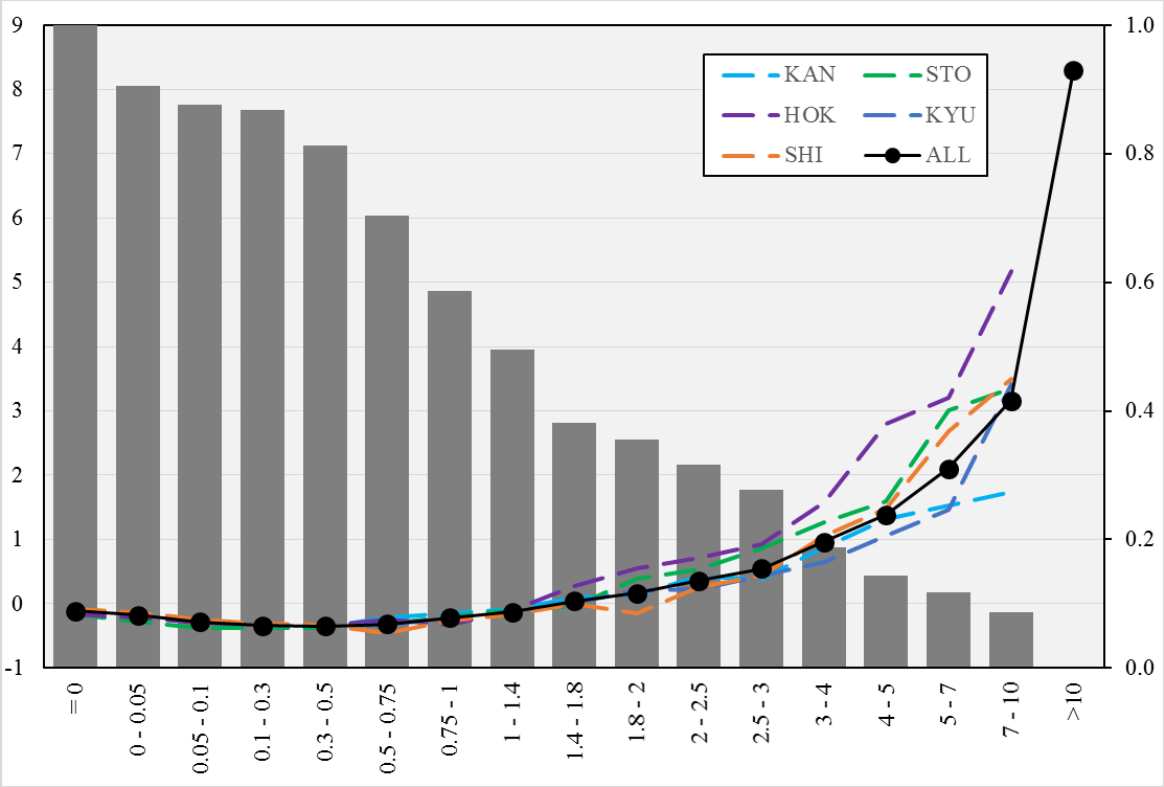


Figure 4.6 Analysis of TMM bias

Vertical left axis: Values of $\bar{\bar{\boldsymbol{b}}}_{\boldsymbol{TR}}$ (colored dashed lines for each analysis site, and black line for the ensemble) corresponding to intervals of $\bar{\bar{\boldsymbol{R}}}_{\mathbf{TR}}$. Vertical right axis: Probability of underestimation of the TRMM product.

4.2.2 Subroutine SR2

SR2 consists in the generation of the fluctuation fields at each scale during the disaggregation process. The main input of this subroutine are the stable parameters that define the marginal distributions of directional fluctuations. The flowchart of SR2 is detailed in Figure 4.8. Using the description of self-similarity, the value of the scale parameter corresponding to the scale m for which the disaggregation is being made ($\gamma_{m,i}$) can be calculated using $\gamma_{1,i}$ and H_i (Eq. 3.23). Then, by using the values of α_i and $\gamma_{m,i}$ in the CMS algorithm, random samples from a population with $S\alpha S(\alpha_i, 0, \gamma_{m,i}, 0)$ distribution are generated for each direction i. Subsequently, for all cells of the rainfall field at scale m (i.e., for all combinations of the pair

of location indexes n'_1, n'_2), values of $|X'_{m,i}|$ are randomly selected from the samples (in absolute value).

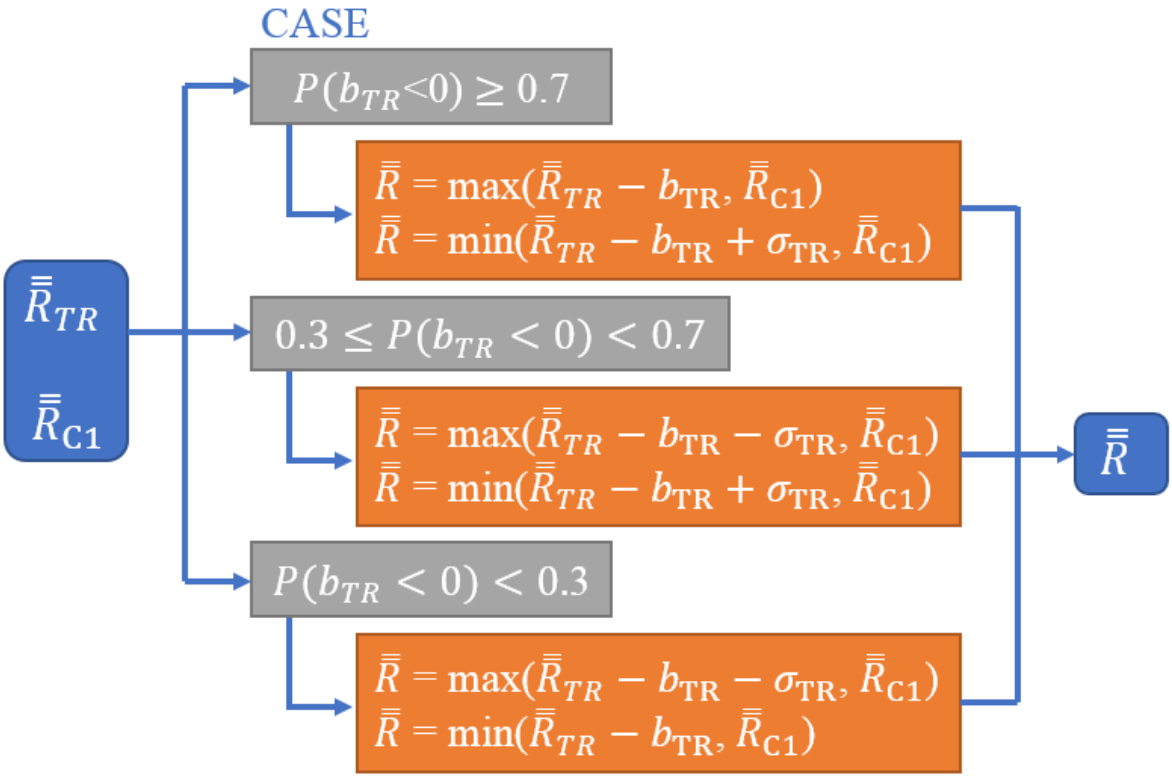


Figure 4.7 Flowchart for the estimation of $\bar{\bar{R}}$ (SR1 case 2)

The selection of the random samples, however, is within a defined interval that depends on the relative position of the cell in terms of intensity (s_m), the scale m and α_i. This dependence was found after an analysis of the relationship between $\bar{X}_m$ and the corresponding values of $|X'_{m,i}|$. For an event that occurred at the Kanto site, Figure 4.9 shows the representation of the rainfall field at scale $m = 1$ ($\bar{X}_1$) computed from the Radar-AMeDAS data. The same figure shows the corresponding directional sets of wavelet coefficients $X'_{1,i}$ (fluctuation fields). The three-dimensional representation of the rainfall field $\bar{X}_1$ shows high-intensity values in dark blue and low-intensity values in light blue. The three-dimensional representations of the fluctuation fields show the absolute values of wavelet coefficients in the east-west direction ($i = 1$) in shades of red, the wavelet coefficients in the north-south direction ($i = 2$) in shades of green, and the wavelet coefficients in the diagonal direction ($i = 3$) in shades of purple. Dark colors denote large absolute values of wavelet coefficients, and light colors denote values of wavelet coefficients approximating zero. Because areas depicting high-intensity rainfall may cause sudden variations of intensity, values of high-intensity in the rainfall field may share the same location as large absolute values of wavelet coefficients in the fluctuation fields.

Given the Radar-AMeDAS dataset of some event, the rainfall field is decomposed at all possible scales of representation into the sets $\bar{X}_m$ and $X'_{m,i}$. Then, for each value of intensity

(cell) of the representation of the rainfall field at scale m ($\bar{X}_m(n'_1, n'_2)$), the corresponding relative position in terms of intensity ca be computed as

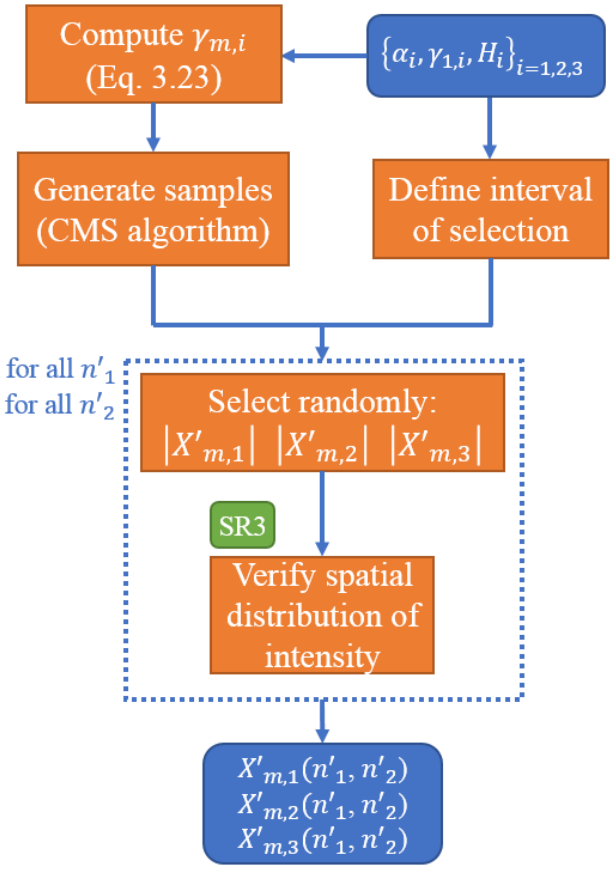


Figure 4.8 Flowchart for the generation of directional fluctuation fields (SR2 subroutine)

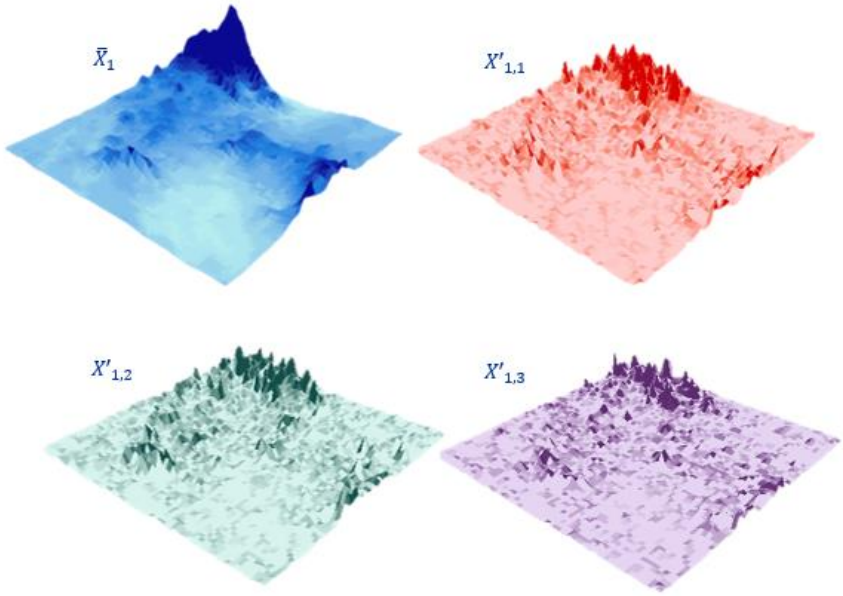


Figure 4.9 Example of a rainfall field represented at scale $m = 1$ and the corresponding directional fluctuation fields.

The dark shades in the fluctuation fields represent large absolute values of wavelet coefficients.

$$s_m(n'_1, n'_2) = \bar{X}_m(n'_1, n'_2)/\max(\bar{X}_m). \quad (4.4)$$

Separately, the parameters of the α-stable distributions $\{\alpha_i, \gamma_{m,i}\}_{i=1,2,3}$ of the wavelet coefficients comprised in the fluctuation fields $X'_{m,i}$ can be estimated at all possible scales of representation ($1 \leq m \leq 7$) using the TSM estimation-method. The cdf of the absolute values of the α-stable distribution ($F_{R_{m,i}}$) evaluated at each $|X'_{m,i}(n'_1, n'_2)|$ can be computed using the integral-based algorithms of Nolan (1997), which are function of the corresponding α_i and $\gamma_{m,i}$. The values of $s_m(n'_1, n'_2)$ and the correspondent values of $F_{R_{m,i}}|X'_{m,i}(n'_1, n'_2)|$ can be plotted for all values of n'_1 and n'_2 with the purpose of observing the kind of relationship existing between the rainfall field $\bar{X}_m$ and the fluctuation fields $X'_{m,i}$.

Figures 4.10, 4.11 and 4.12 show respectively examples of the procedure explained above for events in which α_i was estimated equal to 1.0, 1.3 and 1.5 in some direction i. The figures show that for a given value of s_m the correspondent absolute value of rainfall fluctuation $|X'_{m,i}|$ is within a defined interval of the pdf. For example, for $s_m = 0.01$, the lower limit of the pdf is equal to 0 at all scales for all values of α. For the same value of $s_m = 0.01$, the upper limit approximates 1.0 as α becomes smaller, and approximates 0.0 as m becomes larger. The lower limit appears to be 0.0 for almost all cases, yet, it shows values larger than 0.0 for α approximating 1.0.

The lower and upper limit of the pdf of the absolute value of an α-stable distribution, from which the rainfall fluctuations $|X'_{m,i}|$ can be selected, have been documented in a lookup table for $1 \leq m \leq 6$, $0.9 \leq \alpha \leq 1.70$, and $1.0 \times 10^{-4} \leq s_m \leq 1.0$. For events with $\alpha < 0.9$, the limits correspondent to $\alpha = 0.9$ are used. For events with $\alpha > 1.70$, the limits correspondent to $\alpha = 1.7$ are used. Therefore, in the flowchart of SR2 shown in Figure 4.8, this lookup table is used to define the interval of selection.

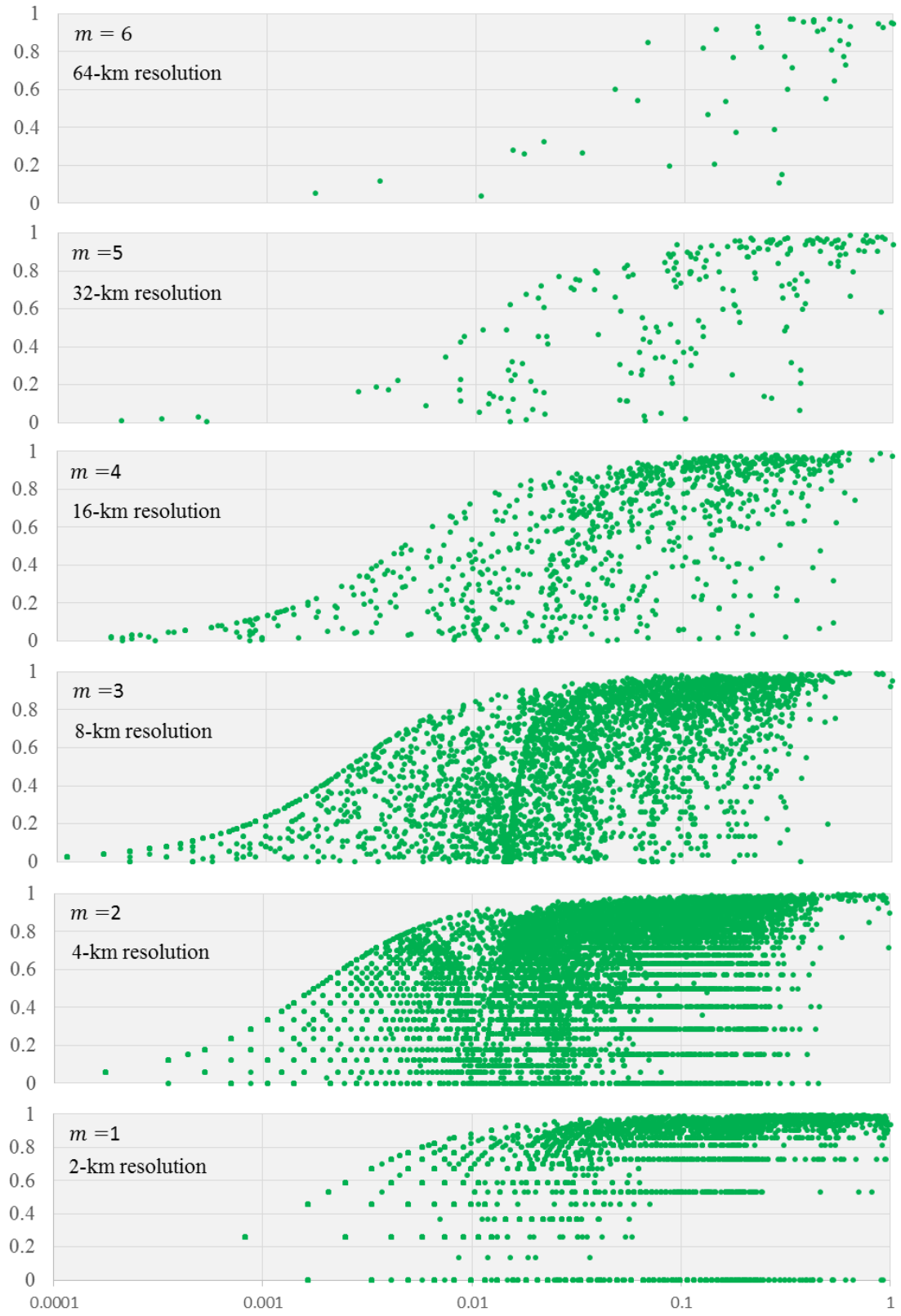


Figure 4.10 Relationship between rainfall intensity and fluctuations (1)

Plot of the values of $F_{R_{m,i}}\left|X'_{m,i}(n'_1, n'_2)\right|$ (horizontal axis) for the corresponding values of $s_m(n'_1, n'_2)$ (vertical axis). The data in this figure was computed from the event KAN_2009.08.07.07 in the north-south direction ($i = 2$), for which the estimated value of $\alpha_2 = 1.007$.

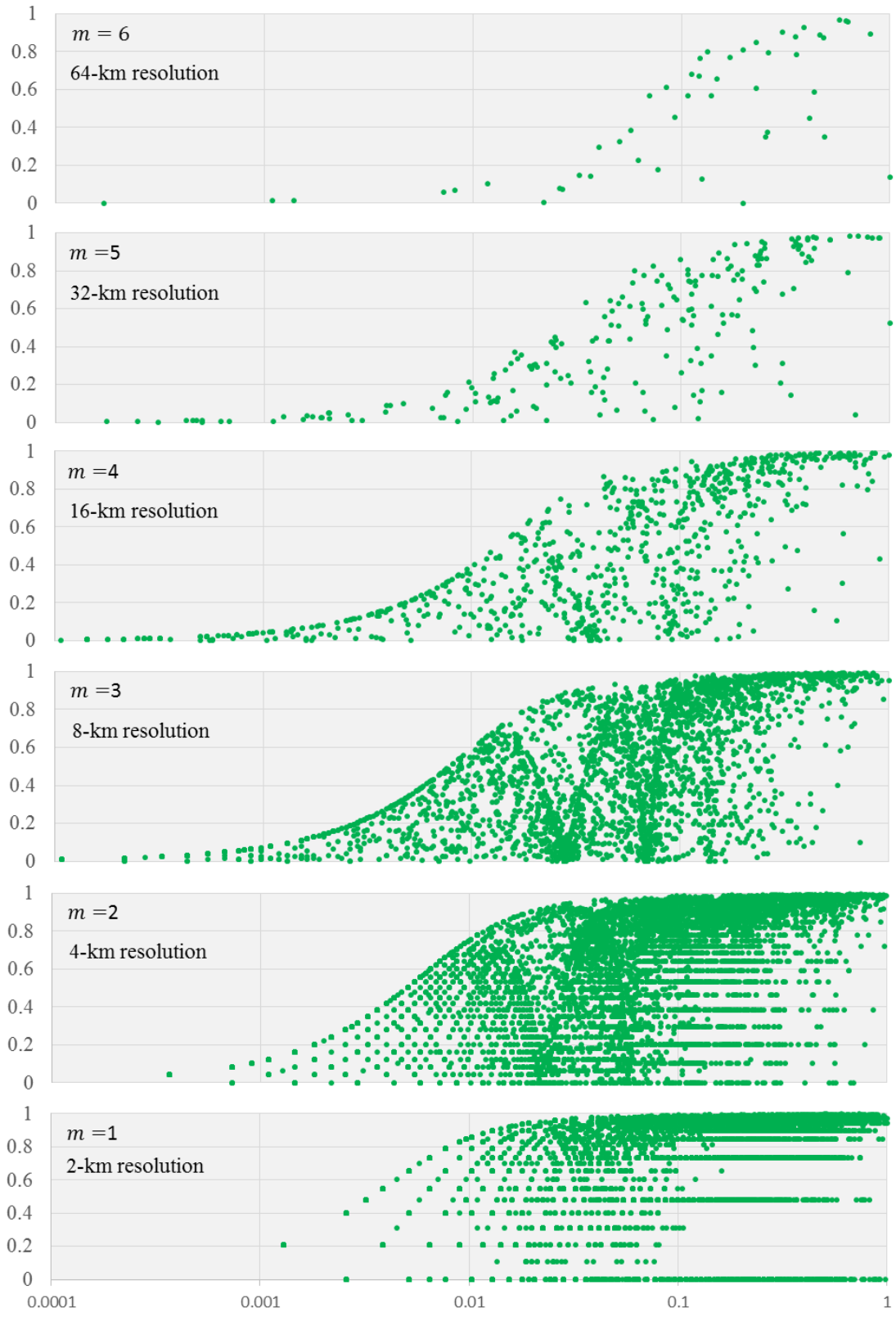


Figure 4.11 Relationship between rainfall intensity and fluctuations (2)

Figure 4.11 Plot of the values of $F_{R_{m,i}}\left|X'_{m,i}(n'_1, n'_2)\right|$ (horizontal axis) for the corresponding values of $s_m(n'_1, n'_2)$ (vertical axis). The data in this figure was computed from the event KAN_2009.08.30.23 in the east-west direction ($i = 1$), for which the estimated value of $\alpha_2 = 1.299$.

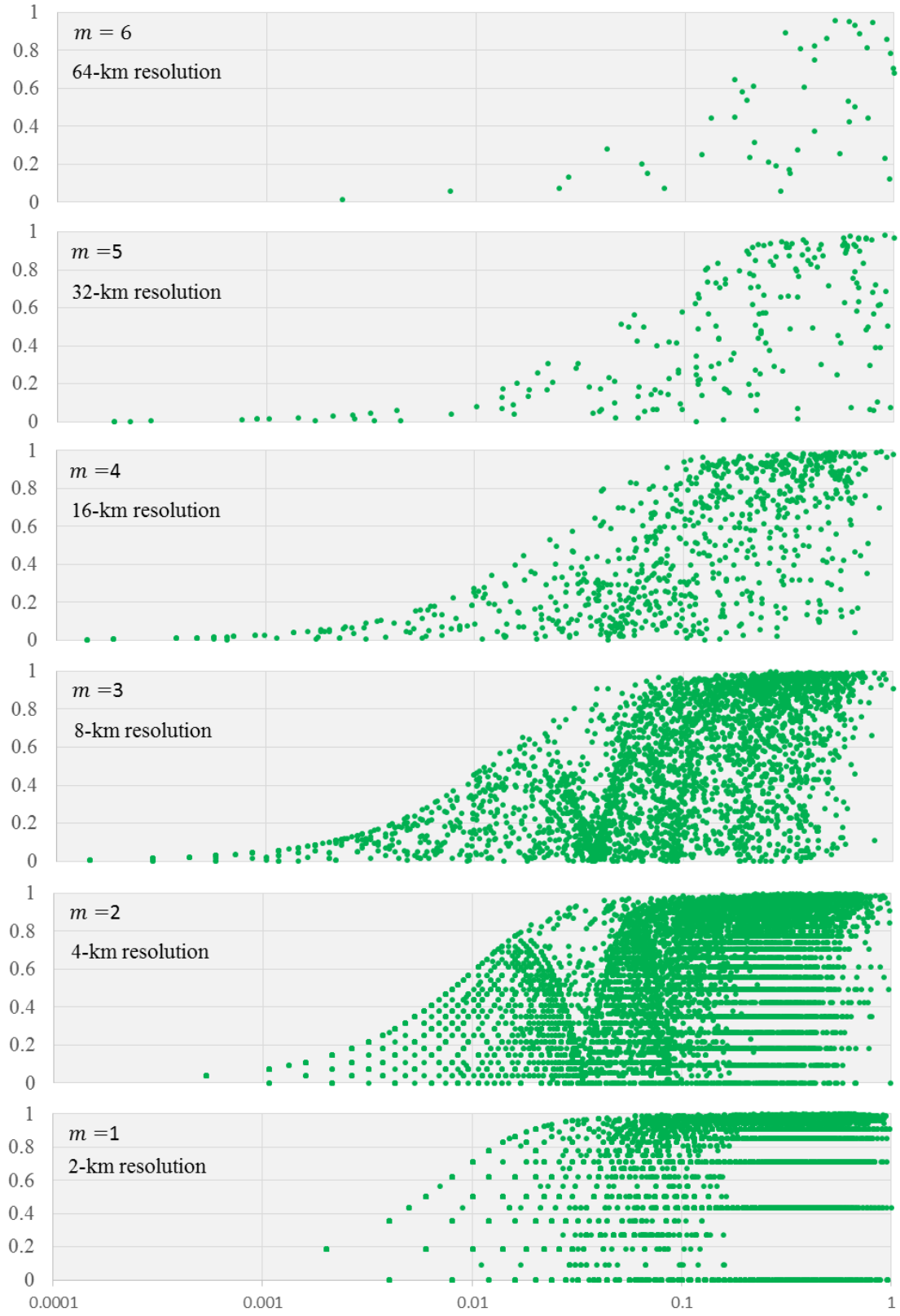


Figure 4.12 Relationship between rainfall intensity and fluctuations (3)

Figure 4.12 Plot of the values of $F_{R_{m,i}}\left|X'_{m,i}(n'_1, n'_2)\right|$ (horizontal axis) for the corresponding values of $s_m(n'_1, n'_2)$ (vertical axis). The data in this figure was computed from the event KAN_2008.06.21.23 in the east-west direction ($i = 1$), for which the estimated value of $\alpha_1 = 1.500$.

4.2.3 Subroutine SR3

In the flowchart of SR1 case 1 (Figure 4.3) and in the flowchart of SR2 (Figure 4.8) the SR3 is used to verify the "most appropriate" random selection of values of wavelet coefficients ($X'_{m,i}$). The appropriateness of the selection is based on a verification of the spatial configuration of the 2×2 sub-grid of values $\bar{X}_{m-1}$ obtained from various combinations of three randomly selected values of: $|X'_{m,1}|$, $|X'_{m,2}|$ and $|X'_{m,3}|$. The flowchart of SR3 is shown in Figure 4.13.

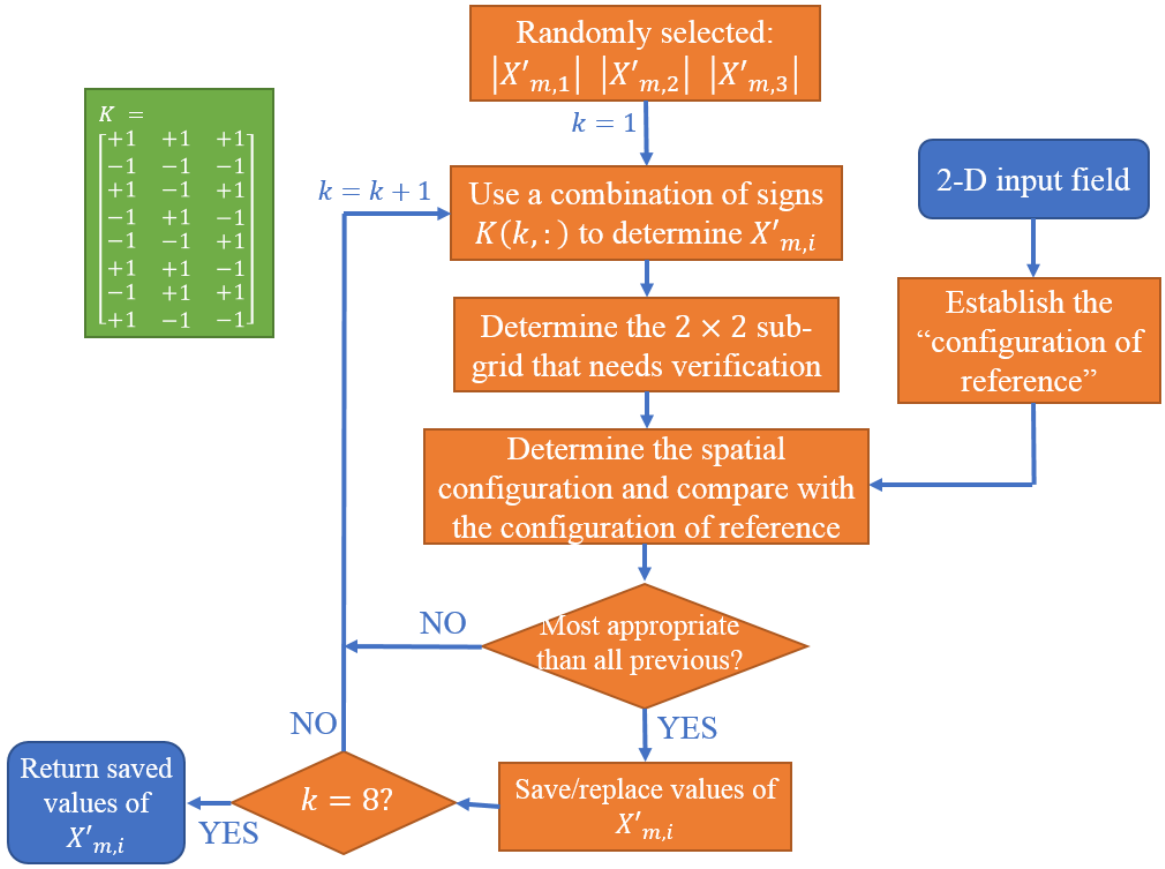


Figure 4.13 Flowchart for the verification of spatial distribution (SR3)

The four cells of a 2×2 sub-grid are labeled: NW, NE, SW and SE, as shown in Figure 4.14. Determining the spatial configuration means to order the four labels according to their respective values of intensity (see Figure 4.13). The spatial configuration (ordered vector) determined for a 2×2 sub-grid belonging to a two-dimensional input field becomes the "configuration of reference".

4.2.3.1 Application of SR3 in SR1

For the use of SR3 in the SR1 case 1, the two-dimensional input field is the LR field with a resolution correspondent to the scale m_{in}.

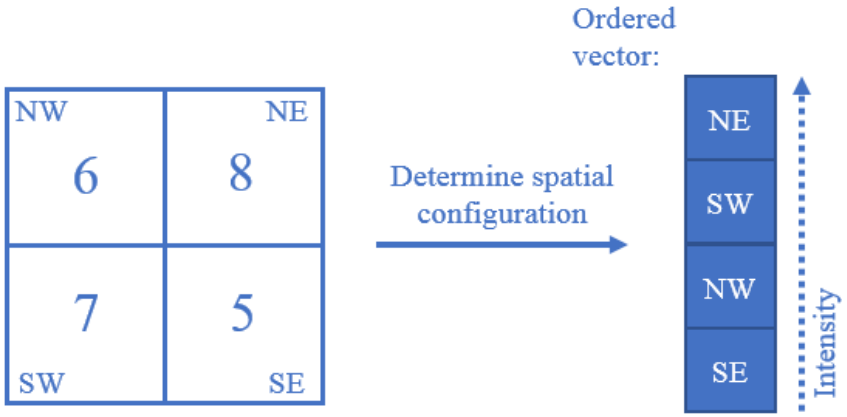


Figure 4.14 Labels of the cells in a 2 × 2 sub-grid and the determination of the spatial configuration.

For $m < m_{in}$, the starting step is to identify the 2 × 2 sub-grid at scale m_{in} that shares the same location as the 2 × 2 sub-grid at scale m for which verification of spatial configuration is needed. For a better understanding, Figure 4.15 shows an example of this identification. The spatial configuration of the 2 × 2 sub-grid at scale m_{in} is determined with the purpose of establishing the "configuration of reference".

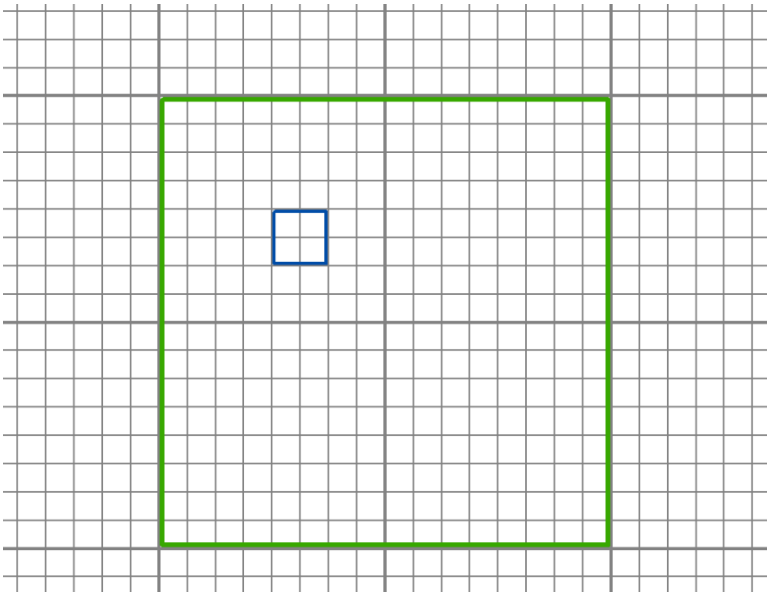


Figure 4.15 Identification of the 2 × 2 sub-grid of reference when $m \leq m_{in}$.

The thick grey lines show the grid at scale m_{in} while the thin grey lines show the grid at scale m. The blue polygon delineates the 2 × 2 sub-grid at scale m for which verification is needed. The green polygon delineates the 2 × 2 sub-grid at scale m_{in} that shares location with the one at scale m.

The SR1 tries to estimate three of the four values of $\bar{X}_{m-1}$ in a 2 × 2 sub-grid at scale m assuming that one of them is known. In order to be able to apply Eq. 2.17, this subroutine also assumes that the three values of wavelet coefficients are known. The three absolute values of wavelet coefficients $|X'_{m,i}|$ can be randomly selected from a sample generated

with the CMS algorithm. Because the three $|X'_{m,i}|$ are absolute values, the arithmetic signs are determined by verifying the spatial configuration. There are eight possible combinations of arithmetic signs. For each combination, the spatial configuration of the 2×2 sub-grid at scale m is obtained, and then compared to the "configuration of reference". From the eight combinations, the one that returns the most appropriate spatial configuration is selected. This procedure is repeated 100 times with the purpose of assuring that the most appropiate combination of signs are obtained.

If $m \geq m_{in}$, the "configuration of reference" is obtained from a 2×2 sub-grid of reference at scale m aggregated from the LR field.

4.2.3.2 Application of SR3 in SR2

The SR2 is used for generating the directional fluctuation fields ($X'_{m,i}$) that are needed for applying the IDWT in the disaggregation process.

Similarly to what was explained in the previous section, the SR3 algorithm needs a 2×2 sub-grid of a two-dimensional input field to generate the "configuration of reference". If $m \geq m_{in}$, the "configuration of reference" is obtained from a 2×2 sub-grid of reference at scale m aggregated from the LR field. If $m < m_{in}$, the "configuration of reference" is determined using the rainfall field $\bar{X}_m$ that is being disaggregated.

4.2.4 Check for blockiness

In the Figure 4.2, the structure of the algorithm shows a process called "check for blockiness" before starting the disaggregation at the next smaller scale. An example of the output of the algorithm with blocky representation is shown in Figure 4.16a, which depicts a kind of pixelated rainfall field without smooth transitions. This issue was reported as a disadvantage of most stochastic downscaling models (Ferraris et al., 2003; Ebtehaj et al., 2012). Unfortunately, the IDWT has the same effect as it does not consider the relationship of cells between two contiguous 2×2 sub-grids, resulting in very large differences of intensity between contiguous cells.

The "check for blockiness" process was designed with the purpose of lessening this deteriorating effect. For a representation of the rainfall field $\bar{X}_m(n'_1, n'_2)$, using Eq. 2.17, the wavelet coefficients $X'_{m,i}(n'_1, n'_2)$ are computed only for those 2×2 sub-grids in which the location indexes n_1 and n_2 are even integer numbers. If the wavelet coefficients are 1.85 times larger than the expected value of the distribution of absolute values, a weighted correction factor is applied to all four cells in the 2×2 sub-grid. In order to preserve the degree of variability, the correction factor was defined as a function of the characteristic exponent of the α-stable distribution of wavelet coefficients. As the value of α becomes smaller (larger variability), the correction factor also becomes small. The value of 1.85 mentioned above and the correction factor were calibrated with the purpose of having

the minimum impact on the statistical metrics of the output of the algorithm and on the distribution of rainfall fluctuations. Figure 4.16b shows an example of the application of this procedure.

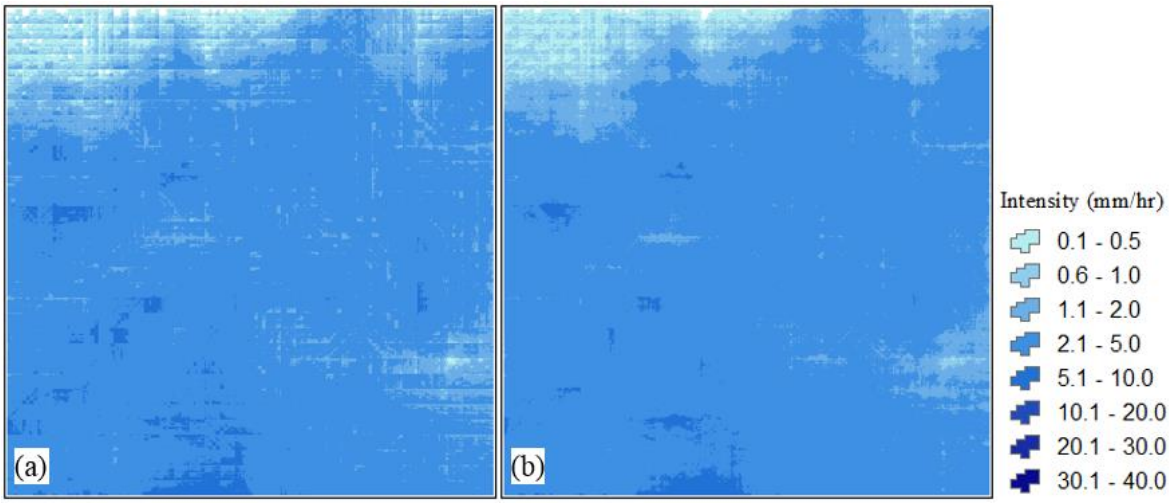


Figure 4.16 Effect of the check for blockiness

(a) Output of the disaggregation algorithm without applying the "checking for blockiness" procedure. (b) Output of the disaggregation algorithm applying the "checking for blockiness" procedure

4.3 VALIDATION OF THE ALGORITHM OF DISAGGREGATION

Because this algorithm uses random samples for the generation of wavelet coefficients, the results presented in this section are the average of an ensemble. No considerable differences were found in the results if an ensemble of 1000 repetitions are considered for the estimation of $\bar{\bar{R}}$ (SR1 case 1), and ensemble of 500 repetitions are considered in the disaggregating process.

4.3.1 Analysis sites

The validation of the algorithm was firstly applied to sites with dimensions 128 km × 128 km located as shown in Figure 4.17. Each of these sites is named according to the rain gauge station of the AMeDAS network from which rainfall measurements are used as input of the algorithm.

Table 4.3 Characteristics of the analysis sites and number of analyzed events

Site	Code	UTM coordinates of the Southwest corner		UTM Zone	Nº of events
Maebashi	MAE	264500E	3976500N	54N	191
Niigata	NII	311000E	4083500N	54N	181
Sapporo	SAP	481000E	4718000N	54N	209
Miyazaki	MIY	612500E	3482500N	52N	223
Nagasaki	NAG	531000E	3574000N	52N	150

For each selected rain gauge station, in the period 2006 – 2009, rainfall events in which the hourly intensity measurement was above some predetermined threshold were selected for the validation of the algorithm. During storms of long duration, only the events in which the rain gauge measured peaks of intensity were considered. The number of events for each station as well as characteristics of the sites are presented in Table 4.3.

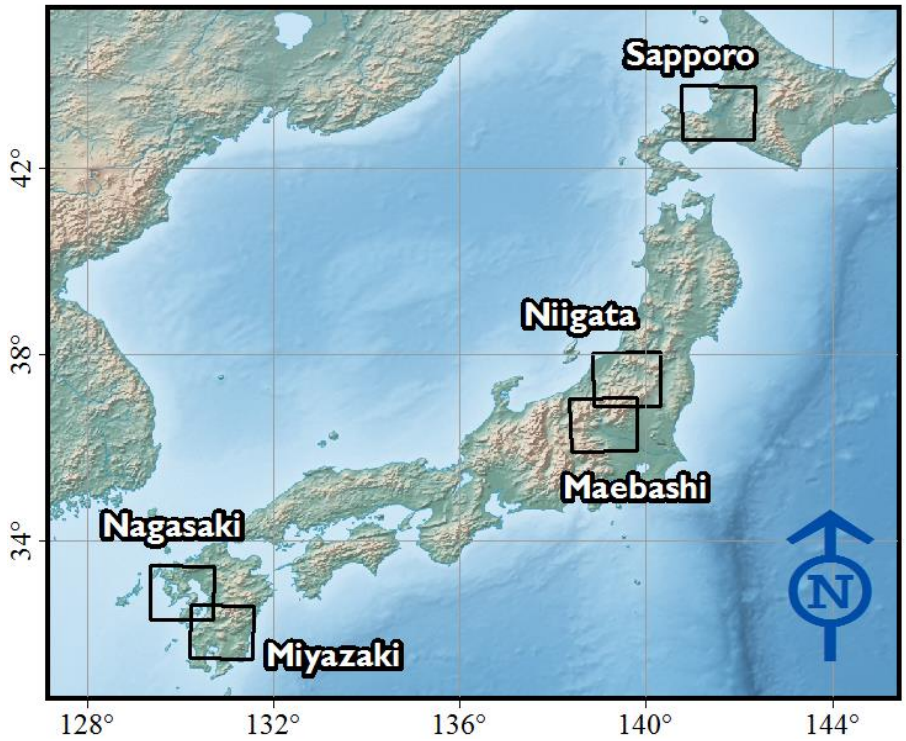


Figure 4.17 Location of the sites with dimensions 128 km × 128 km

4.3.2 Fidelity measures for statistical evaluation

The validation of the algorithm should evaluate how well the subroutine (SR1) for the estimation of $\bar{\bar{R}}$ works and how well the spatial structure of the rainfall field is reproduced after the disaggregation. For this purpose, fidelity measures that take into account the first and second order statistics of the error are considered. Let us denote the error matrix $\mathcal{E} = \widehat{\mathcal{X}} - \mathcal{X}$, where $\mathcal{X}$ is the grid (matrix) of intensity values corresponding to the Radar-AMeDAS data of a rainfall event and $\widehat{\mathcal{X}}$ is the matrix of intensity values corresponding to the output of the algorithm for the same event. Then, expressions for the Normalized Bias Ratio (NBR), the Root Mean Square Error ($RMSE$), the Coefficient of Variation (CV) and the Mean Absolute Error (MAE) are given respectively in Eq. 4.5, 4.6, 4.7 and 4.8, where $\overline{(\cdot)}$ is the entry-wise mean and $|\cdot|$ is the entry-wise absolute value.

$$NBR = \bar{\mathcal{E}}/\bar{\mathcal{X}} \tag{4.5}$$

$$RMSE = \sqrt{\overline{\mathcal{E}^2}} \tag{4.6}$$

$$CV = RMSE/\bar{\mathcal{X}} \tag{4.7}$$

$$MAE = \overline{|\mathcal{E}|} \tag{4.8}$$

Excellent results should yield values of NBR approximating zero, however, it is necessary to establish an interval for which the bias of the output can be considered acceptable. Since the value of NBR depends on $\bar{\mathcal{X}}$, and since $\bar{\mathcal{X}}$ may range from 0.1 mm/hr to 15 mm/hr, it is difficult to establish an interval of NBR that can be assumed acceptable for all events. Regardless, the interval of acceptance established herein is $0.5 \leq NBR \leq 0.5$ and it should be evaluated alongside the value of MAE.

Regarding the evaluation of the reproduction of the characteristic spatial structure of the rainfall field, an acceptable output should yield fidelity measures such that

$$\bar{\hat{\mathcal{X}}} = \bar{\mathcal{X}} \pm 1.0 \times RMSE.$$

This expression and the definition of CV imply that an acceptable output should yield values of $CV \leq 1.0$.

4.3.3 Results

4.3.3.1 Performance of the disaggregation process

The disaggregation process is evaluated under the following assumptions/considerations:

- The parameters $\{\alpha_i, \gamma_{1,i}, H_i\}_{i=1,2,3}$ are known, which were estimated using the DWT with a combined sampling of wavelet coefficients according to the description given in Section 3.4.5.
- The input LR field is obtained as an aggregation of the Radar-AMeDAS data into a 16-km resolution grid. Therefore, the disaggregation process is applied from scale $m = 4$ (input) to scale $m = 0$ (output).
- The resolution of the output is the same as the resolution of the Radar-AMeDAS data for statistical evaluation purposes.

Examples of the performance are shown in Figures 4.18 and 4.20, which show respectively the output rainfall field for the events MAE_2006.07.18.13 and MAE_2009.06.23.02. Additionally, Figures 4.19 and 4.21 show the scatterplots of the output rainfall field $\hat{\mathcal{X}}$ on the Radar-AMeDAS data $\mathcal{X}$ and the computed metrics, respectively for the same events.

Note that the values of $\bar{\mathcal{X}}$ and $\bar{\hat{\mathcal{X}}}$ should be equal because of the invariance of the mean that implies the IDWT algorithm. However, because of the "checking for blockiness" process, in events with no-rain areas, $\bar{\hat{\mathcal{X}}}$ was found to differ in the second or third decimal point with respect to $\bar{\mathcal{X}}$.

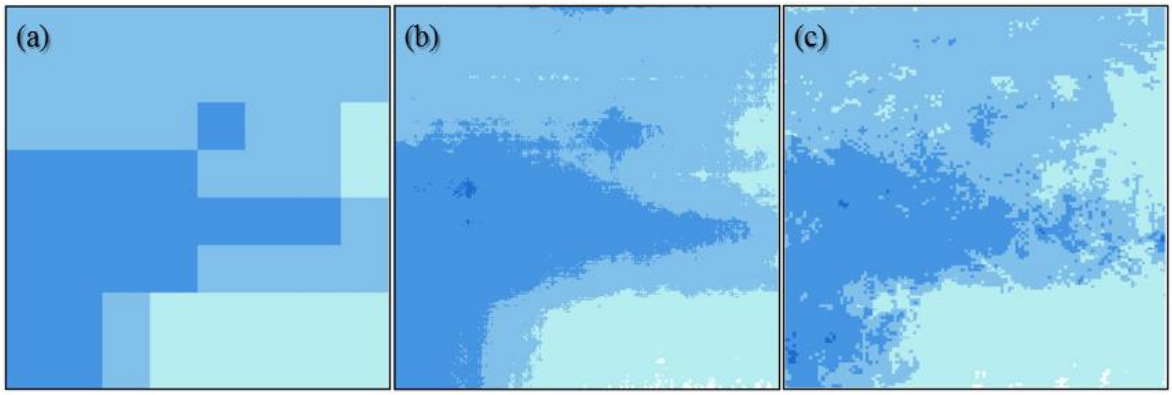


Figure 4.18 Performance of the disaggregation process for the event MAE_2006.07.18.13

(a) Input LR field with 16-km resolution. (b) Output of the algorithm with 1-km resolution. (c) Radar-AMeDAS data.

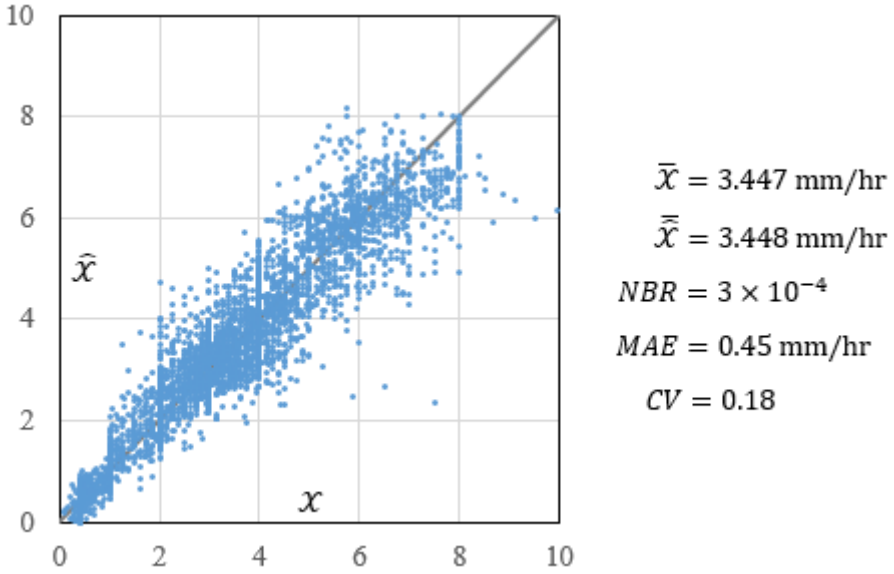


Figure 4.19 Scatterplot of $\hat{x}$ on x and fidelity measures for the event MAE_2006.07.18.13

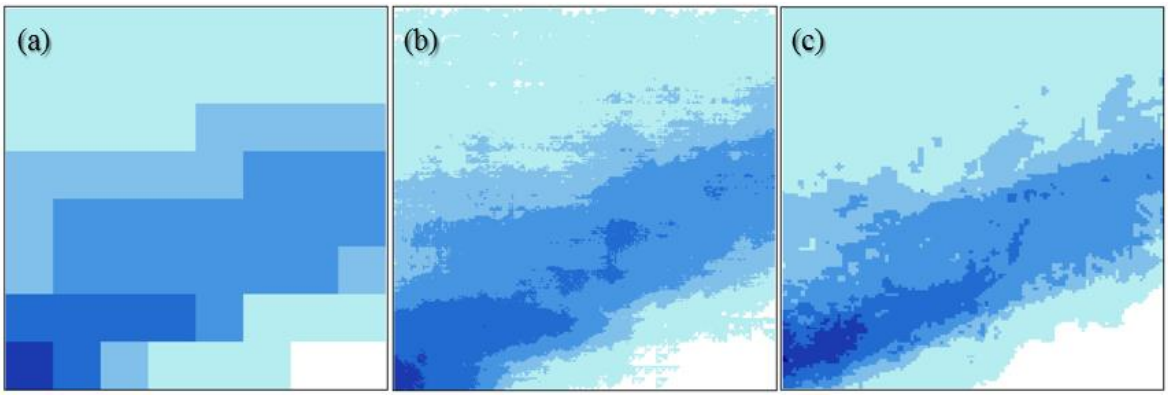


Figure 4.20 Performance of the disaggregation process for the event MAE_2009.06.23.02

(a) Input LR field with 16-km resolution. (b) Output of the algorithm with 1-km resolution. (c) Radar-AMeDAS data.

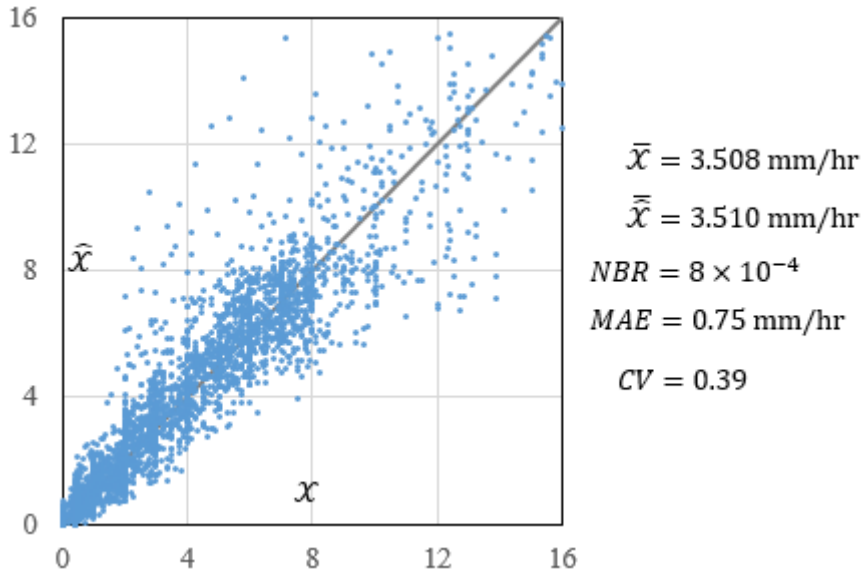


Figure 4.21 Scatterplot of $\widehat{\mathcal{X}}$ on $\mathcal{X}$ and fidelity measures for the event MAE_2009.06.23.02

The results for all the events considered in the five $128\ km \times 128\ km$ sites are presented in Table 4.4. Since Radar-AMeDAS data aggregated into a 16-km resolution grid is used as the input LR field, all events returned values of NBR within the range $-0.5 \leq NBR \leq 0.5$. On the other hand, reproducing the characteristic spatial configuration of rainfall and small-scale variation, which is measured by the CV, was not accomplished for all events. The lowest percentage of success is shown by the events of the Maebashi site, where convective rainfall events depicting sudden changes of intensity over short distances are more frequent than in the other sites, and happen to be more difficult to reproduce. An example of an event that returned a value of CV larger than 1.0 is shown in Figure 4.22, and its respective scatterplot and metrics are presented in Figure 4.23. The characteristic exponents of the α-stable distributions of rainfall fluctuations of this event were estimated as $\alpha_{i=1,2,3} = \{1.09, 1.05, 1.08\}$, denoting a quite high degree of variability. Therefore, the events most difficult to reproduce are the ones showing values of α_i approximating 1.0.

Table 4.4 Percentage of events in each site that returned fidelity measures within the acceptable ranges

Site	Nº Events	$\lvert NBR \rvert \leq 0.5$	$CV \leq 1.0$
Maebashi	191	100	77
Niigata	181	100	86
Miyazaki	223	100	83
Nagasaki	209	100	91
Sapporo	150	100	90

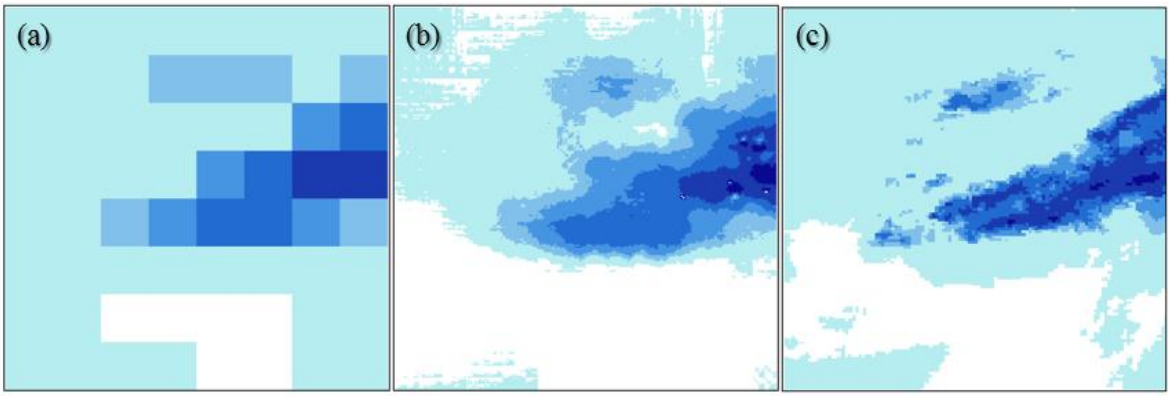


Figure 4.22 Performance of the disaggregation process for the event MAE_2009.07.16.22

(a) Input LR field with 16-km resolution. (b) Output of the algorithm with 1-km resolution. (c) Radar-AMeDAS data.

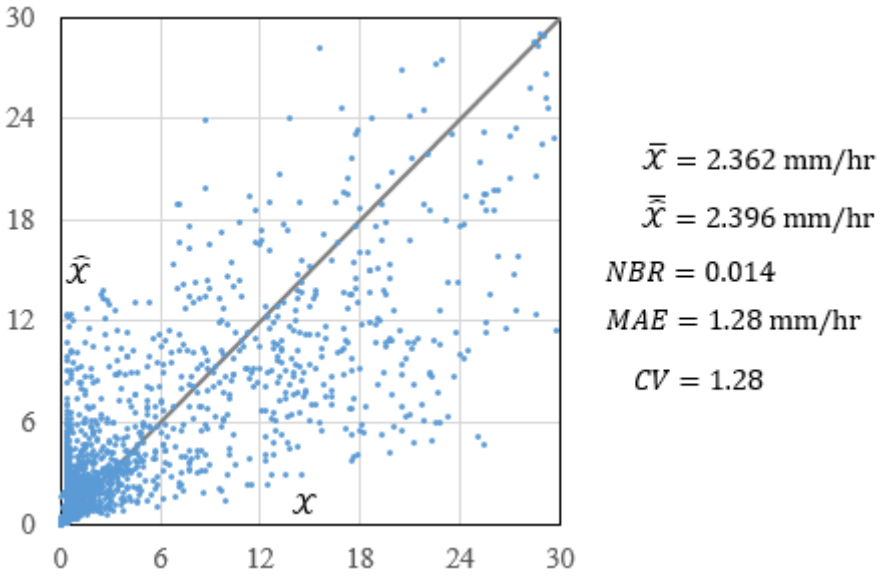


Figure 4.23 Scatterplot of $\widehat{X}$ on X and fidelity measures for the event MAE_2009.07.16.22

4.3.3.2 Performance of the algorithm using SR1 case 1

The SR1 case 1 estimates the mean rainfall intensity of the analysis site ($\bar{\bar{R}}$) by stochastically expanding into larger scales the measurement of only 1 rain gauge station. Aside from evaluating the performance of this subroutine, the effects of having a biased value of $\bar{\bar{R}}$ may also be evaluated by running the whole algorithm and computing the fidelity metrics of the output. The following considerations are in order:

- The measurement of the selected rain gauge station at each site will be used as input in the SR1.
- The parameters $\{\alpha_i, \gamma_{1,i}, H_i\}_{i=1,2,3}$ are known, which were estimated using the DWT with a combined sampling of wavelet coefficients according to the description given in Section 3.4.5.

- The input LR field is obtained as an aggregation of the Radar-AMeDAS data into a 16-km resolution grid.
- The resolution of the output is the same as the resolution of the Radar-AMeDAS data for statistical evaluation purposes.

Examples of the performance are shown in Figures 4.24 and 4.26, which show respectively the output rainfall field for the events MAE_2006.07.18.13 and MAE_2009.06.23.02. Additionally, Figures 4.25 and 4.27 show the scatterplots of the output rainfall field $\widehat{\mathcal{X}}$ on the Radar-AMeDAS data $\mathcal{X}$ and the computed metrics, respectively for the same events. The scatterplots are shown at 16-km resolution and at 1-km resolution, which allows to appreciate the functioning of the algorithm while disaggregating the estimated $\bar{\bar{R}}$ from scale $m = 7$ (128-km resolution) into smaller scales (higher resolutions).

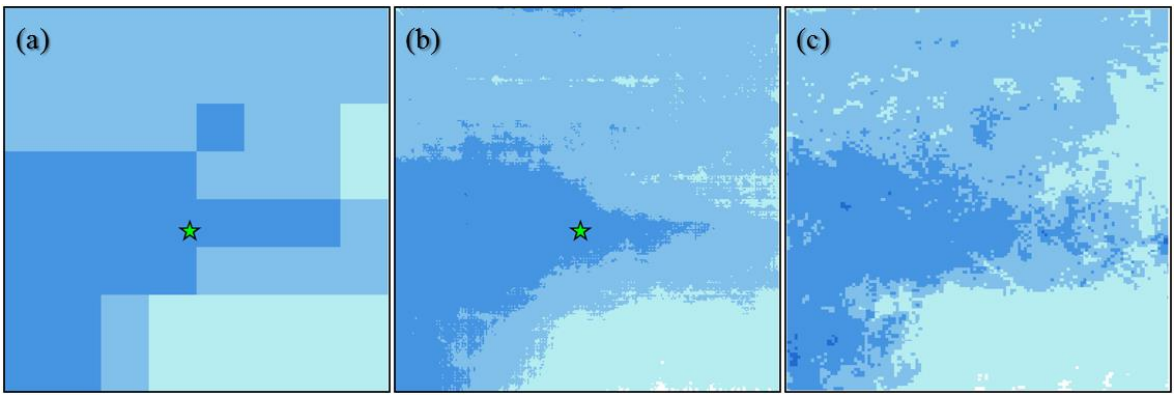


Figure 4.24 Performance of the whole algorithm for the event MAE_2006.07.18.13

(a) Input LR field with 16-km resolution. The green star represents the location of the Maebashi station. (b) Output of the algorithm with 1-km resolution. (c) Radar-AMeDAS data.

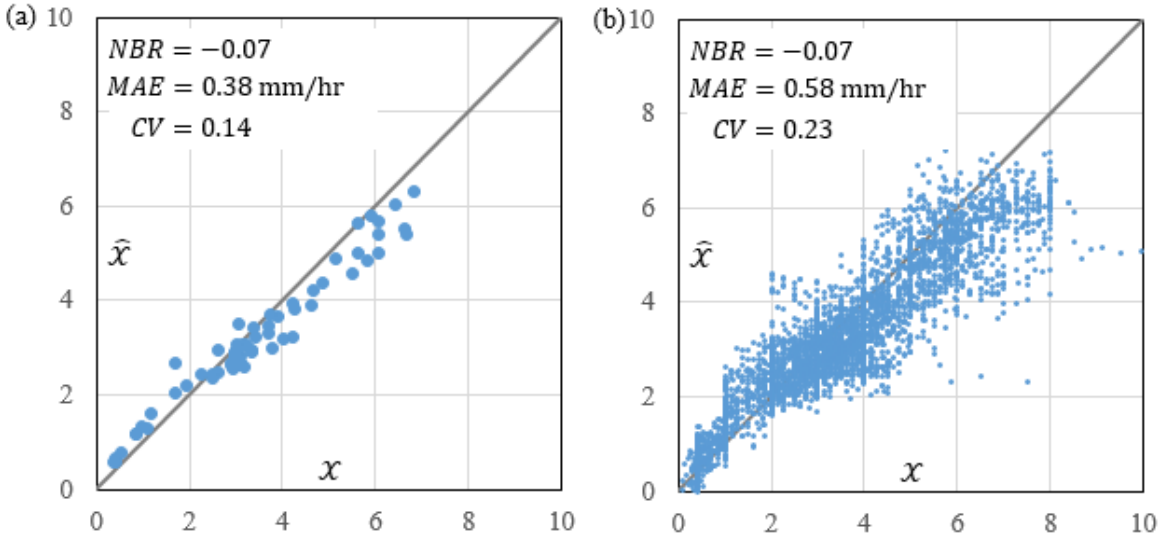


Figure 4.25 Scatterplot of $\widehat{\mathcal{X}}$ on $\mathcal{X}$ and fidelity measures for the event MAE_2006.07.18.13

(a) Comparison at 16-km resolution. (b) Comparison at 1-km resolution.

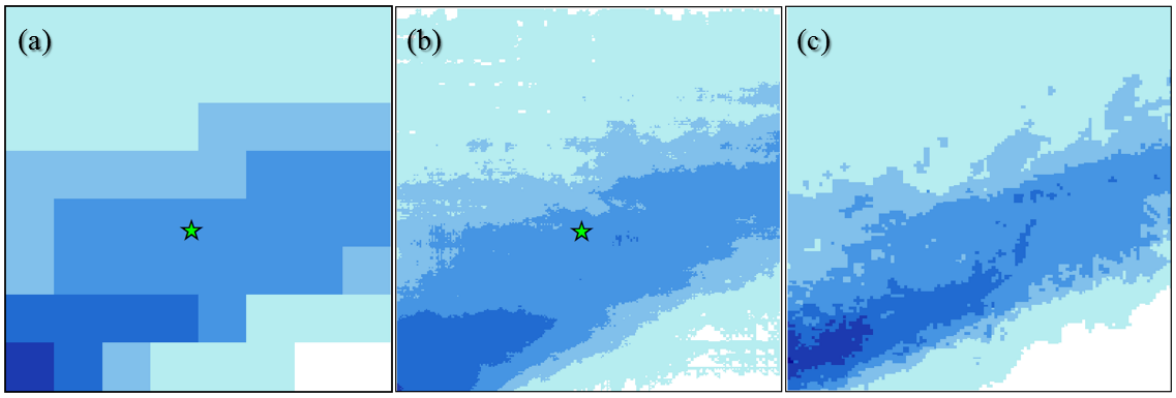


Figure 4.26 Performance of the whole algorithm for the event MAE_2009.06.23.02

(a) Input LR field with 16-km resolution. The green star represents the location of the Maebashi station. (b) Output of the algorithm with 1-km resolution. (c) Radar-AMeDAS data.

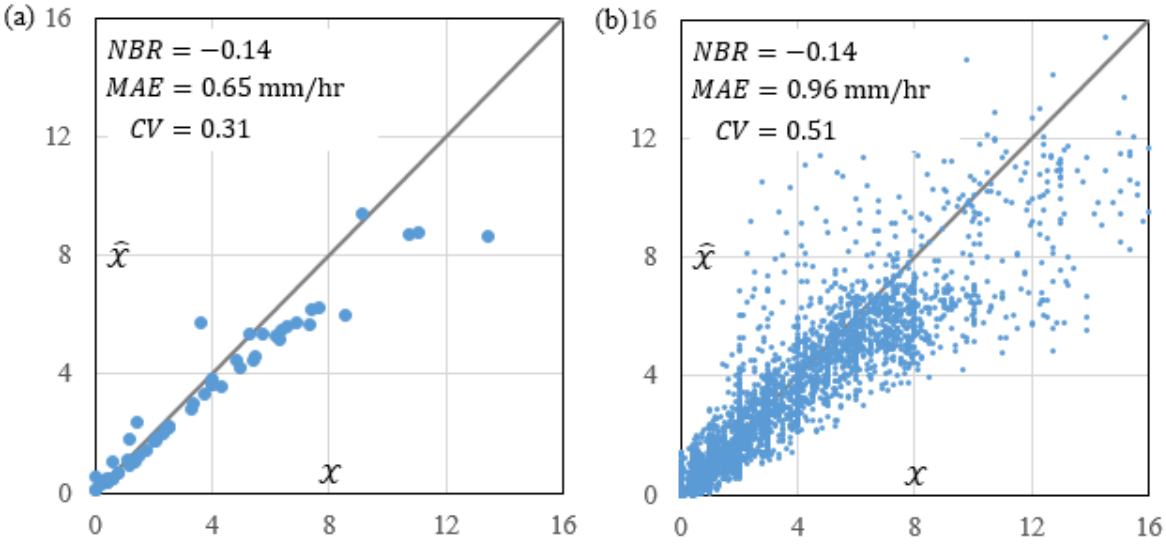


Figure 4.27 Scatterplot of $\hat{X}$ on X and fidelity measures for the event MAE_2009.06.23.02

(a) Comparison at 16-km resolution. (b) Comparison at 1-km resolution.

Table 4.5 Percentage of events in each site that returned fidelity measures within the acceptable ranges

			$CV \leq 1.0$	
Site	Nº Events	$\|NBR\| \leq 0.5$	16-km res.	1-km res.
Maebashi	191	78	82	66
Niigata	181	85	90	73
Miyazaki	223	71	82	64
Nagasaki	209	83	92	78
Sapporo	150	73	81	73

The results for all the events considered in the five $128\ km \times 128\ km$ sites are presented in Table 4.5. From a total of 954 events (considering all five sites), in 78% of them the SR1 case 1 was able to estimate the mean intensity $\bar{\bar{R}}$ yielding values of NBR within the range of acceptance. However, the bias carried in the estimation of $\bar{\bar{R}}$ produced a reduction of the

percentage of events in which the small-scale variability and characteristic spatial configuration of rainfall can be reproduced.

4.4 TRMM RAINFALL PRODUCTS AS INPUT OF THE ALGORITHM

The TRMM 3B42 version 7.0 product has a $0.25° \times 0.25°$ resolution and produces 3 hourly estimates between the latitudes $50°$S to $50°$N. Using this rainfall estimates as the input LR field, the following results are intended to verify the functioning of the subroutine detailed in Section 4.2.1.2 and to evaluate the conditions in which the input generates acceptable outcomes.

The results presented herein correspond to events that happened over the Kanto analysis site, which has dimensions 256 km $\times$ 256 km. The following considerations are in order:

- Since we have been studying the datasets over sites with dimensions equal to powers of 2, the TRMM products were first resampled to a grid with 0.5-km resolution via the Nearest Neighbors method, and then aggregated to a grid with 16-km resolution. Figure 4.28a shows the original data and Figure 4.29b shows the resampled data for the event KAN_2009.11.13.21.
- The parameters $\{\alpha_i, \gamma_{1,i}, H_i\}_{i=1,2,3}$ are known, which were estimated using the DWT with a combined sampling of wavelet coefficients according to the description given in Section 3.4.5.
- Let us define the error matrix of the TRMM products with respect to the Radar-AMeDAS datasets as $\mathcal{E}_{\mathcal{Y}} = \mathcal{Y} - \mathcal{X}$, where $\mathcal{Y}$ is the matrix of intensity values correspondent to the TRMM products. Replacing $\mathcal{E}$ by $\mathcal{E}_{\mathcal{Y}}$ in Eq. 4.5 to 4.8, the same fidelity measures are used for determining the bias between these two datasets.

Because of how the algorithm was constructed, an acceptable performance can be expected if the input LR field is able to identify correctly the location of high and low intensity areas. The spatial correlation of TRMM products and Radar-AMeDAS datasets at 16-km resolution was analyzed in events with a mean intensity of conditional rainfall greater than 1 mm/hr. Only 11% of the events between 2006 and 2009 showed a coefficient of determination greater than 0.5, and only 6% showed a coefficient of determination greater than 0.6. The results for some events that showed different degrees of spatial correlation is shown below.

Figure 4.29 shows the input LR field, the resampled input LR field, the output of the algorithm and the Radar-AMeDAS data for the event KAN_2007.07.17.06, which yielded a coefficient of determination equal to 0.78 for the spatial correlation with Radar-AMeDAS data. Figure 4.30 shows the scatterplot of the TRMM product on the Radar-AMeDAS data, and the scatterplot of the output of the algorithm on the Radar-AMeDAS data.

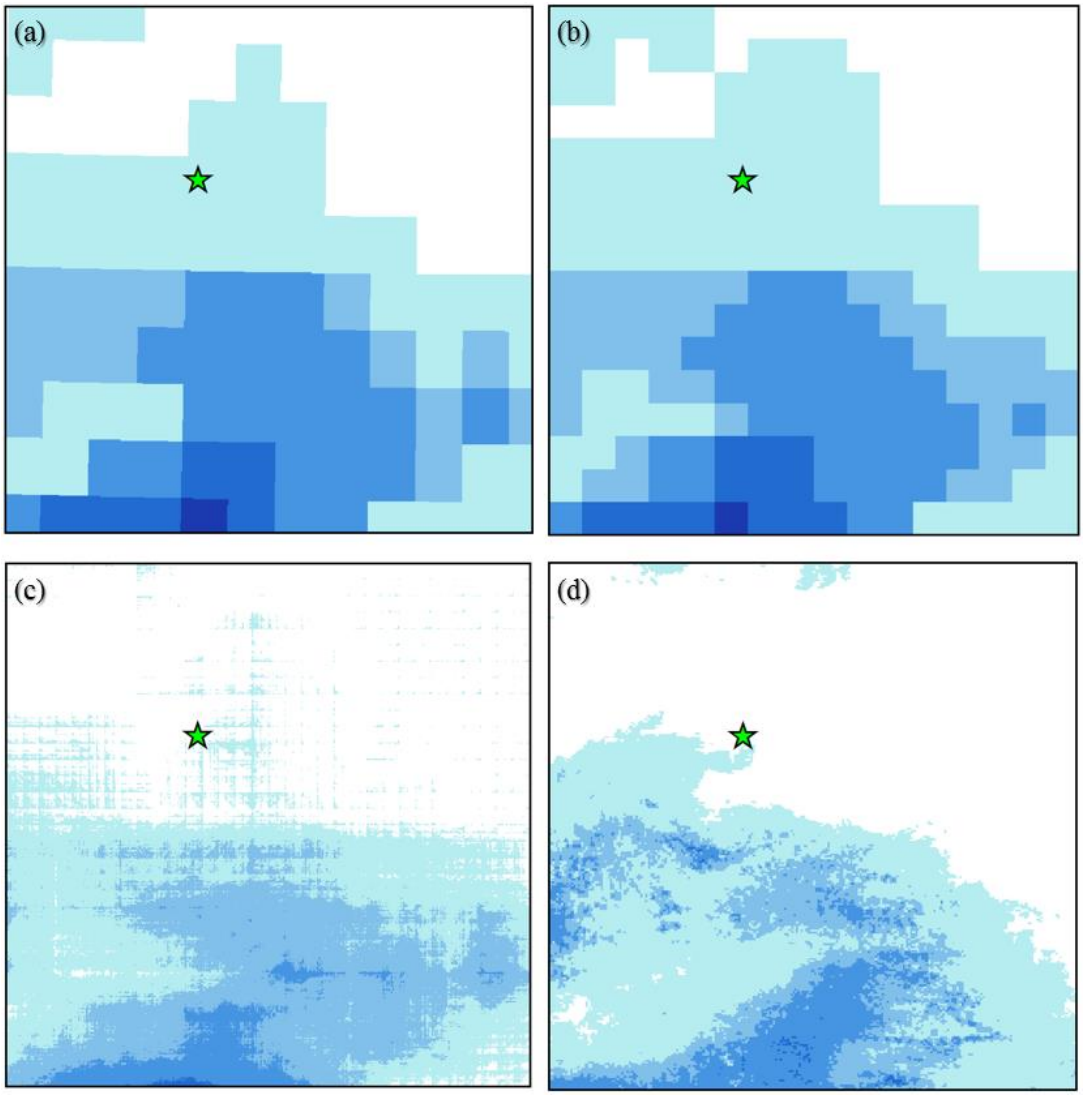


Figure 4.29 Input and Output data for the event KAN_2007.07.17.06.

(a) TRMM data. (b) TRMM data resampled to 16-km resolution. (c) Output of the algorithm. (d) Radar-AMeDAS data.

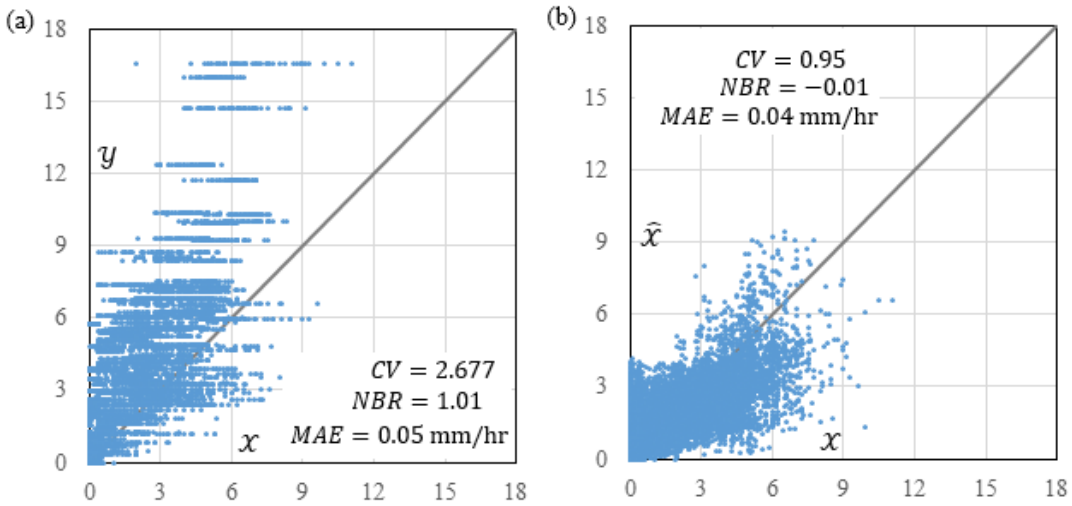


Figure 4.30 Scatterplots for the event KAN_2007.07.17.06

(a) y on x and (b) $\hat{x}$ on x.

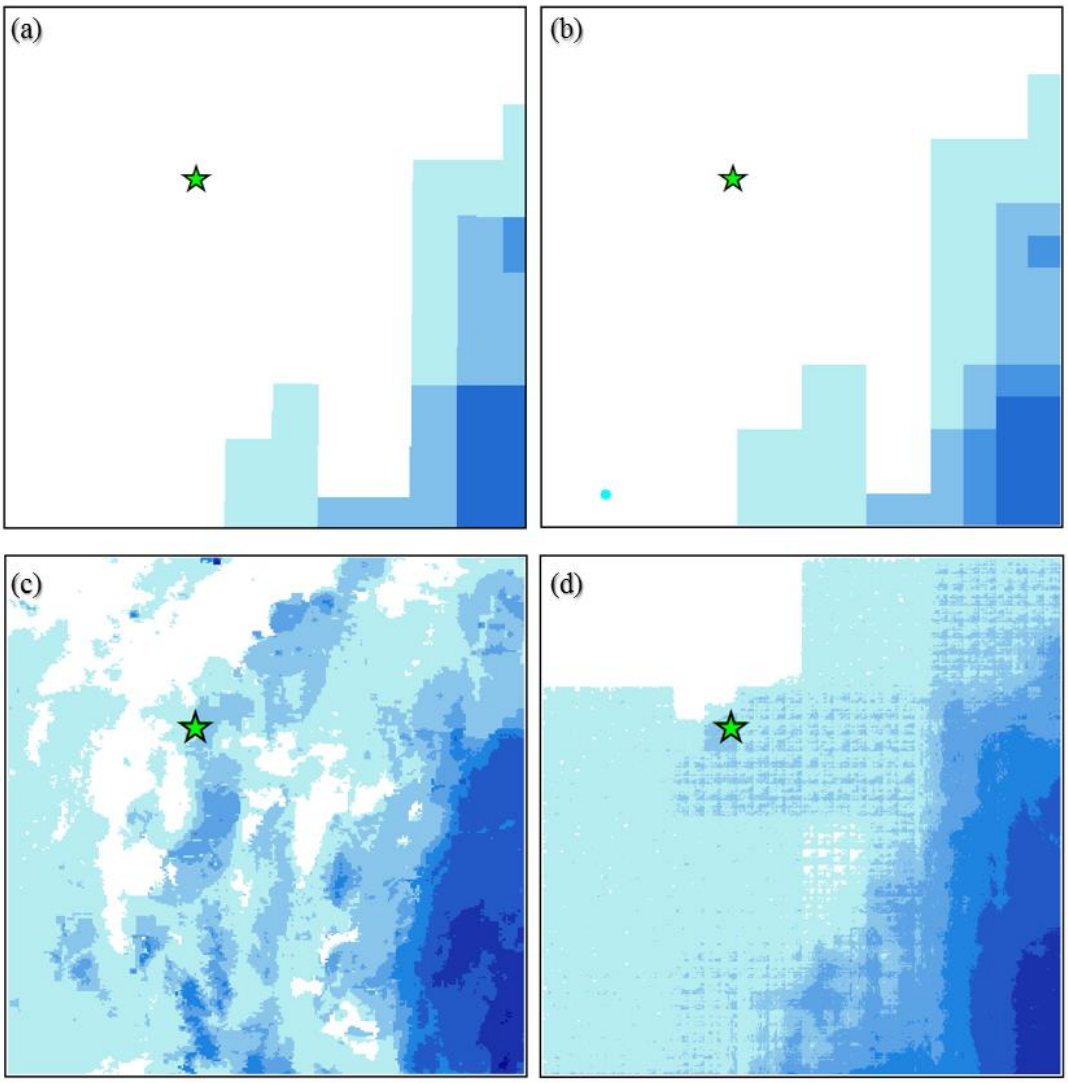


Figure 4.31 Input and Output data for the event KAN_2009.11.13.21

(a) TRMM data. (b) TRMM data resampled to 16-km resolution. (c) Output of the algorithm. (d) Radar-AMeDAS data.

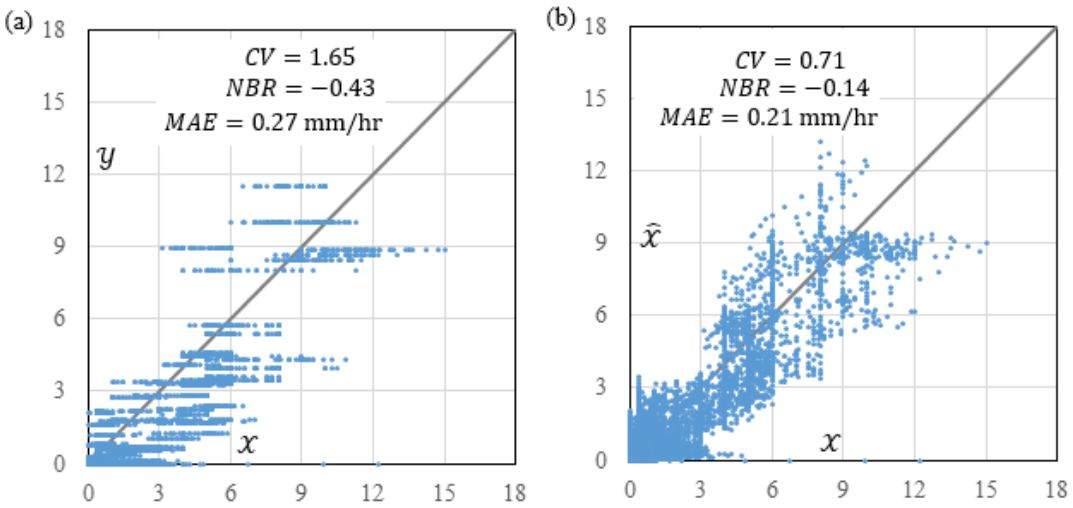


Figure 4.32 Scatterplots for the event KAN_2009.11.13.21

(a) y on x and (b) $\hat{x}$ on x.

Figure 4.31 shows the input LR field, the resampled input LR field, the output of the algorithm and the Radar-AMeDAS data for the event KAN_2009.11.13.21, which yielded a coefficient of determination equal to 0.57 for the spatial correlation with Radar-AMeDAS data. Figure 4.32 shows the scatterplot of the TRMM product on the Radar-AMeDAS data, and the scatterplot of the output of the algorithm on the Radar-AMeDAS data.

5. CONCLUSIONS

High-resolution rainfall fields are only available in regions that can afford the operation of weather radars and/or dense rain gauge networks. Most developing regions lack this kind of equipment and, for hydro-meteorological simulations, the main source of rainfall intensity is sparse rain gauge measurements or satellite-based estimations of rainfall. Normally, schemes for the bias-correction of satellite-based products try to find ways of distributing the error with respect to the few in-situ measurements. However, it was recognized that the best approach should focus on studying the types of rainfall and their differences in structure rather than trying to impose a numerical schemes based on local conditions (Ebert et al., 2007). This research aimed to make a stochastic analysis of the spatial structure of rainfall fields with the purpose of characterizing the scale-to scale dependence of some statistical property. A direct consequence of accomplishing such characterization is the construction of algorithms that link the spatial structure of fields at low and high resolutions with the least possible number of parameters. Stochastic disaggregation models should be able to transmit the structure of low-resolution fields to high-resolution fields while adding the details that are characteristic of spatial rainfall. The detail information can be seen as variations of the field with respect to local averages. In this sense, previous studies proposed the decomposition of rainfall fields into large-scale components and small-scale components using discrete orthonormal wavelet transformations (Kumar and Foufoula-Georgiou, 1993).

Datasets of Radar-AMeDAS were extensively studied for five analysis sites located in Japan with different precipitation regimes and weather conditions. The two-dimensional orthonormal wavelet transformation allows to separate three sets of wavelet coefficients from the rainfall field. Each set represents the local fluctuations of the field in the east-west, north-south and diagonal directions. These sets contain the necessary detail-information to transform low-resolution fields into the original high-resolution fields. Previous stochastic analysis of the sets of wavelet coefficients evidenced dependence between scales of the distribution parameters (Perica and Foufoula-Georgiou, 1996; Ebtehaj and Foufoula-Gerogiou, 2011). If these parameters can be related to physical measures of the storm environment, it would imply that the small-scale structure of rainfall fields can be statistically derived from physical models that provide large-scale information of the variables controlling the formation of storms.

In this study, we proposed to adjust the sets of wavelet coefficients to symmetric α-stable ($S\alpha S$) distributions, which are able to model the high frequencies of extremes found in the pdfs better than Gaussian models. The setback of working with α-stable distributions, however, is the lack of closed-form expressions of the pdf and cdf, which hinders the estimation of the parameters. Non-conventional methods of estimation tend to adjust different "features" of the sample, yielding different results. Moreover, the large amount of zeros and limited amount of data has shown to deteriorate the performance of the estimation-methods. The characteristic exponent of $S\alpha S$ distributions defines the rate of decay of the

tails towards infinity, defining as well the frequency of extreme values. Since extreme variations are usually observed in the rainfall field, we set the hypothesis that α could be a measure of the degree of variability of the rainfall event.

The properties of the $S\alpha S$ distributions were revised with the purpose of designing a method of estimation that is able to overcome the difficulties that come naturally in the samples of rainfall fluctuations. Two methods were proposed: the FAM method and TSM method. Through Montecarlo simulations, the FAM method showed a performance that is comparable to methods regarded as most accurate while employing less computational time. This and the other existing methods normally present problems when trying to estimate the parameters of samples that contain large number of zeros at high resolutions. The TSM method, on the other hand, proved to have the capacity to overcome this issue.

The scale-to-scale dependence of the estimated parameters were analyzed. Independently at each direction, the estimators of the characteristic exponent α showed an almost invariant behavior between the 1-km and the 16-km resolutions (reaching the 32-km resolution in several cases). At the 32-km and 64-km resolutions, the estimated characteristic exponent α tends to have smaller values than the ones estimated at higher resolutions. This situation can be caused by the insufficient information of the small samples at these scales or simply because self-similarity does not hold up to these scales. The former claim might be more probable since the samples come from events that develop at larger atmospheric scales. The scale parameter showed a power-law behavior in the scale range between 1-km and 64-km for a large portion of the considered events. These two conditions suggest that the processes of rainfall fluctuations can be qualified as self-similar and the distribution at any scale of representation can be estimated if the distribution at a specific scale is known.

The hypothesis that the characteristic exponent α is related to the degree of variability is supported by the results presented in this chapter. Events that depict short-range changes of intensity, clustered structure and intermittency tend to have values of α smaller than 1.3. The parameter γ, on the other hand, is directly related to the magnitude of the fluctuations at each scale. Different values of γ at different directions might suggest anisotropy in the fluctuations, and therefore related to external forces that induce the movement and formation of the precipitating clouds. The self-similarity index H, which is the logarithmic slope of the power-law behavior of the scale parameter γ, denotes the rate of growth of the magnitude of the fluctuations. Large values of H were found in events that show small rain areas, intermittency and vast areas with drizzle. This observation suggests that H could be related to the degree of variability of the event, but also to some additional fractal measure of the areal intermittency.

Based on the inverse of the operation that was used to separate the wavelet coefficients from the rainfall field, a disaggregation algorithm was proposed that has as main input only one rain gauge measurement, a Low-resolution Rainfall (LR) field and the parameters that the define the scale-to scale dependence of the distributions of rainfall fluctuations.

The algorithm initiates by estimating the mean rainfall intensity for the analyzed area by expanding the rain gauge measurement into larger scales. The growth of intensity during the expansion, however does not have a boundary since the mean behavior at larger scales is unknown. This issue produces overestimations of the mean intensity under certain circumstances. An analysis of the variation of intensity between scales for several different rainfall patterns, showed that the growth is caused mainly by the degree of variability, the relative location of the station within the storm and the magnitude of the measurement. If a correction factor depending on these variables is applied between the scale at which the expansion starts and the scale of the LR field, the estimation of the mean intensity improves for a large percentage of the rainfall events.

A comparison of the mean intensity computed from the TRMM-LR field and the mean intensity computed from the Radar-AMeDAS showed that there is a tendency for overestimation if the mean intensity computed from TRMM is approximately larger than 1.8 mm/hr. This finding allows to fix a set of rules that help controlling the estimation of the mean intensity more accurately.

Once the mean intensity is estimated, the algorithm proceeds to disaggregate this value scale by scale until reaching the desired high-resolution. This procedure includes an analysis of the relationship between the local average and the local fluctuations of the rainfall field. It was found that extreme values of intensity produce extreme fluctuations, and lower values of intensity do not produce extreme fluctuations. This conditions help to limit the access to random samples at the time of choosing the right combination of wavelet coefficients that are used in the Inverse Discrete Wavelet Transform.

The results of validation was done using simple fidelity metrics to measure the first and second order moments of the matricial error. Acceptable metrics were observed in 78% of 954 events that encompass different rainfall patterns from different locations. The events in the remaining 22 % are mostly events in which the $S\alpha S$ distribution of wavelet coefficients have a characteristic exponent smaller than 1.30. These results demonstrate that the algorithm, as it is currently devised, is not able to entirely reproduce the sudden changes of intensity in events with characteristic exponent smaller than 1.30.

When using TRMM rainfall data as the LR input field, the results showed that the rain gauge station and the analysis of section 4.2.1.2 contribute to correct the existing bias. However, the reproduction of the actual rainfall event depends on how well the areas of high and low intensity are detected by the TRMM product.

The proposed procedure, in which the field is decomposed, is a first attempt to take advantage of the self-similarity in the distribution of rainfall fluctuations and the degree of variability expressed by the parameter α. Nevertheless, we consider that the framework has the potential for improvement if the inter-scale dependence of the rainfall fluctuations are analyzed differently (e.g., the work of Huang and Mumford, 1999).

Being able to estimate the mean intensity for a $256\text{km} \times 256\text{km}$ site from only one rain gauge is a remarkable accomplishment, and contributes to the correction of the bias of the LR field to some degree. Undoubtedly, including one or two more stations (which is a situation normally found in most regions of the world) may improve the performance of the algorithm.

6. REFERENCES

Akgiray, V., Lamoureux, C.G., 1989. Estimation of stable-law parameters: a comparative study. J. Bus. Econ. Stat. 7, 85–93.

Alfieri, L., Claps, P., D'Odorico, P., Laio, F., Over, T.M., 2008. An analysis of the soil moisture feedback on convective and stratiform precipitation. J. Hydrometeorol. 9, 280–291.

Anagnostou, E. N., Maggioni, V., Nikolopoulos, E. I., Meskele, T., Hossain, F., Papadopoulos, A., 2010. Benchmarking high–resolution global satellite rainfall products to radar and rain gauge rainfall estimates. IEEE Trans. Geosci. Remote Sens. 48, 1667–1683.

Arad, R.W., 1980. Parameter-estimation for symmetric stable distribution. Int. Econ. Rev. 21, 209-220.

Arnaud, P., Bouvier, C., Cisneros, L., Dominguez, R., 2002. Influence of rainfall spatial variability on flood prediction. J. Hydrol. 260, 216–230.

Arnaud, P., Lavabre, J., Fouchier, C., Diss, S. Javelle, P., 2011. Sensitivity of hydrological models to uncertainty of rainfall input. Hydrol. Sci. J. 56 (3), 397–410.

Besbeas, P., Morgan, B.J.T., 2008. Improved estimation of stable laws. Stat. Comput. 18, 219-231.

Bell, V. A., Moore, R. J., 2000. The sensitivity of catchment runoff models to rainfall data at different spatial scales. Hydrol. Earth Syst. Sc. 4 (4), 653–667.

Beven, K., 2001. How far can we go in distributed hydrological modelling? Hydrol. Earth Syst. Sc. 5 (1), I–12.

Borak, S., Härdle, W., Weron, R., 2011. Models for heavy-tailed asset returns. In: Čížek, P., Härdle, W., Weron, R. (eds.) Statistic tools for finance and insurance, second edition, 21-55. Springer, Berlin.

Bouchaud, J.P., Potters, M., Meyer, M., 2000. Apparent multifractality in finantial time series. Eur. Phys. J. B. 13, 595–599.

Brcich, R. F., Iskander, D. R., Zoubir, A. M., 2005. The stability test for symmetric alpha stable distributions. IEEE T. Signal Proces. 53, 977-986.

Chambers, J.M., Mallows, C.L., Stuck, B.W., 1976. A method for simulating stable random variables. J. Am. Stat. Assoc. 71, 340–344.

Dance, C. R., Kuruoğlu, E. E., 1999. Estimation of the parameters of skewed α -stable distributions. In: Nolan, J. P., Swami, A., (eds.) Proceedings of the ASA-IMS Conference on Heavy Tailed Distributions. Applications of heavy tailed distributions in economics, engineering and statistics. Washington, DC.

Daubechies, I., 1992. Ten lectures on wavelets, second edition. Society for industrial and applied mathematics. Philadelphia, Pennsilvania.

Deidda, R, 2000. Rainfall downscaling in a space-time multifractal framework. Water Resour. Res. 36 (7), 1779–1794.

Deidda, R., Badas, M. G., Piga, E., 2006. Space ‐ time multifractality of remotely sensed rainfall fields. J. Hydrol. 322, 2–13.

Dominicy, Y., Veredas, D., 2013. The method of simulated quantiles. J. Econometrics 172 (2), 235-247.

Ebert, E. E., Janowiak, J., Kidd, C., 2007. Comparison of near-real-time precipitation estimates from satellite observations and numerical models. Bull. Amer. Meteor. Soc. 88, 47–64.

Ebtehaj, M., Foufoula-Georgiou, E., 2011. Statistics of precipitation reflectivity images and cascade of Gaussian-scale mixtures in the wavelet domain: A formalism for reproducing extremes and coherent multiscale structures. J. Geophys. Res. 116 (D14110).

Ebtehaj, A. M., Foufoula-Georgiou, E., Lerman, G., 2012. Sparse regularization for precipitation downscaling. J. Geophys. Res. 116, D22110.

Embrechts, P., Klüppelberg, C., Mikosch, T., 1997. Modelling Extremal Events for Insurance and Finance. Springer-Verlag, Berlin, Germany.

Fama, E.F. Roll, R., 1971. Parameter estimates for symmetric stable distributions. J. Am. Stat. Assoc. 66 (334), 331-338.

Ferraris, L., Gabellani, S., Rebora, N., Provenzale, A., 2003. A comparison of stochastic models for spatial rainfall downscaling. Water Resour. Res. 39 (12), 1368.

Fofack, H., Nolan, J.P., 1999. Tail behavior, modes and other characteristics of stable distributions. Extremes 2 (1), 39–58.

Foufoula-Georgiou, E., Ebtehaj, A. M., Zhang, S. Q., Hou, A. Y., 2014. Downscaling satellite precipitation with emphasis on extremes: A variational ℓ_1-Norm regularization in the derivative domain. Surv. Gephys. 35, 765–783.

Gebregiorgis, A. S., Hossain, F., 2012. Understanding the dependency of satellite rainfall uncertainty on topography and climate for hydrologic model simulation. IEEE Trans. Geosci. Remote Sens. 51, 704–718.

Gebregiorgis, A. S., Hossain, F., 2013. Estimation of satellite rainfall error variance using readily available geophysical features. IEEE Trans. Geosci. Remote Sens. 52, 288–304.

Guillot, G., Lebel, T., 1999. Disaggregation of Sahelian mesoscale convective system rain fields: Further developments and validation, J. Geophys. Res. 104, 31533– 31551.

Gupta, V. K., Waymire, E. 1990. Multiscaling properties of spatial rainfall and river flow distributions. J. Geophys. Res. 95 D3, 1999–2009.

Gupta, V.K., Waymire, E., 1993. A statistical analysis of mesoscale rainfall as a random cascade. J. Appl. Meteorol. 32, 251–267.

Haberlandt, U., 2007. Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large–scale extreme rainfall event. J. Hydrol. 332, 144–157.

Harris, D., Menabde, M., Seed, A., Austin, G., 1996. Multifractal characterization of rain fields with a strong orographic influence. J. Geophys. Res.: Atmos. 101 (D21), 26, 405–414.

Harris, D., Foufoula–Georgiou, E., 2001. Subgrid variability and stochastic downscaling of modeled clouds: Effects on radiative transfer computations for rainfall retrieval. J. Geophys. Res. 106 (D10), 10349–10362.

Harris, D., Foufoula–Georgiou, E., Droegemeier, E. E., Levit, J. J., 2001. Multiscale statistical properties of a high-resolution precipitation forecast, J. Hydrometeorol. 2 (4), 406–418.

Houze, R. A. Jr., 1993. Cloud Dynamics. Academic Press, San Diego, California.

Huang, J., Mumford, D., 1999. Statistics of natural images and models. IEEE Conference on Computer Vision and Pattern Recognition, IEEE, New York.

Huffman, G. J., and Coauthors, 2007. The TRMM Multisatellite Precipitation Analysis (TMPA): Quasi-global, multiyear, combined-sensor precipitation estimates at fine scales. J. Hydrometeor. 8, 38–55.

Huffman, G. J., Adler, R. F., Bolvin, D. T., Nelkin, E. J., 2010. The TRMM Multi-Satellite Precipitation Analysis (TMPA). In: Gebremichael, M., Hossain, F., (Eds.) Satellite Rainfall Applications for Surface Hydrology. Springer, 3–22.

Jiang, H., Zipser, E.J., 2010. Contribution of tropical cyclones to the global precipitation from eight seasons of TRMM data: regional, seasonal, and interannual variations. J. Climate 23, 1526–1543.

Joseph, R., Ting, M., 2000. Multiple-scale spatio-temporal variability of precipitation over the coterminous United States. J. Hydrometeorol. 1 (5), 373–392.

Kamahori, H., 2012. Mean features of tropical cyclone precipitation from TRMM/3B42. SOLA 8, 017–020.

Klein, P. M., Harr, P. A., Elsberry, R. L., 2000. Extratropical transition of Western North Pacific Tropical Cyclones: An overview and conceptual model of the transformation stage. Weather Forecast. 15 (4), 373–395.

Kogon, S.M., Williams, D.B., 1998. Characteristic function based estimation of stable parameters. In: Adler, R.J., Feldman, R.E., Taqqu, M.S. (eds.) A practical guide to heavy tails, 311-335. Birkhäuser, Boston.

Koutrouvelis, I., 1980. Regression-type estimation of the parameters of stable laws. J. Am. Stat. Assoc. 75, 918-928.

Kumar, P., Foufoula-Georgiou, E., 1993a. A multicomponent decomposition of spatial rainfall fields: 1. Segregation of large- and small-scale features using wavelet transforms. Water Resour. Res., 29 (8), 2515–2532.

Kumar, P., Foufoula-Georgiou, E., 1993b. A multicomponent decomposition of spatial rainfall fields: 2. Self-similarity in fluctuations. Water Resour. Res. 29 (8), 2533–2544.

Kusaka, H., Kitahata, H., 2009. Synoptic-scale climatology of cold frontal precipitation systems during the passage over the Central Japan. SOLA 5, 061–064.

Lepore, C., Veneziano, D., Molini, A., 2015. Temperature and CAPE dependence of rainfall extremes in the Eastern United States. Geophys. Res. Lett. 42, doi:10.1002/2014GL062247.

Liang, Y., Chen, W., 2013. A survey on computing Lévy stable distributions and a new MATLAB toolbox. Signal Process. 93, 242-251.

Liu, C., Zipser, E.J., Cecil, D.J., Nesbitt, S.W., Sherwood, S., 2008. A cloud and precipitation feature database from nine years of TRMM observations. J. Appl. Meteorol. Clim. 47, 2712–2728.

Liu, C., Zipser, E.J., 2009. "Warm Rain" in the trpics: Seasonal and regional distributions based on 9 yr of TRMM data. J. Climate 22, 767–779.

Liu, C., 2011. Rainfall contributions from precipitation systems with different sizes, convective intensities, and durations over the tropics and subtropics. J. Hydrometeorol. 12, 394–411.

Liu, J., Bray, M., Han, D., 2012. Sensitivity of the weather research and forecasting (WRF) model to downscaling ratios and storm types in rainfall simulation. Hydrol. Process. 26, 3012–3031.

Lovejoy, S., 1982. Area perimeter relations for rain and cloud areas. Science 187, 1035–1037.

Lovejoy, S., Mandelbrot, B. B., 1985. Fractal properties of rain, and a fractal model. Tellus 37A, 209-232.

Lovejoy, S., Schertzer, D., 1986. Scale invariance, symmetries, fractals and stochastic simulation of atmospheric phenomena, Bull Am. Meteorol. Soc. 67, 21–32.

Lovejoy, S., Schertzer, D., 1990. Multifractals, universality classes and satellite and radar measurements of cloud and rain fields. J. Geophys. Res. 95 (D3), 2021–2034.

Lovejoy, S., Schertzer, D., 2006. Multifractals, cloud radiances and rain, J. Hydrol., 322 (1–4), 59–88.

Lovejoy, S., Schertzer, D., 2013. The weather and climate: Emergent laws and multifractal cascades. Cambridge University Press, New York.

Lovejoy, S., Schertzer, D., Allaire, V. C., 2008. The remarkable wide range spatial scaling of TRMM precipitation. Atmos. Res. 90, 10–32.

Maggioni, V., Sapiano, M. R. P., Adler, R. F., 2016. Estimating uncertaintities in high–resolution satellite precipitation products: Systematic or random error? J. Hydrometeorol. 17, 1119–1129.

Mallat, S., 1989. A theory for multiresolution signal decomposition: The wavelet representation. IEEE T. Pattern Anal. 11,674–693.

Mallat, S., 2009. A Wavelet Tour of Signal Processing: The Sparse Way. Academic, San Diego, California.

Mandelbrot, B., 1963. The variation of certain speculative prices. J. Business 36 (4), 394-419.

Mandelbrot, B., 1982. The fractal geometry of nature. W.H. Freeman and Company, New York.

Matsui, M., Takemura, A., 2006. Some improvements in numerical evaluation of symmetric stable density and its derivatives. Commun. Stat. Theory 35 (1), 149-172.

Matsui, M., Takemura, A., 2008. Goodness-of-fit tests for symmetric stable distributions—Empirical characteristic function approach. Test 17, 546-566.

Matsui, M., Pawlaz, Z., 2014. Fractional absolute moments of heavy tailed distributions. ArXiv e-prints. 1301.4804v2.

McCulloch, J.H., 1986. Simple consistent estimators of stable distribution parameters. Commun. Stat. Simul. Comput. 15, 1109-1136.

Menabde, M., Harris, D., Seed, A., Austin, G., Stow, D., 1997. Multiscaling properties of rainfall and bounded random cascades. Water Resour. Res. 33, 2823–2830.

Menabde, M., Seed, A., Harris G., Austin, G., 1999. Multiaffine random field model of rainfall. Water Resour. Res. 35, 509– 514.

Menn, C., Rachev, S. T., 2006. Calibrated FFT-based density approximations for α-stable distributions. Comput. Stat. Data An. 50 (8), 1891–1904.

Mittnik S., Doganoglu T., Chenyao D., 1999. Computing the probability density function of the stable paretian distribution. Math. Comput. Model 29, 235–240.

Mittnik, S., Rachev, S.T., Doganoglu, T., Chenyao, D., 1999. Maximum likelihood estimation of stable paretian models. Math. Comput. Model 29, 275–293.

Nason, G. P., Silverman, B. W., 1995. The stationary wavelet transform and some statistical applications. Lect. Notes Stat. 103, 281–299.

Nikias, C. L., Shao, M., 1995. Signal Processing with Alpha-Stable Distributions and Applications. Wiley-Interscience, New York.

Nogueira, M., Barros, A.P., Miranda, P.M., 2013. Multifractal properties of embedded convective structures in orographic precipitation: toward subgrid-scale predictability. Nonlinear Process. Geophys. 20 (5), 605–620.

Nogueira, M., Barros, A.P., 2016. Transient stochastic downscaling of quantitative precipitation estimates for hydrological applications. J. Hydrol. 529, 1407–1421.

Nolan, J.P., 1997. Numerical calculation of stable densities and distribution functions. Commun. Stat. Stoch. Models 13 (4), 759-774.

Nolan, J.P., 1999. Fitting data and assessing goodness-of-fit with stable distributions. In: Nolan, J.P., and Swami, A. (eds.) Proceedings of the ASA-IMS Conference on Heavy

Tailed Distributions. Applications of heavy tailed distributions in economics, engineering and statistics. Washington, DC.

Nolan, J.P., 2001. Maximum likelihood estimation and diagnostics for stable distributions. In: Barndorff-Nielsen, O.E., Mikosch, T., Resnick, S. (eds.) Lévy Processes. Birkhäuser, Boston.

Nolan, J.P., 2013. Multivariate elliptically contoured stable distributions: theory and estimation. Computation Stat. 28 (5), 2067-2089.

Nykanen, D. K. 2008. Linkages between orographic forcing and the scaling properties of convective rainfall in mountainous regions. J. Hydrometeorol. 9, 327–347.

Nykanen, D. K., Foufoula-Georgiou, E., 2001. Impact of small-scale rainfall variability on larger-scale spatial organization of land-atmosphere fluxes. J. Hydrometeorol. 2 (2), 105–121.

Nykanen, D. K., Harris, D., 2003. Orographic influences on the multiscale statistical properties of precipitation. J. Geophys. Res. 108 (D8), 8381.

Ohba, M., Kadokura, S., Yoshida, Y., Nohara, D., Toyoda, Y., 2015. Anomalous weather patterns in relation to heavy precipitation events in Japan during the Baiu Season. J. Geophys. Res. 16 (D2), 688–701.

Over, T.M., Gupta, V.K., 1994. Statistical analysis of mesoscale rainfall: dependence of a random cascade generator on large scale forcing. J. Appl. Meteorol. 33, 1526–1542.

Over, T.M., Gupta, V.K., 1996. A space-time theory of mesoscale rainfall using random cascades. J. Geophys. Res. 101 (D21), 26319–26331.

Parodi, A., Foufoula–Georgiou, E., Emanuel, K., 2011. Signature of microphysics on spatial rainfall statistics, J. Geophys. Res. 116 (D14119), doi:10.1029/2010JD015124.

Perica, S., Foufoula–Georgiou E., 1996a. Linkage of scaling and thermodynamic parameters of rainfall: Results from midlatitude mesoscale convective systems. J. Geophys. Res. 101 (D3), 7431–7448.

Perica, S., Foufoula–Georgiou E., 1996b. Model for multiscale disaggregation of spatial rainfall based on coupling meteorological and scaling descriptions. J. Geophys. Res. 101 (D21), 26347–26361.

Pombo S,. Proença de Oliveira, R., Mendes, A., 2015. Validation of remote–sensing precipitation products for Angola. Meteorol. Appl. 22, 395–409.

Rebora, N., von Ferraris, L., Hardenberg, J., Provenzale, A., 2006. RAINFARM: rainfall downscaling by a filtered autoregressive model. J. Hydrometeor. 7, 724–738.

Ruelland, D., Ardoin–Bardin, S., Billen, G., Servat, E., 2008. Sensitivity of a lumped and semi-distributed hydrological model to several methods of rainfall interpolation on a large basin in West Africa. J. Hydrol. 361, 96–117.

Samorodnitsky, G., Taqqu, M.S., 1994. Stable non-Gaussian random processes: Stochastic models with infinite variance. Chapman and Hall, New York.

Sapiano, M. R. P., Arkin, P. A., 2009. An intercomparison and validation of high-resolution satellite precipitation estimates with 3-hourly gauge data. J. Hydrometeor. 10, 149–166.

Sapozhnikov, V.B., Foufoula-Georgiou, E, 2007. An exponential Langevin-type model for rainfall exhibiting spatial and temporal scaling. Nonlinear Dyn. Geosci. 87–100.

Schertzer, D., Lovejoy, S., 1987. Physical modeling and analysis of rain and clouds by anisotropic scaling of multiplicative processes. J. Geophys. Res. 92, 9693–9714.

Schertzer, D., Lovejoy, S., 2011. Multifractals, generalized scale invariance and complexity in geophysics. Int. J. Bifurcat. Chaos 21 (12), 3417–3456.

Schuurmans, J., Bierkens, M., 2007. Effect of spatial distribution of daily rainfall on interior catchment response of a distributed hydrological model. Hydrol. Earth Syst. Sci. 11, 677–693.

Smith, M. B., Koren, V. I., Zhang, Z., Reed, S. M., Pan, J. J., Moreda, F., 2004. Runoff response to spatial variability in precipitation: an analysis of observed data, J. Hydrol. 298 (1-4), 267–286.

Song, X., Zhang, J., Zhan, C., Xuan, Y., Ye, M., Xu, C., 2015. Global sensitivity analysis in hydrological modeling: Review of concepts, methods, theoretical framework, and applications. J. Hydrol. 523, 739–757.

Tang, L., Hossain, F., 2009. Transfer of satellite rainfall error from gaged to ungaged locations: How realistic will it be for the Global Precipitation Mission? Geophys. Res. Lett. 36, L10405.

Tao, K., Barros, A.P., 2010. Using fractal downscaling of satellite precipitation products for hydrometeorological applications. J. Atmos. Oceanic Technol. 27, 409–427.

Tessier, Y., Lovejoy, S., Schertzer, D., 1993. Universal multifractals: theory and observations for rain and clouds. J. Appl. Meteorol. 32, 223–250.

Thatcher, L., Takayabu, Y.N., Yokoyama, C., Pu, Z., 2012. Characteristics of the tropical cyclone precipitation features over the western Pacific warm pool. J. Geophys. Res. 117, D16208, doi:10.1029/2011JD017351.

Tian, Y., Peters-Lidard, C. D., Choudhury, B. J., Garcia, M., 2007. Multitemporal analysis of TRMM-based satellite precipitation products for land data assimilation applications. J. Hydrometeor. 8, 1165–1183.

Tian, Y., Peters-Lidard, C. D., Eylander, J. B., Joyce, R. J., Huffman, G. J., Adler, R. F., Hsu, K., Turk, F. J., Garcia, M., Zeng, J., 2009. Component analysis of errors in satellite-based precipitation estimates. J. Geophys. Res. 114, D24101.

Tian, Y., Huffman, G. J., Adler, R. F., Tang, L., Sapiano, M., Maggioni, V., Wu, H., 2013. Modeling errors in daily precipitation measurements: Additive or multiplicative? Geophys. Res. Lett. 40, 2060–2065.

Uchaikin, V. V., Zolotarev, V. M., 1999. Chance and stability, stable distributions and their applications, second ed. VSP, Utrecht, Netherlands.

Veillette, M., 2012. MATLAB toolbox: alpha-stable distributions in MATLAB. URL: http://math.bu.edu/people/mveillet/html/alphastablepub.html.

Veneziano, D., Bras, R. L., Niemman, J., D., 1996. Nonlinearity and self-similarity of rainfall in time and a stochastic model. J. Geophys. Res. 101, 26371–26392.

Venugopal, V., Foufoula-Georgiou, E., Sapozhnikov, V., 1999a. A space-time downscaling model for rainfall. J. Geophys. Res. 104 (D16), 19705–19721.

Venugopal, V., Foufoula-Georgiou, E., Sapozhnikov, V., 1999b. Evidence of dynamic scaling in space-time rainfall. J. Geophys. Res. 104 (D24), 31599–31610.

Verrier, S., de Montera, L., Barths, L., Mallet, C., 2010. Multifractal analysis of African monsoon rain fields, taking into account the zero rain-rate problem. J. Hydrol. 389 (12), 111–120.

Weron, R., 1996. On the Chambers-Mallows-Stuck method for simulating skewed stable random variables. Stat. Probabil. Lett. (28), 165–171.

Weron, R., 2011. Computationally intensive value at risk calculations. In: Gentle, J. E., Härdle, W., Mori, Y. (eds.) Handbook of computational statistics, second edition, 911–950. Springer, Berlin.

Weijian, G., Chuanhai, W., Xianmin, Z., Tengfei, M., Hai, Y., 2015. Quantifying the spatial variability of rainfall and flow routing on flood response across scales. Environ. Earth Sci. 74, 6421–6430.

Yokoyama, C., Takayabu, Y.N., 2012. Relationships between rain characteristics and environment. Part I: TRMM precipitation features and the large-scale environment over the tropical Pacific. Mon. Weather Rev. 140, 2831–2840.

Younger, P. M., Freer, J. E., Beven, K. J., 2009. Detecting the effects of spatial variability of rainfall on hydrological modelling within an uncertainty analysis framework. Hydrol. Process. 23, 1988–2003.

Yoshida, A., Asuma, Y., 2004. Structures and environment of explosively developing extratropical cyclones in the northwestern Pacific region. Mon. Weather Rev. 132 (5) 1121–1142.

Zhang, S., Foufoula-Georgiou, E., 1997. Subgrid-scale rainfall variability and its effects on atmospheric and surface variable predictions. J. Geophys. Res. 102 (D16), 19559–19573.

Zhou, Y., Lau, W.K.M., Liu, C., 2013. Rain characteristics and large-scale environments of precipitation objects with extreme rain volumes from TRMM observations. J. Geophys. Res-Atmos. 118, 9673–9689, doi:10.1002/jgrd.50776.

Zolotarev, V.M., 1986. One-dimensional stable distributions, Translations of Mathematical Monographs, Vol. 65. American Mathematical Society, Providence, RI. (Transl. of the original 1983 russian).