Variability of the Rain Drop Size Distribution: Stochastic Simulation and Application to Telecommunication Microwave Links

THÈSE N^{O} 5345 (2012)

PRÉSENTÉE LE 20 AVRIL 2012

À LA FACULTÉ DE L'ENVIRONNEMENT NATUREL, ARCHITECTURAL ET CONSTRUIT
LABORATOIRE DE TÉLÉDÉTECTION ENVIRONNEMENTALE
PROGRAMME DOCTORAL EN ENVIRONNEMENT

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Marc SCHLEISS

acceptée sur proposition du jury:

Prof. M. Parlange, président du jury
Prof. A. Berne, directeur de thèse
Dr C. Frei, rapporteur
Prof. W. F. Krajewski, rapporteur
Prof. A. Rinaldo, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2012

Below the Sun above the ground,
inside a cloud that's where I'm found.
Then thunder rolling, lightning flashes,
breaking free in downward dashes.

Falling quickly to Earth,
flying the way the wind blows.
While off in the distance
arise sparkling rainbows.

Then suddenly stopping with a splashing sound,
softly bathing the Earth and all around.
Rushing together in puddles we lay,
as the children come out to play.

Jumping, splashing, swishing,
making waves and ripples ever so neat.
As we caress and wash over their tiny feet.

Then the Sun begins to heat the day
and our puddles start to fade away.
A vapor, a mist as upward we begin to soar,
gathering together again in the sky as before.

Below the Sun above the ground
inside a cloud once again I'm found.

— Robert Renfro (2010)

To my parents, my friends and the love of my life ...

Acknowledgements

The author would like to acknowledge the financial support by the Swiss National Science Foundation (SNSF), grants 200021-118057 and 200020-132002.

The author also thanks the jury members for their precious time and their helpful comments and suggestions. Special thanks to Alexis Berne, who did an excellent job supervising his first PhD students and showing them all the beauty and excitement of research.

Lausanne, the 8th of February 2012 M. S.

Abstract

Precipitation is an important component of the Earth's water cycle and needs to be carefully monitored. Its large variability over a wide range of spatial and temporal scales must be taken into account. For example, hydrological models require accurate rainfall estimates at high spatial and temporal resolutions (e.g., 1 km and 5 min or higher). Obtaining accurate rainfall estimates at these scales is known to be difficult. So far, the only instruments capable of measuring rainfall over extended domains at such resolutions are weather radars. Their estimates are, however, affected by large errors and uncertainties partly due to the spatial and temporal variability of the drop size distribution (DSD). Major progress in the field is slowed down by the lack of knowledge about the spatial and temporal variability of DSD at scales that are relevant in remote sensing. This lack of reference data can be addressed through two different methods : (1) experimental investigations and (2) stochastic simulation.

In this thesis, a comprehensive framework for the stochastic simulation of DSD fields at high spatial and temporal resolutions is proposed. The method is based on Geostatistics and uses variograms to describe the spatial and temporal structures of the DSD. The simulator's ability to generate large numbers of DSD fields sharing the same statistical properties provides a very useful theoretical framework that nicely complements experimental approaches based on large networks of weather sensors. To illustrate its potential, the simulator is applied to different rain events and validated using data from a network of disdrometers at EPFL. The results show that the simulator is able to reproduce realistic spatial and temporal structures that are in adequacy with ground measurements.

The second part of this thesis focuses on the simulation and parametrization of intermittency (i.e., the alternating between dry/rainy periods). Simple scaling functions that can be used to downscale/upscale intermittency at different spatial and temporal resolutions are proposed and used to parametrize a new disaggregation method that includes the DSD as an output. Finally, different methods to identify dry and rainy periods and to quantify rainfall intermittency using telecommunication microwave links are proposed. The false dry/wet classification error rates of each method are estimated using data from a new and innovative experimental set-up located in Dübendorf, Switzerland. The results show that the dry/wet classification is significantly improved when data from multiple channels are used.

Keywords : precipitation microphysics, raindrop size distributions, stochastic simulation, intermittency, telecommunication microwave links.

Résumé

La pluie est une importante composante du cycle hydrologique sur Terre et doit être mesurée avec grande précision. Sa grande variabilité à plusieurs échelles spatiales et temporelles doit être prise en compte. Pour l'instant, seuls les radars météo sont capables de mesurer la pluie sur de grands domaines à des échelles pertinentes en hydrologie (e.g., 1 km et 5 min ou davantage). Leurs estimations sont cependant affectées par de grandes erreurs et incertitudes dues à la variabilité de la microstructure des précipitations (DSD ci-dessous). L'ordre de grandeur de ces erreurs est difficile à quantifier à cause du manque d'information sur la variabilité de la DSD à ces échelles. Ce manque d'information peut être comblé (1) à l'aide d'une approche expérimentale ou (2) à l'aide une approche basée sur la simulation.

Dans cette thèse, un cadre théorique pour la simulation stochastique de champs de DSD est proposé. La méthode se base sur des Géostatistiques et utilise des variogrammes pour décrire la structure spatiale et temporelle de la DSD. La capacité du simulateur de générer de grandes quantités de champs de DSD avec les mêmes propriétés statistiques constitue un cadre théorique très utile qui complète des approches plus expérimentales basées sur de larges réseaux de capteurs météo. Afin d'illustrer le potentiel de l'approche, le simulateur est appliqué à différents événements pluvieux et évalué à l'aide de données issues d'un réseau de disdromètres à l'EPFL. Les résultats montrent que le simulateur est capable de reproduire des structures spatiales et temporelles réalistes qui sont cohérentes avec des mesures prises au sol.

La seconde partie de cette thèse se concentre sur la simulation et la paramétrisation de l'intermittence (i.e., l'alternance de périodes sèches/pluvieuses). Des modèles de changement d'échelle permettant de représenter l'intermittence à différentes échelles spatiales et temporelles sont proposés et utilisés pour paramétrer une nouvelle méthode de désagrégation qui fournit la DSD en sortie. Finalement, différentes méthodes permettant d'identifier des périodes sèches/pluvieuses à l'aide de faisceaux hertziens de télécommunication sont proposées. Les erreurs de classification sont évaluées pour chaque méthode grâce aux données d'un nouveau dispositif expérimental installé à Dübendorf. Les résultats montrent que la classification peut être améliorée si les données de plusieurs canaux sont combinées.

Mots clés : microphysique des précipitations, DSD, simulation stochastique, intermittence, faisceaux hertziens de télécommunication.

Contents

List of Figures

List of Tables

List of symbols

$A_B(t)$	attenuation baseline at time t [dB]	95
A_h	specific attenuation on propagation at horizontal polarization [$\mathrm{dB\,km^{-1}}$]	8
$A_R(t)$	rain-induced attenuation at time t [dB]	95
A_v	specific attenuation on propagation at vertical polarization [$\mathrm{dB\,km^{-1}}$]	8
$A(t)$	total attenuation at time t [dB]	95
D	equivolume spherical drop diameter [mm]	4
D_0	median volume drop diameter [mm]	6
D_m	average drop diameter [mm]	6
D_{max}	maximum drop diameter [mm]	6
D_{min}	minimum drop diameter [mm]	6
E	kinetic energy rate [$\mathrm{J\,m^{-2}\,h^{-1}}$]	7
$I(x,t)$	rainfall intermittency [-]	38
K	dielectric factor of water [-]	10
K_{dp}	specific differential phase shift on propagation [$^\circ\,\mathrm{km^{-1}}$]	8
m	complex refractive index of water [-]	7
N_0	concentration parameter of the DSD [$\mathrm{mm^{-1}\,m^{-3}}$]	5
N_t	concentration parameter of the DSD [$\mathrm{m^{-3}}$]	6
$N(D)$	drop size distribution [$\mathrm{m^{-3}\,mm^{-1}}$]	6
$\tilde{N}_t$	concentration parameter of the DSD after anamorphosis [-]	24
p	percentage of rainy pixels at the fine scale [-]	63
$p(k,\tau)$	probability that an area of size k remains dry during τ hours [-]	83
P	percentage of rainy blocks at the coarse scale [-]	64
$\bar{P}_r$	average power received from a target at range r [W]	10
R	rain rate [$\mathrm{mm\,h^{-1}}$]	2
r_{ab}	drop axis ratio [-]	7
S_{hh}	forward scattering amplitude at horizontal polarization [m]	8
S_{vv}	forward scattering amplitude at vertical polarization [m]	8
$S_W(t)$	variability of link signal at time t [dB]	95
$v(D)$	terminal fall speed of a drop of diameter D [$\mathrm{m\,s^{-1}}$]	7
w	wavelength [cm]	7
Z	radar reflectivity factor [$\mathrm{mm^6\,m^{-3}}$]	10
Z_{dr}	differential reflectivity [dB]	8

Z_h	radar reflectivity at horizontal polarization [$\mathrm{mm^6\,m^{-3}}$]	7
Z_v	radar reflectivity at vertical polarization [$\mathrm{mm^6\,m^{-3}}$]	7
γ	variogram	23
γ_S	spatial variogram	39
γ_T	temporal variogram	39
Λ	scale parameter of the DSD [$\mathrm{mm^{-1}}$]	5
$\tilde{\Lambda}$	scale parameter of the DSD after anamorphosis [-]	24
μ	shape parameter of the DSD [-]	6
$\tilde{\mu}$	shape parameter of the DSD after anamorphosis [-]	24
$\rho_S(h,\tau)$	spatial autocorrelation of rainfall intermittency [-]	86
$\rho_T(\tau,k)$	temporal autocorrelation of rainfall intermittency [-]	87
ρ_w	density of liquid water [$\mathrm{kg\,m^{-3}}$]	7
σ_0	rain detection threshold [dB]	95
$\sigma_{B_h}(D)$	backscattering cross section at horizontal polarization [$\mathrm{cm^2}$]	7
$\sigma_{B_v}(D)$	backscattering cross section at vertical polarization [$\mathrm{cm^2}$]	7
$\sigma_{E_h}(D)$	extinction cross section at horizontal polarization [$\mathrm{cm^2}$]	8
$\sigma_{E_v}(D)$	extinction cross section at vertical polarization [$\mathrm{cm^2}$]	8

List of acronyms

AMSL	Above Mean Sea Level	46
DSD	Drop Size Distribution	2
EPFL	École Polytechnique Fédérale de Lausanne	45
FG	Factor Graph	106
GPM	Global Precipitation Measurement	2
HIRE	Hydromet Integrated Radar Experiment	25
MSM	Multivariate Markov Switching model	109
MSU	Univariate Markov Switching model	107
MW	Moving Window	106
MWL	MicroWave Link	11
PARSIVEL	PARticle SIze and VELocity	45
PIA	Path-Integrated Attenuation	105
RADAR	RAdio Detecting And Ranging	10
SGS	Sequential Gaussian Simulation	43
SIS	Sequential Indicator Simulation	43
SNMP	Simple Network Management Protocol	110
ST	Simple Threshold	106
TRMM	Tropical Rainfall Measurement Mission	2

1 Introduction

1.1 Motivation

There is probably nothing better than drinking a cup of hot tea (or coffee if you prefer) while listening to the sound of the raindrops falling against the window during a rainy Sunday afternoon. Watching millions of tiny drops, so small they can hardly be seen, landing onto leaves, splashing on the streets, flowing into rivers and gathering in boundless seas. Imagining them, rising to the sky only to fall again, continuing their endless journey on ever changing paths. A journey always beginning, yet without start. A journey with many ends, yet no real end. The sheer complexity of the hydrological cycle: a never ending movement of water on, above and below the surface of the earth. And precipitation is a key component of it.

From the scientific point of view, precipitation is a very difficult process to model and predict. It results from a combination of many different physical mechanisms (e.g., nucleation, diffusional growth, collisional growth, evaporation and breakup) that are all fairly well understood on their own (Pruppacher and Klett, 1997), but remain poorly documented as a whole, dynamical interacting system (Khain et al., 2000). It is a fundamentally discrete process, consisting of a (very) large number of falling hydrometeors (e.g., raindrops, snow flakes or ice crystals) with different positions, sizes, shapes and fall speeds. Unfortunately, modeling and simulating such an interacting particle system at its highest level of detail remains largely infeasible, even for the fastest state-of-the-art supercomputers. In fact, the only reasonable approach for efficiently describing precipitation at larger scales is by considering statistics and average properties over sufficiently large volumes and time intervals. Knowledge of the exact positions, sizes and fall speeds of the individual hydrometeors is then no longer necessary. Small scale characteristics and variabilities are summarized and represented by means of probability distributions and stochastic processes in space and time. This process is called parameterization.

For hydrologists, the major parameter of interest is the rain rate R , i.e., the average mass flux density of water (expressed in $\mathrm{mm\,h^{-1}}$) falling over a certain area. The total rain amount (expressed in mm) is then obtained by integrating the rain rate in time. Although it is very important in many applications, the rain rate is not the only and not necessarily the most suitable quantity to represent rainfall at the macroscopic level. In soil erosion problems for example, the average kinetic energy of rainfall (Rosewell, 1986), a quantity closely related to the size and speed of the raindrops impacting the soil, can be more important than the rain rate. In remote sensing of precipitation, precise knowledge of the number, shape and size distributions of the raindrops (which determine the interactions between rainfall and electromagnetic waves) is also more important than the the rain rate. Because of this wide range of applications and interests, it can sometimes be preferable to work with alternative, more general and more detailed descriptions of rainfall, such as the (rain)drop size distribution (DSD hereinafter).

The DSD is a statistical description of the microstructure of rainfall. It describes the expected number of drops with equivolume spherical diameters between D and $D + dD$ per unit volume ($\mathrm{m^3}$) of air, independently of the positions, shapes and fall speeds of the considered drops. More specifically, it involves a combination of two different quantities: (1) the spatial concentration of the raindrops in the sampling volume and (2) the probability distribution of the raindrop sizes. The spatial concentration of raindrops is very variable and covers several orders of magnitude, from 0 (no rainfall) to more than 10^4 per $\mathrm{m^3}$. Similarly to the drop concentration, drop sizes also exhibit significant variability in space and time. Their size distribution depends both on the local climatology and the type of precipitation. Most of the drops observed in temperate regions have diameters between 0.1 and 3 mm. Larger drops are possible (up to 6 or 7 mm in diameters) but very rare and usually associated to strong convective rainfall events. It is also important to mention that because of the physical mechanisms responsible for drop formation, there are always significantly more small drops than large ones (Marshall and Palmer, 1948).

Because it reflects the fundamental microphysical processes involved in the formation of rain, knowledge of the drop size distribution is essential in many hydrological and meteorological applications. It is, for example, intensively used in remote sensing of precipitation using ground based weather radars (e.g., Battan, 1973; Wilson and Brandes, 1979; Jameson, 1989; Bringi et al., 1990; Ulbrich and Atlas, 1998; Krajewski and Smith, 2002) and spaceborne weather satellites like TRMM (Simpson et al., 1988; Iguchi et al., 2000; Robertson et al., 2003) and the upcoming global precipitation measurement mission GPM (http://gpm.nasa.gov). Its importance has also been acknowledged in other research areas like cloud and precipitation microphysics (Srivastava, 1971; Sauvageot and Lacaux, 1995; Testud et al., 2001), soil detachment, landslides and erosion problems (Rosewell, 1986; Finlay et al., 1997; Van Dijk et al., 2002; Kinnell, 2005), aerosol scavenging and atmospheric deposition processes (Radke et al., 1980; Beier and Hansen, 1993; Chate et al., 2003), and rainfall estimation using microwave links (Rincon and Lang, 2002; Rahimi et al., 2003; Krämer et al., 2005; Messer et al., 2006; Leijnse et al., 2007b).

The fundamental problem with rainfall (and with DSD) is its large variability across a wide range of different spatial and temporal scales (Koster and Suarez, 1995; Roe, 2005). This variability is a significant source of uncertainty in hydrological and remote sensing applications and needs to be carefully investigated. At the synoptic scale (several hundreds of kilometers), the variability of precipitation is relatively low. At this scale, the rainfall process mostly depends on large storm systems and weather fronts determined by well-known global circulation patterns. Rainfall that is observed at the meso-scale (5-50 km) exhibits much higher variability and spatial heterogeneity. At this scale, the rainfall process is mostly characterized by the birth, growth and decay of individual rain cells and rain bands with different lifetimes, possibly enhanced by orographic effects. The largest variabilities, however, are at the microscale. At this scale, rainfall is "patchy" and chaotic and mostly controlled by turbulence and complex microphysical processes (Jameson and Kostinski, 2001; Lovejoy and Schertzer, 2008). At very small scales (in the order of a few m^3 and a few seconds), the large sampling uncertainties (due to the discrete nature of rainfall) further enhance this variability and make it particularly difficult to obtain reliable rainfall estimates.

Unfortunately, there is currently no perfect method for accurately measuring rainfall at scales that are relevant for most hydrological and meteorological applications (e.g., between the microscale and the meso-scale). Rain gauges and disdrometers provide good point estimates but are not representative over larger domains. High density rain gauge networks provide good spatial estimates but they are relatively expensive and difficult to set up in highly populated areas or complex terrain. Microwave radiometers and microwave links provide interesting measurements at the path scale, but their estimates are affected by large uncertainties and do not capture the full spatial variability either. So far, only weather radars are capable of providing rainfall estimates over large domains at (relatively) high spatial and temporal resolutions. Quantitative estimation of rainfall from radar observations is, however, a very difficult process that involves a large number of issues (Krajewski and Smith, 2002). For hydrological and meteorological applications based on weather radar, it is essential to characterize these different sources of uncertainty and to quantify their impact on the quality (i.e., bias and random errors) of the final rainfall products.

One way to assess the uncertainties associated to radar rainfall estimation is by comparing radar rain-rate estimates to surface data collected using nearby rain gauges (Collier, 1986). Such comparisons are interesting but suffer from several fundamental limitations (e.g., scale incompatibility and vertical variability) discussed in detail in Zawadzki (1975). Another, more theoretical way to study the error structure and contributions related to radar rainfall estimation is to use simulation. The motivation behind the simulation approach is that most of the progress in hydrological modeling and radar remote sensing of precipitation is slowed down by the lack of accurate reference data at relevant scales. Simulation is capable of addressing this problem by providing large amounts of known reference data at high spatial and temporal resolutions. It also offers the advantage of

reproducibility, i.e., the fact that alternative realizations and rain events with similar statistical properties can be generated. This is a clear advantage over direct observations for which only a single realization is usually available. In particular, reproducibility allows to investigate important issues related to rainfall scaling and nonlinear error propagations in hydrological and climatic models.

Many different models and techniques for the simulation of synthetic rainfall with given statistical and physical properties are reported in the literature (see Section 1.4.2 for a brief review). Most of them focus on the simulation of rain rate values. Recently, there has been an increased interest in a new class of stochastic rainfall simulators that are capable of generating rainfall fields with a higher level of detail, for example by including the drop size distribution. Such DSD simulators are more difficult to implement but offer unprecedented advantages over traditional rainfall simulators, especially for applications involving polarimetric weather radars for which most quantities of interest (e.g., the radar reflectivity, the path-integrated attenuation and the phase shift) explicitly depend on the DSD or its weighted moments. So far, however, only a few, very simple DSD simulators are available. Most of them are limited to spatial or temporal range profiles and do not include rainfall intermittency. Their utility in remote sensing applications is therefore very limited and a lot of improvements can be expected if more realistic DSD fields can be generated.

In the remaining of this chapter, a more detailed description of the most important concepts and topics that are needed to understand this thesis are provided. Section 1.2 provides more details about the microstructure of rainfall and its importance in remote sensing applications. More information about remote sensing instruments like weather radars and microwave links are provided in Section 1.3. For a brief overview of stochastic rainfall simulation techniques, see Section 1.4. Finally, the outline of the thesis is given in Section 1.5.

1.2 The microstructure of rainfall

1.2.1 The drop size distribution (DSD)

The drop size distribution is a very general and detailed statistical description of the microstructure of rainfall. Its purpose is to efficiently summarize (using probability theory and statistics) the huge amount of information about the drops contained in a given volume of air (typically, 1 m^3). A schematic illustration of this very important concept is showed in Figure 1.1. In its traditional definition, the DSD is used to describe the expected number of drops with equivolume spherical drop diameters between D and $D + dD$ per unit volume (m^3) of air. Note that because raindrops can have different shapes depending on their sizes, the raindrop diameter D is always described by the equivolume spherical diameter, i.e., the diameter of a sphere with equal volume.

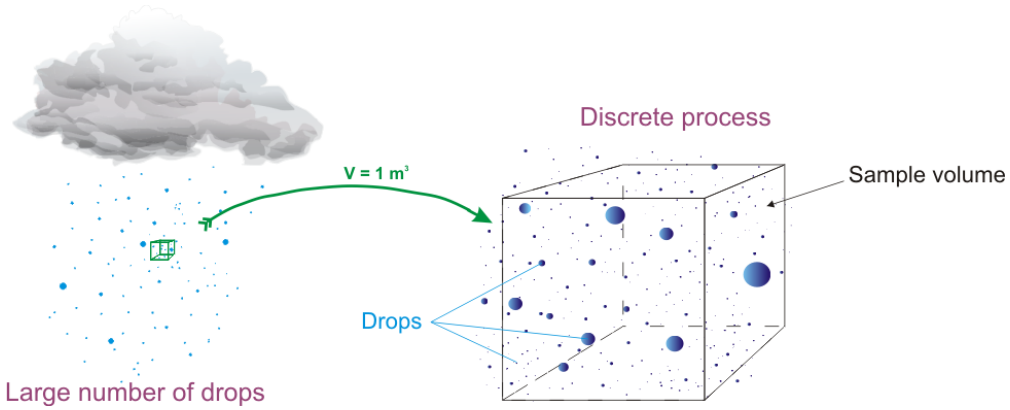

Figure 1.1: Visual illustration of the concept of drop size distribution (DSD)

On average, 1 m^3 of air usually contains 10^2-10^3 raindrops with various diameters between 0.1 and 6 mm. The average distance between two raindrops is in the order of 10 cm. Although this is rather large compared to the average drop size, it is not sufficient to prevent the raindrops from occasionally colliding with each other. When this happens, the two raindrops can either coalesce and form a new and bigger drop (collisional growth), bounce off and continue their way separately or break up into many different and smaller drops (collisional breakup). If for any reason (e.g., diffusional or collisional growth) a drop becomes too big, it gets unstable and rapidly breaks up into smaller and more stable droplets. Very small drops on the other hand can evaporate or can be caught by small updrafts or turbulences. The raindrop size distribution is a very valuable piece of information because it contains the signature of the major microphysical processes (e.g., nucleation, diffusional growth, collisional growth, evaporation and breakup) at work in the formation and evolution of rain in the atmosphere. Its properties and dynamics must therefore be studied carefully.

Since the early work by Marshall and Palmer in 1948, many efforts have been devoted to improving our understanding of the physical and mathematical properties of DSD's. Because of instrumental limitations, DSD was originally modeled using an exponential size distribution (Marshall and Palmer, 1948):

$$N(D) = N_0 \exp(-\Lambda D) \tag{1.1}$$

where N_0 [$mm^{-1}\,m^{-3}$] is a drop concentration parameter and Λ [mm^{-1}] a scale parameter. Applications of this model can be found (among many others) in Sekhon and Srivastava (1971); Waldvogel (1974); Austin (1987) and Uijlenhoet et al. (1999a). The values of concentration N_0 and scale Λ characterizing the raindrop size distribution have been shown to fluctuate in space and time depending on the rain rate R and the type of rainfall (stratiform vs convective). Because of this natural variability, it is common to consider N_0 and Λ as two (correlated) stochastic processes in space and time. Of course, other models and parameterizations of the DSD have been proposed in the literature.

Popular alternative DSD models are given by the log-normal distribution (Bradley and Stow, 1974), the Weibull distribution (Assouline and Mualem, 1989) and the Gamma distribution (Ulbrich, 1983; Willis, 1984).

Nowadays, the most common representation of the DSD in the literature is the Gamma model, possibly normalized or transformed in order to minimize the scale-dependency of the parameters (Sempere-Torres et al., 1994; Testud et al., 2001; Lee et al., 2004). This choice is supported by a large amount of theoretical and experimental results for different climatological regimes, types of rainfall, instruments and space-time resolutions. In its simplest formulation, the Gamma DSD model is given by:

$$N(D) = \alpha N_t D^{\mu} \exp(-\Lambda D) \tag{1.2}$$

where $N(D)dD$ [m^{-3}] denotes the number of drops per unit volume with diameters between D [mm] and $D + dD$, $\mu > -1$ [-] is a shape parameter, $\Lambda > 0$ [mm^{-1}] a rate parameter, $N_t > 0$ [m^{-3}] a concentration parameter and $\alpha = (\int_{D_{min}}^{D_{max}} D^{\mu} \mathrm{e}^{-\Lambda D}\, dD)^{-1}$ [-] a normalization factor taking into account the finite range of possible drop sizes between D_{min} and D_{max} . Note that the Exponential DSD model given in Equation (1.1) is a particular case of the more general Gamma model with shape parameter $\mu = 0$.

Similarly to the parameters of the Exponential model, μ, Λ and N_t also fluctuate in space and time and can therefore be interpreted as (correlated) stochastic processes describing the natural variability of the DSD.

1.2.2 Important quantities related to the DSD

Knowledge of the DSD in a given volume of air is very useful. It allows to derive characteristic drop sizes like the mean and median drop diameter (which are related to the type of precipitation) and many other important physical quantities like the average drop speed, the average kinetic energy and the rain rate. Most importantly, the DSD can also be used to derive most of the quantities related to remote sensing of rainfall (e.g., reflectivity, attenuation and phase shift).

The average drop diameter D_m [mm] is given by:

$$D_m = \frac{\int_{D_{min}}^{D_{max}} DN(D)dD}{\int_{D_{min}}^{D_{max}} N(D)dD} \tag{1.3}$$

Note that because DSDs are skewed distributions, D_m is often slightly larger than the mode of the size distribution (i.e., the diameter with the highest density). Also, for remote sensing applications, other quantities like the median volume diameter D_0 that are more representative of higher order moments are usually preferred. The median volume diameter, i.e., the diameter D_0 [mm] which divides the distribution of the rain

volume over all raindrop sizes in two equal parts is given by:

$$\int_{D_{min}}^{D_0} D^3 N(D) dD = \int_{D_0}^{D_{max}} D^3 N(D) dD = \frac{1}{2} \int_{D_{min}}^{D_{max}} D^3 N(D) dD. \tag{1.4}$$

For applications related to soil erosion, the average kinetic energy rate E [$\mathrm{J\,m^{-2}\,h^{-1}}$] on the ground is given by:

$$E = \frac{3\pi}{10^7} \rho_w \int_{D_{min}}^{D_{max}} D^3 v(D)^3 N(D) dD \tag{1.5}$$

where $v(D)$ represents the average terminal fall speed (in still air) of a drop with equivolume spherical diameter D [mm] and ρ_w [$\mathrm{kg\,m^{-3}}$] is the density of liquid water. In general, $v(D)$ is not known, but a very large number of experimental and theoretical investigations using wind tunnels, falling towers and high-speed cameras showed that it can be approximated using a power law of the drop diameter (e.g., Sekhon and Srivastava, 1971; Beard, 1976; Atlas and Ulbrich, 1977). Typical raindrop velocities are between 1 $\mathrm{ms^{-1}}$ for $D \sim 0.25$ mm and 8 $\mathrm{ms^{-1}}$ for $D \sim 3$ mm. The density of liquid water ρ_w slightly depends on the temperature and is approximatively 10^3 kg per $\mathrm{m^3}$ (Kell, 1975).

Similarly to the kinetic energy, the rain rate R [$\mathrm{mm\,h^{-1}}$] also depends on the average terminal fall speed of the individual drops:

$$R = \frac{6\pi}{10^4} \int_D D^3 v(D) N(D) dD \tag{1.6}$$

For applications involving weather radars, the radar reflectivity $Z_{h|v}$ [$\mathrm{mm^6\,m^{-3}}$] at horizontal and vertical polarization is given by:

$$Z_{h|v} = \frac{10^6 w^4}{\pi^5 |\frac{m^2-1}{m^2+2}|} \int_D \sigma_{B_{h|v}}(D) N(D) dD \tag{1.7}$$

where w [cm] represents the radar wavelength, m [-] the complex refractive index of water (at a given temperature) and $\sigma_{B_{h|v}}(D)$ [$\mathrm{cm^2}$] the backscattering cross section (e.g., the area which, when multiplied by the incident intensity gives the total backscattered power) of a drop of diameter D at horizontal and vertical polarization. Note that because raindrops are not spherical, $\sigma_{B_h}(D)$ and $\sigma_{B_v}(D)$ are different in general. Small raindrops ($D < 1$ mm) falling at their terminal velocities have nearly spherical shape (Pruppacher and Pitter, 1971; Beard and Chuang, 1987). Larger drops, however, are increasingly flattened at their base and can be approximated by oblate spheroids (Andsager et al., 1999; Beard et al., 2010). It is therefore common to model raindrops by oblate spheroids with an equivolume drop diameter D and an average axis ratio r_{ab} (e.g., the ratio between the vertical and the horizontal dimension of the drop). The T-matrix method by Mishchenko et al. (2002) can then be used to compute the backscattering cross sections $\sigma_{B_{h|v}}(D)$ based on this drop shape model. Note that once Z_h and Z_v have been computed, it is

also possible to derive the differential reflectivity Z_{dr} [dB]:

$$Z_{dr} = 10 \log_{10} \left(\frac{Z_h}{Z_v} \right) \tag{1.8}$$

Another, very important quantity, especially for X-band radars and microwave links is the specific attenuation on propagation $A_{h|v}$ [dB km^{-1}] :

$$A_{h|v} = \frac{1}{\ln(10)} \int_D \sigma_{E_{h|v}}(D) N(D) dD \tag{1.9}$$

where $\sigma_{E_{h|v}}(D)$ [cm^2] denotes the extinction cross section (e.g., the area which, when multiplied by the incident intensity gives the total absorbed or scattered power) of a drop of diameter D. Note that both $\sigma_{E_h}(D)$ and $\sigma_{E_v}(D)$ can be computed (similarly to the backscattering cross sections) using the T-matrix method for a given drop shape model.

The specific differential phase shift on propagation K_{dp} [$^\circ$ km^{-1}] , a very important quantity for rain rate estimation and attenuation correction algorithms is given by:

$$K_{dp} = \frac{1800w}{\pi} \int_D \mathrm{Re} \left[S_{hh}(D) - S_{vv}(D) \right] N(D) dD \tag{1.10}$$

where $\mathrm{Re}(S_{hh|vv})$ [m] denote the real part of the forward scattering amplitudes at horizontal and vertical polarizations (also computed using the T-matrix code).

1.2.3 Intermittency

A major problem in the mathematical description of precipitation is spatial and temporal intermittency, i.e., the constant alternating between dry and rainy periods. Like the DSD, intermittency is also highly variable in space and time. This variability has important consequences that must be taken into account. For example, rainfall intermittency and inhomogeneous water availability significantly influences vegetation cover (Baudena et al., 2007; Kletter et al., 2009) and controls important natural processes like stream flow, runoff, soil moisture and soil erosion (Pitman et al., 1990).

The major difficulty with rainfall intermittency is the fact that it significantly varies with respect to the spatial and temporal scales at which it is observed (Kundu and Siddani, 2011). Typically, it is higher at finer scales and lower at coarser scales. This scale dependency has important consequences in many practical applications in hydrology, meteorology and remote sensing of precipitation. It is, for example, crucial in rainfall interpolation and disaggregation techniques and must be taken into account when upscaling/downscaling the outputs of numerical weather models or weather radar data (Seo, 1998; Lanza et al., 2001). Other examples can be found in the field of stochastic rainfall simulation (Bárdossy and Plate, 1992; Kang and Ramirez, 2010), partial beam

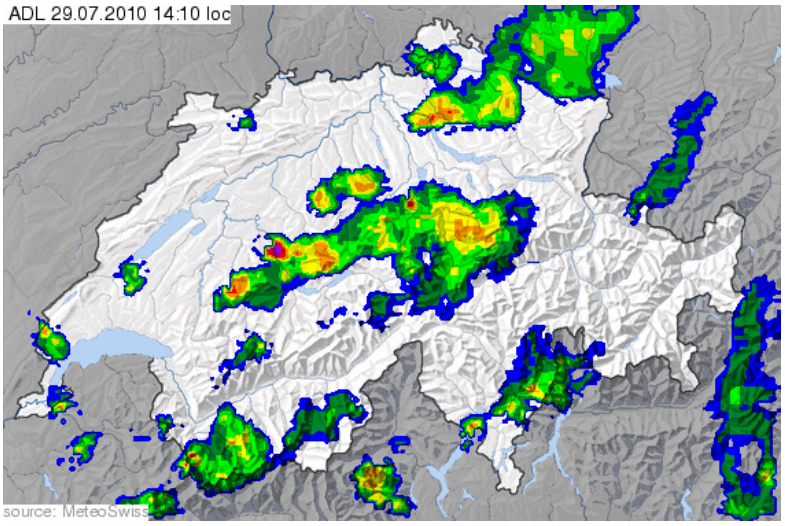

Figure 1.2: Visual illustration of rainfall intermittency.

filling problems (e.g., Durden et al., 1998; Gosset and Zawadzki, 2001) and attenuation correction for weather radars (Bringi et al., 2001; Matrosov et al., 2002; Gorgucci and Chandrasekar, 2005).

So far, several studies have investigated rainfall intermittency at different spatial and temporal scales. Among the studies focusing more on spatial intermittency, Braud et al. (1993) and Jeannin et al. (2008) analyzed the relations that exist between the mean areal rain-rate and the fractional area where it rains above a fixed threshold. Using mathematical morphology, Kumar and Foufoula-Georgiou (1994) proposed different algorithms for downscaling/upscaling intermittent rainfall fields. Using disdrometer measurements and high temporal resolutions, Lavergnat and Golé (1998) analyzed the inter-arrival times of raindrops and proposed to model the time intervals between raindrops using a Bi-Pareto law. Using radar data, Pavlopoulos and Gritsis (1999) and Pavlopoulos and Gupta (2003) analyzed the durations and scaling of wet and dry periods. Among the studies focusing more on the stochastic simulation of intermittent rainfall fields, one can mention the work of Lanza (2001) who proposed a space-time stochastic rainfall simulation method to generate intermittent rainfall fields conditional on rain gauge observations. Based on radar data, Wojcik et al. (2009) proposed a combination between Geostatistical methods for the simulation of intermittency and truncated multiplicative random cascade for the rain rate. More recently, Kleiber et al. (2012) proposed a simulation method for weather generators that relies on latent Gaussian processes for intermittency and transformed Gaussian processes for the rain intensity. Starting from tipping bucket rain gauges, Molini et al. (2009) and Rigby and Porporato (2010) investigated different relations that exist between rainfall intermittency and turbulence across a large range of time scales and climatic regimes. Using a different approach based on the maximum entropy principle, Koutsoyiannis (2006) noted that,

under some circumstances, the probability that a time interval is dry, follows a scaled exponential function of timescale. This is consistent with the results obtained by Kundu and Siddani (2011) who showed that dry probabilities (both spatial and temporal) can be modeled using simple scaled exponential functions. Their study, however, did not investigate the spatial and temporal autocorrelation of intermittency.

1.3 Remote sensing of precipitation

1.3.1 A brief introduction to weather radars

For over 50 years now, weather radars have been the unchallenged tools to accurately measure precipitation at high spatial and temporal resolutions over large domains. The term "radar" stands for "RAdio Detecting And Ranging" and was first introduced by S. Taylor and F. Furth of the U.S. Navy in 1940. At this early stage, radars were mostly used for military purposes, to detect aircrafts and ships during World War II. The very origin and first use of radar as a new tool to detect clouds and precipitation is difficult to trace back because of wartime secrecy but a significant amount of information can be found in Atlas and Ulbrich (1990). As this thesis is not about weather radars, only the most important notions will be mentioned here. For a more detailed and technical description of weather radars and their applications in hydrology and meteorology, the reader is referred to Bringi and Chandrasekar (2001).

A weather radar is an instrument used to locate precipitation, estimate its type (e.g., rain, snow, hail), calculate its motion and forecast its future position and intensity. Its basic principle is rather simple: an electromagnetic wave is transmitted in the direction of a meteorological target and the backscattered signal from the volume of interest is recorded. Most of the time, the targets are raindrops but the same technique can be used for other hydrometeors (e.g., snow flakes, hail stones or ice crystals) or non-meteorological targets (e.g., insects, birds and planes). The relationship between the transmitted and the received signals is described by the weather radar equation for a volume target (Battan, 1973):

$$\bar{P}_r = C\frac{|K|^2}{r^2}Z, \tag{1.11}$$

where $\bar{P}_r$ [W] denotes the average power received from the target at range r [km], $|K|^2$ [-] is a coefficient related to the dielectric constant of water (about 0.93), C is called the radar constant and Z denotes the radar reflectivity factor [$\mathrm{mm^6\,m^{-3}}$]. Note that while the radar constant depends on the radar properties, the radar reflectivity is a purely meteorological quantity that is independent of the radar properties and only depends on the characteristics and properties of the targets. In case of liquid precipitation, it can be expressed as a function of the raindrop size distribution according to Equation 1.7.

Unfortunately, weather radars do not measure directly the rain rate R (which is the quantity of interest in most applications) but only the reflectivity Z. Methods that allow to estimate R given Z are called $Z - R$ relationships. The most common $Z - R$ relationship is a simple power law (Marshall et al., 1955):

$$Z = \alpha R^{\beta} \tag{1.12}$$

where α and β are two parameters that depend on the time, the location and the type of rainfall. In fact, the optimal values of α and β strongly depend on the drop size distribution (which is usually unknown in practical applications). A common solution to this problem is to rely on climatological relationships and to estimate α and β using large data sets of different rain events that are supposed to be representative of the local climatology. For example, Marshall et al. (1955) suggested

$$Z = 200R^{1.6} \tag{1.13}$$

The problem with climatological $Z - R$ relationships is that they only represent the average relationship between Z and R. As a result, rain rates derived from radar measurements are always affected by large uncertainties and can be strongly biased due to the non-linearity of Equation 1.12. Moreover, traditional radar rain-rate estimation techniques always rely on the assumption that the radar sampling volume (which can be larger than 1 km^3) is uniformly filled with rainfall. In case of strong intermittency however, this assumption of uniform beam-filling is often violated and results in additional errors in the rain rate estimation.

Nowadays, weather radars are increasingly polarimetric, i.e., they transmit two different signals: one at horizontal and one at vertical polarization. The information contained in the amplitude, phase and polarization states of the scattered electromagnetic waves can then be used to improve the rain rate and snowfall estimates (Bringi and Chandrasekar, 2001). After more than 30 years of successful research and development, basic and applied research in polarimetric radar meteorology continues to be strong world-wide but suffers from the lack of knowledge about the raindrop size distribution and its variability at small spatial and temporal scales. A stochastic simulator capable of generating synthetic DSD fields (like the one presented in this thesis) is therefore a very valuable tool because it allows to investigate and quantify the errors and uncertainties associated to rain rate estimation using polarimetric weather radars.

1.3.2 Microwave links (MWL)

Microwave links (MWL hereinafter) are widely used for wireless data exchange between base stations of mobile phone networks. Because of the high frequencies used for data transmission, the link signal can be significantly attenuated when rainfall occurs along the path of the link (see Figure 1.3). This attenuation can be related to the path-integrated

rain rate and used to provide valuable information at an intermediate scale between measurements from rain gauges and radar measurements (Rahimi et al., 2004; Grum et al., 2005; Upton et al., 2005).

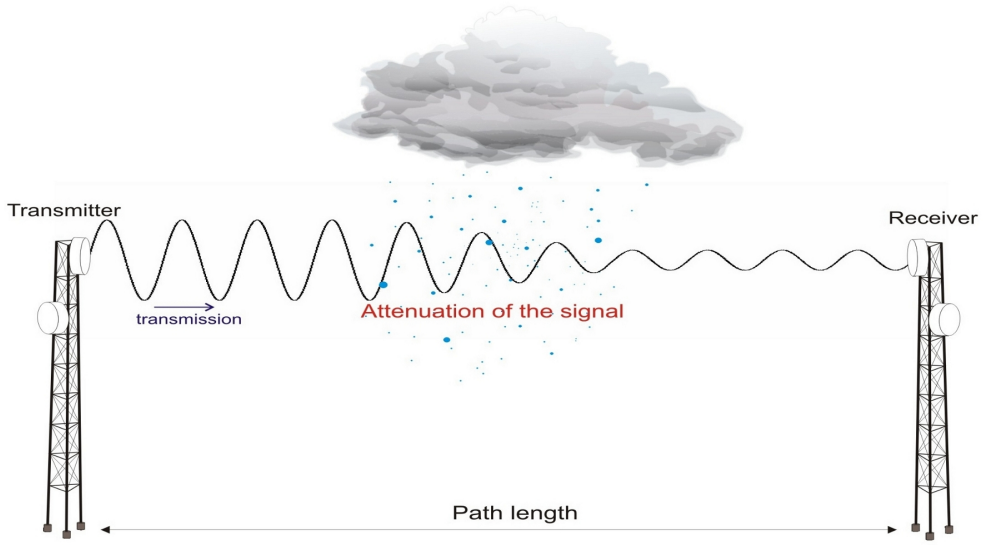

Figure 1.3: Visual illustration of a terrestrial microwave link.

Recently, Messer et al. (2006) and Leijnse et al. (2007c) proposed to use operational microwave links employed in telecommunication networks for quantitative rainfall estimation. They showed that the potential applications of MWLs in hydrology and meteorology are considerable and will benefit from the rapid growth of cellular communication in the years ahead. Some new studies have already investigated the potential of such highly dense networks for spatial rainfall reconstruction techniques (Zinevich et al., 2008), stochastic interpolation of rain rate estimates (Goldshtein et al., 2009) and state-space modeling (Zinevich et al., 2009). However, the ability to obtain accurate and reliable precipitation estimates using MWL strongly depends on the spatial variability of rainfall (Berne and Uijlenhoet, 2007; Leijnse et al., 2010) and the uncertainties affecting the link measurements (Zinevich et al., 2010). A particularly critical issue concerns the ability to distinguish between the rain-induced attenuation and other sources of attenuation like water vapor, wind effects on the antennas, interferences and wet-antenna (see Chapters 6-7).

In the last decade, significant efforts have been devoted to exploiting the full potential of microwave links for rainfall estimation. Important progresses can still be expected for applications including dual-polarization or dual-frequency microwave links. Indeed, by exploiting the fact that large raindrops are not spherical and by comparing the attenuations at different polarizations or frequencies, one is able to provide more information about the underlying microstructure of rainfall. For example, Rincon and Lang (2002) was able to show that it is possible to retrieve the path-integrated drop size distribution from the

measurements of a dedicated dual-frequency link. In 2009, (Berne and Schleiss, 2009) proposed a similar approach for commercial dual-polarization links but their technique must still be validated using ground data.

1.4 Stochastic simulation of rainfall

1.4.1 The rainfall measurement problem

Modern hydrological models often rely on accurate rainfall estimates at high spatial and temporal resolutions (in the order of 1 km and 5 min or higher). Measuring rainfall at these scales is known to be very difficult. Rain gauges and disdrometers provide good point estimates at high temporal resolutions but their measurements are not very representative at larger scales because of the large spatial variability of rainfall and DSD (Ciach and Krajewski, 2006). High density rain gauge and disdrometer networks partially overcome this problem but cannot be set up over large domains at reasonable costs. Moreover, such networks cannot capture the vertical variability of rainfall, which is very important in meteorological models and satellite applications. Telecommunication microwave links constitute an interesting alternative to traditional rain gauge networks by providing path-integrated rainfall measurements at high temporal resolutions (a few meters above the ground). However, because these networks were not specifically built for rainfall estimation, the estimated rain rates are usually affected by large errors and uncertainties. Furthermore, their spatial resolution and representativity are limited by the length of the links and density of the network.

So far, only weather radars are capable of providing quantitative rainfall estimates that meet all the requirements of hydrological and meteorological models: extended coverage with a single instrument (including the vertical dimension), rapid access to data for real-time applications and relatively high spatial and temporal resolutions (e.g., sampling volume of about 1 km^3). Quantitative estimation of rainfall from radar observations is, however, a very difficult problem (Krajewski and Smith, 2002). As a result, radar rain-rate estimates can be affected by significant errors and uncertainties. These errors and sources of uncertainties must be carefully quantified and taken into account in the models. For example, 100 mm of rain falling on 10 % of a model grid square will produce a very different hydrological response from 10 mm uniformly distributed over the entire square. Such issues of scaling and non-linear error propagation are often encountered in hydrology and meteorology and must be taken into account when working with radar rainfall estimates.

One way to assess the uncertainties associated to radar rainfall estimation is by comparing radar rain-rate estimates to surface data collected using nearby rain gauges or disdrometers (Collier, 1986). Such comparisons are interesting but suffer from several fundamental limitations like scale incompatibility and vertical variability (e.g., Zawadzki,

1975; Krajewski, 1987; Creutin et al., 1988). Another interesting way to study the error structure and contributions related to radar rainfall estimation is to use stochastic rainfall simulators. These simulators are not predictive tools like those used in weather forecasting, but are designed to generate a large number of synthetic rain events that share the same statistical properties (e.g., rain rate distributions, spatial and temporal structures, intermittency). These generated fields can then be used to investigate important issues related to input uncertainties and biases in rainfall estimation in hydrological models, an approach known in the satellite community as an Observing System Simulation Experiment (OSSE) (Arnold and Dey, 1986).

1.4.2 A brief overview of rainfall simulators

Models and techniques for the simulation of synthetic rainfall events with certain statistical and physical properties have a long history. Initially, rainfall simulators were mostly developed with the objective of simulating single variables, most often daily rain amounts at a given location (e.g., Gabriel and Neumann, 1962). Soon, a new family of models that included more climatic variables (e.g., temperature and humidity) emerged and became known as weather generators. Most of the first weather generators were based on chain-dependent processes (Katz, 1977; Richardson, 1981) that reproduced the statistical properties of observed weather time series. The main reasons behind the development of these generators were that observed time series of meteorological parameters were often not long enough to accurately evaluate the long-term effects of proposed, man-made hydrological changes. Moreover, the use of observed data only provided estimates for one single realization of the weather process and could not be used to evaluate the result for different inputs and scenarios. The second purpose of weather generators was to provide a powerful tool to extend observed weather time series to unobserved locations. Typically, in order to generate precipitation data at unobserved locations, the statistical parameters of a weather generator would first be calibrated using the data from the nearest meteorological stations and subsequently interpolated to the unobserved locations (Hutchinson, 1995). For more details about the development of such statistical weather models for single and multi-site rainfall simulation and their use in water engineering design, agricultural, ecosystem and hydrological impact studies, the reader is referred to Wilks and Wilby (1999).

Weather generators are undoubtedly a powerful tool to generate precipitation data at single locations at the daily or hourly scale. Recent interest in the field of stochastic simulation has, however, moved away from individual locations to more complicated, space-time models of rainfall at higher spatial and temporal resolutions. Clearly, this requires to make certain assumptions regarding the physical and statistical properties of rain events. For example, Bras and Rodríguez-Iturbe (1976) emphasized that the spatial variations should be linked to the temporal variations in the direction of the storm movement according to Taylor's hypothesis of frozen turbulence. Moreover, the

event should be given a non-stationary temporal structure by allowing the values of the model parameters to vary with time. This necessity for physically realistic simulation frameworks is also pointed out by Waymire et al. (1984) who proposed a conceptual rainfall simulation model based on point processes with many physical features that can be observed in extratropical cyclonic storms (e.g., clustering of the rain cells, cell birth, cell decay and cell motion relative to the ground). These ideas where further developed by Rodríguez-Iturbe and Eagleson (1987) and Sivapalan and Wood (1987). Using a transformed AR(1) process whose parameters are linked to atmospheric circulation patterns, Bárdossy and Plate (1992) proposed a space-time stochastic simulation model for daily precipitation that explicitly acknowledges the importance of intermittency and non-stationarity of the rainfall process. However, as pointed out by Sivapalan and Wood (1987), parameter estimation constitutes the largest problems in all these simulation models. In their opinion, good rainfall simulators should be as simple as possible and facilitate parameter estimation using minimal data such as observations from rain gauge networks, disdrometers or weather radars. In order to retain their validity in different conditions, the simulators should also incorporate, whenever possible, simple physically-based structures and principles.

As a matter of fact, many different stochastic rainfall simulators have been proposed during the last 20-30 years. The objective of this section is not to dress a list of all the proposed techniques. Instead, a summary and overview of the major ideas and statistical techniques behind these simulators are provided. Each technique is then illustrated through one or two chosen references to the literature.

Since Lovejoy and Mandelbrot (1985) established the applicability of fractals in meteorology and their capability to generate complex structures like “bands”, “fronts” and “clusters”, a large body of literature on rainfall simulation algorithms based on (multi)fractals and self-similarity concepts has emerged (e.g., Gupta and Waymire, 1993; Menabde et al., 1997). In order to study sahelian storms, Guillot (1999) proposed to represent the spatial structure of the rainfall field using Geostatistics. Based on parameterizations from rain gauges, Shah et al. (1996) showed that it is possible to generate synthetic precipitation fields with realistic spatial structure using the turning bands method. Later, a new procedure for conditioning the simulated fields with rain gauge observations while preserving the intermittency was proposed by Lanza (2001). Using principal component decomposition based on rain gauge observations, Bouvier et al. (2003) proposed a method to simulate daily rainfall fields by separating the temporal distributions and the spatial correlations. More recently, (Wojcik et al., 2009) proposed a combination of Geostatistical methods for the simulation of intermittency and a truncated multiplicative random cascade model to simulate rain rates in each cluster.

A common characteristic of all the simulation methods mentioned above is that they all focus on the simulation of rain rate values. Only a few studies have investigated the possibility to simulate rainfall fields with higher level of detail, for example by including

the drop size distribution. Such DSD simulators offer unprecedented advantages over traditional rainfall simulators and can be used to investigate various issues related to polarimetric weather radars (Krajewski et al., 1993; Anagnostou and Krajewski, 1997). So far, however, only a few, very simple DSD simulators are available. Using simulated range profiles of DSDs, Berne and Uijlenhoet (2005), Uijlenhoet and Berne (2008) and Montopoli et al. (2008) investigated different uncertainties associated with rainfall estimation using weather radars. Using point processes, Lavergnat and Golé (2006) proposed a rainfall simulator based on a drop release mechanism for series of drop diameters and arrival times. Following a different approach, Lee et al. (2007) proposed to use measured radar reflectivity fields in combination with ground-based DSD measurements to generate synthetic DSD fields. In this case, however, the highest spatial resolution is imposed by the resolution of the radar, which is usually of the order of 1 km^2. Hence this approach does not allow to investigate the radar subgrid variability of the DSD.

1.5 Thesis outline

The main objective of this thesis is to develop a stochastic simulator capable of generating realistic fields of raindrop size distributions in space and time. Special emphasis is put on the reproduction of the spatial and temporal structures as well as the intermittency of observed DSD fields. An experimental approach based on a network of 16 optical disdrometers covering an area of about 1 km^2 in the vicinity of Lausanne, Switzerland (Jaffrain et al., 2011; Jaffrain and Berne, 2012) is used to derive meaningful parameterizations of the simulator that are representative of the local climatology. The proposed DSD simulator is then used to develop a new and innovative disaggregation method of rainfall fields that includes the drop size distribution as an output. The final part of the thesis is devoted to the study of rainfall intermittency at small spatial and temporal scales. It presents new results about the downscaling and upscaling of rainfall intermittency that are necessary to parameterize the DSD simulator and investigates the possibility to identify dry and rainy periods using telecommunication microwave links.

This thesis is structured as follows: In Chapter 2, a first and (relatively) simple DSD simulator based on a Gamma DSD model is presented. It is based on Geostatistics and uses variograms to model the spatial structure of the DSD parameters μ, Λ and N_t. The simulator is limited to the static case (i.e., the simulated fields are frozen in time) and does not address the issue of rainfall intermittency. It is, however very important because it provides the general ideas and methodologies used to build the more elaborate simulation techniques presented in Chapter 3 and 4.

Chapter 3 presents a second, more general and complex DSD simulator. It includes the possibility to simulate both in space and in time and explicitly considers rainfall intermittency. The generated DSD fields are more realistic and their evolution in time or space can be partially controlled through the use of external drifts. Spatial and temporal

variations are linked and described using a separable space-time variogram. Chapter 4 further generalizes the simulator by including the possibility to generate intermittent DSD fields with realistic space-time structures that are conditioned by average rain-rate values at a coarser resolution. Further extensions of this new technique to dual-frequency and dual-polarization weather radars are also discussed.

Chapter 5 provides a short statistical analysis of rainfall intermittency at small spatial and temporal scales. Its objective is to provide simple methods and guidelines for accurate downscaling and upscaling of intermittency in space and time. Such scaling laws are relevant to parameterize rainfall intermittency in the DSD simulator and to estimate the percentage of dry regions and periods during the disaggregation process. It is also useful for further investigations on rainfall intermittency using microwave links in Chapter 6 and 7. The objective of these last chapters is to investigate different methods to identify and quantify rainfall intermittency using commercial microwave links. Significant efforts are devoted to the identification of dry and rainy periods based solely on the path-integrated signal attenuation. Chapter 7 further describes a new and innovative experimental setup deployed in Dübendorf and used to investigate different aspects of rainfall retrieval using commercial microwave link networks. Finally, a summary of the most important results in this thesis as well as some perspectives for future research are given in Chapter 8

Note that this thesis is a compilation of published or submitted articles and might therefore contain small repetitions (in particular in the introduction and in the description of the models and methods in each chapter). The author apologizes for any inconvenience caused by this particular format.

2 Geostatistical simulation of 2D fields of raindrop size distributions at the meso gamma scale [1]

2.1 Summary

The large variability of the raindrop size distribution (DSD) in space and time must be taken into account to improve remote sensing of precipitation. The ability to simulate a large number of 2D fields of DSDs sharing the same statistical properties provides a very useful simulation framework that nicely complements experimental approaches based on DSD ground measurements. These simulations can be used to investigate radar beam propagation through rain and to evaluate different radar retrieval techniques.

The proposed approach uses Geostatistical methods to provide structural analysis and stochastic simulation of DSD fields. First, the DSD is assumed to follow a Gamma distribution with three parameters. As a consequence, 2D fields of DSDs can be described as a multivariate random function. The parameters are normalized using a Gaussian anamorphosis and simulated by taking advantage of fast Gaussian simulation algorithms. Variograms are used to characterize the spatial structure of the DSD fields. The generated fields have identical spatial structure and are consistent with the observations. Because intermittency cannot be simulated using this technique, the size of the simulation domain is limited to the meso-γ scale (2-20km).

To assess the proposed approach, the method is applied to data collected during intense Mediterranean rainfall. Taylor's hypothesis is invoked to convert time series into 1D range profiles. The anisotropy of the fields is derived from radar measurements. Simulated and measured reflectivity fields are in good agreement with respect to the mean, the standard deviation and the spatial structure, demonstrating the promising potential of the proposed stochastic model of DSD fields.

1. This chapter is a slightly modified version of the article by Schleiss, M., A. Berne and R. Uijlenhoet, 2009: Geostatistical simulation of 2D fields of raindrop size distributions at the meso gamma scale, *Water Ressour. Res.*, **45**, W07415, doi:10.1029/2008WR007545.

2.2 Introduction

Because of the complex interactions between atmospheric dynamics and cloud microphysics, precipitation is highly variable over a large range of space and time scales (e.g., Berndtsson and Niemczynowicz, 1988; Groisman and Easterling, 1994; Xie and Arkin, 1997). This variability is a significant source of uncertainty for the measurement, the simulation and the forecasting of precipitation as well as of the environmental processes influenced by it. For example, the variability of land surface hydrology is strongly controlled by the variability of precipitation (Syed et al., 2004).

To investigate aspects of the large space-time variability of precipitation, a simulation approach is able to provide known reference data from which a variety of sources of uncertainty can be studied quantitatively. Hence a lot of attention has been devoted to the development of techniques to simulate 2D or 3D precipitation fields (Foufoula-Georgiou and Krajewski, 1995; Pegram and Clothier, 2001). These techniques can be divided in two main categories: (1) physical approaches that aim at simulating the physical processes involved in precipitation (e.g., ARPS, Xue et al., 2000); (2) statistical approaches that consider precipitation as a random variable in space and time. Within the latter category, different techniques have been applied: point processes and clustering (e.g., Waymire et al., 1984; Onof et al., 2000); self-similarity (e.g., Gupta and Waymire, 1993; Menabde et al., 1997); and Geostatistics (e.g., Guillot, 1999; Bouvier et al., 2003). The proposed methods focus solely on the simulation of rain rate values.

Remote sensors do not directly measure the rain rate but rather some observables which are related to the electromagnetic properties of the ensemble of drops within the considered sampling volume. In the case of weather radars, the conversion of radar reflectivity values into rain rate values is strongly influenced by the microstructure of rainfall (mainly by the size, the shape and the fall velocity of individual rain drops). The fall velocity and the shape of rain drops are closely related to their equivolume diameter (Beard, 1976; Andsager et al., 1999). Therefore the raindrop size distribution (DSD hereafter) is of critical importance for the quantitative interpretation of radar measurements. Similarly to rain rates, the DSD is highly variable in space and time (Tokay and Short, 1996; Jameson and Kostinski, 2001; Uijlenhoet et al., 2003). This variability must be taken into account to improve radar rain-rate estimates.

To analyze the different sources of uncertainty in radar rain-rate estimates using a simulation approach, it is necessary to include information about the spatial variability of the DSD. Simulation methods based on point processes have been developed (e.g., Lavergnat and Golé, 2006), but so far cannot provide 2D or 3D fields of DSDs. Conditional simulation methods have also been proposed. Starting from simulated rain rate fields, Krajewski et al. (1993) derived a consistent Gamma DSD assuming uniform and uncorrelated distributions of the DSD parameters. These are strong assumptions, not supported by our own data. More recently, Lee et al. (2007) proposed to use measured

radar reflectivity fields in combination with ground-based DSD measurements to generate DSD fields. In this case, the highest spatial resolution is imposed by the radar resolution, which is usually of the order of 1 km^2. Hence this approach does not allow to investigate the radar subgrid variability of the DSD. Simulations of correlated DSD fields based on point DSD measurements do not have this limitation. However, to date such techniques have only been proposed to simulate time series or 1D range profiles (Berne and Uijlenhoet, 2005; Montopoli et al., 2008).

The main objective of this chapter is to develop a stochastic simulation framework allowing for conditional as well as non-conditional simulation of 2D fields of DSDs using point DSD measurements. Geostatistics provides useful tools for the analysis and the simulation of random fields with complex spatial structures (Chilès and Delfiner, 1999; Lantuéjoul, 2002). The ability to generate a large number of statistically homogeneous fields can be used to obtain reliable statistical characterizations of a variety of issues related to radar beam propagation through rain as well as radar retrieval techniques. It is worth mentioning that such a simulation framework may also be useful for other domains dealing with the propagation of radio waves in the atmosphere. Satellite or ground-based microwave communication is a relevant illustration (Dissanayake et al., 1997; Fong et al., 2003).

The chapter is organized as follows: Section 2.3 describes the modeling of the DSD. The simulation framework is detailed in Section 2.4. The simulator is applied and evaluated using data collected during an intense Mediterranean rain event in Section 2.5. The conclusions and perspectives are given in Section 2.6.

2.3 Modeling the DSD

2.3.1 The Gamma model

The DSD describes the number of drops per unit volume and per unit size interval of equivolume spherical drop diameter. It is supposed to be adequately described by a Gamma distribution (Ulbrich, 1983; Willis, 1984) given by the following expression:

$$N(D) = \alpha N_t D^{\mu} \exp(-\Lambda D), \tag{2.1}$$

where $N(D)dD$ [m^{-3}] denotes the number of drops per unit volume with diameters between D [mm] and $D + dD$ and $\alpha = (\int_{D_{min}}^{D_{max}} D^{\mu} \mathrm{e}^{-\Lambda D}\, dD)^{-1}$ is a normalization factor taking into account the finite range of possible drop sizes between D_{min} and D_{max}. The Gamma DSD depends on three parameters: the shape $\mu > -1$ [-], the rate $\Lambda > 0$ [mm^{-1}] and the concentration $N_t > 0$ [m^{-3}].

2.3.2 Gaussian anamorphosis of DSD parameters

DSD parameters (μ,Λ,N_t) can be interpreted as correlated random functions in space and time with a theoretical multivariate distribution function F. Simulating realistic DSD fields means generating independent realizations of (μ,Λ,N_t) according to F. In theory, this requires the complete knowledge of all the finite-dimensional distributions of F, which is rarely the case in practical applications. Furthermore, finding a simulation algorithm for any given distribution function F is known to be a very difficult problem, as pointed out by Lantuéjoul (2002).

A possible solution is to transform the original distribution F into a Gaussian distribution for which a variety of simulation algorithms have been developed (e.g. sequential simulation, turning bands, spectral decomposition). Such a transformation is called a Gaussian anamorphosis (Journel and Huijbregts, 1978; Guillot, 1999). At the end of the simulation, the inverse transformation is applied to retrieve the original parameters.

For multivariate continuous distributions, a possible Gaussian anamorphosis is given by the stepwise conditional transformation (Leuangthong and Deutsch, 2003). The advantage of stepwise conditional transformation is that it creates independent jointly Gaussian variables that can be simulated separately. The relations between the original variables (e.g. the correlations and higher order moments) are preserved in the back-transformation process.

For n-variate problems, the nth variable is transformed conditionally to the first $n-1$ variables, as follows:

$$Y_n = \Phi^{-1}[F_{n|1,...,n-1}(z_n|z_1,...,z_{n-1})] \tag{2.2}$$

where $F_{n|1,...,n-1}$ is the conditional distribution function of the nth component given $z_1,...,z_{n-1}$ and Φ is the cumulative distribution function of a standardized Gaussian random variable. For $n=1$ this reduces to $Y_1 = \Phi^{-1}[F_1(z_1)]$. If it exists, the inverse transformation is given by $Z_n = F^{-1}_{n|1,...,n-1}[\Phi(y_n)|y_1,...,y_{n-1}]$.

In practical applications where $F_{n|1,...,n-1}$ is unknown, the conditional distributions must be estimated from the sample. This can be done empirically by discretizing the space of parameters or by applying more complex methods (e.g. Diciccio et al., 1993; Hall et al., 1999). In each case, the inverse transformation is approximated using a correspondence table between the original and the transformed variables, which implies that simulated fields will *in fine* be composed of measured values only. Inversion problems may arise when many data share the same value. This is unlikely with continuous distributions but may happen for mixed distributions like those produced by intermittent rain fields where the marginal distribution of N_t has an atom at zero. Therefore the presented simulator is limited to non-intermittent rainfall fields, which is a reasonable assumption up to scales of the order of 20 km, also referred as the meso-γ scale (Orlanski, 1975).

2.4 Geostatistical simulation of DSD fields

2.4.1 Modeling the spatial structure of DSD fields

Previous studies have shown that the 3 parameters μ, Λ and N_t that describe the Gamma DSD model are highly variable in space and time (e.g. Tokay and Short, 1996; Ulbrich and Atlas, 1998). This must be taken into account for accurate radar rain-rate estimation. Because of their variability, μ, Λ and N_t can be seen as realizations of random variables. In theory, rainfall is an intrinsically discrete process consisting of individual drops. Nevertheless, it can be seen as a continuous quantity when considering bulk variables like the rain rate or the radar reflectivity factor integrated over scales larger than a few cubic meters. At these scales, DSD parameters can also be considered continuous in space and time and the values of (μ,Λ,N_t) can be interpreted as realizations of a multivariate random function (Yaglom, 2004). Geostatistics has been developed to provide a mathematical framework for the analysis of such random functions (Matheron, 1965).

The fundamental tool for the analysis of the spatial structure of a random function $Z(x)$ is the semi-variogram (called variogram in the following) :

$$\gamma(h) = \frac{1}{2}\mathrm{E}[(Z(x+h) - Z(x))^2], \tag{2.3}$$

where E denotes the expectation, $x \in \mathbb{R}^2$ is a position vector and $h \in \mathbb{R}^2$ is a separation vector. The variogram is only defined if $Z(x)$ is an intrinsic random function, meaning that its increments $Z(x+h)-Z(x)$ must be second-order stationary (Chilès and Delfiner, 1999, pp.16-17). This assumption, however, is less restrictive than second-order stationarity of the variable $Z(x)$ itself.

If $\gamma(h)$ only depends on the norm of h, the random function is said to be isotropic. In general, however, random functions are anisotropic, meaning that $\gamma(h)$ depends both on the norm of h and on its direction. Finally, a variogram must satisfy some mathematical properties: in particular, $-\gamma(h)$ must be an even and positive definite function with $\gamma(0) = 0$ (Chilès and Delfiner, 1999, pp.57-63). In general, variograms are preferred to covariances for two reasons: (1) they are more general than covariances because they do not assume finite variance of the random function and (2) they do not require any knowledge about the mean of the sample.

In practice, variograms must be estimated from the sample. This can be done by using the following standard expression:

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{x_k - x_l \sim h} [z(x_k) - z(x_l)]^2 \tag{2.4}$$

where $N(h)$ represents the number of observations separated by a vector h. This estimate

is known to be asymptotically unbiased. However, it is sensitive to the presence of outliers and measurement errors in the sample. Therefore, more robust alternatives to the standard variogram estimate have been proposed (e.g. Cressie and Hawkins, 1990).

Note that in general $-\hat{\gamma}(h)$ is not a positive definite function and hence does not verify the mathematical properties of a variogram. Therefore, a common approach is to fit a theoretical model on the sample variogram. Popular variogram models are exponential, spherical or Gaussian functions (Chilès and Delfiner, 1999, pp.80-93). Several variograms can be combined to create nested structures. In particular, the sum of two variograms is still a valid variogram.

2.4.2 Simulation of the DSD fields

This section explains how to generate realistic 2D fields of DSDs using the Geostatistical methods introduced in the previous section.

First, the DSD parameters (μ,Λ,N_t) are fitted on measured DSD spectra and normalized using a Gaussian anamorphosis (see Section 2.3.2) to obtain a new set ($\tilde{\mu}$,$\tilde{\Lambda}$,$\tilde{N}_t$) of independent and centered parameters. Sample variograms are computed on the transformed parameters and fitted using a theoretical variogram model. A Gaussian simulation algorithm is used to generate independent fields of ($\tilde{\mu}$,$\tilde{\Lambda}$,$\tilde{N}_t$) with spatial structure given by the fitted variogram models. At the end of the simulation, the inverse anamorphosis is applied to retrieve the original DSD parameters. Bulk variables characterizing rainfall such as the radar reflectivity factor Z [$\mathrm{mm^6\,m^{-3}}$] can be derived from the fields at both horizontal and vertical polarization using the following expression:

$$Z_{H|V} = \frac{10^6 w^4}{\pi^5 |K|^2} \int_{D_{min}}^{D_{max}} \sigma_{B_{H|V}}(D) N(D) dD \tag{2.5}$$

where $\sigma_{B_{H|V}}(D)$ [$\mathrm{cm^2}$] is the backscattering cross-section of a drop with diameter D [mm] at the given polarization, w [cm] is the wavelength, and K is the dielectric factor of liquid water, a function of the complex relative permittivity.

The simulation method described above can be applied to produce both conditional and non-conditional simulations. Conditional simulations are particular realizations which honor observed values at some specified locations within the considered domain, whereas non-conditional simulations are just independent realizations of the same random function. Since non-conditional simulations can be easily transformed into conditional ones (e.g. Chilès and Delfiner, 1999, p.465), the following application will focus solely on the simulation of non-conditional fields.

2.5 Application

This section presents an application of the proposed DSD simulator using the statistical software package "R" (*http://www.r-project.org*) together with the "gstat" package by Pebesma (2004).

2.5.1 Data

In theory, the description of the spatial and temporal variability of DSD fields requires large data sets of DSD measurements in space and time. Such data sets are not yet available at a resolution that is sufficient to capture the variability of DSD fields over a large range of scales. Therefore, the following application has been parameterized using time series of DSD measurements instead of spatial measurements. The considered time series was collected on 11 September 1998 during the HIRE'98 experiment that took place in Marseille, southern France (Uijlenhoet et al., 1999b). A rain event of about 2 h and representative of intense Mediterranean precipitation was simultaneously observed by an optical disdrometer and by an S-band weather radar (10 cm wavelength) operated by Météo France and located at about 100 km from Marseille. The optical disdrometer was operated at a 20-s time resolution and collected 415 DSD measurements during the event. The total rain amount seen by the disdrometer over the 2 hours was 27 mm for a maximum intensity of 80 $\mathrm{mm\,h^{-1}}$. The choice of the 20-s time resolution is a trade-off between availability of enough data for the structural analysis and limitation of the sampling effects due to high temporal resolution.

2.5.2 DSD fitting

For each of the 415 DSD spectra recorded by the disdrometer, a three-parameter Gamma DSD was fitted using the maximum likelihood method. This method has already been employed in previous investigations including Haddad et al. (1997) and Kliche et al. (2008). Maximum likelihood estimators are known to be asymptotically unbiased, efficient and Gaussian distributed (e.g., van der Vaart, 1998). For small samples, however, they must be handled with care since significant uncertainty can be introduced in the estimates (Uijlenhoet et al., 2006). Figure 2.1 shows the time series of the fitted DSD parameters together with their empirical marginal distributions. The average number of drops per fit is 242. Only 7 DSD spectra contained less than 30 drops and could not be fitted properly. Note the large correlation (0.99) between μ and Λ and the negative correlation (-0.40) between μ and N_t, respectively between Λ and N_t (-0.37).

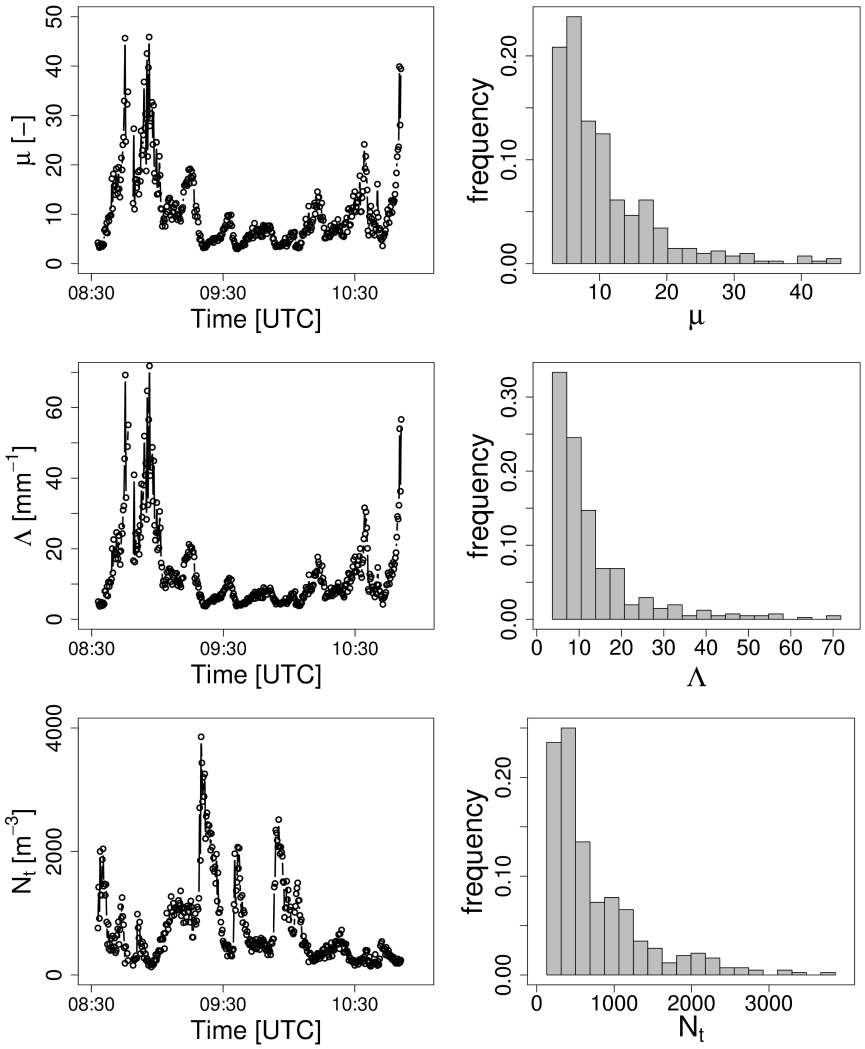

Figure 2.1: Time series and empirical distributions of the DSD parameters (μ,Λ,N_t) before the anamorphosis.

2.5.3 Gaussian anamorphosis of DSD parameters

Clearly, the marginal distributions of the DSD parameters are not Gaussian. Therefore, a Gaussian anamorphosis (see Section 2.3.2) is performed to obtain a new set of independent parameters ($\tilde{\mu}$,$\tilde{\Lambda}$,$\tilde{N}_t$) with joint Gaussian distribution. The conditional distributions needed for the transformation are approximated by discretizing the space of parameters into 30 regularly spaced bins. Here we can take advantage of the strong correlation (0.99) between μ and Λ, which makes it particularly easy to estimate their conditional distributions. The number of discretization bins has been chosen after testing several alternative possibilities and is a trade-off between availability of enough points for the estimation and limitation of the bias. The normality of the joint distribution is confirmed by performing a Mardia test (Mardia, 1970) and a multivariate version of the Shapiro-

Wilk test (Srivastava and Hui, 1987). The independence assumption is confirmed by performing a standard t-test on the correlations. Figure 2.2 shows the transformed DSD parameters together with their empirical marginal distributions. A correspondence table between original and transformed variables is used to approximate the inverse transformation.

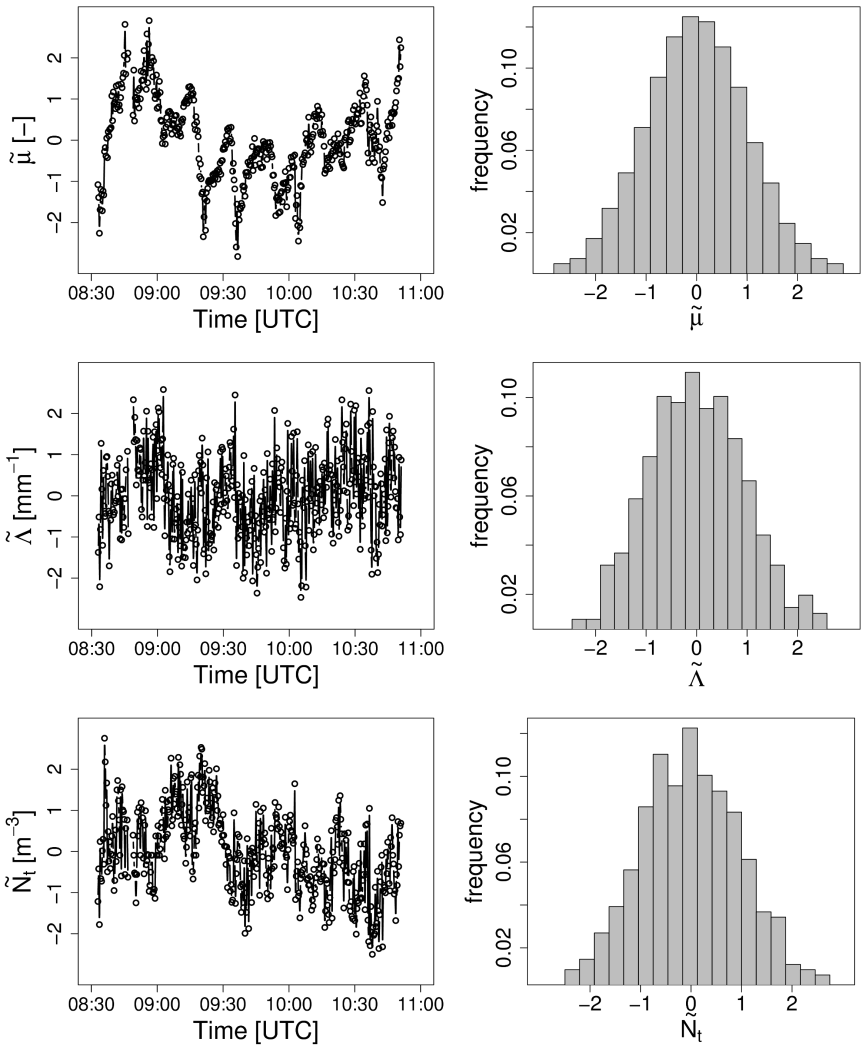

Figure 2.2: Time series and empirical distributions of the DSD parameters ($\tilde{\mu}$,$\tilde{\Lambda}$,$\tilde{N}_t$) after the anamorphosis.

2.5.4 Testing Taylor's hypothesis

According to Berne et al. (2004), the average rainstorm movement velocity during the described event was about 12.5 $\mathrm{m\,s^{-1}}$ in the N-E direction (i.e., azimuth of 50 degrees). Time series of ($\tilde{\mu}$,$\tilde{\Lambda}$,$\tilde{N}_t$) can thus be converted into a 1D range profile along the direction of advection by assuming Taylor's hypothesis of frozen turbulence. The spatial resolution

(250 m) of this profile is determined by the time resolution (20 s) and the average advection speed (12.5 $\mathrm{m\,s^{-1}}$).

Note that Taylor's hypothesis is a strong assumption that is not required by the simulator but necessary in absence of spatial DSD data. Reflectivity measurements taken by the radar can be used to quantify the quality of Taylor's hypothesis. The method consists in shifting the fields in time and space for a given advection speed and direction (in this case 12.5 $\mathrm{m\,s^{-1}}$ and 50 degrees with respect to the North). If the fields result from pure advection by a constant wind, the shifted pixels should overlap on average, meaning that in the (Z,Z_{shift}) space, Taylor's hypothesis is represented by a straight line with slope 1 and intercept 0. The quality of Taylor's hypothesis can thus be verified by looking at the correlation between Z and Z_{shift} and at the ratio of their means. Figure 2.3 represents these values for time shifts between 5 and 20 minutes. It can be seen that for a time shift of 5 min, both the correlation (0.95) and the ratio of means (0.98) stay close to 1, indicating that Taylor's hypothesis is a good approximation over short time periods. For time shifts of 10-20 min, the correlation (0.80,0.63,0.50) decreases rapidly. The ratio of means (0.96,0.94,0.91) also decreases slightly. The conclusion is that Taylor's hypothesis is not a good approximation for time shifts larger than 20 min and implies that consistent structural analysis is limited to time shifts less or equal to 20 min, corresponding to 15 km in the spatial domain.

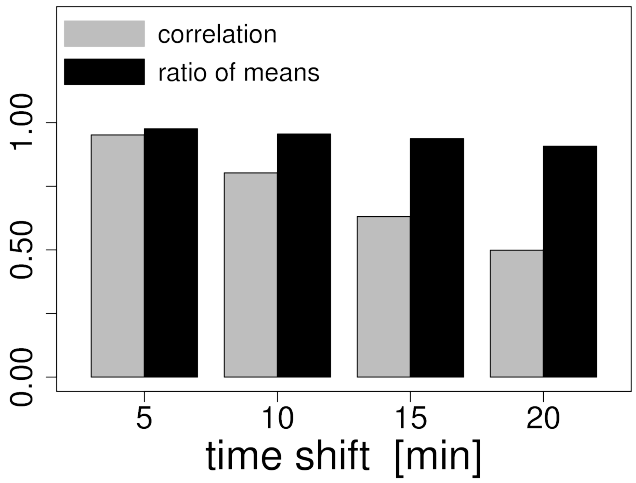

Figure 2.3: Verification of Taylor's hypothesis for time shifts between 5 and 20 min. Under the hypothesis of pure advection, both the correlation and the ratio of means should be equal to 1.

2.5.5 Fitting a variogram model

Using Taylor's hypothesis with 12.5 $\mathrm{m\,s^{-1}}$ advection speed, the DSD time series ($\tilde{\mu}$,$\tilde{\Lambda}$,$\tilde{N}_t$) are converted into range profiles and their sample variograms are computed. The sample

variograms are fitted by combining two spherical variogram models, one for the short-range variability and one for the long-range variability. A spherical variogram model is given by:

$$\gamma_{sph}(h) = \begin{cases} C_0 + C_1 \left(\frac{3}{2}\frac{h}{a} - \frac{1}{2}\frac{h^3}{a^3} \right) & h < a \\ C_0 + C_1 & h \geq a \end{cases} \tag{2.6}$$

where C_0 denotes the nugget, C_1 the partial sill, h the distance lag and a the range. The values of the fitted parameters are given in Table 2.1. Figure 2.4 shows the fitted variograms for each DSD parameter. The total variogram for each DSD parameter is the sum of the two spherical models.

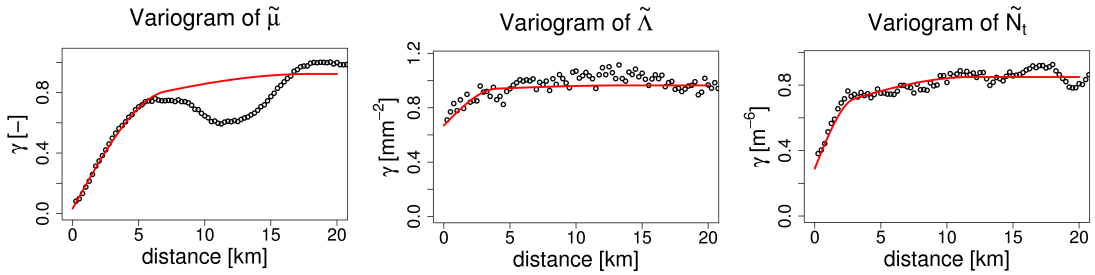

Figure 2.4: Sample variograms of ($\tilde{\mu}$,$\tilde{\Lambda}$,$\tilde{N_t}$) after Gaussian anamorphosis. The values of the parameters are given in Table 2.1.

Table 2.1: Nugget [-], range [km] and partial sill [-] values of two nested spherical models fitted to the sample variograms of ($\tilde{\mu}$,$\tilde{\Lambda}$,$\tilde{N_t}$).

	Nugget	Range 1	Partial Sill 1	Range 2	Partial Sill 2
$\tilde{\mu}$	0.03	7	0.63	18	0.26
$\tilde{\Lambda}$	0.67	4	0.25	14	0.04
$\tilde{N_t}$	0.29	3	0.35	12	0.21

2.5.6 Estimation of the anisotropy

The variograms defined in Table 2.1 only describe the spatial structure along the average direction of advection. The complete 2-dimensional structure of the fields can be expressed in terms of the 1D variograms by adding information on anisotropy derived from radar reflectivity measurements. For the considered event, 28 reflectivity measurements were collected by an S-band weather radar from 08h35 to 10h50 UTC at a 5-min temporal resolution over the considered area. The analysis of these reflectivity fields as well as their corresponding 2D variograms indicates that the fields have geometric anisotropy (i.e., the 2D variograms exhibit elliptic structure). The average azimuth of smallest variability (the major axis of the ellipse) is 320 degrees. The average anisotropy ratio (the ratio between

the minor and the major axis of the ellipse) is about 0.7. As an example, Figure 2.5 shows the radar reflectivity field and the corresponding 2D variogram map at 09h35 UTC. The major axis of the anisotropy ellipse is approximatively in the N-W direction (azimuth of 315 degrees). The ratio between the minor and the major axis of the ellipse is about 0.5. In the rest of the application, the average anisotropy values (azimuth of 320 degrees and anisotropy ratio of 0.7) are used to characterize the 2D variograms.

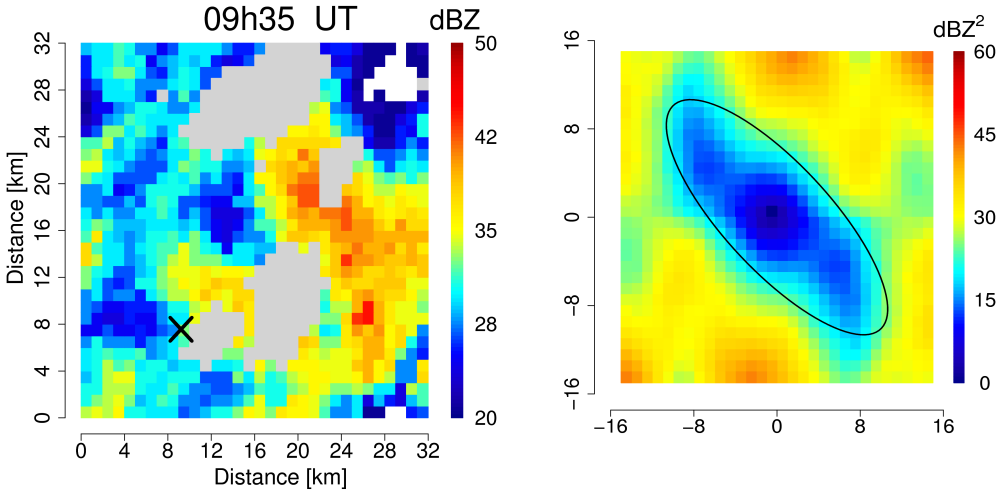

Figure 2.5: Observed radar reflectivity field at 09h35 UTC over the considered 32 × 32 km^2 domain (left panel) together with its corresponding 2D variogram map (right panel). Grey pixels correspond to ground echos. Dry regions are represented in white. The location of the disdrometer is indicated by the black cross in the lower left-hand corner. The anisotropy direction is given by an azimuth of 315 degrees. The anisotropy ratio is about 0.5.

2.5.7 Simulation of DSD fields

Using the variogram model described in Table 2.1 together with the average geometrical anisotropy described above, 200 Gaussian DSD fields have been simulated on a 32 × 32 km^2 domain consisting of 16'384 pixels of size 250 × 250 m^2. Such an area adequately represents the meso-γ scale and is large enough to contain all the measured variability in the DSD without being too large to avoid problems related to intermittency. Using inverse anamorphosis, the Gaussian fields are back-transformed into the original DSD fields. The backscattering cross-sections $\sigma_{H|V}(D)$ are computed at both horizontal and vertical polarization using the T-matrix code (Mishchenko and Travis, 1998) and used in Equation 2.5 to derive the corresponding reflectivity fields at S-band (negligible attenuation).

Figure 2.6 shows an example of a simulated DSD field together with the corresponding rain rate R, horizontal reflectivity Z_h and differential reflectivity $Z_{dr} = Z_h - Z_v$ (in dB).

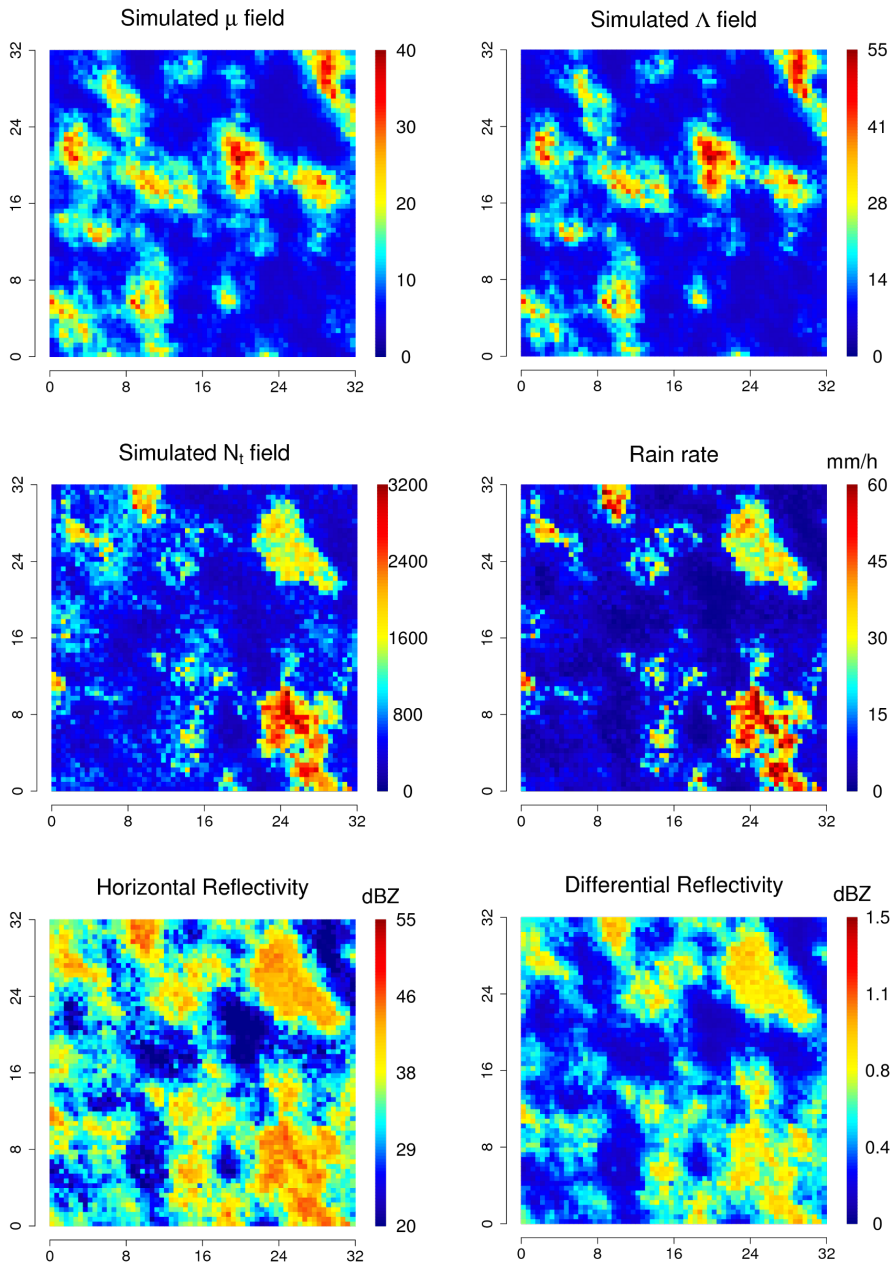

Figure 2.6: Example of a simulated DSD field (μ,Λ,N_t) together with its corresponding rain rate, horizontal and differential reflectivity fields.

2.5.8 Comparison with radar measurements

In order to assess the quality of the simulation, the simulated reflectivity fields Z_h derived using Equation 2.5 are compared to the observations taken by the radar. Because the radar resolution is 1×1 km^2 and the simulation resolution is 250×250 m^2, the

simulations are averaged over blocks of 4 × 4 pixels before comparison. Moreover, all ground echos are removed from the data before the comparison. To avoid issues due to intermittency, the comparison is restricted to radar measurements taken between 09h15 and 09h55 UTC (10 measurements), for which intermittency is negligible. The selected fields have mean values ranging from 30.5 to 34.2 dBZ, with an average of 33.8 dBZ. The standard deviation of the observed fields is between 4.6 dBZ and 7.6 dBZ, with an average of 5.6 dBZ. Figure 2.7 shows that these values are in good agreement with the simulated fields. Indeed, the simulated reflectivity fields have mean values between 29.3 and 36.4 dBZ, with an average of 32.9 dBZ and their standard deviations are between 5.3 and 7.9 dBZ, with an average of 6.3 dBZ. This shows that the simulated fields adequately reproduce the first order moments (mean and standard deviation) of the observed reflectivity fields. Furthermore, the simulations also exhibit a similar asymmetry in the distributions as shown in Figure 2.7.

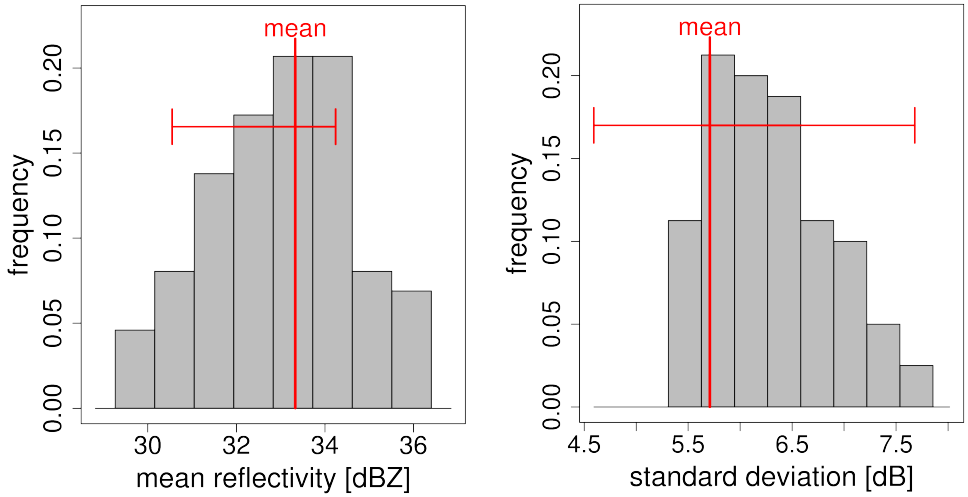

Figure 2.7: Histogram of the mean and of the standard deviation of 200 simulated reflectivity fields. Simulated values are represented by grey boxes. The red cross represents the minimum, mean and maximum values observed by the S-band radar during the event.

Figure 2.8 shows a comparison between the spatial structure of the simulations and the spatial structure of the observations. In order to keep the figure readable, an average variogram has been plotted for both simulated and observed reflectivity fields. The dispersion about the average variogram is represented (for each distance class) by the 10 % and 90 % quantiles.

It can be seen that the simulated and observed variograms are in good agreement except for the first 2-3 km where the simulations exhibit slightly too much variability. Different explanations can be given for this: (1) there is a smoothing effect in the observed reflectivity fields due to the power distribution in the radar beam. Hence, radar pixels that are far away from the antenna are averaged over large overlapping volumes and

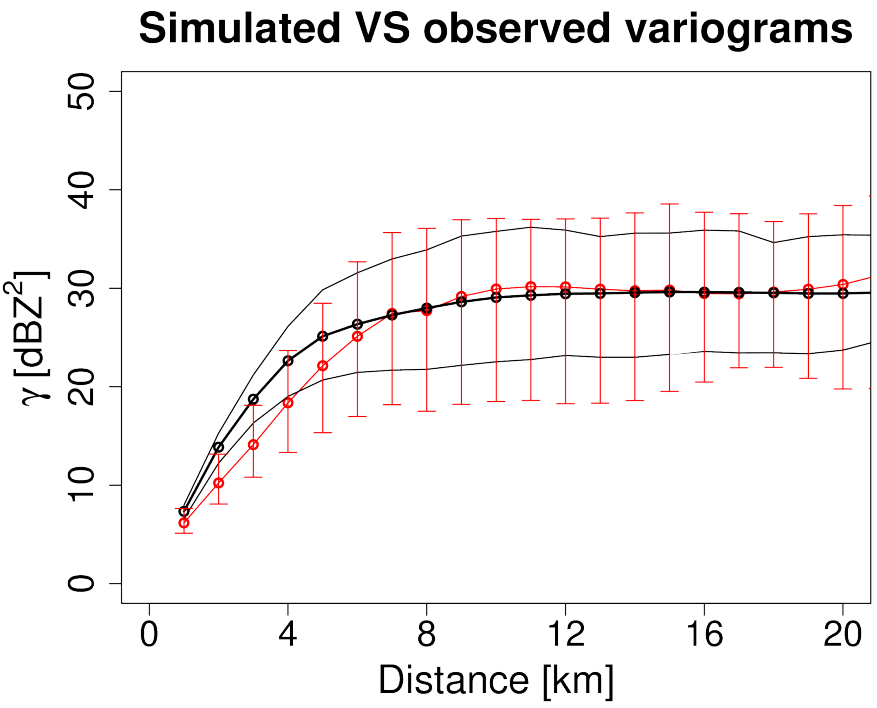

Figure 2.8: Average radar reflectivity variogram (in red with vertical bars) and average simulated reflectivity variogram (in black). The lower and upper bounds represent the 10 % and 90 % variogram quantiles for each distance lag computed on 28 radar pictures and 200 simulations.

exhibit more correlation. This implies a smaller slope in the variograms computed using radar observations, especially at small distance lags where the averaging effect is the strongest; (2) the Gaussian anamorphosis described in Section 2.3 was computed on a small sample of 415 DSD observations. The inverse transformation was approximated using a correspondence table between the original and transformed variables. The discretization introduced by this technique may explain the rather strong variability at short ranges in the simulations; (3) Taylor's hypothesis of frozen turbulence was shown to be acceptable but not perfect. Furthermore, parameters like the advection speed, the anisotropy direction and anisotropy ratio were estimated using radar data and supposed constant over the entire event, which is not exactly true. Despite the above mentioned problems, the proposed simulator produces very encouraging results that are consistent with the observations. The simulated 2D fields are in good agreement in terms of first and second order moments. The spatial structure is also in good agreement with the observations.

2.6 Conclusions

DSD fields are highly variable in space and time. To investigate issues related to this variability in radar rain-rate estimation, a stochastic simulation framework has been proposed. It is based on Geostatistics and considers the parameters (μ,Λ,N_t) of the Gamma DSD as realizations of a multivariate random function.

In order to take advantage of simple and fast Gaussian simulation algorithms, the distribution of the three DSD parameters is normalized using a Gaussian anamorphosis technique. The spatial structure of the fields is quantified using variograms. Gaussian fields with identical spatial structure are generated and then back-transformed into the original parameter space. In this way, realistic conditional and non-conditional 2D fields of DSDs can be generated.

The proposed approach is applied to DSD measurements collected during an intense Mediterranean rainfall event. As only DSD time series are available, Taylor's hypothesis is used to convert time series into range profiles. The anisotropy direction and anisotropy ratio of the fields are derived from measurements taken by an S-band weather radar. The reflectivity fields derived from the simulated DSD fields are compared to radar measurements. Both the mean and the standard deviation are in very good agreement. The spatial structure is also coherent with the observations, indicating that the proposed simulator generates realistic 2D fields of DSDs.

The main limitation of the proposed simulator is its inability to simulate intermittent DSD fields. This limits the size of the simulated domains to the meso-γ scale (about 20 km) at which non-intermittent rain fields are plausible. Taylor's hypothesis is not required to run the simulator, but is used because of the lack of spatial DSD measurements. Better results can even be expected when spatial DSD data will be available.

3 Stochastic simulation of intermittent DSD fields in time[1]

3.1 Summary

This chapter generalizes previous results presented in Chapter 2 and introduces a method for the stochastic simulation of intermittent fields of raindrop size distributions in space and time. Rainfall intermittency is modeled using an indicator field and simulated using sequential indicator simulation. The raindrop size distribution is described by a Gamma distribution with 2 or 3 stochastic parameters and simulated using sequential Gaussian simulation. Separable space-time variograms are used to model the spatial and temporal structures of all these variables. A simple and user-oriented procedure for the parameterization of the simulator is proposed. The only data required are DSD time series and radar rain-rate (or reflectivity) measurements. The proposed simulation method is illustrated for frontal and convective precipitation using real data collected in the vicinity of Lausanne, Switzerland. The spatial and temporal structures of the simulated fields are evaluated and validated using DSD measurements from 8 independent disdrometers.

1. This chapter is a slightly modified version of the article by Schleiss, M., J. Jaffrain and A. Berne, 2012: Stochastic simulation of intermittent DSD fields in time, *J. Hydrometeorol.*, vol.13, No.3, 621-637

3.2 Introduction

Precipitation is an important and complex part of the Earth's water cycle and climate system. Although it is very beneficial in every day's life (it produces most of the fresh water on the Earth's surface and is therefore vital for plants and animals), it can also have dramatic effects on humans and infrastructures (e.g., floods, droughts, landslides, avalanches). Consequently, a lot of efforts have been devoted over the last decades to measure, quantify, model and predict precipitation with increasing accuracy. Nowadays, most of the individual physical mechanisms (e.g., nucleation, diffusional growth, collisional growth, evaporation and breakup) involved in the formation of clouds and precipitation are fairly well understood (Pruppacher and Klett, 1997). The study and modeling of their complex interactions is, however, still an active field of research. At larger scales, additional complexity is introduced through the influence of land and sea surface processes (e.g., changes in temperature and evaporation) which significantly contribute to the spatial and temporal variability of precipitation (Koster and Suarez, 1995; Roe, 2005). As a result, rainfall is generally non-stationary, inhomogeneous, intermittent and largely variable in space and time.

Modern hydrological models require accurate rainfall estimates with high spatial and temporal resolutions. Weather radars provide relatively good inputs to these models but their estimates can be affected by large uncertainties (Wilson and Brandes, 1979; Krajewski and Smith, 2002; Villarini and Krajewski, 2010). More robust and better data can be expected if measurements coming from several independent rainfall sensors (e.g., satellites, ground-based radars, rain gauges, disdrometers, microwave links) are combined. Finding robust and reliable merging techniques that take into account the relative uncertainties and error contributions of each instrument, is known to be a very difficult problem. In particular, there is a strong lack of reference data against which the different instruments and retrieval algorithms could be compared and evaluated objectively. A common approach therefore consists in using stochastic rainfall simulators. The latter are very interesting and useful because they allow to generate large numbers of statistically homogeneous fields at high spatial and temporal resolutions (far beyond the capabilities of traditional rainfall sensors) that can be used as reference data in further analysis.

In this chapter, a stochastic rainfall simulator at the mesoscale (domain size between 10-100 km) is presented which adequately reproduces the main statistical features of a given rain event, i.e., the spatial and temporal structures, the rainfall intermittency and, most importantly, the (rain)drop size distribution (DSD). From the DSD it is then possible to derive or approximate most physical quantities of interest, i.e., the shape, fall speed and energy of the individual drops. In addition, quantities like the rain rate, the median drop diameter, the radar reflectivity, the differential reflectivity, specific attenuation and differential phase shift (which are very important for remote sensing of precipitation and explicitly depend on the DSD) can also be derived. As a result,

simulated DSD fields are very valuable for many practical applications including soil erosion problems (Kinnell, 2005), runoff and infiltration (Smith et al., 2009), atmospheric deposition processes (Beier and Hansen, 1993), the $Z - R$ relationship for weather radars (Battan, 1973; Ulbrich, 1985), attenuation correction algorithms at C- and X-band (e.g., Bringi et al., 1990; Delrieu et al., 1999; Testud et al., 2000), rainfall measurement using satellites (e.g., TRMM and GPM) and microwave links from telecommunication networks (Messer et al., 2006; Leijnse et al., 2007c).

So far, most of the stochastic rainfall simulation techniques have been focusing on the simulation of daily, monthly or annual precipitation amounts using chain-dependent stochastic processes (e.g., Katz, 1977; Richardson, 1981; Wilks, 1999). Also, significant efforts have been devoted to the simulation of spatially correlated rain rate fields using, for example, principal components (Bouvier et al., 2003), clustering processes (Waymire et al., 1984; Rodríguez-Iturbe and Eagleson, 1987; Sivapalan and Wood, 1987), self similarity (Gupta and Waymire, 1993; Menabde et al., 1997) and Geostatistics (Guillot, 1999). Only a few studies have investigated the possibility to simulate rainfall fields (or rain profiles) with higher level of detail, for example by including the (rain)drop size distribution (e.g., Krajewski et al., 1993; Berne and Uijlenhoet, 2005; Lavergnat and Golé, 2006; Lee et al., 2007). In Chapter 2, a (relatively) simple DSD simulation method based on Geostatistics has been proposed. The method was, however, limited to small, non-intermittent and instantaneous rainfall fields. In this chapter, the method presented in Chapter 2 is extended and a new DSD simulator capable of generating intermittent DSD fields in space and time is proposed.

This chapter is structured as follows: Section 3.3 is devoted to the description of the models and mathematical tools needed for the simulation. Section 3.4 explains how to parameterize the simulator. Section 3.5 provides a detailed step by step description of the simulation algorithm and explains how the main variables of interest can be computed from the simulated DSD fields. In Section 3.6, the capabilities of the simulator are illustrated and evaluated using real data collected in the vicinity of Lausanne, Switzerland. The advantages and limitations of the proposed method are discussed in Section 3.7 and the conclusions and perspectives are given in Section 3.8.

3.3 Modeling

A substantial amount of material presented in this section is similar to Section 2.3 and 2.4 in Chapter 2 and will not be repeated here. Instead, the major differences between the two approaches are highlighted.

3.3.1 Drop Size Distribution

The drop size distribution is described using a Gamma distribution (Ulbrich, 1983; Willis, 1984; Ulbrich and Atlas, 2007) with three parameters (μ, Λ and N_t), exactly as in Section 2.3.1.

$$N(D) = \alpha N_t D^{\mu} e^{-\Lambda D}. \tag{3.1}$$

Because of the natural variability of rainfall, the DSD parameters (μ, Λ, N_t) can be seen as realizations (in time and space) of an underlying multivariate random function. Furthermore, some studies (Zhang et al., 2001; Seifert, 2005) suggest that there a deterministic relation between the shape and the rate parameter:

$$\Lambda = f(\mu), \tag{3.2}$$

where f depends on the type of precipitation and on the local climatology. Such relations are interesting because they allow to reduce the number of stochastic parameters to be simulated. However, Moisseev and Chandrasekar (2007) argued that $\mu - \Lambda$ relations must be handled with extreme care because they might be the result of statistical errors and data filtering of disdrometer measurements. In any case, this is not a critical issue because the proposed simulation method can be easily adapted to include an additional third stochastic parameter, as in Schleiss et al. (2009). But for simplicity, only the first case (two stochastic parameters μ and N_t and a deterministic relationship between Λ and μ) is presented. Finally, it must be noted that the proposed simulation technique can easily be adapted to any other parametric DSD model (e.g., exponential or log-normal).

3.3.2 Intermittency

Intermittency, i.e., the presence or absence of rainfall, is modeled using an indicator field

$$I(x) = \begin{cases} 1 \text{ if } R(x) > 0 \\ 0 \text{ else} \end{cases} \tag{3.3}$$

where $R(x)$ [$\mathrm{mm\,h^{-1}}$] represents the instantaneous rain rate at location x. Using this notation, $R(x)$ can be seen as the product of two random functions

$$R(x) = I(x)R^{+}(x), \tag{3.4}$$

where $R^{+}(x) > 0$ represents the nonzero rain rate at location x (for consistency, R^{+} is only defined for rainy locations). Using the same decomposition, Barancourt et al. (1992) derived an interesting scheme for the interpolation and simulation of intermittent rain rate fields. Following the same idea, one can extend this result to the simulation of intermittent DSD fields by applying the same decomposition to the two DSD parameters:

$$\mu(x) = I(x)\mu^+(x) \qquad N_t(x) = I(x)N_t^+(x). \tag{3.5}$$

Using this notation, an intermittent DSD field (at a given time t and location x) is given by a triplet (I, μ^+, N_t^+) where I represents the rainfall indicator field and (μ^+, N_t^+) the DSD parameters. But for conciseness (and because the DSD is not defined for dry locations anyway), we will drop the "+" sign and simply denote the DSD field by (I, μ, N_t).

3.3.3 Anamorphosis

For now, suppose that a given rainfall indicator field I has been simulated and that the rainy and dry locations are known. The only parameters that remain to be simulated are μ and N_t. Simulating realistic DSD fields means generating independent realizations of (μ, N_t) according to some unknown bivariate distribution, a problem known to be very difficult in general. A possible solution, known as Gaussian anamorphosis (Leuangthong and Deutsch, 2003), is to transform the original parameters (μ, N_t) into independent Gaussian variables ($\tilde{\mu}$,$\tilde{N}_t$) for which a variety of efficient simulation algorithms are known, e.g., turning bands (Journel and Huijbregts, 1978; Guillot, 1999) and sequential simulation (Ripley, 1987). For details about the anamorphosis procedure, see Section 2.3.2.

3.3.4 Space-time structure

The space-time structure of I, $\tilde{\mu}$ and $\tilde{N}_t$ is modeled using a space-time variogram (Matheron, 1965; Chilès and Delfiner, 1999):

$$\gamma(h, \tau) = \frac{1}{2}\mathrm{E}[(Z(x + h, t + \tau) - Z(x, t))^2], \tag{3.6}$$

where $h \in \mathbb{R}^n$ is a separation vector, $\tau \in \mathbb{R}$ a given time lag and E denotes the expectation. Unfortunately, fitting a valid space-time variogram can be very difficult in general, especially when few data are available. Therefore, a common approach consists in separating the spatial and the temporal variations:

$$\gamma(h, \tau) = \gamma_S(h) + \gamma_T(\tau), \tag{3.7}$$

where $\gamma_S(h)$ and $\gamma_T(\tau)$ represent the spatial respectively temporal variogram of $Z(x, t)$. For the rainfall indicator field I, both $\gamma_S(h)$ and $\gamma_T(\tau)$ can be derived and fitted using radar rain-rate or reflectivity data. For the DSD parameters, this is usually not possible because disdrometer networks are usually not dense and large enough for a direct estimation of $\gamma_S(h)$. Hence most of the time, the only variogram that can be computed is the one in the time domain. In this case, the spatial variogram must be approximated

using external information (e.g., the speed and direction of advection and the anisotropy of the rainfall field). If available, polarimetric radar quantities like the differential reflectivity Z_{dr} and the reflectivity at horizontal polarization Z_h can also be used to estimate the spatial structure of $\tilde{\mu}$ and $\tilde{N}_t$.

In the following, a method based on Taylor's hypothesis of frozen turbulence (Taylor, 1938) that allows to approximate the spatial structure of $\tilde{\mu}$ and $\tilde{N}_t$ without using polarimetric radar data is proposed. Taylor's hypothesis states that the rainfall field moves, during short periods of time, with constant velocity v, and that its evolution for time lags up to 20-30 min is small compared to the advection process (Li et al., 2009). As a consequence, the temporal covariance of a variable at time lag τ is equal to the spatial covariance at space lag $h = \tau v$, or equivalently, $\gamma_S(\tau v) = \gamma_T(\tau)$. Hence (short) variations in time can be converted into small variations in space, but only in the direction of advection. For all other directions, the relation between temporal and spatial variations is not known a priori. However, it is reasonable to assume that similar relations exist and can be described using a geometric anisotropy parameter derived from radar rain-rate or reflectivity measurements.

$$\gamma_S(h) = \gamma_T\left(\alpha(h)\frac{\|h\|}{\|v\|}\right) \tag{3.8}$$

$$\alpha(h) = \frac{1}{\alpha} + (1 - \frac{1}{\alpha})\left|\frac{< h, a >}{\|h\|\|a\|}\right| \tag{3.9}$$

where $0 < \alpha \leq 1$ is called the anisotropy ratio (i.e., the ratio between the minor and major axis of the anisotropy ellipse), $a \in \mathbb{R}^2$ is the vector giving the direction of minimum variability (not necessarily identical to the direction of advection), $< h, a >$ denotes the standard scalar product between h and a and $\|\cdot\|$ denotes the euclidean norm in $\mathbb{R}^2$. The anisotropy factor $\alpha(h)$ is supposed to reflect a geometrical property of the rainfall field and is therefore assumed to be identical for the rain rate, the reflectivity, $\tilde{\mu}$ and $\tilde{N}_t$. Under these assumptions, the complete space-time variograms of $\tilde{\mu}$ and $\tilde{N}_t$ is given by:

$$\gamma(h, \tau) = \gamma_T\left(\alpha(h)\frac{\|h\|}{\|v\|}\right) + \gamma_T(\tau). \tag{3.10}$$

Note that it is possible to generalize Equation 3.10 by taking into account a time-dependent speed and direction of advection. In this case, each time period (e.g., 20-30 min) is represented by an average advection vector v_i, an average anisotropy direction a_i and an average anisotropy ratio $\alpha_i(h)$. The space-time variogram of each time period is then given by

$$\gamma_i(h, \tau) = \gamma_T\left(\alpha_i(h)\frac{\|h\|}{\|v_i\|}\right) + \gamma_T(\tau). \tag{3.11}$$

3.4 Parameterization

3.4.1 Required data

Optimal parameterization of the simulator is achieved if both spatial and temporal DSD data are available. Unfortunately, such data are not yet available. In the meantime, alternative parameterizations using fewer data must be considered. In the following, a method that allows to parameterize the simulator using solely DSD time series (collected using one or several disdrometers) and radar rain-rate or reflectivity data for the considered domain and event of interest is proposed. The temporal sampling resolution of the disdrometers must be high enough to capture most of the natural variations occurring in the DSD but not too high to avoid strong sampling effects. Typically, temporal resolutions between 30 s and 1 min are adequate. The size of the simulation domain must be large enough to catch the spatial structure of the rainfall field but not too large for the variogram to be representative over the entire domain. Also, the size of the simulation domain must be consistent with the length of the collected DSD times series and the average speed of advection. A 30-min times series with an average speed of advection of 5 $\mathrm{m\,s^{-1}}$ only corresponds to 9 km along the direction of advection and is unlikely to be representative of a 50 $\times$ 50 $\mathrm{km^2}$ domain. Depending on these parameters, domain sizes between 10 and 100 km can be considered.

3.4.2 Parameters derived from radar data

Radar data are used to estimate (1) the rainfall intermittency, (2) the spatial and temporal variograms of the rainfall indicator field, (3) the speed and direction of advection and (4) the anisotropy parameters of the rainfall field. All these parameters are assumed constant during the considered event or, alternatively, time dependent. For simplicity, the procedure is only illustrated for constant parameters.

The rainfall intermittency is estimated using the radar data by computing the proportion p_W of wet and $p_D = 1 - p_W$ of dry locations within the simulation domain. Note that if the intermittency changes significantly with time, it is better to consider separate time periods with varying intermittency. The spatial structure of the rainfall indicator field is then estimated (for each time period) by computing its average spatial sample variogram $\hat{\gamma}_I^S(h)$ (aggregating the fields in time) and by fitting a valid variogram model $\gamma_I^S(h)$ on it. The temporal structure of the rainfall indicator field is estimated by computing the average sample time variogram $\hat{\gamma}_I^T(\tau)$ (aggregating the time series in space) and by fitting a valid variogram model $\gamma_I^T(\tau)$ on it. This defines the complete space-time variogram $\gamma_I(h,\tau) = \gamma_I^S(h) + \gamma_I^T(\tau)$ of the rainfall indicator field.

The average speed and direction of advection v is determined using a cell or echo-tracking algorithm on each pair of successive radar images (e.g., Rinehart, 1979; Crane, 1989;

Tuttle and Foote, 1990). The anisotropy parameters a and α given in Equations 3.8 and 3.9 are estimated by computing the spatial sample (2D) variogram $\hat{\gamma}_R^S(h)$ of the radar rain-rate or reflectivity data. It is then possible to identify the (average) direction of minimum variability a and the corresponding (average) anisotropy ratio α.

3.4.3 Parameters derived from DSD data

First, the 3-parameter DSD model (μ, Λ, N_t) given in Equation 3.1 is fitted on each collected DSD spectrum. The scatterplot between μ and Λ is used to investigate if there is a relation (possibly nonlinear) between μ and Λ. For simplicity, and because of the high correlation that is usually observed, only linear relations between μ and Λ are considered. The generalization of this technique to nonlinear relations (e.g. power-laws or exponential models) is straightforward. For linear relations, the standard Pearson correlation coefficient is used to decide weather two or three stochastic parameters should be included in the model. As a rule of thumb, a correlation of 0.9 or higher is considered sufficient to drop one of the parameters. Otherwise, all three DSD parameters are kept and the simulation scheme is adapted to include an additional stochastic parameter. Note that this also includes the case of (complex) nonlinear relations which are then captured automatically by the Gaussian anamorphosis and do not need to be parameterized. However, it must be emphasized that 3 parameter simulations require far more DSD data (at least ten times more than with 2 parameters) and significantly increase the simulation time. It is therefore highly recommended to work with 2 stochastic parameters whenever it is possible.

In the next step, the DSD parameters (μ, N_t) are detrended following a procedure described in the Appendix. This is particularly important for highly intermittent rain fields where the constant alternating between dry and rainy periods significantly affects the variograms of μ and N_t. The detrended time series of (μ, N_t) are then transformed into independent Gaussian variables $\tilde{\mu}$ and $\tilde{N}_t$ using a Gaussian anamorphosis (see Section 3.3.3). The temporal variograms of $\tilde{\mu}$ and $\tilde{N}_t$ are estimated and fitted using two valid temporal variogram models $\gamma_{\tilde{\mu}}^T(\tau)$ and $\gamma_{\tilde{N}_t}^T(\tau)$. Finally, the complete space-time variograms $\gamma_{\tilde{\mu}}(h, \tau)$ and $\gamma_{\tilde{N}_t}(h, \tau)$ needed for the simulations are estimated using Equation 3.10.

3.5 Simulation

Generating a simulation means choosing a simulation algorithm. The choice of this algorithm depends on the type of variables to be simulated. An intermittent DSD field consists of two parts: (1) a rainfall indicator field $I(x)$ and (2) a bivariate (μ, N_t) DSD field (for rainy locations only) which can be transformed into two independent Gaussian fields $\tilde{\mu}$ and $\tilde{N}_t$ with zero mean and unit variance. Although different simulation methods

can be considered, we focus on a very general technique called sequential simulation (Ripley, 1987).

3.5.1 Sequential simulation

Consider the case of a random function $Z(x)$ with known values $z_1, ..., z_M$ at locations $x_1, ..., x_M$. If $M = 0$, the simulation is said to be non-conditional. Suppose that $Z(x)$ needs to be simulated at N new locations $x_{M+1}, ..., x_{M+N}$, conditionally to the previous values $z_1, ..., z_M$. The sequential simulation paradigm states that $Z(x_i)$ can be simulated sequentially by randomly sampling from the conditional distribution of $Z(x_i)$ given all the previously simulated or prescribed values and by including the outcome z_i in the conditioning data set for the next step. The practical difficulty is that in general, the conditional probabilities are not known, except for the Gaussian case (with known mean) for which the conditional distribution of $Z(x_i)$ knowing $\{z_j\}_{j<i}$ is Gaussian with mean $z_i^\star$ (the simple kriging estimator at x_i) and variance $\sigma_i^{\star 2}$ (the associated kriging variance) (e.g., Chilès and Delfiner, 1999, p.164). Note also that for indicator fields, i.e., fields which take only 0 and 1 values, the conditional distribution is equal to the conditional expectation $\mathrm{E}[Z(x_i)|Z(x_1) = z_1, \ldots, Z(x_{i-1}) = z_{i-1}]$, i.e., the probability that $Z(x_i)$ is equal to 1. This probability is not known in general but can be estimated like in the Gaussian case by the simple kriging estimate $z_i^\star$ (Alabert, 1987). The only theoretical problem with this method is that $z_i^\star$ can be less than 0 or greater than 1, even for very simple variogram models. If this is the case, the corresponding probabilities are set to 0 or 1 accordingly. At the end of the simulation, the variogram of the simulated indicator field is checked against the model to discard possible artifacts due to these thresholding.

3.5.2 DSD Simulation algorithm

This paragraph provides a detailed step by step description of the algorithm used to generate intermittent DSD fields in space and time. For a visual diagram of the simulation algorithm, see Figure 3.1.

First, sequential indicator simulation (SIS) is used to generate an indicator field with mean p_W and space-time structure given by $\gamma_I(h, \tau)$. Locations and periods for which the outcome of the simulation is 1 are considered wet. The others are considered dry. Only the wet periods and locations are used to simulate the DSD parameters. In the next step, sequential Gaussian simulation (SGS) is used to generate a Gaussian field for $\tilde{\mu}$ with mean 0 and space-time structure given by $\gamma_{\tilde{\mu}}(h, \tau)$. A second Gaussian field with mean 0 and space-time structure given by $\gamma_{\tilde{N}_t}(h, \tau)$ is generated for $\tilde{N}_t$. At the end of the simulations, an inverse anamorphosis is applied to back-transform ($\tilde{\mu}$,$\tilde{N}_t$) into the original parameter space (μ,N_t). If the original time series of μ and N_t were detrended prior to analysis (see Appendix), the external drifts (mainly caused by the transitions

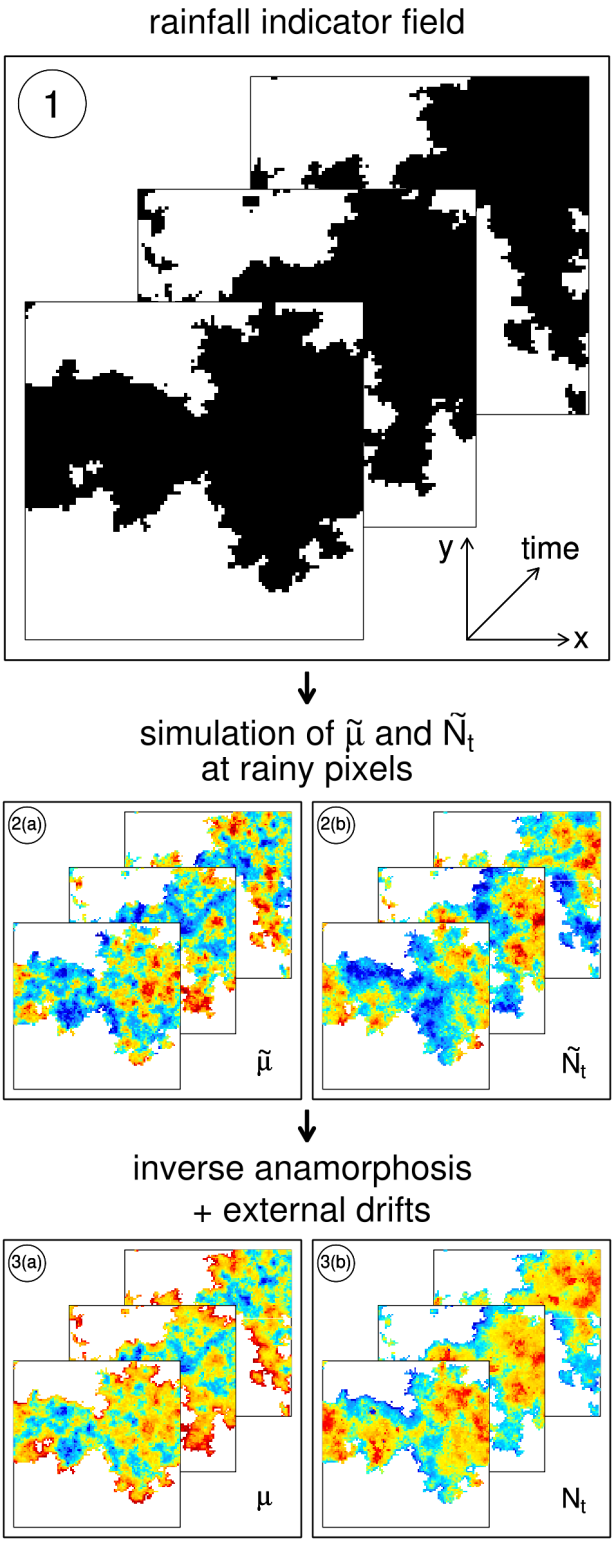

Figure 3.1: Visual illustration of the DSD simulation algorithm (from top to bottom).

between dry and rainy periods or locations) are added back to the simulated fields at this stage. Finally, Λ is derived through its deterministic relation with μ. If desired, The simulation procedure can be repeated several times to obtain different realizations in space and time.

3.5.3 Post-processing

One of the main advantages of simulating DSD fields is that they can be used to derive most quantities of interest related to remote sensing. This includes the rain rate R [mm h^{-1}], the median drop diameter D_0 [mm], the radar reflectivity $Z_{h|v}$ [dBZ] at horizontal and vertical polarization, the differential reflectivity Z_{dr} [dB], the specific attenuation $A_{h|v}$ [dB km^{-1}] at horizontal and vertical polarization and the differential phase shift on propagation K_{dp} [$^\circ$ km^{-1}]. For more details on these quantities and on how to compute them from the DSD, the reader is referred to Section 1.2.2.

3.6 Illustration using real data

This section illustrates the capabilities of the previously described DSD simulator using real data collected in the vicinity of Lausanne, Switzerland. Two rain events (frontal and convective) which are very different in terms of magnitude and temporal dynamics are selected to better illustrate the simulator's capability of reproducing different rainfall types and structures. Note that the goal is neither to predict nor to reproduce the observed events but to generate synthetic rainfall fields sharing the same statistical properties, i.e., spatial and temporal structures, intermittency and raindrop size distributions.

3.6.1 Data

The DSD data are collected using a network of 16 optical disdrometers of type Parsivel (Löffler-Mang and Joss, 2000) deployed over EPFL campus, Lausanne, Switzerland. The entire network covers an area of approximatively 1 km^2 (Jaffrain et al., 2011). The distances between the stations are between 80 m and 800 m. The sampling temporal resolution is 30 s. The uncertainty on the collected measurements has been quantified and extensively documented in Jaffrain and Berne (2011). For validation purposes, the disdrometer data are divided into two groups: 8 disdrometers are used for the parameterization and the remaining 8 for the validation.

The radar data are provided by MeteoSwiss. The complete rain rate map is of size 610 $\times$ 538 km^2 with 1 $\times$ 1 km^2 resolution. The temporal resolution is 5 min. The rain rates are estimated by combining the measurements of three C-band weather radars at different elevations, correcting for the main sources of errors (ground clutter, beam shielding, vertical variability) according to the procedure described in Germann et al.

(2006). The estimated rain rates are then coded using 16 (irregularly spaced) intensity classes from 0-100 mm h^{-1}.

3.6.2 Simulation domain

A simulation domain of size 50 × 50 km^2 covering the city of Lausanne and the disdrometer network described in Section 3.6.1 is considered. The domain mostly lies in the so-called "Swiss Plateau", between the "Jura mountain" and the Swiss Alps. The minimum and maximum altitudes are 370 m and 2010 m AMSL but most of the domain (75 %) is lower than 800 m AMSL and only 10 % is higher than 1100 m AMSL. The simulation domain mostly avoids regions like the mountains of "Savoie" in the South, the Swiss Alps in the East and the "Jura mountains" in the North West where radar data are known to be noisy and affected by larger uncertainties.

3.6.3 Considered events

Two events that occurred on 17th June and 5th August 2010 are selected. The first event is classified (based on visual inspection of disdrometer and operational radar data) as frontal, the second is convective. The radar rain-rate maps corresponding to these events are shown in Figure 3.2 at 00:50 UTC respectively 13:55 UTC. The main characteristics of the selected events are given in Table 3.1. For simulation purposes, only the time periods during which the events passed over the disdrometer network are considered, i.e., from 00:50 UTC to 03:45 UTC for the first event and from 13:55 UTC to 14:55 UTC for the second.

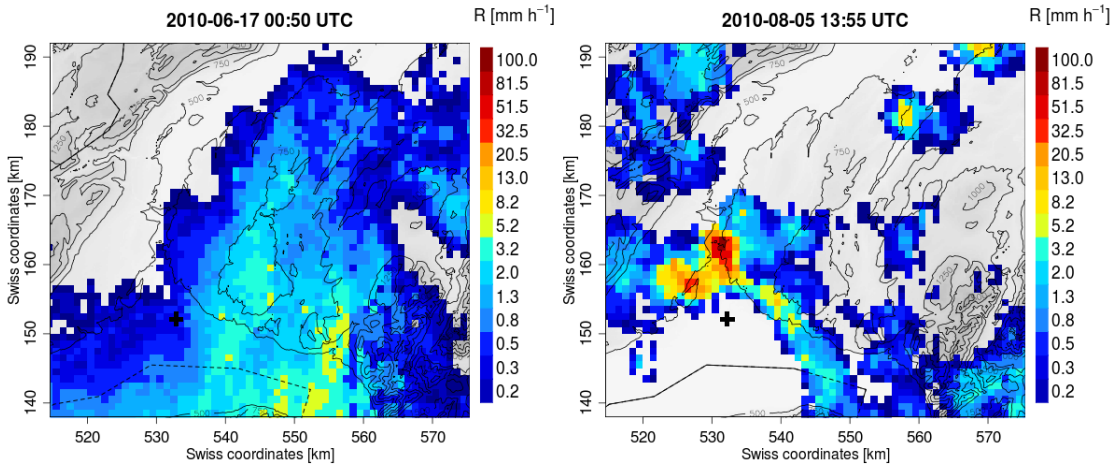

Figure 3.2: Radar rain-rate maps for event 1 (left panel) and event 2 (right panel). The size of the domain is 50 × 50 km^2. The location of the 16 disdrometers is represented by a black cross in the lower left corner.

The time series of μ, N_t and R corresponding to these time periods are shown in

Table 3.1: Rainfall type, intermittency, advection and anisotropy of each selected event.

	Event 1	Event 2
type	frontal	convective
intermittency (p_D)	21 %	70 %
direction of advection	to the North-West	to the South-East
speed of advection	9.5 $\mathrm{m\,s^{-1}}$	12.5 $\mathrm{m\,s^{-1}}$
anisotropy direction a	North-West	North-West
anisotropy ratio α	0.95	0.90

Figure 3.3 and Figure 3.4. Note that Λ is not shown because it is strongly correlated with μ (correlation coefficient of 0.96) and is therefore expressed as a linear model of μ in the simulations. It is also important to point out the differences between the two events (represented on different scales for better readability). The frontal event is characterized by small to moderate values of N_t, moderate to large values of μ (small drops) and limited temporal variability. The convective event on the other hand exhibits a lot of temporal variability, large values of N_t (lots of drops) and small values of μ (large drops).

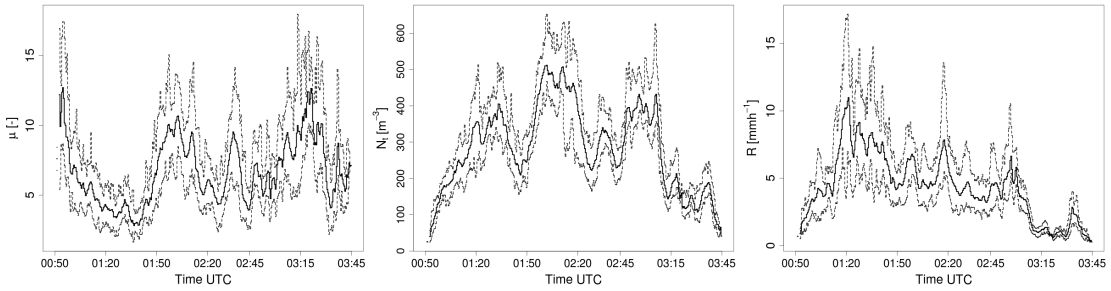

Figure 3.3: Time series of μ [-], N_t [$\mathrm{m^{-3}}$] and R [$\mathrm{mm\,h^{-1}}$] at 30 s time resolution for event 1 (frontal). The 3 lines represent the minimum, average and maximum values measured by the 16 disdrometers at each time step.

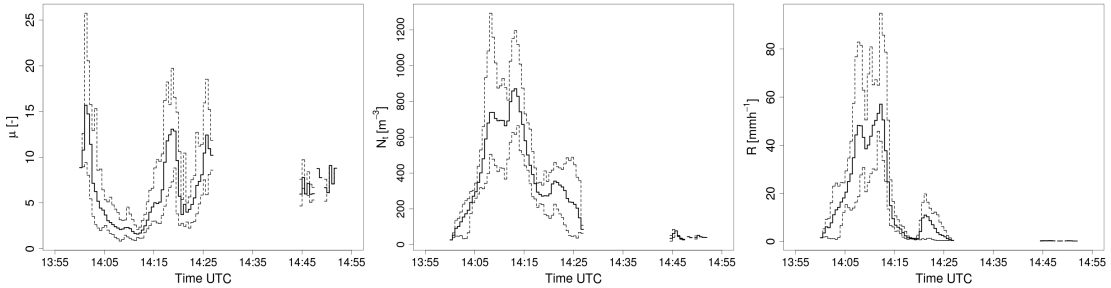

Figure 3.4: Time series of μ, N_t and R for event 2 (convective). Same format as Figure 3.3.

Table 3.2: Fitted values of nugget, partial sill and range for the rainfall indicator variograms

Type	Event	Nugget	Partial sill	Range
spatial variogram	1	0.005	0.16	26 km
	2	0.001	0.21	20 km
temporal variogram	1	0.003	0.16	40 min
	2	0	0.20	25 min

3.6.4 Parameterization

For each selected event, the spatial and temporal sample variograms of the rainfall indicator field are computed using the radar rain-rate data (see Section 3.3.2). Each sample variogram is fitted using a spherical variogram model:

$$\gamma(h) = \begin{cases} n + s\left(\frac{3h}{2r} - \frac{h^3}{2r^3}\right) & \text{for h} < \text{r} \\ n + s & \text{for h} \geq \text{r} \end{cases} \tag{3.12}$$

where n is called the nugget, s the (partial) sill and r the range. The fitted values of nugget, sill and range for each variogram are given in Table 3.2.

Next, the time series of $\tilde{\mu}$ and $\tilde{N}_t$ are used to compute the temporal variograms of the DSD parameters. A sum of two spherical variograms, one for the small-scale variability and one for the large-scale variability, is fitted to each sample variogram (see Figure 3.5). The fitted values of nugget, sills and ranges are given in Table 3.3. The temporal variograms are then used to define the complete space-time variograms of $\tilde{\mu}$ and $\tilde{N}_t$ as described in Section 3.3.4.

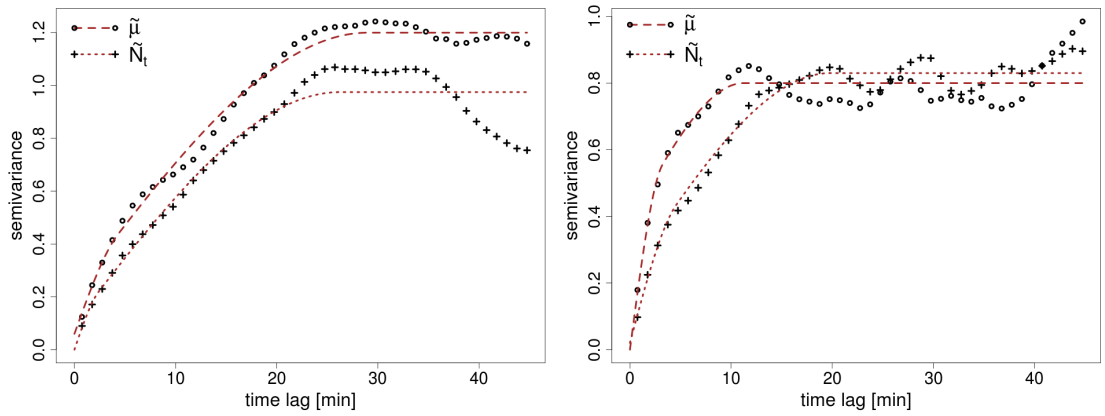

Figure 3.5: Fitted temporal variograms of $\tilde{\mu}$ and $\tilde{N}_t$ for event 1 (left panel) and event 2 (right panel). For details about the variogram parameters, see Table 3.3.

Table 3.3: Fitted values of nugget, partial sills and ranges of the temporal variograms of $\tilde{\mu}$ and $\tilde{N}_t$.

Parameter	Event	Nugget	Partial sills	Ranges [min]
$\tilde{\mu}$	1	0.06	0.16 and 0.98	4.3 and 29
	2	0	0.38 and 0.42	3.3 and 11.3
$\tilde{N}_t$	1	0	0.1 and 0.88	2.4 and 26.3
	2	0.02	0.21 and 0.6	4.6 and 19.8

3.6.5 Simulated DSD fields

For illustration purposes, two 1-h simulations with a temporal resolution of 1 min and a spatial resolution of 500 × 500 m^2 have been generated. This turned out to be a good trade-off between the required simulation time (a few minutes on a standard desktop computer) and the ability to reproduce most of the space-time dynamics occurring in rainfall and DSD at these scales. All the simulations were performed using the statistical computing software "R" (R Development Core Team, 2011) and the "gstat" package by Pebesma (2004).

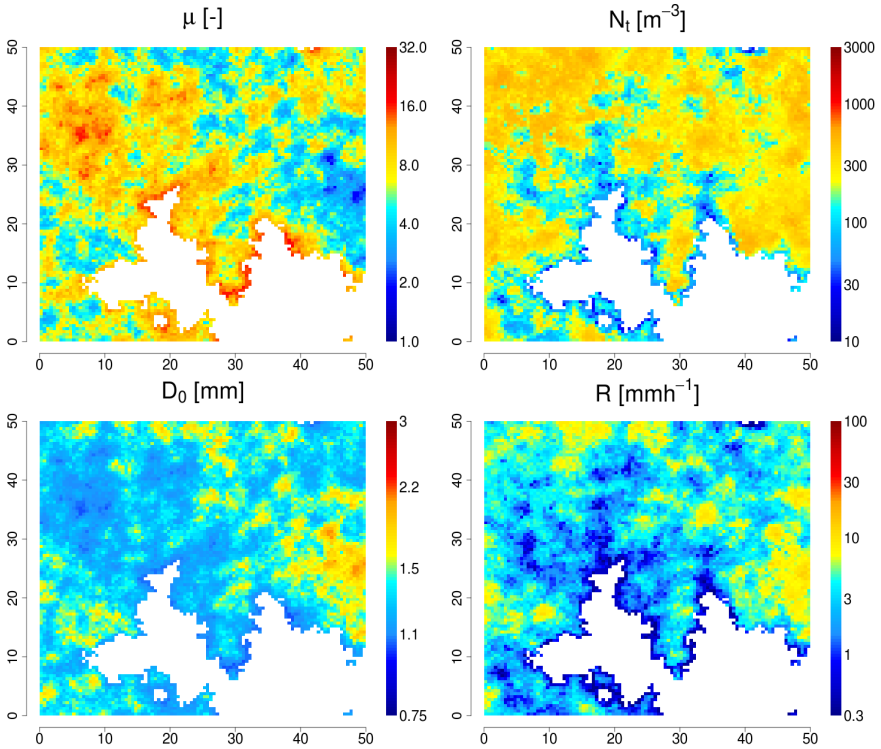

Figure 3.6: Snapshot of a simulated field of μ, N_t, D_0 and R for the frontal event. The simulation size is 50 × 50 km^2 and corresponds to the domain shown in Figure 3.2. The pixel resolution is 500 × 500 m^2.

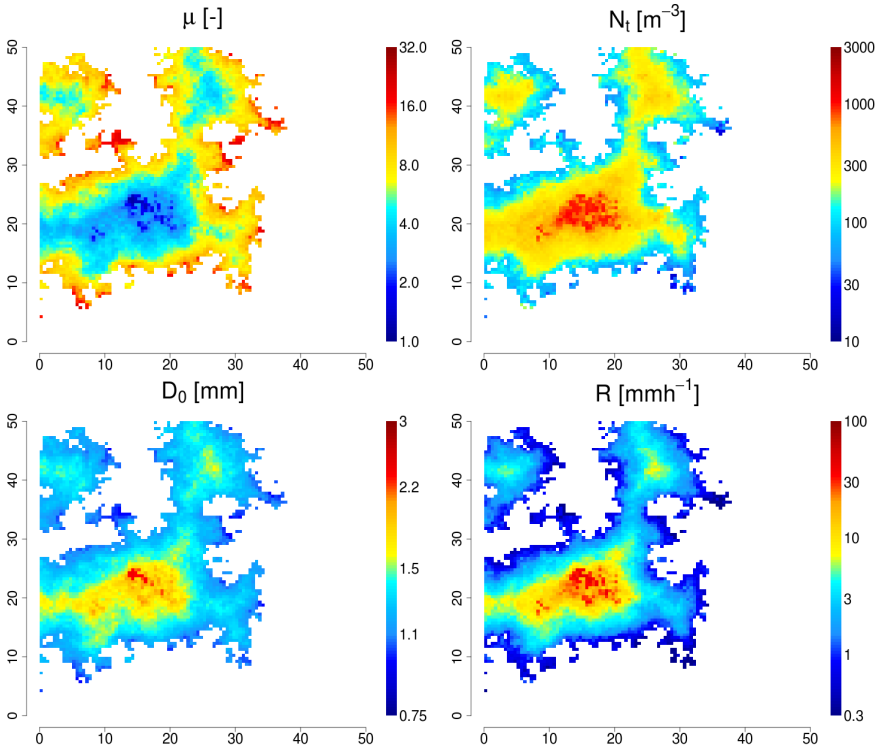

Figure 3.7: Snapshot of a simulated field of μ, N_t, D_0 and R for the convective event. Same format as in Figure 3.6.

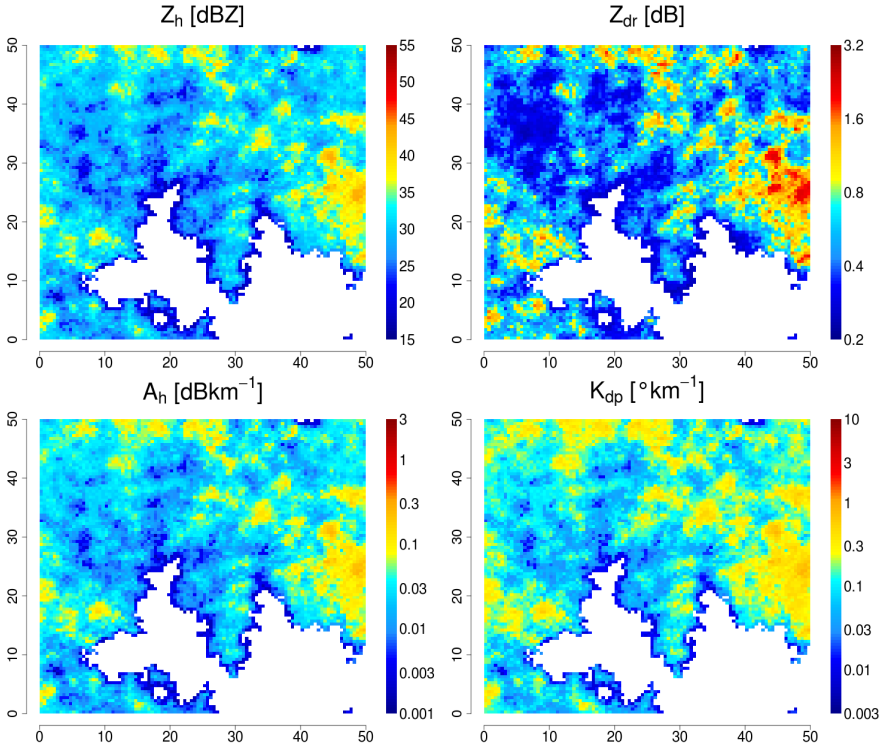

Figure 3.8: Fields of Z_h, Z_{dr}, A_h and K_{dp} corresponding to Figure 3.6.

Snapshots illustrating the simulated fields at a given time are provided in Figure 3.6 and Figure 3.7. The associated radar bulk variables Z_h, Z_{dr}, A_h and K_{dp} at X-band (9.41 GHz) and 20 ° C are shown in Figure 3.8 and Figure 3.9.

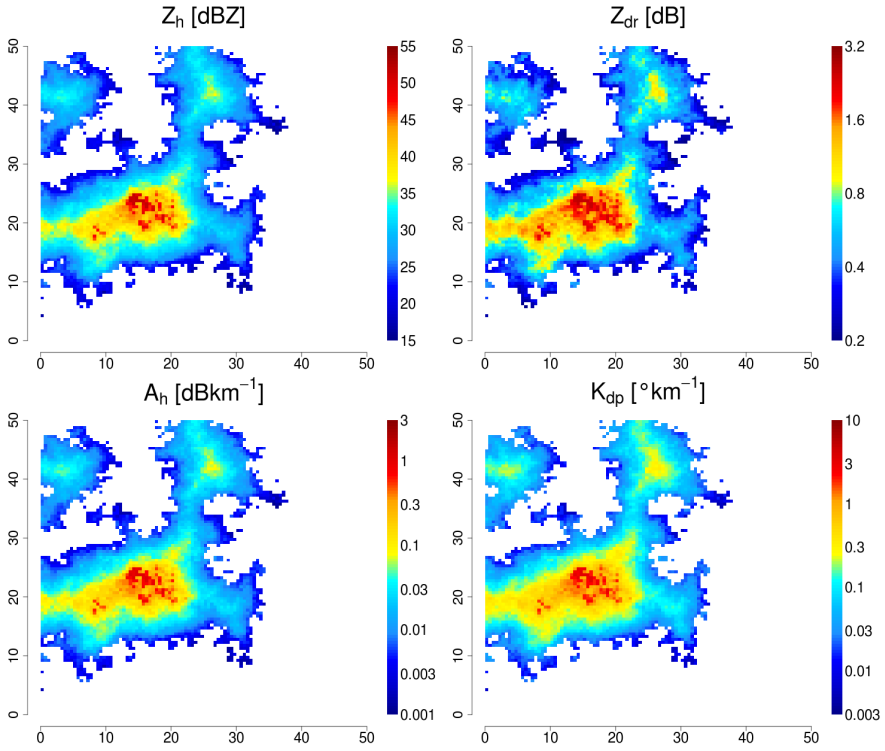

Figure 3.9: Fields of Z_h, Z_{dr}, A_h and K_{dp} corresponding to Figure 3.7.

3.6.6 Evaluation

The quality and realism of the simulated DSD fields are difficult to evaluate quantitatively. It is, for example, not possible to compare the simulated fields with observed radar data because the latter were used to parameterize the simulator. An "indirect" evaluation of the simulated fields based on the measurements of the 8 remaining disdrometers discarded during the parameterization is nevertheless proposed.

The idea is to compare the statistical properties and the temporal structure of the simulated and observed time series of μ and N_t. For each location x in the simulation domain, the simulated values of μ and N_t (at this particular location) are extracted. This gives, for each location, a 1-h time series of simulated DSD parameters. For each location, the mean $\bar{\mu}$ and $\bar{N}_t$ of the extracted time series is computed. To validate the simulations, the distributions of $\bar{\mu}$ and $\bar{N}_t$ are compared to their equivalents obtained using the time series from the 8 "control" disdrometers (see Figure 3.10). One can see that the simulated and observed values of $\bar{\mu}$ and $\bar{N}_t$ are in good agreement but that the simulated mean values have a much larger dispersion than the control values. This

can however be explained by the fact that there are far more simulated time series (one for each pixel in the simulation domain) than measured time series (only 8 control disdrometers).

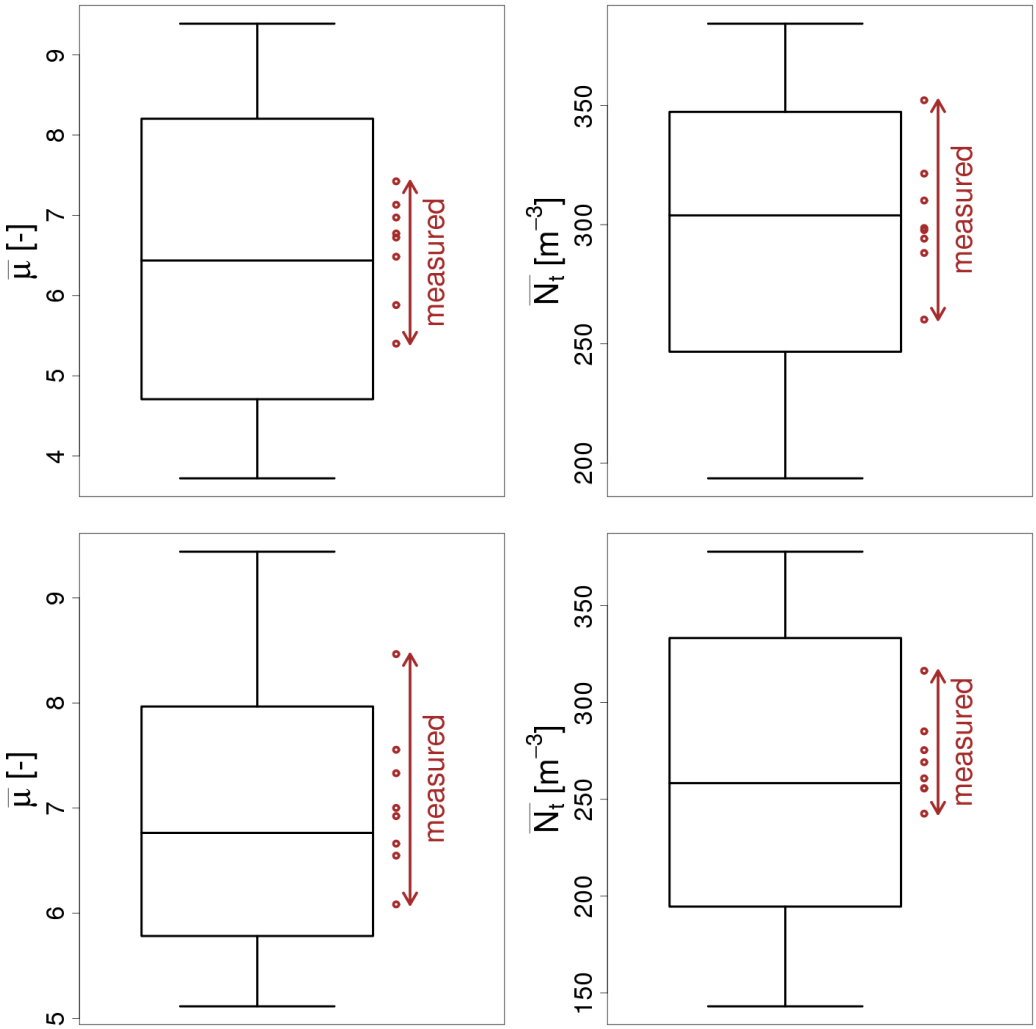

Figure 3.10: Simulated distributions of $\bar{\mu}$ (left) and $\bar{N}_t$ (right) for event 1 (top) and event 2 (bottom). Each box plot represents the 10 %, 25 %, 50 %, 75 % and 90 % quantiles of all simulated values. The arrow next to each box plot represents the range of $\bar{\mu}$ and $\bar{N}_t$ for the 8 control disdrometers.

In order to evaluate the simulator's capabilities to reproduce correct temporal structures, the variograms of the simulated time series of $\tilde{\mu}$ and $\tilde{N}_t$ are compared to the variograms obtained from the 8 control disdrometers (see Figure 3.11). Again, the simulated variograms exhibit a larger dispersion than the control variograms but the overall agreement between the variograms is good. This suggests that the simulated fields exhibit realistic temporal structures that are consistent with the control disdrometers. Finally, it is also important to point out that both events (frontal and convective) were

statistically well reproduced by the simulator despite their different nature.

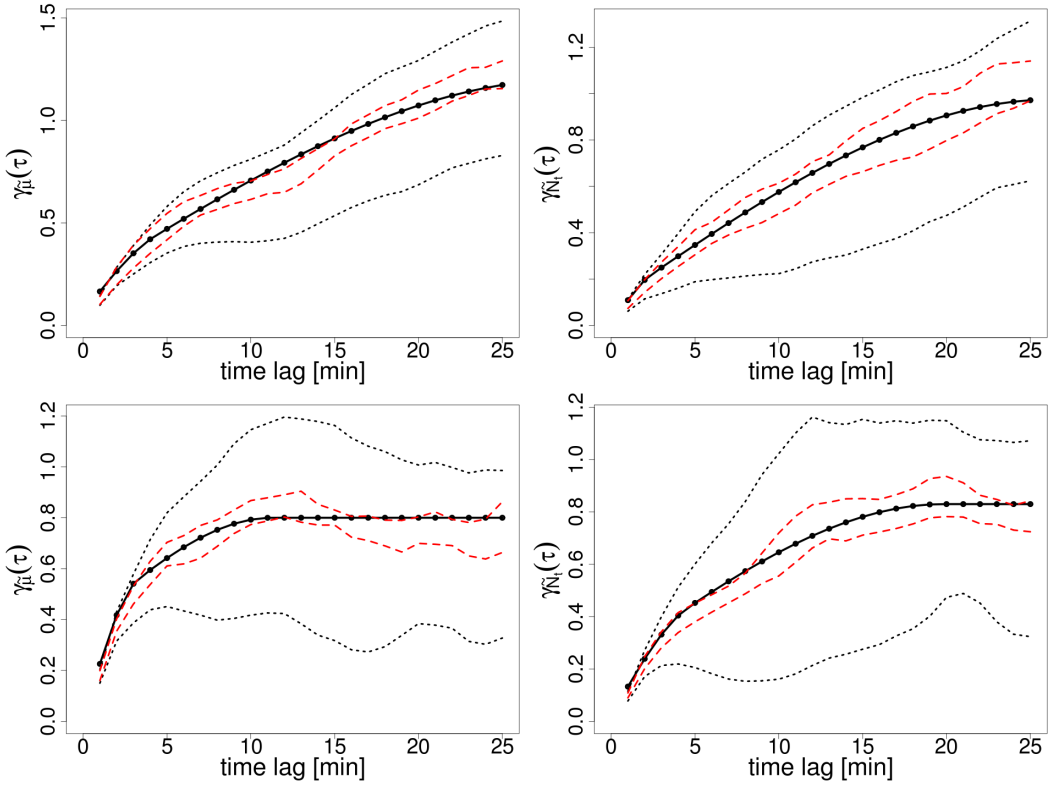

Figure 3.11: Simulated and measured temporal variograms of $\tilde{\mu}$ (left) and $\tilde{N}_t$ (right) for event 1 (top) and event 2 (bottom). The black dotted lines represent the 10 % and 90 % quantiles of the simulated variograms for each pixel in the simulation domain. The continuous black line represents the median of all simulated variograms. The dashed red lines represent the minimum and the maximum of the 8 variograms computed using the control disdrometers.

3.7 Discussion

Clearly, the proposed DSD simulator offers several advantages. It is relatively easy to parameterize and only requires DSD time series (from a single disdrometer) and radar rain-rate or reflectivity data. Once all the parameters have been estimated, hundreds of simulations can be performed in a reasonable time on a single desktop computer. There are, however, some limitations which have to be mentioned.

The size of the simulation domain is limited by the fact that the variograms of $\tilde{\mu}$ and $\tilde{N}_t$ must be representative of the variability of the DSD over the entire domain. Areas with different variabilities (due for example to orographic effects) should be separated and treated individually. As a rule of thumb, a reasonable simulation domain for convective events should not exceed 50×50 km^2 and 100×100 km^2 for stratiform events. Another

factor that strongly limits the size of the simulation domain is the particular form of the space-time structure assumed in Equation 3.10 which essentially states that the advection field must be constant over the entire domain. Local changes in advection within the simulation domain (for example due to mountains or a coastal line) cannot be reproduced correctly. The only possible generalization is to consider an average advection field that changes over periods of 20-30 min and to simulate each time block separately.

The presented simulation technique only considers 2-dimensional fields of DSD at the ground level (+ an additional temporal dimension). No assumptions with respect to the vertical variability and structure of the DSD are made. In theory, 3-dimensional fields of DSD could also be simulated using the technique presented in this chapter. In practice, however, this turns out to be far more difficult because the horizontal and vertical dimensions exhibit completely different dynamics and structures. Nevertheless, if adequate parameterizations of the vertical variability (i.e., variogram) of μ and N_t can be provided, an extension to 3-dimensional fields of DSD (using the generated 2D + time fields as a starting point) is possible. This is particularly interesting for applications in which the vertical properties of the DSD in the atmosphere are more important than on the surface (e.g., for space-borne weather radars and telecommunication applications).

Finally, it must be noted that there is currently no direct way of controlling complex physical processes like the growth, the decay, the merging and the splitting of individual rain cells in the simulations. Individual rain cells may emerge or split up during the simulations by "chance" (i.e., without any particular forcing nor additional parameterization) but this is essentially due to "side effects" (e.g., a rain cell that is advected outside the simulation domain causes another cell to appear elsewhere) and cannot be related to any microphysical processes or atmospheric dynamics. Introducing such a forcing into the simulations is difficult and far beyond the scope of this study. In the simplified case with just one rain cell, additional external drifts on I, $\tilde{\mu}$ and $\tilde{N}_t$ (similar to the one presented in the Appendix), could be used to introduce such dynamics into the simulations. For example, a temporal drift on I can be used to control the percentage of dry and wet locations in the simulation domain and therefore the size of the considered cell with respect to its lifetime. Similarly, a temporal drift on $\tilde{\mu}$ can be used to control the increase or decrease of the average drop diameter within the cell. Finally, a temporal drift on $\tilde{N}_t$ allows to control the average drop concentration with respect to the lifetime of the cell. It is however not clear how this technique could be extended to the case of multiple interacting rain cells with different lifetimes and dynamics.

3.8 Conclusions

Precipitation is an essential part of the hydrological, atmospheric and climatic system and must therefore be monitored and measured with great accuracy. Yet the large spatial and temporal variability of precipitation appears to be an inevitable source of uncertainty

in many practical applications. Modern hydrological and atmospheric models require accurate rainfall estimates with high spatial and temporal resolutions, sometimes far beyond the capabilities of current rainfall sensors. In these cases, a stochastic rainfall simulator can be a very valuable tool because it provides large amounts of reference data at high spatial and temporal resolutions. These reference data can be used to investigate various aspects and algorithms in the models. For example, they can be used to evaluate the performance of new algorithms for the merging of data collected using different sensors and techniques (e.g., weather radars, rain gauges, disdrometers and microwave links). Other possible applications include the study of the $Z - R$ relationship for weather radars, attenuation correction algorithms at C- and X-band, downscaling of precipitation and reflectivity fields, soil erosion problems, atmospheric deposition and signal attenuation of telecommunication microwave links.

In this chapter, a new method for the stochastic simulation of intermittent DSD fields in time has been presented. The method generalizes previous work done by Barancourt et al. (1992) and Schleiss et al. (2009) presented in Chapter 2 and uses a combination of sequential indicator simulation and sequential Gaussian simulation. A multivariate Gaussian anamorphosis is used prior to the simulation in order to transform the original DSD parameters into independent normalized Gaussian variables. External drifts are used to ensure the continuity between dry and rainy locations. In Section 3.6, the simulator's ability to reproduce complex and different rainfall patterns using real data collected in Lausanne, Switzerland has been illustrated. An indirect evaluation of the simulations suggests that the simulated DSD fields exhibit realistic space-time structures that are in good agreement with independent DSD measurements.

The main advantage of the presented simulator is its simplicity and its short computation time. The main limitation is the inability to reproduce the interactions (growth, decay, merging and splitting) of multiple rain cells. Future work will mainly focus on practical applications of the DSD simulator to problems related to the downscaling of DSD fields and to the statistical representativity of radar and satellite measurements.

Appendix: detrending the DSD time series

Recall that the variogram is only defined for intrinsic random functions (see Chilès and Delfiner (1999)). Time series of μ and N_t usually do not satisfy this condition because $\mathrm{E}[\mu(t)]$ and $\mathrm{E}[N_t(t)]$ depend (non-linearly) on t. Typically, the beginning and the end of each event is characterized, on average, by lower rain rates. Therefore, the expected rain rate drifts towards zero as one approaches the next dry period. Similar continuity conditions also affect each of the DSD parameters. The characteristics of this drift are event dependent and must be estimated and removed before computing the sample variograms. Finding appropriate methods to estimate these drifts can be difficult and should be done with great care. In the following, a possible method for detrending DSD time series that takes into account the proximity (in time) to the nearest dry period is proposed. First, a logarithmic transform is applied to $\mu > -1$ and $N_t > 0$:

$$\mu \mapsto \log_{10}(\mu + 1) \tag{3.13}$$

$$N_t \mapsto \log_{10}(N_t). \tag{3.14}$$

The distribution of the log-transformed parameters is usually more symmetric and hence better suited for statistical analysis. For a measured DSD time series $(\mu,\ N_t,\ I)(t_i)$, the sets $D = \{i \,|\, I(t_i) = 0\}$ and $W = \{\, i|\, I(t_i) = 1\}$ of all dry and wet time periods are identified. For all $i \in W$, the distance $\tau_i = \min_{j \in D}\{|t_i - t_j|\}$ to the nearest dry period is computed. Finally, the drifts $m_\mu(\tau)$ and $m_{N_t}(\tau)$ are estimated for each (transformed) DSD parameter:

$$m_\mu(\tau) = \frac{1}{N(\tau)} \sum_{\tau_i \sim \tau} \mu(t_i) \tag{3.15}$$

$$m_{N_t}(\tau) = \frac{1}{N(\tau)} \sum_{\tau_i \sim \tau} N_t(t_i) \tag{3.16}$$

where $N(\tau)$ represents the number of observations at an (approximate) time lag τ from the nearest dry period. A (simple) theoretical drift model is fitted to the sample estimates and used to detrend the time series of μ and N_t (in the logarithmic space):

$$\mu(t_i) \mapsto \mu(t_i) - m_\mu(\tau_i) \tag{3.17}$$

$$N_t(t_i) \mapsto N_t(t_i) - m_{N_t}(\tau_i) \tag{3.18}$$

An example of this technique is shown in Figure 3.12 where the estimated drift functions $m_\mu(\tau)$ and $m_\mu(\tau)$ are shown for the convective rainfall event presented in Section 3.6. Note that $m_{N_t}(\tau)$ usually increases with τ whereas for $m_\mu(\tau)$ it is the opposite.

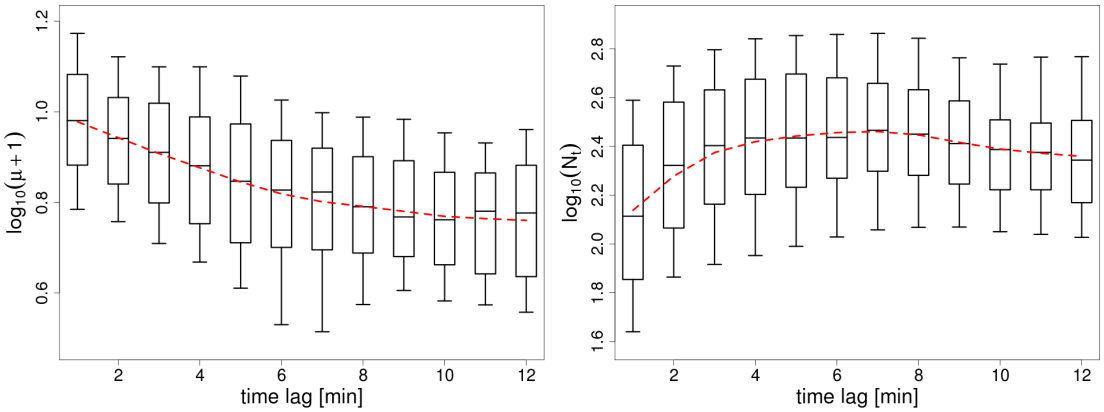

Figure 3.12: Estimated drift functions for $\log_{10}(\mu + 1)$ and $\log_{10}(N_t)$ for the convective event.

After removing the drift, a Gaussian anamorphosis is applied to the DSD time series. All the simulations are performed in the Gaussian space. At the end, the simulated variables are back-transformed using an inverse anamorphosis and the drift (depending on the distance to the nearest simulated dry location or period) is added. Finally, the inverse transform corresponding to (3.13) and (3.14) is applied to retrieve the parameters in the original parameter space. The logarithmic transform ensures that only valid values of $\mu > -1$, and $N_t > 0$ are produced after adding the drifts to the simulations.

4 Stochastic space-time disaggregation of rainfall into DSD fields [1]

4.1 Summary

In this chapter, a stochastic method to disaggregate rainfall fields into DSD fields is proposed. It is based on the stochastic DSD simulator presented in Chapter 3 that has been modified in order to take into account prescribed mean values of rain rates provided at the coarser scale. The proposed method is illustrated and validated using radar rain-rate data provided by MeteoSwiss for two rain events of contrasted type. The space-time structure of the DSD fields at the fine scale is parameterized using data collected with a network of 16 disdrometers. The rainfall intermittency is modeled using climatological relationships derived from radar data at different spatial and temporal scales. The results show that the two types of rain events can be correctly disaggregated, although the general agreement in terms of rain rate distributions, intermittency and space-time structures is much better for the stratiform case. At the end of the chapter, possible extensions and generalizations of the technique (e.g., using radar reflectivities at two different frequencies or polarizations to drive the disaggregation process) are discussed.

1. This chapter is a slightly modified version of the article by Schleiss, M. and A. Berne, 2012: Stochastic space-time disaggregation of rainfall into DSD fields, submitted to *J. Hydrometeorol.*, on the 17th of January 2012.

4.2 Introduction

Meaningful investigations of many hydrological processes like flash floods, landslides and urban water management, require accurate rain rate estimates at high spatial and temporal resolutions (typically in the order of 1 km and 5 min and higher). Given the impracticality of deploying extended rain gauge networks at the ground, one often has to rely on remotely sensed rainfall estimates provided by passive and active sensors like weather radars. The major problem with this approach is that the spatial and temporal resolutions at which theses measurements are provided may not be relevant for hydrological and meteorological applications. A common solution to this problem, called rainfall disaggregation, consists in downscaling the initial measurements prior to their use as input into the models. This requires to make certain assumptions about the physics and the statistical properties of the considered rainfall fields. As a consequence, disaggregated rainfall fields should not be taken as a faithful representation of the "true" rainfall field, but as a particular realization of a stochastic process satisfying a set of prescribed physical and statistical properties. A distinctive feature of this approach is that each disaggregated rainfall field looks different at the fine scale but cannot be distinguished from the others at the coarser scale.

So far, many different disaggregation techniques of various degrees of complexity have been proposed in the literature. Among the few deterministic disaggregation methods, one can mention the linear model by Ormsbee (1986) for time series disaggregation and the chaotic dynamical approach by Sivakumar et al. (2001). Examples of spatial deterministic disaggregation techniques are given by the Haar wavelet transform proposed by Perica and Foufoula-Georgiou (1996) and the QQ transforms by Bárdossy and Pegram (2011). Stochastic disaggregation techniques on the other hand are much more common. Most of them focus on the temporal disaggregation of rain rate time series using stochastic point processes (e.g., Hershenhorn and Woolhiser, 1987; Bo et al., 1994; Cowpertwait et al., 1996; Koutsoyiannis and Onof, 2001) or (multi)fractal random cascades (Olssen and Berndtsson, 1998; Guntner et al., 2001). Different approaches based on simulated annealing (Bárdossy, 1998) and master target scaling techniques (Jennings et al., 2010) are also worth mentioning. A lot of efforts have also been devoted to develop stochastic spatial and space-time disaggregation methods, like the morphological algorithms by Kumar and Foufoula-Georgiou (1994) and Mackay et al. (2001) that have the advantage of explicitly taking into account the rainfall intermittency during the downscaling procedure. Other examples include the Gibbs sampler by Onibon and Lebel (2004), filtered autoregressive models (also called meta-Gaussian models) by Rebora et al. (2006b) and spectral multi-scaling by Pavlopoulos (2011).

A common characteristic of all the above mentioned disaggregation techniques is that they all focus on the disaggregation of rain rates. So far, no attempts have been made to include the (rain)drop size distribution (DSD) into the disaggregation process. Yet generating DSD fields is potentially more interesting and general than simulating rain

rate fields because most of the important quantities in remote sensing of precipitation (like the rain rate, the median drop diameter, the radar reflectivity, the differential reflectivity, specific attenuation and differential phase shift) explicitly depend on the DSD and its weighted moments. Such disaggregated DSD fields would constitute a very valuable input for many practical applications in hydrometeorology. They can, for example be used to investigate issues related to soil detachment, landslides and erosion problems (Rosewell, 1986; Finlay et al., 1997; Kinnell, 2005), flash floods (Rebora et al., 2006a), urban water management (Onof and Arnbjerg-Nielsen, 2009), rainfall estimation using telecommunication microwave link networks (Messer et al., 2006; Leijnse et al., 2007c) and remote sensing of precipitation using weather radars (e.g., Bringi and Chandrasekar, 2001; Krajewski and Smith, 2002), satellites like TRMM (Simpson et al., 1988; Iguchi et al., 2000) and the upcoming global precipitation measurement mission GPM (http://gpm.nasa.gov).

In this chapter, a stochastic disaggregation technique that allows to generate realistic space-time intermittent DSD fields with given space-time integral rain rates (or any other variable related to the DSD) is described. It is based on Geostatistics and uses climatological variograms to model the space-time structure of the DSD parameters. Sequential indicator and sequential Gaussian simulation methods are used iteratively until the generated DSD fields satisfy all the prescribed conditions at the coarse scale. The method is illustrated and evaluated qualitatively using two different rainfall events and MeteoSwiss radar rain-rate data.

This chapter is structured as follows: first, the models and methods needed for the disaggregation are described in Section 4.3. The disaggregation technique is then detailed in Section 4.4 and illustrated in Section 4.5. The conclusions are given in Section 4.6.

4.3 Models

This section describes the models used to represent the DSD and the rainfall intermittency in the disaggregation process. Only the most important and relevant parts are explained. For more details about the models and the stochastic simulation method used to generate the DSD fields, see Chapter 2 and Chapter 3.

4.3.1 DSD model

The DSD is assumed to be adequately represented by a Gamma distribution (Ulbrich, 1983; Willis, 1984):

$$N(D) = \alpha N_t D^{\mu} e^{-\Lambda D}, \tag{4.1}$$

where $N(D)dD$ [m^{-3}] denotes the number of drops per unit volume with diameters between D [mm] and $D + dD$ and $\alpha = (\int_{D_{min}}^{D_{max}} D^{\mu} e^{-\Lambda D}\, dD)^{-1}$ is a normalization factor taking into account the finite range of possible drop sizes between D_{min} and D_{max}. The Gamma DSD model has three parameters: the shape $\mu > -1$ [-], the rate $\Lambda > 0$ [mm^{-1}] and the concentration $N_t > 0$ [m^{-3}]. Because of the natural variability of rainfall, (μ, Λ, N_t) can be seen as realizations (in space and time) of an underlying multivariate random function. Simulating DSD fields is therefore equivalent to simulating a (tri-variate) random field $(\mu, \Lambda, N_t)(x, t)$ at each point x in space and t in time. Note that other parameterizations corresponding for example to normalized DSDs (Bringi and Chandrasekar, 2001) can also be considered.

Simulating DSD fields is more interesting and general than simulating rain rate fields because the rain rate can be derived from the DSD, through direct numerical integration:

$$R = \frac{6\pi}{10^4} \int_D D^3 v(D) N(D) dD \tag{4.2}$$

where R [$mm\,h^{-1}$] denotes the rain rate and $v(D)$ [$m\,s^{-1}$] the terminal fall speed of a drop of diameter D (e.g., Beard, 1977). Similar expressions exist for many other quantities of interest like the median drop diameter D_0 [mm], the radar reflectivity at horizontal or vertical polarization $Z_{h|v}$ [dBZ], the differential reflectivity $Z_{dr} = Z_h - Z_v$ [dB] and the specific attenuation on propagation $A_{h|v}$ [$dB\,km^{-1}$]. Consequently, any quantity that is directly related to the DSD can be used as a driving variable in the disaggregation process. More generally, any combination of moments or quantities related to the DSD (e.g., the radar reflectivities at two different frequencies or polarizations) can be used to further refine the disaggregation process (see Section 4.5.4 for more details). For illustration, only the rain rate will be considered.

It is worth mentioning that several studies on the Gamma DSD mention the existence of deterministic relationships between the shape parameter μ and the rate parameter Λ in Equation (4.1) (e.g., Zhang et al., 2001; Seifert, 2005). Although they are not necessary in the disaggregation process, these relationships are very interesting because they allow to reduce the number of stochastic parameters to be simulated and therefore significantly increase the computational efficiency. For more conciseness, the disaggregation method will only be explained for two stochastic parameters (μ, N_t) instead of three, therefore assuming a deterministic relationship between μ and Λ. The extension of the method to 3 stochastic parameters is straightforward and does not change the general idea behind the approach.

4.3.2 Gaussian anamorphosis of DSD parameters

Simulating a bivariate random field $(\mu, N_t)(x,t)$ with given space-time structure is very difficult in general. There are, however, very efficient simulation algorithms for the case where (μ, N_t) are jointly Gaussian with known correlation structure (Journel and Huijbregts, 1978; Ripley, 1987; Guillot, 1999). A common approach in stochastic simulation therefore consists in transforming the original parameters (μ,N_t) into jointly Gaussian variables $(\tilde{\mu},\tilde{N}_t)$ prior to their simulation, a procedure also known as "Gaussian anamorphosis". At the end of the simulation, the values of $(\tilde{\mu},\tilde{N}_t)$ are back-transformed into the original parameter space using the inverse transform. A particularly interesting Gaussian anamorphosis is the so-called stepwise conditional Gaussian transformation technique (Leuangthong and Deutsch, 2003) that ensures that the transformed parameters are independent Gaussian variables with zero mean and unit variance. Because they are independent, the parameters can be simulated separately. Cross-correlations and higher order moments between the variables are then re-introduced after the simulation, during the back-transformation.

4.3.3 Intermittency

Rainfall intermittency, i.e., the presence or absence of rainfall, is described by an indicator function

$$I(x,t) = \begin{cases} 1 \text{ if } R(x,t) > 0 \\ 0 \text{ else} \end{cases} \tag{4.3}$$

where $R(x,t)$ [$\mathrm{mm\,h^{-1}}$] represents the instantaneous rain rate at location $x \in \mathbb{R}^2$ and time $t \in \mathbb{R}$. By extension, we define

$$I(X,T) = \begin{cases} 1 \text{ if } \exists\,(x,t) \in (X,T) \text{ with } I(x,t) = 1 \\ 0 \text{ else} \end{cases} \tag{4.4}$$

where X is a given domain in $\mathbb{R}^2$ and T a given time period. For simplicity, only square areas of size $k \times k$ [$\mathrm{km^2}$] and continuous time periods $[t, t+\tau]$ of duration τ [h] are considered.

The problem with rainfall intermittency is that it strongly depends on the scale at which it is considered. This dependency must be taken into account during the disaggregation process. More specifically, the amount of dry regions and locations must increase when going from the coarse scale to the fine scale. Analytical relationships describing the average amount of intermittency with respect to the considered spatial and temporal scales can be found in Kundu and Siddani (2011) and Chapter 5. These relationships can be used to estimate the percentage p [-] of rainy pixels at the fine scale (k,τ) as a

function of the observed percentage of rainy blocks P [-] at the coarse scale (K, T). This is a difficult problem that has not been addressed in detail in the literature so far. A possible approach is to estimate p using the climatological relationships between rainfall intermittency at different scales given in Chapter 5:

$$p = P + p(k, \tau) - p(K, T) \tag{4.5}$$

where

$$p(k, \tau) = (p_0 + p_1 k) \exp\left\{ - \left[\frac{\tau}{a k^b} \right]^\beta \right\} \tag{4.6}$$

for some given parameter values p_0, p_1, a, b and β provided in Table 5.1. For example, a simple numerical application of Equation (4.5) shows that the percentage of rainy periods and locations decreases by about 5 percent points on average when disaggregating from 5×5 km^2 and 15 min to 1×1 km^2 and 5 min. In other words, if the observed percentage of rainy periods and locations at 5×5 km^2 and 15 min is 95 %, the expected value at 1×1 km^2 and 5 min resolution will be 90 %. However, because these are only climatological relationships, the "real" value of rainfall intermittency for a given event can be significantly different from its expectation, as shown in Section 4.5. Consequently, it is suggested to generate many, different realizations with different values of p within a given confidence interval (e.g., 5-10 %) around the expected climatological value.

4.3.4 Space-time structure

The space-time structures of I, $\tilde{\mu}$, and $\tilde{N}_t$ are modeled using variograms (Matheron, 1965), a key tool in the field of Geostatistics (Chilès and Delfiner, 1999). The space-time variogram of a random field Z is defined as

$$\gamma_Z(h, \tau) = \frac{1}{2} \mathrm{E}\left[(Z(x + h, t + \tau) - Z(x, t))^2 \right], \tag{4.7}$$

where $h \in \mathbb{R}^2$ is a spatial separation vector, $\tau \in \mathbb{R}$ is a given time lag and E denotes the expectation.

In theory, any valid space-time variogram can be used to model the space-time structure of I, μ, and N_t. In Section 4.5, different variogram models that have been fitted using a large amount of DSD data collected in Lausanne, Switzerland and are representative of the local climatology and rainfall types (e.g., stratiform vs convective) are proposed. The general space-time structure of these variograms is closely related to Taylor's hypothesis of frozen turbulence (Taylor, 1938), which states that the rainfall field moves, during short periods of time, with constant velocity v, and that its evolution for time lags up to 20-30 min is small compared to the advection process (Li et al., 2009). Hence the following variogram model can be used to represent the space-time structure of the

rainfall field:

$$\gamma_Z(h,\tau) = \gamma_Z^{(S)}(h - \tau v) + \gamma_Z^{(T)}(\tau) \tag{4.8}$$

where $\gamma_Z^{(S)}(h)$ and $\gamma_Z^{(T)}(\tau)$ are two valid variogram models representing the spatial and the temporal structure of Z. In other words, the space-time variogram of Z can be decomposed in spatial and temporal variations, linked by a speed and direction of advection v. Note that $\gamma_Z^{(S)}(h)$ can be isotropic (i.e., it only depends on the norm of h) or anisotropic (it depends on both the norm and the direction of h).

The fact that, in the absence of any external source of information the space-time variograms of I, μ, and N_t must be "guessed" or estimated from the observed structure at the coarse scale is an inherent property of any disaggregation method and should not be seen as a major limitation. Climatological variograms are useful but will not always provide optimal results, especially if the "true" variogram of the field is significantly different. In case the spatial and temporal structures are highly uncertain, it is always possible to generate different realizations of the same event using different space-time variogram models covering a large range of plausible structures.

4.3.5 Sequential Gaussian and indicator simulation

The simulation methods used to generate Gaussian and indicator fields with a given space-time structure are the sequential Gaussian simulation (SGS) and the sequential indicator simulation (SIS) algorithms (Ripley, 1987). These methods were chosen because they are intuitive and efficient, even for large simulation domains (through the use of local neighborhoods). Because SGS and SIS are very similar, the implementation details are only provided for SGS.

1. Start with a previously simulated or prescribed set of (Gaussian) values $z_1, ..., z_M$ at locations x_1,...,x_M. If $M = 0$, the simulation is said to be non-conditional.
2. Define the locations x_{M+1},...,x_{M+N} at which Z will be simulated.
3. Choose a random path Γ that visits each location $x_{M+1}, ..., x_{M+N}$ exactly once. For simplicity, we will assume that $\Gamma = (M+1, ..., M+N)$ (otherwise the remaining locations are re-arranged in the order given by Γ).
4. Go through Γ. At location i, estimate the conditional distribution F_i of $Z(x_i)$ given all the previously simulated or prescribed values. Generate a realization z_i according to F_i and include the outcome in the set of previously simulated values. In the Gaussian case, this is very easy because

$$Z(x_i) \mid Z(x_{i-1}) = z_{i-1}, ..., Z(x_1) = z_1 \ \sim \ N(z_i^\star, {\sigma_i^\star}^2) \tag{4.9}$$

   where $z_i^\star$ and ${\sigma_i^\star}^2$ represent respectively the simple kriging estimate at location i and the associated kriging variance (e.g., Chilès and Delfiner, 1999, p.164). Hence the

problem reduces to solving the kriging equations and drawing a value at random from a Gaussian distribution with known mean and variance.

5. Go to the next location and repeat the procedure for the remaining locations.

Note that, in its simplest formulation (given above), SGS is not computationally efficient because the size of the kriging system increases at each iteration. Consequently, the time needed to solve the kriging system and to derive $z_i^\star$ and $\sigma_i^{\star 2}$ rapidly becomes prohibitive. A common solution to this problem, called sequential Gaussian simulation with local neighborhood selection, is to exploit the screening effect (e.g., Chilès and Delfiner, 1999, p.202) and to consider only a fixed amount of nearest neighbors at each iteration step when solving the kriging system. In this way, the algorithm stays computationally efficient, even for large simulation domains. The size of the neighborhood is selected as a trade-off between computational cost and accuracy of the kriging estimates.

Sequential indicator simulation (SIS) is very similar, except that the considered variable Z follows a discrete probability distribution with two states (0/1). The major difference compared to the Gaussian case is that the conditional distribution F_i is now Bernoulli and that the simple kriging mean $z_i^\star$ is supposed to accurately represent the conditional expectation of $Z(x_i)$ (i.e., the probability of $Z(x_i)$ to be equal to 1). The estimated kriging variance is not needed anymore and does not need to be computed. Once the kriging system has been solved and $z_i^\star$ determined, the problem reduces to drawing a value at random from a Bernoulli distribution with known mean. The only theoretical problem with this method is that $z_i^\star$ can, in theory, be less than 0 or greater than 1. If this is the case, the corresponding probabilities are set to 0 or 1 accordingly and the outcome z_i is purely deterministic. At the end of the simulation, the variogram of the simulated indicator field is checked against the model in order to discard possible artifacts caused by this thresholding.

4.4 Disaggregation

This section describes the proposed algorithm to generate realistic DSD fields with prescribed space-time structure, intermittency and block-averaged rain rates. A simplified visual illustration of the disaggregation procedure (without the temporal dimension) is provided in Figure 4.1.

4.4.1 Conditions and constraints

In the following, the conditions that need to be satisfied by the generated DSD fields are described. For simplicity, the driving variable is supposed to be the rain rate. Suppose a target rain rate field $R = (R_1, ..., R_n)$ consisting of n space-time blocks of equal size and duration is provided at the coarse scale (K,T). In the classical disaggregation framework,

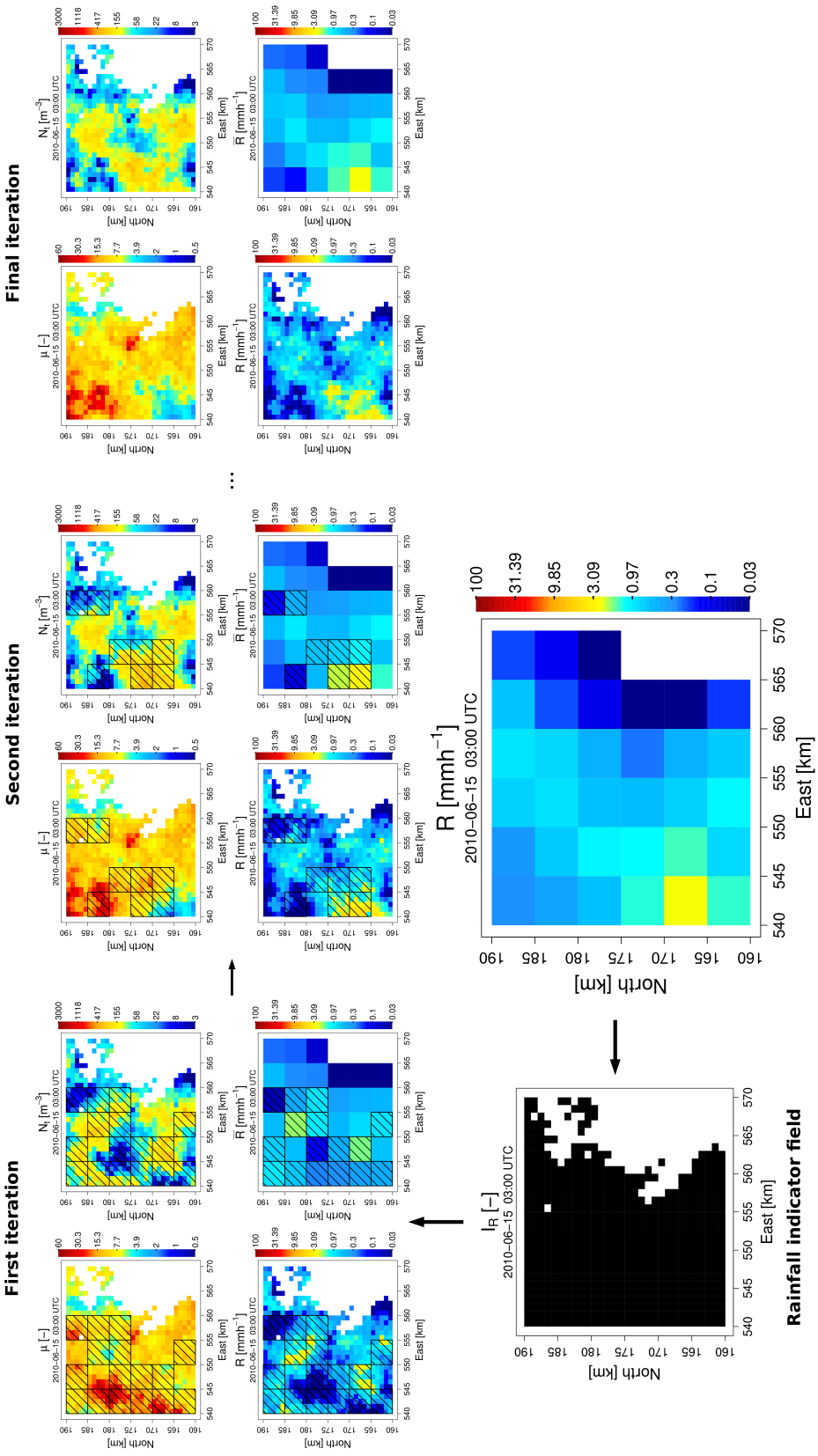

Figure 4.1: Simplified visual illustration of the disaggregation procedure (without temporal dimension). Blocks with incorrect values of rain rate are shaded in black. Their amount decreases with the number of iterations. The procedure continues until all blocks in the considered domain satisfy the conditions given in Equations (4.11), (4.12) and (4.13).

the objective is to generate a new rain rate field $r = (r_1^{(1:m)}, ..., r_n^{(1:m)})$, at a smaller scale (k, τ), such that

$$\frac{1}{m}\sum_{j=1}^{m} r_k^{(j)} = R_k \qquad \forall\, k = 1, \ldots, n \tag{4.10}$$

where m represents the number of space-time pixels of size $k \times k \times \tau$ per block of size $K \times K \times T$. In other words, the generated rain rate field r must aggregate exactly into R when going from the fine to the coarse scale. In practical applications, however, Equation (4.10) only needs to be verified up to a given tolerance level $\varepsilon_R^{\text{abs}}$ [mm h^{-1}] or relative tolerance $\varepsilon_R^{\text{rel}}$ [-] related to the measurement uncertainty on the rain rates at the coarse scale.

A more general approach consists in generating a DSD field $N(D) = \{N(D)_1^{(1:m)} ..., N(D)_n^{(1:m)}\}$ at the fine scale and to derive the corresponding rain rate field r using Equation (4.2). The simulated DSD field must satisfy the same conditions as above:

$$\frac{6\pi}{10^4 m}\sum_{j=1}^{m}\int_{D_{min}}^{D_{max}} D^3 v(D) N(D)_k^{(j)} dD = R_k \qquad \forall\, k = 1, \ldots, n \tag{4.11}$$

where $v(D)$ is the terminal fall speed of a drop of diameter D. Note that for consistency, the DSD is assumed to be defined only for rainy periods and locations.

Finally, recall that the generated DSD fields are intermittent. If $I = (I_1, ..., I_n)$ denotes the intermittency at the coarse scale and $i = (i_1^{(1:m)}, ..., i_n^{(1:m)})$ the intermittency at the fine scale, the following conditions related to the presence and absence of rainfall need to be satisfied:

$$\begin{cases} \sum_{j=1}^{m} i_k^{(j)} = 0 & \forall \text{ blocks for which } I_k = 0 \\ \sum_{j=1}^{m} i_k^{(j)} > 0 & \forall \text{ blocks for which } I_k = 1. \end{cases} \tag{4.12}$$

Also, the percentage p of rainy pixels at the fine scale must be related to the observed intermittency at the coarse scale, in accordance to Equation (4.5)

$$p = \frac{1}{nm}\sum_{k=1}^{n}\sum_{j=1}^{m} i_k^{(j)} = \frac{1}{n}\sum_{k=1}^{n} I_k + p(k,\tau) - p(K,T) \tag{4.13}$$

where $p(k, \tau)$ respectively $p(K, T)$ represent the average climatological rain probabilities at the fine and the coarse scale.

4.4.2 Disaggregation algorithm

This subsection describes the algorithm used to generate intermittent DSD fields satisfying the disaggregation conditions given in Equations (4.11), (4.12) and (4.13). For more details on the underlying models and methods (e.g., Gaussian anamorphosis, SIS and SGS), see Section 4.3. For a (simplified) visual illustration of the disaggregation algorithm, see Figure 4.1.

1. Use SIS to generate a conditional indicator field $i = (i_1^{(1:m)}, ..., i_n^{(1:m)})$ with mean p and space-time structure given by $\gamma_I(h, \tau)$. Note that it might be necessary to draw more than 1 realization until a field satisfying all the conditions given in Equation (4.12) is found.
2. Assign a zero rain rate value to all simulated dry pixels ($i_k^{(j)} = 0$). Keep only the wet pixels for the next simulation steps.
3. For the wet pixels only, use SGS to generate 2 independent Gaussian fields $\tilde{\mu} = (\tilde{\mu}_1^{1:m}, ..., \tilde{\mu}_n^{1:m})$ and $\tilde{N}_t = (\tilde{N}_{t_1}^{(1:m)}, ..., \tilde{N}_{t_n}^{(1:m)})$ with zero mean and space-time variograms given by $\gamma_{\tilde{\mu}}(h, \tau)$ and $\gamma_{\tilde{N}_t}(h, \tau)$.
4. Apply an inverse anamorphosis to back-transform the simulated values ($\tilde{\mu}_k^{(j)}$,$\tilde{N}_{t_k}^{(j)}$) into the original parameter space ($\mu_k^{(j)}$,$N_{t_k}^{(j)}$).
5. Compute the rain rates $r_k^{(j)}$ from the simulated DSD fields using Equations (4.1-4.2).
6. Aggregate the simulated rain rate field $r = (r_1^{(1:m)}, ..., r_n^{(1:m)})$ at the coarse scale.
7. Identify all the blocks that satisfy Equation (4.11). If all the block-averaged rain rates are correct, stop.
8. Generate two new realizations of $\tilde{\mu}$ and $\tilde{N}_t$ on the sub-domain defined by all the (wet) pixels with incorrect block-averaged rain rates. Use the previously simulated values of $\tilde{\mu}_k^{(j)}$ and $\tilde{N}_{t_k}^{(j)}$ on the pixels with correct block-averaged rain rates as conditioning values for the new simulations.
9. Continue until all blocks satisfy Equation (4.11) or until a predefined maximum number of iterations is reached. If there are remaining blocks with incorrect rain rates, go back to 1 and start a new simulation or continue and slightly decrease/increase the range/sill of $\gamma_{\tilde{\mu}}(h, \tau)$ and $\gamma_{\tilde{N}_t}(h, \tau)$.

This defines the method to simulate a realistic intermittent DSD field with given space-time integral rain rate values. Because the proposed procedure is stochastic, an infinite number of DSD fields satisfying the same conditions can be generated by repeating the previous simulation steps. Note that this can be done using the same rainfall indicator field (i.e., starting at step 2) or, more generally, by simulating another indicator field at each iteration.

4.5 Application

In this section, the disaggregation procedure is illustrated using two different events. The first is a rather stratiform event that occurred on the 30$^{\text{th}}$ of March 2010 and the second is a rather convective event that occurred on the 13$^{\text{th}}$ and 14$^{\text{th}}$ of June 2010. The rain rate measurements used to drive the disaggregation process are taken from the operational MeteoSwiss radar rain-rate product (1×1 km^2 and 5 min resolution), for a chosen domain of size 30×30 km^2 in the North East of the city of Lausanne, Switzerland, for which the radar data are known to be of good quality. For validation purposes, the radar data are first aggregated at 5×5 km^2 and 15 min and then disaggregated back to their original resolution, i.e., 1×1 km^2 and 5 min. In this way, it is possible to rigorously compare the statistics of the measured and disaggregated fields and to draw relevant conclusions with respect to the performance of the disaggregation process (see Section 4.5.3 for more details). Snapshots illustrating the two considered events are given in Figure 4.2 and Figure 4.3.

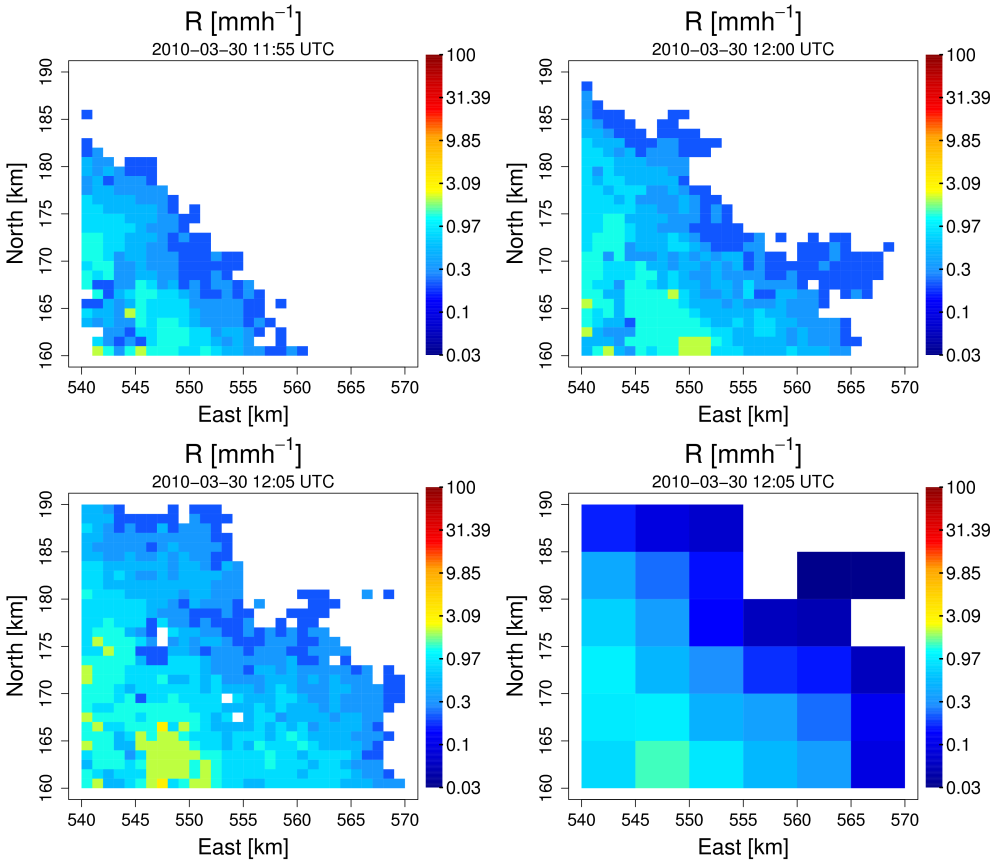

Figure 4.2: Radar rain rate measurements (in mm h^{-1}) for the first 15 min of event 1, the 30$^{\text{th}}$ of March 2010, between 11:55 UTC and 12:05 UTC. The lower right panel shows the aggregated rain rates at 5×5 km^2 for the considered 15 min. The domain size is 30×30 km^2.

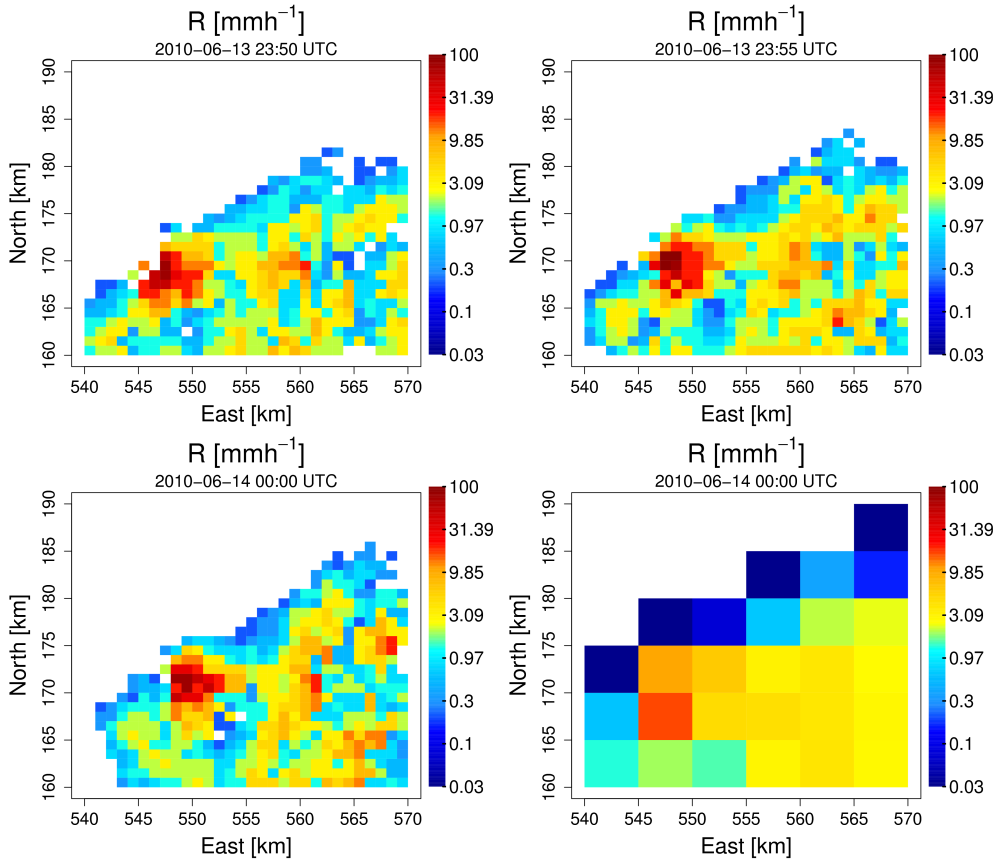

Figure 4.3: Radar rain rate measurements for the first 15 min of event 2, between the 13$^{\text{th}}$ of June 2010 at 23:50 UTC and the 14$^{\text{th}}$ of June 2010 at 00:00 UTC. Same format as in Figure 4.2.

The DSD database used for disaggregation consists of 5239 DSD spectra at 5 min temporal resolution and 1 km^2 spatial resolution collected using a network of 16 optical disdrometers on EPFL campus, Lausanne, Switzerland (Jaffrain et al., 2011). The selected database represents a total of 87.3 hours of rainfall for 33 different events. Note that because DSD parameters are scale dependent, it is important to collect them at scales that are similar to the one used in the disaggregation. A 3-parameter Gamma distribution (μ,Λ,N_t) given in Equation (4.1) has been fitted on each observed DSD spectrum using the maximum likelihood method (correcting for drop censoring at the lower end of the distribution). For simplicity, the rate parameter Λ is expressed as a linear function of the shape parameter μ, therefore reducing the number of stochastic parameters to be simulated. A bivariate Gaussian anamorphosis is applied to transform the observations (μ,N_t) in the original parameter space into independent Gaussian variables $\tilde{\mu}$ and $\tilde{N}_t$ (see Section 4.3.2). A total of 6 rather stratiform events and 9 rather convective events were selected and used to compute climatological variograms representative of the spatial and temporal variability of the transformed DSD parameters. The remaining 18 events (representative of transitional scenarios) were not considered.

4.5.1 Stratiform event

The selected stratiform event lasts several hours but only the first hour between 11:55 and 12:55 UTC is considered for disaggregation. This represents 144 blocks of size 5 $\times$ 5 km^2 and 15 min duration, decomposed into 10800 pixels of size 1 $\times$ 1 km^2 and 5 min duration (i.e., each block is decomposed into 75 pixels). The percentage of dry blocks at the coarse scale is 3.47 %. Based on Equation (4.5), the percentage of dry pixels at the fine scale should therefore be about 8.5 %. Note that this is different from the "true" intermittency value of 12.97 % derived from the radar rain-rate data at the highest resolution. However, because the "true" value of intermittency at the fine scale is usually unknown in practical applications, it has been decided to illustrate the disaggregation procedure using the non-optimal climatological estimate of 8.5 %.

Next, the speed and direction of advection $v \in \mathbb{R}^2$ for this event was estimated using an echo-tracking algorithm (Rinehart, 1979). The results show that for the considered hour, the rainfall field moved at approximatively 12 m s^{-1} towards the North-East. The average advection vector v was then used to estimate the parameters of the corresponding space-time variogram models described in Equation (4.8). In this case, both γ_S and γ_T were well represented by an isotropic spherical variogram model:

$$\gamma(h) = \begin{cases} n + (s-n)\left(\frac{3||h||}{2r} - \frac{||h||^3}{2r^3}\right) & \text{if } ||h|| < r \\ s & \text{if } ||h|| \geq r \end{cases} \tag{4.14}$$

where n is called the nugget, s the sill, r the range and $||\cdot||$ denotes the Euclidean norm in $\mathbb{R}^2$. The fitted values of n, s and r are given in Table 4.1.

Table 4.1: Variogram parameters for event 1 (stratiform). Both γ_S and γ_T are isotropic spherical variogram models. The nugget is assumed to be zero.

	$\gamma^{(S)}$		$\gamma^{(T)}$	
	sill [-]	range [km]	sill [-]	range [min]
I	0.034	24	0.044	33
$\tilde{\mu}$	0.43	18	0.57	24
$\tilde{N}_t$	0.43	12	0.57	16

No particular problems occurred during the disaggregation of this event. The algorithm needed between 30 and 50 iterations on average (corresponding to a few minutes on a standard desktop computer) until all conditions were satisfied. An illustration of the final, disaggregated rain rate field (with DSD parameters μ and N_t) is shown in Figure 4.4. Note that for a better visibility, only the first 15 min are shown.

One can see that the simulated DSD parameters μ and N_t are plausible with respect to stratiform rainfall. There are many large values of μ (small drops) and the spatial

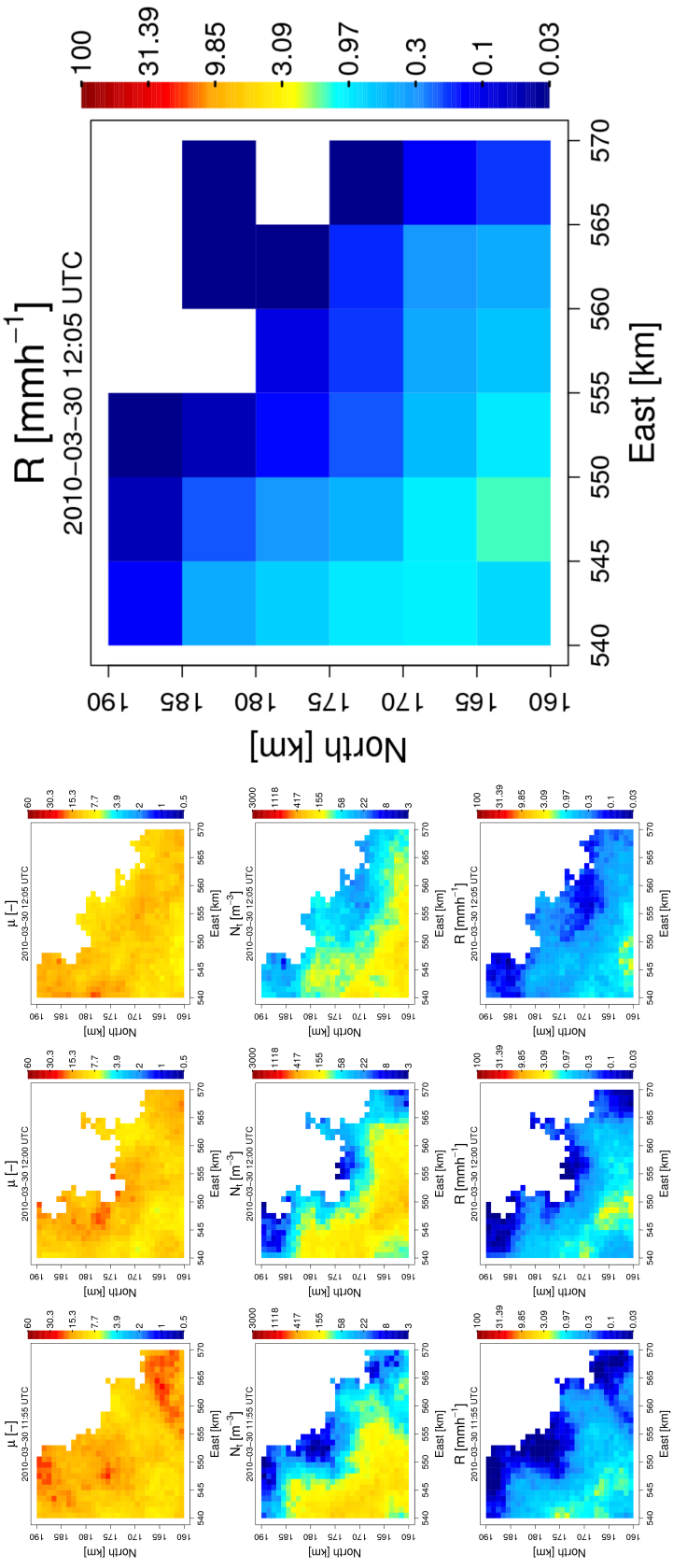

Figure 4.4: Disaggregated fields of μ (top row), N_t (medium row) and corresponding rain rates R (bottom row) for the first 15 min of event 1 (stratiform). Each column represents a 5-min time period between 11.55 UTC and 12:05 UTC. The coarse-scale rain rate field (at 5×5 km^2 and 15 min) used to drive the disaggregation process is shown on the right. Note that by construction, the 3 disaggregated rain rate fields at the bottom aggregate exactly into the coarse scale rain rate field shown on the right.

variability is relatively small. The rain rate seems to be mostly influenced by changes in the drop concentration N_t which exhibits a moderate spatial variability, mostly due to the presence of intermittency. The corresponding rain rates (bottom panel) exhibit limited spatial and temporal variability and look very realistic.

4.5.2 Convective event

Disaggregating convective events is more difficult than stratiform cases. Their variability is larger and their space-time structures can be significantly different from the average climatological variogram. If the difference between the climatological and the "true" (unknown) variogram of the field is too large, the algorithm might not converge at all. Most of the time, however, it just needs more iterations until an acceptable solution satisfying all the prescribed block-averaged rain-rates is found. Another reason why convective events are generally more difficult to disaggregate is because they sometimes exhibit small spatial or temporal non-homogeneities. For example, convective cells with high rain rates will be more difficult to generate, especially if these are surrounded by very low rain rates or dry regions. If the transition between small and large rain rates is too abrupt, the algorithm might not be able to generate a DSD field satisfying both the prescribed spatial structure and the block-averaged rain rates. Fore more details on this problem (and possible solutions), see Section 4.5.4.

The considered convective event is slightly longer than 1 hour but only the first hour between the 13$^{\text{th}}$ of June 2010 at 23:50 UTC and the 14$^{\text{th}}$ of June 2010 at 00:50 UTC is considered. The percentage of dry blocks at the coarse scale is 25 % and is therefore estimated at 30 % at the fine scale. The "true" value, however, is 44.91 %. This clearly shows the current limitations of using climatological relationships for the transformation of dry probabilities from one scale to another. Next, the speed and direction of advection v is estimated for the considered hour and found to be about 6 $\mathrm{m\,s^{-1}}$ along the 30° direction (clockwise from the North). The nugget, sill and range values of the corresponding space-time variogram models defined in Equations (4.8) and (4.14) are given in Table 4.2. Note that similar variogram models are used for the stratiform and convective events, but with different values of sill and range.

Table 4.2: Variogram parameters for event 2 (convective). Both γ_S and γ_T are isotropic spherical variogram models. The nugget is assumed to be zero.

	$\gamma^{(S)}$		$\gamma^{(T)}$	
	sill [-]	range [km]	sill [-]	range [min]
I	0.05	12	0.16	33.3
$\tilde{\mu}$	0.26	7.2	0.74	20
$\tilde{N}_t$	0.26	5.8	0.74	16.1

As expected, the disaggregation algorithm needed slightly more iterations on average (between 50 and 100) until all the conditions were satisfied. An illustration of a disaggregated rain rate field (with DSD parameters μ and N_t) is shown in Figure 4.5. As previously, only the first 15 min are shown.

The generated DSD fields look realistic and are plausible with respect to convective rainfall: the μ values are low (large drops) and exhibit strong spatial and temporal variabilities. The values of N_t are large (high drop concentrations) and also exhibit strong variabilities. The rain rate fields derived from the simulated DSD parameters (bottom row) also look realistic and satisfy all the prescribed conditions at the coarse scale. They are, however, slightly too smooth compared to the radar estimates at the same scale (see Figure 4.3 for comparison). This discrepancy is confirmed by additional analyses presented in Section 4.5.3 that show that both the spatial and temporal ranges of the simulated fields are overestimated.

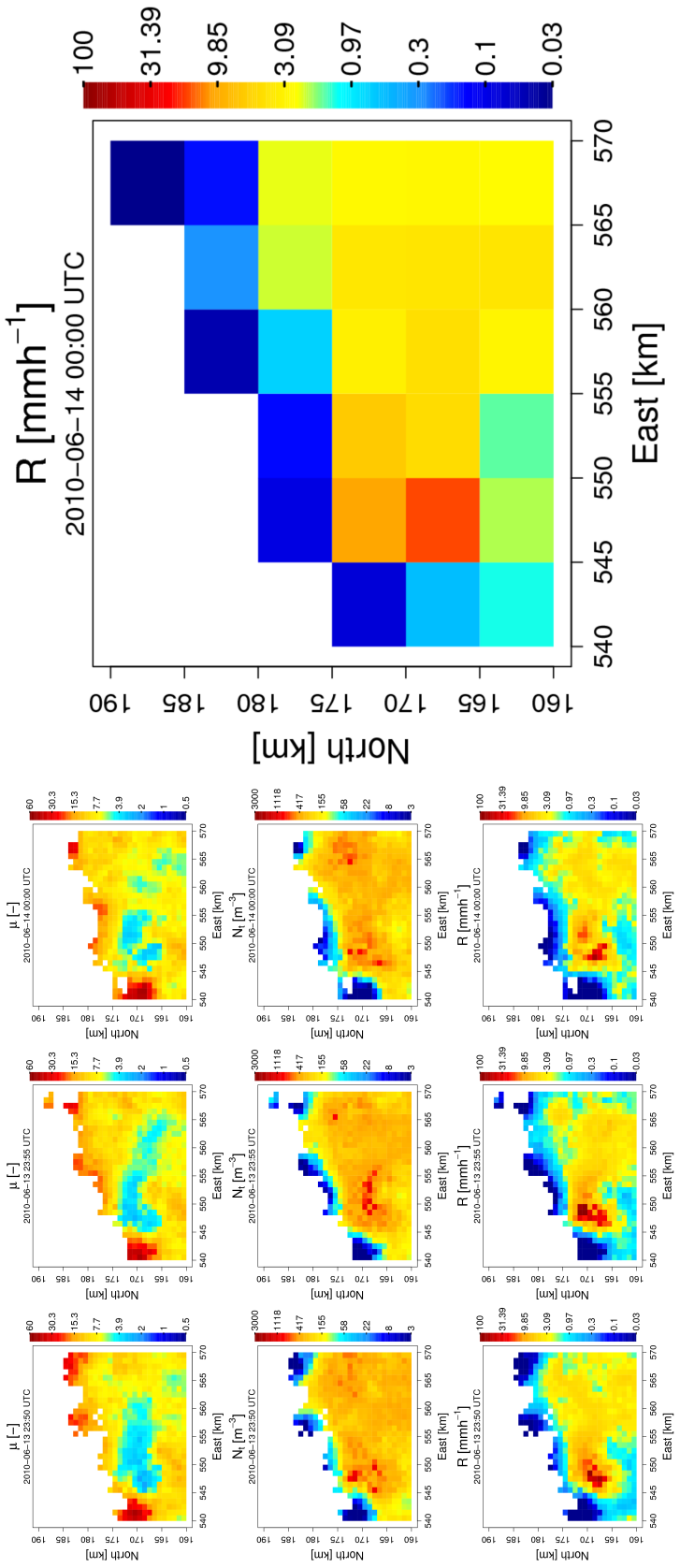

Figure 4.5: Disaggregated fields of μ, N_t and R for the first 15 min of event 2 (convective). Same format as in Figure 4.4.

4.5.3 Evaluation

Figures 4.4 and 4.5 nicely illustrate the disaggregation process but only represent 1 possible realization for each considered event. In order to better evaluate the performance of the proposed algorithm, 100 different realizations have been generated for each type of rainfall (stratiform and convective). The statistical properties of the generated fields have then been analyzed and compared to the ones derived from the radar rain-rate data at the same spatial and temporal resolutions. Note that the goal is not to reproduce the same rain rate values as the radar but to generate an ensemble of realizations that are consistent with the radar data and have similar statistical properties. Here, the criteria used to assess the consistency between the simulated and observed fields are (1) the rain rate distributions over the considered region, (2) the percentage of dry locations over the considered domain, (3) the spatial structure of the rain rate fields and (4) the temporal structure of the rain rate fields.

The results for criteria 1-2 are shown in Figure 4.6 (stratiform event) and in Figure 4.7 (convective event).

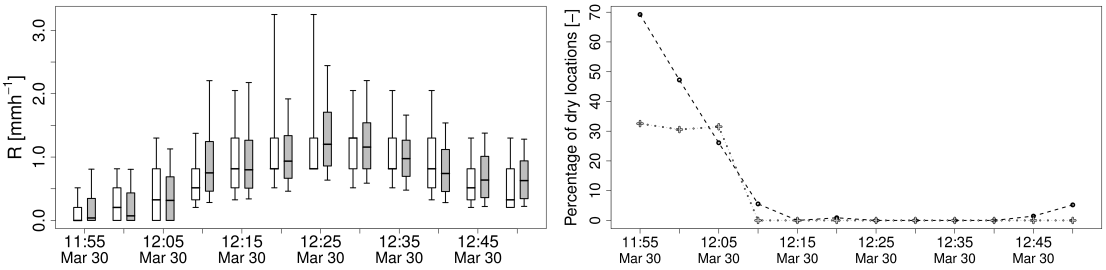

Figure 4.6: Statistical analysis of 100 simulated realizations for event 1 (stratiform). The left panel shows the measured (black) and simulated (plain gray) rain rate distributions using box plots (10 %, 25 %, 50 %, 75 % and 90 % quantiles). On the right, the measured (black points) and simulated (grey crosses) percentage of dry locations with respect to time.

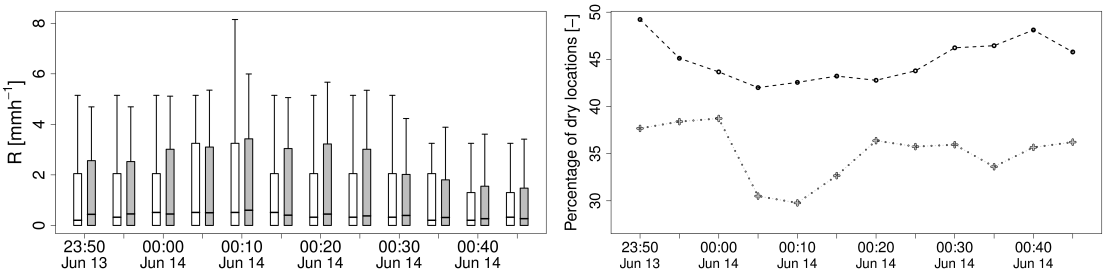

Figure 4.7: Statistical analysis of 100 simulated realizations for event 2 (convective). Same format as in Figure 4.6.

In both cases, the simulated rain rate distributions are consistent with the observations. The percentage of dry locations is underestimated in both cases because the intermittency

at the small scale is estimated using a climatological relationship which is not optimal for the considered events. Note that this is not a limitation of the disaggregation method but rather a consequence of our lack of knowledge about the scaling of rainfall intermittency.

The spatial and temporal structures of the disaggregated rain rate fields are shown in Figure 4.8 (stratiform event) and Figure 4.9 (convective event). Note that the analysis focuses on the small spatial and temporal scales only. Larger scales are less interesting because they are mostly conditioned by the structure of the rain rate field used to drive the disaggregation process (and are therefore very similar). Note that because the rain rate distributions are highly skewed, all the sample variograms were estimated using Cressie's robust estimator (Cressie and Hawkins, 1990). Figure 4.8-4.9 show that the small-scale structures of the disaggregated fields are consistent with the "true" structures observed by the radar.

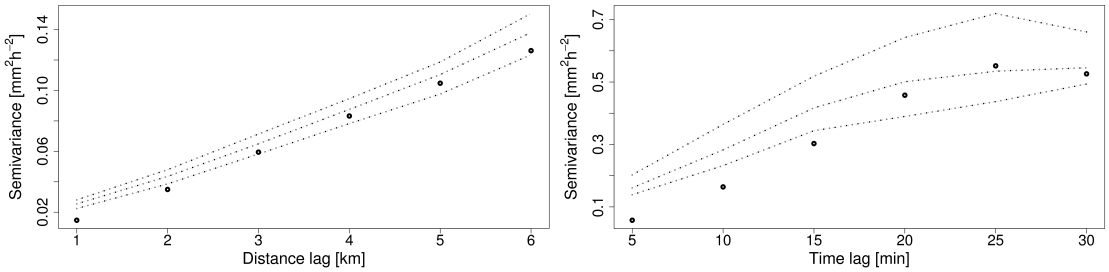

Figure 4.8: Statistical analysis of the spatial and temporal structures of 100 simulated realizations for event 1 (stratiform). The left panel shows the 10 %, 50 % and 90 % quantiles of the spatial variogram of the disaggregated rain rate fields (grey dotted lines). The black dots represent the "true" spatial variogram computed using the radar rain rates at the fine scale. The right panel shows a similar comparison but for the temporal structure of the disaggregated rain rate fields.

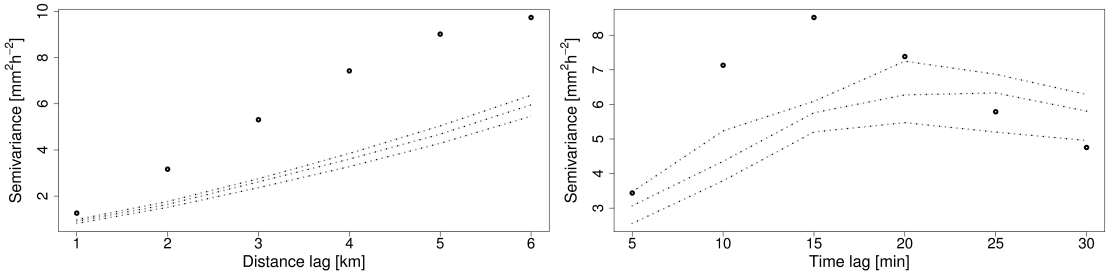

Figure 4.9: Statistical analysis of the spatial and temporal structures of 100 simulated realizations for event 2 (convective). Same format as in Figure 4.8.

The spatial and temporal structures of the stratiform event are particularly well reproduced. The convective event, on the other hand, is too smooth compared to the radar data, which can also been seen when comparing Figure 4.3 and Figure 4.5. Clearly, the climatological space-time variograms for convective events are not optimal in this case, although one must be careful because radar rain-rate estimates are not perfect either. A

significant part of the observed spatial variability in Figure 4.3 could be due to noise affecting the radar measurements. If this is the case, then the disaggregated rain rate fields (that are free from any noise because computed directly from the DSD) are more realistic.

4.5.4 Discussion

The simulated DSD fields presented in this section are all plausible in terms of structure and dynamics and their associated rain rates are consistent with radar measurements at the same scales. In general, stratiform events are easier to disaggregate than convective events because they are smoother and more homogeneous in space and time. Convective events, on the other hand, are more difficult to generate using a single climatological variogram. Furthermore, if there are regions that exhibit a significantly different behavior and variability than the rest of the domain, the disaggregation algorithm might not converge at all. In this case, it is better to break down the simulation domain into smaller sub-domains represented by different variabilities and variograms. This naturally limits the size of the simulation domain which should be in the order of 50×50 km^2 and a few hours in time, after which non-stationarities and non-homogeneities may become too important to be ignored. Instead of breaking down the simulation domain into smaller pieces, another, maybe more interesting approach is to slowly increase/decrease the sill/range of the variograms after each iteration step. In this way, the DSD fields are generated in ascending order of variability, starting with the large scale structures (with small variability) and finishing with the regions that exhibit the strongest spatial and temporal variabilities. Intuitively, this is equivalent to generate a smooth field representative of the trend over the entire domain and to add successive layers of (increasing) noise to this initial field.

Note that the average rain rate at the coarse scale was the only variable used to drive the disaggregation process in this application. The presented disaggregation method is, however, far more general and can be applied to other and more complex inputs and driving variables. It is, for example, possible to drive the disaggregation process using radar reflectivity fields instead of rain rates in Equation (4.10). The relationship between the DSD and the driving variable is then given by:

$$Z_{h|v} = \frac{10^6 w^4}{\pi^5 |\frac{m^2-1}{m^2+2}|} \int_D \sigma_{B_{h|v}}(D) N(D) dD \tag{4.15}$$

where w [cm] represents the radar wavelength, m [-] the complex refractive index of water (at a given temperature) and $\sigma_{B_{h|v}}(D)$ the backscattering cross sections at horizontal and vertical polarizations for a drop of diameter D. Such an extension is easy, straightforward and particularly interesting if reflectivity measurements are available at different frequencies (e.g., GPM) or polarizations because in this case, the generated DSD fields

must satisfy the prescribed space-time integral reflectivity values at both frequencies or polarizations which significantly reduces the range of possible solutions compared to the traditional disaggregation framework. This is, however, beyond the scope of the present study.

4.6 Conclusions

In this chapter, a stochastic method for the disaggregation of rainfall fields into intermittent fields of DSD with given space-time integral rain rates has been presented. The feasibility of the method has been illustrated using radar rain-rate data for two different events and types of rainfall (convective vs stratiform). The results show that the generated DSD fields exhibit realistic spatial and temporal structures and that their associated rain rates are consistent with radar measurements at the same scales. The rainfall intermittency at the small scale is slightly underestimated but coherent with the available observations at the coarse scale. Improvement can be expected if other, more advanced methods are used to estimate the amount of intermittency at the fine scale based on the observations at the coarse scale. The application further showed that stratiform events are easier to disaggregated than convective events because they are smoother and exhibit fewer non-stationarities. However, for convective events, convergence can still be obtained if the sill/range is slightly increased/decreased at each iteration.

The main limitations of the proposed disaggregation method are that the space-time variograms used to generate the DSD fields at the fine scale must be (more or less) representative over the entire simulation domain (which limits the size of the domain that can be simulated) and that the disaggregation scheme is limited to 2-dimensional fields of liquid precipitation. Issues related to mixed-phase precipitation and vertical variabilities (which are very important for satellites applications) are not addressed in this version but will be investigated in future work.

The main advantages of the proposed approach are its generality and flexibility with respect to the variables and constraints that can be used to drive the disaggregation process and the broad range of potential applications. The DSD can be parameterized using 2 or 3 stochastic parameters for a standard or normalized Gamma distribution (or any parametric distribution in general). Furthermore, any weighted moment of the DSD (or combination of different moments) can be used to drive the disaggregation process. Most of the time, the target variable is the rain rate but other, more general approaches can be considered. Future work will mainly focus on disaggregation using dual-frequency reflectivity constraints, vertical variability and methods to deal with non-stationarities in larger simulation domains.

5 Statistical analysis of rainfall intermittency at small spatial and temporal scales [1]

5.1 Summary

Rainfall intermittency is analyzed and quantified at small spatial and temporal scales using 2 years of radar and disdrometer data collected in Switzerland. Analytical models are fitted and used to describe the intermittency for spatial scales between 0 and 30 km and temporal resolutions between 30 s and 6 h, providing climatological parameterizations for efficient and accurate upscaling/downscaling of intermittent rainfall fields. First, the zero rainfall probability is analyzed with respect to the considered spatial resolution. Second, the spatial autocorrelation of rainfall intermittency is quantified with respect to the temporal resolution. Finally, the temporal autocorrelation is analyzed with respect to the spatial resolution. The results show that all these different aspects of rainfall intermittency can be accurately described by a scaled exponential function with a fixed shape parameter and a variable scale parameter. Models describing this variability are provided.

1. This chapter is a slightly modified version of the article by Schleiss, M., J. Jaffrain and A. Berne, 2011: Statistical analysis of rainfall intermittency at small spatial and temporal scales, *Geophys. Res. Lett.*, **38**, doi:10.1029/2011GL049000

5.2 Introduction

Precipitation is a highly variable non-continuous process in space and time. It is characterized by relatively long dry periods punctuated by shorter rain events with complex spatial and temporal structures. This constant alternating between dry and rainy periods, called rainfall intermittency, significantly affects the environment and the ecosystems. For example, vegetation cover in arid or semi-arid regions is highly sensitive to rainfall intermittency which limits the available resources (water and nutrient) and controls their abundance in time (Kletter et al., 2009). In more temperate regions, the soil and surface hydrology are known to be strongly sensitive to rainfall intermittency which influences important natural processes like stream flow, runoff, soil moisture and soil erosion (Pitman et al., 1990).

The major difficulty with rainfall intermittency is the fact that it varies significantly with respect to the considered spatial and temporal scales. Short time periods are more likely to be completely dry than long ones and small areas are more likely to be dry than large ones. Yet the ability to quantify the probability of zero rainfall at multiple space and timescales is crucial for many practical applications in hydrology, meteorology and remote sensing of precipitation. For example, it plays a major role in rainfall interpolation and disaggregation techniques and must be taken into account when upscaling/downscaling the outputs of numerical weather models or weather radar data (e.g., Seo, 1998; Lanza et al., 2001). Other examples can be found in the field of stochastic rainfall simulation (Kang and Ramirez, 2010), partial beam filling and path-integrated attenuation using ground based weather radars or satellites.

Several studies have investigated rainfall intermittency at different spatial or temporal resolutions. Among the studies focusing on spatial intermittency, Braud et al. (1993) and Jeannin et al. (2008) analyzed the relations between the mean areal rainfall and the fractional area where it rains above a fixed threshold. Using mathematical morphology, Kumar and Foufoula-Georgiou (1994) proposed different algorithms for downscaling/upscaling intermittent rainfall fields. Using radar data, Pavlopoulos and Gritsis (1999) and Pavlopoulos and Gupta (2003) analyzed the durations and scaling of wet and dry periods. Using time series from tipping bucket rain gauges, Molini et al. (2009) and Rigby and Porporato (2010) investigated different relations between rainfall intermittency and turbulence across a large range of time scales and climatic regimes. Their study, however, did not include any information on the spatial structure of rainfall intermittency. Using a different approach based on the maximum entropy principle, Koutsoyiannis (2006) noted that, under some circumstances, the probability that a time interval is dry, follows a scaled exponential function of timescale. Using disdrometer measurements and high temporal resolutions, Lavergnat and Golé (1998) analyzed the inter-arrival times of raindrops and proposed to model the time intervals between raindrops using a Bi-Pareto law. Finally, recent work by Kundu and Siddani (2011) shows that dry probabilities (both spatial and temporal) can be modeled using a scaled exponential function.

This chapter extends previous results by providing a complete statistical analysis of rainfall intermittency (including spatial and temporal structures) for a large range of scales. Section 5.3 provides some definitions and Section 5.4 presents the data used for the analysis. In Section 5.5, the zero-rainfall probability is quantified using 2 years of data collected in Switzerland. In Sections 5.6 and 5.7, the spatial and temporal autocorrelation of rainfall intermittency are analyzed. The conclusions and perspectives are given in Section 5.8.

5.3 Modeling the intermittency

Rainfall intermittency, i.e., the presence or absence of rainfall, can be described by an indicator function

$$I(x,t) = \begin{cases} 1 \text{ if } R(x,t) > 0 \\ 0 \text{ else} \end{cases} \tag{5.1}$$

where $R(x,t)$ [$\mathrm{mm\,h^{-1}}$] represents the instantaneous rain rate at location x and time t. By extension, we define

$$I(X,T) = \begin{cases} 1 \text{ if } \exists\,(x,t) \in (X,T) \text{ with } I(x,t) = 1 \\ 0 \text{ else} \end{cases} \tag{5.2}$$

where X is a given area and T a given time period. For simplicity, only square areas of size $k \times k$ and continuous time periods $[t_0, t_0 + \tau]$ of duration τ are considered. For convenience, k is chosen to be expressed in kilometers and τ in hours. The probability that a randomly chosen area of size k remains dry for at least τ hours is denoted by $p(k,\tau)$. For consistency, it is supposed that $p(k,\tau)$ is well-defined and does not depend on the location nor on the timing, i.e., that $I(X,T)$ is stationary over the considered period and area. Note that if this is not the case, the analysis can always be confined into appropriate areas and periods (e.g., months or seasons) for which $I(X,T)$ can be considered stationary.

5.4 Data

The results in this chapter are mostly based on the analysis of nearly 2 years of operational radar rain-rate data provided by MeteoSwiss. These data are available from 30th April 2009 to 2nd February 2011, at a spatial resolution of 1×1 km^2 and a temporal resolution of 5 min. There are almost no missing data for the considered time period. The estimated rain rates are obtained by combining the measurements of 3 C-band weather radars at different elevations, correcting for the main sources of errors (ground clutter, beam shielding, vertical variability) according to a procedure described in Germann et al. (2006). For the purpose of this study, only the rain/no-rain information is retrieved from

the estimated rain-rate maps, which limits the influence of the uncertainties associated with rain rate retrieval from radar measurements. Note that the minimum detectable rain rate is 0.16 mm h^{-1} which defines the threshold for rain/no-rain detection. Two 30×30 km^2 areas located in the North-East and in the South-West of Geneva are selected for statistical analysis. These are the biggest square areas with no evident signs of data contamination (e.g., ground clutter or beam shielding) that were found for the selected time period. Note that both areas are relatively flat and close (less than 60 km) to one of the 3 C-band radars. It is therefore assumed that false rain and dry detections are negligible over the selected areas and do not affect significantly the statistical and structural analysis.

At very small spatial scales (less than 1 km^2), rainfall intermittency is analyzed using the data from a network of disdrometers deployed over EPFL campus, Lausanne, Switzerland (Jaffrain et al., 2011). The distance from the network to the 2 areas defined above is less than 100 km. For comparison with radar data, the same time period between 30th April 2009 and 2nd February 2011 is considered. During this period, a total of 6 disdrometers sampling at 30 s temporal resolution and separated by 80 m to 800 m were available. For consistency with radar data, the same rain/no-rain detection threshold of 0.16 mm h^{-1} (corresponding to the 20 % rain-rate quantile) is applied to the disdrometer data.

5.5 Zero-rainfall probability

In this section, the data presented in Section 5.4 are used to investigate the probability $p(k, \tau)$ that a randomly chosen area of size k remains completely dry for at least τ hours. Because they are very similar for the 2 considered areas, the results are only shown and discussed for the first area. For details about the second area, see Table 5.1. During this analysis, the largest spatial scale that can be considered is given by the size of the studied area, i.e., $k = 30$ km. The smallest possible scale is given by the radar resolution (k =1 km). For smaller scales, the rain/no-rain information derived from the 6 disdrometers are used to estimate $p(k, \tau)$ at the point-scale, i.e., $k = 0$.

The values of $p(k, \tau)$ for different spatial and temporal scales are displayed in Figure 5.1. It shows that, for a fixed spatial scale k, $p(k, \tau)$ can be described by a scaled exponential function:

$$p(k, \tau) = p(k) \exp\left\{ - \left[\frac{\tau}{\lambda(k)} \right]^{\beta(k)} \right\} \tag{5.3}$$

where $\lambda(k) > 0$ [h^{-1}] is a scale parameter, $\beta(k) > 0$ [-] a shape parameter and $p(k) \in [0, 1]$ represents the dry probability for $\tau = 0$. The fact that scaled exponential functions can be used to describe zero-rainfall probabilities has been pointed out previously by Kundu and Siddani (2011). Another parallel can be found in the field of survival analysis, where

scaled exponentials are used to express the probability that the time of "death" occurs later than some specified time. Hence another way of seeing rainfall intermittency is through a marked point process characterized by the "birth" and "death" of individual dry periods.

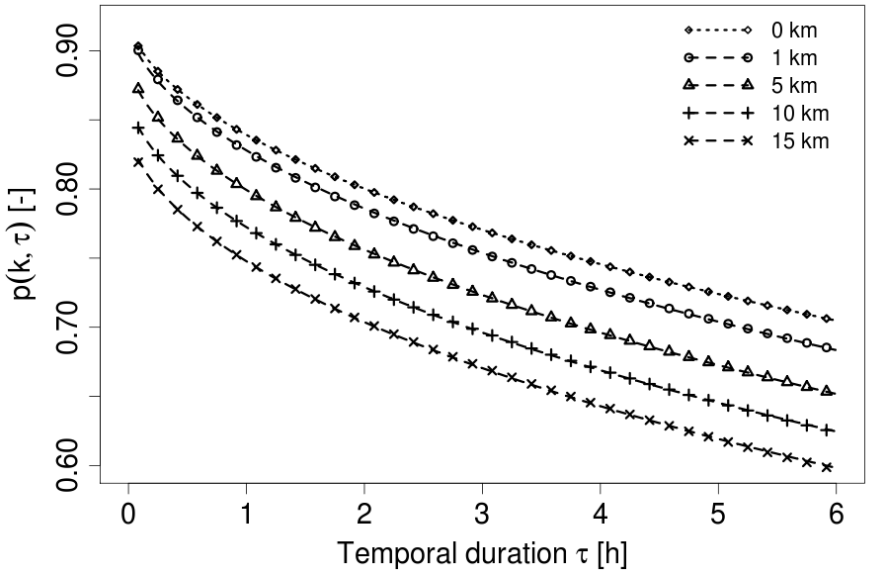

Figure 5.1: Estimated values of $p(k, \tau)$ at k=0, 1, 5 , 10 and 15 km spatial resolution. All estimates are obtained using radar data except k=0 which represents the average value computed from the 6 disdrometers. The dashed lines represent the fitted model from Equation 5.3.

In order to investigate the dependence of $p(k)$, $\lambda(k)$ and $\beta(k)$ to the considered spatial scale, the radar estimates of $p(k, \tau)$ corresponding to each value of k between 1 and 15 km, were fitted using non-linear least squares. The scatter plots of $p(k)$, $\lambda(k)$ and $\beta(k)$ with respect to k suggest that $p(k)$ is a linear function, $\lambda(k)$ a power law and that $\beta(k)$ can be assumed constant.

$$\begin{cases} p(k) = p_0 + p_1 k \\ \lambda(k) = a k^b \\ \beta(k) = \beta \end{cases} \tag{5.4}$$

Figure 5.2 shows the fitted scatter plots of $p(k)$ and $\lambda(k)$ with respect to k. The triangles at $k = 0$ represent the average parameter estimates from the 6 disdrometers. Note that these were not used to fit the relations in Equation 5.4 but are in very good agreement with their values predicted using radar measurements only. The fitted values of p_0, p_1, a, b and β are given in Table 5.1. Note that they are likely to be specific to the considered region and must therefore be adapted to the local climatology. It is conjectured, however, that the functional forms for $p(k)$ and $\lambda(k)$ are generally valid (at least for the considered

range of spatial scales) and can be used to describe rainfall intermittency for other regions and climatologies.

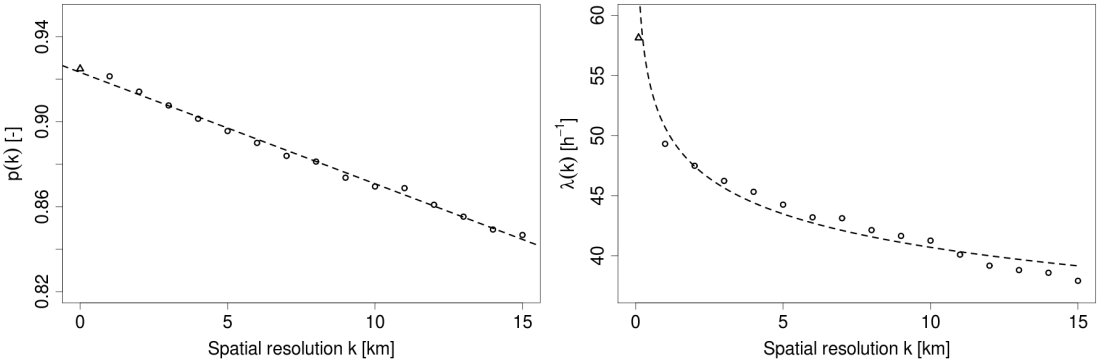

Figure 5.2: Estimated values of $p(k)$ (left panel) and $\lambda(k)$ (right panel) at spatial resolutions between 1 km and 15 km. The triangles at $k = 0$ represent the average parameter estimates from the 6 disdrometers. The dotted lines represent the fitted models given in Equation 5.4.

The sensitivity of p_0, p_1, a, b and β to the rain/no-rain detection threshold has been investigated by considering the next possible radar rain detection threshold, i.e., 0.25 mm h^{-1} (corresponding to the 30 % rain rate quantile at the point-scale). Higher thresholds would not be reasonable as too many rain rate values would be below this threshold. The results show that the model parameters are not very sensitive to the rain threshold (less than 2 % variation for p_0, a and β and 15 % for b and p_1). Moreover, the functional forms in Equations 5.3 and 5.4 remained unchanged. Finally, the stationarity hypothesis was tested by splitting the data into 4 periods (summer 2009, winter 2009-2010, summer 2010 and winter 2010-2011) and by computing $p(k, \tau)$ for each season. No statistically significant differences between these periods could be observed.

5.6 Spatial autocorrelation structure

In this section, the spatial autocorrelation structure $\rho_S(h, \tau)$ of the rainfall intermittency is investigated at different distance lags h and temporal resolutions τ using the data presented in Section 5.4. In order to ensure the highest possible resolution in the spatial domain, this analysis is performed at the highest available spatial resolution, i.e., 1×1 km^2.

Figure 5.3 shows the spatial autocorrelation for $\tau = 5$ min, 15 min, 45 min, 120 min and 360 min. The data show that, for each temporal resolution τ, the spatial autocorrelation at lag h can also be described by a scaled exponential function:

$$\rho_S(h, \tau) = \exp\left\{ -\left[\frac{h}{\lambda_S(\tau)} \right]^{\beta_S(\tau)} \right\} \tag{5.5}$$

where $\lambda_S(\tau) > 0$ [km^{-1}] is a scale parameter and $\beta_S(\tau) > 0$ [-] a shape parameter. The scatter plots of $\lambda_S(\tau)$ and $\beta_S(\tau)$ versus τ show that $\lambda_S(\tau)$ can be described as a power law of τ and that $\beta_S(\tau)$ can be assumed constant for all temporal scales:

$$\begin{cases} \lambda_S(\tau) = a_S \tau^{b_S} \\ \beta_S(\tau) = \beta_S \end{cases} \tag{5.6}$$

where a_S, b_S and β_S depend on the local climatology.

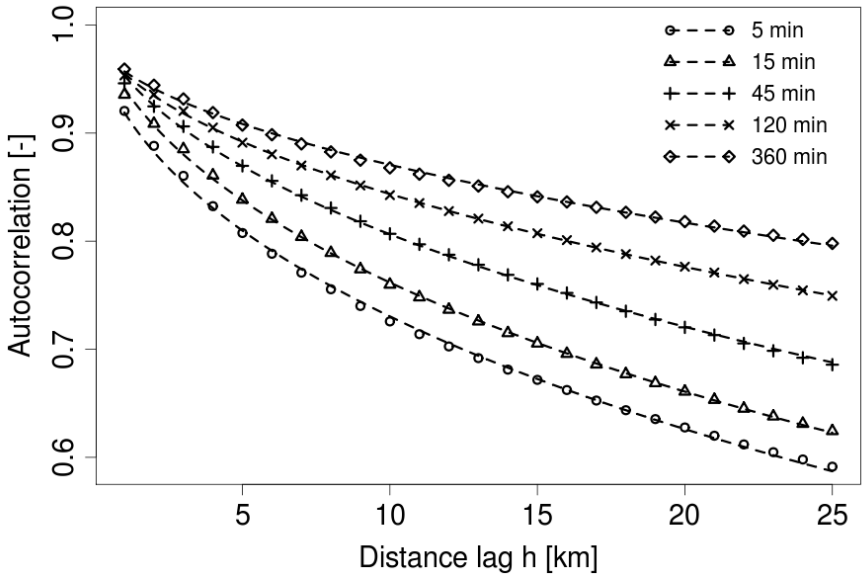

Figure 5.3: Estimated spatial autocorrelation $\rho_S(h, \tau)$ of rainfall intermittency at 5, 15, 45, 120 and 360 min temporal resolutions. For each temporal resolution, a scaled exponential function as given in Equation 5.5 has been fitted.

The fitted values of a_S, b_S and β_S for the considered areas are given in Table 5.1. The fitted model for $\lambda_S(\tau)$ is shown in Figure 5.4. The sensitivity of a_S, b_S and β_S to the radar rain detection threshold (see Section 5.5) has been investigated and is in the order of 3 % for a_S and β_S and 15 % for b_S. The functional forms in Equations 5.5 and 5.6 were also preserved.

5.7 Temporal autocorrelation structure

In this section, the temporal autocorrelation structure $\rho_T(\tau, k)$ of the rainfall intermittency is investigated at different spatial resolutions using the data presented in Section 5.4. The approach is similar to the previous section and, for conciseness, only the results without any further illustrations are given. In order to ensure the highest possible resolution in the temporal domain, this analysis is performed at the highest available temporal resolution, i.e., 5 min. The results show that, for each fixed spatial resolution k, the

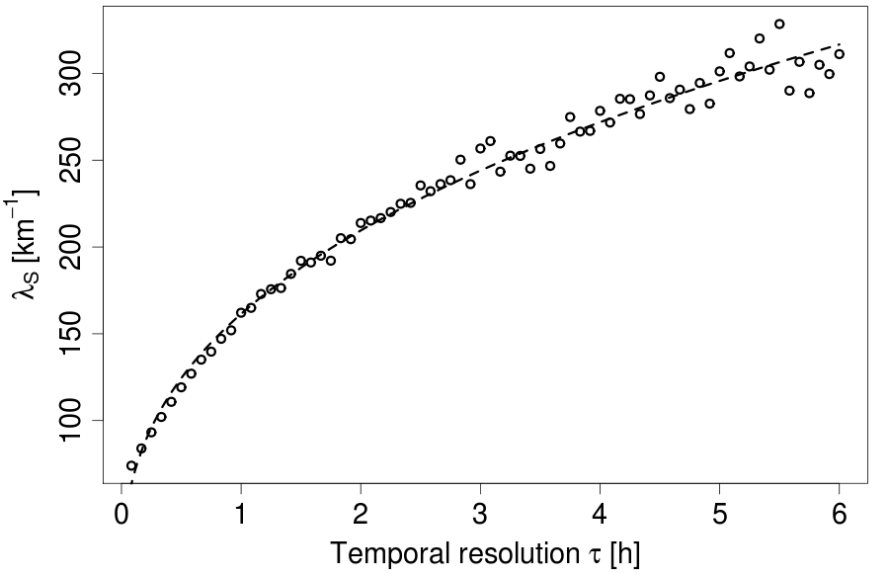

Figure 5.4: Estimated values of $\lambda_S(\tau)$ for τ between 5 min and 6 h. The dotted line represents the fitted power law given in Equation 5.6.

temporal autocorrelation $\rho_T(\tau, k)$ of rainfall intermittency can be described by a scaled exponential function:

$$\rho_T(\tau, k) = \exp\left\{-\left[\frac{\tau}{\lambda_T(k)}\right]^{\beta_T(k)}\right\} \tag{5.7}$$

for a given scale parameter $\lambda_T(k) > 0$ [h^{-1}] and shape parameter $\beta(k) > 0$ [-]. In this case, the scatter plots of $\lambda_T(k)$ and $\beta_T(k)$ versus k show that $\lambda_T(k)$ is a linear function of k and that $\beta_T(k)$ can be assumed constant at all spatial scales.

$$\begin{cases} \lambda_T(k) = a_T + b_T k \\ \beta_T(k) = \beta_T \end{cases} \tag{5.8}$$

The fitted values of a_T, b_T and β_T are provided in Table 5.1. Similarly to the zero-rainfall probability, the values of $\lambda_T(0)$ and $\beta_T(0)$ obtained from the disdrometer data (i.e., $k = 0$) are in very good agreement with their predicted values using radar data only. The sensitivity of a_T, b_T and β_T to the rain detection threshold (similarly to the previous sections) is lower than 2 % for b_T and β_T and in the order of 15 % for a_T.

5.8 Conclusions

Rainfall intermittency significantly varies with respect to the spatial and temporal scale at which it is considered. It is larger at finer scales and its spatial correlation is larger at lower

Table 5.1: Estimated parameter values.

	Parameter	Area 1	Area 2
Zero rainfall probability	p_0	0.925	0.936
	p_1	-0.005	-0.007
	a	50.8	45.1
	b	-0.096	-0.173
	β	0.57	0.64
Spatial autocorrelation	a_S	161.5	147.2
	b_S	0.376	0.362
	β_S	0.579	0.527
Temporal autocorrelation	a_T	2.977	2.958
	b_T	0.156	0.138
	β_T	0.514	0.501

temporal resolutions (i.e., longer time periods). In this chapter, a statistical analysis of rainfall intermittency at different spatial and temporal resolutions has been performed. It shows that zero-rainfall probabilities can be adequately described by a scaled exponential function (of the timescale) with parameters depending on the considered spatial scale. Models describing this dependency have been provided. The estimated parameters are likely to depend on the considered region and climatology but the functional forms and models are assumed to be generally valid. The choice of models is supported by the fact that independent disdrometer measurements (at scales that can not be observed by the radar) are in very good agreement with their radar-based predictions. Using the same approach, the spatial and temporal autocorrelations of rainfall intermittency are also quantified. The model parameters are not very sensitive to the rain/no-rain detection threshold (up to 15 % for the considered thresholds). The functional forms describing $p(k, \tau)$, $\rho_S(h, \tau)$ and $\rho_T(\tau, k)$ are not influenced by this threshold and can be used to improve the modeling of rainfall intermittency (including its spatial and temporal structure) in hydroclimatological models, to upscale and downscale the outputs of numerical weather prediction models and to parameterize stochastic rainfall simulators.

6 Identification of dry and rainy periods using telecommunication microwave links [1]

6.1 Summary

Microwave links are widely used for wireless data exchange, in particular between base stations of mobile phone networks. Because of the range of frequencies used for data transmission, the link signal is attenuated when rainfall occurs along the path of the link. This attenuation can be related to the path-averaged rain rate. A critical issue in this procedure is the ability to separate the attenuation due to rainfall from the attenuation occurring during dry periods.

In this chapter, a technique to separate dry and rainy periods and to estimate a time-varying attenuation baseline using path-integrated attenuation measurements from operational telecommunication microwave links is presented. Dry and rainy periods are separated by analyzing the local variability of the link signal. The attenuation baseline is estimated in real-time through the classification into dry and rainy periods. The method is applied to 4 different links and 10 different rain events. Measurements from a near-by C-band weather radar are used to evaluate the performance of the algorithm. The new method performs well, identifying about 92 % of all rainy periods, 86 % of all dry periods and 93 % of the total rain amount on average.

1. This chapter is a slightly modified version of the article by Schleiss, M. and A. Berne, 2010: Identification of dry and rainy periods using telecommunication microwave links, *IEEE Geosci. Remote Sens. Lett.*, **7**, No.3, pp. 611-615

6.2 Introduction

The possibility to use operational microwave links employed in telecommunication networks for rain estimation has recently been demonstrated (Messer et al., 2006; Leijnse et al., 2007c). Because of the range of frequencies (roughly from 10 to 50 GHz) used for data transmission, the link signal is attenuated when rainfall occurs along the link path. This attenuation can be related to the path-integrated rain rate and used to provide valuable information at an intermediate scale between point measurements and radar measurements Rahimi et al. (2004). A critical issue in this approach is the ability to distinguish the attenuation due to rainfall from the attenuation occurring during dry periods, referred to as the attenuation baseline (Leijnse et al., 2007c; Rahimi et al., 2003). The difficulty lies in the fact that the attenuation baseline changes with time and fluctuates due to changes in water vapor concentration, temperature, wind effects on the antenna, losses during transmission and reception, interferences, wet-antenna and possible multipath effects. Moreover, attenuation measurements are quantized (usually at 1 dB) which introduces additional variability in the process.

Using dual-frequency links specifically designed for rainfall estimation, Rahimi et al. (2003) and Holt et al. (2003) proposed to identify dry and rainy periods using the correlation between the link signals at 2 different frequencies. Following the same idea, Goldshtein et al. (2009) proposed to identify dry and rainy periods using the correlation between bidirectional signals (each antenna acting both as a transmitter and a receiver). For dual-polarization links, Ruf et al. (1996) and Aydin and Daisley (2002) showed that the attenuation baseline can be removed by considering the difference of attenuation between the two polarizations. Unfortunately, many operational telecommunication microwave links only use single-polarization signals and the frequencies of local and remote signals are often too close (about 1 GHz apart) for dual-frequency considerations.

In this chapter, a method to identify and separate dry from rainy periods and to fit a time-varying attenuation baseline using only path-integrated attenuation measurements from single-polarization single-frequency telecommunication microwave links is presented. Once the attenuation baseline has been estimated, the attenuation due to rainfall can be retrieved and used to estimate the path-integrated rain rate. Furthermore, the fact that the proposed technique does not require any calibration is of advantage in locations where rain gauge and radar monitoring is sparse or hindered. Note that the focus of this study is on liquid precipitation only. Issues related to solid or mixed-phase hydrometeors as well as antenna icing are not addressed.

This chapter is organized as follows: Section 6.3 presents the microwave link and radar data. Section 6.4 describes the proposed method and Section 6.5 presents an application for different rain events and different links. The summary and conclusions are given in Section 6.6.

6.3 Data

Measurements of instantaneous (up to a few ms) transmitted and received powers, expressed in dBm, from 4 operational telecommunication links are provided by Bouygues Telecom, a French telecommunication company. The length, frequency and polarization of each link are given in Table 6.1. Note that link 13 and 14 are dual-polarization links, but only the horizontal polarization is used for this study. Link 1 and link 2 are co-located parallel links but with distinct transmitters and receivers. The attenuation along the link path is obtained by subtracting the received power to the transmitted one (in dB). The instantaneous transmitted and received power are measured every 30 s from May to August 2008 and every 6 s since January 2009. No link data are available between September and December 2008. All power measurements are rounded at 1 dB.

Table 6.1: Length [km], frequency [GHz] and polarization of the considered telecommunication microwave links.

Link	Length	Frequency	Polarization
1	3.7	26	V
2	3.7	26	H
13	7.1	19	H/V
14	2.4	26	H/V

Ten rain events of various intensities and durations have been selected and are considered as representative of the local rainfall climatology. The dates, durations, rain amounts and maximum rain rates (as measured along the link paths by a near-by C-band weather radar) of these events are given in Table 6.2. The selected events represent about 90 h of rain and about 150 mm of rain amount.

Table 6.2: Date, duration [h], rain amount [mm] and maximum rain rate [$\mathrm{mm\,h^{-1}}$] of the considered events.

Event	Date	Duration	Amount	R_{max}
1	27-28 May 2008	8	40	60
2	22 Aug 2008	23	25	20
3	18 Jan 2009	10	12.5	10
4	27 Apr 2009	7	11.5	5.5
5	29 Apr 2009	3	10	18
6	25-26 May 2009	1.5	5	12
7	06-07 Jun 2009	2	1.5	2.5
8	07 Jun 2009	2	3.5	12.5
9	09-10 Jun 2009	16	18.5	12.5
10	10-11 Jun 2009	15	21.5	20

The performance of the proposed method is investigated by comparing the link estimates with independent radar rain-rate estimates provided by Météo France. The radar rain-rate maps are provided at a spatial resolution of 1×1 km^2 and temporal resolution of 5 min. They are derived from measurements of an operational C-band weather radar located in Trappes, about 20-30 km from the links. All rain rate maps are obtained by combining different scans at different elevation angles using the technique described in Tabary (2007) which corrects for the main sources of error (e.g., ground clutter, beam blocking, vertical variability and advection).

Since microwave links provide path-integrated measurements, the corresponding radar path-averaged rain-rate values are calculated by averaging the radar pixels crossed by the link beam with weights given by the length of the link in each pixel. A period is considered rainy if at least one radar pixel along the link-path has a strictly positive rain rate value.

6.4 Method

Figure 6.1 shows the raw attenuation measurements recorded on 22 August 2008 by link 14 (2.4 km and 26 GHz). Rainy periods are clearly characterized by a local increase of the signal attenuation (e.g., between 17:00 and 18:00 GMT). However, it is also important to note the significant (54-56 dB) and time-varying attenuation of the signal during dry periods (e.g. between 09:00 and 12:00 GMT).

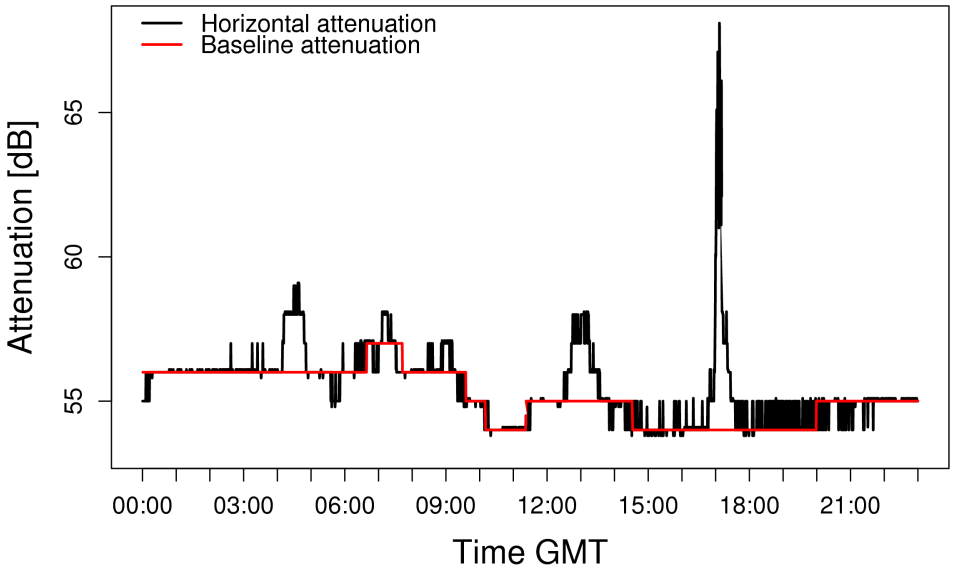

Figure 6.1: Attenuation measurements and attenuation baseline for link 14 (horizontal polarization) and event 2 (22 August 2008).

Let $A(t)$ be the total attenuation affecting a given link at time t. Because of the natural variability affecting the link measurements, $A(t)$ can be seen as a stochastic process. Furthermore, we suppose that $A(t)$ is the sum of two underlying stochastic processes $A_B(t)$ and $A_R(t)$ representing respectively the attenuation baseline and the attenuation caused by rainfall.

$$A(t) = \begin{cases} A_B(t) & \text{for every dry period,} \\ A_B(t) + A_R(t) & \text{for every rainy period.} \end{cases} \tag{6.1}$$

For a moving time window $W_t = [t - w, t]$ of duration $w > 0$, we define the following test statistic related to the local variability of the link signal :

$$S_{W_t} = \left[\frac{1}{N_W} \sum_{k \in W_t} \left(A(k) - \bar{A}_{W_t} \right)^2 \right]^{1/2} \tag{6.2}$$

where

$$\bar{A}_{W_t} = \frac{1}{N_W} \sum_{k \in W_t} A(k) \tag{6.3}$$

is the local average of the signal and N_W the number of observations in W_t. Note that since $A(t)$ is non-stationary (hence non-ergodic), there is no convergence between S_{W_t} and the local standard deviation $\sigma(t) = (\mathrm{Var}[A(t)])^{1/2}$. The proposed method to separate dry from rainy periods is based on the assumption that S_{W_t} is small during dry periods and large during rainy periods.

$$\mathrm{P}[S_{W_t} \leq \sigma \mid \text{dry}] \geq \mathrm{P}[S_{W_t} \leq \sigma \mid \text{rainy}] \quad \forall \sigma > 0 \tag{6.4}$$

where P denotes the probability. Hence the following decision rule provides a simple classification into dry and rainy periods:

$$\textbf{Decision rule:} \begin{cases} \text{rain if } S_{W_t} > \sigma_0 \\ \text{dry if } S_{W_t} \leq \sigma_0 \end{cases} \tag{6.5}$$

for a given threshold value σ_0 . The accuracy of the classification depends on the threshold value σ_0 but also on the underlying conditional distributions of S_{W_t} with respect to dry and rainy periods. In particular, light rainfall might not be distinguishable from dry periods because both can exhibit similar variability.

There are different methods to estimate σ_0 in practical applications. If rain gauges or radar data are available near the link, σ_0 can be calibrated using these measurements. If there are no external data, σ_0 must be estimated from the link data. Two different techniques are proposed here, depending on the amount of link data. If a large amounts

of data are available (typically several months of attenuation measurements), σ_0 can be estimated on a climatological basis, taking advantage of the fact that rainy periods represent only a small fraction r of all time periods. For Paris and its surroundings, this fraction is between 5 % and 15 % per year for a time resolution of 5 min to 60 min. Hence a reasonable threshold value is given by

$$\sigma_0 = q_{1-r}\{\, S_{W_t} \,\}, \tag{6.6}$$

where q_{1-r} denotes the $1 - r$ quantile. The fact that this method does not require any particular dry or rainy period for calibration is a clear advantage. Furthermore, a sensitivity analysis has been conducted and shows that the choice of r has limited influence on the classification procedure as long as it stays in a reasonable range (5 % to 15 %). Larger values for r reduce the number of type II errors (dry detections during rainy periods) but increase the number of false rain detections.

If few data are available, σ_0 can be estimated by considering the attenuation measurements from a collection of dry periods (typically several dry days). If D denotes these dry periods, a possible value for σ_0 is given by:

$$\sigma_0 = q_{99}\{\, S_{W_t} |\ t \in D \,\}, \tag{6.7}$$

where q_{99} denotes the 99 % quantile. This choice of σ_0 is preferred to the maximum because it limits the effect of outliers and link failures. Smaller quantiles can be used depending on the desired detection sensitivity. Combining several non-consecutive dry days is recommended because single or consecutive dry days may not adequately represent all the variability of the attenuation baseline.

The choice of the window size w is a delicate trade-off and is related to the natural variability of rainfall. Short time windows are desirable because they capture more details about the local variability of the signal. On the other hand, they are less powerful for detecting rain because rainfall (especially stratiform) can be nearly constant over short periods of time. Hence the attenuation caused by rain and the attenuation baseline cannot be separated effectively and the algorithm will miss a significant amount of rain. Larger time windows do not have this limitation because it is unlikely to have nearly constant rainfall over extended periods of time. On the other hand, large time windows are less powerful for detecting dry periods because w also determines the duration of the shortest dry period that can be identified using this algorithm. For Paris, time windows between 15 min and 35 min adequately capture the dynamics of the considered rainfall climatology.

Once w has been chosen and σ_0 estimated, the previously described decision rule can be applied to the entire data set to identify dry (D) and rainy time periods (R). The attenuation baseline $A_B(t)$ can then be estimated in real-time using the following algorithm:

(1) For each time index $t \in D$, set $A_B(t) = \bar{A}_{W_t}$.
(2) For each $t \in R$, set $A_B(t) = A_B(t-k)$ where k is the smallest value such that $t-k \in D$.

Finally, the baseline is used as a reference to remove all rain detections for which the total attenuation is below the baseline. This step is necessary since the proposed algorithm only considers the variability of the link signal without making any assumption on the absolute level of attenuation. Hence no distinction is made between strongly increasing/decreasing link signals above and below the baseline.

This defines the proposed method to identify dry and rainy periods and to fit the attenuation baseline using solely single-frequency and single-polarization microwave link data. Note that this approach does not take into account possible changes of the baseline during rainy periods. Such changes can be estimated by interpolating baseline estimations from two consecutive dry periods. However, this cannot be done in real-time anymore.

6.5 Performance evaluation

In this section, the proposed method is applied to 4 different microwave links and 10 different rain events between May 2008 and June 2009 (see Section 6.3). Independent radar rain-rate maps provided by Météo France are used to validate the classification into dry and rainy periods obtained with the links. These maps are used to separate dry and rainy periods and to quantify the total amount of rain identified by the links. Note that link and radar measurements cannot be in perfect agreement because they have different sampling volumes and altitudes. For comparison with radar estimates, a 5-min time step is considered rainy if at least one link measurement during this period is identified as rainy. Otherwise, the period is considered dry.

In this application, the test statistic S_{W_t} is computed using a 25-min moving time window which adequately captures the rainfall dynamics for the considered climatology. The corresponding rain detection thresholds σ_0 as described in Section 6.4 are 0.50 (link 1), 0.51 (link 2) and 0.48 (link 13 and 14) for the first method and 0.53 (link 1), 0.52 (link 2), 0.55 (link 13) and 0.50 (link 14) for the second method. For the first method, we assumed that rainy periods represent 10 % of all periods. For the second method, three different dry days (23 August 2008, 21 January 2009 and 16 April 2009) were used. Although both methods do not produce the exact same threshold values, the performance of the algorithm is similar and differences between the two methods did not exceed more than 10 % in the worst case. Hence, only the results for the first method are shown.

In order to quantify the performance of the algorithm, 5 different criteria are used: (1) the percentage of rain detections during rainy periods, (2) the percentage of dry detections during dry periods, (3) the percentage of type I errors (rain detections during

dry periods), (4) the percentage of type II errors (dry detections during rainy periods) and (5) the detected rain amount assuming the rain rate estimates from the links are identical to the radar rain-rate estimates. While the first 4 criteria define the contingency table between the link estimates and the radar data, criterion 5 is important for hydrological applications for which the total rain amount plays a major role.

Finally, the proposed algorithm is compared to two alternative baseline methods proposed by Leijnse et al. (2007c) and Rahimi et al. (2003). In the first method, the attenuation baseline is assumed constant and equal to the mode (the most observed value) of all attenuation measurements. All periods for which the attenuation exceeds the mode are considered rainy. The rest is considered dry. The second method was originally designed for dual-frequency links but can also be applied to bidirectional signals (each antenna acting both as a transmitter and a receiver). It is based on the cross-correlation over a 15 min time window between attenuation measurements from opposite signals. Time periods for which the cross-correlation is larger than 0.6 are considered rainy. Otherwise, they are considered dry (Rahimi et al., 2003; Goldshtein et al., 2009).

Because of its rather simplistic approach, the performance for the fixed "mode" baseline is low and its contingency table is not shown here. The median percentage of true rain detections is between 68 % and 83 % with worst case about 40 % (link 13, event 2). The percentage of true dry detections is between 67 % and 83 % with worst case down to 14 % (link 13, event 7). The type I errors are between 15 % and 33 % with worst case up to 83 % (link 13, event 7) and the type II errors are between 21 % and 42 % with worst case about 59 % (link 13, event 5). The median rain amount captured using this method is about 95 %.

Figure 6.2 shows the performance for criteria 1-4 for both the new method and the cross-correlation method. For the new method, the median percentage of correct rain detections for the four considered links is between 90 % and 95 % and the median percentage of correct dry detections is between 84 % and 88 %. The median type I errors (rain detection during dry periods) are between 11 % and 16 % and the median type II errors are between 4 % and 10 %.

For the cross-correlation method, the median percentage of true rain detections is between 76 % and 81 % and the median percentage of true dry detections is between 93 % and 97 %. The median type I errors (rain detection during dry periods) are between 1 % and 3 % and the median type II errors are between 19 % and 24 %.

Finally, the proposed algorithm detected about 93 % of the total rainfall amount compared to 83 % for the cross-correlation method.

It is important to note that the comparison between the correlation-based method and the proposed one is done at different error rates. Therefore, it is not clear which method is more accurate. To address this question, a cost function is introduced to compare

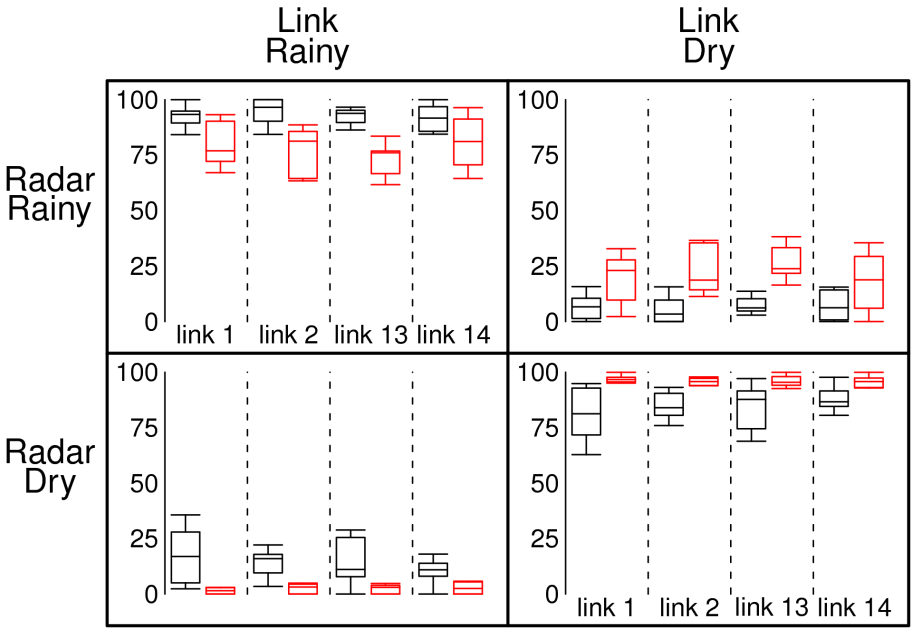

Figure 6.2: Performance for all 4 links and all 10 rain events. The results of the proposed method are represented by black box-plots. Red box-plots represent the performance for the cross-correlation method. Each box-plot represents the 10 %, 25 %, 50 %, 75 % and 90 % quantiles. All values are in percentage.

the performances of the two methods. For hydrological applications, a reasonable cost function is given by

$$C(x) = \frac{R_{tot} - R_{capt}}{R_{tot}} + \frac{F_p}{N_{dry}} + \alpha \frac{F_n}{N_{rainy}} \tag{6.8}$$

where x represents the detection threshold, F_p and F_n denote respectively the number of false positives and false negatives, R_{capt} and R_{tot} the captured and total amount of rainfall, N_{dry} and N_{rainy} the number of dry and rainy periods and $\alpha > 0$ is the relative weight attributed to type II errors (with respect to type I errors). For the baseline estimation problem, α is usually larger than one because type II errors (dry detection during rainy periods) have more serious consequences than type I errors, as pointed out by Rahimi et al. (2003). Figure 6.3 shows the cost function for all links and methods ($\alpha = 2$). It shows that the new method performs better than the cross-correlation method for all links and all type II error rates smaller than 15 %, except for link 1 which suffers from larger dry-air attenuation variability. This is likely due to the transmission-reception chain of link 1 as the proposed method gives much better results for link 2 which has the exact same path. Similar results are obtained for any value of $\alpha \geq 1$.

The major problem with the cross-correlation method is that it produces large type

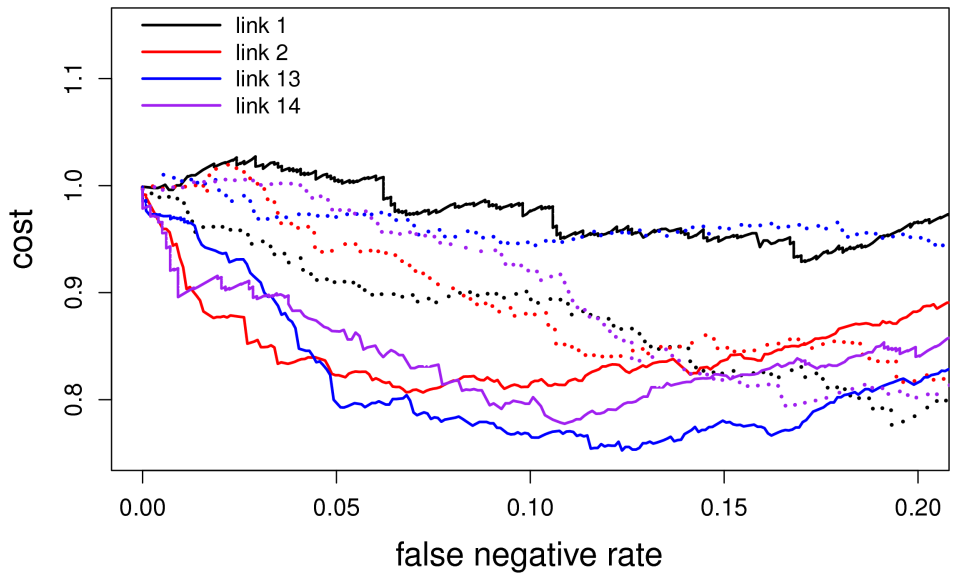

Figure 6.3: Cost function (for $\alpha = 2$) for all links and both methods. The solid lines correspond to the proposed method and the dotted lines to the cross-correlation method.

II error rates, even for low threshold values. This is mainly due to the fact that the correlation coefficient often drops to zero, even during strong rainfall. The new method based on the local variability measure S_{W_t} produces less type II errors and strong rainfall is never misclassified as being dry. The value of S_{W_t} increases rapidly as soon as it starts to rain but takes some time to decrease at the end of each event. Consequently, false rain detections often occur at the end or between two rain peaks.

The performance of the algorithm with respect to the average rain rate $\bar{\mathrm{R}}$ and maximum rainfall intensity $\mathrm{R_{max}}$ is investigated in Table 6.3. In this classification, events at the top of the table can mostly be associated with convective precipitation whereas events at the bottom of the table mostly represent stratiform precipitation.

The results show that the algorithm performs better during strong rainfall and worse during light rainfall. In particular, the algorithm produces less type II errors (dry detections during rain) for strong rainfall (about 2 % type II errors) than for light rain (about 13 % type II errors). As a consequence, the amount of rain captured by the algorithm is higher during intense precipitation (about 98 %) than during light rainfall (about 84 %). Table 6.3 concerns only link 14, but similar results were obtained for the other links (not shown for more conciseness). Note that further studies involving more links and more events may be necessary to quantify precisely the differences in performance for different types of precipitation, links and climatologies. However, the results obtained in this section clearly demonstrate the feasibility of the method and its improvements on other techniques employed so far.

Table 6.3: Detailed performance for link 14.

Event	$\bar{R}$ [mm h^{-1}]	R_{max} [mm h^{-1}]	type I errors	type II errors	Rain amount captured
1	10.7	60	0 %	1 %	97.4 %
5	3.5	18	17 %	3 %	94.8 %
10	3.0	20	11 %	4 %	99.5 %
3	3.0	10	14 %	0 %	98.0 %
6	2.8	12	11 %	0 %	98.1 %
2	2.2	20	0 %	17 %	89.3 %
9	1.9	12.5	26 %	9 %	94.6 %
8	1.6	12.5	13 %	15 %	88.5 %
4	1.8	5.5	0 %	13 %	65.5 %
7	0.8	2.5	8 %	13 %	83.7 %

The limitations of the proposed algorithm are illustrated in Figure 6.4 which shows the attenuation signal recorded by link 14 for event 4. This event is particular because it is characterized by a long period of nearly uniform rainfall (17:45-18:15 GMT) preceding a strong rainfall peak (18:30-19:00 GMT). After 25 min of nearly constant rainfall (at about 18:00 GMT), S_{W_t} drops below σ_0 and the corresponding period is declared dry (type II error).

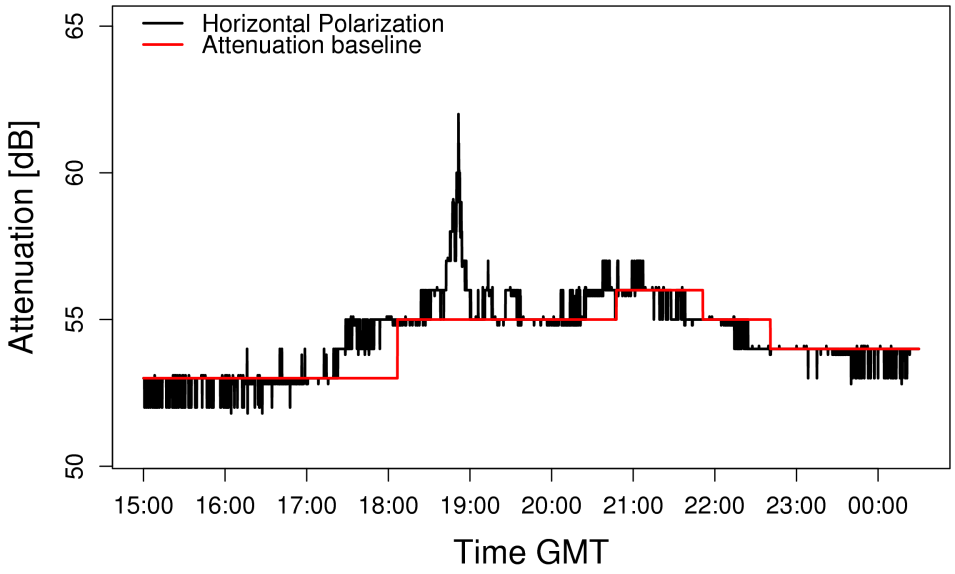

Figure 6.4: Attenuation measurements and attenuation baseline for link 14 (horizontal polarization) during event 4 (27 Apr 2009).

As a consequence, an erroneous baseline is fitted at 55 dB, about 2 dBs above the correct baseline. As a result, about 35 % of the total rain amount is lost during the rest of the event. This is a particular case where a 25-min moving time-window does not adequately capture all the dynamics of the considered rainfall. This is confirmed by the fact that better results are obtained for larger time windows. A 60-min time window captures 82 % of the total rainfall amount and a 90-min time window captures more than 90 % of the total rainfall with less than 5 % type I/II errors. This shows that the performance of the algorithm could be improved by choosing different time windows for each event or by adapting the time window to seasonal variations in rainfall dynamics.

6.6 Summary and conclusions

Telecommunication microwave links can be used for rainfall estimation. A critical issue in this procedure is the ability to separate the attenuation during dry periods (called the attenuation baseline) from the attenuation due to rainfall. In particular, this requires to identify and separate dry from rainy periods.

In this chapter, a method to identify dry and rainy periods and to fit a time-varying attenuation baseline has been presented. The proposed method can be applied in real-time on a single-polarization single-frequency link and does not require any additional data for calibration. Rain identification is performed by analyzing the local variability of the link signal. The attenuation baseline is estimated through the classification into dry and wet periods. The new method detected about 92 % of all rainy periods, 86 % of all dry periods and 93 % of the total rain amount on average for 10 rain events.

The proposed algorithm provides an interesting and efficient alternative to existing techniques like the “mode” and the cross-correlation method. It is easy to calibrate and can be employed in regions were rain gauge or radar data are sparse.

The main limitation of the proposed method is its inability to separate light rain from dry-air attenuation and to identify changes in the attenuation baseline during rainy periods. It may not be suitable for a limited number of links with large dry-air attenuation variability. Future work will mainly focus on the analysis of the intermittency as seen by microwave links and on daily and seasonal variations in the link signal.

7 Using Markov switching models to infer dry and rainy periods from telecommunication microwave link signals [1]

7.1 Summary

A Markov switching algorithm is introduced to classify attenuation measurements from telecommunication microwave links into dry and rainy periods. It is based on a simple state-space model and has the advantage of not relying on empirically estimated threshold parameters. The algorithm is applied to data collected using a new and original experimental set-up in the vicinity of Zürich, Switzerland. The false dry and false rain detection rates of the algorithm are evaluated and compared to 3 other algorithms from the literature. The results show that, on average, the Markov switching model outperforms the other algorithms. It is also shown that the classification performance can be further improved if redundant information from multiple channels is used.

1. This chapter is a slightly modified version of a manuscript by Wang, Z., M. Schleiss, J. Jaffrain, A. Berne and J. Rieckermann: Using Markov switching models to infer dry and rainy periods from telecommunication microwave link signals, submitted to *Atmos. Meas. Tech.*, on the 16$^{\text{th}}$ of December 2011.

7.2 Introduction

Precipitation is an important component of the Earth's water cycle and needs to be accurately measured. So far, several techniques have been proposed to measure rainfall with different spatial and temporal resolutions, ranging from traditional point measurements from rain gauges to observations from weather radars and satellites. Not surprisingly, each of these techniques has its advantages but also its limitations (Sevruk, 1999; Upton et al., 2005; Germann et al., 2006).

Recently, microwave links (MWL), which are commonly used in telecommunication networks for wireless data transmission, have been suggested as a novel tool to monitor rainfall in urban areas (Messer et al., 2006; Leijnse et al., 2007c). The main idea behind this technique is to relate the rain-induced signal attenuation to the path-averaged rain-rate along the considered link. The potential of this technique has been demonstrated using microwave links specifically designed for rainfall estimation (Ruf et al., 1996; Rahimi et al., 2003; Holt et al., 2003; Upton et al., 2005; Krämer et al., 2005) and commercial microwave links operated by telecommunication companies (Messer et al., 2006; Zinevich et al., 2009). Note that, in addition to estimating rain rates, MWL can also be used to measure evaporation (Leijnse et al., 2007a) and water vapor (David et al., 2009). In fact, MWL nicely complement traditional rainfall sensors because they provide rain rate measurements (near the ground level) at an intermediate scale between point measurements from rain gauges and weather radars with sampling volumes up to several km^3. The fact that MWL networks can be very dense can also be used to improve rain rate estimates using spatial interpolation techniques (Zinevich et al., 2008).

A very important issue that needs to be addressed prior to rainfall estimation using MWL is the so-called baseline estimation problem (Rahimi et al., 2003; Leijnse et al., 2007c). It consists in identifying and separating the attenuation occurring during dry periods from the rain-induced attenuation (which is the quantity of interest in most applications). Depending on the link characteristics, this problem can be very difficult (Upton et al., 2005). Dry-weather signal attenuations can exhibit significant variability caused, for example by changes in water vapor, wind effects on the antennas, birds or insects crossing the beam, losses during transmission or reception, interferences, wet-antenna and multi-path effects (Zinevich et al., 2010). Moreover, attenuation measurements are often quantized, which introduces additional variability in the process. It is only after the attenuation baseline has been properly estimated for each time step that the corresponding path-averaged rain-rate can be retrieved. A very important step in this procedure is the ability to identify the dry and rainy periods using solely the measurements of the MWL. The measured attenuation levels during the dry periods can then be used to better estimate the attenuation baseline during rainy periods.

So far, various techniques have been suggested to solve this identification problem (see Section 7.3 for a detailed description). The simplest of them uses a global attenuation

threshold, i.e., all the periods for which the path-integrated attenuation (PIA hereinafter) is above a given threshold are considered rainy and vice versa (Leijnse et al., 2007b). A slightly more sophisticated procedure was suggested by Schleiss and Berne (2010) who proposed to use a threshold on the temporal variability of the PIA (see Chapter 6). A more complex algorithm by Reller et al. (2011) investigates the possibility to identify dry and wet periods using a Bayesian approach based on Factor Graphs (Loeliger et al., 2010). The underlying idea is that the variations of the baseline are learned during dry weather and propagated to wet periods. Although it is more sophisticated and flexible, the algorithm also requires the choice of a subjectively estimated threshold for classification. Finally, special algorithms have been developed for situations where signals from dual-frequency MWLs or simultaneous PIA's from two communication channels are available. For example, Holt et al. (2003) and Rahimi et al. (2003) proposed a method based on the assumption that the correlation between the attenuations from two different frequencies tend to be higher during rainy periods. Such methods, however, also rely on empirically adjusted thresholds for the classification, which is not necessarily optimal.

In this chapter, a new classification algorithm based on Markov switching models is introduced. It is based on a simple state-space model and has the advantage of not relying on any empirically estimated threshold parameters. Also, the proposed algorithm can be easily generalized to multivariate inputs, i.e., inputs from different channels or frequencies. A real-world application of the algorithm (see Section 7.5) shows that it performs better than other existing techniques and that its performance can be improved if multiple channel inputs are considered. The proposed algorithm thus improves the signal processing of MWL data and helps estimating better attenuation baselines required for accurate rainfall retrieval.

This chapter is structured as follows: Section 7.3 describes some of the existing classification methods and introduces the Markov switching model. In Section 7.4, the experimental set-up used to quantify the performances of the classification algorithms is described. Section 7.5 evaluates and compares the performances of the different algorithms for two very different datasets. Possible improvements of the algorithm are then discussed in Section 7.6. The conclusions are given in Section 7.7.

7.3 Methods

This section briefly describes some of the existing methods proposed in the literature to identify dry and rainy periods using single channel MWL attenuation measurements. Then, a new classification method based on Markov switching models is introduced and discussed in detail.

7.3.1 Existing algorithms

Three popular classification methods have been chosen: the simple threshold method (ST), the moving window method (MW) and the Factor Graph (FG). The simple threshold algorithm (Leijnse et al., 2007b) is straightforward and computationally efficient. It uses a global threshold on the path-integrated attenuation to distinguish between dry and rainy periods. Each time period for which the PIA is above the threshold is classified as rainy, and vice versa.

$$\text{Decision rule for ST:} \begin{cases} \text{rainy if } A_t > a_0 \\ \text{dry if } A_t \leq a_0 \end{cases} \tag{7.1}$$

where A_t [dB] denotes the path-integrated attenuation at time t and a_0 [dB] is a given threshold value. The method has shown to produce good results in practical applications and can be applied in real-time. Finding the optimal detection threshold a_0 is, however, difficult. Moreover, the performance (in terms of false dry and rain detections) of this algorithm can be very sensitive to the value of the threshold. Finally, this method is only appropriate for datasets for which the dry-weather attenuation is more or less constant. This is not always the case as can be seen in the right panel of Figure 7.2. In some situations, the dry-weather attenuation exhibits clear daily cycles and a strong temporal drift in the PIA, possibly due to changes in temperature between day and night and hardware instabilities. Obviously, the simple threshold is not appropriate for such types of signals and alternative classification methods have been suggested.

A slightly more complex approach (hereinafter referred to as the moving window algorithm) which is better suited for non-stationary dry-weather attenuations has been proposed in Chapter 6. The method is based on the assumption that the temporal variability of the PIA is small and bounded during dry weather. On the other hand, rainy periods are characterized by larger signal fluctuations. Each time period is classified according to the following decision rule:

$$\text{Decision rule for MW:} \begin{cases} \text{rainy if } S_{W_t} > \sigma_0 \\ \text{dry if } S_{W_t} \leq \sigma_0 \end{cases} \tag{7.2}$$

where S_{W_t} [dB] represents the local (temporal) variability of the signal attenuation for a moving window $[t-w, t]$ and σ_0 [dB] is a rain detection threshold estimated using one of the two approaches described in Section 6.4. The moving window algorithm is also computationally efficient and can be applied in real-time to non-stationary time series of attenuations. However, finding the optimal detection threshold σ_0 can be very difficult without appropriate calibration data over extended periods of time. Moreover, one of the main disadvantages of the moving window algorithm is its inability to separate light rain from dry periods because both signals exhibit similar variability.

The Factor Graph algorithm proposed by Reller et al. (2011) can also be applied to

non-stationary MWL signals but does not require large datasets for model calibration. A Factor Graph is a particular type of graphical model, with applications in Bayesian inference, which computes marginal distributions through the sum-product message passing algorithm (Kschischang et al., 2001). More specifically, the Factor Graph algorithm models the attenuation baseline during dry weather using a line model whose parameters can vary slowly over time together with periodicity constraints. In this, it assumes a smoothly varying baseline and, where the signal exceeds a certain threshold, the algorithm identifies that the system enters another state. The Factor Graph algorithm possesses several advantages, as it can deal with irregular time series and not only identifies dry and rainy periods, but simultaneously estimates the baseline and thus delivers the rain-induced attenuation. However, it also relies on several tuning parameters that need to be estimated subjectively prior to the classification into dry and rainy periods.

In the following, a new algorithm for the identification of dry and rainy periods based on MWL attenuation measurements is introduced. It uses Markov switching models to estimate the state of the system (i.e., dry or rainy).

7.3.2 Univariate Markov switching model (MSU)

A Markov switching model combines dynamic linear system behavior with a Markov process, which models the transitions between different states. It belongs, similarly to the Factor Graph, to a very general class of so-called state-space models. Such models are commonly used to model a change in behavior with respect to different regimes. The regimes themselves can be related to certain events, often stochastic, such as a financial crisis or changes in government policy. Practical applications of such models can be found (among others) in the fields of Economics (Hamilton, 1989) and Physics (Yue and Han, 2005; Metzner et al., 2007). Markov switching models have also been used in weather generators to model rainfall patterns (Weiss, 1964).

For simplicity, the details of the algorithm are only given for the univariate case, i.e., a single channel input. The multivariate case is briefly described at the end of this section. For more details on Markov switching models, the reader is referred to Hamilton (1989, 1990) and Kim (1994). Note that Rayitsfeld et al. (2011) proposed a similar approach based on a hidden Markov model with a slightly different implementation. They did, however not compare their method with previously proposed classification techniques.

The underlying assumption of the Markov switching algorithm is that the magnitude and the variability of the PIA are fundamentally different during dry and rainy periods. During dry periods, the PIA mildly fluctuates around a given value, while for rainy periods it is much larger and variable. This additional variability is caused by the scattering and absorption of the transmitted signal by the raindrops along the path of the link. Hence, it should be possible to identify two fundamentally different states of the

system (dry/rainy) from the different behavior of the PIA. For example, the following, very simple model can be used to describe the data:

$$A_t = \begin{cases} \mu_0 + \varepsilon_0 & \text{for every dry period} \\ \mu_1 + \varepsilon_1 & \text{for every rainy period} \end{cases} \tag{7.3}$$

where A_t [dB] represents the path-integrated attenuation at time t, μ_0 [dB] and μ_1 [dB] represent the average value of the attenuation during dry and rainy periods. The noise terms ε_0 [dB] and ε_1 [dB] are assumed to be independent Gaussian random variables with zero mean and standard deviations given by σ_0 [dB] and σ_1 [dB]. The transitions between the dry and the rainy periods are modeled using a stationary hidden random variable $S_t \in \{0,1\}$ where

$$S_t = \begin{cases} 0 & \text{for every dry period} \\ 1 & \text{for every rainy period} \end{cases} \tag{7.4}$$

The unconditional probability of the system being in the dry state is denoted by $p_0 = \mathbf{Pr}(S_t = 0) = 1 - p_1$. Combining Equations 7.3 and 7.4, it is possible to write A_t using a single expression given by

$$A_t = \mu_{S_t} + \varepsilon_{S_t} \tag{7.5}$$

with 5 model parameters $\mathbf{\Theta} = (\mu_0, \mu_1, \sigma_0, \sigma_1, p_0)$. The maximum likelihood technique is then used to infer the optimal model parameters for a given set of observations $\{A_t = a_t\}$:

$$\hat{\Theta} = \text{argmax}\,\{l(\Theta)\} \tag{7.6}$$

where the log-likelihood function $l(\Theta)$ is given by

$$l(\Theta) = \sum_t \log\left[\sum_{k=0}^{1} f_k(a_t, \Theta) p_k(a_t, \Theta)\right] \tag{7.7}$$

with

$$f_k(x, \Theta) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{(x-\mu_k)^2}{2\sigma_k^2}\right) \qquad k = 0, 1$$

and

$$p_k(x, \Theta) = \frac{p_k f_k(x, \Theta)}{p_0 f_0(x, \Theta) + (1 - p_0) f_1(x, \Theta)} \qquad k = 0, 1$$

where $f_k(x, \Theta)$ denotes the probability density for a given state k and $p_k(x, \Theta)$ the associated state probability (for a given set of model parameters Θ). The maximization of $l(\Theta)$ is performed using a standard Newton-type algorithm. In order to be valid, the

solution must satisfy some simple conditions. Specifically, one must have $1 > p_0 > 0$, $\sigma_1 > \sigma_0 > 0$ and $\mu_1 > \mu_0 > 0$. Once $\hat{\Theta}$ has been estimated, the classification into dry and rainy periods can be easily derived from the estimated state probabilities $p_0(a_t, \hat{\Theta}) = 1 - p_1(a_t, \hat{\Theta})$.

$$\hat{s_t} = \begin{cases} 0 & \text{if } \; p_0(a_t, \hat{\Theta}) > \frac{1}{2} \\ 1 & \text{else} \end{cases} \tag{7.8}$$

Note that it is also possible to choose another threshold depending on the relative cost associated to each of the classification errors. One of the advantages of the Markov switching model is that it can be easily generalized to include multivariate inputs from different channels or frequencies.

7.3.3 Multivariate Markov switching model (MSM)

Telecommunication microwave links are usually operated using multiple channels such as two directions, frequencies or polarizations. This redundant information can be used to improve the classification performance.

In the multivariate case with N channels, the attenuation at time step t is given by a vector $A_t = (A_t^{(1)}, ..., A_t^{(N)})$ where

$$A_t^{(j)} = \mu_{S_t}^{(j)} + \varepsilon_{S_t}^{(j)} \qquad \forall\, j = 1, ..., N. \tag{7.9}$$

Note that the vector of model parameters is now significantly longer and given by $\Theta = (\mu_0^{(1:N)}, \mu_1^{(1:N)}, \sigma_0^{(1:N)}, \sigma_1^{(1:N)}, p_0)$, that is, $4N + 1$ variables to estimate. The major difference with respect to the univariate case concerns the difficulty to estimate the joint densities $f_0(A_t, \Theta)$ and $f_1(A_t, \Theta)$, although significant simplifications occur if the channels are assumed independent. While this is certainly not the case for rainy periods, it is, at least, reasonable during dry periods (which usually represent the majority of all the periods). In the absence of any further information, a pragmatic solution therefore consists in assuming that all channels are independent and that the log-likelihood function is given by

$$l(\Theta) = \sum_t \log \left[\sum_{k=0}^{1} \prod_{j=1}^{N} f_k^{(j)}(a_t^{(j)}, \Theta) p_k(a_t, \Theta) \right] \tag{7.10}$$

where

$$f_k^{(j)}(x, \Theta) = \frac{1}{\sqrt{2\pi(\sigma_k^{(j)})^2}} \exp\left(-\frac{(x - \mu_k^{(j)})^2}{2(\sigma_k^{(j)})^2} \right) \qquad k = 0, 1$$

and

$$p_k(a_t, \Theta) = \frac{p_k \prod_{j=1}^{N} f_k^{(j)}(a_t^{(j)}, \Theta)}{\sum_{i=0}^{1} p_i \prod_{j=1}^{N} f_i^{(j)}(a_t^{(j)}, \Theta)}$$

Maximizing $l(\Theta)$ yields, similarly to the univariate case, the maximum likelihood estimate $\hat{\Theta}$. The classification into dry and rainy periods can then be derived from the estimated state probabilities $p_k(a_t, \hat{\Theta})$. Possible extensions to correlated attenuation values, at least during rainy periods, and more general expressions for the joint density $f_1(A_t, \Theta)$ will not be discussed. For simplicity, only the independent case is presented in Section 7.5.

7.4 Experimental set-up

The experimental site is located in Dübendorf, near the city of Zürich, Switzerland (see Figure 7.1). It consists of a 1.85 km long commercial dual-polarization microwave link, 5 disdrometers and 3 rain gauges placed approximatively at equal distances along the path of the link. The dataset is complemented by climatic and meteorological data from two weather stations. The experiment is designed to investigate different aspects of rainfall monitoring using microwave links in the context of a humid continental climate, such as the retrieval of path-averaged rain-rates, the influence of the drop size distribution, the characteristics of dry-weather attenuation and wet-antenna effects. In particular, the horizontally and vertically polarized signals could be used to retrieve the effective drop size distribution along the link path.

7.4.1 Microwave link

The installed microwave link is an "Ericsson Mini-link TN ETSI", a widely used system in commercial telecommunication applications. The MWL is operated at about 38 GHz in a dual-polarization set-up with the specific characteristics given in Table 7.1. For more redundancy, the link provides measurements on 4 different channels (2 polarizations and 2 directions). In its original configuration, the link only records the transmitted and received powers every 15 min. This is clearly not sufficient for accurate rainfall monitoring at scales relevant for modern hydrological and meteorological applications. Therefore, a stand-alone data logging application using the SNMP protocol has been developed and implemented to record the power measurements in much shorter intervals. For the purposes of this project, a 4 s temporal resolution has been chosen.

The experiment started in March 2011 and is continuing into the first half of 2012. It is divided into two parts. During the first part of the project, i.e., until the 10th of October

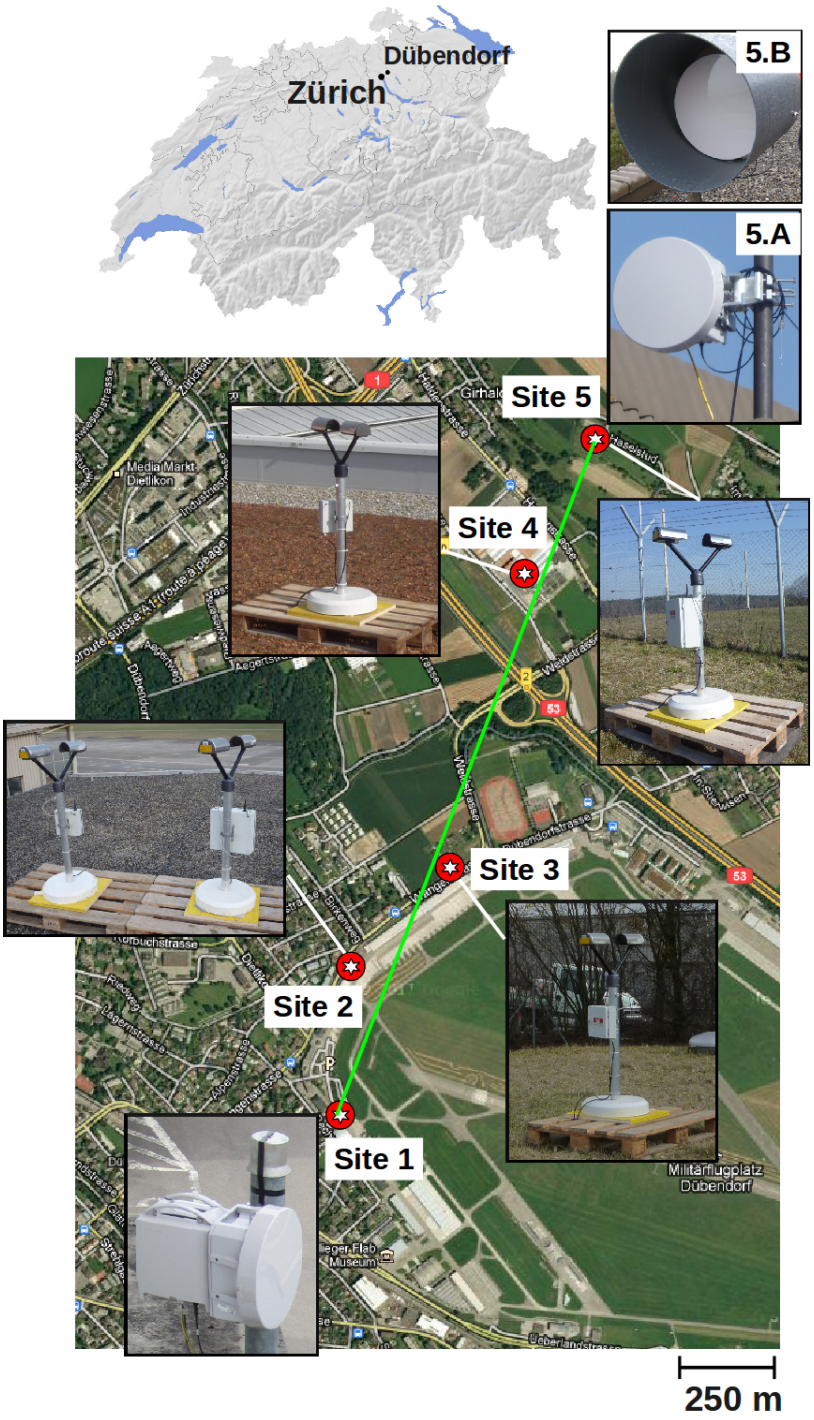

Figure 7.1: Experimental set-up deployed in Dübendorf, Switzerland. The Disdrometers are located at sites 2 (2 collocated stations), 3, 4 and 5. The rain gauges at sites 2, 4 and 5. Pictures 5.A and 5.B show the MWL (at site 5) with (respectively without) the rain shields.

Table 7.1: Longitude, latitude, height and frequencies of the installed microwave link.

Parameters	Site 1: Dübendorf	Site 5: Wangen
Longitude	8°37'43.10" E	8°38'16.26 E
Latitude	47°24'4.80" N	47°25'0.25" N
Height	436 m AMSL	486 m AMSL
Freq. (horizontal)	38'657.5 MHz	37'397.5 MHz
Freq. (vertical)	38'650.5 MHz	37'390.5 MHz
Length	1.85 km	

2011, the antennas of the link were fully exposed to the rain. Consequently, during rainy periods, a thin film of water was formed on the surface of the antennas, causing additional attenuation in the order of several dB (Kharadly and Ross, 2001; Leijnse et al., 2008). Preliminary data analysis suggest that the antennas sometimes remained wet for several hours after the rain had stopped. In the second part of the experiment, i.e., after the 10th of October 2011, the antennas were shielded from rain using plastic shields specifically designed for this experiment (see 5.A in Figure 7.1). Visual inspection of the antennas proved that these shields effectively protect the surface of the antennas, even during strong rainfall and moderate wind speeds.

In addition to wet-antenna effects, the experimental set-up also revealed unexpected fluctuations in the transmitted power levels. According to the manufacturer, the received power is measured with an accuracy of 0.1 dB and the transmitted power with an accuracy of 1 dB. Additional measurements of the transmitted power using a power meter showed that the transmitted power was accurate within a range of approximatively 0.35 dB over a period of 11 days, for temperatures between 7° C and 23° C and relative humidities between 37 % and 100 %. This is confirmed by independent measurements collected in the laboratory, with (more or less) constant temperatures and humidities and for which the uncertainty on the transmitted power was found to be 0.3 dB.

7.4.2 Disdrometers and rain gauges

5 optical disdrometers of type Parsivel (1st generation, manufactured by OTT) have been deployed at 4 different sampling locations (sites 2-5) along the 1.85 km path of the link (see Figure 7.1). For more details on the principle of these optical disdrometers, see Löffler-Mang and Joss (2000). All the disdrometers are designed to be autonomous in terms of power supply and data transmission (Jaffrain et al., 2011). They provide measurements of particle sizes and velocities at a 30-s temporal resolution. Note that sampling point 2 is equipped with two collocated disdrometers in order to quantify the measurement uncertainty associated with Parsivel disdrometers (Jaffrain and Berne, 2011). The 4 sampling locations have been chosen as a trade-off between a regular distribution of the instruments, the distance to the path of the link, line of sight for data

transmission between the different instruments and minimum probability of disturbance and vandalism.

In addition to the 5 disdrometers, 3 tipping-bucket rain gauges from Précis Mécanique (model 3029) have been deployed at sampling locations 2, 4 and 5. The tipping-bucket rain gauges have a catching area of 400 cm^2 and are connected to data loggers that record the tipping time with an accuracy of 0.1 s. One tip corresponds to 0.1 mm of rain. Note that the 3 rain gauges are not transmitting the data in real time. The collected data are used to check the calibration of the disdrometers and to identify possible biases between the sensors.

7.4.3 Additional data

The rainfall measurement network is complemented by operational radar data provided by MeteoSwiss. Processed maps of rain rate and radar reflectivities are available at a spatial resolution of 1×1 km^2 and a temporal resolution of 5 min. In addition, meteorological and climatic data (e.g., temperature, relative humidity, pressure, wind speed and wind direction) are collected using a MIDAS IV weather station (manufactured by Vaisala) located at the airport in Dübendorf. The MIDAS IV system collects data from two sensors situated at both ends of the runway. The temporal resolution depends on the considered parameter and can vary between 3 and 60 s.

7.4.4 Originality

Several other studies involving simultaneous measurements of microwave links, rain gauges, disdrometers and weather radar can be found in the literature. Rincon and Lang (2002) proposed a method to estimate the drop size distribution from the measurements of a dedicated 2.3 km, dual-frequency research link and validated their results using 6 rain gauges and a single 2D video disdrometer placed along the path of the link. Rahimi et al. (2003) used a 23.3 km, dual-frequency research link with 22 rain gauges and radar data. However, only 4 or 5 rain gauges were reasonably close to the considered link. More recently, Leijnse et al. (2007c) used a 4.89 km, 27 GHz research link with 6 rain gauges placed along the path of the link. Finally, Zinevich et al. (2010) compared the rain estimates from 23 commercial microwave links with 5 nearby rain gauges. The experimental set-up presented above is original because it combines attenuation measurements from a dual-polarization commercial microwave link with a sufficiently dense network of disdrometers to accurately estimate the path-averaged DSD. This provides a platform to develop and validate new methods for rainfall retrieval using MWL and to evaluate their respective performances as outlined above. In particular, it can be used to investigate if the redundancy between the different channels and polarizations can be used to improve the rain rate estimates. Furthermore, it might also be of interest

to radio engineers concerned with better predictions of rain-induced attenuation and MWL simulation methods (Paulson, 2002; Callaghan et al., 2008).

7.4.5 Selected datasets

Two datasets have been selected from the experimental observational record to evaluate the performance of the algorithms described in Section 7.3 under fundamentally different conditions. A visual illustration of these datasets is given in Figure 7.2. Note that for a better visibility, the attenuation measurements are only shown for one channel. The first dataset covers the period between the 17$^{\text{th}}$ of May 2011 and the 12$^{\text{th}}$ of June 2011 and is representative of a (more or less) constant dry-weather attenuation baseline (hereinafter referred to as the stationary case). This period is also characterized by small variations in the PIA during dry weather. The second dataset covers the period between the 17$^{\text{th}}$ of March 2011 and the 26th of April 2011 and illustrates a very different behavior (hereinafter referred to as the non-stationary case). This period is characterized by a highly-variable attenuation baseline with a strong temporal drift and daily cycles in the PIA, due to changes in temperature and humidity. A preliminary analysis of the current observational record suggests that the non-stationary cases represent a non-negligible amount (about 10-20 %) of all the time periods and must therefore be considered carefully.

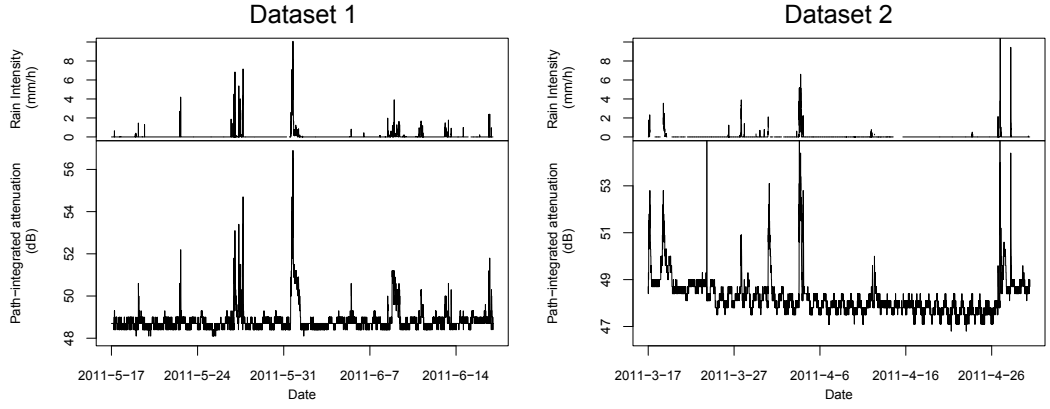

Figure 7.2: Rain rates [mm h^{-1}] and path-integrated attenuations [dB] for dataset 1 (stationary) and 2 (non-stationary). For better illustration, only the attenuation of channels 1 (dataset 1), and 4 (dataset 2) are shown. The time is given in UTC.

7.5 Results and model comparison

7.5.1 False rain and dry detections

The performances of the algorithms described in Section 7.3 are evaluated and compared using two criteria:

$$\text{type I error: } \frac{\#\text{dry periods classified as rainy}}{\#\text{dry periods}}$$

$$\text{type II error: } \frac{\#\text{rainy periods classified as dry}}{\#\text{rainy periods}}$$

In other words, type I errors correspond to false rain detections and type II errors to false dry detections. A perfect classification algorithm has 0 type I error and 0 type II error. In practical applications, however, both types of errors are usually competing against each other, i.e., if the type I error decreases, the type II increases and vice versa. Finding an optimal trade-off between both errors is difficult and depends on the underlying application and the cost associated to each type of error. However, this is beyond the scope of this study.

For comparison purposes, it is assumed that the path-averaged rain-rate measured by the 5 disdrometers along the path of the link (see Section 7.4) is representative of the "true" weather state. If the path-averaged rain rate is greater than zero, the period is considered rainy. Otherwise, it is supposed to be dry. In order to analyze the sensitivity of the results with respect to this rain-detection threshold, a slightly higher rain detection threshold of 0.1 $\mathrm{mm\,h^{-1}}$ is also considered. All periods for which the path-averaged rain rate is smaller than 0.1 $\mathrm{mm\,h^{-1}}$ are considered dry and vice versa. The value of 0.1 $\mathrm{mm\,h^{-1}}$ was chosen as a threshold because it approximatively corresponds to the hardware-induced measurement uncertainty of 0.1 dB in the path-integrated attenuation (ITU-R P.838-3, 2005). In other words, rainy periods with rain rates smaller than 0.1 $\mathrm{mm\,h^{-1}}$ cannot be distinguished from dry periods because of the uncertainty on the power measurements. Finally, note that because the disdrometer data are provided at a 30-s temporal resolution, the corresponding MWL data (at a 4-s temporal resolution) are averaged at 30-s prior to the analysis. Periods for which one of the instruments was not working are not considered for the comparison.

7.5.2 Stationary dry-weather attenuation baseline

The results for the first dataset (stationary case) are shown in Table 7.2. For better illustration of important details, a small subset of dataset 1 (a 5-days period between the 8^{th} of June 2011 and the 12^{th} of June 2011) is plotted in Figure 7.3.

Table 7.2: Classification performances (in percentages) for the simple threshold (ST), the moving window (MW), the Factor Graph (FG), the univariate Markov switching (MSU) and the multivariate Markov switching (MSM) algorithms for dataset 1 (stationary case). For the univariate algorithms, the value given in the table corresponds to the average classification performance for all 4 channels. In parentheses the associated standard deviation. Note that no model parameters could be fitted for the MSU algorithm on channel 2.

Models	rain detection threshold 0 mm h^{-1}		rain detection threshold 0.1 mm h^{-1}	
	type I error	type II error	type I error	type II error
ST	2.74 (0.88)	23.07 (4.44)	4.52 (0.98)	13.06 (4.26)
MW	12.35 (0.46)	39.04 (2.44)	14.43 (0.14)	11.84 (1.78)
FG	12.00 (2.13)	38.78 (5.75)	12.80 (2.04)	27.07 (6.06)
MSU	2.11 (0.10)	23.32 (2.50)	3.89 (0.18)	13.28 (2.53)
MSM	2.46	21.97	4.27	11.82

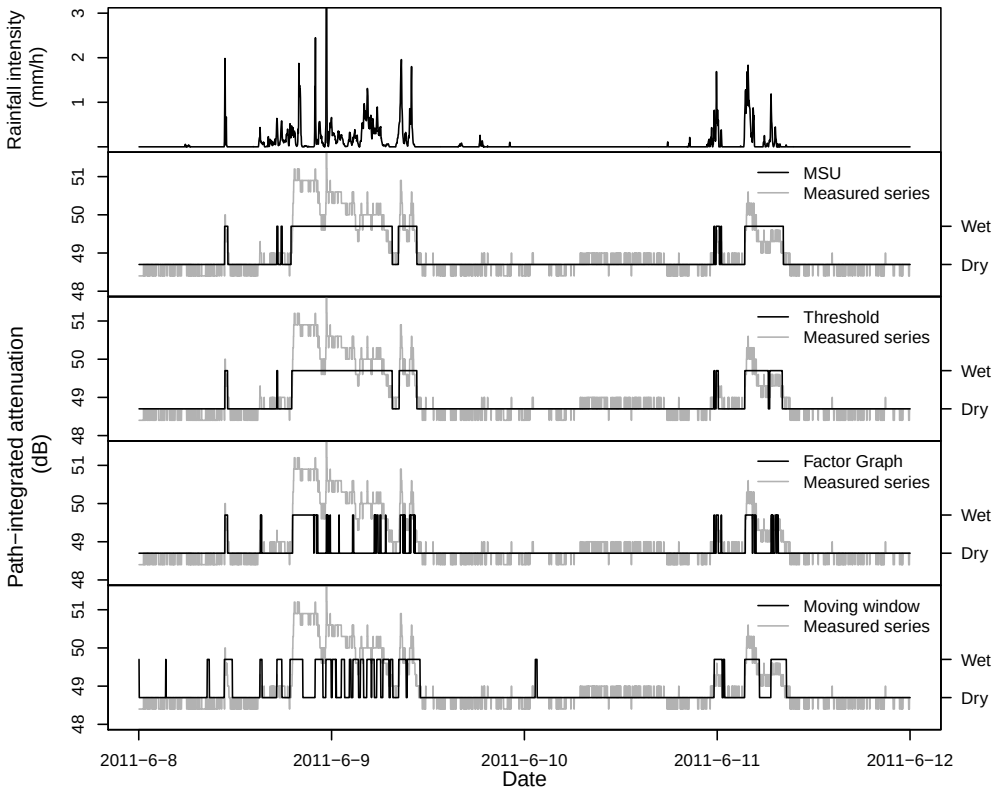

Figure 7.3: Illustration of the classification performances for the univariate Markov switching algorithm (MSU), the simple threshold, the Factor Graph and the moving window on a subset of dataset 1 (stationary case). Displayed are the observations from channel 1. The time is given in UTC.

The univariate Markov switching model (MSU) clearly produced the best classification performances among the univariate models, closely followed by the simple threshold

method. The good performance of the simple threshold algorithm is explained by the fact that the dry-weather attenuations over this time period are (more or less) constant with very low fluctuations. The moving window and the Factor Graph, on the other hand, have significantly higher values of type I and type II errors. This can be partially explained by the fact that these models rely on pre-defined threshold parameters which were not necessarily optimal over the considered time period. For example, it is possible to decrease the type II error rate in the moving window algorithm by increasing the value of σ_0. This will, however, also result in an increased type I error rate. Additional tests with different threshold parameters confirmed that the moving window algorithm produces, on average, less reliable classifications than the threshold and the Markov switching algorithm. Not surprisingly, the multivariate Markov switching model outperformed all the other univariate models in terms of type I and type II errors. Its false rain/dry detection rates are 2.46 % and 21.97 % for a rain detection threshold of 0 mm h^{-1} and 4.27 % and 11.82 % for a rain detection threshold of 0.1 mm h^{-1}(not shown). This confirms the intuitive idea that the state of the system can be estimated more accurately using 4 channels rather than 1. The improvement is, however, only minor because the univariate Markov switching model already produced good and similar classifications for all the considered channels (except for channel 2 for which no valid model parameters could be fitted). The fact that the univariate Markov switching model provides realistic classifications can also be seen in Figure 7.3, which shows the estimated states (dry/rainy) for all the considered algorithms. A qualitative evaluation suggests that the best classifications are obtained for the threshold method and the univariate Markov switching model (MSU). The classifications obtained using the Factor Graph and the moving window algorithm are not satisfactory. Both the Factor Graph and the moving window produce considerable false dry detections. The moving window algorithm also produces some false rain detections at the beginning of the period. Clearly, the threshold parameters (which were subjectively estimated for the entire dataset) are not optimal for this period.

7.5.3 Non-stationary dry-weather attenuation baseline

The results for the second dataset are shown in Table 7.3. As for the first dataset, the classification performances is illustrated in Figure 7.4, where the results are plotted for a 11-days subset from the 27$^{\text{th}}$ of March 2011 to the 7$^{\text{th}}$ of April 2011.

The first point to notice is that all the considered models have a very high rate of type II errors (about 50-60 % for the first rain detection threshold and 20-35 % for the second rain detection threshold). This is due to the large variability of the attenuation baseline during dry periods, which makes it difficult for the models to separate dry periods from light rainfall. Consequently, more rainy periods are classified as dry. This is also confirmed by the low type I error rates, meaning that very few dry periods are actually classified as rainy. The “best” average performance (among the univariate algorithms)

Table 7.3: Classification performances for dataset 2 (non-stationary case). Same format than in Table 7.2. Note that no model parameters could be fitted for the MSU algorithm on channels 2 and 4.

Models	rain detection threshold 0 mm h^{-1}		rain detection threshold 0.1 mm h^{-1}	
	type I error	type II error	type I error	type II error
ST	7.30 (2.74)	48.91 (2.35)	7.51 (2.69)	17.72 (3.57)
MW	11.54 (0.28)	65.74 (0.77)	11.97 (0.34)	17.54 (2.16)
FG	1.80 (0.47)	61.81 (1.40)	1.93 (0.46)	34.78 (2.57)
MSU	1.83 (0.43)	56.10 (2.17)	2.06 (0.47)	27.09 (2.27)
MSM	3.24	50.55	3.53	20.08

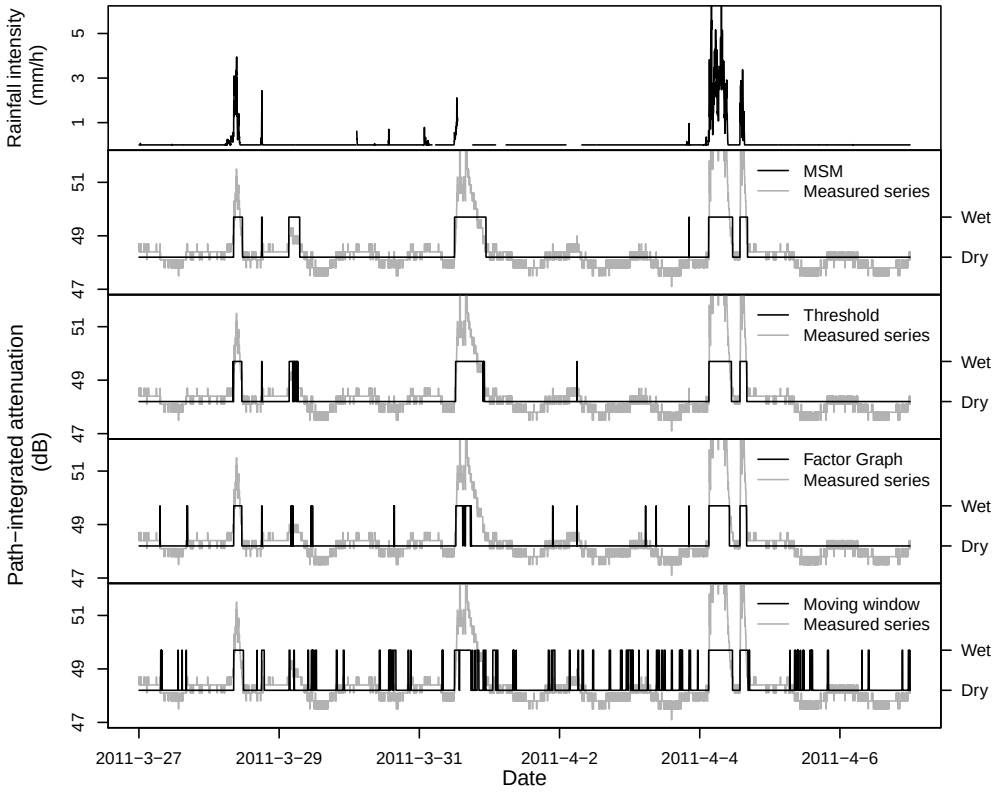

Figure 7.4: Illustration of the classification performances of the algorithms for a subset of dataset 2 (non-stationary case). Displayed are the observations from channel 4. Note that the results of the MSU, which did not converge for this channel, have been replaced by the results of the MSM.

is again obtained for the univariate Markov switching model and the simple threshold method, although these two models do not have the same type I,II error rates. As can be seen in Table 7.3, the threshold method produces less false dry detections but more false rain detections. The moving window algorithm has the highest rate of type II errors (65.74 % on average for a rain detection threshold of 0 mm h^{-1}), but most of

these false dry detections correspond to very light rain rates. This is indicated by the fact that, for a rain detection threshold of 0.1 $\mathrm{mm\,h^{-1}}$, which essentially removes light rainfall, a much lower type II error rate of 17.54 % is obtained. In fact, for the higher rain detection threshold, the moving window algorithm performs similarly to the simple threshold and the univariate Markov switching model. Again, the multivariate Markov switching algorithm outperformed (on average) the univariate algorithms in terms of false dry and rain detections. In particular, it is worth mentioning that no valid model parameters could be fitted for the univariate Markov switching model for channels 2 and 4 whereas the multivariate Markov switching model (using all 4 channels) was still able to provide valid parameter estimates for all channels. The threshold method and the multivariate Markov switching algorithm (MSM) produce very good and similar results for this time period. The classifications obtained using the Factor Graph and the moving window algorithm do not look very good. In particular, the strong variability in the attenuation baseline causes the moving window algorithm to produce a large amount of false rain detections. This problem could be (partially) solved by considering a lower detection threshold σ_0 for this time period, but there is currently no easy way of doing this automatically in the absence of any control data from nearby weather stations.

7.6 Discussion and possible developments

The Markov switching model proposed in Section 2 already provides good results at a reasonable computational cost. It remains, however, very simple in its formulation and does not exploit the full potential of state space models. As a possible extension, the performance of an autoregressive state space model of order 1 was also investigated. Although more elaborate, the autoregressive model of order 1 only showed little improvement in performance compared to the much simpler AR(0) model. Because autoregressive state space models are longer and more difficult to fit, the AR(0) was preferred for practical applications. Next, the authors investigated how the Gaussian error assumption in Equation (7.3) affects the dry/wet classification performance. It is well known that the distribution of rain rate values (and consequently path-integrated attenuation) is skewed and closer to a log-normal distribution than to a Gaussian distribution. An alternative model formulation with non-Gaussian error structure was therefore considered:

$$A_t = \begin{cases} \mu_0 + \varepsilon_0 & \text{for every dry period} \\ \mu_0 + \varepsilon_0 + \varepsilon_1 & \text{for every rainy period} \end{cases} \tag{7.11}$$

where ε_0 is a Gaussian random variable with zero mean and standard deviation σ_0 and ε_1 is a positive random variable with log-normal distribution representing the rain-induced attenuation. The major drawback of such a formulation is that it has no analytical expression for the conditional density of A_t knowing $S_t = 1$. It is, at the expense of additional computation costs, however still possible to fit this model using numerical approximations. Surprisingly, the more complex and physical error structures did not

improve significantly the classification performance. The reason for this can be seen in Figure 7.5 which shows the probability density functions of attenuation values for dry and rainy periods. The sample distributions are not exactly Gaussian but the fact that the tails of the distributions are not correctly reproduced is not critical with respect to the classification problem. In fact, the optimal classification threshold which is at the intersection between the two empirical probability density functions (i.e., about 49 dB) is very close to the threshold derived from the Gaussian model (i.e., the intersection between the two Gaussian densities). Similar results are obtained for all channels and all considered datasets and show that the Gaussian error assumption is not critical with respect to the classification problem.

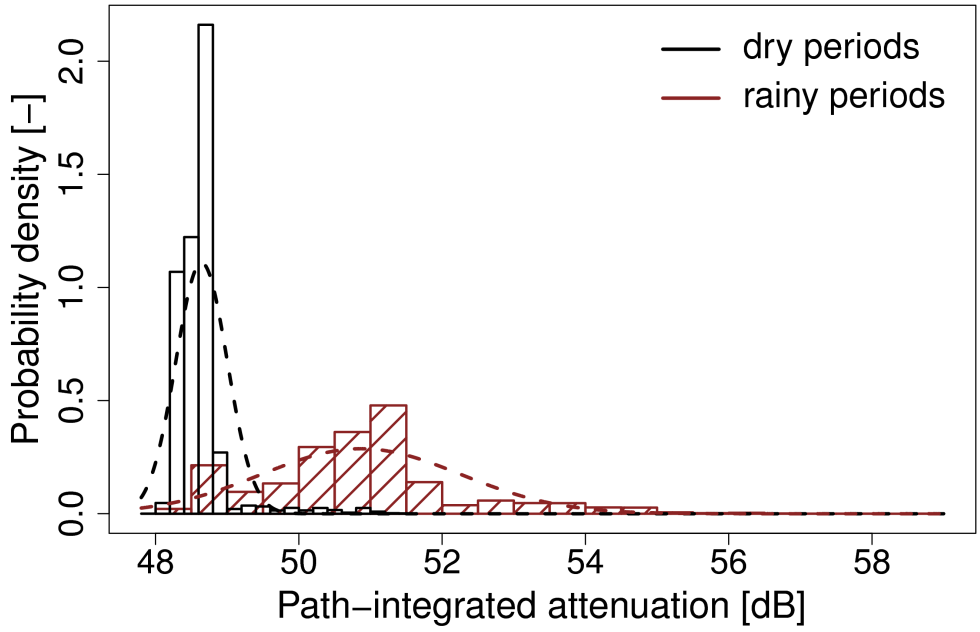

Figure 7.5: Empirical probability density functions of attenuation values for dry and rainy periods (for dataset 1). The dashed lines represent the fitted densities of a Gaussian distribution with same mean and variance as the samples. The dry and rainy periods are derived from the disdrometer data.

Another fundamental problem that needs to be addressed in future studies concerns the problem of the wet antenna effects on the classification into dry and wet periods. Most commercial microwave links do not have shielded antennas. Consequently, they experience some additional attenuation due to a thin water film formed on the surface of the antennas. This effect can be in the order of several dB and must be taken into account when estimating dry and rainy periods, especially during and immediately after a given rain event where the antenna can stay wet for several hours. Future investigations could consider two different states for dry periods, depending on the state of the antenna.

$$S_t = \begin{cases} 0 & \text{for every dry period with dry antenna} \\ 1 & \text{for every dry period with wet antenna} \\ 2 & \text{for every rainy period (with wet antenna)} \end{cases} \tag{7.12}$$

In this case, a possible attenuation model could be given by

$$A_t = \begin{cases} \mu_0 + \varepsilon_0 & \text{for every dry period with dry antenna} \\ \mu_1 + \varepsilon_1 & \text{for every dry period with wet antenna} \\ \mu_2 + \varepsilon_2 & \text{for every rainy period} \end{cases} \tag{7.13}$$

It must be noted, however, that such a model might be poorly identifiable, i.e., the parameters and states can not be identified without ambiguity because of the uncertainty affecting the power measurements and because of the strong dependence between the model parameters.

7.7 Conclusions

A new algorithm based on a Markov switching model has been introduced to classify attenuation measurements from commercial microwave links into dry and rainy periods. The performance of the algorithm has been evaluated using real data from a new and original experimental set-up and compared to 3 other classification methods. The results show that the Markov switching algorithm performs well and that its classification performance can be increased if multiple channel inputs are considered. Clearly, this is a big advantage compared to other univariate algorithms from the literature which cannot be generalized easily to the multivariate case. The fact that the Markov switching model does not require any empirically estimated threshold parameters is also of advantage.

The experimental set-up described in Section 7.4 provides a unique platform from which various aspects of rainfall retrieval using MWL can be investigated. For example, it is now possible to rigorously evaluate and compare the classification performances of the different algorithms presented in Section 7.3. The potential applications and scientific value of this experiment go, however, far beyond the simple application presented in this study. Future studies will, for example, investigate the effect of wet antenna on retrieved rain rates, explore how attenuation of orthogonal polarizations can be used to retrieve the effective drop size distribution (DSD) along the link path, and the possibility to use multiple channels in order to improve the accuracy of the rain rate estimates.

8 Conclusions and perspectives

8.1 Summary

Precipitation is an essential part of the hydrological cycle and must be carefully monitored. Knowledge of its microstructure (e.g., the drop size distribution) is essential in many hydrological and remote sensing applications. The large spatial and temporal variability of rainfall (and DSD) makes it particularly difficult to accurately measure rainfall at scales that are relevant for hydrological applications (e.g., 1 km and 5 min). Stochastic simulation addresses this issue by providing the ability to generate large numbers of synthetic rain events with similar structures and statistical properties. Such simulated fields provide useful reference data that can be used to quantify the errors and uncertainties associated to remote sensing of precipitation (e.g., using weather radars or microwave links).

In Chapter 2, a (relatively) simple 2-dimensional DSD simulation method (without intermittency and temporal structure) based on Geostatistics has been presented. It uses variograms to characterize the spatial and temporal structures of the DSD parameters. A multivariate Gaussian anamorphosis is used to transform the correlated DSD parameters μ, Λ and N_t into independent Gaussian variables $\tilde{\mu}$, $\tilde{\Lambda}$ and $\tilde{N}_t$ prior to their simulation. Cross-correlations are re-introduced during the back-transformation at the end of the simulation. The simulator is parameterized using DSD data collected from a network of disdrometers at EPFL campus, Switzerland and radar rain-rate data from MeteoSwiss. The potential of the simulator to address important issues in hydrology and remote sensing applications is illustrated for two events of contrasted types.

In Chapter 3, a new and more general DSD simulator has been described. It includes the temporal dimension and proposes to model rainfall intermittency using an indicator field (0 = dry ; 1 = rainy). The space-time structures of the DSD parameters (and of the intermittency field) are described using space-time variograms. Taylor's hypothesis of frozen turbulence is used to link the temporal variations and the spatial variations along

the direction of advection. The variability in all the other directions is described using a geometrical anisotropy parameter (estimated from radar data). External drifts are used to ensure that the transitions between dry and rainy locations are realistic. They can also be used to (partially) control the growth and decay of individual rain cells.

In Chapter 4, the DSD simulator has been applied to the problem of rainfall disaggregation. A modified version of the simulator described in Chapter 2 and Chapter 3 is used to generate DSD fields at high spatial and temporal resolutions that satisfy a set of conditions at the coarse scale (e.g., prescribed areal rain-rate values or radar reflectivities). The procedure is illustrated using MeteoSwiss radar rain-rate data and two types of rainfall (stratiform vs convective). Further extensions of the method to dual-frequency and dual-polarization weather radars are discussed.

In Chapter 5, a statistical analysis of rainfall intermittency at different spatial and temporal scales has been performed. The analysis shows that intermittency can be adequately represented by scaled exponential functions with fixed shape parameters and variable scale parameters. Similar results are found for the spatial and temporal autocorrelation functions of intermittency. Climatological scaling laws and intermittency models based on radar and disdrometer data were fitted and subsequently used to parameterize the DSD simulator and the rainfall disaggregation algorithm.

In Chapter 6, a new algorithm for the identification of dry and rainy periods using attenuation measurements from single-polarizations, single-frequency telecommunication microwave links has been proposed. Rain identification is achieved by analyzing the local (temporal) variability of the link signal over short periods of about 30 min. If the variability exceeds a given threshold, the period is declared rainy. Two different methods (depending on the amount of data that are available) for the estimation of the rain detection threshold are provided. Once dry and rainy periods have been identified, the attenuation baseline is estimated during dry periods and subsequently interpolated during rainy periods. Comparisons with a nearby C-band radar show that the dry/wet classification error rate is about 10 %.

In Chapter 7, a new and better algorithm for the identification of dry and rainy periods using MWL attenuation measurements has been provided. It is based on a Markov switching model whose parameters are estimated using the maximum likelihood method. The method is very general and can be easily extended to multivariate inputs (from different channels). Data from a new and innovative set-up deployed in Dübendorf in the vicinity of Zürich are used to evaluate the performance of the algorithm. The false dry/wet detection rates are compared to 3 other algorithms from the literature, including the method described in Chapter 6. On average, the Markov switching model outperforms the other algorithms. More general Markov switching models that take into account wet antenna effects are discussed.

8.2 Important contributions of this thesis

The most important contributions of this thesis can be summarized as follows:

(1) A stochastic DSD simulator capable of generating large numbers of DSD fields in space and time with realistic space-time structures and intermittency has been proposed.

(2) A new rainfall disaggregation method (based on the DSD simulator) has been presented. It can be used to downscale rain rate or reflectivity fields provided at a coarse resolution into high resolution DSD fields with realistic space-time structures and intermittency.

(3) Different scaling laws that can be used to downscale/upscale rainfall intermittency (including spatial and temporal autocorrelations) have been provided. The results were used to parameterize the DSD simulator and the rainfall disaggregation method.

(4) Two new methods (i.e., the moving window algorithm and the Markov switching model) for the identification of dry and rainy periods using attenuation measurements from telecommunication microwave links were proposed. A new experimental set-up to investigate the potential of commercial microwave links for quantitative rainfall estimation has been presented.

8.3 Perspectives

The research presented in this thesis has shown that there are a few points that merit further attention. First, it must be noted that the proposed simulation techniques only mention 2-dimensional fields of DSD at the ground level (+ an additional temporal dimension). No assumptions with respect to the vertical variability and structure of the DSD fields are made. Finding a way to include this vertical variability in the simulations would be extremely useful, especially for satellite applications. In theory, 3-dimensional fields of DSD can be generated using the techniques described in Chapters 2-4. In practice, however, this requires to parameterize and characterize the spatial and temporal structures of the DSD in the vertical dimension. In order to be realistic, such characterizations would have to take into account many different and complex issues. For example, one must take into account the fact that large drops fall at higher speeds than small ones. Other issues related to the melting layer, the effect of winds and updrafts and finally, the fact that rainfall intermittency in the vertical dimension is poorly documented must also be addressed. Nevertheless, if adequate parameterizations of the vertical variability can be provided, an extension to 3-dimensional fields of DSD (using the generated 2D + time fields as a starting point) is possible.

Another, very important issue that needs to be addressed in future work concerns the extension of the simulation domain (e.g., to domain sizes of 100×100 km^2 or more).

At these scales, spatial inhomogeneities (e.g., due to orographic effects) and temporal non-stationarities (e.g., due to the birth, growth, merging and decay of individual rain cells) play an important role and must be carefully taken into account. As suggested in the Appendix of Chapter 3, a possible way to include such phenomena into the simulations is by adding external drifts to the simulated parameters. Another more pragmatic way would be to break down the simulation domain into smaller pieces and to simulate each sub-domain separately. Finding a way to combine the individual pieces together is, however, not trivial. Another, maybe more promising idea suggested in Section 4.5.4 is to build the simulations from "bottom" to "top", i.e., by adding successive layers of (increasing) noise to an initial and very smooth field that represents the general trend over the considered domain. However, it is not clear how such methods could be parameterized using observations from disdrometers, rain gauges or weather radars.

Future work on disaggregation techniques will mainly focus on the possibility to include data from dual-frequency or dual-polarization weather radars in the disaggregation process. Also, data from different sensors can be easily combined in order to produce DSD simulations that satisfy multiple conditions at different scales (e.g., areal rain rates prescribed by radar, path-integrated attenuations in adequacy with measurements from MWL and specific values of DSDs at the point scale imposed by nearby disdrometers). The evaluation of the disaggregated fields presented in Section 4.5.3 also showed that more efforts are necessary in order to better describe the relationships between rainfall intermittency at the coarse and the fine scales.

A lot of future and interesting results can also be expected from the data collected using the experimental set-up presented in Chapter 7. So far, about 1 year of data have been collected. First results have already been obtained (e.g., see Chapter 7) but a lot of issues still need to be investigated. For example, issues related to wet antenna effects and their consequences in terms of rainfall estimation errors and uncertainties. Other, more elaborate methods (e.g., wavelets) for the identification of dry and rainy periods and the estimation of an attenuation baseline are also worth investigating.

Bibliography

Alabert, F. G., 1987: Stochastic imaging of spatial distributions using hard and soft information. M.S. thesis, Stanford University, 198 pp., Dept of Applied Earth Sciences.

Anagnostou, E. N. and W. F. Krajewski, 1997: Simulation of radar reflectivity fields: Algorithm formulation and evaluation. *Water Resour. Res.*, **33 (6)**, 1419–1428.

Andsager, K., K. V. Beard, and N. F. Laird, 1999: Laboratory measurements of axis ratios for large rain drops. *J. Atmos. Sci.*, **56 (15)**, 2673–2683.

Arnold, C. and C. Dey, 1986: Observing-Systems Simulation Experiments: Past, Present, and Future. *Bull. Amer. Meteor. Soc.*, **67 (6)**, 687–695.

Assouline, S. and Y. Mualem, 1989: The similarity of regional rainfall - a dimensionless model of drop size distribution. *Trans. ASAE*, **32 (4)**, 1216–1222.

Atlas, D. and C. W. Ulbrich, 1977: Path and area integrated rainfall measurement by microwave attenuation in the 1-3 cm band. *J. Appl. Meteor.*, **16 (4)**, 327–332.

Atlas, D. and C. W. Ulbrich, 1990: Early foundations of the measurement of rainfall by radar. *Radar Meteorology*, D. Atlas, Ed., American Meteorological Society, 86–97.

Austin, P. M., 1987: Relation between measured radar reflectivity and surface rainfall. *Mon. Weather Rev.*, **115 (5)**, 1053–1071.

Aydin, K. and S. E. Daisley, 2002: Relationships between rainfall rate and 35-ghz attenuation and differential attenuation: modeling the effects of raindrop size distribution, canting, and oscillation. *IEEE T. Geosci. Remote Sens.*, **40 (11)**, 2343–2352.

Barancourt, C., J.-D. Creutin, and J. Rivoirard, 1992: A method for delineating and estimating rainfall fields. *Water Resour. Res.*, **28 (4)**, 1133–1144.

Bárdossy, A., 1998: Generating precipitation time series using simulated annealing. *Water Resour. Res.*, **34 (7)**, 1737–1744, doi:10.1029/98WR00981.

Bárdossy, A. and G. Pegram, 2011: Downscaling precipitation using regional climate models and circulation patterns toward hydrology. *Water Resour. Res.*, **47 (W04505)**, doi:10.1029/2010WR009689.

Bárdossy, A. and E. J. Plate, 1992: Space-time model for daily rainfall using atmospheric circulation patterns. *Water Resour. Res.*, **28 (5)**, 1247–1259.

Battan, L. J., 1973: *Radar observation of the atmosphere*. University of Chicago Press, 324 pp.

Baudena, M., G. Boni, L. Ferraris, J. von Hardenberg, and A. Provenzale, 2007: Vegetation response to rainfall intermittency in drylands: Results from a simple ecohydrological box model. *Adv. Water Resour.*, **30 (5)**, 1320–1328, doi:10.1016/j.advwatres.2006.11.006.

Beard, K. V., 1976: Terminal velocity and shape of cloud and precipitation drops aloft. *J. Atmos. Sci.*, **33 (5)**, 851–864.

Beard, K. V., 1977: Terminal velocity adjustment for cloud and precipitation drops aloft. *J. Atmos. Sci.*, **34 (8)**, 1293–1298.

Beard, K. V., V. N. Bringi, and M. Thurai, 2010: A new understanding of raindrop shape. *Atmos. Res.*, **97 (4)**, 396–415, doi:10.1016/j.atmosres.2010.02.001.

Beard, K. V. and C. Chuang, 1987: A new model for the equilibrium shape of raindrops. *J. Atmos. Sci.*, **44 (3)**, 1509–1524.

Beier, C. and K. Hansen, 1993: Spatial variability of throughfall fluxes in a spruce forest. *Environ. Pollut.*, **81**, 257–267.

Berndtsson, R. and J. Niemczynowicz, 1988: Spatial and temporal scales in rainfall analysis, some aspects and future perspectives. *J. Hydrol.*, **100 (1-4)**, 293–313.

Berne, A., G. Delrieu, J.-D. Creutin, and C. Obled, 2004: Temporal and spatial resolution of rainfall measurements required for urban hydrology. *J. Hydrol.*, **299 (3-4)**, 166–179.

Berne, A. and M. Schleiss, 2009: Retrieval of the rain drop size distribution using telecommunication dual-polarization microwave links. *Proc. 34th AMS Conf. Radar Meteorol.*, Williamsburg, USA, AMS.

Berne, A. and R. Uijlenhoet, 2005: A stochastic model of range profiles of raindrop size distributions: application to radar attenuation correction. *Geophys. Res. Lett.*, **32 (10)**, L10803, doi:10.1029/2004GL021899.

Berne, A. and R. Uijlenhoet, 2007: Path-averaged rainfall estimation using microwave links: Uncertainty due to spatial rainfall variability. *Geophys. Res. Lett.*, **34 (7)**, L07403, doi:10.1029/2007GL029409.

Bo, Z., S. Islam, and A. B. Eltahir, 1994: Aggregation-disaggregation properties of a stochastic rainfall model. *Water Resour. Res.*, **30 (12)**, 3423–3435.

Bouvier, C., L. Cisneros, R. Dominguez, J.-P. Laborde, and T. Lebel, 2003: Generating rainfall fields using principal components (PC) decomposition of the covariance matrix: a case study in Mexico City. *J. Hydrol.*, **278 (1-4)**, 107–120.

Bradley, S. G. and C. D. Stow, 1974: The measurement of charge and size of raindrops: Part II. Results and analysis at ground level. *J. Appl. Meteor.*, **13 (1)**, 131–147.

Bras, R. L. and I. Rodríguez-Iturbe, 1976: Rainfall generation - nonstationary time-varying multidimensional model. *Water Resour. Res.*, **12 (3)**, 450–456.

Braud, I., J. D. Creutin, and C. Barancourt, 1993: The relation between the mean areal rainfall and the fractional area where it rains above a given threshold. *J. Appl. Meteor.*, **32**, 193–202.

Bringi, V. N. and V. Chandrasekar, 2001: *Polarimetric doppler weather radar*. Cambridge University Press, 662 pp.

Bringi, V. N., V. Chandrasekar, N. Balakrishnan, and D. S. Zrnič, 1990: An examination of propagation effects in rainfall on radar measurements at microwave frequencies. *J. Atmos. Oceanic Technol.*, **7 (6)**, 829–840.

Bringi, V. N., T. D. Keenan, and V. Chandrasekar, 2001: Correcting C-band radar reflectivity and differential reflectivity data for rain attenuation: A self-consistent method with constraints. *IEEE T. Geosci. Remote Sens.*, **39 (9)**, 1906–1915.

Callaghan, S., B. Boyes, A. Couchman, J. Waight, C. Walden, and S. Ventouras, 2008: An investigation of site diversity and comparison with ITU-R recommendations. *Radio Sci.*, **43 (RS4010)**, doi:10.1029/2007RS003793.

Chate, D. M., P. S. P. Rao, M. S. Naik, G. A. Momin, P. D. Safai, and K. Ali, 2003: Scavenging of aerosols and their chemical species by rain. *Atmos. Environ.*, **37 (18)**, 2477–2484.

Chilès, J.-P. and P. Delfiner, 1999: *Geostatistics: Modeling spatial uncertainty*. Probability and statistics, Wiley, 695 pp.

Ciach, G. J. and W. F. Krajewski, 2006: Analysis and modeling of spatial correlation structure in small-scale rainfall in central Oklahoma. *Adv. Water Resour.*, **29 (5)**, 1450–1463, doi:10.1016/j.advwaters.200511.003.

Collier, C. G., 1986: Accuracy of rainfall estimates by radar, part II: comparison with raingauge network. *J. Hydrol.*, **83 (1-4)**, 224–235.

Cowpertwait, P. S. P., P. E. O'Connell, A. V. Metcalfe, and J. A. Mawdsley, 1996: Stochastic point process modelling of rainfall. ii. regionalization and disaggregation. *J. Hydrol.*, **175**, 47–65.

Crane, R. K., 1989: Automatic cell detection and tracking. *IEEE T. Geosci. Electronics*, **17 (4)**, 250–262.

Cressie, N. and D. M. Hawkins, 1990: Robust estimation of the variogram. *J. Math. Geology*, **12 (2)**, 115–125.

Creutin, J.-D., G. Delrieu, and T. Lebel, 1988: Rain measurement by raingage-radar combination: a geostatistical approach. *J. Atmos. Oceanic Technol.*, **5 (1)**, 102–115.

David, N., P. Alpert, and H. Messer, 2009: Technical Note: Novel method for water vapour monitoring using wireless communication networks measurements. *Atmos. Chem. Phys.*, **9 (7)**, 2413–2418, doi:10.5194/acp-9-2413-2009.

Delrieu, G., L. Hucke, and J.-D. Creutin, 1999: Attenuation in rain for X- and C-band weather radar systems: sensitivity with respect to the drop size distribution. *J. Appl. Meteor.*, **38 (1)**, 57–68.

Diciccio, T., M. Martin, and G. Young, 1993: Analytical approximations to conditional distribution functions. *Biometrika*, **80 (4)**, 781–790.

Dissanayake, A., J. Allnutt, and F. Haidara, 1997: A prediction model that combines rain attenuation and other propagation impairments along Earth-satellite paths. *IEEE T. Antenn. Propag.*, **45 (10)**, doi:10.1109/8.633864.

Durden, S. L., Z. S. Haddad, A. Kitiyakara, and F. K. Li, 1998: Effects of nonuniform beam filling on rainfall retrieval for the TRMM precipitation radar. *J. Atmos. Oceanic Technol.*, **15 (3)**, 635–646.

Finlay, P. J., R. Fell, and P. K. Maguire, 1997: The relationship between the probability of landslide occurrence and rainfall. *Can. Geotech. J.*, **34 (6)**, 811–824.

Fong, B., P. B. Rapajic, G. Y. Hong, and A. C. M. Fong, 2003: Factors causing uncertainties in outdoor wireless wearable communications. *IEEE Pervas. Comput.*, **2 (2)**, 16–19, doi:10.1109/MPRV.2003.1203748.

Foufoula-Georgiou, E. and W. F. Krajewski, 1995: Recent advances in rainfall modeling, estimation and forecasting. *Rev. Geophys.*, **33 (S1)**, 1125–1137.

Gabriel, K. R. and J. Neumann, 1962: A markov chain model for daily rainfall occurrence at Tel-Aviv. *Q. J. Roy. Meteor. Soc.*, **88 (375)**, 90–95.

Germann, U., G. Galli, M. Boscacci, and M. Bolliger, 2006: Radar precipitation measurement in a mountainous region. *Q. J. Roy. Meteor. Soc.*, **132 (618)**, 1669–1692, doi:10.1256/qj.05.190.

Goldshtein, O., H. Messer, and A. Zinevich, 2009: Rain rate estimation using measurements from commercial telecommunication links. *IEEE T. Signal Processing*, **57 (4)**, 1616–1625.

Gorgucci, E. and V. Chandrasekar, 2005: Evaluation of attenuation correction methodology for dual-polarization radars: Application to X-band sytems. *J. Atmos. Oceanic Technol.*, **22 (8)**, 1195–1206, doi:10.1175/JTECH1763.1.

Gosset, M. and I. Zawadzki, 2001: Effect of nonuniform beam filling on the propagation of the radar signal at X-band frequencies. Part I: changes in the *k(Z)* relationship. *J. Atmos. Oceanic Technol.*, **18 (7)**, 1113–1126.

Groisman, P. Y. and D. R. Easterling, 1994: Variability and trends of total precipitation and snowfall over the United States and Canada. *J. Climate*, **7 (1)**, 184–205.

Grum, M., S. Krämer, H. R. Verworn, and A. Redder, 2005: Combined use of point rain gages, radar, microwave link and level measurements in urban hydrological modelling. *Atmos. Res.*, **77 (1-4)**, 313–321.

Guillot, G., 1999: Approximation of Sahelian rainfall fields with meta-Gaussian random functions - Part 1: model definition and methodology. *Stoch. Env. Res. Risk A.*, **13 (1-2)**, 100–112.

Guntner, A., J. Olsson, A. Calver, and B. Gannon, 2001: Cascade-based disaggregation of continuous rainfall time series: the influence of climate. *Hydrol. Earth Syst. Sci.*, **5 (2)**, 145–164.

Gupta, V. K. and E. Waymire, 1993: A statistical analysis of mesoscale rainfall as a random cascade. *J. Appl. Meteor.*, **32 (2)**, 251–267.

Haddad, Z. S., D. A. Short, S. L. Durden, E. Im, S. Hensley, M. B. Grable, and R. A. Black, 1997: A new parametrization of the rain drop size distribution. *IEEE T. Geosci. Remote Sens.*, **35 (3)**, 532–539.

Hall, P., R. Wolff, and Q. Yao, 1999: Methods for estimating a conditional distribution function. *J. Amer. Stats. Assoc.*, **94**, 154–163.

Hamilton, J., 1989: A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, **457 (2)**, 357–384.

Hamilton, J., 1990: Analysis of time series subject to changes in regime. *J. Econometrics*, **45 (1-2)**, 39–70, doi:10.1016/0304-4076(90)90093-9.

Hershenhorn, J. and D. A. Woolhiser, 1987: Disaggregation of daily rainfall. *J. Hydrol.*, **95**, 299–322.

Holt, A. R., G. G. Kuznetsov, and A. R. Rahimi, 2003: Comparison of the use of dual-frequency and single-frequency attenuation for the measurement of path-averaged rainfall along a microwave link. *IEE Proc-H*, **150 (5)**, 315–320.

Hutchinson, M. F., 1995: Stochastic space-time weather models from ground-based data. *Agr. and Forest Meteor.*, **73 (3-4)**, 237–264.

Iguchi, T., T. Kozu, R. Meneghini, J. Awaka, and K. Okamoto, 2000: Rain-profiling algorithm for the TRMM precipitation radar. *J. Appl. Meteor.*, **39 (12)**, 2039–2052.

ITU-R P.838-3, 2005: Specific attenuation model for rain for use in prediction methods. International Telecommunication Union Radiocommunication Recommendations, International Telecommunication Union Radiocommunication Recommendations.

Jaffrain, J. and A. Berne, 2011: Experimental quantification of the sampling uncertainty associated with measurements from Parsivel disdrometers. *J. Hydrometeor.*, **12 (3)**, doi:10.1175/2010JHM1244.1.

Jaffrain, J. and A. Berne, 2012: Quantification of the small-scale spatial structure of the raindrop size distribution from a network of disdrometers. *J. Appl. Meteor. Climate*, in press.

Jaffrain, J., A. Studzinski, and A. Berne, 2011: A network of disdrometers to quantify the small-scale variability of the raindrop size distribution. *Water Resour. Res.*, **47**, W00H06, doi:10.1029/2010WR009872.

Jameson, A. R., 1989: Theoretical analysis and meteorological interpretation of the role of raindrop shape on microwave attenuation and propagation phase shifts: implication for the radar measurement of rain. *J. Atmos. Oceanic Technol.*, **6 (1)**, 76–88.

Jameson, A. R. and A. B. Kostinski, 2001: What is a raindrop size distribution? *Bull. Amer. Meteor. Soc.*, **82 (6)**, 1169–1177.

Jeannin, N., L. Féral, H. Sauvageot, L. Castanet, and J. Lemorton, 2008: Statistical distribution of the fractional area affected by rain. *J. Geophys. Res.*, **113 (D21120)**, doi:10.1029/2008JD009780.

Jennings, S. A., M. F. Lambert, and G. Kuczera, 2010: Generating synthetic high resolution rainfall time series at sites with only daily rainfall using a master-target scaling approach. *J. Hydrol.*, **393 (3-4)**, 163–173, doi:10.1016/j.jhydrol.2010.08.013.

Journel, A. G. and C. J. Huijbregts, 1978: *Mining geostatistics*. Academic Press, London, 600 pp.

Kang, B. and J. Ramirez, 2010: A coupled stochastic space-time intermittent random cascade model for rainfall downscaling. *Water Resour. Res.*, **46 (W10534)**, doi:10.1029/2008WR007692.

Katz, R. W., 1977: Application of chain-dependent processes to meteorology. *J. Appl. Prob.*, **14 (3)**, 598–603.

Kell, G. S., 1975: Density, thermal expansivity, and compressibility of liquid water from 0 degrees to 150 degrees - correlations and tables for atmospheric-pressure and saturation reviewed and expressed on 1968 temperature scale. *J. Chem. Eng. Data*, **20 (1)**, 97–105.

Khain, A., M. Ovtchinnikov, M. Pinsky, A. Pokrovsky, and H. Krugliak, 2000: Notes on the state-of-the-art numerical modeling of cloud microphysics. *Atmos. Res.*, **55 (3-4)**, 159–224, doi:10.1016/S0169-8095(00)00064-8.

Kharadly, M. M. Z. and R. Ross, 2001: Effect of wet antenna attenuation on propagation data statistics. *IEEE T. Antenn. Propag.*, **49 (8)**, 1183–1191.

Kim, C. J., 1994: Dynamic linear models with Markov-switching. *J. Econometrics*, **60**, 1–22.

Kinnell, P. I. A., 2005: Raindrop-impact-induced erosion processes and prediction: a review. *Hydrol. Processes*, **19**, 2815–2844.

Kleiber, W., R. Katz, and B. Rajagopalan, 2012: Daily spatiotemporal precipitation simulation using latent and transformed gaussian processes. *Water Resour. Res.*, **48 (W01523)**, doi:10.1029/2011WR011105.

Kletter, A. Y., J. Hardenberg, E. Meron, and A. Provenzale, 2009: Patterned vegetation and rainfall intermittency. *Journal of Theoretical Biology*, **256**, 574–583.

Kliche, D., P. Smith, and R. Johnson, 2008: L-moment estimators as applied to gamma drop size distributions. *J. Appl. Meteor. Climate*, doi:10.1175/2008JAMC1936.1.

Koster, R. D. and M. J. Suarez, 1995: Relative contributions of land and ocean processes to precipitation variability. *J. Geophys. Res.*, **100 (D7)**, 775–790.

Koutsoyiannis, D., 2006: An entropic-stochastic representation of rainfall intermittency: The origin of clustering and persistence. *Water Resour. Res.*, **42 (W01401)**, doi:10.1029/2005WR004175.

Koutsoyiannis, D. and C. Onof, 2001: Rainfall disaggregation using adjusting procedures on a poisson cluster model. *J. Hydrol.*, **246 (1-4)**, 109–122, doi:10.1016/S0022-1694(01)00363-8.

Krajewski, W. F., 1987: Cokriging radar-rainfall and rain-gauge data. *J. Geophys. Res.*, **90 (D8)**, 9571–9580.

Krajewski, W. F., R. Raghavan, and V. Chandrasekar, 1993: Physically based simulation of radar rainfall data using a space-time rainfall model. *J. Appl. Meteor.*, **32 (2)**, 268–283.

Krajewski, W. F. and J. A. Smith, 2002: Radar hydrology: rainfall estimation. *Adv. Water Resour.*, **25 (8-12)**, 1387–1394, doi:10.1016/j.advwatres.2005.03.018.

Krämer, S., H.-R. Verwon, and A. Redder, 2005: Improvement of x-band radar rainfall estimates using a microwave link. *Atmos. Res.*, **77**, 278–299.

Kschischang, F. R., B. J. Frey, and H.-A. Loeliger, 2001: Factor graphs and the sum-product algorithm. *IEEE Trans. Inf. Theory*, **47 (2)**, 498–519.

Kumar, P. and E. Foufoula-Georgiou, 1994: Characterizing multiscale variability of zero intermittency in spatial rainfall. *J. Appl. Meteor.*, **33**, 1516–1525.

Kundu, P. K. and R. K. Siddani, 2011: Scale dependence of spatiotemporal intermittency of rain. *Water Resour. Res.*, **47 (W08522)**, doi:10.1029/2010WR010070.

Lantuéjoul, C., 2002: *Geostatistical simulation, models and algorithms*. Springer, 256 pp.

Lanza, L., 2001: A conditional simulation model of intermittent rain fields. *Hydrol. Earth Syst. Sci.*, **4 (1)**, 173–183.

Lanza, L., J. Ramirez, and E. Todini, 2001: Stochastic rainfall interpolation and downscaling. *Hydrol. Earth Syst. Sci.*, **5 (2)**, 139–143.

Lavergnat, J. and P. Golé, 1998: A stochastic raindrop time distribution model. *J. Appl. Meteor.*, **37 (8)**, 805–818.

Lavergnat, J. and P. Golé, 2006: A stochastic model of raindrop release: application to the simulation of point rain observations. *J. Hydrol.*, **328 (1-2)**, 8–19.

Lee, G., A. W. Seed, and I. Zawadzki, 2007: Modeling the variability of drop size distributions in space and time. *J. Appl. Meteor. Climate*, **46 (7)**, 742–756, doi:10.1175/JAM2505.1.

Lee, G., I. Zawadzki, W. Szyrmer, D. Sempere-Torres, and R. Uijlenhoet, 2004: A general approach to double-moment normalization of drop size distributions. *J. Appl. Meteor.*, **43 (2)**, 264–281.

Leijnse, H., R. Uijlenhoet, and A. Berne, 2010: Errors and uncertainties in microwave link rainfall estimation explored using drop size measurements and high-resolution radar data. *J. Hydrometeor.*, **11 (6)**, 1330–1344, doi:10.1175/2010JHM1243.1.

Leijnse, H., R. Uijlenhoet, and J. N. M. Stricker, 2007a: Hydrometeorological application of a microwave link: 1. evaporation. *Water Resour. Res.*, **43 (4)**, W04416, doi:10.1029/2006WR004988.

Leijnse, H., R. Uijlenhoet, and J. N. M. Stricker, 2007b: Hydrometeorological application of a microwave link. Part II: precipitation. *Water Resour. Res.*, **43 (4)**, W04417, doi:10.1029/2006WR004989.

Leijnse, H., R. Uijlenhoet, and J. N. M. Stricker, 2007c: Rainfall measurement using radio links from cellular communication networks. *Water Resour. Res.*, **43 (3)**, W03201, doi:10.1029/2006WR005631.

Leijnse, H., R. Uijlenhoet, and J. N. M. Stricker, 2008: Microwave link rainfall estimation: effects of link length and frequency, temporal sampling, power resolution, and wet antenna attenuation. *Adv. Water Resour.*, **31 (11)**, 1481–1493.

Leuangthong, O. and C. V. Deutsch, 2003: Stepwise conditional transformation for simulation of multiple variables. *Math. Geology*, **35 (2)**, 155–173.

Li, B., A. Murthi, K. Bowman, G. North, M. Genton, and M. Sherman, 2009: Statistical tests of taylor's hypothesis: An application to precipitation fields. *J. Hydrometeor.*, **10**, 254–265.

Loeliger, H. A., J. Dauwels, J. Ju, S. Korl, L. Ping, and F. R. Kschischang, 2010: The factor graph approach to model-based signal processing. *Proc. of the IEEE*, **95 (6)**, 1295–1322.

Löffler-Mang, M. and J. Joss, 2000: An optical disdrometer for measuring size and velocity of hydrometeors. *J. Atmos. Oceanic Technol.*, **17 (2)**, 130–139.

Lovejoy, S. and B. B. Mandelbrot, 1985: Fractal properties of rain, and a fractal model. *Tellus A*, **37A (3)**, 209–232.

Lovejoy, S. and D. Schertzer, 2008: Turbulence, raindrops and the $l^{1/2}$ number density law. *New J. Phys.*, **10 (075017)**, doi:10.1088/1367-2630/10/7/075017.

Mackay, N. G., R. E. Chandler, C. Onof, and H. S. Wheater, 2001: Disaggregation of spatial rainfall fields for hydrological modelling. *Hydrol. Earth Syst. Sci.*, **5 (2)**, 165–173.

Mardia, K., 1970: Measures of multivariate skewness and kurtosis with applications. *Biometrika*, **57**, 519–530.

Marshall, J. S., W. Hitschfeld, and K. L. S. Gunn, 1955: Advances in radar weather. *Adv. Geophys.*, **2**, 1–56.

Marshall, J. S. and W. M. Palmer, 1948: The distribution of raindrops with size. *J. Meteor.*, **5**, 165–166.

Matheron, G., 1965: *Les variables régionalisées et leur estimation.* Masson et Cie, Paris, 305 pp.

Matrosov, S. Y., K. A. Clark, B. E. Martner, and A. Tokay, 2002: X-band polarimetric radar measurements of rainfall. *J. Appl. Meteor.*, **41 (9)**, 941–952.

Menabde, M., A. Seed, D. Harris, and G. Austin, 1997: Self-similar random fields and rainfall simulation. *J. Geophys. Res.*, **102 (D12)**, 13 509–13 515.

Messer, H., A. Zinevich, and P. Alpert, 2006: Environmental monitoring by wireless communication networks. *Science*, **312 (5774)**, 713–713, doi:10.1126/science.1120034.

Metzner, P., I. Horenko, and C. Schuette, 2007: Generator estimation of Markov jump processes based on incomplete observations nonequidistant in time. *Phys. Rev. E*, **76 (6)**, doi:10.1103/PhysRevE.76.066702.

Mishchenko, M., L. Travis, and A. Lacis, 2002: *Scattering, Absorption, and Emission of Light by Small Particles*. Cambridge University Press, 445 pp.

Mishchenko, M. I. and L. D. Travis, 1998: Capabilities and limitations of a current FORTRAN implementation of the T-matrix method for randomly oriented, rotationally symmetric scatterers. *J. Quant. Spectrosc. Radiat. Transfer*, **60 (3)**, 309–324.

Moisseev, D. N. and V. Chandrasekar, 2007: Examination of the μ-δ relation suggested for drop size distribution parameters. *J. Atmos. Oceanic Technol.*, **24 (5)**, 847–855, doi:10.1175/JTECH2010.1.

Molini, A., G. Katul, and A. Porporato, 2009: Revisiting rainfall clustering and intermittency across different climatic regimes. *Water Resour. Res.*, **45 (W11403)**, doi:10.1029/2008WR007352.

Montopoli, M., F. S. Marzano, and G. Vulpiani, 2008: Analysis and synthesis of raindrop size distribution time series from disdrometer data. *IEEE T. Geosci. Remote Sens.*, **46 (2)**, 466–478, doi:10.1109/TGRS.2007.909102.

Olssen, J. and R. Berndtsson, 1998: Temporal rainfall disaggregation based on scaling properties. *Water Sci. Technol.*, **37 (11)**, 73–79.

Onibon, H. and T. Lebel, 2004: Gibbs sampling for conditional spatial disaggregation of rain fields. *Water Resour. Res.*, **40 (W08401)**, doi:10.1029/2003WR002009.

Onof, C. and K. Arnbjerg-Nielsen, 2009: Quantification of anticipated future changes in high resolution design rainfall for urban areas. *Atmos. Res.*, **92**, 350–363.

Onof, C., R. E. Chandler, A. Kakou, P. Northrop, H. S. Wheater, and V. Isham, 2000: Rainfall modelling using Poisson-cluster processes: a review of developments. *Stoch. Env. Res. Risk A.*, **14 (6)**, 384–411.

Orlanski, I., 1975: A rational subdivision of scales for atmospheric processes. *Bull. Amer. Meteor. Soc.*, **56 (5)**, 527–530.

Ormsbee, L. E., 1986: Rainfall disaggregation model for continuous hydrological modeling. *J. Hydraul. Eng.*, **115 (4)**, 507–525.

Paulson, K. S., 2002: The spatial-temporal statistics of rain rate random fields. *Radio Sci.*, **37 (5)**.

Pavlopoulos, H., 2011: A stochastic framework for downscaling processes of spatial averages based on the property of spectral multiscaling and its statistical diagnosis on spatio-temporal rainfall fields. *Adv. Water Resour.*, **34 (8)**, 990–1011, doi:10.1016/j.advwatres.2011.05.006.

Pavlopoulos, H. and J. Gritsis, 1999: Wet and dry epoch durations of spatially averaged rain rate, their probability distributions and scaling properties. *Environ. Ecol. Stat.*, **6**, 351–380.

Pavlopoulos, H. and V. K. Gupta, 2003: Scale invariance of regional wet and dry durations of rain fields: A diagnostic study. *J. Geophys. Res.*, **108(D8)**, doi:10.1029/2002JD002763.

Pebesma, E. J., 2004: Multivariate geostatistics in S: the gstat package. *Comput. Geosci.*, **30 (7)**, 683–691, doi:10.1016/j.cageo.2004.03.012.

Pegram, G. G. S. and A. N. Clothier, 2001: High resolution space time modelling of rainfall: the string of beads model. *J. Hydrol.*, **241 (1-2)**, 26–41.

Perica, S. and E. Foufoula-Georgiou, 1996: Model for multiscale disaggregation of spatial rainfall based on coupling meteorological and scaling descriptions. *J. Geophys. Res.*, **101 (D21)**, 26 347–26 362.

Pitman, A. J., A. Henderson-Sellers, and Z.-L. Yang, 1990: Sensitivity of regional climates to localized precipitation in global models. *Nature*, **346**, 734–737.

Pruppacher, H. R. and R. L. Klett, 1997: *Microphysics of clouds and precipitation*. 2d ed., No. 18 in Atmospheric and Oceanographic Sciences Library, Kluwer Academic Press, 954 pp.

Pruppacher, H. R. and R. L. Pitter, 1971: Semi-empirical determination of shape of cloud and rain drops. *J. Atmos. Sci.*, **28 (1)**, 86–94.

R Development Core Team, 2011: *R: A Language and Environment for Statistical Computing*. Vienna, Austria, R Foundation for Statistical Computing, URL http://www.R-project.org/, ISBN 3-900051-07-0.

Radke, L. F., P. V. Hobbs, and M. W. Eltgroth, 1980: Scavenging of aerosol-particles by precipitation. *J. Appl. Meteor.*, **19 (6)**, 715–722.

Rahimi, A. R., G. J. G. Upton, and A. R. Holt, 2004: Dual-frequency links - a complement to gauges and radar for the measurement of rain. *J. Hydrol.*, **288 (1-2)**, 3–12.

Rahimi, A. R., G. J. G. Upton, A. R. Holt, and R. J. Cummings, 2003: Use of dual-frequency microwave links for measuring path-averaged rainfall. *J. Geophys. Res.*, **108 (D15)**, 4467, doi:10.1029/2002JD003202.

Rayitsfeld, A., R. Samuels, A. Zinevich, U. Hadar, and P. Alpert, 2011: Comparison of two methodologies for long term rainfall monitoring using a commercial microwave communication system. *Atmos. Res.*, **104-105 (0)**, 119–127, doi:10.1016/j.atmosres.2011.08.011.

Rebora, N., L. Ferraris, J. von Hardenberg, and A. Provenzale, 2006a: Rainfall downscaling and flood forecasting: a case study in the mediterranean area. *Nat. Hazard. Earth Sys. Sci.*, **6**, 611–619.

Rebora, N., L. Ferraris, J. von Hardenberg, and A. Provenzale, 2006b: RainFARM: Rainfall downscaling by a filtered autoregressive model. *J. Hydrometeor.*, **7**, 724–738.

Reller, C., H. Loeliger, and J. Diaz, 2011: A model for quasi-periodic signals with application to rain estimation from microwave link gain. *Proc. 19th European Signal Proc. Conf. (EUSIPCO)*, Barcelona, Spain.

Richardson, C., 1981: Stochastic simulation of daily precipitation, temperature and solar radiation. *Water Resour. Res.*, **17 (1)**, 182–190.

Rigby, J. and A. Porporato, 2010: Precipitation, dynamical intermittency, and sporadic randomness. *Adv. Water Resour.*, **33**, 923–932.

Rincon, R. S. and R. H. Lang, 2002: Microwave link dual-wavelength measurements of path-average attenuation for the estimation of drop size distribution and rainfall. *IEEE T. Geosci. Remote Sens.*, **40 (4)**, 760–770.

Rinehart, R. E., 1979: Internal storm motions from a single non-Doppler weather radar. Ph.D. thesis, Colorado State University, 280 pp.

Ripley, B. D., 1987: *Stochastic Simulation*. Wiley, 237 pp.

Robertson, F. R., D. E. Fitzjarrald, and C. D. Kummerow, 2003: Effects of uncertainty in TRMM precipitation radar path integrated attenuation on interannual variations of tropical oceanic rainfall. *Geophys. Res. Lett.*, **30**, L016416, doi:10.1029/2002GL016416.

Rodríguez-Iturbe, I. and P. S. Eagleson, 1987: Mathematical models of rainstorm events in space and time. *Water Resour. Res.*, **23 (1)**, 181–190.

Roe, G. H., 2005: Orographic precipitation. *Ann. Rev. Earth Planet Sci.*, **33**, 645–671.

Rosewell, C. J., 1986: Rainfall kinetic energy in Eastern Australia. *J. Climate Appl. Meteor.*, **25**, 1695–1701.

Ruf, C. S., K. Aydin, S. Mathur, and J. P. Bobak, 1996: 35-GHz dual-polarization propagation link for rain-rate estimation. *J. Atmos. Oceanic Technol.*, **13 (2)**, 409–425.

Sauvageot, H. and J. P. Lacaux, 1995: The shape of averaged drop size distributions. *J. Atmos. Sci.*, **52 (8)**, 1070–1083.

Schleiss, M. and A. Berne, 2010: Identification of dry and rainy periods using telecommunciation microwave links. *IEEE Geosci. Remote Sens. Lett.*, **7 (3)**, 611–615, doi: 10.1109/LGRS.2010.2043052.

Schleiss, M., A. Berne, and R. Uijlenhoet, 2009: Geostatistical simulation of 2d fields of raindrop size distributions at the meso-γ scale. *Water Resour. Res.*, **45**, W07415, doi:10.1029/2008WR007545.

Seifert, A., 2005: On the shape-slope relation of drop size distributions in convective rain. *J. Appl. Meteor.*, **44**, 1146–1151.

Sekhon, R. S. and R. C. Srivastava, 1971: Doppler radar observations of drop-size distributions in a thunderstorm. *J. Atmos. Sci.*, **28 (6)**, 983–994.

Sempere-Torres, D., J. M. Porrà, and J.-D. Creutin, 1994: A general formulation for raindrop size distribution. *J. Appl. Meteor.*, **33 (12)**, 1494–1502.

Seo, D.-J., 1998: Real-time estimation of rainfall fields using rain gauge data under fractional coverage conditions. *J. Hydrol.*, **208**, 25–36.

Sevruk, 1999: Adjustment of tipping-bucket precipitation gauge measurements. *Atmos. Res.*, **42 (1-4)**, 237–246.

Shah, S. M. S., P. E. O'Connel, and J. R. M. Hosking, 1996: Modelling the effects of spatial variability in rainfall on catchment response. 1. Formulation and calibration of a stochastic rainfall field model. *J. Hydrol.*, **175 (1)**, 67–88.

Simpson, J., R. F. Adler, and G. R. North, 1988: A proposed Tropical Rainfall Measuring Mission (trmm) satellite. *Bull. Amer. Meteor. Soc.*, **69 (3)**, 278–295.

Sivakumar, B., S. Sorooshian, H. V. Gupta, and X. Gao, 2001: A chaotic approach to rainfall disaggregation. *Water Resour. Res.*, **37 (1)**, 61–72.

Sivapalan, M. and E. F. Wood, 1987: A multidimensional model of nonstationary space-time rainfall at the catchment scale. *Water Resour. Res.*, **23 (7)**, 1289–1299.

Smith, J. A., E. Hui, M. Steiner, M. L. Baeck, W. Krajewski, and A. A. Ntelekos, 2009: Variability of rainfall rate and raindrop size distributions in heavy rain. *Water Resour. Res.*, **45 (W04430)**, doi:10.1029/2008WR006840.

Srivastava, M. and T. Hui, 1987: On assessing multivariate normality based on Shapiro-Wilk W statistic. *Statistics and Probability Letters*, **5**, 15–18.

Srivastava, R., 1971: Size distributions of raindrops generated by their breakup and coalescence. *J. Atmos. Sci.*, **28 (3)**, 410–415.

Syed, T. H., V. Lakshmi, E. Paleologos, D. Lohmann, K. Mitchell, and J. S. Famiglietti, 2004: Analysis of process controls in land surface hydrological cycle over the continental United States. *J. Geophys. Res.*, **109**, D22 105, doi:10.1029/2004JD004640.

Tabary, P., 2007: The new french operational radar rainfall product. Part I: methodology. *Weather Forecast.*, **22 (3)**, 393–408, doi:10.1175/WAF1004.1.

Taylor, G. I., 1938: The spectrum of turbulence. *Proc. Roy. Soc. London*, **164 (919)**, 476–490.

Testud, J., E. Le Bouar, E. Obligis, and M. Ali-Mehenni, 2000: The rain profiling algorithm applied to polarimetric weather radar. *J. Atmos. Oceanic Technol.*, **17 (3)**, 332–356.

Testud, J., S. Oury, R. A. Black, P. Amayenc, and X. Dou, 2001: The concept of "normalized" distribution to describe raindrop spectra: a tool for cloud physics and cloud remote sensing. *J. Appl. Meteor.*, **40 (6)**, 1118–1140.

Tokay, A. and D. A. Short, 1996: Evidence from tropical raindrop spectra of the origin of rain from stratiform versus convective clouds. *J. Appl. Meteor.*, **35 (3)**, 355–371.

Tuttle, J. D. and G. B. Foote, 1990: Determination of the boundary layer airflow from a single doppler radar. *J. Atmos. Oceanic Technol.*, **7**, 218–232.

Uijlenhoet, R. and A. Berne, 2008: Stochastic simulation experiment to assess radar rainfall retrieval uncertainties associated with attenuation and its correction. *Hydrol. Earth Syst. Sci.*, **12 (2)**, 587–601.

Uijlenhoet, R., J. M. Porrà, D. Sempere Torres, and J.-D. Creutin, 2006: Analytical solutions to sampling effects in drop size distribution measurements during stationary rainfall: estimation of bulk rainfall variables. *J. Hydrol.*, **328**, 65–82, doi:10.1016/j.jhydrol.2005.11.043.

Uijlenhoet, R., M. Steiner, and J. A. Smith, 2003: Variability of raindrop size distributions in a squall line and implications for radar rainfall estimation. *J. Hydrometeor.*, **44 (4)**, 43–61.

Uijlenhoet, R., J. N. M. Stricker, P. J. J. F. Torfs, and J.-D. Creutin, 1999a: Towards a stochastic model of rainfall for radar hydrology: testing the Poisson homogeneity hypothesis. *Phys. Chem. Earth*, **24 (6)**, 747–755.

Uijlenhoet, R., et al., 1999b: HYDROMET Integrated Radar Experiment (HIRE): experimental setup and first results. *Proc. 29th AMS Conf. Radar Meteorol.*, Montréal, Canada, 926–930.

Ulbrich, C. W., 1983: Natural variations in the analytical form of the raindrop-size distribution. *J. Climate Appl. Meteor.*, **22 (10)**, 1764–1775.

Ulbrich, C. W., 1985: The effects of drop size distribution truncation on rainfall integral parameters and empirical relations. *J. Climate Appl. Meteor.*, **24**, 580–590.

Ulbrich, C. W. and D. Atlas, 1998: Rainfall microphysics and radar properties: analysis methods for drop size spectra. *J. Appl. Meteor.*, **37 (9)**, 912–923.

Ulbrich, C. W. and D. Atlas, 2007: Microphysics of raindrop size spectra: tropical continental and maritime storms. *J. Appl. Meteor. Climate*, **46**, 1777–1791.

Upton, G. J. G., A. R. Holt, R. J. Cummings, A. R. Rahimi, and J. W. F. Goddard, 2005: Microwave links: the future for urban rainfall measurements? *Atmos. Res.*, **77 (1-4)**, 300–312.

van der Vaart, A. W., 1998: *Asymptotic Statistics*. Cambridge University Press, 443 pp.

Van Dijk, A. I. J. M., L. A. Bruijnzeel, and C. J. Rosewell, 2002: Rainfall intensity-kinetic energy relationships: a critical literature appraisal. *J. Hydrol.*, **261 (1-4)**, 1–23.

Villarini, G. and W. F. Krajewski, 2010: Review of the different sources of uncertainty in single polarization radar-based estimates of rainfall. *Surveys in Geophysics*, **31 (1)**, 107–129.

Waldvogel, A., 1974: The N_0 jump of raindrop spectra. *J. Atmos. Sci.*, **31 (4)**, 1067–1078.

Waymire, E., V. K. Gupta, and I. Rodriguez-Iturbe, 1984: A spectral theory of rainfall intensity at the meso-β scale. *Water Resour. Res.*, **20 (10)**, 1453–1465.

Weiss, L. L., 1964: Sequences of wet or dry days described by a Markov chain probability model. *Mon. Weather Rev.*, **92**, 169–176.

Wilks, D. S., 1999: Simultaneous stochastic simulation of daily precipitation, temperature and solar radiation at multiple sites in complex terrain. *Agr. and Forest Meteor.*, **96 (1-3)**, 85–101.

Wilks, D. S. and R. L. Wilby, 1999: The weather generation game: a review of stochastic weather models. *Prog. Phys. Geog.*, **23 (3)**, 329–357.

Willis, P. T., 1984: Functional fits to some observed drop size distributions and parameterization of rain. *J. Atmos. Sci.*, **41 (9)**, 1648–1661.

Wilson, J. M. and E. A. Brandes, 1979: Radar measurement of rainfall - a summary. *Bull. Amer. Meteor. Soc.*, **60 (9)**, 1048–1058.

Wojcik, R., D. McLaughlin, A. G. Konings, and D. Entekhabi, 2009: Conditioning stochastic rainfall replicates on remote sensing data. *IEEE T. Geosci. Remote Sens.*, **47 (8)**, 2436–2449.

Xie, P. and P. A. Arkin, 1997: Global precipitation: a 17-year monthly analysis based on gauge observations, satellite estimates and numerical model outputs. *Bull. Amer. Meteor. Soc.*, **78 (11)**, 2539–2558.

Xue, M., K. K. Droegemeier, and V. Wong, 2000: The Advanced Regional Prediction System (ARPS); A multi-scale nonhydrostatic atmospheric simulation and prediction model. Part I: model dynamics and verification. *Meteorol. Atmos. Phys.*, **75 (3-4)**, 161–193, doi:10.1007/s007030070003.

Yaglom, A., 2004: *An Introduction to the Theory of Stationary Random Functions*. Dover Publications, 256 pp.

Yue, D. and Q. Han, 2005: Delay-dependent exponential stability of stochastic systems with time-varying delay, nonlinearity, and Markovian switching. *IEEE Trans. Autom. Control*, **50 (2)**, 217–222.

Zawadzki, I., 1975: On radar-raingauge comparison. *J. Appl. Meteor.*, **14 (8)**, 1430–1436.

Zhang, G., J. Vivekanandan, and E. Brandes, 2001: A method for estimating rain rate and drop size distribution from polarimetric radar measurements. *IEEE T. Geosci. Remote Sens.*, **39 (4)**, 830–841.

Zinevich, A., P. Alpert, and H. Messer, 2008: Estimation of rainfall fields using commercial microwave communication networks of variable density. *Adv. Water Resour.*, **31 (11)**, 1470–1480, doi:10.1016/j.advwatres.2008.03.003.

Zinevich, A., P. Alpert, and H. Messer, 2009: Frontal rainfall observation by commercial microwave communication network. *J. Appl. Meteor. Climate*, **48**, 1317–1334.

Zinevich, A., H. Messer, and P. Alpert, 2010: Prediction of rainfall intensity measurement errors using commercial microwave communication links. *Atmos. Meas. Tech.*, **3**, 1385–1402.

Marc Schleiss

Education

2008–2012 **PhD in hydrometeorology**, *École Polytechnique Fédérale de Lausanne*, Switzerland.

2006–2008 **M.Sc. in mathematics**, *École Polytechnique Fédérale de Lausanne*, Switzerland.

2003–2006 **Bachelor in mathematics**, *École Polytechnique Fédérale de Lausanne*, Switzerland.

2000–2003 **Maturité fédérale scientifique**, *Gymnase de Burier*, Switzerland.

PhD thesis

Title "Variability of the Rain Drop Size Distribution: Stochastic Simulation and Application to Telecommunication Microwave Links"

Supervisor Prof. A. Berne

Other research projects

Autumn 2007 Master thesis: "R-estimators in regression and extension to generalized linear models"

Spring 2007 Semester project: "Random algorithms"

Autumn 2006 Semester project: "Optimal frequency allocation in telecommunication networks"

Spring 2006 Semester project: "Generalized linear models"

Autumn 2005 Semester project: "Interacting particle systems"

Experience

2008-2011 **Teaching assistant**, *École Polytechnique Fédérale de Lausanne*, Switzerland.
Teaching assistant for "Spatial statistics and analysis", Prof. Golay and Prof. Berne

Summer 2006 **Examination supervisor**, *École Secondaire de Montreux*, Switzerland.
External expert for maths exams.

Summer 2004 **Military service**, Bern, Switzerland.
Appointed to private first class. Service accomplished successfully.

Languages

French **Mother tongue**

Swiss German **Mother tongue**

German **Fluent**

English **Fluent**

Portuguese **Intermediate**

✉ *Vers-chez-Cochard 4 – 1807 Blonay, Switzerland*
• *+41 79 545 36 26* • ☎ *+41 21 943 39 35* • *marc.schleiss@epfl.ch*

Computer skills

OS	Windows (XP/Vista/7), Linux
Office	Word, Excel, Power Point, Open Office, LaTex
Programming	R, C++, Matlab, Fortran-77

Other skills

Mathematics	Applied statistics, multivariate statistics, data mining, modeling, stochastic simulation, time series analysis, spatial statistics, Geostatistics, Bayesian statistics, optimization, statistical programming in "R".
Remote sensing	Rainfall estimation using polarimetric weather radars, rain gauges, disdrometers, microwave links and satellites. Rainfall simulation and disaggregation.

Publications

2012	"Polarimetric X-band radar network simulator": Domaszczynski, P., W. Krajewski, A. Kruger, M. Schleiss, A. Berne and R. Uijlenhoet, submitted to *Water Ressour. Res.*, Feb 2012
2012	"Stochastic space-time disaggregation of rainfall into DSD fields": Schleiss, M. and A. Berne, submitted to *J. Hydrometeorol.*, Jan 2012
2012	"Using Markov switching models to infer dry and rainy periods from telecommunication microwave link signals": Wang, Z., M. Schleiss, J. Jaffrain, A. Berne and J. Rieckermann, submitted to *Atmos. Meas. Tech.*, Dec 2011
2012	"Stochastic simulation of intermittent DSD fields in time": Schleiss, M., J. Jaffrain and A. Berne, *J. Hydrometeorol.*, vol.13, No.2, 621-637
2012	"Scaling analysis of the variability of the rain drop size distribution at small scale": Berne, A., J. Jaffrain and M. Schleiss, *Adv. Water Resour.*, in press
2011	"Statistical analysis of rainfall intermittency at small scales": Schleiss, M., J. Jaffrain and A. Berne, *Geophys. Res. Lett.*, vol.38, L18403
2010	"Identification of dry and rainy periods using telecommunication microwave links": Schleiss, M. and A. Berne, *IEEE Geosci. Remote Sens. Lett.*, vol.7, No.3, 611-615
2009	"Geostatistical simulation of 2D fields of raindrop size distributions at the meso-γ scale": Schleiss, M., A. Berne and R. Uijlenhoet, *Water Ressour. Res.*, vol.45, W07415

Awards & Media

March 2011	I was interviewed and filmed by the "Télévision suisse romande" about my work on urban rainfall estimation using telecommunication microwave links, in collaboration with the Swiss Aquatic Research (EAWAG) in Dübendorf, Switzerland.
January 2011	I was interviewed by the "Radio suisse romande" about the potential of rainfall estimation using telecommunication microwave links, Lausanne, 26 Jan 2011.
September 2010	Finalist for the best article at the European Conference on Radar in Meteorology and Hydrology, Sibiu, Romania, 6-10 Sept 2010.

✉ *Vers-chez-Cochard 4 – 1807 Blonay, Switzerland*
• *+41 79 545 36 26* • ☎ *+41 21 943 39 35* • *marc.schleiss@epfl.ch*