

**SPATIO-TEMPORAL MODELING OF PRECIPITATION IN A  
HIGH TEMPORAL RESOLUTION FOR URBAN  
HYDROLOGICAL APPLICATIONS**

**VON DER FAKULTÄT FÜR BAUINGENIEURWESEN UND GEODÄSIE  
DER  
GOTTFRIED WILHELM LEIBNIZ UNIVERSITÄT HANNOVER**

**ZUR ERLANGUNG DES GRADES**

**Doktor der Ingenieurwissenschaften  
Dr.-Ing.**

**genehmigte Dissertation**

**von**

**Ana Claudia Callau Poduje, M. Sc.**

**geboren am 25.09.1979 in Rio de Janeiro**

**2018**

**Referent: Prof. Dr. -Ing. Uwe Haberlandt**  
**Korreferent: Prof. Dr. rer. nat. Dr.-Ing. András Bárdossy**

**Tag der Promotion: 20.03.2018**

## **ACKNOWLEDGEMENTS**

To my doctoral Thesis supervisor Prof. Haberlandt, for giving me the opportunity to learn from his experience and for being an outstanding researcher to which I will always look up to. For his structured, precise and very reliable way of working, for taking care of the Doctor students, and for always pushing us to take the best out of our work. For teaching me the highest level of making research. I am very grateful to have such a great scientific mentor.

To Rafael Seoane for his valuable scientific guidance during the very early steps of my career. For introducing me to the amazing world of research through statistics applied to hydrology.

To my colleagues and friends in WAWI, for sharing a harmonic day to day life together, for their solidarity in contributing with the common tasks of our Institute and for the fruitful discussions not only about research but also about life. Special gratefulness goes to Ross Pidoto, Hannes Müller, Bora Shehu, Stefan Plötner and Christian Berndt for proof-reading an earlier version of this document and for their valuable comments.

I would like to express my thankfulness to Benedikt Gräler for his constructive suggestions regarding the sections involving Vine copulas.

I am extremely grateful with the students that have helped me with many of the tasks involved during my research, like data pre-processing or running things over and over. Special thanks go to: Sören Leimbach, Raúl Villanueva, Tam Nguyen and Bora Shehu, who were all great people to work with and always willing to help with any task I would ask, no matter how tedious it would be.

To my family, to my parents for their unconditional support and reliance, to my sister Paula for being the best friend I could ask for. To my husband Matthias for his never-ending love, patience and charming company, for his motivating energy and positive spirit always encouraging this venture.



## DECLARATION

I, Ana Claudia Callau Poduje, hereby declare that:

1. I know the regulations for doctoral candidates and have met all the requirements. I agree with an examination under the provisions of the doctoral regulations.
2. I have completed the Thesis independently, and where I have consulted other works, to the best of my knowledge and belief, this is always clearly attributed.
3. I have not paid any third party to contribute to the content of my dissertation (this means that the scientific work may has not been acquired or conveyed, either in part or in full, by a third party for payment or any other consideration).
4. This work contains no material which has been accepted for the award of any other degree or diploma at any other university or tertiary institution.
5. The same or partly similar work has not been submitted at any other university or academic institution.
6. I agree that my Thesis can be used for verifying the compliance with scientific standards, in particular using electronic data processing programs.

Finally, I would like to state that part of the material presented within this Thesis has already been published in advance. This is the publication by Callau Poduje & Haberlandt (2017).

22.03.2018.....

Date

  
.....  
Ana Claudia Callau Poduje



*“HERE THE IMPORTANCE OF EVERY SINGLE DROP OF RAIN FALLING ON A CITY IS POINTED OUT. WHEREVER IT FALLS IS RIGHT. THAT IS A SUPREME LAW WHICH CANNOT BE EMULATED BY MAN. THE CITY IN THE BACKGROUND IS SEEN FROM ABOVE. HERE, TOO, THERE IS AN ABSOLUTE HARMONY BETWEEN RAINDROPS AND THE CITY WHICH HAS GROWN ORGANICALLY AND CONSISTS OF INDIVIDUAL CELLS.” (HUNDERTWASSER, 1996)*

**HUNDERTWASSER**



## ABSTRACT

Long and continuous series of rainfall data in a high temporal resolution (sub-hourly) are relevant for several purposes, namely for quantifying erosivity, hydrological applications involving urban or small, steep rural catchments, engineering design of flash flood control structures, etc. Unfortunately, available observations with such resolutions are usually short in time and restricted to some locations. Precipitation models can be used to overcome these limitations by generating long time series which are not constrained by the length of the observed data and for locations without observations. Here, a stochastic model is developed for this purpose. This model involves an alternating renewal process that describes a system consisting of rainfall events which are differentiated by two system states, wet or dry. Events are further characterized by variables describing amounts, durations and peak intensities which are simulated stochastically.

The model developed in this Thesis is based on an existing one and includes the introduction of major improvements, namely i) the incorporation of multivariate distribution functions called copulas, ii) seasonality, iii) modeling of small events, and iv) multi-site synthesis. A special focus is given to properly reproduce the extreme values. The generation of rainfall time series using this type of models is straightforward for single sites and the potential of using copulas to model the joint behavior of some of the involved variables is analyzed. An extension of the model for simulating rainfall in several sites simultaneously is presented; the proposed extension involves Vine copulas. Finally, a copula-based approach is developed for regionalization of the model, i.e. for estimating different model parameters in locations without observations based on site descriptors.

Rainfall series registered in 104 stations located in different regions in Germany are used to develop and test the proposed methodologies. The available data consists of registers with high temporal resolution records, i.e. rainfall data with a temporal resolution of 5 minutes, and a temporal coverage of 6 to 21 years. The descriptors include non-climatic and climatic information available for entire Germany.

The evaluation of the model is performed based on ensembles of numerous long synthetic time series which are compared to observed ones. The different applications of the proposed methodologies are evaluated in terms of their capability to reproduce long term rainfall properties along with extreme value statistics. Results from the single site application show that properly modeling the joint behavior of amount and duration of rainfall events is essential for reproducing the observed properties, especially for the extreme events. Copulas are

found to be an advantageous tool in terms of properly reproducing this joint behavior; however selecting the proper copula is a crucial aspect. The proposed development of the model to multi-site applications involves additional information that accounts for the spatial extension of the rainfall events. Vine copulas enable modeling multiple variables via pair-copulas which results in a high flexibility to reproduce different dependence structures. Therefore this type of copulas is found to be an appropriate tool for simulating events in several stations due to their capability of representing the joint behavior of rainfall characteristics for the different cases. Furthermore, spatial consistency criteria resulting from long synthetic time series are compared with observations indicating that the proposed method represents a valuable extension of the model for multi-site simulations. Regionalization of the model is evaluated by cross-validation. Results indicate that modeling the continuous time series in locations without observations is challenging, whereas extreme events are reproduced with good agreement. The copula-based regionalization technique proved to be remarkably robust to the inclusion of new stations. The proposed techniques are compared with commonly used methods applied by the hydrological community for similar purposes, namely multi-linear regression (for single site model and regionalization), regional frequency analysis (for regionalization) and simulated annealing (for multi-site model). A comparison with the current design practice is as well included. Finally, the application of the models to simulate time series for urban hydrological purposes is provided.

Overall, this Thesis contributes to provide more reliable rainfall time series for applications which require rainfall data in a high temporal resolution, in particular urban hydrological applications. The copula-based methods have shown to perform very satisfactory in the simulation of long time series of rainfall events and have the advantage that their complexity is not affected by the temporal resolution. The developed models outperform the available design practices for particular events; furthermore the proposed methods exhibit a robust behavior both for extreme value estimations and for flood simulations.

**Keywords:** stochastic modeling of precipitation, copulas, regionalization, multi-site synthesis, continuous simulation, high temporal resolution.

## KURZFASSUNG

Lange, kontinuierliche Niederschlagszeitreihen in hoher zeitlicher Auflösung (d.h. weniger als eine Stunde) werden für verschiedenste Anwendungen benötigt, z.B. für die Quantifizierung der Erosivität, für die Untersuchung urbanhydrologischer Fragestellungen, der Modellierung kleiner Einzugsgebiete mit großen Höhenunterschieden oder der Dimensionierung von baulichen Strukturen zur Kontrolle von Sturzfluten. Häufig existieren solch hochaufgelöste Zeitreihen jedoch nur an wenigen Standorten und für kurze Zeiträume. Niederschlagsmodelle (oder -generatoren) werden angewandt, um diese Limitationen zu beheben. So bieten diese Modelle die Möglichkeit, lange Niederschlagszeitreihen, die nicht von der Länge der verfügbaren Beobachtungsdaten abhängig sind, zu generieren. Darüber hinaus können hochaufgelöste Daten auch für unbeobachtete Standorte erzeugt werden.

In dieser Doktorarbeit wird ein stochastischer Ansatz vorgestellt, um kontinuierliche, zeitlich hochaufgelöste Niederschlagsdaten zu generieren. Das entwickelte Modell nutzt einen 'Alternating Renewal Process', der ein System als eine Aneinanderreihung von Niederschlagsereignissen beschreibt, die durch zwei verschiedene Zustände – trocken oder nass – charakterisiert sind. Darüber hinaus werden die Ereignisse durch stochastisch simulierte Variablen wie Niederschlagsmenge, Niederschlagsdauer und maximale Niederschlagsintensität beschrieben.

Das in dieser Arbeit entwickelte Modell basiert auf einem bereits existierenden Niederschlagsgenerator, beinhaltet aber zahlreiche Erweiterungen und Verbesserungen, namentlich i) die Einbeziehung multivariater Verteilungsfunktionen (Copulas); ii) die Betrachtung von Saisonalität; iii) der Möglichkeit kleinere Niederschlagsereignisse zu modellieren und iv) einer räumlichen Synthese ('Multi-site'). Ein Hauptschwerpunkt wurde auf der korrekten Reproduktion von Extremereignissen gesetzt. Weiterhin wurde das Potential von Copulas zur simultanen Berücksichtigung mehrerer Variablen detailliert untersucht. Die Arbeit beinhaltet darüber hinaus eine Erweiterung des bestehenden Modells durch *Vine Copulas* für die simultane Simulation von Niederschlag an mehreren Standorten. Abschließend wird ein Copula-basierter Ansatz für Regionalisierung des Modells via geeigneter Deskriptoren entwickelt, d.h. es wird eine Ableitung verschiedener Modellparameter für unbeobachtete Standorte ermöglicht.

Niederschlagszeitreihen von 104 Stationen aus ganz Deutschland werden verwendet, um die vorgeschlagenen Methoden zu entwickeln und testen. Dabei werden lediglich Daten verwendet, die hochaufgelöst, d.h. in einer zeitlichen Auflösung von 5 Minuten und einer

Länge von 6 bis 21 Jahren, vorliegen. Die Deskriptoren beinhalten sowohl klimatische als auch nicht-klimatische Informationen, die für ganz Deutschland verfügbar sind. Die Evaluierung des Modells basiert auf Ensembles zahlreicher langer, synthetischer Zeitreihen, die mit beobachteten verglichen werden. Verschiedene Anwendungen der entwickelten Methoden werden auf Grundlage der Fähigkeit bestimmte Niederschlagscharakteristiken zu reproduzieren und Extremwertstatistiken beurteilt.

Die Ergebnisse dieser Untersuchung zeigen, dass für einzelne Standorte das simultane Betrachten von Niederschlagsmenge und Niederschlagsdauer essentiell ist, um das beobachtete Verhalten korrekt zu reproduzieren. Dies ist insbesondere für Extremereignisse der Fall. Es zeigt sich, dass Copulas eine wertvolle Methode darstellen, die dieses simultane Betrachten ermöglicht; der wichtigste Aspekt dabei ist die Wahl einer geeigneten Copula. *Vine Copulas* erlauben die Modellierung mehrerer Variablen gleichzeitig mittels Paar-Copulas (pair copulas); dadurch bieten diese eine hohe Flexibilität in Bezug auf verschiedene Abhängigkeitsstrukturen. Dieser Typ Copulas ist daher basierend auf den Ergebnissen dieser Arbeit ein geeignetes Werkzeug für die Niederschlagssimulation an mehreren Standorten, da für verschiedene Fälle mehrere Niederschlagscharakteristiken simultan betrachtet werden können. Weiterhin zeigt die Auswertung der Kriterien in Bezug auf räumliche Konsistenz, dass das entwickelte Modell eine wertvolle Erweiterung für die Einbeziehung vieler Standorte darstellt („Multi-site“). Anschließend wurden die Ergebnisse regionalisiert und mittels Kreuzvalidierung evaluiert. Hier zeigt sich, dass die Generierung kontinuierlicher Zeitreihen an Standorten ohne Beobachtungen schwierig realisierbar ist. Extremereignisse können allerdings gut reproduziert werden. In Bezug auf eine Einbeziehung neuer Stationen ist die Copula-basierte Methode außerordentlich robust. Um dies zu zeigen, wurde ein Vergleich zwischen routinemäßig verwendeten hydrologischen Evaluierungsmethoden (z.B. multilineare Regression, regionale Häufigkeitsanalyse und simulated annealing) durchgeführt. Abschließend wird die Anwendung des Modells für die Simulation von Zeitreihen für urbanhydrologische Fragestellungen untersucht.

Diese Doktorarbeit trägt dazu bei, um verlässlichere Niederschlagszeitreihen für alle Fragestellungen, in denen eine hohe zeitliche Auflösung benötigt wird, bereitzustellen. Dies ist besonders hilfreich bei der Untersuchung von urbanhydrologischen Aspekten. Copula-basierte Methoden sind enorm vorteilhaft für die Generierung von Niederschlagsereignissen; die Komplexität ist dabei nicht beeinflusst von der zeitlichen Auflösung, was einen weiteren Vorteil darstellt. Die in dieser Arbeit erstellten Modelle zeigen eine bessere Performance im Vergleich zu den derzeit verfügbaren Methoden zur Abschätzung von Niederschlagsereignissen bestimmter Wiederkehrintervalle (z.B. KOSTRA). Außerdem sind

die vorgeschlagenen Ansätze äußerst robust für sowohl die Abschätzung von Extremwerten als auch die Simulation von Überflutungen.

**Schlagworte:** Stochastische Niederschlagsmodellierung, Copulas, Regionalisierung, Räumliche Synthese, Kontinuierliche Simulation, hohe zeitliche Auflösung.



## ACRONYMS

<b>AIC</b>	Akaike Information Criterion
<b>ARP</b>	Alternating renewal process
<b>BAWU</b>	German state of Baden Württemberg
<b>BL</b>	Barlett-Lewis model
<b>CvM</b>	Cramér-von Mises test
<b>DEM</b>	Digital Elevation Map
<b>DSD</b>	Dry spell duration
<b>DSDmin</b>	Minimum dry spell duration for event definition
<b>DWD</b>	Deutscher Wetterdienst (German National Weather Service)
<b>DWF</b>	Dry weather flow
<b>EDK</b>	External drift kriging
<b>IDF</b>	Intensity duration frequency curves
<b>IID</b>	Independent identically distributed
<b>KG</b>	Köppen-Geiger climate classification
<b>KOSTRA</b>	Koordinierte Starkniederschlagsregionalisierung
<b>KS</b>	Kolmogorov-Smirnov test
<b>LMI</b>	L-moment order i
<b>LOOCV</b>	Leave-one out cross-validation
<b>MAP</b>	Mean Annual Precipitation
<b>ML</b>	Maximum likelihood
<b>MLR</b>	Multiple linear regression
<b>m.a.s.l.</b>	Meters above sea level referred to tide gauge Amsterdam (DHHN92)
<b>mNN</b>	Meters above German Standard Zero (Normalnull)
<b>MW</b>	Mann-Whitney test
<b>NN</b>	Nearest Neighbor
<b>NS</b>	German state of Lower Saxony (Niedersachsen)
<b>OK</b>	Ordinary kriging
<b><math>p_{00}</math></b>	Joint probability of no-rain (0) and no-rain (0) at two stations
<b><math>p_{01}</math></b>	Joint probability of no-rain (0) and rain (1) at two stations
<b><math>p_{10}</math></b>	Joint probability of rain (1) and no-rain (0) at two stations
<b><math>p_{11}</math></b>	Joint probability of rain (1) and rain (1) at two stations
<b>PDF</b>	Probability distribution function
<b>RFA</b>	Regional frequency analysis
<b>RSE</b>	Relative standard error
<b>SCS</b>	Soil Conservation Service curve number method
<b>SA</b>	Simulated annealing
<b>SD</b>	Site Descriptors
<b>SWMM</b>	Storm Water Management Model
<b>SWT</b>	Storm-water tank
<b>Tr</b>	Return period
<b>WSA</b>	Wet spell amount
<b>WSAmin</b>	Minimum wet spell amount for event definition
<b>WSD</b>	Wet spell duration
<b>WSI</b>	Wet spell intensity
<b>WSImin</b>	Minimum wet spell intensity for event definition
<b>WSPeak</b>	Wet spell peak amount
<b>WSTpeak</b>	Wet spell time to peak
<b>WWTP</b>	Waste Water Treatment Plant

## CONTENTS

<u>1</u>	<u>INTRODUCTION</u>	<u>1</u>
1.1	BACKGROUND AND MOTIVATION	1
1.2	OBJECTIVES	3
1.3	STRUCTURE	4
<u>2</u>	<u>STATE OF ART</u>	<u>5</u>
2.1	PRECIPITATION MODELS	5
2.2	REGIONALIZATION OF PRECIPITATION	9
2.3	OPEN QUESTIONS	11
<u>3</u>	<u>METHODS</u>	<u>13</u>
3.1	ALTERNATING RENEWAL PROCESS	13
3.2	COPULAS – GENERAL CONCEPT	14
3.3	SINGLE SITE PRECIPITATION MODEL	21
3.3.1	EXISTING MODEL (OLD MODEL)	21
3.3.2	PROPOSED MODEL	22
3.4	MULTI-SITE PRECIPITATION MODEL	27
3.4.1	PROPOSED MODEL	27
3.4.2	ALTERNATIVE APPROACH: SIMULATED ANNEALING	32
3.5	REGIONALIZATION OF PRECIPITATION	33
3.5.1	GENERAL CONCEPTS	33
3.5.2	PROPOSED METHOD	34
3.5.3	ALTERNATIVE APPROACHES: MLR AND RFA	37
3.6	APPROACH FOR VALIDATING THE MODELS	38
3.6.1	DIRECT VALIDATION (RAINFALL BASED)	39
3.6.2	CROSS-VALIDATION	43
3.6.3	UNCERTAINTY ANALYSIS	44
3.6.4	APPLICATION TO URBAN HYDROLOGY	47
3.7	SUMMARY COMPARING EXISTING AND PROPOSED PRECIPITATION MODEL	49
<u>4</u>	<u>DATA</u>	<u>51</u>
4.1	STUDY REGION	51
4.2	STATIONS FOR SINGLE SITE MODEL	58
4.3	STATIONS FOR MULTI-SITE MODEL	62
4.4	STATIONS FOR REGIONALIZATION	64
4.5	SITE DESCRIPTORS (SDs)	65

4.6	DESIGN PRACTICE: IDF'S KOSTRA	67
4.7	ARTIFICIAL URBAN SYSTEM	69
<u>5</u>	<u>RESULTS</u>	<u>71</u>
5.1	GENERAL RESULTS	71
5.2	SINGLE SITE PRECIPITATION MODEL	74
5.3	MULTI-SITE PRECIPITATION MODEL	83
5.4	REGIONALIZATION OF PRECIPITATION	96
5.4.1	PRECEDING ANALYSIS	97
5.4.2	COPULA-BASED METHOD	99
5.4.3	MULTI-LINEAR REGRESSION (MLR)	101
5.4.4	REGIONAL FREQUENCY ANALYSIS (RFA)	101
5.4.5	COMPARISON OF DIFFERENT METHODS	102
5.5	UNCERTAINTY ANALYSIS	107
<u>6</u>	<u>APPLICATION TO URBAN HYDROLOGY</u>	<u>113</u>
6.1	UNCERTAINTY OF INPUT DATA IN THE RESPONSE OF URBAN CATCHMENTS	113
6.2	EFFECT OF THE INTERNAL DISTRIBUTION IN THE RESPONSE OF URBAN CATCHMENTS	115
6.3	SINGLE SITE PRECIPITATION MODEL	118
6.4	REGIONALIZATION OF PRECIPITATION	120
<u>7</u>	<u>SYNTHESIS</u>	<u>122</u>
7.1	SUMMARY AND CONCLUSIONS	122
7.2	FUTURE RESEARCH	130
	<u>REFERENCES</u>	<u>131</u>
	<u>LIST OF PUBLICATIONS</u>	<u>139</u>
	<u>APPENDIX</u>	<u>141</u>

## LIST OF FIGURES

FIGURE 3.1: CONCEPTUAL ILLUSTRATION OF THE ALTERNATING RENEWAL PROCESS APPLIED FOR RAINFALL MODELING.	14
FIGURE 3.2: CONCEPTUAL ILLUSTRATION OF A COPULA IN A BIVARIATE CONTEXT (UNDER $N(0,1)$ MARGINS FOR VISUALIZATION PURPOSES).	15
FIGURE 3.3: ABILITY OF DIFFERENT COPULAS TO MODEL TAIL DEPENDENCIES, SYMMETRIC AND ASYMMETRIC RELATIONSHIP BETWEEN VARIABLES. RED LINE REPRESENTS THE AXIS OF SYMMETRY.	17
FIGURE 3.4: EXAMPLE OF COMPONENTS OF A VINE COPULA IN A TRIVARIATE CONTEXT. LEFT: MARGINAL DENSITIES, UNCONDITIONAL AND CONDITIONAL PAIR COPULA DENSITIES (UNDER $N(0,1)$ MARGINS FOR VISUALIZATION PURPOSES). RIGHT: VINE TREE STRUCTURE.	20
FIGURE 3.5: EXTERNAL (LEFT) AND INTERNAL (RIGHT) STRUCTURES OF RAINFALL EVENTS FOR OLD MODEL.	21
FIGURE 3.6: EXTERNAL (LEFT) AND INTERNAL (RIGHT) STRUCTURES OF RAINFALL EVENTS FOR PROPOSED MODEL.	23
FIGURE 3.7: STOCHASTIC MODELING OF DSD (LEFT) AND WSD-WSA (RIGHT).	24
FIGURE 3.8: STOCHASTIC MODELING OF WSTPEAK (LEFT) AND WSI-WSPEAK (RIGHT).	25
FIGURE 3.9: TEMPORAL LOCATION OF SMALL EVENTS WITH RESPECT TO EVENTS (LEFT) AND PAIRS OF WSA AND WSD DESCRIBING SMALL EVENTS (RIGHT).	26
FIGURE 3.10: PAIRS PLOT OF PSEUDO-VALUES OF EVENTS REGISTERED IN TWO STATIONS SIMULTANEOUSLY. DIAGONAL: HISTOGRAM OF MARGINALS. UPPER TRIANGULAR MATRIX: BIVARIATE SCATTER PLOTS WITH KENDALL'S TAU BETWEEN PAIRS IN RED. LOWER TRIANGULAR MATRIX: CONTOUR PLOTS WITH STANDARD NORMAL MARGINS.	28
FIGURE 3.11: COMPONENTS OF A VINE COPULA TO MODEL EVENTS OCCURRING SIMULTANEOUSLY IN TWO STATIONS. LEFT: MARGINAL DENSITIES, UNCONDITIONAL AND CONDITIONAL PAIR COPULA DENSITIES (UNDER $N(0,1)$ MARGINS FOR VISUALIZATION). RIGHT: VINE TREE STRUCTURE.	29
FIGURE 3.12: PREDEFINED TREE STRUCTURE FOR THE VINE-COPULAS SIMULATING RAINFALL EVENTS OCCURRING SIMULTANEOUSLY IN TWO STATIONS (LEFT PLOT) AND THREE STATIONS (RIGHT PLOT).	30
FIGURE 3.13: CHARACTERISTICS OF RAINFALL EVENTS OCCURRING IN 1, 2 OR 3 STATIONS SIMULTANEOUSLY.	31
FIGURE 3.14: GRAPHICAL EXPLANATION OF THE PROCEDURE FOR A CASE WITH 3 SITE DESCRIPTORS.	36
FIGURE 3.15: EXAMPLE OF P-VALUES RESULTING FROM THE MANN-WHITNEY TEST (LEFT) AND CRAMÉR-VON MISSES TEST (RIGHT).	40

FIGURE 3.16: EXAMPLE OF MODELS UNBIASED (RED), OVERESTIMATING (ORANGE) AND UNDERESTIMATING (PURPLE) A PARTICULAR ATTRIBUTE ASSESSED ON ONE STATION (LEFT PLOTS) AND ON MANY STATIONS (RIGHT PLOTS).	46
FIGURE 3.17: EVALUATION OF INTERNAL STRUCTURE IN TERMS OF HYDROLOGICAL MODELING.	48
FIGURE 4.1: DIGITAL ELEVATION MAP OF GERMANY WITH LOCATION OF RAIN GAUGE STATIONS USED FOR DEVELOPING (24 CIRCLES) AND TRANSFERRING (22 TRIANGLES) THE SINGLE SITE MODEL, AND FOR REGIONALIZATION (81 CROSSES).	52
FIGURE 4.2: LEFT: MEAN ANNUAL PRECIPITATION (1981-2010) AND RIGHT: KÖPPEN-GEIGER CLIMATE CLASSIFICATION IN GERMANY WITH LOCATION OF ALL RAIN GAUGE STATIONS.	53
FIGURE 4.3: DESCRIPTION OF DATA AVAILABLE FOR SINGLE SITE MODEL DEVELOPMENT AND REGIONALIZATION.	54
FIGURE 4.4: AUTOCORRELATION OF VARIABLES DESCRIBING RAINFALL EVENTS OBSERVED IN NS AND BAWU. LEFT SEMI-PLOTS CORRESPOND TO SUMMER AND RIGHT ONES TO WINTER EVENTS.	55
FIGURE 4.5: AUTOCORRELATION OF WSD RESULTING FROM EVENTS OBSERVED IN NS ASSOCIATED TO DIFFERENT VALUES OF DSDMIN. LEFT SEMI-PLOTS CORRESPOND TO SUMMER AND RIGHT ONES TO WINTER EVENTS.	56
FIGURE 4.6: AUTOCORRELATION OF DSD RESULTING FROM EVENTS OBSERVED IN NS ASSOCIATED TO DIFFERENT VALUES OF DSDMIN. LEFT SEMI-PLOTS CORRESPOND TO SUMMER AND RIGHT ONES TO WINTER EVENTS.	56
FIGURE 4.7: CORRELATION COEFFICIENTS BETWEEN DIFFERENT CHARACTERISTICS DESCRIBING RAINFALL EVENTS OBSERVED IN NS AND BAWU. LEFT SEMI-PLOTS CORRESPOND TO SUMMER AND RIGHT ONES TO WINTER EVENTS.	57
FIGURE 4.8: TEMPORAL LOCATION OF SMALL EVENTS WITH RESPECT TO EVENTS FOR NS (LEFT PLOT) AND BAWU (RIGHT PLOT).	58
FIGURE 4.9: LOCATION OF RAIN GAUGE STATIONS USED FOR DEVELOPING THE SINGLE SITE MODEL (SEE NUMBERING IN TABLE 4.1). SIZE OF CIRCLES IS PROPORTIONAL TO MEAN ANNUAL PRECIPITATION.	60
FIGURE 4.10: DISTRIBUTION OF DAILY RAINFALL FOR STATIONS USED FOR DEVELOPING THE SINGLE SITE MODEL (SEE NUMBERING IN TABLE 4.1).	60
FIGURE 4.11: LOCATION OF RAIN GAUGE STATIONS AND GROUPS USED FOR DEVELOPING THE MULTI-SITE MODEL.	62
FIGURE 4.12: LOCATION OF RAIN GAUGE STATIONS USED FOR REGIONALIZATION OF THE MODEL.	64
FIGURE 4.13: MAIN CHARACTERISTICS OF TIME SERIES USED FOR REGIONALIZATION OF THE MODEL.	65
FIGURE 4.14: EXAMPLE OF GRID MAPS WITH SITE DESCRIPTORS USED FOR REGIONALIZATION OF THE MODEL.	66
FIGURE 4.15: DISTRIBUTION OF SOME SITE DESCRIPTORS FOR ALL STATIONS USED FOR REGIONALIZATION OF THE MODEL.	67

FIGURE 4.16: EXTREME VALUES PROVIDED BY KOSTRA FOR A RETURN PERIOD OF 10 YEARS AND A DURATION OF 5 MINUTES (LEFT) AND 24 HOURS (RIGHT).	68
FIGURE 4.17: KOSTRA REDUCTION FACTORS FOR DIFFERENT SIZES OF AREAS AND DURATIONS OF EVENTS.	69
FIGURE 4.18: SCHEMATIC REPRESENTATION FICTIONAL URBAN SYSTEM.	70
FIGURE 5.1: PAIRS OF PSEUDO-VALUES OF WSA AND WSD FOR ALL EVENTS (GRAY) AND ANNUAL EXTREME EVENTS (COLORS ACCORDING TO DURATIONS) RECORDED IN ALL STATIONS.	72
FIGURE 5.2: LEFT: PAIRS OF PSEUDO WSI AND WSPEAK FOR ALL EVENTS OBSERVED IN ALL STATIONS (GREY CIRCLES) AND SIMULATED WITH THE COPULA (DARK GREY SQUARES), KENDALL'S TAU CORRELATION COEFFICIENT IS SHOWN. RIGHT: HISTOGRAM OF WSTPEAK/WSD EXTRACTED FROM OBSERVED EVENTS WITH THE UNIFORM MODEL USED FOR THEIR SIMULATION	73
FIGURE 5.3: EXTERNAL STRUCTURE: SCHEMATIC REPRESENTATION OF DIFFERENT BIVARIATE MODELS FOR WSA-WSD SYNTHESIS. KENDALL'S TAU CORRELATION COEFFICIENTS ARE SHOWN FOR THE 3 COPULA MODELS.	75
FIGURE 5.4: P-VALUES RESULTING FROM APPLYING THE MANN-WHITNEY TEST TO VALIDATE THE SINGLE SITE MODEL BASED ON SUMMER (LEFT PLOTS OF VIOLINS) AND WINTER (RIGHT PLOTS OF VIOLINS) EVENTS CHARACTERISTICS FOR ALL STATIONS.	76
FIGURE 5.5: P-VALUES RESULTING FROM APPLYING THE GINI TEST TO VALIDATE THE SINGLE SITE MODEL BASED ON SUMMER (LEFT PLOTS OF VIOLINS) AND WINTER (RIGHT PLOTS OF VIOLINS) EVENTS CHARACTERISTICS FOR ALL STATIONS.	77
FIGURE 5.6: VALIDATION OF SINGLE SITE MODEL BASED ON ERRORS OF TOTAL SEASONAL RAINFALL FOR ALL STATIONS. FOR THE "OLD MODEL" AN ADDITIONAL VARIANT IS CONSIDERED (SEMI-TRANSPARENT PLOTS) IN WHICH SMALL EVENTS ARE INCLUDED.	78
FIGURE 5.7: VALIDATION OF SINGLE SITE MODEL FOR DIFFERENT TEMPORAL RESOLUTIONS BASED ON ERRORS OF MEAN PRECIPITATION FOR WET TIME STEPS (LEFT PLOTS OF VIOLINS) AND FRACTION OF WET TIME STEPS (RIGHT PLOTS OF VIOLINS) FOR ALL STATIONS.	79
FIGURE 5.8: VALIDATION OF PRECIPITATION MODEL BASED ON AUTOCORRELATION FOR 5 MINUTE TIME STEPS AND ALL STATIONS (MEDIANS).	80
FIGURE 5.9: P-VALUES RESULTING FROM APPLYING THE CRAMÉR-VON MISES TEST TO VALIDATE THE PRECIPITATION MODEL BASED ON EXTREME VALUES OBSERVED DURING THE SUMMER FOR ALL STATIONS.	81
FIGURE 5.10: COMPARISON OF PERFORMANCE FROM PRECIPITATION MODEL AND KOSTRA TO REPRODUCE OBSERVED EXTREMES FOR DIFFERENT RETURN PERIODS (20, 10, 5 AND 2 YEARS).	81
FIGURE 5.11: EVALUATION OF TRANSFERABILITY OF THE SINGLE SITE MODEL: P-VALUES RESULTING FROM APPLYING THE MANN-WHITNEY (SUMMER AND WINTER EVENTS: LEFT AND RIGHT PLOTS OF VIOLINS) AND CRAMÉR-VON MISES TESTS TO 22 STATIONS IN BADEN-WÜRTTEMBERG.	82

FIGURE 5.12: CHARACTERISTICS OF EVENTS ACCORDING TO THEIR OCCURRENCE IN 1, 2 OR 3 STATIONS SIMULTANEOUSLY FOR ALL CASES. LEFT: KENDALL'S TAU CORRELATION BETWEEN WSA-WSD FOR EACH STATION. RIGHT: KENDALL'S TAU CORRELATION BETWEEN WSA-WSA AND WSD-WSD FOR PAIRS OF STATIONS.	84
FIGURE 5.13: PAIRS OF PSEUDO-VALUES OF SUMMER EVENTS OCCURRING IN THREE STATIONS SIMULTANEOUSLY FOR THE BAWU_A CASE (KENDALL'S TAU IN RED). UPPER TRIANGULAR MATRIX: OBSERVED. LOWER TRIANGULAR MATRIX: SYNTHETIC.	85
FIGURE 5.14: CUMULATIVE DISTRIBUTION FUNCTIONS (LEFT) AND BIAS CORRECTION (RIGHT) FOR WSD OF SUMMER EVENTS CORRESPONDING TO BAWU_A CASE.	86
FIGURE 5.15: ERRORS OF MEAN VALUES OF DSD RESULTING FROM DIFFERENT ALTERNATIVES OF INCORPORATING THEM WITHIN THE TIME SERIES FOR ALL STATIONS.	89
FIGURE 5.16: BIVARIATE SPATIAL CONSISTENCY MEASURES FROM OBSERVED (CONTINUOUS LINES) AND OBSERVED WITHOUT SMALL EVENTS (DASHED LINES) TIME SERIES USED AS CRITERIA FOR SA OPTIMIZATION.	90
FIGURE 5.17: BIVARIATE SPATIAL CONSISTENCY MEASURES OF OBSERVED AND SIMULATED TIME SERIES WITH AND WITHOUT SMALL EVENTS.	91
FIGURE 5.18: KOSTRA REDUCTION FACTORS FOR THE CASE STUDIES (LEFT) AND GRAPHICAL EXPLANATION OF CRITERIA ADOPTED FOR AREAL ESTIMATION AND PROPORTION OF STATIONS (RIGHT).	91
FIGURE 5.19: ERRORS OF TOTAL SEASONAL RAINFALL BASED ON AREAL AND SINGLE SITES RESULTING FROM TWO MODELS USED FOR MULTISITE SYNTHESIS FOR SUMMER (LEFT PLOTS OF VIOLINS) AND WINTER (RIGHT PLOTS OF VIOLINS) FOR ALL STATIONS.	92
FIGURE 5.20: ERRORS OF PROPORTION OF TIME STEPS WITH RAINFALL BASED ON AREAL AND SINGLE SITES RESULTING FROM TWO MODELS USED FOR MULTISITE SYNTHESIS FOR SUMMER (LEFT PLOTS OF VIOLINS) AND WINTER (RIGHT PLOTS OF VIOLINS) FOR ALL STATIONS.	93
FIGURE 5.21: ERRORS OF MEAN VALUES OF EVENT CHARACTERISTICS BASED ON AREAL AND SINGLE SITES RESULTING FROM TWO MODELS USED FOR MULTISITE SYNTHESIS FOR SUMMER (LEFT PLOTS OF VIOLINS) AND WINTER (RIGHT PLOTS OF VIOLINS) EVENTS FOR ALL STATIONS.	93
FIGURE 5.22: COMPARISON OF MODELS REGARDING THEIR ABILITY TO REPRODUCE EXTREME EVENTS FOR DURATIONS RANGING FROM 5 MINUTES TO 24 HOURS FOR ALL STATIONS.	95
FIGURE 5.23: COMPARISON OF PERFORMANCE FROM MULTISITE MODELS (LEFT: SA AND RIGHT: VINE COPULA) AND KOSTRA - REDUCED TO REPRODUCE OBSERVED EXTREMES FOR DIFFERENT RETURN PERIODS (10, 5 AND 2 YEARS).	96
FIGURE 5.24: COMPARISON OF VARIABLES TO REGIONALIZE IN TERMS OF REPRODUCING THE PARAMETERS OR LMS.	98
FIGURE 5.25: EXAMPLE OF RELATIONSHIP BETWEEN PSEUDO VALUES OF SOME SITE DESCRIPTORS AND TARGET VARIABLES. NUMBERS IN RED INDICATE THE KENDALL'S TAU CORRELATION COEFFICIENTS.	100

FIGURE 5.26: P-VALUES RESULTING FROM APPLYING THE MANN-WHITNEY TEST TO COMPARE THE DIFFERENT REGIONALIZATION MODELS BASED ON SUMMER (LEFT PLOTS OF VIOLINS) AND WINTER (RIGHT PLOTS OF VIOLINS) EVENTS CHARACTERISTICS FOR ALL STATIONS.	103
FIGURE 5.27: P-VALUES RESULTING FROM APPLYING THE CRAMÉR-VON MISES TEST TO COMPARE THE DIFFERENT REGIONALIZATION MODELS BASED ON EXTREME VALUES OBSERVED DURING THE SUMMER FOR ALL STATIONS.	104
FIGURE 5.28: COMPARISON OF REGIONALIZATION MODELS REGARDING THEIR ABILITY TO REPRODUCE EXTREME EVENTS FOR DIFFERENT DURATIONS AND A RETURN PERIOD OF 10 YEARS.	104
FIGURE 5.29: RATIO BETWEEN RELATIVE STANDARD ERRORS RESULTING FROM USING ALL STATIONS AND CROSS VALIDATION FOR THE FOUR MOMENTS AND ALL EVENT CHARACTERISTICS (LEFT PLOTS OF VIOLINS) AND EXTREME EVENTS FOR ALL DURATIONS AND RETURN PERIODS (RIGHT PLOTS OF VIOLINS) FOR ALL STATIONS.	106
FIGURE 5.30: COMPARISON OF PERFORMANCE FROM REGIONALIZATION MODEL USING COPULAS AND KOSTRA TO REPRODUCE OBSERVED EXTREMES FOR DIFFERENT RETURN PERIODS (20, 10, 5 AND 2 YEARS).	107
FIGURE 5.31: MEAN VALUES (TOP) AND EXTREME EVENTS WITH A $T_r=2$ YEARS (BOTTOM) RESULTING FROM OBSERVATIONS (RED LINES) AND 100 REALIZATIONS (VIOLINPLOTS) USING THE SINGLE SITE MODEL BASED ON ALL OBSERVED EVENTS (AR), AND RESAMPLING OF DIFFERENT PERCENTAGES OF OBSERVED EVENTS (100, 75 AND 50%).	108
FIGURE 5.32: UNCERTAINTY OF MEAN VALUES: <i>RSE</i> FOR DIFFERENT QUANTILES RESULTING FROM 100 REALIZATIONS USING THE SINGLE SITE MODEL BASED ON ALL OBSERVED EVENTS (RED DOTS) AND ON RESAMPLING OF DIFFERENT PERCENTAGES OF EVENTS (BLUE CURVES) FOR 81 STATIONS.	110
FIGURE 5.33: UNCERTAINTY OF EXTREME EVENTS: <i>RSE</i> FOR DIFFERENT QUANTILES RESULTING FROM 100 REALIZATIONS USING THE SINGLE SITE MODEL BASED ON ALL OBSERVED EVENTS (RED DOTS) AND ON RESAMPLING OF DIFFERENT PERCENTAGES OF EVENTS (BLUE CURVES) FOR 23 STATIONS.	111
FIGURE 6.1: AUTOCORRELATION OF SUMMER (LEFT PLOTS OF VIOLINS) AND WINTER (RIGHT PLOTS OF VIOLINS) EVENTS FOR ALL STATIONS RESULTING FROM ORIGINAL TIME SERIES (OBS.) AND RESAMPLING WITHOUT REPLACEMENT BASED ON OBSERVED TIME SERIES (RES. 1 TO 10).	114
FIGURE 6.2: ERRORS OF FLOOD EVENT CHARACTERISTICS FOR ALL STATIONS RESULTING FROM RESAMPLING THE OBSERVED TIME SERIES. SOURCES OF ERROR: SERIAL AUTOCORRELATION (LEFT) AND NATURAL VARIABILITY (RIGHT).	115
FIGURE 6.3: ERRORS OF OVERFLOW EVENT CHARACTERISTICS FOR ALL STATIONS RESULTING FROM RESAMPLING THE OBSERVED TIME SERIES. SOURCES OF ERROR: SERIAL AUTOCORRELATION (LEFT) AND NATURAL VARIABILITY (RIGHT).	115
FIGURE 6.4: OBSERVED AND THEORETICAL RAINFALL EVENTS (LEFT) AND RESULTING FLOOD EVENTS (RIGHT) FOR ONE IMPORTANT FLOOD IN STATION BRAUNLAGE. THE	



NUMBERS [%] INDICATE THE OVERESTIMATION OF FLOODING VOLUME COMPARED TO RESULTS FROM OBSERVATIONS.	116
FIGURE 6.5: ERRORS OF FLOOD EVENTS FOR ALL STATIONS RESULTING FROM DIFFERENT INTERNAL DISTRIBUTIONS BASED ON OBSERVED EVENTS.	117
FIGURE 6.6: ERRORS OF OVERFLOW EVENTS FOR ALL STATIONS RESULTING FROM DIFFERENT INTERNAL DISTRIBUTIONS BASED ON OBSERVED EVENTS.	117
FIGURE 6.7: VALIDATION OF PRECIPITATION MODEL BASED ON ERRORS OF FLOOD EVENTS FOR ALL STATIONS. DOTTED LINES INDICATE MAXIMUM AND MINIMUM ERROR OF FLOODING VOLUME FOR RESAMPLING OBSERVATIONS.	119
FIGURE 6.8: COMPARISON OF PERFORMANCE FROM PRECIPITATION MODEL AND KOSTRA TO REPRODUCE VOLUME OF FLOOD EVENTS.	119
FIGURE 6.9: VALIDATION OF PRECIPITATION MODEL BASED ON ERRORS OF OVERFLOWING EVENTS FOR URBAN SYSTEM WITH T20 AND ALL STATIONS. FOR THE "OLD MODEL" AN ADDITIONAL VARIANT IS CONSIDERED (SEMI-TRANSPARENT PLOTS) IN WHICH SMALL EVENTS ARE INCLUDED.	120
FIGURE 6.10: VALIDATION OF REGIONALIZATION OF THE PRECIPITATION MODEL BASED ON ERRORS OF FLOOD EVENTS FOR STATIONS WITH LONG REGISTERS.	121
FIGURE 6.11: VALIDATION OF REGIONALIZATION OF PRECIPITATION MODEL BASED ON ERRORS OF OVERFLOWING EVENTS FOR URBAN SYSTEM WITH T20 AND STATIONS WITH LONG REGISTERS.	121
FIGURE 7.1: SUMMARY OF RESULTS OF THE PROPOSED MODEL FOR SINGLE SITE SYNTHESIS IN TERMS OF MEAN VALUE OF ALL EVENT CHARACTERISTICS (LEFT) AND EXTREME VALUES (RIGHT).	124
FIGURE 7.2: SUMMARY OF RESULTS OF THE PROPOSED MODEL FOR MULTI-SITE SYNTHESIS IN TERMS OF MEAN VALUE OF ALL EVENT CHARACTERISTICS (LEFT) AND EXTREME VALUES (RIGHT).	126
FIGURE 7.3: SUMMARY OF RESULTS OF THE PROPOSED MODEL FOR REGIONALIZATION IN TERMS OF MEAN VALUE OF ALL EVENT CHARACTERISTICS (LEFT) AND EXTREME VALUES (RIGHT).	127

## LIST OF TABLES

TABLE 3.1: MAIN DIFFERENCES BETWEEN EXISTING AND IMPROVED MODEL.	50
TABLE 4.1: ATTRIBUTES OF RAINFALL STATIONS USED FOR DEVELOPING THE SINGLE SITE MODEL.	59
TABLE 4.2: ATTRIBUTES OF RAINFALL STATIONS USED FOR TRANSFERING THE SINGLE SITE MODEL.	61
TABLE 4.3: ATTRIBUTES OF RAINFALL STATIONS USED FOR MULTISITE SYNTHESIS.	63
TABLE 4.4: BASIC INFORMATION DESCRIBING THE MODEL AND FICTIONAL URBAN SYSTEM.	70
TABLE 5.1: DESCRIPTION OF SELECTED CRITERIA AND COMPONENTS INVOLVED IN THE SINGLE SITE, MULTI-SITE AND REGIONALIZATION OF THE MODEL.	74
TABLE 5.2: EVALUATION OF STATISTICS OF DIFFERENT VARIABLES DESCRIBING THE EXTERNAL STRUCTURE OF RAINFALL EVENTS RESULTING FROM TWO SINGLE SITE MODELS AND FOR ALL STATIONS.	77
TABLE 5.3: EVALUATION OF PERFORMANCE BASED ON WET TIME STEPS FOR DIFFERENT TEMPORAL RESOLUTIONS RESULTING FROM TWO SINGLE SITE MODELS AND PRESENTED FOR ALL STATIONS AS <i>RSE</i> .	79
TABLE 5.4: PERCENTAGES OF OBSERVED EVENTS OCCURRING IN SEVERAL STATIONS SIMULTANEOUSLY (1, 2, 3) AND IN EACH OF THE STATIONS (A, B, C) FOR DIFFERENT SEASONS AND CASE STUDIES	83
TABLE 5.5: EXAMPLE OF SHORT TIME SERIES RESULTING FROM HYBRID MODEL FOR BAWU_A SUMMER CASE	87
TABLE 5.6: AREAS, PROPORTION PER STATION AND <i>KOSTRA</i> REDUCTION FACTORS FOR THE DIFFERENT CASE STUDIES.	92
TABLE 5.7: EVALUATION OF STATISTICS OF DIFFERENT VARIABLES DESCRIBING THE EXTERNAL STRUCTURE OF RAINFALL EVENTS RESULTING FROM TWO MODELS FOR MULTISITE SYNTHESIS BASED ON AREAL PRECIPITATION AND PRESENTED FOR ALL CASES AS <i>RSE</i> .	94
TABLE 5.8: COMPARISON OF REGIONALIZATION BASED ON PARAMETERS AND LMS IN TERMS OF PERCENTAGE OF CASES IN WHICH THE TESTS (MW: ALL EVENTS OR CVM: EXTREME EVENTS) ARE NOT REJECTED.	98
TABLE 5.9: NUMBER OF BIVARIATE RELATIONSHIPS CONSIDERED IN THE REGIONALIZATION OF DIFFERENT LMS USING COPULA BASED APPROACH.	100
TABLE 5.10: VARIABLES (SD OR LM) WITH HIGHEST CORRELATIONS WITH LMS TO BE REGIONALIZED.	100
TABLE 5.11: DISTRIBUTIONS SELECTED FOR MODELING THE MARGINAL BEHAVIOR OF THE DIFFERENT LMS.	101
TABLE 5.12: NUMBER OF EXPLANATORY VARIABLES CONSIDERED IN THE REGIONALIZATION OF DIFFERENT LMS USING MLR.	101

TABLE 5.13: DISTRIBUTIONS AND NUMBER OF CLUSTERS SELECTED FOR MODELING THE REGIONAL DATA SETS INVOLVED IN THE RFA.	102
TABLE 5.14: COMPARISON OF REGIONALIZATION BASED ON DIFFERENT METHODS IN TERMS OF PERCENTAGE OF CASES IN WHICH THE TESTS (MW: ALL EVENTS OR CvM: EXTREME EVENTS) ARE NOT REJECTED.	105
TABLE 5.15: RATIOS BETWEEN <i>RSE</i> ESTIMATED BASED ON RESAMPLING 100, 75 AND 50% OF OBSERVED EVENTS OVER <i>RSE</i> ESTIMATED WITH THE ORIGINAL AR MODEL. THE VALUES CORRESPOND TO MEDIAN AND MAXIMUM RATIOS.	112

# 1 INTRODUCTION

## 1.1 BACKGROUND AND MOTIVATION

Rainfall is characterized by high temporal and spatial variability. Since ancient times, human settlement has been largely influenced by the dramatic consequences of variable rainfall in the form of floods and droughts. As reported by CRED (2017) natural disasters related to meteorological and hydrological extreme conditions show an increasing number of occurrences per year causing millions of people to leave their homes, diseases and large economic losses. Consequences of drastic droughts and floods due to extreme rainfall events around the world indicate a deficiency in understanding the natural process governing them. These events are specified by several characteristics and their inter-dependencies are fundamental descriptors for both practical and scientific purposes (Gaál *et al.*, 2014). Even in different places around Europe (like Germany), in which the spatial and temporal coverage of rainfall measurements is considerably adequate compared to other remote regions of the world, fatalities resulting from excessive rainfall events are frequent, causing severe impacts for the society. Barredo (2007) provides a catalogue of major European flood disasters in the last half century and states that an increase of these events is very likely; in particular cities are becoming increasingly exposed to the occurrence and impacts of pluvial flooding.

According to WMO (2012) accurate precipitation data is required for various aspects of water management, namely hydrological characterization, flood management and control, irrigation and drainage, groundwater, navigation, power generation, water supply, water quality, fisheries and conservation, and amenity. The required temporal resolution differs according to the case. In particular sub-daily rainfall data is essential for the design of protection and retention measures, like dams, different types of diversions, river bank and infrastructure protection structures and for floodplain zoning. The required accuracy is particularly high due to the critical nature of design of these structures. Ochoa-Rodriguez *et al.* (2015) have shown the need of rainfall in high temporal resolutions (sub-hourly) to properly capture the variability of events for urban hydrological applications. Moreover, Mosthaf & Bárdossy (2017) state that the available precipitation records in high temporal resolution are usually short and contain erroneous measurements. It is therefore propitious to use stochastic precipitation models to generate synthetic time series and replace the observations.

The design of hydraulic structures and delineation of flood-risk maps involve the definition of synthetic hydrographs associated to different return periods. As stated by Grimaldi *et al.*

(2012), such hydrographs can be estimated based on events or continuous modeling. Intensity Duration Frequency curves (IDFs) are widely available and used for event-based estimations. However, limitations regarding the shape of the hyetograph (temporal pattern of rainfall intensity), the definition of the critical duration of the rainfall event, initial soil moisture condition and relationship between rainfall and discharge return periods result in inaccuracies that are difficult to quantify. On the other hand, continuous modeling relies on long synthetic rainfall time series which are used as input for continuous rainfall-runoff models, avoiding the above-mentioned limitations.

Furthermore, during its service life a hydraulic structure is exposed to many different loads and environmental conditions. Therefore it is propitious to include several possible loads during the design and planning of such structures. Different hydrological scenarios can be established and evaluated by combining a precipitation model providing continuous rainfall, and a rainfall-runoff model. By continuous simulation a broad range of hydrological loads differing in duration, intensity, amount, etc. is available and can be used when structures, for instance for flood protection, are designed. Moreover controlling flood characteristics differ according to the structure: gross flood volume is the main driver for dimensioning flood storages whereas for flood control structures the peak flood plays a fundamental role.

The accuracy of the input data plays a fundamental role in all of the before mentioned applications. In order to establish a robust precipitation model, it should reproduce the observed characteristics of rainfall sufficiently. Therefore, this PhD Thesis aims to contribute by developing a precipitation model that is able to jointly capture the main characteristics of rainfall events, i.e. amounts, intensities, peaks and durations, for a high temporal resolution. The model is to be used for generating long time series of rainfall in a short time increment of 5 minutes, which are required for several purposes, namely, for quantifying rainfall erosivity, hydrological applications involving urban or small, steep rural catchments, engineering design of flash flood control structures, etc. The developed model should be able to be applied

- i) at individual sites, i.e. temporal simulation of long continuous time series at particular location is possible;
- ii) at multiple sites, i.e. the simultaneous temporal and spatial simulation of long continuous time series for several sites is possible; and
- iii) at sites without observations, i.e. the temporal simulation of long continuous time series for any location is possible.

Another important challenge is related to one of the major inadequacies of present day hydrology which is the insufficient transfer of knowledge from research to practice. It is

therefore a sub-goal of this Thesis to directly transmit and provide the findings to engineers and planners involved in the design and management of hydraulic structures. Consequently a comparison between current design practice regarding rainfall analysis, IDFs and the developed rainfall models is included along with the application of these models for urban hydrological purposes.

## **1.2 OBJECTIVES**

The precipitation model presented in this Thesis is based on an existing one developed by Haberlandt (1996) for a different region. It is important to remark that the purpose of the original one was to model load-pollution, i.e. average overflow properties was of interest; therefore extreme event statistics were not of major concern in the development. The focal research areas and the specific objectives of the studies presented in this Thesis are:

- To develop a model able to mimic rainfall time series for one location and in a high temporal resolution. The aim of the model is to generate long time series for the particular location and to use these time series for planning and design purposes.
- To extend the model to mimic rainfall time series in several locations simultaneously and in a high temporal resolution. The aim of the model is to generate long time series in several locations simultaneously and provide with time series which are longer than the available ones.
- To develop a methodology to mimic rainfall time series for single locations without observations. The aim of this part of the work is to be able to generate reliable long time series of rainfall in a high temporal resolution for locations without any available measurement of rainfall.
- To evaluate the proposed methodologies both in terms of reproducing the average events properties of rainfall along with the extreme event statistics.
- To compare the presented methods with current design practice commonly adopted by the engineering community for designing hydraulic structures.
- To evaluate some of the proposed methods in terms of urban response with hydrological modeling and to compare event based with continuous modeling.

## 1.3 STRUCTURE

This Thesis is organized as follows. After an introduction to the topic presented in Chapter 1, Chapter 2 gives a brief description of many of the existing works dealing with sub-daily and in particular sub-hourly precipitation modeling. The challenges of modeling this type of data along with limitations of the existing models are discussed.

Chapter 3 describes the main assumptions and concepts related to the precipitation model used here. A sub-Chapter (3.2) describing the general concept of Copulas is as well included, since this mathematical tool has been used for most of the developments presented in the Thesis. Different following sub-Chapters describe the proposed methodologies to model precipitation for single sites, multi-sites and regionalization. The alternative standard approaches used for assessing the capability of the proposed methods are as well briefly described. Finally the different tests used for validating the proposed methodologies are explained, along with the uncertainty analysis for evaluating some of the methods.

In Chapter 4, a brief description of the study area is provided along with the available data set that is used for developing and validating the different methodologies. Furthermore, sub-Chapter 4.6 describes the current design practice in the study region and sub-Chapter 4.7 provides details of a fictitious urban model used for the application of the model to urban hydrology.

Following, Chapter 5 includes the overall results describing the final models and validations of the different proposed techniques. These validations are performed in comparison with existing standard techniques. The chapter is divided into different sub-Chapters to show the efficiency of the models in terms of average and extreme event statistics for a single site, multi-sites and sites without observations. The advantages and limitations of using copulas for the different applications are discussed. Additionally uncertainties associated to the variability of the stochastic process and the model parameter estimation are presented.

The application of precipitation time series to urban hydrological modeling is presented in Chapter 6, in which the ability of the single site model and regionalization of this model are evaluated in terms of overflow and flood events resulting from an artificial urban network.

Finally, Chapter 7 includes a broad discussion of the main findings of this Thesis. The advantages and limitations of the different developed methodologies are discussed taking into consideration different applications. Some ideas regarding further studies are presented in the sub-Chapter 7.2. All the details regarding the references mentioned in the work can be found in the "References" section. The last section of this work includes the Appendixes.

## 2 STATE OF ART

### 2.1 PRECIPITATION MODELS

Stochastic modeling of precipitation consists of conceptually representing the natural phenomenon through some mathematical relationships. The model is used to generate long synthetic time series that should adequately reproduce the historical statistics of the phenomenon. These synthetic series are applied for different purposes, e.g. design and planning, deriving long extended time series or regionalization of rainfall to areas without measurements. Of particular interest is the use of these models for urban applications. Urban catchments are usually characterized by small sized contribution areas with high imperviousness and therefore low concentration times. Furthermore, the proportion of rainfall producing surface runoff is high compared to rural catchments. Due to such fast response, rainfall in high temporal resolution is a crucial input for urban modeling.

Among the existing precipitation models, the following can be mentioned: data driven generators (e.g. resampling and disaggregation), time series models (e.g. Markov chains), and event based methods such as point and cluster process-based (e.g. Neyman-Scott, Barlett-Lewis) or alternating renewal models. The spatio-temporal scale of such models is determined by the hydrological application and becomes more challenging for high temporal resolutions. For sub-daily resolution the complexity of some models increases, as is the case of time series models which result in more parameters (see Verhoest *et al.*, 1997), and their performance declines as a result of the difficulty to reproduce historical properties. Thus, many investigations focus on the generation of daily precipitation time series, whereas models for hourly or sub-hourly time series are less frequent. Nevertheless clear progress has been made and therefore a review of some of the existing models applied to high temporal resolution data is presented in the following paragraphs. A discussion focusing on their performance regarding extreme events is included.

#### *Data driven approaches*

Licznar *et al.* (2011) applied a disaggregation model based on microcanonical cascade to 4 rain gauges in Germany. The disaggregation was performed from daily data up to 5 minutes and the outcome was overestimation of precipitation amounts, especially for high intensities. Another disaggregation model based on daily data combined with posterior resampling for multi-sites application was proposed by Müller & Haberlandt (2016). The microcanonical cascade model was as well used for rainfall observed in Germany; results indicated an



overestimation of extreme values for 5 minutes durations and most of the cases, whereas the extremes were better reproduced for 1 hour. Beck (2013) suggested some improvements to an existing synthetic rainfall generator called NiedSim (see Bárdossy, 1998), which follows a resampling approach. The modifications were tested based on stations in Germany and indicated underestimation of most extreme events on a daily aggregation and a slight underestimation of moderate extremes at hourly time scales.

For urban applications even hourly simulated time series need to be disaggregated to higher temporal resolutions, thus an extra disaggregation model is necessary. Cowpertwait *et al.* (2004) presented a stochastic model to disaggregate spatio-temporal data from hourly into 5 minutes. The model involved a moderate number of parameters and was validated with data registered in New Zealand. Overflow volume simulations indicated good agreement between historical and disaggregated data.

Most of the data driven approaches are restricted to the length of the observation period and therefore limited for high return periods. Furthermore, most of these models have the drawback that they rely on observed historical data for generating realizations and cannot generate unrecorded rainfall patterns, which is possible by other stochastic rainfall generators.

#### *Time series models*

As was mentioned, the application of time series models is not recommended for high temporal resolution requirements. Nevertheless, Katz & Parlange (1995) proposed an extension of chain-dependent process commonly applied to daily data for hourly simulations. The model assumes dependency between successive intensities and involves diurnal cycles and higher-order Markov chains for the occurrence process, thus a high number of parameters is involved. Unfortunately the evaluation of the model is based on 12 and 24 hours aggregation levels. The number of parameters required for high temporal resolutions discourage the application of these models.

#### *Event based methods*

Two types of cluster process-based rectangular pulses models are Barlett-Lewis (BL) and Neyman-Scott. Several works explore the capability of modified versions of BL model to reproduce high resolution precipitation for single sites. Rodriguez-Iturbe *et al.* (1988) modeled hourly data from two cities in the United States which delivered a good agreement between observations and simulated values; however the number of historical extreme values was larger than derived by the model, especially for hourly time resolution. Cowpertwait *et al.* (2007) used 60 years of 5 minutes rainfall data registered at a site in New

Zealand and compared annual maxima of historical and simulated data which resulted in a good fit at both 1 and 24 hour levels of aggregation, but some underestimation at the 5 minutes level. Based on a similar model Kaczmarska *et al.* (2014) used 5 minutes data registered at a site in Germany and introduced an improvement by including the dependence between event intensity and duration. They pointed out that the performance was improved, particularly for short temporal resolutions; however the model underestimated the extremes. Based on a long 10 minutes rainfall time series observed in one station in Belgium, Verhoest *et al.* (1997) compared three versions of BL; unfortunately none of them was able to reproduce the extreme values at short durations. Vandenberghe *et al.* (2011) used the same data set and pointed out that the BL model generated more and shorter storms and severe clustering. Consequently simulated extreme events did not reproduce observed ones. A similar analysis focusing on drought characteristics was performed by Pham *et al.* (2013), for which the over-clustering of rainfall events failed to preserve the extreme characteristics. Vernieuwe *et al.* (2015) used multivariate functions called Vine copulas to model continuous time series of rainfall for the same station resulting in a well reproduction of aggregated yearly rainfall statistics, however extreme values were overestimated at short durations, whereas long durations showed an underestimation.

Bernadara *et al.* (2007) coupled an alternating renewal process (ARP) to mimic the sequence of wet and dry weather states with a fractional noise to model the intensity. The results from 2 Italian stations revealed an overestimation of mean and mean maximum annual volumes as well as extreme values when short durations were evaluated. Based on ARP, Haberlandt *et al.* (2008) and Haberlandt & Radke (2014) modeled hourly rainfall for some stations in Germany. Results showed a slight overestimation of extreme rainfall amounts for high return periods and long durations. Furthermore, annual maximum flows derived from runoff modeling using the synthetic rainfall as input were overestimated. Nevertheless, the authors included a comparison of this model with an alternative random cascade disaggregation model and showed that the ARP outperformed the alternative one.

Other event based models for single sites are proposed by Gyasi-Agyei & Melching (2012) and Zhang & Switzer (2007). Gyasi-Agyei & Melching (2012) analyzed hourly data observed in some stations in the United States to model the joint behaviour of 3 major properties of storm events: duration, total and maximum amount. The model was set up on a monthly basis with harmonic Fourier functions to reduce the number of parameters. In this case Intensity-Duration-Frequency curves showed a well reproduction of observed values which were in most of the cases within the 95% of the prediction limits, except for short and long durations, for which the extreme values are over- and underestimated. Zhang & Switzer

(2007) used hourly data registered in 8 stations in United States to simultaneously model pairs of variables describing storm characteristics (velocities and sizes, radius and average intensity). The results showed a reasonable agreement between simulated and observed maximum rainfall intensity. However this model needs to be further combined with a frequency model in order to simulate continuous rainfall series.

Single site models are useful when spatial variability of rain is low and/or cases in which a uniform representation of the rainfall field is sufficient. However, for some applications the spatial attributes of rainfall can have an important impact in the outputs. Therefore, multi-site models are used to generate time series at different sites, while preserving temporal and spatial cross-correlation of the process. Many examples of these models can be found in the literature; however most of them are based on monthly, weekly and up to daily time scales. For high temporal resolution applications some of them become very complex and the number of parameters increase drastically.

Cowpertwait *et al.* (2002) proposed a model based on Neyman-Scott process combined with a generator of spatial circular rain cells, to simulate hourly rainfall in several stations in Italy. This model was implemented by Burton *et al.* (2008) in a generator called RainSim, which consists of seven parameters that are calibrated on a monthly basis based on single and dual-site statistics. The model was able to simulated rainfall up to 1 hour time resolution. However, as the model is monthly based, a high number of parameters is involved. Tarpanelli *et al.* (2012) described a simpler approach as well based on Neyman-Scott but combined with a rearrangement algorithm. The results from many stations in Italy with hourly data indicated a good performance of the model in reproducing the frequency of annual maximum rainfall for 1 and 24 hours. The method is simple to interpret and apply, however the rearranging is based on the hourly time series and it is not clear whether the structure of the storms is conserved within this process, which is crucial for events causing urban flooding.

Paschalis *et al.* (2013) presented a model called SPREAP and explored the capability of the model to mimic high resolution space-time rainfall. The model was compared with a Neyman-Scott based one by using gauges and radar data from a region located in the Mediterranean side of the Alps. The proposed model consisted of 3 stages: storm arrival, within-storm temporal evolution and two-dimensional temporal evolution of storm structure. Both models underestimated the precipitation depths in single stations for sub-hourly durations, especially for warm seasons with convective intense events; nevertheless SPREAD outperformed the Neyman-Scott method.

Willems (2001) proposed a spatial rainfall generator consisting of rain cells and rain storms. Rain cells are described by the shape and movement in space and time of events, whereas rain storms are the clustering of cells. The model was set up on a dense network of stations in Belgium with high temporal resolution data. The validation based on intensity-duration-frequency curves revealed an acceptable general performance for different temporal scales. However the proposed model is complex, as it requires many stepwise analysis and information, thus an adaptation to new cases studies and interpretation is not easy.

Haberlandt *et al.* (2008) presented a method to simulate hourly rainfall in multiple stations. The methodology was based on alternating renewal process (ARP) and applied to single sites independently, and was thereafter combined with a multi-site resampling procedure to reproduce the spatial dependence structure of rainfall by reordering the events. Extreme values showed to be well reproduced for short durations, whereas a systematic overestimation was obtained for long durations.

The performance of event based models depends on the proper modeling of the event characteristics. Several studies state the fact that as these characteristics are generated by the same physical phenomenon, their statistical dependence structure should be included in the modeling. Gaál *et al.* (2014) state that, along with the statistical properties of event characteristics, their inter-dependencies are fundamental descriptors of the phenomenon. Joint probability distributions are able to model these structures, in particular copula functions. Copulas have the advantage of modeling the dependencies of random variables; independently of their marginal distributions (see e.g. Nelsen, 2006; Genest & Favre, 2007; Salvadori *et al.*, 2007). Application of copulas to model joint behavior of variables describing storm characteristics can be found in Grimaldi & Serinaldi (2006), Salvadori & De Michele (2006), Zhang & Singh (2007), Kao & Govindaraju (2008), Vandenberghe *et al.* (2010), Balistocchi & Bacchi (2011), Gyasi-Agyei & Melching (2012), Ariff *et al.* (2012), Serinaldi & Kilsby (2013), Xiong *et al.* (2014) and Vernieuwe *et al.* (2015). Many of these applications focus on the modeling of single storm events, i.e. the occurrence of rainfall is neglected and thus a continuous modeling is not considered.

## **2.2 REGIONALIZATION OF PRECIPITATION**

The existing precipitation models involve parameters that are estimated based on long historical records. In order to apply these models in sites with short or no available data, a regionalization procedure must be developed. Estimation of rainfall in regions without observations is widely studied by several authors resulting in several methods that can mimic

the main features of precipitation acceptably. However many of the existing works focus mainly on regionalization of either total amounts (monthly, yearly) or extreme values associated to different durations.

Generally the existing regionalization methods involve two different steps: first regions with homogeneous rainfall regimes are identified, and subsequently an analysis of the rainfall within each identified region is performed. Different criteria exist for grouping or clustering stations with homogeneous statistical properties, see e.g. K-means partitioning (MacQueen, 1967), Partitioning around medoids (Kaufman & Rousseeuw, 1990), Self-organizing Maps (Kohonen, 1990), Ward's Hierarchical Classification (Ward, 1963), Hierarchical Bootstrapped Classification (Suzuki & Shimodaira, 2006), Model-Based Classification (Fraley & Raftery, 2007), Fuzzy C-Means (Asong *et al.* 2015), Cluster Probability Model (Cowpertwait, 2011) and Random Forest (Breiman, 2001). For the subsequent analysis several methods commonly applied are regional frequency analysis (Hosking & Wallis, 1997; Alila, 1999; Modarres, 2010; Sveinsson *et al.* 2010; Asong *et al.* 2015), regional vector method (Rau *et al.*, 2016), sample algorithm to generate data series for the ungauged locations (Mehrotra *et al.*, 2012).

The estimation of rainfall characteristics required for generating long time series in regions without observations is more limited. Nevertheless some attempts of regionalizing point-process stochastic models (for single sites) exist and are briefly described in the following paragraphs.

Some of the works dealing with regionalization focus on the estimation of amounts of rainfall for every single time step in regions without observations. These works rely on the distributions of precipitation amounts that need to be regionalized. The regionalization approach can either involve the interpolation of the rainfall time series from surrounding stations followed by the estimation of the distribution based on this interpolated data set or the interpolation of the distribution functions from the surrounding stations to the target point (see e.g. Wilks, 2008; Kleiber *et al.* 2012; Mosthaf & Bárdossy, 2017; both for parametric and nonparametric cases). Most of the studies focus on the daily time series except for Mosthaf & Bárdossy (2017) which applied the proposed regionalization method to an hourly temporal resolution.

Regionalization of the cluster process-based Neyman-Scott model was proposed by Cowpertwait *et al.* (1996) in which the model parameters were related to site variables (altitude, distance to coast, etc.) by regression equations. Over 100 stations in the United Kingdom were used. Regionalization errors were lower than sampling error expected from a 20 year long historical record, suggesting the use of the regionalized rainfall with reasonable

confidence. Mohud Daut *et al.* (2016) proposed an alternative methodology for regionalizing the same model, which consisted of merging all stations (previously scaled on individual mean values) and estimating model parameters for the merged time series treated as a single station. The model was applied on hourly basis to some stations located in Malaysia and showed a tendency to underestimate autocorrelation and auto covariance although the main statistical profiles were preserved.

The Barlett-Lewis Rectangular model was regionalized by Kim *et al.* (2013) by interpolating the model parameters using Ordinary Kriging based on over 3000 stations registering hourly rainfall in United States. Interpolated parameters showed to be smoothed over the space. Cross-validation results indicated that the mean and variance of rainfall were well reproduced, whereas the regionalized model failed to reproduce the autocorrelation along with the probability of no rainfall for some cases.

Another method to model rainfall events in regions without observations was presented by Haberlandt (1998). It was applied in the south of Germany and involved the estimation of the parameters of probability distributions describing rainfall characteristic by Kriging methods. The proposed method was validated by comparing the model results without and with regionalization. A systematic underestimation of precipitation characteristics was obtained that, as stated in the work, most likely results from the nonlinearity between regionalized parameters and target variables. Another work dealing with rainfall events was presented by Hernaéz & Martín-Vide (2011). The north of Spain was used for regionalizing empirical probability distributions of wet spells along with rainfall persistence using geostatistical interpolators. Despite the complex interaction between topographic, geographic and climatic factors, the resulting maps reflected the general expected trends of precipitation behavior.

## **2.3 OPEN QUESTIONS**

The overview of the existing literature reveals some limitations of the available models. Some of these methods need to be further combined with other models to provide long time series of precipitation, especially if the following aspects are pursuit: i) continuous simulations, ii) high temporal resolution and/or iii) several sites simultaneously. Most of the data driven methods are limited to observed attributes, like patterns of rainfall and length of observed period. Multi-site models are usually complex, hard to interpret and involve several stages of simulation which result in many parameters, especially for applications which require high temporal resolutions. Copulas have shown to efficiently mimic joint behavior of event characteristics; however their application for continuous modeling has been limited. Methods

for regionalizing event based characteristics are uncommon. This general outlook leaves some open questions:

- Is it possible to reproduce average event properties and extreme event statistics of observed rainfall with one single model in a high temporal resolution?
- Are copulas efficient tools for modeling continuous precipitation in a high temporal resolution?
- Is it possible to generate reliable long time series of rainfall with a high temporal resolution in areas without observations?

The objective is to contribute to the possibility of obtaining long rainfall series that can properly reproduce different characteristics of the observed ones, whilst keeping the model as simple as possible. For this reason a model based on Alternating Renewal process is tested, due to the fact that it is considered easy to interpret and transfer to new regions of interest. Furthermore, as it is event based, the number of parameters is not affected by the temporal resolution and thus benefits the regionalization procedure. This leads to an additional question:

- Is an extension of the Alternating Renewal based process model to a multi-site application feasible?

It is the aim of this Thesis to address these questions by developing a stochastic continuous precipitation model for simulating long time series of rainfall in a high temporal resolution. The application of the model for single, multi-sites and sites without observations is to be assessed.

### 3 METHODS

This chapter includes the explanation of the different methodologies involved in the Thesis. First in sub-Chapter 3.1 the general concept of the basis process for the precipitation model is described in order to present the main principles, which are important for understanding the proposed methods. Following, in sub-Chapter 3.2, a general definition of copulas is included to clarify the main concepts, limitations and characteristics of these models used for the multivariate analysis within different steps of calculation. A description of the existing model for simulating rainfall in single locations is included in sub-Chapter 3.3, which was developed previous to this Thesis and was used as a basis for developing the methods that are detailed afterwards. The following sub-Chapter 3.4 describes the proposed methodology for simulating rainfall in several stations simultaneously by using the vine copulas. The existing approach based on Simulated Annealing and used for comparison purposes is as well concisely described. Afterwards the concept of regionalizing the model with the aid of copulas and site descriptors is described in sub-Chapter 3.5. The alternative approaches used for comparison are as well briefly explained. In sub-Chapter 3.6 the different criteria used for evaluating the proposed methods are presented and explained, followed by the procedure applied to assess the uncertainty of the different proposed models along with the measures adopted for comparing the models based on urban hydrological applications. Finally in sub-Chapter 3.7 a summary comparing the existing and new proposed model is presented.

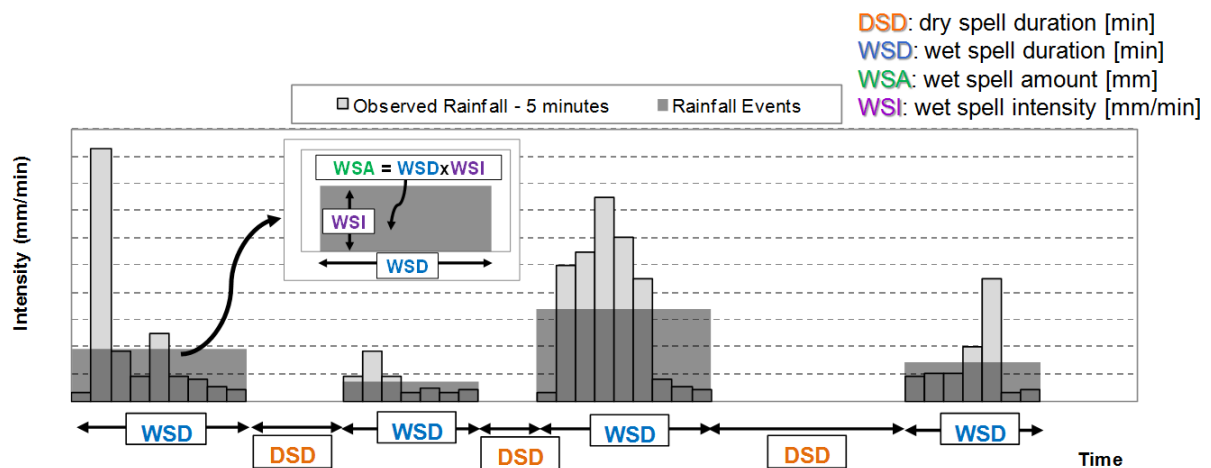
#### 3.1 ALTERNATING RENEWAL PROCESS

The precipitation model presented in this Thesis is based on the theory of alternating renewal process (ARP). A renewal process is, as defined by Serfozo (2009), a stochastic model in which an event occurs repeatedly over time and times between occurrences are independent and identically distributed (IID). Suppose  $0 \leq T_1 \leq T_2 \leq \dots$  are finite random times at which a certain event occurs with at most one occurrence at any instant. A process is defined as renewal if the inter-occurrence times  $\xi_n = T_n - T_{n-1}$ , for  $n \geq 1$ , are independent and have a common distribution  $F$ , where  $F(0) = 0$  and  $T_0 = 0$ . The  $T_n$  are called renewal times, referring to the independent or renewed stochastic information at these times. The  $\xi_n$  are the inter-renewal times, and  $N(t)$  is the number of renewals in  $(0, t]$ . A renewal process is therefore defined by specifying a distribution  $F$  with  $F(0) = 0$  for the inter-renewal times. This type of process can for instance define the random times at which a stochastic process enters a



special state of interest. An alternating renewal process is a continuous-time stochastic process  $X(t)$  that cycles through two possible states, say 0 and 1 or *on* and *off*, again and again. A complete cycle of the process  $X(t)$  goes from state 0 back to 0. With a clever definition of *on* and *off* states, many stochastic processes can be turned into alternating renewal processes.

In the case analyzed in this Thesis, the two states are described by rainfall and no-rainfall, which are described by wet spell durations (WSD) and dry spell durations (DSD) respectively. Figure 3.1 shows the application of ARP to mimic observed rainfall.



**Figure 3.1: Conceptual illustration of the alternating renewal process applied for rainfall modeling.**

The figure indicates some additional variables characterizing rainfall events: total volume of rainfall falling during the event or wet spell amount (WSA) and wet spell intensity (WSI). All these variables constitute the external structure of the precipitation model which is complemented by an internal structure and is explained later. The ARP is used to model the behavior of this external structure, i.e. the succession of independent rain events.

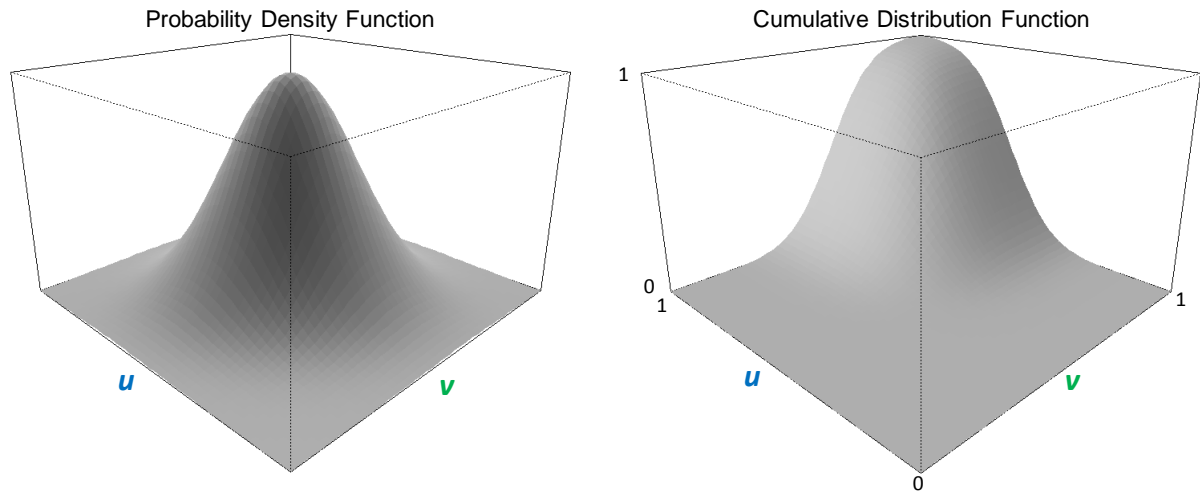
### 3.2 COPULAS – GENERAL CONCEPT

A copula is a function that enables modeling the dependency structure of random variables, independently of their marginal distributions. These functions are special multivariate distributions which model the joint behavior of variables which follow uniform distributions. They can therefore model the dependency of uniformly transformed observations (e.g. in form of empirical cumulative distributions) denoted as pseudo-observations. Copulas are used for simulating samples of pseudo-values, which are then transformed to values by the inverse marginal distributions describing the different involved variables. The joint distribution is therefore decomposed into the marginal ones and the copula which connects them. For further details the reader may refer to Nelsen (2006), Genest & Favre (2007) or Salvadori *et al.* (2007).

For a multivariate case with  $n$  variables, the link between a copula, denoted as  $C(u_1, u_2, \dots, u_n)$ , and the multivariate distribution is provided by Sklar's Theorem (see Nelsen, 2006) with the following equation:

$$F(x_1, x_2, \dots, x_n) = C[F_1(x_1), F_2(x_2), \dots, F_n(x_n)], \quad (\text{Eq. 1})$$

where  $F(x_1, x_2, \dots, x_n)$  is the joint cumulative distribution function with the continuous marginal distribution functions of the random variables:  $F_1(x_1), F_2(x_2), \dots, F_n(x_n)$ . The model representing the joint behaviour of the random variables is defined by the selection of the copula to describe the relationship between the random variables and the marginal distributions representing each of them. The main advantage of this approach is that the selection of the appropriate copula to represent the dependence structure can be done independently from the choice of the marginal distributions. The bivariate case involves a total of  $n=2$  variables, and the following marginal distributions  $u=F_1(x_1)$  and  $v=F_2(x_2)$ . A conceptual illustration of this case is presented in Figure 3.2. In a bivariate context, a copula function is defined as mapping  $C:[0,1]^2$  to  $[0,1]$ , with the properties that if either one of the marginal distributions is zero then the joint distribution will be zero, and if either one of the marginal is equal to one then the joint distribution will behave as a univariate distribution of the opposite variable.



**Figure 3.2: Conceptual illustration of a copula in a bivariate context (under  $N(0,1)$  margins for visualization purposes).**

The best sampled-based representation of the copula is the so called empirical copula which in the bivariate domain is defined by:

$$C_n(u, v) = \frac{1}{n} \sum_{i=1}^n 1 \left( \frac{R_i}{n+1} \leq u, \frac{S_i}{n+1} \leq v \right), \quad (\text{Eq. 2})$$

where  $n$  is the number of pairs,  $1(A)$  is an indicator function of set  $A$ ,  $u$  and  $v$  are the marginal distributions and  $R_i$  and  $S_i$  stand for the ranks of each observation among the complete time series.

A well known nonparametric measure of dependency structure is the Kendall's Tau, which is rank based and has the following empirical version:

$$\tau_n = \frac{4}{n(n-1)} P_n - 1, \quad (\text{Eq. 3})$$

where  $n$  is the number of pairs and  $P_n$  is the number of concordant pairs. Two pairs  $(X_i, Y_i), (X_j, Y_j)$  are concordant if  $(X_i - X_j) \cdot (Y_i - Y_j) > 0$ . And since two pairs are concordant if and only if the ranks are concordant, i.e.  $(R_i - R_j) \cdot (S_i - S_j) > 0$ , then the measure is rank based and can be expressed as function of the empirical copula  $C_n$  as reported by Genest & Favre (2007) by:

$$\tau_n = \frac{4n}{n-1} \int_{[0,1]^2} C_n(u, v) dC_n(u, v) - \frac{n+3}{n-1} \quad (\text{Eq. 4})$$

and more general for  $n \rightarrow \infty$  and  $C_n \rightarrow C$ ,

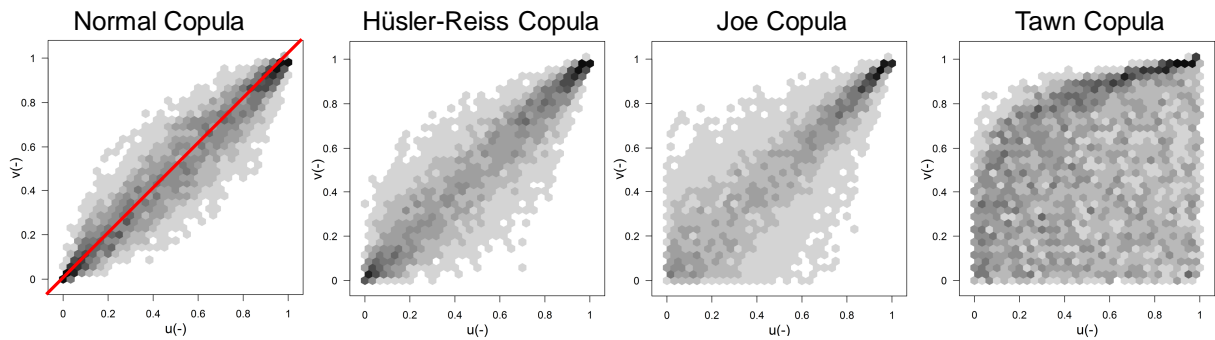
$$\tau = 4 \int_{[0,1]^2} C(u, v) dC(u, v) - 1. \quad (\text{Eq. 5})$$

Thereafter relationships between Kendall's Tau and parameters describing different copulas can be derived, either analytically or numerically, some analytical solutions are listed in Chowdhary *et al.* (2011). As this coefficient is nonparametric and rank based Genest & Favre (2007) refer to this method of parameter estimation as nonparametric adaptation of the method of moments or moment-like. Another estimation method is the so called pseudo-likelihood method, which again relies completely on the relative ranks of joint variates.

The literature review reveals a list with several copula functions which are commonly applied to analyze multivariate hydrological data. Different copulas are able to model different ranges of dependence structures, so the suitability of a particular copula depends on the type of dependence structure between variables. In order to evaluate the performance of a particular copula, the empirical and parametric probabilities are compared with a Cramér-von Mises type of test (Genest and Favre 2007, Chowdhary *et al.* 2011). Large values of this statistic can lead to the rejection of the model under consideration. The objective of a copula model is to adequately represent the dependence of the observed data. A special aspect that these models must meet for some applications, e.g. extreme value analysis, is to ensure suitability in terms of tail dependence characteristics. Upper and lower tails are the pairs of variables corresponding to either very high or very low probabilities. Certain copulas may exhibit

similar overall dependence features while showing different lower and upper tail dependence characteristics. Tail dependencies may be visually compared with the observed data to decide which model is more suitable. Some coefficients to quantify these dependencies are as well provided by e.g. Dupuis (2007) and Vandenberghe *et al.* (2011).

Several symmetric copulas commonly used in hydrological applications are considered in this Thesis. The symmetry refers to the exchangeability of pairs along the main diagonal (see red line in Figure 3.3). In a bivariate context these models are usually characterized by one parameter and as can be seen in the figure some examples are: Normal, Hüsler-Reiss and Joe, for which any pair of values has a corresponding one when swapping them with respect to the red line axis. If random variables are exchangeable then  $C(u, v) = C(v, u)$ . The Tawn copula is as well included in the figure to illustrate an asymmetric model, these copulas are discussed later.



**Figure 3.3:** Ability of different copulas to model tail dependencies, symmetric and asymmetric relationship between variables. Red line represents the axis of symmetry.

The pairs of values shown in Figure 3.3 are visualized using hexagon binning, which is a form of bivariate histogram which shows the joint-structure of big samples. The number of points falling within each hexagon is counted and a color ramp is used to show the proportion of points, light-colored hexagons indicate less points, whereas dark indicates a higher concentration of pairs. These plots are done using the *hexbin* R package (see Carr *et al.*, 2015).

From these symmetric examples, the bivariate Normal copula is defined as:

$$C_{\theta}(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\theta^2}} \exp\left(-\frac{x^2-2\theta xy+y^2}{2(1-\theta^2)}\right) dx \cdot dy, \quad (\text{Eq. 6})$$

where  $\theta$  is the dependence parameter,  $u$  and  $v$  are the marginal distributions and  $\Phi^{-1}$  is the inverse of the Standard Normal distribution. The univariate Standard Normal cumulative distribution function is:

$$\Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^z \exp\left(-\frac{\varepsilon^2}{2}\right) d\varepsilon. \quad (\text{Eq. 7})$$

Regarding the tail dependency, the Normal copula shows to mimic a similar behavior for the upper and lower tails as the distribution of pairs looks similar for the upper-right and lower-left corners (see Figure 3.3). This is not the case for the Hüsler-Reiss, which shows a different behavior for the two tails, the upper tail pairs show to be very concentrated around the main diagonal, whereas the lower ones are as well concentrated but in a lower degree compared to the upper tail. The Joe copula shows an even stronger difference between degrees of dependence for the two tails; with very spread pairs for the lower tail. The Hüsler-Reiss copula is described by one parameter. The equation for the bivariate case is defined as:

$$C_{\theta}(u, v) = \exp\left\{\ln(u) \Phi\left[\frac{1}{\theta} + \frac{1}{2}\theta \cdot \ln\left(\frac{\ln(u)}{\ln(v)}\right)\right] + \ln(v) \Phi\left[\frac{1}{\theta} + \frac{1}{2}\theta \cdot \ln\left(\frac{\ln(v)}{\ln(u)}\right)\right]\right\}, \quad (\text{Eq. 8})$$

where  $\theta$  is the dependence parameter,  $u$  and  $v$  are the marginal distributions describing each of the variables and  $\Phi$  stands for the univariate Standard Normal distribution.

Additional asymmetric models described by more than one parameter (Tawn, Skew-t and Khoudraji transformation of Archimedean copulas) are as well included in the analysis to mimic the dependency of non-exchangeable variables. Such variables present some asymmetry with respect to the main diagonal, so that one variable cannot be exchanged by the second one, i.e.  $C(u, v) \neq C(v, u)$ ; therefore some of the pairs of points cannot be captured by a symmetric copula. This is the case of the Tawn copula, shown in Figure 3.3 (right plot), which indicates the presence of pairs in the lower right corner, whereas in the upper left corner no pairs are present. Non-exchangeable dependence structures have been reported for some hydrological variables; see e.g. Vandenberghe *et al.* (2010) and Genest & Nešlehová (2013). The Tawn copula (see Tawn, 1988) is an asymmetric extension of the Gumbel copula, with two additional parameters that add flexibility to the model and has the following bivariate form:

$$C_{\theta}(u, v) = (uv)^{A\left(\frac{\ln u}{\ln(uv)}\right)}, \quad (\text{Eq. 9})$$

where  $A(t)$  is the so called Pickands dependence function, defined for the Tawn copula as:  $A(t) = (\psi_2 - \psi_1)t + (1 - \psi_2) + [(\psi_2(1-t))^{\theta} + (\psi_1 t)^{\theta}]^{1/\theta}$  and  $\theta$ ,  $\psi_1$  and  $\psi_2$  are the dependence and two additional asymmetry parameters. For the Gumbel copula  $\psi_1 = \psi_2 = 1$  (for further details see Bernard & Czado, 2015).

Multivariate copulas have not been used for hydrological applications as often as bivariate ones, for which a well investigated rich variety of families is available. The increase of dimensionality makes the application and interpretation of results more complicated.

Modeling the dependence structure of several variables involved in a multivariate phenomenon is more challenging for high-dimensional cases. Recent research introduces pair-copula constructions to build flexible multivariate distributions (see Czado, 2010). The resulting model is called Vine copula and consists of decomposition of multivariate probability density into bivariate copulas. Bivariate models are combined as a cascade or nested set of trees to model multiple variables simultaneously. Each pair-copula is selected independent from the others allowing for a very high flexibility in modelling different dependence structures (Brechmann & Schepsmeier, 2013).

The use of Vine copulas for a n-dimensional application requires the decomposition of the problem into products of pairs, i.e. bivariate, copula densities and marginal densities. The pair-copula construction is defined by Joe (1996), Bedford and Cooke (2001), Aas *et al.* (2009) and Czado (2010) and the density is represented as:

$$f(x_1, \dots, x_n) = \prod_{j=1}^{n-1} \prod_{i=1}^{n-j} c_{i,(i+j)|(i+1),\dots,(i+j-1)} \times \prod_{k=1}^n f_k(x_k), \quad (\text{Eq. 10})$$

where the first double product of the equation corresponds to pair copula densities (conditional and unconditional pairs) whereas the third product, i.e. the  $f_k$ , are the marginal densities of the n variables. If three variables are involved in the analysis, the density is presented as:

$$f(x_1, x_2, x_3) = c_{13|2} \times c_{12} \times c_{23} \times f_1(x_1) \times f_2(x_2) \times f_3(x_3), \quad (\text{Eq.11})$$

where  $c_{13|2}$  is the conditional pair,  $c_{12}$  and  $c_{23}$  are the unconditional pairs and  $f_1$ ,  $f_2$  and  $f_3$  are the marginal densities of each of the variables. A graphical representation of this structure is shown in Figure 3.4, which includes a nested set of trees used to depict the decomposition. The Vine tree structure (Bedford and Cooke, 2001) indicates the order of dependency of the variables for a regular vine structure.

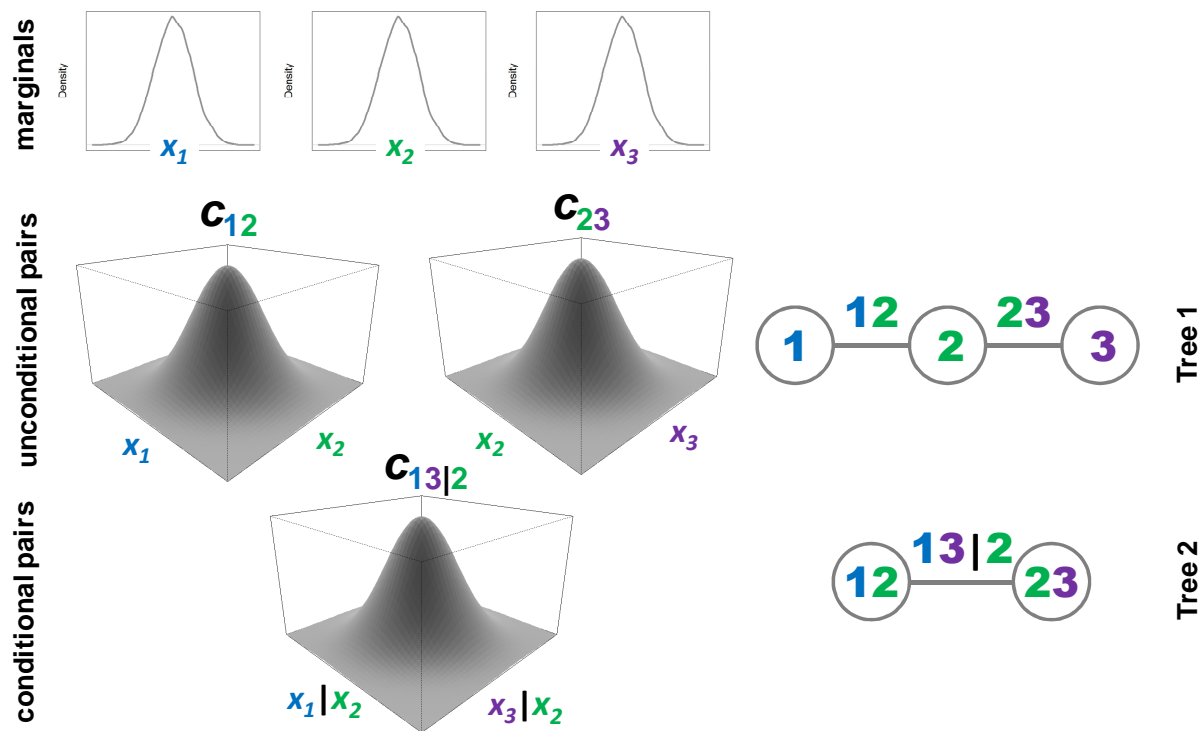


Figure 3.4: Example of components of a Vine copula in a trivariate context. Left: Marginal densities, unconditional and conditional pair copula densities (under  $N(0,1)$  margins for visualization purposes). Right: Vine tree structure.

Brechmann & Schepsmeier (2013) resume the following sequential steps to set up a Vine copula model:

- Determination of adequate tree structure by estimation of empirical dependence measure for each pair and selection of the structure that maximizes the sum of absolute dependence structure (see Dißmann *et al.*, 2013).
- Selection of appropriate pair-copula families (for both unconditional and conditional pairs) by information criteria such as Akaike Information Criterion (see Brechmann *et al.*, 2012; Dißmann *et al.*, 2013).
- Estimation of the corresponding parameter for each family by statistical inference (maximum likelihood, Bayesian approach, etc.).

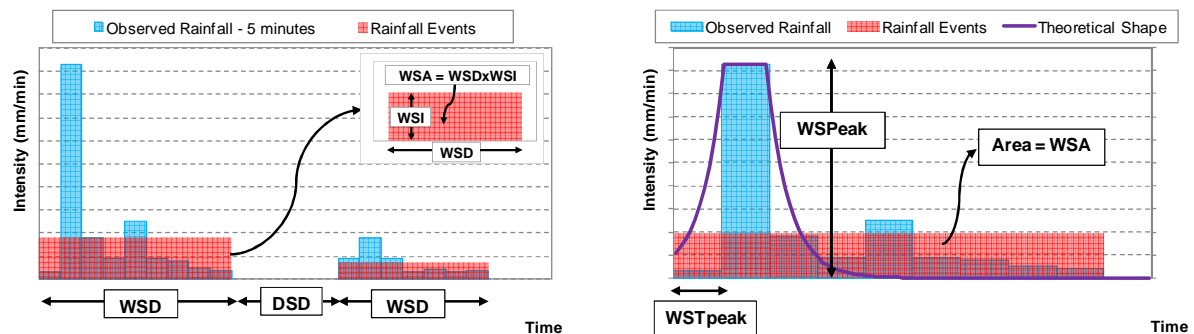
### 3.3 SINGLE SITE PRECIPITATION MODEL

The single site model involves the *temporal* simulation of rainfall time series for single locations.

#### 3.3.1 EXISTING MODEL (OLD MODEL)

The existing model was developed by Haberlandt (1996). The aim of the model is to generate long synthetic 5-min time series based on the simulation of variables describing rainfall events. It is based on the theory of renewal processes and rainfall is described as two structures: external and internal.

The external structure is the succession of independent rain events, each one described by i) time between two events or dry spell duration (DSD), ii) time of event in which rain occurs or wet spell duration (WSD), iii) total volume of rainfall falling during the event or wet spell amount (WSA) and iv) wet spell intensity (WSI) which is the ratio of WSA divided by WSD. The internal structure describes the distribution of the total rainfall within the wet spell and is defined by v) intensity of the peak (WSPeak) and vi) the time of occurrence of the peak (WSTpeak). The different variables can be visualized in Figure 3.5.



**Figure 3.5: External (left) and internal (right) structures of rainfall events for Old model.**

Defining rainfall events from the continuous series requires the setting of the following minimum values: wet spell intensity ( $WSI_{min}=0.01\text{mm}/5\text{min}$ ), wet spell amount ( $WSA_{min}=1\text{mm}$ ) and dry spell duration ( $DSD_{min}=5\text{min}$ ). These criteria provide events without dry time steps within the WSD and result in the exclusion of small events.

Two of the variables involved in the external structure are directly modeled by probability distributions which are fitted to the observed time series of DSD and WSD. The WSA is derived from the WSD by simple linear regressions relating the WSD with different moments of WSA. The moments are then used for fitting a probability distribution from which a particular WSA is modeled. The WSI is estimated from pairs of WSA and WSD.



As the WSI is defined in the external structure of the model, this variable is used to estimate the following variable involved in the internal structure: WSPeak. For this purpose a simple linear regression between these two variables is used to derive the WSPeak from the WSI. The regression model is based on all observed events from all stations. To estimate a WSTpeak for each event, a simple model is applied which involves the use of a uniform distribution to randomly generate a value between 0 and 1, which is then multiplied by the WSD to derive the value of the time to peak.

The internal structure of the rainfall event is modeled by a mixture of two exponential functions, both described by one parameter  $\lambda$ . The intensity is calculated for every time step with the following equation:

$$i(t) = \text{WSPeak} \cdot \exp(c\lambda(t - \text{WSTpeak})) \quad \begin{cases} c = +1, t < \text{WSTpeak}, \\ c = -1, t \geq \text{WSTpeak}. \end{cases} \quad (\text{Eq. 12})$$

The parameter  $\lambda$  is estimated on-line, i.e. for each event, by integrating the Eq. 12 over WSD which leads to the following equation:

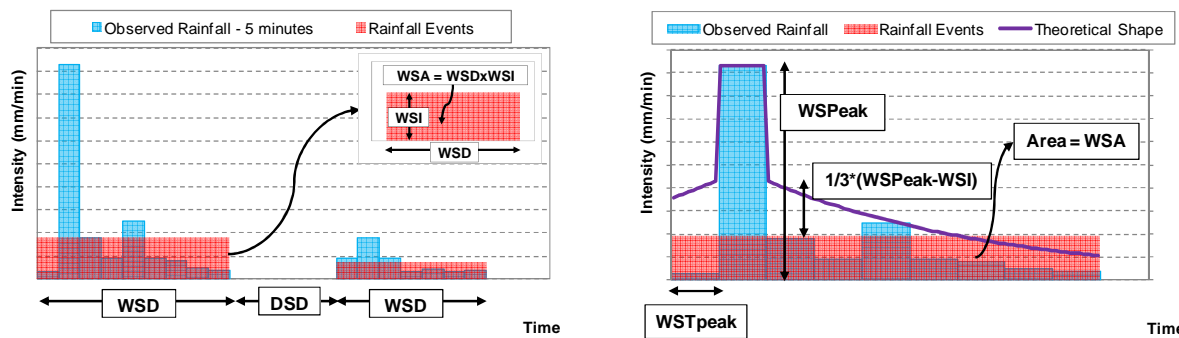
$$f(\lambda) = \frac{1}{\lambda} \text{WSPeak} \cdot [2 - \exp(-\lambda(\text{WSTpeak})) - \exp(\lambda(\text{WSTpeak} - \text{WSD}))] = \text{WSA}. \quad (\text{Eq. 13})$$

The  $\lambda$  is therefore estimated based on the parameters describing the external (WSA, WSD) and internal (WSPeak, WSTpeak) structures of the event.

For further details the readers can refer to Haberlandt (1996), only available in German, and a more compact explanation can be found in Haberlandt (1998), available in English.

### 3.3.2 PROPOSED MODEL

As was mentioned, the aim of the precipitation model is to generate long synthetic 5-min time series based on the simulation of variables describing rainfall events. Therefore most of the concepts described in this sub-Chapter are valid as well for the multi-site and regionalization of the model. The proposed model is derived from the “Old Model” and some important improvements are implemented which are described in the following paragraphs. The same variables are used to describe events and the external and internal structures are presented in Figure 3.6. The definition of rainfall events from the continuous series is done following the same criteria, thus events consist of continuous rainfall within the WSD and result in the exclusion of small events that are later added back to the model.



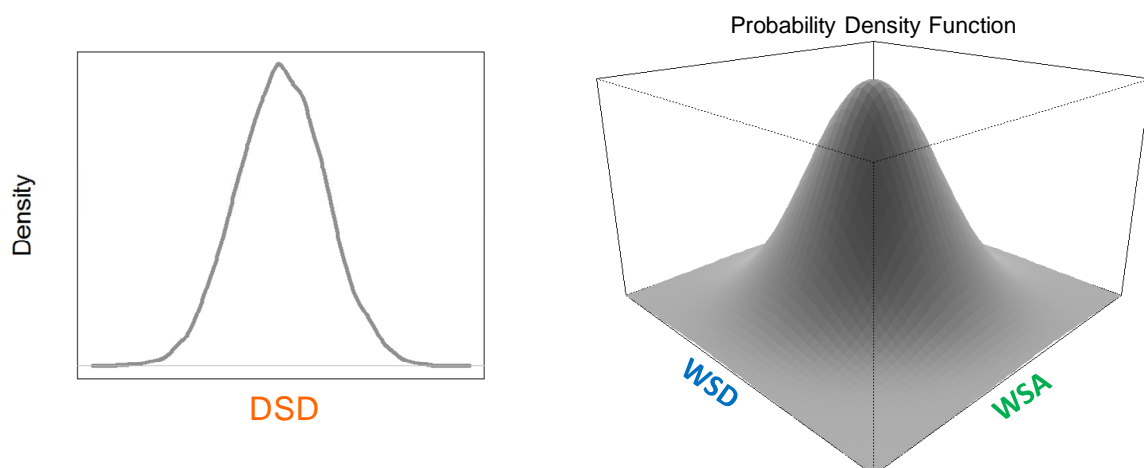
**Figure 3.6: External (left) and internal (right) structures of rainfall events for proposed model.**

The proposed external structure is modeled by probability distributions which are fitted to the following variables: DSD, WSD and WSA. The WSI is estimated from pairs of WSA and WSD. Seasonality is included by modeling the marginal behaviour of the different variables separately for summer and winter, i.e. dividing the year into two seasons according to the nature of the physical processes causing the events. Several probability distributions are included in the frequency analysis of the different variables. These distributions are described by different number of parameters and several methods of estimating the parameters are applied, i.e. method of moments, maximum pseudo-likelihood and L-moments. L-moments are equivalent to moments but estimated by a linear combination of order statistics (see Hosking, 1990). The considered distribution functions include: Exponential (1 and 2 parameters), Normal (2 and 3 parameters), Lognormal (2 and 3 parameters), Weibull (2 and 3 parameters), Gamma, Gumbel, Generalized Extreme Value, Generalized Pareto, Pearson type 3, Rayleigh, Reverse Gumbel, Generalized Logistic, Generalized Lambda, Wakeby, Generalized Normal and Kappa. In the case of WSA a mixture model is also considered in the analysis which is a combination of a Weibull distribution (for the bulk model) and a Generalized Pareto (for the tail) with a transition function described by a Cauchy distribution.

The aim of the internal structure is to distribute the total amount of rainfall within the wet spell, by modeling a peak-type structure of the event and guaranteeing the conservation of volume. Different popular models exist for estimating the shape of hyetographs for design purposes, see Nguyen *et al.* (2002) or Alfieri *et al.* (2014) for a review and comparison of some of them. Some approaches suggest simplifying the shape, e.g. Garcia-Guzman & Aranda-Oliver (1993) propose the use of beta distribution functions, whereas an Euler type 2 distribution is applied according to German design standards (DWA, 2006). In this work a mixture of two exponential functions similar to the one involved in the “Old Model” is used. An advantage of applying this type of shape is that the event duration, volume and peak intensity are exactly reproduced. The adoption of this shape implies a simplification of the real internal behavior, but was shown to be acceptable by Haberlandt (1996). Experiments

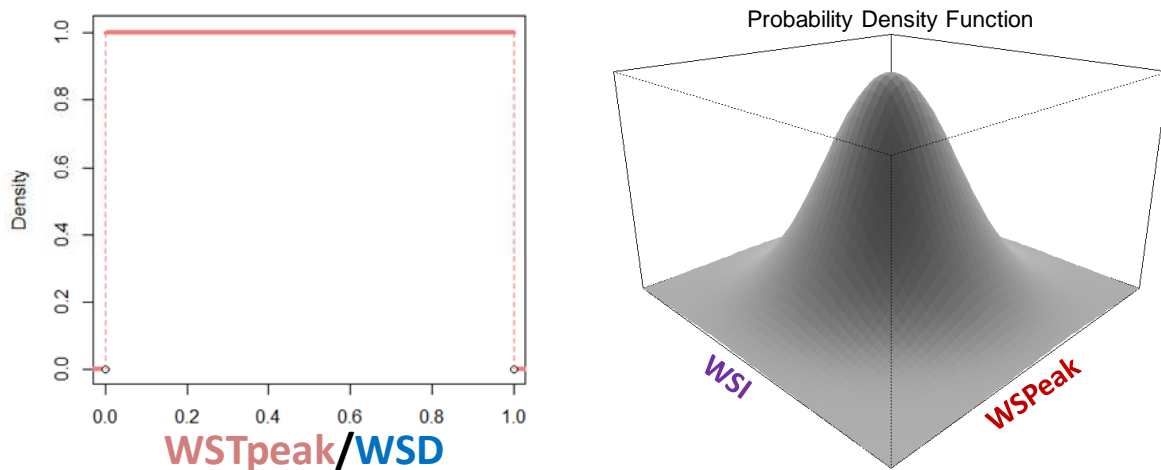
carried out within this Thesis have shown that profiles with rainfall concentrated around the peak lead to flood events which occur in shorted periods of time and therefore to overestimation of their volume. For this reason the shape parameter of the exponential functions is estimated considering a fraction of the peak intensity, i.e. one third of peak intensity minus the event intensity (see right image of Figure 3.6). This leads to events with rainfall volume spread along their durations which are more consistent with observations.

Several works state the fact that the variables describing a rainfall event, like total amount, intensity and peak, are generated by the same physical phenomenon, and therefore the dependence structure between these variables should not be neglected (Grimaldi & Serinaldi, 2006; Serinaldi & Kilsby, 2013). Significant dependence structures linking WSD and WSI or WSA have been reported by several authors (see e.g. Salvadori & De Michele, 2006; Balistrocchi & Bacchi, 2011; Vandenberghe *et al.* 2011; Kaczmarska *et al.* 2014). To include the dependence structure between some of the variables describing rainfall events, bivariate frequency analysis is performed with the aid of the copula models. Copulas are used here for modeling the joint behavior of WSA-WSD and WSI-WSPeak (see right images in Figures 3.7 and 3.8). As was mentioned earlier copulas are used for simulating pairs of pseudo-observations, which are then transformed to observations by the inverse marginal distributions describing the different involved variables (for further details see Genest & Favre, 2007). The following symmetric copulas commonly used in hydrological applications are considered in this work: Clayton, Frank, Galambos, Gumbel, Hüsler-Reiss and Normal. They are described by 1 parameter which is estimated using the moment-like and pseudo-likelihood methods. Additional asymmetric models described by more than 1 parameter (Tawn, Skew-t and Khoudraji transformation of Archimedean copulas) are as well included in the analysis to mimic the dependency of non-exchangeable variables. The DSD is modeled directly and not related or conditioned to other variables (see left image in Figure 3.7) with the assumption that the dependency between wet and dry spell durations is not significant.



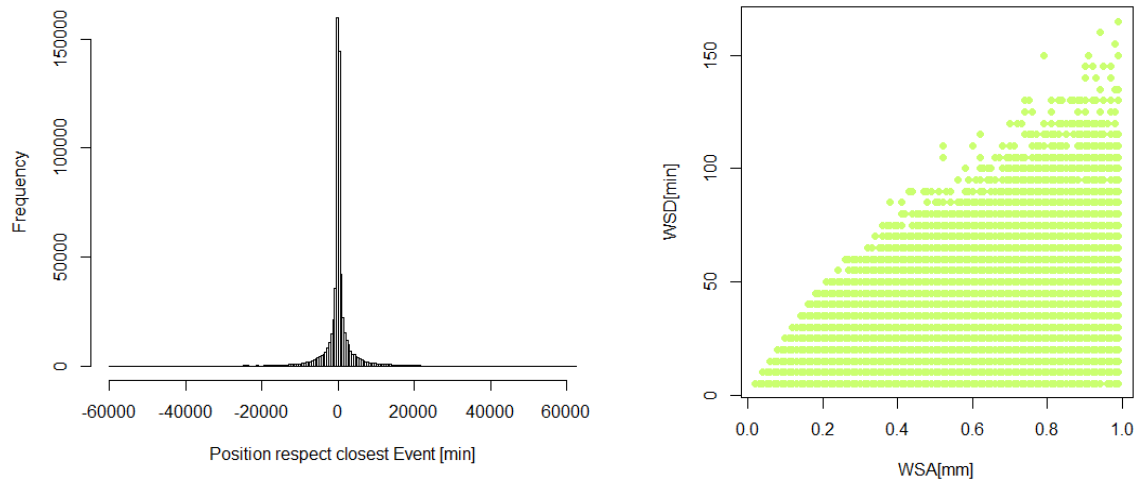
**Figure 3.7: Stochastic modeling of DSD (left) and WSD-WSA (right).**

As in the “old” model WSTpeak is estimated for each event with a uniform distribution that provides values between 0 and 1 for the ratio WSTpeak/WSD (see left image Figure 3.8). This model results in time series that are irreversible in time, as only events with a ratio of 0.5 would have the same shape for the original and reversed direction. Temporal irreversibility is a natural property in precipitation time series and as shown by Müller *et al.* (2017) can have significant implications on flow simulations.



**Figure 3.8: Stochastic modeling of WSTpeak (left) and WSI-WSPeak (right).**

The small events with WSA lower than 1 mm can have an important impact in urban hydrological simulations, especially for simulations performed in a continuous way. To include them, their total seasonal contribution is calculated as a volume for each station and a set of small events from all stations is created (see right image in Figure 3.9). They are incorporated in the times series by randomly selecting DSDs and assigning small events from the observed set until the total contribution is reached. Most of the small events occur directly before or after larger events ( $WSA \geq 1\text{mm}$ ) as can be seen in Figure 3.9 (left image). Therefore the histogram is used for allocating each small event within a DSD by attributing higher probabilities to time steps close to large events.



**Figure 3.9: Temporal location of small events with respect to events (left) and pairs of WSA and WSD describing small events (right).**

All the estimations for univariate and copula analysis and synthesis presented in this work are performed using R (R Core Team, 2016) a free software environment for statistical computing. The parameters of the marginal distributions are estimated by the L-moments Method using the *Lmomco* package (see Asquith, 2016), and Method of Moments and Maximum Likelihood with the *Fitdistrplus* package (see Delignette-Muller *et al.*, 2010). The mixture model is applied with the aid of the *Evmix* package (see Scarrott & MacDonald, 2012; and Hu, 2013). The *Copula* (see Yan, 2007; Kojadinovic & Yan, 2010; Hofert & Maechler, 2011; Hofert *et al.*, 2015), *CVine* (see Brechmann & Schepsmeier, 2013), *VineCopula* (Schepsmeier *et al.*, 2016), *sn* (see Azzalini, 2015) and *GenSA* (see Xiang *et al.*, 2013) packages are used for the parameter estimation of the copula models and the synthesis of random pairs of dependent variables.

A summary and graphical explanation of different steps involved in the synthesis of rainfall time series for single sites with the developed method is included in Appendix A.

## 3.4 MULTI-SITE PRECIPITATION MODEL

The multi-site model involves the *temporal* and *spatial* simulation of rainfall time series for several locations simultaneously.

### 3.4.1 PROPOSED MODEL

In this Thesis a multivariate approach using copulas is presented to extend the single site precipitation model to a multi-site synthesis. To generate several spatially correlated time series a methodology is proposed which involves firstly a hybrid model that defines the locations in which rainfall events are occurring simultaneously and a subsequently multivariate copula models used to generate synthetic events for the different locations. Vine copulas are used for this purpose which, as was mentioned previously, consists of decomposition of multivariate probability density into bivariate copulas. Thus the method presents the advantage of using 2-dimensional models for which a well-investigated variety of families is available and results in high flexibility to reproduce different dependence structures (Brechmann & Schepsmeier, 2013). The Vine copulas are used to simulate pseudo-values of WSA and WSD for the different locations.

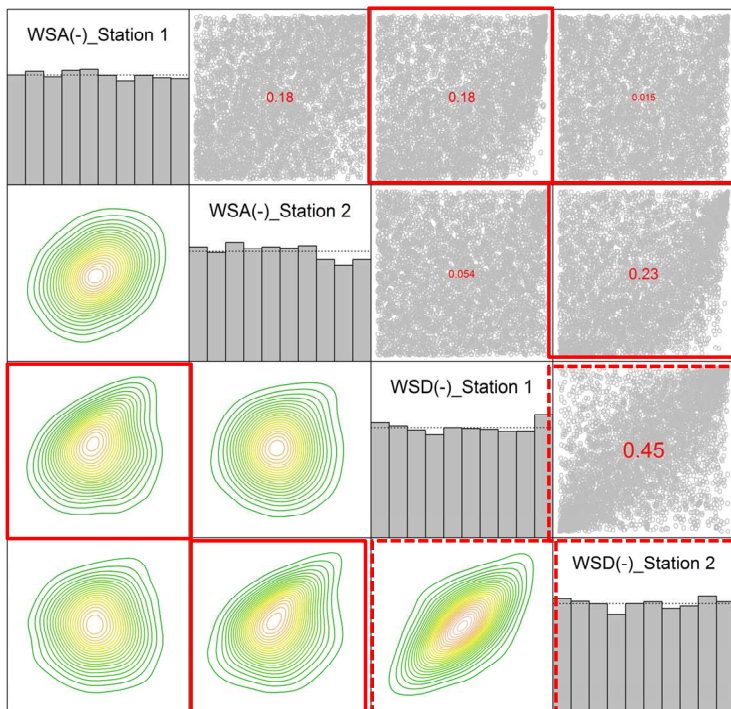
Consequently pseudo-values are bias corrected and transformed to real values using probability distributions fitted to all events observed in each of the stations. In order to obtain continuous time series dry spells must be introduced between rainfall events. Under the hypothesis that wet and dry spell durations are not correlated, the DSD are introduced into the model randomly for one of the stations, and then adjusted for the rest of the stations according to the wet spell durations in order to ensure that events occurring in several stations simultaneously are consistent in the temporal occurrence. Different subsamples of this variable are considered for the random sampling. The general procedure for incorporating DSD is explained in more detail in the Results Chapter (Sub-Chapter 5.3). In the last step the continuous series of events are converted to time series of 5 minutes in a similar way as for the single site model.

Bacchi & Kottegoda (1995) discuss about spatio-temporal correlation structures of rainfall events, and point out that these patterns are affected by topography, orography, coastal influences and the type of storm according to the season. Intense storms have an elliptical shaped field influence, whereas long lasting duration storms show a multi-cellular shape. In this Thesis, the identification of events occurring simultaneously is performed with a moving window of  $\pm 1.5$  times the WSD which is applied at the starting of each event in order to

decide whether events registered in different stations occur simultaneously or not. As in the single site model, summer and winter events are handled separately.

The hybrid model is set up to simulate long time series of events occurring in 1, 2, ... ,  $k$  stations. Percentages of events occurring in 1, 2, ... ,  $k$  stations along with percentage of events occurring at each of the stations estimated from observations are included within the model. The model results in a long time series of event occurrences in 1, 2, ... ,  $k$  stations along with the assignment of the event to particular station/s. Therefore a time series of event occurrences consisting of 0s and 1s is assigned to each of the  $k$  involved stations, with 0 meaning no event and 1 meaning event occurring at a particular station.

The next step is to assign event characteristics to the cases designated as with occurrences. For this step different Vine copula models are set up and used according to the number of stations in which events occur simultaneously. Figure 3.10 shows a pairs plot of the joint distribution of pseudo-values of hypothetical events occurring in two stations simultaneously, with scatter plots above and contour plots (under  $N(0,1)$  margins) below the diagonal. This information is used for setting up the Vine copula to model events occurring in two stations. Events are characterized by WSA and WSD for each station which are transformed to pseudo-values using the corresponding empirical distribution function. Pairs of pseudo-values of WSA-WSD corresponding to each of the two stations are marked with red boxes (with continuous lines) and indicate non-exchangeable behaviour.



**Figure 3.10: Pairs plot of pseudo-values of events registered in two stations simultaneously. Diagonal: histogram of marginals. Upper triangular matrix: bivariate scatter plots with Kendall's Tau between pairs in red. Lower triangular matrix: contour plots with standard normal margins.**

The structure of the vine trees are set up to guarantee that the first one contains the copulas modeling the pairs of WSA and WSD for each of the stations. Additionally highly correlated pairs of variables (high Kendall's Tau correlations) are as well included in the first tree. This type of tree structure results in bivariate copulas that are unconditionally modeling the pairs of variables WSA-WSD for each of the stations along with the highly correlated pairs, i.e. the pairs involved in the first level. The selection of the type of copula and estimation of parameters for each of the pair of variables is done in a sequential way and the algorithm is provided in the VineCopula R package (see Brechmann *et al.*, 2012; Dißmann *et al.* 2013). In order to guarantee a suitable modeling of WSA-WSD, the selected copula is supervised to see whether it is capable of modeling non-symmetric pairs, as is the case of the Tawn copula (see sub-Chapter 3.2), in case of presence of non-exchangeable variables. The general structure of the vine-copula model set up for the hypothetical case involving 2 stations previously presented is shown in Figure 3.11. All pairs of variables marked with red boxes are involved in the first tree of the structure. From these the pairs of WSA-WSD for each of the stations are pursued to be modeled by a Tawn copula. The pairs of WSD for both stations shows to have the highest correlation (see Kendall's Tau in Figure 3.10). The subsequent trees (2 and 3) involve the conditional pairs.

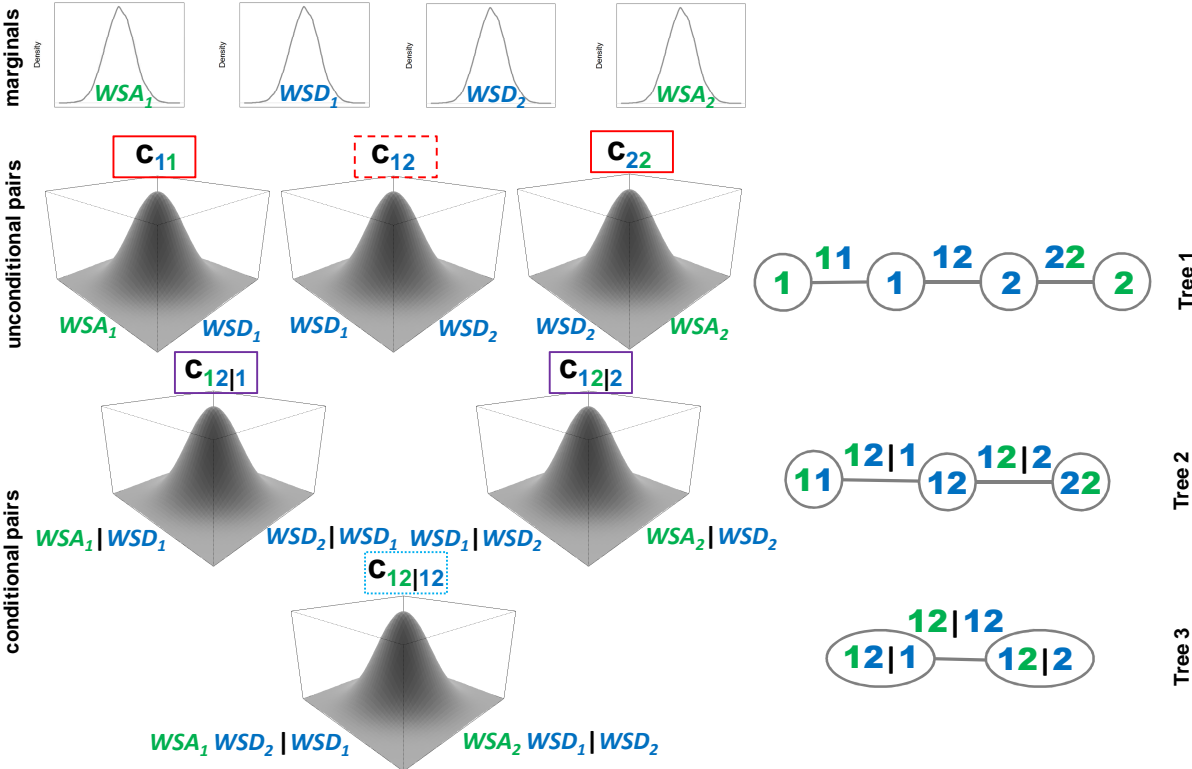
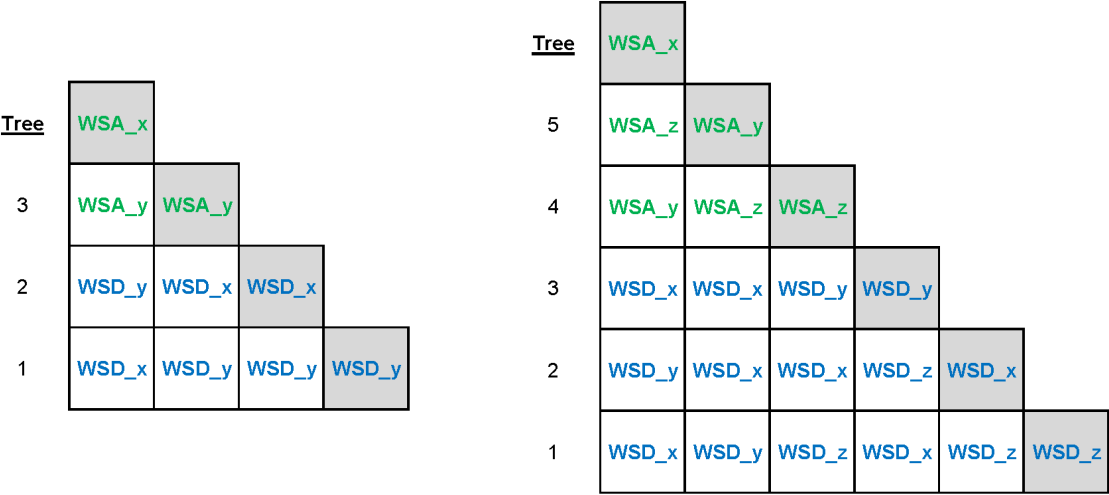


Figure 3.11: Components of a Vine copula to model events occurring simultaneously in two stations. Left: Marginal densities, unconditional and conditional pair copula densities (under  $N(0,1)$  margins for visualization). Right: vine tree structure.

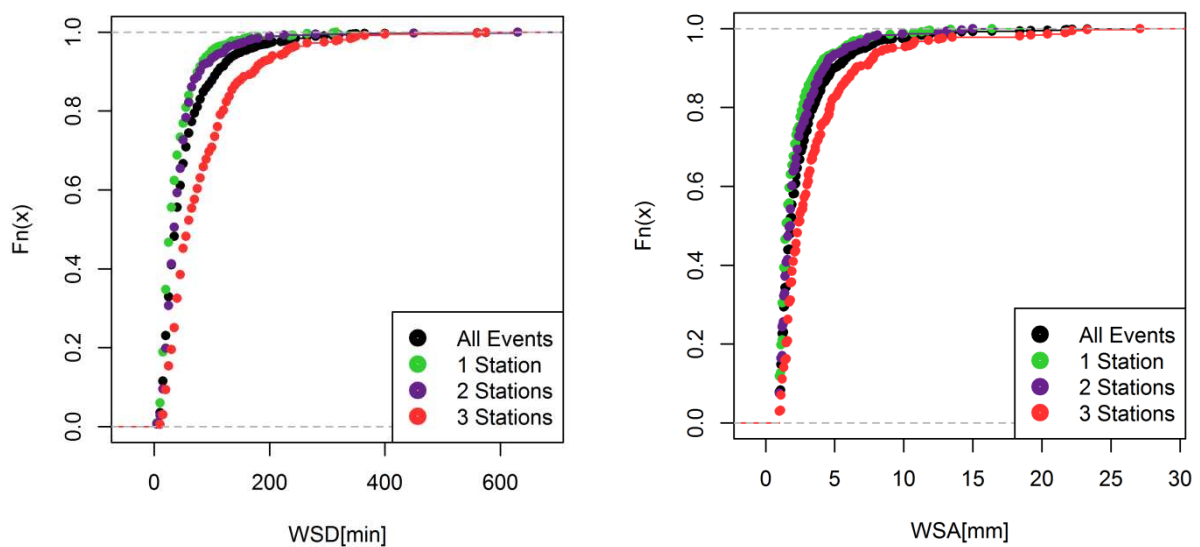


Tree structures are pre-defined for different cases. Events occurring in only one station comprise two variables WSA and WSD and are therefore modeled by a Vine copula consisting of one tree with one bivariate model. As shown in the preceding figures events occurring in 2 stations simultaneously, involve four variables, i.e. the WSA and WSD corresponding to each of the stations, and 3 tree levels. Cases involving occurrence in 3 stations, the dependence structure between 6 variables is to be modeled, resulting in 5 levels of trees and 15 different bivariate models. For cases including more than 2 stations, the conceptual illustration gets complex, and therefore lower triangular arrays established by Morales-Nápoles (2008) are advantageous for tree structure definition. The pre-defined tree structures for two cases are presented in Figure 3.12, in which WSA<sub>i</sub> and WSD<sub>i</sub> are the amounts and durations of events registered in the i station, with i being one of three stations x, y or z. These lower triangular arrays are read as follows: the boxes in gray color indicate the variables that are modeled by a bivariate copula with the different variables in the corresponding column and different tree levels. So for the 2 stations case (left plot) and first tree structure, the following pairs are modeled by bivariate copulas: (WSA<sub>x</sub>,WSD<sub>x</sub>), (WSA<sub>y</sub>,WSD<sub>y</sub>) and (WSD<sub>x</sub>,WSD<sub>y</sub>). In the second tree the bivariate copulas are conditioned to variables used in the first tree and the following pairs are modeled: (WSA<sub>x</sub>,WSD<sub>y</sub>) conditioned to WSD<sub>x</sub> and (WSA<sub>y</sub>,WSD<sub>x</sub>) conditioned to WSD<sub>y</sub>. In the third tree a bivariate model relating (WSA<sub>x</sub>,WSA<sub>y</sub>) is used conditioned to WSD<sub>x</sub> and WSD<sub>y</sub>. The same logic is followed to interpret the structure predefined to model events occurring in three stations (right plot). So in the first tree the following variables are modeled by a pair-copula: (WSA<sub>x</sub>,WSD<sub>x</sub>), (WSA<sub>y</sub>,WSD<sub>y</sub>) , (WSA<sub>z</sub>,WSD<sub>z</sub>), (WSD<sub>x</sub>,WSD<sub>y</sub>) and (WSD<sub>x</sub>,WSD<sub>z</sub>). It is the aim to directly model the variables of WSA and WSD for each of the stations in the first step of simulation.



**Figure 3.12: Predefined tree structure for the vine-copulas simulating rainfall events occurring simultaneously in two stations (left plot) and three stations (right plot).**

The probability distribution describing the different event characteristics, i.e. WSA, WSD or DSD, are estimated based on all events registered in each of the stations (see black dots in Figure 3.13). Therefore in order to transform the pseudo-values resulting from the different Vine copula models to values of events occurring in one isolated site, or in several sites simultaneously a Bias correction needs to be introduced. Isolated events are usually characterized by short durations, whereas events registered in several stations are usually longer and characterized with a bigger amount of rainfall. This behavior can be seen in the following figures (see Fig 3.13) for a case in which 3 stations are involved and therefore events are assigned to occur in either 1, 2 or 3 stations.



**Figure 3.13: Characteristics of rainfall events occurring in 1, 2 or 3 stations simultaneously.**

To summarize the proposed method involves different steps which are sequential for setting up the model, i.e. data based analysis and then application of the model to generate long time series. The analysis of data is described by the following steps:

1. Rainfall events are identified for each of the stations.
2. Events are compared for the multiple sites to define events occurring simultaneously in several sites.
3. Events are separated according to the season and the occurrence in 1, 2, ... ,  $k$  stations.
4. For each group of events a multivariate frequency analysis, ranging from 2-variate for events occurring in 1 station (WSA-WSD) up to  $2 \cdot k$ -variate for events occurring in  $d$  stations (WSA-WSD for each station), is performed with the aid of Vine copulas.

5. A Bias Correction is performed for each station by comparing the complete time series of events with the events belonging to the different groups, i.e. events occurring at 1, 2, ... ,  $k$  stations.

The simulation of time series for multi-sites involves several steps:

1. A hybrid model is used to generate the occurrence of rainfall events in simultaneous stations.
2. Pseudo-values of WSA and WSD are generated either for one or several stations simultaneously using bivariate or Vine copulas.
3. Pseudo-values are bias corrected and transformed to real values by using probability distributions fitted to events registered in each station.
4. Dry spells are introduced into the time series for one of the stations and adjusted for the surrounding ones.
5. Finally the continuous series of events are converted to time series.

A summary and graphical explanation of different steps involved in the synthesis of rainfall time series for multi-sites with the developed method is included in Appendix B.

### **3.4.2 ALTERNATIVE APPROACH: SIMULATED ANNEALING**

An alternative method is used to compare the proposed methodology. This method is called simulated annealing, explained and applied to hourly rainfall time series by Haberlandt *et al.* (2008). The method consists of two steps, first a temporal stochastic synthesis of rainfall is performed in different stations as single sites, i.e. neglecting the spatial consistency between them. In the second step the rainfall events from the different stations are resampled to reproduce the spatial dependency among the stations according to the distance between them. The resampling consists of a non-linear discrete optimization method and the aim is to minimize an objective function which takes into account the difference of spatial dependency structure between observed and simulated time series.

Three different spatial consistency measures are taken into account, all comparing 5-minutes time series of continuous rainfall from two sites, i.e. bivariate criteria. Different weights are assigned to each measure within the objective function. The first criterion was proposed by Wilks (1998) and consists of a continuity measure, the second is the Pearson's coefficient of correlation and the third one accounts for the probability of rainfall occurrence at both stations (p11). These measures have the following equations:

$$Continuity[-] = \frac{E(z_i|z_i>0, z_j=0)}{E(z_i|z_i>0, z_j>0)}, \quad (Eq. 14)$$

$$Correlation[-] = \frac{cov(z_i|z_i>0, z_j|z_j>0)}{\sqrt{Var(z_i|z_i>0) \times Var(z_j|z_j>0)}} \text{ and} \quad (Eq. 15)$$

$$p11[-] \approx \frac{n11}{n}, \quad (Eq. 16)$$

where  $z_i$  and  $z_j$  are the 5-minute rainfall time series at stations  $i$  and  $j$ .  $E(.)$  is the expectation,  $cov(.)$  is the covariance and  $Var(.)$  is the variance operator.  $n11$  is the number of time steps with rainfall occurrence in both stations and  $n$  is total number of time steps. The continuity expresses the ratio of expected amount of rainfall at one station for time steps without rainfall at the neighboring station, to the expected amount at the station for time steps with rainfall at the neighboring one. The lower this continuity measure the higher the interrelation between stations, and the opposite for the correlation and  $p11$ .

### 3.5 REGIONALIZATION OF PRECIPITATION

The regionalization of the model involves the *temporal* simulation of rainfall time series for single locations without observations.

#### 3.5.1 GENERAL CONCEPTS

The regionalization of the model, i.e. the synthesis of rainfall time series in areas without observations, consists of estimating the different functions involved within the model for any location by using information from the sites with observations along with additional data available for the whole region.

The different groups of variables to regionalize considered in this work are divided into 4 cases: the parameters describing the different probability distributions, the L-moments (LM) describing the rainfall characteristics (3 or 4 LMs according to the characteristic), the “orthogonal” set of parameters and the “orthogonal” set of LM. The last two cases consist of a transformation of the original parameters and LMs which removes the significant correlation between the sets to perform the regionalization into this transformed set. The orthogonalization method applied here is the Gram–Schmidt process, which uses projection to transform the original variables to the new sets (for further details see Anton, 2000). After the orthogonalized parameters/LMs are regionalized, the transformed variables must be retransformed to the original system. The projection operator is assumed to remain constant

for this purpose, i.e. the same proportion for the projection is applied to the regionalized values.

It is important to mention that for any of the 4 cases considered it is necessary to set up a priority of the variables to regionalize. The inclusion of the remaining parameters/LM to be used as additional variables in case of significant correlation is performed in a stepwise way, i.e. the first variable to estimate does not use any additional information from the rest of the parameters/LM, whereas the second uses the first one as additional, the third uses the first and second and so on. On the other hand, the first step of the orthogonalization process involves the selection of a first vector that remains constant and is used to project and transform a second one, the first and second are used for transforming a third one, and so on. Therefore for either of the cases a first set of variables will be regionalized deprived of the benefit of using the information from the rest. To define which variable to estimate in the first, second step and so on, a sensitivity analysis is performed to study the effect of over/under-estimating each parameter/LM into the resulting model. The parameters and LM are estimated based on a long time series of observed events and then each value is altered ( $\pm 10$ , 50 and 100% of the original values) to derive new model parameters. The alterations are performed individually for each case, i.e. the correlation between the different variables is neglected in this analysis. The models resulting from the different sets of parameters are used for generating long time series of events which are compared with the observed ones and the maximum distance between the two distribution functions is computed. The aim is to decide for each of the models the parameter/LM for which the deviation between both curves is the lowest for the different considered alterations (the least sensitive), and then this parameter/LM is set as the first one within the interpolation procedure. The parameters/LMs that show the highest sensitivity are interpolated in the last step of calculation, and favor from using the additional information of the ones interpolated previously.

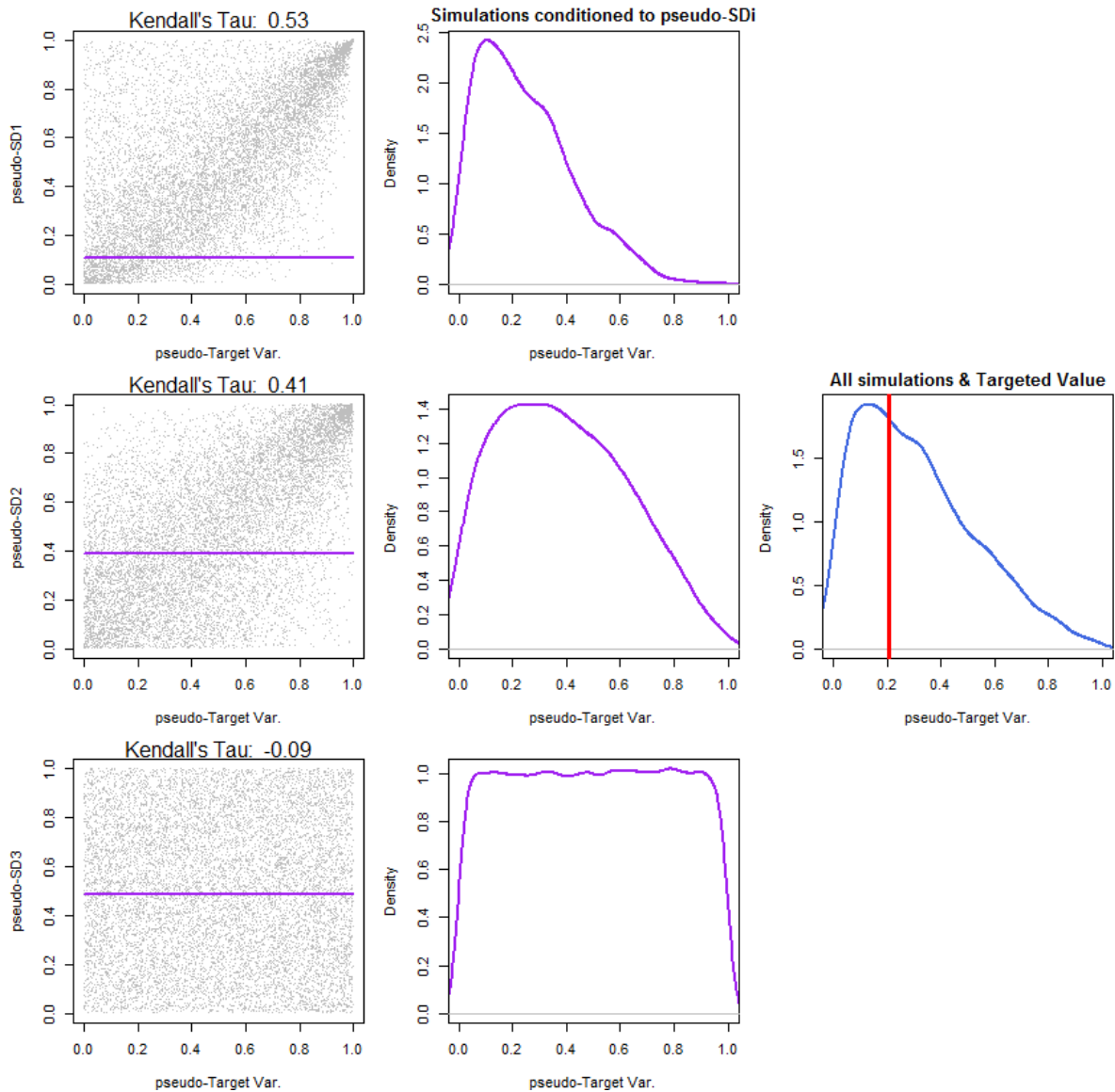
### **3.5.2 PROPOSED METHOD**

The proposed method for regionalizing the precipitation model consists of estimating the LMs of the characteristics describing rainfall events based on other information describing the sites (SDs) which are available for the whole space to be regionalized. Different order of LMs describing the components of the external structure (DSD, WSD and WSA) are estimated for each station and season. Then SDs which provide with different type of information describing the locations are used for estimating the LMs. The joint behavior between LMs and SDs is modeled with the aid of different copulas which result in high flexibility to reproduce different dependence structures. As some LMs describing each characteristic

show to be significantly correlated, they are also included in the regionalization in a sequential way, i.e. only in cases in which they have been previously estimated.

The dependence between the target LMs and each SD is measured by means of Kendall's Tau coefficient using the information from all stations. For each target LM bivariate models are fitted to each SD. As the bivariate models are copulas the different variables are related as pseudo-values. Each fitted copula is used to generate a sample of the target LM for a particular station based on the SD corresponding to that station. The amount of values to be generated by each copula is proportional to the Kendall's Tau correlation between the target variable and the descriptor. The descriptors used for simulating possible values are limited to the ones with absolute Kendall's Tau higher than 0.4, this way only variables with some degree of correlation are included. For some of the target LMs previously estimated LMs are as well included in the analysis in a similar way as SDs.

Figure 3.14 provides an example of the proposed method for one particular LM to be estimated based on three SDs. The first column shows the joint behavior and Kendall's Tau of target LM and each SD based on all stations. Each pair of variables is used to fit a copula model. The purple horizontal lines indicate the value of pseudo-SDs for a particular location. In the second column possible values of the target variable are shown as density functions, which are generated from each of the copula models conditioned to the pseudo-SD corresponding to the location of interest. In the third column the final result of all possible target LM is presented by mixing values generated using the different descriptors. The vertical red line indicates the targeted value of the LM, i.e. the one estimated from observations in the particular location. From all possible values any random value could be taken, in this Thesis the median is taken as estimated pseudo-LM. As mentioned a limiting value of  $\pm 0.4$  for Kendall's Tau is set up, so the SD shown in the third line would be excluded from the analysis. From the other two SDs the one shown in the first line shows the highest correlation and will therefore provide with more possible values of target LM, i.e. 56.4% of values resulting from the proportion of correlations  $0.53/(0.53+0.41)$ , whereas 43.6% will be provided by the second SD ( $0.41/(0.53+0.41)$ ).



**Figure 3.14: Graphical explanation of the procedure for a case with 3 Site Descriptors.**

Note that the copula based model provides regionalized pseudo-LMs which need to be back transformed to the original LM space by probability distributions. The univariate fitting of these models is based on all stations and following a procedure similar to the one presented in section 3.3.2. The regionalized LMs are then used to estimate the parameters of the functions describing the characteristics of the rainfall events. Some LMs need to be modified to fulfill minimum possible values (5 min for DSD and 1 mm for WSA).

The regionalization of the model explained here includes the probability functions of the variables describing the external structure of the precipitation model. The rest of the elements involved in the model are taken as regional elements based on all stations and therefore do not need to be regionalized. That is the case of the elements involved in the internal structures, the set of small events and the Copula model relating the WSA-WSD which are defined for each season and the whole region. The total volume and seasonal

proportion of small events are regionalized for each station by Nearest Neighbor (NN), i.e. by taking the value of the corresponding closest station.

A summary and graphical explanation of different steps involved in the synthesis of rainfall time series for sites without observations with the developed regionalization method is included in Appendix C.

### **3.5.3 ALTERNATIVE APPROACHES: MLR AND RFA**

In order to evaluate the copula based proposed approach other existing methods commonly used for regionalization of hydrological variables are applied. The final objective of the regionalization methods is to estimate long time series of rainfall events that properly reproduce the properties of the observed ones. Two alternative methods are applied for this purpose which involve either the estimation of LMs by multi-linear regression (MLR) or the estimation of the whole probability distribution by regional frequency analysis (RFA). All approaches are based on the same set of site descriptors used for the copula approach.

The LMs are regionalized using multi-linear regression (MLR). The site descriptors are used as explanatory variables along with significantly correlated LMs, which are included in the MLR in a sequential way, i.e. only in cases in which they have been estimated previously. To avoid colinearity between SDs partial correlations are analyzed (see Cox & Wermuth, 1996) and only SDs with values lower than 0.7 are included in the models estimation. A stepwise method is applied to define which variables should be included in the MLR model (stepwise model selection by Akaike Information Criterion, for more details see Venables & Ripley, 2002).

A method that has become very popular for several hydrological applications is the Regional Frequency Analysis (RFA). Therefore the variables describing the external structure of the model have been analyzed to test this popular method for generating time series in locations without observations. A brief description of the main steps involved within the method are resumed here and further details can be found in Leimbach (2017) where the method is applied to the same data sets used in this Thesis. The steps involved in the RFA follow the approach proposed by Hosking & Wallis (1997) and are performed with the following particular methods:

1. Selection of the Site Descriptors (SDs) to be included in the analysis which is performed with a non-linear approach called Random Forest



2. Grouping/clustering of the stations performed based on the selected SDs and applying the Partitioning Around Medoids method.
3. Normalization of time series by dividing each value with the flood Index, which in this case is the LM1 of the corresponding rainfall characteristic and station.
4. Normalized time series from stations belonging to one cluster are grouped together and Regional Frequency distributions are fitted for each of the clusters and rainfall characteristics.
5. Regionalization of LM1 for each station and rainfall characteristics using MLR method.

Three different regionalization methods are compared in this Thesis: copula-based, MLR and RFA. The first two methods aim to regionalize LMs, whereas RFA relies on the LMs for the grouping of the stations and results in a regional probability distribution for each of the groups. Despite of the methodological difference between these methods, they are all set up based on the same available information; hence a comparison among them is appropriate.

### **3.6 APPROACH FOR VALIDATING THE MODELS**

Given that the precipitation model is stochastic the performance is assessed on the basis of ensembles of many long synthetic time series which are compared with observed rainfall for a set of rainfall stations. Direct validations are performed by comparing different rainfall characteristics. Some of the tests are evaluating the external structure of the model, i.e. event based validation, whereas the final time series in 5 minutes temporal resolution, i.e. including the external and internal structures, are as well validated. Extreme values are of main interest in urban hydrology and are therefore evaluated separately for both cases. Event based validations have the advantage that they can be applied much faster compared to the ones based on the final 5 minutes time series, and are therefore very useful for comparing several models. Since a high temporal resolution is to be assessed, synthetic time series are indirectly validated based on hydrological modeling using a fictional urban system.

Several characteristics of the rainfall events are of interest for the present work, and thus should be evaluated to define whether the model is properly reproducing them. Validation is done based on i) event statistics, ii) total rainfall, iii) statistical properties for different temporal resolutions, iv) temporal correlation, v) extreme values and vi) intensity duration frequency curves. For the multisite modeling some additional spatial consistency properties are as well included. Indirect validation of the model evaluates the response of the hydrological model in terms of the ability to simulate the duration and volume of i) overflow and ii) flooding events.

Deviations between observed and simulated variables are calculated as relative percentages for each station (Equation 17) or as errors averaged over all stations, i.e. relative standard error (Equation 18). These errors are computed as:

$$Error[\%] = \frac{Z_i^* - Z_i}{Z_i} \cdot 100, \quad (\text{Eq. 17})$$

$$RSE[-] = \sqrt{\frac{1}{N} \sum_{i=1}^N \left( \frac{Z_i^* - Z_i}{Z_i} \right)^2}, \quad (\text{Eq. 18})$$

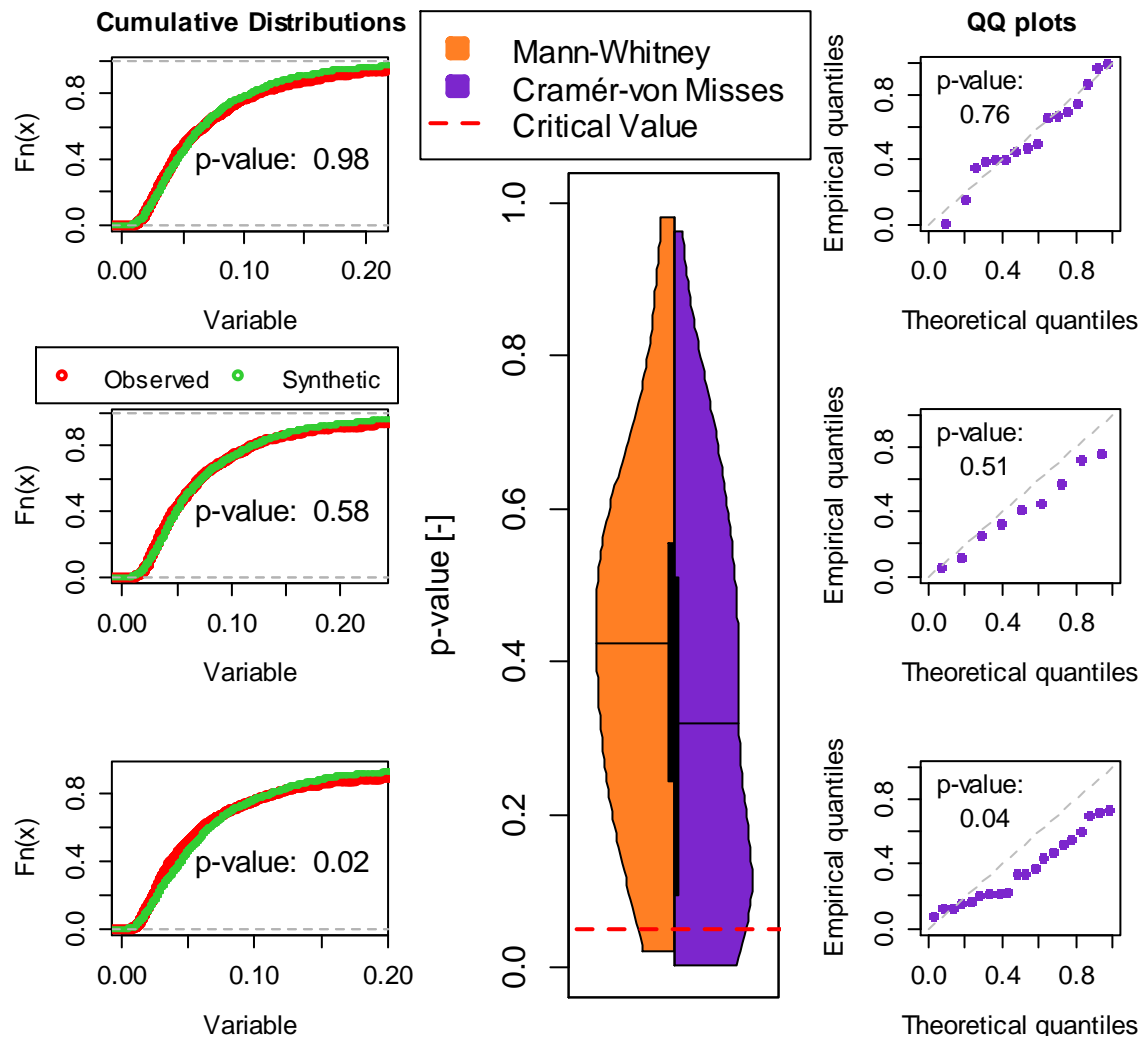
where  $N$  is the total number of stations,  $Z_i^*$  and  $Z_i$  are characteristics (e.g. moments, total seasonal rainfall, etc.) estimated using either the long synthetic time series or the observed one for a particular station  $i$ . Positive errors resulting from Equation 17 indicate overestimation by synthetic series.

Values resulting from several stations (e.g. errors) are presented as violin plots, which are a combination of box plots (black boxes showing the 0.25 and 0.75 quantiles) and density traces (see e.g. Figure 3.15). The combination of summary statistics and density shape into one diagram allows for quick and clear comparison of several model results (for further information the reader can refer to Hintze & Nelson, 1998). The density trace can be plotted symmetrically to the left and right of the box plot, or when the data is split into two groups (e.g. summer and winter) the violins can be split in half for an easy comparison between the distributions belonging to the groups. All violin plots are done using the *vioplot* R package (see Adler, 2015).

### 3.6.1 DIRECT VALIDATION (RAINFALL BASED)

#### 3.6.1.1 All Events – External Structure

Statistics of different variables describing the external structure of rainfall events (DSD, WSD, WSA and WSI) are compared. The Mann-Whitney (MW) test is used for this purpose, which according to Fagerland & Sandvik (2009), assesses to decide whether 2 populations are identical, or whether they differ in some way, i.e. in the mean, variance, skewness or a combination of them. For each station 100 synthetic realizations with length equal to observed time series are used. Cases in which the p-value of the statistic is lower than 5% indicate that the two compared series are not identical (a model with p-value higher than this critical value is acceptable). Examples of p-values resulting from different deviations between distributions are shown in Figure 3.15 (left plot).



**Figure 3.15: Example of p-values resulting from the Mann-Whitney test (left) and Cramér-von Misses test (right).**

The same procedure is followed to apply the Gini test and assess the performance of the model in terms of reproducing the variance. The test is applied to the cases for which the MW test shows acceptable results (for further details see Sordo *et al.*, 2016). Different moments (mean, standard deviation, skewness and kurtosis) describing time series of event characteristics are as well evaluated and presented as *RSE* for all stations (see Equation 18).

### 3.6.1.2 All Events – External and Internal Structure

Total seasonal rainfall is evaluated and compared with observed values and errors are computed using Equation 17. The performance of the model for increased temporal resolutions is as well evaluated based on long time series (observed and synthetic) which are aggregated to different temporal resolutions (original 5 minutes, 1 hour, 6 hours and 1 day). Thereafter wet time steps are identified for each time series, using the following thresholds for defining a time step as wet: 0.01 mm (5 minutes), 0.1 mm (1 and 6 hours) and 1 mm (1 day). Mean values of wet time steps along with the fraction of these steps are

estimated and deviations between synthetic and observed values are estimated using Equation 17. Temporal persistence of the original 5 minutes time series is another characteristic that is of interest when modeling rainfall. It is important to stress that the internal structure of the model proposed here does not aim to mimic the observed autocorrelations and thus overestimation is expected due to this fixed geometric structure (double exponential function). Nevertheless this characteristic is evaluated.

The multi-site precipitation model is evaluated in terms of reproducing the spatial dependence structure of the rainfall process. Different measures are used to compare observed and synthetic rainfall series based on single stations and areal precipitation. Three spatial consistency measures are used in this Thesis, all comparing 5-minutes time series of continuous rainfall from two sites, i.e. bivariate criteria. The first criterion is the continuity measure (see Eq. 14), the second is the Pearson's coefficient of correlation (see Eq. 15) and the third one was proposed by Mehrotra *et al.* (2006) and is denominated the Log-odds ratio. This last measure has the following equation:

$$\text{Log-Odds}[-] = \log \frac{p00_{i,j} \times p11_{i,j}}{p10_{i,j} \times p01_{i,j}}, \quad (\text{Eq. 19})$$

where  $p11_{i,j}$ ,  $p00_{i,j}$ ,  $p10_{i,j}$  and  $p01_{i,j}$  are the joint probabilities of rain (1) or no-rain (0) at stations  $i$  and  $j$ , i.e.  $p11_{i,j}$  indicates the probability of rain at both stations. These probabilities are calculated by the number of time steps for each of the cases over the total number of time steps (see e.g. Eq. 16). The Log-odds ratio takes into account the probabilities of either rain or no-rain in two stations simultaneously. High values of this ratio indicate high spatial interrelation between two stations, as was the case for the correlation coefficient, whereas the opposite is valid for the continuity measure.

An additional measure of spatial dependence is the comparison of the spatially averaged rainfall amount over the region. For this purpose each of the stations involved in the multi-site modeling is assigned an area of influence within the region by simple Thiessen polygon method. Time series resulting from this areal rainfall estimation are used for evaluating the model in terms of reproducing event characteristics, proportion of time steps with rainfall and total seasonal amounts.

### 3.6.1.3 Extreme Events – External Structure

Extreme values are evaluated applying the Cramér-von Mises (CvM) goodness of fit test. For this purpose a Gumbel distribution (see Equation 20) is fitted to the maximum annual values resulting from 1000 years of synthetic series, and then the maximum annual observed values

are tested to decide whether they belong to the fitted model. The cumulative distribution function of the Gumbel distribution is defined as follows:

$$F(x) = \exp \left[ -\exp \left( -\frac{x - \xi}{\alpha} \right) \right], \quad (\text{Eq. 20})$$

where  $\xi$  and  $\alpha$  are the location and scale parameters, respectively. P-values of this statistic which are lower than a critical limit (5% of confidence level) indicate that the extreme observed events do not belong to the theoretical model describing the long synthetic series. It is important to mention that the extreme values are associated to different durations of events, ranging from 5 minutes to 3 hours. For each case, the analysis is performed including all events with WSD equal or longer than the particular duration to find the maximum values and events are treated as blocks, i.e. only the external structure is involved. A model that performs satisfactorily would indicate p-values higher than the critical limit. Examples of p-values resulting from different deviations between empirical and theoretical distributions are shown in Figure 3.15 (right plot).

For the regionalization an additional analysis is included that involves the estimation of rainfall intensities associated to different durations and return periods and estimated from the fitted Gumbel distributions. The selected durations are 5, 15, 30, 60 and 180 minutes and the return periods 2, 5, 10 and 20 years. The procedure is done based on observed and synthetic time series. Errors are then calculated by comparing resulting extreme intensities from both time series.

#### 3.6.1.4 **Extreme Events – External and Internal Structure**

A comparison based on intensity duration frequency curves (IDF), which are commonly used for urban hydrological design (see Koutsoyiannis *et al.* 1998; Ariff *et al.*, 2012), is as well included. The curves are constructed by univariate rainfall frequency analysis from observed and synthetic time series based on moving windows of different durations (from 5 minutes up to 1 day) and for the return periods of 1, 2, 3, 5, 10 and 20 years. Stations with long observations are selected for this comparison. The analysis is performed according to design standards available for the study region (DWA, 2012). i.e. Gumbel distributions (see Eq. 20) are fitted to time series for the different durations. The synthetic time series comprise here ten realizations with length equal to observed time series. Errors between observed and synthetic intensities (median from realizations) are computed for each duration and return period using Equation 17.  $Z_i^*$  and  $Z_i$  are amount associated to one return period and one duration estimated using the Gumbel distribution fitted either to the synthetic or observed time series. Positive errors indicate overestimation by synthetic series.

For comparison, a similar analysis is performed based on IDFs available for design purposes. Available IDF are used to extract design rainfall events corresponding to the selected durations and return periods which are compared with IDFs constructed based on observations. These values are as well included since it is a common worldwide practice to use this data for dimensioning sewer networks, rainwater retention basins, infiltration systems, etc.

For the multi-site model the spatially averaged rainfall time series are used for constructing the IDFs. For the IDFs available for design practice an average is as well calculated based on the area where the multiple sites are located and a Reduction Factor is applied (more details are presented in the Sub-chapter 4.6).

### **3.6.2 CROSS-VALIDATION**

The leave-one out cross-validation (LOOCV) is performed in order to evaluate the regionalization of the model. The principle is very simple and consists of estimating the value of a variable for a particular location based on information from all other locations, i.e. excluding the sample value at that location. Thereafter the value is replaced by its original value and used for estimating variables for other locations. So for each station, the parameters of the models (regional frequency distributions, flood indexes or copula models along with the marginal distributions) are re-estimated using all the remaining stations.

As was explained different regionalization methods are included within this Thesis, i.e. the copula-based proposed method and the two alternative methods RFA and MLR. As the methods involve different models within the regionalization procedure, the cross-validation varies as well. For the MLR method the selection of explanatory variables is not redefined within each cross validation step, but the parameters involved in the regressions are recalculated. For the RFA the grouping of the stations is not redefined in each step of calculation, and the type of functions used for fitting the parameters either. What changes in each step are the parameters describing the regional functions as they are re-estimated for each cross validation step. Lastly for the copula based method the type of functions describing the marginal behavior of the different LMs are not changed, but the parameters are re-estimated within each step of cross validation. The type of copulas along with their parameters and variables used for the estimation (which is determined by the Kendall's Tau values) are re-estimated in each step of cross validation.

A robustness analysis of cross-validation is to be performed by applying the different regionalization methods and estimating *RSE* errors based on different events characteristics, i.e. 4 moments (mean, standard deviation, skewness and kurtosis) of DSD, WSD, WSA and WSI corresponding to summer and winter, and extreme events associated to different durations and return periods (5, 15, 30, 60 and 180 minutes and 2, 5, 10 and 20 years). The *RSE* are estimated based on cross validation (*LOOCV*) and as well based on an additional analysis (*ALL*) in which all the stations are considered in the regionalization, i.e. excluding the dropping of one station in each step. The robustness is evaluated based on the ratio between the two *RSE*:

$$RATIO\_RSE[-] = \frac{RSE_{ALL}[-]}{RSE_{LOOCV}[-]} \quad (\text{Eq. 21})$$

values of this ratio close to one indicate that the method is robust to new stations.

### 3.6.3 UNCERTAINTY ANALYSIS

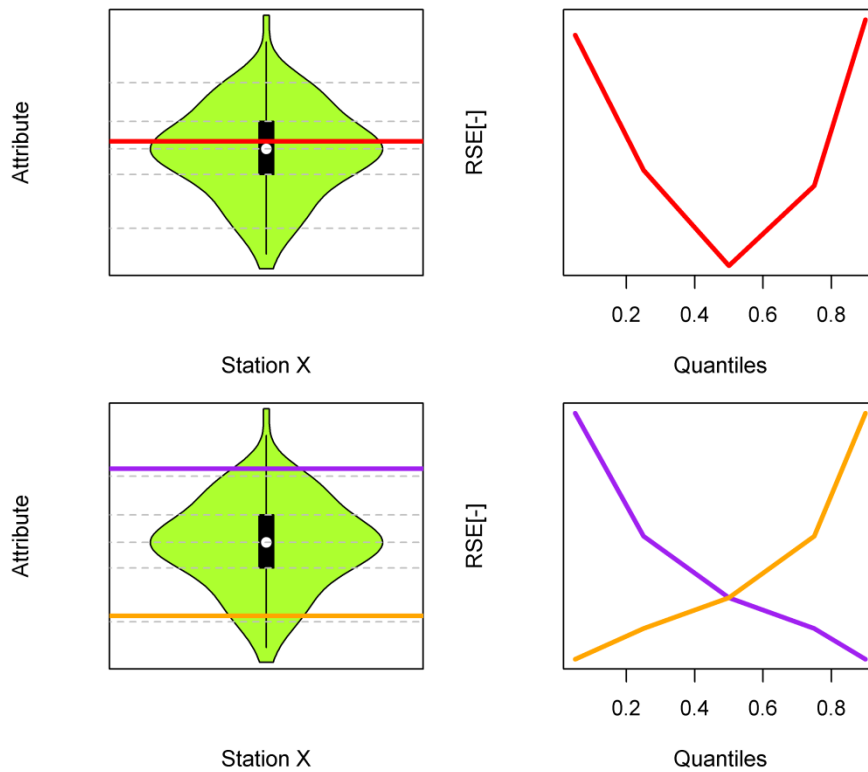
As exposed by Muller *et al.* (2009) results of models are uncertain as they can only provide a partial representation of reality. According to the authors, these uncertainties depend on the quality and quantity of available data, i.e. measurement and sampling errors, model errors due to imperfect structure and uncertainties in model parameters. In this Thesis, two different sources of uncertainty are considered namely the natural variability of the stochastic process and the uncertainty in the parameter estimation. Given that the precipitation model is stochastic and the performance is assessed based on many realizations, the aim is to provide some assessment in terms of model uncertainty for a range of rainfall attributes extracted from many realizations and considering all stations. Model results are always compared with information obtained from observed time series to assess the uncertainty.

The uncertainty analysis is performed based on 100 simulations each with a total length of 100 years. From each simulation different attributes characterizing the generated rainfall are studied, i.e. the mean, standard deviation, skewness and kurtosis of DSD, WSD, WSA and WSI corresponding to both summer and winter events. These attributes are estimated based on the same number of events as the observations. Extreme values are as well analyzed and the intensities associated to different return periods (2, 5, 10 and 20 years) and durations (5, 15, 30, 60 and 180 minutes) are included. All the analyses are based on time series that involve the external structure of the rainfall model.

For each station the attributes resulting from the 100 simulations are compared with the ones estimated based on the complete observed time series, i.e. with the aimed values. Five different quantiles (0.05, 0.25, 0.5, 0.75 and 0.95) characterizing the attributes resulting from

the simulations are used for calculating the deviation between synthetic and target values (estimated from all observations). Figure 3.16 (top row, left graph) shows an example of model results for one station, in which the horizontal red line is the value of the attribute calculated from observations; the violin indicates the distribution of the attribute corresponding to 100 simulations and the gray lines indicate the different quantiles for which the deviations from the red line are calculated. The procedure is done for all stations and *RSE* (see Equation 18) is calculated for each quantile considering all stations. This is done for the different attributes separately. Figure 3.16 (top row, right graph) shows an example of resulting *RSE* by considering several stations for the different quantiles. The red lines indicate two models that result in minimum *RSE* for the median value and increasing values in a symmetric way as the quantile is deviated from the median. These results indicate non-biased models when all stations are considered. A non-biased result indicates that overall for all stations results are not biased to either over or underestimation. Information regarding the deviation from target values can be inferred by comparing the continuous and non-continuous lines. Both results indicate non-biased models; however the dashed line indicates higher *RSE* values and therefore a model that produces larger deviations from the target values, even though these deviations can be positive for some of the stations and negative for others. Other possible cases are shown in the lower row of Figure 3.16 in which the observed value could be either the orange or the purple line. If this is the case for most of the stations (orange or purple) then the resulting *RSE* would be like the ones shown in the right plot. The orange line shows an increase in the *RSE* as the quantile increases, indicating a model that overestimates the attributes for most of the stations. The purple line indicates the opposite behavior, as lower quantiles indicate higher *RSE* (and therefore larger deviations) for most of the stations; this model is often resulting in underestimations.





**Figure 3.16: Example of models unbiased (red), overestimating (orange) and underestimating (purple) a particular attribute assessed on one station (left plots) and on many stations (right plots).**

As was mentioned, two different sources of uncertainty are considered here, i.e. the natural variability of the stochastic process and the uncertainty in the parameter estimation. The first uncertainty is analyzed using fixed parameters which are estimated based on all available observations and are assumed to be correct. The uncertainty of model parameter estimation is analyzed based on sample size, i.e. parameters are estimated based on different scenarios of available data.

First the uncertainty of the Alternating Renewal (AR) process is analyzed by using the fixed parameters to generate synthetic time series for single sites and evaluating the behavior of resulting *RSE* based on all stations. The aim is to say whether the model is unbiased or whether there is a tendency of under/overestimating a particular attribute. Thereafter the uncertainty of the parameter estimation due to input data is analyzed. This is done by considering different percentages of available data, namely 100, 75 and 50%, which are randomly sampled with replacement from all observed events. For each scenario of data availability, the random sampling is performed 100 times and used for estimating model parameters and thereafter generating long time series of events. The resulting *RSE* based on all stations and for different quantiles are compared with the ones from the AR model (with fixed parameters). The aim is to assess the robustness of the model to input data, to see whether the results have a similar or different behaviour as the original one and how are the *RSE* values differing when the different input data sets are considered.

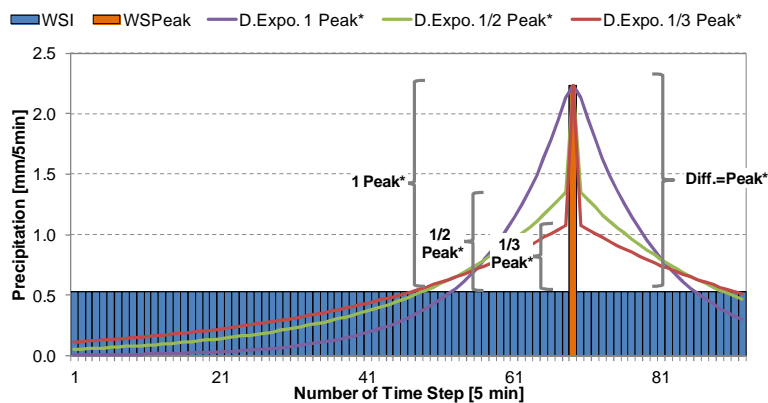
### 3.6.4 APPLICATION TO URBAN HYDROLOGY

Application of the different rainfall time series is assessed by urban hydrological modeling using a fictional drainage system. Fictional catchments have been used in the past for evaluating rainfall input influence (see e.g. Arnaud *et al.* 2002) and omit the calibration procedure. Simulations are conducted in a continuous way for uninterrupted 5 minutes precipitation data. Continuous hydraulic modeling is only possible for low complexity urban systems, which could represent a portion of the sewage system of a city. The advantage of continuous simulations is that they allow for an overall assessment since the whole time series of rainfall and flood events are included in the analysis, whereas a more complex urban system would only allow for event based hydraulic simulations, thus limiting the assessment to selected events. Furthermore, no pre-definition of antecedent soil moisture conditions or duration of rainfall events is required as it would be for event based analysis. Some limitations of using a simple fictional catchment are that it represents a portion of a whole system and is only considering one possible representation.

Del Giudice *et al.* (2013) state that the input uncertainty and its propagation deliver urban hydrological simulations which are biased. To assess the possible uncertainty introduced by observed time series, an analysis is performed by randomly sampling events and creating new time series with length equal to original observed time series. This procedure is performed 10 times. The observed and new time series are used as input to the hydrological model and simulation results are compared. Two different assessments are carried out one in which the random sampling is performed without replacement of events and an additional one in which all events are involved within each sampling step (sampling with replacement). The former analysis is meant to evaluate the source of error due to a possible existence of serial autocorrelation in the time series; whereas the latter is meant to assess the natural variability of the stochastic process in terms of hydrological response.

The role of the shape of events in the reproduction of flood events is additionally evaluated. As was mentioned the internal structure is modeled by a mixture of two exponential functions described by one single parameter. A drawback of this model is that it consists of a single peak and thus the synthetic events are always distributed around it. To minimize this shortcoming, different alternatives of shapes are tested. All alternatives are based on the exponential shape, but the peak used for estimating the parameter of the internal structure differs. The peak intensity ( $WSP_{peak}$ ) is in all cases the same, only the concentration of the volume around it changes. To estimate the shape parameter three different alternatives are

considered which are shown in Figure 3.17. The “D. Expo. 1 Peak\*\*” case corresponds to the internal distribution used for the “Old Model”. The other two alternatives result in rainfall events with volume that is distributed along the whole duration of the events, and less concentrated around the peak. For the three alternatives the total WSA, WSD, WSPeak and WSTpeak describing the events are exactly reproduced, the only thing that changes is the parameter describing the shape.



**Figure 3.17: Evaluation of internal structure in terms of hydrological modeling.**

Observed time series are used for modeling flood events and the resulting values are used as a basis for comparing the different alternatives. Rainfall events causing flooding are identified from the observed time series. Characteristics describing these events are extracted and used for fitting the different theoretical shape for each of the events. Observed events are then replaced by the new ones with the theoretical internal distributions. The new time series are used for modeling flood events which are thereafter compared with the ones resulting from the original observed time series. The internal structure that provides the lowest overall errors is preferable.

The performance of the precipitation models is assessed by urban hydrological modeling. Results from the simulations using observed data and ten synthetic realizations with length equal to observed time series are compared for each station both for single site and regionalization models. Annual runoff statistics are compared with observed values and an error is computed (see Equation 17). This comparison includes errors of extreme statistics, i.e. volume and duration of flood events within the urban system, as well as long-term mean annual overflow properties, i.e. volume and duration of spilling to a receptor water body within a combined sewage system. The errors resulting from the assessment of the natural variability of the process are as well included.

Finally for the single site model a comparison between observed, synthetic time series and events provided by IDFs available for design purposes is performed based on urban modeling. Stations with long observations are used for simulating and extracting flood events, which are ranked based on the volume and associated to different return periods (1,

2, 3, 5, 10 and 20 years). A similar analysis is performed based on 10 realizations of synthetic time series and median values are compared to observations. Available design events associated to different return periods and with duration equal to the time of concentration of the system are used for estimating flood volumes. Design events resulting from IDF curves can only be modeled on an event based mode and therefore require the set up of initial soil moisture conditions. Uncertainty of antecedent soil moisture conditions and precipitation are considered following the procedure proposed by Haberlandt & Radke (2014), i.e. taking 3 different initial soil moisture conditions (dry, average and wet) and accounting for an error of precipitation up to  $\pm 15\%$  adopted as design tolerance ranges. All together 9 cases are run for each return period. The return period of the precipitation is assumed to be equal to the flooding event, which is commonly done in practice. Errors are calculated based on both data sets, i.e. synthetic and design events, using Equation 17. Note that the evaluation of long-term overflow properties is only possible for continuous simulations and therefore available design events cannot be used for assessing this property.

All results related with urban hydrological modeling are presented in the Chapter 6 in which application to urban hydrology is discussed.

### **3.7 SUMMARY COMPARING EXISTING AND PROPOSED PRECIPITATION MODEL**

As mentioned the model presented here is based on an existing one which is further developed by introducing several modifications like multivariate analysis, seasonality, simulation of small events and multisite synthesis. A special focus is given to properly reproduce the extreme values. Seasonality is included by modeling the marginal behaviour of the different variables separately for summer and winter events. To include the dependence structure that relates some of the variables, like WSA and WSD or WSI and WSPeak, bivariate frequency analyses are performed with the aid of the copula models. To summarize all the changes introduced to the model the main differences between the existing and new proposed one are listed in the following Table (see Table 3.1), features that have been explained in more detail in the previous sub-chapters.

**Table 3.1: Main differences between existing and improved model.**

	<b>Old Model</b>	<b>Updated Model</b>
<b>Seasonality</b>	Neglected	Considered
<b>External structure</b>	WSD is directly simulated by a probability distribution WSA moments are derived from linear regressions based on WSD WSA moments are used for estimating a specific probability distribution from which the WSA is simulated	Bivariate copula is used for simulating pairs of pseudo WSA-WSD, which are then transformed to real values by their probability distributions
<b>Internal structure</b>	WSPeak is derived from a simple linear regression based on WSI Estimation of the $\lambda$ parameter is based on the total peak intensity	Bivariate copula is used for estimating WSPeak conditioned to WSI Estimation of the $\lambda$ parameter is based on a portion of the peak intensity
<b>Small Events</b>	Neglected	Considered
<b>Regionalization</b>	Multi-linear regression or External Drift Kriging to estimate the parameters of marginal distribution and linear regressions	Estimation of LMs describing the rainfall characteristics. Copulas are used for the estimation conditioned to site descriptors and previously estimated LMs
<b>Multisite Synthesis</b>	Neglected (Simulated Annealing only applied for a temporal resolution of 1 hour)	Vine-copulas are used to model events characteristics in several stations simultaneously
<b>All events evaluation</b>	Considered	Considered
<b>Extreme Value Analysis</b>	Neglected	Considered
<b>Overflow in urban areas</b>	Considered	Considered
<b>Flood events in urban areas</b>	Neglected	Considered

## 4 DATA

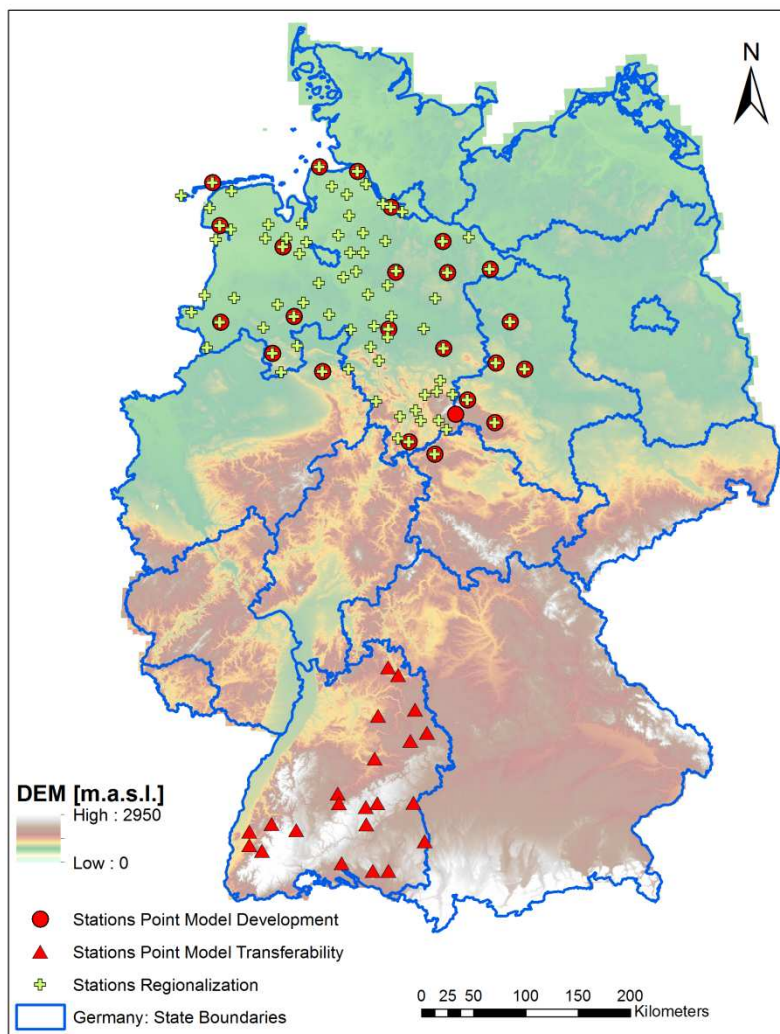
The data used for developing and evaluating the different methods presented in this Thesis are described in this Chapter. In the first sub-Chapter (4.1) a description of the study area along with a general presentation of all stations is included. The following sub-Chapters (4.2 to 4.4) include a more detailed analysis of the stations used for the development of the single site, multi-site and regionalization of the model. Site descriptors, which are involved in the regionalization, are as well presented (sub-Chapter 4.5). Thereafter (sub-Chapter 4.6) a description of KOSTRA, i.e. IDF's available as design practice for the study area, is presented. Finally (sub-Chapter 4.7) the fictional urban system used for validating the precipitation models is described.

### 4.1 STUDY REGION

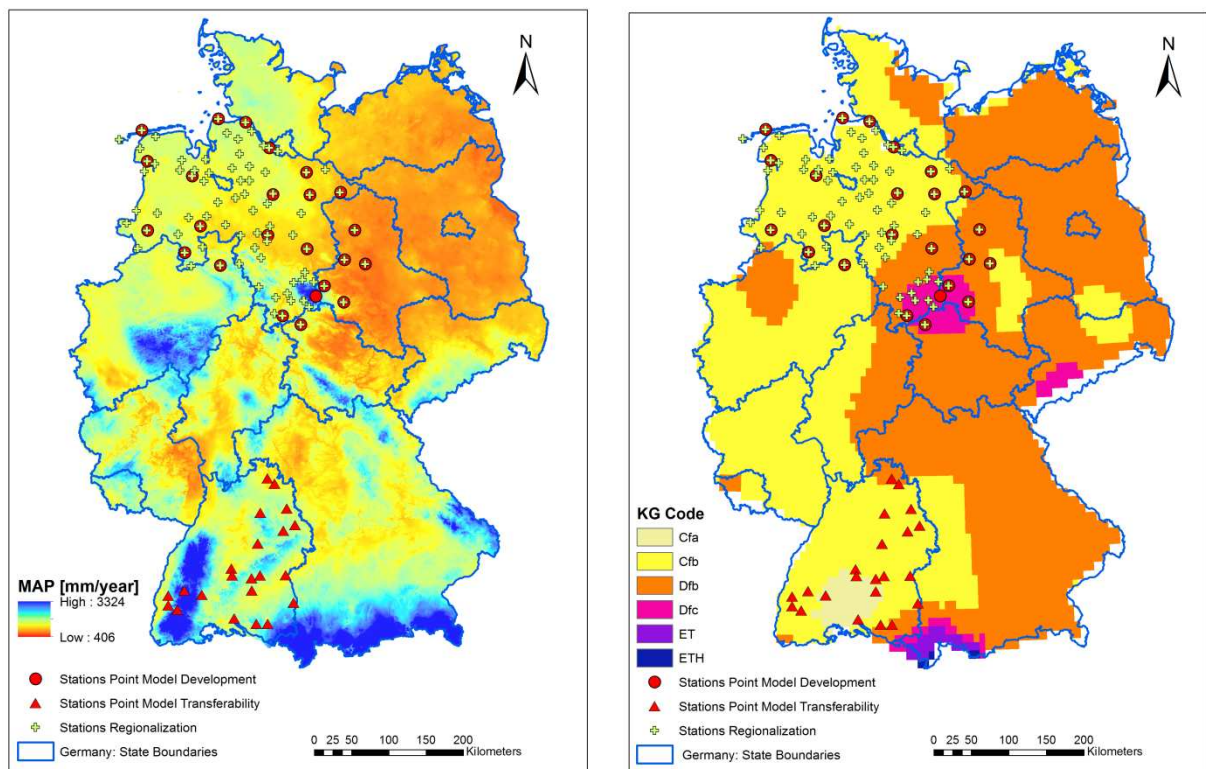
The state of Lower Saxony (NS) and surrounding areas, located in the northwest of Germany is used to develop the single site precipitation model and the regionalization. The stations belonging to the state of Baden Württemberg (BAWU), located in the southwest of the country, are used for evaluating the transferability of some of the methods proposed in this Thesis (see Figure 4.1). As can be seen from the figure 24 stations are used for developing and 22 for transferring the single site model, whilst 81 stations are used for its regionalization. The multisite model is based on some of these stations (11 in the North and 10 in the South). Further details describing all stations can be found in Appendix D.

The elevation of the NS study area ranges from sea level for the stations located in the northwest region bordered by the North Sea, to up to 600 mNN (meters above German Standard Zero), for Harz Mountains in southeast, with most of the stations in a height of 50 mNN. As described by Liu *et al.* (2013) BAWU consists of two mountain ranges: in the east the Swabian Alps with lower elevation and the Black Forest with medium elevation in the west which resembles a high plateau, rising abruptly in the northwest to a flat or gently hilled upland. For this study area the stations are located around the two mountain ranges and their elevation ranges from 200 mNN up to around 800 mNN, with a median value of 500 mNN. The total annual rainfall registered during the period from 1980-2010 in the regions where the stations are located ranges from 400 mm/year in the northeast to up to 1500 mm/year in the southwest (see left map in Figure 4.2). Rainfall occurs all year round and summer precipitation events are dominated by convective type, whereas the occurrence of stratiform events is more frequent during the winter (see Eggert *et al.*, 2015).

After Köppen–Geiger climate classification, see Peel *et al.* 2007, Germany presents the following climates: Temperate without dry season and summers hot and warm (Cfa and Cfb), cold without dry season and summers warm and cold (Dfb and Dfc), and Polar-Tundra with elevations lower and higher than 1500 meters above sea level (ET and ETH). As can be seen in Figure 4.2 (right map), the Polar-Tundra climates are only present in the Alps region (south of the country) not covering the areas of study. Most of the stations in the flatlands of NS are belonging to the Cfb climate caused by warm westerly winds from the North Sea, whilst some stations towards the southeast around the Harz Mountains are belonging to the Dfb or Dfc climates, caused by the continental effect combined with the higher altitudes. BAWU stations are characterized mostly by Temperate climate (Cfa and Cfb) and a few stations as Dfb, the topography and valley effect have a strong influence on the structure of the local climate. The two regions (NS and BAWU) show to be very different in terms of topography, total rainfall and climate. Therefore BAWU is considered to be a challenging region to test the transferability of the precipitation model.



**Figure 4.1:** Digital elevation map of Germany with location of rain gauge stations used for developing (24 circles) and transferring (22 triangles) the single site model, and for regionalization (81 crosses).



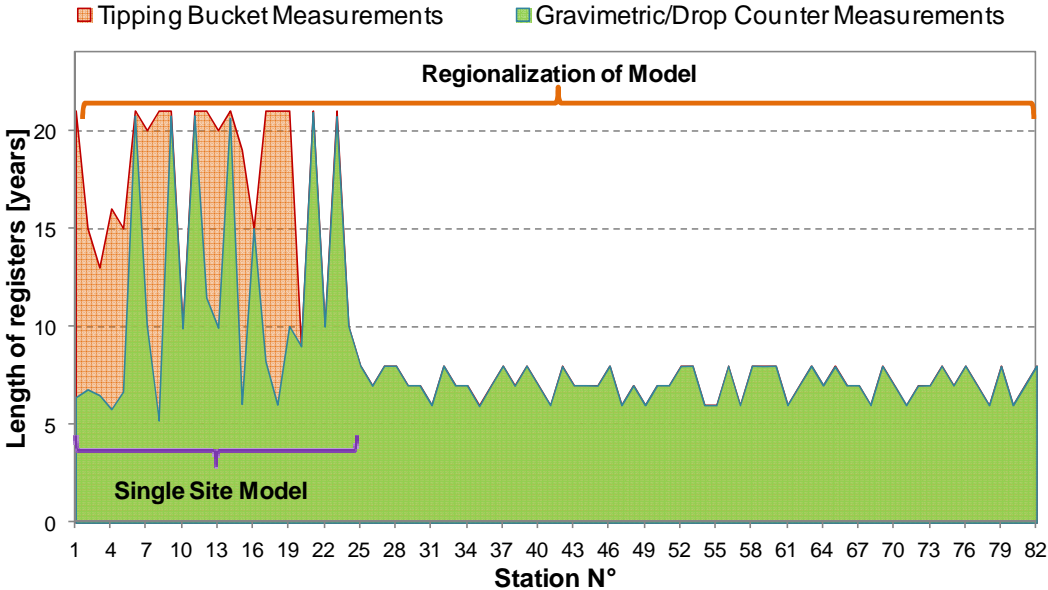
**Figure 4.2: Left: Mean Annual Precipitation (1981-2010) and Right: Köppen-Geiger climate classification in Germany with location of all rain gauge stations.**

The available data for the Northern part of Germany is provided by the German National Weather Service (Deutscher Wetterdienst DWD) and corresponds to stations belonging to the new network (automated weather stations) that is operating since the mid 1990s. The measuring devices are either gravimetric/drop counters (Pluvio-Ott) with depth resolution of 0.01 mm or tipping buckets with resolutions of 0.1 mm. The different resolutions have some impact when rainfall events are identified; nevertheless to achieve a data set with long records in several stations this shortcoming has to be accepted. Around 84% of the total registers correspond to the drop counter and the rest to tipping bucket as shown in Figure 4.3. The temporal resolution is for all cases 1 minute and time series are aggregated to 5 minutes for event identification and final model evaluations. Regarding the stations located in BAWU, this information was provided in a 5 minutes temporal resolution by the University of Stuttgart within the SYNOPSE project. No detailed information regarding the measuring mechanism of these stations is available for this Thesis.

Measuring precipitation in such short time intervals is very challenging and small losses can result in high relative errors. Mechanic errors, such as device blocking or data transmission, are controlled from the original 1 minute time series (if available). Some unrealistic high



values (e.g. a value of 55 mm in 1 minute is present in one station in NS) and repeated measurements are detected for some of the stations, and thereafter discarded.

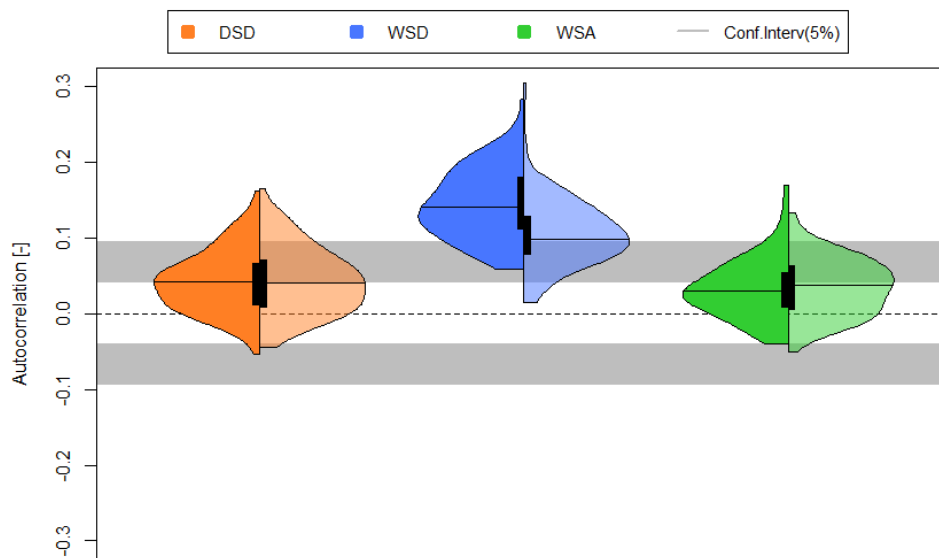


**Figure 4.3: Description of data available for single site model development and regionalization.**

Homogeneity of the data is checked based on monthly values. A double sum analysis is performed for each of the regions and for the period in which observations for all stations is available, i.e. January 2007 to December 2012. The cumulative sum of monthly values for each station is plotted against the sums corresponding to all stations together (for each region separately). The plots are presented in the Appendix E, indicating that data are overall homogeneous. For the NS case only two cases show some atypical behavior at the end of the registers (stations E501 and 10215), but this is due to some months with missing values at the end of the analyzed period. Similar results are obtained for two stations in BAWU (stations 71573 and 90307) which have missing values at the beginning of the analyzed period.

As was mentioned the precipitation model is based on the alternating renewal process (ARP), i.e. the modeling of succession of independent rainfall events. Therefore some analyses based on rainfall events are presented here. These events are derived from each of the time series based on the criteria adopted in this Thesis (DSDmin, WSAmin and WSImin) and are separated into Summer (April to September) and Winter (October to March), same seasonal selection was adopted by Eggert *et al.* (2015). The main hypothesis behind frequency analysis and the ARP is that the variables are independently identically distributed (IID). A sample of data is IID if each random variable belongs to the same probability distribution and all are mutually independent. Statistics derived from random samples, such as mean, standard deviation, etc. are assumed to be representative of the whole population of possible values. Therefore the autocorrelation of the different variables describing the

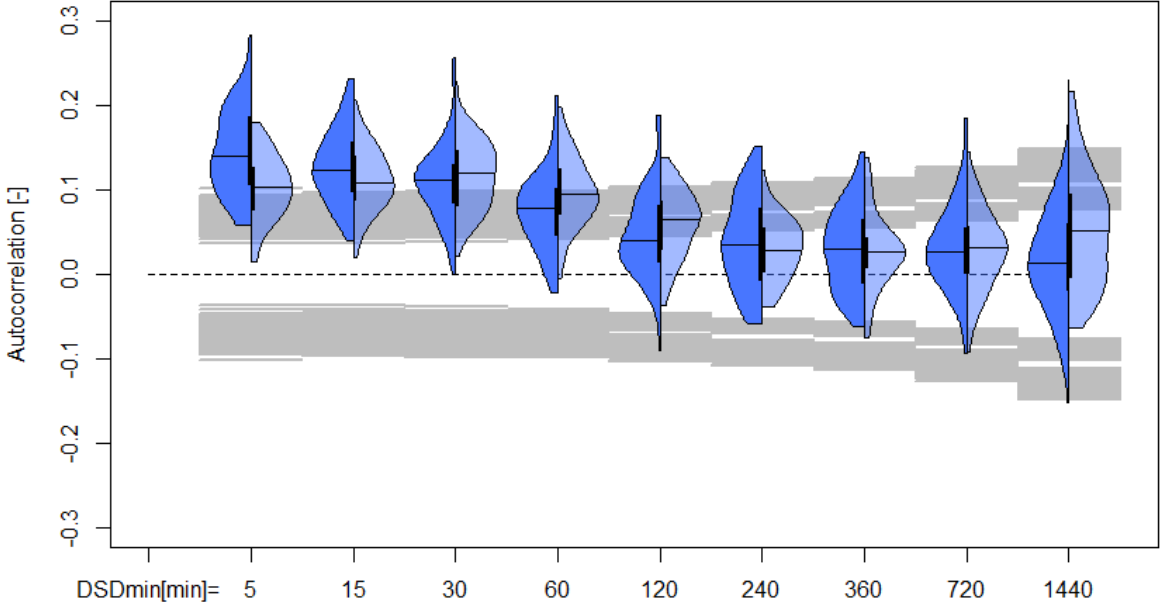
rainfall events is analyzed. Results are shown in Figure 4.4 in which the gray bands indicate the 95% confidence levels approximated by  $\pm 1.96/\sqrt{N}$  (N is the total number of observations and varies for each station and season). As can be seen for some of the cases the autocorrelations are significantly different from zero, especially for the WSD, for which 85% of the cases indicate an autocorrelation different from zero, whereas for DSD and WSA cases these percentages are only 24 and 15 respectively.



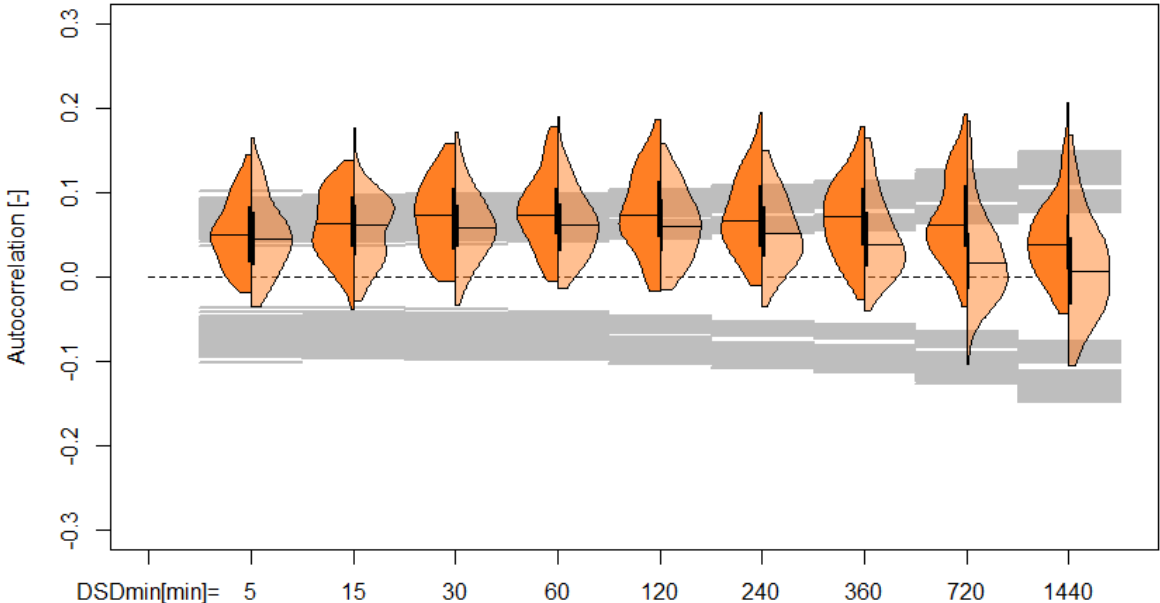
**Figure 4.4: Autocorrelation of variables describing rainfall events observed in NS and BAWU. Left semi-plots correspond to summer and right ones to winter events.**

An increase in the DSDmin is expected to lead to events that are independent and hence to decrease the autocorrelations. Therefore an analysis is performed in order to decide how this value affects the autocorrelation of the duration variables, i.e. WSD and DSD. The results show that this variable indeed has an effect on the autocorrelation results (see Figures 4.5 and 4.6). Note that as DSDmin increases the number of identified events decreases and consequently the confidence intervals (gray bands) change. As the DSDmin increases the autocorrelation of WSD variables decreases and the opposite is observed for the DSD (only up to 6 hours). If both variables are considered together, then a DSDmin of 360~720 minutes (6~12 hours) would deliver non-significant autocorrelations for most of the analyzed cases. These would mean that maximum 2 or 4 events per day would be identified. The aim of the Thesis is the development of a precipitation model for urban applications, and as urban catchments are characterized by very fast responses, the values of 6 or 12 hours between events do not seem to be consistent with the desired application. It is therefore adopted a DSDmin of 5 minutes with the drawback that some significant autocorrelation is present for some of the variables and stations. Gaál *et al.* (2014) discourage selecting long DSDmin as it results in negative effects on event properties by introducing extensive intra-event gaps and

leading to strong biases in WSD and WSI. Regarding this issue a further analysis is performed and presented in Sub-chapter 6.1 where a discussion about the effect of significant autocorrelation in urban applications is included.



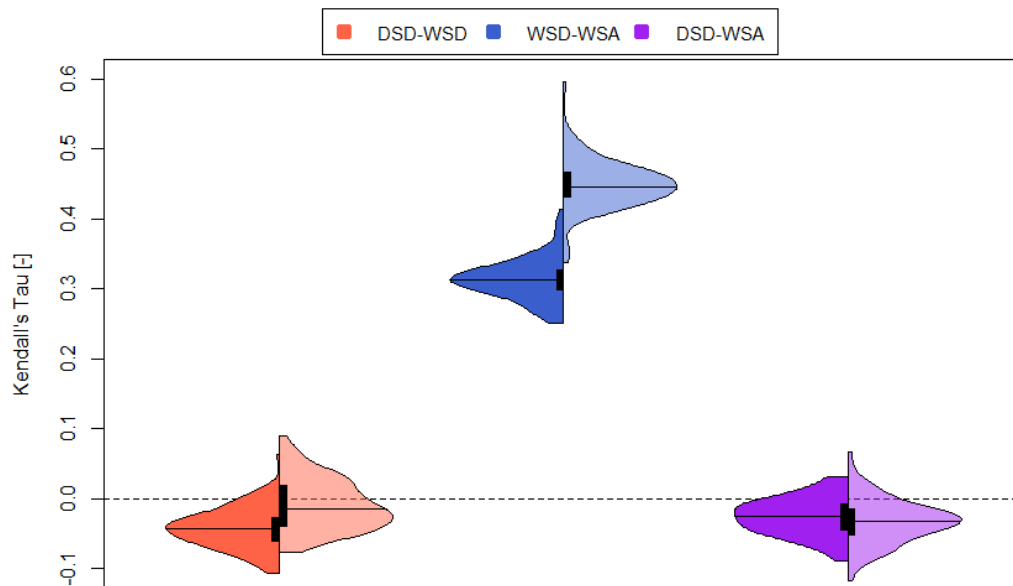
**Figure 4.5:** Autocorrelation of WSD resulting from events observed in NS associated to different values of DSDmin. Left semi-plots correspond to summer and right ones to winter events.



**Figure 4.6:** Autocorrelation of DSD resulting from events observed in NS associated to different values of DSDmin. Left semi-plots correspond to summer and right ones to winter events.

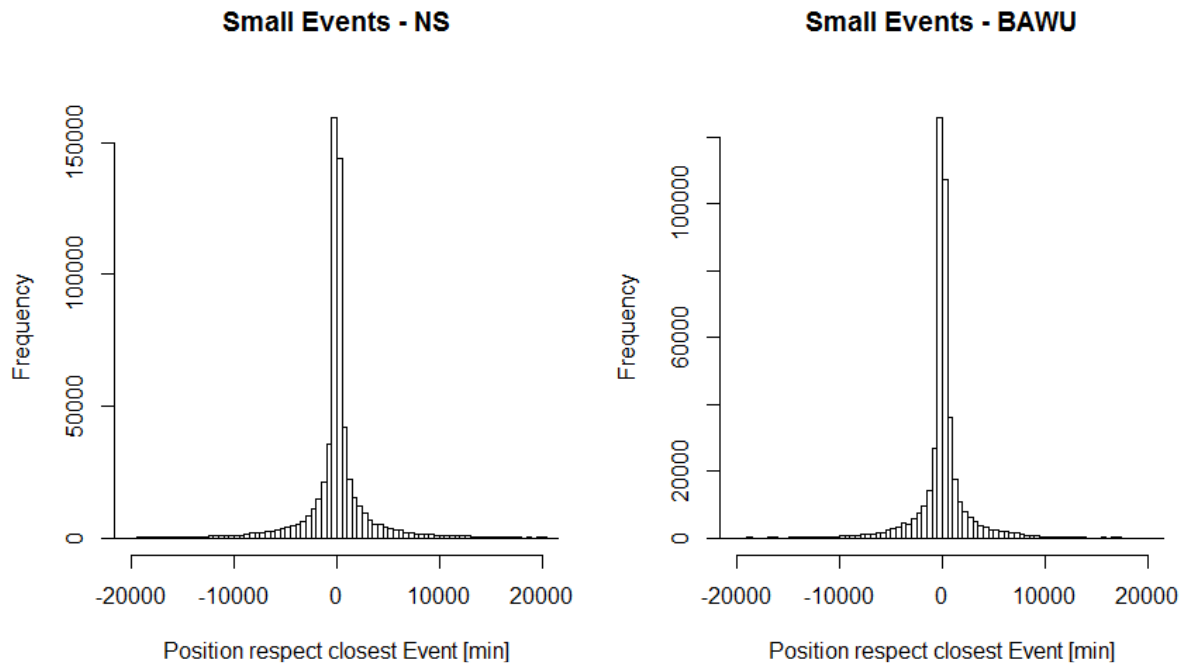
Correlation between different variables describing the rainfall events is as well analyzed. The Kendall's Tau correlation coefficient (see Eq. 3) is used for this purpose, since it is rank based and therefore useful information for the copula models. The analysis is performed based on pairs of variables, i.e. DSD-WSD, WSD-WSA and WSA-DSD. Results show that the WSD-WSA correlations are for all cases positive and as well significantly different from

zero. As can be seen in Figure 4.7 Winter events show to have higher correlations compared to summer ones (median values of 0.45 and 0.31, respectively). For the other compared pairs only a percentage of 32% (DSD-WSD) and 26% (DSD-WSA) from all stations show to be significantly different from zero, so for these pairs of variables no correlation is considered within the further analysis. Regarding the WSD-WSA correlations, differences between two correlations are evaluated with a test proposed by Cohen *et al.* (2003), resulting in more than 95% of the stations with significantly different correlations for summer and winter events. Furthermore it was mentioned that convective events are more frequent in summer and stratiform in winter (see Eggert *et al.*, 2015) so it is reasonable to model the pairs corresponding to the two seasons separately.



**Figure 4.7:** Correlation coefficients between different characteristics describing rainfall events observed in NS and BAWU. Left semi-plots correspond to summer and right ones to winter events.

Small events with  $WSA < WS_{Amin}$  are as well studied. These events account for between 19% and up to 37% (median 27%) of the total volume of annual rainfall and should therefore not be neglected. Small events registered during winter account for more volume for all stations, with a ratio of winter over summer that ranges from 1.1 up to 1.9 with median value of 1.6. The following plots (see Figure 4.8) show the distribution of occurrence (temporal location) of small events with respect to the occurrence of events, i.e. events with an amount equal or bigger than  $WS_{Amin}$  which are modeled by the ARP. Half of these events occur within a window of 7~8 hours around the modeled ones and the most frequent temporal locations are of only 10 minutes for small events observed before and of 70~100 minutes for events observed after. This information is relevant for the inclusion of small events into the model, especially as it is the aim to use it for continuous simulation of an urban system.



**Figure 4.8: Temporal location of small events with respect to events for NS (left plot) and BAWU (right plot).**

## 4.2 STATIONS FOR SINGLE SITE MODEL

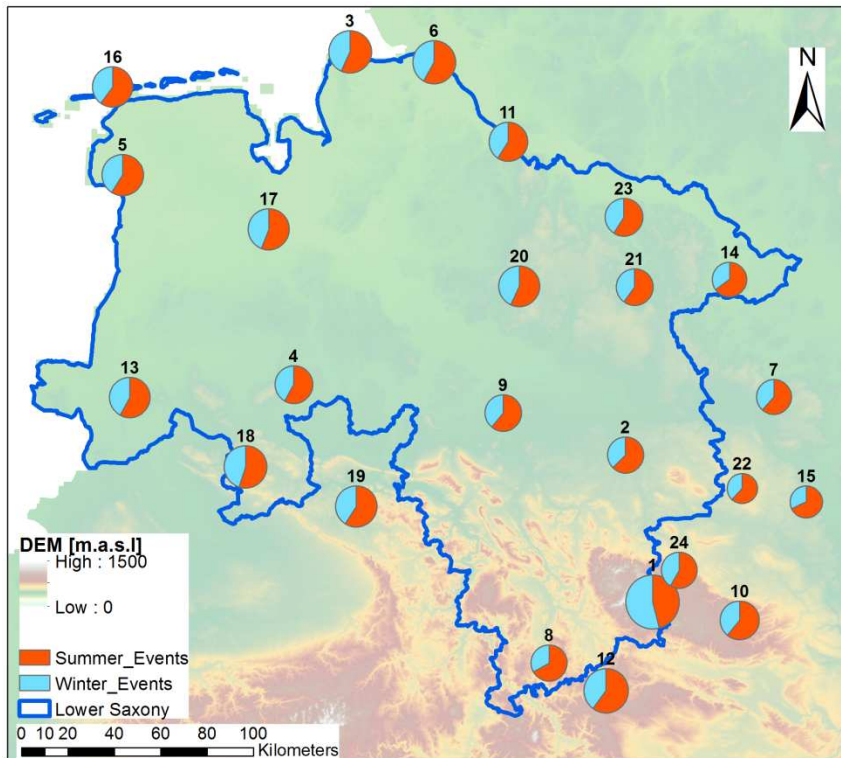
A total of 24 stations located in NS and surroundings are used to set up and evaluate the single site model and some basic information can be found in the Table 4.1. The final data set includes stations with at least 9 years of record lengths for a period finalizing in years 2012-13. In this case around 65% of the total registers correspond to the gravimetric/drop counter and the rest to tipping bucket. The mean annual precipitation ranges from 440 mm/year to 1380 mm/year, with highest values corresponding to high elevations, but lowest ones not directly related with the height of the station. Stations with lowest values are located in the mid-east region of the study area. For most stations, except for Braunlage (N° 1), summer events attribute to more than half of the total amount of rainfall, as shown in the pies for each station included in Figure 4.9. The stations are numbered after alphabetical order and the numbering of the table corresponds to the one in Figures 4.9 and 4.10.

Consistency of the data is checked based on daily values to identify possible discrepancies. Violinplots showing the distribution of daily rainfall for each of the stations are presented in Figure 4.10. From the violinplots some days with extremely high rainfall compared with the rest of the days are identified and labeled as Identified Days Summer/Winter. The identified days are compared with daily data provided by DWD for the corresponding station. For station N° 18 (Osnabrück) an outlier is detected for the summer days, this value is as well present in the DWD daily data base. Station N° 12 (Leinefelde) indicates two outliers one for summer and one for winter sample. The daily data base does not show a register of the

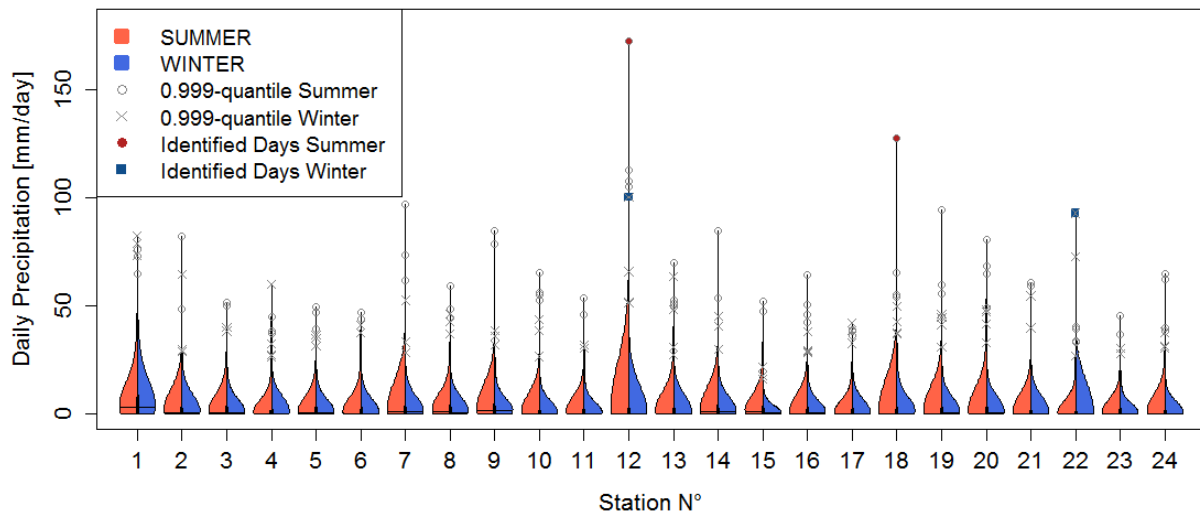
summer extreme, whereas the winter case is registered but indicates a total volume of half of the magnitude. For station N° 22 (Ummendorf) one winter day is found to have very high amount of rainfall, value which is marked as missing in the DWD daily data base. All these cases are checked using the original 1 minute data base and as no indicators of strange issues are detected they are not eliminated.

**Table 4.1: Attributes of rainfall stations used for developing the single site model.**

N°	ID	NAME	POSITION			DATA	Mean Annual Precip. [mm/y]
			Latitude [°]	Longitude [°]	Elevation [mNN]	Length [y]	
1	3984	BRAUNLAGE	51.73	10.6	607	21	1382
2	54251_10348	BRAUNSCHWEIG	52.3	10.45	81	15	633
3	31121_10131	CUXHAVEN (WEWA)	53.87	8.7	5	13	864
4	56329_10321	DIEPHOLZ(WEWA)	52.58	8.35	39	16	686
5	60820_10200	EMDEN (WEWA)	53.38	7.23	0	15	824
6	E082	FREIBURG/ELBE	53.83	9.25	2	10	888
7	3169	GARDELEGEN	52.51	11.4	47	21	581
8	10444	GÖTTINGEN	51.5	9.95	167	20	631
9	1538	HANNOVER	52.47	9.68	55	21	640
10	3193	HARZGERODE	51.65	11.14	404	21	726
11	E188	JORK-MOORENDE	53.52	9.73	1	10	721
12	3400	LEINEFELDE	51.39	10.3	356	21	941
13	1132	LINGEN	52.52	7.31	22	21	788
14	10253	LÜCHOW	52.97	11.14	17	20	569
15	3177	MAGDEBURG	52.1	11.58	76	21	496
16	32126_10113	NORDERNEY	53.72	7.15	11	19	767
17	10215	OLDENBURG	53.18	8.18	11	15	808
18	1516	OSNABRÜCK	52.26	8.05	95	21	869
19	1525	SALZUFLEN, BAD	52.11	8.75	135	21	824
20	10235	SOLTAU	52.96	9.79	76	21	803
21	E475	UELZEN	52.95	10.53	50	9	643
22	3173	UMMENDORF	52.16	11.18	162	21	440
23	E298	WENDISCH EVERN	53.22	10.47	62	10	686
24	3180	WERNIGERODE	51.85	10.77	234	21	616



**Figure 4.9:** Location of rain gauge stations used for developing the single site model (see numbering in Table 4.1). Size of circles is proportional to mean annual precipitation.



**Figure 4.10:** Distribution of daily rainfall for stations used for developing the single site model (see numbering in Table 4.1).

The transferability of the point model is performed based on 22 station located in the southwest of Germany, and some details are presented in Table 4.2. The stations are selected according to the length and period with data availability, to be consistent with the NS stations. These stations are located in higher altitudes and therefore have higher annual precipitation amounts compared with the previous set of stations.

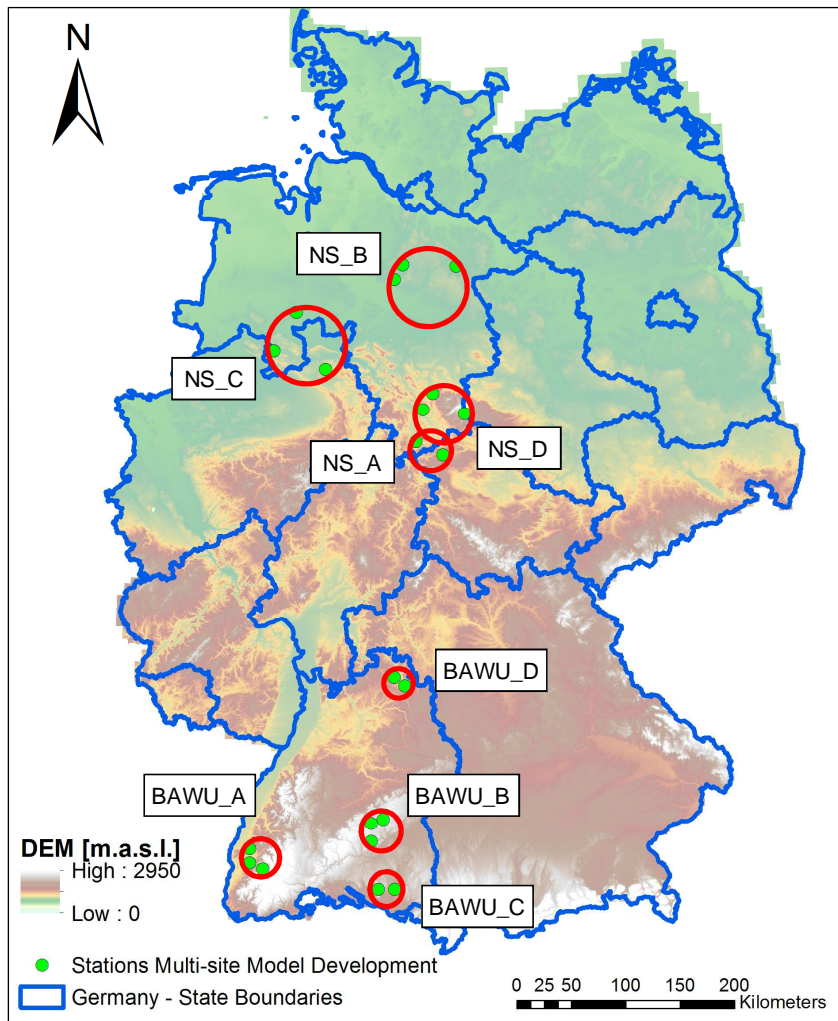
**Table 4.2: Attributes of rainfall stations used for transferring the single site model.**

N°	ID	NAME	POSITION			DATA	Mean Annual Precip. [mm/y]
			Latitude [°]	Longitude [°]	Elevation [mNN]	Length [y]	
1	71525	ABTSGMUEND-U.	48.92	9.92	432	10	810
2	71155	BALTMANNSSWEILE	48.77	9.45	457	16	917
3	90284	BERKHEIM	48.05	10.08	570	16	947
4	70323	BUCHENBACH	47.97	8	445	11	1114
5	70153	DEGGENHAUSERTA	47.8	9.42	708	16	1022
6	71605	ELLWANGEN-R.	48.98	10.13	460	11	824
7	70304	ELZACH-F.	48.2	8.12	440	10	1316
8	70314	EMMENDINGEN-M.	48.13	7.83	201	12	819
9	70354	FREIBURG	48.04	7.82	236	19	916
10	71058	HECHINGEN	48.38	8.98	522	16	860
11	90164	HOHENSTEIN-B.	48.35	9.33	740	16	933
12	71615	KIRCHBERG/JAGST-	49.18	9.98	426	16	800
13	71005	KOENIGSFELD/S.	48.15	8.43	730	10	1005
14	90156	LANGENENSLINGE	48.2	9.33	777	16	831
15	73942	LAUDA-K.-H.	49.55	9.63	324	16	745
16	73930	MERGENTHEIM,BAD	49.48	9.77	250	16	707
17	90163	MUENSINGEN-A.	48.38	9.48	750	16	965
18	71064	ROTTENBURG-K.	48.47	8.97	360	10	696
19	70173	STOCKACH	47.87	9.02	532	10	823
20	90307	ULM	48.38	9.95	567	20	723
21	70145	WEINGARTEN	47.8	9.62	440	16	912
22	71573	WUESTENROT-O.	49.13	9.5	392	16	853



### 4.3 STATIONS FOR MULTI-SITE MODEL

A total of 8 groups of stations located in the north (NS\_A, NS\_B, NS\_C and NS\_D) and in the south (BAWU\_A, BAWU\_B, BAWU\_C and BAWU\_D) are selected to develop and test the proposed model. The stations are selected based on the length of data available for a common period within each group. The groups of stations consist of either 2 or 3 stations located at distances ranging from 11 up to 59 km. Some basic information from the stations as well as their location can be found in Figure 4.11 and Table 4.3.



**Figure 4.11: Location of rain gauge stations and groups used for developing the multi-site model.**

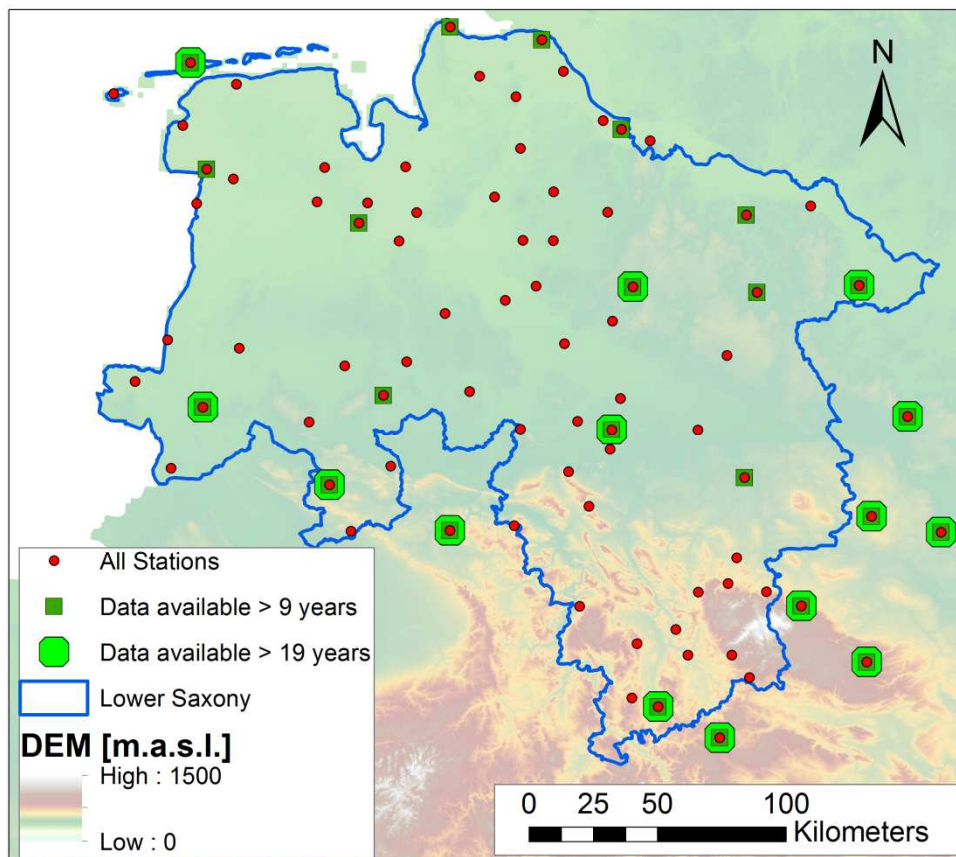
As can be seen in Table 4.3, the available time series with rainfall registered in all stations for each case study range from 8 up to 19 years. The case studies corresponding to NS indicate stations with longer distances between each other, and therefore bigger areas involved for the evaluation based on areal precipitation.

**Table 4.3: Attributes of rainfall stations used for multisite synthesis.**

Stat.	ID	Name	Latitude [°]	Longitude [°]	Elevation [mNN]	Data Availability [y]	Distance [km]	
							A	B
<b>BAWU_A</b>								
A	70354	FREIBURG	48.04	7.82	236	10	-	-
B	70314	EMMENDINGEN-MUNDINGEN	48.13	7.83	201	10	13	-
C	70323	BUCHENBACH	47.97	8	445	10	13.6	22.3
<b>BAWU_B</b>								
A	90164	HOHENSTEIN-BERNLOCH	48.35	9.33	740	11	-	-
B	90163	MUENSINGEN-APFELSTETTEN	48.38	9.48	750	11	11.7	-
C	90156	LANGENENSLINGEN-ITTENHAUSEN	48.2	9.33	777	11	16.7	23.2
<b>BAWU_C</b>								
A	70145	WEINGARTEN	47.8	9.62	708	13	-	-
B	70153	DEGGENHAUSERTAL-AZENWEILER	47.8	9.42	440	13	15	-
<b>BAWU_D</b>								
A	73930	MERGENTHEIM,BAD-NEUNKIRCHEN	49.48	9.77	250	14	-	-
B	73942	LAUDA-KOENIGSHOFEN-HECKFELD	49.55	9.63	324	14	12.2	-
<b>NS_A</b>								
A	2925	LEINEFELDE	51.39	10.3	356	19	-	-
B	1691	GOETTINGEN	51.5	9.95	167	19	27.2	-
<b>NS_B</b>								
A	4745	SOLTAU	52.96	9.79	75.6	8	-	-
B	1336	FALLINGBOSTEL,_BAD	52.85	9.68	70	8	15.5	-
C	5146	UELZEN	52.95	10.53	50	8	49.1	58.1
<b>NS_C</b>								
A	3815	OSNABRÜCK	52.26	8.05	95.4	14	-	-
B	4371	SALZUFLEN, BAD	52.11	8.75	134.6	14	50.7	-
C	963	DIEPHOLZ(WEWA)	52.58	8.35	39	14	41	59
<b>NS_D</b>								
A	656	BRAUNLAGE	51.73	10.6	607	8	-	-
B	3650	NORTHEIM-IMBSHAUSEN	51.77	10.05	212	8	38.2	-
C	4651	SEESSEN	51.9	10.18	186	8	34.4	17.5

## 4.4 STATIONS FOR REGIONALIZATION

The regionalization method is set up and validated based on 81 rainfall stations. As mentioned the available data is provided by the DWD and corresponds to stations belonging to the new network (automated weather stations). The location of the stations is shown in the following figure (see Figure 4.12) and the ones with registers longer than 9 and 19 years are marked with green squares. The latter 23 and 14 stations respectively, are used for evaluation of the regionalization methodology in terms of extreme events (based on the external structure) and for assessment regarding the IDFs.



**Figure 4.12: Location of rain gauge stations used for regionalization of the model.**

In the following plots (see Figure 4.13) some characteristics of the time series observed in the 81 locations are presented. The first plot shows the length of the available time series. It can be seen that most of the stations have a total observed period of 8 years, some have only 6 years and the maximum is 20 years. The elevation plot, along with the map, shows that most of the data was collected in the flat region, with topographic heights lower than 100 mNN. The distance to shore line shows a quite uniform distribution that goes from 0 up to 250 km. Finally the mean annual precipitation values indicate that most of the stations have a total rainfall of around 730 mm/year, and the range goes from 450 (stations located on the east) up to 1100 (one station located on the south west of the highest peaks of Harz mountains) mm/year.

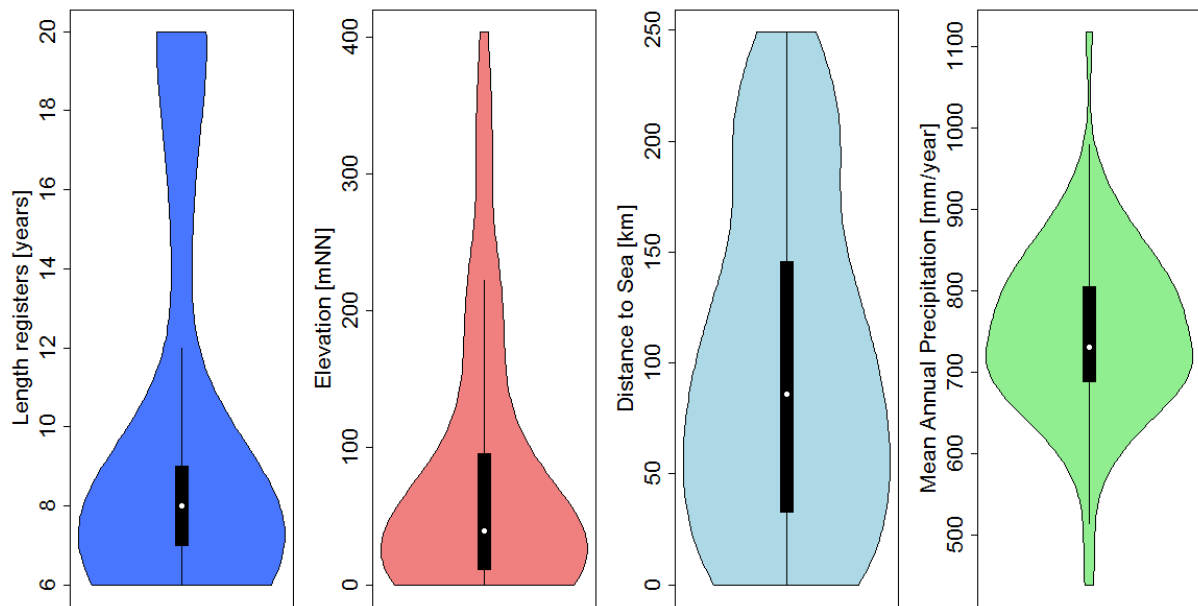


Figure 4.13: Main characteristics of time series used for regionalization of the model.

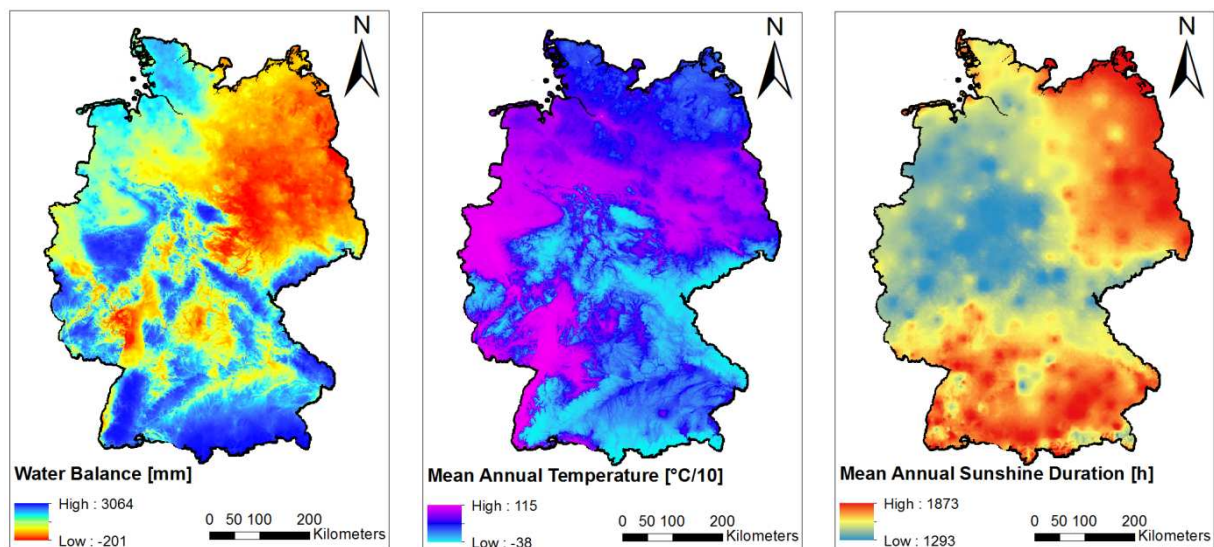
## 4.5 SITE DESCRIPTORS (SDs)

Site descriptors include different types of information available for all stations used in the regionalization procedure. This information is not only available for the location with stations but also for any other location, therefore it is possible to apply the regionalization for any point. SDs include general descriptors like location (latitude, longitude), distance to the shore and elevation. Several climatological and hydrological parameters, provided by the German Weather Service (DWD, 2015) on a raster basis for whole Germany and freely accessible online are as well included in the set of descriptors. Among the information the following statistical description of the climate system are included, which are based on the period of time 1981 - 2010 unless a different period is indicated:

- mean rainfall: yearly, summer and winter
- min/mean/maximum temperature: yearly, summer and winter
- mean number of days per year with a precipitation amount exceeding 10/20/30 mm
- mean sunshine duration: yearly, summer and winter
- mean solar radiation
- water balance: mean difference between precipitation and potential evapotranspiration (1971 - 2000)
- mean start day of the vegetation period (1992 - 2015)
- mean number of days classified as: Summer, hot, snow, frost, ice

Additionally, rainfall information available in a raster basis for whole Germany and on a daily temporal resolution (REGNIE) for 1981-2010 which is provided by the German Weather Service is used. From this data base the mean number of rainy days (defined by four different thresholds: 0, 1, 2 and 3 mm) is extracted for the summer, winter and year.

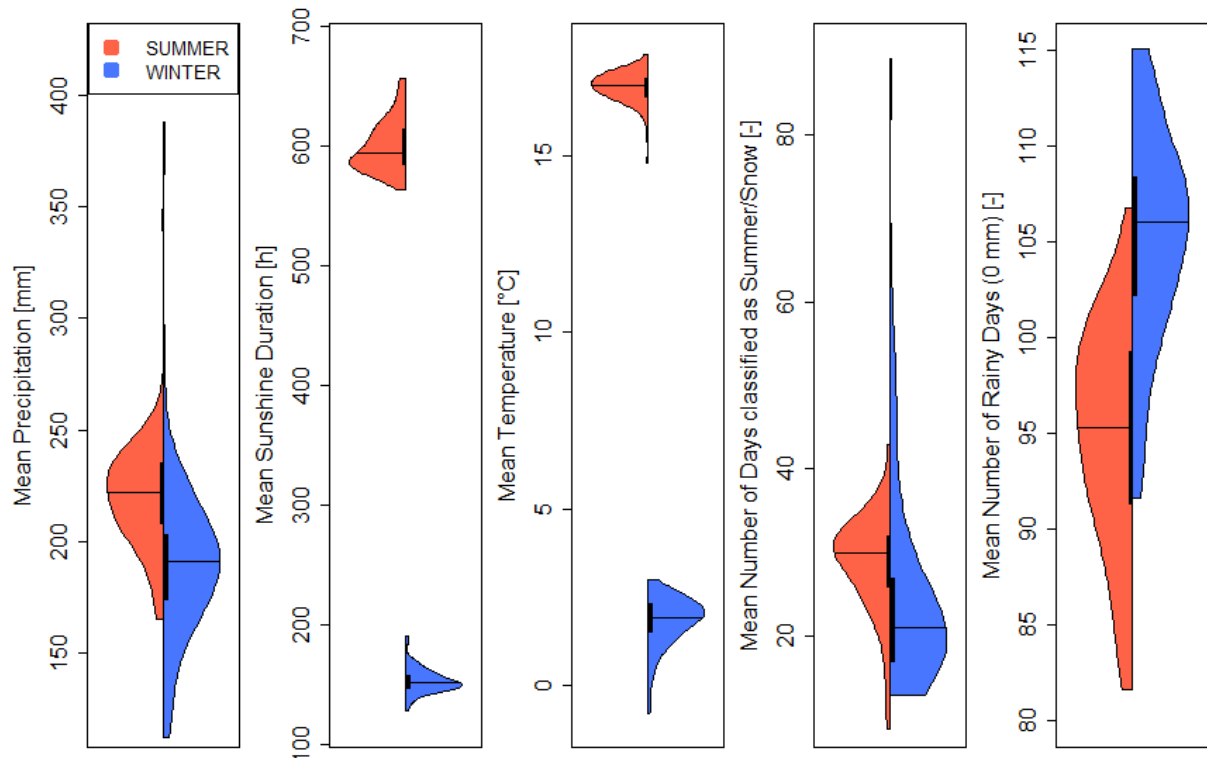
The raster data bases provided by the German Weather Service consist of maps with grids for the whole country (for some examples see Figure 4.14). These maps were developed by the provider based on different regionalization techniques. Simple and multiple linear regressions were used for regionalizing the temperature based on elevation (Müller-Westermeier, 1995) and the start of the vegetation period based on elevation along with latitude and longitude (DWD, 2016). Solar radiation was regionalized with the aid of satellite data (Möser & Raschke, 1984) which was corrected with measurements at the earth surface. The REGNIE data was regionalized in two steps procedure. First a multiple linear regression was applied based on elevation, latitude, longitude, direction and absolute value of exposition (effect of inflow of air masses) to estimate background fields. Precipitation was scaled by the values of background field and these dimensionless variables were interpolated to all raster locations and later rescaled by estimated fields. For further details the reader can refer to Rauthe *et al.* (2013).



**Figure 4.14: Example of grid maps with site descriptors used for regionalization of the model.**

A total of 26 and 27 SDs are available for summer and winter, respectively. Most of the SDs are included in the regionalization of events registered both in summer and winter, however some are only included for the summer cases (average number of days per year classified as summer and hot, i.e. with more than 25°C and 30° C) whereas others only for the winter (mean number of days per year classified as snow, frost and ice, i.e. with more than 50% of area covered with snow, and minimum/maximum daily temperature below 0°C). The temperature values correspond to a height of 2 meters above ground. The summer and winter variables provided by DWD describe the calendar period, i.e. June-July-August and

December-January-February, and differs from the summer and winter definition used in this Thesis. Only for the data derived from the REGINE data set, the same period is used. The distribution for some of the SDs are presented in Figure 4.15, it can be seen that some of them are symmetrically distributed whereas others present a skewed behavior. All SD used in this Thesis are listed in Appendix F.



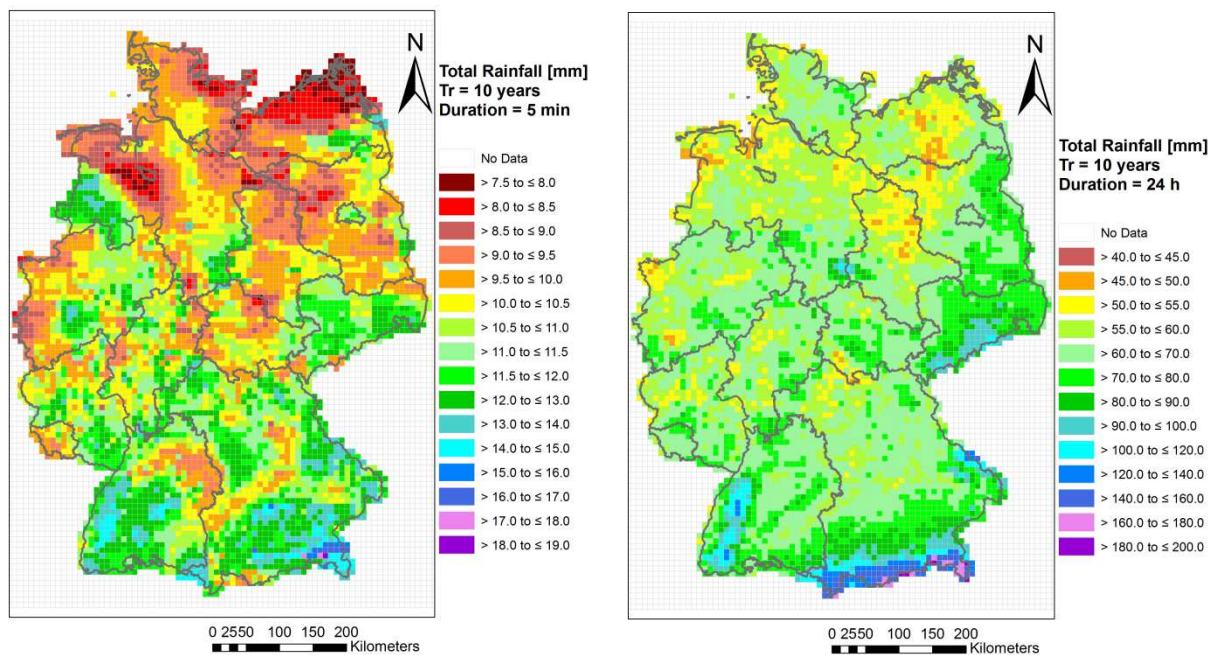
**Figure 4.15: Distribution of some site descriptors for all stations used for regionalization of the model.**

## 4.6 DESIGN PRACTICE: IDFS KOSTRA

KOSTRA DWD-2000 consists of an atlas which provides IDF curves on a raster for all Germany with cell sizes of 8.45 kmx8.45 km (Bartels *et al.*, 2005). These curves consist of regionalized values of extreme rainfall amounts associated to different durations and return periods. The curves are constructed based on statistical frequency analysis of annual maxima series registered by recording (digitized pluviograph records) and non-recording stations with long registers from the old German network, and thereafter regionalized for points without registers. The automated weather stations belonging to the new network that are used for developing and evaluating the different methods presented in this Thesis were not involved in the construction of the KOSTRA curves. These IDFs are included in the analysis since they are commonly used by engineers for dimensioning sewer networks, rainwater retention basins, infiltration systems, etc. in the study region and are therefore



used as a comparison criterion for the assessing the performance of the different models. As an example of information provided by KOSTRA, the gridded maps of Germany for two different durations and one return period are presented in Figure 4.16. For each grid a value of total rainfall is provided as extreme value for the selected duration and return period. Similar maps are available for return periods ranging from 0.5 up to 100 years and for durations starting from 5 minutes and up to 3 days.



**Figure 4.16: Extreme values provided by KOSTRA for a return period of 10 years and a duration of 5 minutes (left) and 24 hours (right).**

As shown in the example KOSTRA provides extreme values of rainfall amounts for different durations and return periods. The total amounts are then distributed within the durations, i.e. disaggregated into 5 minutes intervals, using an Euler type 2 model in which the peak intensity is located at one third of the total duration of the event (for details see DWA, 2006). KOSTRA provides tolerance ranges for different return periods which are used for accounting uncertainty of precipitation. The uncertainties are of  $\pm 10$ , 15 and 20% for return periods of up to 5, 50 and 100 years.

For the evaluation of extreme events by multi-site modeling based on areal precipitation, the average over rasters included in the area of interest must be calculated. As KOSTRA was developed on a point basis, the application of these values on an areal basis requires some reduction factors. Verworn (2008) provides reduction factors of KOSTRA values based on the area and duration which are used for reducing the averages. In this Thesis the final values are denominated as KOSTRA-Reduced. The areas included in the mentioned publication reach a maximum of 1000 km<sup>2</sup>; therefore reduction factors must be extrapolated for larger areas. This extrapolation is performed by simple linear regressions for each of the durations which are shown as gray dashed lines in Figure 4.17. For events with durations of

5 minutes the Reduction Factor\* values corresponding to 15 minutes are taken, since this is the minimum available duration.

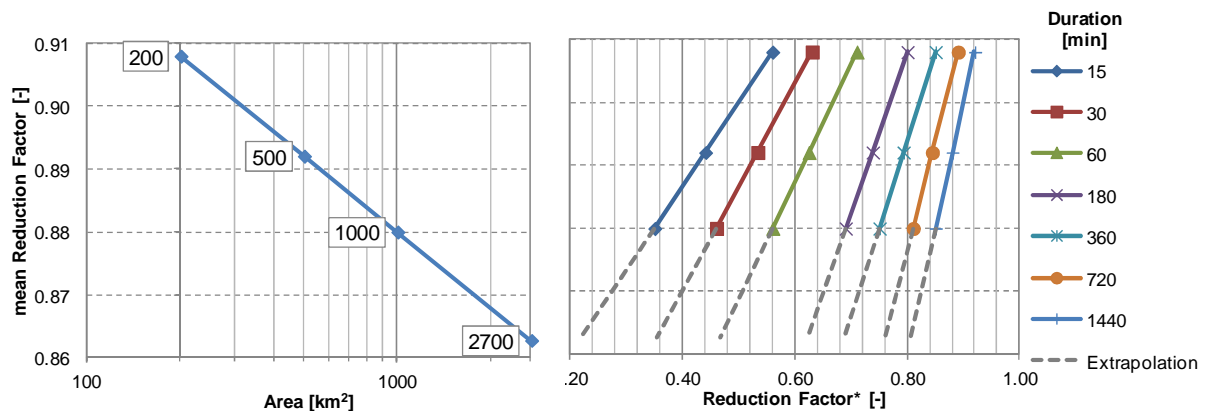


Figure 4.17: KOSTRA reduction factors for different sizes of areas and durations of events.

## 4.7 ARTIFICIAL URBAN SYSTEM

To evaluate the rainfall model in terms of runoff statistics a fictional urban system is set up. EPA Storm Water Management Model (SWMM) is applied for the modeling (see Rossman, 2010), which is a dynamic hydrology-hydraulic simulation model used primarily for urban applications. Several authors have used this model for urban studies (see e.g. Wu *et al.* 2013, Meierdiercks *et al.* 2010 and Hsu *et al.* 2000). The processes included in this work are rainfall-runoff (runoff generation and concentration) and flow routing (for details see Table 4.4). Precipitation, evaporation and infiltration are modeled at a sub-basin scale which results in a runoff load that is transported within the system through pipes, channels, storage/regulator devices and pumps. The infiltration is modeled following the Soil Conservation Service one-parameter curve number method (SCS method), for details refer to Maidment (1993). A curve number of 65 is used which describes residential areas. For event based modeling curve numbers of 45 and 82 are as well included to account for antecedent dry and wet conditions (see U.S. Soil Conservation Service, 1985). Monthly averages of evaporation for the region are given as input to the SWMM.

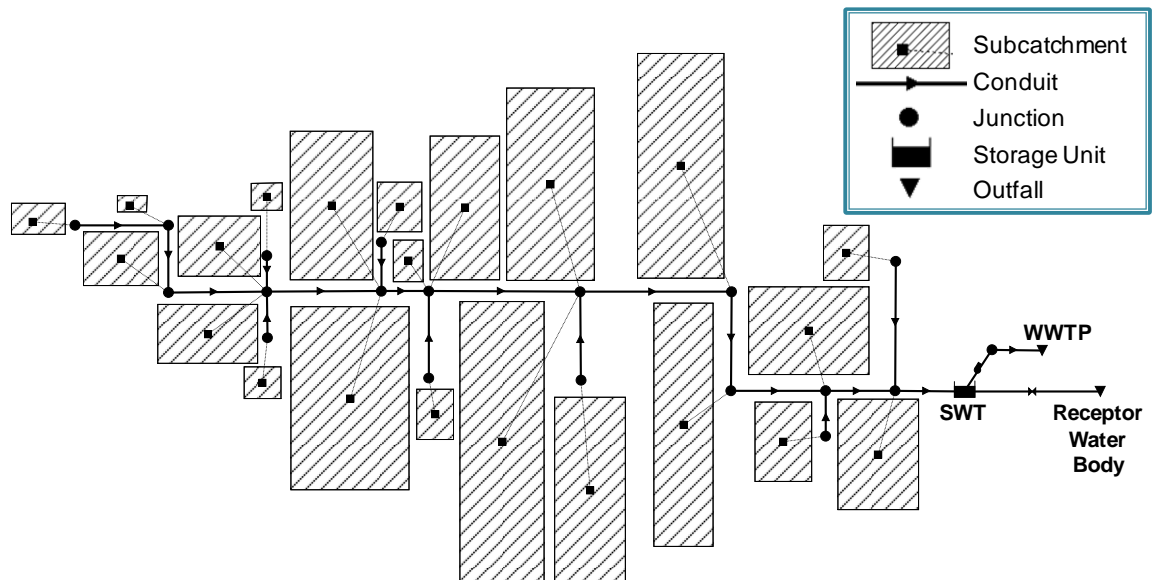
The total surface of the fictional system is of 168 Ha with the geometrical configuration shown in Figure 4.18 which consists of a combined sewer system, i.e. storm and waste water are carried together by one pipe system to a Waste Water Treatment Plant (WWTP). There is a storm-water tank (SWT) connected to an overflow structure ahead the WWTP, to account for big volume storm events which would not be treated but rather spilled directly to a receptor water body. As the sewer system is combined a dry weather flow (DWF) must be specified which is estimated for each sub-catchment according to the population and is



constantly been discharged to the system and pumped to the WWTP. Two different sizes of SWT are considered according to German standard practices (Imhoff & Imhoff, 2007), T20 and T40 which have specific volumes of 20 m<sup>3</sup>/ha and 40 m<sup>3</sup>/ha of impervious area respectively.

**Table 4.4: Basic information describing the model and fictional urban system.**

	Rainfall-runoff	Flow routing
<b>Model: Methods</b>	<b>Infiltration:</b> SCS (Curve Number: 65) <b>Subarea Routing:</b> Runoff from impervious area flows to pervious area <b>Interflow with groundwater, snow water accumulation/melting effects, quality analysis:</b> Neglected	<b>Routing:</b> Dynamic Wave Flow Routine <b>Friction loss:</b> Normal conditions: Manning's Under pressure conditions: Hazen-Williams' <b>Surface Ponding:</b> allowed <b>Definition of Supercritical flow:</b> Slope and Froude Number <b>Energy loss entrance/exit:</b> Neglected
	Sub-catchments	Pipes
<b>Urban System: Parameters</b>	<b>Total areas:</b> 1.1 to 16 ha <b>Impervious sub-areas (65%)</b> Manning's coefficient: 0.014 Depression storage depth: 2 mm <b>Pervious sub-areas (35%)</b> Manning's coefficient: 0.035 Depression storage depth: 3.5 mm <b>Surface slope:</b> 0.25% <b>Potential evapo-transpiration:</b> mean monthly values for the region <b>DWF: Discharge:</b> 0.13 m <sup>3</sup> /day per inhabitant <b>Population density:</b> 45 inhabitants/ha	<b>Manning coefficient:</b> 0.014 <b>Shape:</b> circular <b>Diameters:</b> 0.7 to 1.1 meters <b>Length:</b> 100 to 500 meters <b>Average slope:</b> 5 ‰



**Figure 4.18: Schematic representation fictional urban system.**

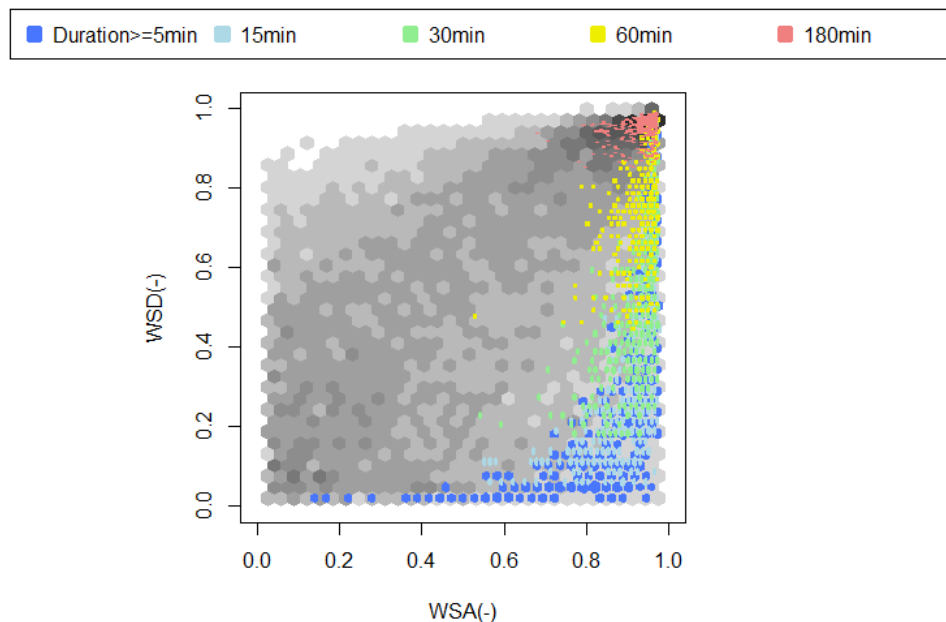
## 5 RESULTS

In this chapter the results of the different precipitation models proposed within this Thesis are presented. A description of selected methods and probability distributions for the particular region of study is included along with a more detailed explanation of some of the methodologies involved based on real case studies. Different results are presented which involve various criteria for evaluating precipitation time series. The proposed models are in all cases compared with alternative ones and/or with the design practice for the study region. First in sub-Chapter 5.1, some general results which are valid for all precipitation models, i.e. for single site, multiple sites and regionalization, are described. Thereafter (sub-Chapter 5.2), the comparison of the precipitation models for single sites is presented and discussed. Different parameters involved in the multi-site model are included in the following sub-chapter 5.3 along with results comparing the different methods involved for this purpose. A description of the final regionalization model is then presented (sub-Chapter 5.4), along with validation of the proposed method and comparison with alternative approaches. Finally in sub-Chapter 5.5 the results of the uncertainty analysis are included with a discussion involving all event characteristics along with extreme values for the different methods used for single site synthesis and regionalization.

### 5.1 GENERAL RESULTS

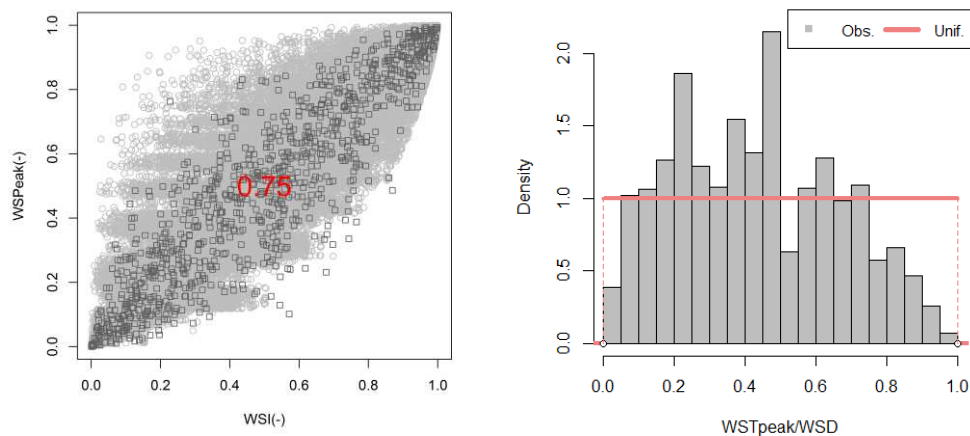
Rainfall events are estimated for the 24 stations with more than 9 years of data located in the state of Lower Saxony (NS) and surroundings. As was mentioned in the Methods Chapter, several distribution functions are tested for modeling the different variables describing rainfall events. The final selected distributions of the model are listed in Table 5.1. Figure 5.1 shows the pairs of pseudo-values of WSA-WSD from events registered in all stations (events are first normalized and then assembled, pseudo-values are calculated for the entire region). The figure indicates a positive correlation between these two variables and different tail dependencies with more concentrated pairs in the upper tail. The annual extreme events associated to different durations are as well indicated. It can be seen that the location of these extreme events is concentrated to the right side of the graph, corresponding to high values of WSA, along the whole data set of WSD, as different durations are considered. Important to note are the extreme values corresponding to short durations, which are located in the lower-right corner of the graph. This plot suggests that pseudo-pairs of WSA-WSD are non-exchangeable, since many of the extreme events are located in specific regions which

present the asymmetric feature. The role of the modeling pairs of values WSA-WSD is of special interest, due to the implication regarding the extreme events. More details regarding this particular component of the models are presented and discussed in the following sub-chapters for the single and multi-site precipitation models.



**Figure 5.1: Pairs of pseudo-values of WSA and WSD for all events (gray) and annual extreme events (colors according to durations) recorded in all stations.**

The analysis of variables involved within the internal structure indicates that pairs of pseudo-values WSI-WSPeak have a strong positive correlation and similar tail dependency for the upper and lower tails and are therefore well reproduced by a Normal copula (see left image in Figure 5.2). This copula is used to estimate WSPeak conditioned on WSI (which results from the external structure). The uniform distribution used for modeling the ratio WSTpeak/WSD is presented in the following figure (see right image in Figure 5.2) along with the histogram of ratios from observed events. It can be seen that the uniform model is a simplification of reality, as the observed events are more concentrated for values lower than 0.5, nevertheless this simple model is applied for the rest of the work.



**Figure 5.2:** Left: Pairs of pseudo WSI and WSPeak for all events observed in all stations (grey circles) and simulated with the copula (dark grey squares), Kendall's Tau correlation coefficient is shown. Right: Histogram of WSTpeak/WSD extracted from observed events with the Uniform model used for their simulation

In the Data Chapter (see last section of Sub-chapter 4.1) it was mentioned that small events account for 19% to 37% of total annual rainfall for all stations. All small events are grouped into a regional data set, which is then used for their synthesis, as a random sampling of these events until the total seasonal contribution is fulfilled. Small events are allocated within dry spells, considering the location from events with  $WSA \geq 1$  mm using histograms based on observations (see Figure 4.8).

The different criteria and components of the final model are listed and described in Table 5.1.

**Table 5.1: Description of selected criteria and components involved in the single site, multi-site and regionalization of the model.**

	Variable	Model	Distribution	Equation	Parameters	Estimation Method
External Structure	DSD	Univariate	Kappa	$F(x) = \left(1 - h \left(1 - \frac{\kappa(x - \xi)}{\alpha}\right)^{1/\kappa}\right)^{1/h}$	$\xi$ : location $\alpha$ : scale $\kappa$ : shape 1 $h$ : shape 2	L-moments
	WSA	Joint distribution $C_\theta(u_{WSA}, v_{WSD})$	Kappa	(See DSD)		
	WSD	<b>Copula</b> (see sub-Chapters 5.2 and 5.3)	Log Normal 3	$F(x) = \Phi\left(\frac{\log(x - \xi) - \mu}{\sigma}\right)$	$\xi$ : lower bounds (RS) $\mu$ : mean (NLS) $\sigma$ : standard deviation (NLS)	L-moments
Internal Structure	WSI*	Joint conditional distribution	Kappa	(See DSD)		
	WSPeak	$C_\theta(u_{WSPeak}, v_{WSI})$ <b>Normal Copula</b> (see Eq. 6)	Generalized Normal	$F(x) = \Phi\left(-\kappa^{-1} \log\left(1 - \frac{\kappa(x - \xi)}{\alpha}\right)\right)$	$\xi$ : location $\alpha$ : scale $\kappa$ : shape	L-moments
	WSTpeak	Univariate	Uniform	$F(x) = \begin{cases} 0, & x < a \\ \frac{x - a}{b - a}, & a \leq x \leq b \\ 1, & x > b \end{cases}$	$a$ : 0 $b$ : 1	-
	Profile model	Univariate	Exponential	$i(t) = \text{WSPeak} \cdot \exp(c\lambda(t - \text{WSTpeak}))$ $\begin{cases} c = +1, t < \text{WSTpeak} \\ c = -1, t \geq \text{WSTpeak} \end{cases}$	$\lambda$ : exponent	on-line for each events
Small events	<b>Variable</b>		<b>Description</b>			
	WSA and WSD		Regional set of pairs			
	Total Amount		Estimated for each station and season			
	Position within DSD		Probabilities according to regional observations			
Event Definition	<b>Variable</b>	<b>Values</b>	<b>Unit</b>			
	WSA <sub>min</sub>	1.0	mm/event			
	WSI <sub>min</sub>	0.01	mm/5min			
	DSD <sub>min</sub>	5	min			

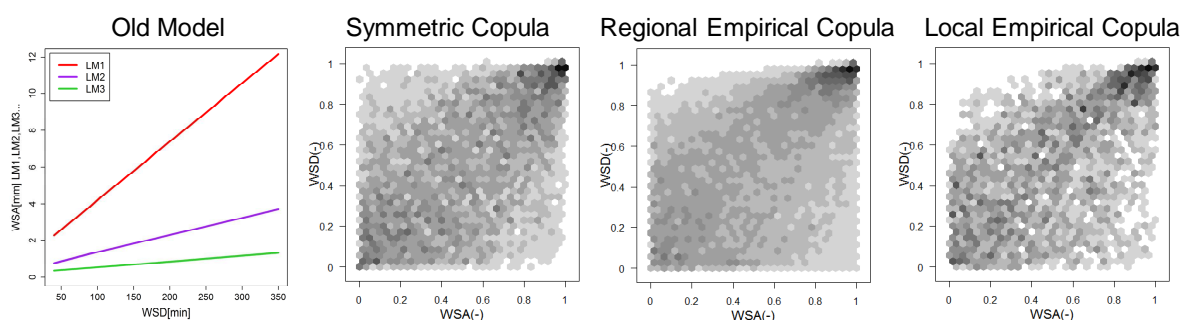
\*WSI is estimated from the external structure (WSA/WSD). Kappa is the marginal distribution for Copula application.  
 $u$  and  $v$ : marginal distributions describing the variables  
 RS and NLS: Real and Natural Logarithmic space  
 $\Phi$ : Standard Normal distribution

## 5.2 SINGLE SITE PRECIPITATION MODEL

Different characteristics are evaluated to identify the best precipitation model for generating long synthetic series of rainfall for single sites. The role of joint modeling of the WSA-WSD pairs in generating synthetic events is analyzed. For this purpose 3 different bivariate probability models are included in the analysis and compared with the “Old Model”. The first model is based on a “Symmetric copula”, often used in hydrologic applications. The estimation and choice of the copula is based on events observed in all stations (normalized and then assembled). The Hüsler-Reiss copula is chosen here (see Eq. 8) which, as was mentioned in the Methods chapter, is able to mimic pairs with a more concentrated upper tail.

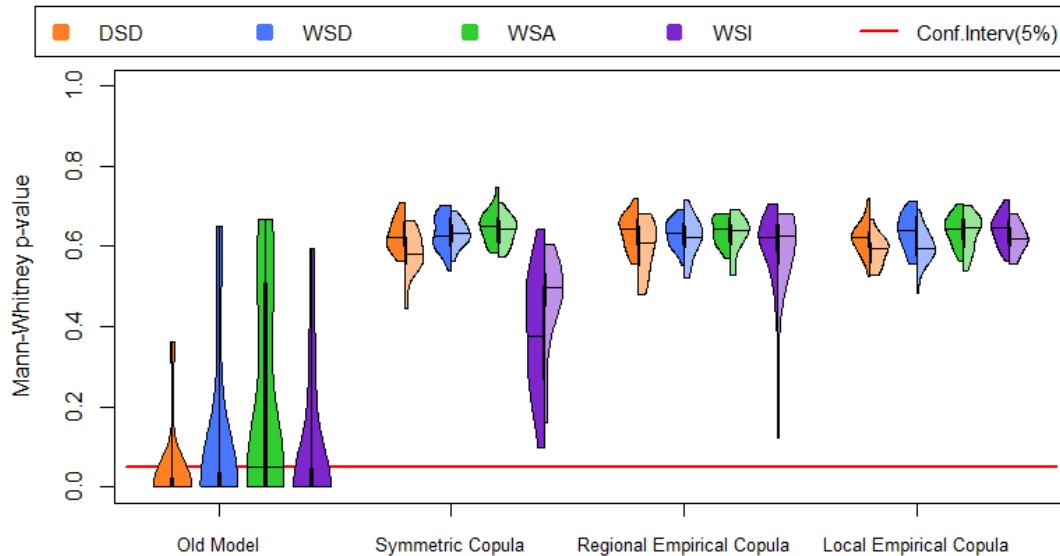
Symmetric copulas fitted to each station separately are as well analyzed; results are similar to the ones using a regional symmetric copula and are therefore not included here.

For the second and third cases, the asymmetric behaviour of the pairs of variables is taken into account. As mentioned in the Methods section, several theoretical asymmetric models are considered; however none of them could deliver satisfactory results. For this reason empirical copulas are used to model the pairs of variables (see Eq. 2). The second model is an “Empirical Regional Copula” based on all stations (same data set of pseudo pairs used for estimating the symmetric model). The “Local Empirical Copula” is the third proposed model which is based on the second one, but involves a local selection of pairs of pseudo observations according to each specific station. The empirical copulas are considered because of the asymmetric structure of the pseudo-observations, especially since the non-exchangeable region is where most of the extreme events are located (see Figure 5.1). Vandenberghe *et al.* (2010) found non-exchangeable behaviour among variables describing rainfall events, and they stress that extreme storm simulations are problematic even if asymmetric copulas are used. Schematic representations of the 4 compared cases, i.e. “Old Model” and 3 copula based alternatives, are shown in Figure 5.3. For the three models which involve a copula, the pairs of values are visualized using hexagon binning, i.e. bivariate histograms. As was mentioned the color ramp shows the proportion number of points falling within each hexagon, light-colored hexagons indicate less points, whereas dark indicates a higher concentration of pairs. It is important to emphasize that both the “Old Model” and the “Local Empirical Copula” are specific for each station, whereas the “Symmetric Copula” and the “Regional Empirical Copula” are general for the whole region. All copula models are estimated for summer and winter seasons separately.



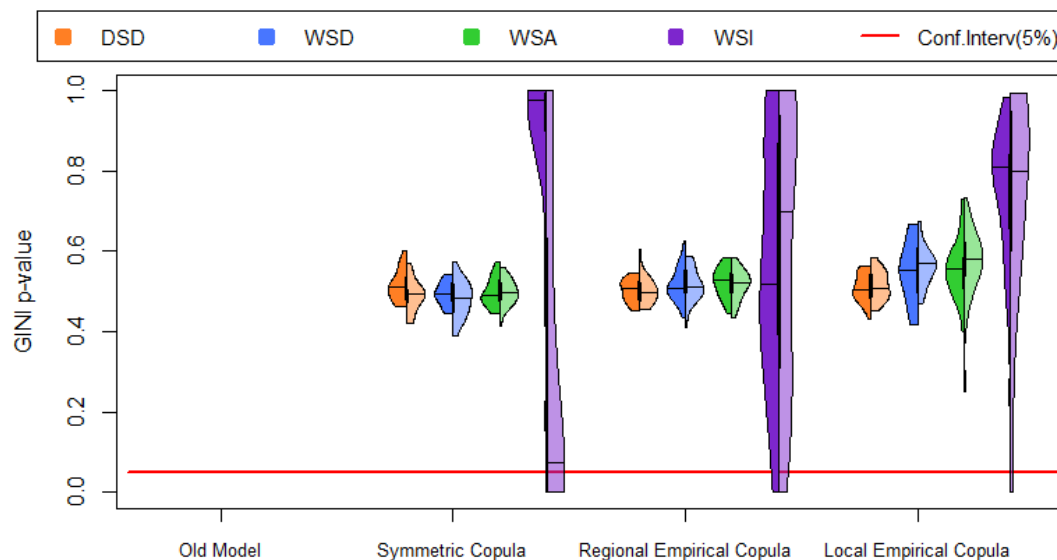
**Figure 5.3:** External Structure: Schematic representation of different bivariate models for WSA-WSD synthesis. Kendall’s Tau correlation coefficients are shown for the 3 copula models.

In the following figures the “Old Model” is compared with the new techniques to assess their potentials. The different models are used for validating the basic statistics using the Mann-Whitney (MW) test. The test is run 100 times based on every station and a median value is extracted for each of the variables. Results of p-values are presented for the 24 stations as violin plots, both for winter and summer and for the different event variables (see Figure 5.4). Note that values lower than 0.05 indicate that the observed distributions are statistically significant different from the simulated ones.



**Figure 5.4:** P-values resulting from applying the Mann-Whitney test to validate the single site model based on summer (left plots of violins) and winter (right plots of violins) events characteristics for all stations.

The “Old Model” does not consider seasonality therefore the violin plots are symmetric. Furthermore, p-values are lower than 0.05 for most of the stations and variables. For the three new models a significant improvement of performance can be seen. In all cases WSD and DSD are modeled directly, WSA is derived after WSD for the “Old Model” case and is directly modeled for the 3 copula based new models. WSI is derived from WSD and WSA; therefore the simultaneous generation of WSA-WSD plays an important role. The modeling of DSD, WSD and WSA with different copulas shows similar results for the three new models; only in case of WSI p-values resulting from the Symmetric copula are lower compared to the other two copula cases. Nevertheless the values are acceptable for all 3 copula cases and variables. Complementary to the MW test, the Gini test is applied to the cases for which the MW test shows acceptable results, i.e. the “Old Model” is excluded from this analysis. Results are shown in Figure 5.5 and indicate that for most of the variables the variance is also reproduced with a significance value of 5% and are shown in the following plot (only some for a few cases the WSI variance is not reproduced).



**Figure 5.5: P-values resulting from applying the GINI test to validate the single site model based on summer (left plots of violins) and winter (right plots of violins) events characteristics for all stations.**

Moreover the capability of the models to reproduce other moments is evaluated. Median of different order of moments is estimated for each station based on 100 simulations. Results are presented in Table 5.2 as *RSE* (see Eq. 18) for all stations and two of the models for comparison, i.e. “Old Model” and “Local Empirical Copula”. Errors increase as the order of the moment evaluated increases, as is expected. The “Local Empirical Copula” model outperforms the old one, especially for the WSI. The *RSE* values indicate a good performance of the copula model, especially for the first and second order moments, and acceptable results for the higher-order ones. The “Old Model” shows lower errors only for a few cases (marked with italic).

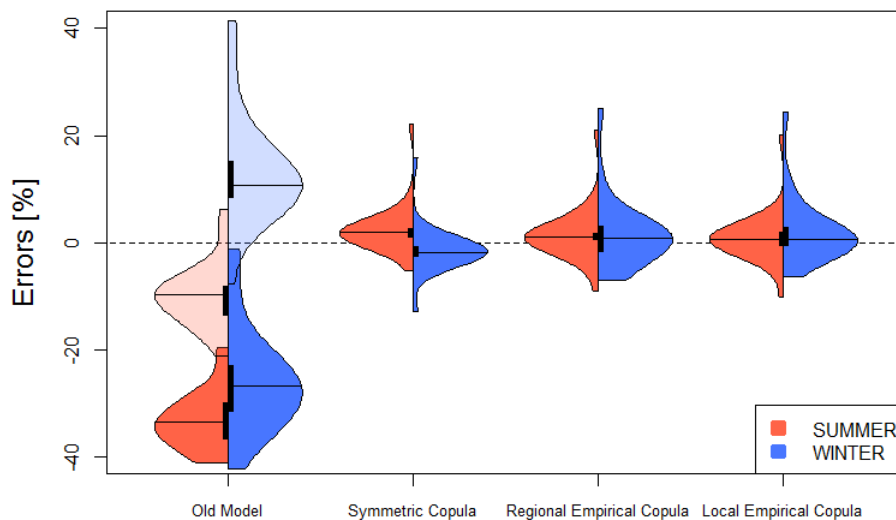
**Table 5.2: Evaluation of statistics of different variables describing the external structure of rainfall events resulting from two single site models and for all stations.**

<i>RSE</i> [-]		Mean		Standard Deviation		Skewness		Kurtosis	
Variable	Season	Old Model	Local Emp. Copula	Old Model	Local Emp. Copula	Old Model	Local Emp. Copula	Old Model	Local Emp. Copula
DSD	Summer	0.09	0.01	0.17	0.02	0.33	0.13	0.89	0.36
	Winter	0.09	0.01	0.10	0.05	<i>0.12</i>	<i>0.27</i>	<i>0.30</i>	<i>0.71</i>
WSD	Summer	0.21	0.01	0.21	0.06	0.23	0.23	0.55	0.54
	Winter	0.16	>0.01	0.07	0.03	0.35	0.16	0.76	0.37
WSA	Summer	0.11	0.01	0.08	0.05	0.35	0.32	0.90	0.88
	Winter	0.17	0.01	0.45	0.08	0.45	0.27	1.26	0.71
WSI	Summer	0.08	0.04	0.61	0.10	0.89	0.21	2.52	0.53
	Winter	0.98	0.05	3.04	0.19	0.82	0.26	1.90	0.53

When total seasonal rainfall values are compared (see Figure 5.6) the importance of including the small events in the model can be appreciated as total rainfall is underestimated

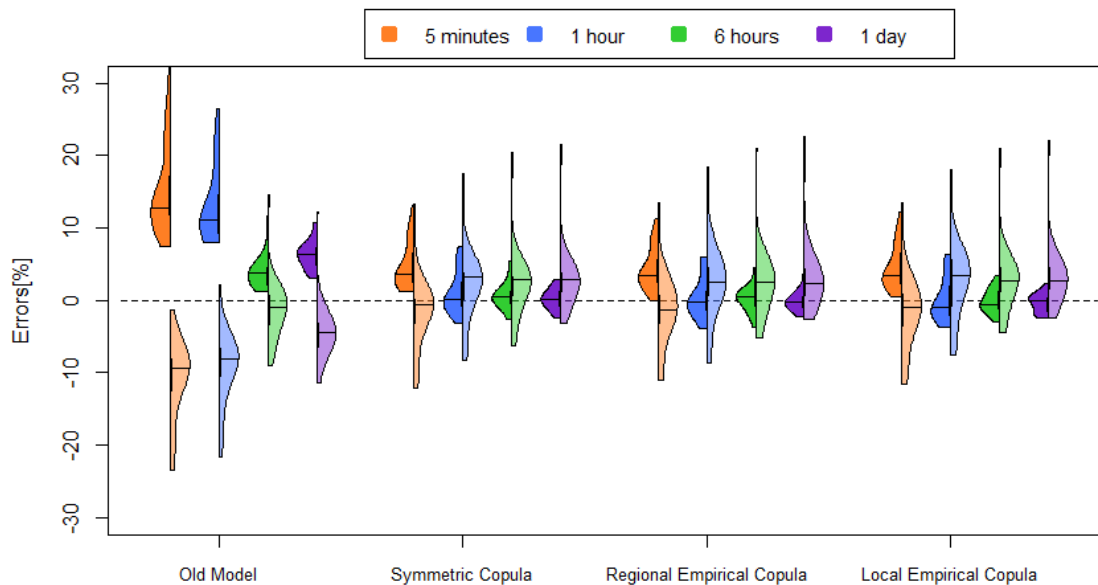


for the “Old Model” case. If small events are included in the “Old Model” (semi-transparent plots), results show an underestimation for summer and overestimation for winter, due to the effect of excluding seasonality. Overall results using the “Symmetric Copula” show a slight overestimation of summer values, and the opposite for winter. The regional and local Empirical Copulas deliver similar results with mean errors close to zero indicating overall satisfactory results.



**Figure 5.6: Validation of single site model based on errors of total seasonal rainfall for all stations. For the “Old Model” an additional variant is considered (semi-transparent plots) in which small events are included.**

The performance of the different models for different temporal resolutions, i.e. the scaling behavior, is shown in Figure 5.7 as errors of mean precipitation for wet time steps (left semi-violins) and fraction of wet time steps (right semi-violins). Range of errors of mean values decreases with increase of temporal aggregation for all models. The “Old Model” shows a systematic overestimation of mean values for all resolutions, whereas the fractions shows to be underestimated for low aggregations and is improved as it increases. No significant differences can be seen when the three models which include copulas are compared. These models show lower range of errors for the mean values compared to intermittency; nevertheless the errors are acceptable as they are lower than  $\pm 20\%$  for all temporal resolutions.



**Figure 5.7: Validation of single site model for different temporal resolutions based on errors of mean precipitation for wet time steps (left plots of violins) and fraction of wet time steps (right plots of violins) for all stations.**

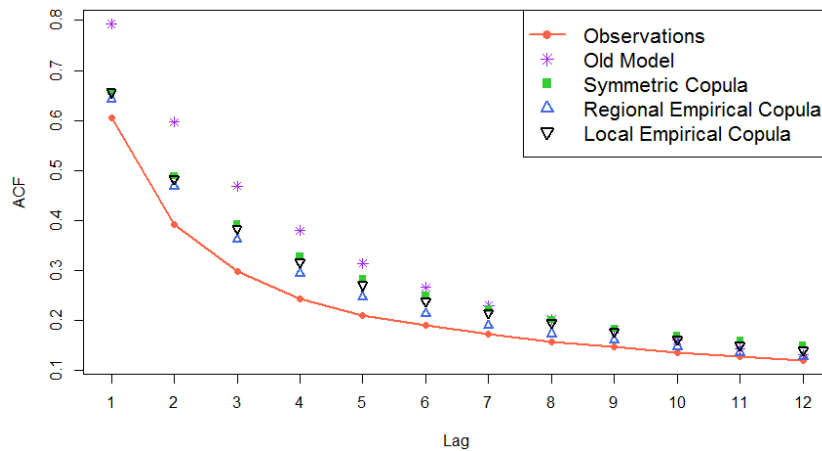
The performance of two of the models regarding other moments is additionally evaluated and presented as *RSE* for all stations in Table 5.3 (see Eq. 18). Errors increase as the order of the moment increases. The “Old Model” shows lower errors only for a few cases (marked with *italic*), otherwise the “Local Empirical Copula” model outperforms the old one.

**Table 5.3: Evaluation of performance based on wet time steps for different temporal resolutions resulting from two single site models and presented for all stations as *RSE*.**

<i>RSE</i>	Fraction		Mean		Standard Deviation		Skewness		Kurtosis	
	Old Model	Local Emp. Copula	Old Model	Local Emp. Copula	Old Model	Local Emp. Copula	Old Model	Local Emp. Copula	Old Model	Local Emp. Copula
<b>5 min.</b>	0.12	0.05	0.16	0.05	<i>0.08</i>	<i>0.14</i>	0.35	0.23	1.18	0.82
<b>1 hour</b>	0.11	0.06	0.14	0.03	0.22	0.07	0.52	0.37	2.61	1.47
<b>6 hours</b>	<i>0.05</i>	<i>0.06</i>	0.04	0.02	0.15	0.04	0.86	0.34	4.26	1.15
<b>1 day</b>	0.06	0.06	0.07	0.01	0.10	0.07	0.62	0.29	2.78	0.92

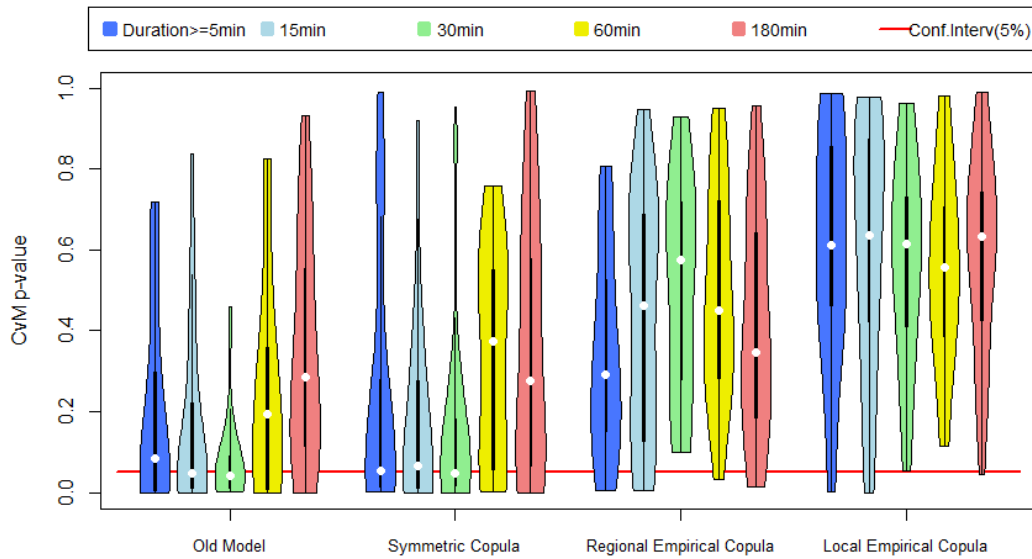
Observed and synthetic time series are used to estimate autocorrelations based on a 5 minute time step. Figure 5.8 shows overestimation by synthetic time series for all lags. The “Old Model” results in deviations that go up to 0.2, whereas for the copula models these are in all cases lower than 0.1. As was mentioned, this overestimation was expected due to the construction of the internal structure, nevertheless this temporal correlation shows to be improved with the new models. Further improvements could be explored by introducing a random noise to the final 5 minutes time steps with rainfall. Nevertheless the effect of this

noise on the other characteristics should be evaluated thus further research is required which goes beyond the scope of the present work.



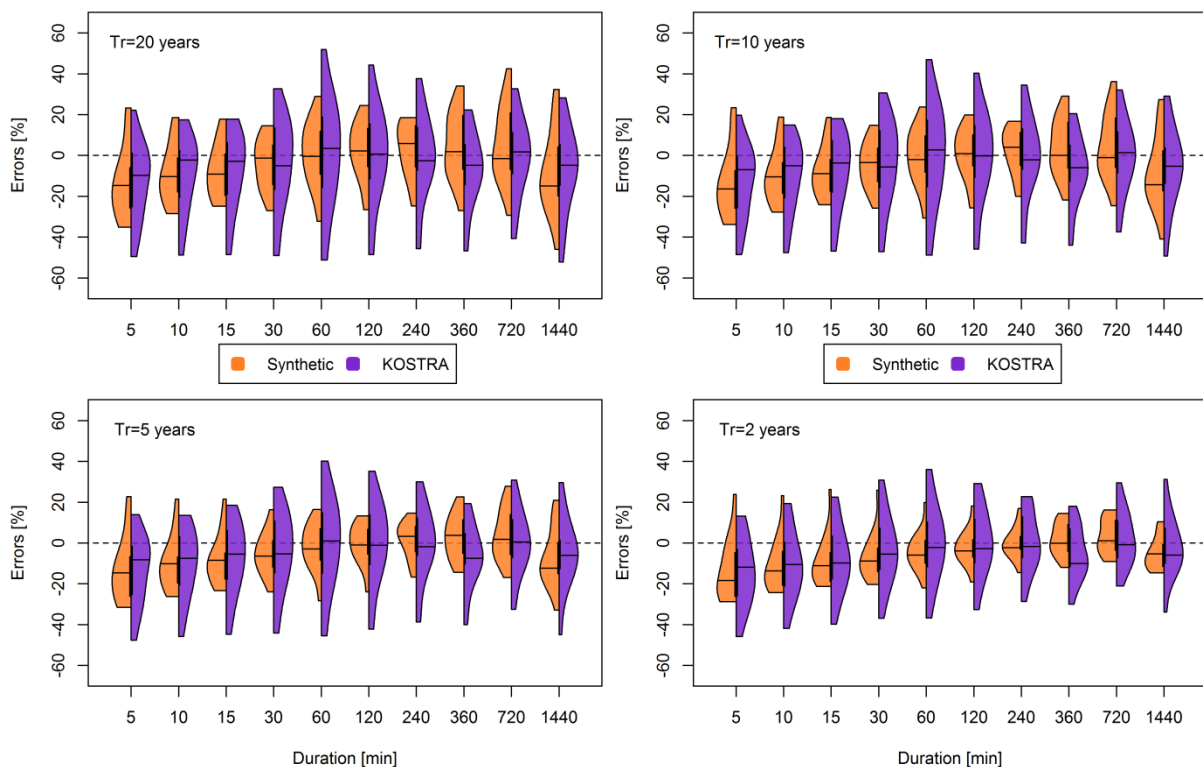
**Figure 5.8: Validation of precipitation model based on autocorrelation for 5 minute time steps and all stations (medians).**

Extreme values are compared for the different models using the CvM test. Results are shown for different durations and all stations as p-values in Figure 5.9. Extreme values are very sensitive to the WSA-WSD modeling. For all durations, results indicate an improvement as the copulas are introduced to the model. The asymmetric copulas show even better performance as the symmetric one, especially for short durations. In order to understand the reasons behind these results it is useful to have a closer look at the observed extreme events. Figure 5.1 shows that the location of the extreme events is concentrated to the right side of the graph, corresponding to high values of WSA, along the whole range of WSD. Important to notice are the extreme values corresponding to short durations, which are located in the lower-right corner of the graph. In case of a symmetric copula, the model fails to reproduce high (extreme) values of WSA corresponding to low durations, whereas the empirical copulas are capable of modeling pairs in this region. For the “Old Model”, short durations correspond to low values of L-moments (LMs) of WSD used for estimating WSA. Therefore the model is not able to deliver high values of WSA when the durations are short.



**Figure 5.9: P-values resulting from applying the Cramér-von Mises test to validate the precipitation model based on extreme values observed during the summer for all stations.**

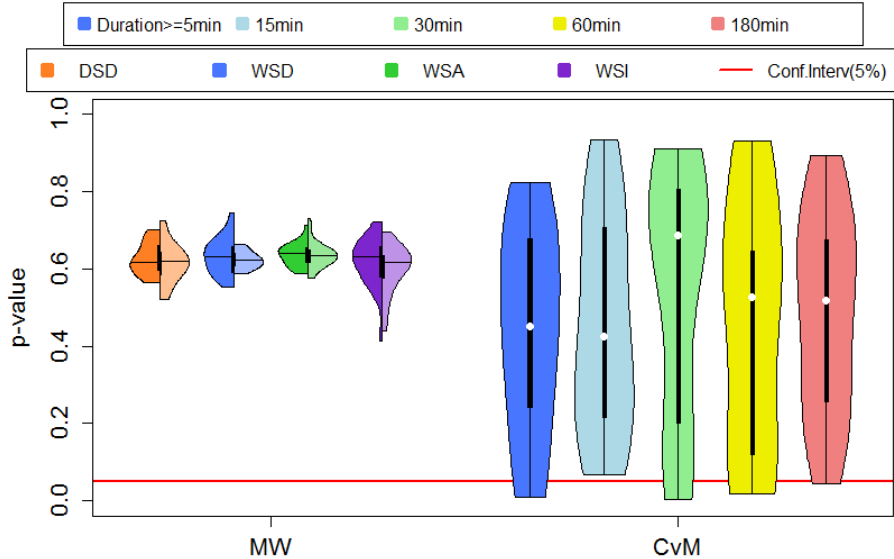
Intensity duration frequency curves (IDFs) derived from observations of 15 stations with registers longer than 19 years are compared with the ones resulting from synthetic time series and KOSTRA values. Synthetic time series are generated using the “Local Empirical Copula” due to the superior performance in estimating extreme events. Errors are presented in Figure 5.10 for four different return periods.



**Figure 5.10: Comparison of performance from precipitation model and KOSTRA to reproduce observed extremes for different return periods (20, 10, 5 and 2 years).**

The synthetic time series show on average underestimation of extreme events for durations up to 30~60 minutes and all return periods, these results improve for longer durations. On the other hand, KOSTRA shows on average errors closer to zero for most durations, however the ranges of errors are larger than the synthetic ones for all cases. As the return period decreases, the synthetic time series show smaller range of errors and therefore more robust results. Both data sets underestimate daily extreme values. Summarizing, synthetic time series only outperform KOSTRA design storms for durations higher than 60 minutes; however the time series appear to be more robust for all durations.

The transferability of the model to other regions is tested by validating the external structure of the model. For this purpose, 22 stations in Baden Württemberg with record lengths equal or longer than 10 years are used. The regional empirical copula is used for validating the model and is estimated based on summer and winter events in a similar way as for Lower Saxony (NS). The selected marginal models are the same probability distributions as for NS, except for WSA summer, which a Weibull distribution is preferred over the Kappa. Mann-Whitney and Cramér-von Mises tests are applied in a similar way but based on the new set of stations. The results are shown here in Figure 5.11. The Mann-Whitney test indicates that none of the distributions resulting from long simulations are statistically different to the observed ones for all stations and event characteristics. The CvM shows that for some stations extreme value simulations are not corresponding to observations; nevertheless the test shows results comparable to NS.



**Figure 5.11: Evaluation of transferability of the single site model: P-values resulting from applying the Mann-Whitney (Summer and winter events: left and right plots of violins) and Cramér-von Mises tests to 22 stations in Baden-Württemberg.**

### 5.3 MULTI-SITE PRECIPITATION MODEL

In this section different results related with multi-site synthesis are presented. As mentioned in the Method chapter, a vine-copula model is proposed in this Thesis and is thereafter compared with the Simulated Annealing (SA) method. Different parameters and results involved in the vine-copula model are discussed followed by an evaluation of the proposed model in comparison with SA.

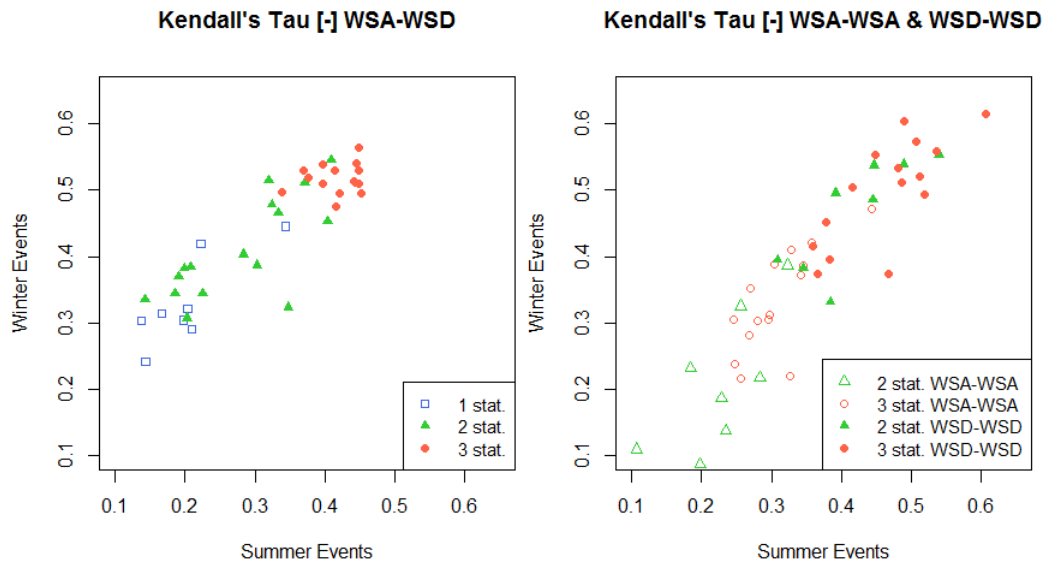
The first step involved in the proposed model is to identify events occurring in several stations simultaneously. The 8 case studies presented in the Data Chapter are used to identify events classified as isolated or as occurring in several stations. In Table 5.4 the percentage of events occurring in either 1, 2 or 3 stations simultaneously are presented in the first three lines, whereas the percentage of event in each of the stations are presented in the following 3 lines. As some of the cases only include 2 stations, the corresponding cells are marked with a “-“. As can be seen most of the events (50-78%) are registered in one station, i.e. they are isolated events. Another feature is that for the BAWU cases the percentage of events registered in 2 or 3 stations are higher than for the NS cases. This is due to the fact that the stations in BAWU are closer to each other with distances ranging from 12 to 23 km (see Table 4.3 in Data Chapter), whereas for the NS cases these distances go from 15 to 60 km. Additionally from Table 5.4 it can be seen that the percentage of events registered in each of the stations is close to balanced.

**Table 5.4: Percentages of observed events occurring in several stations simultaneously (1, 2, 3) and in each of the stations (A, B, C) for different seasons and case studies**

Number or ID Stations	BAWU_A		BAWU_B		BAWU_C		BAWU_D		NS_A		NS_B		NS_C		NS_D	
	Summer	Winter	Summer	Winter	Summer	Winter	Summer	Winter	Summer	Winter	Summer	Winter	Summer	Winter	Summer	Winter
1	61	57	58	50	65	57	63	59	77	73	68	58	78	68	68	63
2	20	21	21	23	35	43	37	41	23	27	21	23	15	21	19	21
3	19	21	22	27	-	-	-	-	-	-	11	20	6.6	11	13	16
A	56	59	64	76	71	77	74	74	73	83	56	70	47	61	65	85
B	58	54	68	75	79	81	72	78	57	49	55	66	50	54	47	42
C	70	81	63	58	-	-	-	-	-	-	44	54	42	47	52	48

The events classified as occurring in 1, 2 or 3 stations are used for estimating Kendall's Tau correlations (see Eq. 3) between different event characteristics. The resulting values for all case studies are presented in Figure 5.12. Correlation between WSA and WSD shows to be stronger for events occurring in winter compared to summer. This correlation is stronger for

events observed in several stations and weaker for isolated events. This supports the separation of events into seasons and subgroups for selecting the Vine copula models. The correlation between variables describing the WSA shows to be weaker compared to the one resulting from the WSD variables. These two cases do not show a clear difference between summer and winter events.



**Figure 5.12: Characteristics of events according to their occurrence in 1, 2 or 3 stations simultaneously for all cases. Left: Kendall's Tau correlation between WSA-WSD for each station. Right: Kendall's Tau correlation between WSA-WSA and WSD-WSD for pairs of stations.**

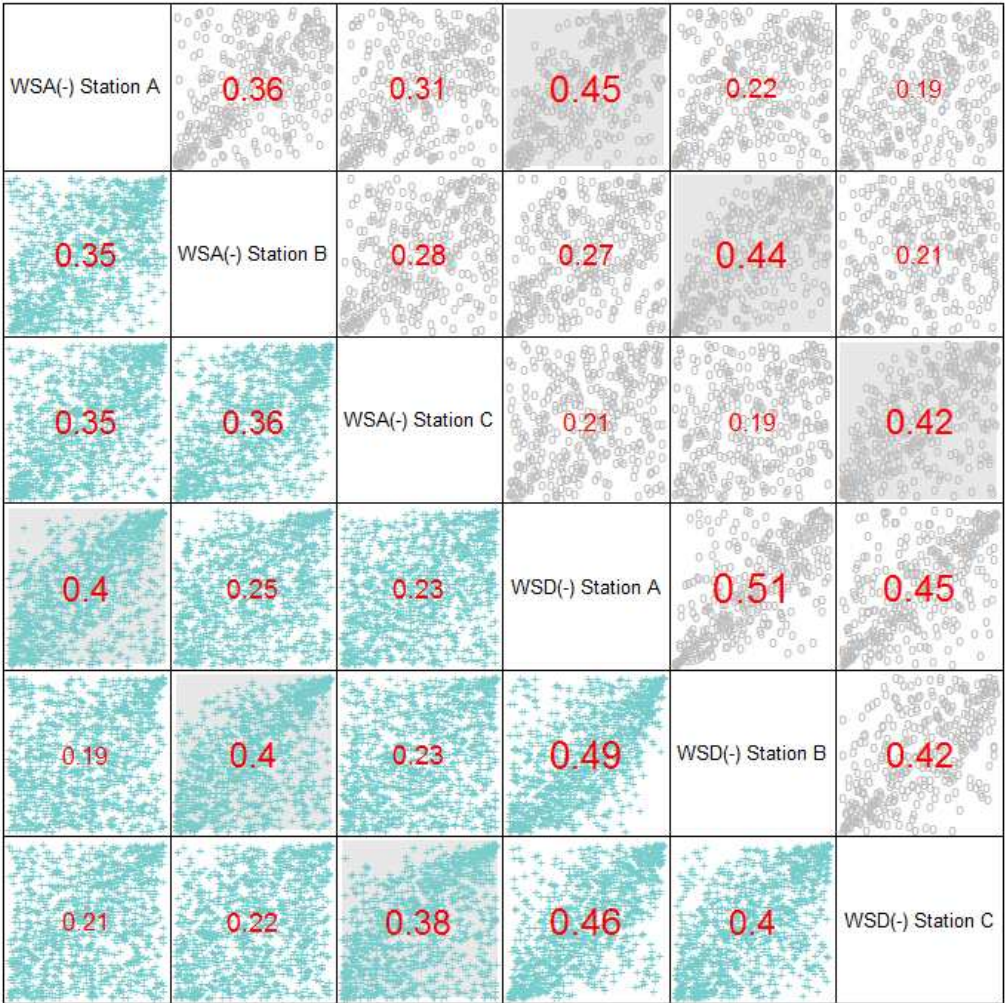
The following step of calculation involves the analysis of time series occurring in either 1, 2 or 3 stations for selecting the Vine-copula models. As explained in the Methods section, the vine copulas are used to mimic the correlations involving WSA and WSD for several stations. An empirical copula as the one used for the single site model is not possible for multi-sites due to the fact that the events are now subdivided into different groups and therefore the samples available for each copula model consist of less pairs.

For events occurring in only one station, the Vine copula consists of one tree with one bivariate model, as only two variables WSA and WSD are involved. A Tawn copula (see Eq. 9) is used for all cases. For events occurring in 2 and 3 stations simultaneously, four and six variables are involved, i.e. the WSA and WSD corresponding to each of the involved stations. The pre-defined tree structures for events occurring in 2 or 3 stations are used for the 8 case studies. These structures were presented in Figure 3.12 (Methods Chapter) and guarantee that pairs of WSA-WSD for each of the stations are directly modeled in the first step of simulation. Based on the pre-defined tree structures, different pair copula families and parameters are selected and estimated for each study case and season separately in a sequential way. The parameters are estimated using the Maximum Likelihood method. Many different families are chosen for the different structures of the tree. The bivariate copulas selected to model the WSA-WSD are in more than half of the cases the Tawn copulas,



followed by Clayton. For the WSD-WSD, the t copula was selected for more than 40% of the cases, followed by the Gumbel, Frank and Frank-Joe. For the rest of the cases, i.e. excluding the first tree, the conditional models include different copulas with Independent, Gaussian, Gumbel and Frank accounting for around 70% of the selected families.

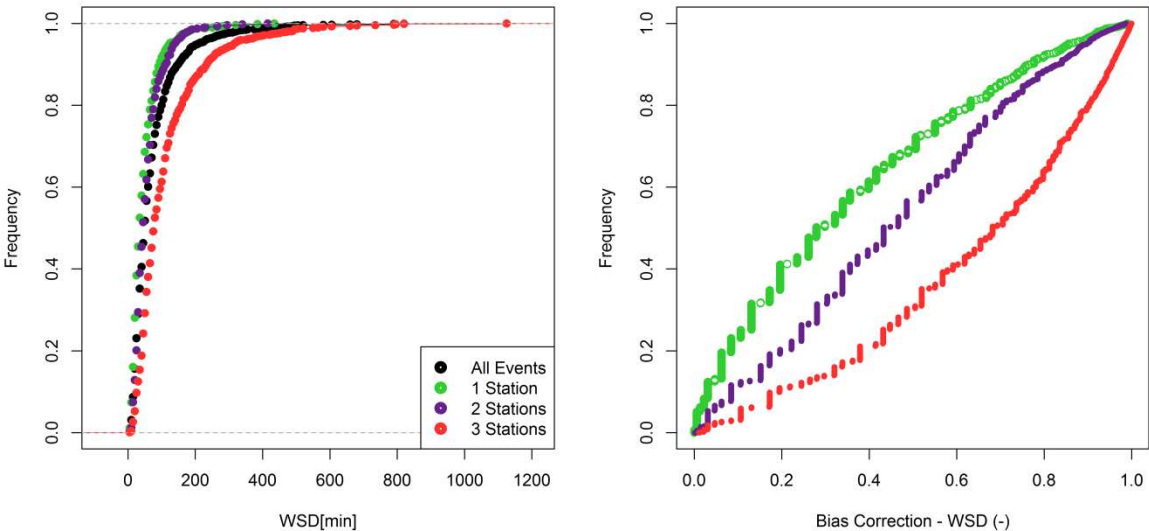
In Figure 5.13 observed and modeled pairs of pseudo-values describing events occurring in three stations during summer for the BAWU\_A case are presented. Pairs of WSA-WSD observed in each of the stations (gray circles with gray background) suggest non-exchangeable behavior that is reproduced by the Vine model (turquoise crosses with gray background). Overall joint structures between pairs of variables are properly modeled by the Vine copula as can be inferred from Kendall's Tau coefficients from upper and lower triangular matrixes.



**Figure 5.13: Pairs of pseudo-values of summer events occurring in three stations simultaneously for the BAWU\_A case (Kendall's Tau in red). Upper triangular matrix: Observed. Lower triangular matrix: Synthetic.**



The probability distributions used for the single site model are fitted to observed time series of event characteristics, i.e. summer and winter DSD, WSD and WSA. The Vine copulas are used for generating random samples of pseudo-characteristics that need to be back transformed to the real space by the marginal distributions. As the copulas are fitted based on events occurring in either 1, 2 or 3 stations, the marginal distributions used for the back transformation should be based on subsamples of events occurring in each of the considered cases. Therefore the probability distributions that are fitted to the whole sample of events can be used for the back transformation by using an additional bias correction that is aiming to transform the pseudo-value from a subsample corresponding to say an event occurring in only 1 station to the total sample of events. Figure 5.14 (left plot) shows the cumulative distributions of WSD corresponding to events classified as occurring in 1, 2 or 3 stations along with the distribution considering all events together for the summer events observed in BAWU\_A case. As can be seen the WSD, corresponding to events observed in 3 stations (red points) are, for similar frequency values, longer than the ones observed in 1 or 2 stations. This is logical since events lasting longer, i.e. with longer WSD, are expected to be registered by several stations. The bias correction is performed based on the probability distribution of the 3 stations together and a curve for correction is provided for each of the number of station cases. If for example a Vine copula modeling events in one station gives as a result a pseudo-WSD value of 0.5, in order to convert it to WSD a value of 0.3 should be used (as can be inferred from the green curve of the right plot in Figure 5.14). On the other hand, if a copula fitted to events occurring in 3 stations generates a pseudo-WSD of 0.5, then a value of 0.7 should be used (see red curve of the same plot) when the pseudo-value is to be transformed using the marginal distribution.



**Figure 5.14: Cumulative distribution functions (left) and bias correction (right) for WSD of summer events corresponding to BAWU\_A case.**

As was mentioned, the method based on Vine copulas proposed in this Thesis consists of different steps of calculation that are sequential and as such, each step is built on the

solution of the previous one. The so called hybrid model is used as a first step of synthesis to randomly generate events occurring in 1, 2 or 3 stations and assign the events to each particular station based on the observed percentages for each case which are shown in Table 5.4. The result of this first step of calculation are long time series of occurrences for each of the involved stations, i.e. either a 0 meaning no event or a 1 meaning event occurring at a particular station. As an example, a short time series of 20 occurrences is generated (see Table 5.5) for the BAWU\_A Summer case; therefore percentages presented in first column of Table 5.4 are used. The occurrences of events assigned to 1, 2 or 3 stations are presented in the first line of Table 5.5 and result of considering the percentages of 61, 20 and 19, respectively. Accordingly, the assignment of occurrence of event to stations A, B or C results from the percentages 56, 58 and 70 for each of the stations.

**Table 5.5: Example of short time series resulting from hybrid model for BAWU\_A Summer case**

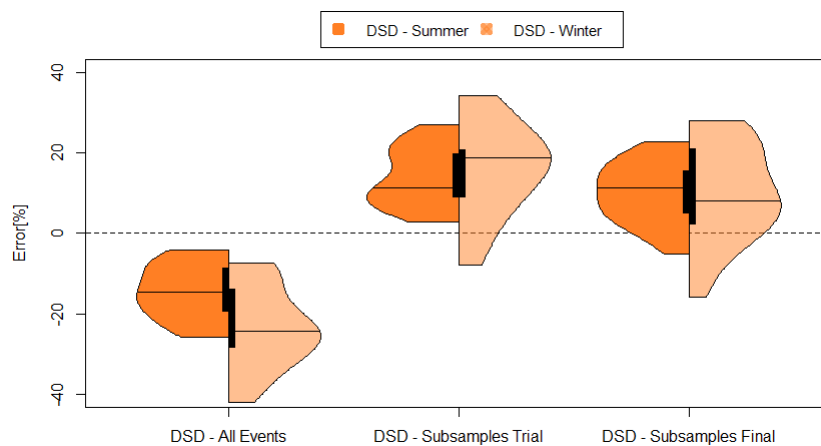
Number of stations in which event occurs	1	3	3	2	1	1	2	2	3	3	1	1	3	2	1	2	1	2	3	1	
Station A	1	1	1	0	0	0	1	0	1	1	0	0	1	1	0	1	0	1	1	1	
Station B	0	1	1	1	1	0	0	1	1	1	0	1	1	1	0	1	0	1	1	0	
Station C	0	1	1	1	0	1	1	1	1	1	1	0	1	0	1	0	1	0	1	0	
		<i>Consecutive</i>							<i>Consecutive</i>												

Thereafter the fitted Vine copula models and bias correction method are applied to assign events characteristics, i.e. WSA and WSD, to each of the time steps in which an event is occurring (assigned as a 1 in the Table presented above). The Vine models (and consequently bias correction curves) to be used are determined by the number of stations in which events are occurring. For the first simulated occurrence shown in Table 5.5, as the event is occurring only in Station A, then the Vine model for 1 isolated station is used. Whereas for the second simulated occurrence, the Vine model fitted to events occurring in 3 stations is applied. This results in long time series of events characterized by WSA and WSD for each of the stations.

The following step is to generate long time series of DSD for the different stations simultaneously, that means to introduce values randomly generated by the probability distribution and to make sure that events occurring in 2 or 3 stations are temporally matching. For this purpose and in order to guarantee that events modeled as occurring simultaneously in several stations occur in the same temporal period, the DSD are generated stochastically for one of the stations and thereafter adjusted for the rest of the stations. The station used for DSD generation changes for the different steps of calculation. For example for the time series generated and presented in Table 5.5 it can be seen that events

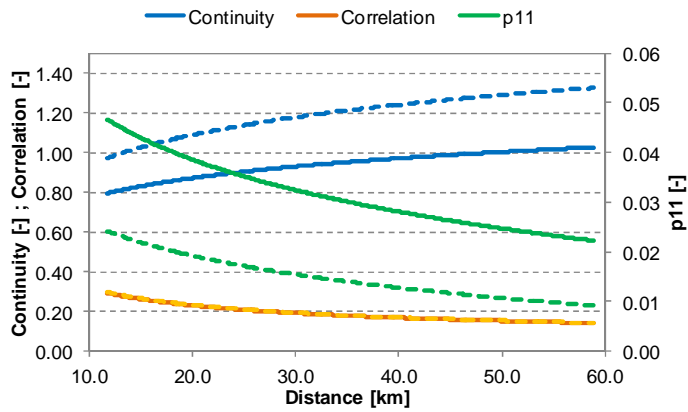
corresponding to columns 2, 3, 9, 10, 13 and 19 must occur simultaneously in the 3 stations, and thus must have the same temporal location. For columns 2 and 3, events occur in all stations for the two consecutive time steps, and therefore the DSD is assigned for one of the stations that is selected randomly and thereafter adjusted for the rest of the stations to fulfill the total duration. These events are denominated “consecutive” cases. On the other hand, for cases like the ones corresponding to columns 3 and 9, many events are occurring and hence many DSD are to be introduced in between (1, 3 and 4 for Stations A, B and C). For these cases the station to be used for DSD generation is selected according to the maximum number of DSD to be generated, i.e. Station C for the exemplified period. These events are denominated “longest” cases. The DSD corresponding to the other two stations are then adjusted to match the total duration between events occurring in all stations for that period.

The described methodology is used for generating continuous time series of events. A first attempt is done using the probability distributions fitted to DSD characterizing all observed events for each of the involved stations. Results show a systematic underestimation of the mean value of DSD when the final time series are evaluated (see Figure 5.15). For this reason some alternatives are evaluated which consist of using subsamples of DSD for introducing either the “consecutive” or “longest” cases events. Events corresponding to “longest” cases are identified from observed time series for each station and a probability distribution is fitted to each subsample of DSD. For the “consecutive” cases two alternatives are considered here. The first one consists of a probability distribution that is fitted to the merging of all DSD from all the stations involved in the multi-site model (events identified as “longest” cases are excluded). This alternative results in a systematic overestimation of DSD mean values and is presented as “DSD – Subsamples Trial” in Figure 5.15. The last alternative consists of identifying events corresponding to “consecutive” cases from observed time series for each station and fitting a probability distribution to the sample consisting of “consecutive” events for all stations involved in the multi-site synthesis. In this case results show overestimation for most of the stations, although some underestimation is as well present (See “DSD – Subsamples Final” in Figure 5.15). This last alternative shows the lowest overall errors and is therefore preferred for incorporating the DSD within the time series and is used for the rest of the analysis.



**Figure 5.15: Errors of mean values of DSD resulting from different alternatives of incorporating them within the time series for all stations.**

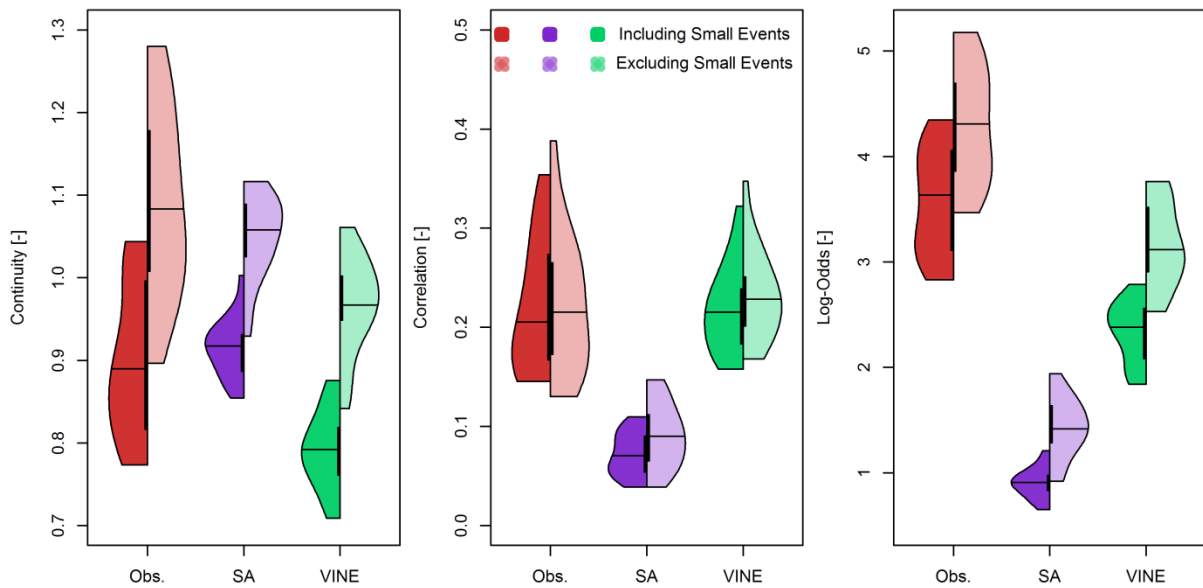
To evaluate the capability of the proposed method to reproduce different event characteristics, an alternative method is applied, namely simulated annealing (SA). For this method the original model for single sites is applied to generate long time series in every station, without taking into account the surrounding stations. Thereafter the events are resampled for each of 8 groups in order to match the spatial consistency measures (see definitions with Eq. 14, 15 and 16). The aim of the method is to find the resampling positions that maximize a criterion. The criterion is composed of different spatial consistency measures which are weighted after a sensitivity analysis. The final measures (weights) are: probability of rain at both stations or p11 ( $2e6$ ), correlation (100) and continuity (10). Different parameters involved in the annealing optimization criteria are as well set up after a sensitivity analysis and include: the number of temperature changes set up to 1000 and number of iterations for each temperature change 500. An initial temperature of 0.1 and temperature decrease rate of 0.9 are as well set up. The resampling is performed considering seasons, i.e. summer events can only be swapped with summer events and the same for winter. The resampling is limited to a window of 50 years, i.e. it is done for different 50 years simulations separately. Finally the resampling is based on events, i.e.  $WSA \geq 1\text{mm}$ , so the spatial consistency measures to be met must be based on time series which only include these events. In Figure 5.16 the final target values used within the optimization algorithm are presented, i.e. the dashed lines for the three measures. As can be seen, the exclusion of small events has a considerable impact in some of these measures. The SA results in a final chronology of rainfall events for each of the stations involved in each group case.



**Figure 5.16: Bivariate spatial consistency measures from observed (continuous lines) and observed without small events (dashed lines) time series used as criteria for SA optimization.**

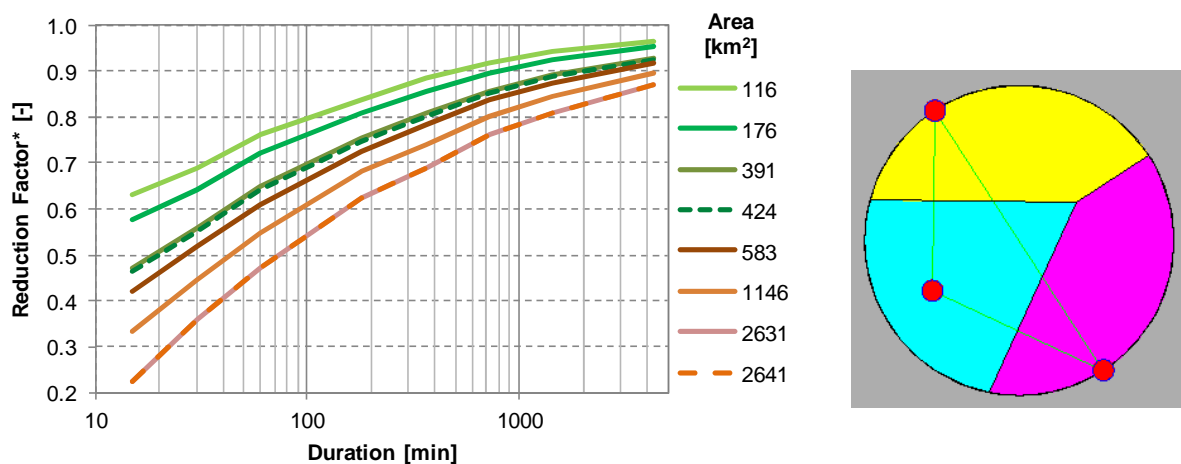
Both the Vine copula model and SA provide long time series of event characteristics for each of the stations involved in the multi-site synthesis. These time series are thereafter transformed into 5 minutes continuous series by applying the internal structure model and introducing the small events as for the single site model. Each of the stations is handled independently for this last step of calculation.

The final synthetic time series consist of 100 years of rainfall and are compared with observations. Some of the evaluations are based on single stations and others on areal precipitation. Spatial consistency measures, i.e for single stations compared in a pair-wise way, are estimated based on the 5 minutes long continuous time series. The different measures are estimated based on observed and synthetic time series with and without small events; the latter analysis is included as it consists of the optimization criteria for the resampling algorithm involved in the SA. Some of the results are shown in Figure 5.17 and include the measures of a total of 18 pairs of stations with distances ranging from 11.7 km to 59 km. It can be seen that the SA is better reproducing the continuity measure, especially if time series without small events are compared; whereas the correlation is better reproduced by the Vine copula. The Log-odds are underestimated by both models; however the Vine copula outperforms the SA as the results are closer to observed ones. Overall, Vine copula appears to reproduce these measures better, even though they are not directly involved within the model set up, as is the case for the SA optimization criteria, so these values are validating the results from the Vine Copula model.



**Figure 5.17: Bivariate spatial consistency measures of observed and simulated time series with and without small events.**

The areal precipitation is estimated including all stations involved for each case study. Observed and synthetic time series from multiple stations (2 or 3) are used and the Thiessen polygon is applied for this purpose. In the case of 2 stations, the areal rainfall is just the average of the 2 time series. For 3 stations the proportion corresponding to each station is calculated as shown in the right plot in Figure 5.18. A description of different variables involved in the estimation and analysis of areal precipitation for the different considered cases is presented in Table 5.6. The reduction factors corresponding to the case studies are as well presented (left plot in Figure 5.18 and Table 5.6) and are applied at the end of this sub-Chapter for IDF evaluation. Note that for 3 of the 8 case studies, i.e. the cases with areas larger than 1000 km<sup>2</sup>, the factors are resulting from the extrapolation by simple linear regressions (shown in Figure 4.17).

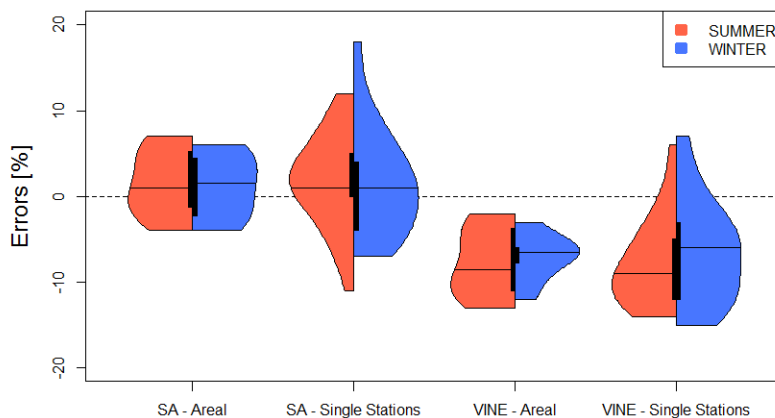


**Figure 5.18: KOSTRA reduction factors for the case studies (left) and graphical explanation of criteria adopted for areal estimation and proportion of stations (right).**

**Table 5.6: Areas, proportion per station and KOSTRA reduction factors for the different case studies.**

CASE	AREA [km <sup>2</sup> ]	Proportions	KOSTRA Reduction Factors* [-] for the following durations [min]:							
			5, 15	30	60	120	240	360	720	1440
BAWU_A	391.1	0.37/0.31/0.32	0.47	0.56	0.65	0.70	0.77	0.81	0.86	0.89
BAWU_B	423.7	0.34/0.30/0.32	0.46	0.55	0.64	0.69	0.77	0.80	0.85	0.89
BAWU_C	176.3	0.5/0.5	0.58	0.64	0.72	0.77	0.83	0.86	0.90	0.93
BAWU_D	116.4	0.5/0.5	0.63	0.69	0.76	0.80	0.85	0.88	0.92	0.94
NS_A	582.8	0.5/0.5	0.42	0.52	0.61	0.67	0.75	0.78	0.84	0.87
NS_B	2641.2	0.37/0.22/0.41	0.22	0.36	0.47	0.55	0.65	0.69	0.76	0.81
NS_C	2630.6	0.27/0.38/0.35	0.22	0.36	0.47	0.55	0.65	0.69	0.76	0.81
NS_D	1146.5	0.43/0.33/0.24	0.33	0.45	0.55	0.61	0.70	0.74	0.80	0.84

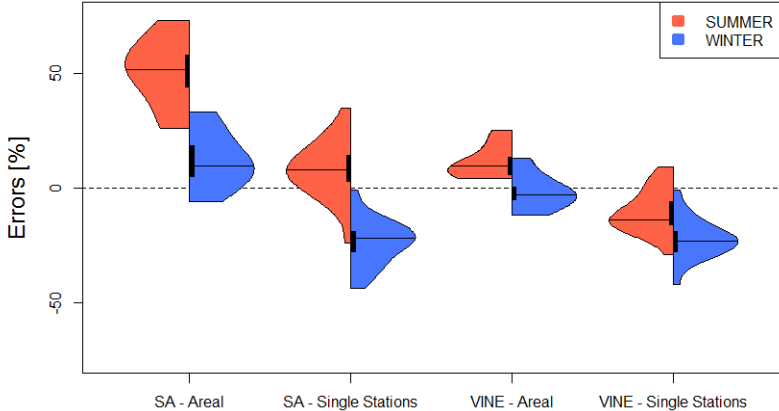
Areal precipitation is estimated from 5 minutes time series corresponding to observations and resulting from each of the multi-site models according to the proportions indicated in Table 5.6. Total seasonal rainfall and proportion of time steps with rainfall are calculated both based on single stations and areal precipitation. The values resulting from the models are compared with observations and are presented as errors in the following figures (Figure 5.19 and 5.20). The SA model shows an acceptable performance when total seasonal rainfall is evaluated both for single stations and areal precipitation as the errors show to be around zero for all stations, with smaller range for the areal precipitation. The Vine Copula underestimates the seasonal rainfall, especially the areal rainfall for which the underestimation is shown for all the cases.



**Figure 5.19: Errors of total seasonal rainfall based on areal and single sites resulting from two models used for multisite synthesis for summer (left plots of violins) and winter (right plots of violins) for all stations.**

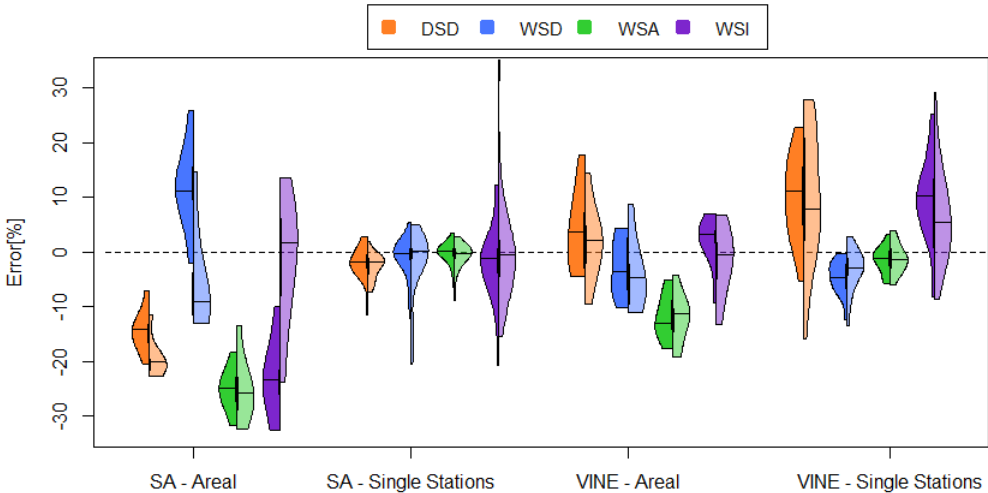
The proportion of time steps with rainfall appears to be better reproduced by the Vine Copula model, especially the areal case for which the range of errors is much lower than the rest of the cases. The SA on the other hand results in an overestimation for most of the cases included in the areal comparison. These results suggest that the Vine Copula model is better

in reproducing the simultaneous occurrence of rainfall in several stations, whereas the SA fails to mimic this simultaneous behavior. For the SA the overestimation of time steps with rainfall results in an acceptable estimation of total rainfall, so the effects are somehow compensated. In the following paragraphs the reasoning behind this compensation is discussed.



**Figure 5.20: Errors of proportion of time steps with rainfall based on areal and single sites resulting from two models used for multisite synthesis for summer (left plots of violins) and winter (right plots of violins) for all stations.**

The capability of the models to reproduce different event characteristics is additionally evaluated. As mentioned, the synthetic time series are 100 years long, and due to the conception of the multisite models, especially for the SA, it is not so straight forward to generate long time series several times to apply tests like MW or GINI. Therefore, their capability to reproduce different moments is evaluated based on the available data. Different moments are considered and ranges of errors for the first order moment and different rainfall characteristics are presented in Figure 5.21.



**Figure 5.21: Errors of mean values of event characteristics based on areal and single sites resulting from two models used for multisite synthesis for summer (left plots of violins) and winter (right plots of violins) events for all stations.**



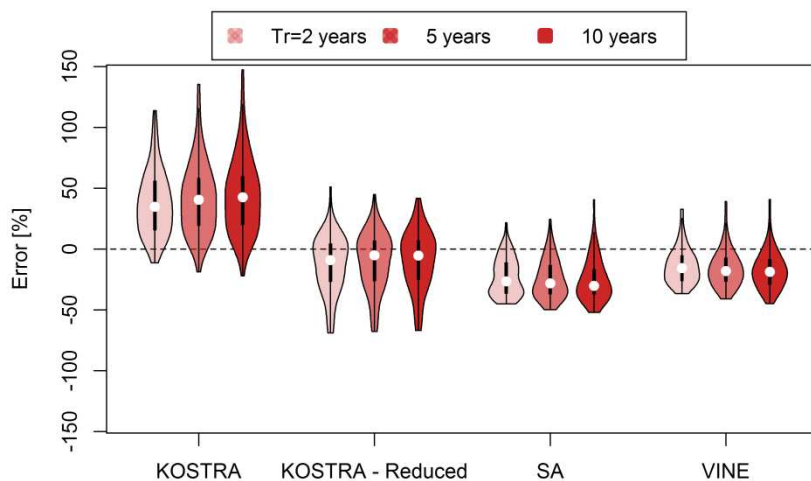
The comparison based on single stations indicates that the SA model is reproducing the first order moments in an acceptable way, as the errors are around zero for all stations. This was expected as the SA is using the single site model for simulating events that are only resampled, so the final results should be similar. For the model based on Vine copulas the single stations results show some deviations. The DSD are for most of the stations overestimated, whereas the WSD and WSA are slightly underestimated but range of errors are lower and the results for these variables are therefore acceptable. The WSD results in higher underestimation compared to WSA, and the combination results in an overestimation of WSI for most of the stations. The overestimation of the DSD is a result of the way that the dry spells are incorporated after the long time series of events occurring in either 1, 2 or 3 stations are simulated. On the other hand even if the WSA and WSD are not directly modeled from a probability distribution as in the SA case the resulting values are acceptable. Event identification based on areal precipitation shows very different results compared with single sites. These results show the outperformance of the Vine Copula model over the SA for most event characteristics. Vine copula delivers very acceptable errors for all characteristics except for the WSA, for which values are underestimated for all cases. This is then causing the underestimation of total seasonal rainfall. The SA model results in an even stronger underestimations of WSA mean values, but it is compensated with the underestimation of DSD, i.e. an overestimation of rainfall events per year and proportion of time steps with rain (see Figure 5.20), and results in an acceptable value of total seasonal rainfall (as shown in Figure 5.19).

Results for different moments are presented as *RSE* in Table 5.7 (see Eq. 18) based on areal precipitation for all cases and the two models for comparison. Errors appear to increase as the order of the moment evaluated increases as would be expected. The Vine Copula model shows for most of the characteristics and moments lower errors compared to the SA when areal precipitation is evaluated.

**Table 5.7: Evaluation of statistics of different variables describing the external structure of rainfall events resulting from two models for multisite synthesis based on areal precipitation and presented for all cases as *RSE*.**

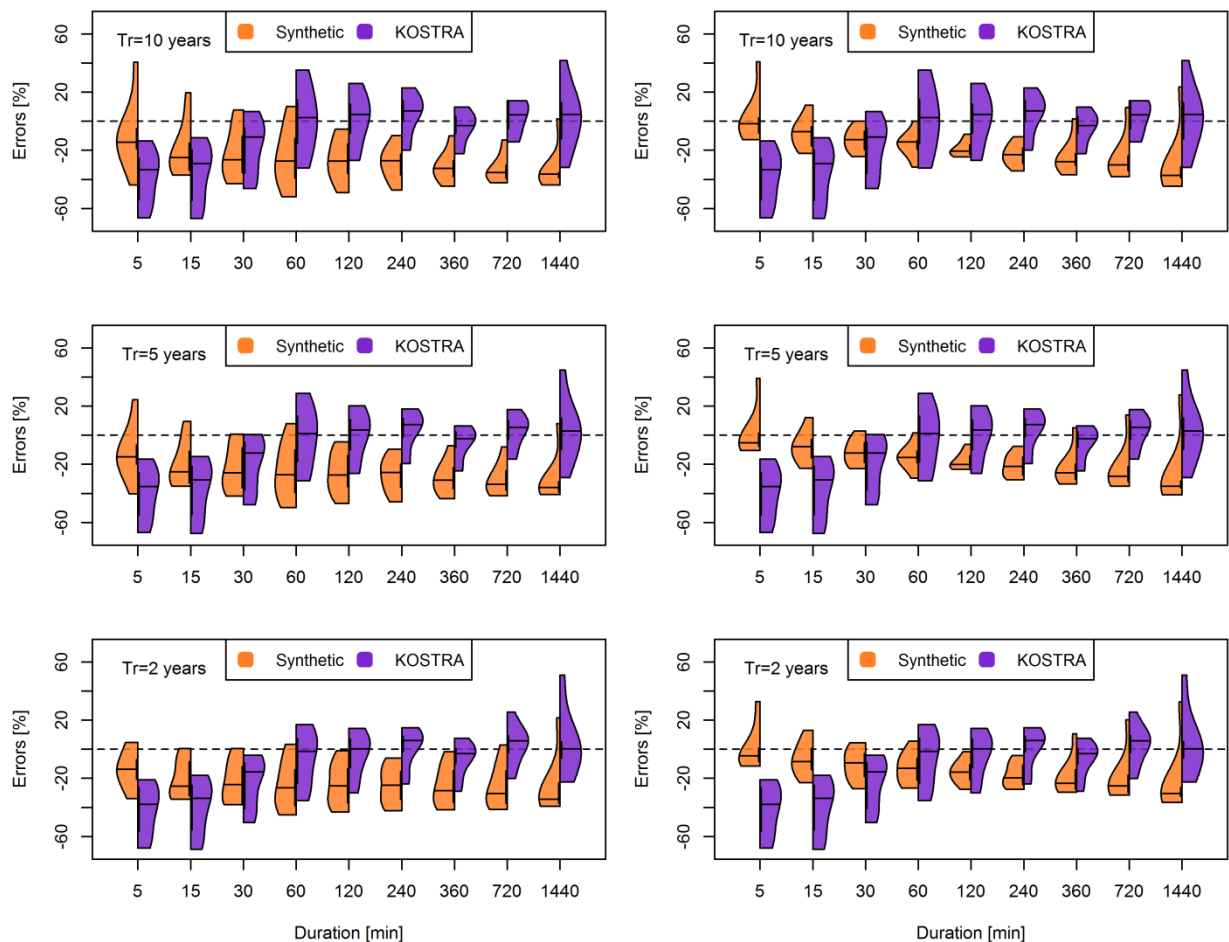
<i>RSE</i> [-]		Mean		Standard Deviation		Skewness		Kurtosis	
Variable	Season	SA	Vine Copula	SA	Vine Copula	SA	Vine Copula	SA	Vine Copula
<b>DSD</b>	Summer	0.15	0.08	0.30	0.07	0.12	0.80	0.31	4.40
	Winter	0.20	0.07	0.36	0.05	0.14	0.45	0.18	1.71
<b>WSD</b>	Summer	0.14	0.06	0.05	0.15	0.18	0.27	0.28	0.54
	Winter	0.10	0.07	0.12	0.13	0.26	0.27	0.81	1.47
<b>WSA</b>	Summer	0.25	0.12	0.38	0.28	0.23	0.26	0.61	0.45
	Winter	0.26	0.13	0.34	0.25	0.34	0.46	1.11	2.81
<b>WSI</b>	Summer	0.24	0.05	0.17	0.27	1.21	0.64	7.29	1.95
	Winter	0.11	0.06	1.15	0.59	3.58	1.65	24.41	6.87

IDFs derived from areal precipitation for all cases are compared with the ones resulting from synthetic time series (SA and Vine Copulas) and KOSTRA (without and with reduction). The factors used for deriving the reductions are presented in the left plot of Figure 5.18 and Table 5.6. Errors between IDFs derived from observations and the different considered cases are presented for return periods that go up to 10 years due to the data availability in the following figure (Figure 5.22). The violin-plots include errors resulting from all cases and durations that range from 5 minutes up to 1 day. It can be seen that if no reduction factor is included, KOSTRA results in an important overestimation of the IDFs for most of the cases. The reduction factors improve the performance of KOSTRA significantly (see “KOSTRA – Reduced”), the errors are around zero, with overall some underestimation especially for low return periods. Range of errors show to be lower for the multisite models, both show underestimation of IDFs but the Vine Copulas model shows median errors closer to zero.



**Figure 5.22: Comparison of models regarding their ability to reproduce extreme events for durations ranging from 5 minutes to 24 hours for all stations.**

A more detailed comparison between each of the multi-site models and KOSTRA-Reduced is presented here. Errors are shown in Figure 5.23 for both cases and the 3 return periods. The ranges of errors resulting from the Vine Copulas model show to be smaller than the ones from the SA and KOSTRA. KOSTRA shows to underestimate extreme values for all return periods and sub-hourly durations. These underestimations may be related with the reduction factors, which range between 0.22 and 0.56 for the corresponding durations as was shown in Table 5.6. On the other hand both multi-site models show acceptable results for low durations with a slight underestimation for all stations, which is more marked as the duration analyzed increases. Summarizing, the multisite models would be preferable for low durations and KOSTRA for long ones, and Vine copulas outperforms SA as the ranges of errors are lower.



**Figure 5.23: Comparison of performance from multisite models (left: SA and right: Vine Copula) and KOSTRA - Reduced to reproduce observed extremes for different return periods (10, 5 and 2 years).**

For the case studies presented here the mean reduction factors (KOSTRA) range from 0.86 to 0.9. If the duration of the events is considered then these factors can get a value as low as 0.2 for short durations (for 15 minutes) resulting in significant underestimations of values provided by KOSTRA Reduced for all cases. The reduction factors for low durations could be revised to improve the performance of KOSTRA.

## 5.4 REGIONALIZATION OF PRECIPITATION

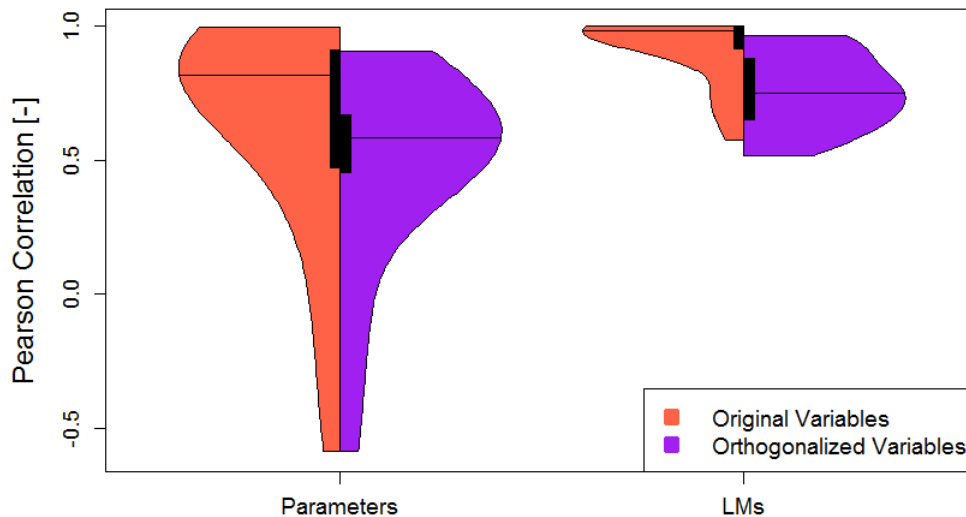
The aim of the regionalization is to estimate the distributions describing rainfall characteristics for any location. These distributions are defined by parameters and are characterized by L-moments (LMs). It was mentioned in the methodology section that some parameters/LMs to be regionalized can be significantly correlated, and therefore four different alternatives are considered. Two of these involve the regionalization of either parameters or LMs directly in a sequential way so that the previously regionalized parameter/LM can be used in the subsequent estimation as additional information. The other two alternatives regionalize the orthogonalized parameters or LMs, so that the correlation among them is eliminated; the regionalized values are then back transformed to the original

space. The regionalization of LMs/orthogonalized LMs requires a post step of calculation to estimate the parameters.

#### **5.4.1 PRECEDING ANALYSIS**

In order to decide which alternative is more appropriate for regionalizing the model, a preprocessing is done based on the four different sets of variables which are used for estimating parameters/LMs. The 24 stations used for single site model development are used for this preliminary analysis. A multiple linear regression (MLR) is used with the following explanatory variables: position of the station, yearly rainfall, gradient of the yearly rainfall, elevation and gradient of the elevation. The same method and variables were used for regionalizing the “Old Model”. It was mentioned that for the cases in which the original parameters or LMs are regionalized, the significantly correlated parameters (or LMs) are as well included as explanatory variables. The sequence is defined after a sensitivity analysis and higher order LMs (or parameters associated with them) are estimated first and then considered as additional information for the consequent order LMs (or parameters). The sensitivity analysis is based on events from all stations that are treated as one station, and then checked for some stations separately.

The two cases that involve parameters/orthogonalized parameters are used for estimating them and thereafter comparing these values with the ones derived from observations for the 24 stations. A similar procedure is followed for cases involving LMs. The final comparison is presented as Pearson correlations as a measure of linear correlation between observed and estimated parameters or LMs. As different numbers of parameters/LMs are involved, along with different event characteristics and seasons, all resulting correlations are presented together as semi-violins and values close to 1 indicate a good performance of the method. The resulting correlations for the different cases in terms of parameters or LMs are shown in Figure 5.24. The two semi-violins located on the left of the figure correspond to the two alternatives which regionalize parameters and the ones on the right the ones that deal with the LMs. The results reveal that either the parameters or the LMs are better reproduced when the original values are regionalized rather than the orthogonalized ones. Therefore, from these results the orthogonalized cases can be excluded from the analysis.



**Figure 5.24: Comparison of variables to regionalize in terms of reproducing the parameters or LMs.**

A further analysis is performed in order to decide the implication of regionalizing either the parameters or the LMs in the synthesis of long time series. The time series are generated using both regionalization methods for 24 stations and all events characteristics along with extreme events are evaluated. The overall results are shown in Table 5.8. As the regionalization of LMs indicates more cases in which the tests are not rejected especially for all events, then this method is preferable. The following analysis is therefore based on the regionalization of the LMs.

**Table 5.8: Comparison of regionalization based on Parameters and LMs in terms of percentage of cases in which the Tests (MW: all events or CvM: extreme events) are not rejected.**

Test	Parameters	LMs	Total cases
MW	21	58	24 stations x 4 variables x 2 seasons
CvM	58	66	24 stations x 5 Durations

In the following analysis the six probability distributions describing the different rainfall characteristics, i.e. DSD, WSD and WSA observed in Summer and Winter, are to be regionalized based on the proposed and alternative methodologies, namely the copula-based one, multiple linear regression (MLR) and regional frequency analysis (RFA). As was mentioned the first two methods aim to regionalize LMs, whereas RFA relies on the LMs for the grouping of the stations and results in a regional probability distribution for each of the groups. According to the probability distributions selected for each of the variables involved in the precipitation model different numbers of LMs need to be estimated, these are: DSD(LM1,LM2,LM3), WSD(LM1,LM2), WSA(LM1,LM2,LM3) for summer and winter separately, i.e. 16 parameters must be estimated for the copula-based and MLR regionalization methods. Despite the methodological difference between the 3 methods they all are based on the same stations and period of registers along with site descriptors (SDs),

hence a comparison among them is appropriate. A total of 26 and 27 SDs are available for the regionalization of LMs describing summer and winter events and were presented in the sub-Chapter 4.5.

## 5.4.2 COPULA-BASED METHOD

The copula-based proposed model is set up on pseudo-values of SDs which are used for simulating possible values of the LMs. The SDs used for the simulation are limited to cases that have a Kendall's Tau correlation (absolute value) equal or higher than 0.4, this way only SDs that indicate high correlations with the target variable are included. As an example in Figure 5.25 the pseudo-values of target variables LM1 describing summer events are shown along with pseudo values of some of the considered SDs. The Kendall Tau correlations among different pairs of variables are shown in red. As can be seen if the DSD (LM1) is to be estimated based on the shown descriptors, the only variable to be used would be the Annual Rainfall, as this is the only SD with an absolute correlation higher than 0.4. For the WSD (LM1) the SDs longitude, elevation and distance to sea would be used, whereas for WSA (LM1) only longitude and distance to sea. Another aspect of the proposed model is that the amount of values to be simulated using each of the SDs is proportional to the Kendall's Tau. This means that for example for the WSD case most of the variables would be generated from the pair WSD (LM1)-Longitude since this pair shows the highest correlation, followed by distance to sea and then elevation. For the shown example from the total number of samples simulated with the first pair, only 87% (0.47/0.54) of values would be generated with the second pair and 77% (0.42/0.54) with the third pair. The three samples would then be combined and a representative value (e.g. the median) of the WSD (LM1) would be taken from this mixed sample.

The final number of bivariate copulas included for regionalizing each of the LMs is presented in Table 5.9. These numbers include copulas fitted to SDs or previously estimated LMs. Different families are selected for the different bivariate relationships, among the most frequently selected are the Normal, Gumbel and t-copulas.



Figure 5.25: Example of relationship between pseudo values of some site descriptors and target variables. Numbers in red indicate the Kendall's Tau correlation coefficients.

Table 5.9: Number of bivariate relationships considered in the regionalization of different LMs using copula based approach.

Variable	Moment	Summer	Winter
DSD	LM1	7	5
	LM2	12	7
	LM3	12	9
WSD	LM1	9	6
	LM2	10	3
WSA	LM1	6	1
	LM2	7	2
	LM3	4	3

The variables with the highest correlation coefficient, and therefore with the highest number of sampled values for the target LMs, are listed in Table 5.10, along with the values of the corresponding Kendall's Tau.

Table 5.10: Variables (SD or LM) with highest correlations with LMs to be regionalized.

Variable	Moment	Summer		Winter	
		Kendall's Tau	SD or LM	Kendall's Tau	SD or LM
DSD	LM1	0.61	mean rainfall: summer	0.67	mean rainfall: yearly
	LM2	0.90	DSD_LM1_Summer	0.67	mean rainfall: yearly
	LM3	0.76	DSD_LM2_Summer	0.85	DSD_LM2_Winter
WSD	LM1	0.53	Longitude	0.71	WSD_LM1_Summer
	LM2	0.79	WSD_LM1_Summer	0.71	WSD_LM1_Winter
WSA	LM1	0.64	WSD_LM1_Summer	0.51	WSD_LM2_Winter
	LM2	0.84	WSA_LM1_Summer	0.9	WSA_LM1_Winter
	LM3	0:80	WSA_LM2_Summer	0.85	WSA_LM2_Winter

As the copula-based method models the pseudo-values, they need to be transformed to the LMs space with marginal distributions; therefore these distributions need to be estimated for each of the 16 variables based on all stations. The final models involve 3 different types of distributions namely Weibull (3 parameters), Generalized Logistic (3 parameters) and Wakeby (5 parameters) and are listed in Table 5.11.

**Table 5.11: Distributions selected for modeling the marginal behavior of the different LMs.**

Variable	Moment	Summer	Winter
DSD	LM1	Wakeby	Wakeby
	LM2	Wakeby	Wakeby
	LM3	Generalized Logistic	Wakeby
WSD	LM1	Wakeby	Wakeby
	LM2	Wakeby	Wakeby
WSA	LM1	Generalized Logistic	Weibull
	LM2	Generalized Logistic	Weibull
	LM3	Generalized Logistic	Weibull

### 5.4.3 MULTI-LINEAR REGRESSION (MLR)

MLR is as well applied and main features are described as follows. As mentioned to avoid colinearity between SDs partial correlations are analyzed and only SDs with partial correlations lower than 0.7 are included in the models estimation. Thereafter for the MLR a stepwise selection of the model by AIC is applied to decide which SDs and LMs are included as explanatory variables. Residuals are supervised to be unbiased with mean values around zero. The final number of variables included in each of the MLR models is presented in Table 5.12. The 1\* indicates that the LM<sub>i</sub> estimated in the previous step of calculation is included as explanatory variables as well.

**Table 5.12: Number of explanatory variables considered in the regionalization of different LMs using MLR.**

Variable	Moment	Summer	Winter
DSD	LM1	12+1*	4+1*
	LM2	5+1*	4+1*
	LM3	9	14
WSD	LM1	9+1*	3+1*
	LM2	7	15
WSA	LM1	4+1*	6+1*
	LM2	1*	1*
	LM3	10	9

### 5.4.4 REGIONAL FREQUENCY ANALYSIS (RFA)

A brief resume of the RFA results are extracted from Leimbach (2017) and are presented here. The clustering of stations resulted in groups that show very similar spatial distributions when the different variables are compared; nevertheless different number of clusters are used according to the variable and range from 4 to 6 clusters, and the number of stations per cluster range from 5 up to 49 (median 17). The probability distributions fitted to the regional



data sets are presented in Table 5.13 and some differ from the ones used in this Thesis (Kappa for DSD and WSA and LN3 for WSD), but are related among each other as special cases, for example the Generalized Pareto is a special case of the Kappa and the Generalized Normal is as well related to the Log Normal. Further details can be found in the original work.

**Table 5.13: Distributions and number of clusters selected for modeling the regional data sets involved in the RFA.**

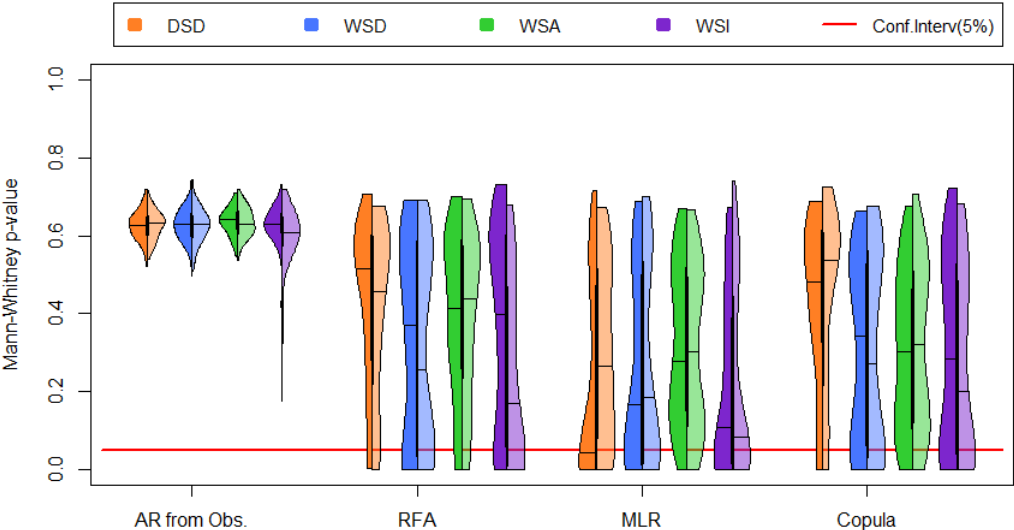
VAR.	Detail	Summer	Winter
<b>DSD</b>	PDF	Pearson	Kappa
	N° clusters	6	5
<b>WSD</b>	PDF	Generalized Normal	Generalized Normal
	N° clusters	5	4
<b>WSA</b>	PDF	Kappa	Wakeby or Generalized Pareto
	N° clusters	4	5

One of the stations located in the Lower-Saxony region was excluded from the regionalization procedure (Station Braunlage, N°1 from Table 4.1). The RFA was not able to allocate this station to any of the clusters, and therefore it had to be excluded from the evaluation of regionalization methods.

**5.4.5 COMPARISON OF DIFFERENT METHODS**

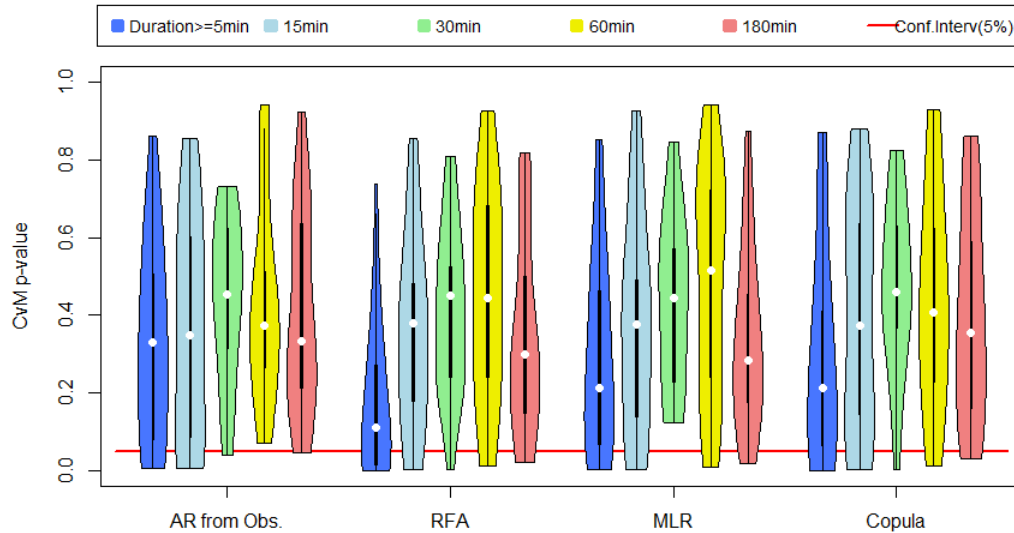
In order to compare the proposed methodology with the alternative commonly used methods, long time series are generated based on the 81 stations and cross-validation, i.e. leaving the station that is to be estimated out of the analysis. The results of the three regionalization methods are as well compared with the single site model without regionalization here referred as “AR from Obs.”. The Regional Empirical Copula is used in all cases for the joint modeling of WSA-WSD. The ability of the different regionalization methods to model the different event characteristics are evaluated by using the MW Test and the results are shown in Figure 5.26. As can be seen, all the regionalization methods result in some of the stations for which the tests are rejected as the p-values are lower than 0.05. A comparison among the different variables to regionalize reveals that the DSD is the variable that shows better results, whereas the WSD is the most difficult one, an effect that is transferred to the WSI. Nevertheless when the three methods are compared the values indicate that both RFA and Copula-based model are superior compared to MLR as the p-values for all variables are higher. A comparison between RFA and Copula-based is not clear from this figure, only the WSA shows to be better reproduced using the RFA method. The capability of the different models to reproduce event characteristics is as well evaluated based on deviations for

different moments. Results are presented as *RSE* (see Eq. 18) in the Appendix G for all stations and the three regionalization models for comparison.



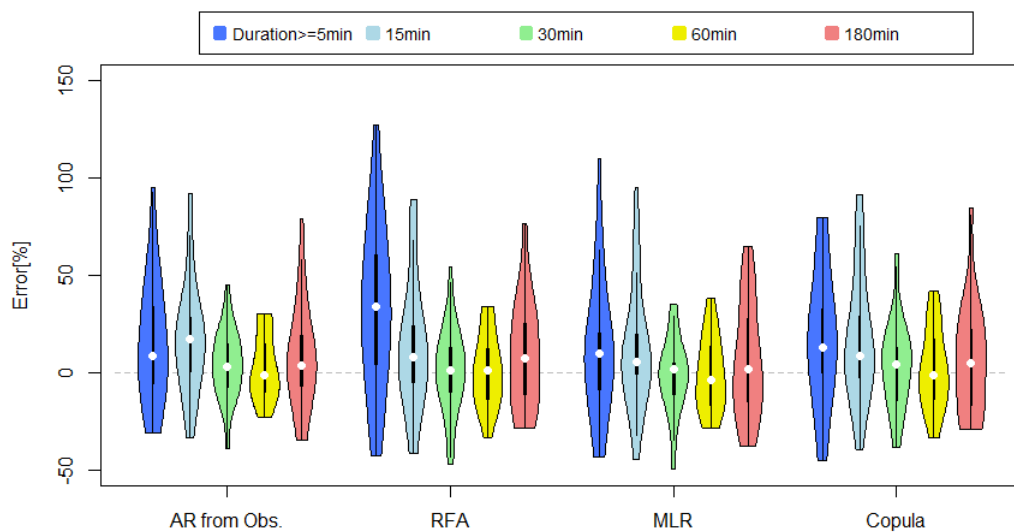
**Figure 5.26: P-values resulting from applying the Mann-Whitney test to compare the different regionalization models based on summer (left plots of violins) and winter (right plots of violins) events characteristics for all stations.**

The different methods are compared in terms of their ability to reproduce extreme events and the CvM Test results are shown in Figure 5.27 for different durations and all stations as p-values. The test is only applied to 23 of the 81 stations, so only to the stations in which the available observed time series is at least 10 years. As expected, the p-values are lower than the ones from the original ARP model. Nevertheless, the decline of the results compared to direct parameter estimation is not as obvious as for the MW test. A comparison between the three regionalization methods indicates that the MLR and copula-based methods perform slightly better over all the durations.



**Figure 5.27: P-values resulting from applying the Cramér-von Mises test to compare the different regionalization models based on extreme values observed during the summer for all stations.**

To further analyze results of extreme values, the Gumbel distribution fitted to annual series of different durations is used for the CvM test and to estimate values associated to different return periods. The same procedure is followed based on observed time series. Errors (see Eq. 17) between observed and estimated values for different durations and return periods were thereafter calculated. The results for the return period of 10 years are presented in Figure 5.28. The errors show to be more concentrated around zero for the copula-based regionalization method. Another aspect that can be seen in the figure is that the errors from the original model (without regionalization) are in the range of errors of regionalized values, especially if the copula-based method is considered. Figures with errors resulting from other return periods (2, 5, 10 and 20) show similar behavior and are presented in the Appendix H.



**Figure 5.28: Comparison of regionalization models regarding their ability to reproduce extreme events for different durations and a return period of 10 years.**

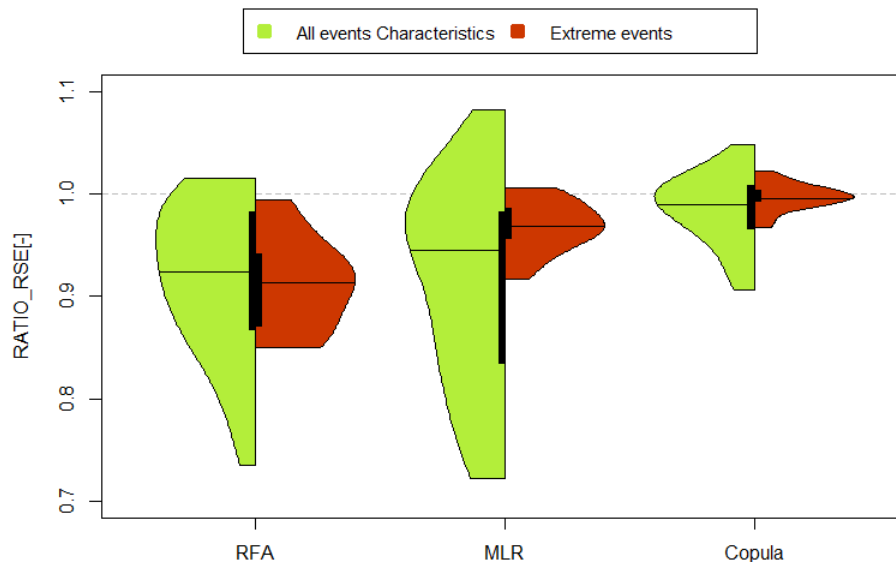
An overall comparison of the different regionalization methods is performed by calculating the percentage of cases in which the tests are not rejected both for all event characteristics

(summer and winter) and for the extreme events. The results are shown in Table 5.14 and indicate that for the MW test both RFA and copula based method are as good, as is deduced from the Figure 5.26, whereas for the extreme events the three methods perform very similarly with MLR slightly outperforming copula based model. Overall the copula based model would be preferred over the other two methods, although none of the methods is clearly outperforming the others.

**Table 5.14: Comparison of regionalization based on different methods in terms of percentage of cases in which the Tests (MW: all events or CvM: extreme events) are not rejected.**

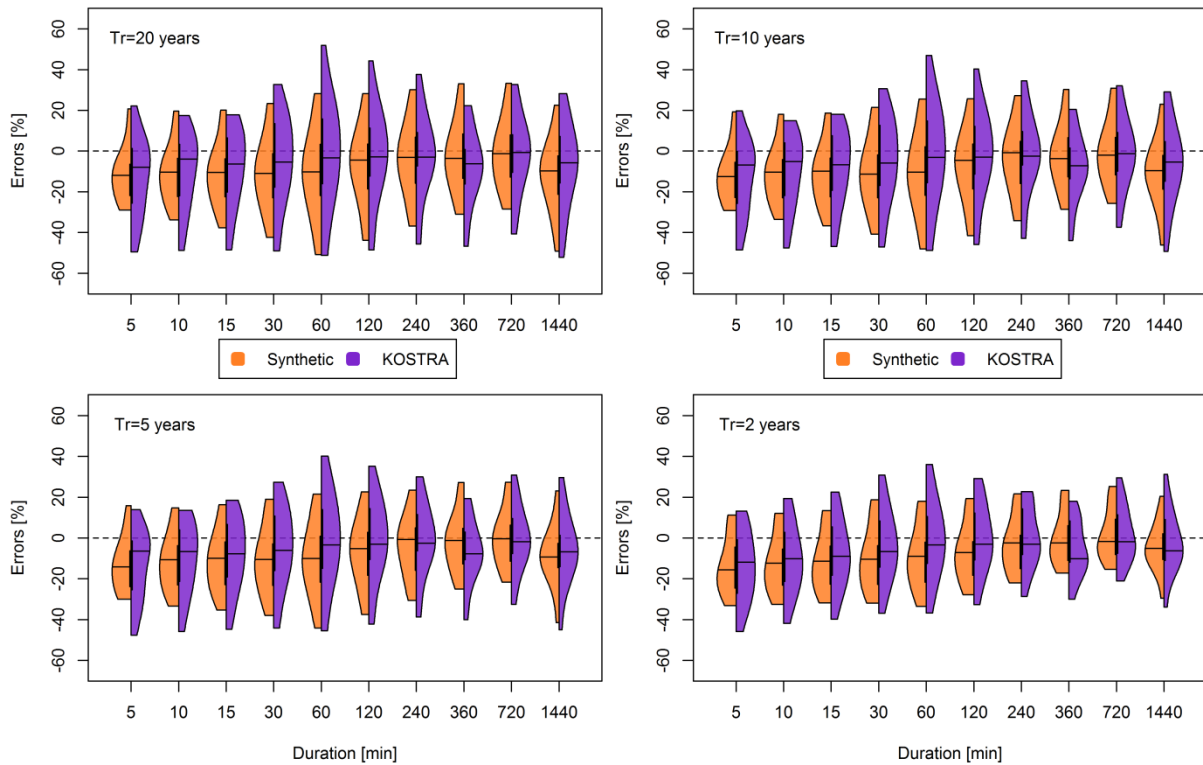
Test	RFA	MLR	Copula	Total cases
<b>MW</b>	81	70	80	81 stations x 4 variables x 2 seasons
<b>CvM</b>	85	90	88	23 stations x 5 Durations

To decide which regionalization method is more robust the  $RSE$  are calculated based on several variables describing events (4 moments of DSD, WSD, WSA and WSI summer and winter events) and for several extreme events (Durations of 5, 15, 30, 60 and 180 minutes and return periods of 2, 5, 10 and 20 years). These  $RSE$  are calculated based on the time series corresponding to the cross validation ( $RSE_{LOOCV}$ ) analysis that were presented in the previous figures. An additional analysis is done with the three regionalization methods but based on all stations ( $RSE_{ALL}$ ), i.e. excluding cross validation. The ratio of  $RSE$  estimated with the inclusion of all stations over the one obtained by cross validation is calculated using Equation 21. Values of the ratio close to 1 show that the method is robust to new stations. The results are shown in the Figure 5.29 both for all and extreme events. The copula-based regionalization method shows a clear superiority compared to the other two methods in terms of robustness to new stations. The RFA method appears to be very sensitive to the addition of more stations.



**Figure 5.29: Ratio between Relative Standard Errors resulting from using all stations and cross validation for the four moments and all event characteristics (left plots of violins) and extreme events for all durations and return periods (right plots of violins) for all stations.**

The different assessment criteria for evaluating the regionalization of the model presented in the previous paragraphs are based on the external structure. A last analysis is included in which the internal structure is as well evaluated. IDFs derived from observations of 14 stations with registers longer than 19 years are compared with the ones resulting from synthetic time series and KOSTRA values. Synthetic time series are generated using the copula-based regionalized method as it showed to be the most robust method. Errors are presented in Figure 5.30 for 4 different return periods. The medians and ranges of errors resulting from both data sets are comparable; nevertheless regionalization shows to be slightly more robust compared to KOSTRA. Regionalized time series results show on average underestimation of extreme events when all stations are considered, except for durations of 240 to 720 for which median of errors are close to zero. Regionalized time series outperform KOSTRA design storms for the mentioned durations, for other durations both data sets are similar.

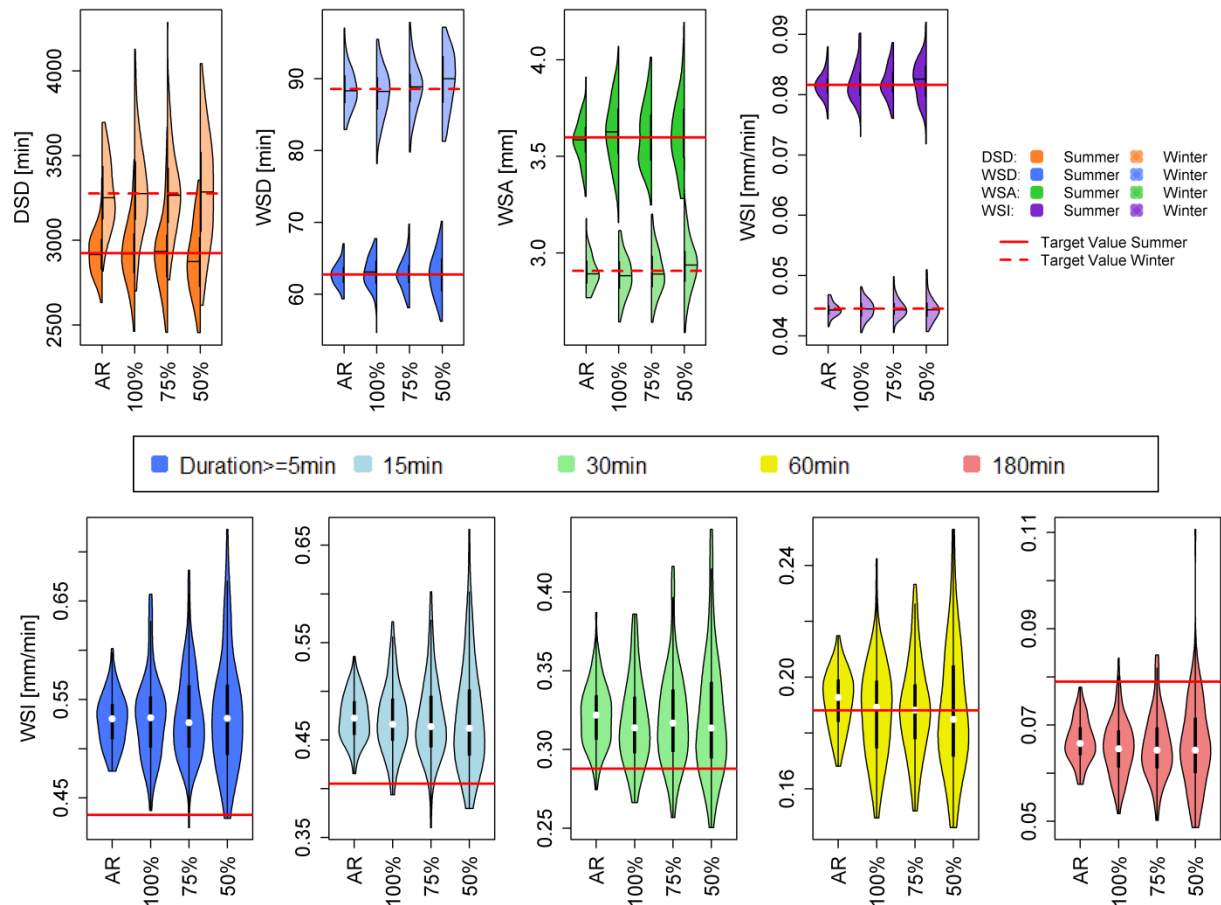


**Figure 5.30: Comparison of performance from regionalization model using Copulas and KOSTRA to reproduce observed extremes for different return periods (20, 10, 5 and 2 years).**

## 5.5 UNCERTAINTY ANALYSIS

In this Thesis two different sources of uncertainty are considered: natural variability of the stochastic process and uncertainty of the parameter estimation. The first uncertainty is analyzed using fixed parameters which are estimated based on all available observations and are assumed to be correct. The uncertainty of model parameter estimation is analyzed considering different percentages of available data which are used for the estimation. These percentages are 100, 75 and 50% and data is randomly sampled with replacement from all observed events. The analysis consists of simulating 100 samples of events and only considers the external structure. The different realizations are used to extract the four moments describing event characteristics (DSD, WSD, WSA and WSI both summer and winter) and the intensities of extreme values associated to 4 return periods (2, 5, 10 and 20 years) and 5 durations (5, 15, 30, 60 and 180 minutes). An example for one station (Lingen) is shown in Figure 5.31 (top: mean values and bottom: extreme events with  $Tr=2$  years). The horizontal red lines (continuous: summer/extreme values and dashed: winter) indicate the attributes estimated from all observations, i.e. the target values. The semi-violins (left: summer and right: winter) and violins (extreme values) indicate the distribution of attributes resulting from each of the 100 realization. The label “AR” stands for results of the alternating

renewal process, i.e. the analysis of natural variability of the stochastic process. The other distributions (“100%”, “75%” and “50%”) result from the simulations using re-estimated model parameters after resampling different percentages from observed events.



**Figure 5.31: Mean values (top) and extreme events with a  $T_r=2$  years (bottom) resulting from observations (red lines) and 100 realizations (Violinplots) using the single site model based on all observed events (AR), and resampling of different percentages of observed events (100, 75 and 50%).**

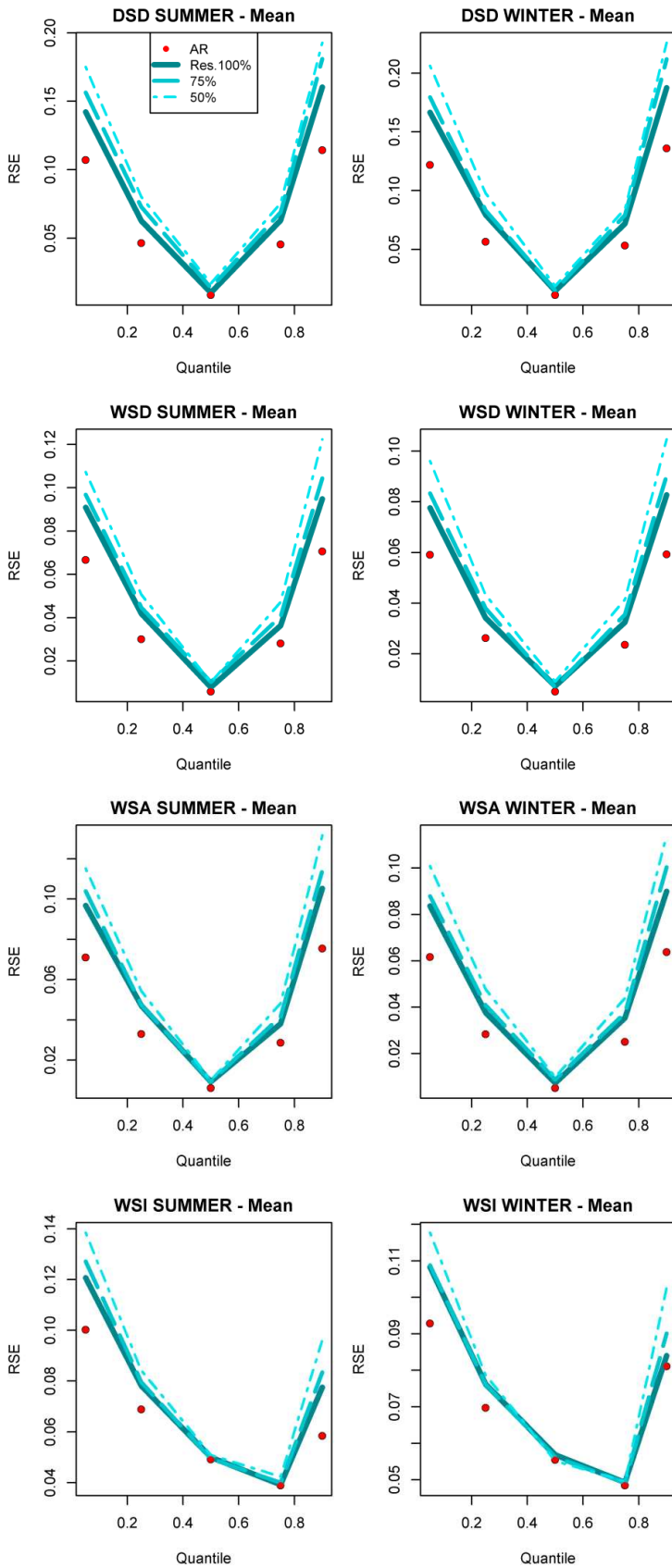
The aim is to compare the resulting values from the different realizations with the target values (estimated from all observations). For this purpose some quantiles representing the 100 realizations are extracted, namely 0.05, 0.25, 0.5, 0.75 and 0.95, and deviations from these quantiles to target values, i.e. to the red horizontal lines in Figure 5.31, are calculated for each station and thereafter the *RSE* is computed based on all stations. The computation involves a total of 81 stations (for moments describing all events) and 23 stations with at least 9 years of observations (for extreme events), i.e. stations located in the state of Lower Saxony and surroundings used for regionalization of the model. The analyses are performed using the regional empirical copula.

*RSE* are presented in Figures 5.32 and 5.33 for mean values of event characteristics and extreme events. The red dots show the results for the natural variability of the process, whereas the different curves indicate the uncertainty of parameter estimation. As can be

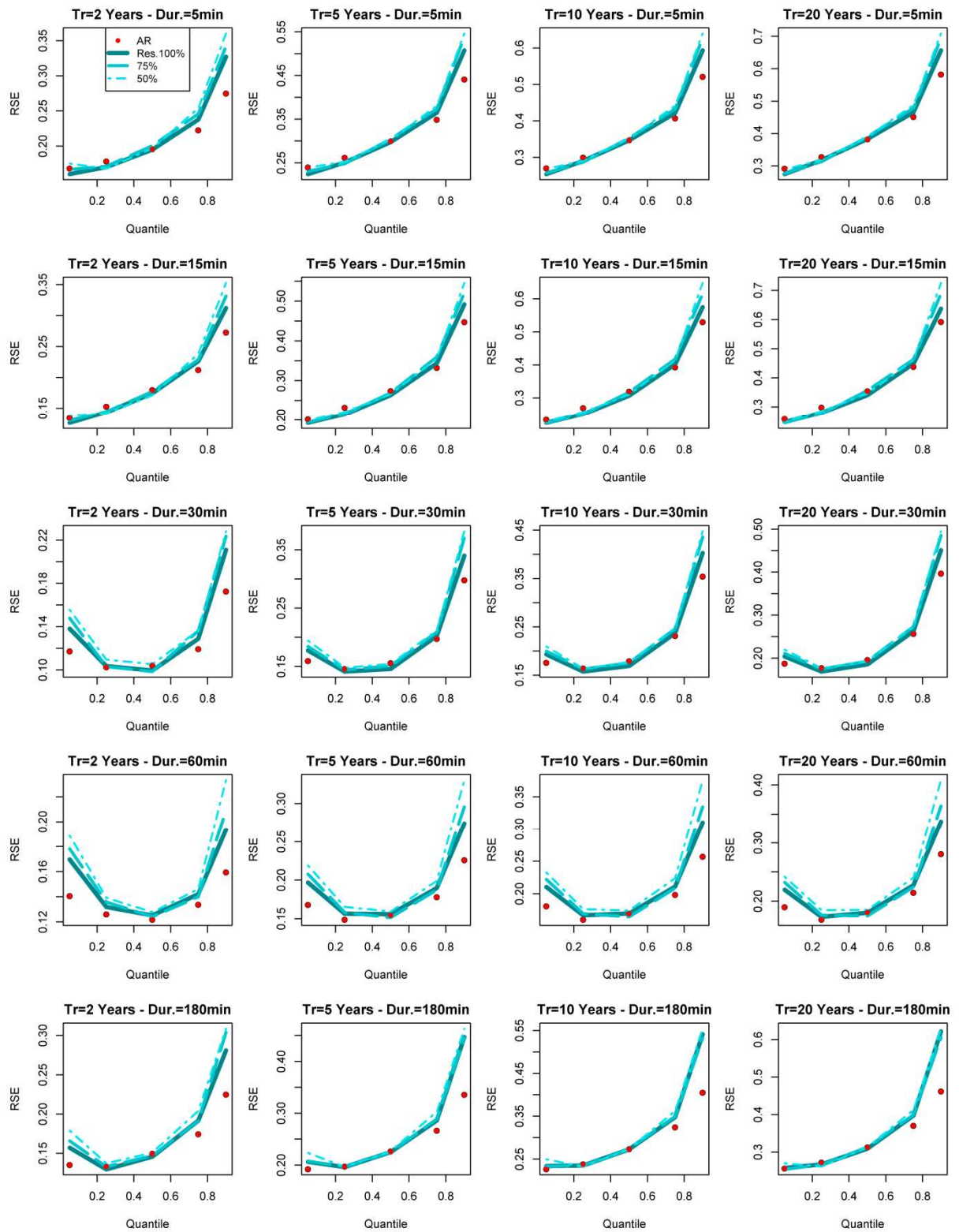
seen from the red dots the results for the mean values of event characteristics that are directly modeled, i.e. DSD, WSD and WSA indicates that the model is unbiased as the lowest *RSE* corresponds to the median and the behavior is symmetric around it, i.e. for low and high quantiles, when all stations are considered in the analysis. For the WSI which is derived from the ratio WSA/WSD the results indicate a tendency of underestimating the mean value for all stations as the lowest *RSE* corresponds to the quantile 0.75. The Standard Deviation (SD) shows a slight overestimation for all characteristics, whereas the overestimation is more obvious for the Skewness and Kurtosis (see Appendix I). When extreme values are analyzed (see red dots in Figure 5.33), the model shows to overestimate intensities, especially for short durations, whereas for durations of 30 and 60 minutes the results are acceptable, as the behavior is closer to unbiased for all stations. As was mentioned, the uncertainty analysis of the input data is studied by considering the different percentages of available events which are used for estimating parameters of the probability distributions and thereafter generating long synthetic time series. As expected the different curves in Figures 5.32 and 5.33 indicate that the *RSE* increases as the different percentages of input data decreases. Nevertheless the resulting values show the same behavior as the ones from the original model for all the analyzed attributes, showing that the model is quite robust to the input data.

Uncertainty of parameter estimation and natural variability of the stochastic process are compared by calculating ratios of *RSE* considering different percentages of available data over *RSE* using all observed events, i.e. "AR". Median values of these ratios both for different event characteristics and for extreme events are presented in Table 5.15. The values indicate that the model is very robust to input data, as the increment in *RSE* shows to be limited, especially for the WSD, WSI and extreme events which have ratios close to 1. Uncertainty of parameter estimation shows to be in the range of natural variability of the process for most of the attributes. Maximum ratios are as well presented and reach a value of 2 for all events and 1.47 for extreme values as maximum deviations.





**Figure 5.32: Uncertainty of mean values: RSE for different quantiles resulting from 100 realizations using the single site model based on all observed events (red dots) and on resampling of different percentages of events (blue curves) for 81 stations.**



**Figure 5.33: Uncertainty of extreme events: RSE for different quantiles resulting from 100 realizations using the single site model based on all observed events (red dots) and on resampling of different percentages of events (blue curves) for 23 stations.**

**Table 5.15: Ratios between *RSE* estimated based on resampling 100, 75 and 50% of observed events over *RSE* estimated with the original AR model. The values correspond to median and maximum ratios.**

<b>Case</b>	<b>Median 100%</b>	<b>Median 75%</b>	<b>Median 50%</b>	<b>Maximum 100, 75, 50%</b>
<b>DSD*</b>	1.18	1.32	1.33	2.02
<b>WSD*</b>	1.05	1.06	1.03	1.80
<b>WSA*</b>	1.12	1.14	1.19	1.91
<b>WSI*</b>	1.00	1.01	1.02	1.65
<b>Extreme Events**</b>	1.03	1.06	1.07	1.47

\*includes 4 moments, summer and winter

\*\* includes 4 return periods and 5 durations

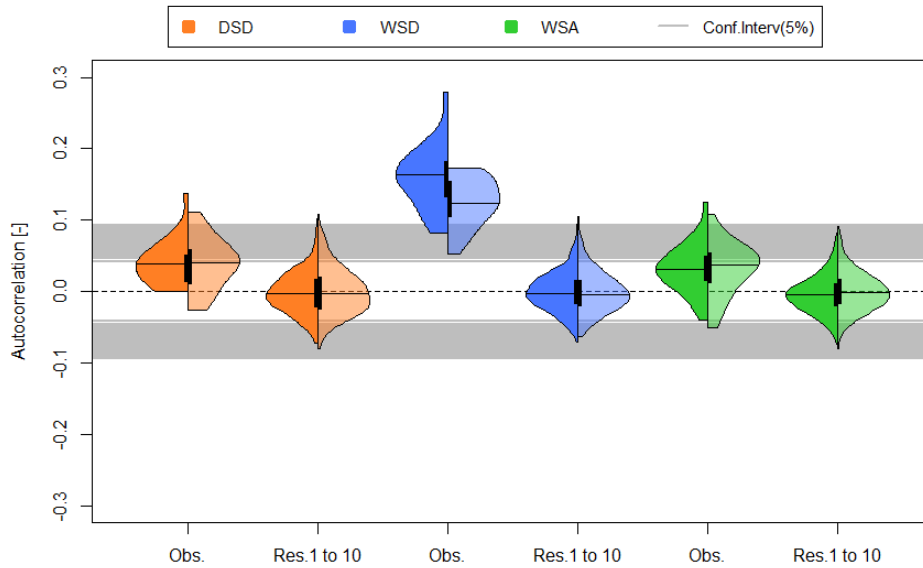
## **6 APPLICATION TO URBAN HYDROLOGY**

Precipitation time series of high temporal resolution are of particular interest for applications in urban hydrology. In this Chapter different analyses based on various rainfall time series are performed regarding the implication for urban hydrological modeling. The small fictional urban hydrological system presented in sub-Chapter 4.7 is used for this purpose. It allows carrying out continuous simulations for assessing the occurrence of floodings, even for long time series.

### **6.1 UNCERTAINTY OF INPUT DATA IN THE RESPONSE OF URBAN CATCHMENTS**

Two different analyses are performed to assess the uncertainty introduced by observed time series and its propagation in urban hydrological simulations. On one hand the source of error due to serial autocorrelation in time series and on the other hand the natural variability of the stochastic process and its effect on the hydrological response.

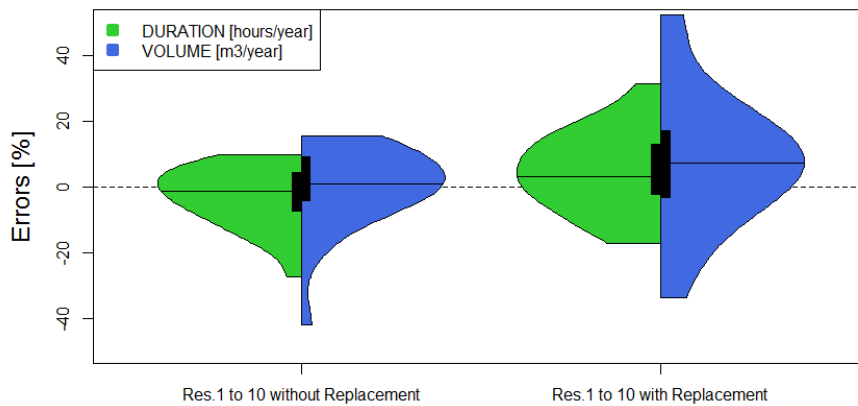
As was mentioned in sub-Chapter 4.1, the variables describing characteristics of rainfall events show significant autocorrelations, especially for WSD. In order to analyze the effect of this autocorrelation in urban response, different time series are generated based on observed ones by randomly resampling the order of rainfall events and therefore removing the autocorrelation. The resampling is performed without replacement for each season separately. A total of 10 resampled time series with length equal to the observed ones are generated for each of the 24 stations used for developing the single site model. Figure 6.1 shows that the autocorrelation is reduced to values around the zero for all event characteristics.



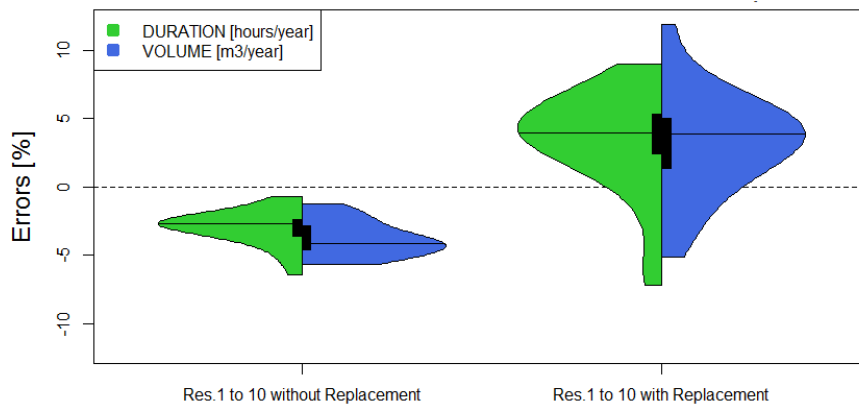
**Figure 6.1: Autocorrelation of summer (left plots of violins) and winter (right plots of violins) events for all stations resulting from original time series (Obs.) and resampling without replacement based on observed time series (Res. 1 to 10).**

The 10 resampled time series are used as the input for the urban system and flooding and overflow events are compared with the ones resulting from the observed time series. Median errors of duration and volume resulting from the 10 simulations are calculated for each station (see Eq. 17). A similar analysis is performed for assessing the natural variability of the process, but in this case the resampling of events is performed with replacement, i.e. some events may be included more than once and consequently others excluded from the generated time series. The errors resulting from both analyses and for all stations are presented in Figures 6.2 and 6.3. Range of errors resulting from the analysis of serial autocorrelation are smaller than the ones representing the natural variability of the process. This is explained by the fact that in the first case, i.e. resampling without replacement, always the same set of events is compared just the order of occurrence is altered.

Clustering of events can induce temporal autocorrelation and therefore cause more severe flood events resulting from the cluster effect of intense events. The removal of autocorrelation of rainfall events does not result in a systematic underestimation of variables describing flood events. The errors show a balance between under- and overestimation that could rather be explained by the variability of the stochastic process. However, removing autocorrelation results in a systematic underestimation of overflow volume and duration, therefore the temporal occurrence of events shows to be very important for the overflow response. Nevertheless, the figures show that the input uncertainty has a stronger effect on the variables describing flood events, especially the volume which shows a larger range of errors.



**Figure 6.2: Errors of flood event characteristics for all stations resulting from resampling the observed time series. Sources of error: Serial autocorrelation (left) and natural variability (right).**



**Figure 6.3: Errors of overflow event characteristics for all stations resulting from resampling the observed time series. Sources of error: Serial autocorrelation (left) and natural variability (right).**

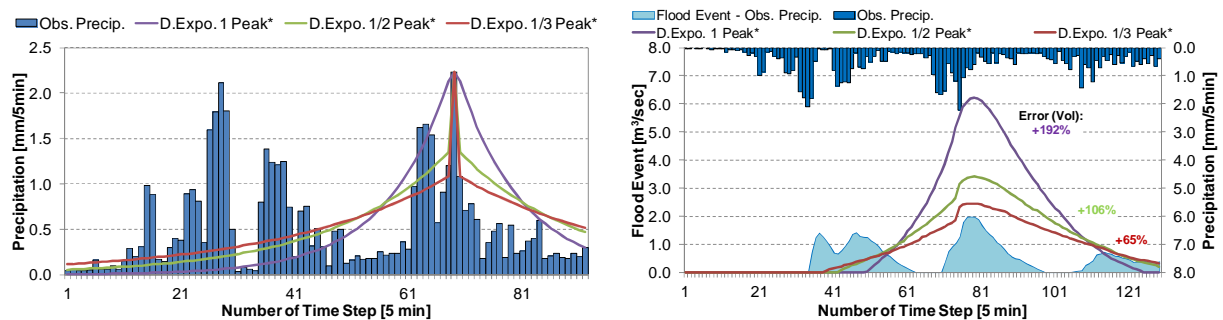
The errors representing the natural variability of the stochastic process are included later when different precipitation models are evaluated. These errors show to be spread around medians (+4% for overflow, +3% and +7% for flood duration and volume) and show under- as well as overestimation of the resulting values for the different stations. They are later referred as “Resampling Obs.” and are used for assessing the uncertainty introduced by the natural variability of the process and to compare the range of errors provided by the different precipitation models.

## 6.2 EFFECT OF THE INTERNAL DISTRIBUTION IN THE RESPONSE OF URBAN CATCHMENTS

The influence of the shape of synthetic events in the reproduction of flood events is analyzed here. For this purpose the 24 stations applied for developing the single site model are used for modeling flood events using the fictional urban hydrological system. The resulting flood events are then used as a reference for comparing synthetic shapes. Rainfall events causing

flooding are identified from the observed time series. Characteristics describing these events (WSA, WSD, WSI, WSPeak and WSTpeak) are extracted. These characteristics describing real events are then used for fitting theoretical shapes. Rainfall events causing flooding are then replaced by the new events with theoretical internal distributions. The new time series are used for modeling flood events, and these results are compared with the ones from the original observed time series. Three different theoretical shapes are considered, as described in Figure 3.15 (Methods Chapter). All are based on the exponential shape with different concentration of rainfall around the peak.

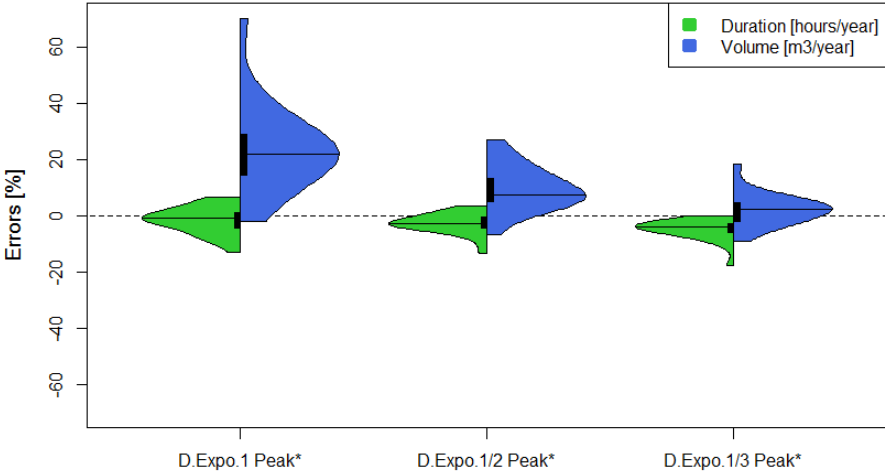
The use of the theoretical shape of the “Old model” (D.Expo.1Peak\*) shows a very low performance in terms of flood events. In particular for the station Braunlage the flood volume shows to be overestimated for more than 80% of the events. To illustrate the case, one of the biggest flood events resulting from observed time series registered in this station is shown in Figure 6.4. The observed rainfall event is used for deriving the three theoretical ones and all are shown in the left plot, whereas the flood events resulting from all these different inputs are shown in the right plot. For the selected event the volume of the flood event is overestimated by 192% (see right plot in Figure 6.4), whereas this overestimation is reduced using the alternative theoretical shapes to 106 and 65%. It can be seen from the figures that the rainfall event is characterized by several peaks that result in multi-peak flooding response.



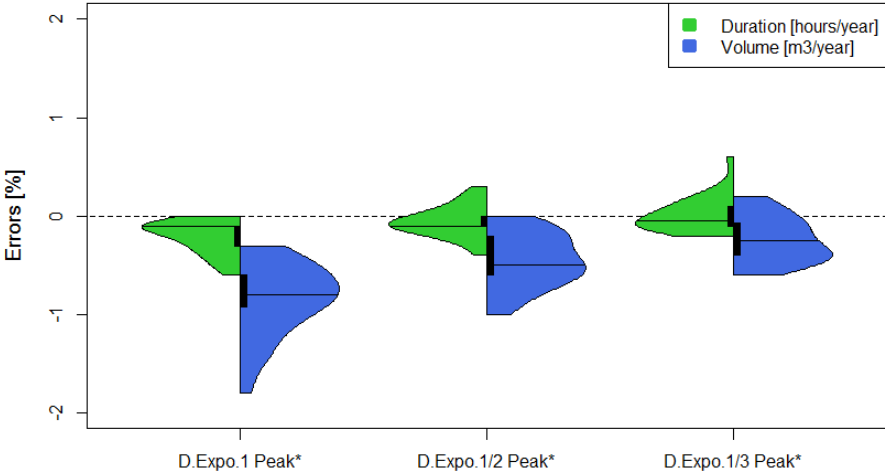
**Figure 6.4: Observed and theoretical rainfall events (left) and resulting flood events (right) for one important flood in Station Braunlage. The numbers [%] indicate the overestimation of flooding volume compared to results from observations.**

As the double exponential theoretical shape is not able to model floods with multiple peaks, results are more accurate when the volume of rainfall is not so concentrated around the peak of the rainfall event delivering flooding events which are more distributed over a longer time period (e.g. case “D:Expo. 1/3 Peak\*”). The comparison of the different theoretical shapes is done based on all rainfall events causing floodings identified from the observed time series of the 24 stations. The results of flood characteristics (mean annual volume and duration) are presented as errors in Figure 6.5 for the three different theoretical shapes. The reduction of errors is significant, especially for the volume characteristic. Absolute errors of volume are lower for 23 out of 24 stations, when the 1 Peak\* and 1/3 Peak\* cases are compared. When

the duration of flood events is compared, more than half of the stations show lower errors when considering the 1/2 or 1/3 Peak\* cases. Nevertheless as the errors resulting from the duration of flood events are much lower than the ones corresponding to the volume, the results from this second variable are considered for deciding which internal distribution to use. Finally, the model that will be used for the internal distribution is the 1/3 Peak\* case. Spilling to a natural river at the outlet of the system is additionally analyzed (both volume and duration). Absolute errors are always lower than 2% for both variables and all analyzed synthetic shapes (see Figure 6.6). The model selected for the internal distribution shows errors closer to a non-biased behavior with median values closer to zero.



**Figure 6.5: Errors of flood events for all stations resulting from different internal distributions based on observed events.**



**Figure 6.6: Errors of overflow events for all stations resulting from different internal distributions based on observed events.**

The fictional urban model used here has the advantage that allows performing a continuous simulation and thus including the complete rainfall time series along with all flood events. The model used is small, simple and meant to represent a portion of a city. The effect of the size of urban model is briefly analyzed by simulating some of the biggest events causing flooding



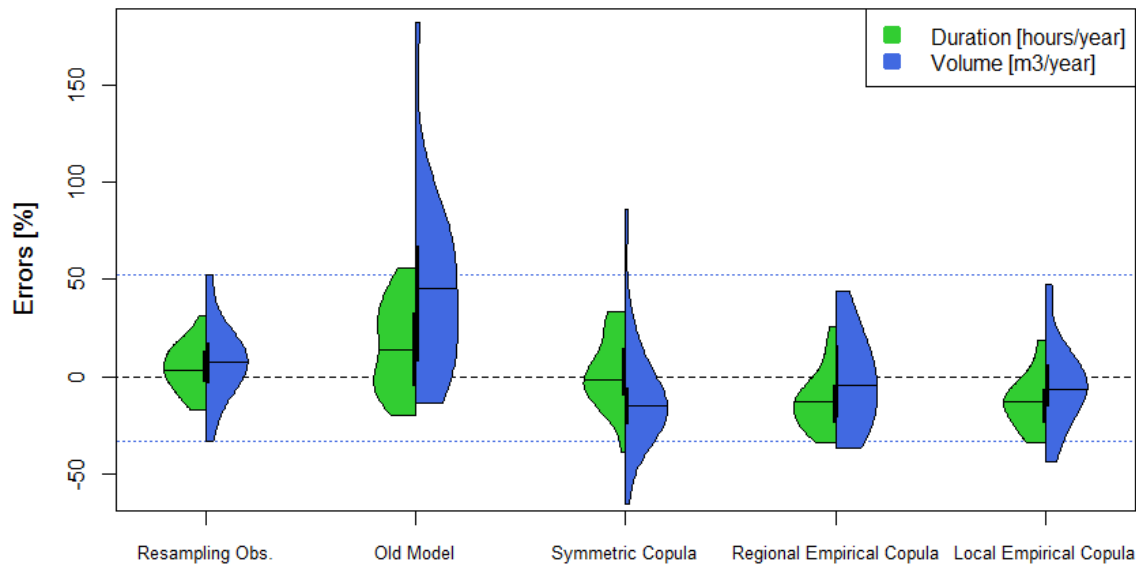
for the Braunlage station using a more complex system and comparing the errors resulting from the different theoretical shapes considered here. The comparison of results from both the simple and complex model is presented in the Appendix J. The same tendency is seen for all events when the different synthetic shapes are compared. The overall errors are much lower for the complex system.

### **6.3 SINGLE SITE PRECIPITATION MODEL**

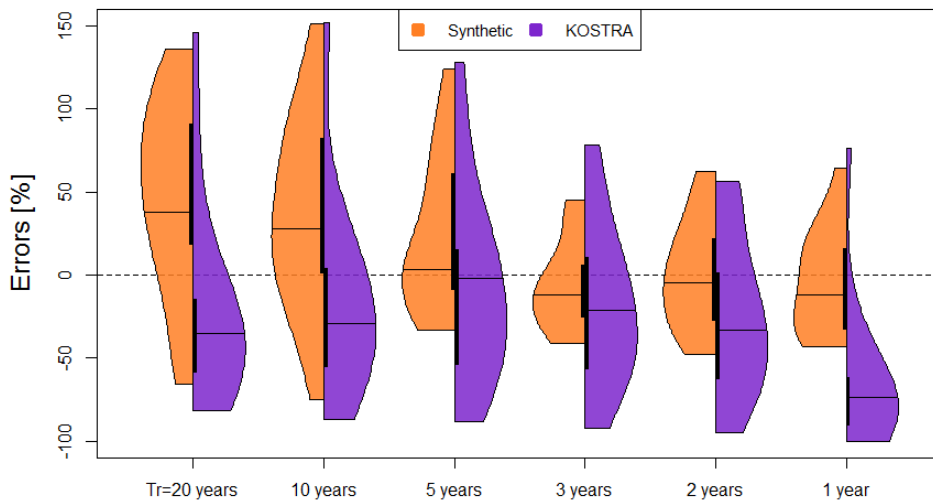
The different single site models presented in sub-Chapter 5.1 are used for generating long synthetic time series which are applied as input for urban hydrological modeling. The urban model is run with a 5 minute time step and therefore the internal precipitation structure plays an important role. Results are presented in terms of errors for the 24 stations and shown in Figure 6.7. The uncertainty introduced by the natural variability of the process (“Resampling Obs.”) is as well included in the plot, which show errors that go up to  $\pm 25\%$  for flood durations and  $-35/+50\%$  for volume (dotted lines), just from randomly sampling events from observed time series.

The “Old Model” results in overestimation of flood event characteristics, especially for the volume. This overestimation results from the internal structure of the model, for which the volume of rainfall is concentrated around the peak intensity of the events. These errors are reduced with the new proposed techniques. The range of errors is smaller when the symmetric copula is replaced by the regional and local empirical copulas. For these two copulas mean values of the error of simulated volume is close to zero indicating a balance between over and under estimation for the different stations. The duration shows more cases in which this variable is underestimated, but the range of errors is smaller compared to the volume error range. The benefit of using the copulas is shown in the plots, as errors can be reduced almost up to the natural variability of the observations.

Figure 6.8 shows the errors for the volume of flood events resulting from synthetic series (median of 10 realizations simulated with the “Local Empirical Copula”) and KOSTRA events (median of 9 cases, see sub-Chapter 3.6.4). For KOSTRA, the duration must be defined therefore concentration time of the system is estimated based on different rainfall hyetographs and runoff hydrographs as around 2 hours. Synthetic time series show on average overestimation of flood volume for return periods of 10 and 20 years, whereas errors are close to zero for the rest of the cases. KOSTRA events with duration of 2 hours show on average underestimation for all return periods. Synthetic time series show smaller range of errors for all return period, especially for low ones. Overall, the precipitation model performs more robust compared to the design practice KOSTRA, as was the case for the IDF curves (see sub-Chapter 5.2).

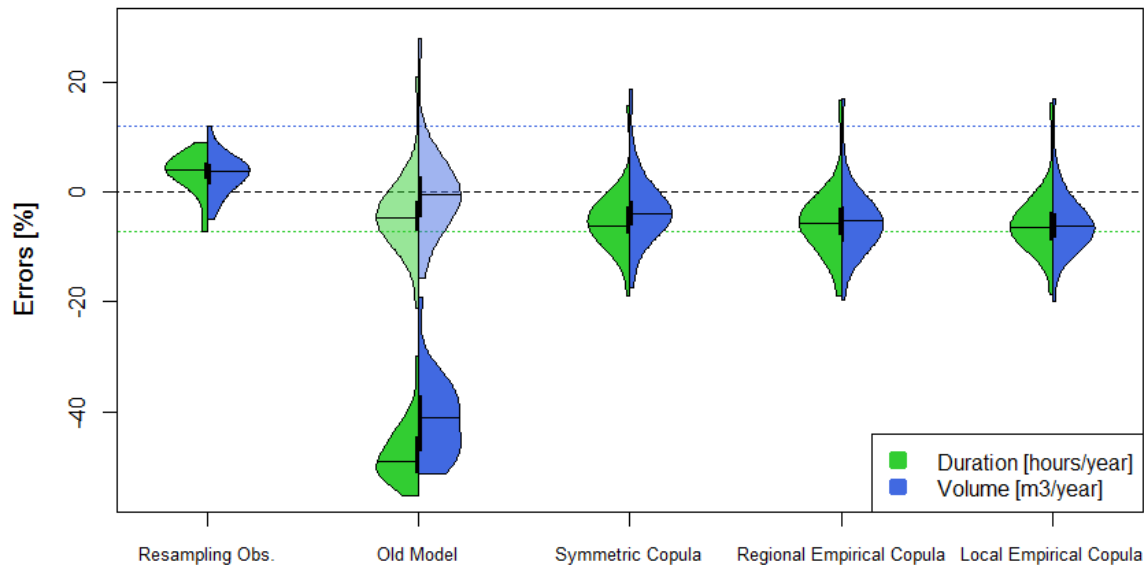


**Figure 6.7: Validation of precipitation model based on errors of flood events for all stations. Dotted lines indicate maximum and minimum error of flooding volume for resampling observations.**



**Figure 6.8: Comparison of performance from precipitation model and KOSTRA to reproduce volume of flood events.**

The ability of the precipitation model to reproduce the average yearly overflow volume and duration is evaluated and results are presented in Figure 6.9. Overflow results are less sensitive to the natural variability of rainfall events showing errors between -7 and + 12% (see “Resampling Obs.”).



**Figure 6.9: Validation of precipitation model based on errors of overflowing events for urban system with T20 and all stations. For the “Old Model” an additional variant is considered (semi-transparent plots) in which small events are included.**

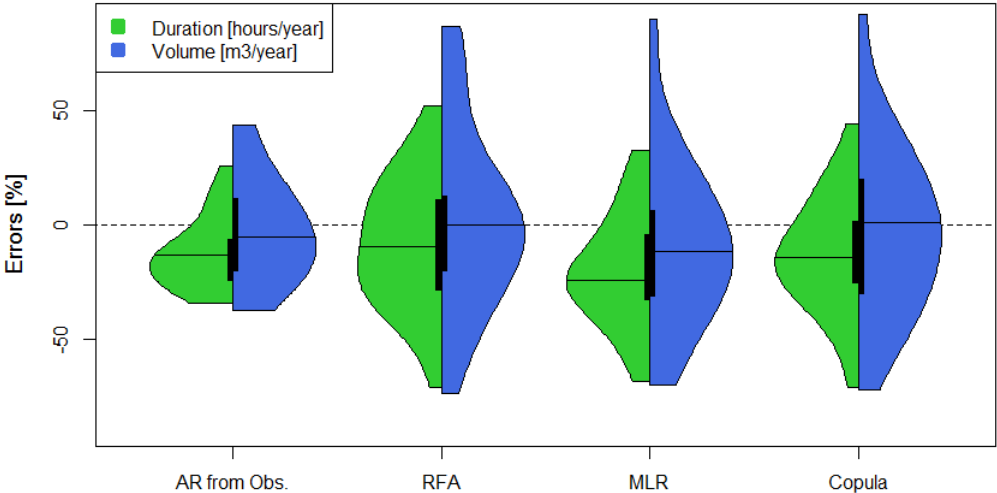
As small events are neglected in the “Old Model”, overflow characteristics are underestimated for all stations, these results are improved with the inclusion of such events (semi-transparent plots). Overflow results do not show to be sensitive to the internal distribution of events and neither to the joint modeling of WSA-WSD as extremes do not play a pivotal role. On the other hand, they are very sensitive to the occurrence of small events. Errors show on average a slight underestimation of duration and volume that could result from the lack of serial autocorrelation in the time series of rainfall events, as was the case of source of error introduced by the presence of autocorrelation (see Figure 6.3). The range of errors increases slightly with increase of retention tank size from T20 to T40 (not shown here). As mentioned before, event based simulations (KOSTRA) cannot be used for assessing these long-term properties.

## 6.4 REGIONALIZATION OF PRECIPITATION

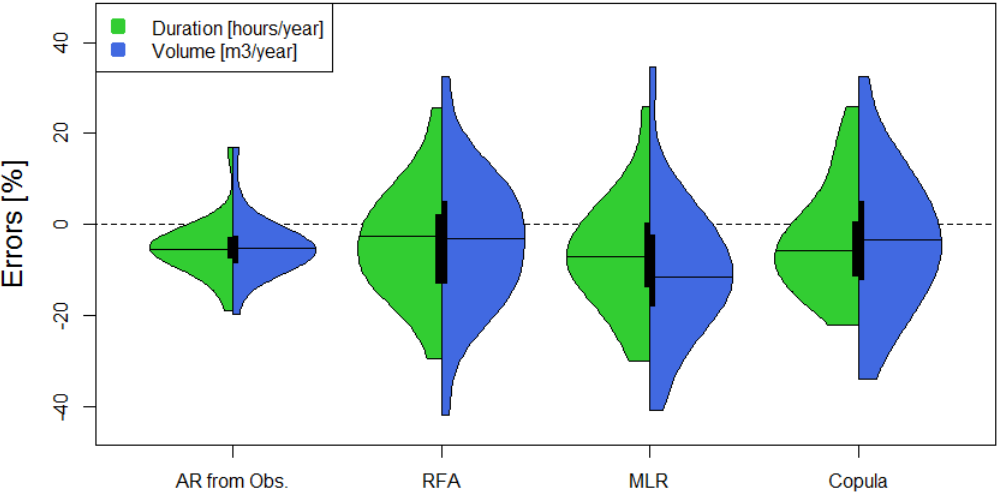
The regionalization of the precipitation model is evaluated in terms of applicability to urban hydrological purposes. A similar approach as for assessing the single site model is performed here. The only difference is that one of the stations (Braunlage) is eliminated from the analysis, as it is not involved in the regionalization procedure. The results from the different regionalization methods are compared with the ones from the original single site model “AR from Obs.” which corresponds to the “Regional Empirical Copula” case presented in the preceding Sub-chapter (with the exclusion of Braunlage). Errors are computed based on different flooding and overflow characteristics and presented in Figure 6.10 and 6.11.

As expected the range of errors resulting from different regionalization methods are bigger than the one resulting from the original model without regionalization. Nevertheless, resulting

values are still acceptable. Multi-linear regression (MLR) tends to underestimate all resulting values for most of the stations, as the median values are negative. On the other hand Regional Frequency Analysis (RFA) and Copula based regionalization techniques show better results, especially for the volumes of flooding and overflow, for which the errors are on average close to zero. This indicates a balance between over- and underestimation for the different stations. A comparison between RFA and Copula based models indicates that the volumes have a similar range of errors for flooding and overflow, whereas the durations are barely better reproduced by the copula based regionalization, as the range of errors is smaller.



**Figure 6.10: Validation of regionalization of the precipitation model based on errors of flood events for stations with long registers.**



**Figure 6.11: Validation of regionalization of precipitation model based on errors of overflowing events for urban system with T20 and stations with long registers.**

## 7 SYNTHESIS

### 7.1 SUMMARY AND CONCLUSIONS

The goal of this work was to develop a stochastic model that can properly reproduce simultaneously mean event properties and extreme value statistics of rainfall required for urban modeling. The model is based on Alternating Renewal process and derived from an existing one developed by Haberlandt (1996) by introducing major improvements to properly synthesize rainfall in a high temporal resolution. The proposed model is applied for single site simulations in locations with observations and a methodology for regionalizing the model is as well presented. Furthermore an extension of the existing model from single site to multi-site applications is developed. The proposed methodologies are compared with alternative ones applied by the hydrological community for similar purposes. An additional validation of the model is performed by using a fictional drainage system to model the response of an urban network. The different developments are tested for different stations in Germany in which registers of rainfall in a 5 minutes resolution are available. A comparison with a common practice design strategy, in this case KOSTRA, is as well included in order to discuss the advantages and limitations of considering a range of possible events into the analysis. The main findings are discussed in the following paragraphs.

The single site model involves the temporal simulation of long continuous precipitation for single stations. The proposed model is compared with the existing one, here denominated as “Old Model”. The major improvements introduced into the model are: i) separation of events into two seasons, ii) inclusion of small events, iii) use of copula functions to jointly model variables involved in the external and internal structures, and iv) modification of the shape for the internal structure. The main results are the following:

- Introducing seasonality into the model, i.e. analyzing events for summer and winter separately, shows to be very important for reproducing the distribution of event characteristic, especially for the variables describing the durations, as shown in Figure 5.4. Seasonality was neglected in the “Old Model”.
- Small events account for an important volume (19% to 37%) of annual rainfall and therefore their inclusion is suitable for properly modeling the total amount of rainfall. Errors are significantly reduced with the updated models which introduce these small events as shown in Figure 5.6.
- The use of copulas to model WSD and WSA jointly showed to be propitious. Direct and simultaneous simulation of WSA and WSD and successive estimation of WSI is benefited by

the bivariate models. Synthetic series show to be similar to observed ones for most of the variables and stations; the proposed models which include copulas have a superior performance compared to the “Old Model” (see Figure 5.4 and Tables 5.2 and 5.3).

- The effect of incorporating appropriate dependence structures to model WSA-WSD is clear when the extreme values are evaluated (see Figure 5.9). These events show to be better reproduced when the copulas able to model asymmetric structures are incorporated into the model, i.e. the empirical copulas. As pointed out by Vandenberghe *et al.* (2010), even the use of some asymmetric copulas could be problematic when extreme storms are to be simulated, due to the fact that the degree of asymmetry is non-homogeneous (changes with the probability). The use of empirical models proposed here is able to capture the particular behaviour of non-exchangeability.

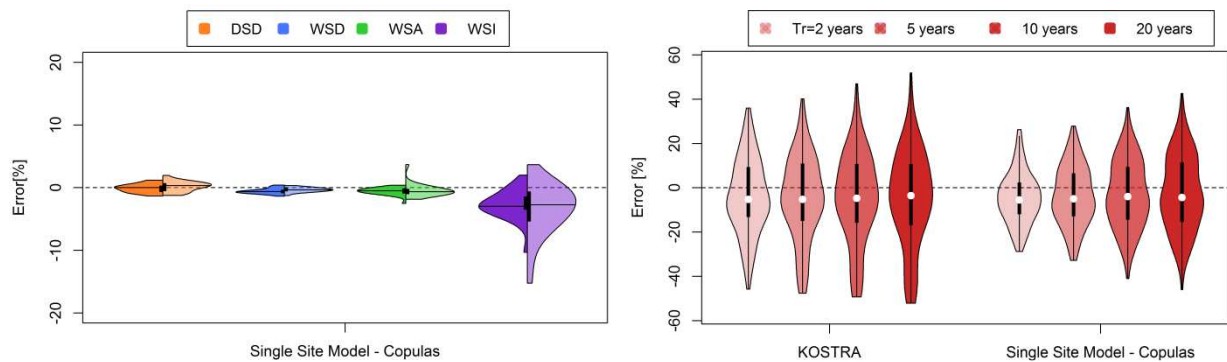
- When IDF curves are evaluated, the proposed precipitation model shows on average underestimation of extreme events for durations up to 30~60 minutes and all return periods. Results improve for longer durations as errors and ranges are reduced as shown in Figure 5.10. Nevertheless errors corresponding to short durations and high return periods are on average lower than the ones reported by Verhoest *et al.* (1997), Bernadara *et al.* (2007) and Vernieuwe *et al.* (2015). The precipitation model outperforms KOSTRA design storms only for durations higher than 60 minutes.

- The transferability of the model to new regions is tested for a set of stations belonging to a region with very different topographic and climatic conditions. Despite of this challenging feature, the transferability was straightforward and shows acceptable results as can be depicted from Figure 5.11, both for all event characteristics and for extreme values.

- Two different sources of uncertainty are analyzed, namely the natural variability of the stochastic process and the uncertainty of parameter estimation. The analysis of the first source of uncertainty depicts the models ability to reproduce the variables DSD, WSD and WSA, whereas there is a tendency to underestimate the WSI. For the extreme events, the model shows to overestimate intensities, especially for short durations, whereas for durations of 30 and 60 minutes the results are more acceptable, as the behavior is closer to unbiased. In terms of uncertainty of model parameters, the model shows to be very robust, as the general behavior is not affected by the percentage of available input data (see Figures 5.32 and 5.33).

The plots in Figure 7.1 summarize the main outputs of the proposed copula based single site model, both in terms of reproducing the mean value of all event characteristics

(summer/winter: left/right semi violins) and extreme events corresponding to durations of 5 minutes, 15 minutes, 30 minutes, 1 hour, 2 hours, 4 hours, 6 hours, 12 hours and 1 day. Both plots demonstrate the ability of the model to mimic different events and its overall outperformance compared with the design practice KOSTRA depicted from the lower range of errors. These errors correspond to the “Local Empirical Copula” applied to the stations used for developing the single site model.



**Figure 7.1: Summary of results of the proposed model for single site synthesis in terms of mean value of all event characteristics (left) and extreme values (right).**

The multi-site model involves the temporal and spatial simulation of long continuous precipitation for several stations simultaneously. The developed method consists of an extension of the Alternating Renewal based process model to a multi-site application. Therefore it is event based and benefits from the fact that the number of parameters is not affected by the temporal resolution. It involves several steps of calculation that are sequential, namely the hybrid model, Vine copulas and incorporation of dry spells. The proposed model has the advantage that it accounts for the spatial extension of events within the hybrid model for simulating rainfall characteristics; however the several stages of simulation make it complex and difficult to interpret. The multi-site model is compared to the alternative Simulated Annealing (SA) method. The main results are the following:

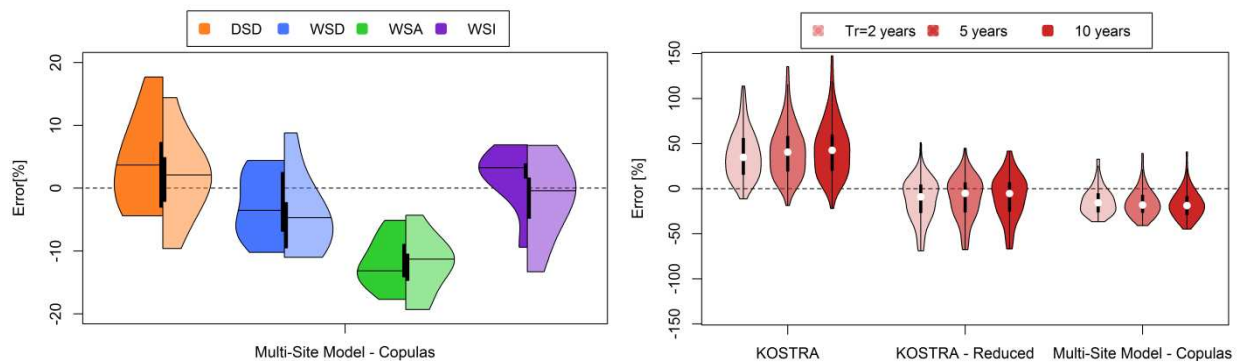
- Synthetic time series are compared in a pair-wise way, i.e. every two stations, to estimate spatial consistency measures. Overall, these measures (as shown in Figure 5.17) are better reproduced by the Vine copula model compared to the alternative method. Furthermore, these measures are not involved within the Vine copula model, as is the case for the SA optimization criteria, so these outcomes emphasize the satisfactory performance of the developed model.
- The capability of the two models to reproduce different event characteristics is additionally compared. The evaluation based on single sites depicts some deficiencies of the proposed model, however these results are improved when the evaluation is performed based on areal precipitation (see Figure 5.21). Overall SA performs better for single stations but fails when the areal precipitation is considered. These results suggest that the Vine copula model

outperforms in reproducing the simultaneous occurrence of rainfall in several stations, whereas the SA misses to mimic this behavior. Unfortunately, the proposed method systematically underestimates total seasonal rainfall as can be depicted from Figure 5.19.

- When IDF curves are evaluated, the proposed precipitation model shows acceptable results as the ranges of errors are smaller than the ones from the SA and KOSTRA (see Figures 5.22 and 5.23). Vine copulas outperforms SA, however these models would be preferable for low durations and KOSTRA for long ones. The reduction factors applied to KOSTRA are provided for different durations and areas of application, though for large areas extrapolation by simple linear regressions was necessary. These factors could be revised, as they result in a systematic underestimation of extreme values. For example for similar sizes of areas of application and durations, Myers & Zehr (1980) provide reduction factors developed for the city of Chicago that indicate a good agreement for durations shorter than 1 hour but these factors are higher (7% to 18%) than the ones applied in this Thesis for longer durations. Additionally the authors provide different factors for different return periods, and for high return periods (100 years), KOSTRA reduction factors show to be higher (10%) than the provided ones for durations lower than 6 hours. The inclusion of the return period in the estimation of the KOSTRA reduction factors could lead to better results.

The plots in Figure 7.2 summarize the general performance of the proposed copula based multi-site model, both in terms of reproducing the mean value of all event characteristics (summer/winter: left/right semi violins) and extreme events for durations ranging from 5 minutes to 1 day. The results are based on areal precipitation and extreme values are compared with average values estimated from KOSTRA. The plot on the left demonstrates the ability and difficulty to mimic the event characteristics for several stations simultaneously, as some of these variables are well reproduced, whereas others like the WSA are systematically underestimated. The plot on the right shows the ability of the method to reproduce extreme values and its outperformance compared with the design practice KOSTRA, despite the underestimation for most of the stations the model delivers a smaller range of errors.



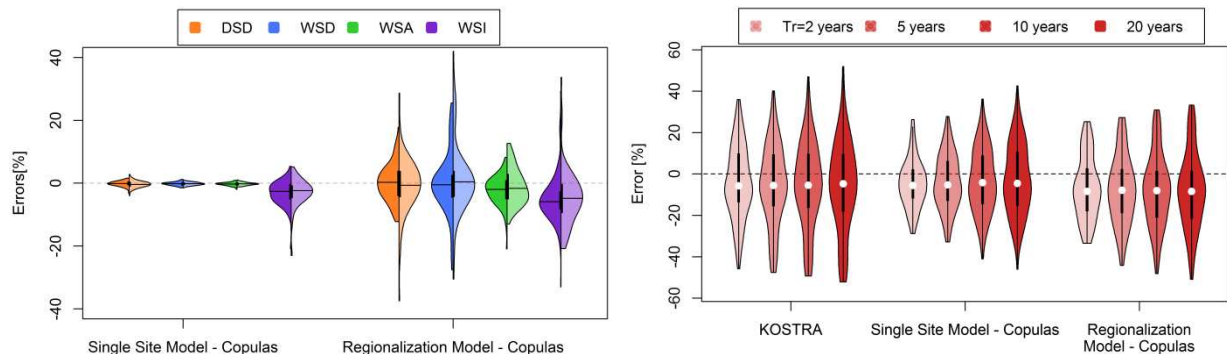


**Figure 7.2: Summary of results of the proposed model for multi-site synthesis in terms of mean value of all event characteristics (left) and extreme values (right).**

The regionalization of the model involves the temporal simulation of long continuous precipitation for single stations without observations. A methodology is developed that involves the use of several site descriptors that are available for any location. Copula models, conditioned to the descriptors, are used to estimate possible values of variables describing rainfall events. The variables are thereafter useful for estimating the model parameters. The proposed methodology has the advantages that it is very fast to implement and easy to interpret. The copula-based method is compared with alternative regionalization methods commonly applied by the hydrological community, namely multi-linear regression (MLR) and regional frequency analysis (RFA). The main results are the following:

- Different evaluations based on all event characteristics and extreme values are performed for the three methods. Overall the copula based model would be preferred over the other two methods, although none of the methods is clearly outperforming the others.
- Regionalization depicts the difficulty of reproducing all event characteristics, whereas extreme values show a better performance compared to non-regionalized results (see Figures 5.26 and 5.27). The three methods considered here point the same conclusion.
- The use of copulas for regionalization clearly outperforms the alternative methods in terms of robustness to new stations as depicted from Figure 5.29. As the copula is derived from ranks or pseudo-values, rather than values itself, the estimation of the models show to be robust to the inclusion/exclusion of some stations. Furthermore, RFA method appears to be very sensitive to the addition of more stations.
- IDF's resulting from the proposed copula-based regionalization method are compared with KOSTRA in Figure 5.30. Ranges of errors indicate a slightly more robust performance of regionalized time series compared to KOSTRA. Regionalization shows on average underestimation of extreme events, except for durations of 4 to 12 hours for which median of errors are close to zero. Regionalized time series outperform KOSTRA design storms for the mentioned durations, for other durations both data sets are similar.

The plots in the following Figure 7.3 summarize the outputs of the proposed copula based regionalization method, both in terms of reproducing the mean value of all event characteristics (summer/winter: left/right semi violins) and extreme events for durations ranging from 5 minutes to 1 day. Regionalization is compared with the original model, i.e. without regionalization. These plots demonstrate the difficulty to regionalize all events and the ability of the method to regionalize extreme values.



**Figure 7.3: Summary of results of the proposed model for regionalization in terms of mean value of all event characteristics (left) and extreme values (right).**

The application of the synthetic time series as input for urban hydrological modeling has been advantageous to assess the implications of some of the proposed techniques. A simple fictional drainage system has been used for this purpose allowing for continuous simulations. An overall assessment is therefore possible since the whole time series of rainfall and flood events are included in the analysis. The main findings are described here:

- As exposed by Smith *et al.* (2013) temporal clustering of flood events is an important element in urban hydrological systems, as many flood events occur within short periods or as response of long rainfall events. Nevertheless, in this Thesis the significant autocorrelation of duration of events did not show to have a relevant effect on the flood events. As clustering of events can induce temporal autocorrelation, a systematic underestimation of flood characteristics was expected by removal of the autocorrelation. However, this is not the case and the errors showed a balance between under- and overestimation that could be explained as a result from the natural variability of rainfall events or just as a sampling error (see Figure 6.2). On the other hand, overflow events showed to be very sensitive to the temporal occurrence of rainfall events, as the removal of the autocorrelation resulted in a systematic underestimation of both duration and volume (see Figure 6.3).
- Uncertainty of rainfall data used as input for urban hydrological simulations shows to have a stronger effect on the variables describing flood events, especially the volume which shows the largest range of errors. Overflow characteristics result in much lower errors.

- The internal structure of the model plays a pivotal role both in terms of properly reproducing the flood and overflow events. The volume of these events is noticeably improved (see Figures 6.5 and 6.6) when the new internal structure is used which provides precipitation profiles with lower concentrations around the peak.
- Regarding the single site model, the hydrological implication of using different dependence structures to model WSA-WSD is assessed by the response urban hydrological. The effect of considering the asymmetry is clear when the flood events are evaluated (see Figure 6.7), in which both volume and duration show to be better reproduced with the use of the empirical copulas. This hydrological interpretation is, according to Vandenberghe *et al.* (2010), a challenge that deserves further research.
- For more localized structures, i.e. the Local Empirical Copula, the range of errors is reduced, in particular for the volume and duration of flood events (see Figure 6.7). Nevertheless the Regional Empirical Copula performs in an acceptable way, suggesting that the joint structure of WSA-WSD can properly be modeled by a unique Copula for all stations. The use of one regional copula provides an important benefit, since the joint structure does not need to be regionalized, hence omitting the uncertainties that would be introduced during this process. Similar results were found by Balistocchi & Bacchi (2011) for 3 long time series recorded in different climate regions in Italy, for which the strength of dependence between volume and duration shows to be independent of the rain gauge location.
- A comparison between the single site model and KOSTRA depicts that, on average, KOSTRA underestimates flood volumes, whereas synthetic time series overestimate these volumes for high return periods and results are improved as the return period decreases. Nevertheless, the proposed precipitation model shows more robust results as the range of errors both for rainfall and flood volume are smaller than KOSTRA, especially for low return periods (see Figures 5.10 and 6.8). Furthermore, as KOSTRA is derived from observations its evaluation is consequently comparing data to data, whereas the synthetic time series consist of a data to model comparison. The performance of the precipitation model is acceptable compared to the design practice with KOSTRA.
- Rainfall-runoff transformation shows a non-linear behavior. If a duration equal to the concentration time of 2 hours is considered, KOSTRA results in a proper estimation of rainfall however the flooded volume is underestimated as depicted from Figures 5.10 and 6.8. Results are more consistent for synthetic time series suggesting an advantage of continuous over event based modeling.
- Overflow behavior does not show to be sensitive to the joint modeling of WSA-WSD and the inclusion of seasonality into the model; whereas the modeling of small events is pivotal,

as was stated by Grodek *et al.* (2011). The inclusion of small events reduces the errors significantly as shown in Figure 6.9. However these errors show on average a slight underestimation that could result from the lack of serial autocorrelation in the time series of rainfall events, nevertheless the underestimation is not systematic as some stations result in overestimation. A systematic underestimation of overflow duration is reported by Haberlandt (1996). KOSTRA is not suitable for the continuous simulation of the overflow dynamics.

- The regionalization of the model results as expected in bigger range of errors, both for flooding and overflow events, compared to the original one (see Figures 6.10 and 6.11). MLR tends to underestimate all resulting values for most of the stations. A comparison between RFA and the copula based model indicates that the volumes have a similar range of errors for flooding and overflow, whereas the durations are barely better reproduced by the copula method, as the range of errors is smaller. Furthermore, the errors of overflow volume are in the range to the ones reported by Cowpertwait *et al.* (1996), which for small steep catchments range between -28% and +25%.

Overall the different proposed modifications to the model show to deliver satisfactory results both for the average event statistics and for the extremes. Compared to other existing models commonly applied by the hydrological community, the copula-based methods proposed in this Thesis have shown to perform very satisfactory in the simulation of rainfall events for high temporal resolutions. Furthermore, the model has several advantages: it is fast to be set up, simple to interpret, easy to transfer to other regions and the number of parameters is not affected by the temporal resolution, which is propitious for its regionalization. The provided model outperforms the available design practices for particular events; moreover it shows a robust behavior both for extreme value estimations and for flooding and overflow resulting from urban hydrological simulations.

## 7.2 FUTURE RESEARCH

Time series of rainfall in a high temporal resolution are very useful for several purposes. Working with this type of data leads to some open questions that are involved in different stages of analysis presented here. One important aspect that was pointed out during this study is regarding the autocorrelation of time series in a 5 minutes temporal resolution. It was shown that the internal structure of the model results in an overestimation of this autocorrelation. A possibility to improve these results could be to introduce some noise to the final time series and therefore reduce the autocorrelation, whilst conserving other aspects of rainfall events like duration and volume. The open question is to explore how this autocorrelation affects the final application of the precipitation model, that could be the quantification of rainfall erosivity, hydrological applications involving urban or small, steep rural catchments, engineering design of flash flood control structures, etc.

Regarding the multi-site model, one limitation of the existing models is that they are complex and hard to interpret, involving several stages of simulation. The model proposed in this Thesis consists as well of several steps that follow each other and is not easy to interpret. Furthermore, the model was only applied to cases involving 2 and 3 stations. Despite the fact that it could be applied to data sets involving more stations, a limitation could be the length of data available in several stations simultaneously for setting up the different modules within the model, especially the Vine copulas. Another important aspect that is subject of further research is the regionalization of this multi-site model, i.e. setting up the model for locations without observations. Future work could involve exploring the capability of the radar data, which is spatially and temporally available in very high resolutions, to extract some of the characteristic which are required for modeling rainfall in multiple sites simultaneously. For instance the radar data could be used for estimating the Vine copulas, which are based on pseudo-observations, or deriving the different bias corrections.

Finally the application of the precipitation model to simulate precipitation for a future scenario could be as well of interest. As the precipitation events are characterized by probability distributions which result from fitting the parameters to LMs describing the events characteristics, a methodology for estimating the LMs for a future climate scenario would be useful. The methodology could follow the regionalization idea proposed here, but using climatic SDs for a future scenario, resulting from a global or regional climate model, and providing estimated LMs for a future condition.

## REFERENCES

- Aas, K., Czado, C., Frigessi, A., Bakken, H., 2009. Pair-copula constructions of multiple dependence. *Insurance, Mathematics and Economics* 44, 182-198.
- Adler, D., 2015. vioplot: Violin plot [online]. R package version 0.2. Available from: <https://cran.r-project.org/web/packages/vioplot>. [Accessed 30 March 2016].
- Alila, Y., 1999. A hierarchical approach for the regionalization of precipitation annual maxima in Canada. *Journal of Geophysical Research*, 104(D24), 31.645-31.655.
- Alfieri, L., Salamon, P., Bianchi, A., Neal, J., Bates, P., Feyen, L., 2014. Advances in pan-European flood hazard mapping. *Hydrological Processes*, 28: 4067-4077. doi: 10.1002/hyp.9947.
- Ariff, N. M., Jemain, A. A., Ibrahim, K., Wan Zin W. Z., 2012. IDF relationship using bivariate copula for storm events in Peninsular Malaysia. *Journal of Hydrology*, 470-471, 158–171.
- Arnaud, P., Bouvier, C., Cisneros, L., Dominguez, R., 2002. Influence of rainfall spatial variability on flood prediction. *Journal of Hydrology*, 260, 216-230.
- Asong, Z. E., Khaliq, M. N., Wheeler, H. S., 2015. Regionalization of precipitation characteristics in the Canadian Prairie Provinces using large-scale atmospheric covariates and geophysical attributes. *Stochastic Environmental Research and Risk Assessment*, 29: 875-892.
- Asquith, W., 2016. Imomco---L-moments, censored L-moments, trimmed L-moments, L-comoments, and many distributions. R package version 2.2.3, Texas Tech University, Lubbock, Texas.
- Azzalini, A., 2015. The R package 'sn': The Skew-Normal and Skew-t distributions (version 1.3-0).
- Bacchi, B., Kottegoda, N. T., 1995. Identification and calibration of spatial correlation patterns of rainfall. *Journal of Hydrology*, 165, 311-348.
- Balistrocchi, M., Bacchi, B., 2011. Modelling the statistical dependence of rainfall event variables through copula functions. *Hydrology and Earth System Sciences*, 15, 1959-1977. (doi:10.5194/hess-15-1959-2011)
- Bárdossy, A., 1998. Generating Precipitation Time Series Using Simulated Annealing. *Water Resources Research*, 34, 1737–1744.
- Barredo, J. I., 2007. Major flood disasters in Europe: 1950-2005. *Natural Hazards*, 42, 125–148. doi:10.1007/s11069-006-9065-2
- Bartels, H., Dietzer, B., Malitz, G., Albrecht, F. M., Guttenberger, J., 2005. Offenbach am Main KOSTRA-DWD-2000, Starkniederschlagshöhen für Deutschland (1951–2000) – Fortschreibungsbericht.
- Beck, F., 2013. Generation of spatially correlated synthetic rainfall time series in high temporal resolution: a data driven approach. Ph.D. Thesis, Universität Stuttgart, Holzgartenstr. 16, 70174 Stuttgart, ISBN 9783942036238.
- Bedford, T., Cooke, R. M., 2001. Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence*, 32, 245-268.

- Bernard, C., Czado, C., 2015. Conditional quantiles and tail dependence. *Journal of Multivariate Analysis*, 138, 104-126.
- Bernardara, P., De Michele, C., Rosso, R., 2007. A simple model of rain in time: An alternating renewal process of wet and dry states with a fractional (non-Gaussian) rain intensity. *Atmospheric Research*, 84, 4, 291-301.
- Brechmann, E. C., Czado, C., Aas, K., 2012. Truncated regular vines in high dimensions with applications to financial data. *Canadian Journal of Statistics*, 40 (1), 68-85.
- Brechmann, E. C., Schepsmeier, U., 2013. Modeling Dependence with C- and D-Vine Copulas: The R Package CDVine. *Journal of Statistical Software*, 52(3), 1-27.
- Breiman, L., 2001. Random Forests. *Machine Learning*, 45(1), 5-32.
- Burton, A., Fowler, H. J., Kilsby, C. G., O'Connell, P. E., 2008. RainSim: a spatial-temporal stochastic rainfall modelling system. *Environmental Modelling and Software*, 1356-1369.
- Carr, D., Lewin-Koh, N., Maechler, M., Deepayan, S., 2015. hexbin: Hexagonal Binning Routines. R package version 1.27.1. <https://CRAN.R-project.org/package=hexbin> [Accessed 15 December 2016]
- Chowdhary, H., Escobar, L. A., Singh, V. P., 2011. Identification of suitable copulas for bivariate frequency analysis of flood peak and flood volume data. *Hydrology Research*, Vol. 42 No. 2-3, 193-216.
- Cohen, J., Cohen, P., West, S.G., Aiken, L.S., 2003. *Applied multiple regression/correlation analysis for the behavioral sciences*. L. Erlbaum Associates, Mahwah, N.J.
- Cowpertwait, P. S. P., O'Connell, P. E., Metcalfe, A. V., Mawdsley, J. A., 1996. Stochastic point process modelling of rainfall. II. Regionalisation and disaggregation. *Journal of Hydrology*, 175, 47-65.
- Cowpertwait, P. S. P., Kilsby, C., O'Connell, P., 2002. A space-time Neyman-Scott model of rainfall: Empirical analysis of extremes. *Water Resources Research*, 38(8), doi:10.1029/2001WR000709.
- Cowpertwait, P. S. P., Lockie, T., Davis, M. D., 2004. A stochastic spatial-temporal disaggregation model for rainfall. *Research Letters in the Information and Mathematical Sciences*, 6, 109-122.
- Cowpertwait, P. S. P., Isham, V., Onof, C., 2007. Point process models of rainfall: developments for fine-scale structure. *Proceeding of Royal Society A*, 463, 2569-2587. (doi: 10.1098/rspa.2007.1889)
- Cowpertwait, P. S. P., 2011. A regionalization method based on a cluster probability model. *Water Resources Research*, 47: W11525.
- Cox, D. R., Wermuth, N., 1996. *Multivariate dependencies*. London: Chapman & Hall.
- CRED, 2017. EM-DAT. The international disaster database. Available from: [www.emdat.be](http://www.emdat.be). [Accessed on August 2017].
- Czado, C., 2010. Pair copula constructions of multivariate copulas. In P. Jaworki, F. Durante, W. Härdle, and T. Rychlik (Eds.), *Workshop on Copula Theory and its Applications*. Springer-Verlag Berlin Heidelberg 2010.
- Delignette-Muller, M.L., Pouillot, R., Denis, J.B., Dutang, C., 2010. fitdistrplus: help to fit of a parametric distribution to non-censored or censored data [online]. R package. Available from: <http://cran.r-project.org/web/packages/fitdistrplus/index.html> [Accessed 30 March 2012].

- Del Giudice, D., Honti, M., Scheidegger, A., Albert, C., Reichert, P., Rieckermann, J., 2013. Improving uncertainty estimation in urban hydrological modeling by statistically describing bias. *Hydrology and Earth System Sciences*, 17, 4209-4225. (doi:10.5194/hess-17-4209-2013)
- Dißmann, J. F., Brechmann, E. C., Czado, C., Kurowicka, D., 2013. Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis*, 59 (1), 52-69.
- Dupuis, D. J., 2007. Using Copulas in Hydrology: Benefits, Cautions and Issues. *Journal of Hydrological Engineering*, 12, 381-393.
- DWA, 2006. Hydraulische Bemessung und Nachweis von Entwässerungssystemen, Arbeitsblatt A 118, Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall e. V., Germany.
- DWA, 2012. Starkregen in Abhängigkeit von Wiederkehrzeit und Dauer, Arbeitsblatt A 531, Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall e. V., Germany.
- DWD, 2015. Climate Data Center. [Online] Available at: <ftp://ftp-cdc.dwd.de/pub/CDC/> [Accessed 01 - 11 - 2016].
- DWD, 2016. Datensatzbeschreibung. Vieljährliche Raster des mittleren Vegetationsbeginns in Deutschland. [Online] Available at: [ftp://ftp-cdc.dwd.de/pub/CDC/grids\\_germany/multi\\_annual/vegetation\\_begin/BESCHREIBUNG\\_gridsgermany\\_multi\\_annual\\_vegetation\\_begin\\_de.pdf](ftp://ftp-cdc.dwd.de/pub/CDC/grids_germany/multi_annual/vegetation_begin/BESCHREIBUNG_gridsgermany_multi_annual_vegetation_begin_de.pdf) [Accessed 02 - 11 - 2016].
- Eggert, B. , Berg, P. , Haerter, J. O. , Jacob, D. , Moseley, C., 2015. Temporal and spatial scaling impacts on extreme precipitation. *Atmospheric Chemistry and Physics*, 15, 5957–5971.
- Fagerland, M. W., Sandvik, L., 2009. The Wilcoxon–Mann–Whitney test under scrutiny. *Statistics in Medicine*, 28: 1487–1497. (doi: 10.1002/sim.3561)
- Fraley C., Raftery A. E., 2007. Model-based methods of classification: Using the mclust software in chemometrics. *Journal of Statistical Software*, 18, paper i06.
- Gaál, L., Molnar, P., Szolgay, J., 2014. Selection of intense rainfall events based on intensity thresholds and lightning data in Switzerland. *Hydrology and Earth System Sciences*, 18, 1561–1573.
- Garcia-Guzman, A., Aranda-Oliver, E., 1993. A stochastic model of dimensionless hyetograph. *Water Resources Research*, 29, 2363-2370.
- Genest, C., Favre, A., 2007. Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask. *Journal of Hydrological Engineering*, 12, 347-368.
- Genest, C., Nešlehová, J. G., 2013. Assessing and modeling asymmetry in bivariate continuous data. In *Copulae in Mathematical and Quantitative Finance. Proceedings of the Workshop Held in Cracow, 10-11 July 2012* (P. Jaworski, F. Durante & W.K. Härdle, Editors). Springer, Berlin, 91-114.
- Grimaldi, S., Serinaldi, F., 2006. Design hyetograph analysis with 3-copula function. *Hydrological Sciences Journal*, 51, 223-238.
- Grimaldi, S., Petroselli, A., Serinaldi, F., 2012. A continuous simulation model for design-hydrograph estimation in small and ungauged watersheds. *Hydrological Sciences Journal*, 57 (6), 1035–1051.
- Grodek, T., Lange, J., Lekach, J., Husary, S., 2011. Urban hydrology in mountainous middle eastern cities. *Hydrology and Earth System Sciences*, 15 (3), 953-966.



- Gyasi-Agyei, Y., Charles, S., 2012. Modelling the dependence and internal structure of storm events for continuous rainfall simulation. *Journal of Hydrology*, 464, 249-261.
- Haberlandt, U., 1996. Stochastische Synthese und Regionalisierung des Niederschlages für Schmutzfrachtberechnungen. PhD Thesis, Universität Stuttgart, Mitteilungen, Heft 88, 1996. ISBN 3921694884.
- Haberlandt, U., 1998. Stochastic rainfall synthesis using regionalized model parameters. *Journal of Hydrologic Engineering*, 3, 160–168.
- Haberlandt, U., Ebner von Eschenbach, A.-D., Buchwald, I., 2008. A space-time hybrid hourly rainfall model for derived flood frequency analysis. *Hydrology and Earth System Sciences*, 12, 1353-1367.
- Haberlandt, U., Radtke, I., 2014. Hydrological model calibration for derived flood frequency analysis using stochastic rainfall and probability distributions of peak flows. *Hydrology and Earth System Sciences*, 18, 353–365.
- Hernaéz, P. F., Martin-Vide, J., 2011. Regionalization of the probability of wet spells and rainfall persistence in the Basque Country (Northern Spain). *International Journal of Climatology*, 32(12), 1909–1920.
- Hintze, J. L., Nelson, R. D., 1998. Violin Plots: A Box Plot-Density Trace Synergism. *The American Statistician*, 52(2), 181-84.
- Hofert, M., Maechler, M., 2011. Nested Archimedean Copulas Meet R: The nacopula Package. *Journal of Statistical Software*, 39(9), 1-20.
- Hofert, M., Kojadinovic, I., Maechler, M., Yan, J., 2015. copula: Multivariate Dependence with Copulas. R package version 0.999-14 URL <http://CRAN.R-project.org/package=copula> [Accessed 15 December 2016].
- Hosking, J. R. M., 1990. L-moments: Analysis and Estimation of Distributions using Linear Combination of Order Statistics. *Journal of the Royal Statistical Society (Series B)*, 52, 105-124
- Hosking, J. R. M., Wallis, J. R., 1997. Regional Frequency Analysis. An Approach Based on L-Moments. Cambridge: Cambridge University Press.
- Hsu, M. H., Chen, S. H., Chang, T. J., 2000. Inundation simulation for urban drainage basin with storm sewer system. *Journal of Hydrology*, 234, 21-37.
- Hu, Y., 2013. Extreme value mixture modelling: An R package and simulation study. MSc (Hons) Thesis, University of Canterbury, New Zealand.
- Hundertwasser, F., 1996. Hundertwasser Architektur. Taschen, Köln. ISBN 3822885940.
- Imhoff, K., Imhoff, K.R., 2007. Taschenbuch der Stadtentwässerung, Oldenburg Industrieverlag, 30. Auflage, 508 p.
- Joe, H., 1996. Families of m-variate distributions with given margins and  $m(m-1)/2$  bivariate dependence parameters. L. Rüschendorf and B. Schweizer and M. D. Taylor (Ed.), *Distributions with Fixed Marginals and Related Topics*.
- Kaczmarska, J., Isham, V., Onof, C., 2014. Point process models for fine-resolution rainfall. *Hydrological Sciences Journal*, Volume 59 (11), 1972-1991.
- Kao, S., Govindaraju, R., 2008. Trivariate statistical analysis of extreme rainfall events via the Plackett family of copulas. *Water Resources Research*, 44, W02415, doi:10.1029/2007WR006261.

- Kaufman, L., Rousseeuw P. J., 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley, New York.
- Katz, R. W., Parlange, M. B., 1995. Generalizations of chain-dependent processes: application to hourly precipitation. *Water Resources Research*, 31 (5), 1331–1341.
- Kim, D., Olivera, F., Cho, H., Socolofsky, S. A., 2013. Regionalization of the Modified Bartlett-Lewis Rectangular Pulse Stochastic Rainfall Model. *Terrestrial, Atmospheric and Oceanic Sciences*, Vol. 24, No. 3, 421-436.
- Kleiber, W., Katz, R. W., Rajagopalan, B., 2012. Daily spatiotemporal precipitation simulation using latent and transformed Gaussian processes. *Water Resources Research*, 48, 1-17.
- Kohonen, T., 1990. The self-organizing map. *Proceedings of the IEEE*, 78(9):1464-1480.
- Kojadinovic, I., Yan, J., 2010. Modeling Multivariate Distributions with Continuous Margins Using the copula R Package. *Journal of Statistical Software*, 34(9), 1-20.
- Koutsoyiannis, D., Kozonis, D., Manetas, A., 1998. A mathematical framework for studying rainfall intensity-duration-frequency relationships. *Journal of Hydrology*, 206 (1-2), 118–135.
- Leimbach, S., 2017. Comparison of regionalization approaches for rainfall characteristics in Germany. Hannover: Master Thesis. Institute of Water Resources Management, Hydrology and Agric. Hydraulic Engineering.
- Liczner, P., Schmitt, T. G., Rupp, D. E., 2011. Distributions of microcanonical cascade weights of rainfall at small timescales. *Acta Geophysica*, 59, 1013–1043. (doi:10.2478/s11600-011-0014-4)
- Liu, M., Bárdossy, A., Zehe, E., 2013. Interaction of valleys and circulation patterns (CPs) on spatial precipitation patterns in southern Germany. *Hydrology and Earth System Sciences*, 17(11): 4685-4699.
- MacQueen, J., 1967. Some methods for classification and analysis of multivariate observations. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics*, 281--297, University of California Press, Berkeley, Calif.
- Maidment, D. R., 1993. *Handbook of Hydrology*. McGraw-Hill, Inc. ISBN 0070397325.
- Mehrotra R, Srikanthan R, Sharma A., 2006. A comparison of three stochastic multi-site precipitation occurrence generators. *Journal of Hydrology*, 331, 280–292, DOI:10.1016/j.jhydrol.2006.05.016.
- Mehrotra, R., Westra, S., Sharma, A., Srikanthan, R., 2012. Continuous rainfall simulation: 2. A regionalized daily rainfall generation approach. *Water Resources Research*, 48(1), 1-16.
- Meierdiercks, K. L., Smith, J. A., Baeck, M. L., Miller, A. J., 2010. Analyses of urban drainage network structure and its impact on hydrologic response. *Journal of the American Water Resources Association*, 46, 932–943.
- Modarres, R., 2010. Regional Dry Spells Frequency Analysis by L-Moment and Multivariate Analysis. *Water Resources Management*, 24(10), 2365-1380.
- Mohd Daud, Z. Mat Rasid, S. M., Abas, N., 2016. A Regionalized Stochastic Rainfall Model for the Generation of High Resolution Data in Peninsular Malaysia. *Modern Applied Science*, Vol. 10, No. 5, 77-86.

- Morales-Nápoles, O., 2008. Bayesian belief nets and vines in aviation safety and other applications. Ph.D. Thesis, Technische Universiteit Delft.
- Möser, W., Raschke, E., 1984. Incident Solar Radiation over Europe Estimated from METEOSAT Data. *Journal of Climate and Applied Meteorology*, Volume 23, 166-170.
- Mosthaf, T., Bárdossy, A., 2017. Regionalizing nonparametric models of precipitation amounts on different temporal scales. *Hydrology and Earth System Sciences*, 21, 2463-2481.
- Müller, H., Haberlandt, U., 2016. Temporal rainfall disaggregation using a multiplicative cascade model for spatial application in urban hydrology. *Journal of Hydrology (SI "Measuring & Modeling Rain")*, accepted
- Müller, T., Schütze, M., Bárdossy, A., 2017. Temporal asymmetry in precipitation time series and its influence on flow simulations in combined sewer systems. *Advances in Water Resources*, 107, 56–64.
- Müller-Westermeier, G., 1995. Numerische Verfahren zur Erstellung klimatologischer Karten. *Berichte des Deutschen Wetterdienstes 193*, Offenbach am Main: Selbstverlag des Deutschen Wetterdienstes.
- Muller , A., Arnaud , P. Lang, M., Lavabre J., 2009. Uncertainties of extreme rainfall quantiles estimated by a stochastic rainfall model and by a generalized Pareto distribution. *Hydrological Sciences Journal*, 54:3, 417-429.
- Myers, V. A., Zehr, R. M., 1980. A methodology for point-to-area rainfall frequency ratio. NOAA Technical Report NWS 24, National Weather Service, NOAA, U. S. Department of Commerce, Washington, D.C..
- Nelsen, R. B., 2006. *An Introduction to Copulas*. 2nd edition. Springer, New York. ISBN 9780387286594.
- Nguyen, V., Peyron, N., Rivard, G., 2002. Rainfall Temporal Patterns for Urban Drainage Design in Southern Quebec. *Global Solutions for Urban Drainage*, 1-16. doi: 10.1061/40644(2002)237.
- Ochoa-Rodriguez, S., Wang, L., Gires, A., Pina, R. D., Reinoso-Rondiel, R., Bruni, G., Ichiba, A., Gaitan, S., Cristiano, E., van Assel, J., Kroll, S., Murla, D., Tisserand, B., Schertzer, D., Tchiguirinskaia, I., Onof, C., Willems, P., ten Veldhuis, M-C., 2015. Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic outputs: A multi-catchment investigation. *Journal of Hydrology*, 531, 389-407.
- Paschalis, A., Molnar, P., Fatichi, S., Burlando, P., 2013. A stochastic model for high-resolution space-time precipitation simulation. *Water Resources Research*, 49 (12), 8400–8417.
- Peel, M. C., Finlayson, B. L., McMahon, T. A., 2007. Updated world map of the Köppen-Geiger climate classification. *Hydrology and Earth System Sciences*, 11 (5), 1633–1644.
- Pham, M. T., Vanhaute, W. J., Vandenberghe, S., De Baets, B., Verhoest, N. E. C., 2013. A copula-based assessment of Bartlett–Lewis type of rainfall models for preserving drought statistics. *Hydrology and Earth System Sciences Discussions*, 10, 7469-7516. (doi:10.5194/hessd-10-7469-2013)
- R Core Team, 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/> [Accessed 15 December 2016].
- Rau, P., Bourrel, L., Labat, D., Melo, P., Dewitte, B., Frappart, F., Lavado, W., Felipe, O., 2017. Regionalization of rainfall over the Peruvian Pacific slope and coast. *International Journal of Climatology*, 37: 143–158. doi:10.1002/joc.4693

- Rauthe, M., Steiner, H., Riediger, U., Mazurkiewicz, A., Gratzki, A., 2013. A Central European precipitation climatology - Part I: Generation and validation of a high-resolution gridded daily data set (HYRAS). *Meteorologische Zeitschrift*, 22(3), 235-256.
- Rodriguez-Iturbe, I., Cox, D. R., Isham, V., 1988. A point process model for rainfall: further developments. *Proceedings of the Royal Society of London A*, 417, 283–298. (doi:10.1098/rspa.1988.0061)
- Rossman, L. A., 2010. Storm Water Management Model Version 5.0 User's Manual Revised July 2010. U.S. Environmental Research Agency. EPA/600/R-05/040.
- Salvadori, G., De Michele, C., 2006. Statistical characterization of temporal structure of storms. *Advances in Water Resources*, 29 (2006), 827–842.
- Salvadori, G., De Michele, C., Kottegoda, N. T., Rosso, R., 2007. *Extremes in Nature: An Approach Using Copulas*. Springer, Dordrecht, The Netherlands. ISBN 9781402044144.
- Scarrott, C. J., MacDonald, A., 2012. A review of extreme value threshold estimation and uncertainty quantification. *REVSTAT - Statistical Journal* 10(1), 33-59.
- Schepsmeier, U., Stoeber, J., Brechmann E. C., Graeler B., Nagler T., Erhardt, T., 2016. VineCopula: Statistical Inference of Vine Copulas. R package version 2.0.1. <https://CRAN.R-project.org/package=VineCopula> [Accessed 15 December 2016].
- Serfozo, R., 2009. *Basics of applied stochastic processes*. Springer, Berlin. ISBN 9783540893318.
- Serinaldi, F., Kilsby, C. G., 2013. The intrinsic dependence structure of peak, volume, duration, and average intensity of hyetographs and hydrographs. *Water Resources Research* 49, 3423-3442, 10.1002/wrcr.v49.6.
- Smith, B. K., Smith, J. A., Baeck, M. L., Villarini, G., Wright, D. B., 2013. Spectrum of storm event hydrologic response in urban watersheds. *Water Resources Research*, 49, 2649-2663.
- Sordo, M. A., de Souza, M. C., Suárez-Llorens, A., 2016. Testing variability orderings by using Gini's mean differences. *Statistical Methodology*, 32, 63–76.
- Suzuki R., Shimodaira H., 2006. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics*, 12: 1540-1542
- Sveinsson, O. G. B., Salas, J. D., Boes, D. C., 2002. Regional Frequency Analysis of Extreme Precipitation in Northeastern Colorado and Fort Collins Flood of 1997. *Journal of Hydrologic Engineering*, 7(1), 49-63.
- Tarpanelli, A., Franchini, M., Brocca, L., Camici, S., Melone, F., Moramarco, T., 2012. A simple approach for stochastic generation of spatial rainfall pattern. *Journal of Hydrology*, 472, 63–76.
- Tawn, J. A., 1988. Bivariate extreme value theory: models and estimation. *Biometrika*, 75(3), 397-415.
- U.S. Soil Conservation Service, 1985. *National Engineering Handbook, Sec. 4, Hydrology* U.S. Department of Agriculture, Washington D.C..
- Vandenberghe, S., Verhoest, N.E.C., De Baets, B., 2010. Fitting bivariate copulas to the dependence structure between storm characteristics: A detailed analysis based on 105 year 10 min rainfall. *Water Resources Research*, 46, W01512. (doi:10.1029/2009WR007857)

- Vandenberghe, S., Verhoest, N.E.C., Onof, C., De Baets, B., 2011. A comparative copula-based bivariate frequency analysis of observed and simulated storm events: a case study on Bartlett-Lewis modeled rainfall. *Water Resources Research*, 47(7). (doi: 10.1029/2009WR008388)
- Venables, W. N., Ripley, B. D., 2002. *Modern Applied Statistics with S*. Fourth edition. Springer.
- Verhoest, N., Troch, P.E., Troch, F.P., 1997. On the applicability of Bartlett–Lewis rectangular pulses models in the modeling of design storms at a point. *Journal of Hydrology*, 202, 108–120.
- Vernieuwe, H., Vandenberghe, S., De Baets, B., Verhoest, N. E. C., 2015. A continuous rainfall model based on vine copulas. *Hydrology and Earth System Sciences*, 19, 2685-2699. doi:10.5194/hess-19-2685-2015.
- Verworn, H. R., 2008. Flächenabhängige Abminderung statistischer Regenwerte. *Korrespondenz Wasserwirtschaft* 9(1): 493-498.
- Ward, J.H., 1963. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58, 236-244.
- Wilks, D. S., 1998. Multisite generalization of a daily stochastic precipitation generation model. *Journal of Hydrology*, 210, 178–191.
- Wilks, D. S., 2008. High-resolution spatial interpolation of weather generator parameters using local weighted regressions. *Agricultural and Forest Meteorology*, 148, 111-120.
- Willems, P., 2001. A spatial rainfall generator for small spatial scales. *Journal of Hydrology*, 252, 126–144.
- WMO, 2012. *Climate and meteorological information requirements for water management*. Technical Report Series N° 1. WMO-No. 1094. ISBN 978926311094 7.
- Wu, J. Y., Thompson, J. R., Kolka, R. K., Franz, K. J., Stewart, T. W., 2013. Using the Storm Water Management Model to predict urban headwater stream hydrological response to climate and land cover change. *Hydrology and Earth System Sciences*, 17, 4743-4758. doi:10.5194/hess-17-4743-2013.
- Xiang, Y., Gubian, S., Suomela, B., Hoeng, J., 2013. Generalized Simulated Annealing for Global Optimization: the GenSA Package. *The R Journal*, 5(1),13-29.
- Xiong, L., Yu, K., Gottschalk, L., 2014. Estimation of the distribution of annual runoff from climatic variables using copulas. *Water Resources Research*, 50, 7134–7152.
- Yan, J., 2007. Enjoy the Joy of Copulas: With a Package copula. *Journal of Statistical Software*, 21(4), 1-21.
- Zhang, L., Singh, V., 2007. Gumbel-Hougaard copula for trivariate rainfall frequency analysis. *Journal of Hydrologic Engineering*, 12, 409-419.
- Zhang, Z., Switzer P., 2007. Stochastic space-time regional rainfall modeling adapted to historical rain gauge data. *Water Resources Research*, 43, W03441. doi:10.1029/2005WR004654.

## LIST OF PUBLICATIONS

### JOURNAL PAPERS

Callau Poduje, A. C., Haberlandt U., 2017. Short time step continuous rainfall modeling and simulation of extreme events. *Journal of Hydrology*, 552, 182-197.



# APPENDIX



# APPENDIX A

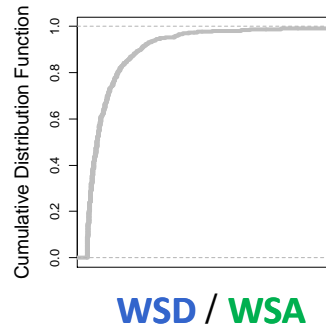
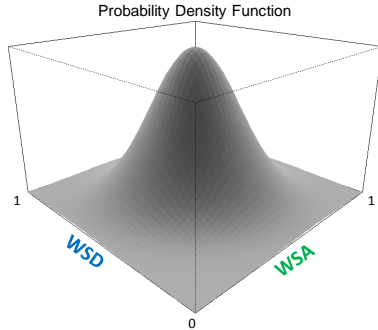
Graphical explanation of steps involved in Synthesis with Single Site Model.

## STEPS

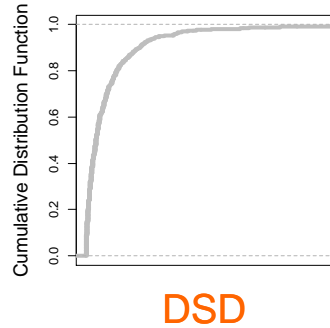
## RESULTS

EXTERNAL STRUCTURE

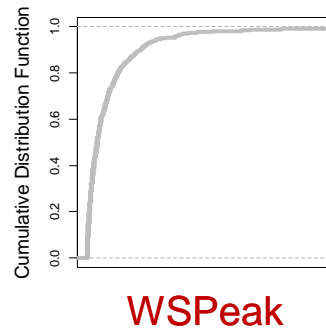
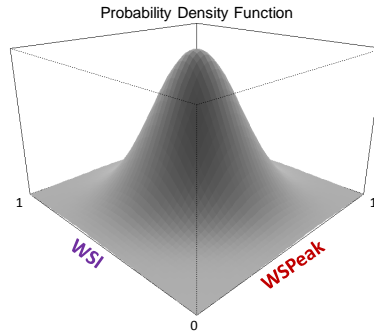
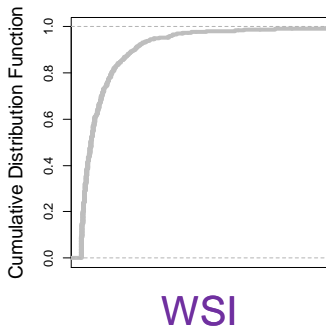
- 1) Simulation of pseudo-pairs of WSA-WSD;  
Converting pseudo-values to values (WSA and WSD);



- 2) Simulation of values of DSD;

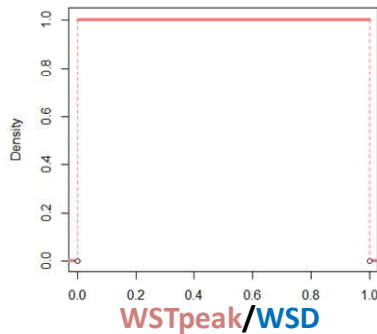


- 3) Calculation of WSI=WSA/WSD, estimation of pseudo-WSI;  
Simulation of pseudo-WSPeak conditioned to pseudo-WSI;  
Converting pseudo-WSPeak to WSPeak;



INTERNAL STRUCTURE

- 4) Estimation of WSTpeak;



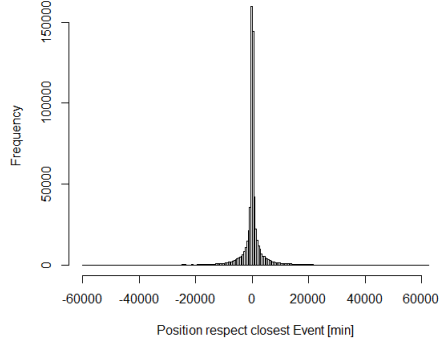
- 5) Estimation of the shape parameter  $\lambda$ ;

$$f(\lambda) = \frac{1}{\lambda} \text{WSPeak} \cdot [2 - e^{-\lambda(\text{WSTpeak})} - e^{\lambda(\text{WSTpeak}-\text{WSD})}] = \text{WSA}$$

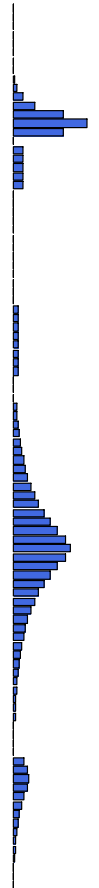
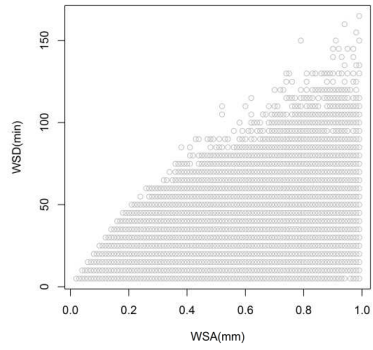


SMALL EVENTS

- 5) Selection of DSDs;  
Selection of locations within DSDs;



- 6) Introducing small events.



# APPENDIX B

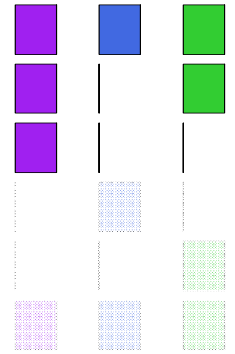
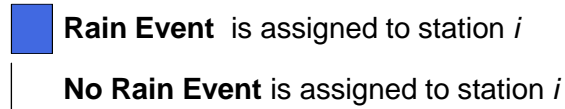
Graphical explanation of steps involved in Synthesis with Multi-Site Model.

## STEPS

## RESULTS

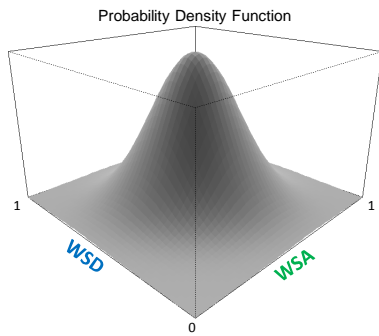
HYBRID MODEL

- Simulation of occurrence of events in 1,2, ... ,  $k$  stations;  
Assignment of event to the different locations;

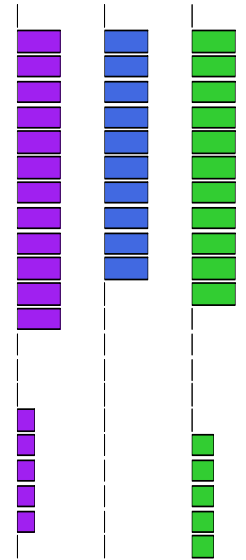
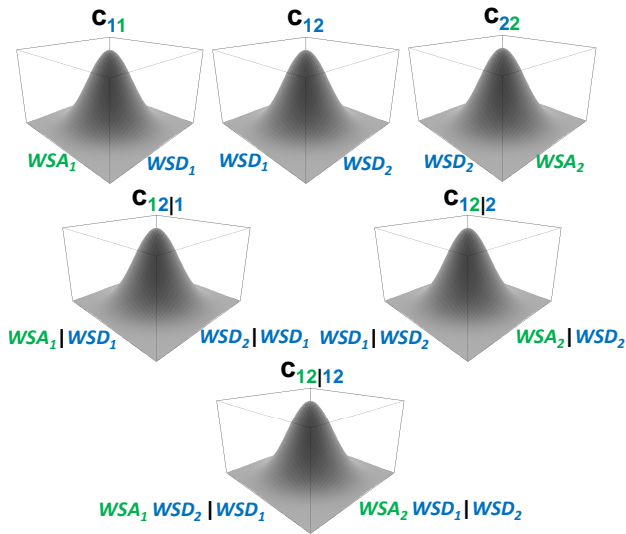


- Simulation of pseudo-pairs of WSA-WSD for events occurring in 1, 2, ... ,  $k$  stations;

### 1 Station

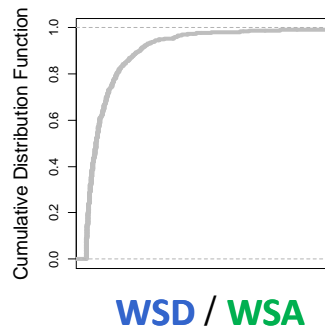
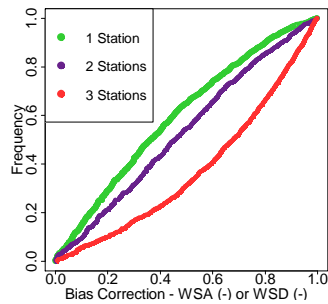


### 2 Stations



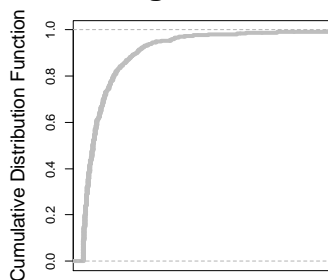
EXTERNAL STRUCTURE

- Bias correction of pseudo values of WSA and WSD;  
Converting pseudo-values to values (WSA and WSD);

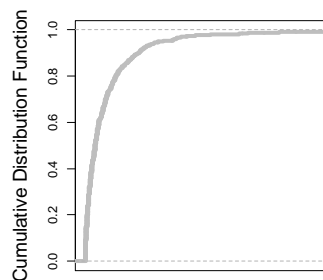


- Simulation of values of DSD for one of the stations;  
Setting of DSD for the other stations to match total duration;

### For "longest" Events

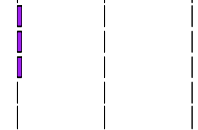


### For "consecutive" Events\*



DSD

DSD \*Merging of  $k$  stations.

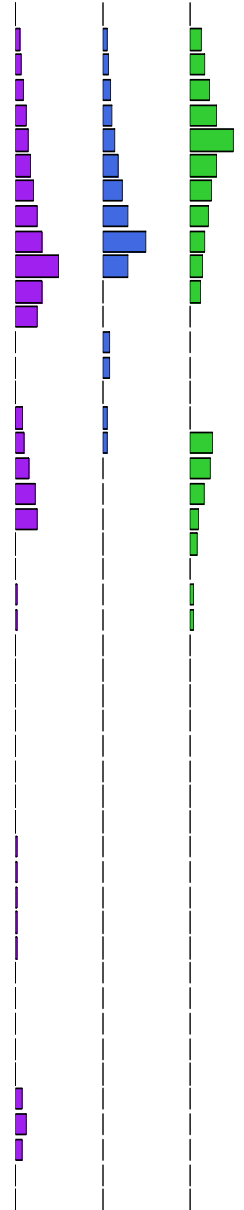


**STEPS**

**RESULTS**

**INTERNAL STRUCTURE & SMALL EVENTS**

- (Similar to Single Site Model)
- 5) Calculation of  $WSI=WSA/WSD$ , estimation of pseudo-WSI; Simulation of pseudo-WSPeak conditioned to pseudo-WSI; Converting pseudo-WSPeak to WSPeak;
- 6) Estimation of  $WST_{peak}$ ;
- 7) Estimation of the shape parameter  $\lambda$ ;
- 8) Selecting DSDs; Selecting locations within DSDs;
- 9) Introducing small events.

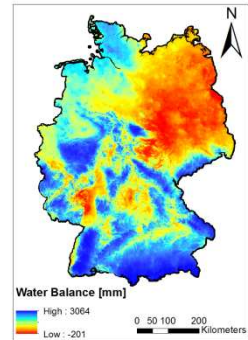
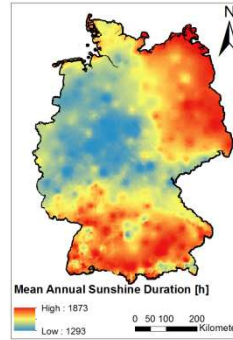
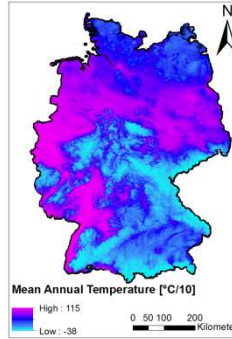


# APPENDIX C

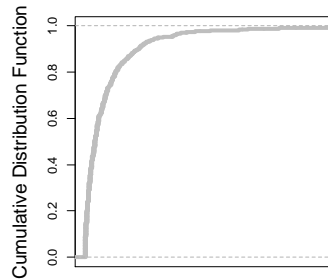
Graphical explanation of steps involved in Regionalization of the model.

## STEPS

- 1) Selection of a location for the regionalization;  
Extraction of site descriptors (SDs) for the location;

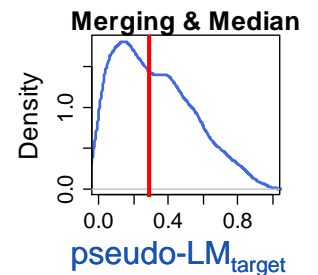
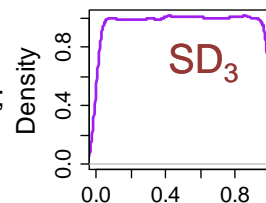
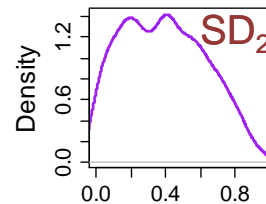
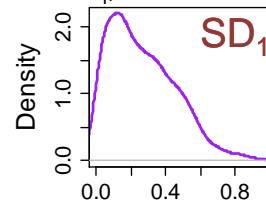
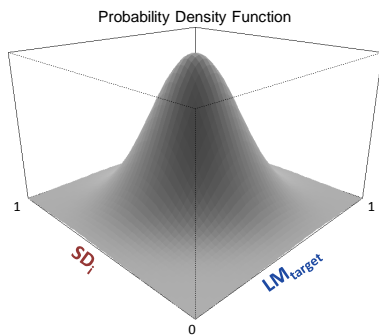


- 2) Estimation of pseudo-SDs;

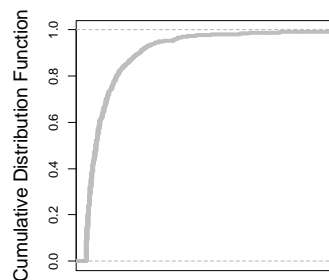


$SD_i$

- 3) Simulation of pseudo- $LM_{target}$  conditioned to pseudo- $SD_i$ ;  
Merging of simulated values using several  $SD_i$ ;



- 4) Converting median of pseudo- $LM_{target}$  to  $LM_{target}$ ;



$LM_{target}$

- 5) Estimation of model parameters using different order of regionalized  $LM_{target}$ .

# APPENDIX D

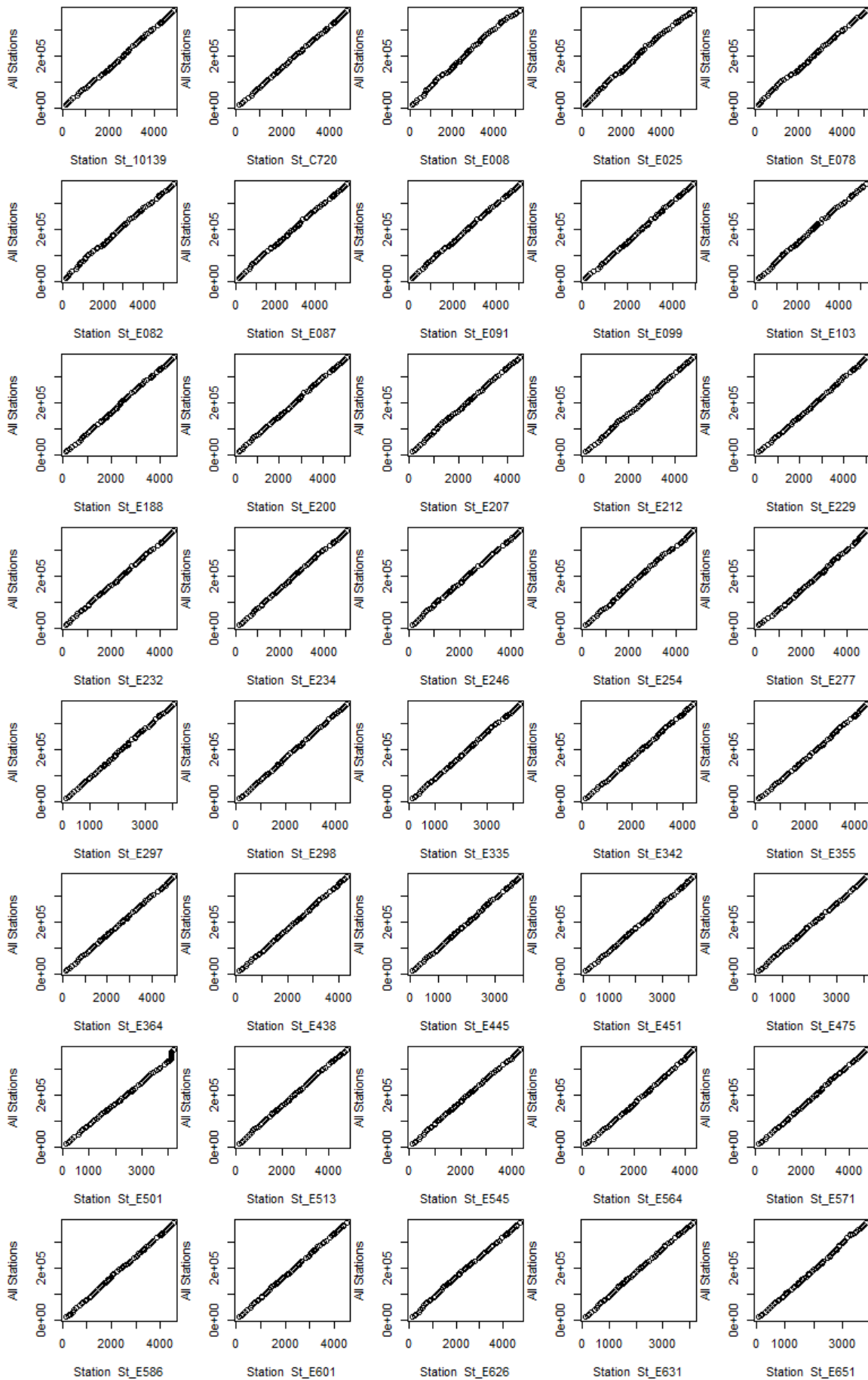
Details of all rain gauge stations available for the different analysis.

N°	NAME	ID	STATE	POSITION			DATA AVAILABILITY			Single Site Model			Multi-Site Model			Regionalization			Transferability		
				Latitude °	Longitude °	Elevation mNN	From -	To -	Length years	MW Test	CvM Test	IDF eval.	MW Test	CvM Test	IDF eval.	MW Test	CvM Test	IDF eval.	MW Test	CvM Test	IDF eval.
1	BRAUNLAGE	3984	Lower Saxony	51.73	10.6	607	1993	2013	21	x	x	x	x	x	x						
2	BRAUNSCHWEIG-VOEL	54251_10348	Lower Saxony	52.3	10.45	81.2	1998	2012	15	x	x					x	x				
3	CUXHAVEN	31121_10131	Lower Saxony	53.87	8.7	5	2000	2012	13	x	x					x	x				
4	DIEPHOLZ	56329_10321	Lower Saxony	52.58	8.35	39	1996	2012	16	x	x		x	x	x	x	x				
5	EMDEN	60820_10200	Lower Saxony	53.35	7.27	0	1998	2012	15	x	x					x	x				
6	FREIBURG/ELBE	E082	Lower Saxony	53.83	9.25	2	2003	2012	10	x	x					x	x				
7	GARDELEGEN	3169	Saxony-Anhalt	52.51	11.4	47	1993	2013	21	x	x	x				x	x	x			
8	GOETTINGEN	10444	Lower Saxony	51.5	9.95	167	1994	2013	20	x	x	x	x	x	x	x	x	x			
9	HANNOVER-LANG.	1538	Lower Saxony	52.47	9.68	55	1993	2013	21	x	x	x				x	x	x			
10	HARZGERODE	3193	Saxony-Anhalt	51.65	11.14	404	1993	2013	21	x	x	x				x	x	x			
11	JORK-MOERENDE	E188	Lower Saxony	53.52	9.73	1	2003	2012	10	x	x					x	x				
12	LEINEFELDE	3400	Saxony-Anhalt	51.39	10.3	356	1993	2013	21	x	x	x	x	x	x	x	x	x			
13	LINGEN	1132	Lower Saxony	52.52	7.3	22	1993	2013	21	x	x	x				x	x	x			
14	LUECHOW	10253	Lower Saxony	52.97	11.13	17	1993	2013	20	x	x	x				x	x	x			
15	MAGDEBURG	3177	Saxony-Anhalt	52.1	11.58	76	1993	2013	21	x	x	x				x	x	x			
16	NORDERNEY	32126_10113	Lower Saxony	53.72	7.15	11	1993	2012	19	x	x	x				x	x	x			
17	OLDENBURG	10215	Lower Saxony	53.18	8.18	11	1998	2013	15	x	x					x	x				
18	OSNABRÜCK	1516	Lower Saxony	52.26	8.05	95.4	1993	2013	21	x	x	x	x	x	x	x	x	x			
19	SALZUFLENBAD	1525	North Rhine-Westphalia	52.11	8.75	134.6	1993	2013	21	x	x	x	x	x	x	x	x	x			
20	SOLTAU	10235	Lower Saxony	52.97	9.8	75.6	1993	2013	21	x	x	x	x	x	x	x	x	x			
21	UELZEN	E475	Lower Saxony	52.95	10.53	50	2003	2012	9	x	x		x	x	x	x	x				
22	UMMENDORF	3173	Saxony-Anhalt	52.16	11.18	162	1993	2013	21	x	x	x				x	x	x			
23	WENDISCH_EVERN	E298	Lower Saxony	53.22	10.47	62	2003	2012	10	x	x					x	x				
24	WERNIGERÖDE	3180	Saxony-Anhalt	51.85	10.77	234	1993	2013	21	x	x	x				x	x	x			
25	ALFHAUSEN	E626	Lower Saxony	52.48	7.92	65	2004	2012	8							x					
26	BARNSTORF	E545	Lower Saxony	52.70	8.48	34	2005	2012	7							x					
27	BARSINGHAUSEN-HOHNENBOST.	E744	Lower Saxony	52.32	9.43	110	2004	2012	9							x					
28	BASSUM	E438	Lower Saxony	52.87	8.70	40	2004	2012	9							x					
29	BENTHEIMBAD	E704	Lower Saxony	52.30	7.13	50	2005	2012	7							x					
30	BERKA	E931	Lower Saxony	51.68	10.12	136	2005	2012	8							x					
31	BEVERN,KR.HOLZMINDEN	E818	Lower Saxony	51.85	9.50	110	2006	2012	7							x					
32	BLECKEDE-WALMSBURG	E297	Lower Saxony	53.25	10.85	12	2004	2012	8							x					
33	BOCKHORN-GRABSTEDE	E200	Lower Saxony	53.37	7.97	11	2005	2012	7							x					
34	BORKUM-FLUGPLATZ	E008	Lower Saxony	53.60	6.70	3	2006	2012	7							x					
35	BREMEROERDE	10139	Lower Saxony	53.45	9.13	10	2006	2012	7							x					
36	DINKLAGE	E631	Lower Saxony	52.68	8.12	25	2005	2012	7							x					
37	DOLLART-KANALPOLDER	E207	Lower Saxony	53.23	7.22	2	2005	2012	8							x					
38	DORNUM	E025	Lower Saxony	53.65	7.43	1	2005	2012	8							x					
39	DRANSFELD-OSSFELD	E970	Lower Saxony	51.53	9.80	317	2004	2012	8							x					
40	DROCHTERSEN	E091	Lower Saxony	53.72	9.38	1	2005	2012	7							x					
41	ESSEN-BROCKHAUSEN_BAD	E718	Lower Saxony	52.33	8.40	48	2006	2012	7							x					
42	FALLINGBOSTEL_BAD	E571	Lower Saxony	52.85	9.68	70	2004	2012	8				x	x	x	x					
43	FRANKENFELD-HEDERN	E564	Lower Saxony	52.77	9.40	19	2005	2012	7							x					
44	GROSS_BERSSEN	E513	Lower Saxony	52.73	7.50	26	2005	2012	7							x					
45	HAMBURG-NEUWIEDENTHAL	C720	Hamburg	53.48	9.90	3	2005	2012	8							x					
46	HANNOVER-HERRNHAUSEN	E755	Lower Saxony	52.40	9.67	50	2004	2012	8							x					
47	HARZBURG_BAD	E897	Lower Saxony	51.90	10.57	201	2006	2012	6							x					
48	HEESLINGEN-WIERSDORF	E364	Lower Saxony	53.30	9.33	27	2005	2012	7							x					
49	HERZBERG-LONAU	E950	Lower Saxony	51.68	10.37	340	2006	2012	6							x					
50	HUDE/OL	E335	Lower Saxony	53.12	8.42	4	2005	2012	7							x					
51	KIRCHDORF_KR.DIEPHOLZ	E651	Lower Saxony	52.60	8.85	34	2005	2012	7							x					
52	KOENIGSMOOR	E277	Lower Saxony	53.23	9.65	40	2004	2012	8							x					

N°	NAME	ID	STATE	POSITION			DATA AVAILABILITY			Single Site Model			Multi-Site Model			Regionalization			Transferability		
				Latitude	Longitude	Elevation	From	To	Length	MW	CvM	IDF	MW	CvM	IDF	MW	CvM	IDF	MW	CvM	IDF
				°	°	mNN	-	-	years	Test	Test	eval.	Test	Test	eval.	Test	Test	eval.	Test	Test	eval.
53	LAMSTEDT	E087	Lower Saxony	53.63	9.10	27	2004	2012	8								x				
54	LANGELSHEIM-ASTFELD	E884	Lower Saxony	51.93	10.35	210	2006	2012	6								x				
55	LAUTERBERG-BAD-BARTOLFEL	E955	Lower Saxony	51.60	10.47	305	2006	2012	6								x				
56	LIEBENBURG-OTHFRESEN	E871	Lower Saxony	52.02	10.40	187	2004	2012	8								x				
57	LOCCUM	E659	Lower Saxony	52.47	9.15	59	2006	2012	6								x				
58	MITTELNKIRCHEN-HOHENFELD	E099	Lower Saxony	53.55	9.62	1	2004	2012	8								x				
59	MOORMERLAND-NEERMOOR	E212	Lower Saxony	53.32	7.43	0	2005	2012	8								x				
60	MORINGEN-LUTTERBECK	E839	Lower Saxony	51.72	9.83	240	2004	2012	9								x				
61	NEUSTADT AM RUEBENBERGE	E667	Lower Saxony	52.50	9.48	40	2006	2012	6								x				
62	NORDEN-LEYBUCHTPOLEDER	E103	Lower Saxony	53.50	7.12	1	2005	2012	7								x				
63	NORTHHEIM-IMBSHAUSEN	E858	Lower Saxony	51.77	10.05	212	2004	2012	8			x	x	x			x				
64	OTTERSBERG-OTTERSTEDT	E342	Lower Saxony	53.13	9.15	17	2005	2012	7								x				
65	OVELGOENNE	E232	Lower Saxony	53.38	8.45	3	2004	2012	8								x				
66	RASTEDE	E234	Lower Saxony	53.25	8.23	16	2005	2012	7								x				
67	RINGE-GROSSRINGE	E601	Lower Saxony	52.60	6.90	13	2005	2012	7								x				
68	RINTELN-VOLKSEN	E738	Lower Saxony	52.13	9.12	212	2006	2012	6								x				
69	ROTENBURG	E355	Lower Saxony	53.13	9.33	32	2004	2012	8								x				
70	ROTHENFELDE BAD	E727	Lower Saxony	52.10	8.18	95	2005	2012	7								x				
71	SCHARNHORST-MARWEDE	E586	Lower Saxony	52.73	10.35	72	2006	2012	7								x				
72	SCHWANEWEDE-NEUENK.	E246	Lower Saxony	53.22	8.52	1	2005	2012	7								x				
73	SCHWARME	E445	Lower Saxony	52.92	9.05	12	2005	2012	8								x				
74	SEESSEN	E864	Lower Saxony	51.90	10.18	186	2004	2012	8			x	x	x			x				
75	SPRINGE	E764	Lower Saxony	52.20	9.55	98	2006	2012	7								x				
76	STEINAU, KR. CUXHAVEN	E078	Lower Saxony	53.70	8.88	1	2004	2012	9								x				
77	TWIST-HEBELERMEER	E501	Lower Saxony	52.75	7.08	19	2005	2012	7								x				
78	UETZE	E688	Lower Saxony	52.47	10.18	60	2006	2012	7								x				
79	VERDEN-DAUELSSEN	E451	Lower Saxony	52.97	9.23	21	2004	2012	8								x				
80	WEDEMARK-ELZE	E672	Lower Saxony	52.58	9.73	39	2006	2012	7								x				
81	WESTERSTEDE	E229	Lower Saxony	53.25	7.93	7	2005	2012	7								x				
82	WORPSWEDE-HUETTENBUSCH	E254	Lower Saxony	53.28	8.98	7	2004	2012	9								x				
83	ABTSGMUEND-UNTERGROENINGEN	71525	Baden-Württemberg	48.92	9.92	432	2003	2012	10								x	x			
84	BALTMANNNSWEILER-HOHENGEHREN	71155	Baden-Württemberg	48.77	9.45	457	1997	2012	16								x	x			
85	BERKHEIM	90284	Baden-Württemberg	48.05	10.08	570	1997	2012	16								x	x			
86	BUCHENBACH	70323	Baden-Württemberg	47.97	8.00	445	2003	2013	11			x	x	x			x	x			
87	DEGGENHAUSERTAL-AZENWEILER	70153	Baden-Württemberg	47.80	9.42	708	1997	2012	16			x	x	x			x	x			
88	ELLWANGEN-RINDELBACH	71605	Baden-Württemberg	48.98	10.13	460	2002	2012	11								x	x			
89	ELZACH-FISNACHT	70304	Baden-Württemberg	48.20	8.12	440	2004	2013	10								x	x			
90	EMMENDINGEN-MUNDINGEN	70314	Baden-Württemberg	48.13	7.83	201	2002	2013	12			x	x	x			x	x			
91	FREIBURG	70354	Baden-Württemberg	48.04	7.82	236	1994	2013	19			x	x	x			x	x			
92	HECHINGEN	71058	Baden-Württemberg	48.38	8.98	522	1997	2012	16								x	x			
93	HOHENSTEIN-BERNLOCH	90164	Baden-Württemberg	48.35	9.33	740	1997	2012	16			x	x	x			x	x			
94	KIRCHBERG/JAGST-HERBOLDSHAUSEN	71615	Baden-Württemberg	49.18	9.98	426	1997	2012	16								x	x			
95	KOENIGSFELD/SCHWARZWALD	71005	Baden-Württemberg	48.15	8.43	730	2003	2012	10								x	x			
96	LANGENENSLINGEN-TTENHAUSEN	90156	Baden-Württemberg	48.20	9.33	777	1997	2012	16			x	x	x			x	x			
97	LAUDA-KOENIGSHOFEN-HECKFELD	73942	Baden-Württemberg	49.55	9.63	324	1997	2012	16			x	x	x			x	x			
98	MERGENTHEIM, BAD-NEUNKIRCHEN	73930	Baden-Württemberg	49.48	9.77	250	1997	2012	16			x	x	x			x	x			
99	MUENSINGEN-APPELSTETTEN	90163	Baden-Württemberg	48.38	9.48	750	1997	2012	16			x	x	x			x	x			
100	ROTTENBURG-KIEBINGEN	71064	Baden-Württemberg	48.47	8.97	360	2003	2012	10								x	x			
101	STOCKACH	70173	Baden-Württemberg	47.87	9.02	532	2003	2012	10								x	x			
102	ULM	90307	Baden-Württemberg	48.38	9.95	567	1993	2012	20								x	x			
103	WEINGARTEN	70145	Baden-Württemberg	47.80	9.62	440	1997	2012	16			x	x	x			x	x			
104	WUESTENROT-OBERHEIMBACH	71573	Baden-Württemberg	49.13	9.50	392	1997	2012	16								x	x			

# APPENDIX E

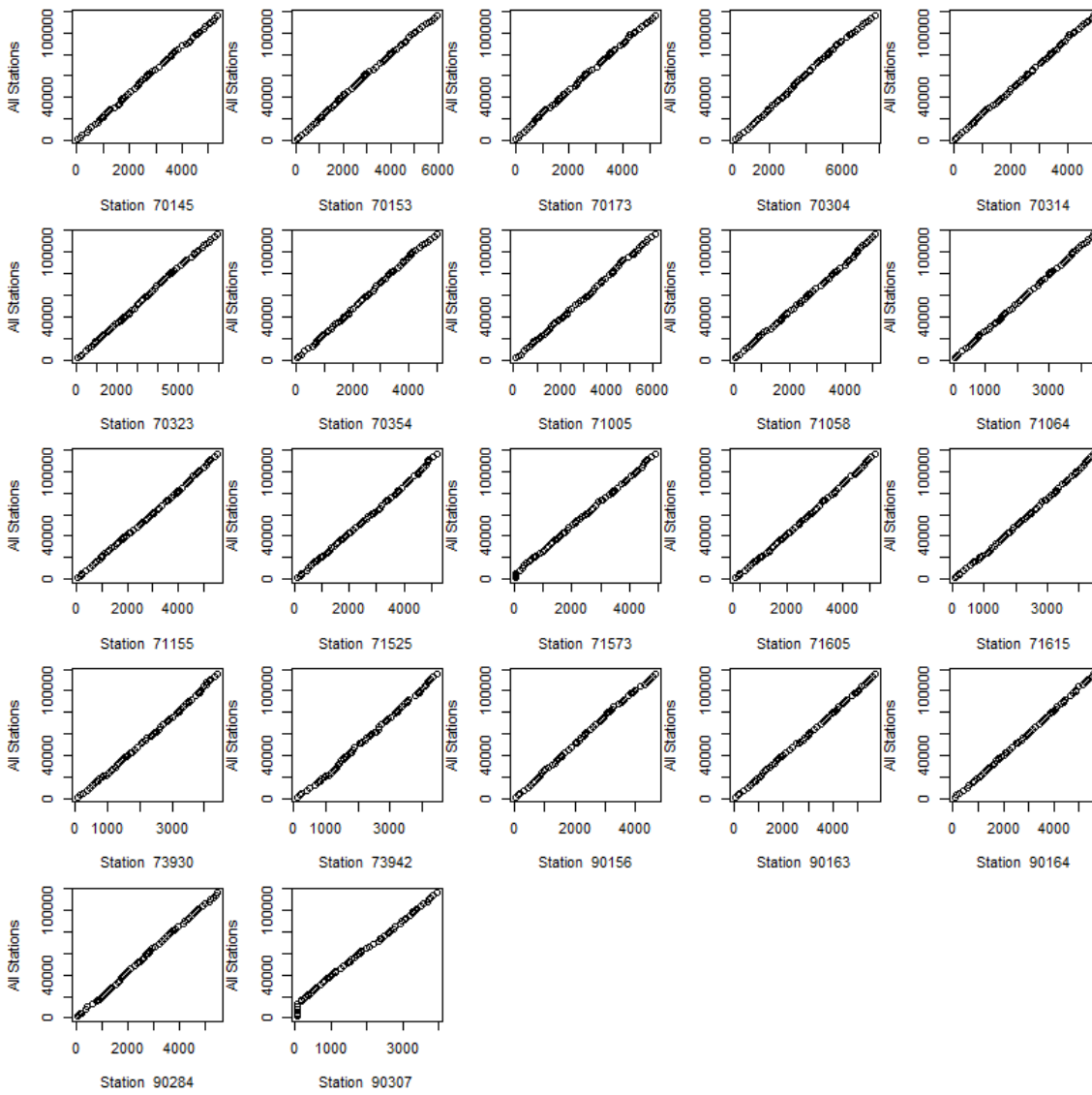
Homogeneity analysis: Double-mass curves (NS, 81 stations, Monthly rainfall 2007-2012)







# Homogeneity analysis: Double-mass curves (BAWU, 22 stations, Monthly rainfall 2007-2012)











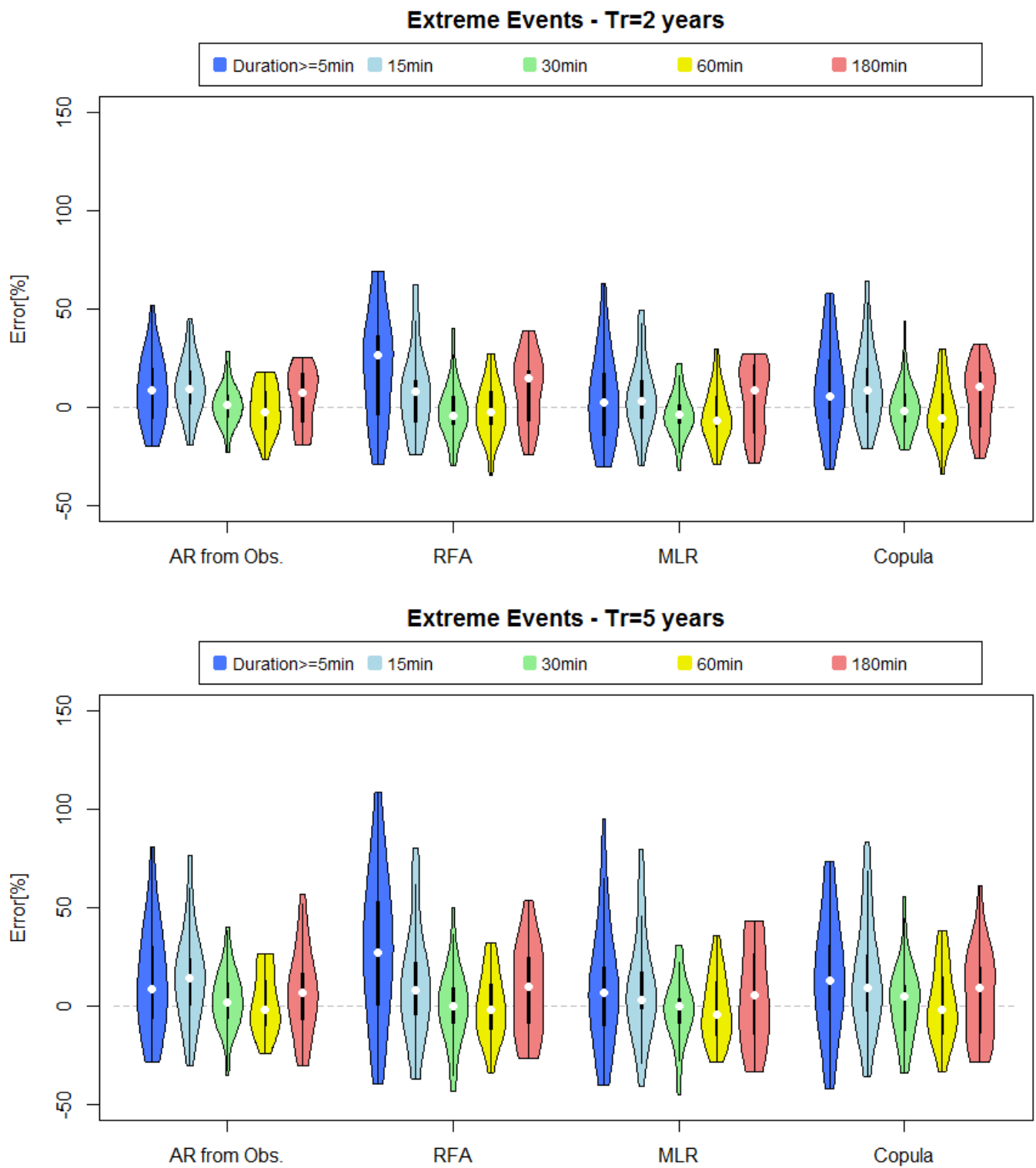
## APPENDIX G

Evaluation of statistics of different variables describing the external structure of rainfall events resulting from synthesis based on observations and three regionalization models presented for all (81) stations as RSE.

RSE [-]		Mean				Standard Deviation				Skewness				Kurtosis			
Variable	Season	AR	RFA	MLR	COPULA	AR	RFA	MLR	COPULA	AR	RFA	MLR	COPULA	AR	RFA	MLR	COPULA
DSD	Summer	0.01	0.07	0.07	0.06	0.02	0.08	0.19	0.06	0.13	0.12	0.45	0.14	0.35	0.26	0.10	0.33
	Winter	0.01	0.10	0.09	0.09	0.03	0.11	0.07	0.08	0.16	0.19	0.19	0.18	0.41	0.46	0.45	0.41
WSD	Summer	0.01	0.11	0.10	0.10	0.07	0.14	0.14	0.15	0.31	0.24	0.33	0.33	0.70	0.55	0.74	0.74
	Winter	0.01	0.12	0.11	0.12	0.05	0.14	0.12	0.15	0.25	0.23	0.28	0.29	0.55	0.50	0.62	0.64
WSA	Summer	0.01	0.05	0.05	0.05	0.03	0.11	0.11	0.11	0.19	0.29	0.27	0.28	0.48	0.73	0.68	0.73
	Winter	0.01	0.05	0.05	0.06	0.04	0.14	0.11	0.13	0.21	0.41	0.26	0.28	0.53	1.09	0.62	0.68
WSI	Summer	0.05	0.09	0.11	0.09	0.13	0.19	0.17	0.14	0.38	0.49	0.37	0.39	0.96	1.32	0.90	0.97
	Winter	0.06	0.09	0.11	0.10	0.22	0.29	0.23	0.20	0.51	0.95	0.61	0.60	1.42	3.08	1.74	1.77

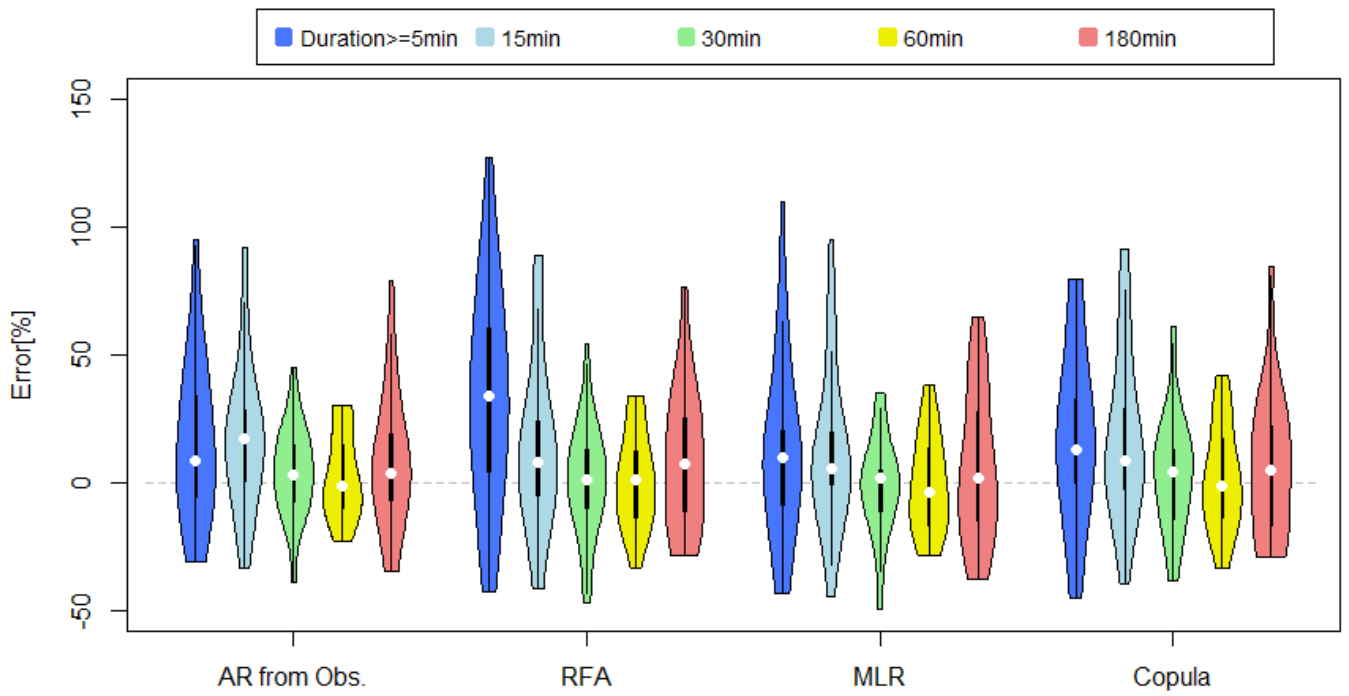
## APPENDIX H

Comparison of regionalization methods regarding their ability to reproduce extreme events for different durations and return periods. Errors for 23 stations.

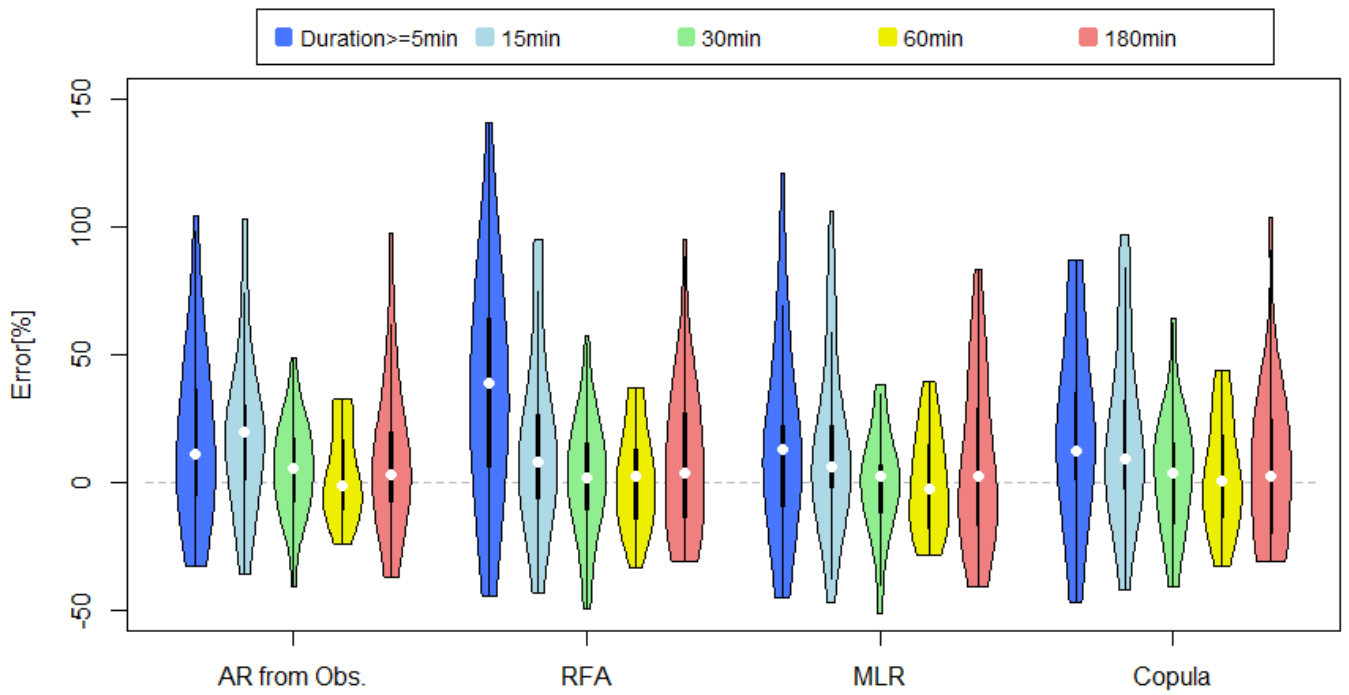




### Extreme Events - Tr=10 years

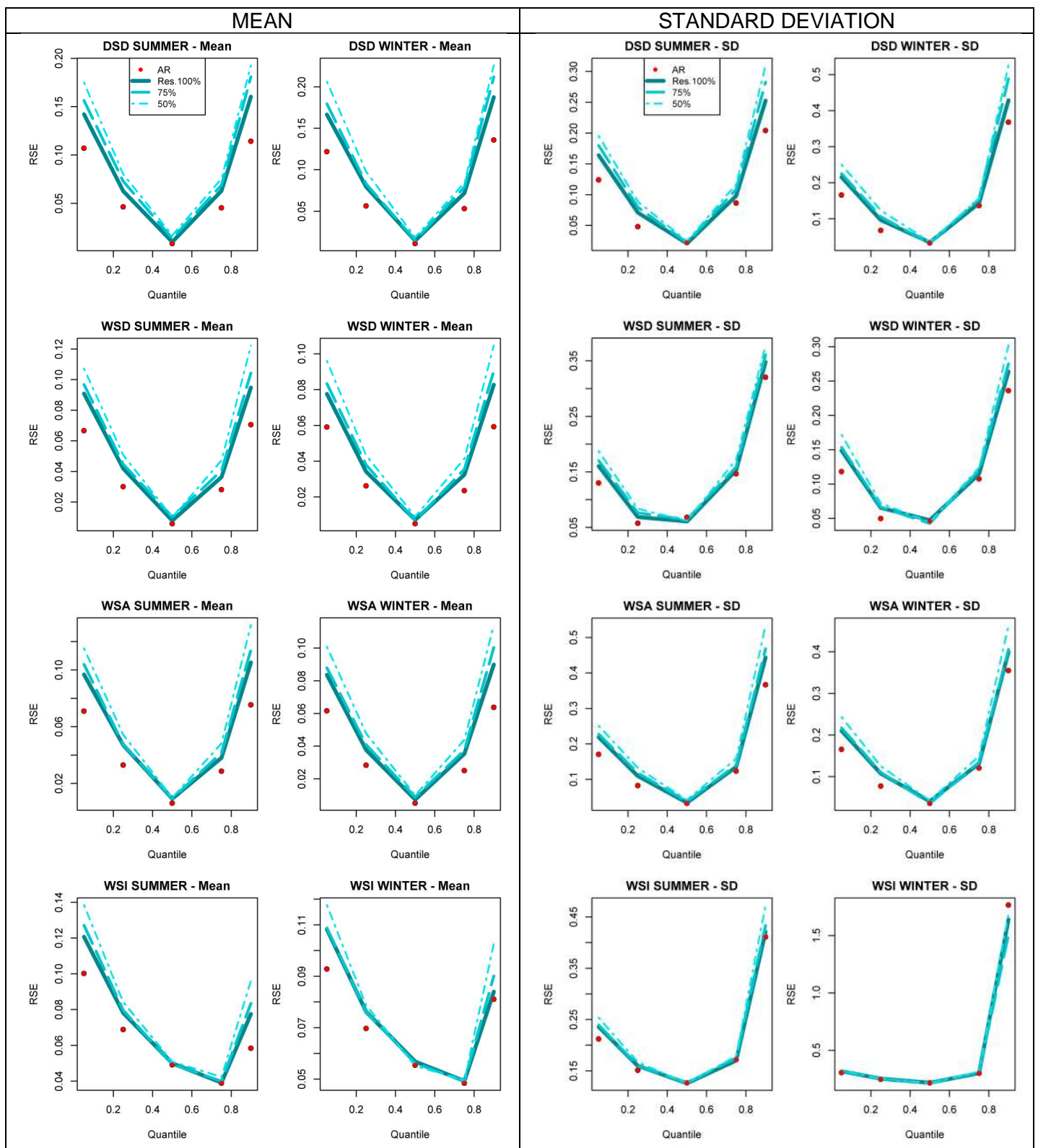


### Extreme Events - Tr=20 years



# APPENDIX I

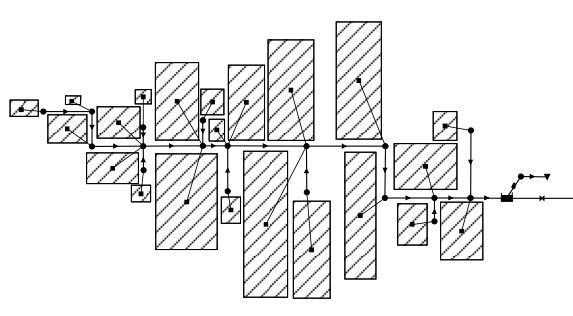
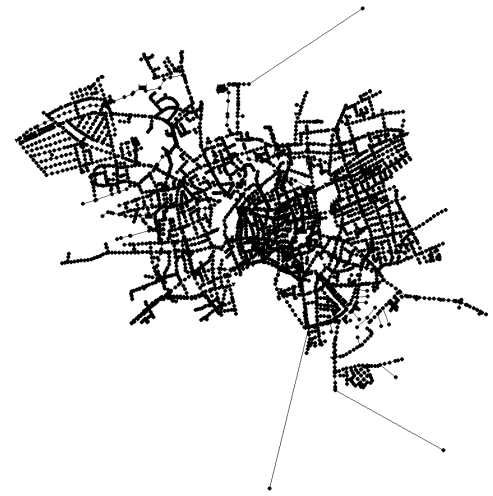
Uncertainty analysis of natural variability of the stochastic process (red dots) and parameter estimation (blue curves).





# APPENDIX J

Effect of size of urban drainage system in the estimation of flood events.

	SIMPLE MODEL	COMPLEX MODEL
Configuration		
Surf (Ha)	168	1733
N° Subcat.	22	3855
Flood Volume	<p>Legend: D.Exp. 1 Peak* (purple), D.Exp. 1/2 Peak* (green), D.Exp. 1/3 Peak* (red), Rect. (blue)</p>	<p>Legend: D.Exp. 1 Peak* (purple), D.Exp. 1/2 Peak* (green), D.Exp. 1/3 Peak* (red), Rect. (blue)</p>
Flood Duration	<p>Legend: D.Exp. 1 Peak* (purple), D.Exp. 1/2 Peak* (green), D.Exp. 1/3 Peak* (red), Rect. (blue)</p>	<p>Legend: D.Exp. 1 Peak* (purple), D.Exp. 1/2 Peak* (green), D.Exp. 1/3 Peak* (red), Rect. (blue)</p>
Flood Event 3	<p>Legend: Flood Event - Obs. Precip. (light blue), Obs. Precip. (dark blue), D.Expo. 1 Peak* (purple), D.Expo. 1/2 Peak* (green), D.Expo. 1/3 Peak* (red), Rectangular (light blue)</p> <p>Error (Vol): +192%, +106%, +65%, -37%</p>	<p>Legend: Flood Event - Obs. Precip. (light blue), Obs. Precip. (dark blue), D.Expo. 1 Peak* (purple), D.Expo. 1/2 Peak* (green), D.Expo. 1/3 Peak* (red), Rectangular (light blue)</p> <p>Error (Vol): +85%, +34%, +16%, -7%</p>

Herausgegeben im Selbstverlag  
des Institutes für Hydrologie und Wasserwirtschaft  
Gottfried Wilhelm Leibniz Universität Hannover

Appelstraße 9a; D-30167 Hannover

Tel.: 0511/762-2237

Fax: 0511/762-3731

E-Mail: [info@iww.uni-hannover.de](mailto:info@iww.uni-hannover.de)

2018

Alle Rechte beim Autor