POLITECNICO
MILANO 1863

SCUOLA DI INGEGNERIA INDUSTRIALE
E DELL'INFORMAZIONE

Spatial Dependence of Extreme Rainfall in the Seveso–Lambro–Olona Basin: An Integrated Analysis Using Citizen Science and Official Data

TESI DI LAUREA MAGISTRALE IN
CIVIL ENGINEERING FOR RISK MITIGATION
INGEGNERIA CIVILE PER LA MITIGAZIONE DEL RISCHIO

Author: **Muhammad Moonis Hafeez**

Student ID: 10889175
Advisor: Alessio Radice
Co-advisor: Ana Maria Rotaru
Academic Year: 2024-25

Abstract in English

This thesis investigates the spatial dependence of extreme rainfall in the Seveso-Lambro-Olona basin by integrating and analyzing data from both institutional (ARPA) and citizen-science (METEONETWORK) stations. Motivated by the need for high-resolution and spatially coherent rainfall data in flood-prone urban areas, the study is structured around three core objectives: validating and completing the non-institutional rainfall records, characterizing spatial dependence among extremes, and generating synthetic events that reflect that dependence.

The first part of the work focuses on validating the METEONETWORK data, based on a preliminary inspection of the spatial correlation they return compared to that returned by the official data of ARPA. Also, incomplete hourly rainfall time series from the METEO network, which are often affected by gaps, reconstructed to have full records. An imputation approach based on donor and receiving stations was used for the purpose, finally checking that reconstruction procedure did not change the spatial correlation, to ensure suitability for further analysis. This step also involved the selection of a subset of reliable stations for use in the final modelling stage.

In the second part, the spatial dependence of extreme rainfall was examined using the conditional extremes model of Heffernan and Tawn, a flexible statistical framework capable of representing spatial links between extremes.

Third, based on the fitted model, synthetic extreme rainfall events were simulated through Monte Carlo methods, producing a 1,000-year catalogue of plausible multivariate scenarios. These simulations demonstrated that combined exceedance of any threshold at multiple stations presents return periods that may be much larger than a target one, contributing to improved understanding and management of flood risk in complex urban basins.

Key-words: Citizen-science data, spatial dependence, rainfall imputation, conditional extremes model, synthetic rainfall events, combined exceedance, return periods.

Abstract in italiano

Questa tesi analizza la dipendenza spaziale degli eventi di pioggia estrema nel bacino del Seveso-Lambro-Olona, integrando e analizzando dati provenienti sia da stazioni istituzionali (ARPA) che da reti di citizen science (METEONETWORK). Motivato dalla necessità di disporre di dati pluviometrici ad alta risoluzione e coerenti nello spazio per le aree urbane soggette a rischio idraulico, lo studio è articolato in tre obiettivi principali: la validazione e il completamento dei dati delle stazioni non istituzionali, la caratterizzazione della dipendenza spaziale tra eventi estremi e la generazione di eventi sintetici che riflettano tale dipendenza.

La prima parte del lavoro si concentra sulla validazione dei dati di METEONETWORK basandosi sul confronto tra la loro correlazione spaziale e quella dei dati ufficiali di ARPA. Inoltre, le serie temporali orarie di pioggia delle stazioni METEONETWORK, spesso affette da lacune sono state ricostruite producendo serie ininterrotte. È stato applicato un metodo di imputazione basato su stazioni donatrici e riceventi per completare le serie mancanti, e si è verificato che la ricostruzione del dataset non modificasse la correlazione spaziale. Questo passaggio ha inoltre consentito di selezionare un sottoinsieme di stazioni affidabili da utilizzare nella fase finale di modellazione.

Nella seconda parte, la dipendenza spaziale degli eventi estremi è stata esaminata mediante il modello dei valori estremi condizionati di Heffernan e Tawn, un approccio statistico flessibile in grado di rappresentare tra gli estremi.

Infine, a partire dal modello calibrato, sono stati simulati eventi estremi sintetici tramite metodi Monte Carlo, generando 1000 anni di scenari multivariati plausibili. Queste simulazioni hanno dimostrato che il superamento di certi livelli di precipitazione a più stazioni ha tempi di ritorno molto maggiori di un certo tempo di ritorno desiderato, contribuendo a una migliore comprensione e gestione del rischio idraulico nei bacini urbani complessi.

Parole chiave: reti da citizen science, dipendenza spaziale, metodo di imputazione, modello dei valori estremi condizionati, eventi estremi sintetici, superamento di certi livelli di precipitazione, tempi di ritorno.

Abstract in Urdu

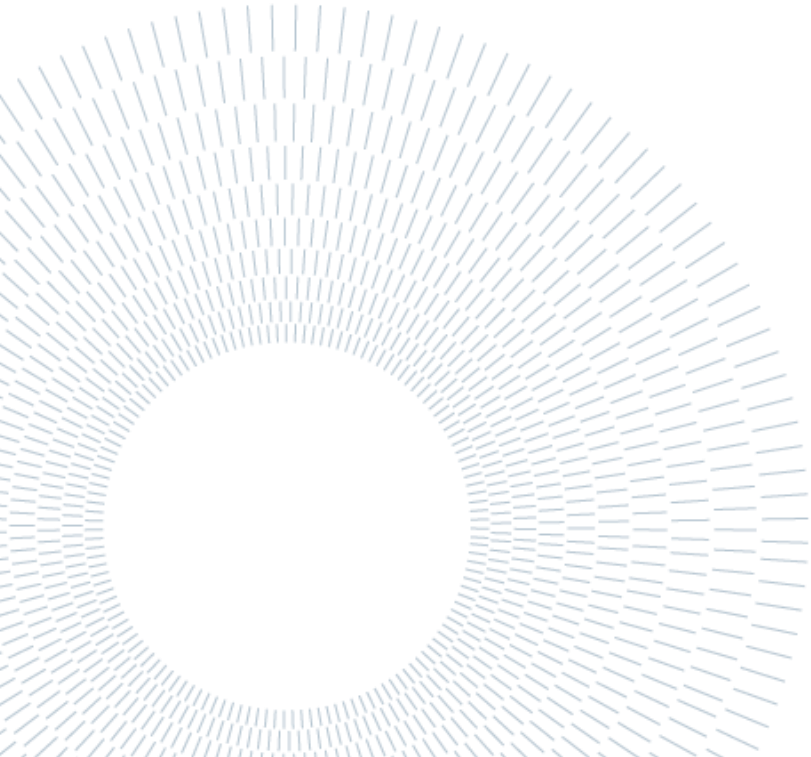
یہ مقالہ ادارہ جاتی (اے آر پی اے) اور سٹیزن سائنس (میٹی نیٹ ورک) دونوں اسٹیشنوں کے اعداد و شمار کو مربوط اور تجزیہ کرکے سیویسو-لیمبرو-اولونا بیسن میں انتہائی بارش کے مقامی انحصار کی تحقیقات کرتا ہے۔ سیلاب زدہ شہری علاقوں میں ہائی ریزولوشن اور مقامی طور پر مربوط بارش کے اعداد و شمار کی ضرورت سے متاثر ہو کر یہ مطالعہ تین بنیادی مقاصد کے گرد تشکیل دیا گیا ہے: غیر ادارہ جاتی بارش کے ریکارڈ کی توثیق اور تکمیل، انتہاؤں کے درمیان مقامی انحصار کی نشاندہی کرنا، اور مصنوعی واقعات پیدا کرنا جو اس انحصار کی عکاسی کرتے ہیں۔

کام کا پہلا حصہ اے آر پی اے کے سرکاری اعداد و شمار کے مقابلے میں واپس کیے جانے والے مقامی باہمی تعلق کے ابتدائی معائنے کی بنیاد پر میٹی نیٹ ورک ڈیٹا کی توثیق پر مرکوز ہے۔ اس کے علاوہ، ایم ای ٹی ای او نیٹ ورک سے نامکمل فی گھنٹہ بارش کے وقت کی سیریز، جو اکثر خلا سے متاثر ہوتی ہے، کو مکمل ریکارڈ رکھنے کے لئے دوبارہ تعمیر کیا گیا ہے. اس مقصد کے لئے عطیہ دہندگان اور وصول کرنے والے اسٹیشنوں پر مبنی ایک الزام تراشی نقطہ نظر کا استعمال کیا گیا ، آخر کار اس بات کی جانچ پڑتال کی گئی کہ تعمیر نو کے طریقہ کار نے مقامی باہمی تعلق کو تبدیل نہیں کیا ، تاکہ مزید تجزیے کے لئے موزونیت کو یقینی بنایا جاسکے۔ اس مرحلے میں حتمی ماڈلنگ مرحلے میں استعمال کے لئے قابل اعتماد اسٹیشنوں کے سب سیٹ کا انتخاب بھی شامل تھا۔

دوسرے حصے میں ، انتہائی بارش کے مقامی انحصار کا جائزہ ہیفرنان اور ٹاون کے مشروط انتہائی ماڈل کا استعمال کرتے ہوئے کیا گیا تھا ، جو ایک لچکدار شماریاتی فریم ورک ہے جو انتہاؤں کے مابین مقامی روابط کی نمائندگی کرنے کی صلاحیت رکھتا ہے۔

تیسرا، نصب کردہ ماڈل کی بنیاد پر، مونٹی کارلو طریقوں کے ذریعے مصنوعی انتہائی بارش کے واقعات کی نقل کی گئی، جس نے قابل قبول کثیر الجہتی منظرناموں کی 1،000 سالہ فہرست تیار کی۔ ان سیمولیشنز سے پتہ چلتا ہے کہ متعدد اسٹیشنوں پر کسی بھی حد سے مشترکہ حد سے تجاوز واپسی کی مدت پیش کرتا ہے جو ہدف سے کہیں زیادہ ہوسکتا ہے ، جس سے پیچیدہ شہری بیسن میں سیلاب کے خطرے کی بہتر تفہیم اور انتظام میں مدد ملتی ہے۔

کلیدی الفاظ :شہری سائنس کے اعداد و شمار، مقامی انحصار، بارش کا تخمینہ، مشروط انتہائی ماڈل، مصنوعی بارش کے واقعات، مشترکہ حد سے تجاوز، واپسی کی مدت.



Contents

List of Figures

List of Tables

List of Equations

1 Chapter One: Introduction

1.1. Background and Motivation

Urban areas around the world are experiencing heightened flood risk, driven by the dual forces of intensifying rainfall due to climate change and the expansion of impervious surfaces due to rapid urbanization. In metropolitan environments, where infrastructure density is high and natural infiltration processes are constrained, even moderate rainfall can result in substantial surface runoff and localized inundation. These conditions are particularly acute in the Seveso-Lambro-Olona (SLO) basin in Northern Italy, which includes the highly urbanized city of Milan. The SLO basin is characterized by complex hydrological networks, historical flood sensitivity, and rapidly evolving land-use patterns, making it a critical site for flood risk assessment and methodological innovation.

Despite the widespread recognition of flood hazards in such urban contexts, many prevailing modelling approaches remain limited in scope. Traditional flood hazard mapping, while essential for land-use planning, insurance regulation, and emergency response, is often based on event-based simulations tied to univariate frequency analyses at the station level. Typically, these models consider the transformation of rainfall into discharge at isolated points, assuming standardized design storms or return period rainfall values. These outputs are then used to drive hydraulic simulations that delineate inundation zones. However, this fragmented approach fails to reflect the spatial heterogeneity and interdependence of extreme rainfall events across a catchment [1], [2].

Moreover, traditional methods rarely consider how the timing and co-occurrence of peak flows across tributaries affect downstream flood risk. For instance, two subcatchments may experience high rainfall independently, but only when these peaks coincide does the downstream system face compounding flood pressure. Modelling such compound interactions requires a framework capable of addressing spatial and temporal dependence, yet this is largely absent in many national flood mapping protocols. Even within Italy, where the Floods Directive [3] mandates comprehensive risk assessment, methodologies differ significantly across

Hydrographic Districts [4], often focusing only on the primary river network and excluding spatially heterogeneous storm scenarios from the analysis.

This limitation is not merely technical; it has critical consequences for decision-making. Flood hazard maps that assume a spatially uniform return period can distort the shape of the risk curve, leading to underestimation of high-damage scenarios or overdesign of low-probability events [5]. Furthermore, at river confluences or shared infrastructure nodes, the superposition of hydrographs from different sources introduces a nonlinear risk profile that cannot be captured by independent rainfall-runoff models [6].

In response to these challenges, a growing body of research has called for the incorporation **of** multivariate and spatial extremes models that move beyond station-level analysis. Among the most promising is the Heffernan and Tawn conditional extremes framework [7], which models how the occurrence of an extreme event at one location conditions the distribution of outcomes at other locations. This model accommodates both asymptotic dependence and asymptotic independence, making it uniquely well-suited to the spatial behavior of convective rainfall systems, which often produce extreme values at one location and moderate values nearby. Studies such as those by Keef et al. [8] and Diederen et al. [9], have demonstrated how this framework can generate realistic synthetic rainfall fields, supporting a new generation of flood hazard maps that are both physically coherent and probabilistically robust.

The broader goal is to demonstrate how spatial extremes modelling, when paired with careful data curation and preprocessing, can overcome the limitations of conventional event-based hydrology and offer more realistic tools for urban flood risk analysis. In doing so, this research supports a shift in flood hazard assessment, from isolated point-level analysis toward a network-aware, spatially integrated framework**,** capable of informing more resilient urban infrastructure design under evolving climatic threats.

1.2. Literature Review

LIMITATIONS OF CONVENTIONAL FLOOD HAZARD MAPPING AND THE ASSUMPTION OF SPATIAL UNIFORMITY

Flood hazard mapping plays a critical role in risk-informed urban planning and infrastructure design. Traditionally, hazard maps are developed using event-based hydrological-hydraulic models, where design rainfall events of specific return periods (e.g., 100-year rainfall) are imposed on hydrological models to simulate runoff and streamflow. This is then routed through hydraulic models to delineate inundation zones. These rainfall inputs are usually derived from univariate extreme value analysis

(EVA) at individual rain gauge stations or adjusted using areal reduction factors (ARFs) or depth-area-duration (DAD) relationships [10].

However, this modelling strategy assumes spatial homogeneity of rainfall, that is, all subcatchments are subjected to identical design storms, which fails to capture the spatial and temporal variability inherent in real precipitation fields. Grimaldi et al. [11] highlighted that such models may be acceptable in ungauged basins when constrained by data limitations, but they simplify complex rainfall-runoff responses and can misrepresent flood magnitude under heterogeneous storm conditions. Wheater [12] similarly critiqued the lack of process realism in lumped models, especially in hydrologically diverse basins where rainfall-runoff relationships vary spatially and interact dynamically.

Moreover, Merz and Theiken [13], and Blöschl et al. [14], highlighted that conventional design-based flood estimation methods often overlook the spatial dependence of rainfall across a catchment. By assuming that extreme rainfall events occur independently at different locations or applying the same return period uniformly across subcatchments, these methods ignore how rainfall extremes tend to cluster spatially due to synoptic-scale weather systems or localized convective cells. This oversight leads to misrepresentation of the spatial coherence of runoff generation and consequently the flood response, particularly in interconnected sub-basins.

In hydrologically complex or urbanized catchments, where tributary inflows may interact non-linearly, the spatial arrangement and synchrony of rainfall events become critically important. A flood may not result from a single extreme at one point but from moderate extremes occurring simultaneously across spatially correlated areas. Blöschl et al. [14] argue that without accounting for these spatial structures, traditional models may produce unrealistic flow volumes and timing, especially in downstream sections where flow convergence occurs. This results in skewed flood hazard estimates and impairs infrastructure design, emergency planning, and insurance assessments.

Beyond their structural limitations, conventional flood models fundamentally misrepresent how extreme rainfall behaves across space. Real precipitation fields often exhibit partial synchrony, where some subcatchments experience peaks while others do not. Ignoring this complexity, traditional models assume simultaneous extremes everywhere, creating artificial alignment of tributary inflows. This not only inflates downstream discharge estimates but skews return period analyses and undermines confidence in infrastructure design thresholds. As flood risk increasingly hinges on spatially correlated hazards, modelling frameworks must evolve to reflect this critical dimension of hydrological reality.

This problem is evident in copula-based studies, such as by Wang et al. [6], who developed a Copula-based Flood Frequency (COFF) model to assess how the timing and correlation of tributary peaks influence joint flood behaviour. They showed that

the joint return period of peak flows varies significantly depending on whether inflows occur independently or concurrently, thereby affecting downstream risk estimations.

Similarly, Brunner et al. [15] applied "Fisher copulas" to assess the spatial dependence of river floods across multiple sites. Their results emphasized that accounting for non-linear dependence structures, particularly tail dependencies in extremes, is vital for accurate joint frequency estimates. Such approaches are crucial in hydrologically complex environments where standard multivariate Gaussian assumptions are inadequate.

Neal et al. [2] advanced this understanding by embedding spatial dependence into probabilistic flood hazard maps through a Monte Carlo simulation framework. By drawing from observed spatial correlation patterns in rainfall, they generated synthetic event sets that more realistically captured floodplain interactions. Their work demonstrated that models incorporating spatial structure aligned more closely with observed inundation footprints, supporting their practical use in risk-based planning.

The need to move beyond spatial homogeneity is echoed by Falter et al. [16], who used a coupled weather-hydrological modelling chain to produce spatially coherent synthetic flood events over Germany. Their research showed that consistent spatial-temporal rainfall patterns substantially influence damage outcomes, particularly in urban and fluvial confluence zones.

At broader scales, Metin et al. [5] employed long-term spatially coherent simulations across the Rhine catchment. Their findings underscored that simplified return-period assumptions, which ignore inter-subcatchment rainfall correlation, lead to misrepresentation of compound flood risk. Specifically, they showed that spatial dependence strongly affects the tail behavior of loss distributions, with implications for extreme event planning and insurance.

Moreover, Schneeberger and Steinberger [17], adapted the Heffernan and Tawn (HT) conditional extremes model to Alpine river basins, demonstrating how realistic, spatially heterogeneous flood scenarios could be generated to improve infrastructure design and risk zoning. They emphasized that traditional flood hazard maps, based on single-site design events, failed to capture joint exceedance probabilities, potentially underestimating the frequency of spatially extensive events.

Additionally, Quinn et al. [18] examined the spatial dependence of flood hazard and risk in the U.S., using national-scale flood simulations. They found that ignoring inter-site dependence systematically underestimates the frequency and magnitude of widespread flood losses, especially for low-probability, high-impact events.

These findings collectively affirm the need for spatially coherent modelling frameworks that preserve both the marginal behavior of rainfall at individual locations and the joint structure across a basin. This shift is essential not only for scientific

accuracy but also for operational relevance in early warning, insurance pricing, and infrastructure resilience planning.

IMPORTANCE OF RAINFALL DATASOURCE SELECTION

The quality and structure of hydrometeorological datasets significantly influence the robustness of flood risk modelling, particularly in spatially heterogeneous urban basins. Traditional institutional networks, such as those maintained by regional environmental protection agencies (e.g., ARPA in Italy), typically offer high-quality, calibrated rainfall data but may suffer from coarse spatial coverage due to the limited number of installed gauges. In contrast, emerging citizen science networks, such as the MeteoNetwork (MNW), offer denser station networks with higher spatial resolution, albeit often with limited metadata and inconsistent quality control procedures.

Recent literature has increasingly highlighted the need to evaluate and incorporate alternative data sources to enhance the spatial representativeness of rainfall inputs. Comparative analyses between MNW and ARPA Veneto datasets [19] have revealed discrepancies in measurement density and data completeness, suggesting both challenges and opportunities in fusing the two sources. A regional study in Emilia-Romagna conducted by CINECA [20] further examined the impact of MNW data on surface weather parameter estimation, demonstrating that integrating citizen-contributed observations can improve spatial granularity but requires careful preprocessing and bias correction. In Lombardy, MNW stations have been used to validate the PRISMA dataset, an institutional product that merges ARPA rain gauges with MeteoSwiss radar observations, demonstrating the applicability of MNW data in the evaluation of high-resolution gridded rainfall products [21]. Notably, manual quality control was performed over 100 hydrologically relevant events, revealing that with proper filtering and cross-validation, MNW can serve as a reliable supplement to traditional networks.

Beyond real-time operational use, MNW data are also being employed in retrospective climatological analyses. A recent thesis [22] initiated the reconstruction of long-term temperature series using MNW stations, with plans to extend this methodology to precipitation. This signals a growing shift in the scientific community toward treating citizen-science data as valid sources not only for nowcasting or public awareness, but for research-grade hydrological modelling as well.

Collectively, these studies underscore the importance of dataset selection in shaping hydrological modelling outcomes, particularly when modelling spatial extremes or compound rainfall events. The inclusion of high-resolution, citizen-sourced data, when appropriately vetted, can strengthen the reliability of spatial dependence modelling and improve the physical realism of synthetic flood scenarios. As such, understanding the limitations and potentials of different data sources is a crucial step before implementing advanced statistical frameworks such as the Heffernan and Tawn conditional extremes model.

METHODOLOGICAL ADVANCES IN SPATIAL DEPENDENCE MODELLING

To address the shortcomings of conventional models that assume uniform or independent rainfall events, a growing body of research has focused on explicitly modelling the spatial dependence of extremes. One of the most advanced and widely adopted frameworks in this domain is the conditional extremes model developed by Heffernan and Tawn [7]. This model estimates the conditional distribution of a multivariate random vector given that one of its components is extreme. It is particularly suited to cases where asymptotic independence (i.e., where extreme values at different locations are not guaranteed to co-occur) is present, making it ideal for modelling convective rainfall systems that often result in highly localized peaks.

The Heffernan and Tawn (HT) model has been successfully applied in various hydrological contexts. Keef et al. [8] utilized the model to estimate the risk of simultaneous flooding across multiple river catchments in the UK. Their work demonstrated that the HT framework could reproduce spatial patterns of flooding more realistically than traditional methods and was particularly effective in capturing the co-occurrence of flood peaks. Keef et al. [23] further developed this approach by incorporating the HT model into a probabilistic spatial risk framework, enabling the estimation of flood extents with quantified uncertainty. Their findings showed that modelling the conditional dependence structure of river flows produced more accurate assessments of joint return periods and flood magnitudes across linked subcatchments.

Recent contributions by Diederen et al. [9] extended the HT approach to generate spatially coherent synthetic flood events over continental scales, integrating it into a stochastic simulation chain for large-scale flood risk analysis. The simulations respected both the marginal behavior at individual locations and the spatial dependence across the network, significantly enhancing the realism of synthetic event sets used in flood hazard modelling.

More recent work by Debusho and Diriba [24] further exemplify the practical implementation of the Heffernan and Tawn conditional extremes model in real-world hydrological contexts. In their study focused on South Africa, they applied a multivariate conditional modelling framework to analyze joint extremes in daily maximum rainfall across multiple weather stations. By employing the Heffernan and Tawn model, they successfully captured both positive and negative extremal dependencies among spatially distributed rainfall events. Notably, their methodology accounted for uncertainties via bootstrap sampling and demonstrated that conditional modelling improved the precision of marginal parameter estimation compared to traditional univariate approaches. These insights reinforce the relevance of conditional dependence models in flood-prone and hydrologically diverse environments.

Compared to other spatial dependence models, such as copula-based frameworks (e.g., by Bevacqua et al. [25]; Brunner et al. [15]), max-stable processes [26], and

Bayesian hierarchical models [27], the HT model stands out for its flexibility in characterizing both weak and strong spatial dependencies and for its relative computational efficiency in high-dimensional settings. It enables researchers to simulate realistic rainfall or discharge fields where extremes may be concentrated in some locations and attenuated in others, a scenario highly relevant for heterogeneous urban basins such as the Seveso-Lambro-Olona (SLO) system.

Despite these advances, challenges remain in applying HT-based models operationally. A major constraint lies in data requirements, i.e., high-resolution, quality-controlled rainfall or discharge time series are necessary to fit the model robustly. Moreover, the treatment of missing or inconsistent data, especially in networks that integrate citizen-science or semi-professional stations, remains a bottleneck to broader adoption.

Nevertheless, the HT model represents a key step forward in flood risk analysis, particularly in catchment-scale applications where traditional models fall short. This thesis builds upon this foundation, applying the HT framework to spatially imputed rainfall data in the SLO basin to explore its value in developing realistic, probabilistically grounded flood hazard scenarios.

1.3. Objective of the Study

In response to the challenges mentioned prior, this study focuses on developing a spatially integrated approach to flood risk assessment, tailored to the complex hydrometeorological context of the Seveso-Lambro-Olona (SLO) basin. By leveraging both citizen-science and institutional rainfall data, modeling spatial dependence in rainfall extremes, and generating synthetic but spatially consistent storm events, the study aims to produce return period estimates for extreme rainfall scenarios (e.g., 5-, 10-, 100-, 500-, and 1000-year events) that better reflect spatial variability, supporting more realistic and data-driven flood hazard characterization in urbanized catchments.

The specific objectives of this study are:

1. To assess the reliability and spatial utility of citizen-contributed MNW rainfall data in comparison with institutional records.
2. To develop a robust, data-driven imputation strategy for addressing missing values and ensuring spatial completeness.
3. To apply univariate extreme value theory to model rainfall intensity distributions at individual stations.
4. To quantify spatial dependence among rainfall extremes using the Heffernan and Tawn conditional extremes framework.

5. To generate spatially consistent synthetic rainfall fields that preserve both marginal and joint behaviors, supporting improved flood risk modelling.

These objectives structure the thesis in a progressive manner, beginning with data acquisition and quality control, moving through marginal and spatial extremes modelling, and culminating in the simulation of synthetic events for scenario-based flood risk analysis.

The remainder of this thesis is organized as follows:

- **Chapter 2** introduces the study area and describes the data sources, including institutional and citizen-science networks.
- **Chapter 3** conducts an exploratory analysis of rainfall data, evaluating station reliability and spatial correlation.
- **Chapter 4** presents the imputation methodology and its validation across multiple scenarios.
- **Chapter 5** introduces the conditional extremes framework and simulates synthetic events to support long-term risk analysis.
- **Chapter 6** synthesizes findings.

This structure reflects a progression from data curation to advanced modelling, supporting a shift from traditional, static flood hazard assessments to dynamic, data-integrated approaches.

2 Chapter Two: Case Study

2.1. Area of Study

The study area encompasses the hydrological basins of the Seveso, Olona, and Lambro rivers, which collectively constitute the Seveso-Lambro-Olona (SLO) basin. This area is located in the northern part of Italy and extends across a latitude range of approximately 45.37°N to 45.93°N and a longitude range from 8.77°E to 9.40°E, covering a surface area of about 1400 km².

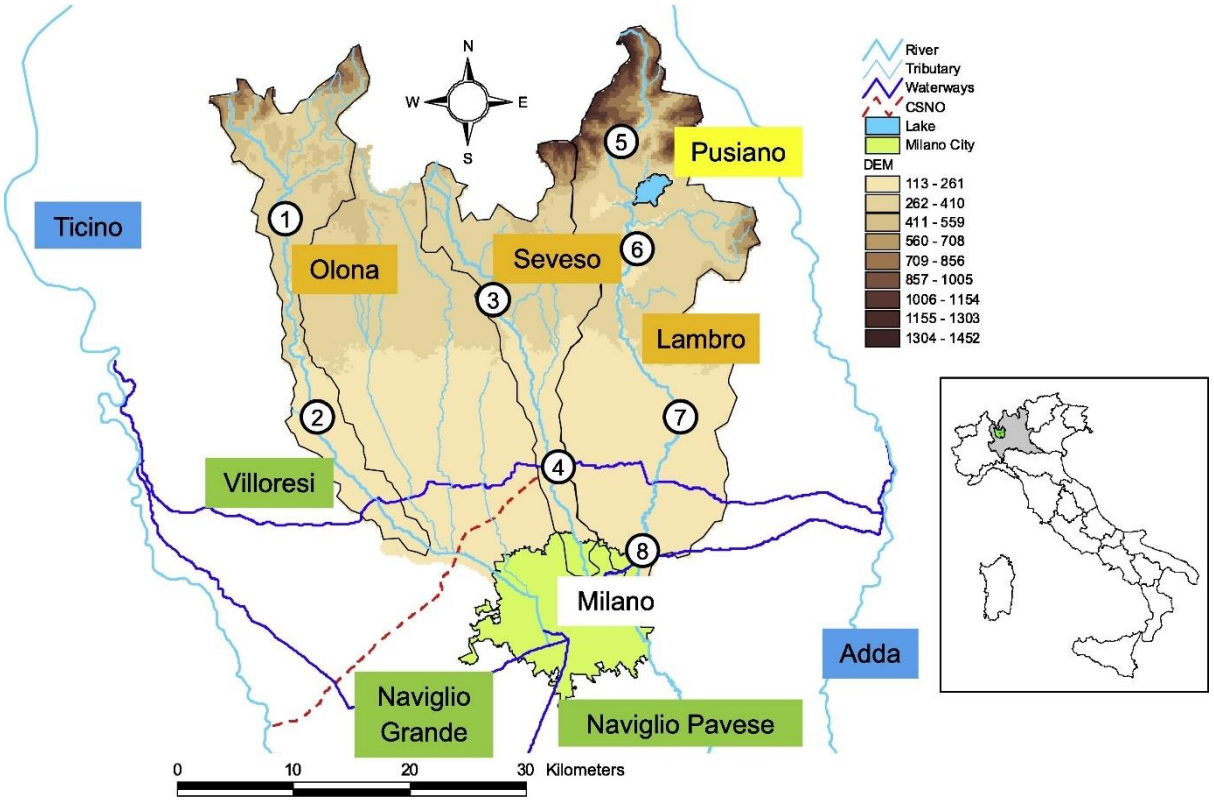


Figure 1. Area of Study (Seveso-Lambro-Olona Basin) - White circles denote hydrometric stations. [28]

Milan, with approximately 1.4 million residents concentrated within an area of 182 km², and nearly 5 million when including the surrounding metropolitan region, stands as one of the most densely populated and economically significant urban centers in Italy. The territory extending to the North of Milan has undergone significant urban area expansion since 1950, ultimately modifying the response of the watershed to precipitation input.

From a climatic perspective, the region is classified as Cfa (Humid Subtropical Climate) according to the Köppen climate classification system [29], indicating a temperate humid subtropical climate. This classification is typical of mid-latitude regions, with hot summers and no significant dry season. The northern part of the area includes the foothills of the Italian Prealps, while the southern portion reaches into the densely urbanized Milan metropolitan area. Between these zones lies a flat to gently undulating terrain, shaped by extensive urbanization and infrastructure growth.

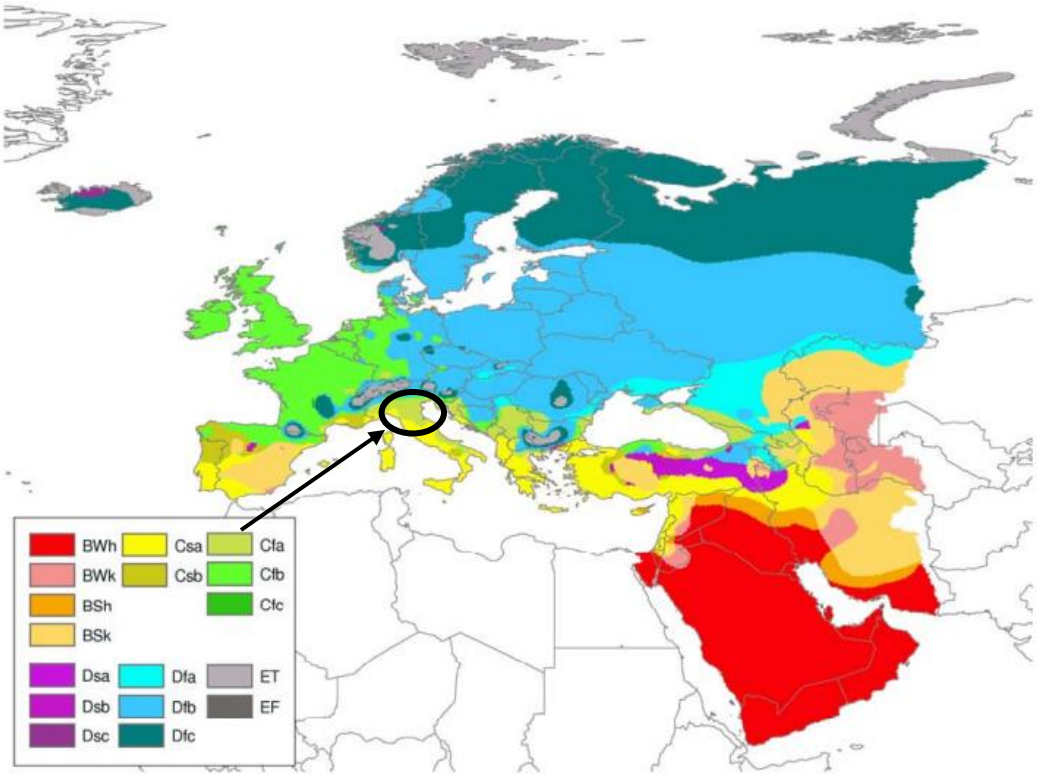


Figure 2. Koppen-Geiger climate type map of Europe [29]

The unique topography and land use of the region contribute to distinct meteorological patterns. The area is often subject to convective precipitation events, particularly during summer months. This is due to humid air masses stagnating in the Po Valley, which are subsequently lifted by thermal convection triggered by surface heating. Additionally, the northern section experiences stronger winds and frequent foehn events (i.e., warm, dry winds that occur on the lee side of mountains) in winter, while the southern urban basin is prone to fog formation and thermal inversions, especially during the colder months.

Average annual precipitation in the area, as seen in Figure 3, ranges between 800 and 2,000 mm, with the highest rainfall amounts occurring in the mountainous regions of the Olona and Lambro basins.

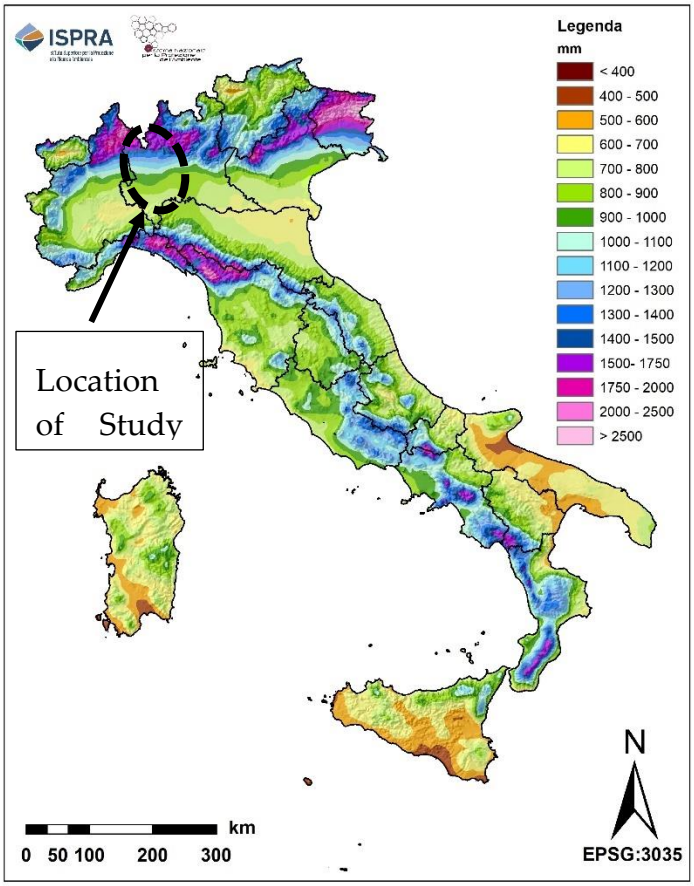


Figure 3. Spatial distribution of long-term mean annual precipitation depth (LTAA, 1951–2023) [30]

2.1.1. Seveso Basin

The Seveso Basin encompasses a watershed area of approximately 227 km², with 100 km² of this area situated within urban zones. The Seveso River, originating near the border with the Canton of Ticino in the province of Como at an elevation of around 490 meters above sea level, flows through various towns in the Brianza region before merging with the Martesana Canal in central area of Milan. Notably, the final 9 kilometers of the river within the urban area of Milan are fully culverted. A significant tributary of the Seveso is the Torrente Certesa, whose basin covers an area of 72 km² [31].

The basin can be conceptually divided into four distinct areas, each with different levels of urbanization and geomorphological characteristics [31]. The northernmost portion, often referred to as "Seveso naturale," extends from the source of the river to the town of Lentate sul Seveso. This section is characterized by steep or moderately steep slopes and remains largely natural, with minimal urbanization, resulting in little alteration to the river's flood processes. To the east lies the area known as "Certesa naturale," which is defined by the Torrente Certesa, the Seveso's main tributary. This includes the watersheds of both Torrente Terrò and Certesa up to their confluence. The

landscape here also features steep slopes and limited development, preserving much of its natural hydrological function.

As the basin extends southward, urban influence increases. The area known as "Certesa urbano" includes the towns of Mariano Comense, Cabiate, Meda, Seveso, and Cesano Maderno. In this zone, from the confluence of Torrente Terrò and Certesa to their merger with the Seveso River, widespread urbanization and gentler slopes have led to significant alterations of the natural flood dynamics. Further downstream, the section referred to as "Seveso urbano" stretches from Lentate sul Seveso to the beginning of the culverted portion in Milan. This segment, characterized by nearly flat terrain and dense urban development, includes municipalities such as Barlassina, Seveso, Cesano Maderno, Bovisio Masciago, Varedo, Paderno Dugnano, Cusano Milanino, Cormano, Bresso, and Cinisello Balsamo.

A significant infrastructure feature of the Seveso Basin is the Canale Scolmatore di Nord-Ovest (CSNO) [31], which diverts floodwaters from the Seveso River at Palazzolo, a district of Paderno Dugnano. Designed in 1954 and completed in 1980, this channel was intended to prevent flooding in Milan by handling up to 30 m^3/s of overflow [28]. However, it has proven insufficient to address the growing flood risks in the area. Other flood defenses present, or currently under design or construction along the Seveso course are expansion areas in the Northern section, as well as water detention basins at Lentate sul Seveso, Paderno Dugnano, and Bresso.

2.1.1 Olona Basin

The Olona River originates at the foothills of the mountains north of Varese and flows for over 60 kilometers before entering the city of Milan. Upon exiting the city, the river is known as the Lambro Meridionale and eventually merges with the Lambro River at Sant'Angelo Lodigiano. The hydrological basin of the Olona River can be divided into two distinct areas: a mountainous section upstream of Ponte Gurone and a flatter section from Ponte Gurone to the city of Milan [32].

The upper basin is Y-shaped, with the western branch encompassing the main Olona River and the eastern branch containing the watersheds of Torrente Bevera, Torrente Clivio, and Rio Ranza. The western branch is significantly urbanized and includes important settlements such as Varese and Induno Olona. In contrast, the eastern branch is more natural, covered mostly by forests and agricultural areas with only a few small towns. Below Ponte Gurone, the basin narrows into an elongated north–south corridor that alternates between densely built-up zones and open land. Up until the Milano-Laghi highway, which marks the end of the Olona Valley, towns are generally built on terrain elevated above the river, while many industrial developments are situated near the riverbed. After exiting the valley, the river continues through increasingly urbanized zones, including the towns of Castellanza and Legnano.

The Ponte Gurone Dam is a key flood mitigation infrastructure in the Olona Basin, constructed in 2010 near the city of Varese. The dam serves an on-stream detention function with a total storage capacity of 1,520,000 m^3, regulating runoff from a drainage area of approximately 3.83 km^2. Equipped with three automated gates, the dam controls discharge to ensure it does not exceed 36 m^3/s, considered the threshold for safe downstream flow [28]. Flood mitigation measures are also water detention basins at Varese and San Vittore Olona.

Beyond Castellanza and Legnano, the landscape continues to alternate between agricultural and urban areas, extending to the boundary of the municipality of Rho. The Olona Basin, therefore, presents a diverse range of land use, from rural and forested zones in the north to highly urbanized areas towards the south, reflecting the ongoing transformation of the landscape as the river progresses toward Milan.

2.1.2 Lambro Basin

The Lambro River originates in the Triangolo Lariano, specifically in the Prealps between the two branches of Lake Como, within the municipality of Magreglio. After a course of approximately 130 kilometers, which includes traversing the city of Milan, the river flows into the Po River at Senna Lodigiana [33]. The Lambro Basin covers an area of around 1,980 km^2, which constitutes about 3% of the total area of the Po River Basin [34].

The basin is characterized by a complex and intricate hydrographic network, with numerous natural watercourses flowing from the north of Milan in a north-south direction. These streams are interconnected through a dense network of artificial canals, built for irrigation purposes and to protect urban areas from flooding. The primary watercourse is the Northern Lambro, which flows to the east of Milan [34].

The discharge from the upper basin is moderated by the lakes of Alserio and, particularly, Lake Pusiano. These lakes, with a considerable surface area of approximately 8 km^2 relative to the size of the surrounding basin, play a significant role in attenuating flood events [34]. The lakes' capacity to store water helps to moderate peak flows, reducing the potential for flooding downstream. This characteristic of the Lambro Basin highlights the importance of natural features in flood regulation, particularly in areas with a dense network of watercourses and urbanization.

Floods are mitigated thanks to the Pusiano Lake, Fornacetta dam and the Brenno quarry that is along the Bevera tributary. On the contrary, there are no major flood defense system in the lower course of the river.

2.2. Hydrological Aspects and Past-Events

The Seveso, Olona and Lambro River basins have a long history of flood events due to combination of intense rainfall, urbanization and geomorphological factors. The Lombardy region, particularly the northern part of Milan, is highly susceptible to flood risk, especially from the Seveso River, which has caused 342 floods over the past 140 years, averaging 2.6 per year [35].

2.2.1. Seveso Basin

The Seveso basin, predominantly urban and located in a densely populated area north of Milan, exhibits a sublitoraneo padano rainfall regime. It means that it experiences two annual precipitation maxima in spring and autumn (monthly averages around 100-110 mm), and two minima in winter and summer (approximately 60mm/month). The total annual rainfall is around 1,100 mm .

The Seveso River Basin is notoriously one of the most flood-prone areas in the Lombardy region. Despite its relatively small catchment area (~930 km²), the Seveso has caused recurrent and often severe flood events, especially in highly urbanized areas like Milan, due to the limited channel capacity, high impervious surface cover, and historical canalization of its course. The river's natural regime has been drastically altered by extensive urban development, particularly from the mid-20th century onwards, which has significantly increased surface runoff and reduced infiltration.

The history of flooding in the Seveso basin dates back over a century, with increasingly frequent and intense events observed in recent decades. This stream has historically caused recurrent flooding in Milan and surrounding areas, with increasing frequency over the decades. Between 1976 and the present, there have been 104 flood events [36].

Even before 1976, significant floods occurred, some with tragic consequences. A particularly devastating event was in 1917, when the collapse of the Paderno Dugnano bridge resulted in 19 fatalities. Other major flood episodes were recorded during the 1925–1935 decade [36]. However, due to substantial urban and hydrological changes in the Seveso basin, particularly after World War II, earlier flood data are considered less relevant for current risk assessments. The modern configuration of the river and surrounding territory differs significantly from earlier periods, making post-1976 events more representative for present-day analyses [36].

Between January 2010 and December 2015, multiple flood events were recorded in the Seveso River basin, leading to overflows within the Municipality of Milan and, in some cases, affecting upstream municipalities as well. Notable rainfall events causing such overflows occurred on eight occasions in 2010, twice in 2011, once each in 2012 and 2013, and eight times again in 2014. The most severe episodes were recorded on 18 September 2010, 8 July 2014, and 12–15 November 2014, which resulted in extensive flooding and damage along the entire river corridor [36].

A major flood event occurred in July 2014, when over 60 mm of rain fell in just 5 hours, leading to the river overflowing and inundating several areas across Como, Monza–Brianza, and Milan. In Milan, around 3 km² of the northern city was submerged, severely disrupting transportation and damaging numerous homes, shops, and infrastructure. The Niguarda and Isola neighborhoods were particularly affected [36]. The July 2014 event was particularly critical, with rainfall return periods ranging from 20 to 50 years across different stations. The flood wave generated at Palazzolo displayed hydrological characteristics consistent with a 100-year return period, including a peak discharge of approximately 160 m^3/s and a volume exceeding the 30 m^3/s threshold derivable into the CSNO (Canale Scolmatore di Nord Ovest) by around 3.3 million cubic meters (Mm^3). This extreme response was exacerbated by soil saturation from a previous event on 29 June 2014, which had already caused a peak discharge of 115 m^3/s and a derived volume of about 1.8 Mm^3. The total damage was estimated at €27.2 million, with private property bearing the greatest losses [35].

A similar hydrological scenario occurred during the November 2014 event, which reached a peak discharge of roughly 140 m^3/s and a flood volume of approximately 4 Mm^3 above the CSNO derivation threshold. The preceding days had experienced significant rainfall, on 5, 6, 10, and 12 November, resulting in cumulative saturation of the catchment. The combined runoff from these antecedent events led to estimated peak discharges of 85, 54, 56, and 105 m^3/s, respectively, with a total wave volume surpassing 6.7 Mm^3.

Figure 4. Flood in Milano, due to Seveso River (14th July 2014)

In response to the increasing frequency and severity of events, the “Programma Seveso” was initiated, which includes the construction of lamination basins (such as those at Senago, Parco Nord, and Bovisio Masciago) and improvements in forecasting and warning systems. Nonetheless, the basin remains sensitive to short-duration, high-intensity rainfall events, which are likely to increase under future climate scenarios.

2.2.2. Olona Basin

The Olona River originates north of Varese and, after traveling through urban and industrial areas, becomes the Lambro Meridionale as it enter Milan. Its basin has a sublitoraneo alpino pluviometric regime, featuring two similar seasonal maxima in spring and autumn (around 130 mm/month), and minima in summer (around 90 mm/month) and winter (around 65 mm/month), with an annual total of around 1,220 mm [37].

The Olona River Basin has a long history of flood events, many of which have caused significant damage to both urban and rural areas along its course. These flood events are largely influenced by the basin's intensely modified hydrological regime, a result of extensive urbanization, the presence of hydraulic infrastructures such as diversion works, and limited natural floodplains downstream. Historically, the Olona's floods have been exacerbated by rapid runoff from upstream mountainous areas, as well as insufficient channel capacity and alterations to natural drainage. One of the earliest documented flood events occurred in October 1801 [37], where the Olona overflowed its banks, causing widespread inundation in several towns including parts of Milan. At that time, the river still maintained a more natural course before extensive human interventions.

A particularly destructive flood followed in December 1910 [37], when continuous rainfall caused the Olona to breach its banks in multiple locations between Varese and Milan, inundating peripheral areas of the city and causing substantial damage to infrastructure and property.

Further severe-flooding occurred on 4–5th June 1936 [37], impacting towns such as Legnano, and again highlighting the vulnerability of urban centers located close to the river channel.

In the 1990s, a series of flood events emphasized the increasing frequency and impact of hydrological extremes in the basin. The 1992 flood [37] primarily affected Castiglione Olona, causing local damage to homes and public infrastructure. Just a year later, in 1993 [37], another flood led to the inundation of Milan's outskirts, reaffirming the ongoing challenges in managing runoff and floodwaters in highly urbanized stretches of the basin.

A major flood took place on 15-16th October 1995 [38] , when the Olona River overflowed in via Milano, causing significant damage to buildings, various sports clubs, and other community organizations. The flood resulted in losses estimated at approximately one hundred million lirre, with extensive destruction of computers, furniture, musical instruments, and other equipment.

Most recently, on 15th July 2009 [39] localized flooding occurred in the Varese area, causing approximately 3 million euros in damages. This event demonstrated persistent

hydrological risk, even with the presence of flood mitigation infrastructure such as the Ponte Gurone retention basin, which was completed in 2010.

2.2.3. Lambro Basin

The Lambro River Basin has historically experienced multiple significant flood events, primarily due to its complex hydrographic network, rapid runoff response, and urbanized areas such as Monza and eastern Milan. These floods are typically caused by intense and persistent rainfall. The rainfall regime of the Lambro basin can be classified as sublitoraneo padano type. It features two maxima and two minima of roughly equivalent magnitude: the spring and autumn peaks have average monthly values of around 110 mm, while the summer and winter lows are around 60 mm. The total average annual precipitation is approximately 1020 mm [40]. The limited retention capacity of the soil, combined with urban impermeability, increases runoff and aggravates the flood risk.

The Lambro River experienced significant flooding events in May 1917, September 1937, 1993 and November 2002. The 1937 flood led to widespread inundation in Monza and the surrounding areas of Milan, while the 1993 flood affected the outskirts of the city [40].

The November 2002 flood is recognized as the most hydrologically extreme event in recent decades within the Lambro River basin. Following an unusually dry start to the month, a series of intense meteorological disturbances originating from North Africa unleashed extraordinary rainfall over the region, particularly from November 12 to 26. In some locations, such as Perticato, cumulative precipitation reached nearly 500 mm, with peak intensities exceeding 30 mm/h. The station at Peregallo (MB) recorded a peak discharge between 120 and 130 m^3/s, corresponding to an estimated return period of 100–200 years. The event caused widespread disruption and significant hydrological impacts. In the historic center of Monza, extensive surface flooding occurred, leading to the temporary closure of commercial activities and the activation of emergency services. In several affected areas, water depths exceeded 1 meter. Economic losses were substantial, with damage estimated in the millions of euros. Residential and commercial zones in municipalities such as Agliate, Fornaci di Briosco, Baggero, and Pusiano were heavily impacted. This catastrophic event triggered a comprehensive reassessment of flood hazard maps, early warning systems, and the resilience of hydraulic infrastructure throughout the basin [41].

.

Figure 5. 26 November 2002: the flooding of the Lambro river in Agliate, a hamlet of Carate Brianza (MI)

These flood events underline the recurring vulnerability of the Lambro Basin, especially in urban zones where land use changes, reduced infiltration, and inadequate hydraulic sections worsen the impact of natural hydrological variability. Continued efforts are being made to integrate early-warning systems, restore natural retention areas, and modernize hydraulic infrastructure, particularly in areas downstream of Lake Pusiano, which plays a partial role in flow regulation through the Cavo Diotti outlet system.

2.3. Monitoring Networks

Understanding the characteristics and origins of the data used in this study is essential for assessing the reliability and applicability of the subsequent analyses. This section provides an overview of the two primary data sources employed: the SIDRO database maintained by ARPA Lombardia, which serves as the official meteorological monitoring system, and data provided by MeteoNetwork, a citizen science initiative. For simplicity, these sources will hereafter be referred to as ARPA and METEO, respectively.

2.3.1. ARPA (Agenzia Regionale per la Protezione Ambientale)

The ARPA meteorological network [42], officially operated by the Regional Hydro-Meteorological Service of ARPA Lombardia since 2004, serves as a critical public resource for meteorological forecasting and real-time monitoring in the Lombardy region. It supports a catchment population of over 10 million inhabitants and provides essential meteorological information to civil protection authorities and the general public.

ARPA's network consists of 318 automatic and continuously operating stations, updated every 10 minutes. These stations are distributed with an average density of approximately 1 station per 200 km², and they function reliably even in complex terrains, ranging in altitude from 10 to 3000 meters above sea level. The system is representative at the mesoscale, capturing regional meteorological variations with high spatial and temporal resolution.

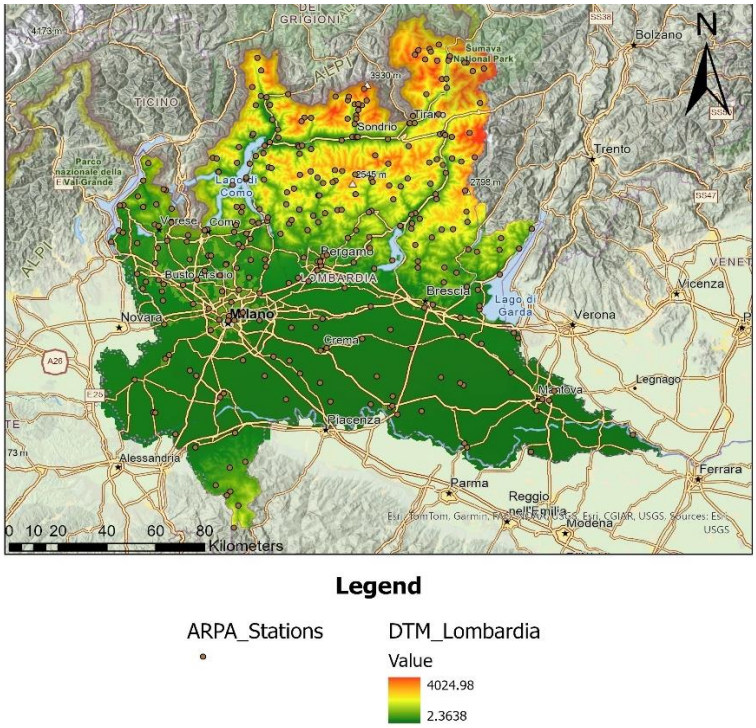


Figure 6. ARPA stations across Lombardia Region

The network records a comprehensive set of parameters, including temperature, precipitation, atmospheric pressure, global and net solar radiation, wind speed and direction, relative humidity, and present weather conditions. Data is transmitted via GPRS (General Packet Radio Service) with dedicated radio backup, ensuring robust data availability.

All acquired data undergo systematic validation and archiving and are made publicly accessible in both raw and processed forms. These processed outputs include daily, weekly, monthly, and annual bulletins. Meteorological forecasts are developed by a team of professionals with expertise in physics, meteorology, engineering, and geology. These forecasts are based on numerical meteorological models and observational data and are regularly updated to meet the specific needs of various users, including those involved in natural hazard mitigation, water resource management, and environmental protection.

2.3.2. METEO (Meteonetwork – Citizen Science Meteorological Network)

The MeteoNetwork (METEO) is a prominent example of citizen science applied to meteorology in Italy. In general, the term "citizen science" refers to scientific research carried out by non-professionals, often volunteers or enthusiasts, who contribute data, observations, or analysis to scientific projects [43]. This model, a form of crowdsourcing, has gained increasing relevance in recent years, supported by the availability of affordable technology and wide access to digital tools [43].

Although citizen science presents an opportunity for public engagement and data collection on a large scale, it also introduces challenges related to data quality, standardization, and reliability. Data from such networks are often collected in non-regulated or non-standardized environments, which may result in inconsistencies or missing data. Therefore, data validation procedures are essential to ensure the scientific credibility and usability of the information collected [44].

MeteoNetwork is a non-profit association, founded in 2002 by atmospheric science and meteorology enthusiasts. According to its mission, MNW aims to "develop and disseminate, for the benefit of the public and the scientific community, knowledge in the fields of meteorology, climatology, environmental science, hydrology, and glaciology" [45].

As of January 2022, the Meteonetwork comprises over 6,500 meteorological stations in 42 countries worldwide, with around 4,780 stations uploading data at least once every 24 hours, and approximately 3,400 stations providing continuous data throughout the day [44]. Most of these stations are operated by individuals or local groups. Over the last two decades, the network has grown steadily, by around 150 new stations per year, offering real-time, high-frequency point observations [44].

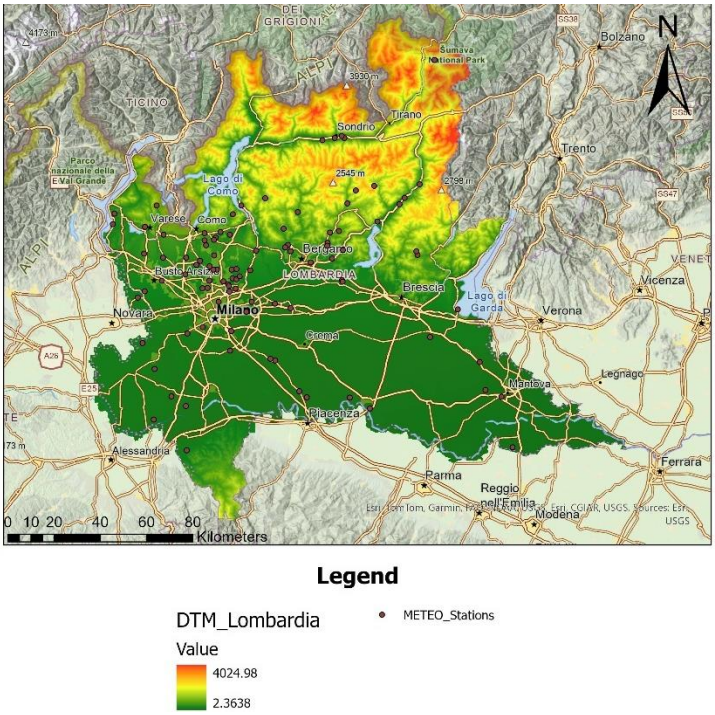


Figure 7. METEO stations across Lombardia Region

The MeteoNetwork platform also provides interpolated weather maps with a 20-minute temporal resolution, as well as daily summaries for various meteorological variables including temperature, precipitations, wind, pressure, relative humidity and dew point.

For this study, only the MNW stations located within the defined study area (see section 2.1) were selected. For each station, hourly cumulative precipitation data were available. However, due to the amateur and volunteer nature of the network, the dataset included a notable number of missing values.

2.3.3. Data Collection

For ARPA data, rainfall time series were retrieved by identifying and selecting monitoring stations located within the SLO basin (Seveso-Lambro-Olona) as delineated on the SIDRO portal. The rainfall stations were selected based on two main criteria: the availability and completeness of historical data, and their spatial distribution across the SLO basin. This approach ensured that the chosen stations not only had consistent and reliable time series but also collectively represented different sectors of the basin, such as upstream, midstream, and downstream areas. This spatial diversity was crucial for capturing the rainfall variability across the entire catchment.

METEO data, which covered a broader portion of the Lombardy region, were obtained in bulk through email correspondence with the organization. These data were then processed to retain only those stations falling within the boundaries of the SLO basin. This allowed for consistency in spatial coverage between the two datasets.

2.3.4. Coverage and Missing Data Assessment

In this study, the use of the ARPA and METEO datasets was essential to achieve a more comprehensive and reliable understanding of rainfall patterns across the Lambro-Seveso-Olona basin. ARPA data offers long-term consistency and reliability, making it ideal for high-resolution temporal analysis. However, the ARPA network might not cover all areas or specific events, particularly in more localized regions. To address these gaps, the METEO dataset was utilized, which, while exhibiting some data gaps, provides broader geographic coverage and a more extensive network of stations. This dual-source approach not only enhanced the geographical coverage but also allowed for cross-validation between the two datasets. Furthermore, the joint use of data allowed a more complete and accurate assessment of rainfall dynamics, leveraging the strengths of each dataset to fill gaps in the other and providing a more robust foundation for the analysis.

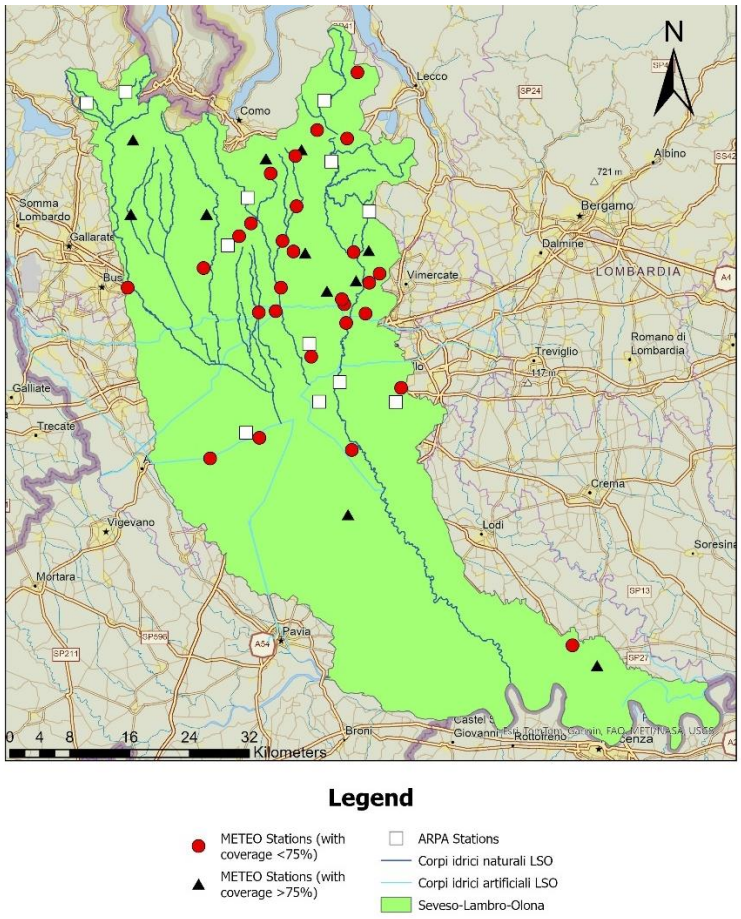


Figure 8. Layout of ARPA and METEO Rain gauges within Study Area

The ARPA dataset presented temporal inconsistencies across different stations. While some stations had data extending back to 1987, others began much later. However, within the operational period of each station, data coverage is generally consistent and of good quality.

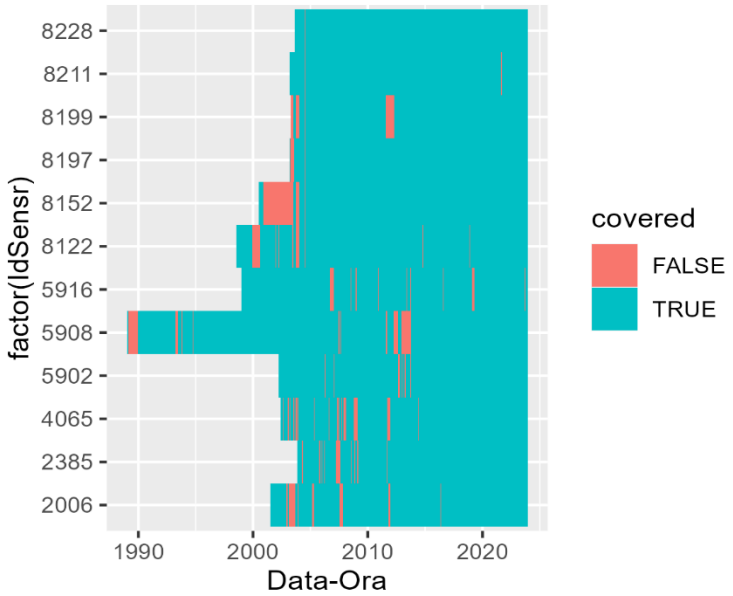


Figure 9. Rainfall Data available for ARPA Stations within Study Area

Conversely, METEO data coverage is limited to the period from 2013 to 2023, and the completeness of records varies significantly across individual stations. Some provide nearly continuous data, while others have frequent gaps or shorter observation periods.

To enable a coherent and fair comparison between the two datasets, the ARPA data was trimmed to cover the same time-period as METEO, i.e., from January 2013 to December 2023. This harmonized time frame ensures that analyses reflect similar temporal dynamics and avoid biases due to uneven data availability.

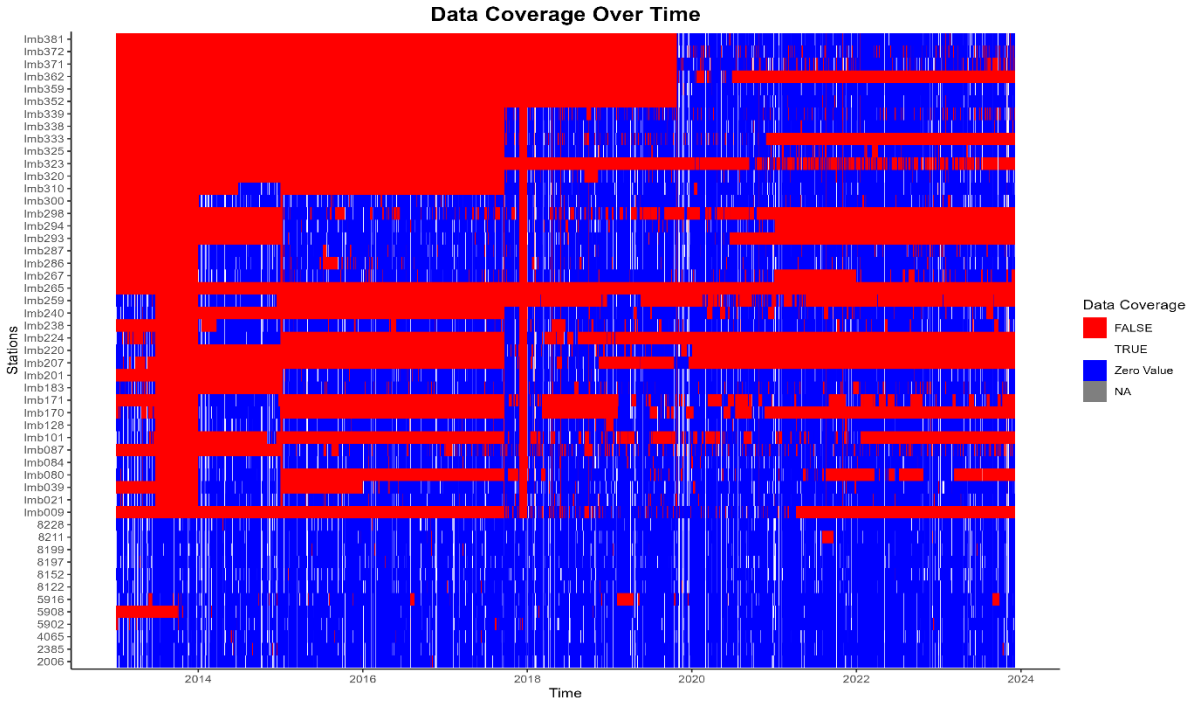


Figure 10. Rainfall Coverage of Stations over Time (2013-2024)

Figure 10 reveals substantial differences in data continuity and completeness between the two sources. ARPA stations demonstrate consistent long-term data availability from 2013 to 2024, with minimal gaps, making them highly reliable for long-term analysis. In contrast, many METEO stations show sparse and fragmented data coverage, particularly before 2018, with extended periods of missing or unavailable data. This disparity highlights the need for careful selection of stations in subsequent analysis and potential use of imputation techniques for missing periods, especially for MeteoNetwork data, ensuring spatial consistency across the SLO basin.

Figure 11 and Figure 12 provide a detailed breakdown of missing data percentages for ARPA and METEO stations, respectively. ARPA stations demonstrate exceptionally low missing data, with most stations showing less than 1% missing data. Only a few stations, such as 5908 and 5916, exceed 4%. In contrast, METEO stations display considerable variability in data completeness. Some stations, like lmb255 and lmb265, have more than 90% of their data missing, while only a few stations have missing data below 20%. This stark contrast underscores the higher reliability of the ARPA dataset for long-term or high-resolution analyses.

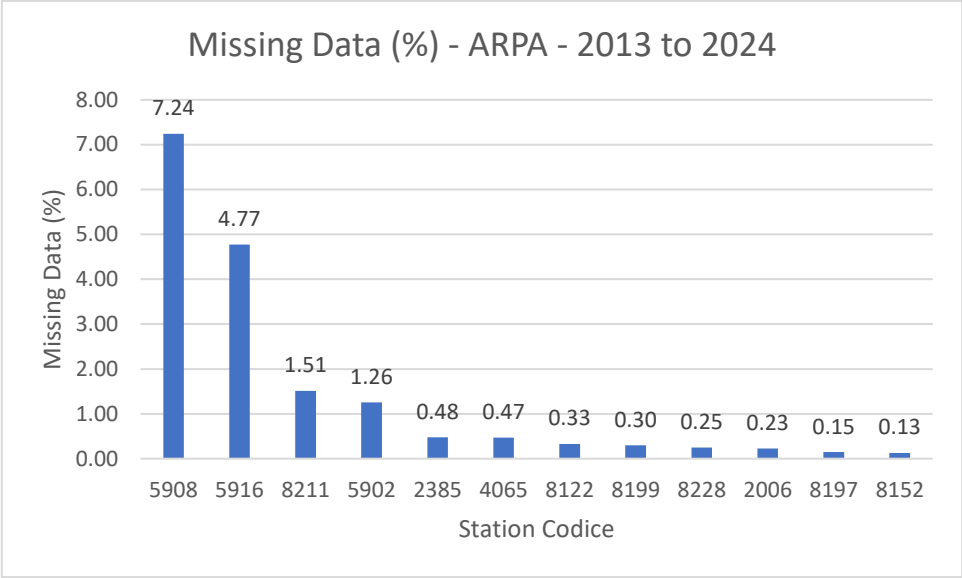


Figure 11. Missing Data Percentage for ARPA Stations

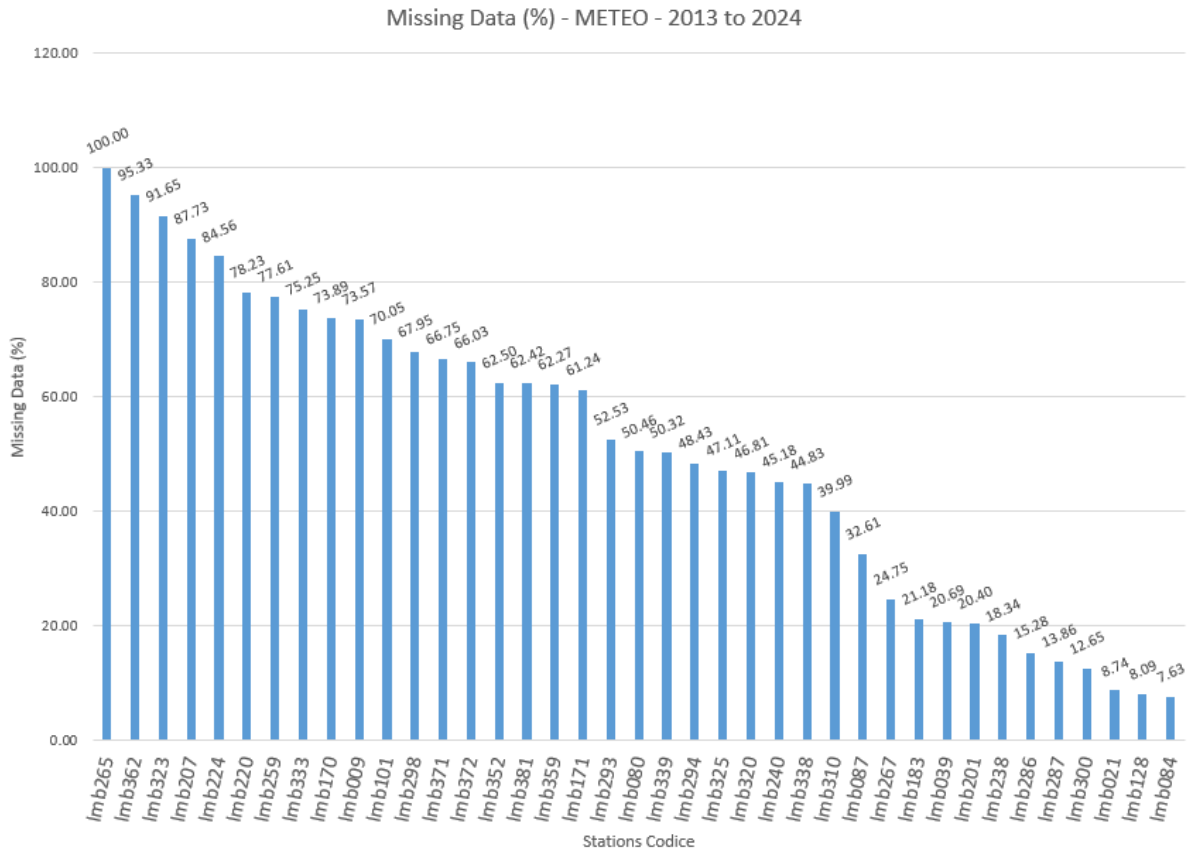


Figure 12. Missing Data Percentage for METEO Stations

Although METEO stations exhibit variable and poorer temporal coverage, the data they provide can still play an important role in supporting the analysis. When present, these data appear to be reasonably consistent and of sufficient quality to potentially support the imputation of missing values at nearby ARPA stations. Moreover, the

greater spatial density of METEO stations enhances the overall spatial resolution of the dataset, allowing for a more detailed representation of local meteorological variability across the study area. These aspects will be further illustrated and discussed in the next chapter.

3 Chapter Three: Spatial Correlation of Rainfall Data

3.1. Rainfall Data Preprocessing

The first step in preparing the rainfall dataset involved filtering the raw hourly data collected from approximately 50 stations across the region, for both ARPA and METEO sources from 2013 to the end of 2023.

In fact, upon inspection, it was observed that certain METEO stations reported implausibly high hourly rainfall values, occasionally exceeding 100 mm/h. Such values are physically unrealistic and were likely the result of sensor malfunctions or recording errors. Cross-validation with nearby ARPA stations, which showed no comparable peaks, supported this conclusion. Considering the regional rainfall characteristics, a decision was made to cap all hourly rainfall values exceeding 60 mm to 60 mm. This threshold was chosen as a conservative upper limit, consistent with extreme yet plausible rainfall intensities in SLO basin, in accordance with the official ARPA data.

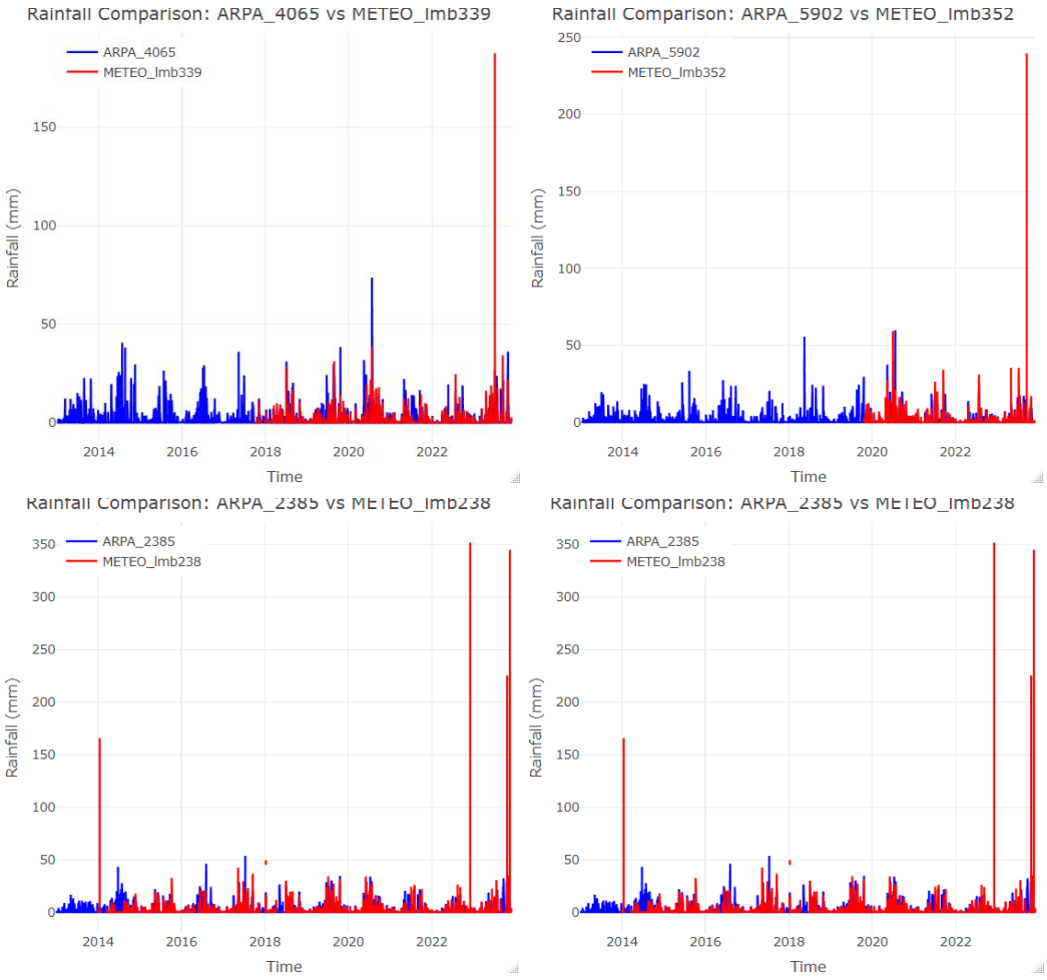


Figure 13. Example comparisons of rainfall time series (mm/h) between nearby ARPA and METEO stations, highlighting anomalous peaks in METEO data

Additionally, the METEO dataset contained some negative rainfall values, which are non-physical and indicate errors in data recording or transmission. Since rainfall cannot be negative, all such values were replaced with zero and considered as representing dry periods.

To better understand the distribution of the outliers, Figure 14 presents a summary of their counts per station, categorized by data source and type. The figure shows that the number of outliers was negligible. These filtering steps were, however, performed to ensure the reliability and consistency of the data before further analysis and inter-station comparisons.

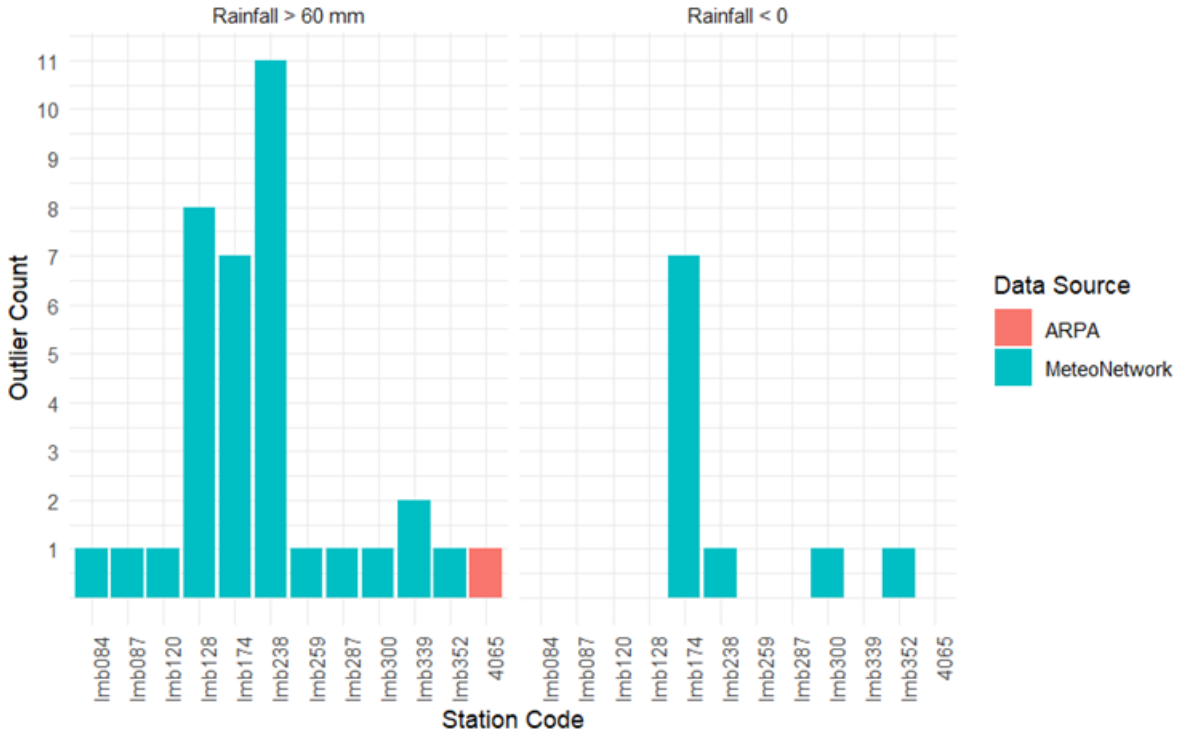


Figure 14. Outlier counts per station from ARPA and METEO sources, categorized by typology: rainfall values exceeding 60 mm/h (left) and negative rainfall values (right)

3.2. Station Pairing, Group Formation and Similarity Analysis

To evaluate spatial correlation patterns across the study region, all stations were paired systematically, thus generating all the possible pairwise combinations of stations within the study area.

Furthermore, the created pairs were divided into three groups based on the data source combinations:

- **ARPA–ARPA**: 66 pairs between ARPA stations only, generally characterized by high-quality and reliable measurements.
- **METEO–METEO**: 741 pairs between METEO stations only, which required closer scrutiny due to occasional data gaps or anomalies.
- **ARPA–METEO**: 467 pairs, consisting of one ARPA station and one METEO station, where differences in instrumentation and quality control procedures may influence the consistency of the recorded data.

The purpose of creating these groupings was to compute and analyze correlation metrics including the Nash–Sutcliffe Efficiency (NSE), its inverse (computed by swapping observed and predicted values to account for NSE's asymmetry), and the Pearson correlation coefficient for each pair. These metrics were selected for their

ability to capture both linear and non-linear relationships, offering complementary perspectives on inter-station agreement. These metrics were then studied in relation to inter-station distance to observe general trends and decay patterns in correlation. With the station pairs organized into distinct groups, computing similarity metrics helped assess how rainfall patterns compared across locations. This analysis aimed to quantify how closely data from one station aligned with that of another, helping to characterize spatial variability in rainfall across the study area.

The following section outlines each metric and presents the formulas used for their computation.

3.2.1. Nash-Sutcliffe Efficiency (NSE)

Basic Definition

The Nash–Sutcliffe Efficiency (NSE) [46] is a statistical measure used to evaluate the predictive skill of a model or the agreement between two time series. Specifically, it quantifies how well the variation in one dataset (typically considered "simulated") aligns with another (typically considered "observed") by comparing the residual variance to the variance of the observed values.

In the context of rainfall data, NSE was used in this work as a tool to assess how closely the time series of one station corresponds to that of another. The NSE value can be interpreted as follows:

- **NSE = 1**: Perfect agreement between the two compared datasets.
- **NSE = 0**: The second dataset is as far from the reference dataset as the reference is from its own average.
- **NSE < 0**: The second dataset is even less similar to the reference dataset than its average is.

The formula is given by:

$$NSE = 1 - \frac{\sum_{t=1}^{n}(O_t - P_t)^2}{\sum_{t=1}^{n}(O_t - O_{mean})^2}$$

Equation 1. Nash-Sutcliffe Efficiency Equation

Where:

- O_t is the observed value at time t
- P_t is the predicted value at time t
- O_{mean} is the mean of the observed values
- n is the number of data points

Here O_t and P_t simply represent two datasets being compared, with O treated as the reference.

Methodology and Use in Station Pair Comparisons

To assess the consistency of rainfall records between different stations, NSE was applied to pairwise comparisons across three types of station pairings mentioned above.

For each pair, the rainfall series from one station was treated as the "observed" dataset and the other as the "simulated" dataset. This assignment was not based on a ground-truth hierarchy, since both series were based on real measurements rather than simulations. Furthermore, it was necessary to test the agreement in both directions.

Bidirectional Evaluation (NSE and NSE Inverse)

Given the absence of a true reference dataset, each station pair was evaluated in both directions:

- The standard NSE was computed by designating the first station as observed and the second as simulated.
- The NSE inverse was computed by swapping the roles—treating the second station as observed and the first as simulated.

This bidirectional approach was used to verify whether the relationship between correlation and distance remained consistent regardless of which station in the pair was treated as the reference. This check was particularly important for ARPA–METEO and METEO–METEO pairs, where potential differences in instrumentation or data quality could introduce directional bias, meaning variation in similarity depending on which station is considered "observed" versus "simulated." Given that ARPA stations are generally more standardized and better maintained, it was expected that NSE values might be slightly higher when ARPA served as the reference. However, the purpose of this analysis was to ensure that the similarity between station pairs remained broadly consistent in both directions. Ensuring that no significant asymmetry or abrupt shift occurred between standard and inverse NSE results provided the confidence needed to proceed with the analysis.

Summary of Purpose

The calculation of both NSE and inverse NSE for all station pairs was designed to establish a quantitative framework for evaluating inter-station similarity across the study area. This bidirectional approach allows for:

- Identifying pairs with strong or weak agreement.
- Understanding how consistently rainfall patterns are captured across stations.
- Evaluating which datasets or networks exhibit greater coherence.

3.2.2. Pearson Correlation Coefficient

Basic Definition

The Pearson Correlation Coefficient (PCC) [47] is a statistical measure that quantifies the linear relationship between two datasets. It assesses how closely the two datasets correlate with each other, indicating the strength and direction of their linear association.

The formula for the Pearson Correlation Coefficient is:

$$PCC(X,Y) = \frac{\sum_{i=1}^{n}(X_i - X_{mean})(Y_i - Y_{mean})}{\sqrt{\sum_{i=1}^{n}(X_i - X_{mean})^2 \sum_{i=1}^{n}(Y_i - Y_{mean})^2}}$$

Equation 2. Pearson Correlation Coefficient Equation

Where:

- X_i and Y_i are the individual data points of datasets X and Y.
- X_{mean} and Y_{mean} are the means of datasets X and Y.
- n is the number of data points.

Unlike the NSE, which was computed in both directions (observed vs simulated and vice versa), the PCC **is** symmetric**.** This symmetry exists because the formula for PCC involves the product of deviations from the mean for both variables in the numerator, and the square roots of their variances in the denominator. Since multiplication is commutative, swapping the variables does not change the value of the correlation.

$$\frac{\sum_{i=1}^{n}(X_i - X_{mean})(Y_i - Y_{mean})}{\sqrt{\sum_{i=1}^{n}(X_i - X_{mean})^2 \sum_{i=1}^{n}(Y_i - Y_{mean})^2}} = \frac{\sum_{i=1}^{n}(Y_i - Y_{mean})(X_i - X_{mean})}{\sqrt{\sum_{i=1}^{n}(Y_i - Y_{mean})^2 \sum_{i=1}^{n}(X_i - X_{mean})^2}}$$

$$PCC(X, Y) = PCC(Y, X)$$

Equation 3. Commutative Property of Pearson Correlation Coefficient

Analysis of Results

After computing the Pearson Correlation Coefficient for each pair of stations, the results are analyzed to evaluate the strength of the linear relationship between the station pairs, the consistency of data quality and behavior across the ARPA and METEO stations, and the overall similarity in rainfall patterns across different stations in the region.

A PCC value close to 1 indicated a strong positive linear relationship, suggesting that the rainfall patterns at the two stations were highly correlated. A PCC value close to 0 indicated no significant linear relationship, meaning that the rainfall patterns at the

two stations were largely independent. Importantly, negative values were virtually absent in the results.

3.3. Correlation Trends Between Station Pairs

To further support the evaluation of similarity metrics, comparative graphs were plotted across various scenarios to visualize how the metrics changed as the distance between station pairs increased, providing insights into spatial patterns of correlation across the region.

3.3.1. Methodological Approach

This section assesses how rainfall similarity varies across stations by comparing time series using the three widely accepted statistical metrics described before. The analysis is structured around a consistent framework defined by five key dimensions: station-pair group, similarity index, temporal aggregation, seasonal window, and zero filtering (see Table 1). These dimensions define the scope and configuration of each plot, allowing for systematic comparison across a wide range of scenarios.

Dimension	Description	Values Considered
Station-Pair Group	Defines the network pairing used in the comparison.	ARPA-ARPA, METEO-METEO, ARPA-METEO
Similarity Index	Statistical metric used to quantify similarity between station time series.	NSE, Inverse NSE, PCC
Temporal Aggregation	Time scale over which rainfall is accumulated before comparison.	Hourly, Daily
Seasonal Window	Time period over which the metric is computed.	Full Year, Summer (June-August)
Zero Filtering	Whether time steps with zero rainfall at both stations are included.	All Time Steps (with zeros), Wet Only (without zeros)

Table 1. Dimensions used in spatial similarity analysis plots

To ensure clarity without oversaturation, figures are designed to vary one or two dimensions at a time while holding others fixed. This approach enables direct comparisons that isolate the effects of factors like station type, temporal resolution, or seasonal conditions.

In addition to analyzing the full time series of rainfall data, similarity metrics (NSE, inverse NSE, and Pearson Correlation Coefficient) were also computed exclusively for the summer months (June, July, and August). This seasonal subset was chosen because summer often features convective rainfall events, which tend to be localized and intense, in contrast to the broader and more uniform precipitation seen in colder months.

The rationale for this approach was twofold:

1. **Capture Seasonal Variability:** Rainfall characteristics differ significantly between seasons. By isolating summer months, we aimed to understand how the spatial correlation structure and inter-station similarity change under different atmospheric conditions.
2. **Test Robustness of Similarity Metrics:** Evaluating metrics under season-specific conditions allows for the assessment of whether certain station pairs maintain strong similarity year-round or only during specific weather patterns.

Rather than relying on scatter plots alone, this section employs probability density functions (PDFs) to statistically summarize the distribution of similarity values across many station pairs within defined distance intervals. For each distance threshold, a range of ±3000 m was applied to include station pairs falling within a reasonable proximity window. This approach balances granularity with sufficient sample size, ensuring that the PDFs represent meaningful and statistically supported trends. PDFs provide a more informative and aggregated view of spatial correlation patterns, particularly when comparing network performance or assessing the influence of data filtering.

To contextualize the reliability of each PDF, station-pair counts corresponding to each distance class are reported separately (see Figure 15). These counts reflect the number of pairings available in each group-distance combination and help interpret how robust the resulting distributions are, especially where group sizes differ. It is encouraged to consider these counts when evaluating differences in PDFs of similarity metrics between networks or distance classes.

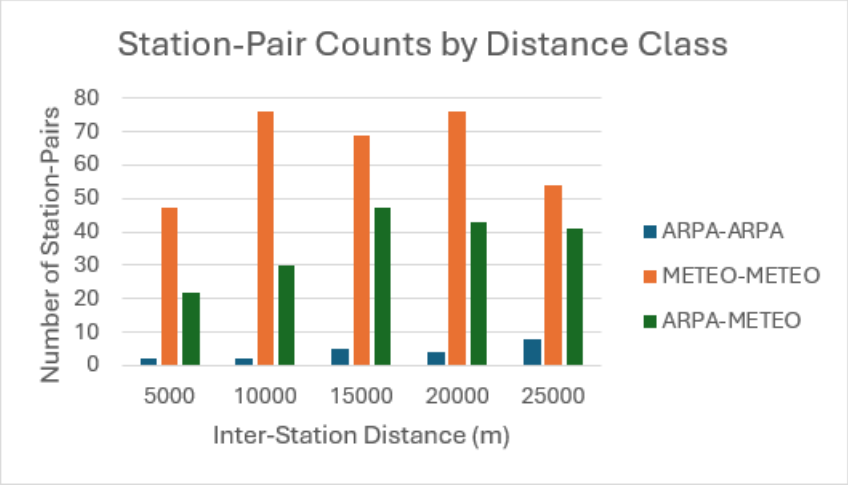


Figure 15. Station-Pair Counts by Distance Class

3.3.2. Influence of Inter-Station Distance on Rainfall Similarity

One of the most fundamental factors influencing the agreement between rainfall stations is their spatial separation. This subsection explores how station similarity changes with inter-station distance, using three well-established statistical metrics, as mentioned before. The analysis is based on the entire available time series, including both rainfall and non-rainfall periods, to reflect the full range of observational behavior across station networks.

Distance-Decay Behavior Across Metrics

Figure 16 - Figure 18 illustrate how station similarity varies with inter-station distance for three different pair types: ARPA–ARPA, ARPA–METEO, and METEO–METEO. Across all metrics, a consistent pattern emerges in which similarity between stations tends to decline as the distance between them increases. However, the rate and clarity of this decline differ depending on the station pairing and the specific similarity metric used.

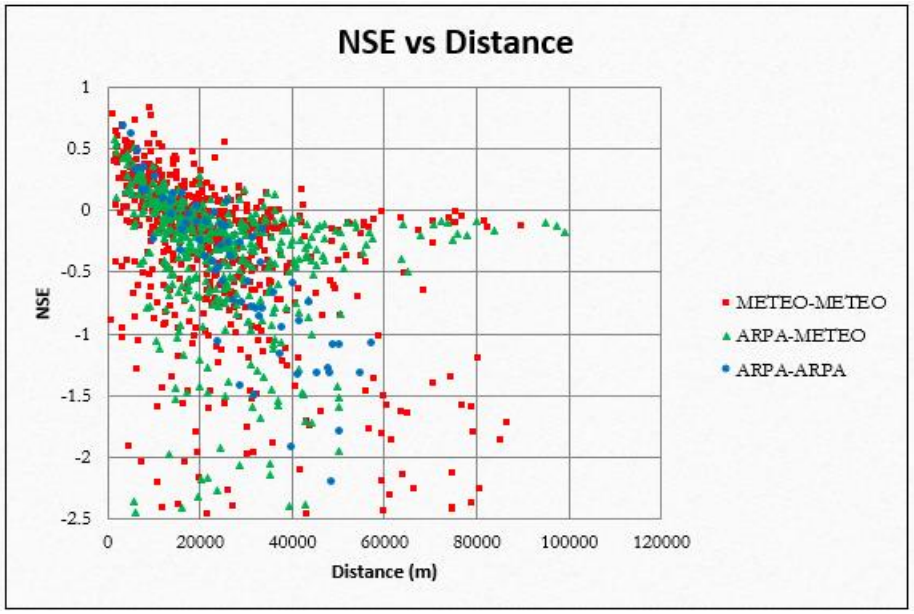


Figure 16. Nash-Sutcliffe Efficiency vs Distance between station pairs (Full Year - Hourly)

Figure 16 demonstrates a clear decline in station similarity as inter-station distance increases. ARPA-ARPA pairs tend to exhibit relatively positive NSE values within the first 20 kilometers. However, it can be seen that as the distance increases, their similarity gradually diminishes, reflecting a weakening temporal alignment. ARPA–METEO pairs follow a similar but more variable pattern, with moderate NSE values that also decay with distance. In contrast, METEO–METEO pairs exhibit a widespread across the entire distance range, with NSE values falling below zero regardless of proximity, indicating persistently poor agreement that is only marginally influenced by distance.

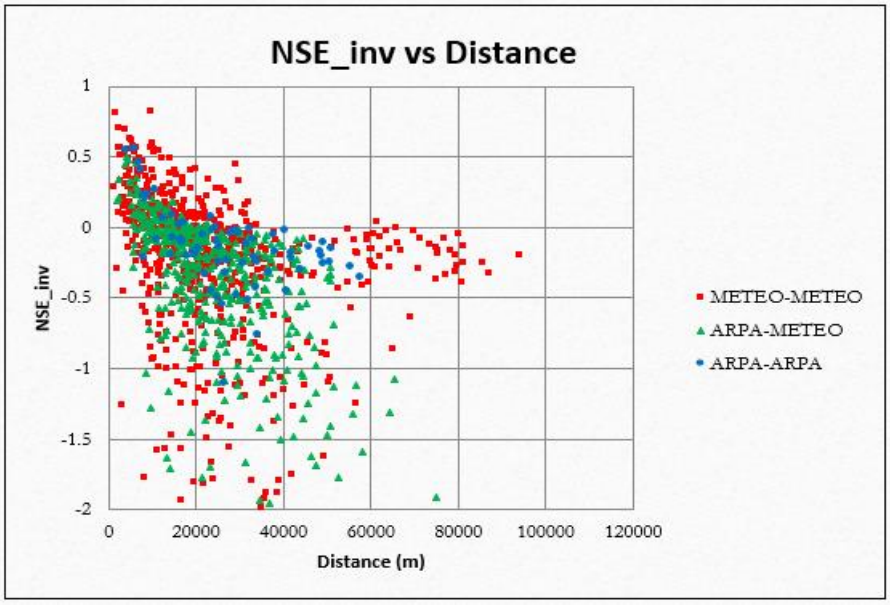


Figure 17. Inverse Nash-Sutcliffe Efficiency vs Distance between station pairs (Full Year - Hourly)

Figure 17 employs an inverse NSE metric, which inverts the direction of the efficiency measure. Despite the transformation, the general structure remains similar. The values tend to decrease with increasing distance, and noticeable clustering is observed at shorter distances.

However, compared to the standard NSE plot (see Figure 16), the distribution for ARPA–METEO pairs (green) appears slightly more dispersed, with a greater number of points falling below zero. ARPA–ARPA pairs (blue) retain a relatively compact distribution, though with a minor downward shift in some cases. METEO–METEO pairs (red) continue to show the widest spread, and their distribution appears largely unchanged from the original NSE plot, with many values remaining below zero even at close distances. The overall visual pattern is consistent with the original NSE-distance relationship, though small directional differences are visible, particularly for pairs involving METEO stations.

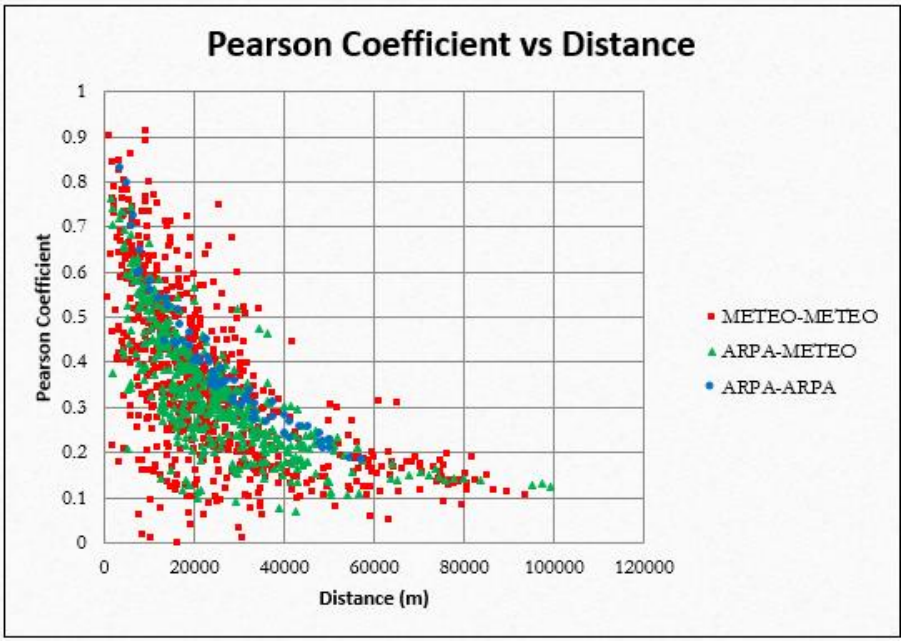


Figure 18. Pearson Correlation Coefficient vs Distance between station pairs (Full Year - Hourly)

In Figure 18, as expected, the highest correlation values are observed at short distances, particularly within 20 kilometers, where ARPA–ARPA pairs (blue) form a dense cluster above 0.6. ARPA–METEO pairs (green) follow a similar pattern, though with slightly lower and more dispersed values. In contrast, METEO–METEO pairs (red) display the widest range of values and generally lower correlations, with many points falling below 0.3 even at short distances, reinforcing their comparatively weaker performance across the analysis. The overall trend reveals a clear negative gradient across all pair types, reflecting the spatial decay of rainfall similarity.

Distributional Patterns by Distance

Figure 19 - Figure 21 display the probability density functions (PDFs) of station similarity metrics across increasing inter-station distance classes (from 5,000 m to 25,000 m), for the three station pair types. These plots offer a statistical perspective on how station similarity is distributed within each distance band and allow for a more nuanced interpretation of how network performance and spatial coherence change with separation.

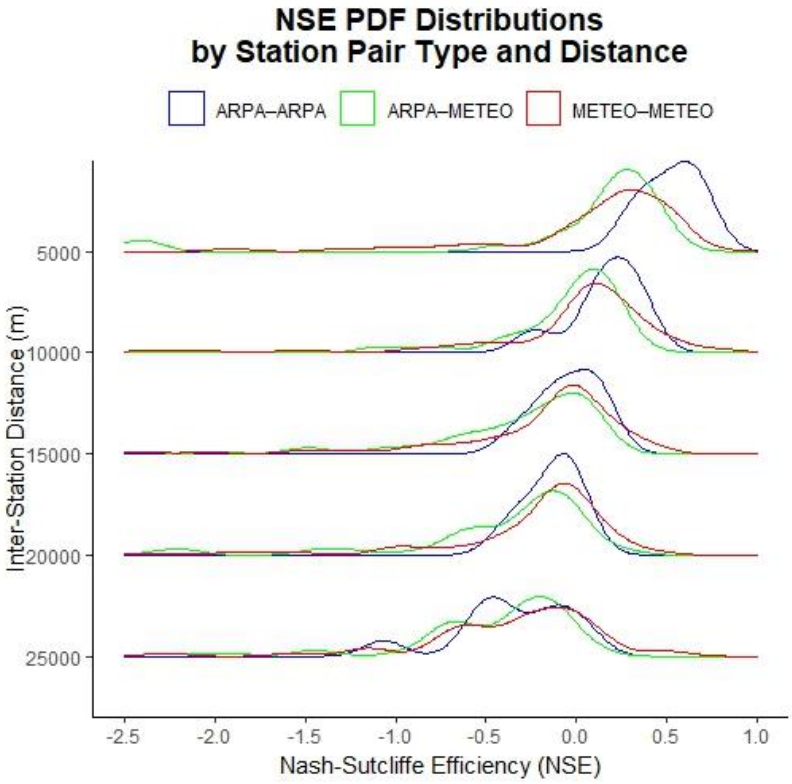


Figure 19. Probability density functions of Nash–Sutcliffe Efficiency (NSE) values for different station pair types (Full Year – Hourly)

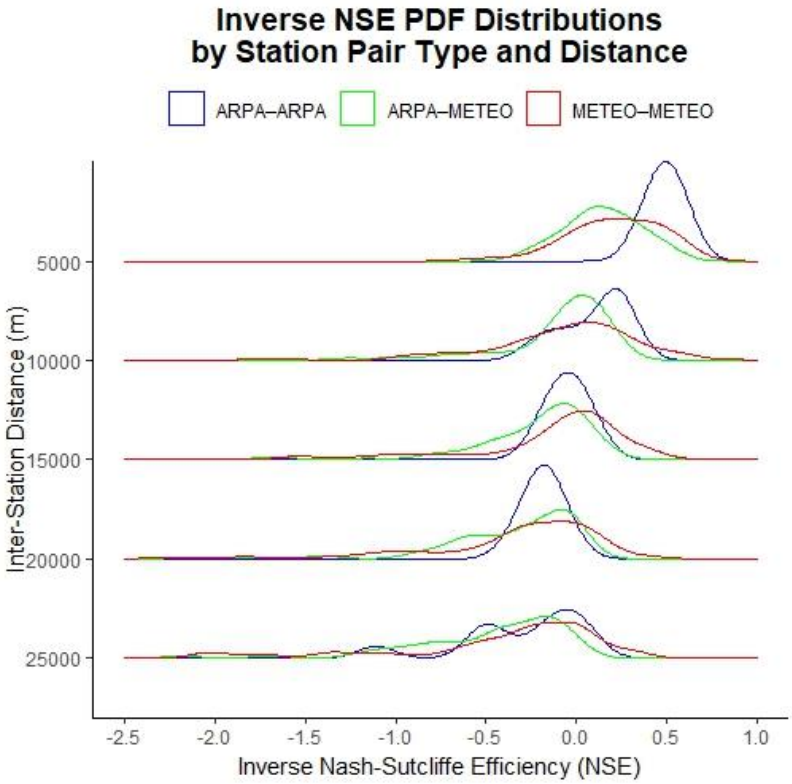


Figure 20. Probability density functions of Inverse Nash–Sutcliffe Efficiency (NSE) values for different station pair types (Full Year – Hourly)

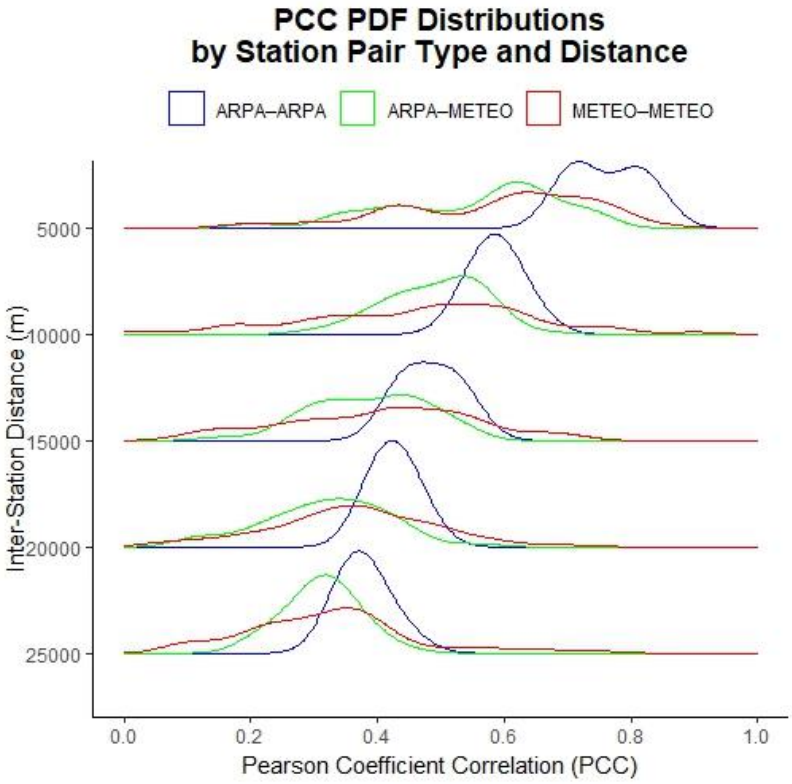


Figure 21. Probability density functions of Pearson Correlation Coefficient (PCC) values for different station pair types (Full Year – Hourly)

Across all three metrics, the evolution of the probability density curves reveals a consistent structural transformation with increasing distance i.e., distributions become progressively flatter, broader, and shift leftwards. This trend is most pronounced for ARPA–ARPA pairs, where tightly peaked curves at short range give way to dispersed profiles at greater distances, indicating a gradual loss of coherence. ARPA–METEO curves remain more variable in shape, reflecting mixed instrumentation, while METEO–METEO distributions appear diffuse and weakly defined across all distance classes. Overall, the shape and spread of these curves provide a visual summary of the spatial fragmentation in station behavior, offering statistical depth beyond point-based scatter plots and emphasizing how distance affects not just average similarity, but its reliability and distribution.

These findings emphasize that spatial proximity is a critical factor in assessing the reliability of station comparisons for data imputation. Since rainfall similarity decreases with distance, using nearby stations is essential when estimating missing values. This distance–similarity relationship provides an objective basis for selecting suitable neighboring stations and ensures that imputed values preserve the spatial structure of rainfall patterns.

3.3.3. Impact of Zero Filtering on Station Similarity

Rainfall datasets often contain extended periods where no precipitation occurs, resulting in valid but zero-valued measurements. While these periods are meteorologically meaningful, their inclusion in similarity analyses can obscure the differences in how stations record actual rainfall events. When both stations register zero simultaneously, similarity metrics may suggest high agreement, even though there is no event to compare. To better assess how stations captured data when rainfall occurred, similarity metrics were recalculated under two conditions: one using the full time series (including zeros) and the other using only time steps where both stations recorded rainfall greater than zero.

This subsection examines how filtering out dry periods influences the behavior of similarity metrics and whether it reveals more consistent and interpretable patterns of inter-station agreement during active precipitation.

Effect on Similarity Metrics

Figure 22 - Figure 33 compares similarity distributions for the three station pair types under filtered conditions.

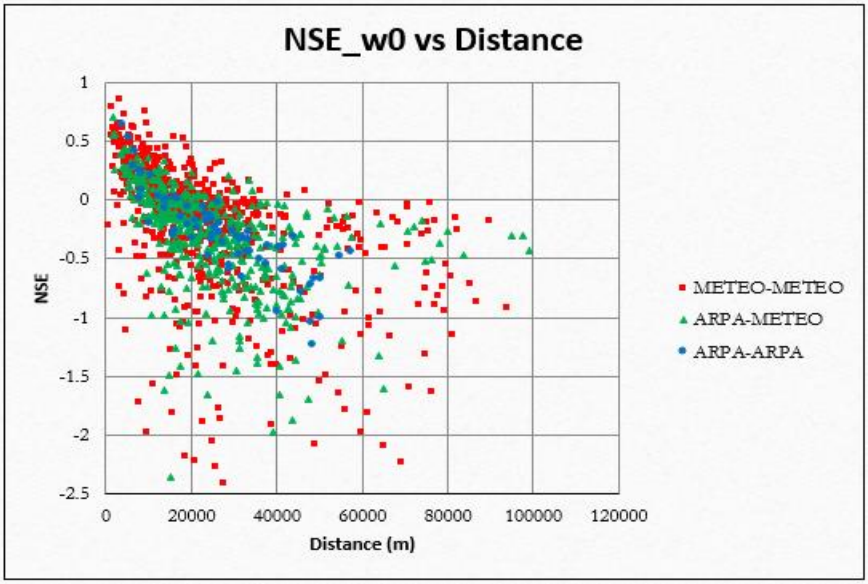


Figure 22. Nash-Sutcliffe Efficiency (without zero rainfall values) vs Distance between station pairs

Figure 22 considers only time steps in which both stations recorded rainfall. Compared to Figure 16, which included all time steps, this filtered analysis results in a slight upward shift in NSE values across all station pair types. The overall structure of the relationship remains consistent, with NSE declining as distance increases. However, the removal of dry-period data appears to improve the clarity of the signal, particularly for ARPA–ARPA pairs, which seem to cluster more tightly in the positive NSE range. ARPA–METEO pairs also show improved alignment, though with greater

variability. METEO–METEO pairs continue to perform the worst, exhibiting a wide spread of mostly negative NSE values, though some short-distance pairs show modest improvement.

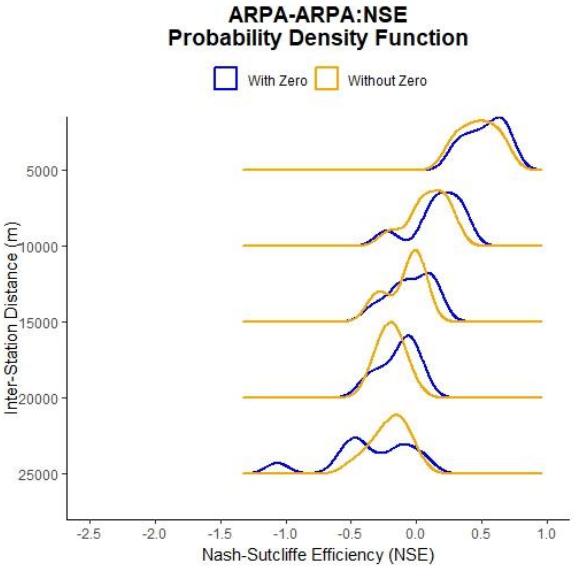


Figure 24. PDFs of NSE values - ARPA–ARPA (Hourly-Yearly-Without Zero)

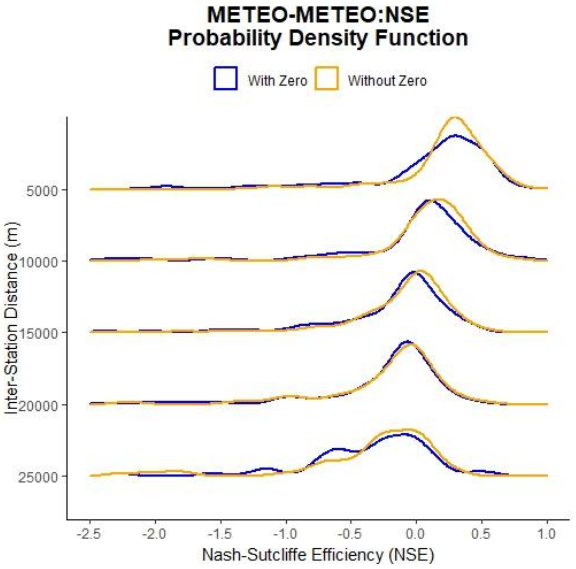


Figure 23. PDFs of NSE values - METEO–METEO (Hourly-Yearly-Without Zero)

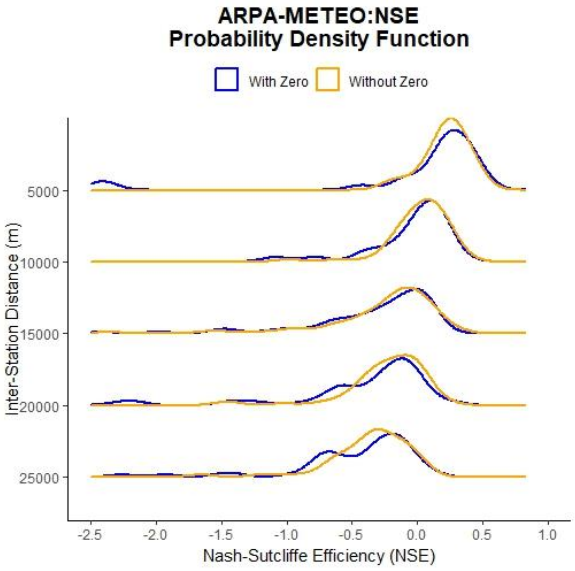


Figure 25. PDFs of NSE values - ARPA–METEO (Hourly-Yearly-Without Zero)

In addition to the scatter view, Figure 23 - Figure 25 present the probability density functions (PDFs) of NSE values across various distance classes, comparing both filtered (wet-only) and unfiltered (all-data) conditions. These distributions complement the upward shift across various station types in scatter plots by statistically summarizing how similarity metrics behave across station types and spatial separations. Zero filtering enhances the clarity of inter-station similarity during

rainfall events, helping to isolate meaningful differences in how stations respond to precipitation.

For ARPA–ARPA pairs, the filtered PDFs at short distances become more peaked and slightly shift leftward compared to the unfiltered case. This subtle left shift can be attributed to the limited number of ARPA stations, which restricts the number of close-range pairs. As inter-station distance increases, rainfall similarity diminishes, not due to poor data quality but due to sparse spatial coverage. In contrast, ARPA–METEO and METEO–METEO pairs show more consistent behavior between filtered and unfiltered conditions. The higher number of METEO stations enables denser spatial pairing, which helps maintain a stable structure in the similarity distributions. While filtering suppresses inflated agreement from dry periods, it does not drastically change the shape or spread of the PDFs, particularly at short and mid-range distances.

Overall, even after zero filtering, though there can be seen slight upward shifts in scatter plots, the overall domain within which NSE values vary at each distance remains largely unchanged across all station types as seen in the PDFs. This stability indicates that the underlying spatial coherence in rainfall patterns is preserved, and filtering primarily serves to enhance the interpretability of the agreement without distorting the fundamental structure of the inter-station similarity.

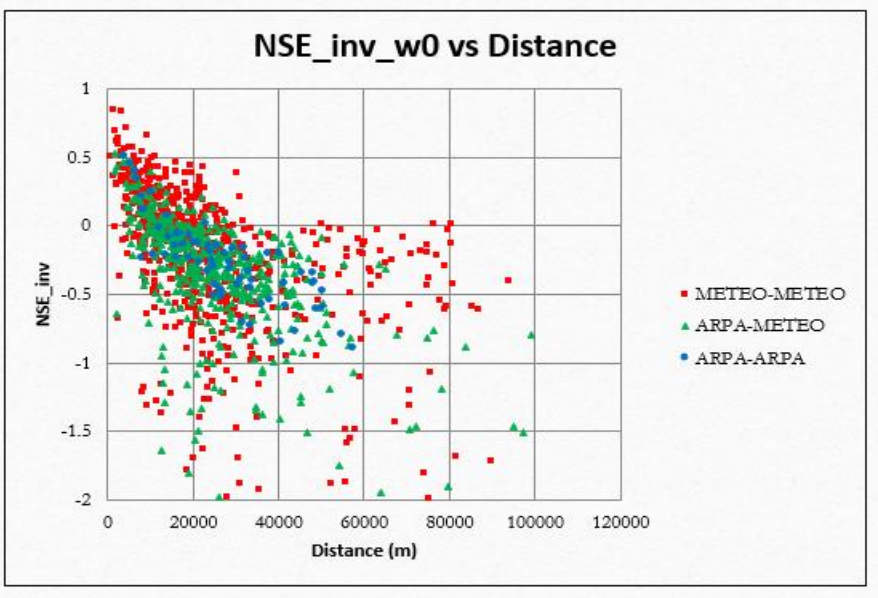


Figure 26. Inverse Nash-Sutcliffe Efficiency vs Distance between station pairs (Hourly-Yearly-Without Zero)

As in the standard NSE case, this filtering results in a general upward shift in NSE values across all station pair types, with a greater concentration of points near or above zero, at shorter distances. Moreover, the exclusion of dry periods yields a cleaner distribution and reduced scatter, particularly for station pairs at shorter distances. ARPA–ARPA pairs remain tightly clustered, while ARPA–METEO pairs show

moderate improvement. METEO–METEO pairs continue to exhibit the widest spread and lowest values, though more points appear closer to zero within the first 20–40 km. However, the downward trend with distance remains persistent, but the filtering sharpens the distance–similarity pattern by minimizing the influence of zero-rainfall intervals improving the interpretability of inter-station similarity metrics.

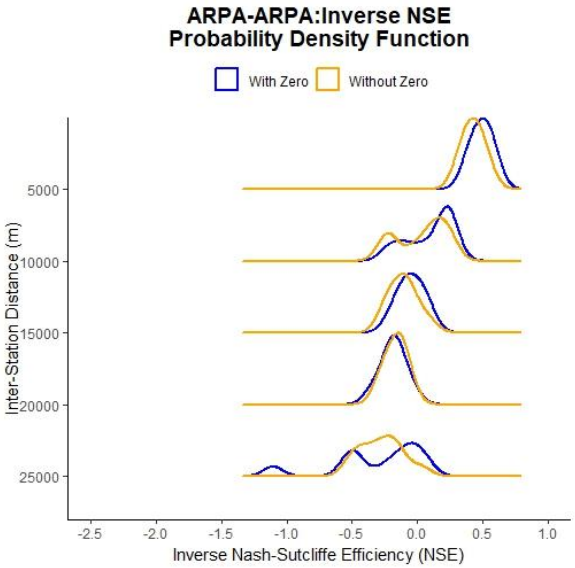


Figure 28. PDFs of Inverse NSE values - ARPA–ARPA (Hourly-Yearly-Without Zero)

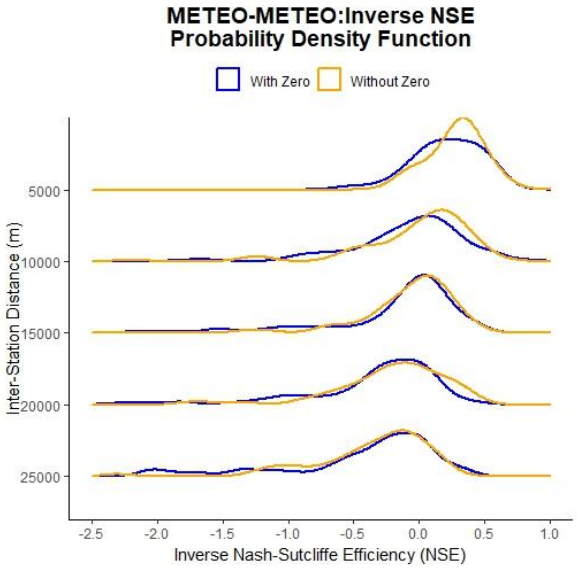


Figure 27. PDFs of Inverse NSE values - METEO-METEO (Hourly-Yearly-Without Zero)

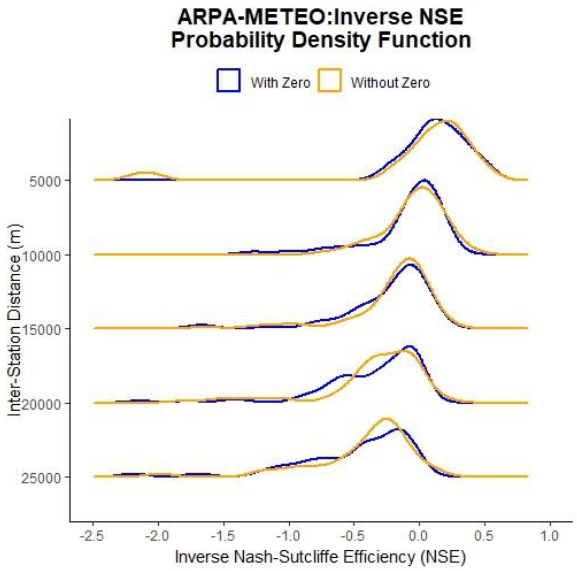


Figure 29. PDFs of Inverse NSE values - ARPA–METEO (Hourly-Yearly-Without Zero)

The results for Inverse NSE closely mirror those observed for standard NSE. The scatter plots (Figure 26) show slight upward shifts, particularly at shorter distances, while the PDFs **(**Figure 27 – Figure 29**)** exhibit overall structural consistency between

filtered and unfiltered data. Overall, the behavior of Inverse NSE statistically complements the NSE analysis, reinforcing that zero filtering enhances interpretability without altering the core spatial coherence of rainfall patterns.

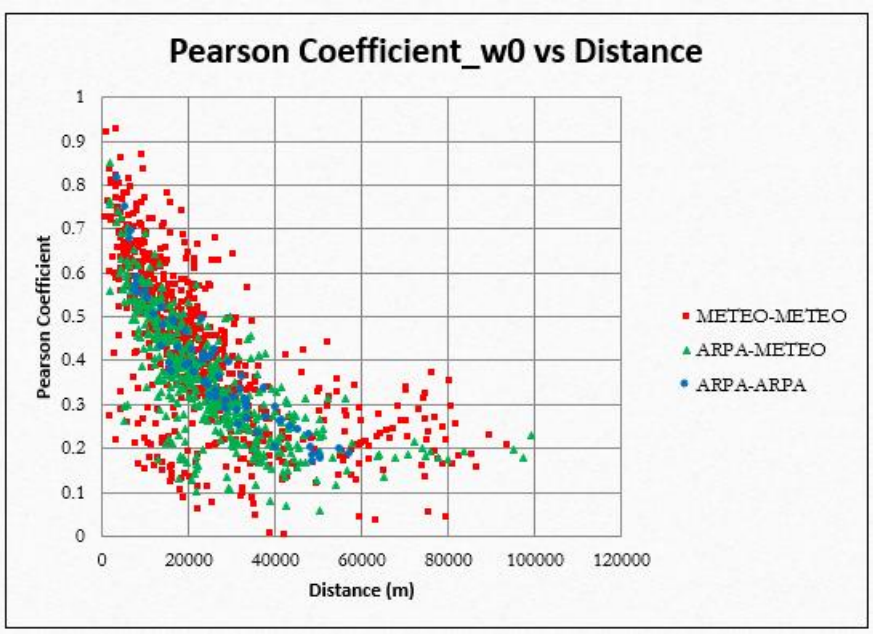


Figure 30. Pearson Correlation Coefficient vs Distance between station pairs (Hourly-Yearly-Without Zero)

Compared to the unfiltered version (see Figure 18), this filtered plot shows a slight shift in correlation values across all station pair types. ARPA–ARPA pairs (blue) remain tightly clustered, with many values above 0.5 even at intermediate distances, reflecting strong linear agreement during shared rainfall events. ARPA–METEO pairs (green) also display improved alignment, though with greater variability, particularly beyond 30 km. METEO–METEO pairs (red) continue to exhibit the widest spread and lowest correlations, but short-distance points show a modest increase in density near the 0.4–0.5 range. The overall pattern retains a clear negative gradient with distance, but the exclusion of dry periods appears to reduce noise and enhance the clarity of the spatial relationship. As in previous metrics, the filtering effect is most noticeable for short-range station pairs, where concurrent rainfall events are more frequent.

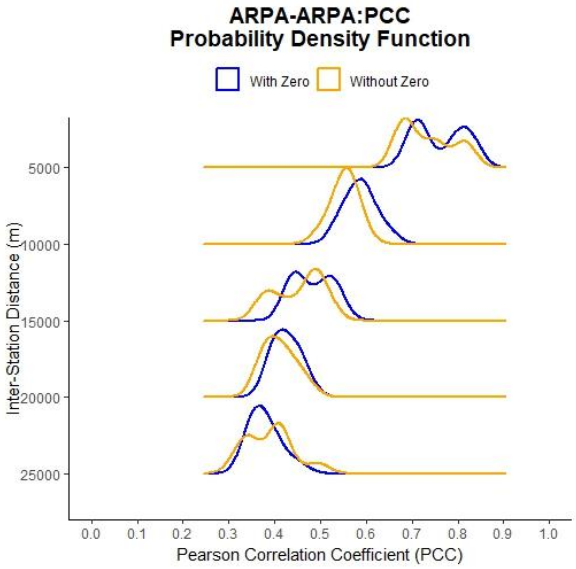


Figure 31. PDFs of PCC values - ARPA–ARPA (Hourly-Yearly-Without Zero)

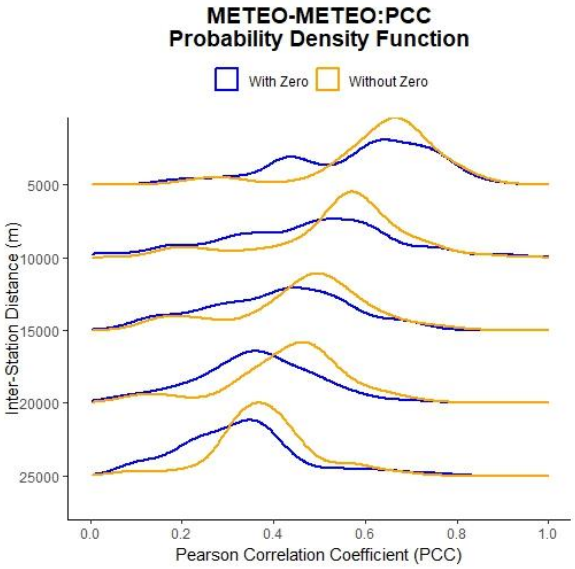


Figure 32. PDFs of PCC values - METEO–METEO (Hourly-Yearly-Without Zero)

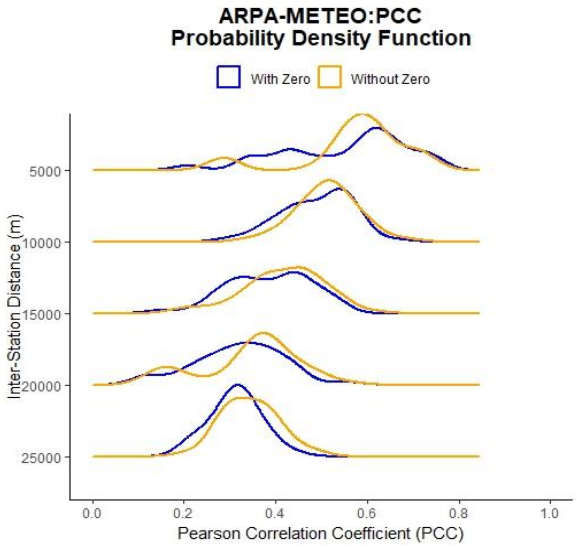


Figure 33. PDFs of PCC - ARPA–METEO (Hourly-Yearly-Without Zero)

Figure 31 - Figure 33 show how the Pearson correlation coefficient (PCC) distributions evolve with inter-station distance under both full and filtered conditions. Compared to NSE-based metrics, the effect of zero filtering on PCC is more subtle but still meaningful. For ARPA–ARPA pairs, excluding zero-rainfall periods leads to a slight leftward shift. This indicates a drop in linear agreement during rainfall events, likely due to the limited station density, which reduces the chance of simultaneous rainfall patterns over longer ranges. In contrast, METEO–METEO pairs show a modest rightward shift, suggesting that the removal of dry periods helps highlight a few coincident rainfall events, slightly increasing apparent correlation, though this may be more reflective of dataset sparsity than true coherence. ARPA–METEO pairs remain relatively stable, with only minor changes, reflecting a moderate but consistent

correlation across filtering conditions. Overall, PCC reinforces the earlier findings i.e., zero filtering improves the interpretability of inter-station relationships, and despite minor shifts, the spatial structure of rainfall coherence remains consistent across all pair types.

3.3.4. Impact of Temporal Aggregation on Station Similarity

Temporal resolution plays a critical role in shaping how similarity metrics behave across rainfall stations. Aggregating data from an hourly to a daily timescale reduces short-term variability and can enhance the stability of similarity measures, but may also obscure fine-scale spatial differences. This subsection compares station similarity across the same set of metrics, computed separately for hourly and daily rainfall totals using the full time series (including zero-rainfall periods).

By examining how metric distributions and trends shift under coarser temporal resolution, we assess whether daily aggregation strengthens spatial signals or suppresses meaningful variability.

Effect on Similarity Metrics

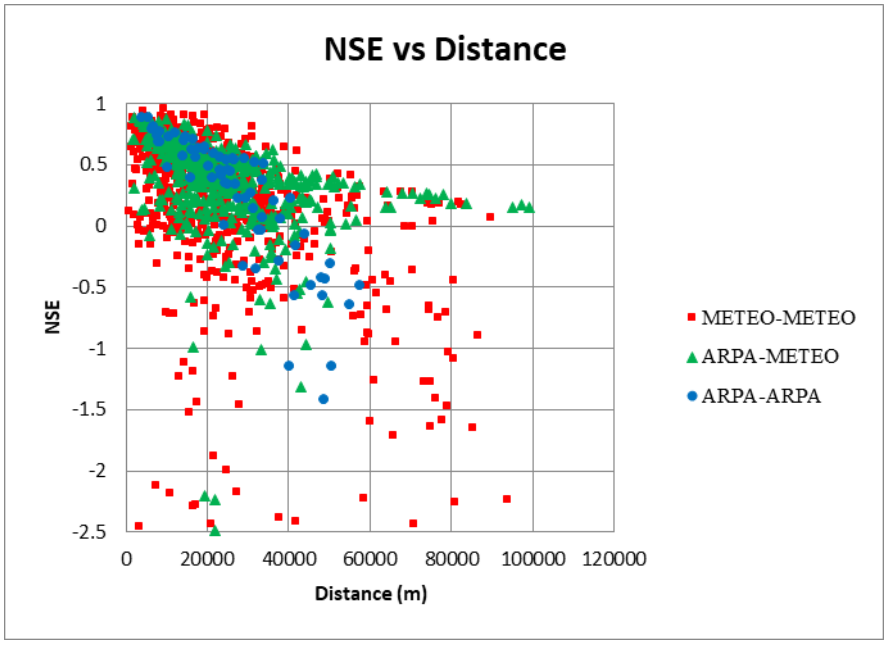


Figure 34. Nash-Sutcliffe Efficiency vs Distance between station pairs (Daily-Yearly-With Zero)

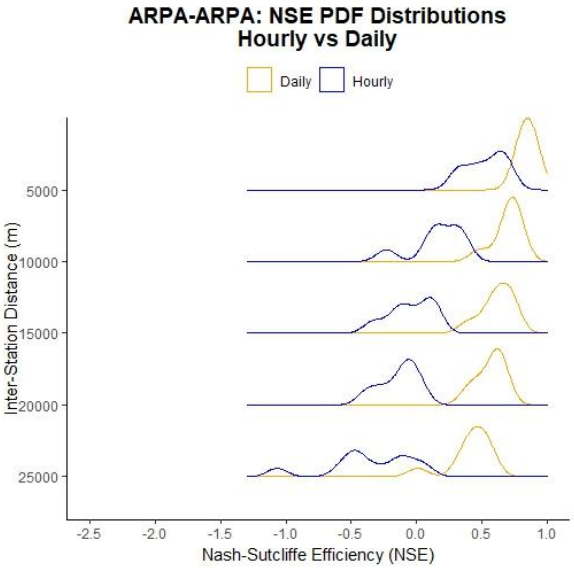


Figure 36. PDFs of NSE - ARPA-ARPA (Daily-Yearly-With Zero)

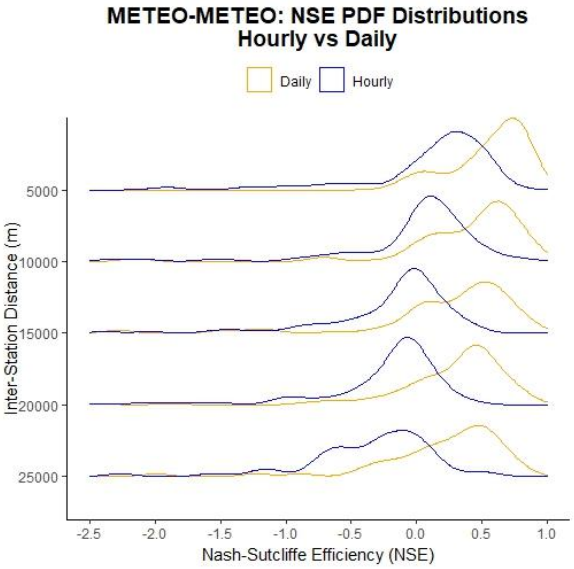


Figure 35. PDFs of NSE – METEO-METEO (Daily-Yearly-With Zero)

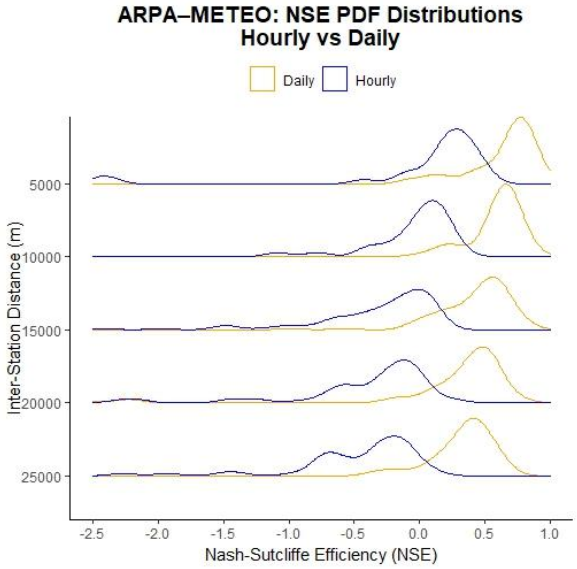


Figure 37. PDFs of NSE - ARPA–METEO (Daily-Yearly-With Zero)

The effect of daily aggregation is very prominent in NSE, as seen in the upward shift of scatter plots (see Figure 34) compared to the hourly case (see Figure 16) and the shifting of PDFs (see Figure 35 – Figure 37). For ARPA–ARPA pairs, similarity values increase and cluster tightly in the positive range, particularly within short distances, indicating strong spatial agreement in daily rainfall totals. ARPA–METEO pairs also show improved coherence, though with more spread, while METEO–METEO pairs display reduced scatter and fewer extreme negatives. The PDFs become more peaked and shift rightward, especially for ARPA-based pairs, confirming that daily aggregation enhances the stability of NSE values across the network.

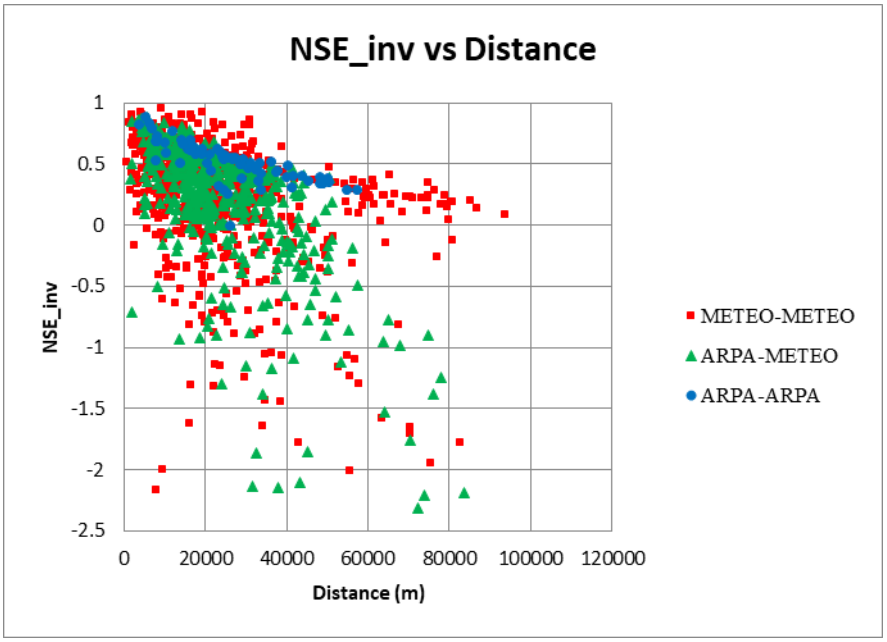


Figure 38. Inverse Nash-Sutcliffe Efficiency vs Distance between station pairs (Daily-Yearly-With Zero)

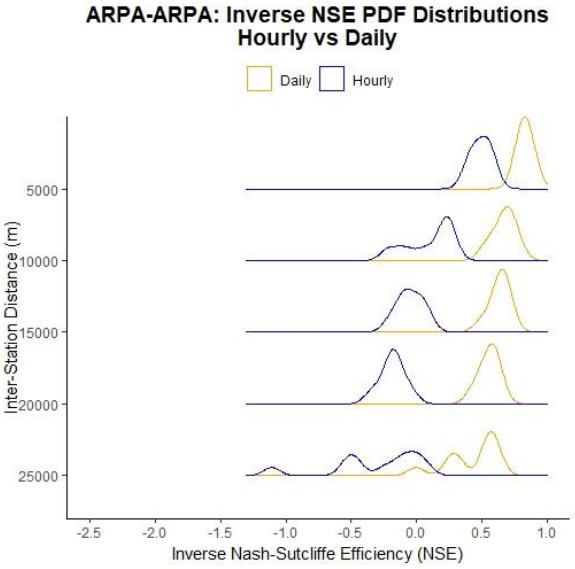


Figure 39. PDFs of Inverse NSE - ARPA–ARPA (Daily-Yearly-With Zero)

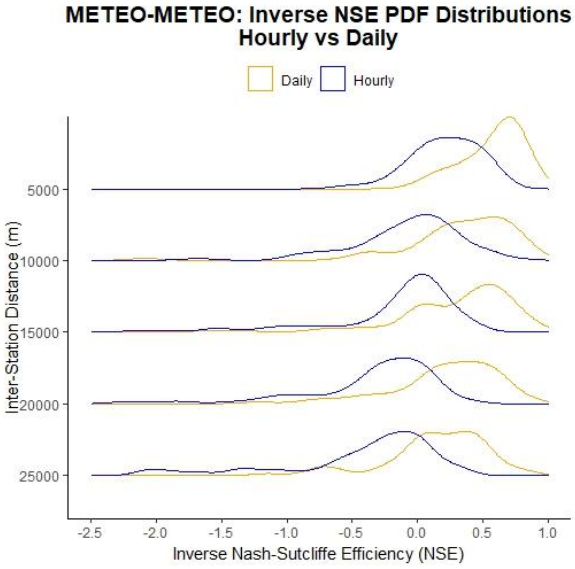


Figure 40. PDFs of Inverse NSE - METEO–METEO (Daily-Yearly-With Zero)

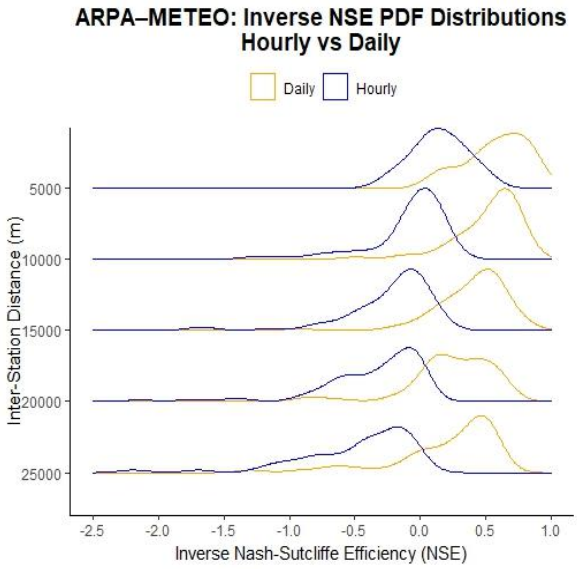


Figure 41. PDFs of Inverse NSE - ARPA–METEO (Daily-Yearly-With Zero)

Similar improvements are seen in Inverse NSE (see Figure 38), with ARPA–ARPA pairs again showing a clear upward shift and tighter clustering, particularly within 20 km. The PDFs (see Figure 39 – Figure 41) mirror those of standard NSE, becoming sharper and more symmetric with daily data. ARPA–METEO pairs exhibit moderate improvement, while METEO–METEO pairs still lag behind but show slightly more concentrated distributions. These changes reaffirm that aggregation reduces the effect of outliers and asymmetry, making inverse NSE more interpretable and consistent.

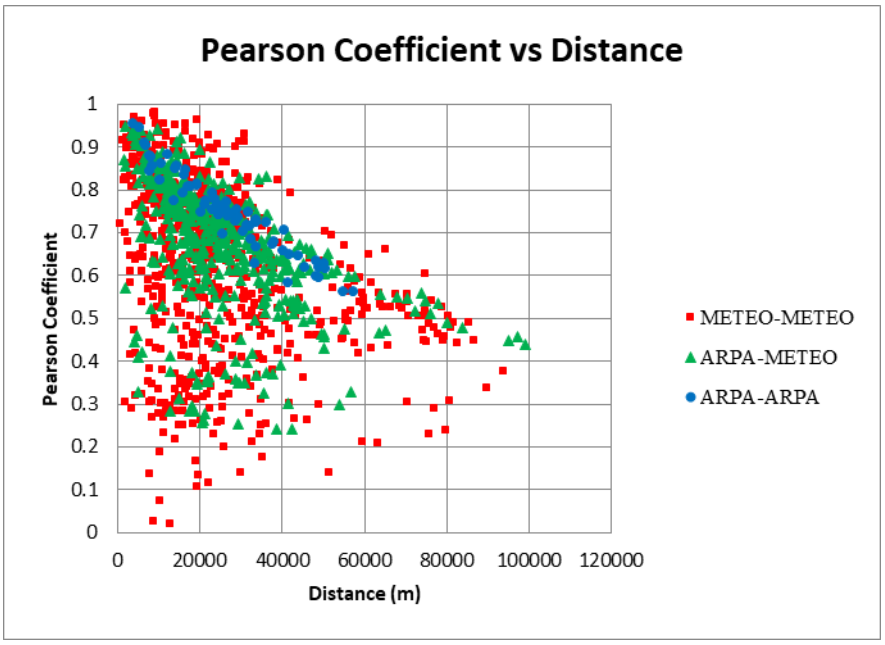


Figure 42. Pearson Correlation Coefficient vs Distance between station pairs (Daily-Yearly-With Zero)

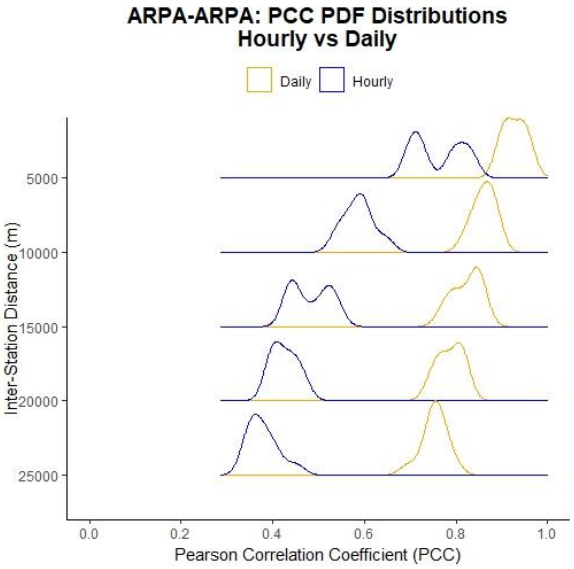


Figure 43. PDFs of PCC - ARPA–ARPA (Daily-Yearly-With Zero)

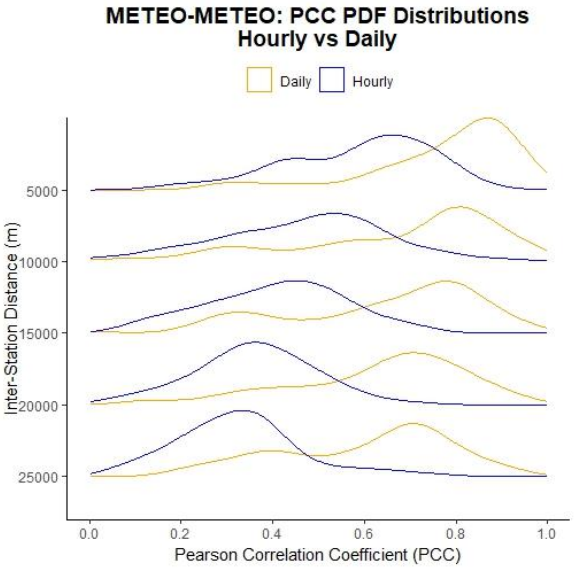


Figure 44. PDFs of PCC - METEO–METEO (Daily-Yearly-With Zero)

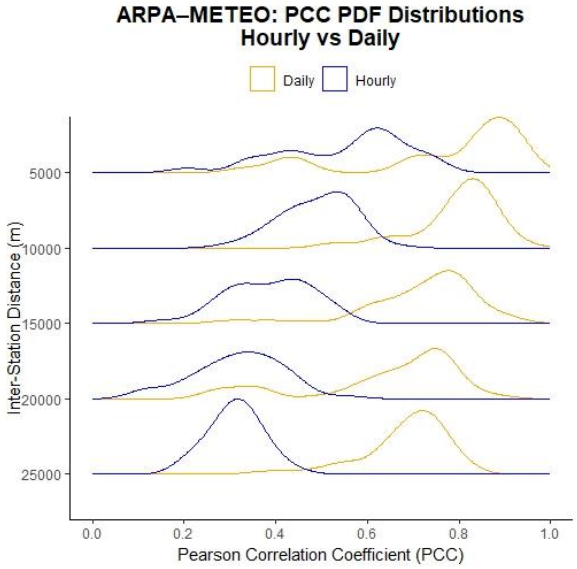


Figure 45. PDFs of PCC - ARPA–METEO (Daily-Yearly-With Zero)

For PCC, daily aggregation leads to a general compression and rightward shift in the scatter plot (see Figure 42). ARPA–ARPA pairs maintain strong correlations even at intermediate distances, while ARPA–METEO pairs show improved linear agreement, especially under 30 km. METEO–METEO pairs benefit modestly, with reduced variability but still lower overall values. The PDFs (see Figure 43 – Figure 45) become narrower and shift toward higher correlations, particularly in ARPA-based pairings, suggesting that daily totals help reveal underlying linear relationships more clearly.

Overall, daily aggregation leads to notable improvements across all similarity metrics, especially for high-quality, closely spaced station pairs. It reduces short-term noise,

enhances the interpretability of inter-station agreement, and reveals broad spatial coherence more clearly. However, this benefit comes at the cost of losing temporal resolution, which can obscure short-lived or high-intensity rainfall events, particularly critical in the analysis of extreme precipitation. Daily totals tend to mask peak rainfall intensities, timing of events, convective signatures, and intra-day variability of rainfall, which are essential to understanding the nature and impact of extreme weather. As a result, while daily aggregation is useful for identifying general spatial trends, it should be applied with caution in studies aimed at capturing event-scale dynamics, where hourly data is essential to preserve the timing, magnitude, and structure of extremes.

3.3.5. Seasonal Effects on Station Similarity

Rainfall characteristics vary substantially across seasons, with important implications for spatial similarity between stations. This subsection examines how station similarity changes when the analysis is restricted to the summer months, typically dominated by convective rainfall events. These events tend to be more intense, short-lived, and spatially localized than those in cooler seasons, potentially reducing correlation even among nearby stations. Results are compared to the full-year analysis to assess how seasonal rainfall patterns influence the behavior of similarity metrics.

Effect on Similarity Metrics

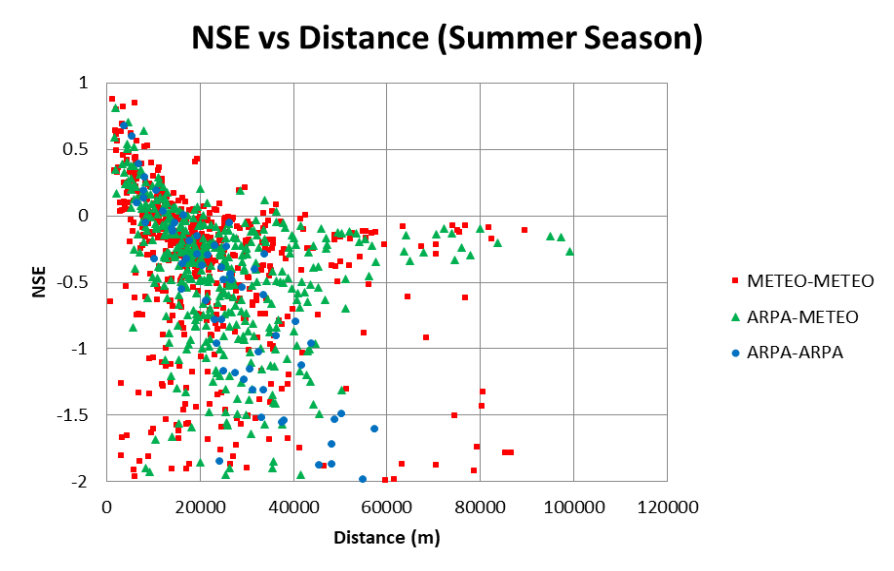


Figure 46. Nash-Sutcliffe Efficiency vs Distance between station pairs (Hourly-Summer-With Zero)

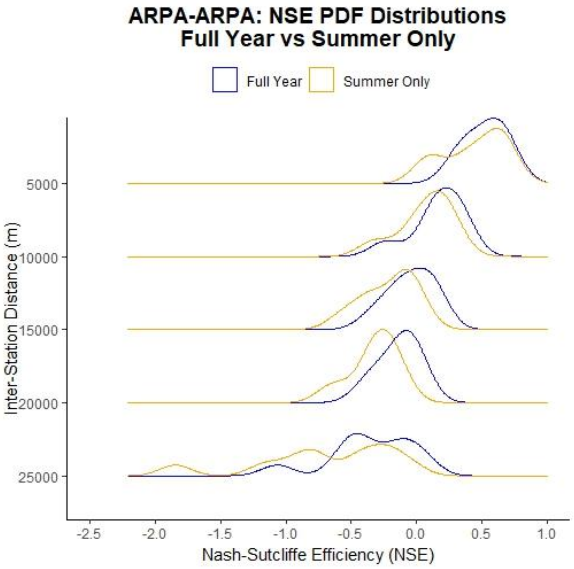


Figure 48. PDFs of NSE - ARPA–ARPA (Hourly-Summer-With Zero)

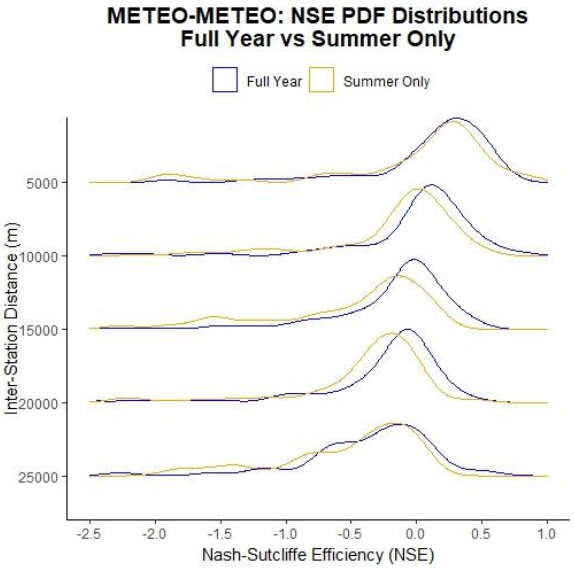


Figure 47. PDFs of NSE - METEO – METEO (Hourly-Summer-With Zero)

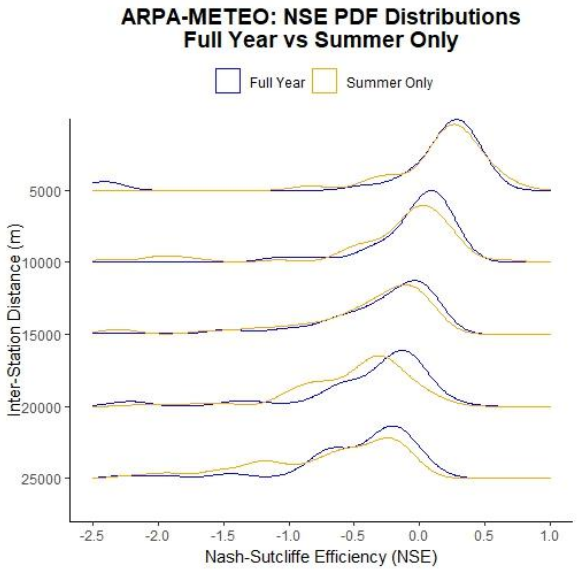


Figure 49. PDFs of NSE - ARPA–METEO (Hourly-Summer-With Zero)

The impact of seasonal variation is evident in the NSE values during the summer months, as shown in Figure 46. Compared to the full-year analysis (see Figure 16), NSE values are generally lower and more scattered. ARPA–ARPA pairs, which typically show strong agreement, now exhibit a downward shift in scatterplot and a leftward shift in PDFs (see Figure 48). This can be attributed to the limited station density in the ARPA network, even with complete data, there is less common rainfall captured between stations, especially during convective summer events that are localized and short-lived. The METEO–METEO network (see Figure 47) exhibits consistently weak

similarity, with NSE values concentrated below zero, regardless of season. The summer-only curves show a slight leftward shift, likely due to the localized and irregular nature of convective rainfall in that season, combined with limited data availability. Despite a high station count, the low data completeness results in summer PDFs that closely resemble full-year patterns, offering limited new insight. The slight shift, however, still reinforces that coherence between METEO stations weakens further in summer, particularly over larger distances. ARPA–METEO pairs benefit from the denser combined network, with a mix of high- and lower-quality data filling spatial gaps. As a result, there is reasonable overlap in PDFs (see Figure 49) up to around 15 km. However, beyond this range, a leftward shift becomes more pronounced, highlighting how the localized and non-uniform nature of summer rainfall weakens inter-station similarity as distance increases.

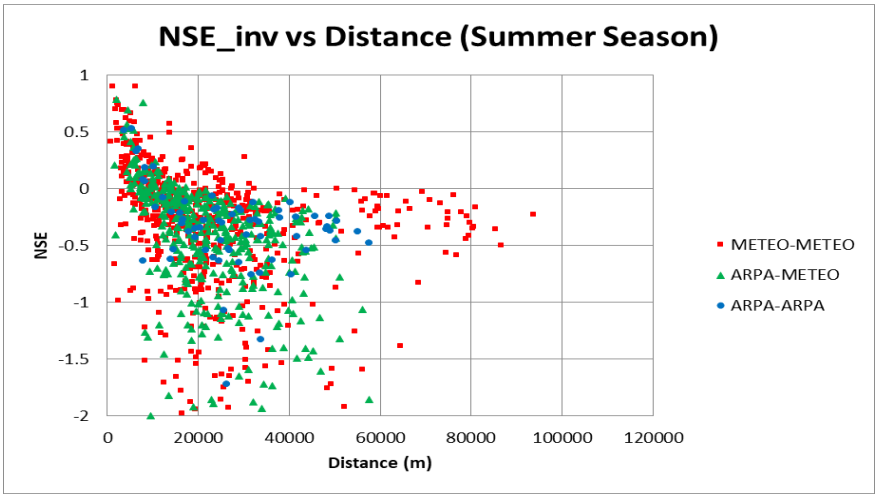


Figure 50. Inverse Nash-Sutcliffe Efficiency vs Distance between station pairs (Hourly-Summer-With Zero)

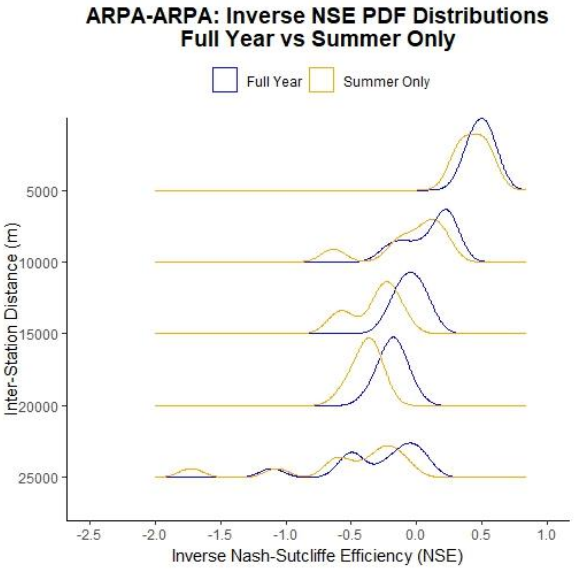


Figure 51. PDFs of Inverse NSE - ARPA–ARPA (Hourly-Summer-With Zero)

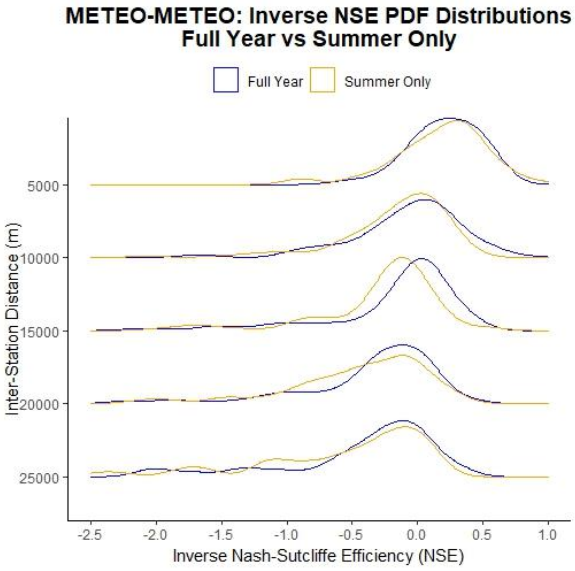


Figure 52. PDFs of Inverse NSE - METEO – METEO (Hourly-Summer-With Zero)

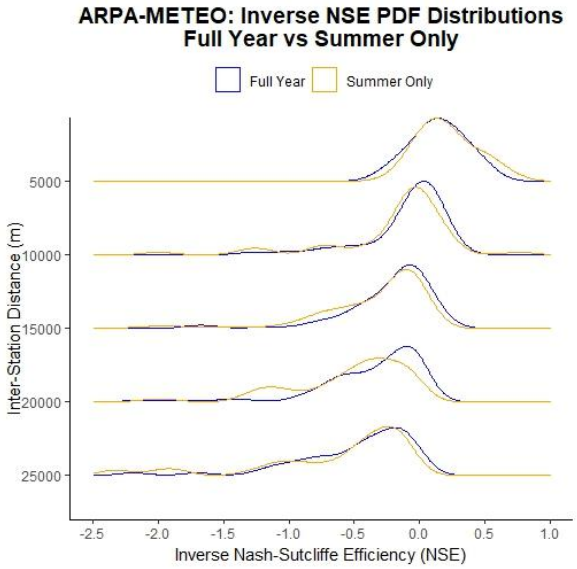


Figure 53. PDFs of Inverse NSE - ARPA–METEO (Hourly-Summer-With Zero)

The Inverse NSE scatter plot (see Figure 50) and PDFs (see Figure 51 - Figure 53) depict a behavior broadly similar to that of standard NSE, with comparable seasonal patterns and variations across station pair types. This consistency across both metrics reinforces the observed trends without introducing significant differences in interpretation.

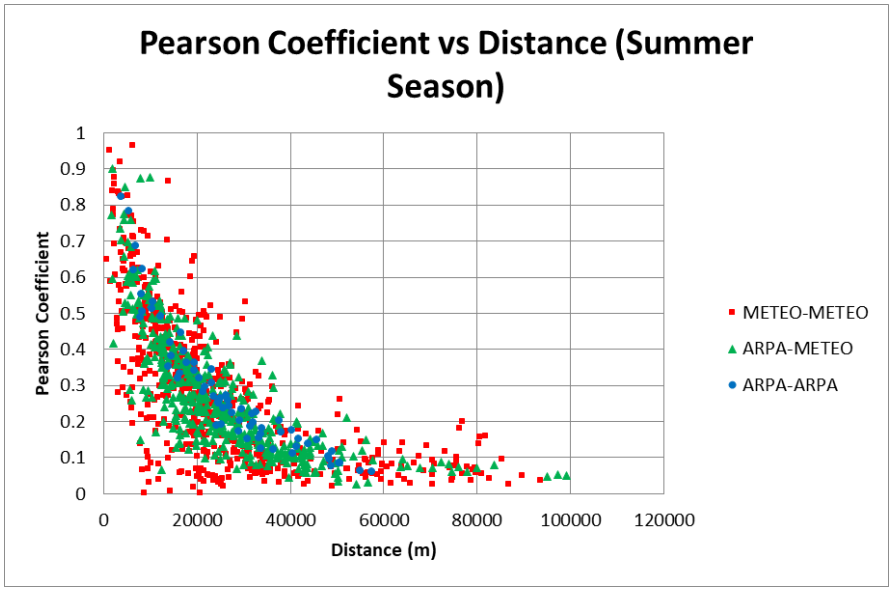


Figure 54. Pearson Correlation Coefficient vs Distance between station pairs (Hourly-Summer-Without Zero)

Figure 54 illustrates the relationship between PCC and inter-station distance during the summer season. Compared to full-year patterns (see Figure 18), PCC values in summer are generally lower and more dispersed, especially beyond 10 km. ARPA–ARPA pairs still show strong correlations at close range, though with more variability, while ARPA–METEO and METEO–METEO pairs exhibit weaker and more scattered

values across all distances. This reflects the typical impact of convective rainfall, which is more localized and less spatially coherent.

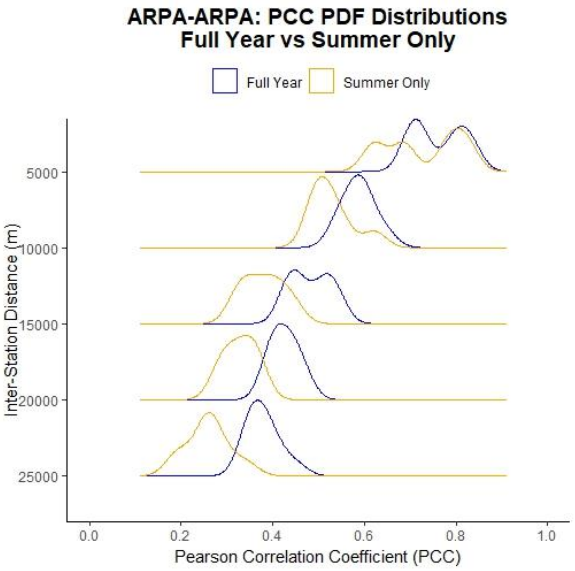


Figure 55. PDFs of PCC - ARPA–ARPA (Hourly-Summer-With Zero)

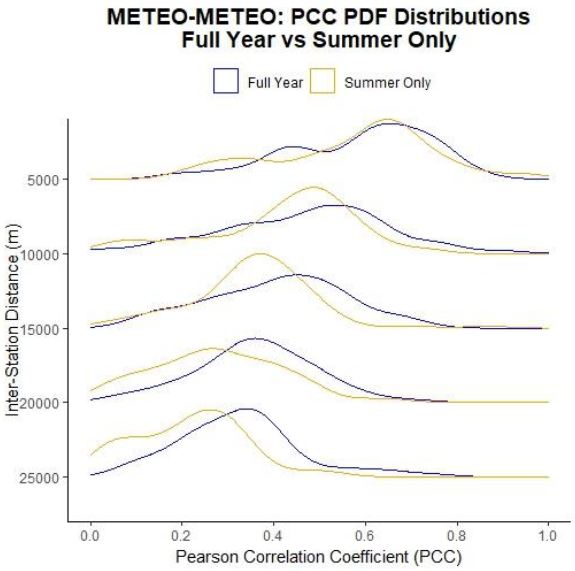


Figure 56. PDFs of PCC - METEO–METEO (Hourly-Summer-With Zero)

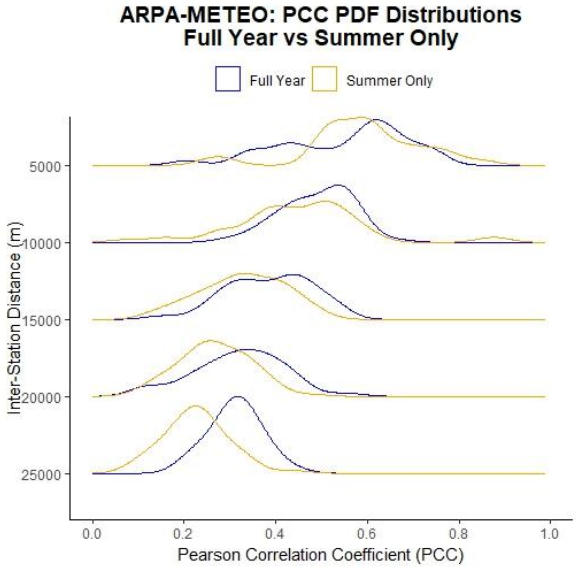


Figure 57. PDFs of PCC - ARPA–METEO (Hourly-Summer-With Zero)

Figure 55 - Figure 57 complement this trend by showing the distribution of PCC values for full-year and summer-only periods. Across all station types, the summer distributions shift left and broaden, confirming the decline in linear agreement. The change is modest but consistent for ARPA–ARPA pairs, more noticeable for ARPA–METEO pairs, and most severe for METEO–METEO, where summer distributions flatten considerably. Together, the scatter and PDFs confirm that spatial coherence in rainfall drops during summer, and PCC, while more stable than NSE, still reflects this seasonal sensitivity.

These seasonal comparisons highlight how convective rainfall in summer significantly disrupts the spatial coherence observed in full-year analyses. Even high-quality station pairs (e.g., ARPA–ARPA) show a decline in similarity with distance during summer, reflecting the localized and sporadic nature of convective events. The broader and left-shifted PDFs complement the scatter plots by illustrating this drop in metric values more clearly, particularly at shorter ranges where similarity would otherwise remain high. These effects are not merely due to data quality but arise from the inherent variability of summer rainfall, where nearby stations may experience markedly different conditions. Moreover, the number of station pairs available within each distance range influences the stability of the distributions, greater representation leads to more reliable patterns, while lower representation, especially at longer distances, may introduce noise or obscure real trends. Finally, isolated instances of high similarity among some METEO–METEO pairs likely result from limited but coinciding wet events, inflating metric values despite low overall agreement. Such cases are to be interpreted cautiously. Overall, this analysis confirms that rainfall spatial dependence is strongly seasonally modulated and can lead to overestimation of spatial consistency.

3.4. Conclusions from the Analysis

This section summarizes the main findings from the spatial correlation analysis performed using three similarity metrics: Nash–Sutcliffe Efficiency (NSE), inverse NSE, and Pearson Correlation Coefficient (PCC). The analysis considered various combinations of station types (ARPA–ARPA, METEO–METEO, ARPA–METEO), distance classes, rainfall filters, time aggregations, and seasonal windows. The key conclusions are as follows:

1. **Convergent Behavior across Metrics**
   Despite their different statistical foundations, the three metrics (NSE, inverse NSE, and PCC) produced analogous trends. All showed a consistent decrease in similarity with increasing inter-station distance, demonstrating that they reliably capture the spatial coherence of rainfall. This convergence reinforces the robustness of the analytical framework (see Figure 16 - Figure 18).

2. **Effect of Filtering for Rainfall Events**
   Limiting the analysis to wet periods, excluding zero-rainfall time intervals, produced generally limited change of the values of the correlation indices (see the distributions in Figure 23 - Figure 25, Figure 27 - Figure 29, Figure 31 - Figure 33); some different behavior might be seen for the PCC for METEO-METEO couples (Figure 31). However, this filtering does not capture the full spatial variability of rainfall, particularly in convective regimes where precipitation can be highly localized and may occur at one station but not at another nearby one.

Considering only cases where both stations experience rain, this approach can artificially inflate similarity metrics, potentially overestimating agreement between stations. As such, while useful for event-focused analysis, wet-only filtering should be applied with care in broader spatial assessments.

3. **Effect of Temporal Aggregation**
   Aggregating rainfall data from hourly to daily resolution led to much higher correlation values and more stable similarity patterns (see Figure 34 - Figure 37, Figure 39 - Figure 41, Figure 43 - Figure 45). However, this is not fully aligned with the goals of this study, which focuses on extreme rainfall events. Such events typically occur on sub-daily timescales, and daily aggregation can obscure their intensity, timing, and spatial structure. As such, short-duration data remains critical for capturing the dynamics of extremes.

4. **Seasonal Effect**
   Figure 47 - Figure 49, Figure 51 - Figure 53 and Figure 55 - Figure 57 return a visible seasonal effect, though weaker than that for temporal aggregation. The PCC appears more sensitive than the NSE indexes, as shown by Figure 55 - Figure 57.

5. **On Reliability of METEO Data**
   The demonstration of how correlation indexes change with different sources is provided by Figure 19 - Figure 21. The source seems to have an effect on the values one obtains; the change in distributions is definitely less than that due to temporal aggregation, more than that due to considering all the values or just those for the wet periods, similar to that for the seasonal consideration. METEO data captures meaningful rainfall patterns and, when used within short-range contexts or in combination with ARPA data, can significantly contribute to robust spatial rainfall analyses. The denser spatial distribution of METEO stations further enhances its value in expanding network coverage and resolving fine-scale variability.

6. **Distance-Similarity Relationship and Threshold for Imputation**
   Correlation metrics decline with distance across all station types. Figure 16 - Figure 21 show that at a distance of around 10 km PCC typically is ~0.5 and NSE approaches zero for all station pair types (apart from some METEO-METEO outliers). This holds even under more challenging conditions, such as convective summer rainfall or wet-only filtering, where spatial coherence typically declines more rapidly. A threshold distance of 11 km will guide gap-filling strategies in Chapter 4 by helping to select nearby stations with sufficiently high similarity to justify data imputation.

4 Chapter Four: Imputation of Missing Rainfall Data

4.1. Motivation for Data Imputation

Rainfall data often contain gaps due to operational disruptions such as sensor malfunctions, power failures, maintenance, or communication errors. This problem is especially pronounced in citizen science datasets, such as those from the METEO network, where observational continuity is less controlled compared to institutional networks (ARPA).

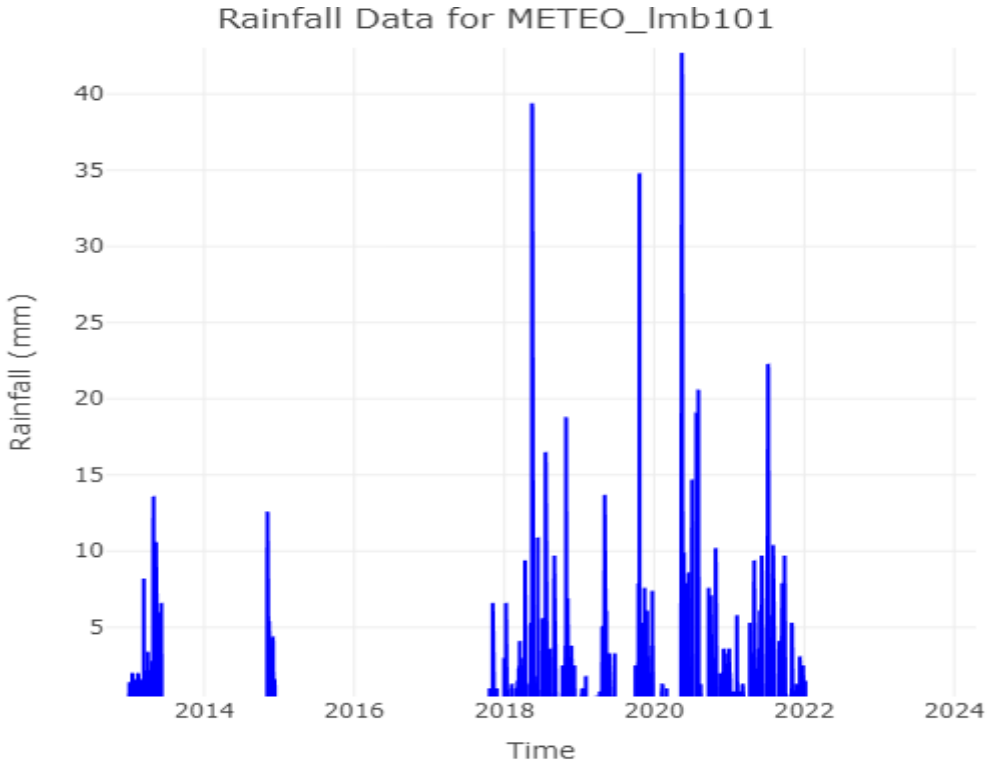


Figure 58. Rainfall time series (2013–2023) from METEO station lmb101, showing frequent data gaps

Figure 58 displays long periods of missing or zero values, interspersed with a few isolated rainfall events in a particular station. This temporal sparsity is common in METEO stations and highlights the difficulty of applying traditional interpolation methods. Conventional techniques such as linear interpolation, spline interpolation, or inverse distance weighting (IDW) assume gradual or spatially smooth variation, which makes them unsuitable for handling abrupt, event-based rainfall patterns. Imputing such data requires a strategy that can handle irregular gaps while preserving the underlying rainfall patterns. The issue is further complicated by the sparse and

intermittent nature of rainfall itself. In many METEO stations, rainfall is infrequent and often appears as isolated events separated by long dry periods. This pattern, combined with missing entries, makes traditional interpolation unreliable and likely to produce misleading values. A more principled and distribution-aware approach to imputation is required, one that preserves both the structure and variability of the original data.

This missingness poses a challenge for statistical modelling. In particular, the Heffernan and Tawn (HT) model, which we aim to apply for the analysis of spatial extremes in rainfall, requires complete input data and cannot accommodate missing values. Therefore, addressing data gaps is a necessary precondition for the valid application of the HT framework.

By reconstructing missing rainfall values, we aim to:

- Improve the temporal coherence of each station's time series,
- Enable the valid application of the HT model for extreme rainfall analysis and flood hazard assessment in the region.

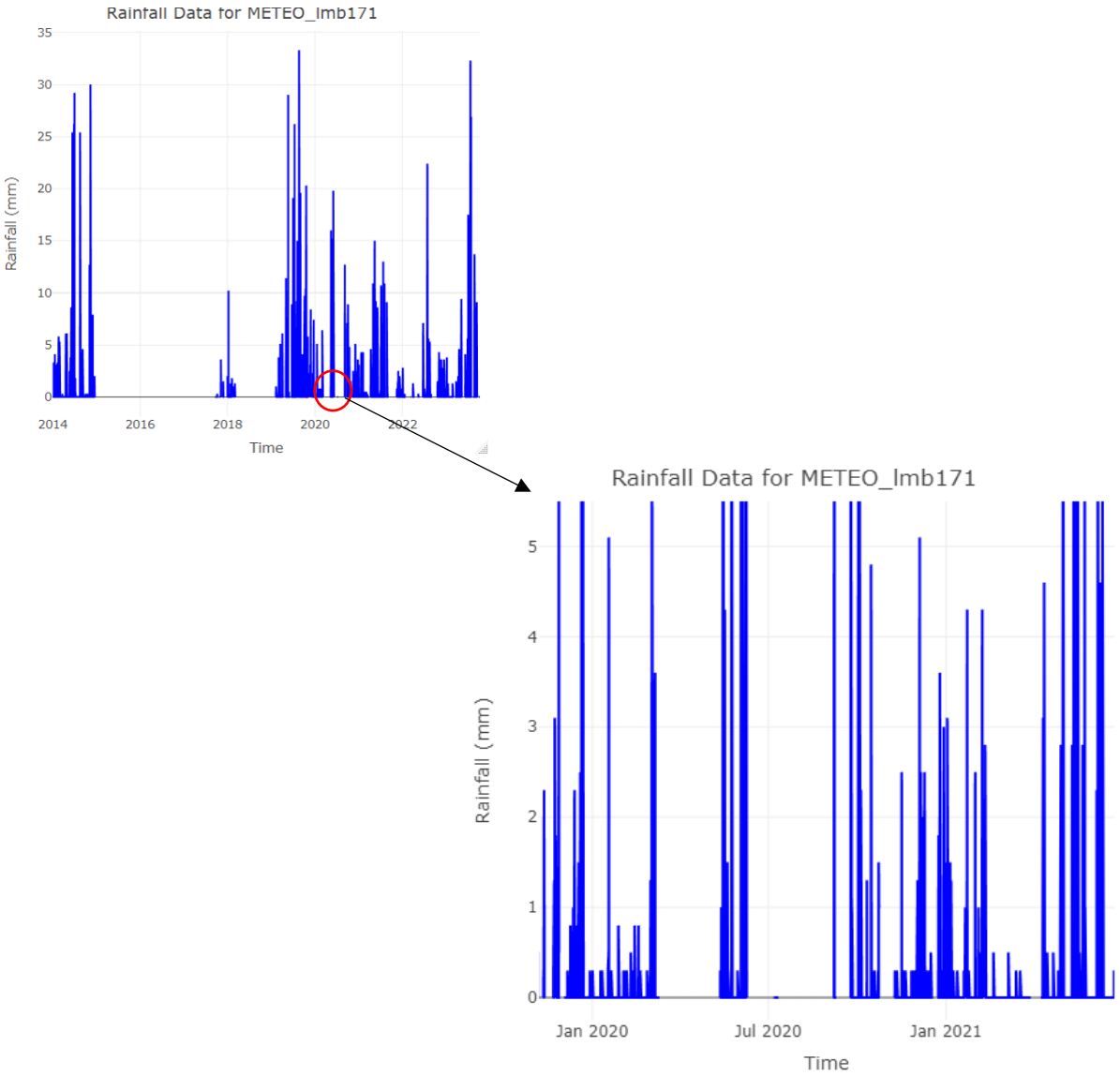


Figure 59. Data gaps in the rainfall time series from METEO station lmb171 during the Summer of 2020

The absence of data not only reduces the pool of usable extreme events but also introduces biases in correlation structures and conditional modelling outcomes. This becomes especially problematic when modelling rare but impactful phenomena, where each data point carries substantial weight in the estimation of tail dependencies.

Therefore, to ensure meaningful application of spatial extremes modelling, and to retain the integrity of rainfall pattern analysis, data imputation becomes an essential step. Through careful reconstruction of missing values, we aim to improve the coherence of the dataset, preserve underlying spatial relationships, and support robust modelling of flood hazard scenarios in the region.

At the same time, it is critical that the imputation process does not distort the underlying data structure. To this end, we implement diagnostic checks to ensure that

gap-filling does not significantly alter the marginal statistics (e.g., distributions, variances) at individual stations.

This chapter outlines the imputation strategy adopted to meet these goals and ensure compatibility with the HT model in the subsequent analysis.

4.2. Multiple Imputations Using MICE

4.2.1. Introduction to Multiple Imputations and MICE

One of the most robust and widely accepted approaches for handling missing data is "Multiple Imputation" (MI), introduced by Rubin [48] [49].

Multiple imputation aims to replace each missing value with a set of plausible values that represent the uncertainty about the right value to impute. Instead of filling in a single value, MI creates several completed datasets, analyzes each of them separately, and combines the results using a framework that preserves variability due to imputation.

Two dominant strategies for performing multiple imputation on multivariate datasets are:

1. **Joint Modelling (JM)**
   This approach defines a multivariate probability distribution for the full dataset and uses methods such as Markov Chain Monte Carlo (MCMC) to draw imputations from the joint posterior predictive distribution – which predicts missing values by considering the relationships between all variables in the dataset [50] . JM can be powerful when the data conform well to a known joint distribution, but specifying such a model becomes difficult when the dataset includes a mix of variable types (e.g., continuous, binary, ordinal) or complex dependencies.
2. **Fully Conditional Specification (FCS) / Multivariate Imputation by Chained Equations (MICE)**
   Introduced and popularized by van Buuren and Groothuis-Oudshoorn [51] [52], MICE offers a more flexible alternative to JM. Rather than modelling the joint distribution, MICE uses a series of univariate regression models to impute each variable with missing data, conditional on all the others. This method cycles through each incomplete variable iteratively, updating imputations based on the most recent values.

The strength of MICE lies in its adaptability. Each variable can be imputed using a model suitable to its data type: for instance, logistic regression for binary variables, predictive mean matching for continuous variables, or ordinal logistic regression for ordered categorical variables. The imputation procedure is iterative, cycling through

each variable and updating its missing values based on the most recent imputations of the others. These models are built conditionally (each one uses the other variables as predictors), but they do not need to be mathematically consistent with each other, that is, they do not need to come from a single joint distribution. In formal statistical terms, the set of conditional models may not correspond to a coherent joint model.

This flexibility has led to the widespread use of MICE in applied research. It is particularly valuable in datasets that exhibit a mixture of variable types or when a well-specified joint model is infeasible. The approach is also known under other names such as regression switching [53], sequential regressions [54], or stochastic relaxation [55], but Multivariate Imputation by Chained Equations "(MICE)" has become the standard terminology in modern statistical practice.

In the context of this study, MICE was employed to impute the missing rainfall values across spatially distributed stations. The specific models and imputation techniques used are tailored to the nature of the data and will be described in detail in the subsequent sections.

4.2.2. General Framework of Multiple Imputation in Mice

Multiple imputation in this study was carried out using the "mice package" in R, developed by van Buuren and Groothuis-Oudshoorn [52]. The package provides a flexible framework for imputing missing values based on conditional models appropriate to each variable's distribution. It is widely used in both clinical and environmental applications due to its extensibility and compatibility with real-world data.

This section describes the structure of multiple imputation as implemented in MICE, introducing the necessary notation, outlining the iterative imputation process, and explaining how different variable types are handled through tailored models.

Notation and Setup

Let $Y = (Y_1, Y_2, \ldots, Y_p)$ be a set of p variables, each of which may contain missing values. For a given variable Y_j, its observed and missing components are denoted Y_j^{obs} and Y_j^{mis}, respectively. Thus, the full dataset can be partitioned into $Y^{obs} = (Y_1^{obs}, \ldots, Y_p^{obs})$ and $Y^{mis} = (Y_1^{mis}, \ldots, Y_p^{mis})$.

The goal of imputation is to generate m≥1 complete datasets, denoted as $Y^{(1)}, Y^{(2)}, \ldots Y^{(m)}$, where each dataset is a filled-in version of the original incomplete data. These datasets are identical in their observed components but differ in their imputed values to reflect the inherent uncertainty in predicting the missing data. The imputation for each variable Y_j is performed conditionally on the other variables, represented as $Y_{-j} = Y_1, \ldots, Y_{j-1}. Y_{j+1}, \ldots, Y_p)$ [52].

Imputation Process

The process begins with the observed, incomplete dataset Y^{obs}. Since reliable statistical inference is typically not possible with incomplete data, MICE aims to generate plausible values for the missing entries based on the observed information. This is done using a series of conditional models, each tailored to the distribution and structure of the variable being imputed. These imputations are performed iteratively and result in multiple imputed datasets. In the mice package, this operation is carried out using the "mice()" function, and the resulting datasets are stored in an object of class "mids" (Multiply Imputed Dataset) [52].

The mice package classifies variables into four main types based on their structure and data type:

1. **Numeric:** continuous variables (e.g., age, income),
2. **Binary:** factors with 2 levels (e.g., yes/no, male/female),
3. **Unordered categorical:** factors with more than 2 levels, where the order does not matter (e.g., blood type: A, B, AB, O),
4. **Ordered categorical:** ordered factors with more than 2 levels, where the order does matter (e.g., education level: high school < bachelor's < master's < PhD).

Each of these types is associated with a default imputation method in MICE, which is automatically selected unless the user specifies otherwise.

Method	Description	Scale Type	Default
Pmm	Predictive Mean Matching	Numeric	Y
Norm	Bayesian Linear Regression	Numeric	
norm.nob	Linear Regression, Non-Bayesian	Numeric	
Mean	Unconditional Mean Imputation	Numeric	
2L.norm	Two-Level Linear Model	Numeric	
Logreg	Logistic Regression	Factor, 2 Levels	Y
Polyreg	Multinomial Logit Model	Factor, >2 Levels	Y
Polr	Ordered Logit Model	Ordered, >2 Levels	Y
Lda	Linear Discriminant Analysis	Factor	
Sample	Random Sample from Observed Data	Any	

Table 2. Built-in univariate imputation techniques (RStudio - mice package) [52]

Selected Imputation Methods used in this Study

Among the various imputation methods listed in the Table 2. Built-in univariate imputation techniques (RStudio - mice package) , only two were employed in the present analysis: Predictive Mean Matching (pmm) for numeric variables and Logistic Regression (logreg) for binary variables. While rainfall data are continuous, the use of a binary method is justified by a double imputation approach introduced later in this study (see Section 4.5), where a binary variable is used to distinguish between wet and dry time steps. This auxiliary variable supports more realistic imputations by first classifying events before imputing rainfall amounts.

1. **Predictive Mean Matching (PMM)**

   Predictive Mean Matching [56], implemented in the mice package using the function "mice.impute.pmm()", is a semi-parametric imputation method designed for numeric variables. Rather than drawing values from a theoretical distribution, this method restricts imputations to observed values from the dataset, ensuring realistic replacements.

   This method proceeds as follows. Let y be the variable with missing values and X the matrix of predictors. A linear regression model is first estimated on the observed data using Equation 4.

   $$\widehat{y_{\text{obs}}} = X_{\text{obs}}\hat{\beta}$$

   Equation 4. Linear regression model fitted on the observed data

   where $\widehat{y_{\text{obs}}}$ represents the predicted values for observed entries based on the predictor matrix X_{obs} and estimated coefficients $\hat{\beta}$.

   The estimated coefficients $\hat{\beta}$ are then used to predict values for both the observed and missing cases using Equation 5.

   $$\widehat{y_{\text{mis}}} = X_{\text{mis}}\hat{\beta}$$

   Equation 5. Predicted values for the missing entries

   where $\widehat{y_{\text{mis}}}$ is obtained using the predictor matrix X_{mis} and the estimated regression coefficients $\hat{\beta}$ from the observed data.

   For each missing case i, a set of k observed cases with predicted values closest to $\widehat{y_{\text{mis,i}}}$ is identified. One of these observed values $y_{obs,j}$ is randomly drawn and used as the imputed value.

   $$\widetilde{y_{\text{mis},i}} \sim \{y_{\text{obs},j} \mid j \in \mathcal{N}_k(i)\}$$

   Equation 6. Imputation step in Predictive Mean Matching

   Where the missing value $\widetilde{y_{\text{mis},i}}$ is randomly drawn from the set of observed values $y_{\text{obs},j}$ whose predicted values are closest to the prediction for case i. The index set $\mathcal{N}_k(i)$ denotes the k nearest neighbours in prediction space.

Its main advantages are:

- **Preservation of original data characteristics**: Because only observed values are used, it avoids creating implausible or out-of-range imputations.
- **Flexibility**: It maintains non-linear relationships even if the underlying regression model is imperfect, making it robust for a wide variety of situations.
- **General-purpose applicability**: It is often considered a reliable default for numeric variables in real-world datasets, especially when the variable distributions are not strictly normal.

2. **Logistic Regression (LOGREG)**

The logreg method, implemented in the mice package using the function "mice.impute.logreg()", is used for binary variables, those that have two possible categories (e.g., 0/1 or yes/no). It works by fitting a logistic regression model using the fully observed cases, with the binary variable as the outcome and the other variables as predictors. The imputed value is then sampled based on the predicted probability from the fitted model.

For example, if the model predicts a 70% probability that a missing value should be '1', the algorithm will randomly assign a '1' with 70% probability and a '0' otherwise. This approach ensures that the association between the binary variable and other covariates is maintained, which is essential for valid inference.

Predictor Matrix

One of the powerful features of the MICE algorithm is its flexibility in selecting predictor variables for the imputation of each incomplete variable. This selection is managed through the "predictorMatrix" argument. The predictorMatrix is a square matrix of 0s and 1s, with dimensions equal to the number of variables in the dataset (i.e., ncol(data) × ncol(data)). Each row of the matrix represents an incomplete variable (the target variable to be imputed), and each column represents a potential predictor. The structure of the predictor matrix can be summarized as follows:

- A value of 1 in cell (i, j) means that variable j is used as a predictor when imputing variable i.
- A value of 0 means variable j is not used to impute variable i.
- The diagonal is always set to 0, as a variable cannot predict itself.
- If a variable has no missing values, its entire row is set to 0 automatically, since there is no need to impute it.

	AGE	BMI	BP	CHOL	SEX
AGE	0	0	0	0	0
BMI	1	0	1	1	1
BP	1	1	0	1	1
CHOL	1	1	1	0	1
SEX	0	0	0	0	0

Table 3. Example of a 5×5 predictor matrix used in the MICE algorithm

To illustrate the structure of a predictor matrix in a more general context, Table 3 presents a simplified example. It demonstrates how predictor matrices encode the relationships used during imputation, with rows representing variables to be imputed and columns indicating which other variables are used as predictors (denoted by 1s). This general structure would also apply in the rainfall setting, though with different variable names and spatial logic.

This matrix can be customized before calling the mice() function to tailor which variables inform the imputations. By default, mice() assume a fully connected system, every variable predicts every other incomplete variable, except itself.

The resulting object from mice() includes the predictor matrix in its "predictorMatrix" component, which can be inspected or modified further. This enables refined control and potentially more accurate imputations, especially when prior knowledge suggests that certain predictors are irrelevant or redundant.

4.3. Imputation Process Implementation in Rstudio

In this section, the steps undertaken to impute missing rainfall data across the stations are detailed. The imputation process leveraged spatial relationships between stations to improve the accuracy of missing data prediction.

4.3.1. Creation of the Distance-Based Predictor Matrix

Given the spatial nature of rainfall data, imputations were conducted using a distance-based predictor matrix. This approach was guided by prior statistical validation in Chapter 3. The predictor matrix was constructed based on the Euclidean distances between stations, calculated using their UTM (Universal Transverse Mercator) coordinates. The previously determined threshold of 11 km was applied, meaning only stations within this distance were considered as potential predictors for each station. This approach ensures that imputations are influenced by stations with similar geographic locations, maintaining the spatial autocorrelation of rainfall patterns. Straightforwardly, the resulting binary matrix assigned a value of 1 to station pairs

within the 11 km threshold, and 0 otherwise. The predictor matrix had dimensions of 50×50, representing all ARPA and METEO stations included in the analysis.

```
# First, get unique station coordinates.
station_coords <- final_clean1 %>%
  select(station_code, utm_easting, utm_northing) %>%
  distinct()

# Convert coordinates to a matrix (using UTM, so Euclidean distance is acceptable)
coords <- as.matrix(station_coords[, c("utm_easting", "utm_northing")])
rownames(coords) <- station_coords$station_code

# Compute pairwise Euclidean distances (in meters)
dist_mat <- as.matrix(dist(coords))

# Set a threshold: use only stations within this distance as predictors.
# (For example, 11 km = 11000 meters)
threshold <- 11000
# Create a binary predictor matrix: 1 if distance is within threshold, 0 otherwise.
pred_matrix <- (dist_mat <= threshold) * 1
diag(pred_matrix) <- 0  # Exclude self-prediction
```

Figure 60. Code (in RStudio) for generating a binary distance-based predictor matrix with an 11 km threshold

	ARPA_2006	ARPA_2385	ARPA_4065	ARPA_5902	ARPA_5908	ARPA_5916	ARPA_8122	ARPA_8152	ARPA_8197	ARPA_8199	ARPA_8211	ARPA_8228	METEO_lmb009	METEO_lmb021	METEO_lmb080	METEO_lmb084	METEO
ARPA_2006	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	
ARPA_2385	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	
ARPA_4065	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
ARPA_5902	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
ARPA_5908	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0	
ARPA_5916	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
ARPA_8122	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	
ARPA_8152	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
ARPA_8197	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	1	
ARPA_8199	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	1	
ARPA_8211	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	
ARPA_8228	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	
METEO_lmb009	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	1	
METEO_lmb021	0	1	0	0	0	0	0	0	0	1	1	0	0	0	1	0	
METEO_lmb080	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	
METEO_lmb084	0	0	0	0	0	0	1	0	1	1	0	0	1	0	0	0	

Figure 61. Excerpt of the binary distance-based predictor matrix (11 km threshold) showing spatial relationships among first 15 stations

4.3.2. Imputation using Predictive Mean Matching (pmm)

To carry out the imputation process, Predictive Mean Matching (pmm) was applied in conjunction with the distance-based predictor matrix. By limiting the pool of potential donors to stations within the 11 km threshold, the imputed values were drawn from geographically relevant sources. The imputation was performed over five imputations and five iterations. Five imputations were deemed sufficient, striking a balance between capturing imputation variability and keeping computational cost manageable. A higher number of imputations would offer only marginal improvements in terms of variance while significantly increasing computational costs, especially considering the scale of the dataset (i.e.., 50 columns and 95,785 rows).

Each imputation consisted of five iterations, during which the algorithm repeatedly updated the imputed values based on the most recent estimates. These iterations allow the underlying models to stabilize, helping parameters converge toward consistent

values. After a few cycles, changes between iterations become very small and progressively less variable, indicating that the process has reached convergence and the imputed values are reliable. This pattern is particularly evident in Figure 63, where the mean and standard deviation of the imputed values across multiple chains begin to stabilize after just a few iterations. While slight fluctuations remain, the absence of systematic drift supports the conclusion that the chosen imputation settings (5 iterations and 5 imputations) were sufficient to achieve stable results. This setup reflects a deliberate balance between statistical reliability and computational efficiency, given the high dimensionality and size of the dataset.

```
# Impute missing data using MICE
imp <- mice(original_data, m = 5, method = "pmm", maxit = 5)
```

Figure 62. Imputation of missing data using Predictive Mean Matching (PMM) with 5 imputations and 5 iterations, performed with the MICE package (RStudio)

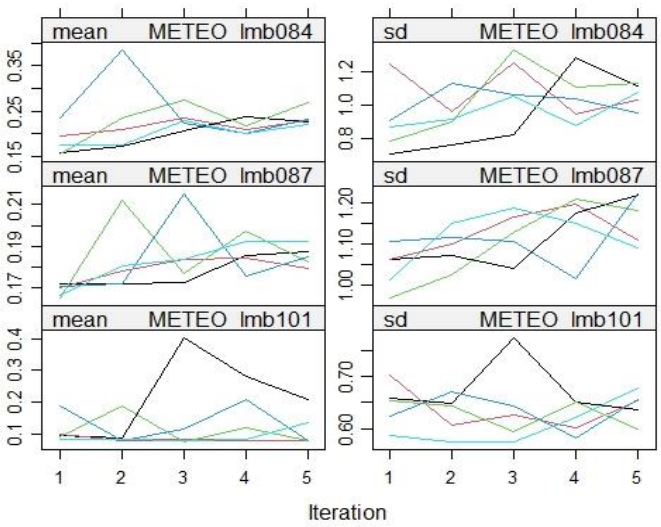


Figure 63. Mean and standard deviation of imputed values across five iterations for three METEO stations

4.3.3. Problem with Averaging the 5 Imputed Datasets

To visualize the imputed rainfall data, rainfall time series were created for each station. These plots allow for the assessment of temporal patterns in the rainfall data, with the imputed values seamlessly integrated into the overall data series. While multiple imputation theory typically discourages averaging across imputations, favoring Rubin's rules to preserve variance and account for imputation uncertainty, it was chosen to average the five imputed datasets in this study for practical and computational reasons. This decision was informed by additional exploratory analyses comparing results derived from single imputations against their average. These comparisons, which included visual inspections (see Figure 64 - Figure 68), showed negligible differences in outcomes for a station with only 30% coverage. Therefore,

averaging was adopted as a pragmatic solution to simplify the data structure while maintaining representational integrity.

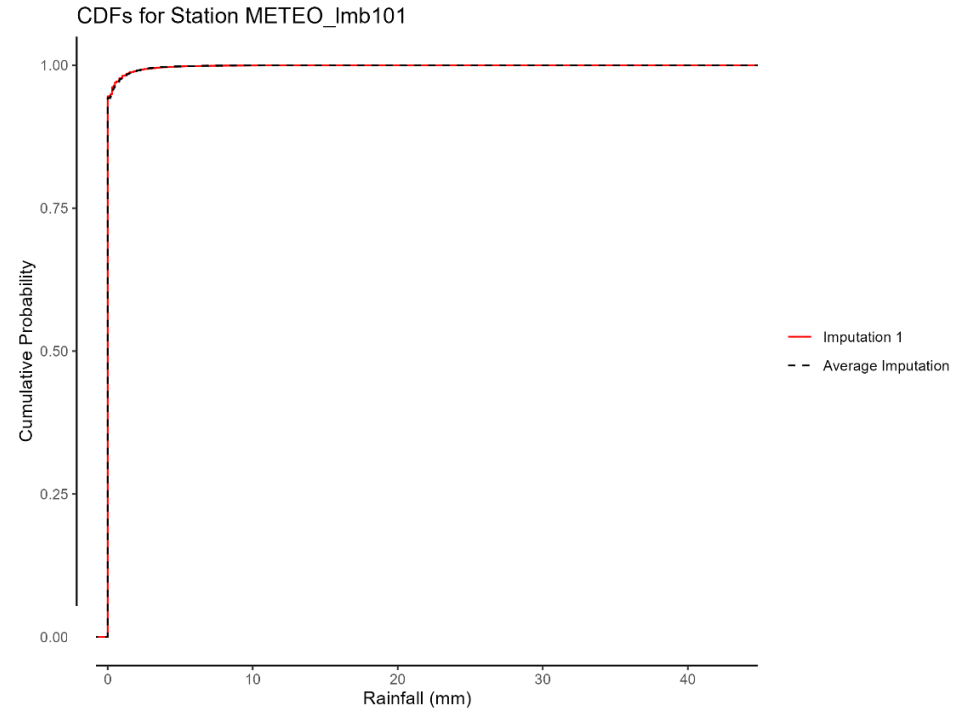


Figure 64. Empirical Cumulative Distribution Functions (CDFs) of Rainfall for Station METEO lmb101 comparing Average Imputation and First Imputation

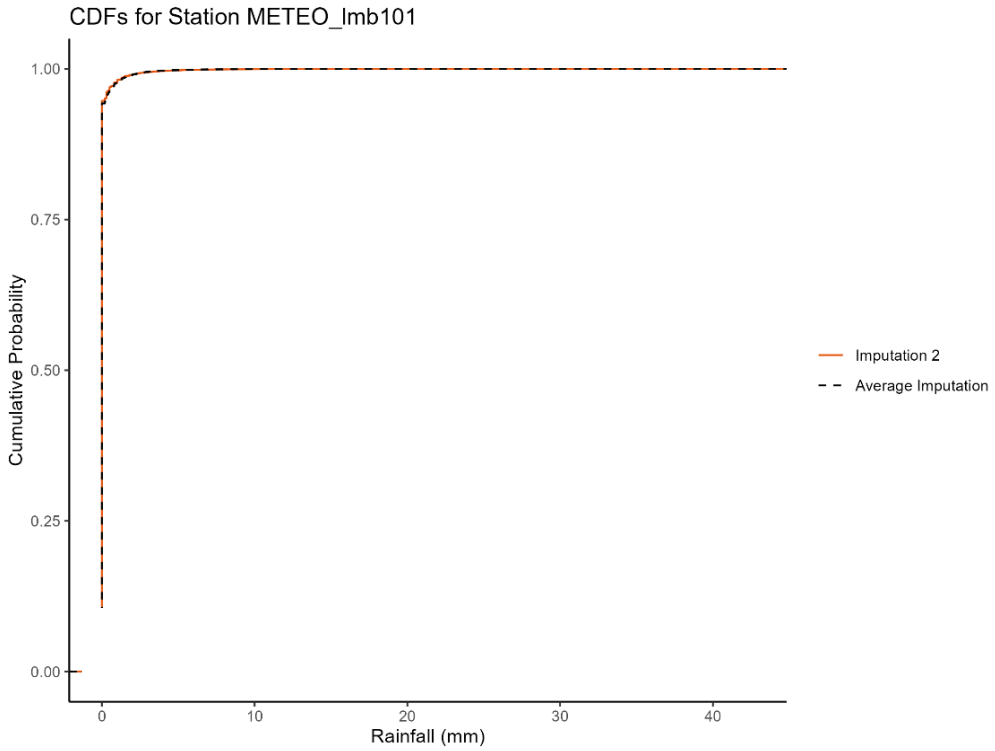


Figure 65. Empirical Cumulative Distribution Functions (CDFs) of Rainfall for Station METEO lmb101 comparing Average Imputation and Second Imputation

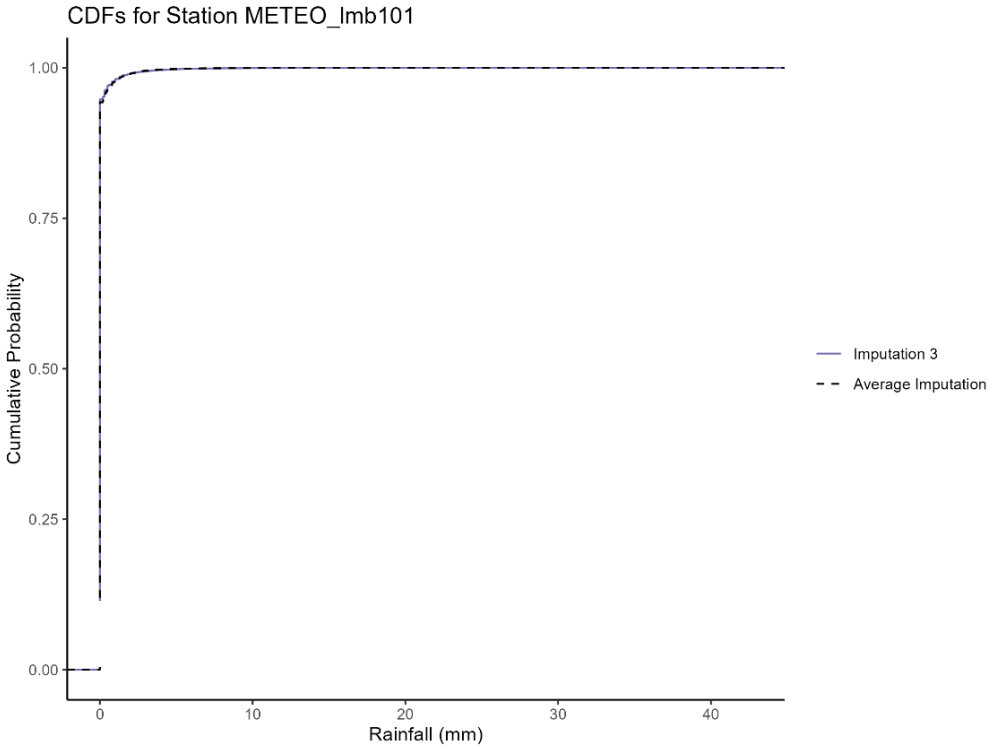


Figure 66. Empirical Cumulative Distribution Functions (CDFs) of Rainfall for Station METEO lmb101 comparing Average Imputation and Third Imputation

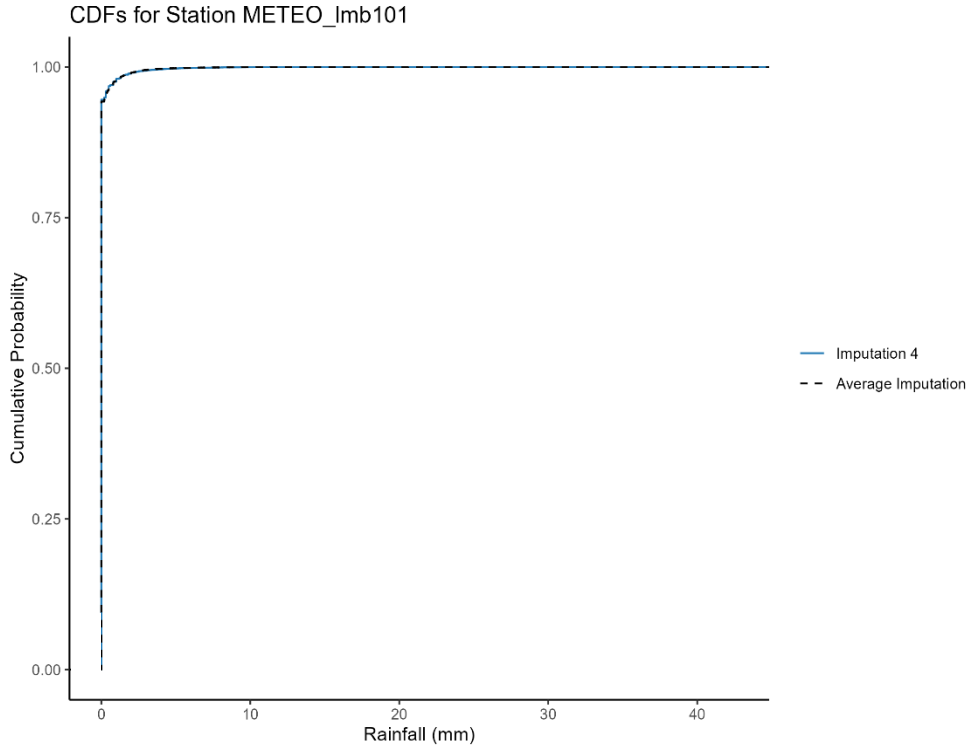


Figure 67. Empirical Cumulative Distribution Functions (CDFs) of Rainfall for Station METEO lmb101 comparing Average Imputation and Fourth Imputation

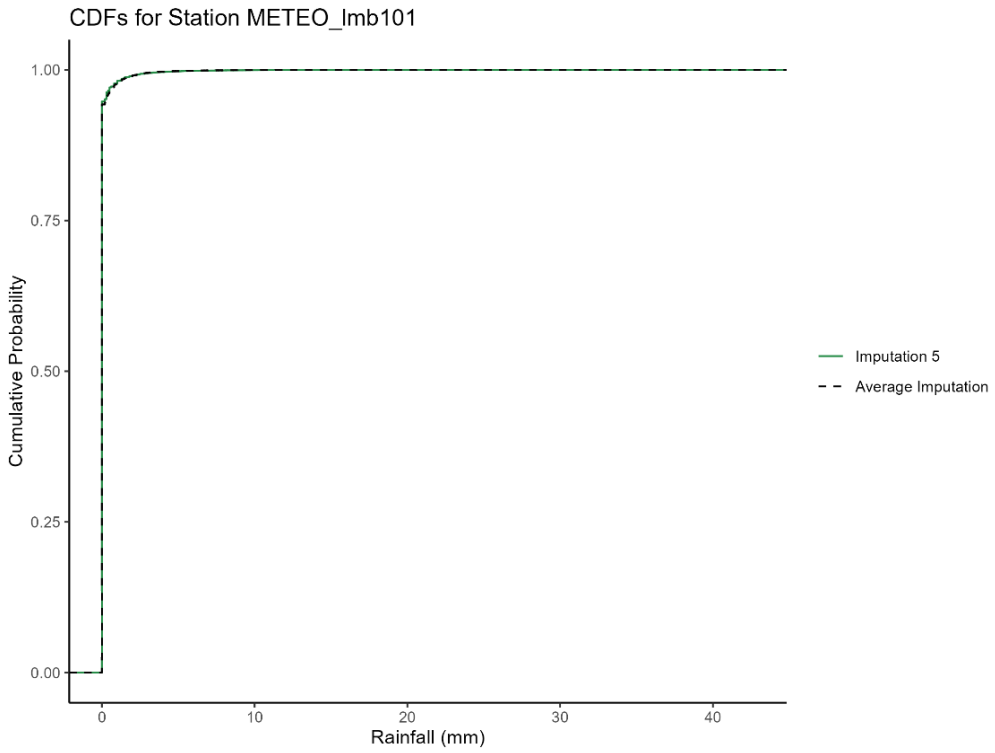


Figure 68. Empirical Cumulative Distribution Functions (CDFs) of Rainfall for Station METEO lmb101 comparing Average Imputation and Fifth Imputation

During the averaging process, values in the imputed dataset could fall below the "0.2 mm" threshold, which is the sensitivity of the rain gauge. For instance, if some imputations return small rainfall amounts (e.g., 0.2 mm) and other return 0 mm, the average may fall below the rain gauge's detection limit of 0.2 mm. This results in imputed values that could not have been recorded by the original instrument, making them inconsistent with the nature of the observed data. Therefore, any imputed values below 0.2 mm were set to 0 to reflect the minimum measurable value that the rain gauges can detect. This ensures that all imputed values align with the physical limits of the measuring equipment and prevents the inclusion of implausibly small values that are not supported by the instrument's sensitivity.

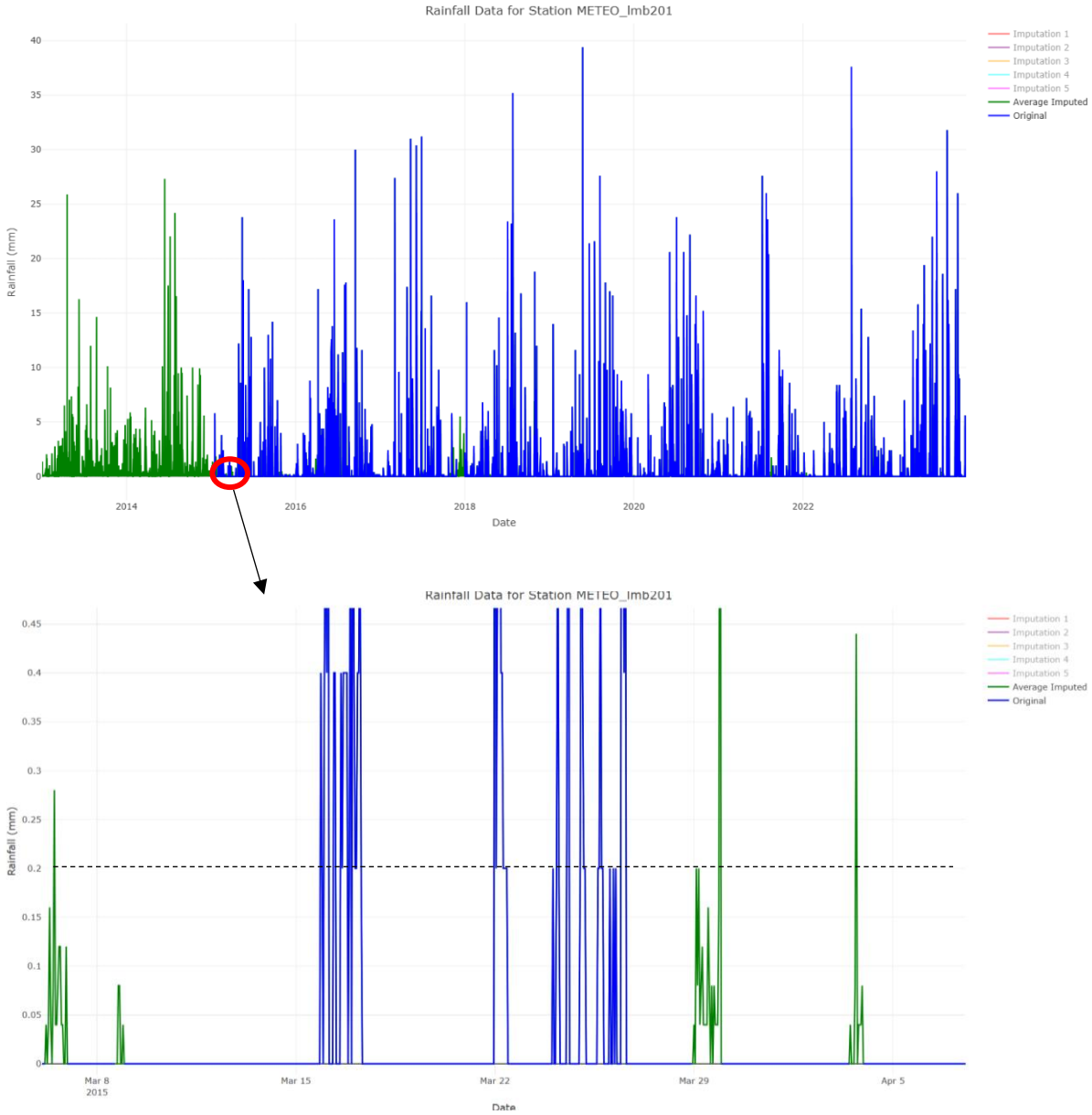


Figure 69. Example of a Station having Imputed Rainfall (Average) less than 0.2 mm

4.4. Issues with Direct Imputation

After the initial phase of imputation using solely the pmm method, the imputed data was examined by visualizing the rainfall time series for the period from 2013 to 2023. Upon analysis, a significant issue became apparent: all five imputations, as well as their average, showed a nearly continuous presence of rainfall throughout the time series. Although some values were as low as 0.2 mm, the imputations lacked the frequent zero-rainfall periods typical of real-world data. In other words, the zero-inflated nature of the original rainfall distribution was lost, resulting in an unrealistic smoothing effect where rain appeared to occur almost constantly, even during typically dry periods.

While some time points in the imputed data did register zero rainfall, they were immediately followed by non-zero values, resulting in an almost uninterrupted sequence of rainfall throughout the period. This did not align with the expected temporal pattern of hourly rainfall, where dry periods, characterized by a lack of rainfall over extended periods, should naturally occur. The direct imputation method failed to capture these dry periods properly, and instead, the data exhibited an unrealistic consistency, suggesting that rainfall was always occurring, even if at very low levels.

This issue was particularly pronounced in stations with low coverage. For stations with limited data availability, the imputation process struggled to correctly represent the variability in rainfall patterns, leading to the unrealistic continuity of rainfall. These stations were more susceptible to showing this false consistency in rainfall, as the imputed values lacked the proper variation and dry intervals that would typically occur in observed data.

Thus, this imputation method did not effectively reproduce this temporal behavior, where periods of dryness (i.e., no rainfall) alternate with periods of rain. This led to the adoption of an original double imputation strategy. This method was implemented to better account for both the occurrence of rainfall events and the corresponding dry periods, producing a more realistic and temporally accurate dataset.

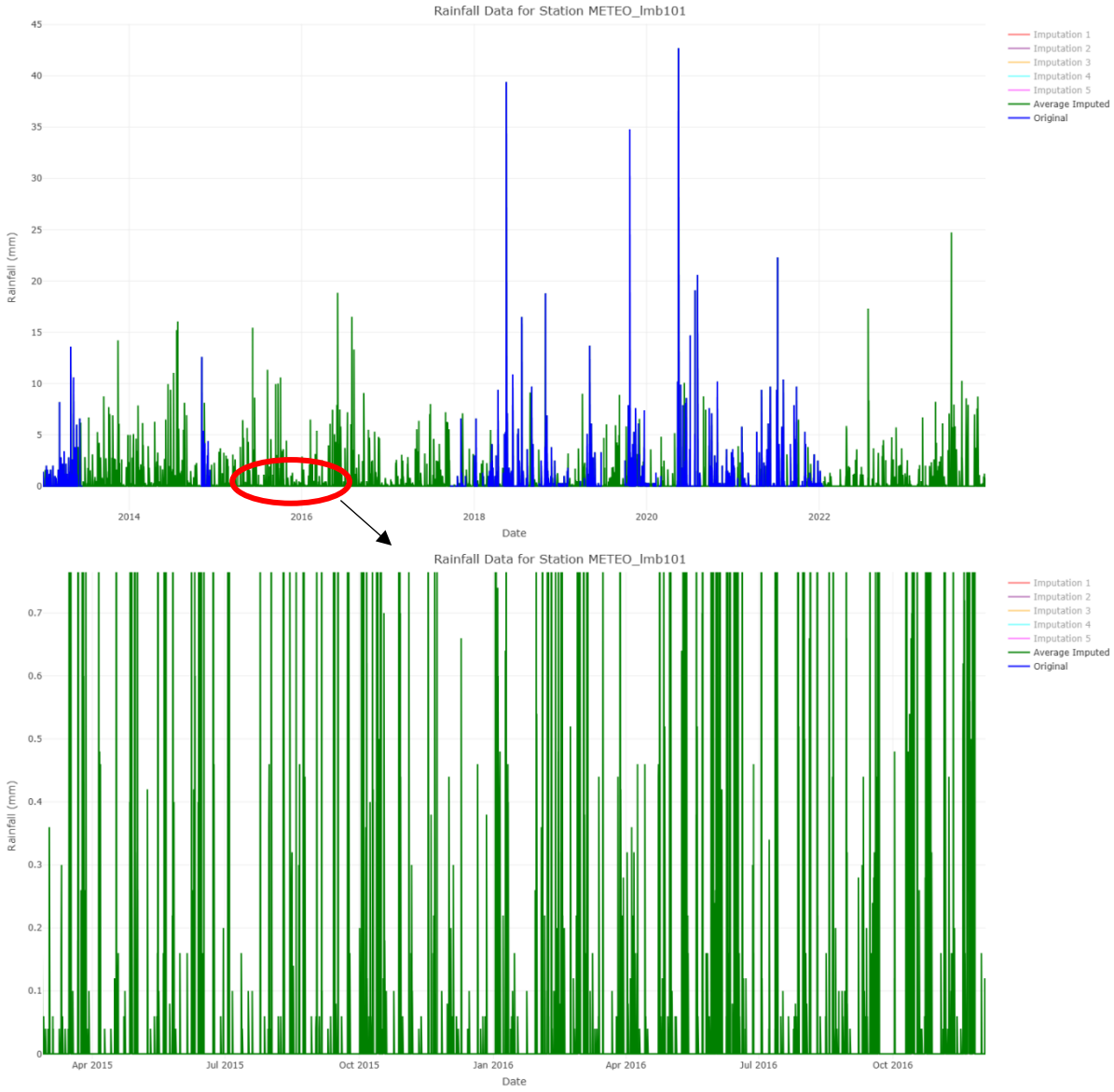


Figure 70. Example of a Station showing unrealistic fluctuations in imputed rainfall

4.5. Development of Double Imputation Approach (MICE-2)

STEP 1: CONVERSION TO BINARY WET/DRY DATA

The original non-imputed rainfall dataset, "data_wide_num", containing hourly rainfall measurements across multiple stations, was first transformed into a binary format named "binary_wetdry" (see Figure 70). This transformation aimed to simplify the data into a wet/dry classification that would later be used to correct the unrealistic persistence of light rainfall in standard imputed datasets. In this step, each value in the dataset was assessed:

- If a value was greater than zero, it was considered an indication of rainfall occurrence (a wet hour) and was thus converted to 1.
- If a value was exactly zero, it was considered a "dry hour" and left as 0.
- Any missing values (NA) in the original dataset were retained as "NA," to be imputed later.

This binary representation captured the fundamental occurrence pattern of rainfall events without regard to their magnitude. It served as the foundation for the next phase of imputation, where the presence or absence of rainfall was imputed separately from the amount, thereby helping to realistically reconstruct dry periods that were previously masked by continuous low-intensity imputed rainfall in the original multiple imputation procedure.

```
binary_wetdry <- data_wide_num

# Turn all values > 0 into 1
binary_wetdry[binary_wetdry > 0] <- 1
```

Figure 71. Code for conversion to binary wet/dry data (RStudio)

STEP 2: IMPUTING THE BINARY WET/DRY DATA

To impute missing values in the binary dataset, the "logreg" method (logistic regression) was used. This method is well-suited for binary outcomes, as it estimates the probability of a wet hour (1) based on a logistic regression model. It models the relationship between the observed values and the missing data, ensuring that imputed values are consistent with the underlying distribution of the binary wet/dry data.

For the binary imputation, the same predictor matrix used in the continuous imputation was applied, based on the previously described 11 km distance threshold. This spatial constraint preserved the local spatial dependence between stations, ensuring that the imputed wet/dry patterns remained geographically relevant.

```
binary_wetdry_imp<-mice(binary_wetdry, method="logreg", predictorMatrix = pred_matrix, seed = 123)
```

Figure 72. Code for imputation using logreg (RStudio)

Once the imputation was performed, the result was stored in "binary_wetdry_comp," which was generated by extracting the completed dataset using the "complete()" function from the mice package.

The final output was a fully populated binary matrix where each entry indicated whether a specific hour was wet (1) or dry (0). This imputation process allowed for the filling of missing entries in the dataset, even at time steps where the original data had no observed values, thus ensuring a more complete dataset for further analysis.

ARPA_5908	ARPA_5916	ARPA_8122	ARPA_8152	ARPA_8197	ARPA_8199	ARPA_8211	ARPA_8228	METEO_lmb009	METEO_lmb021	METEO_lmb080	METEO_lmb084	METEO_lmb087
1	1	1	1	1	1	1	1	0	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	0	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1
1	1	1	0	1	1	1	0	1	1	1	1	1
1	1	0	1	1	1	1	0	0	1	1	1	0
1	1	1	0	1	1	1	0	1	1	0	1	1
1	1	0	0	1	1	1	0	0	0	1	0	0
0	1	0	0	1	0	0	1	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0	0	0	0	0
1	1	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0	0	1

Figure 73. Sample binary mask (wet/dry) for 13 stations over a 24-hour period (2/1/2013), showing imputed wet/dry status

STEP 3: MASKING THE IMPUTED RAINFALL VALUES

After imputing the binary wet/dry structure, the next step was to filter or mask the previously imputed continuous rainfall values using this binary structure. The goal was to ensure that rainfall values were only retained during hours where the binary model had predicted a wet condition (i.e., hours where the binary mask was 1).

The first step involved applying the binary mask to each of the five imputed datasets, which were generated earlier using Predictive Mean Matching (pmm). At each location in the matrix, the binary wet/dry mask was used to determine whether the imputed rainfall value should be retained or removed. The mask provided information about whether each hour was predicted to be wet or dry:

1. When the binary mask had a value of 0 (indicating a predicted dry hour), the corresponding imputed rainfall value was forcibly set to 0. This step ensured that no rainfall values were present during hours where the binary model had predicted dry conditions.

2. When the binary mask had a value of 1 (indicating a predicted wet hour), the imputed rainfall value was left unchanged. This step ensured that the imputed rainfall values remained intact during hours where the binary model had predicted wet conditions. These wet hours were considered valid for rainfall occurrence, and thus, the imputed values were preserved.

```
imputed_list_masked <- lapply(imputed_list, function(df) {
  df_subset <- df[rows_to_keep, ]
  df_subset[binary_mask == 0] <- 0
  return(df_subset)
})
```

Figure 74. Code for Application of the binary wet/dry mask to remove imputed rainfall during predicted dry hours across all datasets

By using this approach, the masking process effectively removed any implausible imputed rainfall values that could have been incorrectly assigned to dry hours. This was crucial for ensuring that rainfall data were properly imputed maintaining their nature.

After applying the masking process to each of the five imputed datasets, a new set of imputed datasets was obtained. Again, the average of these five imputed datasets was computed, resulting in a new average imputation dataset.

Next, again a filter was applied on the average imputed dataset to remove any rainfall values less than 0.2 mm. This step was undertaken to ensure the consistency and accuracy of the imputed data, as previously discussed. By removing these small values, the dataset was further cleaned to reflect more plausible rainfall amounts and to prevent the inclusion of insignificant values that would not have meaningful impacts on the analysis. This step ensured that the imputed rainfall data was both reasonable and relevant for subsequent analyses.

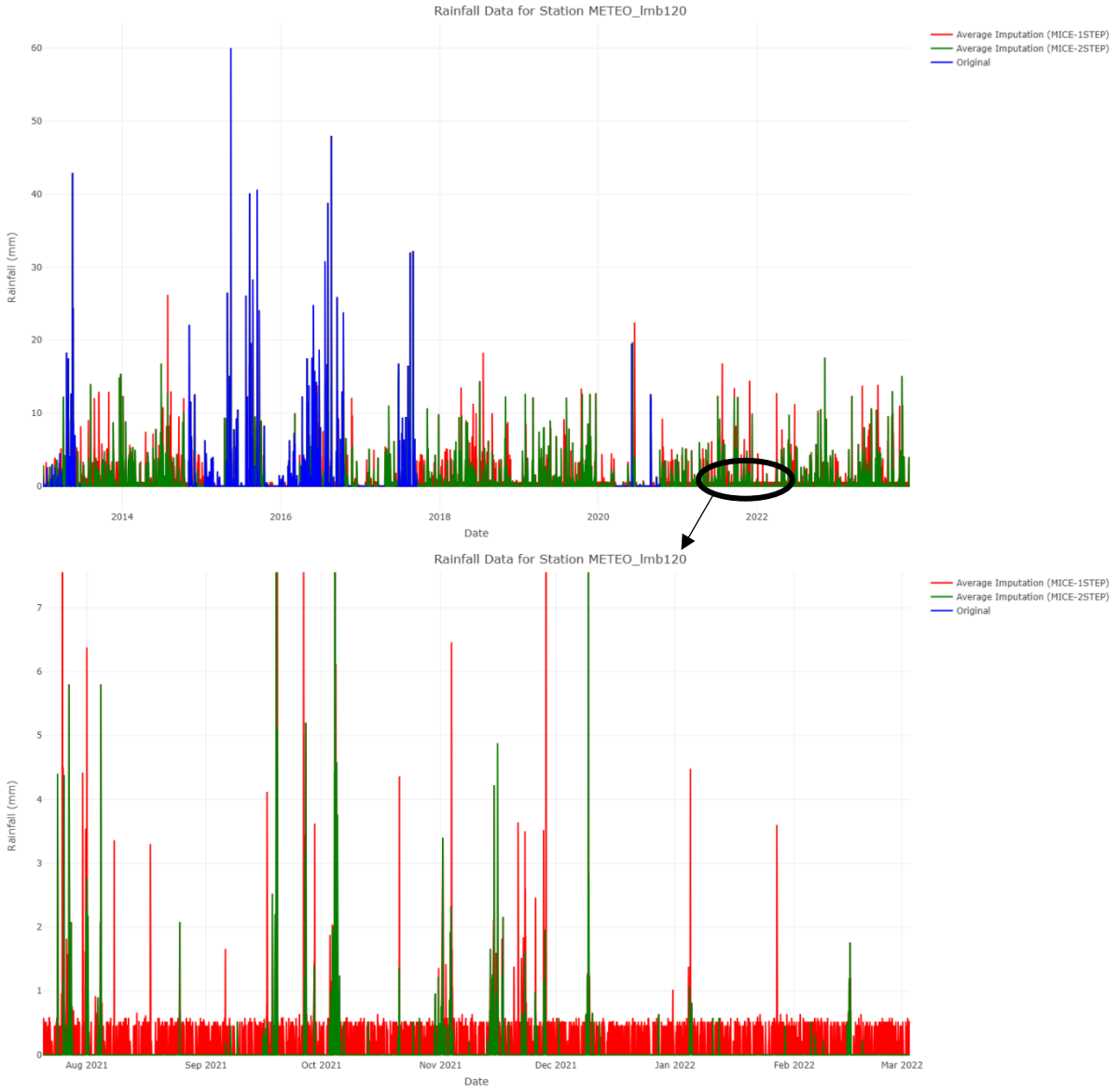


Figure 75. Comparison between MICE-1 and MICE-2

Figure 75 illustrates a time series comparison of rainfall data for Station METEO lmb120, highlighting the differences between the two imputation strategies i.e., MICE-1STEP (in red) and MICE-2STEP (in green).

The red bars, representing the average imputed values from the MICE-1STEP method, are densely distributed, with frequent low-intensity rainfall estimates, even during clearly dry periods. This suggests that the single-step imputation approach tends to overestimate rainfall, often introducing small values where no precipitation likely occurred.

In contrast, the green bars, which depict the results of the MICE-2STEP approach, show a markedly more selective distribution. These imputed values are limited mostly to periods that are more plausibly wet and align more closely with the real-world observations. This pattern results from the additional binary wet/dry filtering applied

before finalizing the imputed rainfall, effectively suppressing spurious drizzle introduced in the first-step imputations.

Overall, the comparison highlights that the double imputation strategy produces a more realistic and conservative estimate of rainfall by reinforcing dry-hour constraints, whereas the one-step method may artificially inflate rainfall occurrence.

4.6. Methods for Statistical Validation

4.6.1. Visual Inspection using Empirical Cumulative Distribution Functions (ECDFs)

As a preliminary step in assessing the quality of imputed rainfall data, empirical cumulative distribution functions (ECDFs) were generated. An ECDF provides a non-parametric estimate of the cumulative distribution function of a dataset, showing the proportion of observations less than or equal to a given value. Mathematically, for a sample of size n, the ECDF at a value x is defined as:

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} I(x_i \leq x)$$

Equation 7. ECDF Equation

where I(·) is the indicator function that equals 1 if the condition is true and 0 otherwise.

The ECDF offers an intuitive and visual means of comparing distributions, making it particularly useful for evaluating how closely the imputed data aligns with the original observations. If the ECDF curves of the original and imputed data match closely, it suggests that the imputation has preserved the underlying distributional characteristics of the rainfall data.

This structured ECDF analysis served as a preliminary yet essential visual validation step prior to formal statistical testing, helping to highlight potential deviations or distributional mismatches between original and imputed data across various conditions.

4.6.2. Formal Testing with the Kolmogorov–Smirnov (Ks) Test

To complement the visual assessment, a Kolmogorov–Smirnov (KS) test was applied for statistical validation. The KS test is a non-parametric method for comparing the distributions of two datasets. It quantifies the maximum difference between the empirical distribution functions (ECDFs) of the observed and imputed data.

Mathematically, the test statistic is defined as:

$$D = sup_x|F_1(x) - F_2(x)|$$

Equation 8. Kolmogorov–Smirnov statistic (D)

where $F_1(x)$ and $F_2(x)$ are the ECDFs of the observed and imputed data respectively, and sup_x is the supremum (i.e., the maximum over all x).

The null hypothesis of the KS test states that the two samples are drawn from the same distribution. Once D is computed, the p-value is derived using the Kolmogorov distribution, which gives the probability of observing a difference as extreme as D, assuming the null hypothesis is true (i.e., both samples are from the same continuous distribution). A p-value indicates the statistical significance of the observed difference. High p-values (typically p>0.05) suggest that the imputed data resemble the distribution of the original values.

For two samples of sizes n_1 and n_2, the test uses the effective sample size:

$$n_{eff} = \frac{n_1 . n_2}{n_1 + n_2}$$

Equation 9. Effective sample size n_{eff} used in the two-sample Kolmogorov–Smirnov test

The p-value is then approximated using:

$$p = Q_{KS}(\sqrt{n_{eff}}.D)$$

Equation 10. Approximate computation of the p-value in the KS test using the Kolmogorov distribution function Q_{KS}

Where Q_{KS} is a function derived from the Kolmogorov distribution, and its exact form involves a converging series (or is approximated numerically in practice).

In R (or most Statistical Software)

KS test and p-value are computed using the function "ks.test(x, y)" which automatically calculates D and then evaluates the corresponding p-value using asymptotic or exact methods, depending on sample size and assumptions.

4.7. Results

This section provides a general description of how the ECDFs were generated, comparing the observed and imputed rainfall data for the different scenarios and cases. To evaluate the quality of imputed rainfall data, a subset of stations was selected for detailed analysis. This subset includes both ARPA and METEO stations, chosen to represent a range of data coverage scenarios and spatial distributions. lmb084 had the highest coverage with 92.24%, followed by lmb183 with 78.82%, lmb080 with 49.44%, lmb333 with 24.75%, and lmb323 with 8.35%.

The stations analyzed in this section are:

- **ARPA 5908** (high coverage – 92.76% - lowest among ARPA stations)
- **ARPA 5916** (high coverage – 95.23% - second lowest among ARPA stations)
- **METEO lmb084** (high coverage – 92.24%)
- **METEO lmb183** (moderate coverage – 78.82%)
- **METEO lmb080** (moderate coverage – 49.44%)
- **METEO lmb333** (low coverage – 24.75%)
- **METEO lmb323** (very low coverage - 8.35%)

These stations were selected to provide a comprehensive view of how imputation quality varies across different conditions. The results for each are presented in the following subsections, organized by temporal resolution and seasonal context.

4.7.1. Moments

To assess whether the imputation process preserved the overall statistical properties of the original rainfall data, we compared the mean and standard deviation of hourly rainfall values before and after imputation across all stations. These basic moments provide insight into how well the imputed datasets maintain both the central tendency and the variability of the original time series.

Station	Coverage	Original Data	Average Imputation	Imputation 1	Imputation 2	Imputation 3	Imputation 4	Imputation 5
ARPA 5908	92.76 %	0.087	0.087	0.087	0.087	0.087	0.087	0.087
METEO lmb084	92.37 %	0.166	0.160	0.161	0.161	0.160	0.160	0.162
METEO lmb183	78.82 %	0.106	0.099	0.098	0.098	0.100	0.098	0.099
METEO lmb087	67.39 %	0.140	0.138	0.139	0.138	0.139	0.140	0.138
METEO lmb293	47.47 %	0.115	0.100	0.100	0.099	0.098	0.099	0.100
METEO lmb333	24.75 %	0.132	0.121	0.121	0.121	0.120	0.121	0.121
METEO lmb323	8.35 %	0.100	0.085	0.083	0.081	0.081	0.081	0.077

Table 4. Mean of hourly rainfall (mm) across five representative stations, comparing original data and imputed datasets

Station	Coverage	Original Data	Average Imputation	Imputation 1	Imputation 2	Imputation 3	Imputation 4	Imputation 5
ARPA 5908	92.76 %	0.660	0.650	0.649	0.655	0.657	0.655	0.659
METEO lmb084	92.37 %	1.144	1.112	1.130	1.122	1.124	1.119	1.128
METEO lmb183	78.82 %	0.875	0.814	0.818	0.839	0.848	0.823	0.850
METEO lmb087	67.39 %	1.064	0.997	1.087	1.045	1.080	1.092	1.045
METEO lmb293	47.47 %	0.981	0.853	0.915	0.889	0.894	0.894	0.924
METEO lmb333	24.75 %	0.943	0.838	0.902	0.887	0.892	0.904	0.915
METEO lmb323	8.35 %	0.722	0.549	0.722	0.665	0.666	0.669	0.665

Table 5. Standard deviation of hourly rainfall (mm) across five representative stations, comparing original data and imputed datasets

Table 4 and Table 5 presents a summary of five representative stations selected to cover a range of data coverage levels:

- **High coverage**: ARPA 5908, METEO lmb084.
- **Moderate coverage**: METEO lmb183, METEO lmb087.
- **Low coverage**: METEO lmb293, METEO lmb333, METEO lmb323.

For each station, the mean and standard deviation of hourly rainfall from the original dataset and the average of the five imputations are reported. This selection illustrates how the imputation process affects basic distributional characteristics across varying levels of data coverage.

Station-Wise Mean Rainfall across Imputations

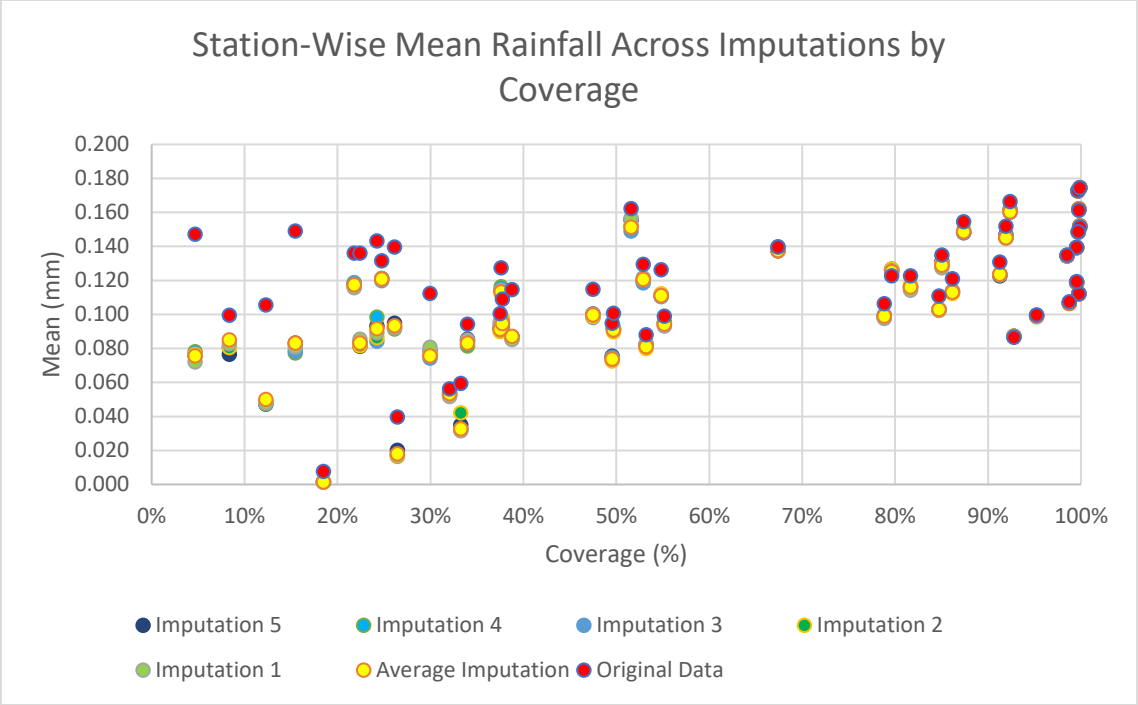


Figure 76. Station-wise mean rainfall as a function of data coverage

Figure 76 displays the mean hourly rainfall for each station plotted against its data coverage percentage, with each point representing one of the five imputations, their average, or the original value. For high-coverage stations (≥90%), the imputed means align almost perfectly with the original data, showing negligible variation across imputations. As we move into the moderate coverage range (approximately 50% to 90%), the means remain largely consistent, although some minor dispersion among imputations becomes visible. Despite this, the averaged imputed value continues to closely reflect the original mean. In contrast, low-coverage stations (below 50%) exhibit greater variability, with a wider spread across imputations and more noticeable deviations from the original values. Even in these cases, however, the results remain broadly reasonable, suggesting that while uncertainty increases, the imputation model still provides plausible estimates of central tendency. Overall, the figure illustrates that above a coverage threshold of around 70%, the mean rainfall is reliably preserved by the imputation process, with increasing uncertainty only becoming apparent as coverage declines further.

Station-Wise Standard Deviation across Imputations

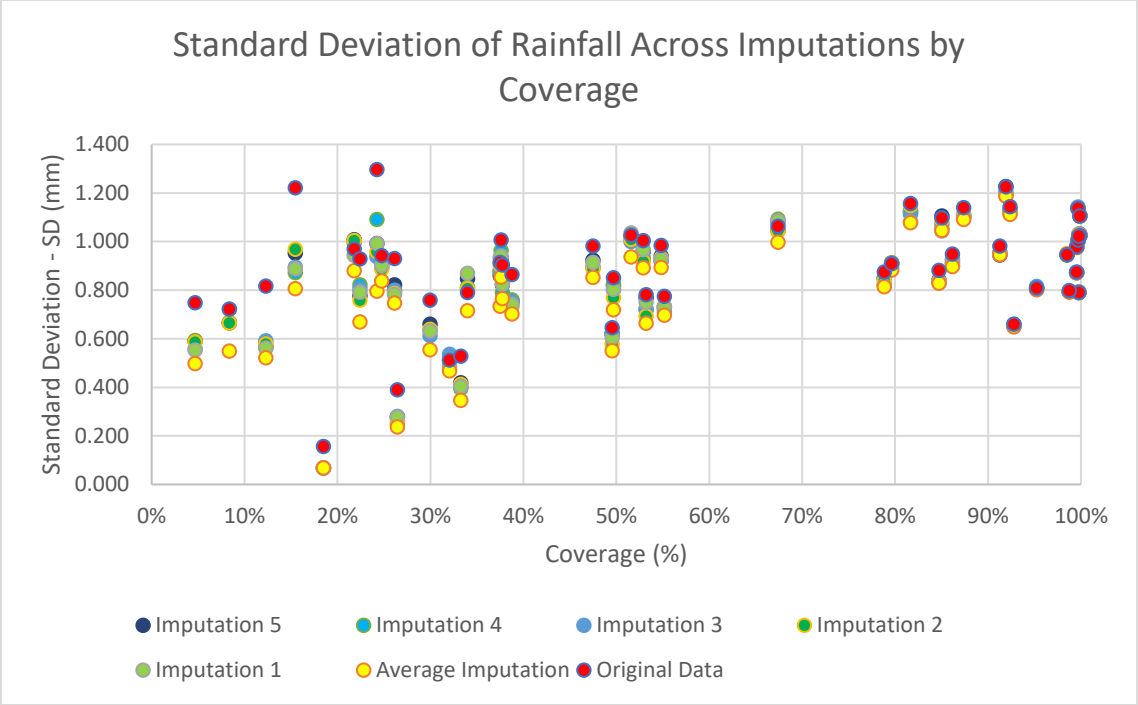


Figure 77. Station-wise standard deviation of rainfall as a function of data coverage

Figure 77 shows the standard deviation of hourly rainfall at each station across imputations, plotted against data coverage percentage. For high-coverage stations (≥90%), the standard deviation remains highly consistent with the original values across all imputations, indicating that the natural variability of rainfall is well preserved. In the moderate coverage range (70% to 90%), the standard deviation still generally follows the original values, though with slight compression in some cases, reflecting the model's tendency to fill gaps with more moderate values. The effect becomes more pronounced in low-coverage stations (below 70%), where a clear reduction in standard deviation is observed. This smoothing is especially evident for stations with less than 30% coverage, where the scarcity of observed data limits the model's ability to reconstruct extreme variability, resulting in a narrower spread of imputed rainfall values. Overall, the figure highlights that the imputation method maintains the underlying variability well in stations with adequate data and that performance gradually declines as coverage becomes sparse.

These figures together illustrate that above approximately 70% coverage, both the mean and standard deviation of the imputed datasets closely resemble the original data. Below this threshold, greater divergence can occur, especially in standard deviation, due to reduced observational support. This observation is consistent with the trends later confirmed by statistical validation in Section 4.7.3.

4.7.2. ECDF Analysis

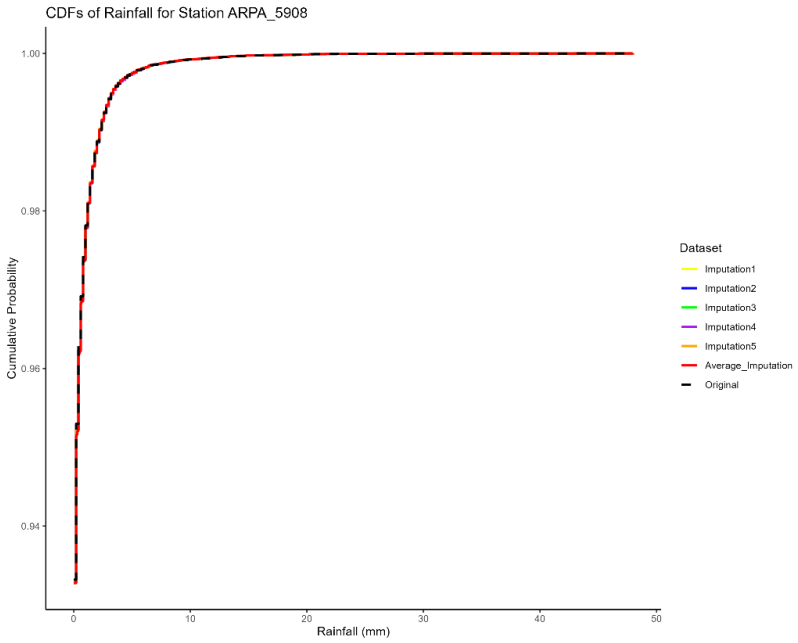


Figure 78. CDF for ARPA Station 5908 (Hourly-Full Year-With Zero)

ARPA Station 5908 was selected for display because, out of the 11 ARPA stations, it had the most missing data i.e., 7.24%. Despite the slight missing data, the CDFs for both original and imputed data overlap almost perfectly, indicating that the imputation method was able to closely replicate the rainfall distribution even with some gaps in the data. The visual validation shows that imputation for ARPA station 5098 performs consistently well. For the remaining ARPA stations, which had nearly complete coverage (around 99%), the CDFs showed a similar behavior, suggesting that the imputation method performed well in cases where data coverage was nearly full, with any small imputation errors having a negligible effect on the overall distribution.

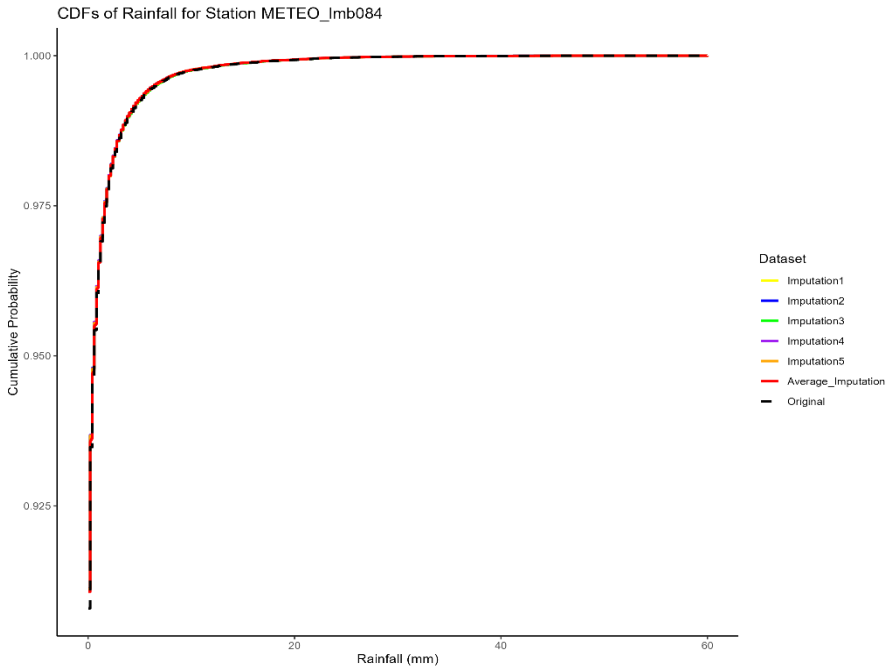


Figure 79. CDF for METEO station lmb084 (Hourly-Full Year-With Zero)

Figure 79 presents the ECDFs for METEO Station lmb084, which has a high data coverage of 92.24%. The curves for all five imputations, their average, and the original data overlap almost perfectly. This high level of agreement confirms that the imputation process preserved the original rainfall distribution with minimal distortion. The ability of the method to maintain both the frequency of zero rainfall and the spread of non-zero rainfall events at this station demonstrates strong performance when observational data are nearly complete.

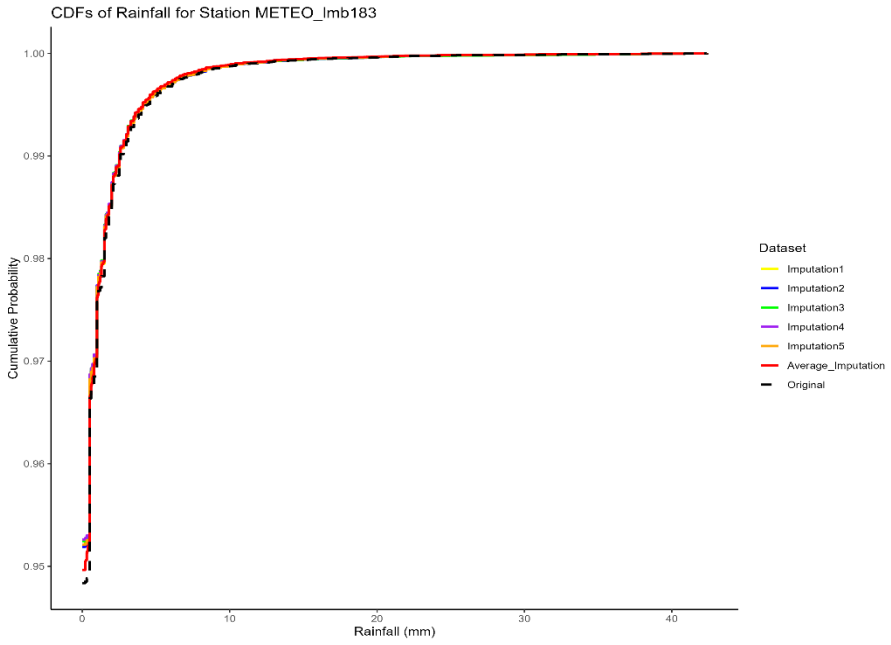


Figure 80. CDF for METEO station lmb183 (Hourly-Full Year-With Zero)

Figure 80 presents the ECDFs for METEO Station lmb183, which has a moderate data coverage of 78.82%. The imputed distributions show a high degree of agreement with the original data, with only a slight upward shift observed in the lower rainfall range (approximately 0–3 mm). This minor divergence reflects the tendency of the imputation process to fill missing values with either dry periods or low-intensity rainfall, which slightly increases the frequency of small values in the dataset. Beyond this range, the curves converge closely, with excellent alignment across the mid and upper portions of the distribution. Overall, the results confirm that the imputation preserved the original distribution well, even under moderate data availability.

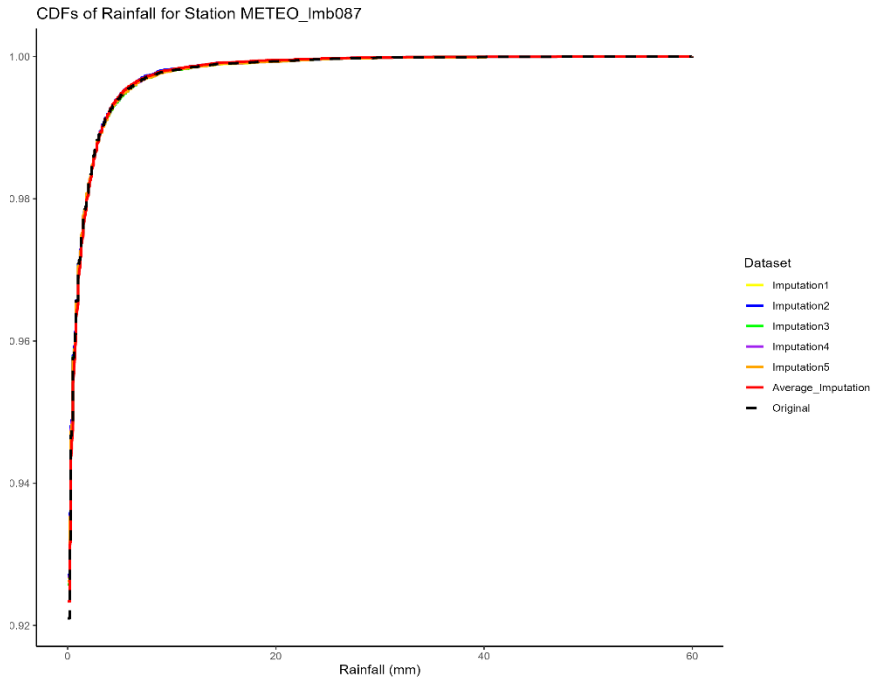


Figure 81. CDF for METEO station lmb87 (Hourly-Full Year-With Zero)

Figure 81 presents the ECDFs for METEO Station lmb087, which has a moderate data coverage of 67.39%. Despite the missing data, the imputed distributions align closely with the original ECDF across the full rainfall range. A slight upward shift is visible in the lower rainfall region, indicating that the imputation process introduced a modest increase in the frequency of dry or low-rainfall values. This behavior reflects the model's effort to maintain realistic rainfall patterns in the absence of observed data. Importantly, the mid to upper portions of the distribution remain nearly identical to the original, demonstrating that even at sub-70% coverage, the method still preserves key distributional characteristics effectively.

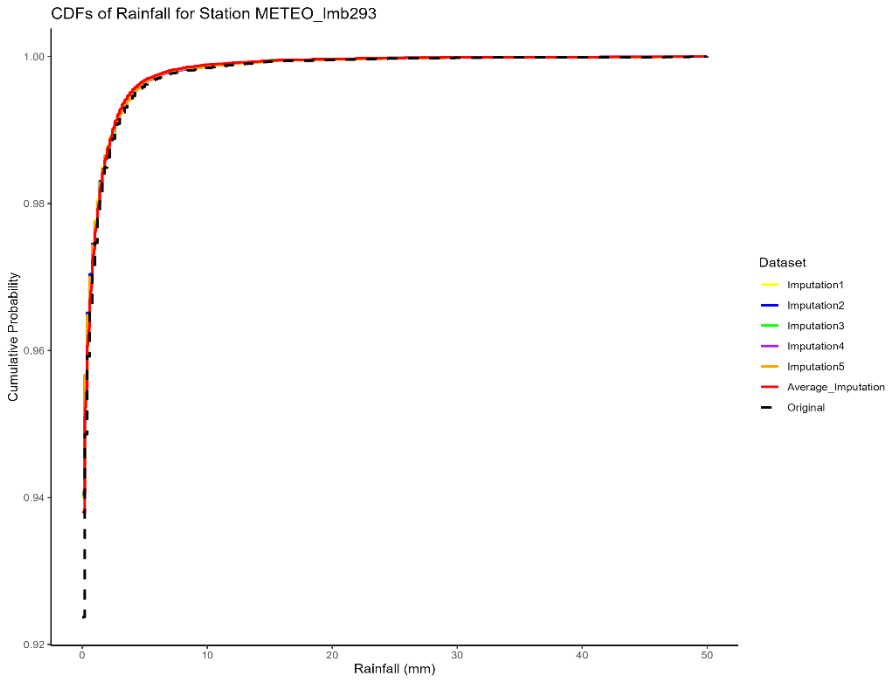


Figure 82. CDF for METEO station lmb293 (Hourly-Full Year-With Zero)

Figure 82 displays the ECDFs for METEO Station lmb293, which has a data coverage of 47.47%. The imputed and original distributions align well overall, particularly in the mid and upper rainfall ranges. However, a clear upward shift is visible near the lower tail, indicating a higher proportion of dry or low-rainfall values in the imputed data. This outcome is typical for low-coverage stations, where a significant portion of the dataset must be estimated and the imputation model tends to favor light rainfall or dry hours in the absence of strong supporting evidence for more extreme events.

While lmb293 has nine stations within the 11 km spatial threshold, including one ARPA and three high-coverage METEO stations, most of these stations are located at distances of 7 km or more. Given the localized nature of convective rainfall, this spatial separation means that extreme events occurring at lmb293 may not be reflected in nearby stations, even those with good data quality. Consequently, the imputation model may underrepresent high-intensity events at lmb293, resulting in a steeper rise in the cumulative distribution near 0 mm and a subtle smoothing of variability overall. Despite this, the upper end of the distribution remains consistent with the original, indicating that the method still retains the general structure of the rainfall distribution under constrained conditions.

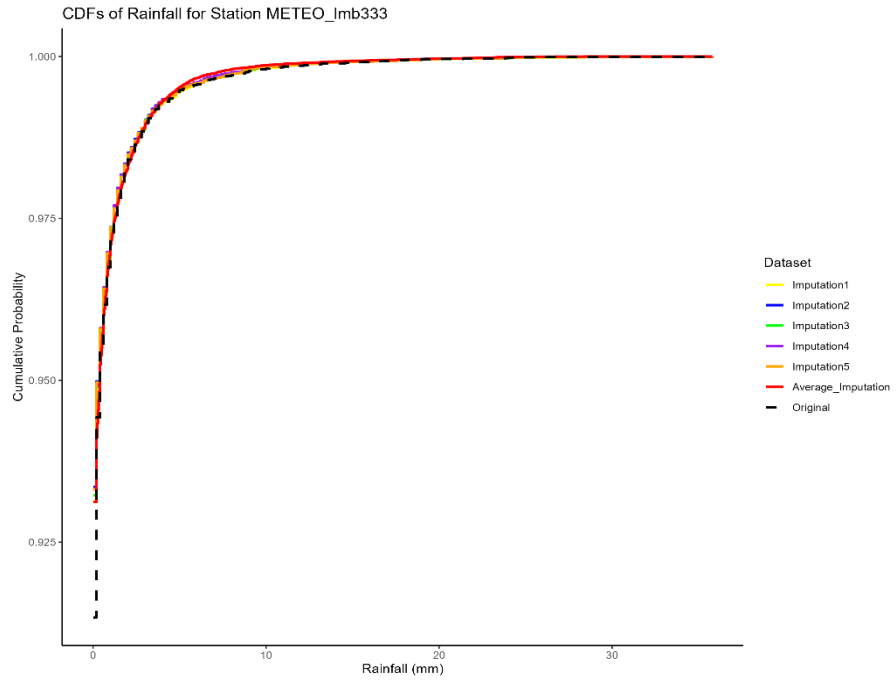


Figure 83. CDF for METEO station lmb333 (Hourly-Full Year-With Zero)

Figure 83 shows the ECDFs for METEO Station lmb333, which has a data coverage of 24.75%, placing it in the low-coverage category. While the overall shape of the imputed distributions broadly mirrors the original, a clear upward shift is seen at the lower end of the rainfall range. This shift indicates the representation of dry or low-intensity rainfall values in the imputed data compared to the original observations.

Given the limited amount of observed data, the model relies heavily on nearby stations to reconstruct missing values. In this case, although neighboring stations may have moderate or high coverage, their spatial separation, or timing differences in capturing rainfall events, may reduce their ability to accurately represent localized extremes at lmb333. As a result, the imputation tends to favor more frequent, low-magnitude values, which explains the steeper ECDF curve near 0 mm.

Despite these deviations in the lower range, the alignment in the upper part of the distribution suggests that the imputation process is still capable of maintaining a reasonable approximation of higher rainfall values. However, the results for lmb333 highlight the limitations of the method when both data coverage and spatial representativeness are low.

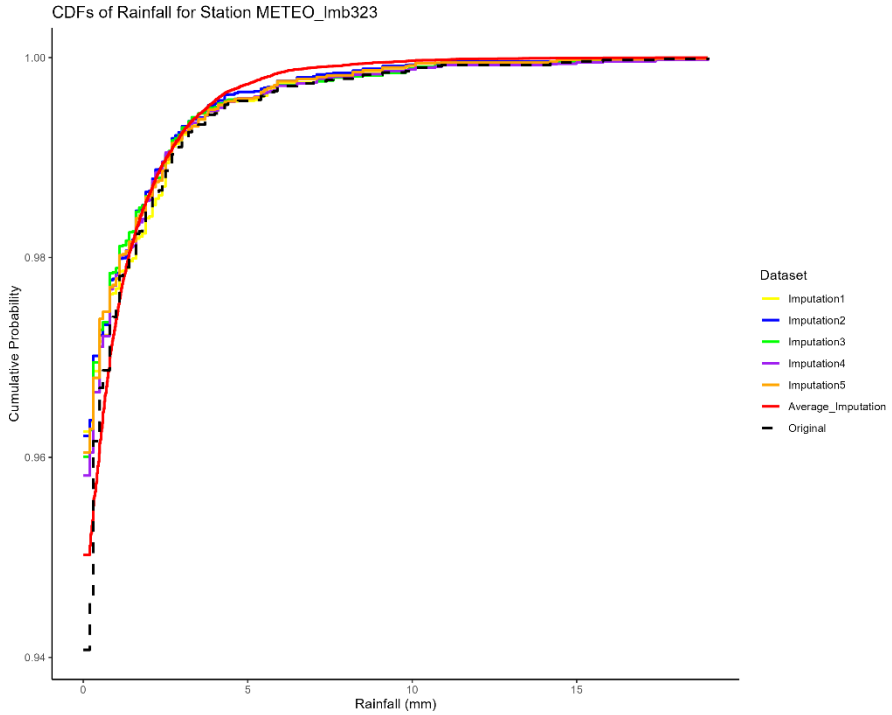


Figure 84. CDF for METEO station lmb323 (Hourly-Full Year-With Zero)

Figure 84 displays the ECDFs for METEO Station lmb323, which has a data coverage of just 8.35%, making it one of the sparsest stations in the network. In this case, the imputed ECDFs show noticeable divergence from the original, particularly in the lower and mid-range of rainfall values. The spread among the five individual imputations is more pronounced compared to other stations, indicating greater uncertainty in reconstructing the rainfall distribution.

The average imputed ECDF exhibits a sharper rise in the lower rainfall region, reflecting an increased frequency of dry or light rain values in the reconstructed data. This pattern arises because, at such low coverage, the model must rely almost entirely on information from neighboring stations. While lmb323 has nine stations within 11 km, only three of them have high coverage, and all are located more than 7 km away. Given the localized nature of intense rainfall events, this spatial separation reduces the chance of coinciding extremes across stations, limiting the model's ability to faithfully reconstruct variability or intensity.

There is a visible discrepancy between the original and imputed curves across a wider range of rainfall values. Nevertheless, the method almost preserves the general shape of the distribution, and the upper tail aligns reasonably well. This figure highlights the practical limitations of imputation under extreme data sparsity, especially when the spatial neighbors are distant or inconsistent.

As we move from high-coverage to low-coverage stations, a consistent pattern emerges across the ECDFs. Stations with high coverage (≥90%) show near-perfect agreement between the original and imputed distributions, with minimal deviation

across all rainfall ranges. In moderate-coverage stations (~50–90%), the alignment remains strong, though slight shifts in the lower tail appear due to increased imputation of dry or low-intensity events. These differences become more pronounced in low-coverage stations (<50%), where the imputed ECDFs tend to rise more steeply near 0 mm and display greater variability between runs. This reflects both the increasing reliance on model assumptions and the reduced representativeness of neighboring stations, especially for capturing local extremes.

The METEO stations, which vary more widely in data coverage (from ~8% to ~90%), provide a more diverse picture. Visual validation indicates that stations with moderate to high coverage (e.g., lmb084, lmb183, lmb080) show strong agreement between original and imputed datasets. Rainfall patterns, particularly when aggregated to daily totals, are well preserved, and the distributions remain coherent. Minor shifts, such as an increase in moderate rainfall or dry periods, appear reasonable and expected.

In low-coverage cases, the results are more mixed and more revealing. Notably, two stations with similarly low coverage produced very different outcomes. Station lmb333 displayed excellent agreement between original and imputed data, while lmb323 showed substantial deviation, with stepped original CDFs and smoothed, shifted imputed curves. This contrast underscores that coverage alone does not determine imputation quality. Instead, the spatial support network, particularly the quality and density of neighboring stations within the 11 km threshold, plays a critical role. If a low-coverage station is surrounded by well-functioning, high-coverage stations, the imputation can still produce coherent results. In contrast, if the surrounding stations are also sparse, noisy, or dissimilar in behavior, the model has less reliable context to draw from, leading to increased deviation.

While the visual validation provided valuable insight into the performance of the imputation model, highlighting where the imputed distributions align well and where deviations occur, it remains a qualitative and interpretative approach. The level of agreement can be influenced by factors such as sample size, data coverage and may not always capture subtle but statistically meaningful differences.

Moreover, as observed in the METEO stations, even similar coverage levels can yield different outcomes depending on spatial context. To objectively assess the similarity between the original and imputed datasets and to support the visual findings with a quantitative, reproducible metric, further validation is necessary.

Therefore, we proceed to the next step i.e.., statistical validation using the Kolmogorov–Smirnov (KS) test. This allows us to evaluate whether the original and imputed distributions differ significantly, based on p-values, offering a formal measure of confidence in the imputation results across different conditions.

4.7.3. Kolmogorov–Smirnov Test Results

To evaluate the statistical similarity between the original and imputed rainfall data, the Kolmogorov–Smirnov (K–S) test was applied. The analysis was limited to the hourly full-year scenario, and p-values were computed for each imputation (including the average) at every station.

The results are presented separately for ARPA and METEO stations, with each plot summarizing the p-values across all imputations, including the average.

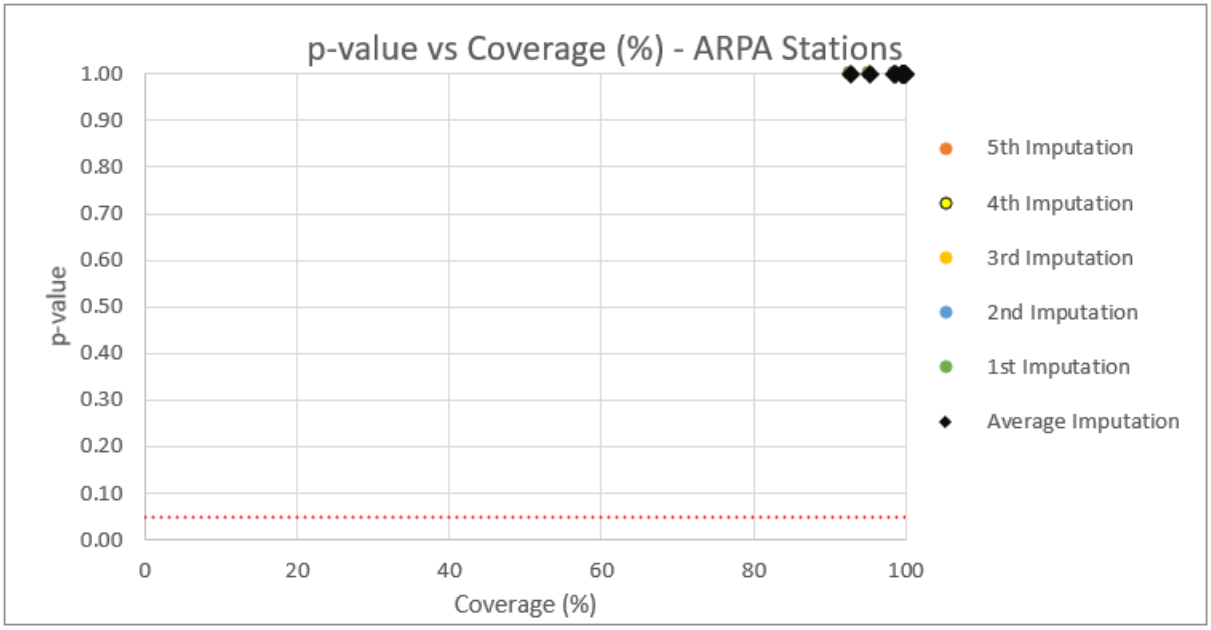


Figure 85. p-value vs Coverage (%) - ARPA Stations – Full Year - Hourly

Figure 85 reinforces the findings established earlier through visual validation by showing that the imputed data for ARPA stations remains statistically consistent with the original data across all scenarios. With coverage levels close to 100%, all p-values, whether from individual imputations or the average, lie well above the conventional significance threshold i.e., 0.05. This confirms that there is no statistically significant difference between the original and imputed distributions, according to the Kolmogorov–Smirnov test. These results align perfectly with the earlier visual interpretation, where ARPA stations exhibited near-perfect overlap in their ECDFs, indicating excellent distributional preservation. The statistical consistency seen here further validates the reliability of the imputed data in these well-observed stations, and the clustering of high p-values suggests the imputation model performs robustly and reproducibly when applied to high-quality data.

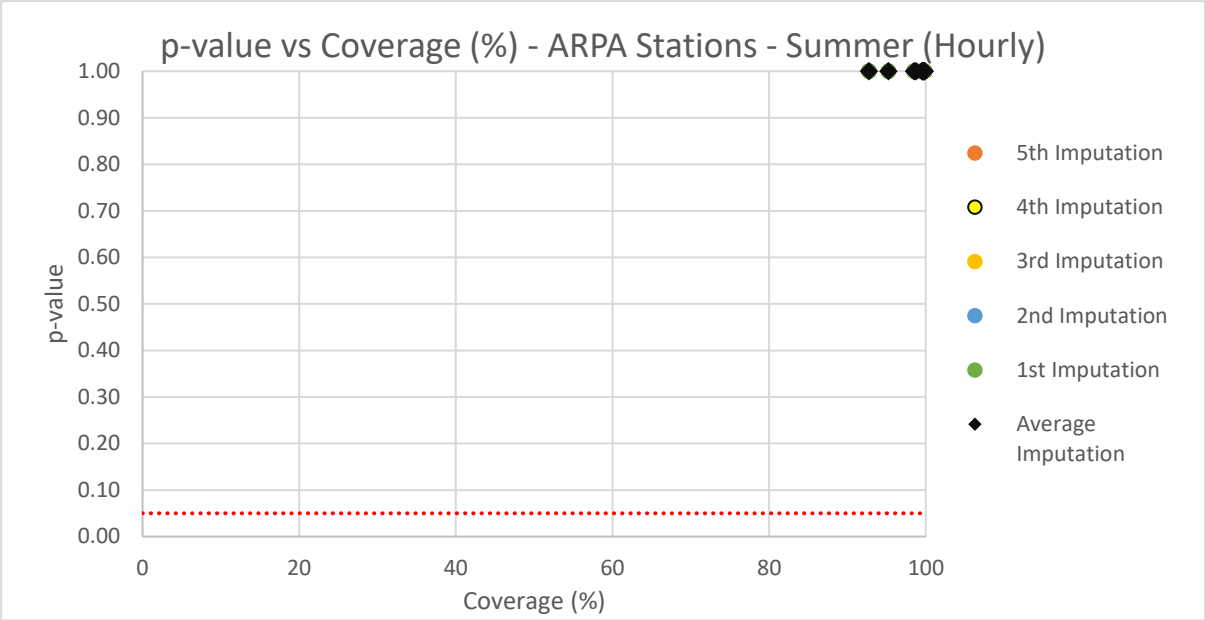


Figure 86. p-value vs Coverage (%) - ARPA Stations – Summer - Hourly

Figure 86 shows p-values against coverage (%) for ARPA stations during summer at hourly resolution. All ARPA stations are clustered around very high coverage (above 90%), and their corresponding p-values, across all five imputations and their average, are consistently well above the 0.05 threshold, often near 1.0. This means the distributions of imputed and original data are almost identical, even during summer, when rainfall is more irregular and difficult to model. The figure essentially confirms that high data coverage ensures reliable imputation, even in challenging conditions. Since ARPA stations have minimal missing data and benefit from consistent observational quality, the imputation process introduces very little distortion, preserving the original hourly rainfall distribution exceptionally well.

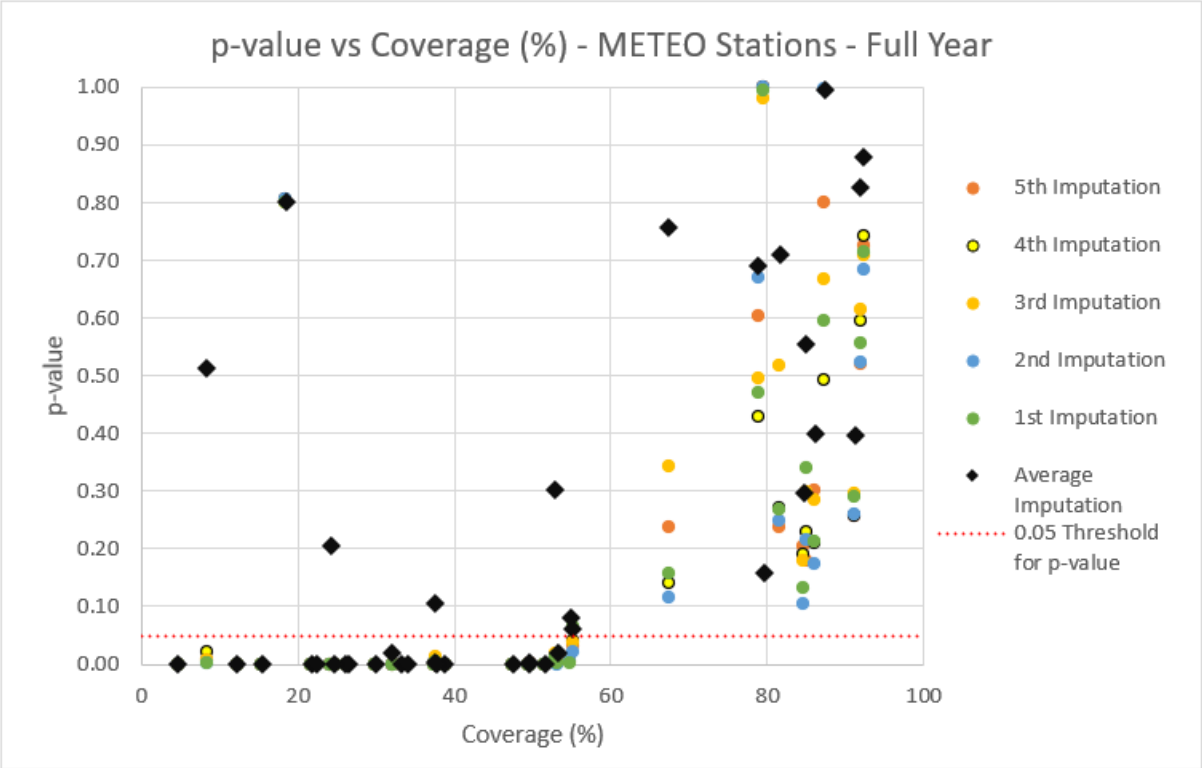


Figure 87. p-value vs Coverage (%) - METEO Stations – Full Year - Hourly

Figure 87, which plots p-values against coverage for METEO stations at hourly resolution across the full year, reveals a clear and interpretable trend. Stations with low data coverage, particularly those below 60%, consistently fail to exceed the 0.05 significance threshold in the Kolmogorov–Smirnov (KS) test, indicating statistically significant differences between the distributions of imputed and observed rainfall. Importantly, these statistical findings strongly support the empirical cumulative distribution function (ECDF) plots discussed earlier, which showed visual discrepancies between curves and a general underrepresentation of rainfall variability. As coverage improves, particularly beyond the 70% threshold, p-values increase markedly, with many stations exceeding 0.5 and some nearing 0.9 or higher, indicating strong statistical alignment with the original data. These high p-values are mirrored by close visual agreement in the ECDFs, where the imputed distributions closely follow the original across the full range of rainfall values. A few outliers with low coverage but high p-values also appear; these likely benefited from strong spatial correlation with well-performing neighboring stations. However, such cases are rare, reinforcing that both high coverage and quality spatial context are crucial for reliable imputation.

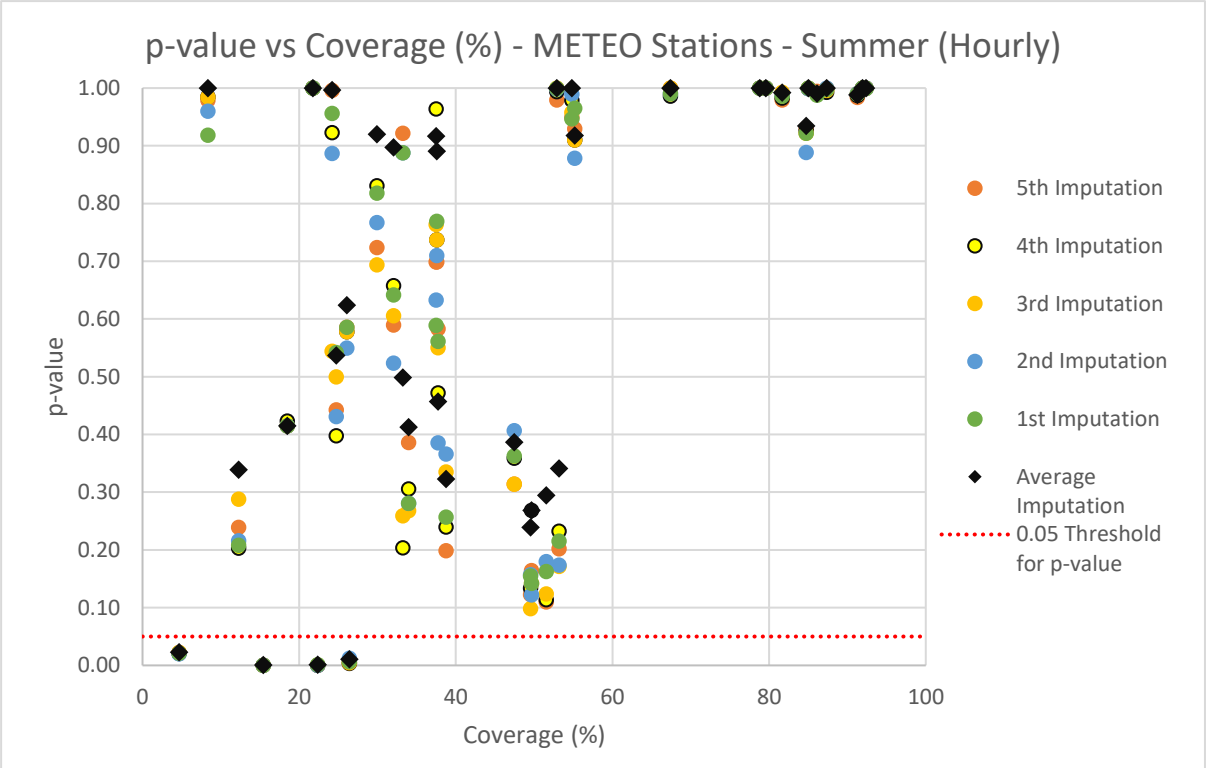


Figure 88. p-value vs Coverage (%) - METEO Stations – Summer – Hourly

Figure 88 focuses on hourly imputation performance during the summer months, when rainfall tends to be more convective, irregular, and spatially localized. This setting poses a greater challenge for imputation models. In this figure, the contrast is striking, METEO stations with coverage above 70% almost uniformly display excellent p-values, many of them clustered tightly near 1.0. These results suggest that where enough data exists, even highly variable summer rainfall can be accurately reconstructed. However, for stations with lower coverage, particularly below 60%, p-values are widely scattered with some of them falling below the 0.05 threshold line. This dispersion reflects the imputation model's difficulty in accurately filling gaps during periods dominated by erratic and short-lived rainfall bursts. Some mid-coverage stations do perform reasonably well, but the variability across imputations increases substantially in this range. Overall, the summer-hourly plot confirms that the imputation framework is highly effective in complex seasonal contexts, but only when data completeness is sufficient.

4.8. Integrated Conclusion on the 75% Coverage Threshold

The decision to include only stations with greater than 75% data coverage in the final application of the Heffernan and Tawn (HT) model (see Chapter 5) was informed by a

consistent body of evidence drawn from both visual and statistical validation techniques, as presented in 4.7.1, 4.7.2, and 4.7.3.

In Section 4.7.1, the comparison of mean and standard deviation between original and imputed datasets revealed that stations with more than 75% data coverage maintained strong agreement in key statistical moments. Below this threshold, particularly under 50%, increasing divergence was observed, especially in standard deviation, suggesting a tendency for the imputation process to underestimate variability and potentially overlook important features such as the intensity and frequency of extreme rainfall events.

Section 4.7.2. complemented these findings through ECDF-based visual comparisons, which showed clear misalignment between original and imputed distributions at stations with low coverage. In contrast, stations with coverage above 75% demonstrated close visual agreement across the full range of rainfall values. These observations were quantitatively supported in Section 4.7.3, where results from the Kolmogorov–Smirnov (KS) test indicated that low-coverage stations frequently produced p-values below the 0.05 significance threshold, reflecting a poor match between imputed and observed data. Conversely, high-coverage stations consistently achieved high p-values, often exceeding 0.5 and, in many cases, reaching or surpassing 0.9, indicating a strong statistical alignment between the imputed and original rainfall distributions.

Section 4.7.3 synthesized these insights to define a practical and conservative threshold for inclusion. A minimum data coverage of 75% emerged as the point above which both the temporal structure (e.g., dry and wet spell dynamics) and distributional features (such as extremes and zero inflation) were reliably preserved. Although a small number of low-coverage stations showed acceptable performance due to strong spatial proximity to high-quality neighbors, these were exceptions. Overall, coverage itself remained the most consistent determinant of imputation reliability.

In conclusion, adopting a 75% coverage threshold represents a data-driven and defensible choice. It ensures that only stations with adequate temporal and spatial integrity are included in the Heffernan and Tawn model, thereby enhancing the robustness of spatial rainfall analysis and the credibility of risk assessment outcomes.

5 Chapter Five: Modelling Spatial Extremes and Dependence Structures

5.1. Introduction

In environmental systems such as rainfall across a river basin, extreme events rarely occur in isolation. Rather, they tend to impact multiple locations simultaneously, especially when driven by large-scale meteorological phenomena like convective systems or frontal passages. In this chapter, we build upon the imputed dataset developed in Chapter IV and proceed to model the spatial structure of extreme rainfall events across the SLO basin. The central aim is to quantify the joint behavior of extremes at multiple locations using a multivariate framework suitable for spatial extremes. Specifically, the Heffernan and Tawn (HT) conditional extremes model is employed due to its flexibility in capturing asymmetric and non-linear dependence structures, particularly in the upper tail of multivariate distributions. This model allows for the estimation of how extreme rainfall at one station influences the distribution of rainfall at other stations, a feature critical to understanding the spatial dynamics of flood-generating events.

Based on the threshold selection and statistical significance testing (using p-values) described in Chapter IV, a subset of 21 stations was selected for spatial extremes modelling. These include both ARPA and METEO stations that initially exhibited data coverage greater than 75%, ensuring reliable imputation quality and meaningful statistical inference. Their spatial distribution, covering key subregions of the Seveso, Lambro, and Olona basins, is illustrated in Figure 89. The map highlights the geographic spread and network structure of the selected stations, which form the analytical backbone of the joint extremes modelling framework employed in this chapter.

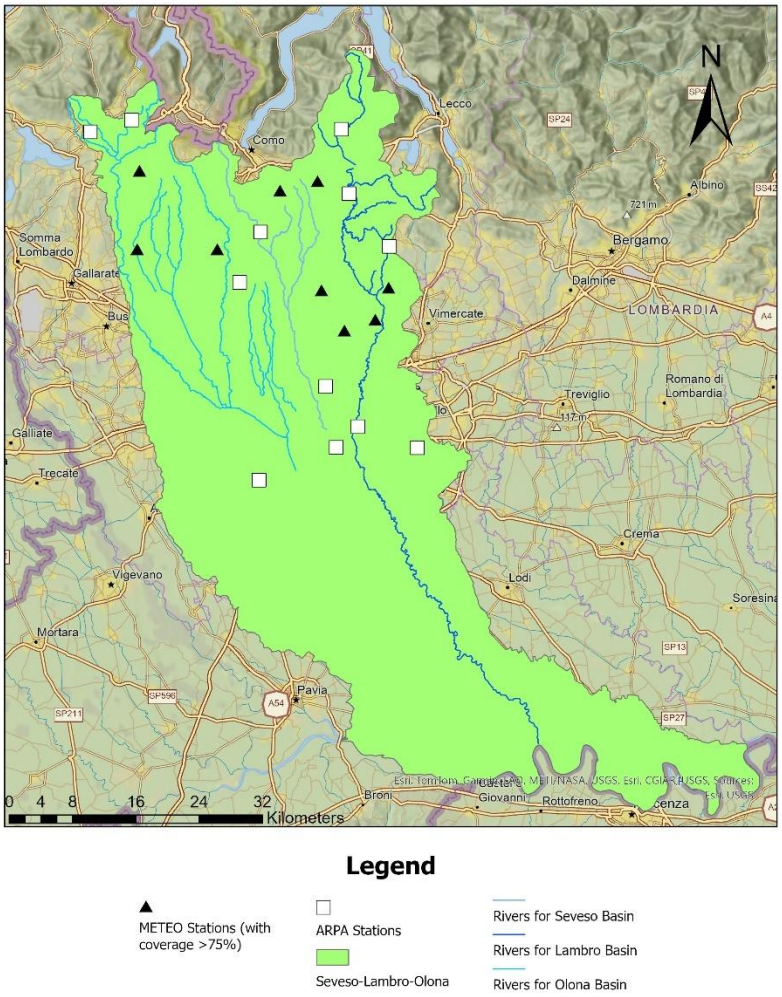


Figure 89. Spatial distribution of the 21 selected rainfall stations across the Seveso-Lambro-Olona (SLO) basin

All statistical modelling and simulation steps in this study were carried out using the "texmex" package in R [57], which is specifically designed for univariate and multivariate extreme value analysis. The package implements the theoretical foundations of extreme value theory (EVT), including marginal modelling using the Generalized Pareto Distribution (GPD), the Heffernan and Tawn conditional extremes model for multivariate dependence, and Monte Carlo simulation for generating synthetic extreme events.

5.2. Theoretical Foundations for Extreme Value Modelling

This section presents the core theoretical concepts underpinning the analysis of extreme rainfall events. The principles of univariate and multivariate extreme value theory (EVT), the Heffernan and Tawn conditional extremes model, and return period

estimation are introduced here to support the methodology and simulations described in later sections.

5.2.1. Univariate Extreme Value Theory

Univariate Extreme Value Theory (EVT) provides the statistical foundation for analyzing rare events in a single variable context, such as extreme rainfall at an individual station. Unlike classical statistical models that focus on average behavior, EVT is designed to model the tail of a distribution, where the most impactful and unusual events reside. In the context of hydrology, this allows for the estimation of return levels, threshold exceedances, and the probability of rare precipitation events.

Among the most widely used EVT approaches is the Peak Over Threshold (POT) method, which focuses on modelling values that exceed a predefined high threshold. For this purpose, the Generalized Pareto Distribution (GPD) is employed, which has been shown to provide a statistically sound model for threshold exceedances under suitable regularity conditions.

Let X be a univariate random variable representing rainfall intensity. For exceedances over a high threshold u, the distribution of excesses $Y = X - u$, conditional on $X > u$, converges to the Generalized Pareto Distribution:

$$P(Y \leq y \mid X > u) = \begin{cases} 1 - \left(1 + \frac{\xi y}{\sigma}\right)^{-1/\xi}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{y}{\sigma}\right), & \xi = 0 \end{cases}$$

Equation 11. Generalized Pareto Distribution (GPD) equation

where $\sigma > 0$ is the scale parameter and ξ is the shape parameter. The scale parameter (σ) controls the spread of the excess values, larger σ means more variability, while the value of ξ determines the tail behavior:

- $\xi > 0$: Heavy-tailed (Fréchet-type)
- $\xi = 0$: Light-tailed (Exponential-type)
- $\xi < 0$: Upper-bounded (Weibull-type)

The application of the Peak Over Threshold (POT) method requires the selection of a threshold to ensure that only the extreme values, those that represent the tail behavior, are modelled. Once a threshold is chosen, the exceedances above this value are assumed to follow a Generalized Pareto Distribution (GPD). The parameters of the GPD are typically estimated using maximum likelihood estimation, and their interpretation provides insight into the tail behavior of the distribution.

The choice of threshold is a critical modelling decision. If the threshold is too low, the GPD assumption may not hold, leading to model bias. If the threshold is too high, too few exceedances may remain for reliable parameter estimation. Therefore, various

diagnostic tools, such as the mean residual life plot and parameter stability plots, are commonly used to guide threshold selection and assess model adequacy.

The result of this univariate modelling step is a set of marginal distributions, one for each station, that accurately describe the behavior of high-intensity rainfall events. These marginals are critical inputs to the multivariate modelling stage, ensuring that each location's extremal behavior is appropriately standardized before assessing dependence across space.

5.2.2. Multivariate Extremes and Dependence Structures

Modelling joint extremes requires a multivariate extension of univariate extreme value theory (EVT), with a focus on the dependence structure between variables in the tails of their distributions.

Multivariate extreme value theory (MEVT) provides the statistical framework to analyze the joint behavior of extremes in multiple variables. At its core, MEVT is concerned with the limiting distribution of component-wise maxima or threshold exceedances in a multivariate context. However, modelling multivariate extremes poses a significant challenge because standard dependence measures, such as correlation, do not accurately describe dependence in the tails. In extreme value theory, tail dependence is the key quantity of interest.

A fundamental concept in multivariate extremes is the distinction between asymptotic dependence and asymptotic independence. Two variables are considered asymptotically dependent if, as the level of one variable becomes extreme, the probability that the other variable also exceeds a similarly high level remains positive in the limit. Conversely, asymptotic independence implies that the joint probability of extreme occurrences decays to zero, even though moderate levels of dependence may still be present. This behaviour is formally described using the tail dependence coefficient, often denoted as χ, which is defined as:

$$\chi = \lim_{u \to 1^-} P\left(F_Y(Y) > u \mid F_X(X) > u \right)$$

Equation 12. Tail dependence coefficient χ

where F_X and F_Y are the marginal distribution functions. A value of $\chi > 0$ indicates asymptotic dependence, while $\chi = 0$ implies asymptotic independence.

This distinction is not merely theoretical; it has practical consequences for risk estimation. Models that assume asymptotic dependence may severely overestimate joint risk when the true dependence structure is asymptotically independent, and vice versa. Unfortunately, many traditional multivariate models, including widely used copula families such as elliptical and Archimedean copulas, are limited in their ability to capture both types of tail behaviour. As a result, there has been increasing interest

in more flexible frameworks that allow for both asymptotic dependence and asymptotic independence, particularly in the context of environmental extremes.

In spatial hydrology, where rainfall fields often show complex, non-linear dependence patterns, the need for such flexibility is critical. The next section introduces one such framework, the Heffernan and Tawn conditional extremes model, which offers a powerful approach for modelling spatial extremes under both dependence regimes.

5.2.3. The Heffernan and Tawn Conditional Extremes Model

The Heffernan and Tawn (2004) conditional extremes model represents a significant advancement in the modelling of multivariate extremes, offering flexibility in capturing both asymptotic dependence and independence. Unlike classical multivariate extreme value models, which focus on joint distributions of maxima and often assume asymptotic dependence, the Heffernan and Tawn (HT) framework adopts a conditional approach. By modelling the behaviour of a random vector given that one of its components is extreme, it enables the representation of a broader range of extremal dependence structures, including scenarios where variables exhibit asymptotic independence.

Before fitting the HT model, all variables must be transformed into a common marginal distribution to ensure consistency and comparability across sites. In this context, the Laplace distribution serves as the standard reference scale. The formulation under Laplace margins and the associated parameter space were developed by Keef et al. in the context of spatial flood risk modeling [23], extending the original conditional extremes framework of Heffernan and Tawn, which was initially formulated under Gumbel margins [7].

To construct the full marginal distribution at each station, a two-part approach is employed. The bulk of the data, the values below a selected high threshold, are modelled using the empirical cumulative distribution function (ECDF), while the exceedances above the threshold are fitted with a Generalized Pareto Distribution (GPD). These two components are combined into a single cumulative distribution function, which is then used to transform the entire marginal distribution to the uniform Laplace scale. This standardization ensures that all stations are evaluated on a consistent scale, allowing the dependence structure to be isolated from marginal effects and satisfying the assumptions of the HT model.

Once the data are standardized, a high threshold is selected for each conditioning variable Y_i, above which extreme behaviour is assumed. The model is then applied by conditioning on the exceedances at a given site and modelling the distribution of the remaining variables. This involves estimating the parameter vectors α_i and β_i, which

respectively control the location and scale of the conditional distribution, along with the distribution of the residuals Z_i, which captures the remaining variability. This procedure is repeated by conditioning on each site in turn, resulting in a set of conditional models that together characterize the spatial dependence structure of extremes across the network.

Let $Y = (Y_1, Y_2, \dots, Y_d)$ denote a d-dimensional random vector with continuous marginals that have been transformed to a common scale, typically Laplace margins. The HT model is constructed by conditioning on one variable Y_i being large and modelling the distribution of the remaining variables Y_{-i} given that $Y_1 > u$, where u is a suitably high threshold. For large values of Y_i, the model approximates the conditional distribution of the remaining variables as:

$$\boldsymbol{Y}_{-i} \mid Y_i = y_i > u \approx \boldsymbol{\alpha_i} y_i + y_i^{\boldsymbol{\beta_i}} \boldsymbol{Z_i}$$

Equation 13. Conditional representation of the Heffernan and Tawn (HT) model

where:

- $\alpha_i \in [-1,1]^{d-1}$ is a vector of location parameters that controls the linear trend or shift induced by the conditioning variable on the remaining components.
- $\beta_i \in [-\infty, 1]^{d-1}$ is a vector of scaling parameters that governs how the residual variability changes as the conditioning variable becomes large.
- Z_i is a residual random vector independent of Y_i.

This formulation allows the model to adapt to both asymptotic dependence and asymptotic independence. When $\alpha_i > 0$ and $\beta_i = 0$, the model reflects asymptotic dependence. When $\alpha_i = 0$ and $\beta_i > 0$, it reflects asymptotic independence. In practice, most environmental phenomena, including rainfall, fall somewhere in between, and the HT model accommodates such mixed behaviour naturally.

An essential feature of the Heffernan and Tawn model is its asymmetry and directional dependence. The conditioning variable Y_i plays a privileged role in the model formulation, which implies that the dependence structure is not assumed to be symmetric across all components. Consequently, the model must be fitted separately for each conditioning station, resulting in a collection of conditional models. Each of these describes how an extreme value at one particular station influences the behaviour of rainfall at all other locations.

5.2.4. Monte Carlo Simulation of Multivariate Extremes

Once a multivariate extremes model has been fitted, a powerful way to assess long-term risk and explore the behaviour of rare events is through Monte Carlo simulation. This technique involves generating large numbers of synthetic realizations from the

fitted model, allowing for probabilistic analysis beyond the limitations of observed data. In the context of spatial rainfall extremes, Monte Carlo simulation enables the creation of synthetic rainfall events at each station that preserve both marginal extreme behaviour and spatial dependence, as learned from the observed dataset.

In the framework of the Heffernan and Tawn model, simulations are performed conditionally. For each event, an extreme value is first sampled at a chosen conditioning site from the Laplace distribution. Given this conditioning value, the corresponding values at all other locations are then simulated using the fitted conditional model. Specifically, the residual vector Z_i is drawn from its estimated distribution, and the conditional structure:

$$Y_{-i} = \alpha_i y_i + y_i^{\beta_i} Z_i$$

Equation 14. Simulation formula under the Heffernan and Tawn model

Equation 14 is used to simulate the remaining components. The process is repeated for multiple conditioning sites and multiple iterations, ensuring that each station contributes to the ensemble of synthetic extreme events.

This simulation strategy allows for the generation of a synthetic catalogue of rare but plausible events, far exceeding the temporal span of the observed record. For example, simulating tens of thousands of multivariate events can correspond to hundreds or thousands of years of synthetic data. This provides a statistically robust basis for estimating the frequency, spatial extent, and intensity of joint extremes, particularly those that lie outside the range of historically observed events.

In particular, the simulation outputs allow for quantile-based analysis across multiple sites, capturing not only the marginal intensity of extremes at individual locations but also their spatial coherence. This is essential for identifying patterns of compound risk, where simultaneous or near-simultaneous high intensities at multiple stations may overwhelm flood control infrastructure or exceed design capacity. Such compound events are often underrepresented in historical data, making synthetic simulation a vital tool for robust design and planning.

Moreover, the ability to simulate thousands of years of plausible extreme scenarios provides a statistical basis for stress-testing hydraulic systems under conditions that are physically realistic but observationally rare. This long-horizon perspective supports decisions related to infrastructure sizing, early warning systems, and climate resilience strategies. The integration of Monte Carlo simulation with a conditional extremes framework like the Heffernan and Tawn model ensures that both marginal behaviour and spatial dependence are preserved, resulting in synthetic rainfall fields that reflect the true joint risk profile of the basin.

5.2.5. Return Levels and Joint Exceedance Risk

An essential application of the synthetic dataset produced using Monte Carlo simulation is the estimation of joint return periods, that quantify the likelihood of simultaneous threshold exceedances across multiple locations. To define meaningful thresholds for these joint assessments, it is necessary to compute univariate return levels at each station by using the Generalized Pareto Distribution (GPD) fitted to threshold exceedances. Given a station-specific exceedance rate λ (i.e., the average number of exceedances per year), the univariate T-year return level is calculated as the rainfall intensity exceeded with probability $\frac{1}{T \cdot \lambda}$ during an individual event.

Mathematically, the return level $RL_{T,j}$ under the GPD is given by:

$$RL_{T,j} = \begin{cases} u_j + \frac{\sigma_j}{\xi_j}\left((\lambda_j T)^{\xi_j} - 1\right), & if\ \xi_j \neq 0 \\ u_j + \sigma_j\, log\left(\lambda_j T\right), & if\ \xi_j = 0 \end{cases}$$

Equation 15. Return level $RL_{T,j}$ for a given return period T

Where,

- u is GPD threshold
- σ, ξ are GPD scale and shape parameters
- λ_j is the station-specific exceedance rate ($\frac{\text{Number of exceedances}}{\text{Years observed}}$)
- T is the target return period in years

These station-specific return levels are then used as thresholds in the simulated multivariate rainfall events to assess how frequently such extremes co-occur. For example, it is possible to estimate the probability that all (or a subset of) stations experience rainfall equal to or exceeding their 100-year return level during the same simulated event.

5.3. Data Preparation

This section outlines the steps taken to prepare the rainfall data for extreme value analysis. Key preprocessing procedures, including time series aggregation, threshold selection, event construction, and multivariate representation, are described. These steps ensure the dataset is suitable for applying the theoretical models discussed in the previous section.

5.3.1. Time Series Aggregation

To prepare the rainfall data for multivariate extreme value analysis, the time series from all stations were aggregated into 3-hour cumulative totals. This step was guided by hydrological considerations based on the concentration time of the catchments

within the study area, which typically falls around 3 hours. This resolution is also commonly used in rainfall extremes studies to reflect short-term accumulation relevant to precipitation events.

The aggregation involved summing rainfall amounts over successive, non-overlapping 3-hour intervals for each station. This temporal scale allowed for effective alignment of rainfall patterns across space, enabling coherent multivariate analysis. Moreover, it reduced the influence of short-term fluctuations and measurement noise, which are less relevant for basin-scale hydrological response.

The resulting dataset preserves the critical features of extreme rainfall events while offering a consistent framework for identifying spatially distributed extremes. It serves as the foundation for the thresholding, declustering and dependence modelling described in subsequent sections.

5.3.2. Data Declustering

With the rainfall series aggregated to 3-hour intervals, the next step was to define extreme events suitable for modeling using the peaks-over-threshold (POT) approach. For each station, a station-specific threshold was computed as the 99th percentile of its aggregated rainfall values. This threshold was used in the declustering process, serving to identify independent exceedance events by ensuring approximate independence and identical distribution (iid) among selected peaks. The 99^{th} percentile is a practical choice in this context, as rainfall time series contain many zero or low values; even at this high quantile, a large number of multivariate events per year is typically retained for meaningful analysis.

Table 6 summarizes the 99th percentile thresholds computed for each of the 21 selected stations based on the 3-hour aggregated rainfall series.

Station ID	Threshold (mm)
ARPA 2006	8.2
APRA 2385	10.2
ARPA 4065	8.6
ARPA 5906	7.6
ARPA 5908	6.4
ARPA 5916	7.2
ARPA 8122	12.2
ARPA 8158	12.2
ARPA 8197	10.6
ARPA 8199	10.328
ARPA 8211	9.8
ARPA 8228	10.928
METEO lmb021	9
METEO lmb084	11.4
METEO lmb128	10
METEO lmb183	7.3
METEO lmb201	8.8
METEO lmb238	7.664
METEO lmb286	7.2
METEO lmb287	8
METEO lmb300	10

Table 6. Station-specific rainfall thresholds

By applying this threshold independently at each site, the spatial heterogeneity of rainfall distributions was preserved, allowing each station to define what constitutes an "extreme" relative to its own historical rainfall distribution.

The selected thresholds are used for the identification of multivariate events in the declustering process. In fact, to prepare the dataset for multivariate extreme value modelling, it was necessary to transform the continuous rainfall time series into a structured set of discrete extreme events. These events represent temporally distinct rainfall events during which at least one station recorded an exceedance above its local 99th percentile threshold.

The construction of these events followed a “sliding window approach combined with a runs-based declustering procedure.” For each threshold exceedance, asymmetric time window of ±3 hours was defined, capturing a total 6-hour period centered on the exceedance. Within each window, the maximum rainfall value was extracted for all stations. This approach ensured that multivariate extremes were captured together in a single multivariate event.

However, rainfall extremes often occur in clusters, particularly during prolonged or frontal weather systems. To ensure temporal dependence between events, a runs procedure was applied. Events whose focal exceedances occurred less than 6 hours apart were grouped together, and only the most intense among overlapping windows was retained. This ensured that the final dataset consists of independent multivariate events, as required by the statistical assumptions of extreme value theory.

The resulting dataset is a matrix of dimensions 756 x 21, where 756 is the number of declustered events and 21 is the number of selected stations. Each row in this matrix corresponds to a spatial snapshot of rainfall maxima across all stations during a distinct extreme event. This matrix forms the core input for both marginal fitting and dependence modelling in the Heffernan and Tawn framework. Table 5.2 displays the first 15 rows for 7 stations of the multivariate event matrix obtained after declustering the rainfall time series. Each row corresponds to a distinct, statistically independent extreme event identified through the declustering procedure, while each column represents selected stations. The values indicate the maximum 3-hour rainfall (in mm) recorded at each station within the corresponding event window.

ARPA_2006	ARPA_2385	ARPA_4065	ARPA_5902	ARPA_5908	ARPA_5916	ARPA_8122
8	0.8	0	0.2	1.36	0.2	0.4
0	0	0	0.2	0	0	0
0	0	0	0	0	0	12.8
16.6	8.8	15	10.8	11.76	8.2	5.8
1.2	0	0.2	7.6	0.28	0	0
10	8.4	9	8.8	5.76	7.8	5.2
7.8	5.2	7	6.6	6.48	6.2	5.4
7	4.8	6.8	9	5	8	3.2
8.6	8.2	8.4	8	6.84	8	6.8
0	3	0	0	0	0	23.4
1.4	8.2	4.2	3	0.84	0	7.8
1	1	0.4	0	0	0.2	1.6
16.4	15.8	15.8	8.6	11.4	16.4	70.4
1.4	1.8	1.2	1.2	1.32	1.2	5.2

Table 7. Sample of the constructed multivariate rainfall event matrix for 7 stations

5.4. Application of Extreme Value Models

This section presents the implementation and outcomes of the modelling approach introduced earlier. It begins with marginal fitting using univariate EVT and proceeds to multivariate dependence estimation and simulation of extreme events.

5.4.1. Marginal Modelling

Each station's rainfall extremes were modelled independently using the Peaks-Over-Threshold (POT) approach, with exceedances fitted to the Generalized Pareto Distribution (GPD). All GPD fits for individual station exceedances were performed using the "migpd()" function from the texmex package, allowing for simultaneous estimation of marginal parameters and standardization to a common Laplace scale. Thresholds corresponding to the 99th, 90th, 85th, and 80th percentiles were explored to assess how the choice of marginal definition affects the modelling process.

For each threshold level, standard diagnostic plots, including threshold stability plots, quantile-quantile plots, and return level plots, were evaluated at all stations. As an illustrative case, diagnostic plots for a "ARPA 8211" station under different thresholds are shown in Figure 90, Figure 91, Figure 92 and Figure 93.

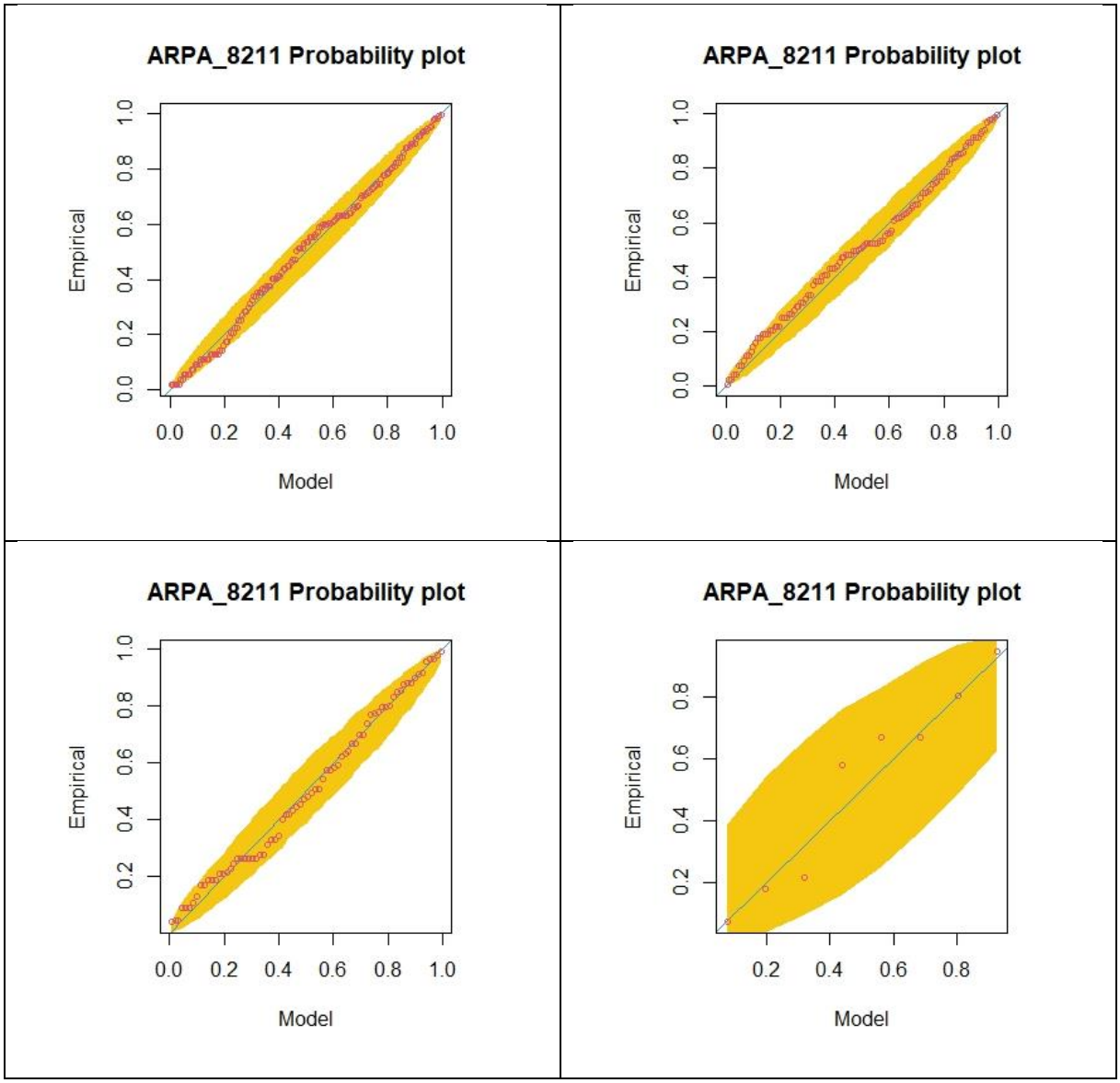


Figure 90. Probability plots for station ARPA_8211 using GPD fits at four marginal thresholds: 0.80 (top left), 0.85 (top right), 0.90 (bottom left), and 0.99 (bottom right)

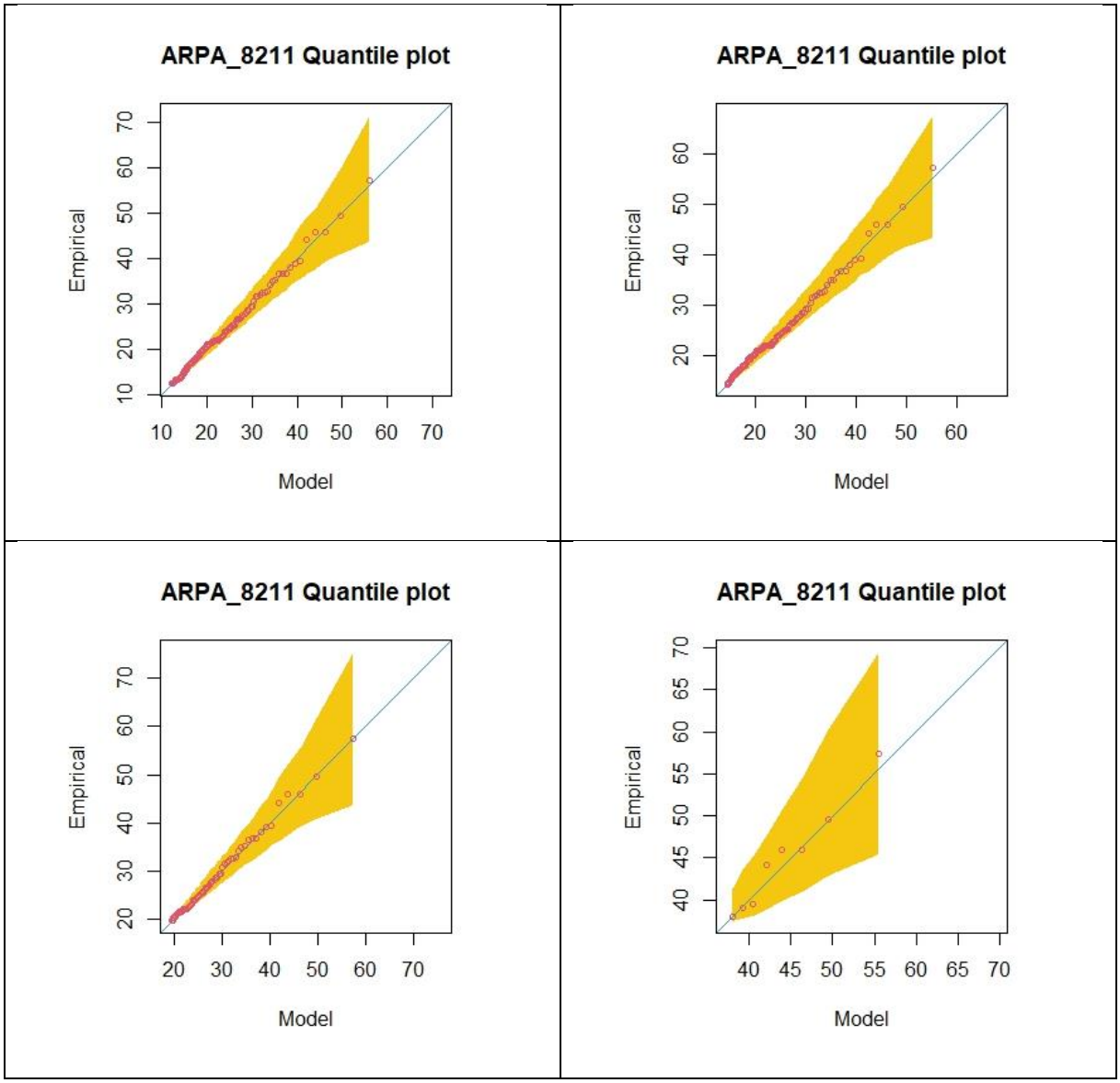


Figure 91. Quantile-Quantile plots for station ARPA 8211 using GPD fits at four marginal thresholds: 0.80 (top left), 0.85 (top right), 0.90 (bottom left), and 0.99 (bottom right)

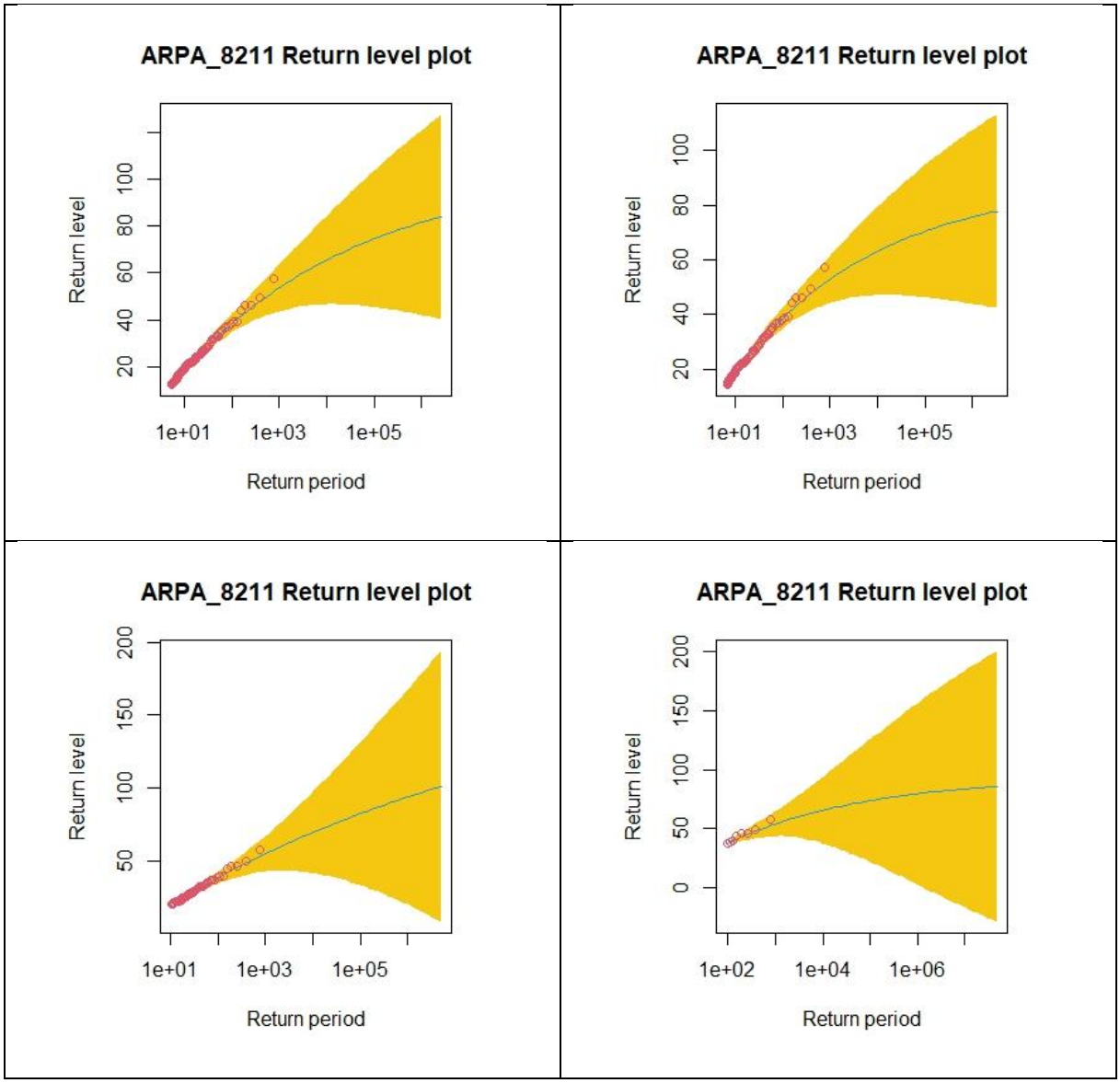


Figure 92. Return Level plots for station ARPA 8211 using GPD fits at four marginal thresholds: 0.80 (top left), 0.85 (top right), 0.90 (bottom left), and 0.99 (bottom right)

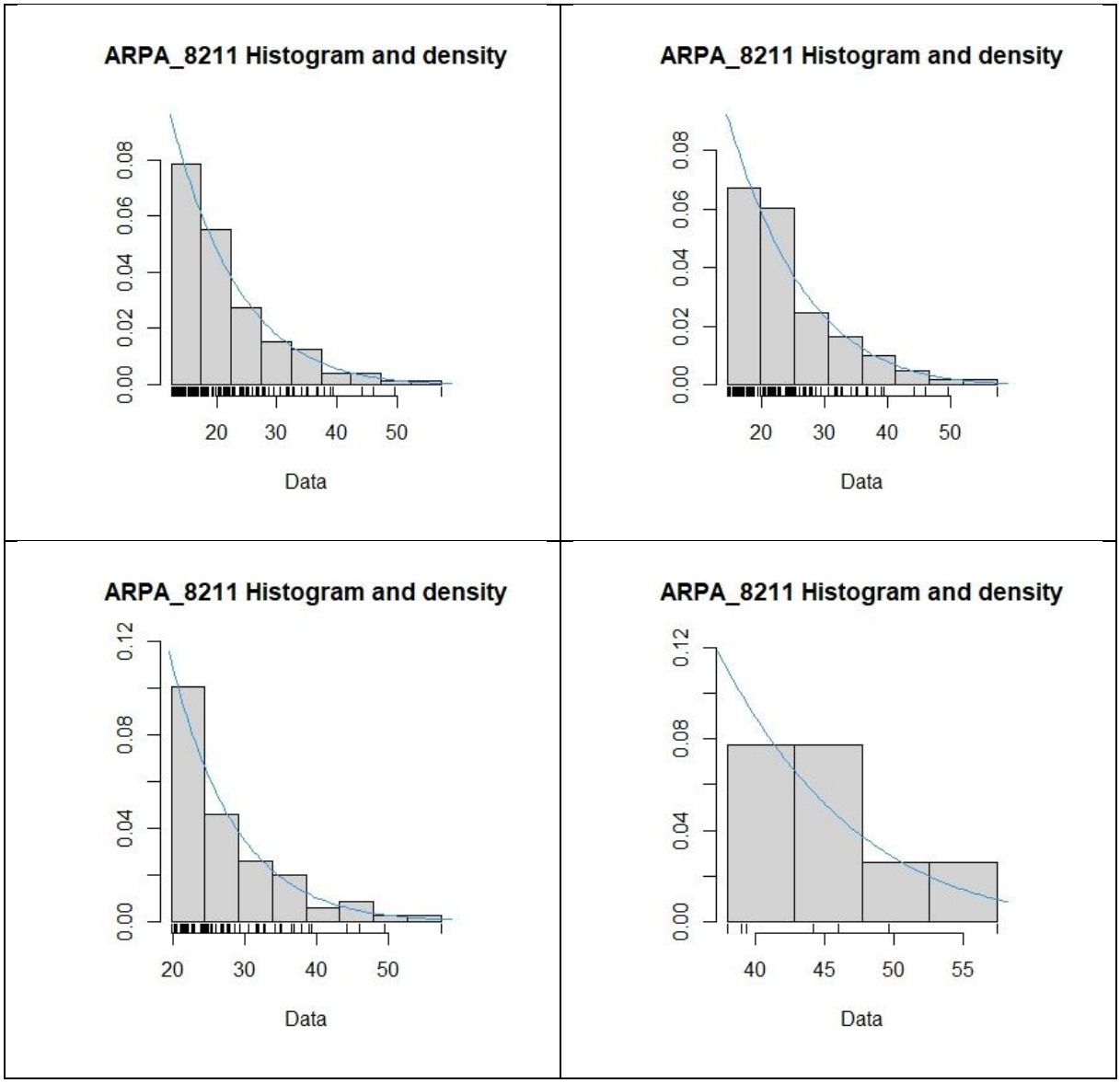


Figure 93. Histogram and Density plots for station ARPA 8211 using GPD fits at four marginal thresholds: 0.80 (top left), 0.85 (top right), 0.90 (bottom left), and 0.99 (bottom right)

The diagnostic plots reveal how the quality of GPD fits evolves across increasing marginal thresholds. At lower thresholds (0.80 and 0.85), the probability and quantile-quantile plots demonstrate tight adherence to the 1:1 line with narrow confidence bands, suggesting reliable model performance and a good volume of exceedances. As the threshold increases to 0.90, this alignment slightly loosens, and by 0.99, the diagnostic plots show visibly wider confidence intervals and greater scatter, reflecting reduced data support and increased uncertainty. Despite these differences, all fits remain within broadly acceptable bounds. However, the final evaluation of threshold suitability is deferred to the simulation stage as it is described in 5.4.3. There, the

impact of marginal choice on the quality and realism of the synthetic rainfall fields will be directly assessed.

5.4.2. Dependence Modelling Using the Heffernan and Tawn Framework

After transforming all marginal exceedances to the Laplace scale, the Heffernan and Tawn (HT) model was applied to quantify the spatial dependence of extreme rainfall across the selected stations.

The model was implemented using the texmex package in R. Each station was treated as a conditioning site in turn, and the conditional model was fit using the "mexDependence()" function. This function estimates the dependence parameters (α, β) and the residual structure for each conditioning site. Fitting was conducted on the 21 × 765 matrix of declustered 3-hour rainfall extremes.

```
ht_dep_struct_2006_08 <- mexDependence(
  x = margins,
  which = "ARPA_2006",
  dqu = 0.8,
  margins = "laplace",
  constrain = TRUE,
  v = 10,
  maxit = 1000000,
  start = c(0.01, 0.01),
  marTransform = "mixture",
  nOptim = 1,
  PlotLikDo = TRUE,
  PlotLikTitle = "Profile Log-Likelihood\n Surface"
)
```

Figure 94. Sample of R code used to fit the Heffernan and Tawn conditional extremes model using the mexDependence() function from the texmex package

In Figure 94, the object "ht_dep_struct_2006_08" is created by applying the "mexDependence()" function to the Laplace-transformed dataset "margins". These margins were generated using a mixture transformation, combining the empirical cumulative distribution function (ECDF) below a high threshold with a Generalized Pareto Distribution (GPD) for exceedances, ensuring compatibility with the conditional extremes model. The threshold quantile "dqu = 0.8" defines exceedances of the conditioning variable (in this case, the ARPA 2006 station), which are used to fit the conditional model. For each conditioning station, the HT model estimates two key parameters: α_i and β_i , which describe the scaling and decay of dependence between the conditioning site and all other sites. In addition, the residual distributions Z_i, representing the remaining variation after accounting for the linear and multiplicative structure, are estimated using likelihood optimization.

The marginal standardization applied prior to modelling is specified through "marTransform = mixture", indicating that each station's marginal distribution was constructed using a mixture of the empirical cumulative distribution function (ECDF) below a threshold and a Generalized Pareto Distribution (GPD) above it. This ensures

the full marginal distribution is appropriately captured and transformed to a Laplace scale, as required by the model.

Likelihood optimization was conducted with constraints , "constrain = TRUE", to promote numerical stability and to ensure estimates remained within feasible bounds. Initial parameter values were set to 0.01 for both α and β, and a maximum of one million iterations, "maxit = 1,000,000", was permitted to aid convergence. The parameter "v = 10" controls the grid resolution for the profile likelihood surface used in diagnostic plotting, which was enabled using "PlotLikDo = TRUE." The function was executed once per conditioning site, "nOptim = 1, and a custom plot title was provided for visual clarity.

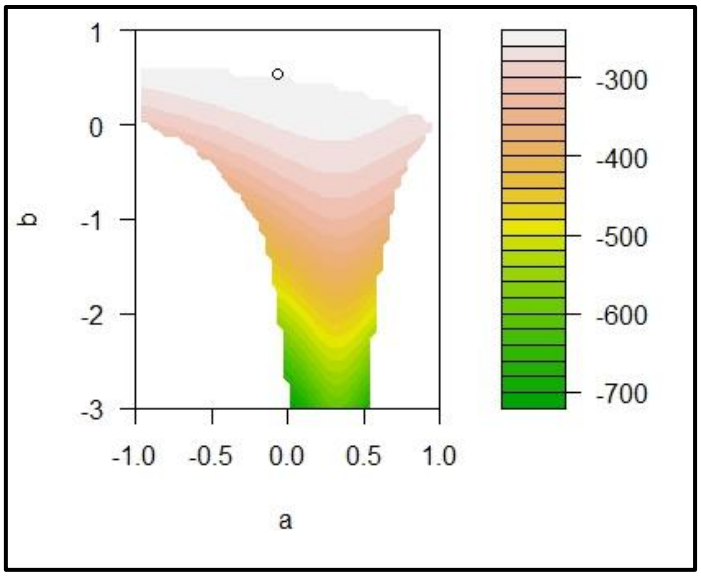


Figure 95. Example of Log-likelihood profile for the fit dependence model at quantile level 0.8, illustrating the relationship between the alpha (a) and beta (b) parameters.

This fitting procedure was repeated across all stations in the study, treating each in turn as the conditioning variable. This produced a collection of conditional models, allowing the full directional dependence structure to be inferred and visualized across the spatial network of stations.

Threshold Sensitivity Testing

A critical modelling choice in the Heffernan and Tawn framework is the selection of a suitable conditioning threshold u, above which the model is fitted. In practical applications, "dqu" in the texmex package, this threshold is expressed as a quantile (e.g., 0.9) rather than a raw value. The threshold defines what is considered "extreme" and directly influences the estimation of the conditional dependence parameters. This threshold determines what is considered "extreme" and directly influences the estimation of the conditional dependence parameters.

To evaluate the robustness of the model to this choice, a systematic sensitivity analysis was conducted. The model was fitted at four conditioning quantiles: 0.6, 0.7, 0.8, and

0.9. For each threshold, the full set of conditional models was estimated across all 21 stations. The resulting parameters, specifically the scaling coefficients α and tail decay coefficients β, were compared across threshold levels to assess their stability and consistency.

Comparative Analysis of Parameters

To examine the sensitivity of parameter estimates to the choice of threshold, each set of $\alpha^{(q)}$ and $\beta^{(q)}$ values obtained at different thresholds was individually compared to the corresponding estimates from a reference threshold. Specifically, separate plots were generated for each threshold, showing:

- $\alpha^{(q)}$ against $\alpha^{(q_{\text{ref}})}$, and
- $\beta^{(q)}$ against $\beta^{(q_{\text{ref}})}$,

where q_{ref} is the baseline threshold under comparison (e.g., 0.6, 0.7, 0.8, or 0.9). In each plot, the identity line was included to assess deviation from perfect agreement.

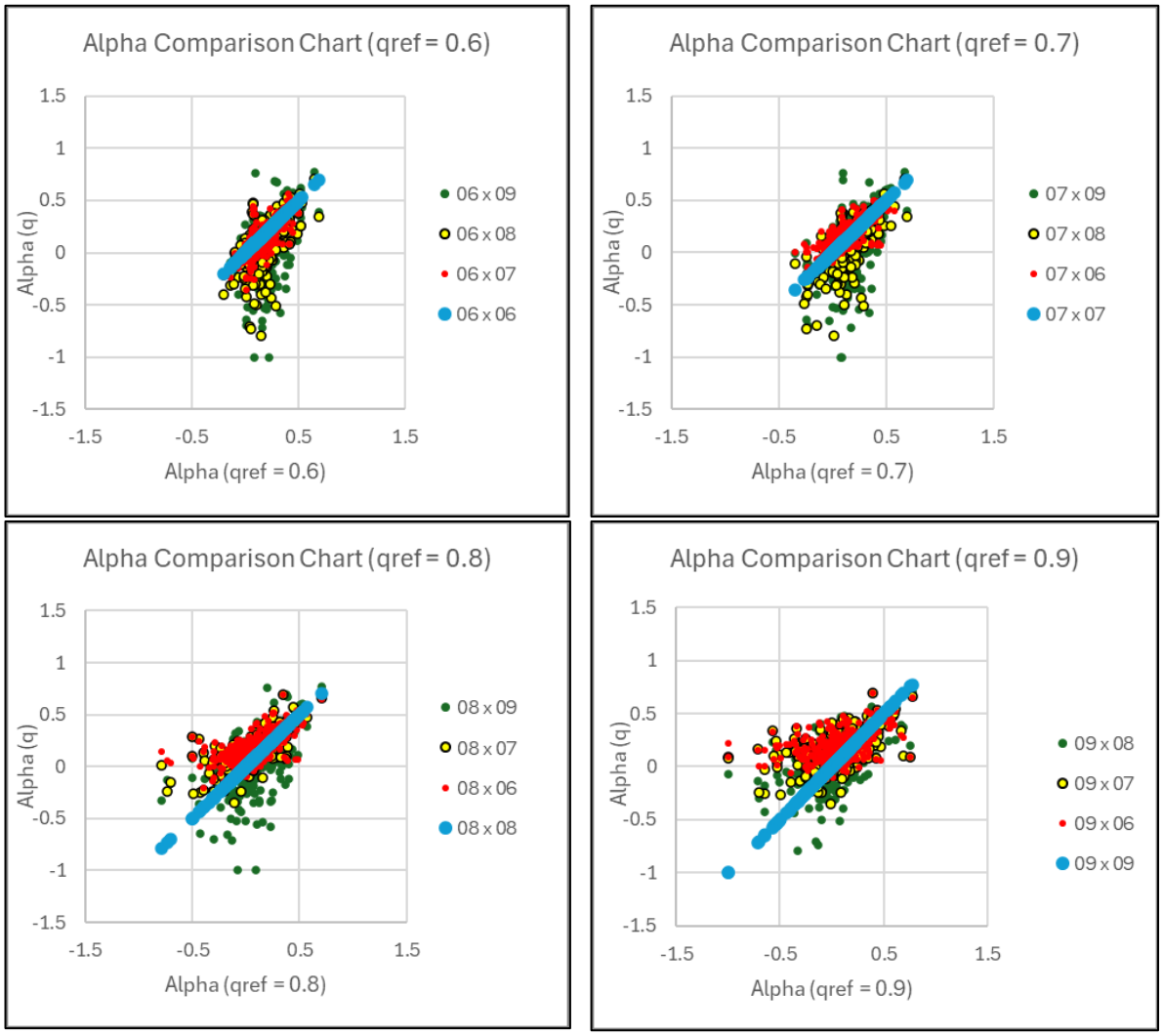


Figure 96. Alpha comparison charts across reference quantiles $q_{ref} \in \{0.6, 0.7, 0.8, 0.9\}$

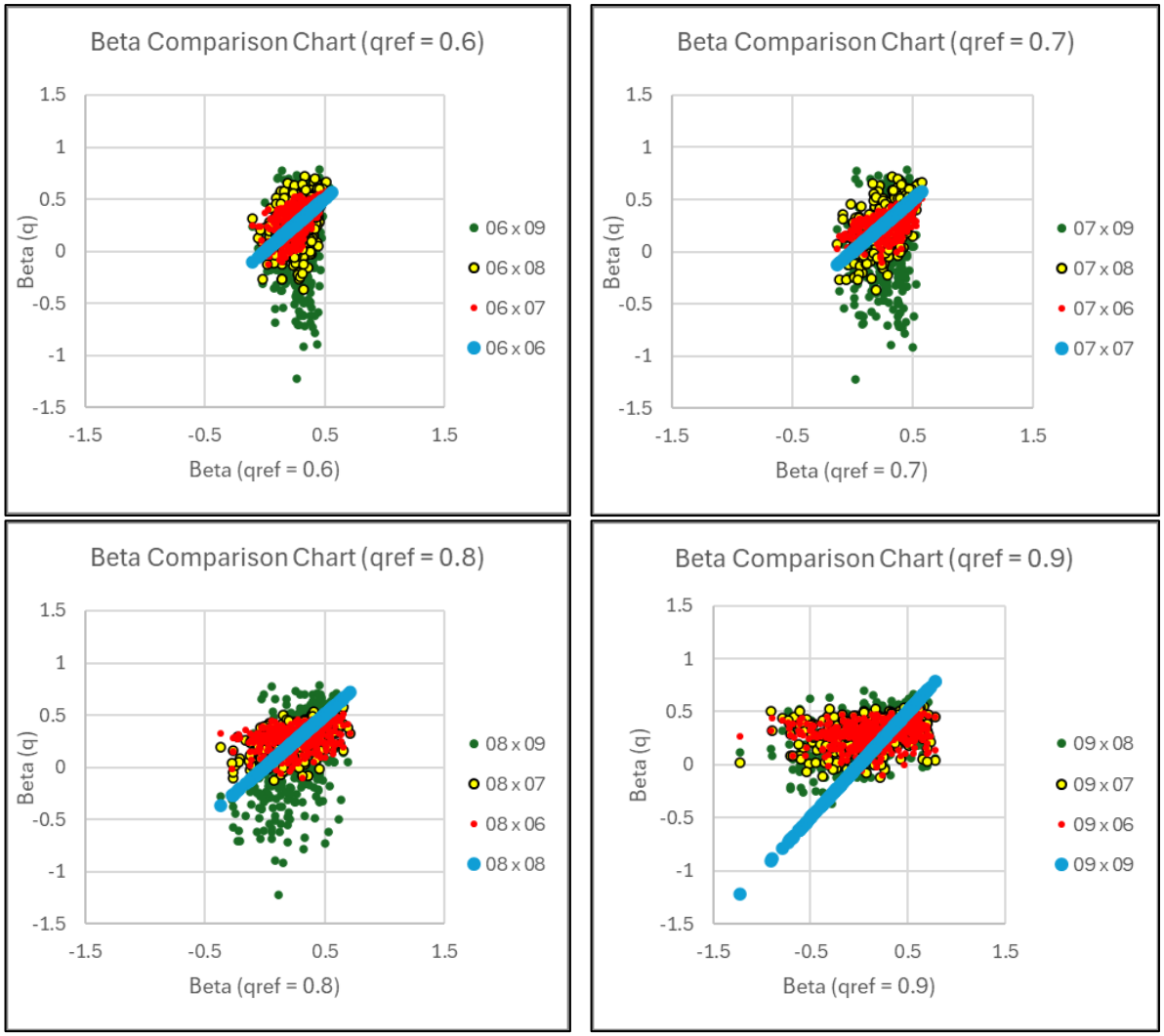


Figure 97. Beta comparison charts across reference quantiles $q_{ref} \in \{0.6, 0.7, 0.8, 0.9\}$

As shown in Figure 96 and Figure 97, estimates from lower thresholds (especially 0.6 and 0.7) exhibit tighter clustering along the identity line, suggesting more stable and consistent parameter estimates. In contrast, estimates from higher thresholds (e.g., 0.8 and 0.9) display greater scatter, indicating increased variability.

This increasing scatter at higher thresholds is likely due to fewer extreme data points available for fitting, which reduces the precision of the parameter estimates. These plots suggest that quantiles around 0.8 strike a balance: they are sufficiently high to target tail behavior, while still retaining enough data to yield stable estimates.

Residual Diagnostic Evaluation

To assess the adequacy of the conditional model fit across varying thresholds, residual diagnostic plots were examined for selected conditioning sites. These included:

- Residual structure plots (residuals Z_i against $F(Y_i)$,
- Normalized residuals against $F(Y_i)$,

- Conditional quantile plots comparing empirical and model-implied behaviour.

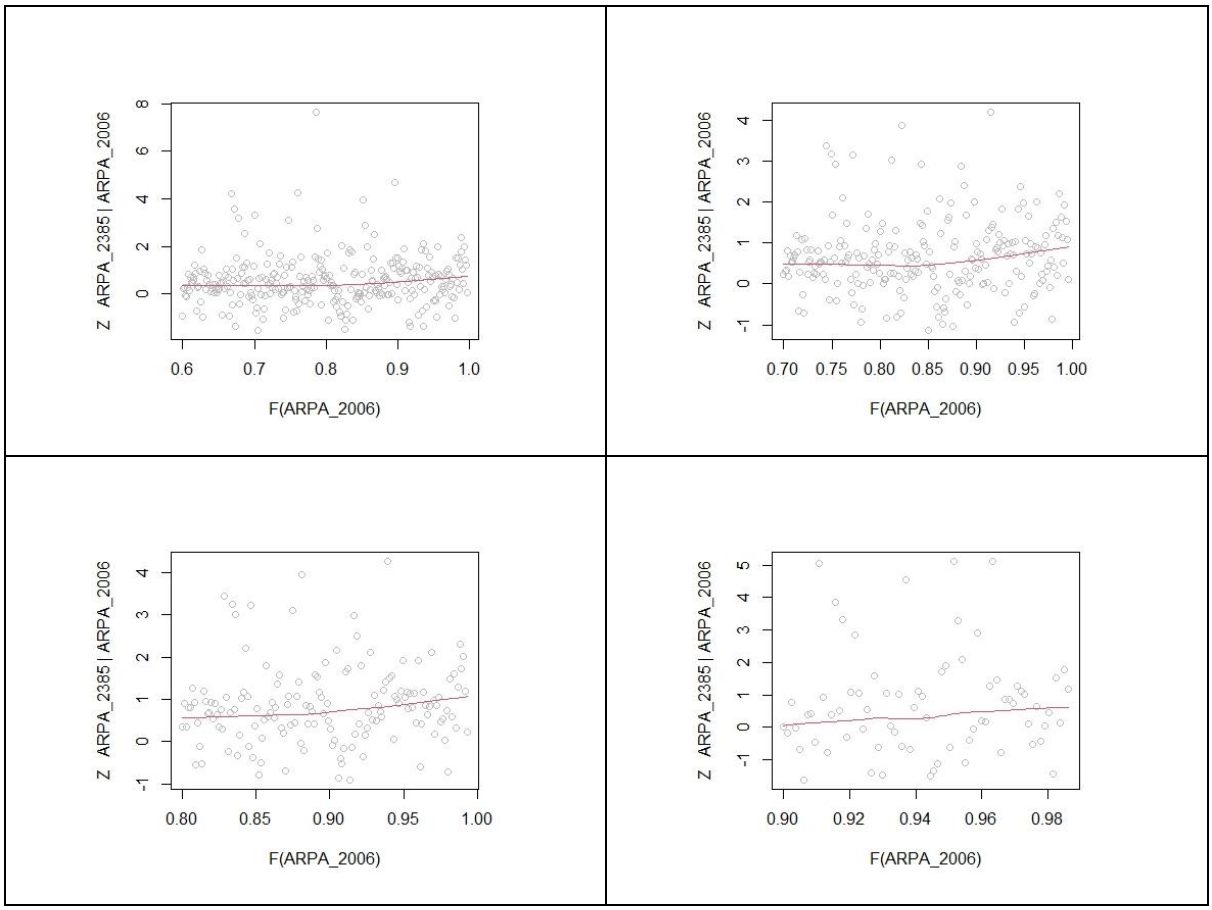


Figure 98. Residual structure plots for ARPA 2006 conditioning on ARPA 2385 at thresholds 0.6 (top left), 0.7 (top right), 0.8 (bottom left), and 0.9 (bottom right)

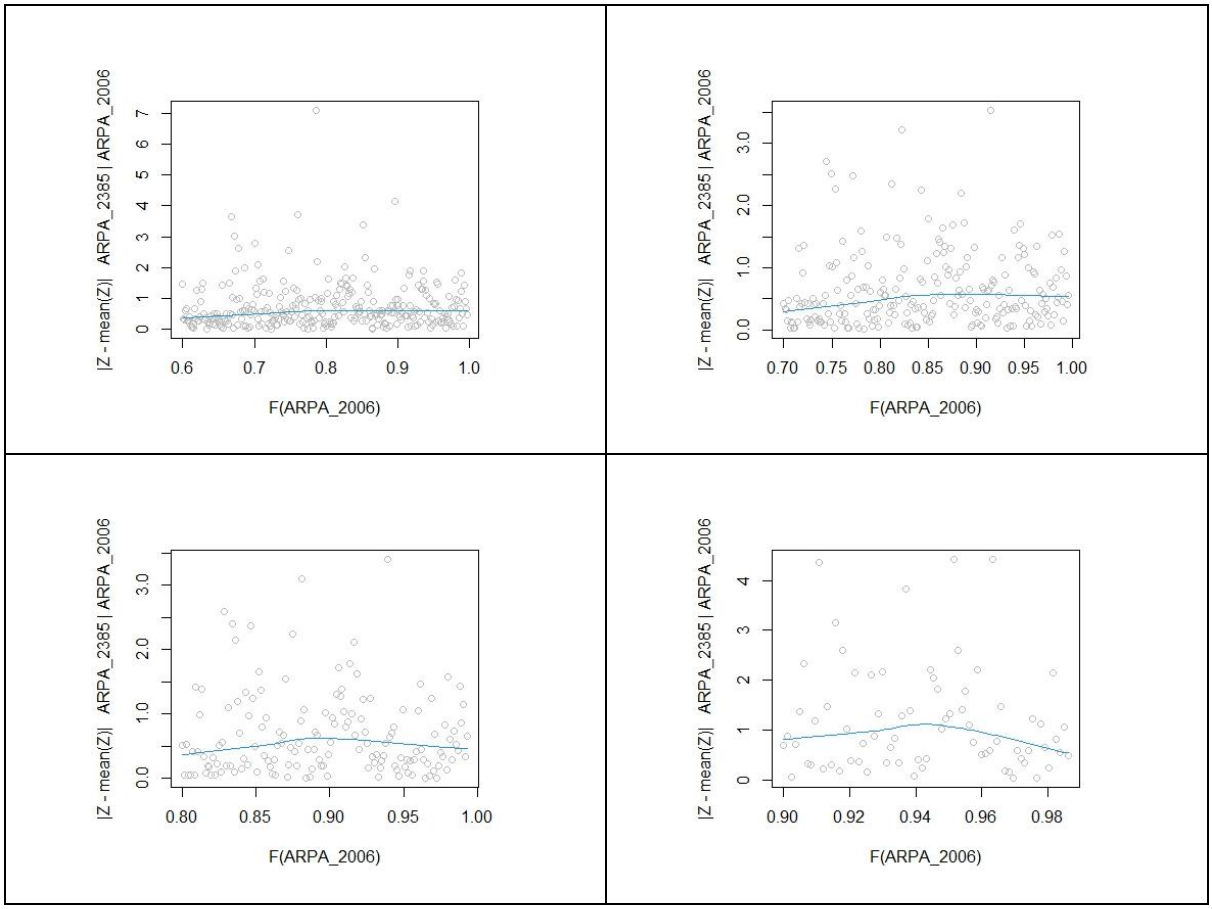


Figure 99. Normalized residual plots for ARPA 2006 conditioning on ARPA 2385 at thresholds 0.6 (top left), 0.7 (top right), 0.8 (bottom left), and 0.9 (bottom right)

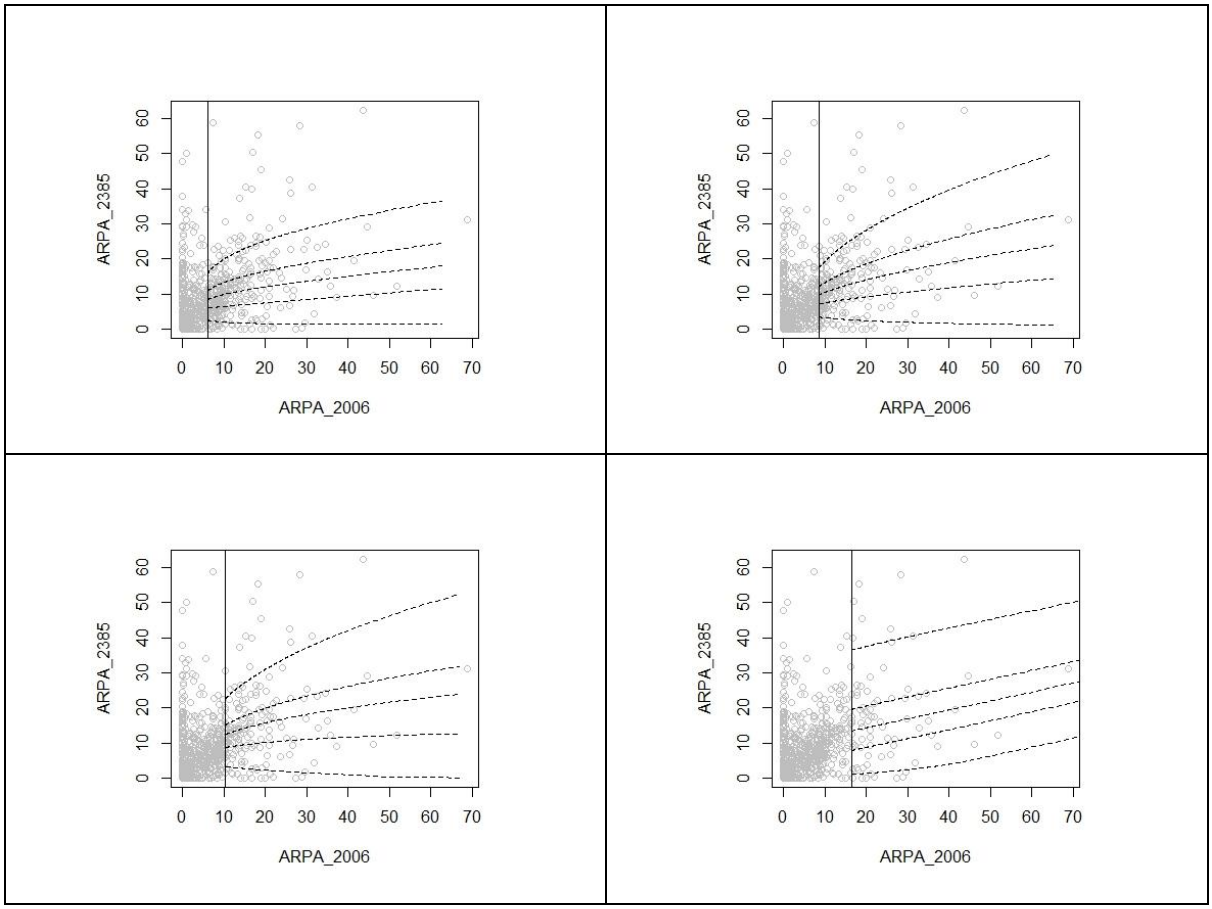


Figure 100. Conditional quantile plots comparing observed vs. model-implied quantiles for ARPA 2385 conditioned on ARPA 2006, across thresholds 0.6 (top left), 0.7 (top right), 0.8 (bottom left), and 0.9 (bottom right)

The diagnostic plots for thresholds 0.6 and 0.7 show no significant structure in the residuals or normalized residuals, indicating the conditional model performs reasonably well in removing dependence. However, the conditional quantile plots at these levels show weaker alignment between observed and modeled quantiles, suggesting the fit is less sharp in capturing the upper tail.

At threshold 0.8, the residuals remain well-behaved, and the quantile plot shows a strong match between observed values and model predictions, with most points lying within the uncertainty bands. In contrast, the 0.9 threshold suffers from increased noise and wider uncertainty due to fewer exceedances.

Final Threshold Selection

Based on combined evidence from the residual diagnostic plots and comparative dependence parameter evaluations, the 80th percentile threshold emerged as a well-supported choice for fitting the conditional Heffernan & Tawn model. The residual diagnostics at this level showed minimal structure and stable variance, indicating strong adherence to model assumptions. The comparative alpha-beta plots demonstrated that dependence estimates at threshold 0.8 exhibited the most favorable

trade-off between stability and extremity. Lower thresholds (e.g., 0.6 and 0.7) yielded more stable estimates due to the larger volume of data, but likely included non-extreme values that diluted the model's tail sensitivity. In contrast, higher thresholds (e.g., 0.9) suffered from greater scatter and instability, reflecting increased estimation uncertainty due to limited data. Collectively, these results support the selection of the 0.8 threshold as a balanced and statistically robust choice for dependence modelling. Based on combined evidence from the residual diagnostic plots and comparative dependence parameter evaluations, the 80th percentile threshold emerged as a well-supported choice for fitting the conditional Heffernan & Tawn model.

5.4.3. Monte Carlo Simulation of Spatial Extremes

Following the successful estimation of the Heffernan and Tawn conditional dependence model, a Monte Carlo simulation framework was employed to generate synthetic realizations of extreme rainfall events. The objective was to explore joint exceedance behavior, quantify spatial risks, and estimate return periods for both univariate and compound extremes across the network of rainfall stations.

Simulation Strategy

To simulate synthetic realizations of extreme rainfall events under the fitted dependence structure, a Monte Carlo approach was employed using the Heffernan and Tawn conditional framework. Once the full set of conditional models was estimated across all stations, these models were used to drive the generation of new multivariate extremes that reflect both the marginal behavior at each location and the spatial dependence among them.

The simulation was conducted using the "mexMonteCarlo()" function, applied directly to the fitted "mexAll" object. The "mexMonteCarlo()" function draws samples under the fitted conditional model, preserving both marginal and spatial extremal behaviour. The "mexAll" object is a structured list that contains the estimated model components necessary for simulation and inference. These include the fitted dependence parameters (α, β), the marginal transformation details, threshold exceedances, and the residuals from the conditional distribution. A total of 69,546 synthetic multivariate events were generated in a single operation. This number was selected based on the fact that the observed dataset spanned 11 years and contained 765 extreme events after declustering. Thus, 69,546 simulated events approximate the number expected from 1,000 years of comparable storm occurrences, assuming a constant rate of occurrence. This provides a sufficiently long synthetic record to support robust estimation of rare-event behavior and joint exceedance probabilities.

Initial simulations were performed using marginal models fitted at the 99th percentile (0.99), paired with a dependence threshold of 0.8. However, upon comparing the empirical cumulative distribution functions (ECDFs) of the simulated and observed

data, the mismatch was immediately apparent. The simulated distributions showed significant divergence with median, mean shifts and distorted variability. Visually, the ECDF plots revealed that the synthetic data consistently over- or underestimated observed extremes, suggesting that marginals at the 0.99 level were too sparse to capture the underlying distribution accurately.

```
############## MONTE CARLO SIMULATION ################

# Load necessary library
library(texmex)

# 1. Fit all conditional models
fit_all <- mexAll(
  station_rainfall_matrix_clean,
  mqu = 0.8,
  dqu = rep(0.8, ncol(station_rainfall_matrix_clean))
)

mc_results <- mexMonteCarlo(nSample=69546,mexList=fit_all)
```

Figure 101. R code used to perform Monte Carlo simulation of conditional extremes

Validation of Simulated Distributions

In light of the results obtained with the first simulation where the marginal threshold was 0.99 quantile, an exploratory simulation was conducted by generating 75 extreme events for four candidate marginal thresholds. This limited set was sufficient to produce empirical CDFs and visually compare the closeness of simulated distributions to the observed rainfall data across stations. This step was necessary due to the high computational cost associated with simulating thousands of years of data for multiple thresholds. Based on this preliminary comparison, thresholds that produced poorly aligned ECDFs were discarded, allowing the subsequent long-run simulation (covering 1,000-year equivalent samples) to focus on the most promising thresholds,

Empirical cumulative distribution functions (ECDFs) were constructed for each station using both the observed and simulated rainfall values. These ECDFs were then plotted side-by-side for direct visual comparison, with particular attention paid to the alignment of the tails and the central mass of the distributions. Each plot included vertical lines marking the median of each distribution to highlight potential shifts.

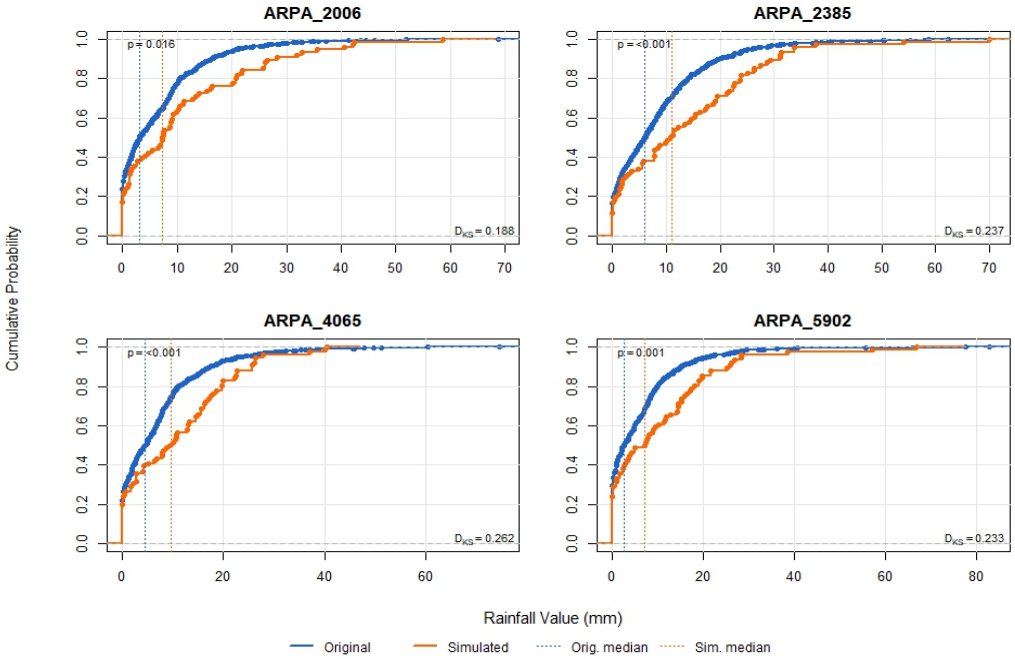


Figure 102. ECDF comparisons between observed and simulated rainfall data for four stations (ARPA 2006, ARPA 2385, ARPA 4065, ARPA 5902) at 0.99 threshold

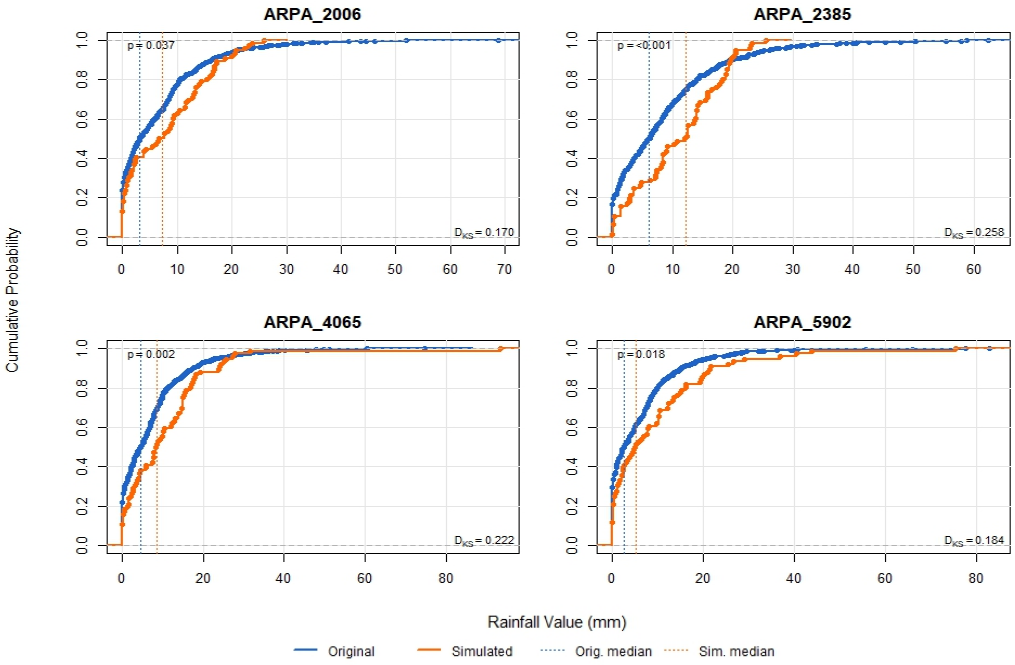


Figure 103. ECDF comparisons between observed and simulated rainfall data for four stations (ARPA 2006, ARPA 2385, ARPA 4065, ARPA 5902) at 0.90 threshold

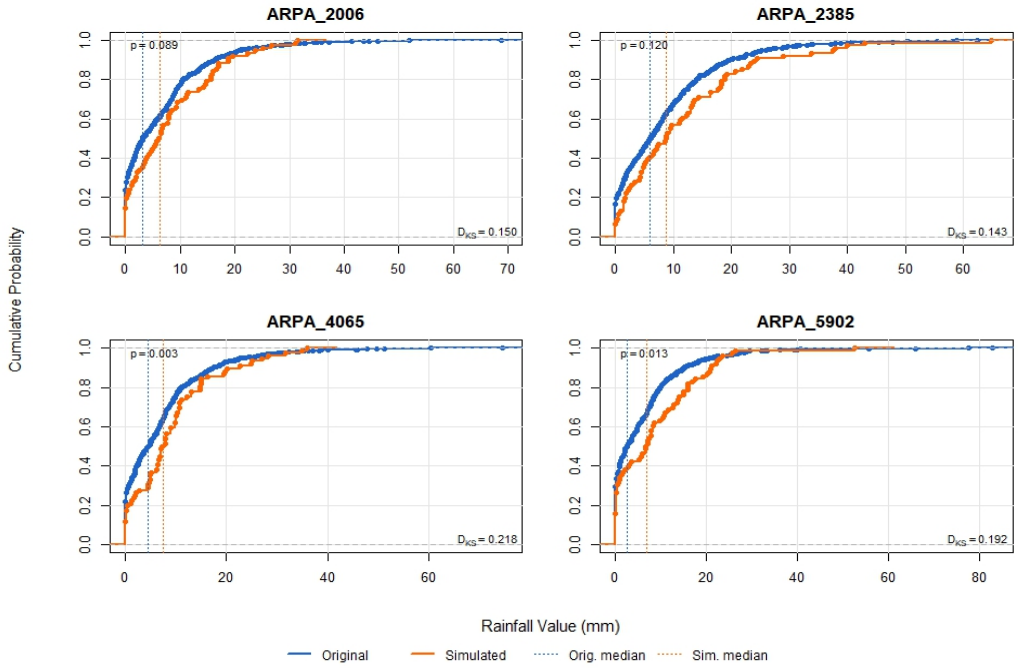


Figure 104. ECDF comparisons between observed and simulated rainfall data for four stations (ARPA 2006, ARPA 2385, ARPA 4065, ARPA 5902) at 0.85 threshold

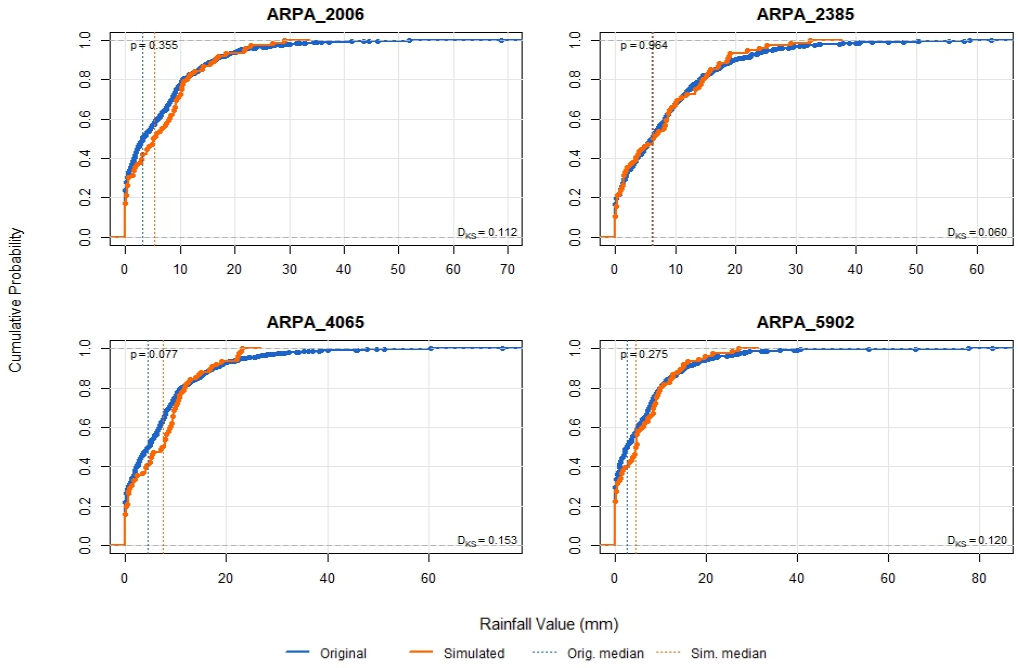


Figure 105. ECDF comparisons between observed and simulated rainfall data for four stations (ARPA 2006, ARPA 2385, ARPA 4065, ARPA 5902) at 0.80 threshold

With each adjustment, a clear trend emerged; the distributional alignment between observed and simulated data improved. At 0.90, while still not ideal, the simulated ECDFs began to resemble their observed counterparts more closely. At 0.85, further

convergence was observed, particularly in the bulk of the distribution, though the tails remained slightly misaligned.

At the 0.80 threshold, a consistent and convincing match was achieved. The ECDFs across stations showed strong agreement between observed and simulated rainfall distributions, with minimal bias in the medians and well-aligned tails. The diagnostic plots indicated that both variability and extremal behavior were being represented with greater accuracy. This aligns with observations made during the marginal modelling stage, where GPD fits at the 0.80 level already demonstrated stable and broadly acceptable behavior across stations. Based on this outcome, the 0.80 threshold was selected for the final simulation phase, wherein a synthetic dataset representing 1,000 years of extreme rainfall was produced using the fitted conditional model. This dataset provided the foundation for all subsequent spatial and hydrological analyses.

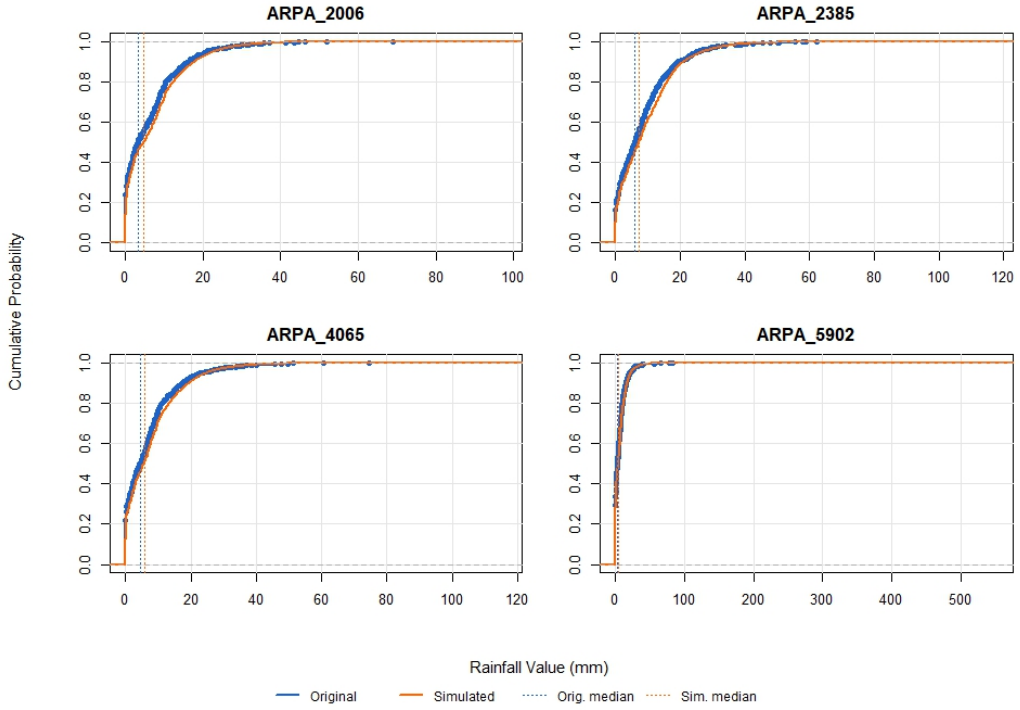


Figure 106. ECDF comparisons between observed and completely simulated rainfall data for four stations (ARPA 2006, ARPA 2385, ARPA 4065, ARPA 5902) at 0.80 threshold

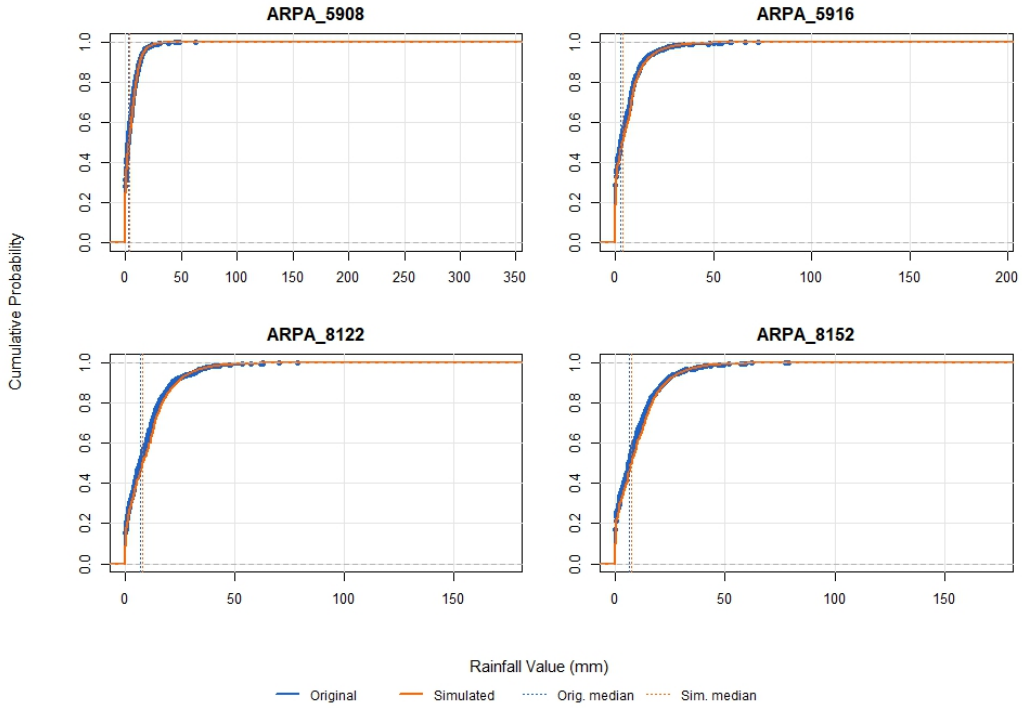


Figure 107. ECDF comparisons between observed and completely simulated rainfall data for four stations (ARPA 5908, ARPA 5916, ARPA 8122, ARPA 8152) at 0.80 threshold

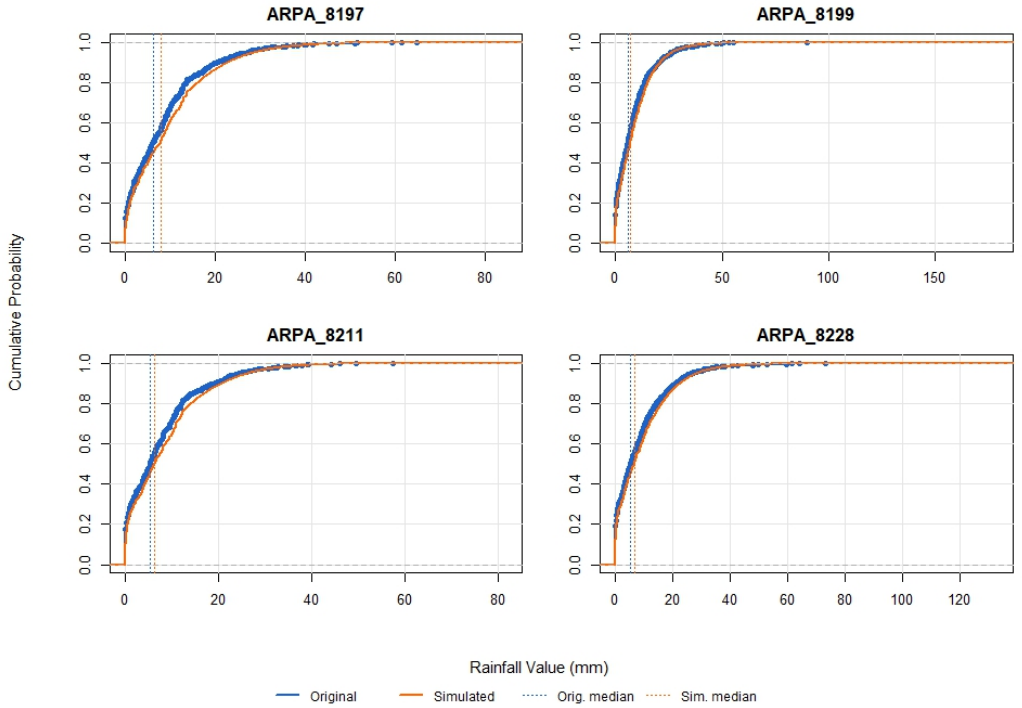


Figure 108. ECDF comparisons between observed and completely simulated rainfall data for four stations (ARPA 8197, ARPA 8199, ARPA 8211, ARPA 8228) at 0.80 threshold

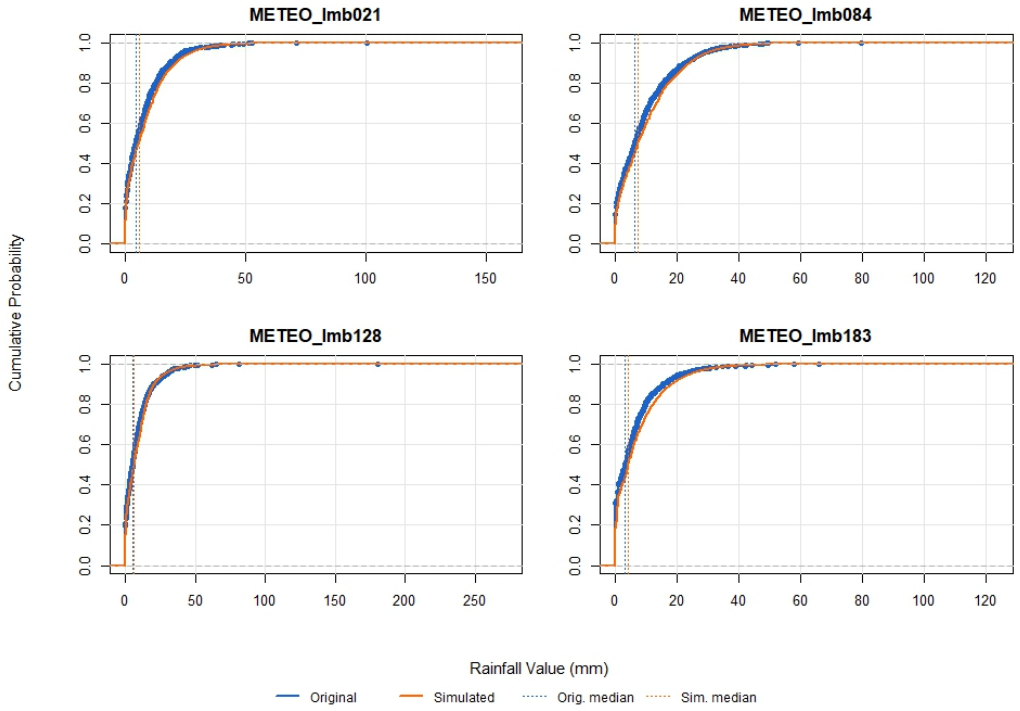


Figure 109. ECDF comparisons between observed and completely simulated rainfall data for four stations (METEO lmb021, METEO lmb084, METEO lmb128 and METEO lmb183) at 0.80 threshold

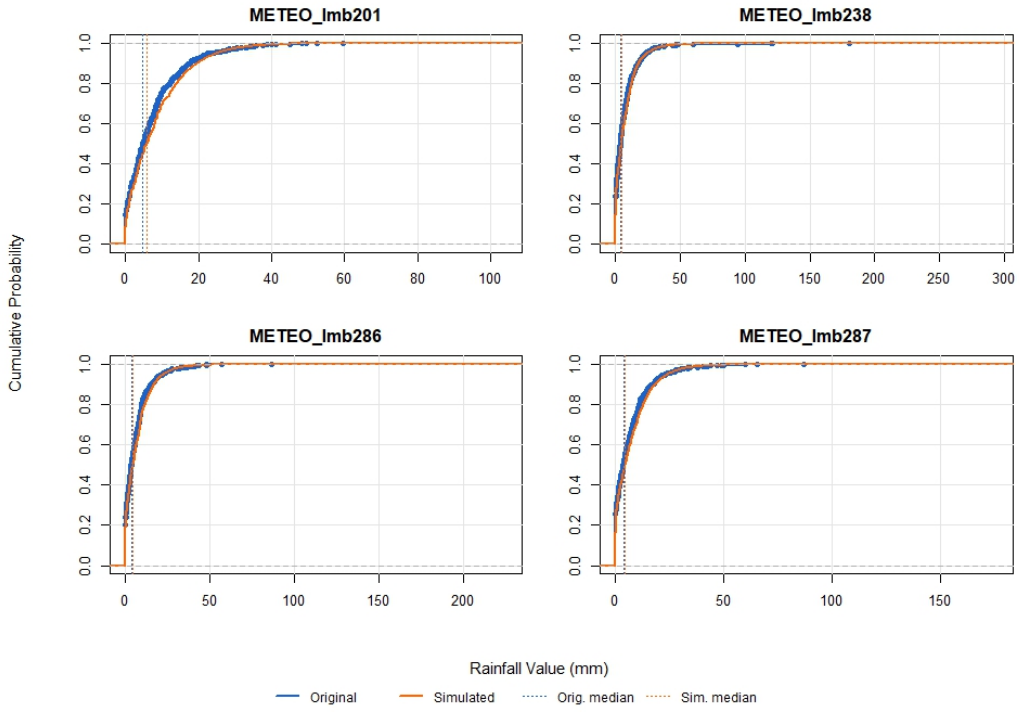


Figure 110. ECDF comparisons between observed and completely simulated rainfall data for four stations (METEO lmb201, METEO lmb238, METEO lmb286 and METEO lmb287) at 0.80 threshold

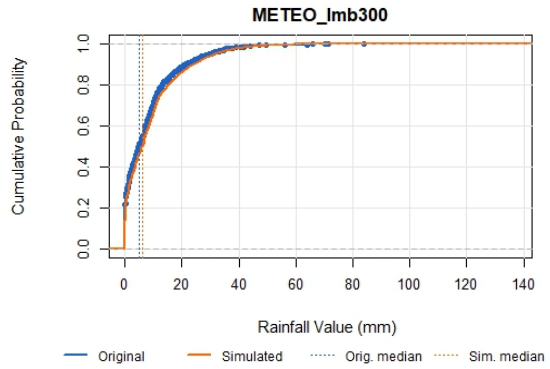


Figure 111. ECDF comparisons between observed and completely simulated rainfall data for stations METEO lmb300 station at 0.80 threshold

Each simulated event represents a spatial configuration of rainfall conditioned on an extreme at one of the stations, as specified by the HT model. The simulation framework probabilistically samples from the conditional structures to replicate realistic tail dependence patterns observed in the data, ensuring that the resulting synthetic events are not merely replicates of past extremes, but extrapolations informed by the underlying statistical structure.

The resulting simulated dataset provides a comprehensive basis for exploring scenarios of compound extremes, estimating marginal and joint return levels, and assessing spatial coherence in extreme rainfall. This level of detail is critical for applications in hydrological risk assessment and infrastructure design, where decisions must often be made based on events more severe than those historically observed.

5.4.4. Return Period Estimation from Simulated Extremes

With 69,546 multivariate extreme rainfall events simulated under the fitted Heffernan and Tawn model, a detailed return level analysis was carried out for both univariate and joint behavior across the rainfall network. The goal was to assess how often events of a given severity are expected to occur individually at each station, and collectively across space.

Methodological Workflow

This section quantifies both univariate and joint return periods using the 1,000-year synthetic rainfall dataset generated through the conditional multivariate Heffernan & Tawn model. The methodology proceeds in two parts:

1. Estimation of station-specific return levels
2. Calculation of univariate and joint return periods from simulated exceedances

Station-specific return levels

To determine rainfall thresholds for return periods T = 5, 10, 50, 100, 200 years, the Generalized Pareto Distribution (GPD) was fitted to exceedances above a fixed threshold at each station. For each return period T, the corresponding rainfall return level RL_T was computed using Equation 15. Return level $RL_{T,j}$ for a given return period T.

These return levels represent the rainfall depth expected to be exceeded, on average, once every T years at a given station.

Exceedance Detection and Return Period Calculation

The return levels were then applied as station-specific thresholds to both the observed and simulated rainfall matrices. Exceedance matrices were generated using:

$$Exceedance_{i,j} = \begin{cases} 1, & \text{if } R_{i,j} > RL_T^{(j)} \\ 0, & otherwise \end{cases}$$

Equation 16. Construction of the exceedance matrix for return period analysis

Where $R_{i,j}$ is the rainfall value at event i and station j, and $RL_T^{(j)}$ is the return level for return period T at station j.

Univariate return periods were estimated using:

$$\mathrm{RP}_j = \frac{1}{\lambda_{\text{events}} \cdot p_j}$$

Equation 17. Univariate return period estimation at station j

Where p_j is the simulated probability of exceeding the threshold at station j, and λ_{events} is the average annual rate of extreme events (estimated as 68.72 events/year from observed data).

Although the return levels RL_T were initially estimated using the theoretical GPD-based formula, we then validated them empirically by applying those return levels as thresholds to the synthetic dataset and computing the actual frequency of exceedance across simulated events. This allowed us to derive an empirical univariate return period from the Month Carlo simulations, which is shown alongside the theoretical one. The empirical return periods tend to be slightly higher due to sampling variability and finite sample size in the simulation.

Stations	Univariate RP = 5 years	Univariate RP = 10 years	Univariate RP = 50 years	Univariate RP = 100 years	Univariate RP = 200 years
ARPA 2006	11.705 years	25.510 years	142.130 years	248.726 years	994.907 years
ARPA 2385	7.262 years	15.090 years	55.273 years	82.909 years	142.130 years
ARPA 4065	9.212 years	16.863 years	142.130 years	198.981 years	331.636 years
ARPA 5902	7.834 years	20.513 years	124.363 years	994.907 years	994.907 years
ARPA 5908	13.629 years	25.510 years	76.531 years	124.363 years	198.981 years
ARPA 5916	10.257 years	20.304 years	165.818 years	994.907 years	Inf years
ARPA 8122	9.212 years	16.863 years	62.182 years	99.491 years	198.981 years
ARPA 8152	8.804 years	17.153 years	71.065 years	124.363 years	198.981 years
ARPA 8197	9.298 years	17.766 years	142.130 years	331.636 years	Inf years
ARPA 8199	9.851 years	19.074 years	106.544 years	123.158 years	331.636 years
ARPA 8211	7.896 years	15.545 years	55.682 years	198.981 years	331.636 years
ARPA 8228	7.370 years	13.211 years	76.531 years	165.818 years	248.727 years
METEO lmb021	8.155 years	16.168 years	90.446 years	198.981 years	497.473 years
METEO lmb084	8.577 years	14.863 years	58.825 years	94.605 years	198.981 years
METEO lmb128	9.128 years	21.168 years	110.545 years	198.981 years	198.981 years
METEO lmb183	7.209 years	15.996 years	71.065 years	102.481 years	198.981 years
METEO lmb201	7.158 years	13.431 years	71.065 years	124.363 years	198.981 years
METEO lmb238	11.179 years	22.959 years	198.981 years	994.907 years	Inf years
METEO lmb286	8.804 years	13.068 years	62.182 years	198.981 years	198.981 years
METEO lmb287	12.456 years	22.612 years	124.363 years	176.916 years	497.173 years
METEO lmb300	17.455 years	36.848 years	198.981 years	994.907 years	198.981 years

Table 8. Univariate return period estimates for 21 rainfall stations across the Seveso-Lambro-Olona (SLO) basin

Table 8 presents the estimated univariate return periods for 21 rainfall stations across the Seveso-Lambro-Olona (SLO) basin, corresponding to theoretical return periods of 5, 10, 50, 100, and 200 years. Although the Monte Carlo simulations are based on the GPD fits, the results do not always align perfectly with the expected theoretical return periods.

For lower return periods, such as RP = 5 years, the simulated values are generally close to the target value; for example, at station ARPA 2385, the simulated return period is 7.26 years, which is reasonably consistent. However, discrepancies become more noticeable at higher return periods. At station ARPA 5916, for instance, the 200-year return period is estimated as infinity, indicating a lack of observed events or poor model support in the extreme tail.

These deviations may arise from limitations in the GPD fitting due to sample size or threshold selection, or they may indicate a need for more Monte Carlo simulations to better capture the tail behavior and stabilize the estimates. In some cases, model refinement or re-evaluation of the threshold for the GPD fit might also be necessary to improve consistency, especially for the estimation of rare events.

Joint return periods for exceedance at n or more stations were computed as:

$$\mathrm{RP}_{\geq n \text{ stations}} = \frac{1}{\lambda_{\text{events}} \cdot P_n}$$

Equation 18. Joint return period for exceedance at n number of stations

Where P_n is the empirical proportion of simulated events where $\geq n$ stations exceeded their return level threshold.

These joint return periods demonstrate how rare concurrent extremes become as both the return level and number of stations increase. Table 9 shows the return periods (in years) for events in which at least 1, 2, 5, 10, 15, or all 21 stations simultaneously exceed their station-specific rainfall return level thresholds corresponding to return periods (RP) of 5, 10, 50, 100, and 200 years.

	RP = 5 years	RP = 10 years	RP = 50 years	RP = 100 years	RP = 200 years
Atleast One Station	0.603 years	1.105 years	5.407 years	11.568 years	22.611 years
Atleast Two Stations	2.900 years	6.257 years	36.848 years	90.446 years	142.129 years
Atleast Five Stations	26.889 years	66.327 years	994.906 years	994.906 years	994.906 years
Atleast Ten Stations	497.453 years	994.906 years	Inf years	Inf years	Inf years
Atleast Fifteen Stations	Inf years	Inf years	Inf years	Inf years	Inf years
All Stations	Inf years	Inf years	Inf years	Inf years	Inf years

Table 9. Estimated joint return periods for multivariate rainfall extremes. "Inf" denotes that no such joint exceedance occurred within the 1,000-year simulation window.

Figure 112 illustrates the relationship between the univariate return period (T) and the return period associated with the combined exceedance at multiple stations, based on

a 1,000-year synthetic catalogue. For cases involving one or two stations, joint return period is often lower than T, since the event is allowed to occur at any station, increasing the likelihood of exceedance and thus reducing the combined return time. However, starting from relatively low return levels (e.g., T = 5 years), joint return period becomes significantly greater than T, particularly when simultaneous exceedances at multiple stations (e.g., 5 or 10) are required. This reflects the increasing rarity of spatially concurrent extremes. At high return periods, the values of joint return periods tend to saturate and appear as discrete plateaus (e.g., 994.906, 497.453), a direct consequence of the finite simulation window: with only 1,000 years of data, the maximum resolvable return period is limited, and events rarer than this upper bound cannot be reliably distinguished.

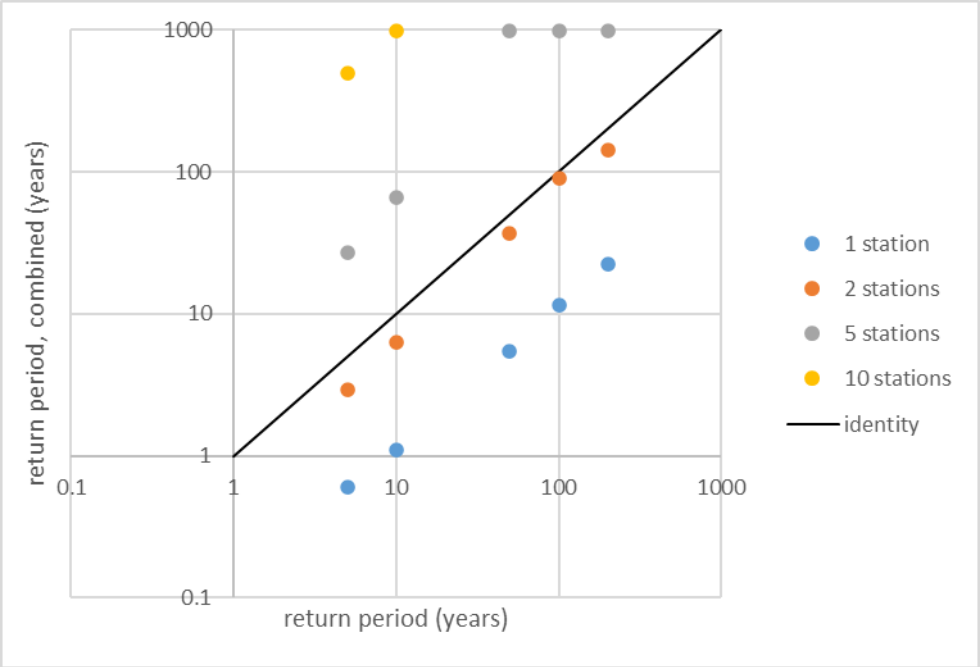


Figure 112. Joint return period analysis showing the maximum number of stations exceeding their respective return levels as a function of return period

In the analysis of joint return periods across the 21 selected rainfall stations, it was observed that no single event within the 1,000-year simulation window resulted in all stations concurrently exceeding their respective rainfall return level thresholds. Specifically, even at relatively modest thresholds such as the 5-year return level, the estimated return period for simultaneous exceedance across all stations remained infinite. To assess whether this outcome was influenced by the spatial extent of the study area, a supplementary analysis was undertaken by repeating the analysis over a smaller subregion.

This involved selecting a geographically concentrated subset of stations to test whether more localized synchrony of extremes could be captured by the same simulation

framework. A circular buffer with a radius of 11 kilometers was drawn around a chosen station located in the bulk of the domain, yielding a cluster of 7 spatially proximate stations. The joint return period analysis was then repeated for this subset using the same methodology and a 5-year rainfall return level as the threshold.

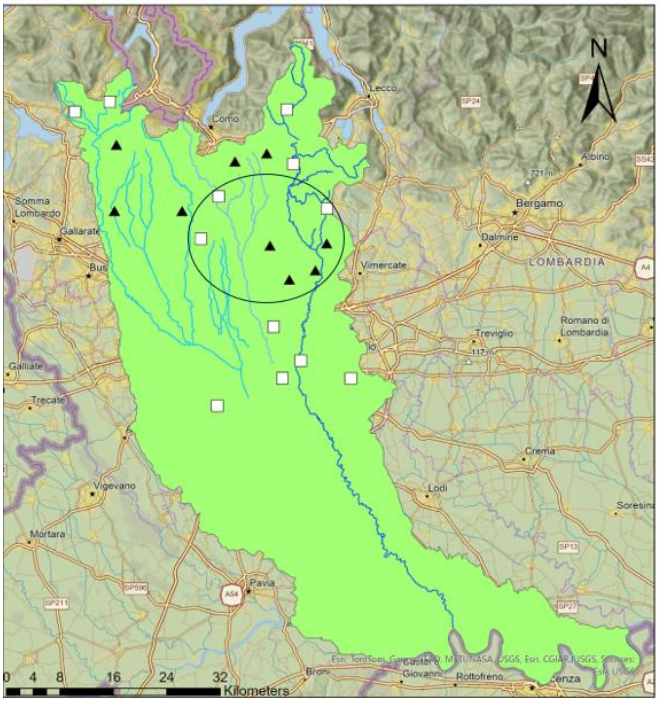


Figure 113. Spatial subset of the Seveso-Lambro-Olona (SLO) basin used for localized joint exceedance validation

The resulting estimates provided insight into the influence of spatial scale on the frequency of joint exceedances. The return period for at least one station exceeding its threshold was approximately 1.09 years, and for two stations it was around 9.66 years. While the return period for exceedances at four stations rose to approximately 110.5 years, no events were recorded in which all 7 stations simultaneously exceeded their return levels, yielding an effectively infinite return period. These results suggest that, even within a relatively small subregion, the spatial co-occurrence of extreme rainfall remains limited, reinforcing the notion that joint extremes become increasingly rare as the number of stations increases, even over compact areas.

Given the spatial spread of the full station network, covering distinct hydrological units across the Seveso, Lambro, and Olona basins, the absence of full-basin joint exceedance is both realistic and expected. It highlights the physical and meteorological heterogeneity of the study area, as well as the importance of sub-basin hydrological independence in driving the spatial behaviour of extreme rainfall events. Therefore, the lack of full-station exceedance within the simulation period is not an anomaly, but a statistically plausible outcome consistent with the basin's spatial dynamics, rainfall

climatology and the broad areal extent covered by the 21 monitored stations across the Seveso, Lambro, and Olona sub-basins.

6 Chapter Six: Conclusions and Future Development

6.1. Summary of Findings

This thesis has presented a comprehensive analysis of rainfall field within the Seveso-Lambro-Olona (SLO) basin, with a focus on spatial dependence, data reconstruction, and the simulation of compound flood-triggering events. The work unfolded across five chapters, each contributing a critical layer of understanding and methodological development.

- **Chapter I** established the motivation, emphasizing the inadequacies of conventional flood hazard mapping in urbanized basins. The need for spatially coherent rainfall analysis was argued through both scientific and policy lenses.
- **Chapter II** contextualized the study area, offering a detailed hydrological, climatic, and urban characterization of the SLO basin. The discussion of past flood events illustrated the urgency of spatially aware risk modelling.
- **Chapter III** focused on spatial correlation among rainfall stations, revealing a clear decay of similarity with distance and identifying consistent patterns across station types. This empirical analysis validated the rationale for spatially informed imputation and dependence modelling.
- **Chapter IV** addressed data gaps, a major obstacle in leveraging citizen science (METEO) stations. A robust imputation strategy was implemented using MICE and a spatially guided predictor matrix. The process enhanced data completeness without compromising marginal or spatial structure.
- **Chapter V** introduced a spatial extremes modelling framework based on the Heffernan and Tawn conditional extremes model. The simulation of synthetic rainfall events using Monte Carlo methods provided a high-resolution catalogue of plausible multivariate extremes, capturing both marginal intensities and spatial co-occurrence. These outputs offer a crucial input for future hydraulic simulations and probabilistic flood risk mapping.

6.2. Contributions to Knowledge

This work offers several contributions to the domain of hydrology and urban flood risk:

1. **Spatial Correlation of Rainfall**
   Rainfall in the basin exhibits significant spatial coherence, which diminishes with increasing inter-station distance. A separation threshold of approximately 11 km was identified as the limit for maintaining moderate correlation (PCC ≈ 0.5, NSE ≈ 0), reinforcing the value of dense monitoring networks in urban hydrology.
2. **Temporal Aggregation Effects**
   Correlation metrics increase substantially when aggregating data from hourly to daily scales, due to temporal smoothing. A smaller but consistent gain is also observed when considering full-year datasets compared to seasonal or wet-period-only analyses.
3. **Validation of Citizen-Science Data**
   Strong similarity in spatial correlation patterns between institutional (ARPA) and citizen-science (METEO) data supports the inclusion of METEO sources in spatial rainfall studies, provided that robust validation and preprocessing steps are applied.
4. **Data Imputation and Quality Control**
   Gaps in METEO datasets were addressed using the Multiple Imputation by Chained Equations (MICE) method, with stations below 75% data completeness excluded to preserve statistical reliability. Good agreement between observed and imputed values was confirmed through ECDF comparisons and Kolmogorov-Smirnov tests, though fidelity decreases with lower initial coverage.
5. **Novel Imputation Framework for Sparse and Zero-Inflated Data**
   Traditional MICE with Predictive Mean Matching was insufficient to reproduce realistic rainfall intermittency, especially at low-coverage stations. This led to the development of a double imputation approach (MICE-2), which reconstructs rainfall occurrence using logistic regression before imputing conditional amounts. This method better preserves the statistical and physical characteristics of hourly rainfall and offers a replicable strategy for handling similar environmental datasets.
6. **Refined Modelling of Rainfall Extremes**
   For marginal fitting within the Heffernan and Tawn (HT) conditional extremes framework, several thresholds (0.80, 0.85, 0.90, 0.99) were tested. While higher thresholds are considered theoretically more appropriate for GPD fitting, they

resulted in poorer agreement between observed and simulated extremes. A threshold of 0.80 offered the best trade-off between tail representation and model stability, yielding spatially realistic synthetic events and better alignment with empirical distributions.

7. **Return Period Inflation due to Spatial Dependence**
   As expected, the recurrence of exceeding a certain T at multiple stations corresponds to much more than T. The joint return periods are multiple of a chosen one, as demonstrated by the fact that the points in Figure 112 were aligned along parallel lines. The coefficient of proportionality is an increasing function of the number of stations at which we require a certain level to be exceeded. This expected result represents the proof of concept that we sought.

6.3. Limitations

While the study achieves its main objectives, some limitations remain:

1. **Computational Constraints on Simulation Scale**
   A major constraint encountered during the modelling phase was the computational intensity of the Monte Carlo simulations required to generate synthetic rainfall events. In theory, simulating a 1000-year catalogue of hourly rainfall would have involved 69,546 extreme events, each conditioned on one of the 21 selected stations. However, attempting to simulate all these events simultaneously proved infeasible, as the process was computationally demanding and time-intensive due to the dimensionality of the spatial dependence structure. The high number of conditioning scenarios and the complexity of fitting the Heffernan and Tawn model across 21 locations significantly slowed down execution, even on multi-core systems. As a result, simulations were limited to a subset of approximately 75 events per threshold trial, roughly equivalent to a one-year synthetic dataset, used to assess model performance. The final model was therefore calibrated using a threshold of 0.8, which showed the best alignment with observed data within this limited simulation window.

2. **Physical Interpretability of Advanced Models**
   As the modelling framework progressed from simple correlation analysis to the more abstract formalism of conditional extremes, the physical interpretability of results diminished. Unlike earlier analyses, the HT model offers no direct validation through observational analogs or physical mechanisms. While this is expected for stochastic models of this nature, it highlights a trade-off between statistical sophistication and interpretability.

6.4. Future Directions for Research

Several avenues for future research emerge:

1. **Coupling Synthetic Rainfall with Hydraulic Models**
   Integrate synthetic rainfall scenarios with hydraulic models to develop a complete flood risk framework. An important extension of this work is to couple the spatially coherent, multivariate rainfall events generated here with distributed hydrological or hydrodynamic models. This would enable the translation of probabilistic rainfall scenarios into flood extents, inundation depths, and damages, creating a comprehensive, end-to-end framework for urban flood risk assessment in the Seveso-Lambro-Olona basin.

2. **Improving Inter-Station Similarity Metric (NSE)**
   One limitation observed in the use of the Nash–Sutcliffe Efficiency (NSE) and its inverse is their directional sensitivity; results can vary depending on which station is treated as the reference. Although computing both directions helps highlight this asymmetry, it complicates the interpretation and aggregation of similarity scores. A potential future improvement could involve testing a symmetric version of NSE, such as averaging NSE and its inverse for each station pair. While it is not yet clear whether this approach would yield more robust or meaningful results, it could help standardize similarity assessment by eliminating reference bias. This direction may be worth exploring in further analyses or publications that rely heavily on inter-station similarity for model construction or spatial imputation.

3. **Basin-Specific Modelling of Spatial Dependence**
   Assess model behavior by applying it separately within each sub-basin using the existing 21-station network. As an extension of this work, future studies could divide the selected 21 rainfall stations according to their location in the Seveso, Lambro, and Olona basins, and apply the conditional extremes model independently within each group. This would allow for a detailed comparison of spatial dependence structures, return levels, and joint exceedance probabilities at the basin scale, potentially uncovering localized risk patterns that may be obscured in a full-network approach. Such basin-specific analysis would also support more targeted flood risk assessments and infrastructure planning.

4. **Threshold Sensitivity in Conditional Extremes Simulation**
   Extend this work by generating 1000 year synthetic rainfall events under multiple threshold levels in the conditional extremes model. Although 0.99 was

found to perform poorly, intermediate thresholds such as 0.85 or 0.9 could not be evaluated at full scale, leaving some uncertainty as to whether a more optimal threshold might have produced better tail behavior or spatial dependence.

Bibliography

[1] S. Vorogushyn, P. D. Bates, K. De Bruijn, A. Castellarin, H. Kreibich, S. Priest, K. Schröter, S. Bagli, G. Blöschl, A. Domeneghetti, B. Gouldby, F. Klijn, R. Lammersen, J. C. Neal, N. Ridder, W. Terink, C. Viavattene, A. Viglione and Zanar, "Evolutionary leap in large-scale flood risk assessment needed," *Wiley Interdisciplinary Reviews: Water,* vol. 5, no. 2, 2018.

[2] J. Neal, C. Keef, P. Bates, K. Beven and D. Leedal, "Probabilistic flood risk mapping including spatial dependence," *Hydrological Processes,* vol. 27, no. 9, p. 1349–1363, 2013.

[3] E. C. "Directive 2007/60/EC on the assessment and management of flood risks," European Commission, 2007.

[4] A. Trigila, C. Iadanza, B. Lastoria, M. Bussettini and A. Barbano, "Dissesto idrogeologico in Italia: pericolosità e indicatori di rischio," ISPRA (Istituto Superiore per la Protezione e la Ricerca Ambientale), 2021.

[5] A. D. Metin, N. Viet Dung, K. Schröter, S. Vorogushyn, B. Guse, H. Kreibich and B. Merz, "The role of spatial dependence for large-scale flood risk estimation," *Natural Hazards and Earth System Sciences,* vol. 20, no. 4, p. 967–979, 2020.

[6] C. Wang, N. B. Chang and G. T. Yeh, "Copula-based flood frequency (COFF) analysis at the confluences of river systems," *Hydrological Processes,* vol. 23, no. 10, p. 1471–1486, 2009.

[7] J. E. Heffernan and J. A. Tawn, "A conditional approach for multivariate extreme values," *Journal of the Royal Statistical Society: Series B (Statistical Methodology),* vol. 66, no. 3, p. 497–546, 2004.

[8] R. Lamb, C. Keef, J. Tawn, S. Laeger, I. Meadowcroft, S. Surendran, P. Dunning and C. Batstone, "A new method to assess the risk of local and widespread flooding on rivers and coasts," *Journal of Flood Risk Management,* vol. 3, no. 4, p. 323–336, 2010.

[9] D. Diederen, Y. Liu, B. Gouldby, F. Diermanse and S. Vorogushyn, "Stochastic generation of spatially coherent river discharge peaks for continental event-based flood risk assessment," *Natural Hazards and Earth System Sciences,* vol. 19, no. 5, p. 1041–1053, 2019.

[10] D. Grebner, "Spatial and temporal patterns of precipitation fields of extreme events in Switzerland and concepts for precipitation scenarios," in *Part of a collected volume: Generation of Hydrometeorological Reference Conditions for the Assessment of Flood Hazard in Large River Basins*, P. Krahe and D. Herpertz, Eds., Koblenz, CHR (International Commission for the Hydrology of the Rhine basin), 2001, p. 21–29.

[11] S. Grimaldi, A. Petroselli, E. Arcangeletti and F. Nardi, "Flood mapping in ungauged basins using fully continuous hydrologic-hydraulic modeling," *Journal of Hydrology,* p. 39–47, 2013.

[12] H. S. Wheater, "Progress in and prospects for fluvial flood modelling," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences,* vol. 360, no. 1796, p. 1409–1431, 2002.

[13] B. Merz and A. H. Thieken, "Flood risk curves and uncertainty bounds," *Natural Hazards and Earth System Sciences,* vol. 9, no. 2, pp. 623-633, 2009.

[14] G. Blöschl, A. Viglione, R. A. P. Perdigão, J. Hall, J. Parajka and B. Merz, "Integrated flood risk management: balancing structural and non-structural measures in the case of Austria," *International Journal of Water Resources Development,* vol. 31, no. 3, p. 325–343, 2015.

[15] M. I. Brunner, R. Furrer and A. C. Favre, "Modeling the spatial dependence of floods using the Fisher copula," *Hydrology and Earth System Sciences,* vol. 23, no. 1, p. 107–124, 2019.

[16] D. Falter, K. Schröter, N. V. Dung, S. Vorogushyn, H. Kreibich, Y. Hundecha, H. Apel and B. Merz, "Spatially coherent flood risk assessment based on long-term continuous simulation with a coupled model chain," *Journal of Hydrology,* vol. 524, p. 182–193, 2015.

[17] K. Schneeberger and T. Steinberger, "Generation of spatially heterogeneous flood events in an Alpine region—Adaptation and application of a multivariate modelling procedure," *Hydrology,* vol. 5, no. 1, p. 1–14, 2018.

[18] N. Quinn, P. D. Bates, J. C. Neal, A. Smith, O. E. J. Wing, C. C. Sampson and J. E. Heffernan, "The spatial dependence of flood hazard and risk in the United States," *Water Resources Research,* vol. 55, no. 3, p. 1890–1911, 2019.

[19] S. Ceschin, "Studio sull'integrazione dei dati della rete MeteoNetwork con i dati ARPAV: valutazione e confronto statistico," 2022.

[20] CINECA, "Studio statistico sull'impatto della rete MeteoNetwork nella stima dei parametri meteorologici al suolo in Emilia-Romagna," 2023.

[21] F. Pilotti, "Confronto tra dati MeteoNetwork e dataset PRISMA per la validazione di eventi idrologici estremi," 2023.

[22] F. Rampinelli, "Using machine learning to reconstruct temperature and precipitation climatologies in an italian citizen science weather station network," 2024.

[23] C. Keef, J. A. Tawn and R. Lamb, "Estimating the probability of widespread flood events," *Environmetrics,* vol. 24, no. 1, p. 13–21, 2013.

[24] L. K. Debusho and T. A. Diriba, "Conditional modelling approach to multivariate extreme value distributions: Application to extreme rainfall events in South Africa," *Environmental and Ecological Statistics,* vol. 28, no. 3, pp. 469-501, 2021.

[25] E. Bevacqua, D. Maraun, I. Hobæk Haff, M. Widmann and M. Vrac, "Multivariate statistical modelling of compound events via pair-copula constructions: Analysis of floods in Ravenna (Italy)," *Hydrology and Earth System Sciences,* vol. 21, no. 6, p. 2701–2723, 2017.

[26] P. Asadi, A. C. Davison and S. Engelke, "Extremes on river networks," *Annals of Applied Statistics,* vol. 9, no. 4, p. 2023–2050, 2015.

[27] H. Yan and H. Moradkhani, "A regional Bayesian hierarchical model for flood frequency analysis," *Stochastic Environmental Research and Risk Assessment,* vol. 29, no. 3, p. 1019–1036, 2015.

[28] G. Ravazzani, A. Amengual, A. Ceppi, V. Homar, R. Romero, G. Lombardi and M. Macini, "Potentialities of ensemble strategies for flood forecasting over the Milano urban area," *Journal of Hydrology,* vol. 539, pp. 237-253, 2016.

[29] W. Köppen, Das geographische System der Klimate, Berlin: Gebrüder Borntraeger, 1936.

[30] "ISPRA," [Online]. Available: https://indicatoriambientali.isprambiente.it/sites/default/files/indicatori_ambientali/2024-05-28/Figura3_PrecipitazioneMedia_LTAA_2023.jpg. [Accessed 08 05 2025].

[31] "Interventi per l'assetto idrogeologico del Torrente Seveso," [Online]. Available: https://www.regione.lombardia.it/wps/portal/istituzionale/HP/DettaglioRedazionale/servizi-e-informazioni/Enti-e-Operatori/territorio/interventi-per-l-assetto-idrogeologico/fiumi-sicuri/interventi-assetto-idrogeologico-torrente-seveso/interventi-assetto-id. [Accessed 07 05 2025].

[32] "Interventi per l'assetto idrogeologico – Fiume Olona," [Online]. Available: https://www.regione.lombardia.it/wps/portal/istituzionale/HP/DettaglioRedazionale/servizi-e-informazioni/Enti-e-Operatori/territorio/interventi-per-l-assetto-idrogeologico/fiumi-sicuri/interventi-assetto-idrogeologico-fiume-olona/interventi-assetto-idroge. [Accessed 07 05 2025].

[33] "Interventi per l'assetto idrogeologico del fiume Lambro," [Online]. Available: https://www.regione.lombardia.it/wps/portal/istituzionale/HP/DettaglioRedazionale/servizi-e-informazioni/Enti-e-Operatori/territorio/interventi-per-l-assetto-idrogeologico/fiumi-sicuri/interventi-assetto-idrogeologico-fiume-lambro/interventi-assetto-idrog. [Accessed 07 05 2025].

[34] A. d. B. d. F. Po, "Linee generali di assetto idrogeologico e quadro degli interventi, Bacino del Lambro," [Online]. Available: http://www.adbpo.it/PAI/3%20-%20Linee%20generali%20di%20assetto%20idraulico%20e%20idrogeologico/3.2%20-%20Elaborato%20Lombardia/Lambro.pdf. [Accessed 07 05 2025].

[35] L. Carrera, A. Petaccia and G. Becciu, "A methodology to assess flood vulnerability in urban areas based on public loss data: The case of the Seveso River, Milan (Italy)," *Natural Hazards and Earth System Sciences*, vol. 22, no. 11, pp. 3543-3561, 2022.

[36] A. d. B. D. d. F. Po, "Relazione Tecnica – Variante alle fasce fluviali del torrente Seveso," Autorità di Bacino Distrettuale del Fiume Po, 2017.

[37] A. d. B. d. Po, "Elaborato Lombardia – Olona," [Online]. Available: https://www.adbpo.it/PAI/3%20-%20Linee%20generali%20di%20assetto%20idraulico%20e%20idrogeologico/3.2%20-%20Elaborato%20Lombardia/Olona.pdf. [Accessed 08 05 2025].

[38] VareseFocus, "Troppe alluvioni, il Varesotto va sempre K.O.," November 2000. [Online]. Available: https://web.archive.org/web/20050328214012/http://www.univa.va.it/varesefocus/VF5/Varesefocus/pag/ter_05_01.htm. [Accessed 08 05 2025].

[39] Varesenews, "Alluvione 2009, dalla Regione tre milioni di euro per Varese"," [Online]. Available: https://www.varesenews.it/2010/09/alluvione-2009-dalla-regione-tre-milioni-di-euro-per-varese/138531/. [Accessed 08 05 2025].

[40] A. d. B. d. f. Po, "Linee generali di assetto idraulico e idrogeologico - Lombardia, Bacino del Lambro," Parma, 2001.

[41] C. M. Lombardo, "Alluvione in Brianza – Novembre 2002," November 2002. [Online]. Available: http://www.centrometeolombardo.com/Files/reportage/OldFoto/Alluvioni/brianza2002/all.brianzanov2002.htm. [Accessed 08 05 2025].

[42] "ARPALOMBARDIA," [Online]. Available: https://www.arpalombardia.it/chi-siamo/cosa-fa-arpa/meteorologia/. [Accessed 06 05 2025].

[43] K. Vohland, A. Land-Zandstra, L. Ceccaroni, R. Lemmens, J. Perelló, M. Ponti, R. Samson and K. Wagenknecht, The Science of Citizen Science, Cham: Springer, 2021.

[44] B. Balázs, P. Mooney, E. Nováková, L. Bastin and J. Jokar Arsanjani, "Data Quality in Citizen Science," in *The Science of Citizen Science* , Cham, Springer, 2021, pp. 139-157.

[45] "METEONETWORK," [Online]. Available: https://www.meteonetwork.it/wp-content/uploads/2023/04/Statuto-aggiornato-al-08-ottobre-2022-Registrato.pdf. [Accessed 06 05 2025].

[46] J. Nash and J. Sutcliffe, "River flow forecasting through conceptual models part I - A discussion of principles," *Journal of Hydrology,* vol. 10, no. 3, pp. 282-290, 1970.

[47] K. Pearson, "Notes on regression and ingeritance in the case of two parents," *Proceedings of the Royal Society of London,* vol. 58, pp. 240-242, 1895.

[48] D. Rubin, Multiple Imputation for Nonresponse in Surveys, New York: Wiley, 1987.

[49] D. Rubin, "Statistical matching using file concatenation with adjusted weights and multiple imputations," *Journal of Business & Economic Statistics,* vol. 14, no. 1, pp. 87-94, 1996.

[50] J. Schafer, Analysis of Incomplete Multivariate Data, Chapman & Hall/CRC, 1997.

[51] S. van Buuren and C. Groothuis-Oudshoorn, "Missing data imputation by multivariate chained equations," *Statistical Methods in Medical Research,* vol. 9, no. 3, pp. 453-464, 2000.

[52] S. van Buuren and C. Groothuis-Oudshoorn, "MICE-Multivariate Imputation by Chained Equations," *Journal of Statistical Software,* vol. 45, no. 3, pp. 1-67, 2011.

[53] S. van Buuren, H. Boschuizen and D. Knook, "Multiple imputation of missing blood pressure covariates in survival analysis," *Statistics in Medicine,* vol. 18, no. 6, pp. 681-694, 1999.

[54] T. Raghunathan, J. Lepkowski, J. Van Hoewyk and P. Solenberger, "A multivariate technique for multiply imputing missing values using a sequence of regression models," *Survey Methodology,* vol. 27, no. 1, pp. 85-96, 2001.

[55] A. Kennickell, "Imputation of the 1989 Survey of Consumer Finances: Stochastic relaxation and multiple imputation," in *Proceedings of the Section on Survey Research Methods*, 1991.

[56] R. Little, "Missing-data adjustments in large surveys," *Journal of Business & Economic Statistics,* vol. 6, no. 3, pp. 287-296, 1988.

[57] H. Southworth, J. E. Heffernan, P. D. Metcalfe, Y. Papastathopoulos, A. Stephenson and S. Coles, *texmex: Statistical Modelling of Extreme Values,* CRAN (Comprehensive R Archive Network), 2024.

List of Abbreviations

ARPA – Agenzia Regionale per la Protezione Ambientale
COFF – Copula-based Flood Frequency
CSNO – Canale Scolmatore di Nord-Ovest
Cfa – Humid Subtropical Climate (Köppen climate classification)
DAD – Depth-Area-Duration
ECDF – Empirical Cumulative Distribution Function
EVA – Extreme Value Analysis
EVT – Extreme Value Theory
FCS – Fully Conditional Specification
GPD – Generalized Pareto Distribution
GPRS – General Packet Radio Service
HT – Heffernan and Tawn (Conditional Extremes Model)
IDW – Inverse Distance Weighting
JM – Joint Modelling
KS – Kolmogorov–Smirnov (Test)
LOGREG – Logistic Regression
MB – Province of Monza e della Brianza (Italy)
MCMC – Markov Chain Monte Carlo
MEVT – Multivariate Extreme Value Theory
MI – Multiple Imputation
MICE – Multivariate Imputation by Chained Equations
MNW – MeteoNetwork
NA – Not Available / Missing Data
NSE – Nash–Sutcliffe Efficiency
PCC – Pearson Correlation Coefficient
PMM – Predictive Mean Matching
POT – Peak Over Threshold
RP – Return Period
RL – Return Level
SLO – Seveso-Lambro-Olona (basin)
UTM – Universal Transverse Mercator (coordinate system)

Acknowledgments

First and foremost, I am deeply grateful to Almighty Allah for granting me the strength, patience, and perseverance to complete this thesis.

I would like to express my heartfelt gratitude to my thesis supervisor, Prof. Alessio Radice, for his invaluable guidance, continuous support, and inspiring mentorship throughout this research. I am also sincerely thankful to my co-supervisor, Ana Maria Rotaru, for her constructive feedback. encouragement and particularly for her significant assistance with the R software and data analysis aspects of this work, her contributions have been instrumental to its completion.

I owe my deepest gratitude to my beloved parents, Kanwal and Hafeezullah, for their unwavering love, prayers, and sacrifices that have been the foundation of all my achievements. Your faith in me has been my greatest motivation.

To my siblings, Hafsa, Wamiq, Ovais, and Samee, thank you for your constant support, encouragement, and belief in my abilities.

I am also fortunate to have incredible friends who have been an integral part of my life and academic journey. From Pakistan: Hamza, Saif, Uzair, Ehsan, Sawaiz, and Aizaz, your companionship, laughter, and motivation kept me grounded and focused. From Italy: Francesca, Saba, and Mohammadreza, thank you for your kindness, friendship, and for making my time abroad so enriching and memorable.

Each one of you has played a vital role in this journey, and I will always remain grateful for your presence and support.