UNIVERSITÀ DEGLI STUDI DI PADOVA

SCUOLA DI SCIENZE

Dipartimento di Geoscienze

Direttore Prof. Nicola Surian

TESI DI LAUREA MAGISTRALE IN

Earth Dynamics

The trade-off between spatial variability and estimation uncertainty of extreme precipitation at multiple scales: comparing weather radar and rain gauge observations

Relatore: Prof. Francesco Marra

Correlatore: Dott. Katharina Lengfeld

Laureanda: Sara Poorakbar

Matricola: 2081256

ANNO ACCADEMICO 2024-2025

Don’t let the fear of losing be greater than the excitement of winning.

Robert Kiyosaki

Contents

Abstract

Extreme precipitation events exhibit significant spatial variability, which needs to be accurately assessed for effective risk management and infrastructure planning. This study compares two methods: KOSTRA-DWD-2020, developed by Deutscher Wetterdienst (DWD), and the Simplified Metastatistical Extreme Value (SMEV) approach, to assess their effectiveness in representation of spatial variability across various return periods and durations. By integrating radar and rain gauge data over Germany, the study analyzes how spatial scale, station density, climatology, and spatial aggregation contribute to influencing the representation of spatial variability and estimation uncertainty for extreme precipitation. The results show that both methods identify regional patterns, notably highlighting the Alps, where spatial variability is more pronounced while SMEV offers a more detailed representation of spatial variability for short-duration, low-return period events. The findings underscore the importance of selecting appropriate estimation methods based on data characteristics, application context, and spatial scale, ensuring that methodological assumptions do not introduce biases or misrepresent variability in extreme precipitation estimates.

Keywords: precipitation, climate extremes, weather radar, SMEV, KOSTRA-DWD-2020

Riassunto

Gli eventi di precipitazioni estreme presentano una importante variabilità spaziale, va accuratamente valutata per una gestione efficace del rischio e una pianificazione delle infrastrutture. Questo studio confronta due metodi: KOSTRA-DWD-2020, sviluppato dal Deutscher Wetterdienst (DWD), e l'approccio Simplified Metasttistical Extreme Value (SMEV), per valutare la loro efficacia nel rappresentare la variabilità spaziale in vari periodi di ritorno e durate. Integrando dati radar e misure da pluviometro in Germania, lo studio esamina come la scala spaziale, la densità delle stazioni, la climatologia e l'aggregazione spaziale influenzino la rappresentazione della variabilità spaziale e l'incertezza della stima delle precipitazioni estreme. I risultati mostrano che entrambi i metodi identificano pattern regionali, evidenziando in particolare le Alpi, dove la variabilità spaziale è più pronunciata, mentre SMEV offre una rappresentazione più dettagliata della variabilità spaziale per eventi di breve durata e basso periodo di ritorno. I risultati evidenziano l'importanza di selezionare metodi di stima appropriati in base alle caratteristiche dei dati, al contesto applicativo e alla scala spaziale, assicurando che le ipotesi metodologiche non introducano distorsioni o travisino la variabilità nelle stime delle precipitazioni estreme.

Parole chiave: Precipitazioni, estremi climatici, radar meteorologico, SMEV, KOSTRA-DWD-2020

1 Introduction

The estimation of extreme precipitation is fundamental to safeguarding human lives, ecosystems, and infrastructure while ensuring sustainable development in the face of climatic uncertainties (Frei et al., 2000). In recent decades, the frequency of natural disasters induced by extreme precipitation has risen significantly. A notable example is the catastrophic floods in Germany and Belgium in July 2021, where rainfall averaged between 100 and 150 mm within 24 hours—equivalent to more than a month's worth of rain. Similarly, the Elbe flood in 2002, triggered by a week of intense rainfall, caused widespread flooding across much of Europe. These events underscore the critical need to understand, estimate, and prepare for extreme precipitation to mitigate its devastating impacts.

1.1 The importance of spatial variability of extreme precipitation

The spatial variability of extreme precipitation plays a crucial role in hydrological and climatological studies, as it directly influences flood risk assessments, infrastructure design, and water resource management. Extreme rainfall events can exhibit significant spatial heterogeneity due to variations in topography, atmospheric dynamics, and local land-use characteristics. This variability affects runoff generation, soil saturation levels, and flood peak magnitudes, making it essential to account for spatial patterns when estimating precipitation return level.

Assessing the probability of extreme precipitation events at different locations is vital for accurate flood forecasting and risk mitigation. The spatial distribution of extreme events determines the extent of flood-prone areas and informs decision-making for urban planning, dam safety, and emergency response strategies. Inconsistent or inadequate consideration of spatial variability can lead to underestimations or overestimations of flood hazards, resulting in either excessive infrastructure costs or insufficient protection against extreme events (McGlynn et al., 2013).

A major challenge in this context is the uncertainty in the estimation of extreme precipitation, which may mask the spatial variability of precipitation. This is particularly important when relying on gridded datasets, weather radar, or sparsely distributed station measurements because these sources may introduce biases due to interpolation methods, measurement errors, or spatial averaging effects, impacting the reliability of return period estimates for precipitation return levels (Koutsoyiannis, 2004; Westra et al., 2014). The

implications of estimation uncertainty will be discussed in detail in a later section, where potential approaches for reducing uncertainty and improving spatial rainfall estimates will be explored.

1.2 Precipitation processes

A key factor in this understanding is recognizing the spatial variability of extreme rainfall, which plays a crucial role in determining how these events are distributed across regions and the differing impacts they have on various landscapes. This variability is shaped by (i) large scale atmospheric dynamics, such as the prevalent global atmospheric circulation, (ii) local atmospheric dynamics and thermodynamics, such as wind patterns and weather systems, and (iii) local spatial features, such as topography, land use, and proximity to water bodies. For example, the convergence of trade winds around the equator causes convective precipitation all year around in a large area at tropical latitudes, known as intertropical convergence zone. Conversely, precipitation patterns at the mid latitudes are dominated by low-pressure systems that originate extratropical cyclones. Mountain ranges enhance rainfall through orographic lifting, where moist air is forced upwards by the mountains, cooling and condensing to produce precipitation on the windward side, while the leeward side remains dry.

Wind patterns play a significant role in shaping weather systems and thus affecting extreme rainfall distribution. For instance, most weather systems in Europe typically move from the southwest to the northeast because of the prevailing westerly winds. Westerly winds are one of the most important wind regimes in Europe, which blow from west to east between 30° and 60° latitude in both hemispheres. These winds are driven by the Earth's rotation and the distribution of heat from the sun.

For example, in Germany, the westerlies bring moist air from the Atlantic Ocean, making the western parts of the country wetter compared to the east. As the westerly winds move inland, they encounter the topography of the land. The mountains in the west and south of Germany, such as the Alps and the Black Forest, force the air to rise, cooling and condensing the moisture, which leads to precipitation through orographic lifting (Lengfeld et al., 2019). As a result, the western regions of Germany, which are closer to the Atlantic, receive more precipitation, while the influence of the westerlies weakens toward the east. In eastern Germany, the westerlies are less effective in providing

moisture, leading to drier conditions, especially in the rain shadow areas behind mountain ranges.

Other atmospheric processes also impact spatial variability by creating different kinds of rainfall: convective and stratigraphic rainfall. Convective rainfall, driven by localized instability and rapid vertical motion, produces intense but localized downpours, in the form of thunderstorms or heavy showers. It is typically short-lived and occurs when warm, moist air rises rapidly, cooling and condensing as it ascends to higher altitudes, forming clouds and precipitation. Due to the localized nature of convection, rainfall can be highly concentrated in small areas, leading to significant variations in precipitation over short distances (Houze Jr & Robert A, 2014). In general, convective rainfall is responsible for extreme precipitation at short durations, up to ~3-6 hours.

In contrast, stratigraphic rainfall occurs when large air masses are slowly lifted while moving horizontally across a region, carrying moisture from one area to another. As moist air is transported across a colder region or lifted, it can cool and condense, resulting in widespread, evenly distributed rainfall. Stratigraphic rainfall tends to cover larger areas and exhibits less spatial variability, with precipitation typically occurring more steadily and over longer periods of time compared to convective rainfall. In general, stratigraphic rainfall is responsible for extreme precipitation at durations longer than ~6-12 hours.

Considering all the factors that influence precipitation, such as topography, wind patterns, and atmospheric dynamics, accurate estimation of extreme rainfall events requires robust historical data. However, precipitation remains one of the most challenging elements to observe and quantify in our climate system due to its high variability in space and time (Lengfeld et al., 2021). Extreme precipitation events pose even greater difficulties, as their rarity and the limited data available make it challenging to assess changes in their frequency or intensity. The rarer the event, the more difficult it becomes to detect and analyze. Consequently, confidence in identifying changes in extreme precipitation heavily depends on the quality and quantity of data, as well as the availability of studies analyzing these data (IPCC, 2012).

1.3 Measuring Precipitation

1.3.1 Rain gauges

The earliest approach to measuring rainfall involves the use of rain gauges, which collect precipitation and determine its volume over a specific period of time. Rain gauges

operate on various measurement principles and can be categorized into different types, including storage gauges, tipping-bucket rain gauges, weighing gauges, drop counting gauges, and non-catching instruments such as disdrometers (Fig 1-1). The volumetric method involves a container with a known surface area where the rainfall depth is calculated as the water volume divided by the collector's area. Storage gauges operate on this principle, requiring manual or automatic water level measurement and periodic emptying. Tipping-bucket rain gauges use a funnel to direct rain into a pair of small buckets balanced on a pivot; when one bucket fills to a set volume, it tips, empties, and resets, with each tip recorded as a fixed rain volume. Weighing gauges collect precipitation in a container and determine rainfall by measuring weight changes over time, offering high sensitivity but requiring corrections for systematic biases. Drop counting gauges use a calibrated nozzle to release droplets, which are optically counted to calculate total rainfall based on drop size and frequency. Non-catching instruments, like disdrometers, rely on optical, acoustic, or microwave methods to sense precipitation without collecting it, providing additional details such as particle size, velocity, and type. These gauges require accurate calibration, environmental corrections, and regular maintenance to ensure reliability, as factors like wind, debris, and undercatch errors can impact measurements, necessitating adjustments for precise data.

In all regions of the world, available rainfall data are characterized by different time-resolutions, mainly on the basis of the specific objective of the network manager and also of the technologic progress of the adopted recording devices (Morbidelli et al., 2022). The time resolution of rain gauges refers to the frequency at which they record rainfall measurements, which can range from one minute to one day. Another important aspect of rain gauges is the length of their historical record and the availability of data. For instance, in the UK, the Royal Botanic Gardens at Kew in London hosts one of the oldest continuous weather monitoring stations in the country. Established in 1840, the Kew Gardens weather station has been collecting data on precipitation and other meteorological variables. Some of the original rain gauges and instruments from the 19th century are still preserved and are used for historical comparisons.

Spatial extent of rainfall data refers to the density and distribution of rain gauges within a network. Designing a rain gauge network involves strategically placing gauges to effectively capture the spatial and temporal variability of rainfall, which is essential for

Fig 1-1 Different types of rain gauges: From left to right: storage gauge, tipping bucket, weighing guage, and disdrometer.

hydrological applications. Factors influencing the density and placement of the network include the terrain, climate characteristics, and the specific purpose of the data collection. Multi-purpose networks are often preferred, as they cater to a variety of users and applications while maintaining a focus on accuracy and traceability of the collected data. A well-designed network ensures reliable information for both scientific research and practical decision-making.

Some examples of common networks include: MeteoSwiss in Switzerland, Deutscher Wetterdienst in Germany, the Walnut Gulch Experimental Watershed in the USA, and WegenerNet Feldbach Region (FBR) in Austria. Among these, the last two have the highest density of rain gauges relative to the area. Walnut Gulch covers about 149 km² and includes approximately 90 rain gauges located at altitudes ranging from 1100 to 2300

meters (Tan et al., 2018), resulting in an average density of about 0.6 gauges per square kilometer. In the WegenerNet FBR, 155 hydrometeorological stations are distributed over an area of about 300 km², with altitudes ranging from 257 to 520 meters. The average density in this region is about 0.5 stations per square kilometer (Fuchsberger et al., 2021).

Although rain gauges still represent the most direct and accurate way to measure precipitation, they fail to capture the spatial variability of precipitation, particularly in extreme events. While rain gauges provide high temporal resolution, their spatial resolution is limited to point measurements, meaning they capture rainfall data for a specific location. This localized nature makes them highly accurate for the exact point but less representative of rainfall distribution over larger areas.

1.3.2 Weather radar

Weather radar plays a crucial role in studying the spatial variability of extreme precipitation. Unlike rain gauges, which provide point-based measurements, radar offers continuous spatial coverage (e.g., every kilometer or even smaller), enabling a more comprehensive representation of precipitation patterns across a variety of scales, from the most localized events to larger regional systems. Radar data are essential for quantifying rainfall patterns at various spatial scales, making them particularly valuable for analyzing the variability of extreme events (Marra & Morin, 2015).

The history of weather radar begins in the early 20th century, with radar technology initially developed for military purposes during World War II. Scientists discovered that radar could detect precipitation when rain, snow, or hail caused interference in military radar signals. This finding led to the adaptation of radar for meteorological applications. Radar measures precipitation by emitting electromagnetic waves and measuring the signal that is scattered back, by raindrops or other hydrometeors. Under some assumption the intensity of the returned signal can be converted to reflectivity, which is used to estimate rainfall rates. However, radar-based rainfall estimates are fraught with uncertainties and errors due to factors such as beam attenuation, ground clutter, variations in drop size distribution, and the influence of non-precipitation particles like insects or dust, which all lead to violations of the aforementioned assumptions. These challenges highlighted the need for advancements in radar technology to improve accuracy. Significant progress was made with the introduction of volumetric scanning, signal digitization, and error-reduction techniques, which enhanced radar-based rainfall

Fig 1-2 Weather radar, NEXRAD in South Dakota

estimation.

At its core, weather radar operates by transmitting short pulses of electromagnetic waves into the atmosphere and analyzing the energy scattered back by precipitation particles. The strength of the returned signal, known as the radar reflectivity, is determined by the size, number, and type of precipitation particles within the radar's sampling volume. This is quantified using the reflectivity factor (Z), which represents the cumulative scattering power of the raindrops and is mathematically expressed as the sixth power of their diameters, weighted by their distribution:

$$Z = \int_0^\infty N\,(D)\,D^6\,dD \qquad \textit{Eq 1}$$

where Z is the reflectivity factor, N (D) is the drop size distribution, representing the number of raindrops with diameter D per unit volume and per unit diameter interval, and D is the diameter of the raindrop (in millimeters). The term D^6 reflects the fact that larger raindrops contribute significantly more to the radar reflectivity than smaller drops. The rainfall rate (R) is derived from the drop size distribution and the velocity of falling

raindrops v (D) , and under the condition of ignoring wind, turbulence, raindrop interaction effect is as follows:

$$R = 6 \times 10^{-4} \int_0^{\infty} N\,(D)\; v\,(D)\, D^3\, dD \qquad \textit{Eq 2}$$

where v (D) represents the functional relationship between the raindrop terminal fall speed in still air and the equivalent spherical raindrop diameter D. In practice, the drop size distribution is rarely known, and reflectivity is empirically (Battan, 1973) related to rainfall rate using a power-law relationship:

$$Z = aR^b \qquad \textit{Eq 3}$$

where a and b are coefficients that vary with geography and rainfall type. Battan (1973) seminal work on radar meteorology includes a compilation of 69 empirical Z–R relationships, each tailored to different climatic conditions across the globe. Building on this Uijlenhoet (2001) analyzed these relationships and derived a mean power-law formula, which closely aligns with the widely adopted Z–R relationship:

$$Z = 200R^{1,6} \qquad \textit{Eq 4}$$

(Marshall et al., 1955). This is widely used because it works well for many types of rainfall.

The shape of raindrops plays a critical role in radar rainfall estimation, as it influences how the radar signal interacts with precipitation particles. Raindrops are not perfect spheres; their shape changes depending on their size. Small drops, typically less than 1 mm in diameter, remain spherical due to surface tension. However, as drops grow larger, air resistance causes them to flatten into an elliptical shape, with the major axis oriented horizontally. This flattening becomes more pronounced as the size increases. This represents a problem for single polarization radars, as the quantification of radar reflectivity from the measured electromagnetic power assumes that all hydrometeors are spherical.

The ratio of the drop's vertical to horizontal axis, known as the axis ratio, helps quantify the degree of flattening. Understanding these shapes is essential for polarimetric radars, which use differences in horizontal and vertical radar signals to extract additional details about raindrop sizes and distributions. These differences enable more precise rainfall measurements, especially for heavy or mixed precipitation types.

Weather radars transmit electromagnetic waves in short pulses from a directional antenna enclosed in a protective radome, typically elevated above obstructions on a tower (Fig 1-2). These pulses travel at the speed of light and interact with precipitation particles, scattering energy back to the radar. The radar calculates the distance to the target by measuring the time delay of the returned signal, known as the radar echo. Scanning at different angles allows volumetric atmospheric observations.

The radar's resolution is determined by the pulse volume, which depends on the pulse length and the antenna's beam width. As range, the distance from the radar, increases, the beam widens, reducing spatial resolution. For example, at 50 km, the beam spreads to a width of 0.9 km, at 100 km, the width increases to 1.7 km.

Weather radars operate in three primary frequency bands—S-band, C-band, and X-band—each selected based on the balance between range, sensitivity to precipitation, signal attenuation, and cost. S-band radars (2.7–2.9 GHz) are ideal for long-range detection, capable of measuring heavy rainfall up to 300 km away with minimal attenuation, making them highly reliable for severe weather monitoring. However, this reliability comes at a high cost, making S-band radars the most expensive option. On the other hand, X-band radars (9.3–9.5 GHz) are more affordable and highly sensitive to small precipitation particles, making them effective for short-range observations up to 50 km. Despite their sensitivity, X-band radars are significantly affected by attenuation in heavy rain, which limits their accuracy for quantitative precipitation estimation. Positioned between these two is the C-band radar (5.6–5.65 GHz), which offers a balanced compromise. It provides moderate range, reliability, and cost, making it a versatile option for general weather monitoring. The choice of radar frequency ultimately depends on the specific requirements, such as the area of coverage, typical weather conditions, and budget considerations.

The radar equation describes how weather radars measure precipitation by relating the power of the reflected signal to the physical properties of precipitation particles. Radars emit short pulses of microwave radiation, and the energy scattered back by precipitation particles is received as reflected power. This received power is then linked to the size, shape, and concentration of the particles.

Under the Rayleigh scattering approximation (valid for spherical droplets of liquid water with diameters much smaller than the radar wavelength, $D \ll \lambda$), the radar equation can be expressed as:

$$P_r = \frac{\pi^3 P_t g^2 \theta \varphi h |K|^2 \sum D_i^{\sigma}}{1024 ln2 \lambda^2 r^2} \quad \textit{Eq 5}$$

Where P_r is the received power, P_r is the transmitted power, g is the antenna gain (measures how effectively a radar antenna focuses energy in a specific direction), θ and φ are the Horizontal and vertical beamwidths. h is the pulse length, K is the dielectric factor of the scatterers, D_i is diameter of the scatterer. λ is radar wavelength and r is distance to the target. For a unit volume, the equation simplifies to:

$$P_r = \frac{C|K|^2 Z}{r^2} \quad \textit{Eq 6}$$

Where C is the radar constant (which includes fixed radar characteristics), and Z is the radar reflectivity factor. The equation assumes Rayleigh scattering, where particles are much smaller than the radar wavelength. However, at higher frequencies (e.g., X-band), this assumption may break down, and Mie scattering occurs ($D \geq \lambda$). In such cases, the reflectivity factor Z is adjusted to the equivalent radar reflectivity factor (Z_e).

The radar equation relies on several assumptions, and deviations from these can result in measurement errors. For instance, if the radar beam contains mixed precipitation types or is not fully filled, the reflectivity measurements may become inaccurate due to target non-uniformity. Instrument-related issues, such as incorrect values for antenna gain, waveguide loss, or pulse length, can introduce systematic biases in the radar constant, leading to errors in converting power to reflectivity. Additionally, environmental variability, such as incorrect assumptions about the dielectric properties of precipitation particles and how they scatter and absorb radar waves, may further impact accuracy. These errors can impact both qualitative applications, like identifying precipitation patterns, and quantitative tasks, such as estimating rainfall intensity, emphasizing the need for precise calibration and error correction.

Radar rainfall estimates provide spatially dense precipitation measurements, but they are prone to significant errors due to various factors. One major issue is the variability in the Z–R relationship ,which connects reflectivity (Z) to rainfall rate (R). This relationship depends on the drop size distribution, which varies by location, season, and storm type (e.g., Steiner & Smith, 2000; Uijlenhoet, 2001; Uijlenhoet et al., 2003; Smith et al., 2009),

making fixed Z–R models (Brandes et al., 1999; Villarini & Krajewski, 2010) prone to errors. Ground clutter, caused by terrain, buildings, and other obstacles, interferes with radar signals, but clutter maps and algorithms help mitigate this issue . Beam blocking from physical obstructions, such as mountains, leads to underestimation of rainfall, corrected through visibility maps and digital elevation models (Pellarin et al., 2002). Range-related errors also pose challenges, as radar beams sample higher altitudes at longer distances, introducing biases due to the vertical profile of reflectivity and bright band effects. Attenuation of radar signals by heavy rain or wet radomes, especially for X- and C-band radars, weakens signals and requires corrections using power-law models or mountain reference techniques (Marra et al., 2014; Marra & Morin, 2018). Additionally, electrical instabilities in radar systems, such as calibration drift or hardware issues, can compromise accuracy, necessitating regular calibration with reference targets like the sun. These challenges highlight the complexity of radar rainfall estimation and the need for advanced correction methods to improve measurement reliability. For a more comprehensive review on radar challenges and the solution to them, see Borga et al. (2022).

1.4 Estimation of extreme precipitation

Once rainfall data is collected, either through radar or rain gauges, the next critical step is the estimation of extreme rainfall events. These estimates are essential for understanding the potential impacts of extreme weather on infrastructure, urban planning, and flood risk assessment (Koutsoyiannis et al., 2008). Accurate estimations of extreme rainfall are particularly crucial in the face of climate change, which is expected to alter rainfall patterns and intensities (IPCC, 2023). Over the years, various models have been developed to predict extreme rainfall and estimate precipitation return levels for specific return periods and durations. These models aim to represent the statistical behavior of rainfall extremes under changing climate conditions and are often based on observed rainfall data, regional characteristics, and advanced statistical methods (Lengfeld & Marra, 2024).

Understanding how extreme precipitation is estimated requires a clear understanding of the key terminology involved. Precipitation intensity is expressed in millimeters per hour (mm h^{-1}) and refers to the amount of precipitation collected per unit time interval. Meanwhile, rainfall depth, measured in millimeters (mm), represents the total amount of

precipitation that reaches the ground over a stated period. It is expressed in terms of the vertical depth of water that would cover a horizontal projection of the Earth's surface. Duration refers to the length of the time window of interest, for example the concentration time of a study catchment. This can range from a few hours to several days, depending on the case. Return period (or recurrence interval) is defined as the inverse of the probability of exceedance in a year. It expresses how often, on average, a rainfall event of a certain intensity is expected to occur within a given timeframe, usually expressed in years. For instance, a 100-year rainfall event describes an event with a 1% chance of occurring in any given year, such as the Elbe flood event. It does not mean the event will happen precisely once every 100 years, but rather conveys the probability of its occurrence. A return level is the magnitude of an extreme precipitation, that is expected to be met or exceeded, on average, once within a given return period.

When estimating extreme precipitation based on rain gauges, several common statistical methods and models are used to derive reliable estimates for return level values, often for specific return periods (e.g., 10-year, 50-year, 100-year events). One widely used approach is extreme value theory, which involves fitting distributions such as the Generalized Extreme Value (GEV) or Generalized Pareto Distribution to the extremes of the dataset. The extreme value theorem was first formulated by Fisher & Tippett (1928), Fréchet (1927), and Gnedenko (1943) while the Generalized Pareto Distribution was later developed by Pickands, (1975). These methods, often applied through Block Maxima or Peak Over Threshold (POT) techniques, allow for the estimation of return periods for extreme events (Jenkinson, 1955). The GEV distribution is used when the analyses are based on annual maximum values, for example, the maximum daily rainfall over a specific year or the maximum rainfall over a given time window. The idea is to fit the GEV distribution to these maximum values and use it to extrapolate the probability of rare and potentially unobserved events. The Generalized Pareto Distribution is often used in the POT method. Instead of using annual maxima, the POT method focuses on exceedances above a specified high threshold. To link these estimates across a range of different durations, Intensity-Duration-Frequency curves are used. They relate rainfall intensity, duration, and frequency for specific return periods (Bernard, 1932). These curves can be derived from long-term data at individual stations or regionally through spatial interpolation.

1.4.1 Application of weather radar in extreme precipitation

Weather radar is an essential tool for capturing the spatial variability of extreme precipitation, providing high-resolution spatial coverage that is crucial for hydrological and climatological studies. Extreme rainfall events often exhibit significant spatial heterogeneity due to variations in topography, atmospheric dynamics, and local land-use characteristics. These spatial variations directly influence flood risk assessments, infrastructure design, and water resource management, making it vital to account for spatial patterns when estimating extreme precipitation (Marra, Nikolopoulos, et al., 2019).

Several studies have highlighted the benefits of radar in estimating precipitation. Lengfeld et al. (2020) emphasized the indispensable role of radar in observing short-term extreme precipitation, highlighting its superiority over rain gauges, which struggle to capture events lasting less than an hour and covering areas of only a few km². The characteristics of extreme precipitation, as described by Isidoro Orlanski (1975), further reinforce this, noting that frontal systems typically produce prolonged (>1 day) rainfall over scales exceeding 100 km, whereas cumulus clouds and deep convection, which drive short-term extremes (10 minutes to a few hours), are confined to scales of ~1 km. Borga et al. (2008) underscored the importance of radar for monitoring convective cells that gauge networks fail to capture effectively. Similarly, Terink et al. (2018) estimated that replicating radar's ability to monitor spatial variability would require at least 12 gauges per km². Peleg et al. (2018) in their examination of extreme rainfall at radar subpixel scales, demonstrated that radar-based estimates exhibit greater spatial variability for shorter durations, emphasizing radar's capacity to detect fine-scale localized phenomena. Smith et al. (2024) further noted that the sparse distribution of rain gauges often results in missed localized variations, leading to an underrepresentation of spatial gradients and variability compared to radar, which provides spatially continuous rainfall fields.

Despite the potential, radar data has limitations, including estimation errors and relatively short records (typically 10-20 years). This can make it difficult to use them directly in frequency analysis, which traditionally relies on long historical datasets to estimate extreme events (Villarini & Krajewski, 2010b; Kidd & Levizzani, 2011). Therefore, precipitation frequency analysis methods that maximize the use of available data are preferred.

One major application of precipitation frequency analysis from radar is the derivation of high quantiles, statistical thresholds for rare, extreme rainfall events, used in design applications (Overeem et al., 2009; Wright et al., 2013; Gado et al., 2017; Marra, Nikolopoulos, et al., 2019). However, their short archive length often lacks enough examples of extreme events, leading to significant uncertainty (Papalexiou & Koutsoyiannis, 2013) in statistical models, particularly in predicting the tails where rare events occur. POT method, which analyzes all rainfall events above a chosen intensity threshold, can reduce this uncertainty by using more data points compared to the Annual Maxima Series (AMS), which only considers the largest event each year. However, implementing POT is complex because the threshold selection must be done consistently across all grid points to maintain spatial coherence which is difficult to automate (Langousis et al., 2016a; Langousis et al., 2016b). Validation of the high quantiles against rain gauge data often reveals systematic biases in radar data, including underestimation (Overeem et al., 2009; Eldardiry et al., 2015; Pombo & de Oliveira, 2015) due to retrieval errors in radar data (Berne & Krajewski, 2013) and overestimation caused by random errors and short records (Marra & Morin, 2015). Regionalization, which pools data from similar nearby locations, has proven effective in reducing these uncertainties and improving predictions for ungauged areas.

Another key application is monitoring and predicting extreme rainfall events in real time, especially in areas with little or no direct measurement data (ungauged or poorly gauged regions). By analyzing these datasets, researchers can calculate extreme quantiles—statistical thresholds used to identify the likelihood of rare events like heavy rainfall. These quantiles are important because they set warning levels for systems designed to monitor extreme weather events. When the same dataset is used for both setting these thresholds and monitoring real-time conditions, systematic biases (errors caused by differences in data) can be reduced (Zhou et al., 2015).

Radar data is also useful for estimating subdaily precipitation quantiles, which describe rainfall intensities over short time intervals (e.g., minutes to hours). Temporal downscaling techniques use statistical relationships to estimate high-resolution extremes from coarser precipitation data (Awadallah et al., 2011; Awadallah, 2013). For instance, data from radar can provide relationships between short-term rainfall patterns (e.g., hourly rainfall) and longer-term statistics (e.g., daily rainfall). These relationships can

then be combined with ground observations to refine the estimates and make them more locally accurate. Another approach involves using weather radar data, which already captures short-term precipitation distributions, and combining it with high daily rainfall amounts from gauges (Bárdossy & Pegram, 2017). This allows for a more direct estimation of short-scale precipitation behavior.

Radar data enhances our understanding of the climatology of extreme precipitation by revealing the spatial distribution of rainfall patterns with high precision (Demirdjian et al., 2018). By analyzing radar data, researchers can identify spatial gradients, changes in precipitation frequency and intensity across regions, that help explain the factors influencing extreme rainfall, such as geography or climate systems.

However, critical issues could compromise the accuracy of frequency analyses based on radar based estimates. Radar precipitation estimates are indirect and subject to various errors, particularly when analyzing extremes (Villarini & Krajewski, 2010; AghaKouchak et al., 2011; Nastos et al., 2013; Marra et al., 2014; Miao et al., 2015; Prakash et al., 2016; Eldardiry et al., 2017). These errors can lead to systematic underestimation and estimation uncertainty (random errors). One of the primary concerns is the large uncertainty that arises from limited data points used to estimate the parameters of extreme value distributions. For instance, Marra et al. (2018) showed that using short records (e.g., 5-20 years) with AMS and POT leads to large uncertainties in the estimation of 100-year return period quantiles. Eldardiry et al. (2015) shows that short records of radar data can lead to significant variability in quantile estimates, although they do not necessarily cause tangible biases in the estimated quantiles.

To address these issues, regionalization methods (Overeem et al., 2009; Eldardiry et al., 2015; Gado et al., 2017) and stochastic storm transposition (Wright et al., 2013; Wright et al., 2017) can help, but they rely on assumptions about spatial homogeneity, which may be difficult to support in regions with strong climatic gradients or ungauged areas (Marra & Morin, 2015). Another major challenge is validating high rainfall quantiles when direct measurements are unavailable. In such cases, rain gauge data are often used as a reference, with radar-based quantiles compared to gauge-derived quantiles using extreme value methods (Overeem et al., 2009; Eldardiry et al., 2015). However, this comparison is complicated by the scale mismatch between radar spatial distributions and rain gauges. This scale mismatch can lead to discrepancies in the resulting quantiles.

Furthermore, radar data provide instantaneous snapshots, whereas rain gauges offer temporal averages. This inherent difference in data types and temporal scales can contribute to differences in the distributional properties of the estimated quantiles.

1.4.2 Non-asymptotic approaches

While traditional approaches to estimating extreme precipitation are widely used, they face challenges in handling uncertainties and capturing small-scale variability. Addressing these gaps is critical for more reliable risk assessments and better planning for extreme rainfall events. For example non-asymptotic methods (Marani & Ignaccolo, 2015; Marra, Zoccatelli, et al., 2019) offer a promising approach for estimating return levels (Miniussi & Marra, 2021 ; Marra et al., 2022). Unlike traditional extreme value theory, which typically assumes that extreme events are rare and that their distribution can be modeled using asymptotic methods, non-asymptotic approaches make different assumptions. These methods assume that the parent distribution of the process (e.g., precipitation) is known and treat extreme events as rare samples from this known distribution.

By using a larger fraction of the available observations, these methods help to reduce uncertainties that arise when data records are short. This is particularly important in the case of weather radar data, which may have relatively short records compared to traditional rain gauge data (Zorzetto et al., 2016; Marra et al., 2018). For example, Lengfeld & Marra (2024) demonstrated that a non-asymptotic approach, the Simplified Metastetical Extreme Value (SMEV) (Marra, Zoccatelli, et al., 2019), offers a significant advantage due to its low sensitivity to statistical outliers and the length of the time series. This characteristic makes SMEV particularly suitable for estimating extreme precipitation from radar data, especially in situations where the available time records are relatively short.

1.5 Research gaps and study aims

Despite the increasing use of radar data for estimation of precipitation return level, accurately capturing the spatial variability of extreme precipitation especially for short duration events, remains a challenge. While weather radar offers high-resolution spatial coverage, its short record length, inherent biases, and methodological assumptions introduce uncertainties that affect the reliability of extreme precipitation estimates. These challenges are further compounded by the spatial distribution and density of rain gauges,

which play a critical role in validating radar-derived estimates. In regions with sparse gauge networks, radar-based estimates can suffer from higher uncertainty due to the lack of ground-truth data, highlighting the importance of integrating both sources for more accurate extreme precipitation studies.

A key challenge in using radar and rain gauge data together lies in their differing spatial scales. Radar provides continuous spatial coverage over large areas, but at a coarser resolution compared to the point-based measurements of rain gauges. This discrepancy complicates direct comparisons and introduces uncertainties when evaluating the accuracy of radar-based precipitation estimates. The impact of spatial aggregation on radar-derived extreme rainfall estimates remains insufficiently understood, particularly in heterogeneous landscapes where local variations in precipitation intensity may be smoothed out by radar's spatial averaging. Understanding how these differences affect extreme rainfall estimation is crucial for improving integration strategies between the two data sources.

Both traditional and emerging statistical methods for extreme precipitation estimation depend on specific assumptions and parameter choices, which influence their ability to represent spatial variability. While newer methods have shown promising results, the extent to which they represent spatial representation, particularly across different spatial scales, station densities, and climatological conditions, is still not fully investigated. In particular, the methods' ability to account for the sparse coverage of rain gauges in certain regions, and how this affects its representation of extreme precipitation variability, requires further exploration.

In this study, we evaluate the effectiveness of capturing spatial variability in extreme precipitation estimates derived from radar data, using two methods: KOSTRA-DWD-2020, developed by Deutscher Wetterdienst (DWD, in English: German Weather Service), and the Simplified Metastatistical Extreme Value (SMEV) approach (Marra, Zoccatelli, et al., 2019). Our analysis focuses on suitability of each method for capturing spatial variability of extreme precipitation across different return periods and durations, particularly how methodological assumptions, spatial scale, station density, and climatology influence the representation spatial variability of extreme precipitation. This research examines their effectiveness, particularly for short duration events, where spatial variability of extreme precipitation is more pronounced.

Furthermore, we examine the impact of differing spatial scales between radar and rain gauge data, exploring how this discrepancy affects the representation of spatial variability of extreme precipitation. We also investigate the effect of spatial aggregation on radar-based extreme precipitation estimates, particularly in heterogeneous landscapes where fine-scale variability may be smoothed out, as well as its influence on estimation uncertainty. By analyzing regional differences in spatial variability, this study seeks to determine whether observed variations in extreme precipitation patterns are intrinsic to precipitation dynamics or artifacts introduced by the estimation techniques.

By addressing these uncertainties, this research aims to provide insights into the suitability of different estimation methods under varying spatial and temporal conditions, ultimately contributing to improved extreme precipitation assessments using radar data.

2 Study Area and Data

Our study covers the entire Germany, 357,021 km^2 extending from 47° N to 55° N and from 5° E to 16° E, selected for its dense network of rainfall stations (averaging 1 per 60 km^2), the good coverage with a weather radar network, and the region's diverse precipitation patterns and generation mechanisms. Most precipitation in Germany results from westerly circulation patterns, whose influence diminishes toward the east, while topography plays a crucial role in shaping spatial precipitation patterns as well (Fig 2-1).

Fig 2-1 Topography of Germany with five main landscape units (the North German Plain, Central Uplands, South German Scarplands, and Alpine Foreland regions). Other minor natural areas in Germany are indicated, to ease the discussion. From (Miniussi & Marra, 2021).

Additionally, the availability of both radar data and a comprehensive network of rain gauges significantly enhances the reliability of precipitation data by leveraging the strengths of each measurement system while compensating for their individual limitations (Fig 2-2 & Fig 2-3).

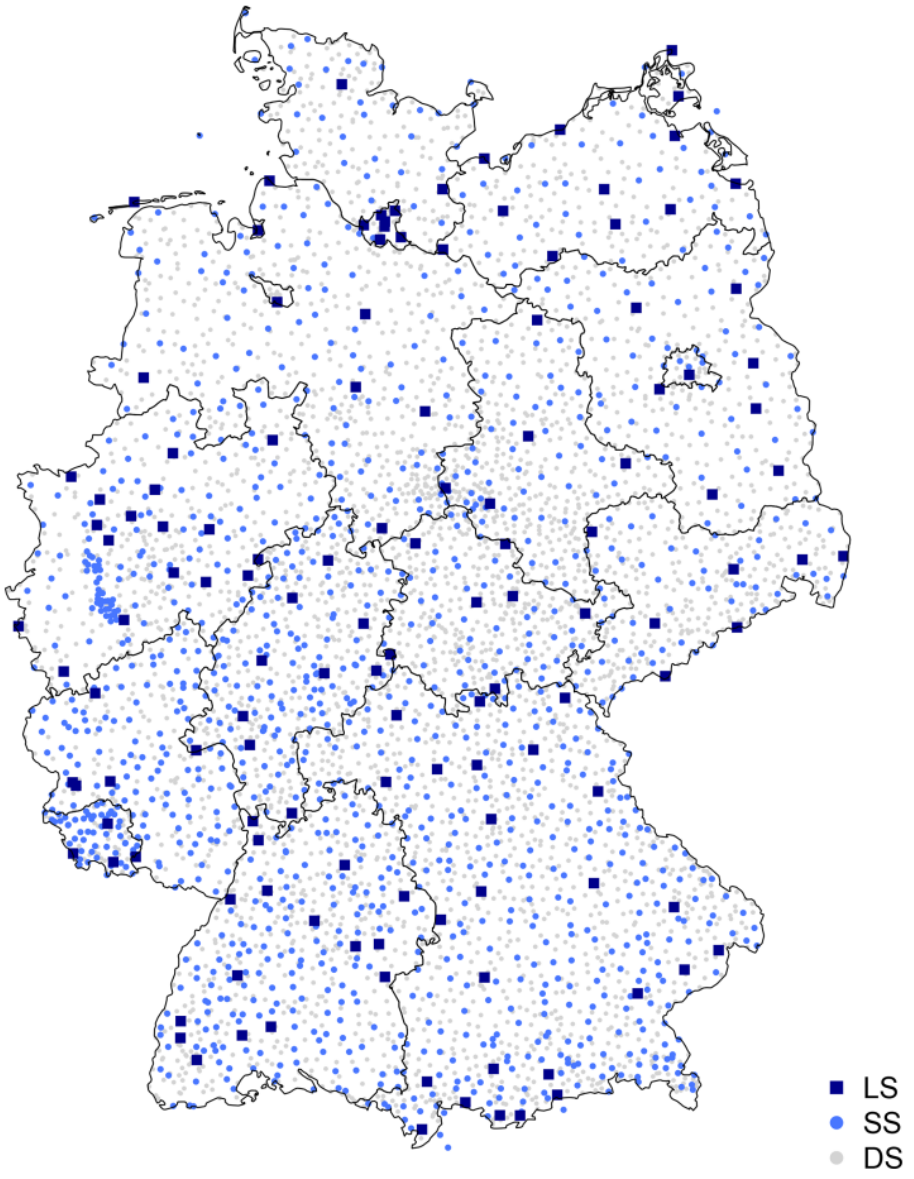

Fig 2-2 Available rainfall data types in Germany categorised in three groups: long series (LS), short series (SS) and daily series (DS). The black lines illustrate the borders of German federal states; from (Shehu et al., 2023).

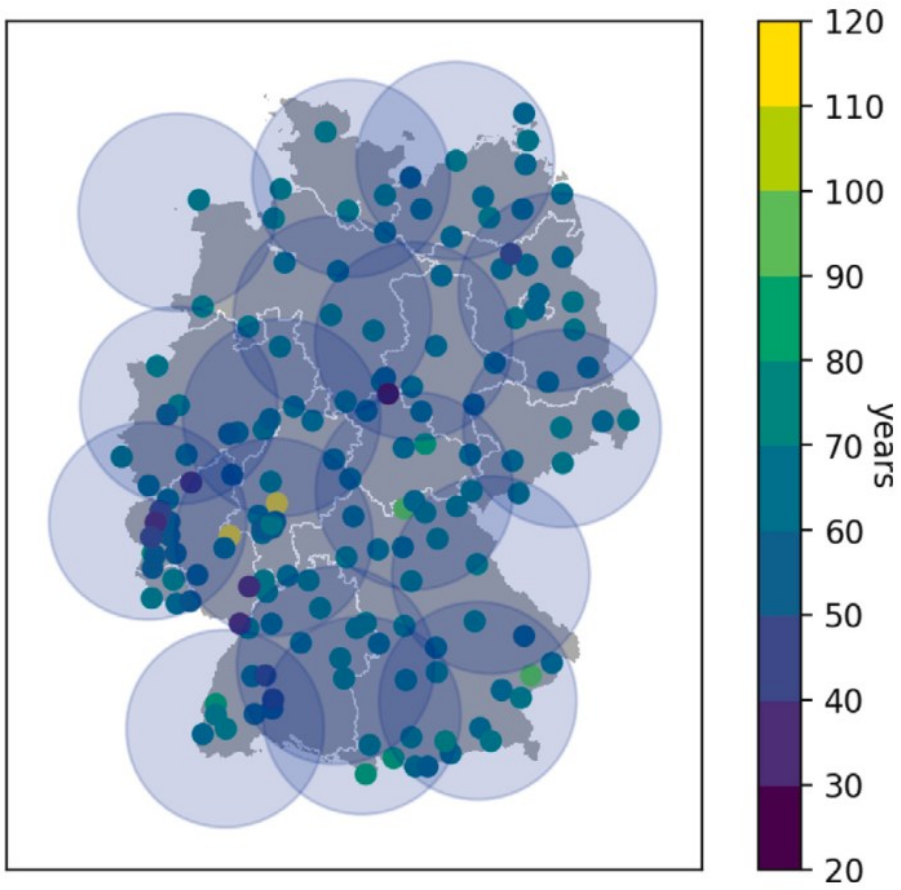

Fig 2-3 Locations of precipitation stations. The colours represent the lengths of the station records in years, circles mark the coverage of the radar network; from (Lengfeld & Marra, 2024).

In this study, precipitation return level estimates are derived using parameters from KOSTRA-DWD-2020 and SMEV, applied to rainfall data from 155 ground-based stations and radar observations. Details on these methods are provided below. The underlying precipitation datasets include:

i) **Radar Data**: DWD operates a network of 17 C-band radar systems (Fig 2-3). The radar network, consisting of 11 single-polarisation Doppler radars and 6 upgraded Doppler systems (with 5 initially single-polarisation systems), began providing nationwide radar products in 2001. Reflectivity data collected by these radars are calibrated using ground-based measurements to ensure accuracy.

To account for technical changes and upgrades, the data are homogenized and corrected for radar-specific errors. These corrections include improved clutter detection, adjustments for signal reduction due to overshooting and attenuation, volume increases, and the correction of spokes resulting from partially shaded radar beams (Winterrath et al., 2017). These processes caused data coverage gaps in the north, east, and south of

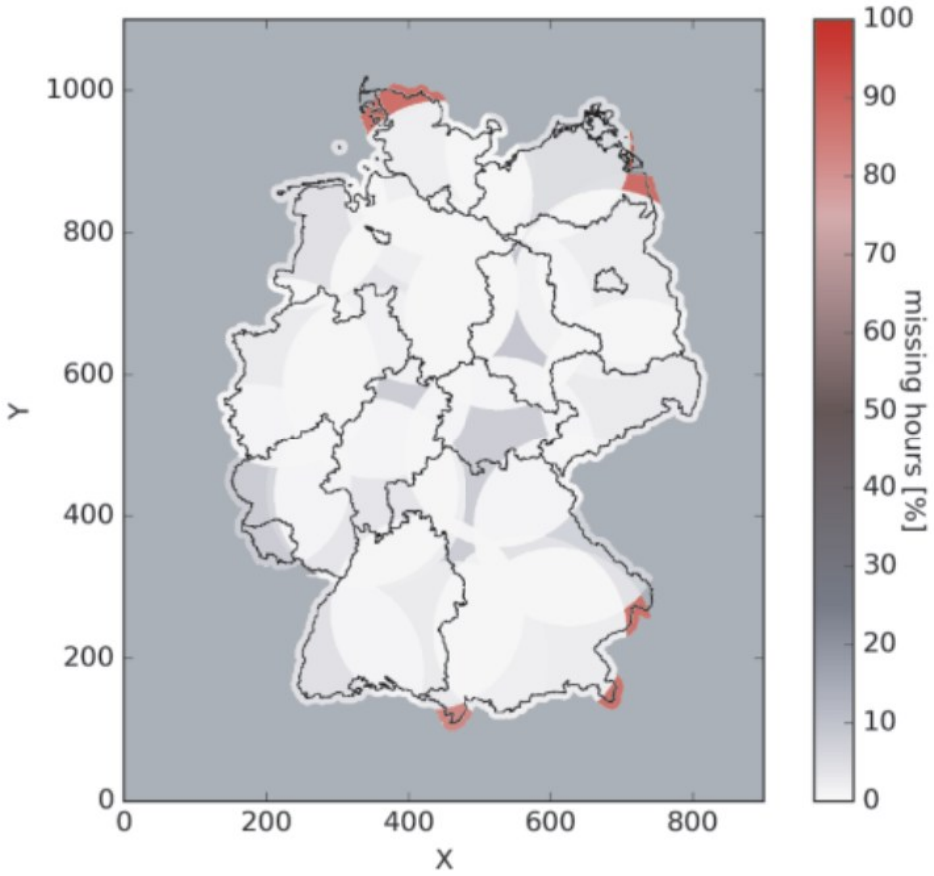

Fig 2-4 Missing hours in the radar composite in percent from 2001 to 2016.

Germany, highlighted in Fig 2-4. The reflectivity data, with a maximum range of 128 km, are compiled into a nationwide mosaic using a polar stereographic projection. Precipitation estimates are adjusted to hourly totals based on in-situ gauge measurements through the real-time RADOLAN method, which has been operational since June 2005 (Winterrath et al., 2012). The daily gauge data are disaggregated to increase the number of station data points available for the hourly adjustment of radar data. This allows for a more comprehensive comparison with the operational RADOLAN product, which lacks daily data. Five-minute radar data are quantified using a scaling factor derived from the ratio of adjusted to non-adjusted hourly totals. This dataset, provided by DWD, is publicly available in (Winterrath et al., 2018).

ii) **Rain Gauge Data**: The rain gauge dataset, also provided by DWD, consists of 155 stations, with record lengths ranging from 29 to 120 years and hourly accumulated precipitation data. All stations cover at least the period from 2001 to 2020, ensuring maximum overlap with the radar dataset.

3 Methods

3.1 KOSTRA-DWD-2020

KOSTRA-DWD-2020 (Koordinierte Starkregenregionalisierung und Auswertung), developed by DWD, is a rainfall statistical framework that operates on three rainfall measuring networks with varying temporal resolutions and station distributions across Germany (Fig 2-2). The Daily Series (DS) includes 4,068 stations and records daily rainfall, covering periods ranging from 10 to 120 years. The Long Series (LS) consists of 133 stations, measuring rainfall at 1-minute intervals with a 0.1 mm resolution and a 2% uncertainty, providing an average of 65 years of observation. The Short Series (SS) consists of 1,261 stations, recording rainfall at 1-minute intervals with a 0.01 mm resolution over a period of 18 years. The daily dataset (DS) exhibits a denser distribution compared to both long (LS) and short (SS) datasets, encompassing extended observation periods beyond those of the LS dataset.

Disaggregating the daily time series using an appropriate methodology significantly enhances the density of data points available for regionalization and estimation of extreme precipitation. In this methodology, disaggregation is performed through the application of a cascade model, as initially proposed by Olsson (1998). This approach leverages the RADOLAN weather radar time series at the grid-cell level to achieve finer temporal granularity in the precipitation data, thereby providing a robust basis for regional analyses and improving the accuracy of return level estimation. The Annual Maximum Series (AMS) is derived using a 5-minute time series by selecting the maximum rainfall values for each year based on durations of 5 minutes, 10 minutes, 15 minutes, 30 minutes, 1 hour, 2 hours, 6 hours, 12 hours, 1 day, 2 days, 3 days, and 7 days, with a minimum dry duration of 4 hours between consecutive events. The next step is fitting the GEV distribution, which is motivated by the Extreme Value Theorem (Fisher & Tippett, 1928; Gnedenko, 1943).

This theorem provides the theoretical foundation for modeling the behavior of extreme values, such as block maxima, by showing that they converge to one of three limiting distributions (Fréchet, Weibull, or Gumbel), all encompassed within the GEV framework. This theorem provides a theoretical foundation for modeling the extreme values of a dataset, such as the maximum values within blocks of time (e.g., annual maxima). It states that, under certain conditions, the block maxima of infinitely large blocks converge to

one of three limiting distributions, which are all unified under the GEV framework. The GEV cumulative distribution function (CDF) is defined as:

$$F(x;\ \mu,\sigma,\gamma) = \ exp\left\{-\left[1+\ \gamma\frac{(x-\mu)}{\sigma}\right]^{-\frac{1}{\gamma}}\right\} 1+\frac{\gamma(x-\mu)}{\sigma} > 0\ ,\ \gamma \neq 0 \qquad Eq\ 7$$

$$F(x;\ \mu,\sigma) = \ exp\left\{-exp\left\{\frac{x-\mu}{\sigma}\right\}\right\}\ ,\ \ \gamma = 0 \qquad Eq\ 8$$

where μ is the location parameter, σ is the scale parameter, and γ is the shape parameter, which determines the tail behavior of the distribution. When $\gamma > 0$ (Eq 7), the Fréchet distribution is a heavy-tailed distribution where the probability of extreme events decreases slowly, indicating that values are unbounded on the upper end. In contrast, when $\gamma = 0$ (Eq 8), the Gumbel distribution is a light-tailed distribution where extreme values decay exponentially, making it versatile for a variety of applications involving less heavy-tailed extremes. The third limiting type (reversed Weibull, for $\gamma < 0$) is upper bounded and is not considered physically plausible for precipitation. For this study, the GEV parameters (μ, σ, γ) are fitted to the AMS of each duration level and station separately, using the L-moments method. The shape parameter (γ) is either estimated or fixed at 0.1, following the recommendation of (Koutsoyiannis, 2004) for return periods up to 100 years. Then, using the Koutsoyiannis method, the intensity for each duration level is generalized by calculating a duration-specific scaling factor,

$$i = i_d . b_d \ with\ b_d = (d+\theta)^{\eta} \qquad Eq\ 9$$

where i is the generalised intensity in mm h^{-1}, i_d is the intensity in mm h^{-1} observed at each duration level, d is the duration level in hours, θ and η are Koutsoyiannis parameters optimized for each station. These parameters are estimated by minimizing the Kruskal-Wallis statistic, a non-parametric method that ensures robustness to extreme values (Shehu et al., 2023).

The next step in the methodology is the regionalisation of the extreme rainfall statistics across Germany. Regionalisation is a process aimed at estimating extreme rainfall at unobserved locations using local data. This involves a series of interpolation and statistical methods. First, Ordinary Kriging is employed to interpolate extreme rainfall parameters such as GEV parameters. A 5x5 km grid is used for the interpolation to ensure consistency with other datasets and to strike a balance between resolution and station density. Next, Kriging with External Drift is applied, where the interpolation incorporates

secondary variables, such as site characteristics like elevation, to explain spatial trends. However, in this study, the correlation between the target variables and site characteristics was found to be weak, so only interpolated local parameters are used as external drift. Finally, Index-based Regionalisation is used, which groups stations into homogeneous regions based on site characteristics like latitude and elevation. The regions are tested for homogeneity, and local statistics are normalized by a scaling factor called the index. L-moments are estimated for each region, and a regional GEV growth curve is fitted to derive extreme rainfall estimates for each region. These regionalisation methods together enable the generation of precipitation return level maps, providing estimates of extreme rainfall for different durations and return periods across Germany. A comprehensive explanation of the data and methodology can be found in (Junghänel et al., 2022; Shehu et al., 2023; Shehu & Haberlandt, 2023).

3.2 Simplified Metastatistical Extreme Value (SMEV) approach

The Simplified Metastatistical Extreme Value (SMEV) describes using parameters directly derived from the ordinary events, which are all the independent realization of the process of interest (Marra, Zoccatelli, et al., 2019). SMEV assumes that extreme values arise from finite sampling of ordinary events, which are here identified as the maximum intensity that is observed over the duration of interest in each independent storm. Storms are defined as consecutive wet time intervals separated by dry period, with the length of the dry period determined based on the region's climatology (Marra et al., 2020); 24 hours in this study (Miniussi & Marra, 2021). Ordinary events are extracted using moving windows of specified durations (e.g., 10, 30, 60, 360, 1440 minutes), with each storm corresponding to an individual ordinary event for each duration. This approach has two key advantages: it ensures the number of ordinary events remains consistent across durations, enabling the examination of their properties and distributions, and it directly relates ordinary events to meteorological features, allowing for consistent application across regions and analysis at multiple scales (Marra et al., 2020).

SMEV assumes that all these ordinary events follow a Weibull distribution(not to be confused with the *reversed* Weibull of the extreme value theorem), whose cumulative distribution function (CDF) is expressed as:

$$W(x;\ \lambda;\ \kappa) = 1 - exp\left(-\left(\frac{x}{\lambda}\right)^{\kappa}\right), \qquad \textit{Eq 10}$$

Where λ is the scale parameter, and κ is the shape parameter. Larger values λ correspond to more intense average events and larger κ to a lighter upper tail. SMEV then accounts for the average number of ordinary events *n* that are observed in each year (a finite and known number, instead of assuming an infinite number as in the extreme value theorem). This leads to a three-parameter description (λ, κ, n), with two parameters describing the event intensity (λ, κ) and one representing the average number of yearly occurences (n):

$$SMEV_{(x)} = \left[1 - exp\left(-\left(\frac{x}{\lambda}\right)^{\kappa}\right)\right]^{n}, \qquad \textit{Eq 11}$$

The parameters of the Weibull distribution are derived using the least-squares method on the left-censored data (Marani & Ignaccolo, 2015). Left censoring (i.e., treating the observations below a selected threshold as non-exceedance) should not be confused with the threshold exceedances of POT methods (in which the values below thresholds are removed from the analyses). This censoring procedure is applied to address the issue of low-intensity ordinary events that may distort the fit of the upper tail distribution. Specifically, we exclude data points that fall below a threshold corresponding to a quantile of the observed ordinary events. In this study, we use the 90th percentile as the left-censoring threshold, meaning that only the 10 top percentiles of the data points are retained for parameter estimation. This ensures that the low-intensity events, which may not accurately represent extreme events due to measurement limitations and physical process differences, do not affect the estimation of the Weibull parameters (Marra et al., 2020).

To ensure clarity for the reader, it is essential to outline the data preparation process, which was conducted in two primary stages. Initially, preprocessing was performed to estimate the parameters for the GEV distribution and the Koutsoyiannis parameters, calculated in accordance with the KOSTRA-DWD-2020 methodology using radar. Simultaneously, the parameters for the Weibull distribution were derived based on the SMEV method, also applied to radar. Subsequently, these estimated parameters were utilized to compute the precipitation return levels. Notably, return levels obtained from station data, as well as those derived using both methods and the SMEV bootstrap values,

were incorporated directly into the analysis without additional preprocessing. Throughout this study, any reference to "KOSTRA" specifically refers to KOSTRA-DWD-2020.

3.3 Spatial aggregation scales

To assess the role of the spatial scales on extreme precipitation and its variability, we reduced the spatial resolution of the 1100 × 900 radar grid data by averaging key parameters—such as the GEV and Koutsoyiannis parameters for KOSTRA, and the Weibull distribution parameters for SMEV—over 1 × 1, 2 × 2, 5 × 5, and 10 × 10 blocks (moving windows). This process served a primary purpose: to assess how the setected spatial variability changes at different radar resolutions, in response to reduced uncertainty expected when reducing the spatial scale of radar data. To ensure the integrity of the coarsening process, we made sure that at least 60% of each moving window (1 × 1, 2 × 2, 5 × 5, or 10 × 10) was filled with valid data. If this threshold was not met, the entire window was filled with NaN values, ensuring that only reliable data was used in further calculations.

After adjusting the spatial scale, we recalculated the return levels based on the new grid of parameters according to each method (KOSTRA-DWD-2020 and SMEV) for the following durations: 1, 3, 6, 12, and 24 hours, and for return periods of 2, 5, 10, 20, 50, and 100 years. The result is precipitation grids with spatial scale coarsened to 1 × 1, 2 × 2, 5 × 5, and 10 × 10 km^2 (Fig 3-1). This process of reducing the spatial scale of the radar grid ensures that the spatial resolution of both the radar-based and station-based data are aligned, which is crucial for performing a valid and meaningful comparison between the two datasets. The alignment of spatial scales enables a more accurate assessment of how the radar estimates correlate with ground-based observations, providing a clearer understanding of their relative strengths and limitations. Furthermore, by reducing the radar grid's spatial scale, we can more effectively account for the inherent uncertainty that comes with radar data while maintaining the statistical validity of the analysis. In the subsequent step of our analysis, we divided the entire grid into 2, 4, 9, and 16 equal blocks (regions) to examine how spatial variability and local patterns in the data change at different regions. To achieve this, we first divided the x-length and y-length of the radar grid into 2, 3, and 4 parts, respectively.

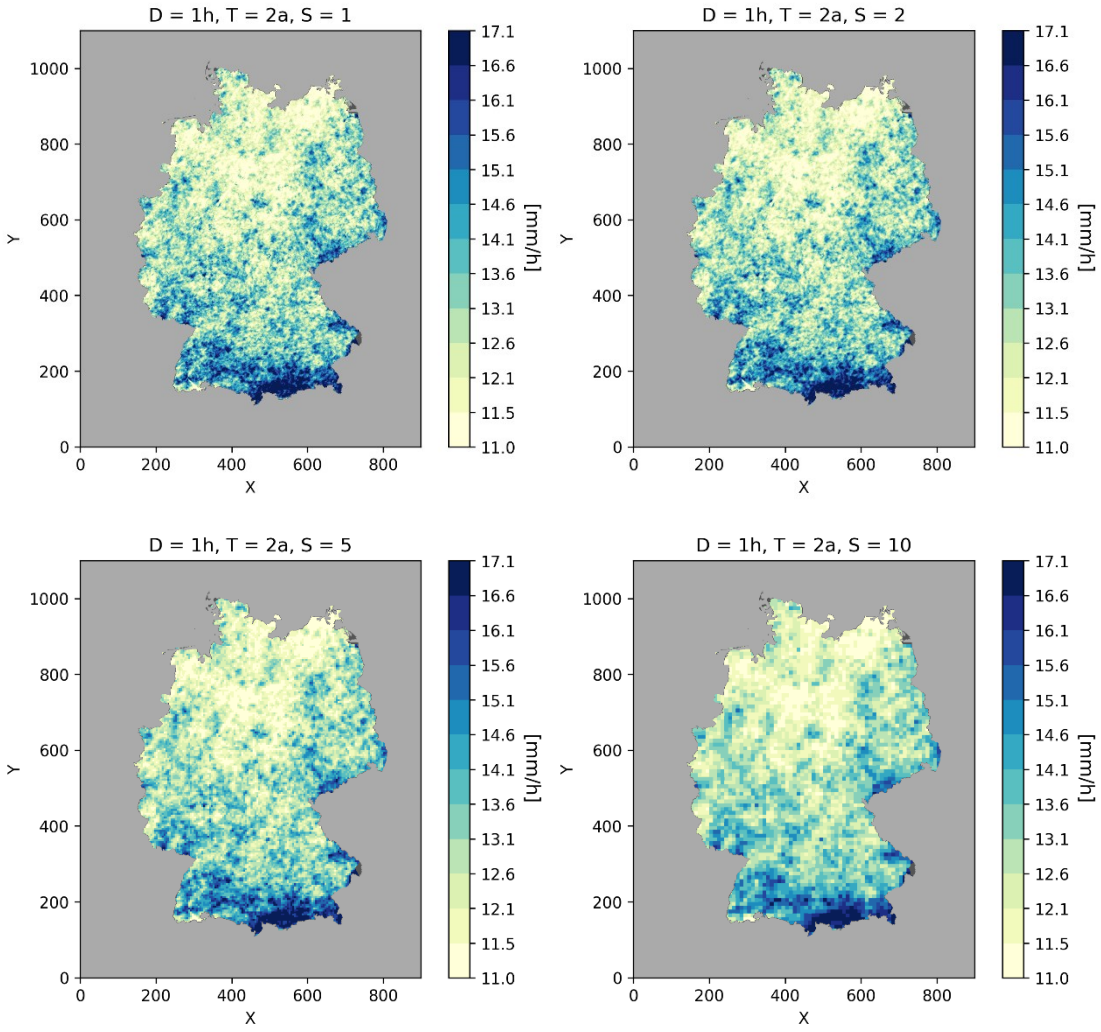

Fig 3-1 SMEV return levels based on radar data reduced to 1,2,5, 10 km spatial scale.

For each division, we considered the pixels corresponding to specified intervals in both the x and y directions (denoted as Δx and Δy) as forming individual blocks. This process was applied to both the radar grid and the station data. For the station data, the corresponding latitude and longitude coordinates of the stations in radar grid were provided by DWD. Using these coordinates, we assigned each station to a block on the radar grid based on its geographical location. Specifically, we grouped all stations located within a defined latitude and longitude of each region as one block (Fig 3-2 & Table 1).

This approach enabled us to compare the spatial variability at multiple scales and assess how the methods perform when capturing local variability. By subdividing the grid, we observed how spatial scale influences the analysis, whether the methods exhibit different levels of spatial variability at finer or coarser scales, and how they relate to varying orographic climatology.

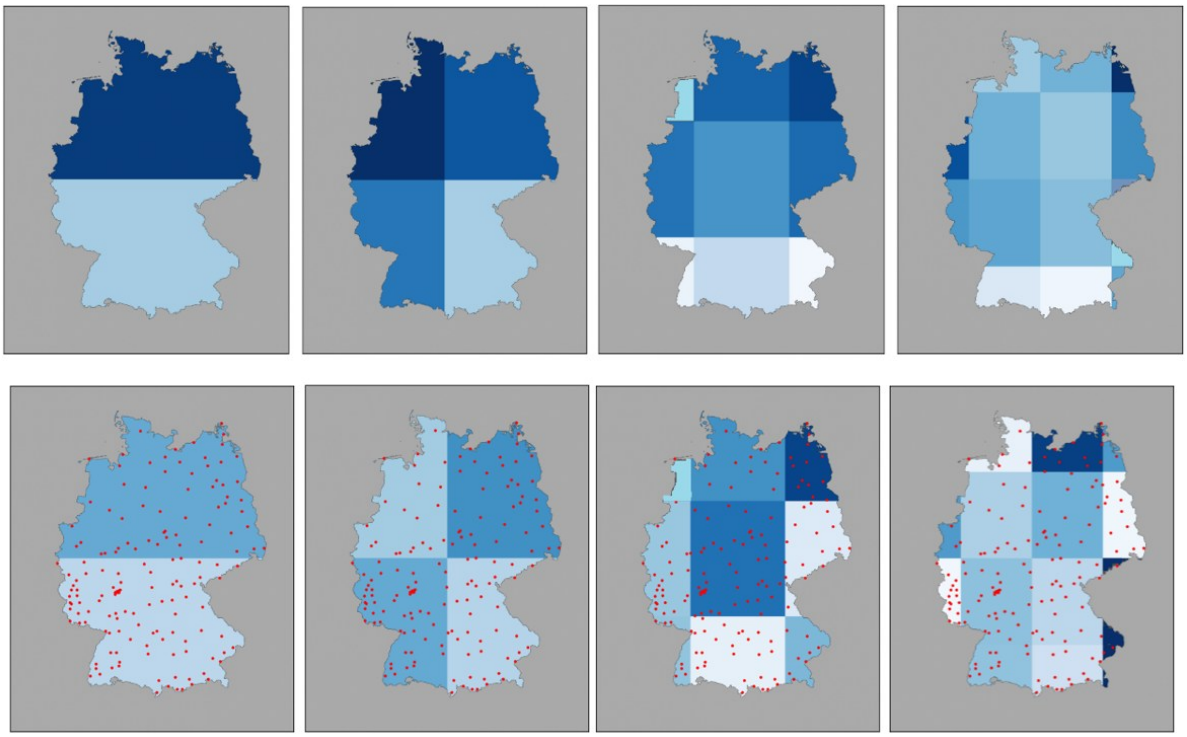

Fig 3-2 Radar grid (first row) and stations (second row) divided to two, four, nine and sixteen regions

Number of stations in each region

two regions

60
95

four regions

21	39
55	38

nine regions

1	13	13
25	44	16
7	28	8

sixteen regions

0	5	10	2
3	13	17	10
15	31	21	1
0	11	14	2

Table 1 Number of stations in each region

3.4 Quantification of Spatial Variability

The spatial variability in our analysis is quantified using the coefficient variation. The coefficient variation (CV) of a given set is defined as the ratio between the standard deviation relative to the mean and is calculated as:

$$CV = \frac{\sigma}{\mu}, \text{ where } \sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i-\mu)^2}{n}} \qquad Eq\ 12$$

Where σ is the standard deviation and μ is the mean, x_i is the i-th data point and n is the number of data points. A large CV for a set of precipitation values measured in different locations in space indicates that the spatial variability is large. The CV of the entire precipitation based on radar with different spatial scale was calculated with a 10*10 moving window. CV was calculated for the set of values based on their position in the 10*10 block. This means that each time the window was set to the dataset only values on one position was collected (Fig 3-3). For example, in the first instance, CV was calculated using all the data within the upper-left corner of the window. For a 1100×900 grid coarsened by 1, 2, 5, and 10 km, the number of CVs calculated per resolution was 100, 25, 5, and 1, respectively. In cases where multiple CV values were obtained, the 25th quantile, 75th quantile, and median were selected. This method was applied to the entire grid first, then to individual regions, enabling the calculation of spatial variability for each region. This approach was exclusive to the radar grid, as it was not feasible for the station data. The CV for the entire set of stations was first calculated using Eq 12, followed by regional calculations based on subsets of stations. For each region, CV was computed using the return level values from the stations within that region, resulting in a regional CV. This process was applied to both KOSTRA-DWD-2020 and SMEV estimates at the stations.

3.5 Quantification of the estimation uncertainty

To quantify the uncertainty in our analysis, we apply the bootstrapping method (Overeem et al., 2008) to the rain gauge datasets. Bootstrapping is a resampling technique that allows us to assess the variability or uncertainty of a dataset or model parameter by generating multiple simulated samples, known as "bootstrap samples," randomly drawn from the original data. The key idea is to create several new datasets, each representing a random sample of the original dataset, by sampling with replacement. This approach enables us to estimate the variability of model outputs without making strong parametric

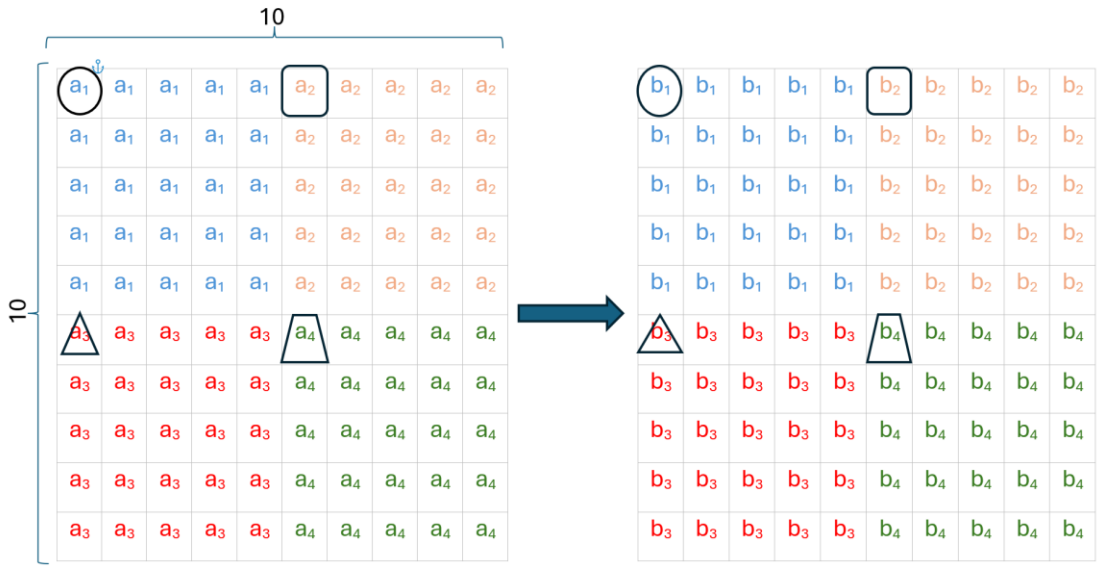

Fig 3-3 Example of Coefficient of Variation calculation using a 10×10 moving window on a grid coarsened to 5×5: each time the 10×10 window is applied to the grid, the parameters are captured based on their unique locations and stored in separate arrays. For instance, values with index 1 are stored in arr1, values with index 2 in arr2, and so forth. After scanning the entire grid, the coefficient of variation(CV) is computed for each array as detailed in Eq 12. This process results in four CV values. Subsequently, the 25th percentile, 75th percentile, and median are calculated for these CV values.

assumptions. For our analysis, we created 1,000 bootstrap samples from the original rain gauge dataset. In each bootstrap sample, individual data points are randomly selected from the original dataset, allowing for repetition of values. For each of the 1,000 bootstrap samples, we re-estimated the parameters of the Weibull distribution which is used to model the extreme rainfall events in our study. Using the re-estimated parameters from each bootstrap sample, we then re-estimated 1,000 different return level values for various durations and return periods. These 1,000 return level values represent different plausible realizations of rainfall for each combination of duration and return period, accounting for the uncertainty inherent in the estimation of extreme values. The distribution of these 1,000 return level estimates provides a quantifiable measure of uncertainty, allowing us to understand the variability in the estimates of extreme rainfall events under different conditions and assumptions.

Here, we used SMEV bootstrap values. Unfortunately, bootstrap values for KOSTRA-DWD-2020 were not available; however, information regarding the uncertainty of KOSTRA-DWD-2020 can be found in Shehu & Haberlandt (2023). CV was calculated for 1000 values per station across durations and return periods, yielding a CV per station. For each region, station CVs were averaged to obtain a regional CV.

4 Results and Discussions

4.1 Spatial variability across different scales

The first set of results compares the spatial variability of precipitation return level derived from radar data at different spatial scales (1, 2, 5, and 10 km) with estimates based on data from 155 stations. To ensure the robustness of the findings, the estimation uncertainty of SMEV estimates from the station data was also calculated. For this analysis, we assume that the station data has lower uncertainty and can serve as the benchmark for comparing the radar-based estimates. Our assumption is motivated by the fact that stations have longer records with respect to the radar.

Fig 4-1 illustrates the spatial variability of estimates for durations of 1, 6, and 24 hours and return periods of 5, 20, and 100 years. Estimation uncertainty is reported (pink line) for reference. In all plots, the estimation uncertainty is consistently the lowest, meaning that the reported spatial variability is not due to estimation uncertainty only and thus supporting the validity of the comparison. The overall trend between radar estimates shows that the spatial variability derived from radar data using SMEV (orange) is higher than that derived using KOSTRA (green), although no consistent trend is observed in their station data.

As the return periods increase (moving horizontally in each row, Fig 4-1), the estimation uncertainty (pink line) in the estimated values also rises, with the highest uncertainty observed for the 1-hour duration at the 100-year return period. This aligns with the expectation that extreme events associated with higher return periods are more difficult to estimate accurately due to their rarity and limited data.

Moreover, when radar data is aggregated into coarser spatial scales (e.g., 2, 5, and 10 km), the spatial variability of radar data (represented by the orange and green) decreases. The decrease in spatial variability is more pronounced for shorter durations, likely because finer-scale temporal events, such as hourly rainfall, exhibit more localized variability that is smoothed out averaging the spatial scale. In contrast, for longer durations, the reduction in variability due to different spatial scales is less noticeable, as these events are typically more spatially uniform and less further affected by spatial aggregation. High-resolution radar data captures localized extreme rainfall events within smaller grid cells. Lower spatial scale (coarser resolutions), however, average rainfall over larger areas, blending localized extremes with surrounding lower-intensity rainfall

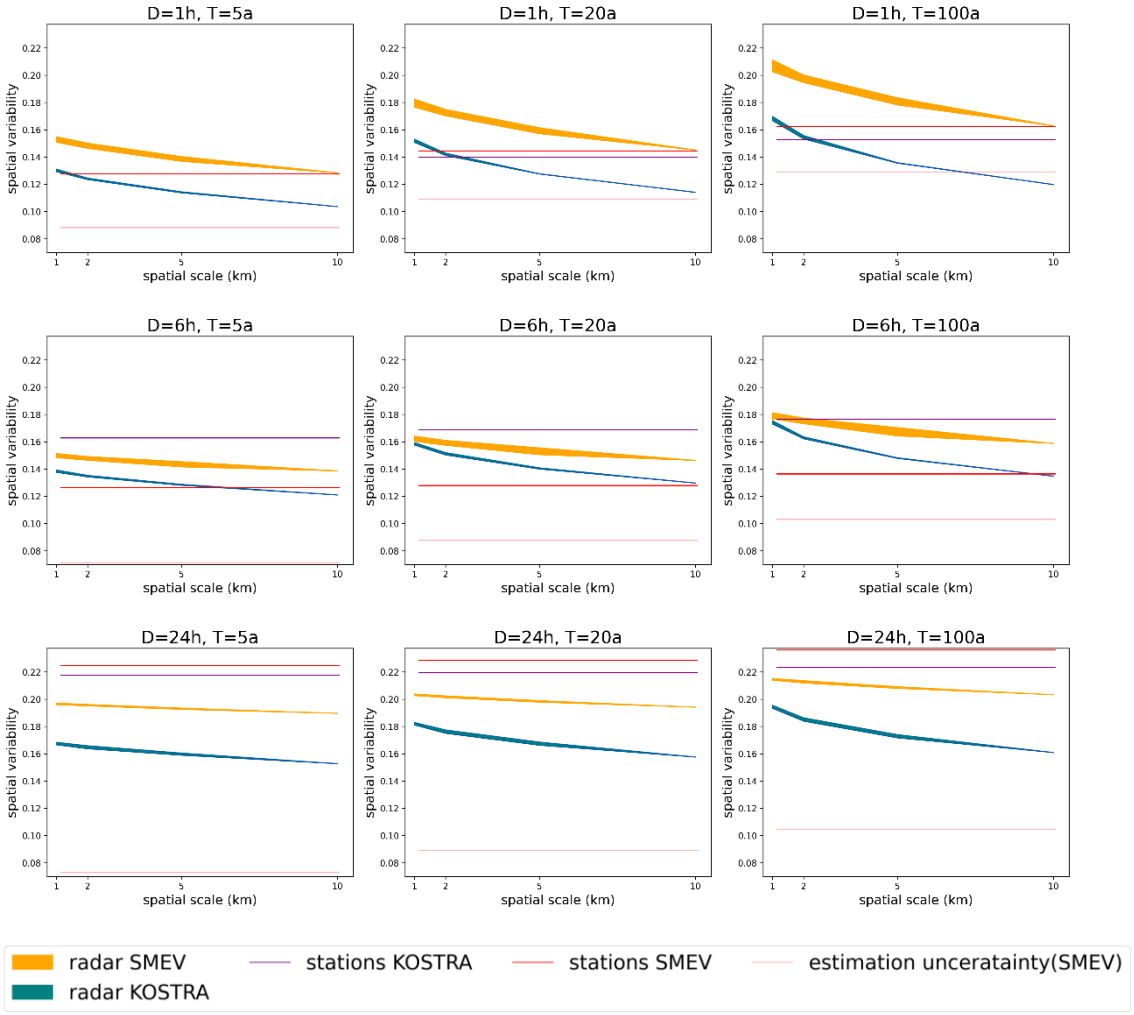

Fig 4-1 Spatial Variability for a 1, 6, 24-Hour Duration and a 5, 20, 100-Year Return Period: based on SMEV and KOSTRA-DWD-2020 estimates applied on radar data with spatial cales of 1 km, 5 km, and 10 km with their 25th and 75th quantiles; based on SMEV and KOSTRA-DWD-2020 estimates from 155 stations for each region; Estimation uncertainty of SMEV estimates from 155 stations.

and thus reducing variability (Marra et al., 2017). This is expected, because extreme precipitation events on the point scale are necessarily greater than (or at most equal to) extreme areal averages. This aligns with findings by Rosin et al. (2024), who observed that short-duration precipitation intensity decreases with increasing area size due to the localized nature of extreme events. For longer durations, differences across areas diminish as multiple rain cells accumulate, averaging spatial variability and producing more homogeneous precipitation patterns.

Comparing the radar with stations, SMEV estimates from radar data (orange) exhibit higher spatial variability for shorter durations (hourly and 6-hour) across all spatial scales and return periods compared to station data (red line). However, for daily durations, this

trend reverses, with radar estimates (orange) showing lower variability than station data (red line).

On the other hand, KOSTRA estimates based on radar data (green line) tend to underestimate the spatial variability of station data (purple line) for hourly durations, especially at lower spatial scales. As the duration increases to 6 hours, the trend fluctuates across different return periods and spatial scales. For daily durations, the trend aligns with that observed for SMEV, with radar estimates (green) showing lower variability than station data (purple line).

The trend observed at shorter durations suggests that radar effectively captures fine-scale variability due to its higher spatial resolution. In contrast, at longer durations, the smoothing effect of larger grid cells reduces radar-based variability, making it lower than that observed in station data.

The discrepancies observed in the hourly results from KOSTRA, when applied to radar data at a higher spatial scale, may be due to estimation uncertainty. However, since KOSTRA radar consistently shows less variability than the station data, it is unlikely that these discrepancies are caused by random errors all aligning in the same direction. Therefore, the lower variability likely reflects a true characteristic of the estimates rather than method uncertainty.

The discrepancies may stem from how KOSTRA handles extreme precipitation data. KOSTRA employs a fixed Koutsoyiannis shape parameter, which might not effectively capture the fine-scale variability of precipitation, particularly for extreme events (e.g., see Marra et al., 2021). This shape parameter governs the tail behavior of the distribution, influencing how rapidly the probabilities of extreme rainfall events decrease as rainfall intensity increases. Relying on a single fixed shape parameter, rather than estimating it for each radar grid cell or duration, can result in the underestimation of extreme rainfall events, especially for shorter return periods, and may fail to represent the true spatial variation in the tail behavior of extreme rainfall distributions across different regions (e.g., see Marra et al., 2022.

The treatment of extremes in the KOSTRA methodology involves smoothing or generalized assumptions that reduce variability. In contrast, SMEV's approach of focusing on ordinary events and their maximum intensities allows for representation of spatial variability across different durations, because no assumptions on the uniform tail

behavior are necessary. By directly relating extreme rainfall to storm characteristics, SMEV avoids the potential pitfalls associated with disaggregation and fixed parameters, leading to results that better represent the variability, especially for hourly data where capturing the spatial dynamics of convective storms is crucial.

4.2 Regional gradients

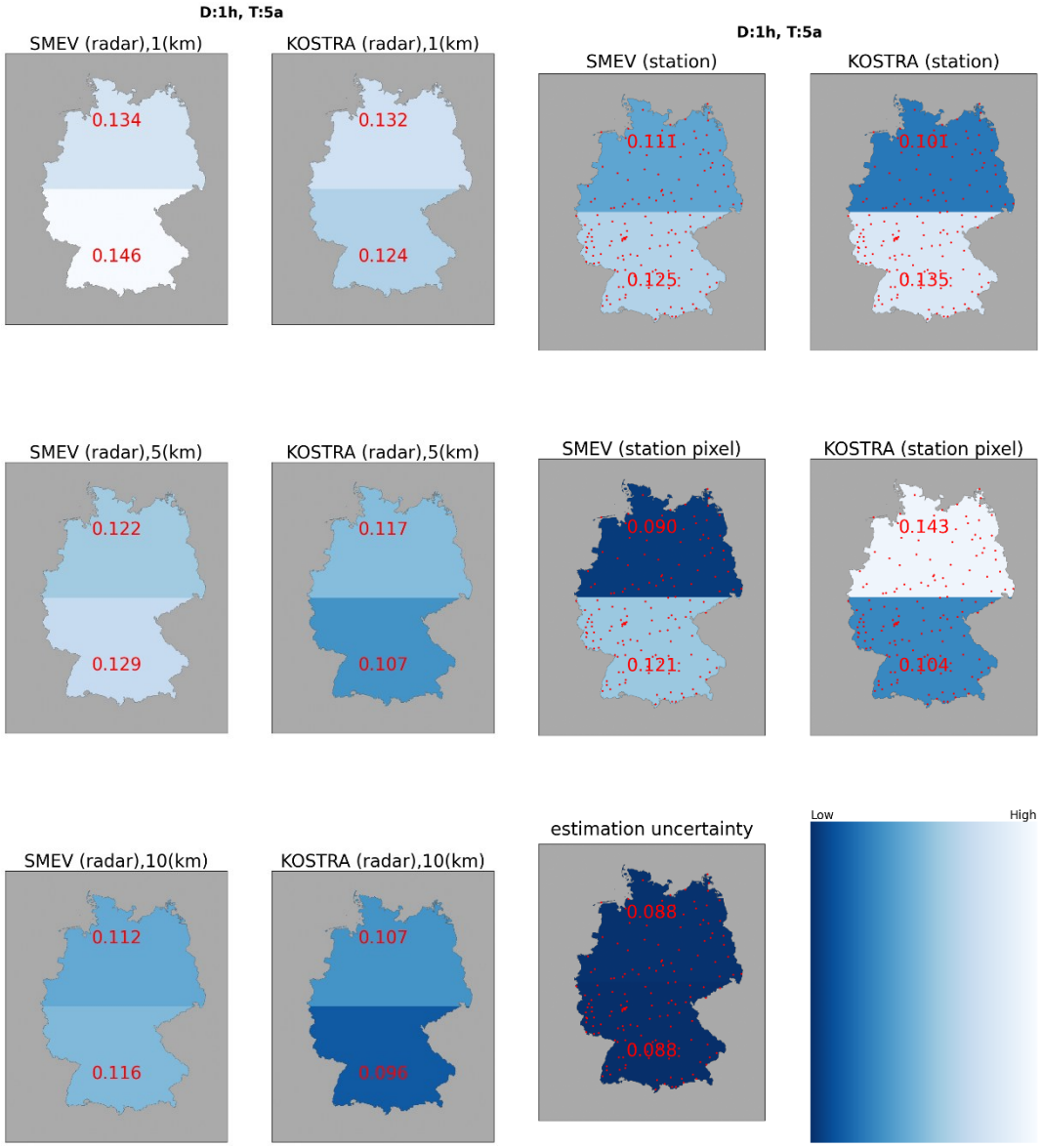

Fig 4-2 Spatial Variability for a 1-Hour Duration and a 5-Year Return Period: (On the left) based on SMEV and KOSTRA-DWD-2020 estimates applied on radar data with spatial cales of 1 km (first row), 5 km (second row), and 10 km (third row), divided into two regions. (on the right) based on SMEV and KOSTRA-DWD-2020 estimates from 155 stations for each region and pixels corresponding to station locations. Estimation uncertainty of SMEV estimates from 155 stations, with the corresponding color scale shown in the last column.

In this section, we compare spatial variability across different region sizes derived from radar and rain gauge data. The estimation uncertainty, indicated by darker shading in the plots, allows for clearer comparisons. As in previous sections, we assume that the station data has lower uncertainty and can serve as the benchmark for comparing radar-

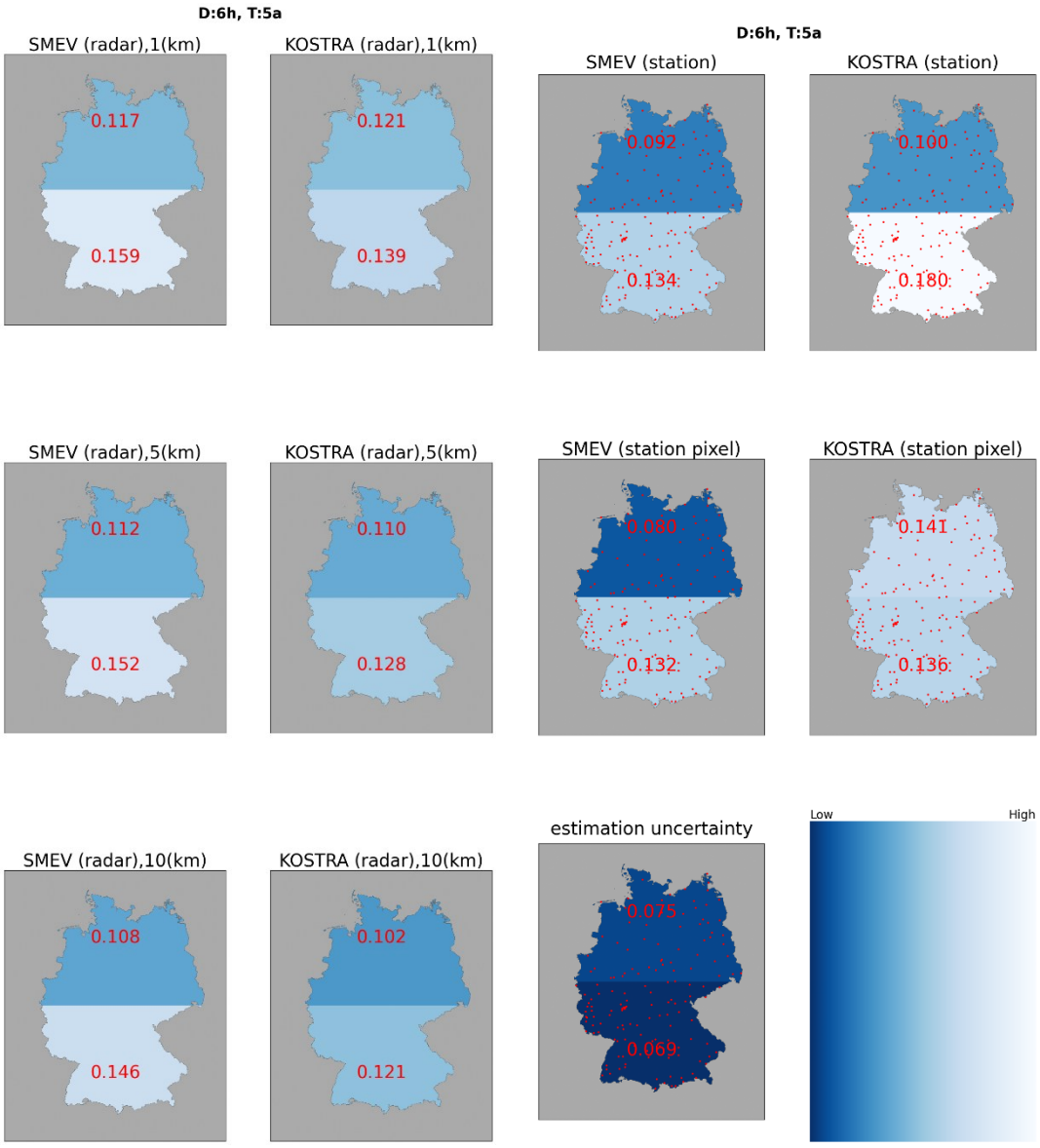

Fig 4-3 Spatial Variability for a 6-Hour Duration and a 5-Year Return Period: (On the left) based on SMEV and KOSTRA-DWD-2020 estimates applied on radar data with spatial cales of 1 km (first row), 5 km (second row), and 10 km (third row), divided into two regions. (on the right) based on SMEV and KOSTRA-DWD-2020 estimates from 155 stations for each region and pixels corresponding to station locations. Estimation uncertainty of SMEV estimates from 155 stations, with the corresponding color scale shown in the last column.

based estimates, although this assumption may not always hold true due to the incomplete spatial sampling offered by stations.

The figures comparing two regions (Fig 4-2 to Fig 4-9) illustrate distinct regional differences in spatial variability and estimation uncertainty. For both KOSTRA and SMEV, station data consistently show higher variability in the southern region and lower variability in the northern region.

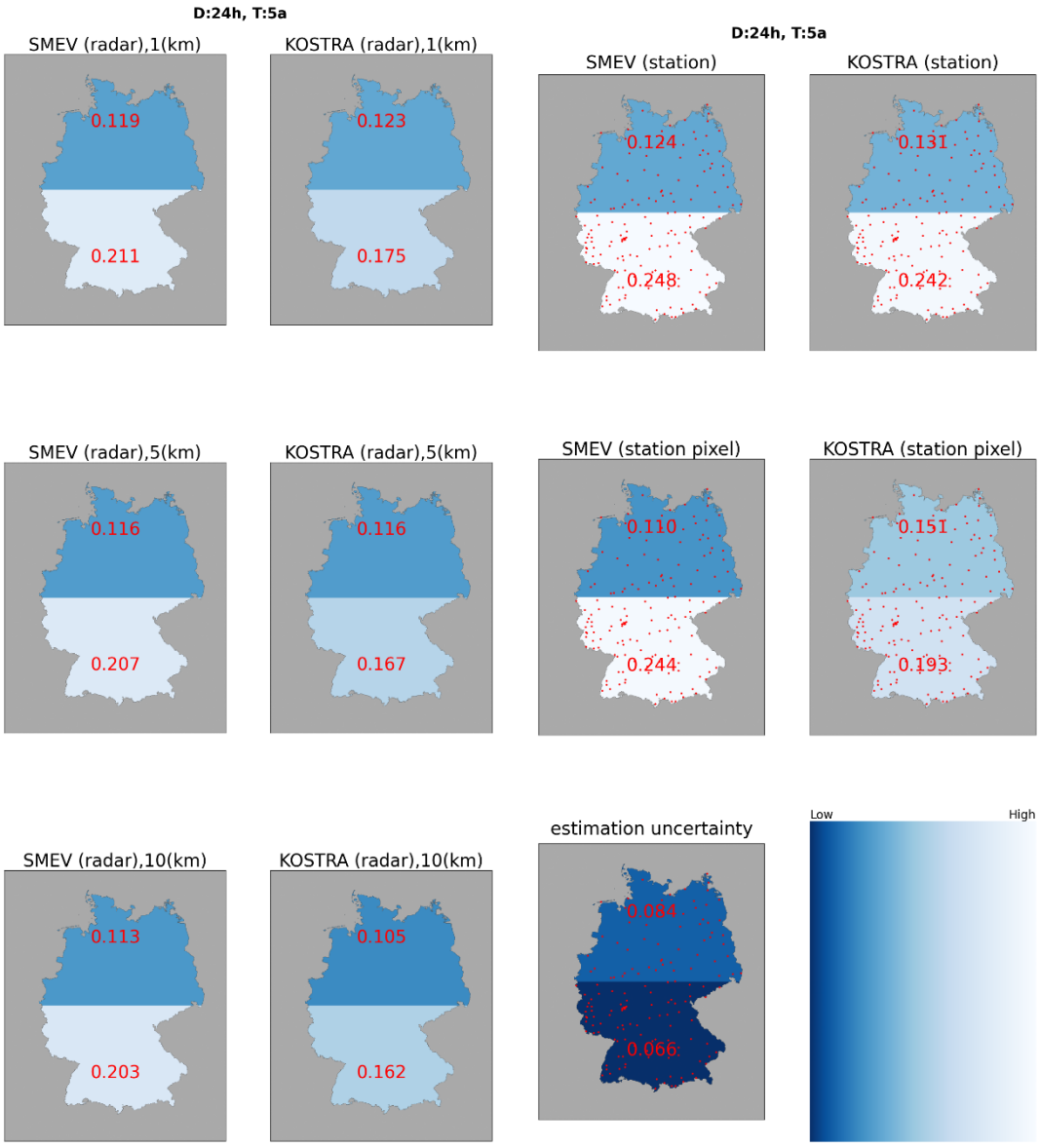

Fig 4-4 Spatial Variability for a 24-Hour Duration and a 5-Year Return Period: (On the left) based on SMEV and KOSTRA-DWD-2020 estimates applied on radar data with spatial cales of 1 km (first row), 5 km (second row), and 10 km (third row), divided into two regions. (on the right) based on SMEV and KOSTRA-DWD-2020 estimates from 155 stations for each region and pixels corresponding to station locations. Estimation uncertainty of SMEV estimates from 155 stations, with the corresponding color scale shown in the last column.

Fig 4-2 (hourly duration with a 5-year return period) shows that, similarly to the stations, SMEV estimates, derived from both radar and the pixels corresponding to the stations, show higher spatial variability in the southern region compared to the north. In contrast, KOSTRA estimates based on radar data, as well as radar pixels corresponding to the stations, exhibit an opposite trend, with higher variability observed in the northern region and lower variability in the southern region.

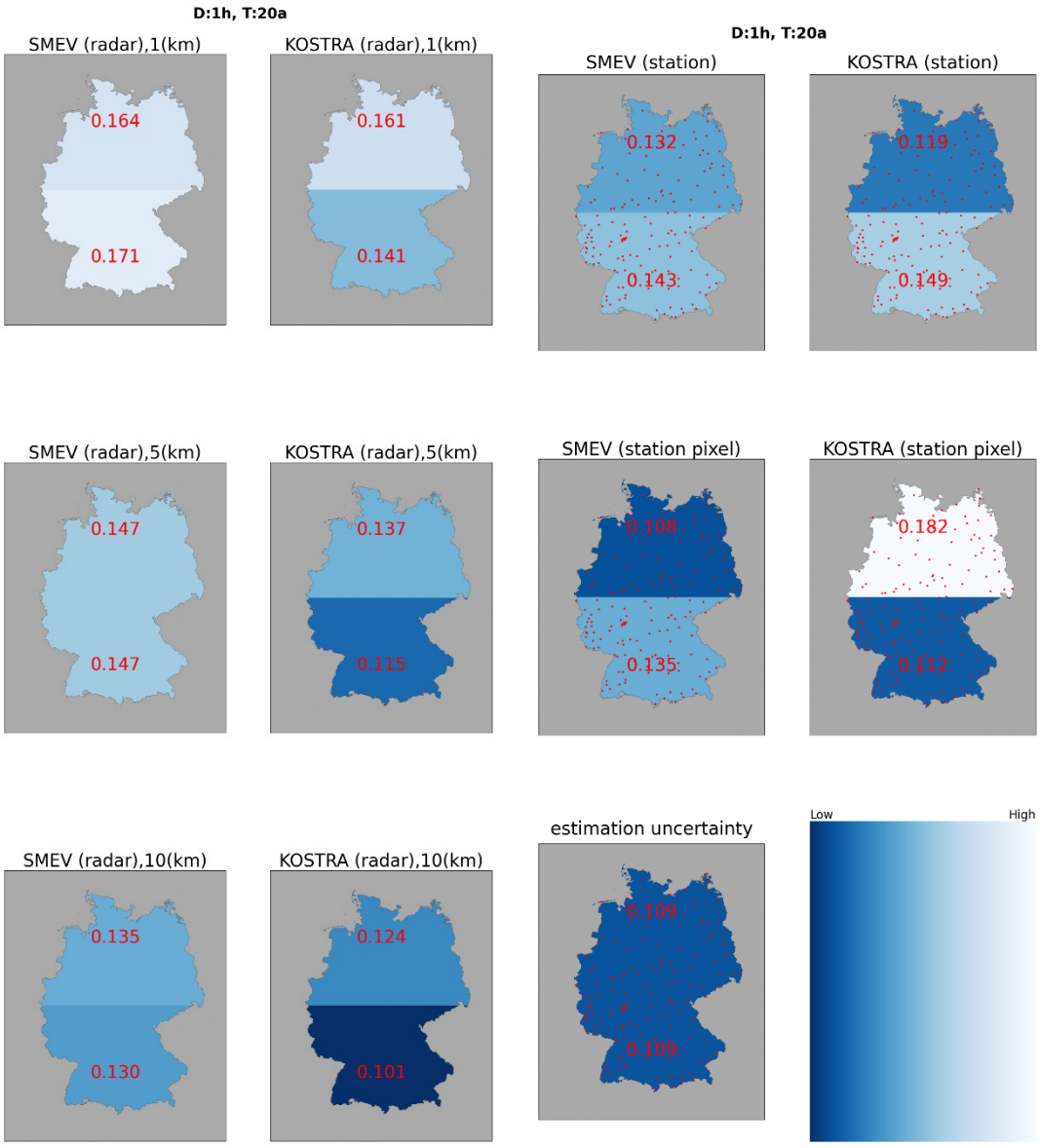

Fig 4-5 Spatial Variability for a 1-Hour Duration and a 20-Year Return Period: (On the left) based on SMEV and KOSTRA-DWD-2020 estimates applied on radar data with spatial cales of 1 km (first row), 5 km (second row), and 10 km (third row), divided into two regions. (on the right) based on SMEV and KOSTRA-DWD-2020 estimates from 155 stations for each region and pixels corresponding to station locations. Estimation uncertainty of SMEV estimates from 155 stations, with the corresponding color scale shown in the last column.

As the duration increases to 6 hours and daily (Fig 4-3 & Fig 4-4), all estimates, including KOSTRA, align in showing higher spatial variability in the southern region and lower variability in the north. However, an exception is observed for KOSTRA on the pixels corresponding to the stations, which still shows the opposite trend for the 6-hour duration.

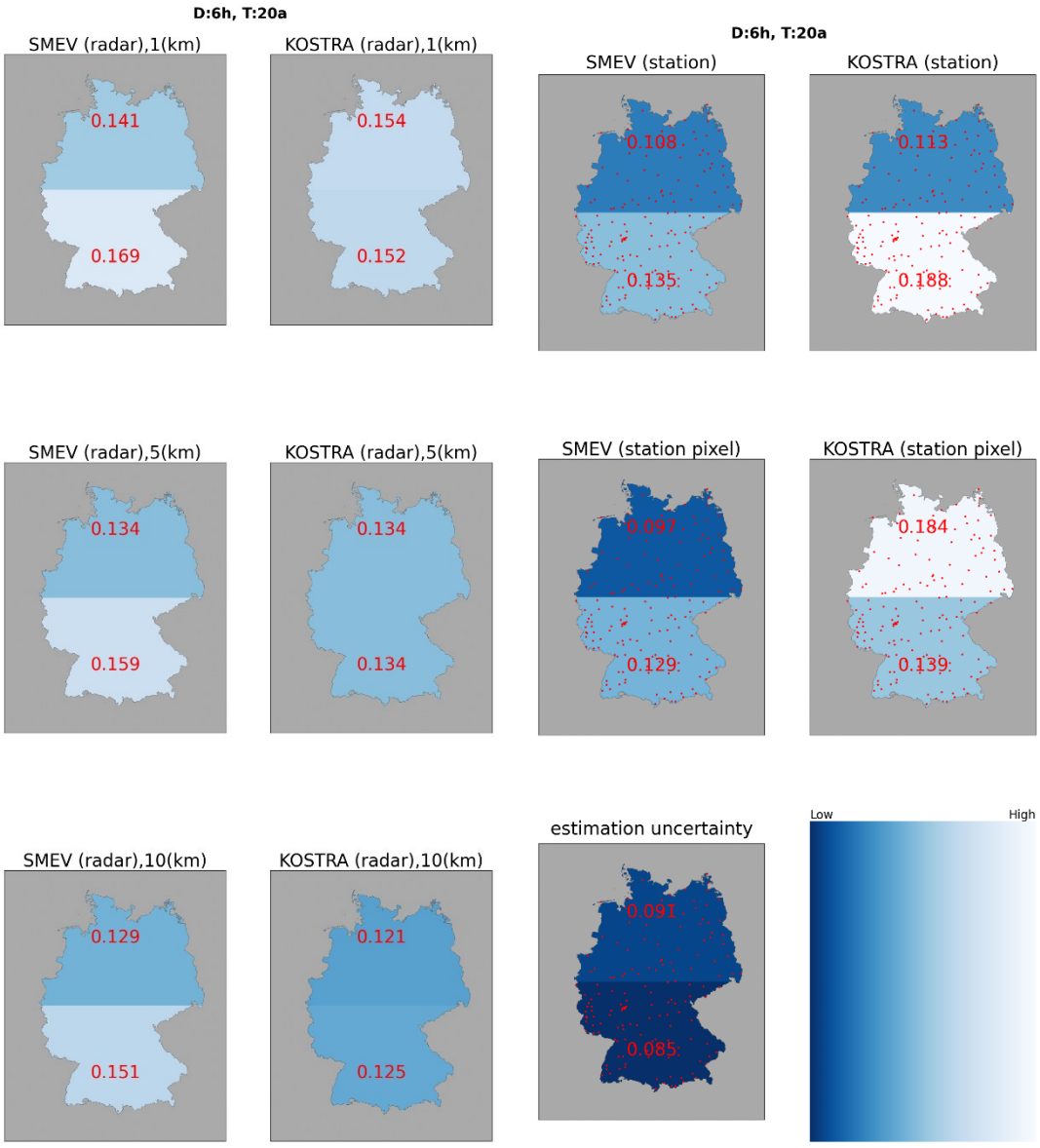

Fig 4-6 Spatial Variability for a 6-Hour Duration and a 20-Year Return Period: (On the left) based on SMEV and KOSTRA-DWD-2020 estimates applied on radar data with spatial cales of 1 km (first row), 5 km (second row), and 10 km (third row), divided into two regions. (on the right) based on SMEV and KOSTRA-DWD-2020 estimates from 155 stations for each region and pixels corresponding to station locations. Estimation uncertainty of SMEV estimates from 155 stations, with the corresponding color scale shown in the last column.

As the return period increases (from 20 years to 100 years in Fig 4-5 & Fig 4-7), SMEV radar-based estimates for lower spatial scales begin to show higher variability in the north, resembling the trend observed in KOSTRA radar-based estimates. However, SMEV estimates derived from station pixels maintain the original pattern, with greater variability in the southern region compared to the north.

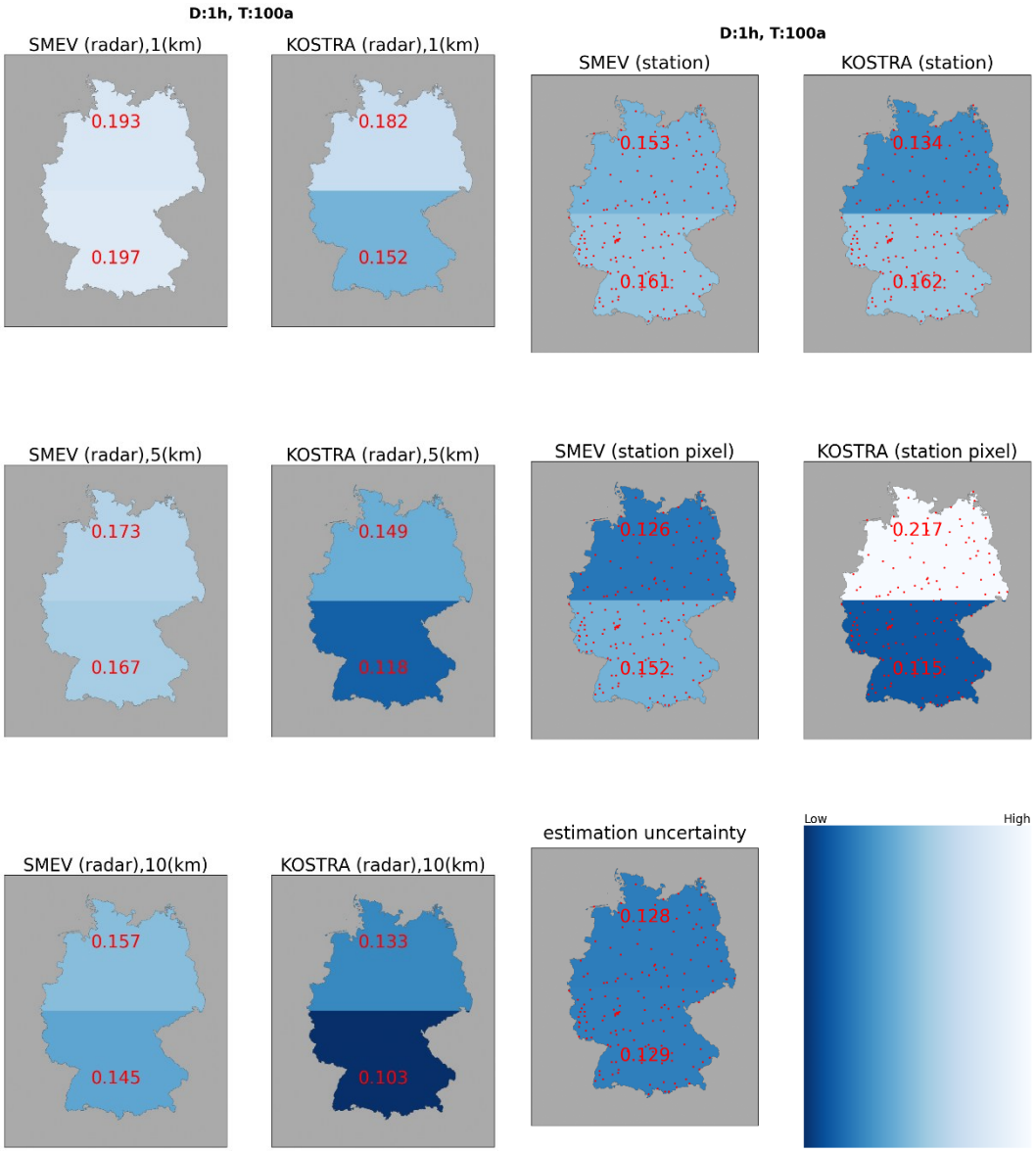

Fig 4-7 Spatial Variability for a 1-Hour Duration and a 100-Year Return Period: (On the left) based on SMEV and KOSTRA-DWD-2020 estimates applied on radar data with spatial cales of 1 km (first row), 5 km (second row), and 10 km (third row), divided into two regions. (on the right) based on SMEV and KOSTRA-DWD-2020 estimates from 155 stations for each region and pixels corresponding to station locations. Estimation uncertainty of SMEV estimates from 155 stations, with the corresponding color scale shown in the last column.

For longer durations and higher return periods (e.g., Fig 4-9 with daily duration and 100-year return period), all methods show higher spatial variability in the southern region compared to the northern region. The exception again is KOSTRA on pixels corresponding to stations, where the trend is reversed, with higher spatial variability in the northern region than in the south.

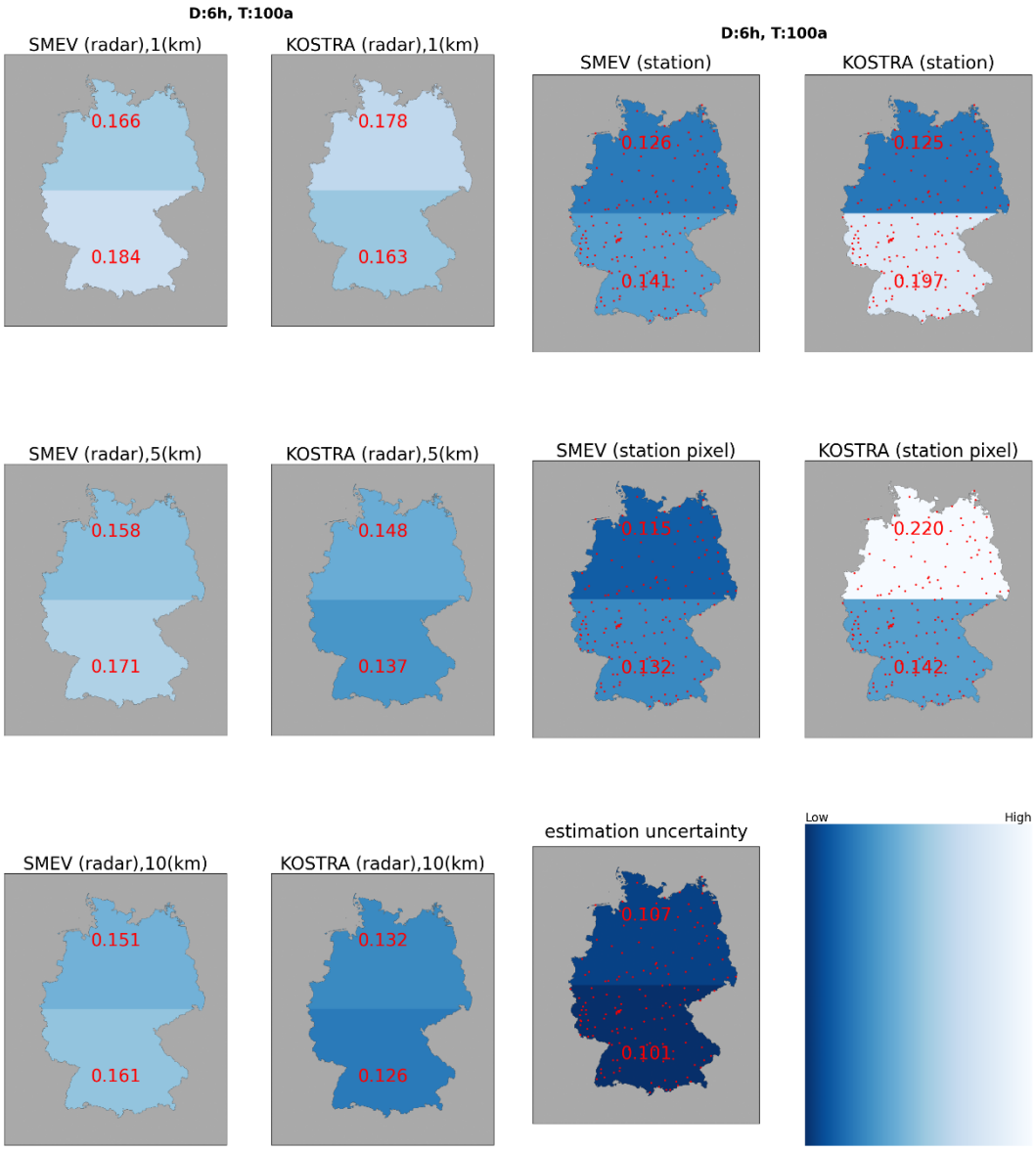

Fig 4-8 Spatial Variability for a 6-Hour Duration and a 100-Year Return Period: (On the left) based on SMEV and KOSTRA-DWD-2020 estimates applied on radar data with spatial cales of 1 km (first row), 5 km (second row), and 10 km (third row), divided into two regions. (on the right) based on SMEV and KOSTRA-DWD-2020 estimates from 155 stations for each region and pixels corresponding to station locations. Estimation uncertainty of SMEV estimates from 155 stations, with the corresponding color scale shown in the last column.

It is also observed that for higher durations and return periods, radar data show higher variability than the station data in the northern part, which is the opposite of the overall trend observed in Section 4.1. where radar data typically showed lower variability than station data. In theory, for longer durations, radar data is expected to exhibit lower variability than station data due to the spatial smoothing effect at larger scales. However, in the northern region, the radar data seem to capture more variability.

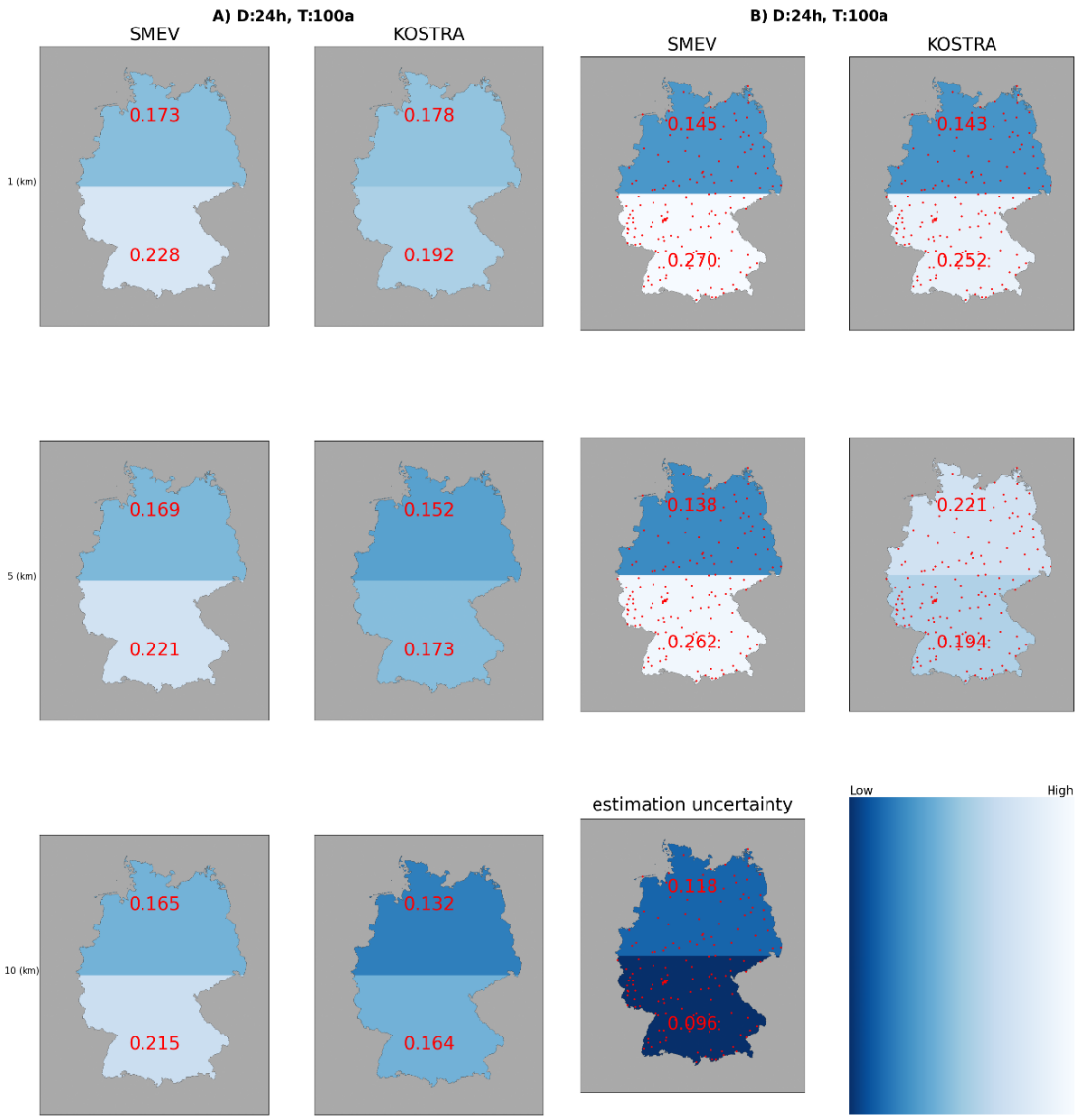

Fig 4-9 Spatial Variability for a 24-Hour Duration and a 100-Year Return Period: (On the left) based on SMEV and KOSTRA-DWD-2020 estimates applied on radar data with spatial cales of 1 km (first row), 5 km (second row), and 10 km (third row), divided into two regions. (on the right) based on SMEV and KOSTRA-DWD-2020 estimates from 155 stations for each region and pixels corresponding to station locations. Estimation uncertainty of SMEV estimates from 155 stations, with the corresponding color scale shown in the last column.

It is also noteworthy that for a constant duration, the alignment between KOSTRA radar data and station data varies with return period. In Fig 4-3, for a 5-year return period, KOSTRA radar data aligns well with station data. However, in Fig 4-6, for a 20-year return period, only KOSTRA radar data at lower spatial scales aligns with the station data. In Fig 4-8, for a 100-year return period, KOSTRA radar data does not align with station data at all. This indicates that estimates based on radar data are more consistent with station data for shorter return periods, but for longer return periods, the discrepancy

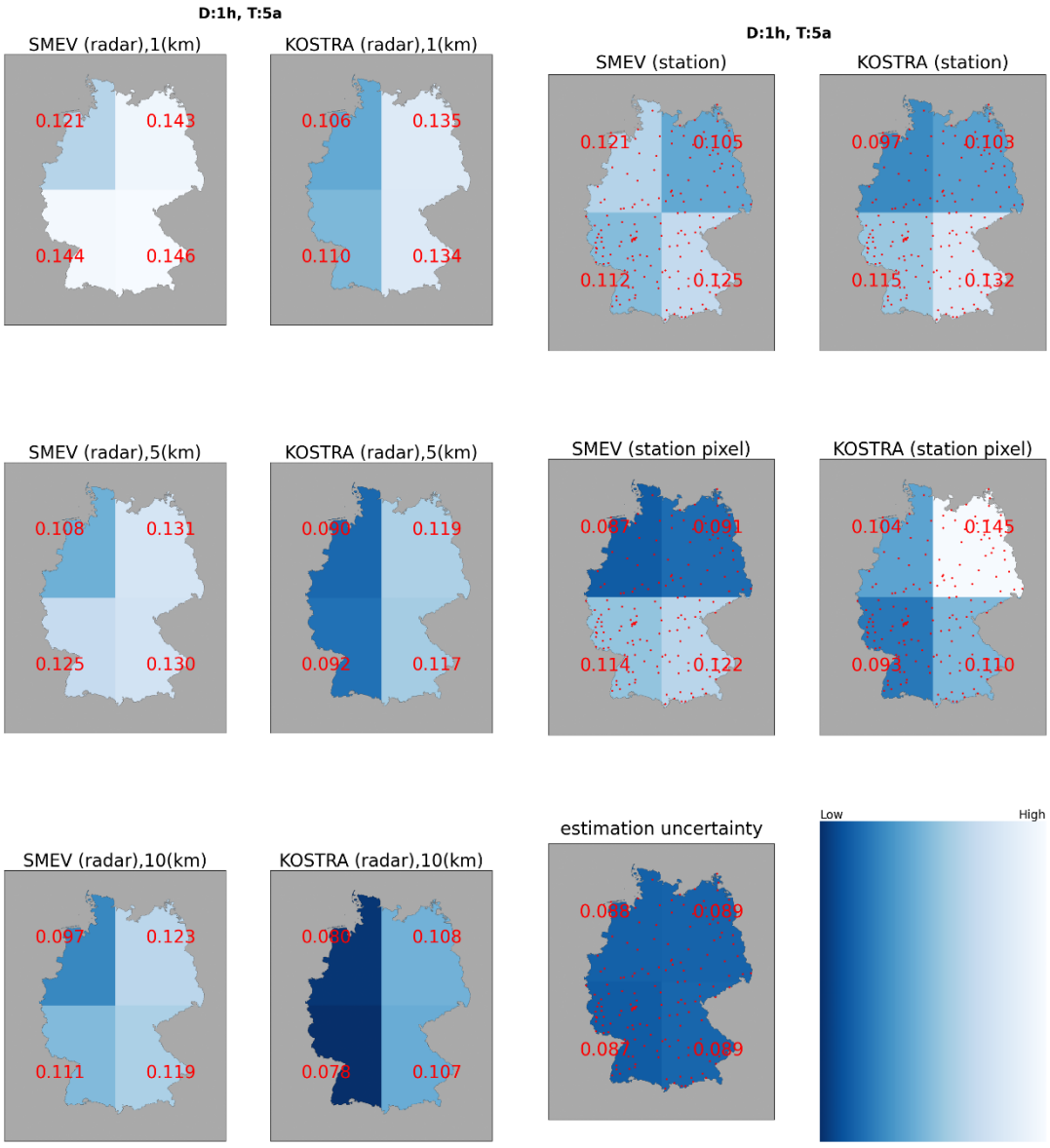

Fig 4-10 Spatial Variability for a 1-Hour Duration and a 5-Year Return Period: (On the left) based on SMEV and KOSTRA-DWD-2020 estimates applied on radar data with spatial cales of 1 km (first row), 5 km (second row), and 10 km (third row), divided into four regions. (on the right) based on SMEV and KOSTRA-DWD-2020 estimates from 155 stations for each region and pixels corresponding to station locations. Estimation uncertainty of SMEV estimates from 155 stations, with the corresponding color scale shown in the last column.

between the radar and station data becomes more pronounced. More precisely, the best alignment is observed for longer durations and lower return periods, suggesting that radar data performs better for capturing spatial variability in these conditions.

Reducing the region size from two divisions to four (Fig 4-10 to Fig 4-12) reveals more detailed spatial patterns of variability, shedding light on region-specific trends. For hourly durations and a 5-year return period (Fig 4-10), the only consistent trend between

Fig 4-11 Spatial Variability for a 24-Hour Duration and a 5-Year Return Period: (On the left) based on SMEV and KOSTRA-DWD-2020 estimates applied on radar data with spatial cales of 1 km (first row), 5 km (second row), and 10 km (third row), divided into four regions. (on the right) based on SMEV and KOSTRA-DWD-2020 estimates from 155 stations for each region and pixels corresponding to station locations. Estimation uncertainty of SMEV estimates from 155 stations, with the corresponding color scale shown in the last column.

radar and station data emerges in the Alps in southeast. SMEV-based station data exhibit higher variability in the southeast and northwest, while KOSTRA-based station data show elevated variability in the southeast and southwest. In contrast, radar estimates from both methods display a west-east gradient, with lower variability in the west and higher values in the south. For daily durations with the same return period (5-year, Fig 4-11), the patterns align with those observed in the two-region plots, showing a clear north-south gradient across all estimates. This consistency reinforces the observation that estimates

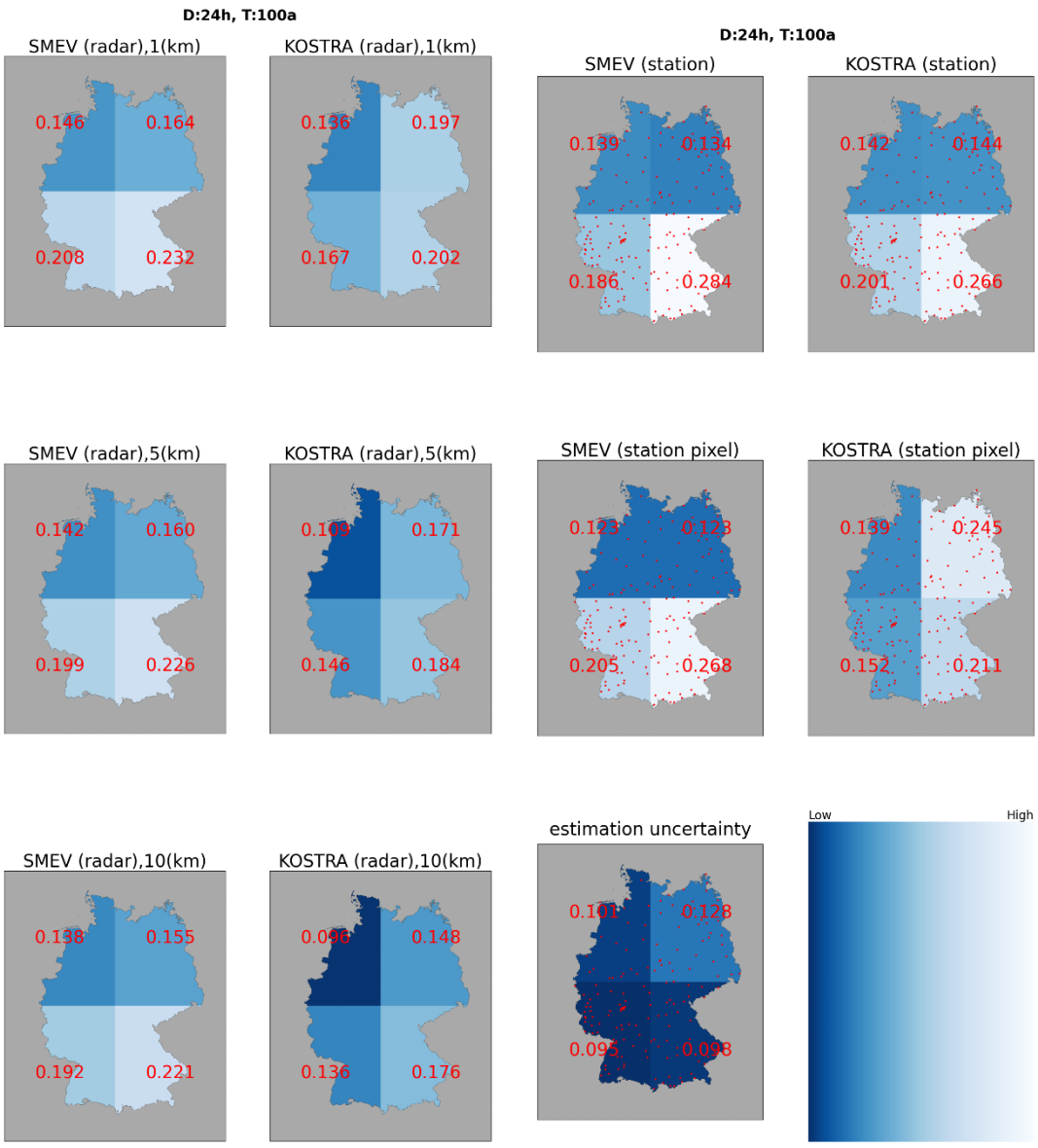

Fig 4-12 Spatial Variability for a 24-Hour Duration and a 100-Year Return Period: (On the left) based on SMEV and KOSTRA-DWD-2020 estimates applied on radar data with spatial cales of 1 km (first row), 5 km (second row), and 10 km (third row), divided into four regions. (on the right) based on SMEV and KOSTRA-DWD-2020 estimates from 155 stations for each region and pixels corresponding to station locations. Estimation uncertainty of SMEV estimates from 155 stations, with the corresponding color scale shown in the last column.

from daily durations capture broader spatial gradients more effectively, with station and radar data showing better agreement.

However, for daily durations with a higher return period (Fig 4-12), a divergence becomes evident. While SMEV shows a weak west-east gradient, its stronger north-south gradient allows it to align more closely with station-derived variability, whereas

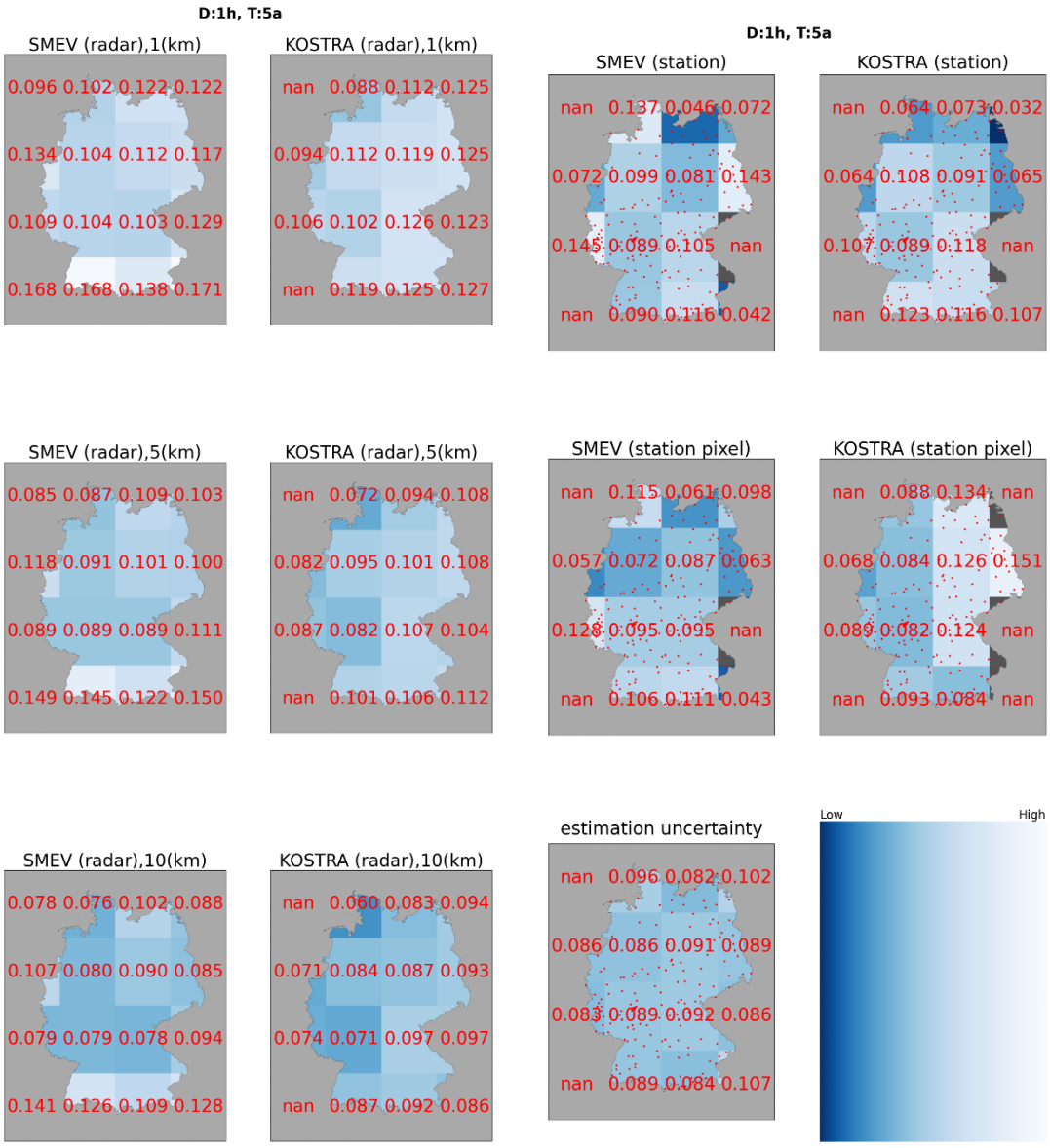

Fig 4-13 Spatial Variability for a 1-Hour Duration and a 5-Year Return Period: (On the left) based on SMEV and KOSTRA-DWD-2020 estimates applied on radar data with spatial cales of 1 km (first row), 5 km (second row), and 10 km (third row), divided into sixteen regions. (on the right) based on SMEV and KOSTRA-DWD-2020 estimates from 155 stations for each region and pixels corresponding to station locations. Estimation uncertainty of SMEV estimates from 155 stations, with the corresponding color scale shown in the last column.

KOSTRA radar estimates persist in showing a dominant west-east gradient, deviating from the north-south pattern observed in the station data. This suggests that, at higher return periods, KOSTRA radar data may be less effective in capturing regional variability patterns and might be influenced by factors unrelated to the observed station-based trends.

In the next set of plots, focusing on 16 regions (Fig 4-13 & Fig 4-14) , the Alps become prominently visible in both radar and station data, with their spatial variability become

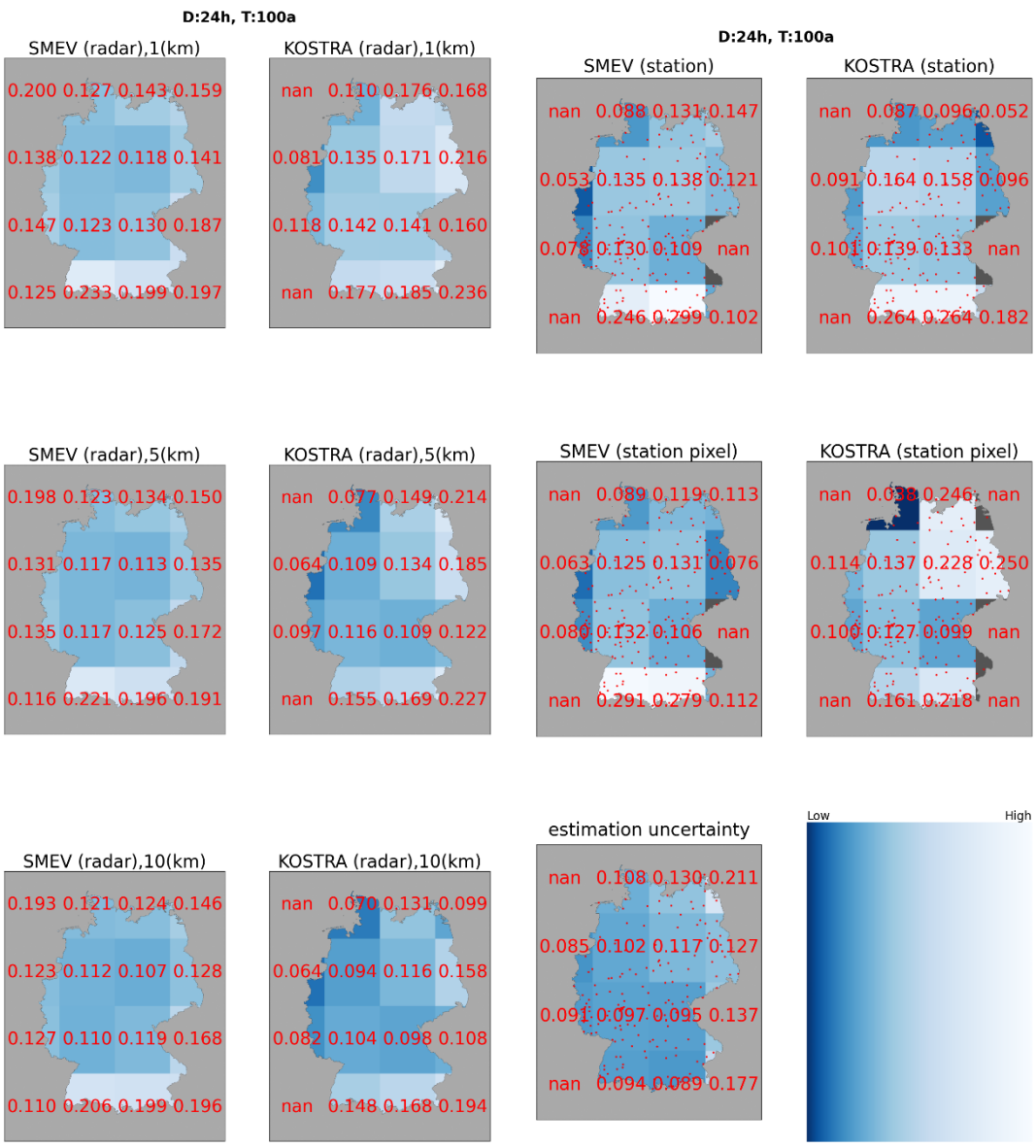

Fig 4-14 Spatial Variability for a 24-Hour Duration and a 100-Year Return Period: (On the left) based on SMEV and KOSTRA-DWD-2020 estimates applied on radar data with spatial cales of 1 km (first row), 5 km (second row), and 10 km (third row), divided into sixteen regions. (on the right) based on SMEV and KOSTRA-DWD-2020 estimates from 155 stations for each region and pixels corresponding to station locations. Estimation uncertainty of SMEV estimates from 155 stations, with the corresponding color scale shown in the last column.

more clearly represented in the daily duration and higher return periods (Fig 4-14). This enhanced visualization underscores the ability of these estimates to capture the intricate orographic effects of the Alps, even for longer durations, higher return periods, smaller regions, fewer stations, and shorter radar record lengths.

The observed pattern of greater spatial variability in southern Germany aligns with expectations and is attributed to the influence of the Alps, which significantly enhance

precipitation through orographic lifting. The duration dependency of spatial variability is well-documented in previous studies (Lengfeld et al., 2019; Shehu & Haberlandt, 2023; Marra et al., 2017; Peleg et al., 2018). For shorter durations, the alignment between radar and station-based estimates tends to be more challenging. Short-duration precipitation events, such as hourly rainfall, are often more localized and influenced by fine-scale variations in the atmosphere and topography. For longer durations, it is generally expected that radar and station-based estimates may align better. This is because longer-duration precipitation events tend to have more averaged-out patterns, reducing fine-scale variability. As a result, both radar and station data may show more consistent trends, leading to improved alignment.

Lengfeld et al. (2019) findings support this obscervation by illustrating that the relationship between precipitation intensity and topography varies with rainfall duration. For daily precipitation, a strong correlation with topographical features is evident and this emphasizes the significant influence of elevation on daily precipitation patterns. Winterrath et al. (2019) also found that the distribution of short-lasting extreme precipitation events is not strictly related to topography. They found that for short-term precipitation events, such as those occurring within an hour, the influence of topography diminishes and they are more spatially localized and less dependent on terrain. This localized nature of short-duration events highlights the challenges in accurately capturing spatial gradients, particularly at lower spatial scale. Pöschmann et al., (2023) also differentiated between the mechanism responsible for short and long duration rainfall; convective and advective respectively. Convective rainfall patterns showed higher maximum values but smaller spatial extent, while advective rainfall exhibited lower maximums but covered larger areas.

The bias between radar and station estimates for shorter durations appears more pronounced in KOSTRA radar-based estimates. This may be due to the use of the Koutsoyiannis parameter on radar data, which assume a uniform spatial structure, explained in Section 4.1. These assumptions may not fully capture the complex topographic effects in regions like southern Germany, where the terrain, especially the Alps, can significantly influence precipitation patterns through orographic lifting. As a result, KOSTRA radar estimates may not reflect the finer-scale variability seen in station data, particularly in areas with pronounced elevation-driven precipitation patterns.

When discussing smaller regions, it is important to recognize that discrepancies between radar and station data can arise from several factors. One of the key reasons is the differing number of stations in each region relative to the area covered (Table 1). Shehu & Haberlandt, (2023) note that regions with lower station density tend to exhibit higher uncertainty ranges. They found that northwest Germany, which has fewer rain gauges, displays higher uncertainty in the return levels. Marra et al. (2022) also observed that a denser network of short-duration rain gauges, particularly in high terrain, could enhance the accuracy of precipitation estimates. Additionally, Germann et al. (2006) emphasize that the limited representativeness of gauge data can affect the validity of results, especially for durations shorter than 24 hours.

Another factor contributing to discrepancies is the estimation uncertainty arising from region-specific characteristics. For example, there is a discrepancy in SMEV station data for the northwestern region and KOSTRA stations in the northeastern region, both of which are relatively flat. This discrepancy further rejects our assumption that station data serve as a reliable benchmark for our comparison in these regions. Previous studies pointed that this region is prone to higher uncertainty. Shehu & Haberlandt, (2023) found that the inclusion of extreme events from 2014 in Münster, a city in northwestern Germany, increases the uncertainty ranges, particularly for short durations. Additionally, Marra et al. (2017) demonstrated that arid climates can lead to large uncertainties due to infrequent rainfall events. Looking at the overall climatology map of Germany (Deutscher Wetterdienst), we can see that the northern parts are more arid than the southern regions, further supporting this observation. Miniussi & Marra (2021) also found when estimating parameters at ungauged locations, underestimation issues appear more significantly in eastern Germany.

Furthermore, the short radar record in certain regions (Fig 2-4) can also contribute to uncertainties. The limited length of these records makes it challenging to accurately estimate the parameters of extreme value distributions, potentially leading to overestimation of the tails of these distributions (Marra, Nikolopoulos, et al., 2019; Marra et al., 2018).

5 Conclusion and Future Work

To assess the spatial variability and estimation uncertainty of extreme rainfall derived from radar data, two methodologies were applied to data from 155 rain gauges and radar records provided by Deutscher Wetterdienst (DWD). The first method, KOSTRA-DWD-2020, relies on the Annual Maximum Series and the Generalized Extreme Value distribution. The second method, the Simplified Metastatistical Extreme Value (SMEV), models extreme rainfall using finite samples of independent ordinary storm events, characterized by a Weibull distribution.

The analysis highlights that both SMEV and KOSTRA-DWD-2020 have distinct strengths and limitations, emphasizing their suitability in different contexts. Radar-based estimates from SMEV demonstrates greater spatial variability than that of KOSTRA-DWD-2020, largely because it accounts for inclusion of a broader range of events leading to a more comprehensive representation of different rainfall intensities and durations. SMEV radar consistently demonstrates better alignment with station data, particularly in regions with complex topography, such as the Alps, where it effectively captures spatial variability. This strength makes SMEV especially suitable for shorter durations and lower return periods. However, the method's performance in certain regions and for short-duration high-return period is influenced by factors such as the spatial scales, the limited availability of rain gauges and the climatology. This highlights the inherent uncertainties in applying SMEV to datasets with incomplete or uneven spatial and temporal coverage.

KOSTRA-DWD-2020 radar estimates for shorter durations tend to deviate more significantly, particularly when compared with station data. These discrepancies may stem from its methodological approach, which includes reliance on Koutsoyiannis parameters and potential disaggregation techniques that are less effective in representing the fine-scale variability of radar data. The use of a fixed shape parameter to stabilize extreme value estimation up to 100-year return periods, while reducing uncertainty for individual stations, may not translate well to radar-based datasets, which inherently smooth out localized precipitation extremes.

For longer durations and higher return periods, no clear distinction emerges between the two methods in terms of superiority. This suggests that both methods can equally capture the localized variability associated with extreme events over extended

timeframes. In such cases, station data, with its longer archival record, may offer a more reliable option for capturing localized precipitation patterns.

In conclusion, radar-based SMEV provides a more detailed representation of spatial variability and aligns better with station data in short duration and low return period, while KOSTRA-DWD-2020 struggles to fully capture the variability in that context. This highlights the importance of carefully evaluating the suitability of each method based on the specific application, dataset characteristics, and scale of analysis. When using precipitation estimates derived from radar data, caution is essential, particularly for short durations and return periods, to ensure that methodological assumptions do not introduce biases or underrepresent variability.

Moreover, when evaluating the differences between radar and station-based precipitation estimates, it is crucial to consider factors such as station density, regional climate, and the length of radar records. These elements significantly influence the accuracy and reliability of the estimates and should be carefully accounted for in any comparative analysis.

In future studies, several areas of improvement and exploration could further enhance the understanding of return level estimation. One key direction is to study the variation of the parameters relationships rather than focusing solely on the variation of return levels. Incorporating information about the landscape into future methods could significantly improve spatial rainfall estimates by better capturing the effects of topography and other regional factors. Additionally, extending the dataset and improving temporal resolution would help refine scaling relationships, enabling more accurate predictions for different durations and return periods. Further regionalization efforts, particularly focusing on more regions and smaller-scale areas, could provide a deeper understanding of precipitation patterns at finer spatial scales. Moreover, including the estimation uncertainty of both KOSTRA estimates (for radar and station data) and SMEV radar estimates would help address the uncertainties in return level calculations and improve model reliability. Finally, narrowing the focus to a specific region could allow for more detailed analysis and a better understanding of local rainfall characteristics, potentially leading to more tailored and accurate estimation methods.

References

AghaKouchak, A., Behrangi, A., Sorooshian, S., Hsu, K., & Amitai, E. (2011). Evaluation of satellite-retrieved extreme precipitation rates across the central United States. *Journal of Geophysical Research*, *116*(D2), D02115. https://doi.org/10.1029/2010JD014741

Awadallah, A. G., & Awadallah, N. A. (2013). A Novel Approach for the Joint Use of Rainfall Monthly and Daily Ground Station Data with TRMM Data to Generate IDF Estimates in a Poorly Gauged Arid Region. *Open Journal of Modern Hydrology*, *03*(01), 1–7. https://doi.org/10.4236/ojmh.2013.31001

Awadallah, A. G., ElGamal, M., ElMostafa, A., & ElBadry, H. (2011). Developing Intensity-Duration-Frequency Curves in Scarce Data Region: An Approach using Regional Analysis and Satellite Data. *Engineering*, *03*(03), 215–226. https://doi.org/10.4236/eng.2011.33025

Bárdossy, A., & Pegram, G. (2017). Combination of radar and daily precipitation data to estimate meaningful sub-daily point precipitation extremes. *Journal of Hydrology*, *544*, 397–406. https://doi.org/10.1016/j.jhydrol.2016.11.039

Battan, L. (1973). *Radar observation of the atmosphere* (Vol. 332p). University of Chicago Press.

Bernard, M. M. (1932). Formulas For Rainfall Intensities of Long Duration. *Transactions of the American Society of Civil Engineers*, *96*(1), 592–606. https://doi.org/10.1061/TACEAT.0004323

Berne, A., & Krajewski, W. F. (2013). Radar for hydrology: Unfulfilled promise or unrecognized potential? *Advances in Water Resources*, *51*, 357–366. https://doi.org/10.1016/j.advwatres.2012.05.005

Borga, M., Gaume, E., Creutin, J. D., & Marchi, L. (2008). Surveying flash floods: gauging the ungauged extremes. *Hydrological Processes*, *22*(18), 3883–3885. https://doi.org/10.1002/hyp.7111

Borga, M., Marra, F., & Gabella, M. (2022). Chapter 5 - Rainfall estimation by weather radar. In R. Morbidelli (Ed.), *Rainfall* (pp. 109–134). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-12-822544-8.00016-0

Brandes, E. A., Vivekanandan, J., & Wilson, J. W. (1999). A Comparison of Radar Reflectivity Estimates of Rainfall from Collocated Radars. *Journal of Atmospheric and Oceanic Technology*, *16*(9), 1264–1272. https://doi.org/10.1175/1520-0426(1999)016<1264:ACORRE>2.0.CO;2

Demirdjian, L., Zhou, Y., & Huffman, G. J. (2018). Statistical Modeling of Extreme Precipitation with TRMM Data. *Journal of Applied Meteorology and Climatology*, *57*(1), 15–30. https://doi.org/10.1175/JAMC-D-17-0023.1

Deutscher Wetterdienst. (n.d.). *Climatological maps of Germany*.

Eldardiry, H., Habib, E., & Zhang, Y. (2015). On the use of radar-based quantitative precipitation estimates for precipitation frequency analysis. *Journal of Hydrology*, *531*, 441–453. https://doi.org/10.1016/j.jhydrol.2015.05.016

Eldardiry, H., Habib, E., Zhang, Y., & Graschel, J. (2017). Artifacts in Stage IV NWS Real-Time Multisensor Precipitation Estimates and Impacts on Identification of Maximum Series. *Journal of Hydrologic Engineering*, *22*(5). https://doi.org/10.1061/(ASCE)HE.1943-5584.0001291

Fisher, R. A., & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, *24*(2), 180–190. https://doi.org/10.1017/S0305004100015681

Fréchet, M. (1927). *Sur la loi de probabilité de l'écart maximum. Annales de la Société Polonaise de Mathématique . 6*, 93–116.

Frei, C., Davies, H. C., Gurtz, J., & Schär, C. (2000). Climate dynamics and extreme precipitation and flood events in Central Europe. *Integrated Assessment*, *1*(4), 281–300. https://doi.org/10.1023/A:1018983226334

Fuchsberger, J., Kirchengast, G., & Kabas, T. (2021). WegenerNet high-resolution weather and climate data from 2007 to 2020. *Earth System Science Data*, *13*(3), 1307–1334. https://doi.org/10.5194/essd-13-1307-2021

Gado, T. A., Hsu, K., & Sorooshian, S. (2017). Rainfall frequency analysis for ungauged sites using satellite precipitation products. *Journal of Hydrology*, *554*, 646–655. https://doi.org/10.1016/j.jhydrol.2017.09.043

Germann, U., Galli, G., Boscacci, M., & Bolliger, M. (2006). Radar precipitation measurement in a mountainous region. *Quarterly Journal of the Royal Meteorological Society*, *132*(618), 1669–1692. https://doi.org/10.1256/qj.05.190

Gnedenko, B. (1943). Sur La Distribution Limite Du Terme Maximum D'Une Serie Aleatoire. *The Annals of Mathematics*, *44*(3), 423. https://doi.org/10.2307/1968974

Houze Jr, & Robert A. (2014). *Cloud dynamics*. Academic Press.

III, J. P. (1975). Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, *3*(1). https://doi.org/10.1214/aos/1176343003

Intergovernmental Panel on Climate Change (IPCC). (2023). *Climate Change 2021 – The Physical Science Basis*. Cambridge University Press. https://doi.org/10.1017/9781009157896

IPCC, 2012. (n.d.). *IPCC, 2012: Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change [Field, C.B., V. Barros, T.F. Stocker, D. Qin, D.J. Dokken, K.L. Ebi, M.D. Mastrandrea, K.J. Mach, G.-K. Plattner, S.K. Allen, M. Tignor, and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, UK, and New York, NY, USA, 582 pp.*

Isidoro Orlanski. (1975). A Rational Subdivision of Scales for Atmospheric Processes. *Bulletin of the American Meteorological Society*, *56*, 327–530.

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, *81*(348), 158–171. https://doi.org/10.1002/qj.49708134804

Junghänel, T., Bär, F., Deutschländer, T., Haberlandt, U., Otte, I., Shehu, B., Stockel, H., Stricker, K., Thiele, L.-B., & Willems, W. (2022). *Methodische Untersuchungen zur Novellierung der Starkregenstatistik für Deutschland (MUNSTAR),. Synthesebericht. 95 pp. Methodological studies on the revision of heavy rain statistics for Germany*.

Kidd, C., & Levizzani, V. (2011). Status of satellite precipitation retrievals. *Hydrology and Earth System Sciences*, *15*(4), 1109–1116. https://doi.org/10.5194/hess-15-1109-2011

Koutsoyiannis, D. (2004). Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records / Statistiques de valeurs extrêmes et estimation de précipitations extrêmes: II. Recherche empirique sur de longues séries de précipitations. *Hydrological Sciences Journal*, *49*(4). https://doi.org/10.1623/hysj.49.4.591.54424

Koutsoyiannis, D., Efstratiadis, A., Mamassis, N., & Christofides, A. (2008). On the credibility of climate predictions. *Hydrological Sciences Journal*, *53*(4), 671–684. https://doi.org/10.1623/hysj.53.4.671

Langousis, A., Mamalakis, A., Puliga, M., & Deidda, R. (2016). Threshold detection for the generalized Pareto distribution: Review of representative methods and application to the NOAA NCDC daily rainfall database. *Water Resources Research*, *52*(4), 2659–2681. https://doi.org/10.1002/2015WR018502

Lengfeld, K., Kirstetter, P.-E., Fowler, H. J., Yu, J., Becker, A., Flamig, Z., & Gourley, J. (2020). Use of radar data for characterizing extreme precipitation at fine scales and short durations. *Environmental Research Letters*, *15*(8), 085003. https://doi.org/10.1088/1748-9326/ab98b4

Lengfeld, K., & Marra, F. (2024). Estimation of design precipitation using weather radar in Germany: A comparison of statistical methods. *Journal of Hydrology: Regional Studies*, *55*, 101952. https://doi.org/10.1016/j.ejrh.2024.101952

Lengfeld, K., Winterrath, T., Junghänel, T., & Becker, A. (2019). *Characteristic spatial extent of rain events in Germany from a radar-based precipitation climatology*.

Marani, M., & Ignaccolo, M. (2015). A metastatistical approach to rainfall extremes. *Advances in Water Resources*, *79*, 121–126. https://doi.org/10.1016/j.advwatres.2015.03.001

Marra, F., Armon, M., Borga, M., & Morin, E. (2021). Orographic Effect on Extreme Precipitation Statistics Peaks at Hourly Time Scales. *Geophysical Research Letters*, *48*(5). https://doi.org/10.1029/2020GL091498

Marra, F., Armon, M., & Morin, E. (2022). Coastal and orographic effects on extreme precipitation revealed by weather radar observations. *Hydrology and Earth System Sciences*, *26*(5), 1439–1458. https://doi.org/10.5194/hess-26-1439-2022

Marra, F., Borga, M., & Morin, E. (2020). A Unified Framework for Extreme Subdaily Precipitation Frequency Analyses Based on Ordinary Events. *Geophysical Research Letters*, *47*(18). https://doi.org/10.1029/2020GL090209

Marra, F., & Morin, E. (2015). Use of radar QPE for the derivation of Intensity–Duration–Frequency curves in a range of climatic regimes. *Journal of Hydrology*, *531*, 427–440. https://doi.org/10.1016/j.jhydrol.2015.08.064

Marra, F., & Morin, E. (2018). Autocorrelation structure of convective rainfall in semiarid-arid climate derived from high-resolution X-Band radar estimates. *Atmospheric Research*, *200*, 126–138. https://doi.org/10.1016/j.atmosres.2017.09.020

Marra, F., Morin, E., Peleg, N., Mei, Y., & Anagnostou, E. N. (2017). Intensity–duration–frequency curves from remote sensing rainfall estimates: comparing satellite and weather radar over the eastern Mediterranean. *Hydrology and Earth System Sciences*, *21*(5), 2389–2404. https://doi.org/10.5194/hess-21-2389-2017

Marra, F., Nikolopoulos, E. I., Anagnostou, E. N., Bárdossy, A., & Morin, E. (2019). Precipitation frequency analysis from remotely sensed datasets: A focused review. *Journal of Hydrology*, *574*, 699–705. https://doi.org/10.1016/j.jhydrol.2019.04.081

Marra, F., Nikolopoulos, E. I., Anagnostou, E. N., & Morin, E. (2018). Metastatistical Extreme Value analysis of hourly rainfall from short records: Estimation of high quantiles and impact of measurement errors. *Advances in Water Resources*, *117*, 27–39. https://doi.org/10.1016/j.advwatres.2018.05.001

Marra, F., Nikolopoulos, E. I., Creutin, J. D., & Borga, M. (2014). Radar rainfall estimation for the identification of debris-flow occurrence thresholds. *Journal of Hydrology*, *519*, 1607–1619. https://doi.org/10.1016/j.jhydrol.2014.09.039

Marra, F., Zoccatelli, D., Armon, M., & Morin, E. (2019). A simplified MEV formulation to model extremes emerging from multiple nonstationary underlying processes. *Advances in Water Resources*, *127*, 280–290. https://doi.org/10.1016/j.advwatres.2019.04.002

Marshall, J. S., Hitschfeld, W., & Gunn, K. L. S. (1955). *Advances in Radar Weather* (pp. 1–56). https://doi.org/10.1016/S0065-2687(08)60310-6

McGlynn, B. L., Blöschl, G., Borga, M., Bormann, H., Hurkmans, R., Komma, J., Nandagiri, L., Uijlenhoet, R., & Wagener, T. (2013). A data acquisition framework for runoff prediction in ungauged basins. In *Runoff Prediction in Ungauged Basins* (pp. 29–52). Cambridge University Press. https://doi.org/10.1017/CBO9781139235761.006

Miao, C., Ashouri, H., Hsu, K.-L., Sorooshian, S., & Duan, Q. (2015). Evaluation of the PERSIANN-CDR Daily Rainfall Estimates in Capturing the Behavior of Extreme Precipitation Events over China. *Journal of Hydrometeorology*, *16*(3), 1387–1396. https://doi.org/10.1175/JHM-D-14-0174.1

Miniussi, A., & Marra, F. (2021). Estimation of extreme daily precipitation return levels at-site and in ungauged locations using the simplified MEV approach. *Journal of Hydrology*, *603*, 126946. https://doi.org/10.1016/j.jhydrol.2021.126946

Morbidelli, R., Corradini, C., Saltalippi, C., & Flammini, A. (2022). Chapter 7 - Time resolution of rain gauge data and its hydrological role. In R. Morbidelli (Ed.), *Rainfall* (pp. 171–216). Elsevier. https://doi.org/https://doi.org/10.1016/B978-0-12-822544-8.00008-1

Nastos, P. T., Kapsomenakis, J., & Douvis, K. C. (2013). Analysis of precipitation extremes based on satellite and high-resolution gridded data set over Mediterranean basin. *Atmospheric Research*, *131*, 46–59. https://doi.org/10.1016/j.atmosres.2013.04.009

Olsson, J. (1998). Evaluation of a scaling cascade model for temporal rain- fall disaggregation. *Hydrology and Earth System Sciences*, *2*(1), 19–30. https://doi.org/10.5194/hess-2-19-1998

Overeem, A., Buishand, A., & Holleman, I. (2008). Rainfall depth-duration-frequency curves and their uncertainties. *Journal of Hydrology*, *348*(1–2), 124–134. https://doi.org/10.1016/j.jhydrol.2007.09.044

Overeem, A., Buishand, T. A., & Holleman, I. (2009). Extreme rainfall analysis and estimation of depth-duration-frequency curves using weather radar. *Water Resources Research*, *45*(10). https://doi.org/10.1029/2009WR007869

Papalexiou, S. M., & Koutsoyiannis, D. (2013). Battle of extreme value distributions: A global survey on extreme daily rainfall. *Water Resources Research*, *49*(1), 187–201. https://doi.org/10.1029/2012WR012557

Peleg, N., Marra, F., Fatichi, S., Paschalis, A., Molnar, P., & Burlando, P. (2018). Spatial variability of extreme rainfall at radar subpixel scale. *Journal of Hydrology*, *556*, 922–933. https://doi.org/10.1016/j.jhydrol.2016.05.033

Pellarin, T., Delrieu, G., Saulnier, G.-M., Andrieu, H., Vignal, B., & Creutin, J.-D. (2002). Hydrologic Visibility of Weather Radar Systems Operating in Mountainous Regions: Case Study for the Ardèche Catchment (France). *Journal of Hydrometeorology*, *3*(5), 539–555. https://doi.org/10.1175/1525-7541(2002)003<0539:HVOWRS>2.0.CO;2

Pombo, S., & de Oliveira, R. P. (2015). Evaluation of extreme precipitation estimates from TRMM in Angola. *Journal of Hydrology*, *523*, 663–679. https://doi.org/10.1016/j.jhydrol.2015.02.014

Pöschmann, J. M., Kronenberg, R., & Bernhofer, C. (2023). Spatial scaling of extreme rainfall from radar QPE in Germany. *Meteorologische Zeitschrift*, *32*(5), 353–365. https://doi.org/10.1127/metz/2023/1179

Prakash, S., Mitra, A. K., Pai, D. S., & AghaKouchak, A. (2016). From TRMM to GPM: How well can heavy rainfall be detected from space? *Advances in Water Resources*, *88*, 1–7. https://doi.org/10.1016/j.advwatres.2015.11.008

Rosin, T., Marra, F., & Morin, E. (2024). Exploring patterns in precipitation intensity–duration–area–frequency relationships using weather radar data. *Hydrology and Earth System Sciences*, *28*(15), 3549–3566. https://doi.org/10.5194/hess-28-3549-2024

Shehu, B., & Haberlandt, U. (2023). Uncertainty estimation of regionalised depth–duration–frequency curves in Germany. *Hydrology and Earth System Sciences*, *27*(10), 2075–2097. https://doi.org/10.5194/hess-27-2075-2023

Shehu, B., Willems, W., Stockel, H., Thiele, L.-B., & Haberlandt, U. (2023). Regionalisation of rainfall depth–duration–frequency curves with different data types in Germany. *Hydrology and Earth System Sciences*, *27*(5), 1109–1132. https://doi.org/10.5194/hess-27-1109-2023

Smith, J. A., Baeck, M. L., Miller, A. J., & Claggett, E. L. (2024). Rainfall Frequency Analysis Based on Long-Term High-Resolution Radar Rainfall Fields: Spatial Heterogeneities and Temporal Nonstationarities. *Water Resources Research*, *60*(3). https://doi.org/10.1029/2023WR035640

Smith, J. A., Hui, E., Steiner, M., Baeck, M. L., Krajewski, W. F., & Ntelekos, A. A. (2009). Variability of rainfall rate and raindrop size distributions in heavy rain. *Water Resources Research*, *45*(4). https://doi.org/10.1029/2008WR006840

Steiner, M., & Smith, J. A. (2000). Reflectivity, Rain Rate, and Kinetic Energy Flux Relationships Based on Raindrop Spectra. *Journal of Applied Meteorology*, *39*(11), 1923–1940. https://doi.org/10.1175/1520-0450(2000)039<1923:RRRAKE>2.0.CO;2

Tan, J., Petersen, W. A., Kirchengast, G., Goodrich, D. C., & Wolff, D. B. (2018). Evaluation of Global Precipitation Measurement Rainfall Estimates against Three Dense Gauge Networks. *Journal of Hydrometeorology*, *19*(3), 517–532. https://doi.org/10.1175/JHM-D-17-0174.1

Terink, W., Leijnse, H., van den Eertwegh, G., & Uijlenhoet, R. (2018). Spatial resolutions in areal rainfall estimation and their impact on hydrological simulations of a lowland catchment. *Journal of Hydrology*, *563*, 319–335. https://doi.org/10.1016/j.jhydrol.2018.05.045

Uijlenhoet, R. (2001). Raindrop size distributions and radar reflectivity–rain rate relationships for radar hydrology. *Hydrology and Earth System Sciences*, *5*(4), 615–628. https://doi.org/10.5194/hess-5-615-2001

Uijlenhoet, R., Steiner, M., & Smith, J. A. (2003). Variability of Raindrop Size Distributions in a Squall Line and Implications for Radar Rainfall Estimation. *Journal of Hydrometeorology*, *4*(1), 43–61. https://doi.org/10.1175/1525-7541(2003)004<0043:VORSDI>2.0.CO;2

Villarini, G., & Krajewski, W. F. (2010a). Review of the Different Sources of Uncertainty in Single Polarization Radar-Based Estimates of Rainfall. *Surveys in Geophysics*, *31*(1), 107–129. https://doi.org/10.1007/s10712-009-9079-x

Villarini, G., & Krajewski, W. F. (2010b). Review of the Different Sources of Uncertainty in Single Polarization Radar-Based Estimates of Rainfall. *Surveys in Geophysics*, *31*(1), 107–129. https://doi.org/10.1007/s10712-009-9079-x

Westra, S., Fowler, H. J., Evans, J. P., Alexander, L. V., Berg, P., Johnson, F., Kendon, E. J., Lenderink, G., & Roberts, N. M. (2014). Future changes to the intensity and frequency of short-duration extreme rainfall. *Reviews of Geophysics*, *52*(3), 522–555. https://doi.org/10.1002/2014RG000464

Winterrath, T., Brendel, C., & Hafer, M. (2018). *RADKLIM version 2017.002: Reprocessed gauge-adjusted radar data, one hour precipitation sums (RW)*.

Winterrath, T., Brendel, C., Hafer, M., Junghänel, T., Klameth, A., Walawender, E., Weigl, E., & Becker, A. (2017). *Erstellung einer radargestützten Niederschlagsklimatologie*.

Winterrath, T., Brendel, T., Junghänel, T., Klameth, A., Lengfeld, K., Walawender, E., Weigle, E., Hafer, M., & Becker, A. (2019). *An overview of the new radar-based precipitation climatology of the Deutscher Wetterdienst – data, methods, products* (N. Peleg & P. Molnar, Eds.).

Winterrath, T., Rosenow, W., & Weigl, E. (2012). *On the DWD quantitative precipitation analysis and nowcasting system for real-time application in German flood risk management.* (Vol. 351, pp. 323–329). IAHS Publ.

Wright, D. B., Mantilla, R., & Peters-Lidard, C. D. (2017). A remote sensing-based tool for assessing rainfall-driven hazards. *Environmental Modelling & Software*, *90*, 34–54. https://doi.org/10.1016/j.envsoft.2016.12.006

Wright, D. B., Smith, J. A., Villarini, G., & Baeck, M. L. (2013). Estimating the frequency of extreme rainfall using weather radar and stochastic storm transposition. *Journal of Hydrology*, *488*, 150–165. https://doi.org/10.1016/j.jhydrol.2013.03.003

Zhou, Y., Lau, W. K. M., & Huffman, G. J. (2015). Mapping TRMM TMPA into Average Recurrence Interval for Monitoring Extreme Precipitation Events. *Journal of Applied Meteorology and Climatology*, *54*(5), 979–995. https://doi.org/10.1175/JAMC-D-14-0269.1

Zorzetto, E., Botter, G., & Marani, M. (2016). On the emergence of rainfall extremes from ordinary events. *Geophysical Research Letters*, *43*(15), 8076–8082. https://doi.org/10.1002/2016GL069445

Acknowledgement

I would like to sincerely thank my supervisors, Francesco Marra and Katharina Lengfeld, for their kind guidance and support throughout this study.

I am also grateful to Deutscher Wetterdienst (DWD) in Offenbach, particularly the Department of Hydrometeorology, for providing me with the opportunity to conduct this study and for granting me access to the necessary data.