https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

Hydrology and Earth System Sciences
Discussions

The accuracy of weather radar in heavy rain: a comparative study for Denmark, the Netherlands, Finland and Sweden

Marc Schleiss[1], Jonas Olsson[2], Peter Berg[2], Tero Niemi[3,5], Teemu Kokkonen[3], Søren Thorndahl[4], Rasmus Nielsen[4], Jesper Ellerbæk Nielsen[4], Denica Bozhinova[2], and Seppo Pulkkinen[5]

[1]Dept. of Geoscience and Remote Sensing, Delft University of Technology, Netherlands
[2]Dept. of Hydrology, Swedish Meteorological and Hydrological Institute SMHI, Norrkoping, Sweden
[3]Dept. of Built Environment, Aalto University, Finland
[4]Dept. of Civil Engineering, Aalborg University, Denmark
[5]Finnish Meteorological Institute FMI, Helsinki, Finland

Correspondence: Marc Schleiss (m.a.schleiss@tudelft.nl)

Abstract. Weather radar has become an invaluable tool for monitoring rainfall and studying its link to hydrological response. However, when it comes to accurately measuring small-scale rainfall extremes responsible for urban flooding, many challenges remain. The most important of them is that radar tends to underestimate rainfall compared to gauges. The hope is that by moving to higher resolution and making use of dual-polarization, these mismatches can be reduced. Each country has developed its own strategy for addressing this issue. But since there is no common benchmark, improvements are hard to quantify objectively. This study sheds new light on current performances by conducting a multinational assessment of radar's ability to capture heavy rain events at scales of 5 min up to 2 hours. The work is performed within the context of the joint experiment framework of project MUFFIN (Multiscale Urban Flood Forecasting), which aims at better understanding the link between rainfall and urban pluvial flooding across scales.

In total, 6 different radar products in Denmark, the Netherlands, Finland and Sweden were considered. The top 50 events for each country were used to quantify the overall agreement between radar and gauges and the errors affecting the peaks. Results show that the overall agreement between radar and gauges in heavy rain is fair, with multiplicative biases in the order of 1.41-1.66 (i.e., radar underestimates by 29-39.8%) and correlation coefficients of 0.71-0.83 across countries. However, the bias increases with intensity, reaching 45.9%-66.2% during the peaks. Only part of the bias (i.e., roughly 13%-30% depending on the radar product) can be explained by differences in measurement areas between gauges and radar. Radar products with higher spatial and temporal resolutions agreed better with the gauges, highlighting the importance of high-resolution radar for urban hydrology. However, for capturing peak intensity and reducing the bias during the most intense part of a storm, the ability to combine measurements from multiple overlapping radars to help mitigate attenuation seemed to play a more important role than resolution. The use of dual-polarization and phase information (e.g., Kdp) in the experimental Finnish OSAPOL product also seemed to provide a slight advantage in heavy rain. But improvements were hard to quantify and similarly good results were achieved in the Netherlands by applying a simple Z-R relation together with a mean field bias-correction.

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

1 Introduction

Today, several high-resolution radar rainfall products for use in hydrology are readily available across the globe (Huuskonen et al., 2014; Thorndahl et al., 2017). Compared with gauges, radar provides superior spatial coverage, leading to more insight into the spatio-temporal characteristics of rain events and their link to hydrological response (Wood et al., 2000; Berne et al., 2004; Smith et al., 2007). Steady improvement in radar technology over the past decades and in particular the switch from single to dual-polarization has lead to significant progress in terms of clutter suppression, hydrometeor classification and attenuation correction, greatly enhancing the accuracy and reliability of operational quantitative radar precipitation estimates (Zrnic and Ryzhkov, 1996; Ryzhkov and Zrnic, 1998; Zrnic and Ryzhkov, 1999; Bringi and Chandrasekar, 2001; Gourley et al., 2007; Matrosov et al., 2007). Polarimetry also fundamentally changed the way we estimate rainfall from radar measurements, with traditional Z-R power law relationships being increasingly replaced by alternative methods based on differential phase shift (Ryzhkov and Zrnic, 1996; Zrnic and Ryzhkov, 1996; Brandes et al., 2001; Matrosov et al., 2006; Otto and Russchenberg, 2011). Despite these encouraging developments, many challenges related to the measurement of small-scale rainfall extremes responsible for urban pluvial flooding remain (Einfalt et al., 2004; Lee, 2006; Krajewski et al., 2010; Villarini and Krajewski, 2010; Berne and Krajewski, 2013). The most important of them is that radar tends to underestimate rainfall peaks compared with rain gauges. This is mainly attributed to signal attenuation and to the large differences in measurement principles and sampling volumes between radar and gauges. In some cases, the underestimation can also be related to calibration issues, range effects or saturation of the receiver channel. Wind effects and vertical variability also play an important role, further complicating the matching between radar and rain gauge data at higher resolutions (Vasiloff et al., 2009; Dai and Han, 2014). The hope is that by moving to higher resolutions and taking advantage of dual-polarization, the average mismatch between radar and gauges will become smaller. However, as highlighted by the studies of Krajewski and Smith (2002) and Seo et al. (2015), this is a very delicate balance as higher resolution and more elaborate retrieval algorithms can also lead to more noise and uncertainties. As a result, accuracy strongly depends on the type of precipitation, its spatio-temporal characteristics and location with respect to the radar(s).

Since radar measurements are inherently uncertain and knowledge about microphysical processes in clouds and rain is limited, post-processing plays an important role. In addition to using better hardware, many weather services now offer higher-level composite rainfall products that combine measurements from different radar systems and have been corrected for various types of biases using rain gauges (Krajewski, 1987; Smith and Krajewski, 1991; Goudenhoofdt and Delobbe, 2009; Stevenson and Schumacher, 2014). If done properly, this can help mitigate attenuation and reduce systematic biases due to calibration issues and natural variability of the raindrop size distribution (e.g., Collier and Knowles, 1986; Young et al., 2000; Gourley et al., 2006; Overeem et al., 2009b). The main limitation of rain gauge adjustments, however, is that they only account for average biases over relatively large spatial and temporal domains. These can be very different from local errors and may not necessarily be very representative of the peaks. Also, one has to keep in mind that rain gauge measurements themselves are prone to biases and errors, the most common of them being an underestimation of the rainfall intensity due to local wind effects around the gauge. These effects have been estimated to be in the order of 5-10% in regular rain events but can reach 25-30%

or more in conditions of extremely heavy rainfall rates over 50-100 mmh^{-1} (Nystuen, 1999; Sieck et al., 2007; Pollock et al., 2018).

Another important problem when studying the performance of radar in heavy rain is the length of the available data records. Due to frequent upgrades in radar hardware, software and data processing, the longest currently available radar records that can be used for analysis span 15-20 years at best. This is significantly shorter than for gauges and makes it hard to draw relevant conclusions about extreme weather events. Thus, so far very few studies have looked at the systematic discrepancies between radar and gauges in times of heavy rain. Using a 12-year archive of 1$\times$1 km and 5-min radar rainfall estimates for Belgium between 2005-2016, Goudenhoofdt et al. (2017) found that hourly radar extremes around Brussels tend to be 30-70% lower than those observed in gauge data. In the Netherlands, Overeem et al. (2009b) compiled a 10-year climatology of radar-based extreme rainfall estimates to derive intensity-duration-frequency curves (Overeem et al., 2009a) and areal extremes (Overeem et al., 2010) for time scales of 15 min to 24 h. The authors concluded that radar data may be suitable to estimate local and regional extreme rainfall statistics, provided that they are carefully quality controlled and bias corrected. In the United States, Smith et al. (2012) and Wright et al. (2014) compiled a 10-year high-resolution radar rainfall dataset at 15 min and 1 km resolution based on the NEXRAD data for the Baltimore and Charlotte metropolitan areas. Their studies highlighted the value of long-term radar observations for characterizing the relationship between rainfall and hydrological response but also pointed out many forms of systematic errors that persist in bias-adjusted radar products such as range-dependent and intensity-dependent multiplicative biases. A few years later, Thorndahl et al. (2014b) developed a storm catalog of 50 heavy rain events as seen by WSR-88D radars in the Milwaukee area between 1996 and 2011. Their analysis covered more than 15 years but the radar data used to derive the statistics were not continuous in time.

Because of the difficulty to get long homogeneous radar archives, the studies published so far mostly focused on regional or national performances. Often, the methodologies used to carry out the analyses were different, which makes it hard to compare the results. Consequently, there is a strong need for systematic, multinational assessments and comparisons of radar's ability to capture heavy rain. This paper sheds new light on this issue by providing a detailed analysis of 6 different radar products across 4 European countries (i.e., Denmark, the Netherlands, Finland and Sweden). Inspired by the approach of Thorndahl et al. (2014b), we selected the 50 most intense events for each country over the last 10 years to study the average agreement between radar and gauges as well as the discrepancies in terms of peak rainfall intensities. The study is performed within the context of the Water JPI funded project MUFFIN: Multiscale Urban Flood Forecasting which aims at better understanding the link between rainfall and urban pluvial flooding across scales. By comparing different types of radar products (C-band vs X-band, single vs dual-polarization) and analyzing error propagation across different spatial and temporal scales, important conclusions and recommendations can be drawn as to the use of radar in hydrology and flood forecasting.

2 Data & Methods

2.1 Event selection

Event selection was done based on rainfall time series from the national networks of automatic rain gauges in Denmark, the Netherlands, Finland and Sweden. Due to data availability and quality, only a smaller subset of all the gauges was used for analysis (i.e., 66 gauges for Denmark, 35 for the Netherlands, 64 for Finland and 10 for Sweden). Table 1 provides an overview of the number of available gauges, their temporal resolutions and length of the observational records for each country. Using the selected gauges, we determined the top 50 rain events (in terms of peak intensity) for each country and observation period. Only events for which both the gauge and radar data were available simultaneously were considered. Also, we imposed the condition that two events for the same location had to be separated by a continuous dry period of at least 6 hours. To increase reliability, all events were subjected to a visual quality control test by human experts, checking both for plausibility and consistency. Cases for which the gauge or radar data were incomplete, obviously wrong or strongly inconsistent with each other were removed and replaced by new events until the total number of events that passed the quality control tests reached 50 for each country. Overall, about 10% of the originally selected events had to be removed and replaced by new ones during these quality control steps, most of them because of incomplete radar data.

The procedure used to extract the radar data was identical for all countries. First, the 4 radar pixels closest to a given rain gauge were extracted. The 4 radar rainfall time series were then aggregated in time (i.e., averaged) to match the temporal sampling resolution of the rain gauge. Then, for each time step, the value among the 4 radar pixels that best matched the gauge was kept for comparison. The motivation behind this type of approach is that it can account for small differences in location and timing between radar and gauge observations due to motion, wind and vertical variability. This leads to a much more conservative approach than pixel-by-pixel comparisons in which we actively try to minimize the differences between radar and gauges as much as possible. Other less favorable ways of extracting the radar data were also tested (e.g., using inverse distance weighted interpolation or the maximum value among the nearest neighbors). But these only resulted in higher discrepancies without changing the main conclusions and were subsequently abandoned.

Figure 1 shows a map with the location of all rain gauges used for the final, quality-controlled rain event catalog for each country. As shown in Figure 2, the final catalog includes a large variety of rain events, ranging from single isolated convective cells to large organized thunderstorms and mesoscale complexes. Additional tables summarizing the starting time, duration, amount and peak rainfall intensity for each event and country are provided in the Appendix (see Tables A1-A5). Note that in Denmark and Finland, each of the top 50 events corresponded to a different rain gauge while in the Netherlands and Sweden, some of the gauges were used for more than one event.

Because events were selected based on peak intensity alone, it is not surprising to see that all 50 of them occurred in the warm season between May and September during which convective activity is at its maximum (see Figure 3). Similar analyses confirm that the events mostly occurred during the afternoon and late evening hours, in agreement with the diurnal cycle of convective precipitation and rainfall intensity at mid-latitudes (Rickenbach et al., 2015; Blenkinsop et al., 2017; Fairman et al., 2017).

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

Hydrology and Earth System Sciences
Discussions

2.2 The radar products

This section gives a brief overview of the different radar products used for the analyses. A short summary of the most important characteristics of each product is provided in Table 2.

2.2.1 Denmark

The Danish radar product is derived from the measurements of the "Stevns" C-band radar located approximately 40 km south of Copenhagen in an area of relatively flat topography with altitudes ranging from -7m to 125m above mean sea level. The radar volume scans at 9 different elevation angles are projected to a pseudo-constant altitude plan position indicator (PCAPPI) at 1000 m height to generate a high-resolution gridded product with 10 min temporal resolution and $500\times$ 500 m^2 grid spacing (Gill et al., 2006). The temporal resolution of the PCAPPI is then enhanced to 1 min using advection interpolation (Thorndahl et al., 2014a; Nielsen et al., 2014). Ground clutter is removed by filtering out echoes with Doppler velocity smaller than 1 ms^{-1}. Rainfall rate R is estimated based on a fixed Z-R relationship given by Z = 200R$^{1.6}$. Rain attenuation correction is estimated as K = 6.9 $\cdot 10^{-5}$ Z$^{0.67}$ [dBZ km^{-1}]. Rain rate values are corrected for mean field bias based on daily data from a network of 66 RIMCO tipping bucket rain gauges operated by the Water Pollution Committee of the Society of Danish Engineers (Madsen et al., 1998). Note that the 500 m, 1 min product used in this study is not operational, but developed for research purposes for Aalborg University.

2.2.2 Netherlands

The used product is a 10-year archive of 5 min precipitation depths at 1×1 km^2 spatial resolution based on a composite of radar reflectivities from 2 C-band radars in De Bilt and Den Helder operated by the Royal Netherlands Meteorological Institute (KNMI). Note that the radar in De Bilt stopped contributing to the composite in the course of January 2017, at which point it was replaced by a new polarimetric radar in the nearby village of Herwijnen (51.837°N,5.138°E). Rainfall estimates are obtained by combining the PCAPPIs of the two radars at 1500 m height and applying a constant Z-R relationship given by Z=200R$^{1.6}$. The rainfall estimates are then adjusted for bias at hourly time scales using 35 automatic weighing rain gauges operated by KNMI. An extensive description and documentation of the radar and gauge products is available on the KNMI website. Note that the Netherlands recently upgraded their radars to dual-polarization. However, the dual-polarization rainfall estimates are not fully operational yet and all rainfall values used in this study were produced with the single-polarization algorithms.

2.2.3 Finland

The Finnish radar product is an experimental product from the FMI OSAPOL-project, which differs from the operational product used by the Finnish Meteorological Institute (FMI) mainly by making a better utilization of dual-polarization and by better taking into account the measurement geometry of the 10 C-band dual-polarization Doppler radars currently available in Finland. The product is based on the years 2013-2016, during which the old single-polarization radars were replaced by newer dual-polarization radars. Since this upgrade took place progressively, the OSAPOL-product combines data from 4 up to 9 dual-polarization radars depending on the number of radars that were available each year. Erroneous echoes and non-meteorological targets are removed using four different techniques. The algorithm used for correcting the vertical profile of reflectivity (VPR) is the same as in the operational product. Rainfall intensity is estimated based on radar reflectivity Z and specific differential propagation phase shift Kdp. For heavy rain, Kdp is used while for low to moderate intensities a fixed Z-R relation given by Z = $223R^{1.53}$ (Leinonen et al., 2012) is used. A PCAPPI at 500 m height with 1×1 km^2 spatial and 5 min temporal resolution is produced from the VPR-corrected radar intensity estimates of 4-6 lowest elevation angles by weighting them with a Gaussian function. The OSAPOL is the only product that is not gauge-adjusted. Since the focus of this study is on heavy convective events, only the radar data between May and September were used.

2.2.4 Sweden

The considered product is the so-called BRDC produced by SMHI. It is a 2×2 km, 15 min composite product of PCAPPIs sourced from 12 operational single-polarization C-band Doppler radars in Sweden (see Figure 1 in Norin et al. (2015)) between the years 2007 and 2016. After that, the product was discontinued and replaced by the newer BALTRAD product (Michelson et al., 2018). In the BRDC, rain rate is estimated by projecting polar reflectivity measurements at 10 different elevation angles between 0.5 and 40 degrees to a PCAPPI at 500 m height (See Section 2.2 in Norin et al. (2015) for more details). Ground clutter is removed by filtering all echoes with radial velocities less than 1 ms^{-1} and all remaining non-precipitation echoes are removed by applying a consistency filter based on satellite observations (Michelson, 2006). The effect of topography is accounted for by applying a beam blockage correction scheme based on the method by Bech et al. (2003). Rainfall rates on the ground are estimated through a constant Z-R relationship $Z=200R^{1.6}$. To reduce errors and biases, a method called HIPRAD (HIgh-resolution Precipitation from gauge-adjusted weather RADar) is applied (Berg et al., 2016). The latter was developed to make radar data more suitable for hydrological modeling by removing both long-term biases and range dependent biases. Note that although several radars are available in Sweden, the system is currently set up such that each radar has a predetermined non-overlapping measurement area. The final rainfall estimates therefore only include information from a single radar (i.e., usually the nearest one) and do not take advantage of possibly overlapping measurement areas. Such methods are being developed but are not yet implemented operationally.

2.2.5 Additional radar products

In addition to the 4 main radar products described above, two additional radar datasets were considered. The first is from a FURUNO WR-2100 polarimetric X-band Doppler research radar system located in Aalborg which scans at a fixed elevation angle of 4° in a radius of about 40 km around Aalborg with a high spatial resolution of 100×100 m^2 and temporal sampling resolution of 1 min. Clutter is removed by applying a filter on the Doppler velocities and a spatial texture filter on reflectivity. Rainfall rates are estimated using a fixed Z-R relationship given by $Z = 200R^{1.6}$. All rainfall rates are corrected for daily mean field bias using gauges using the same procedure as for the C-band data. The main issue with the X-band data is that it only covers a two-year period from 2016-2017 which strongly limits the number of heavy rain events available for the analysis.

Hydrology and Earth System Sciences
Discussions

Consequently, only the 10 most intense events were considered. Despite the low sample size, the hope is that by comparing the performance of the X-band product to the C-band product, valuable insight into the benefits of high-resolution polarimetric rainfall measurements in times of heavy rain can be gained.

The second additional radar product used for comparisons is an international composite derived from the BALTRAD collaboration (Michelson et al., 2018). The version used in this paper is the "tas BALTRAD" and it is essentially identical to the BRDC product used in Sweden except that it does not include the HIPRAD adjustments. Bias correction is done by taking each 15-min time step and scaling it with the ratio of 30-day aggregation of gauge and radar accumulations. The HIPRAD also covers a much larger area than the BRDC product. This extended coverage is made possible thanks to the automatic radar data exchange between neighboring countries around the Baltic sea (i.e., Norway, Finland, Estonia, Latvia and Denmark). The high data availability means that BALTRAD is suitable for evaluation and comparisons of all rain events studied in this paper except the ones that occurred over the Netherlands (which are currently not part of BALTRAD). Nevertheless, by analyzing and comparing the BALTRAD for the 50 top events in Denmark, Sweden and Finland, important conclusions about the advantages and limitations of tailored high-resolution national radar products can be made.

2.3 Performance metrics

Since radar and gauges measure rainfall at different scales based on different measuring principles, one does not expect a perfect agreement between the two. Gauges are more representative of point rainfall measurements on the ground while radar provides volume-averages at several hundreds of meters above the ground. In addition, each sensor has its own measurement uncertainty and limitations in times of heavy rain. For example, gauges are known to underestimate rainfall rates in conditions of high winds (e.g., Sieck et al., 2007; Goudenhoofdt et al., 2017; Pollock et al., 2018) which is common during thunderstorms while radar is known to suffer from signal attenuation, non-uniform beam filling, clutter, hail contamination and overshooting (Krajewski et al., 2010; Villarini and Krajewski, 2010; Berne and Krajewski, 2013). The main goal here is not to make a statement about which measurement is closer to the truth but to quantify the average discrepancies between the gauge and radar measurements as a function of the event, time scale, intensity and radar product. Such information can be used as a benchmark against which further developments in radar products can be assessed or as a very simple way to study the effect of rainfall measurement uncertainty on error propagation in hydrological models.

To assess performance, the average discrepancies between radar and gauges were quantified by calculating standard error metrics such as the linear correlation coefficient (CC) and relative root mean square error (RRMSE):

$$\mathrm{CC} = \frac{1}{N} \cdot \frac{\sum_{i=1}^{N} (X_i - \mu_X)(Y_i - \mu_Y)}{\sigma_X \cdot \sigma_Y} \tag{1}$$

$$\mathrm{RRMSE} = \frac{1}{\mu_Y} \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} (Y_i - X_i)^2} \tag{2}$$

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

where X_i and Y_i represent the radar and rain gauge measurements, N is the number of observations, $\mu_{X|Y}$ the average rainfall intensities and $\sigma_{X|Y}$ their respective standard deviations. All these statistics are calculated on an event-by-event basis at a fixed temporal aggregation scale Δt (omitted in the equations to simplify the notations).

In addition to the CC and RRMSE, we also consider the multiplicative bias (MB) between the gauge and radar measurements. By convention, the multiplicative bias is calculated by taking the gauge measurements Y_i (in mmh^{-1}) as a reference value:

$$Y_i = \mathrm{MB} \cdot X_i \cdot \varepsilon_i \tag{3}$$

where ε_i are random errors drawn from a continuous and positive probability distribution (e.g., a log-normal) with median 1 (Smith and Krajewski, 1991). In the equation above, a value of MB>1 means that the rain gauges tend to give larger rainfall rates than the radar, which is generally the case for heavy rain events. Previous studies have shown that the multiplicative bias model in Equation (3) provides a better, physically more plausible representation of the error structure between in-situ and remotely-sensed rainfall observations than a the additive bias model commonly used in statistics (e.g., Tian et al., 2013).

In this paper, the multiplicative bias is estimated through the so-called G/R method, that is, by taking the mean rainfall value measured by the gauges over an event divided by the mean rainfall value of the radar (Yoo et al., 2014). Other more elaborate estimators (e.g., least squares and maximum likelihood) have been proposed depending on the distribution of ε_i but the G/R ratio has the advantage of providing estimates that are directly related to total rainfall amounts and do not depend on the temporal aggregation scale. This may not necessarily be the optimal way to estimate the multiplicative bias but considerably simplifies the analyses by making it easier to compare values from one country to another, independently of the spatial and temporal resolution of the radar products.

To express the multiplicative bias in terms of a relative error ϵ_{rel} (in percentage relative to the values recorded by the gauge), the following formula is used:

$$\epsilon_{\mathrm{rel}} = 100\% \cdot \mathbb{E}\left[\frac{Y_i - X_i}{Y_i}\right] = 1 - \frac{1}{\mathrm{MB}} \cdot \mathbb{E}\left[\frac{1}{\varepsilon_i}\right] = 1 - \frac{1}{\mathrm{MB}} \tag{4}$$

where $\mathbb{E}$ denotes the expectation and by definition the median of ε_i is assumed to be equal to 1.

While standard error metrics like RRMSE, CC and MB provide an important overview of the average error, they may not necessarily be representative of what happens during the most intense parts of a storm. Therefore the second part of the analyses focuses on assessing the peak rainfall intensity bias (PIB) between radar and gauges. The PIB is defined as:

$$Y_{\max}(\Delta t) = \mathrm{PIB}(\Delta t) \cdot X_{\max}(\Delta t) \tag{5}$$

where $Y_{\max}(\Delta t)$ and $X_{\max}(\Delta t)$ denote the maximum rain rate values recorded by the gauges and radar at temporal aggregation time scale Δt. The PIB values are computed on an event-by-event basis, by aggregating the radar and gauge data to a fixed temporal resolution Δt (using overlapping time windows) and extracting the maximum rain rate over the event at this scale. Note that this is done independently for the gauges and the radar time series, which means that the maximum values may not necessarily correspond to the same time interval. The advantage of this is that it leads to more reliable and robust PIB estimates at high resolutions where statistics would otherwise be strongly sensitive to small timing issues between radar and gauge observations.

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

Hydrology and Earth System Sciences
Open Access
Discussions

3 Results

3.1 Agreement during the most intense events

Figure 4 shows the time series of rainfall intensities at the highest available temporal resolution for the top event in each country. The time series reveal a strong, consistent pattern of underestimation by the radar compared with the gauge values. The multiplicative biases corresponding to these 4 events are 1.66, 1.37, 1.55 and 1.69 for Denmark, the Netherlands, Finland, and Sweden, respectively. In other words, according to equation (4), radar underestimates the rainfall intensity by 27-41% compared with the gauges, which is consistent with previous values reported in the literature. For example, Goudenhoofdt et al. (2017) mentions up to 30% underestimation while Seo et al. (2015) reported up to 50% on individual events.

Note that all 4 events displayed in Figure 4 fall under the category of extremely intense rain, with peak intensities reaching 204 mmh^{-1} in Denmark, 180 mmh^{-1} in the Netherlands, 89.1 mmh^{-1} in Finland and 91.2 mmh^{-1} in Sweden. The July 2, 2011 event in Denmark was particularly violent, affecting more than a million people in the greater Copenhagen region and causing an estimated damage of at least 800 million euros (Wójcik et al., 2013). The third rainfall peak was particularly impressive, with rain rates remaining well above 125 mmh^{-1} for three consecutive time steps, resulting in more than 41 mm of rain (e.g., about one month's worth of rain for the Copenhagen region) in only 15 minutes. During the same time period, the radar only recorded 12.1 mm, underestimating the 15-min peak rainfall intensity by a factor of more than 3. Clearly, the error structure between radar and gauges appears to be time dependent, with increasing discrepancies as we move towards higher intensities. The relatively large peak intensity biases of 2.17, 2.09, 1.98 and 1.73 for Denmark, Finland, the Netherlands and Sweden respectively confirm this hypothesis. During the most intense parts of the storms, radar underestimates by 42-54% compared with the gauges (i.e., about 10-15% more than suggested by the average multiplicative bias).

3.2 Overall agreement between radar and gauges

In the following, the overall agreement between radar and gauges for all 50 top events is analyzed. Figure 5 shows the radar rainfall intensities versus the gauge estimates at the highest available temporal resolution for each country (e.g., 5 min for Denmark, 10 min for the Netherlands and Finland and 15 min for Sweden). Each dot in this figure represents a radar-gauge pair and all 50 events have been combined together into the same graph.

The large scatter and relatively large RRMSE values of 116.4% to 139.1% highlight the strong disagreements between radar and gauge estimates at these scales. This is normal and can be explained by the fact that radar and gauges do not measure at the same height and over the same volume. It is important to note also that the gauge integrates precipitation over time whereas radar takes snapshots. Wind effects, changing microphysics and sampling uncertainties therefore also play an important role at such small scales. Despite the large scatter, linear correlation coefficients are relatively high (i.e., 0.71-0.83), indicating a good agreement in terms of temporal structure. However, the radar clearly underestimates the rainfall intensity compared with the gauges, Multiplicative bias values are 1.59 for Denmark, 1.41 for the Netherlands, 1.56 for Finland and 1.66 for Sweden which corresponds to an underestimation of 37.1%, 29.1%, 35.8%, and 39.8% respectively.

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

Hydrology and Earth System Sciences
Discussions

Figure 6 provides a similar overview of the discrepancies between radar and gauges for the event scale. Each dot in this graph represents the total rainfall accumulation (in mm) over an event. The aggregation to the event scale removes a lot of the noise and scatter that is present at the higher resolutions, providing a much clearer overview of the systematic bias affecting radar estimates. However, values are strongly dependent on the event duration and the measurement frequency of the radar.

Figure 6 shows that when data are aggregated to the event scale, the agreement between radar and gauges tends to improve, as confirmed by the lower relative root mean square errors of 39.4-47.7% and the higher correlation coefficients of 0.86-0.92. The multiplicative bias values, however, remain the same due to the way they were estimated through the G/R ratio (see Section 2.3). The good agreement at the event scale is clearly encouraging but must be interpreted carefully as improvements are mostly due to the inclusion of many lower intensity rainfall periods during which radar and gauges are in relatively good agreement with each other. The latter make up a significant part of an event but may not necessarily be representative of the differences observed in periods of high intensities.

Based on Figures 5 and 6, one could conclude that the Dutch C-band radar product appears to have the best overall agreement with the gauges among all countries, followed by Finland, Denmark and Sweden. However, such direct comparisons would not really be fair, as one also needs to take into account the differences in spatial and temporal resolutions between the radar products. To better separate the two, empirically derived areal-reduction factors (ARFs) proposed by Thorndahl et al. (2019) were used to estimate the theoretical bias between a point measurement and an areal-average from radar (i.e., using Equation (8) in Thorndahl et al. (2019) with $b_1 = 0.31$, $b_2 = 0.38$ and $b_3 = 0.26$). Our calculations show that for the Danish product (0.25 km^2, 5 min), about 12.8% of the underestimation can be explained purely due to differences in measurement support (i.e., the spatio-temporal domain over which measurements are performed). For Finland and the Netherlands (1 km^2, 10 min), the underestimation due to the measurement support is in the order of 18.6% while for Sweden (4 km^2, 15 min), values up to 29.6% can be expected. This means that after accounting for areal-reduction factors, radar only underestimates by about 10-24% compared with the gauges (i.e., 24.3% for Denmark, 17.2% for Finland, 10.5% for the Netherlands and 10.2% for Sweden). Table 3 summarizes the agreement of each product

We see that measurement support bias obviously plays an important role, explaining why lower resolution products such as the BRDC in Sweden tend to have a higher overall bias. But resolution alone does not explain everything. For example, the high 500 m, 5 min resolution in the Danish product does not appear to translate into a clear advantage in terms of multiplicative bias compared with the 1 km, 10 min resolution in the Netherlands and Finland. Taking into account the measurement support biases, the Danish product underestimates by 24.3% while the Finnish and Dutch only underestimate by 17.2% and 10.5% respectively. One possible explanation for this could be that the Finnish and Dutch products combine data from multiple radars to produce the final rainfall estimates (which helps mitigate attenuation and overshooting), whereas the Danish product only considers the measurement from a single radar. Other small differences in the bias-correction schemes and the density of the rain gauge networks used to adjust the radar could also play a role here. Another, simpler explanation could be that the bias increases with the intensity of the rain events, potentially masking the benefits of a higher spatial and temporal resolution. This is a rather important issue to consider when making comparisons between countries given that not all rain events in the database are of the same magnitude. For example, the Danish database contains events that are significantly more intense compared with

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

the Netherlands, Finland and Sweden (see Figure 7). Also, the longest event in the Danish database only lasted 4 hours, which is significantly less than for the other countries.

A deeper analysis of this issue confirms that on average, higher rainfall intensities appear to be linked with slightly larger multiplicative biases. However, the link between the bias and the average intensity remains rather weak, with rank correlation values of 0.33 in the Netherlands, 0.30 in Denmark, 0.04 in Finland and 0.19 in Sweden. Still, there appears to be a strong contrast between the average discrepancies between radar and gauges at the event scale, as shown in Figure 7(a), and the large mismatches in terms of peak rainfall intensities in Figure 7(b). In most cases, the highest intensities measured by the radar over the top 50 events barely match the lowest peak intensities measured by the gauges. The bias therefore appears to be largely influenced by event duration and the presence of lower rainfall intensities for which radar and gauges tend to be in better agreement than during the peaks.

Before diving deeper into the analysis of the peak rainfall intensities, we finish this sub-section by taking a closer look at the overall agreement between radar and gauges as a function of the temporal aggregation time scale. Figure 8 shows the relative root mean square error and correlation coefficient of radar versus gauge measurements for different aggregation time scales up to 2 hours. It shows a strong link between the spatial and temporal resolution of the radar data and its overall agreement with the gauges. When displayed at a similar temporal resolution, the Danish radar product clearly exhibits the lowest relative errors and highest correlation coefficients. It is followed by the Dutch and Finnish products (1 km) which have similar performance overall (e.g., the Finnish product has slightly higher correlation values but the Dutch has slightly lower RRMSE). The Swedish product, which has the lowest spatial resolution (i.e., 2 km) clearly exhibits the lowest agreement with the gauges. These results are not really surprising, only confirming that on average, a higher spatial and temporal resolution in the radar leads to a better agreement between radar and gauges (i.e., a better representativity of point measurements with respect to an areal-average). Still, the fact that the Dutch radar product (which has been bias-adjusted using gauges) performs very similarly to the Finnish OSAPOL product (which has not been bias-corrected) is interesting. One possible reason for this could be that the Finnish product makes use of polarimetry and phase information (e.g., Kdp) to estimate rainfall intensity in times of heavy rain as opposed to reflectivity alone. However, this remains highly speculative at this point as the statistics shown here were calculated over different events and radar configurations. Furthermore, the quality and density of the gauge networks used to perform bias adjustment in the Netherlands also plays an important role.

3.3 Agreement during the peaks

While the previous section heavily focused on the overall agreement between radar and gauges, this section takes a closer look at the peaks. Figure 9 shows the underestimation of peak rainfall intensity between radar and gauges as a function of aggregation time scale for each country. The dashed horizontal lines denote the average underestimation in each country, corresponding to the multiplicative bias in Figures 5 and 6. The data can be divided in two groups depending on the magnitude of the underestimation. The first group (i.e., Netherlands and Finland) is characterized by a median underestimation of peak rainfall intensity (at 10 min scale) of 47.1% and 45.9% respectively, only slightly exceeding the overall bias by 16.8% and 11.2% respectively. Moreover, the bias affecting the peak intensity rapidly decreases with aggregation time scale, converging to the overall bias previously calculated for all 50 events. The hourly mean field bias correction in the Dutch product does not appear to provide a big advantage in terms of peak intensities, which could be expected given that gauge adjustments are applied at a lower resolution and do not specifically target peak intensities. Also, note how in the Finnish product, rainfall peaks tend to be underestimated only slightly more (i.e., +11.2%) than the average. They also appear to converge faster to the average MB value than in the Dutch product. This is interesting and could point to the benefits of polarimetry. But there are many other factors to consider and more analyses are necessary to formally test this hypothesis.

We now turn to the second group of radar products (i.e., Denmark and Sweden) which is characterized by larger biases during the peaks. For Denmark, the median underestimation of the radar compared with the gauges is 66.2% (+29.1% with respect to the average MB). For Sweden, the median value is 54.9% (+15.1% with respect to the MB). The main difference compared with group 1 is that the bias affecting the peaks remains well above the average multiplicative bias across all aggregation time scales. The results for the Danish radar product are particularly interesting. According to our previous analyses, this product has the best overall agreement with gauges in terms of RRMSE and CC, mostly thanks to its high spatial and temporal resolution. It is therefore surprising to see that it contains such strong discrepancies in terms of peak intensities. Even the Swedish product, with its lower spatial and temporal resolution of 2 km and 15 min, shows a better agreement during the peaks. A possible explanation for this surprising result could be that the rain events in the Danish database are more intense and shorter than in the other countries. However, a closer analysis reveals a rank correlation coefficient between the PIB and peak intensity of only 0.20. Therefore, intensity is likely not the dominant factor at play here. Another explanation could be that bias-adjustment in the Danish radar product is performed on the basis of daily rainfall accumulations, which tends to smooth out peaks. Thorndahl et al. (2014a) showed that switching from daily to hourly mean field bias adjustments can slightly improve peak rainfall estimates but pointed out that hourly bias corrections tend to be problematic in times of low rain rates due to the small number of tips in the gauges. Therefore, in order to make a generally applicable adjustment that works for all rain conditions, the authors argued that it was better to use daily adjustments.

Finally, note that an alternative explanation for the higher peak intensity bias values in group 2 could be that Denmark and Sweden currently do not take advantage of multiple overlapping radar measurements during the rainfall estimation process. By contrast, the Dutch and Finnish radar products in group 1 are "true composites" that perform a weighted average of overlapping radar measurements depending on the quality of the measurement and the distance between the radar and the target. This could explain why the bias in peak rainfall intensity is only slightly larger than the overall average. It also suggests that the ability to combine measurements from multiple radars and viewpoints appears to play a crucial role in times of heavy rain, perhaps even more than spatial resolution.

3.4 Sensitivity to temporal aggregation time scale

Another equally interesting result of this study concerns the fact that biases in peak rainfall intensities do not necessarily become smaller when moving to a coarser scale. Figure 10 illustrates this point by showing, for the top event in each country, how much radar underestimates peak rainfall intensity compared with the gauge as a function of the temporal aggregation time scale. The time series corresponding to these 4 events were already shown in Figure 4.

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

While in the Netherlands and Finland the bias exponentially decays with aggregation time scale, the errors in Denmark and Sweden exhibit a much more complicated structure characterized by multiple ups and downs. Looking at the curve for event 1 in Denmark, we see that the peak intensity bias starts at 53.9% at 5 min, decreases to 52.4% at 10 min, increases again to 53.9% at the 15 min time scale, decreases until 43.8% at 35 min only to increase again to 50.2% at 45-50 min. The multiple ups and downs can be explained by the intermittent nature of this event, with 4 successive rainfall peaks separated by approximately 15-45 min (see Figure 4). Each of these peaks is characterized by different random observational errors, causing extremes at certain scales to be captured better than others. Because measurement errors in radar and gauges can be correlated in time, it is possible for the multiplicative bias to amplify over short aggregation time windows instead of converging to the mean value as would be expected if the observations were independent from each other. The same applies to the event in Sweden, where the peak intensity bias starts at 42.2% at 15 min, decreases to 40.1% at 30 min and increases again to 42.9% at 45 min. In this case, there is only one single rainfall peak. However, Figure 4 clearly shows 3 consecutive time steps during which the radar underestimates the rainfall rate. Together, these two examples for Denmark and Sweden show that even though globally speaking, the peak intensity bias between radar and gauges converges to the average multiplicative bias when data are aggregated over longer time periods, this might not always be the case locally and does not necessarily apply to all events.

The notion that multiplicative biases between radar and gauges can amplify when data are aggregated to coarser time scales is not new in itself but has important consequences for the representation of peak rainfall intensities in hydrological models as it affects the choice of the optimal spatial and temporal resolution at which models should be run when making flood predictions. An important finding of our study is that single-radar products are more vulnerable to error amplification due to the strong autocorrelation of the observation errors associated with using a single radar system. This can be verified by identifying, for each event, the time scale at which peak intensity bias was maximum, as shown in Figure 11. We see that out of the top 50 events in Denmark, 21 had maximum peak intensity bias at a scale larger than that of the highest available temporal resolution. Similarly, for the Swedish radar product, 26/50 cases of locally amplifying peak intensity biases could be identified. By contrast, the composite radar products in Finland and the Netherlands only contained 14 and 8 such events, respectively. A deeper analysis reveals that most of the identified cases consist of two or more rainfall peaks separated by 10-30 min, with rapidly fluctuating rainfall intensities between them (i.e., high intermittency). Alternatively, events consisting of one single rainfall peak during which radar was strongly underestimating for two or more time steps in a row are also possible. Most of the time, due to the limited temporal autocorrelation in heavy rain, the time scale of maximum peak intensity bias was limited to 30 minutes or less. However, there were also a few special unexplained cases in which peak intensity biases reached a maximum at time scales above 1-2 hours.

3.5 Results for additional radar products

Figure 12 summarizes the results obtained for the X-band radar system in Denmark. It shows that overall, there is a relatively good agreement between the X-band rainfall estimates and the gauges. The multiplicative bias at 5 min is only 1.20 (i.e., radar underestimates by 16.7%) and the correlation coefficient of 0.81 indicates good agreement in terms of the temporal structure. The relative root mean square error remains high (98.0%) but it is significantly smaller compared with the C-band products

(116-139%). The statistics for the X-band must be interpreted very carefully as only 10 events over 2 years were considered for the analyses (see Table A5 for more details). Still, the top right panel of Figure 12 shows that the peak intensities during these 10 events (i.e., 70-95 mmh^{-1}) were in the same order of magnitude than for the top 50 events in the Netherlands, Finland and Sweden (see Figure 7). The total accumulated rainfall amounts per event (i.e., 10-30 mm) were lower though, suggesting that the events sampled by the X-band system were rather short and localized.

Clearly, the high resolution of the X-band radar and the dual-polarization capabilities seem to improve the overall agreement between the radar and the gauges. Nevertheless, the bias affecting the peaks remains high. The median underestimation of peak rainfall intensity at 5 min was approximately 40%, which is slightly better than for the C-band products in the Netherlands and Finland and significantly better than for the C-band radar in Copenhagen. Still, the peaks appear to be affected by a bias that is more than twice as large as the average multiplicative bias, pointing to serious issues in times of heavy rain. This is consistent with our previous findings and suggests that resolution and polarimetry alone are not sufficient to accurately capture the peaks. Based on the analysis of the C-band products, one way to further reduce these biases during the peaks would be to use 2 or more overlapping X-band systems.

Figure 13 compares the agreement between the individual C-band radar products in Denmark, Finland and Sweden and the BALTRAD composite for the top 50 events in each country. The Netherlands are not included in this graph because they are not covered by the BALTRAD. To avoid sampling issues, all values are compared at the common temporal resolution of 15 min. The spatial resolutions, however, remain unchanged. Looking at the RRMSE, we see that the Finnish and Swedish products agree slightly better with the gauges than BALTRAD (-4.12% and -4.52% respectively) while the Danish agrees slightly worse (+2.47%). There are many possible explanations for these differences and each case needs to be analyzed separately. For Sweden, the interpretation is rather easy: the only major difference between the Swedish BRDC product and the BALTRAD lies in the additional bias-correction scheme implemented in HIPRAD. Otherwise, everything is identical. Thus we can say with high confidence that the reduction in RRMSE between BALTRAD and BRDC is likely due to the use of the bias-adjustment scheme. This, however, does not appear to improve significantly the bias affecting the peak rainfall intensities, as shown by the boxplots in the lower panel of Figure 13. The Finnish product shows similar improvements in RRMSE compared with the BALTRAD as well as a slightly lower spread in terms of peak intensity bias. However, since the Finnish OSAPOL product is not bias adjusted, other factors must be at play here. One of them could be the higher spatial resolution of the OSAPOL product compared with the BALTRAD. The other could be linked to the way rainfall rates are estimated, using polarimetry and phase information. And while it is impossible to say for sure which aspect contributed the most here, given our previous findings, we can say that differences are most likely due to the higher spatial resolution.

Finally, we turn our attention to Denmark. Results are more interesting there. We can see that the BALTRAD composite appears to agree slightly better with the rain gauges than the Danish C-band product. This is rather surprising given that the Danish product has the highest spatial resolution (500 m) of all 4 C-band products, making it the product with the lowest overall RRMSE and highest CC among all 4 considered C-band radar products. Still, the BALTRAD clearly agrees better with the gauges, improving the RRMSE by 2.47% and reducing the median peak intensity bias by 10.9 percentage points from 61.7% to 50.8%. The only negative aspect of the BALTRAD is its slightly higher spread in terms of peak rainfall intensity

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

bias, which is likely due to its lower spatial resolution of 2 km. We think that the main reason BALTRAD agrees better with the gauges in times of heavy rain is because it includes data from multiple radars in the greater Copenhagen region. This offers more flexibility compared with a single-radar setup and makes sure that the closest possible radar gets selected with respect to the position and characteristics of the storm. Note that although BALTRAD includes data from several radars, it is not a "full" composite product in the sense that it does not take advantage of overlapping radar measurements to perform merging and reduce measurement uncertainties. Still, even a simple multi-radar setup already appears to provide a clear advantage, highlighting the importance of designing robust and reliable algorithms for combining overlapping radar measurements in space and time. This is a research area that has been receiving more attention during the last decades but surprisingly, has not yet been implemented operationally in many countries.

4 Conclusions

Rain rate estimates from 6 different radar products in 4 countries (Denmark, Finland, the Netherlands and Sweden) have been analyzed. Special emphasis has been put on quantifying discrepancies between radar and gauges in times of heavy rain, focusing on the top 50 most intense events per country. A relatively good agreement was found in terms of temporal structure (correlation coefficient between 0.7-0.8). However, due to the large differences in sampling volume between gauges and radar, relative root mean square errors remained high (120-150% at 5-15 min). A substantial part of the discrepancies could be attributed to differences in spatial measurement support through the use of areal-reduction factors. The rest was attributed to systematic underestimation of rainfall rates by radar compared with the gauges. Together, the average underestimation reached 37.1% for Denmark, 29.1% for the Netherlands, 35.8% for Finland and 39.8% for Sweden. Furthermore, the underestimation has been shown to increase with intensity, reaching on average 45.9% to 66.2% at the time of the peak. Bias correction using surrounding rain gauges did not appear to have a big impact on peak intensity bias.

On average, the radar products with higher spatial resolutions were in better agreement with the gauges, thereby confirming the importance of high-resolution radar observations in hydrological studies. The X-band data for Denmark showed very promising results, outperforming all other products in terms of accuracy and correlation. However, this last result must be interpreted very carefully as only 10 events over 2 years were considered for the X-band radar analysis. Polarimetry also seemed to provide a slight advantage in times of heavy rain. However, due to the many confounding factors, it is hard to precisely quantify its added-value within the framework of this study. What we can say with high confidence is that dual-polarization and higher resolution alone are not sufficient to get reliable estimates of peak rainfall intensities. Other factors such as the ability to combine data from multiple radars and viewpoints seem to play a much more important role, as demonstrated by the superior performance of the Dutch and Finnish C-band products (despite their slightly lower resolution). By contrast, the single-radar C-band product in Denmark, which had the highest spatial resolution (i.e., 500 m) and lowest overall RRMSE, did not perform well on the peaks at all, exhibiting the highest peak intensity biases of all 6 products. Even the lower resolution BALTRAD composite (2 km, 15 min) over Denmark performed better.

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

Another important finding of this paper was that the largest bias between radar and gauges in terms of peak intensities does not necessarily occur at the highest temporal sampling resolution. Depending on the autocorrelation structure of the errors, multiplicative biases may amplify over time instead of converging to the mean value. This mostly happens at the sub-hourly time scales and roughly affects 40-50% of all events in single-radar products and 15-30% in composite products. Most of these cases were characterized by a succession of multiple rainfall peaks or alternatively, one very intense peak of 15-30 min during which radar strongly underestimated the intensity for 2 or more consecutive time steps. The strong dependence of the error structure on the underlying aggregation time scale has already been pointed out in the past, but still represents a major challenge in terms of how to correctly represent rainfall extremes and rainfall measurement uncertainties in hydrological models.

Finally, like with any statistical analysis, there are a few important limitations in the methodology that need to be mentioned. The first is that all performance metrics provided in this paper are based on the assumption that rain gauges constitute a reliable reference for assessing the radar estimates. In reality, gauges also suffer from measurement uncertainties and errors, the most common being an underestimation of rainfall rates in times of heavy precipitation due to calibration issues and wind effects. Therefore, actual biases and errors might be even larger than suggested by the analyses. No attempt has been made to correct for these additional biases nor to distinguish between gauge and radar-induced errors. Instead, only the differences between the two measurements have been analyzed. This was done with the goal to analyze and compare different radar products without making any statement about which one of the two is closer to the "truth". The second limitation of this study is that differences between gauges and radar likely depend on gauge location and distance from the radar. Such subtle effects could not be documented here as the number of events was too low and most gauges were only used for a single event (see Table 1). The last limitation worth mentioning is the lack of a common denominator for comparing the individual radar products. Because all 6 radar products were different from each other, and events of different duration and intensities were considered, we were not able to precisely quantify the individual merits of high-resolution, polarimetry, compositing and bias adjustments. Future studies involving a larger number of products and different levels of processing (e.g., by switching on/off individual correction schemes) for identical radar systems would help to get a more detailed view into the strengths and weaknesses of individual techniques. Future work will focus on these issues to help national agencies monitor and improve the performance of their precipitation products and make good strategic choices when upgrading their systems.

Acknowledgements. The authors acknowledge funding by the EU within the framework of the ERA-NET Cofund WaterWorks2014 project MUFFIN (Multiscale Flood Forecasting: From Local Tailored Systems to a Pan-European Service). This ERA-NET is an integral part of the 2015 Joint Activities developed by the Water Challenges for a Changing World Joint Programme Initiative (Water JPI). The first author acknowledges funding by the Netherlands Organisation for Scientific Research NWO (project code ALWWW.2014.3). The Finnish partners acknowledge funding by the Maa- ja vesitekniikan tuki ry. foundation. The Optimal Rain Products with Dual-Pol Doppler Weather Radar (OSAPOL) project was funded by the European Regional Development Fund and Business Finland. The authors would like to thank the Danish, Finnish, Swedish and Dutch Meteorological Institutes (i.e., DMI, FMI, SMHI and KNMI) for collecting and distributing the radar and gauge data used in this study.

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

Data availability. The Dutch radar products are available for free in HDF5 format through the FTP of KNMI or in netCDF4 format via the Climate4Impact website. The Danish, Swedish and Finnish products are not open yet but can be made available for research purposes upon request to the authors.

Competing interests. The authors declare that they have no competing interests.

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

References

Bech, J., Codina, B., Lorente, J., and Bebbington, D.: The Sensitivity of Single Polarization Weather Radar Beam Blockage Correction to Variability in the Vertical Refractivity Gradient, J. Atmos. Oceanic Technol., 20, 845–855, https://doi.org/10.1175/1520-0426(2003)020<0845:TSOSPW>2.0.CO;2, 2003.

Berg, P., Norin, L., and Olsson, J.: Creation of a high resolution precipitation data set by merging gridded gauge data and radar observations for Sweden, J. Hydrol., 541, 6–13, https://doi.org/10.1016/j.jhydrol.2015.11.031, 2016.

Berne, A. and Krajewski, W. F.: Radar for hydrology: Unfulfilled promise or unrecognized potential?, Adv. Water Resour., 51, 357–366, https://doi.org/10.1016/j.advwatres.2012.05.005, 2013.

Berne, A., Delrieu, G., Creutin, J.-D., and Obled, C.: Temporal and spatial resolution of rainfall measurements required for urban hydrology, J. Hydrol., 299, 166–179, https://doi.org/10.1016/j.jhydrol.2004.08.002, 2004.

Blenkinsop, S., Lewis, E., Chan, S. C., and Fowler, H. J.: Quality-control of an hourly rainfall dataset and climatology of extremes for the UK, Int. J. Climatol., 37, 722–740, https://doi.org/10.1002/joc.4735, 2017.

Brandes, E. A., Ryzhkov, A. V., and Zrnic, D. S.: An evaluation of radar rainfall estimates from specific differential phase, J. Atmos. Oceanic Technol., 18, 363–375, https://doi.org/10.1175/1520-0426(2001)018<0363:AEORRE>2.0.CO;2, 2001.

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

Collier, C. G. and Knowles, J. M.: Accuracy of rainfall estimates by radar, part III: application for short-term flood forecasting, J. Hydrol., 83, 237–249, https://doi.org/10.1016/0022-1694(86)90154-X, 1986.

Dai, Q. and Han, D.: Exploration of discrepancy between radar and gauge rainfall estimates driven by wind fields, Water Resour. Res., 50, 8571–8588, https://doi.org/10.1002/2014WR015794, 2014.

Einfalt, T., Arnbjerg-Nielsen, K., Golz, C., Jensen, N. E., Quirmbach, M., Vaes, G., and Vieux, B.: Towards a roadmap for use of radar rainfall data in urban drainage, J. Hydrol., 299, 186–202, https://doi.org/10.1016/j.jhydrol.2004.08.004, 2004.

Fairman, J. G., Schultz, D. M., Kirshbaum, D. J., Gray, S. L., and Barrett, A. I.: Climatology of Size, Shape, and Intensity of Precipitation Features over Great Britain and Ireland, J. Hydrometeorol., 18, 1595–1615, https://doi.org/10.1175/JHM-D-16-0222.1, 2017.

Gill, R. S., Overgaard, S., and Bøvith, T.: The Danish weather radar network, in: Proceedings of Fourth European Conference on Radar in Meteorology and Hydrology (ERAD), pp. 1–4, Barcelona, Spain, 2006.

Goudenhoofdt, E. and Delobbe, L.: Evaluation of radar-gauge merging methods for quantitative precipitation estimates, Hydrol. Earth Syst. Sci., 13, 195–203, https://doi.org/10.5194/hess-13-195-2009, 2009.

Goudenhoofdt, E., Delobbe, L., and Willems, P.: Regional frequency analysis of extreme rainfall in Belgium based on radar estimates, Hydrol. Earth Syst. Sci., 21, 5385–5399, https://doi.org/10.5194/hess-21-5385-2017, 2017.

Gourley, J. J., Tabary, P., and Parent-du Chatelet, J.: Data quality of the Meteo-France C-band polarimetric radar, J. Atmos. Oceanic Technol., 23, 1340–1356, https://doi.org/10.1175/JTECH1912.1, 2006.

Gourley, J. J., Tabary, P., and Parent-du Chatelet, J.: A fuzzy logic algorithm for the separation of precipitating from nonprecipitating echoes using polarimetric radar observations, J. Atmos. Oceanic Technol., 24, 1439–1451, https://doi.org/10.1175/JTECH2035.1, 2007.

Huuskonen, A., Saltikoff, E., and Holleman, I.: The Operational Weather Radar Network in Europe, Bull. Amer. Meteor. Soc., 95, 897–907, https://doi.org/10.1175/BAMS-D-12-00216.1, 2014.

Krajewski, W. F.: Cokriging radar-rainfall and rain-gauge data, J. Geophys. Res. Atmos., 90, 9571–9580, https://doi.org/10.1029/JD092iD08p09571, 1987.

Krajewski, W. F. and Smith, J. A.: Radar hydrology: rainfall estimation, Adv. Water Resour., 25, 1387–1394, https://doi.org/10.1016/j.advwatres.2005.03.018, 2002.

Krajewski, W. F., Villarini, G., and Smith, J. A.: RADAR-Rainfall Uncertainties: Where are we after Thirty Years of Effort?, Bull. Amer. Meteor. Soc., 91, 87–94, https://doi.org/10.1175/2009BAMS2747.1, 2010.

Lee, G.: Sources of errors in rainfall measurements by polarimetric radar: variability of drop size distributions, observational noise, and variation of relationships between R and polarimetric parameters, J. Atmos. Oceanic Technol., 23, 1005–1028, 2006.

Leinonen, J., Moisseev, D., Leskinen, M., and Petersen, W. A.: A Climatology of Disdrometer Measurements of Rainfall in Finland over Five Years with Implications for Global Radar Observations, J. Appl. Meteor. Clim., 51, 392–404, https://doi.org/10.1175/JAMC-D-11-056.1, 2012.

Madsen, H., Mikkelsen, P. S., Rosbjerg, D., and Harremoës, P.: Estimation of regional intensity-duration-frequency curves for extreme precipitation, Water Sci. Technol., 37, 29–36, https://doi.org/10.1016/S0273-1223(98)00313-8, 1998.

Matrosov, S. Y., Cifelli, R., Kennedy, P. C., Nesbitt, S. W., Rutledge, S. A., Bringi, V. N., and Martner, B. E.: A comparative study of rainfall retrievals based on specific differential phase shifts at X- and S-band radar frequencies, J. Atmos. Oceanic Technol., 23, 952–963, https://doi.org/10.1175/JTECH1887.1, 2006.

Matrosov, S. Y., Clark, K. A., and Kingsmill, D. E.: A polarimetric radar approach to identify rain, melting-layer, and snow regions for applying corrections to vertical profiles of reflectivity, J. Appl. Meteor. Clim., 46, 154–166, 2007.

Michelson, D.: The Swedish weather radar production chain, in: Proceedings of Fourth European Conference on Radar in Meteorology and Hydrology (ERAD), pp. 382–385, Barcelona, Spain, 2006.

Michelson, D., Henja, A., Ernes, S., Haase, G., Koistinen, J., Ośródka, K., Peltonen, T., Szewczykowski, M., and Szturc, J.: BALTRAD Advanced Weather Radar Networking, Journal of Open Research Software, 6, 1–12, https://doi.org/10.5334/jors.193, 2018.

Nielsen, J. E., Thorndahl, S. L., and Rasmussen, M. R.: A Numerical Method to Generate High Temporal Resolution Precipitation Time Series by Combining Weather Radar Measurements with a Nowcast Model, Atmos. Res., 138, 1–12, https://doi.org/10.1016/j.atmosres.2013.10.015, 2014.

Norin, L., Devasthale, A., L'Ecuyer, T. S., Wood, N. B., and Smalley, M.: Intercomparison of snowfall estimates derived from the CloudSat Cloud Profiling Radar and the ground-based weather radar network over Sweden, Atmos. Meas. Tech., 8, 5009–5021, https://doi.org/10.5194/amt-8-5009-2015, 2015.

Nystuen, J. A.: Relative performance of automatic rain gauges under different rainfall conditions, J. Atmos. Oceanic Technol., 16, 1025–1043, https://doi.org/10.1175/1520-0426(1999)016<1025:RPOARG>2.0.CO;2, 1999.

Otto, T. and Russchenberg, H. W. J.: Estimation of specific differential phase and differential backscatter phase from polarimetric weather radar measurements of rain, IEEE Geosci. Remote Sens. Lett., 8, 988–992, https://doi.org/10.1109/LGRS.2011.2145354, 2011.

Overeem, A., Buishand, T. A., and Holleman, I.: Extreme rainfall analysis and estimation of depth-duration-frequency curves using weather radar, Water Resour. Res., 45, W10424, https://doi.org/10.1029/2009WR007869, 2009a.

Overeem, A., Holleman, I., and Buishand, T. A.: Derivation of a 10-year radar-based climatology of rainfall, J. Appl. Meteor. Clim., 48, 1448–1463, https://doi.org/10.1175/2009JAMC1954.1, 2009b.

Overeem, A., Buishand, T. A., Holleman, I., and Uijlenhoet, R.: Extreme value modeling of areal rainfall from weather radar, Water Resour. Res., 46, W09514, https://doi.org/10.1029/2009WR008517, 2010.

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

Pollock, M. D., O'Donnell, G., Quinn, P., Dutton, M., Black, A., Wilkinson, M., Colli, M., Stagnaro, M., Lanza, L. G., Lewis, E., Kilsby, C. G., and O'Connell, P. E.: Quantifying and Mitigating Wind-Induced Undercatch in Rainfall Measurements, Water Resour. Res., 54, 3863–3875, https://doi.org/10.1029/2017WR022421, 2018.

Rickenbach, T. M., Nieto-Ferreira, R., Zarzar, C., and Nelson, B.: A seasonal and diurnal climatology of precipitation organization in the southeastern United States, Q. J. R. Meteorol. Soc., 141, 1938–1956, https://doi.org/10.1002/qj.2500, 2015.

Ryzhkov, A. and Zrnic, D. S.: Assessment of rainfall measurement that uses specific differential phase, J. Appl. Meteorol., 35, 2080–2090, https://doi.org/10.1175/1520-0450(1996)035<2080:AORMTU>2.0.CO;2, 1996.

Ryzhkov, A. V. and Zrnic, D. S.: Discrimination between rain and snow with a polarimetric radar, J. Appl. Meteorol., 37, 1228–1240, 1998.

Seo, B.-C., Dolan, B., Krajewski, W. F., Rutledge, S. A., and Petersen, W.: Comparison of Single- and Dual-Polarization-Based Rainfall Estimates Using NEXRAD Data for the NASA Iowa Flood Studies Project, J. Hydrometeorol., 16, 1658–1675, https://doi.org/10.1175/JHM-D-14-0169.1, 2015.

Sieck, L. C., Burges, S. J., and Steiner, M.: Challenges in obtaining reliable measurements of point rainfall, Water Resour. Res., 43, W01 420, https://doi.org/10.1029/2005WR004519, 2007.

Smith, J. A. and Krajewski, W. F.: Estimation of the Mean Field Bias of Radar Rainfall Estimates, J. Appl. Meteorol., 30, 397–412, https://doi.org/10.1175/1520-0450(1991)030<0397:EOTMFB>2.0.CO;2, 1991.

Smith, J. A., Baeck, M. L., Meierdiercks, K. L., Miller, A. J., and Krajewski, W. F.: Radar rainfall estimation for flash flood forecasting in small urban watersheds, Adv. Water Resour., 30, 2087–2097, https://doi.org/10.1016/j.advwatres.2006.09.007, 2007.

Smith, J. A., Baeck, M. L., Villarini, G., Welty, C., Miller, A. J., and Krajewski, W. F.: Analyses of a long-term, high-resolution radar rainfall data set for the Baltimore metropolitan region, Water Resour. Res., 48, W04504, https://doi.org/10.1029/2011WR010641, 2012.

Stevenson, S. N. and Schumacher, R. S.: A 10-Year Survey of Extreme Rainfall Events in the Central and Eastern United States Using Gridded Multisensor Precipitation Analyses, Mon. Wea. Rev., 142, 3147–3162, https://doi.org/10.1175/MWR-D-13-00345.1, 2014.

Thorndahl, S., Nielsen, J. E., and Rasmussen, M. R.: Bias adjustment and advection interpolation of long-term high resolution radar rainfall series, J. Hydrol., 508, 214–226, https://doi.org/10.1016/j.jhydrol.2013.10.056, 2014a.

Thorndahl, S., Smith, J. A., Baeck, M. L., and Krajewski, W. F.: Analyses of the temporal and spatial structures of heavy rainfall from a catalog of high-resolution radar rainfall fields, Atmos. Res., 144, 111–125, https://doi.org/10.1016/j.atmosres.2014.03.013, 2014b.

Thorndahl, S., Einfalt, T., Willems, P., Nielsen, J. E., ten Veldhuis, M.-C., Arnbjerg-Nielsen, K., Rasmussen, M. R., and Molnar, P.: Weather radar rainfall data in urban hydrology, Hydrol. Earth Syst. Sci., 21, 1359–1380, https://doi.org/10.5194/hess-21-1359-2017, 2017.

Thorndahl, S., Nielsen, J. E., and Rasmussen, M. R.: Estimation of Storm-Centred Areal Reduction Factors from Radar Rainfall for Design in Urban Hydrology, Water, 11, https://doi.org/10.3390/w11061120, 2019.

Tian, Y., Huffman, G. J., Adler, R. F., Tang, L., Sapiano, M., Maggioni, V., and Wu, H.: Modeling errors in daily precipitation measurements: Additive or multiplicative?, Geophys. Res. Lett., 40, 2060–2065, https://doi.org/10.1002/grl.50320, 2013.

Vasiloff, S. V., Howard, K. W., and Zhang, J.: Difficulties with correcting radar rainfall estimates based on rain gauge data: a case study of severe weather in Montana on 16-17 June 2007, Weather Forecast., 24, 1334–1344, https://doi.org/10.1175/2009WAF2222154.1, 2009.

Villarini, G. and Krajewski, W. F.: Review of the Different Sources of Uncertainty in Single Polarization Radar-Based Estimates of Rainfall, Surveys in Geophysics, 31, 107–129, 2010.

Wójcik, O. P., Holt, J., Kjerulf, A., Müller, L., Ethelberg, S., and Molbak, K.: Personal protective equipment, hygiene behaviours and occupational risk of illness after July 2011 flood in Copenhagen, Denmark, Epidemiology and Infection, 141, 1756–1763, https://doi.org/10.1017/S0950268812002038, 2013.

Wood, S. J., Jones, D. A., and Moore, R. J.: Accuracy of rainfall measurements for scales of hydrological interest, Hydrol. Earth Syst. Sci., 4, 531–543, https://doi.org/10.5194/hess-4-531-2000, 2000.

Wright, D. B., Smith, J. A., Villarini, G., and Baeck, M. L.: Long-Term High-Resolution Radar Rainfall Fields for Urban Hydrology, J. Amer. Water Resour. Assoc., 50, 713–734, https://doi.org/10.1111/jawr.12139, 2014.

Yoo, C., Park, C., Yoon, J., and Kim, J.: Interpretation of mean-field bias correction of radar rain rate using the concept of linear regression, Hydrol. Processes, 28, 5081–5092, https://doi.org/10.1002/hyp.9972, 2014.

Young, C. B., Bradley, A. A., Krajewski, W. F., Kruger, A., and Morrisey, M. L.: Evaluating NEXRAD multisensor precipitation estimates for operational hydrologic forecasting, J. Hydrometeorol., 1, 241–254, 2000.

Zrnic, D. S. and Ryzhkov, A. V.: Advantages of rain measurements using specific differential phase, J. Atmos. Oceanic Technol., 13, 454–464, https://doi.org/10.1175/1520-0426(1996)013<0454:AORMUS>2.0.CO;2, 1996.

Zrnic, D. S. and Ryzhkov, A. V.: Polarimetry for weather surveillance radars, Bull. Amer. Meteor. Soc., 80, 389–406, 1999.

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

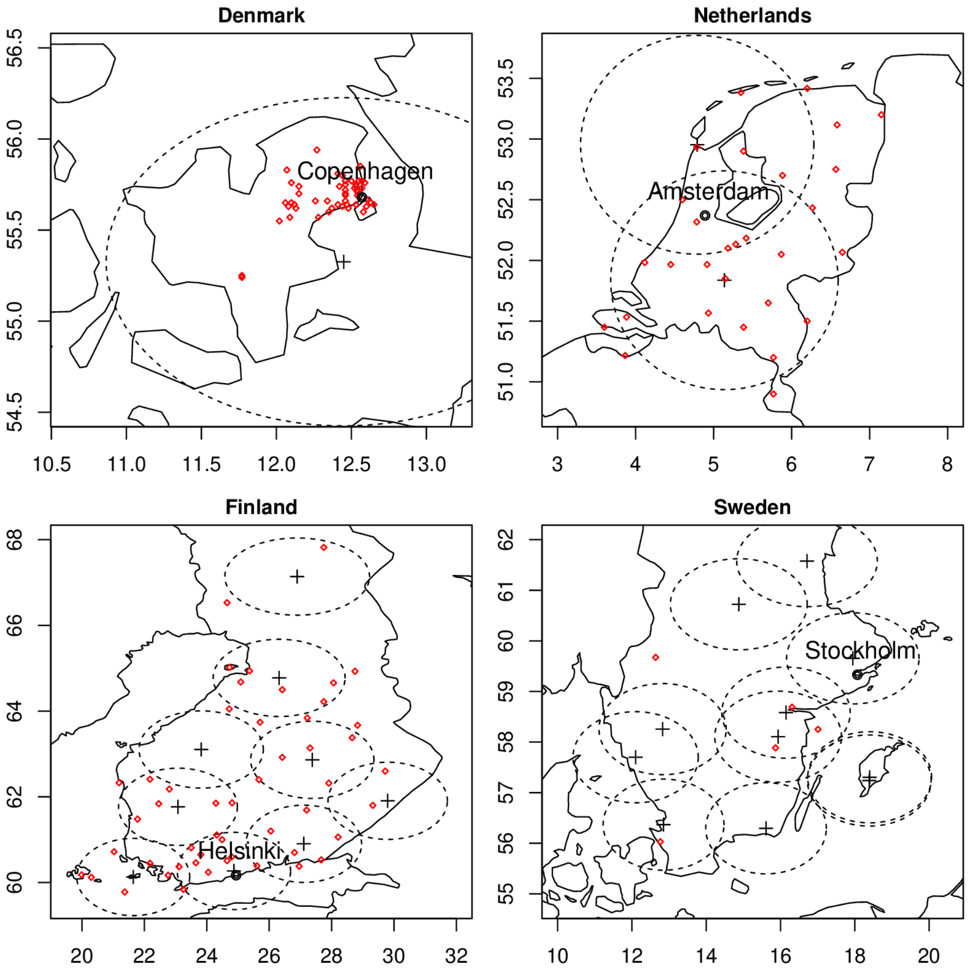

Figure 1. The four considered study areas in Denmark, the Netherlands, Finland and Sweden with the used rain gauges (red diamonds) and the location of the C-band radars (black crosses). The dashed lines denote circles of 100 km radius around each radar.

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

Hydrology and Earth System Sciences
Open Access
Discussions

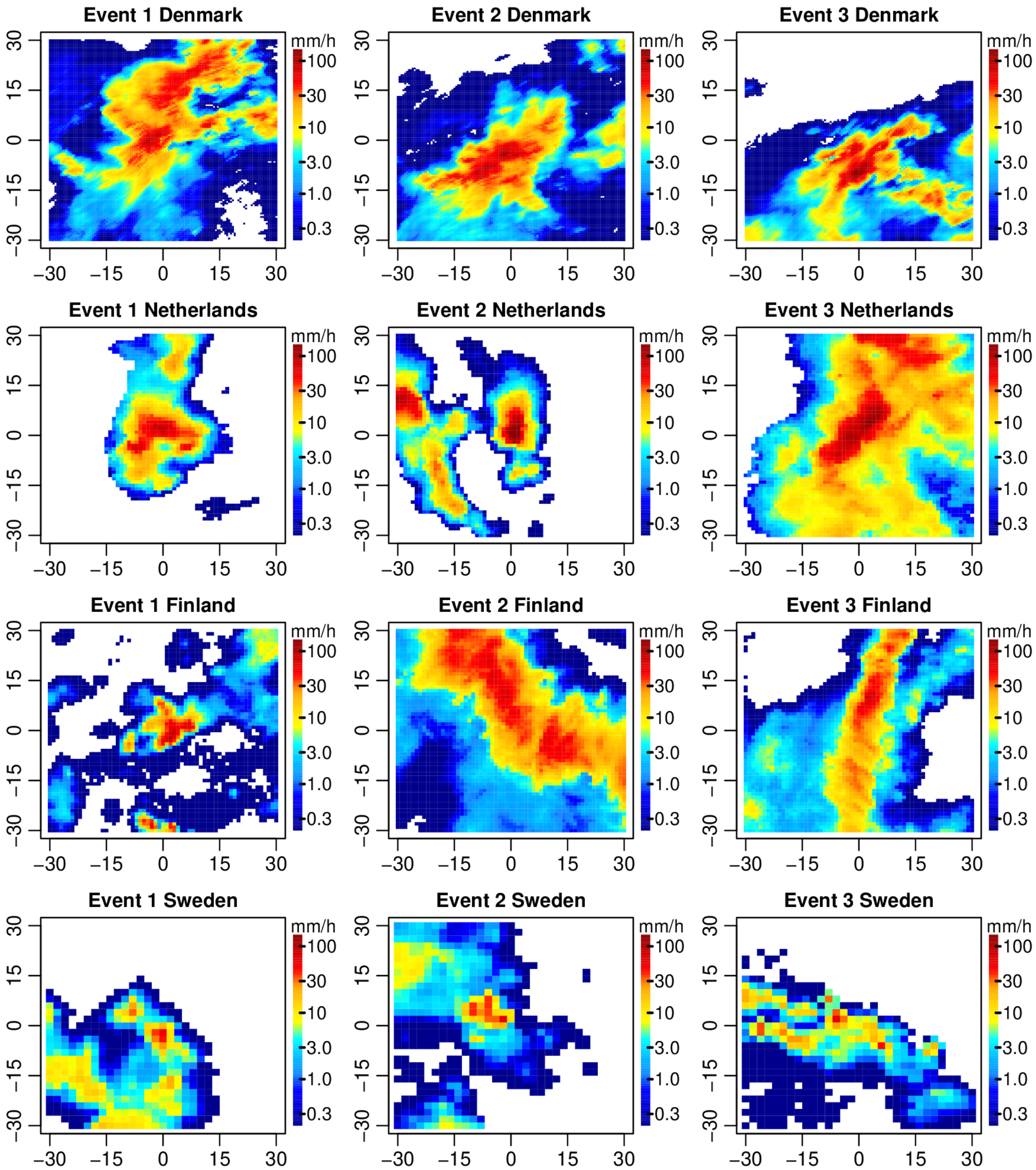

Figure 2. Snapshots of the 3 most intense events for each country at the time of peak intensity (in mm/h). Each map is a square of size 60×60 km^2 with the gauge location in the center.

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

Hydrology and Earth System Sciences
Open Access
Discussions

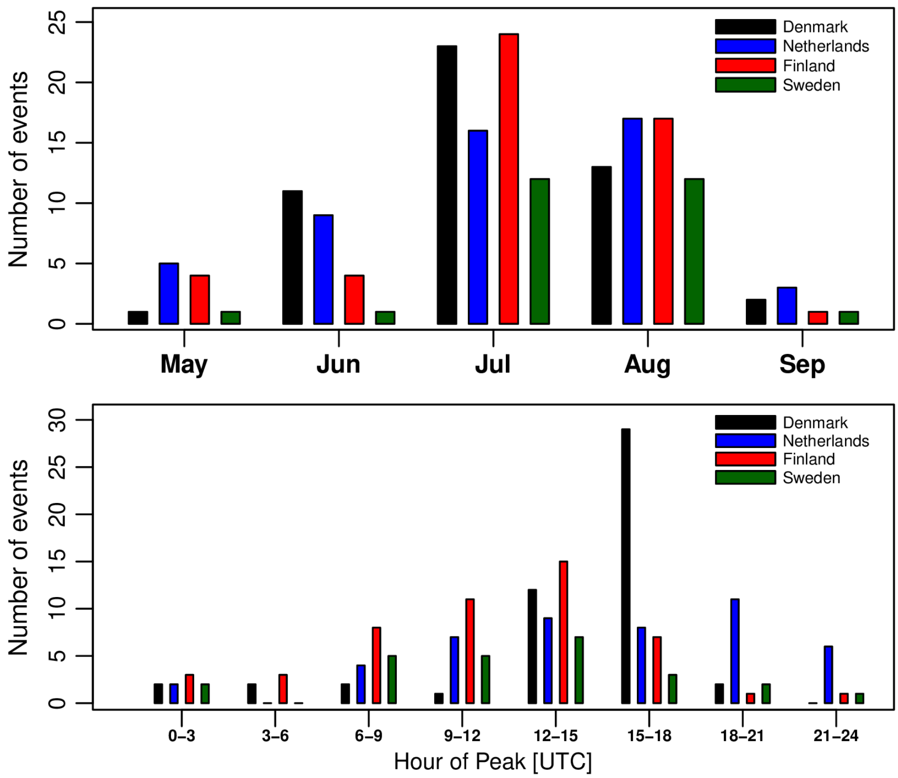

Figure 3. Distribution of the 50 top events over the month (top panel) and hour of the day (bottom panel).

Hydrology and Earth System Sciences
Open Access
Discussions

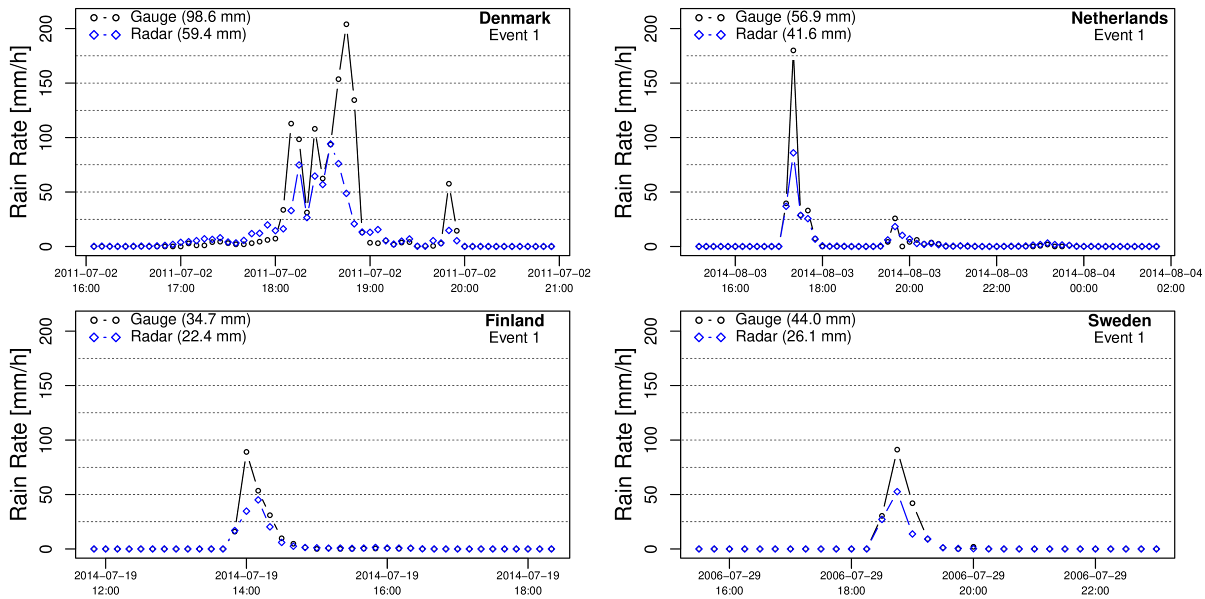

Figure 4. Time series of radar and gauge intensities (in mm/h) at the highest available temporal resolution for the most intense event of each country.

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

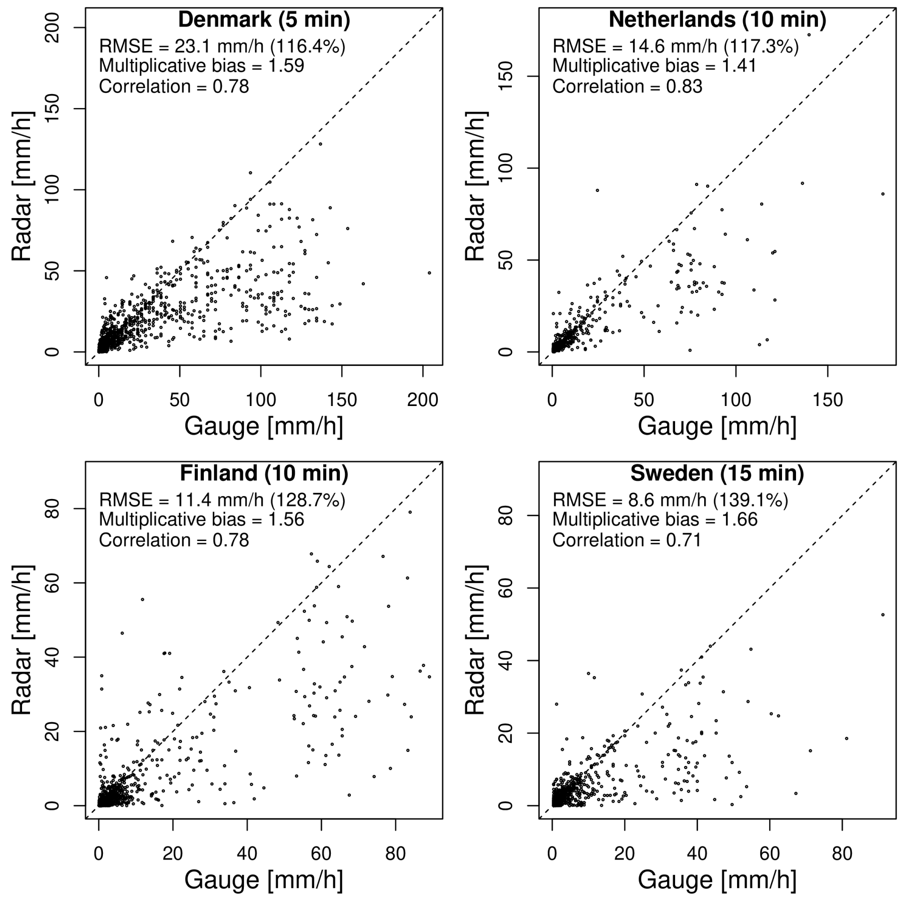

Figure 5. Radar versus gauge intensities (in mm/h) at the highest available temporal resolution for each country (all 50 events combined). The dotted line represents the diagonal.

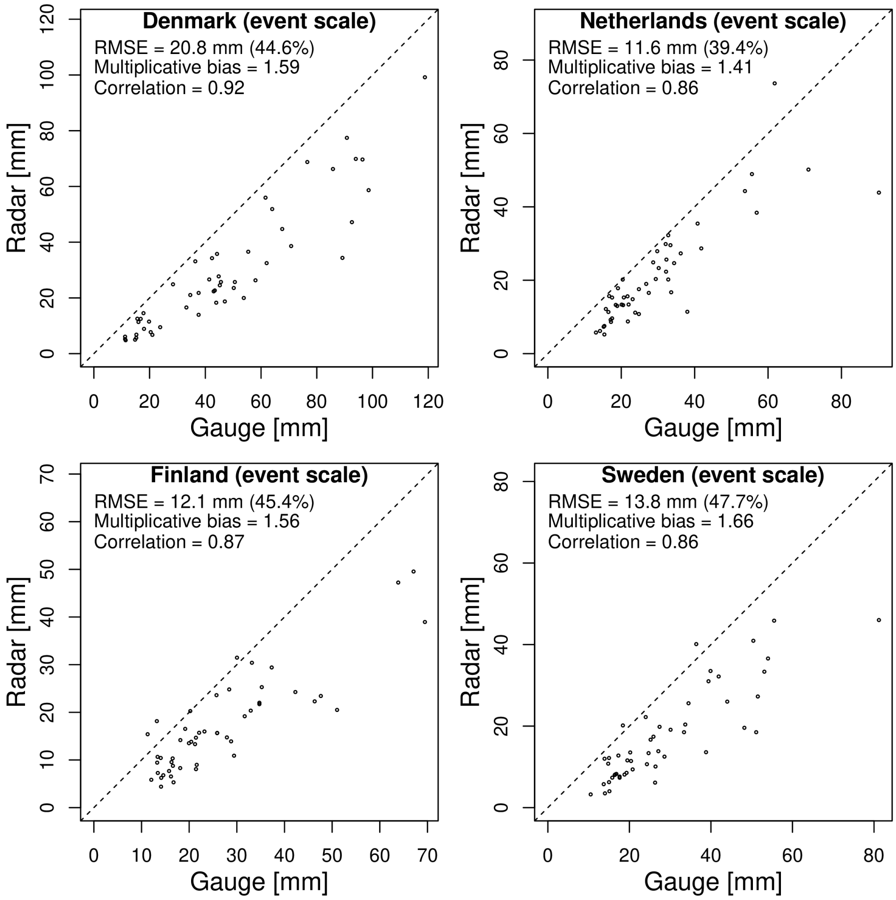

Figure 6. Radar versus gauge accumulations (in mm) at the event scale for each country (i.e., one dot per event). The dotted line represents the diagonal.

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

Hydrology and Earth System Sciences Discussions

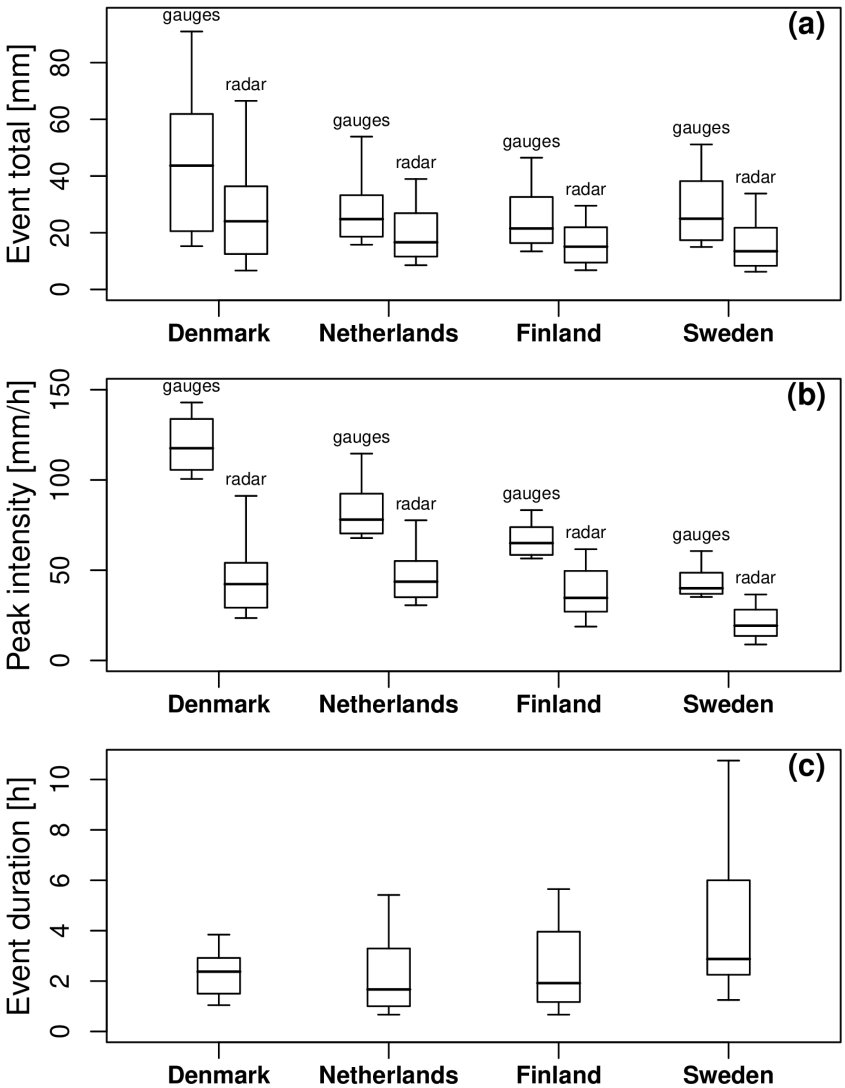

Figure 7. Total rainfall accumulations (in mm), peak rainfall intensities (in mmh^{-1}) and event duration (in hours) for each country. The boxplots denote the 10%, 25%, 50%, 75% and 90% quantiles of the 50 events in each country.

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

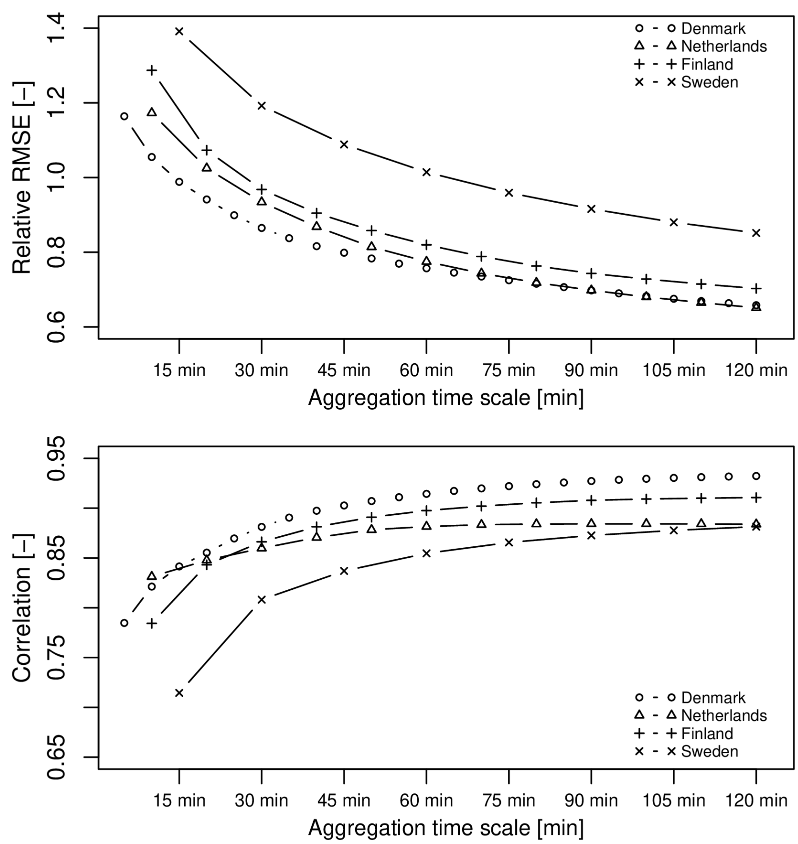

Figure 8. Relative root mean square error and correlation coefficients of radar versus rain gauge estimates at different aggregation time scales between 5 min and 2 h (all 50 events combined).

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

Hydrology and Earth System Sciences
Discussions

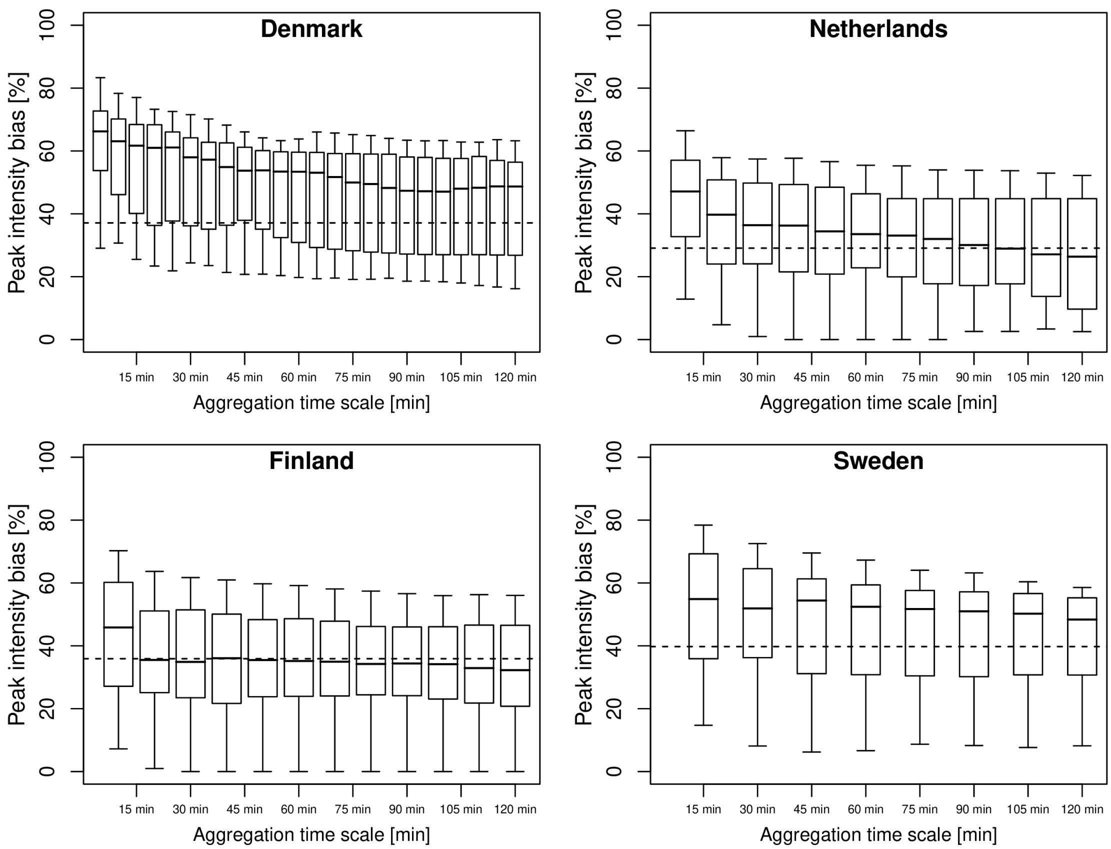

Figure 9. Underestimation of peak rainfall intensity by radar compared with gauges (expressed in %) versus temporal aggregation time scale. Each boxplot represents the 10%, 25%, 50%, 75% and 90% quantiles for the 50 top events in each country. The horizontal lines represent the average multiplicative bias values for each country.

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

Hydrology and Earth System Sciences
Open Access
Discussions

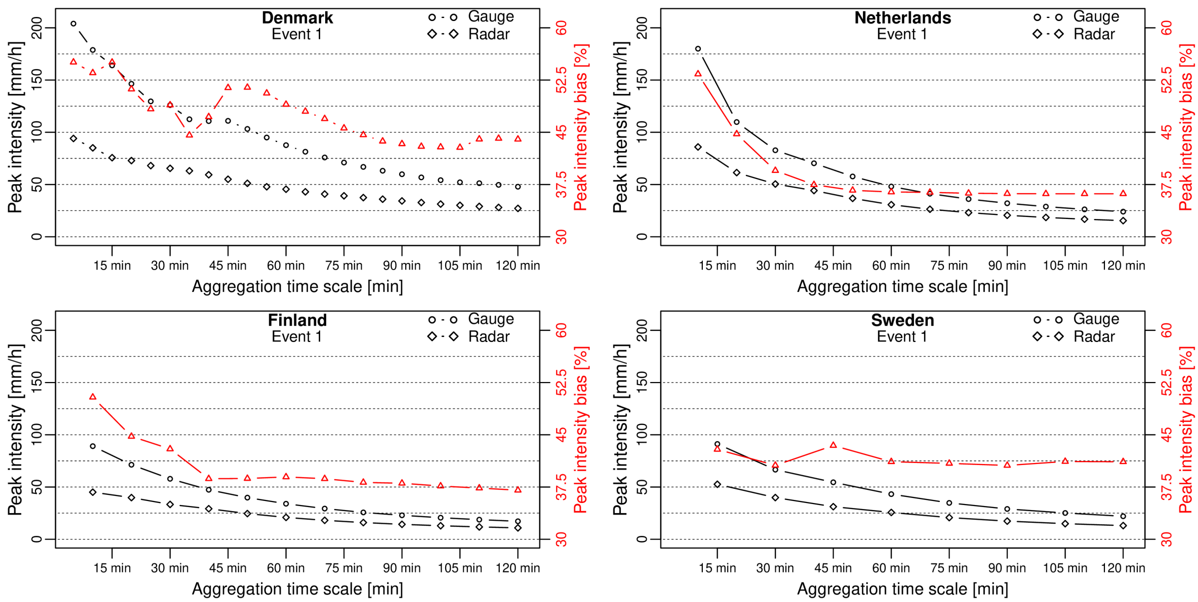

Figure 10. Peak rainfall intensities measured by radar and gauges for the top 1 event in each country. The red triangles show the peak rainfall intensity bias between radar and gauges as a function of the aggregation time scale (axis on the right).

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

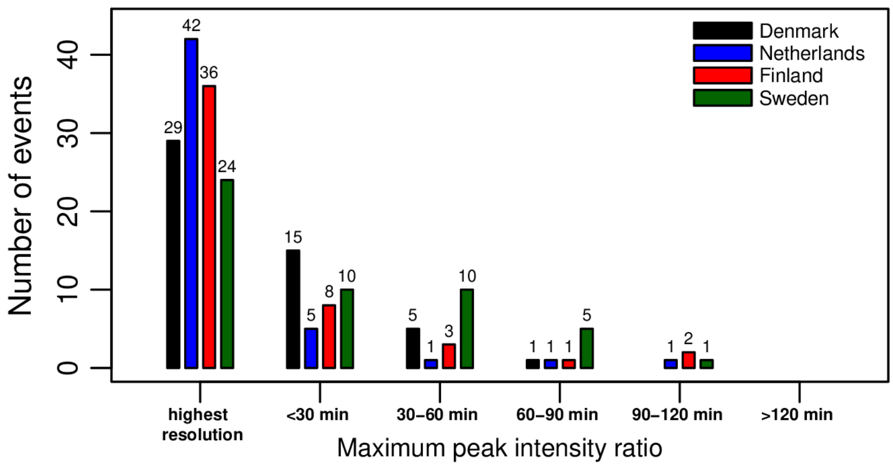

Figure 11. Aggregation time at which the maximum error on peak intensity between gauge and radar occurred.

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

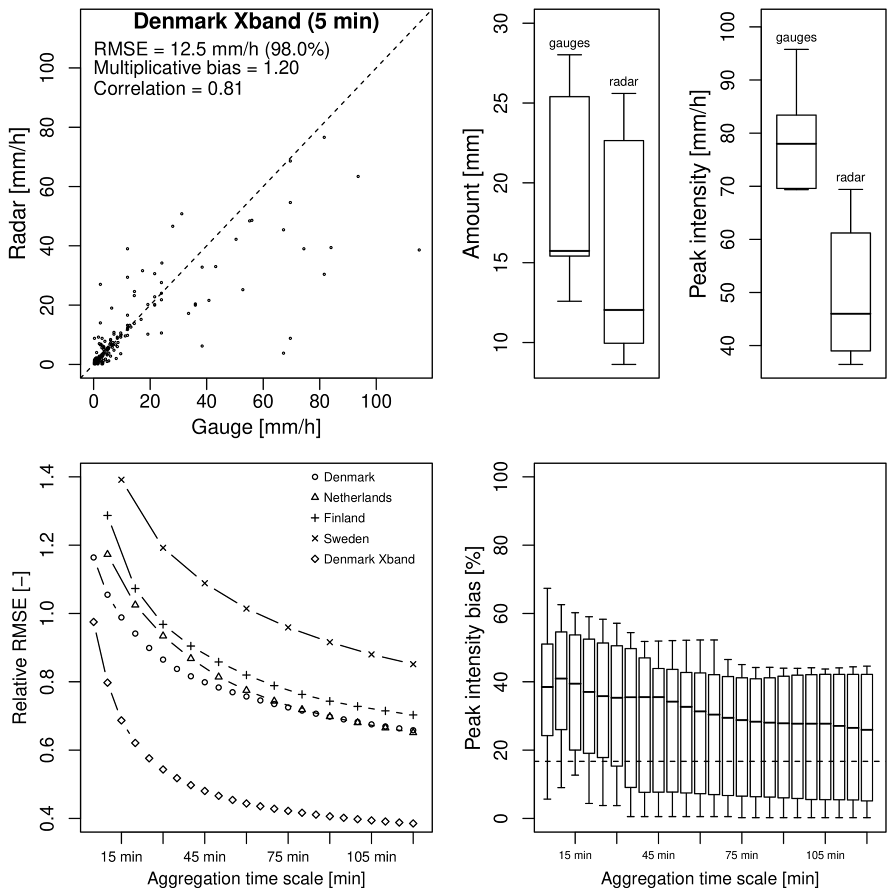

Figure 12. Performance metrics for the Danish X-band radar system (top 10 events).

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

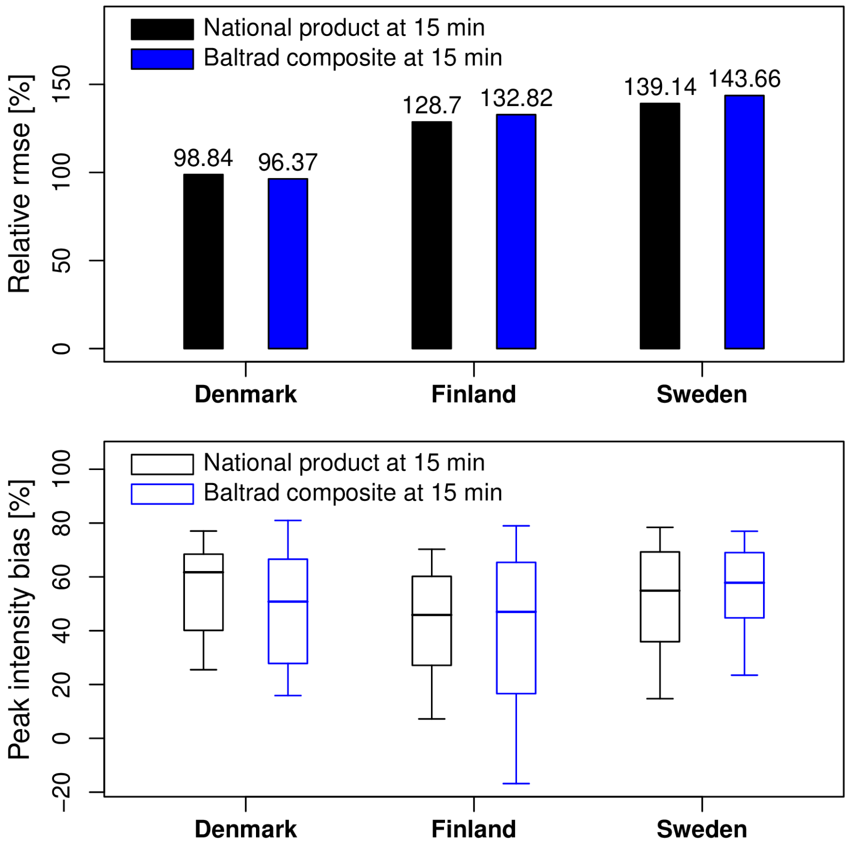

Figure 13. Comparison of relative root mean square error and peak intensity ratios (at 15 min resolution) between the national radar products and the BALTRAD composite.

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

Table 1. Rain gauge datasets used to determine the top 50 rainfall events for each country. The time periods were chosen based on radar data availability.

	Denmark	Netherlands	Finland	Sweden
Number of available gauges	66	35	64	10
Gauges used for top 50 events	50	31	50	5
Time period	2003–2016	2008-2018	2013-2016	2000–2018
Gauge sampling resolution	5 min	10 min	10 min	15 min

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

Hydrology and Earth System Sciences
Open Access
Discussions

Table 2. Radar products used in this study.

Country	Radar type(s)	Resolution	Method	Bias correction
Denmark	1 single-pol C-band	500×500 m, 5 min	Z-R	yes
Netherlands	2 single-pol C-band	1×1 km, 5 min	Z-R	yes
Finland	9 dual-pol C-band	1×1 km, 5 min	Z-R and Kdp	no
Sweden	12 single-pol C-band	2×2 km, 15 min	Z-R	yes
Denmark	1 dual-pol X-band	100×100 m, 1 min	Z-R	yes
Baltic region	C-band (BALTRAD)	2×2 km, 15 min	Z-R	yes

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

Table 3. Summary statistics for the 4 main radar products at the highest available spatial and temporal resolution. Correlation coefficient (CC), relative root mean square error (RRMSE), multiplicative bias (MB) and area reduction factor (ARF) between a point measurement and a radar pixel (expressed as a percentage).

Country	CC	RRMSE	MB	ARF
Denmark (500 m, 5 min)	0.78	116.4%	37.1%	12.8%
Netherlands (1 km, 10 min)	0.83	117.3%	29.1%	18.6%
Finland (1 km, 10 min)	0.78	128.7%	35.8%	18.6%
Sweden (2 km, 15 min)	0.71	139.1%	39.8%	29.6%

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

Hydrology and Earth System Sciences
Open Access
Discussions

1 Appendix: Top 50 events for each country

Table A1. Top 50 events for Denmark

Event	Starting Time [UTC]	Gauge	Duration	Amount [mm]	Peak [mmh^{-1}]
1	2011-07-02 17:05	5805	2h50min	98.6	204.0
2	2011-07-02 17:20	5725	2h10min	92.6	163.2
3	2011-07-02 17:10	5685	2h25min	89.2	148.8
4	2013-08-10 17:25	5675	30min	15.2	144.0
5	2006-08-15 05:55	5901	11h45min	20.4	144.0
6	2011-07-02 17:10	5730	2h25min	94.0	142.8
7	2011-07-02 16:55	5740	2h50min	118.8	141.6
8	2016-07-25 16:30	5590	35min	23.8	139.2
9	2011-07-02 17:00	5785	2h50min	96.4	136.8
10	2011-07-02 17:15	5675	2h15min	37.6	134.4
11	2007-08-11 13:05	5790	2h35min	67.6	134.4
12	2007-08-11 14:50	5650	1h35min	58.0	134.4
13	2007-08-11 13:50	5705	2h25min	42.4	134.4
14	2011-07-02 17:10	5790	2h55min	90.8	132.0
15	2011-07-02 15:45	5745	3h30min	76.6	129.6
16	2005-08-07 09:15	5755	8h35min	53.8	129.6
17	2011-07-02 18:15	5665	2h5min	44.0	127.2
18	2016-06-23 18:45	5675	9h25min	47.0	127.2
19	2007-08-11 13:45	5771	2h5min	37.6	127.2
20	2011-07-02 17:05	5810	2h60min	55.4	127.2
21	2007-06-23 09:15	5655	6h5min	38.8	122.4
22	2007-06-23 09:30	5670	5h60min	30.2	122.4
23	2011-07-02 17:20	5715	2h20min	70.8	120.0
24	2011-07-02 17:25	5710	2h20min	64.0	120.0
25	2011-07-02 17:20	5795	2h20min	61.6	120.0
26	2011-08-08 13:05	5585	3h10min	18.0	117.6
27	2011-07-02 17:20	5804	2h35min	85.8	117.6
28	2013-08-10 10:20	5670	7h30min	16.8	117.6
29	2016-06-23 18:30	5915	9h30min	45.6	115.2
30	2008-06-27 09:25	5620	9h10min	21.0	112.8
31	2011-07-02 17:25	5655	2h10min	43.4	112.8
32	2007-08-11 13:50	5710	1h10min	34.6	112.8
33	2005-07-30 08:10	5570	5h10min	28.4	110.4
34	2013-08-10 17:20	5690	10min	11.2	108.0
35	2009-07-20 09:20	5570	8h30min	15.4	108.0
36	2015-09-04 06:40	5685	1h25min	36.4	108.0
37	2011-07-02 17:20	5694	2h15min	62.0	108.0
38	2016-06-23 18:30	5905	7h20min	44.8	108.0
39	2011-08-09 19:00	5675	20min	11.4	105.6
40	2015-09-04 06:05	5690	1h60min	44.2	105.6
41	2011-07-02 17:20	5660	2h15min	50.2	105.6
42	2016-06-23 18:20	5925	9h40min	50.6	103.6
43	2011-05-22 14:50	5740	2h50min	19.8	103.2
44	2007-08-10 18:20	5855	10min	14.8	103.2
45	2016-06-23 18:30	5930	9h40min	43.0	103.2
46	2008-06-27 09:20	5633	1h10min	11.2	100.8
47	2016-06-23 18:30	5901	7h20min	41.4	100.8
48	2011-07-02 18:20	5650	1h15min	45.2	98.4
49	2011-07-02 18:55	5825	1h5min	33.2	98.4
50	2014-06-20 03:50	5580	5h10min	15.6	96.8

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

Table A2. Top 50 events for Finland

Event	Starting Time [UTC]	Gauge	Duration	Amount [mm]	Peak [mmh^{-1}]
1	2014-07-19 13:50	101787	2h30min	34.7	89.1
2	2014-07-31 09:00	101103	1h20min	18.1	87.5
3	2014-07-30 15:50	101289	19h20min	34.8	86.6
4	2014-05-25 16:40	101555	29h50min	31.6	84.2
5	2014-07-31 11:10	101690	2h60min	51.0	83.9
6	2014-07-18 08:40	101799	1h60min	25.7	83.2
7	2013-08-07 10:10	100951	14h60min	25.9	82.4
8	2014-07-19 09:50	101194	50min	14.6	79.1
9	2014-05-25 09:50	101339	25h60min	48.4	78.6
10	2014-07-31 11:00	101787	3h60min	28.4	78.1
11	2015-07-22 09:00	101603	2h30min	29.4	77.9
12	2014-07-09 14:40	101800	20min	22.1	76.6
13	2014-08-13 21:40	100908	6h50min	28.9	74.2
14	2014-08-09 14:40	101826	30min	16.3	72.8
15	2014-08-11 22:50	100953	3h20min	37.3	71.6
16	2013-08-10 13:50	100917	40min	14.1	69.2
17	2016-07-31 17:20	101572	2h10min	21.2	68.3
18	2016-08-06 16:40	101338	60min	35.2	68.2
19	2016-07-31 09:40	101555	11h20min	27.9	67.5
20	2016-07-03 12:30	101603	7h30min	67.1	66.9
21	2016-06-30 10:10	126736	25h50min	63.9	66.2
22	2014-08-12 23:10	100955	7h60min	20.1	65.6
23	2014-08-11 07:00	101726	4h30min	13.5	65.6
24	2016-07-25 09:00	101743	6h20min	25.9	65.6
25	2014-07-14 11:50	101339	1h30min	23.2	65.0
26	2015-08-30 17:10	100953	20min	15.8	65.0
27	2016-07-12 05:10	101537	3h10min	21.4	64.7
28	2014-08-22 12:20	101805	1h60min	16.3	63.6
29	2015-07-08 14:00	101537	25h10min	46.3	62.9
30	2013-06-27 10:20	101338	8h30min	33.2	62.1
31	2014-06-06 13:00	101690	6h30min	16.7	61.4
32	2013-09-01 06:10	101272	9h30min	33.0	61.2
33	2016-07-31 06:40	100974	3h40min	21.6	61.0
34	2013-08-15 14:00	101124	50min	14.0	60.5
35	2014-05-19 18:40	101537	4h10min	21.4	59.6
36	2015-08-08 16:50	101632	2h30min	11.3	58.9
37	2013-08-31 11:30	100955	3h20min	30.0	58.7
38	2016-07-11 14:30	103794	11h30min	14.1	58.4
39	2014-07-14 13:00	101555	2h10min	20.2	58.1
40	2016-07-31 06:20	101632	6h30min	16.5	58.1
41	2016-08-04 11:10	101194	6h60min	18.1	58.0
42	2016-07-27 14:50	101950	20min	13.2	57.3
43	2014-08-13 16:50	100967	3h40min	12.1	56.8
44	2014-08-11 08:30	126736	3h20min	13.4	56.7
45	2015-07-16 12:20	101103	24h30min	69.5	56.6
46	2016-07-27 04:00	101805	5h20min	16.6	55.5
47	2016-07-14 10:10	101933	60min	20.4	55.2
48	2014-05-19 13:40	100967	20min	13.3	55.1
49	2014-08-11 23:40	101603	12h10min	42.4	53.9
50	2013-06-27 11:00	101150	5h10min	19.2	53.2

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

Table A3. Top 50 events for the Netherlands

Event	Starting Time [UTC]	Gauge	Duration	Amount [mm]	Peak [mmh^{-1}]
1	2014-08-03 17:10	380	6h30min	56.9	180.0
2	2014-07-28 11:30	275	2h60min	61.8	139.8
3	2011-06-28 18:20	356	5h60min	90.2	136.2
4	2016-06-23 01:10	260	60min	36.2	121.2
5	2015-08-30 22:20	283	3h50min	30.2	120.0
6	2013-08-19 11:20	286	2h10min	29.8	114.0
7	2015-08-30 19:40	356	6h20min	55.6	112.8
8	2012-05-20 14:20	375	4h30min	21.8	109.8
9	2013-07-26 12:50	286	30min	22.0	106.2
10	2016-09-15 21:20	375	1h30min	18.9	94.2
11	2011-06-28 19:50	273	11h40min	25.1	93.6
12	2012-08-15 19:40	370	60min	15.4	92.4
13	2011-08-22 23:40	375	11h60min	33.4	92.4
14	2011-08-18 16:30	391	4h10min	29.4	92.4
15	2016-06-23 20:20	380	3h30min	27.5	90.6
16	2015-08-31 14:30	270	2h20min	32.2	88.2
17	2009-07-03 14:10	391	2h10min	38.0	88.2
18	2013-08-05 23:00	280	30min	14.2	84.0
19	2012-06-21 20:00	290	3h10min	17.2	82.2
20	2009-07-21 16:50	269	2h60min	17.2	80.4
21	2016-06-15 10:50	277	7h30min	34.5	80.4
22	2008-08-07 07:10	240	7h10min	32.9	79.2
23	2008-07-26 18:10	270	8h10min	26.8	78.6
24	2015-07-05 09:50	270	6h30min	15.4	78.6
25	2016-06-23	344	10h10min	32.8	78.6
26	2014-07-28 02:20	257	10h20min	71.3	77.4
27	2009-07-14 12:20	286	3h20min	17.5	77.4
28	2012-08-05 13:10	323	6h40min	18.5	77.4
29	2009-05-25 20:50	260	6h30min	23.8	76.8
30	2012-05-10 14:40	375	3h50min	15.3	76.2
31	2014-07-10 23:20	269	50min	20.7	75.6
32	2008-07-06 08:00	277	30min	20.1	75.6
33	2009-06-09 10:50	319	8h20min	24.8	75.6
34	2014-07-10 21:10	391	20min	20.4	75.6
35	2008-09-11 23:50	265	16h40min	41.8	74.4
36	2011-06-05 16:10	286	1h30min	19.1	73.8
37	2015-08-24 15:00	269	3h40min	13.3	70.8
38	2012-05-20 21:30	278	30min	15.8	70.2
39	2013-07-27 21:40	350	2h10min	33.6	70.2
40	2011-08-03 14:00	278	7h50min	40.8	69.0
41	2011-08-23 10:40	283	1h30min	16.5	69.0
42	2008-08-12 23:40	257	12h20min	23.1	68.4
43	2010-07-14 15:50	377	1h30min	16.7	68.4
44	2014-07-27 22:00	240	14h20min	53.7	67.8
45	2009-05-15 05:00	273	16h20min	28.8	67.8
46	2012-08-04 14:40	273	4h10min	17.5	67.8
47	2013-07-27 23:50	278	50min	20.5	67.8
48	2009-07-03 14:30	290	4h10min	32.1	66.0
49	2015-08-14 18:10	310	3h60min	21.7	66.0
50	2011-09-06 10:20	257	11h20min	33.1	64.8

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

Hydrology and Earth System Sciences Discussions
Open Access
EGU

Table A4. Top 50 events for Sweden

Event	Starting Time [UTC]	Gauge	Duration	Amount [mm]	Peak [mmh^{-1}]
1	2006-07-29 18:30	92410	1h30min	44.0	91.2
2	2013-07-26 07:30	87140	3h45min	48.2	81.2
3	2008-07-21 03:15	98490	7h45min	51.5	71.2
4	2010-08-17 04:15	76420	8h15min	26.3	67.2
5	2001-08-26 18:00	97280	19h15min	54.0	62.4
6	2008-07-05 14:15	92410	60min	16.8	60.4
7	2014-08-03 01:00	87140	1h30min	28.6	54.8
8	2008-07-05 20:30	75520	37h45min	53.1	53.6
9	2001-08-26 15:15	86420	19h30min	38.8	52.0
10	2007-09-10 15:30	89230	17h15min	51.1	51.6
11	2015-07-14 18:45	75520	2h60min	25.9	49.6
12	2014-08-11 07:15	89230	2h30min	26.4	49.6
13	2012-08-07 16:45	97280	5h45min	16.5	48.8
14	2011-08-10 11:00	97280	2h45min	33.4	48.0
15	2012-08-08 20:00	89230	9h45min	39.9	47.2
16	2011-07-23 02:30	92410	60min	18.8	45.2
17	2012-07-20 18:15	98490	11h45min	24.7	45.2
18	2018-08-05 13:15	98490	3h45min	15.1	44.8
19	2006-08-22 15:45	62040	20h60min	50.4	41.6
20	2006-08-20 05:30	62040	14h15min	27.4	41.2
21	2013-08-13 07:45	62040	35h15min	81.2	41.2
22	2009-05-20 12:00	76420	7h30min	17.6	41.2
23	2010-07-29 09:45	97280	8h15min	36.4	40.8
24	2001-08-06 12:45	98490	2h60min	17.3	40.4
25	2011-07-22 20:15	86420	8h45min	13.7	40.0
26	2006-09-03 04:15	97280	4h45min	19.5	40.0
27	2010-08-17 14:15	86420	2h45min	20.4	39.6
28	2011-08-18 11:00	98490	4h45min	10.5	39.6
29	2016-07-26 13:15	87140	45min	17.6	38.8
30	2012-05-31 08:30	97280	10h45min	20.8	38.8
31	2008-08-07 17:45	97280	16h15min	34.5	38.4
32	2018-08-24 12:15	77210	3h15min	18.4	37.6
33	2011-06-23 00:45	86420	7h60min	39.4	37.6
34	2009-07-30 14:00	92410	2h30min	24.3	37.6
35	2007-08-10 06:45	98490	5h45min	20.2	37.6
36	2018-08-14 01:45	75520	18h30min	55.5	37.2
37	2008-07-12 09:15	92410	3h30min	19.3	37.2
38	2014-07-28 12:15	76420	2h15min	15.0	36.8
39	2010-07-17 15:45	89230	4h60min	13.9	36.8
40	2008-06-30 06:45	98490	5h45min	14.8	36.8
41	2008-08-02 09:15	97280	13h30min	33.7	36.4
42	2010-08-23 21:15	87140	3h60min	24.0	35.6
43	2006-08-03 00:15	89230	4h60min	41.9	35.6
44	2001-08-10 02:15	92410	26h45min	27.1	35.6
45	2010-08-19 11:45	77210	5h45min	25.2	35.2
46	2015-07-13 08:00	75520	22h15min	30.1	34.8
47	2005-05-04 16:00	86420	60min	14.0	34.8
48	2014-07-28 06:45	89230	1h30min	15.8	34.8
49	2012-06-11 10:15	97280	1h60min	16.4	34.8
50	2010-08-09 06:45	76420	7h60min	15.0	34.0

https://doi.org/10.5194/hess-2019-427
Preprint. Discussion started: 15 August 2019

Table A5. Top 10 events for Danish X-band product

Event	Starting Time [UTC]	Gauge	Duration	Amount [mm]	Peak [mmh^{-1}]
1	2017-08-01 18:15	5058	7h10min	15.6	115.2
2	2016-07-25 13:35	5049	5h10min	25.0	93.6
3	2016-07-25 13:55	5045	4h20min	26.4	84.0
4	2017-08-01 18:20	5057	4h10min	15.6	81.6
5	2017-08-15 18:15	5057	2h5min	31.8	81.6
6	2017-08-15 18:15	5058	1h60min	27.6	74.4
7	2017-06-16 01:15	5052	5min	8.8	69.6
8	2017-08-18 12:50	5054	9h15min	15.8	69.6
9	2017-06-15 21:45	5057	3h40min	13.2	69.6
10	2016-06-16 15:50	5052	2h10min	16.2	67.2