List of Major Changes:

Attached, please find the revised study about the accuracy of radar in time of heavy rain. As you can see, major changes were made to each of the sections in order to accommodate the referee's comments. The list of major changes include:

1. A new, longer introduction that emphasizes the importance of accurate rainfall estimates for hydrological applications.
2. More details about the individual gauge and radar products in Section 2.2
3. A new methodology for estimating the bias and quantifying the part of the bias due to differences in sampling volumes between radar and gauges (Section 2.3)
4. New results (Section 3) about the conditional bias with intensity and range.
5. Stronger, more precise conclusions with clear recommendations for future research (Section 4).
6. Two new figures (Fig 7-8) for investigating the conditional bias with intensity and range.
7. One figure (previously Fig 8, RMSE vs time scale) was removed because it was redundant. Also, we feared that it could be confusing to interpret due to the fact that RMSE combines both information about scatter and bias and can therefore be hard to interpret.

The authors would like to thank all the reviewers again for their constructive feedback and for their numerous suggestions that helped improve the paper. Below, please find a detailed answer to each referee's comments.

Referee 1:

Major comments:

1. (Abstract) L15-19: Throughout the manuscript, supporting materials for urban hydrology and mitigations of attenuation are not presented. Revise this part and reflect what has been presented.

<u>Response:</u> The whole paper has been revised and several new references to urban hydrology have been included. The revised manuscript also contains a much more precise and in-depth discussion about how to mitigate peak intensity biases using rain gauges and polarimetry.

2a. The link with hydrology or urban flooding/forecast:
One of the objectives of this study is to better understand the link between rainfall and urban flooding (L7-9) or/and the use of radar in hydrology and flood forecasting (L84-85). However, very few discussions were presented in this aspect. Add either more supporting materials for flooding parts (link with the presented work) or clarify better the objective of the presented work.

<u>Response:</u> More details about the importance of accurate rainfall measurements for hydrology and flood forecasting were added in the Introduction.

2b. Hydrological model (L171, L205, L397, L472, L490) has been mentioned in several sections without reference cited and the statements are rather generally made, which requires improvement in either writing or strengthening the explanation with more supporting materials (particularly for the statement made in the conclusion).

Response: More references to hydrological modeling have been added and specific numbers are now given in the Introduction.

3. Better clarification and more supporting materials are required in results and conclusions (see the minor comments 16-37).

Response: The results and conclusion sections were completely rewritten during revision. Thanks to the new model in Section 2.3.1 and the new analyses of conditional bias and range-dependent bias, a much more precise quantification of the representativeness errors (areal vs point) and overall accuracy of the radar products is now possible.

Minor Comments:

1. L10-L11: Clarify better “the top 50 events”, “overall agreement”, “the peaks” of what.

Response: Done

2. L44: need clarification of “accuracy” (of what).

Response: Done

3. L46-47: This term “higher-level” composite is less objective and vague. Rephrase it.

Response: Done

4. L59-60: “, the longest...15-20 years at best.” Is it the case for world-wide or those countries presented in the manuscript?

Response: To the best of our knowledge, we believe that this is the case worldwide.

5. L76-78: “Often...the results” This is not clearly written in the context. Specify better. Also, adding more backgrounds/references to support strong needs in multinational assessment and comparisons will be necessary. At least, in Europe, there has been an effort made with BALTRAD products (Michelson et al. 2018, referenced already in the manuscript but in later chapter) and with the OPERA products (e.g., Saltikoff et al. 2019, Park et al 2019), which can be referred in the introduction.

Response: We added the reference to the OPERA product and BALTRAD to the text and included some more details in the introduction to support the need for an international assessment and comparison.

6. Table 2: Clarify the data resolution original vs. used for the comparison, e.g., in the text Line 128, Danish data has been interpolated to 1 min. In Table 3, is the comparison done at 5 min and at 1 min?

Response: The comparisons for Denmark were done at the 5 min resolution to match the resolution of the radar and gauge data. Although 1-min gauge data could be used in theory (using advection interpolation), this is not recommended here as this would add additional uncertainty due to interpolation. Also, the gauges in Denmark are 0.1 mm RIMCO tipping buckets which means that the sampling uncertainty at 1 min would be very large.

7. L153-154: reference missing for the operational product.

Response: The reference to Koistinen et al. 2014 as been added to the text.

8. L164: “Polar radar measurements”. Describe better, it seems a jargon, meaning radar measurement done at polar grid.

Response: Yes, the measurements are made over a polar grid and projected afterwards. The sentence has been reformulated to convex the right meaning.

9. L170: After applying HIPRAD, the temporal/spatial resolution of the data remains the same as shown in Table 2?

Response: Yes, the output has the same spatial and temporal resolution.

10. L178, “Aalborg” add country name and indicate the coverage of this radar in Fig1.

Response: Done

11. L188: what is “tas BALTRAD”?

Response: This product is now simply referred to as “BALTRAD”.

12. L206-208: Add reference

Response: A reference to Rossa et al. (2011) has been added.

13. L290: “the HIPRAD” here, isn’t it BALTRAD?

Response: No, these are two different products.

14. L249: “the highest available temporal” This term is used several times later, but isn’t it the same as gauge sampling resolution (shown in table 1)? Is there any reason for such term? If so, explain better.

Response: The highest available temporal resolution refers to the highest common time resolution at which both radar and gauge data are available. This has been clarified during revision.

15. L 249: “Top event” → Event 1 (fig. 2), where are these gauges located in Fig 1?

Response: We are not allowed to disclose the exact location of the gauges but this is not really important here anyway. Indeed, as shown by Figure 8, there is no clear trend/bias with respect to the distance to the radar.

16. L253-254: Some results presented were already gauge adjusted and one (Finland) not. It is not clear to compare these numbers from literature examples (which is not clearly mentioned either if they were also derived before the adjustment or after?). Is it necessary?

Response: Yes, we believe that comparing them is useful. At the same time, we agree with the referee that this can be rather tricky and misleading if done improperly. We tried our best to clarify this during revision. Most literature values that we could find were for gauge-adjusted products. But a few studies have also looked at biases in non-adjusted products.

17. L258: “The third rainfall peak” indicate here figure 4 (perhaps better with 4a indicating Denmark).

Response: Done

18. L264-265: “the relatively large peak intensity biases of 2.17, 2.09, 1.98 and 1.73 for Denmark, Finland, the Netherlands and Sweden...confirms this hypothesis” if the hypothesis refers the previous sentence, the bias for Netherlands should be larger than that of Finland because the peak intensity is higher for NL than for Finland (L256), isn’t it?

Response: This sentence does no longer exist in the revised version.

19. L272 “at these scales” and L275 “such small scales”. What does it mean? Is it related to storm scale? Or do you mean that the comparison was done with the instantaneous and point estimates (that affects representativeness error)?

Response: It means that the measurements are compared at high temporal resolutions (i.e., 5, 10 or 15 minutes depending on the radar product). At these time scales, sampling effects can have a rather large impact on traditional error metrics such as bias and rmse. The sentence has been reformulated during revision and now reads as follows:

“This is characteristic for sub-hourly aggregation time scales and can be explained by the large spatial and temporal variability of rainfall and the fact that radar and gauges do not measure precipitation at the same height and over the same volumes.”

20. L283: This is redundantly written (merge with L280-282)

Response: This sentence does not exists anymore in the revised version.

21. L300-301: Are these numbers MB after the ARFs reduction applied? is it also shown in Table 3?

Response: This sentence does not exist anymore in the revised version. A new Table (Table 4) now provides a better overview of ARFs and biases before/after correction for ARFs.

22. L302-302: Is the statement made before applying the ARFs? Clarify better. After ARFs, Swedish result shows the best, doesn’t it?

Response: This part of the paper has been completely reformulated during revision and should now be easier to follow.

23. L306-307: This does not support any argument and redundantly written in L300. Rephrase or remove it.

Response: This part has been rewritten. Please see Section 3.4 (other sources of bias) for more details.

24. L324, L405: “deeper analysis” Avoid “deeper” (somewhat subjective word) and revise the sentence.

Response: Done

25. L325: “temporal aggregation time scale” -> aggregation time scale (isn’t it the same as shown in Figures 8-10?)

Response:

26. L338-339: “Furthermore, the quality....an important role”. Add supporting explanation.

Response: Done

27. L359: It is not clear in Table 3 that the Danish products are the best in terms of RRMSE and CC. Revise this part.

Response: Done

28. L363-364: “However, a closer analysis....only 0.2”, what does it mean?

Response: The problematic sentence does not exist anymore.

29. L375-376: Clarify what is “viewpoints”. Apart from the statement, how the attenuation and VPR correction applied to the group 2 data (Yes for Danish C band data, not explicitly indicated for the Swedish) were performed?

Response: More details about this have been added to the Data section.

30. L379: “a coarser scale” in time or/and space?

Response: In time (has been changed during revision).

31. L397-399: add reference. Is there any example run for the presented event?

Response: Some references have been added in the Introduction to explain this.

32. L418: “the same order...than for...” -> the same order...as for

Response: Done

33. L421-L422: This statement needs better supporting explanation, e.g., what dual-polarization capabilities was used in the processing of the data?

Response: This should now be clear thanks to the new information about the individual radar products. See response to major comment 3 of referee 2 for more details.

34. L469-470: “Bias correction...on peak intensity bias”. Is this conclusion derived from all the presented cases for four countries? There are some explanations for the Dutch product (L348-349), but

not easy to find for the others. For Finland, the presented examples are not even bias corrected, so it is not clear what the authors mean.

Response: This sentence does not exist anymore in the revised version.

35. L471-472: Throughout the manuscript, "the importance of high-resolution radar observations in hydrological study" is hardly demonstrated/literature-reviewed with respect to the high-resolution radar products, which makes such conclusive statements weak. Add more solid outputs or references.

Response: More references have been added in the introduction, together with some explanations for why higher resolution is necessary and how it affects the timing and magnitude of predicted peak flows.

36. L488-489: Add references or strengthen supporting material for the referred rainfall uncertainties in hydrological models (e.g., some examples among any of the events 50 events*4 countries as a part of discussion or more explanation in L397-399).

Response: A reference to Bruni et al. (2015) has been added.

Referee 2

This paper compares the accuracy of weather radar rainfall using data from different countries (Denmark, Netherlands, Finland and Sweden). The study focuses on the top 50 heavy rainfall events which are more relevant for urban hydrology. The results showed that 1) radar underestimates rainfall rates; 2) radar products with higher spatial/temporal resolutions agree better with observations; 3) the combination of radar measurements from overlapping radars can improve rainfall rates. Although the results are interesting for the scientific community, there are a number of issues that the authors need to address before the paper is accepted for publication:

Major comments:

1) Rain gauge data quality. The rain gauge measurements used to validate the radar observations come from different operational agencies. It is obvious that the quality of the gauge measurements is not going to be the same among the different agencies and therefore this could impact your results. There is no discussion about this in the paper.

Response: Indeed, data quality plays a big role. More details about the type of rain gauges used in each country have been added to the data section. Systematic biases due to wind and calibration issues are important but unfortunately, we do not have enough information to reliably estimate them on an event-by-event basis. However, some typical values are now provided in the text based on literature. Also, it is important to remind the reviewer that basic visual quality control has been performed on the gauge and radar data for each of the 50 events. Suspicious or obviously wrong measurements were discarded during this step. Finally, note that gauges are not considered as ground truth in this study. Rather, the goal is to describe the overall discrepancies between radar and gauge measurements, combining all sources of errors (i.e., gauges, radars, algorithms, humans) as well as differences in measurement scales.

2) Rain gauge network density (Fig 1). It seems that the gauge network density is playing an important role in your results and there is little discussion on this. For Denmark the gauges are mainly clustered in a particular area (around 40-60km from radar site), for Finland the gauges are further away (beyond 50km) and cover different radars, for Sweden I can only see 4-5 gauges, whereas for the Netherlands all the gauges are more or less evenly distributed between 0-100km in range from the radar sites. This again will have important consequences in your results. For instance, VPR corrections will be important at far ranges. Attenuation due to heavy rain will also play a role. I will expect the radar rainfall error to increase with range and so the results will be better (or worse) depending on the location of the rain gauge network.

Response: Indeed, this was somewhat neglected during the analyses and discussion. Additional analyses were performed to study the range-dependent bias. Also, a new figure (Fig.8) has been added to show the distribution of rain gauges as a function of their distance to the radar(s). The analyses show that range-dependent biases are negligible compared to other factors (such as intensity-dependent bias).

3) Radar data quality. Every operational agency applies different corrections to the radar data. These corrections are extremely important and can help to explain some of the results. However, there is very little detail in the paper on the actual processing steps performed by each operational agency. Some corrections are discussed, but what about corrections for attenuation, VPR, partial beam blockage, etc for some of the countries. How do you ensure that the radar data have good data quality in both rain/no rain conditions? How does the operational agencies monitor the calibration of their radars (I do not mean comparisons with rain gauge observations)? Do the bias corrections include the same (or some) of the gauges that you used for your validation? If so, what are the implications? I think this section deserves a more detailed summary.

Response: We agree this is a very important issue. There are many factors at play here and unfortunately, it is impossible to address all of them in this paper. During revision, we added as much information as possible about each radar product, including how it was derived and what type of post-processing (e.g., bias adjustments) were applied. See Section 2.2 in the revised manuscript for more details. This makes it easier to interpret the different performances between countries.

4) The radars have different spatial/temporal resolutions. This is obviously a challenge when comparing the accuracy across different operational agencies. Would not be better to accumulate to the same spatial/temporal resolution (e.g. 2x2km, 15min) in order to have a fair assessment of the results? It seems to me that the different spatial resolutions have important implications in your comparisons.

Response: Actually, because the Swedish product is at 15 min resolution and the Dutch product is at 10 min resolution, the smallest common resolution (assuming we don't want to interpolate) is 2x2 km and 30 minutes. This is rather coarse compared with the lifetime of convective cells and probably of lesser interest for most readers. Also, aggregation to coarser scales is not recommended as simple arithmetic averaging of processed radar fields does not mimic what a lower resolution radar would see (e.g., due to the non-linear relation between rain rate and reflectivity and the multiple post-processing steps applied to the rainfall estimates). This is now clearly mentioned in the text (see Section 3.4). For all these reasons, we think it is best to work at the highest possible resolution for the main parts of the analyses.

5) The use of ARF can help to explain the discrepancies, but I suggest to compare with the method proposed by Ciach and Krajewski (1999) which actually uses the spatial correlation of the rainfall field within the radar grid resolution to separate (or explain) the variance due to the fact that gauges represent a point whereas radar rainfall is an areal measurement from the total variance (see also Bringi et al, 2011).

Response: Thank you for the suggestion. We carefully looked into the method of Ciach and Krajewski (1999) during revision and included their reference in the text. However, their approach was developed for additive error models and therefore not directly applicable to multiplicative biases. Moreover, our rain gauge networks were not dense enough to properly estimate the extension variance and estimating the nugget from the radar would not have made a lot of sense either since the radar data contain errors and biases. So instead, we opted for a comparatively simpler approach and proposed our own model (see Section 2.3.1) in which the differences in sampling volumes are an integral part of the random error terms (together with all other measurement errors). The two methods give different results, showing how difficult it is to separate the actual bias from bias due to the differences in measurement volumes.

6) Although the focus of heavy rainfall is important, what about the accuracy of radar rainfall for more conventional events (implications of the different corrections for radar errors) or in no rain conditions (e.g. implications of using robust clutter schemes, etc)? Are the results still consistent with those observed during heavy rainfall?

Response: Unfortunately, a systematic assessment of this issue was not feasible as only a small subset of the radar data archives has been processed so far (i.e. the top 100 events for each country, of which the 50 most intense after quality control were kept). However, it is worth pointing out that the 50 top events already contain a lot of "regular" time periods with low to moderate rainfall intensities. Therefore, a lot can be said already about the differences between average performance and performance during the peaks. In the revised paper, this is done by comparing the G/R ratios to the peak intensity bias (PIB) and quantifying the conditional bias with intensity (as done in Figure 7).

Other comments:

Fig 1. x/y labels? is that lat/lon?

Response: This figure has been replaced during revision.

Line 80. There is a reference that it is worth to look at related to the impact of spatial/temporal resolution in hydrodynamic modeling (Ochoa-Rodriguez et al, 2015).

Response: Thanks! The reference has been added to the Introduction.

Line 155. A lot of statements not justified: "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."

Response: Additional information about the radar products has been provided.

Line 160. "BRDC"?

Response: That's the name of the product. BRDC stands for BALTEX Radar Data Center, and BALTEX stands for Baltic Sea Experiment. A little note has been added in the text to explain this.

Fig 5. did you accumulate to 1h? or it is 5min,15min ...and so on?

Response: Fig 5 shows the results at 5, 10 and 15 min (as indicated)

Table 2. can you include more radar specs? e.g. beamwidth, scanning rate, radome type, pulse width, etc that can affect the measurements.

Response: Some additional details were added. However, it should be pointed out that these characteristics were not always constant over time due to hardware and software changes. This is now clearly stated in the text.

Line 420. "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." For x-band radars, sometimes the radar signal might be lost due to attenuation in heavy rain and without signal there is no way to apply any correction. Is this the reason for the lower rainfall amounts? I think signal lost due to rain attenuation at X-band has to be carefully taken in to account.

Response: No, the signal was never lost during these events. We can't guarantee that the attenuation correction worked well but the results seem to suggest that there were no major issues. The lower rainfall amounts are simply due to the fact that there are only 2 years of data and that the events for the X-band radar observations were relatively short and localized compared with the others.

Referee 3

It is a very nice simple-minded but important paper. We need results such as those reported in the paper to monitor our progress in variety of hydrologic problems. Radar-rainfall estimation is one of many of such problems in hydrology. I have very few comments to suggest to improve the paper:

1. The authors say little about the type of rain gauges used in the studies. "Automated" does not define the type and the type has implications for the expected errors (sampling). I suggest including the reference by Ciach (2003) if some of the gauges are tipping buckets.

Response: Thank you for the suggestion. The reference to Ciach (2013) was added during revision and the new paper now contains more details about the gauge type (see Section 2.1) and their distance to the radar(s). Also, a new figure (Fig 8) has been added to the manuscript to show how the bias depends on the distance to the radar.

2. In the Conclusions, the authors say: "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." There are problems with this statement. First, it has been shown by several studies in the past that rain gauges have representativeness errors. The larger the area, the larger the error. Ciach and Krajewski (1999a,b) have established a framework on this that was followed my many subsequent studies. Therefore, it is expected that radar products with coarser resolution will show poorer agreement with rain gauges data. This says nothing regarding importance of high

resolution radar observations in hydrologic studies. In fact, for many applications the resolution is not the most important aspect of the radar-rainfall product.

Response: Thanks for the comment. We completely revised the paper to better account for this issue. The most important change is the new statistical model in Section 2.3 for separating the measurement support bias from the actual bias. Also, we paid more attention to conditional bias with intensity and range. Using this new model, we rewrote the entire results sections and conclusions.

3. The quality of the figures should be improved.

Response: We followed the reviewer's suggestions and did our best to improve the quality of the figures (when judged necessary).

4. I recommend removing the whole story of the X-band radar. Including it seems forced. That's not what the paper is all about. Write another study about the X-band radar performance.

Response: Yes, the story for the X-band radar is a bit different (shorter time period and different frequency). But we don't think it looks forced. The X-band data is not in the focus of the paper but it provides additional interesting results at higher resolutions that strengthen the conclusions of the paper.

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]Hydrology Research Unit, 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 measuring at higher resolutions and making use of dual-polarization radar, these mismatches can be reduced. Each country has developed its own strategy for addressing this issue. But However, 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 in a 10-year database of radar data were used to quantify the overall agreement between radar and gauges and the errors as well as the bias affecting the peaks. Results show that the overall agreement between radar and gauges in heavy rain is fair (correlation coefficient 0.7-0.9), 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 countries1.2-1.8 (17-44% underestimation). However, after taking into account the different sampling volumes of radar and gauges, actual biases could be as low as 10%. Despite being adjusted for bias by gauges, 5 out of 6 radar products still exhibited a clear conditional bias with intensity of about 1-2% per mmh^{-1}. Peak rainfall intensities were therefore severely underestimated (factor 1.8-3.0 or 44-67%). The most likely reason for this is the use of a fixed Z-R relationship when estimating rainfall rates (R) from reflectivity (Z), which fails to account for natural variations in raindrop size distribution with intensity. Differences in sampling volumes between radar and gauges could also play an important role in explaining the bias but are hard to quantify precisely due to the many post-processing steps applied to radar. Based on our findings, 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 radarfor 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 easiest way to mitigate the bias in times of heavy rain is to perform frequent (e.g., hourly) bias adjustments with the help of rain gauges, as demonstrated by the Dutch C-band product. An even more promising strategy that does not require any gauge adjustments is to estimate rainfall rates using a combination of reflectivity (Z) and differential phase shift (Kdp), as done in the Finnish OSAPOL productalso 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. Both approaches lead to approximately similar performances, with an average bias (at 10 min resolution) of about 30% and a peak intensity bias of about 45%.

1 Introduction

Today, several high-resolution radar rainfall products for use in hydrology are readily available across the globe (Huuskonen et al., 2014; Th . Compared with gauges, radar provides superior The ability to measure short-duration, high-intensity rainfall rates is of paramount importance in predicting hydrological response. Indeed, several studies have shown that the resolution of the rainfall data directly impacts the shape, timing and peak flow of hydrographs (Aronica et al., 2005; Löwe et al., 2014; Ochoa-Rodriguez et al., 2015 . Previous research has shown that in order to obtain reliable results in small urban catchments, the rainfall data should have a resolution of at least 10 min and 1 km (Schilling, 1991; Ogden and Julien, 1994; Berne et al., 2004). If the resolution is insufficient compared with what is needed for the runoff simulations, the accuracy of flood predictions is likely to be compromised (Andréassian et al., 2001; Aronica et al., 2005; Bruni et al., 2015; Rafieeinasab et al., 2015).

Another important issue besides resolution is the accuracy of the rainfall data themselves. Currently, only weather radar offers the spatial coverage, leading to more insight into resolution and accuracy needed to study the complex link between the spatio-temporal characteristics of rain events and their link to hydrological response (Wood et al., 2000; Berne et al., 2004; Smith et al., 200 . Steady improvement hydrological response (Wood et al., 2000; Berne et al., 2004; Smith et al., 2007; He et al., 2013; Thorndahl et al., 20 . The most common application of radar in hydrology is the study and characterization of heavy rain events associated with flooding (Baeck and Smith, 1998; Delrieu et al., 2005; Collier, 2007; Ntelekos et al., 2007; Anagnostou et al., 2010; Villarini et al., 2010; . However, there have been many other successful applications of radar in urban hydrology, such as generating detailed runoff predictions or creating flood maps (Wright et al., 2014; Thorndahl et al., 2016; Yang et al., 2016). Steady progress 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 improving the accuracy of radar rainfall 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 This has promoted the development of smaller, cheaper and higher-resolution X-band polarimetric radars for use in urban flood forecasting (Wang and Chandrasekar, 2010; Ruzanski et al., 2011). The hope is that by moving to higher resolutions and taking advantage of dual-polarization, the accuracy of radar-based rainfall estimates and flood predictions will increase. However, this is a delicate process as higher resolution and more elaborate retrieval algorithms also increase sampling uncertainty. A higher resolution therefore does not automatically translate into more accurate rainfall estimates (Krajewski and Smith, 2002; Seo et al., 2015; Cunha et al., 2015). Also, the space/time correlation structure of radar errors and their dependence on precipitation type and distance to the radar means that there are practical limits to what can be achieved in terms of predictive skill in hydrological models (Rafieeinasab et al., 2015; Courty et al., 2018).

Despite decades of research, quantifying individual errors and biases in radar retrievals remains hard (Einfalt et al., 2004; Lee, 2006; Kraj . One aspect that is still poorly documented concerns the overall accuracy of radar in times of heavy rain. Because radar hardware, software and data processing techniques are subject to frequent replacements and updates, most homogeneous radar records currently available for analysis only span 10-15 years. This is likely to improve in the future thanks to open data policies and the automatic exchange of radar data between countries, such as OPERA (Huuskonen et al., 2014; Saltikoff et al., 2019) . However, until now, datasets are limited and studies have mostly looked at performances of individual radar systems and/or national networks. The few results that are available suggest 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 (Smith et al., 1996; Overeem et al., 2009a; Smith et al., 2012; Peleg et al., 2018). For example, based on a 12-year archive of 1×1 km and 5-min radar rainfall estimates for Belgium, Goudenhoofdt et al. (2017) found that hourly radar extremes around Brussels tend to be 30-70% lower than those observed in gauge data. The underestimation is partly attributed to differences in sampling volumes between radar and gauges. In some cases, the underestimation can also be related to But other factors such as calibration issues, range effects, signal attenuation or saturation of the receiver channel . Wind can also play a role. At very high resolutions (e.g., 5 min and 1 km), wind effects and vertical variability also play an important role, further complicating the matching of rainfall can also introduce substantial biases 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)gauge measurements (Dupasquier et al., 2000; Vasiloff et al., 2009; Dai and Han, 2014). Another series of studies in the Netherlands showed that in principle, it is possible to derive robust intensity-duration-frequency curves (Overeem et al., 2009b, a) and areal extremes (Overeem et al., 2010) from long radar data archives. However, the authors clearly mention that the radar data need to be carefully quality controlled and bias corrected first.

Since radar measurements are inherently ~~uncertain~~ prone to errors 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~~ gridded, quantitative 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 Delobb~~ ~~. If done properly, this can help mitigate attenuation and reduce~~ and other sources of information such as elevation, cloud cover and satellite imagery (Krajewski, 1987; Smith and Krajewski, 1991; Goudenhoofdt and Delobbe, 2009; Delrieu et al., 2014; Stevenson and . During post-processing, many 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~~ attenuation, calibration, vertical variability and range effects are mitigated (e.g., Collier and Knowles, 1986; Young et al., 2000; Gourley et al., 2006; Overeem et al., 2009b; Delrieu et al. . However, rain gauge data also contain errors and biases, the most ~~common of them being~~ important of which is an underestimation of the rainfall intensity due to local wind effects~~around the gauge. These effects have been estimated to be~~ . For regular events, errors usually remain 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 1×1 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~~ . However, during heavy rain evens, wind-induced biases can exceed 30% (Nystuen, 1999; Sieck et al., 2007; Pollock et al., 2018). As a result, post-processed radar products might still contain important residual errors (Krajewski et al., 2010). For example, Smith et al. (2012), Wright et al. (2014), Thorndahl et al. (2014b) and Cunha et al. (2015) highlighted several major quality issues affecting post-processed quantitative precipitation estimates from NEXRAD, including range-dependent and ~~2011. Their analysis covered more than 15 years but the radar data used to derive the statistics were not continuous in time.~~ intensity-dependent biases. Quantifying these residual errors and studying their propagation in hydrological models is crucial for improving the timing and accuracy of flood predictions (Cunha et al., 2012; Bruni et al., 2015; Courty et al., 2018; Niemi et al., 2017). For example, in their study, Stransky et al. (2007) estimated that the propagation of biased radar measurements in urban drainage models could result in up to 30-45% errors in terms of peak flow magnitude. To limit error propagation, Schilling (1991) recommended that the bias affecting areal-averaged rainfall intensities should not exceed 10%.

~~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~~ Over the years, each country has developed its own strategy for mitigating errors and biases in operational radar rainfall estimates. However, since there is no common benchmark and few international studies are available, the merits and weaknesses of each approach remain difficult to quantify objectively. This study sheds new light on current performances by conducting a multinational assessment of radar's ability to capture heavy rain ~~. This paper sheds new light on this issue by providing a detailed analysis of~~ events at scales of 5 min up to 2 hours. In total, 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~~ are considered. Special emphasis is put on analyzing the performance during the 50 most intense events ~~for each country~~ over the last ~~10 yearsto 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~~10-15 years. 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~~ identifying the main sources of errors and biases across scales, important ~~conclusions and recommendations~~ recommendations about how to improve the accuracy of quantitative precipitation estimates for flash flood prediction and urban pluvial flooding can be drawn~~as to the use of radar in hydrology and flood forecasting~~. The rest of this paper is organized as follows: Section 2.1 explains the methodology used to select events and extract the gauge and radar data. Section 2.2 gives a detailed description of the radar products used for the analysis. Section 2.3 introduces the statistical models used to quantify the bias between gauges and radar. Section 3 presents the results and Section 4 summarizes the main conclusions.

2 Data & Methods

2.1 Event selection and data extraction methods

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~~ small subset of all the existing 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~~ gauges used, their temporal resolutions and length of the observational records for each country. ~~Using the selected gauges , we determined~~ Note that Denmark has two separate rain gauge networks. The first is operated by the Danish Meteorological Institute DMI and consists of OTT Pluvio2 weighing gauges (Vejen, 2006; Thomsen, 2016). The second belongs to the Water Pollution Committee of the Society of Danish Engineers and consists of RIMCO tipping bucket gauges (Madsen et al., 1998, 2017). For this study, only the RIMCO tipping buckets were used. In the Netherlands, precipitation is measured using the displacement of a float in a reservoir (KNMI, 2000). The 10-min data from 2003-2017 used in this study have been validated internally by the Royal Netherlands Meteorological Institute KNMI using a combination of automatic and manual quality control tests. In Finland, weighting gauges of the type OTT Pluvio2 are used. Observations are made using a wind protector according to World Meteorological Organization regulations (WMO, 2008) . Automatic quality control tests are used to flag suspicious values which are then double checked manually by human experts. In Sweden, gauges are vibrating wire load sensors of the type GEONOR with an oil film to keep evaporation at very low amounts.

Based on the available gauge data, the top 50 rain events (in terms of peak intensity) were determined 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~~ For every gauge, a continuous 6~~hours. To increase reliability, all events were~~ hour dry period was used to separate events from each other. This was done separately for each gauge which means that some events were included multiple times into the dataset given that they were observed by different gauges at different locations. To ensure quality, each identified event was subjected to a visual quality control test by human experts, ~~checking both for plausibility and consistency~~making sure the rainfall rates recorded by the gauges and the radar (see Section 2.2) were plausible and consistent with each other in terms of their temporal structure. 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~~ identified events had to be removed and replaced by new ones during these quality control steps, most of them because of incomplete or erroneous radar data.

The ~~procedure used to extract the radar data was identical for all countries~~radar data for each country were extracted according to the following procedure. 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 considered 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~~ (Dai and Han, 2014). Note that this is a rather conservative and favorable way of comparing gauges with radar that leads to smaller overall discrepancies and more robust results 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~~ However, these only resulted in higher discrepancies ~~without changing~~ and did not change the main conclusions and were ~~subsequently abandoned~~ therefore abandoned in subsequent analyses.

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~~ can be seen 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).

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 Radar data for Denmark

The ~~Danish radar product is derived from the measurements of the "Stevns"~~ weather radar network of the Danish Meteorological Institute (DMI) operates four 5.625 GHz C-band ~~radar located~~ pulse radars with 1 degree beam width and 250 kW peak power located in Rømø, Sindal, Stevns, Virring and Bornholm (Gill et al., 2006; He et al., 2013). New dual-polarization radars have been installed at all sites between 2008 and 2017. However, for this study, only the single-polarization data from the Stevns radar were used. The latter is located near the coast, at 55.326°N 12.449°E and 53 m elevation, approximately 40 km south of Copenhagen in an area of relatively flat topography with altitudes ranging from ~~-7m to 125m~~ -7 m to 125 m above mean sea level. ~~The radar volume scans~~ It was purchased in 2002 from Electronic Enterprise Corporation (EEC) and is operated using a combination of EEC and DMI software. The scanning strategy involves collecting reflectivity measurements at 9 different elevation angles of 0.5, 0.7, 1.0, 1.5, 2.4, 4.5, 8.5, 13.0 and 15.0 degrees with a range resolution of 500 m and a maximum range of 240 km. The reflectivity measurements Z [dBZ] at these 9 elevations 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 500×500 m^2 grid spacing (Gill et al., 2006). The temporal resolution of the PCAPPI is then ~~enhanced to 1~~ statistically enhanced to 5 min using ~~advection interpolation~~ an advection interpolation scheme (Thorndahl et al., 2014a; Nielsen et al., 2014). Ground clutter in the PCAPPI is removed by filtering out echoes with Doppler velocity smaller than 1 ms^{-1}. ~~Rainfall rate R is estimated~~ Rainfall-induced attenuation K is estimated as $K = 6.9 \cdot 10^{-5} Z^{0.67}$ [dBZ km^{-1}] and attenuation-corrected reflectivity estimates are converted to rainfall rates R based on a fixed Marshall-Palmer 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~~ $Z = 200R^{1.6}$. To take into account calibration errors and variations in raindrop size distributions, a daily mean field bias ~~based on daily~~

~~data~~ correction is applied to the high-resolution radar rainfall estimates based on the measurements from a network of 66 RIMCO tipping bucket rain gauges in the region operated by the Water Pollution Committee of the Society of Danish Engineers ~~(Madsen et al., 1998)~~(Madsen et al., 1998, 2017). Note that the final 500 m, ~~1 min~~ 5 min bias-corrected product used in this study is not operational ~~,~~ but developed for research purposes ~~for~~ by Aalborg University.

2.2.2 Radar data for the 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 Netherlands recently upgraded their radars to dual-polarization. However, the dual-polarization rainfall estimates are not fully operational yet and all radar rainfall estimates used in this study were produced with the single-polarization algorithms. Also, 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~~ . For a detailed description of the processing chain, the reader is referred to Overeem et al. (2009b). The radars used in this study were two single-polarization Selex (Gematronik) METEOR 360 AC Pulse radars with a wavelength of 5.2 cm, peak power of 365 kW, pulse repetition frequency of 250 Hz and 3-dB beam width of 1 degree. The scanning strategy consists of four azimuthal scans of 360 degrees at 4 elevation angles of 0.3, 1.1, 2.0, and 3.0 degrees. The data from these scans are combined into 5-min PCAPPI at 800 m height according to the following procedure: for distances up to 60 km from the radar, only the highest elevation angle is used to reduce the risk of ground clutter and beam blockage. For distances of 15-80 km from the radar, the PCAPPI is constructed by bilinear interpolation of the reflectivity values (in dBZ) of the nearest elevations below and above the 800-m height level. For distances of 80-200 km from the radar, only the reflectivity values of the lowest elevation angle are used, whereas it should be pointed out that the 800 m level only stays within the 3-dB beam width of the lowest elevation up to a range of about 150 km. Values beyond 200 km from the radar are ignored. Once the PCAPPI have been constructed, ground clutter and anomalous-propagation are removed using the procedure of Wessels and Beekhuis (1995) also described in Holleman and Beekhuis (2005). Spurious echoes within a radius of 15 km from the radar are mitigated based on the procedure described in Holleman (2007). A fixed Marshall-Palmer Z-R ~~relationship given by Z=200R$^{1.6}$. The rainfall estimates~~ relation of $Z = 200R^{1.6}$ is used to convert the reflectivities in the PCAPPI to rainfall rates. During the conversion, reflectivity values are capped at 55 dBZ to suppress the influence of echoes induced by hail or strong residual clutter. Because of this, the maximum rainfall rate that can be estimated with this approach is 154 mm/h. Individual rainfall estimates from the two radars are then combined into one final composite using a weighting factor as a function of range from the radar, as described in Eq. 6 of Overeem et al. (2009b). During the compositing, accumulations close to the radar are assigned lower weights to limit the impact of bright bands and spurious echoes. The composited rainfall rates 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~~on an hourly basis using a network of 32 automatic rain gauges at 10 min resolution and 322 manual gauges at daily resolutions following the procedures of Holleman (2007) and Overeem et al. (2009b). 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 algorithmsadditional bias correction at daily timescale (downscaled to 10 min scales) is primarily used to improve the large-scale spatial consistency of the radar and gauge estimates and is therefore not extremely important in the context of this study.

2.2.3 Radar data for Finland

The Finnish radar product is an experimental product from the FMI Finnish Meteorological Institute (FMI) OSAPOL-project, which differs from the operational product used by the Finnish Meteorological Institute (FMI) FMI mainly by making a better utilization of dual-polarization and by better taking into account the measurement geometry of the 10 C-band use of dual-polarizationDoppler radars currently available in Finland. The product is based on the data from the years 2013-2016, during which the old single-polarization radars were replaced by newer being replaced by C-band dual-polarization radars. Since this upgrade took place progressively, the OSAPOL-product combines data from 4 up to 9 Doppler radars. The product is therefore based on data from 4-8 dual-polarization radars depending on the number of radars that how many were available each year. Erroneous echoes and The beam width is 1 degree, range resolution is 500 m and the scanning is done in Pulse Pair Processing (PPP) mode. Doppler filtering is done first in the signal processing stage, and reflectivity measurements are calibrated based on solar signals (Holleman et al., 2010). Next, non-meteorological targets are removed using four different techniques. The algorithm used for correcting the vertical profile of reflectivity (VPR) targets are removed using statistical clutter maps and fuzzy-logic-based HydroClass classification by Vaisala (Chandrasekar et al., 2013). The reflectivity Z is the same as in the operational product. Rainfall intensity is estimated based on radar reflectivity Z and specific differential propagation attenuation-corrected (Gu et al., 2011), and the differential phase shift Kdp is estimated using the method described in Wang and Chandrasekar (2009). For hydrometeors classified as liquid precipitation, two alternative rain rate conversions are used. For heavy rain, i.e., Kdp>0.3 and Z>30 dBZ, the R(Kdp) relation given by $R = 21$Kdpis used while for $^{0.72}$ is used (Leinonen et al., 2012). For low to moderate intensitiesa fixed Z-R , i.e., Kdp≤0.3 or Z≤30 dBZ and for radar bins where HydroClass indicates non-liquid precipitation, a fixed Z(R) relation given by Z = 223R$^{1.53}$ (Leinonen et al., 2012) is used . A $Z = 223R^{1.53}$ is used (Leinonen et al., 2012). Using the estimated rainfall rates at the 4 lowest elevation angles, a PCAPPI at 500m height with m height is produced using inverse distance-weighted interpolation with a Gaussian weight function. Finally, a composite VPR correction map (Koistinen and Pohjola, 2014) is applied to the PCAPPI to generate a 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 functionresolution product. The OSAPOL is the only product radar product in this study 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 Radar data for Sweden

The considered product is the so-called BRDC (BALTEX Radar Data Center) 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 between the years 2007 and 2016 (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 Note that Swedish radars are being used for real-time operational production, and therefore prone to frequent changes and re-tuning. For example, the beam width of the radars has changed over time due to hardware upgrades. Also, the scanning strategies, filters and processing chains have been updated several times. Describing all these changes is not feasible within the context of this study. Therefore, the differences between gauge and radar estimates in Sweden include both a technical component (related to the hardware and number of radars) and a component related to the operation strategies over the years (i.e., human and algorithm). The technical aspects of the quantitative precipitation estimation in the BRDC product are explained in Section 2.2 of Norin et al. (2015). Azimuthal scans of reflectivity measurements at up to 10 different elevation angles between 0.5 and 40 degrees to are projected into 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 . 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 described in Bech et al. (2003). Rainfall rates on the ground are estimated from the PCAPPI through a constant Marschall-Palmer 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 applying 30-day mean correction factors to correct for mean field 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 radar-estimated rainfall rates at each location are therefore obtained by only taking into account the data from a single radar (i.e., usually the nearest one) and do not no attempt is made to take advantage of possibly overlapping measurement areas . Such methods are being developed measuring areas (except for bias-correction using gauges). Better radar compositing methods are currently being developed at SMHI 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 These are not the main focus of the paper and are only used to provide additional insights and help with the interpretation of the results. The first additional radar dataset is from a FURUNO WR-2100 polarimetric dual-polarization X-band Doppler research radar system located in Aalborgwhich scans at a fixed elevation angle of 4° , Denmark. The radar performs fast azimuthal scans at 6 different elevation angles 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. However, for this study, only the data from a single elevation angle (i.e., 4°) were used. 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 (after attenuation correction). Similarly to the Danish C-band product, 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 RIMCO tipping bucket rain gauges. Only two years of X-band data is that it only covers a two-year period from radar measurements between 2016-2017 which strongly limits the number of heavy rain events available for the are available for analysis. Consequently, only the 10 most intense events were considered. Despite the low sample size, the hope is that by comparing the performance of the these limitations, the X-band product to the C-band product, data can be used to provide valuable insight into the benefits of advantages and challenges associated with using high-resolution polarimetric rainfall X-band radar measurements in times of heavy raincan be gained.

The second additional radar product used for comparisons in this study is an international composite at 15 min temporal and 2×2 km^2 spatial resolution derived from the BALTRAD collaboration (Michelson et al., 2018). The version used in this paper is the "tas BALTRAD " and it is essentially BALTRAD is almost identical to the BRDC product used in Swedenexcept that it . The main difference is that it covers a much larger area and does not include the HIPRAD adjustments. Bias correction bias adjustments. Instead, bias correction in the BALTRAD 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 extended coverage in the BALTRAD product is made possible thanks to the automatic radar data exchange exchange of radar data between neighboring countries around the Baltic sea (i.e., Norway, Sweden, 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 ofBALTRAD) . Nevertheless, by analyzing and comparing the BALTRAD for the 50 top events fact that the BALTRAD product spans multiple countries makes it particularly interesting for evaluating and comparing performances with respect to tailored national products. This means that direct comparisons with the BALTRAD are available for (most of) the top 50 events identified in Denmark, Sweden and Finland , important conclusions about the advantages and limitations of tailored high-resolution national radar products can be madeFinland and Sweden. Unfortunately, the Netherlands are currently not part of BALTRAD which means that no further comparisons are possible for the Dutch C-band product.

2.3 Performance metricsComparison of radar and gauge measurements

Since radar and gauges measure rainfall at different scales based on using different measuring principles, one does can 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 averages over large resolution volumes 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 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 intensity by up to 25-30% in heavy rain and windy conditions (e.g., Nystuen, 1999; Chang an . On the other hand, 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 Missing data in one or both of the sensors also further complicate the comparison (Vasiloff et al., 2009). Therefore, the main goal here is not will not be to make a statement about which measurement is closer sensor comes closest 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 study useful to monitor the performance and consistency of operational radar and gauge products or study the propagation of rainfall ~~measurement uncertainty~~ uncertainties in hydrological models (Rossa et al., 2011).

~~To assess performance, the average discrepancies~~

2.3.1 Bias estimation

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):~~

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

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

gauge observations are assessed with the help of a multiplicative error model:

$$R_r(t) = \beta \cdot R_g(t) \cdot \varepsilon(t) \tag{1}$$

where ~~X_i and Y_i represent the radar and rain gauge measurements, N is the number of observations~~ $R_r(t)$ (in mmh^{-1}) denote the radar measurements a time t, $R_g(t)$ (in mmh^{-1}) the gauge measurements, ~~$\mu_{X|Y}$ 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,~~ β [-] the multiplicative bias ~~is calculated by taking the gauge measurements Y_i (in mmh^{-1}) as a reference value:~~

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

~~where ε_i are~~ and $\varepsilon(t)$ [-] are independent, identically distributed random errors drawn from a ~~continuous and positive probability distribution (e.g., a~~ log-normal ~~)~~ distribution 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~~ and scale parameter $\sigma_\varepsilon > 0$ (Smith and Krajewski, 1991). The multiplicative bias in Equation (1) can also be expressed in terms of the log-ratios of radar versus gauge values:

$$\ln\left(\frac{R_r(t)}{R_g(t)}\right) = \ln(\beta) + \ln(\varepsilon(t)) \tag{2}$$

where $\ln(\varepsilon(t))$ is a Gaussian random variable with mean 0 and variance σ_ε^2. Equation (2) can be used to detect the presence of conditional bias with intensity by checking whether the expected value of the log-ratio $\ln\left(\frac{R_r(t)}{R_g(t)}\right)$ depends on $R_g(t)$ or not. Note that the multiplicative bias model in ~~Equation~~Equations (1) ~~provides~~ and (2) has been shown to provide a better, physically more plausible representation of the error structure between in-situ and remotely-sensed rainfall observations than ~~a the~~ the classical additive bias model ~~commonly used in statistics (e.g., Tian et al., 2013).~~ used in linear regression (e.g., Tian et al., 2013). It assumes that the discrepancies between radar and gauge measurements are the result of two error contributions: a deterministic component β that accounts for systematic errors in radar and gauge measurements (e.g., due to calibration, wind effects, wrong Z-R relationship, ...) and a random term $\varepsilon(t)$ that represents sampling errors and noise in radar and gauge observations. Since gauges are not seen as ground truth in this study, $\varepsilon(t)$ is assumed to contain all possible sources of errors in both the gauge and radar observations, including the ones due to differences in sampling volumes (Ciach and Krajewski, 1999b). The last point is particularly important as radar sampling volumes can be up to 7 orders of magnitude larger than that of rain gauges (Ciach and Krajewski, 1999a). This means that even if both sensors would be perfectly calibrated, their measurements would still disagree with each other due to the fact that rain gauge measurements made at a particular location within a radar pixel are usually not representative of averages over larger areas. In their paper, Ciach and Krajewski (1999a) proposed a rigorous statistical framework for assessing this representativeness error based on the spatial autocovariance function and the notion of extension variance. However, their approach was developed for an additive error model and can not be directly applied here. Instead, we propose a comparatively simpler approach in which the differences in sampling volumes are already included in the random errors $\varepsilon(t)$. Our approach is based on the assumption that the errors $\varepsilon(t)$ have a log-normal distribution with median 1 and scale parameter $\sigma_\varepsilon > 0$, which means that we must have $\mathbb{E}[\varepsilon(t)] = \exp(\frac{\sigma_\varepsilon^2}{2}) \neq 1$. Furthermore, if we assume that $R_g(t)$ and $R_r(t)$ are second-order stationary random processes with fixed mean μ_g and μ_r and variances σ_g^2 and σ_r^2 and that the random errors $\varepsilon(t)$ are identically distributed and independent from $R_g(t)$, then we get the following system of equations:

~~In this paper, the multiplicative bias is estimated through the so-called G~~

$$\begin{cases} \mathbb{E}[R_g(t)] &= \beta \cdot \mathbb{E}[R_r(t)] \cdot \mathbb{E}[\varepsilon(t)] = \beta \cdot \mu_r \cdot \exp(\frac{\sigma_\varepsilon^2}{2}) \\ \mathrm{Var}[R_g(t)] &= \beta^2 \cdot \mathrm{Var}[R_r(t)] \cdot \mathrm{Var}[\varepsilon(t)] = \beta^2 \cdot \sigma_r^2 \cdot \exp(\sigma_\varepsilon^2) \cdot \big(\exp(\sigma_\varepsilon^2) - 1\big) \end{cases} \tag{3}$$

From the first equation we get $\beta^2 = \frac{\mu_g^2}{\mu_r^2} \cdot \exp(-\sigma_\varepsilon^2)$ which can be plugged into the second equation to get an estimate of the scale parameter $\hat{\sigma}_\varepsilon$:

$$\hat{\sigma}_\varepsilon^2 = \ln\left(1 + \frac{\sigma_g^2 \mu_r^2}{\sigma_r^2 \mu_g^2}\right) = \ln\left(1 + \frac{\mathrm{CV}_g^2}{\mathrm{CV}_r^2}\right). \tag{4}$$

where $\mathrm{CV}_{g|r} = \frac{\sigma_{g|r}}{\mu_{g|r}}$ denotes the coefficient of variation of the gauge and radar values respectively. Substituting, we get the following estimate for β:

$$\hat{\beta} = \frac{\mu_g}{\mu_r} \cdot \exp(-\frac{\hat{\sigma}_\varepsilon^2}{2}). \tag{5}$$

The first term $\frac{\mu_g}{\mu_r}$ in Equation (5) is known as the 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~~ R ratio

(Yoo et al., 2014) and it quantifies the apparent bias between radar and gauge measurements. The second term $\exp(-\frac{\hat{\sigma}_\varepsilon^2}{2})$ is a bias adjustment factor that accounts for the fact that gauge and radar measurements do not have the same mean and variance (e.g., ~~least squares andmaximum likelihood)have been proposed depending on the distribution of ε_i but~~ due to differences in sampling volumes and/or different measurement uncertainties). The "true" underlying model bias β is obtained by multiplying the two terms together. However, it is important to keep in mind that only 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~~is directly observable from the data while β heavily depends on the assumptions that the errors are log-normally distributed with median 1 and independent from the radar observations. Note that σ_ε and β could also be estimated through Equation (2) by calculating the mean and standard deviation of $\ln\left(\frac{R_g(t)}{R_r(t)}\right)$. However, this approach is not recommended as the ratios for small rainfall rates can be very noisy and numerical errors will arise whenever one of the measurements is zero.

~~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:~~ For readers not familiar with the interpretation of multiplicative biases, note that it is also possible to express the G/R ratio and model bias β as an average relative error. In this case, we have:

$$\epsilon_{\mathrm{rel}} Err_{\mathrm{avg}} = 100\% \cdot \mathbb{E}\left[\frac{Y_i - X_i}{Y_i} \frac{R_g(t) - R_r(t)}{R_g(t)}\right] = 1 - \frac{1}{\mathrm{MB}} \frac{1}{\beta} \cdot \mathbb{E}\left[\frac{1}{\varepsilon_i} \frac{1}{\varepsilon(t)}\right] = 1 - \frac{1}{\mathrm{MB}} 1 - \frac{\exp(\sigma_\varepsilon^2) \cdot \left(\exp(\sigma_\varepsilon^2) - 1\right)}{\beta} \quad (6)$$

where ~~$\mathbb{E}$ denotes the expectation and by definition the median of ε_i is assumed to be equal to 1.~~ we used the fact that $\frac{1}{\varepsilon(t)}$ is also a log-normal with median 1 and scale parameter σ_ε. However, for simplicity and robustness, we prefer to report the median relative error which is independent of the variance of $\varepsilon(t)$:

$$Err_{\mathrm{med}} = \mathrm{Med}\left[\frac{R_g(t) - R_r(t)}{R_g(t)}\right] = 1 - \frac{1}{\beta} \cdot \mathrm{Med}\left[\frac{1}{\varepsilon}\right] = 1 - \frac{1}{\beta} \quad (7)$$

~~While standard error metrics like RRMSE, CC and MB provide an important overview of the average error, they may~~

2.3.2 Peak intensity bias

Equation (5) provides a convenient way to estimate the average bias between radar and gauge measurements over the course of an event. However, in reality, the true bias is likely to fluctuate over time as a function of the spatio-temporal characteristics and intensity of the considered events and their location with respect to the radar(s). Consequently, the G/R ratio and model bias β might 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~~ an event. To account for this, we also consider the peak rainfall intensity bias (PIB) between radar and gauges. The PIB is defined as:

$$Y_{\max}(\Delta t) R_g^{\max} = \mathrm{PIB}(\Delta t) \cdot X_{\max}(\Delta t) R_r^{\max} \quad (8)$$

where ~~$Y_{\max}(\Delta t)$ and $X_{\max}(\Delta t)$~~ $R_g^{\max}$ and $R_r^{\max}$ denote the maximum rain rate values recorded by the gauges and radar ~~at temporal aggregation time scale Δt~~over the course of an event. 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~~ gauge and radar time series, which means that the maximum values may not necessarily correspond to the same time interval. The ~~advantage of~~ main reason for this is that it leads to more reliable and robust ~~PIB estimates at high resolutions where statistics would otherwise be strongly sensitive~~ estimate of PIB at high spatial and temporal resolutions and reduces the sensitivity to small timing ~~issues~~ differences between radar and gauge observations due to wind and vertical variability.

2.3.3 Other metrics

To complement the bias analysis and provide a more comprehensive overview of the agreement between gauge and radar measurements, we also calculate standard error metrics such as the Spearman rank correlation coefficient (CC), root mean square difference (RMSD) and relative root mean square difference $\mathrm{RRMSD} = \frac{\mathrm{RMSD}}{\mu_g}$ between gauge and radar values. All these statistics are calculated on an event-by-event basis at a fixed aggregation time scale.

3 Results

3.1 Agreement during the 4 most intense events

Figure 4 shows the time series of rainfall intensities ~~at the highest available temporal resolution~~ for the top ~~event~~ events 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~~(i.e., ~~1.37, 1.55 and 1.69 for~~ Denmark, the Netherlands, Finland ~~, and Sweden , respectively. In other words, according to equation (**??**), 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 intenserain~~and Sweden respectively). Each of these events is highly intense, 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~~ During the third rainfall peak ~~was particularly impressive, with rain rates remaining~~ in Denmark, rain rates remained well above 125 mmh^{-1} for three consecutive 5-min 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~~15 minutes, 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~~ which is 3.39 times less than what was measured by the gauge. Note that this does not necessarily imply that the radar estimates are wrong, as rain gauge data can also suffer from large biases in times of heavy rain and are not directly comparable to radar due to the large difference in sampling volumes. Nevertheless, all 4 depicted events show a strong, systematic pattern of underestimation by radar compared with the gauges. The G/R ratios, as defined in Equation 5, are 1.66, 1.37, 1.55 and 1.68 respectively, which corresponds to a relative difference in rainfall rates 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 of 27-40%. This order of magnitude is consistent with previous values reported in the literature. For example Goudenhoofdt et al. (2017) mentioned a 30% underestimation of radar compared with gauges in Belgium and Seo et al. (2015) found up to 50% underestimation on individual events in the United States.

Despite being biased, radar and gauge measurements are rather consistent with each other in terms of their temporal structure (e.g., rank correlation values of 0.92, 0.75, 0.80 and 0.85 for Denmark, Finlandthe Netherlands, Finland and Sweden respectively). Also, a substantial part of the apparent bias is likely attributable to differences in sampling volumes. According to Equation (5), the bias adjustment factor $e^{-\sigma_\varepsilon^2/2}$ is 0.63, 0.59, 0.66, 0.70 in Denmark, the Netherlands, Finland and Sweden respectivelyconfirm this hypothesis. During the most intense parts of the storms, radar underestimates by 42-54%compared with the gauges (i.e. . The "true" underlying model bias β for the 4 depicted events is therefore estimated to be 1.04, about 10-15% more than suggested by the average multiplicative bias)0.81, 1.02 and 1.18. In other words, once the differences in scale between radar and gauge data have been accounted for, radar only appears to underestimate rainfall rates by a factor 1.04 (3.8%) in Denmark, 1.02 (2.0%) in Finland and 1.18 (15.3%) in Sweden. In the Netherlands, radar values even seem to be overestimated by a factor 1.23 (18.7%). However, it is important to remind the reader that these values should be interpreted very carefully as they heavily depend on the assumption that the random errors between radar and gauges are independent and log-normally distributed with median 1. Figure 4 suggests that this might not be the case, as the bias between radar and gauges appears to considerably fluctuate over time and increase during the peaks (see Section 3.3 for more details). In this case, the peak intensity biases were 2.17 in Denmark, 2.09 in Finland, 1.98 in the Netherlands and 1.73 in Sweden, which are consistently larger than the average G/R ratios.

3.2 Overall agreement between radar and gauges

In the following, we consider the overall agreement between radar and gauges for all 50 top events is analyzedeach country. Figure 5 shows the radar rainfall intensities versus the gauge estimates at the highest available temporal resolution rainfall intensities of radar versus gauges for each country (e.g., 5 min for Denmark, 10 min for the Netherlands and Finland and 15 min for Swedenat the highest temporal resolution). 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 Results show a good consistency between the two sensors (i.e., 0.71-0.83), indicating a good agreement in terms of temporal structurerank correlation coefficients between 0.77-0.91). However, the radar clearly underestimates the rainfall intensity compared with the gauges, Multiplicative bias values intensities measured by radar are clearly lower than that of the gauges. The G/R ratios are 1.59 for Denmark, 1.41 1.40 for the Netherlands, 1.56 for Finland and 1.66 for Swedenwhich corresponds to an underestimation of 37.1%, 29.1%, 35.8, corresponding to median relative differences of 38.8%, 28.4%, 35.9%, and 39.839.7% respectively.

Figure 6 provides a similar overview of the discrepancies between In addition to the bias, we also see a significant amount of scatter with relative root mean squares differences between 116.4% and 139.1% (depending on the country). This is characteristic for sub-hourly aggregation time scales and can be explained by the large spatial and temporal variability of rainfall and the fact that radar and gauges for do not measure precipitation at the same height and over the same volumes.

Since it can be hard to compare gauge and radar measurements over short aggregation time scales, additional analyses were carried out to better understand how resolution affects the discrepancies between the two rainfall sensors. Figure 6 shows the scatter plot of radar versus gauge estimates when the data are aggregated to 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 strongly reduces the scatter (i.e., RRMSD between 38.8% and 47.7%) and further increases the correlation coefficient (i.e., 0.80-0.92), making it easier to see the bias. The G/R ratio (see Section 2.3). The good agreement remains the same, as values only depend on total accumulation and not on the temporal resolution at which the events are sampled. The fact that radar and gauges agree more at the event scale is clearly encouragingbut must be interpreted carefully as improvements are mostly due to the inclusion of many lower intensity rainfall periods during which than at the sub-hourly scale is encouraging. However, improvements are mainly attributed to the fact that many of the large discrepancies affecting the rainfall peaks get smoothed out during aggregation. This leads to an overly optimistic assessment of the agreement between 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 intensitiesthat is not necessarily representative of what happens during the most intense parts of the events.

Based on Figures 5and 6, one could conclude that the values of the G/R ratio in Figure 5, the Dutch C-band radar product composite appears to have the best overall agreement with the gauges among all countrieslowest apparent bias of all products (28.4%), followed by Finland , Denmark and Sweden(35.9%), Denmark (38.8%) and Sweden (39.7%). However, such direct comparisons would not really be are not really fair, as one also needs to they do not take into account the differences in different spatial and temporal resolutions between of the radar products, the number of radars used during the estimation and their distances to the considered rain gauges. They also ignore the fact that the top 50 events in each country do not have the same intensities, durations and spatio-temporal structures. For example, the events in Denmark are significantly more intense compared with the Netherlands, Finland and Sweden, which might help explain some of the differences. Also, the longest event in the Danish database only lasted 4 hours, which is shorter than for the other countries. To better separate the two, empirically derived areal-reduction factors (ARFs) proposed by Thorndahl et al. (2019) were used understand the origin of the bias and interpret the differences between the countries, additional, more detailed analysis are necessary.

The first analysis we did was to estimate the theoretical bias between a point measurement and an areal-average from radar (i.e., using model bias β in 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 5) under the assumption that the errors are log-normally distributed with median 1. Table 3 shows the estimated values of μ_g, μ_r, σ_g, σ_r and σ_ε at the highest available temporal resolution for each radar product (all 50 events combined). The obtained β values are 1.04 for Denmark, 0.94 for the Netherlands, 1.11 for Finland and 1.11 for Sweden. This leads to a radically different assessment of the bias between radar and gauge values than with the G/R ratio. According to the β values, the Danish product has the lowest model bias (3.8%), followed by the Netherlands (-6.4%), Finland (9.9%) and Sweden (9.9%). The Dutch radar product again appears to slightly overestimate the rainfall intensity, which is counter-intuitive given that the actual radar values are 30-40% lower than the gauges on average. However, this can be explained purely due to differences in measurement support by the fact that β accounts for the relative variability of the rain gauge and radar observations around their respective means (see Equations 4-5). Products for which CV_g is larger than CV_r therefore see their bias reduced. This makes sense as gauge measurements are expected to have a larger coefficient of variation than radar due to their smaller sampling volume (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 thatpoint estimate versus areal average). Another reason is that gauges are known to suffer from relatively large sampling uncertainties at sub-hourly time scales. The fact that Denmark uses RIMCO tipping bucket gauges (as opposed to the float gauges in the Netherlands and weighing gauges in Finland and Sweden) therefore also makes a difference when calculating β. The bias adjustment factor $\exp(\frac{-\sigma_\varepsilon^2}{2})$ combines all these different factors together, making it possible to compare the different radar products on a fairer basis. However, one has to keep in mind that β is a theoretical bias that strongly depends on the adequacy of the model proposed in Equation (1). Further analyses presented in the next section show that some of these assumptions might not be very realistic. Still, it is quite encouraging to see that, contrarily to what the G/R ratio suggests, the actual bias in the radar products 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 ?? summarizes the agreement of each product differences in scale could be as low as 10%. Moreover, the products with the highest spatial/temporal resolutions also seem to be affected by the lowest bias (in absolute value).

We see that measurement support biasobviously 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

3.3 Conditional bias with intensity

The analyses performed in Sections 3.1 and 3.2 are useful to understand the overall agreement between radar and gauges over a large number of events but the estimated values strongly depend on the assumption that the bias β in Equation (1) is constant. Our initial analysis in Section 3.1 already showed that in reality, the bias is likely to fluctuate over time, increasing in times of heavy rain. As mentioned in the introduction, time and intensity-dependent biases in radar or gauge estimates are highly problematic because they affect the timing and magnitude of peak flow predictions in hydrological models. Here, we perform a more quantitative assessment of this effect by studying the conditional bias between radar and gauges with respect to the rainfall intensity. Figure 7 shows the log ratio of rain gauge versus radar estimates $\ln(\frac{R_g(t)}{R_r(t)})$ as a function of the rainfall intensity $R_g(t)$ recorded by the rain gauges. Each dot in these graphs represents a measurement (at the highest available temporal resolution) and all 50 events have been combined into a single plot.

The multiplicative bias model in Equation (1) assumes that the average log-ratio is constant (i.e., equal to the log bias). However, Figure 7 shows that three out of the four main radar products exhibit a clear conditional bias with intensity. The only product for which the bias does not increase with intensity is the Finnish OSAPOL. Incidentally, the Finnish OSAPOL is also the only product in which heavy rainfall rates are estimated through differential phase instead of reflectivity, pointing to the advantage of polarimetry over fixed Z-R relationships. The relative rates at which the multiplicative biases β in Equation (5) increase with intensity are 1.09% per mmh^{-1} in Denmark, 0.86% in the Netherlands, 0.09% in Finland and 2.12% in Sweden. This may not seem large but can make a big difference when rainfall intensities vary from 1 mmh^{-1} to more than 100 mmh^{-1}. For example, in Denmark, the multiplicative bias increases from 0.92 at 1 mmh^{-1} to 2.69 at 100 mmh^{-1}. In Sweden, the bias varies from 1.49 at 1 ~~km, 10 min resolution in~~ mmh^{-1} to 11.96 at 100 mmh^{-1}. By contrast, the multiplicative biases at 100 mmh^{-1} for 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~~ only reach values of 2.48 and 2.40 respectively. The fact that both the Danish and Swedish products have large conditional biases also explains why the overall apparent bias (as estimated through the G/R ratio) of these two products is larger than for the Netherlands and Finland.

The most likely explanation for the conditional bias with intensity is the fact that 3 out of the 4 main radar products use a fixed Marshall-Palmer Z-R relationship to estimate rainfall rates from reflectivity. Therefore, 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~~ bias will grow whenever the raindrop size distribution deviates from the ~~density of the rain gauge networks used to~~ Marshall-Palmer, as is usually the case during strong convective precipitation and high rainfall intensities. The mean field bias-adjustments based on rain gauge data can help reduce the overall bias by tuning the prefactor in the Z-R relationship. However mean field bias adjustments are insufficient to account for the rapid changes in raindrop size distributions in heavy rain. Previous studies suggest that the best way to mitigate biases and ensure accurate hydrological predictions is to frequently 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, data over time (Löwe et al., 2014). This might also explain why the Swedish and Danish radar products which are corrected using daily gauge data have a stronger conditional bias with intensity than the Dutch product which uses hourly corrections. Another even better strategy, as demonstrated by the low conditional bias of the Danish database contains events that are significantly more intense compared with the Netherlands, Finland and Sweden (see Figure ??). Also, the longest event in the Danish database only lasted 4 hours, which is significantly less than for the other countries. Finnish OSAPOL product, is to replace the Z-R relation by a R(Kdp) retrieval which is known to be less sensitive to variations in drop size distributions and calibration effects (Wang and Chandrasekar, 2010).

A deeper analysis of this issue confirms that on average, higher rainfall intensities appear to be linked with slightly larger multiplicative biases

3.4 Other sources of bias

The conditional bias with intensity explains a lot of the differences between the radar products. However, this is only one part of the story and other confounding factors such as the distance between the radar(s) and the gauges also need to be considered. Figure 8 shows the log-ratio of gauge versus radar estimates $\ln(\frac{R_g(t)}{R_r(t)})$ as a function of the distance to the nearest radar. Compared with intensity, the trend with distance appears to be much weaker. Out of the 4 considered products, only the Danish C-band exhibits a trend that is significantly different from zero (at the 5% level). This makes sense given that the Danish product only considers data from a single radar and only applies a mean field bias correction, making it more likely to be affected by range effects such as overshooting, non-uniform beam filling and attenuation. Based on our analyses, the multiplicative bias β increases by 0.73% per km. However, since the range of distances between radar and gauges in Denmark is relatively small (from 29.2 to 74.2 km), bias values only vary from 1.06 to 1.47 at minimum and maximum distances respectively. Distance therefore only plays a minor role in explaining the variations in bias compared with intensity. Interestingly, the composite products in the Netherlands and Finland do not seem to suffer from significant conditional biases with distance, highlighting the advantage of combining data from different radars and viewpoints to mitigate range effects. The Swedish product currently does not combine measurements from multiple radars in an optimal way, only using the measurements from the best (i.e., nearest) radar. 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 ??(a) , and the large mismatches in terms of peak rainfall intensities in Figure ??(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. Swedish BRDC also contains an additional range-dependent bias correction (see Section 2.2.4) that appears to be rather efficient at removing large-scale trends with distance. However, the strong conditional bias with intensity in the Swedish BRDC also makes it harder to see potential range-dependent biases in the first place.

Before diving deeper into Another important aspect that needs to be considered when comparing the radar products is the difference in spatial and temporal resolutions. One way to study this would be to aggregate all radar products to a 2×2 km^2

and 30 min time scales before comparing them. However, this is not recommended as simple arithmetic averaging of processed radar fields does not really mimic what a lower resolution radar would see (e.g., due to the non-linear relation between rain rate and reflectivity and the multiple post-processing steps applied to the rainfall estimates). A better approach is to derive so-called areal-reduction factors (ARFs). Several ways to estimate ARFs have been proposed in the literature. ARFs can be estimated through the analysis of the ~~peak rainfall intensities , we finish this sub-section by taking a closer look at the overall agreement~~ spatial correlation structure (Rodríguez-Iturbe and Mejía, 1974; Ciach and Krajewski, 1999a) or more empirically as the ratio between maximum areal-averaged rainfall intensities between radar and gauges ~~as a function of the temporal aggregation time scale. Figure **??** 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,~~ (Thorndahl et al., 2019). Here, the latter approach is used, specifically, Equation (8) in Thorndahl et al. (2019) with $b_1 = 0.31$, $b_2 = 0.38$ and $b_3 = 0.26$. Using the calculated ARFs, we estimated that the average bias between a point measurement and the Danish radar ~~product clearly exhibits the lowest relative errors and highest correlation coefficients. It is followed by the Dutch and Finnish products~~ estimates (0.25 km^2, 5 min) should be in the order of 13%. For Finland and the Netherlands (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.~~ ^{2}, 10 min), ~~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~~ average underestimation should be about 19% and 30% for Sweden (4 km^2, 15 min). Table 4 summarizes the G/R ratios before and after subtracting the areal-reduction factors above. The new multiplicative biases between radar and gauges after taking into account the ARFs are 1.39 in Denmark, 1.14 in the Netherlands, 1.27 in Finland and 1.17 in Sweden. This corresponds to median relative differences of 28%, 12.2%, 21.2% and 14.5% with respect to the gauges. The best products in terms of residual bias after applying the ARF would therefore be the Dutch, followed by the Swedish, Finnish and Danish. However, this is a rather simplistic way of accounting for the difference in scale that does not take into account the spatio-temporal structures and different characteristics of top 50 rain events in each country. Also, it is highly questionable whether it makes sense to apply areal-reduction factors to the radar data in the first place since most of the products (except the Finnish OSAPOL~~product (which has not~~) have 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~~using gauges. Part of the differences in measurement support bias should therefore already have been accounted for during the bias adjustments. Also, the fact that the ARFs used in this paper were derived from Danish radar data only and using a different collection of events might not be optimal. A more elaborate approach with variable ARFs for each country/event might provide a more realistic assessment of the support bias.

Future studies with denser rain gauge networks could take a more detailed look at this. In particular, it would be interesting to know whether the conditional bias in Section 3.3 is mostly due to support bias (with higher rainfall intensities corresponding to higher ARFs) or to natural variations in raindrop size distributions (through the Z-R relation).

3.5 Agreement during the peaks

While the previous section heavily focused on the overall agreement between radar and gauges, this sectiontakes In this section, we take a closer look at the peakshow well the rainfall peaks are captured by the radar. Figure 9 shows the underestimation of peak rainfall intensity 10%, 25%, 50%, 75% and 90% quantiles of peak intensity bias between radar and gauges as a function of aggregation time scalefor 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 apparent bias (i.e., the G/R ratio). We see that the Netherlands and Finland) is characterized by a median underestimation of peak rainfall intensity (have relatively low median peak intensity biases of 1.82 and 1.88 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%) resolution (approximately 1.2-1.3 times higher 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., bias). 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 averageMB). 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 on the other hand have substantially higher median PIB values of 2.96 and 2.24, (1.86 respectively 1.35 times higher than the average). Moreover, the rate at which the PIB decreases with aggregation time scale is different in each country. In Denmark and Sweden, the PIB remains well above the average multiplicative bias across bias for 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 up to 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 peakshours while in the Netherlands and Finland, the PIB converges much faster to the mean bias (i.e., after approx. 60 min for the Netherlands and 20 min for Finland). This is no coincidence and can be explained by the fact that the Netherlands use hourly rain gauge data to bias correct their radar estimates while the Danish and Swedish products use daily bias adjustment factors. Thorndahl et al. (2014a) showed that switching from daily to hourly mean field bias adjustments can slightly improve peak rainfall estimates but also 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~~ Here, we see that this strategy can result in a severe increase of the peak intensity bias ~~values in group 2 could be that~~ at sub-hourly scales, with some of the radar-gauge pairs differing by more than a factor 5. The Dutch radar product also exhibits a rapid increase in PIB at sub-hourly scales. However, the overall bias at 10 min resolution rarely exceeds more than a factor 3. The Finnish product is interesting, as it is the only that has not been bias corrected with gauges. Its strength is that it makes use of polarimetry (i.e., Kdp) to estimate rainfall rates during the peaks. This seems to result in almost identical performances in terms of PIBs than a traditional approach based on Z-R relationship with hourly bias corrections, as used in the Netherlands. The only notable difference is the rate at which the peak intensity bias converges to the average bias, with the Finnish product exhibiting a lower dependence on the aggregation time scale than the Dutch product.

Another explanation for the high peak intensity biases in Denmark and Sweden could be that these two countries 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~~ based on a weighted average of overlapping radar measurements (with weights depending on the ~~quality of the measurement and the distance between~~ distance to 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~~ elevation angle). Clearly, the ability to combine measurements from multiple radars and viewpoints ~~appears to play a crucial role~~ is an advantage in times of heavy rain~~, perhaps even more than spatial resolution.~~

3.6 ~~Sensitivity to temporal aggregation time scale~~

. However, quantifying this would require additional dedicated experiments (e.g., with/without compositing) that are beyond the scope of this study. Moreover, since we have already established that range-dependent biases only play a minor role in this study, the effects of radar compositing on the total bias and peak intensity bias are likely to be small and limited to a few events.

Another equally interesting result ~~of this study concerns~~ is the fact that ~~biases in peak rainfall intensities~~ peak intensity biases for specific events do not necessarily become smaller when ~~moving~~ the data are aggregated to a coarser time scale. Figure 10 illustrates this point by showing ~~,~~ the values of PIBs 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.

While in the Netherlands and Finland the ~~bias~~ PIB 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~~ These examples show that even though globally speaking, the peak intensity bias between radar and gauges converges to the average multiplicative bias when the data are aggregated ~~over longer time periods~~to coarser time scales, 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 with daily rain gauge adjustments 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~~ exhibited their 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~~ Finnish and Dutch radar products only contained 14 and 8 such events, respectively. ~~A deeper~~ Further 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,~~ Some 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~~were also identified. 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.~~

3.6 Results for the additional radar products

~~Figure 12summarizes~~ Figures 12(a)-(d) summarize the results obtained for the X-band radar system in Denmark. ~~It shows that overall,~~ Figure 12a) shows that there is a ~~relatively good agreement between the X-band rainfall estimates and the gauges~~fairly good consistency between the radar and gauge estimates (rank correlation coefficient of 0.87). The multiplicative bias (G/R ratio) 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~~ difference is 12.5 mmh^{-1} (98.0%)~~but it is significantly smaller compared with~~. The scatter is therefore large but slightly lower than for the C-band products (116-139%). ~~The~~ Note that the statistics for the X-band radar 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~~ as for the top 50 events in the Netherlands, Finland and Sweden~~(see Figure ??)~~. The total ~~accumulated~~ rainfall amounts per event (~~i.e.,~~ 10-30 mm) were lower though, ~~suggesting that~~ as the events sampled by the X-band system were rather short and localized.

~~Clearly, the high resolution of~~ The model bias β in Equation (1) is 0.77, which suggests that after accounting for the differences in scales, the X-band radar ~~and the dual-polarization capabilities seem to improve the overall agreement between the radar and~~ overestimates the rainfall rates compared with the gauges. ~~Nevertheless, the bias affecting the peaks remains high. The median underestimation of peak rainfall intensityat~~ However, this is a statistical artifact caused by our initial assumption that the error terms are independent of intensity. This is not true here as 5 min ~~was approximately 40%~~ tipping rain gauge data tend to be affected by larger sampling uncertainties at low rain rates. This causes the rain gauge data to be more variable (CV$_g$=1.61) compared with the radar measurements (CV$_r$=1.34) and results in overestimated noise terms $\varepsilon(t)$ and underestimated bias. In addition to the sampling issue, Figure 12b) also shows that there is a clear conditional bias with intensity (0.88% per mmh^{-1}). One reason for this conditional bias with intensity could be attenuation, which is ~~slightly better than for the~~ known to play a major role at X-band. However, all reflectivity measurements have been corrected for attenuation prior to rainfall estimation. Also, Figure 12c) shows that there is no obvious change in bias with the distance to the radar, as would be expected for attenuated signals. This leads us to conclude that similarly to the Danish and Swedish C-band ~~products in the Netherlands and Finland and significantly better than for the C-band radarin 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~~ products, the conditional bias with intensity is caused by the use of a fixed Z-R relation (together with daily bias adjustments). Figure 12d) shows how the conditional bias with intensity affects the accuracy of the radar in times of heavy rain. The median peak intensity bias at 5 min is 1.64 (39%) with 10% of the PIBs exceeding 3.1 (67.7%). Similarly to the Danish C-band and Swedish C-band products, the peak intensity bias only slowly decreases with aggregation time scale, remaining well above the average G/R ratio up to 2 h. This is consistent with our previous findings and ~~suggests that resolution and polarimetry alone are~~ is a good reminder that resolution alone is not sufficient to accurately capture the rainfall peaks. Based on the analysis of the C-band products, ~~one way to further reduce these biases during the peaks would be to use~~ the most promising way to reduce the conditional bias with intensity is to replace the fixed Z-R relationship with a R(Kdp) estimate in times of heavy rain or to use hourly or sub-hourly rain gauge data for the bias correction. Current research done at KNMI and DMI is also investigating the possibility to retrieve rainfall rates from reflectivity measurements at horizontal and vertical polarizations or to combine polarimetric data from 2 or more overlapping ~~X-band systems.~~ radar systems. However, this is still ongoing research.

Figure 13 compares the agreement between the ~~individual~~ 4 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~~ aggregation time scale of 15 min, which might introduce some additional sampling uncertainty. 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 Overall, the BALTRAD seems to perform rather similarly to the national products. It has slightly lower rank correlation coefficients and higher root mean square differences. The bias (as measured by the G/R ratio) is also very similar, except in Sweden where the BALTRAD appears to underestimate more with respect to the gauges (1.77 versus 1.66). This makes sense given that the BALTRAD does not include the HIPRAD adjustments which results in higher overall bias and conditional bias with intensity. Interestingly, the BALTRAD performs worse than the Danish C-band product in terms of overall bias but better in terms of median peak intensity bias. However, since the Finnish OSAPOL product is not bias adjusted, other factors must be at play here. One of them There are many possible explanations for these differences. One reason 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 difference in spatial resolution (2 km for BALTRAD versus 500 m for the Danish C-band). Another reason could be the differences in the bias adjustment schemes, more specifically the fact that BALTRAD uses monthly gauge data to correct for bias while 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 is adjusted on a daily basis. However, this does not explain why 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 bias, which is likely due to its lower spatial resolution of 2 km. We is lower in the BALTRAD. While this remains rather speculative, 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 BALTRADincludes 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 countriesHowever, this does not seem to result in systematic improvements across all events. Indeed, it is worth pointing out that while the median PIB value is lower in BALTRAD, the average PIB value is slightly larger in BALTRAD (3.0) than for the Danish C-band product (2.63). The same applies to all the other countries as well (2.49 versus 2.05 for Finland and 3.27 versus 2.60 for Sweden). In other words, there are some events in the database for which BALTRAD has significantly larger PIB values than others. These are the events responsible for the strong conditional bias with intensity. For these events, the bias is most likely due large deviations from the theoretical Marshall-Palmer Z-R relationship, which can not be mitigated with the help of compositing.

4 Conclusions

Rain rate estimates from The accuracy of 6 different radar products in 4 countries (Denmark, Finland, the Netherlands and Sweden) have has 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 consistency (correlation coefficient between 0.7-0.8-0.9). However, due to the large differences in sampling volume between gauges and radar, relative root mean square errors remained high (120-150the scatter at sub-hourly time scales remains high (98-144% at 5-15 min). Moreover, all 6 radar products exhibited a clear pattern of underestimation. The multiplicative biases at 5-15 min were between 1.20-1.77, suggesting that radar underestimates rainfall rates by 17-44% compared with gauges. A substantial part of the discrepancies could be attributed to differences in spatial measurement support through the use of bias (i.e., 10-30% according to 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) is likely due to differences in sampling volumes. However, this remains hard to quantify precisely in the absence of dense rain gauge networks. An alternative bias model that accounts for the differences in mean and variance between radar and gauge measurements suggested that the actual bias affecting radar rainfall estimates could be as low as 10%. Moreover, higher resolution radar products seemed to agree better with gauges, which is encouraging. At the same time, these conclusions strongly rely on the assumption that errors are log-normally distributed and independent of intensity, which, as we have seen in this study, is likely not to be true during the peaks.

Based on our analysis, the main issue affecting current operational radar rainfall estimates is the fact that the multiplicative bias increases with rainfall intensity. The most likely reason for this conditional bias is the use of a fixed Marshall-Palmer Z-R relationship to convert reflectivity to rainfall rates, which does not account for the changes in raindrop size distributions during heavy convective precipitation events. One way to mitigate the conditional bias 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,~~ as demonstrated by the Finnish OSAPOL project, is to rely on differential phase shift Kdp instead of reflectivity. Another possibility is to use a fixed Z-R relationship but to perform frequent bias adjustments with the help of rain gauges (as demonstrated by the Dutch C-band product). Here, the temporal resolution of the gauge data appears to play crucial role in controlling the magnitude of the conditional bias, with daily and monthly corrections resulting in an increase of the bias of approximately 2% per mmh^{-1} and hourly adjustments resulting in an increase of about 1% per mmh^{-1}. Nevertheless, even the hourly adjustments appeared to be insufficient for radar to adequately capture the peaks. Regardless of how rainfall rates were estimated, median peak intensity biases systematically exceeded the average G/R ratios, reaching values of 1.8-3.0 (i.e., radar underestimates by 44-67%). Occasionally, 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~~ peak intensity bias even exceeded 80% (factor of 5). We believe that sub-hourly bias adjustments might help further reduce the bias affecting the peaks. However, this only applies to the peaks and is not recommended for low to moderate rainfall intensities due to the large uncertainty affecting rain gauge measurements. Future research should focus on finding better ways to dynamically adjust radar data with the help of rain gauge measurements at different temporal resolutions depending on event dynamics, amounts and intensities.

Overall, the X-band data for Denmark showed ~~very~~ promising results, outperforming all other C-band products in terms of accuracy and correlation, thereby demonstrating the value of high-resolution rainfall observations for urban hydrology. However, ~~this last result must be interpreted very carefully as~~ due to the shorter data record, only 10 events over 2 years ~~were consideredfor 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~~ could be considered. The polarimetric estimates from the Finnish OSAPOL project also showed promising performance, which is remarkable considering the fact that they were not adjusted by any gauges. However, it should also be pointed out that for now, the overall performance of the ~~Dutch and Finnish C-band products (despite their slightly lower resolution). By contrast, the single-radar~~ OSAPOL remains similar to that of the Dutch 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~~product with fixed Z-R relationship and hourly bias correction. Interestingly, the distance between the radar and the gauges did not appear to have a strong effect on peak intensity bias. We explain this by the fact that range-dependent biases tend to be small compared with the large spatial variability of rain at the event scale. Therefore, range effects are masked by other errors and only become visible when the radar data are aggregated over the course of several days or months.

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 and the resolution of the rain gauge data used for the adjustments, 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~~ in radar data depending on aggregation time scale ~~has already been pointed out in the past, but~~ still represents a major challenge ~~in terms of how to correctly represent~~ as it limits our ability to accurately characterize rainfall extremes and ~~rainfall measurement~~ uncertainties in hydrological models across scales (Bruni et al., 2015). One way to partially mitigate this effect is to combine measurements from multiple radars. However, more research is necessary to precisely quantify this part of the error.

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~~little focus has been given to the analysis of the rain gauge data themselves. 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~~ Since the gauge data are likely to be underestimated as well, the actual bias 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~~ sensors might be larger than suspected. The second issue is the relatively short length of the observational record (10-15 years) which meant that only a small number of extreme rain events could be considered. Moreover, it is worth mentioning that some of the events in the database actually occurred on the same day but were captured by different gauges at different locations. The derived statistics might therefore be biased towards characterizing the performance of the radar during these days instead of the average performance over a large number of independent events. Another issue 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~~ identical radar systems and different levels of processing (e.g., by switching on/off individual correction schemes) ~~for identical radar systems would help~~ would be useful to get a ~~more detailed view into~~ better understanding of 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~~retrieval techniques within a more controlled setting. Despite all these limitations, the present study already provided some important insight into the major issues affecting radar-rainfall estimates in times of heavy rain. Also, several useful strategies for mitigating errors and reducing biases were identified. Future research should focus on analyzing more radar products and identifying the most promising strategies for improving performance in each country.

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.

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.

References

Anagnostou, M. N., Kalogiros, J., Anagnostou, E. N., Tarolli, M., Papadopoulos, A., and Borga, M.: Performance evaluation of high-resolution rainfall estimation by X-band dual-polarization radar for flash flood applications in mountainous basins, J. Hydrol., 394, 4–16, https://doi.org/10.1016/j.jhydrol.2010.06.026, 2010.

Andréassian, V., Perrin, C., Michel, C., Usart-Sanchez, I., and Lavabre, J.: Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models, J. Hydrol., 250, 206–223, https://doi.org/10.1016/S0022-1694(01)00437-1, 2001.

Aronica, G., Freni, G., and Oliveri, E.: Uncertainty analysis of the influence of rainfall time resolution in the modelling of urban drainage systems, Hydrol. Processes, 19, 1055–1071, https://doi.org/10.1002/hyp.5645, 2005.

Baeck, M. L. and Smith, J. A.: Rainfall Estimation by the WSR-88D for Heavy Rainfall Events, Weather Forecast., 13, 416–436, https://doi.org/10.1175/1520-0434(1998)013<0416:REBTWF>2.0.CO;2, 1998.

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.

Bruni, G., Reinoso, R., van de Giesen, N. C., Clemens, F. H. L. R., and ten Veldhuis, J. A. E.: On the sensitivity of urban hydrodynamic modelling to rainfall spatial and temporal resolution, Hydrol. Earth Syst. Sci., 19, 691–709, https://doi.org/10.5194/hess-19-691-2015, 2015.

Chandrasekar, V., Keranen, R., Lim, S., and Moisseev, D.: Recent advances in classification of observations from dual polarization weather radars, Atmos. Res., 119, 97–111, https://doi.org/10.1016/j.atmosres.2011.08.014, 2013.

Chang, M. and Flannery, L. A.: Spherical gauges for improving the accuracy of rainfall measurements, Hydrol. Processes, 15, 643–654, https://doi.org/10.1002/hyp.181, 2001.

Ciach, G. J.: Local random errors in tipping-bucket rain gauge measurements, J. Atmos. Oceanic Technol., 20, 752–759, https://doi.org/10.1175/1520-0426(2003)20<752:LREITB>2.0.CO;2, 2003.

Ciach, G. J. and Krajewski, W. F.: On the estimation of radar rainfall error variance, Adv. Water Resour., 22, 585–595, https://doi.org/10.1016/S0309-1708(98)00043-8, 1999a.

Ciach, G. J. and Krajewski, W. F.: Radar-Rain Gauge Comparisons under Observational Uncertainties, J. Appl. Meteorol., 38, 1519–1525, https://doi.org/10.1175/1520-0450(1999)038<1519:RRGCUO>2.0.CO;2, 1999b.

Collier, C. G.: Flash flood forecasting: What are the limits of predictability?, Q. J. R. Meteorol. Soc., 133, 3–23, https://doi.org/10.1002/qj.29, 2007.

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.

Courty, L. G., Rico-Ramirez, M. A., and Pedrozo-Acuna, A.: The Significance of the Spatial Variability of Rainfall on the Numerical Simulation of Urban Floods, Water, 10, 1–17, https://doi.org/10.3390/w10020207, 2018.

Cristiano, E., ten Veldhuis, M.-C., and van de Giesen, N.: Spatial and temporal variability of rainfall and their effects on hydrological response in urban areas – a review, Hydrol. Earth Syst. Sci., 21, 3859–3878, https://doi.org/10.5194/hess-21-3859-2017, 2017.

Cunha, L. K., Mandapaka, P. V., Krajewski, W. F., Mantilla, R., and Bradley, A. A.: Impact of radar-rainfall error structure on estimated flood magnitude across scales: An investigation based on a parsimonious distributed hydrological model, Water Resour. Res., 48, W10515, https://doi.org/10.1029/2012WR012138, 2012.

Cunha, L. K., Smith, J. A., Krajewski, W. F., Baeck, M. L., and Seo, B.-C.: NEXRAD NWS Polarimetric Precipitation Product Evaluation for IFloodS, J. Hydrometeorol., 16, 1676–1699, https://doi.org/10.1175/JHM-D-14-0148.1, 2015.

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.

Delrieu, G., Nicol, J., Yates, E., Kirstetter, P.-E., Creutin, J.-D., Anquetin, S., Obled, C., Saulnier, G.-M., Ducrocq, V., Gaume, E., Payrastre, O., Andrieu, H., Ayral, P.-A., Bouvier, C., Neppel, L., Livet, M., Lang, M., du Châtelet, J., Walpersdorf, A., and Wobrock, W.: The Catastrophic Flash-Flood Event of 8-9 September 2002 in the Gard Region, France: A First Case Study for the Cévennes-Vivarais Mediterranean Hydrometeorological Observatory, J. Hydrometeorol., 6, 34–52, https://doi.org/10.1175/JHM-400.1, 2005.

Delrieu, G., Wijbrans, A., Boudevillain, B., Faure, D., Bonnifait, L., and Kirstetter, P.-E.: Geostatistical radar-raingauge merging: A novel method for the quantification of rain estimation accuracy, Adv. Water Resour., 71, 110–124, https://doi.org/10.1016/j.advwatres.2014.06.005, 2014.

Dupasquier, B., Andrieu, H., Delrieu, G., Griffith, R. J., and Cluckie, I.: Influence of the VRP on High Frequency Fluctuations Between Radar and Raingage Data, Phys. Chem. Earth, 25, 1021–1025, https://doi.org/10.1016/S1464-1909(00)00146-5, 2000.

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.

Gu, J.-Y., Ryzhkov, A., Zhang, P., Neilley, P., Knight, M., Wolf, B., and Lee, D.-I.: Polarimetric Attenuation Correction in Heavy Rain at C Band, J. Appl. Meteor. Clim., 50, 39–58, https://doi.org/10.1175/2010JAMC2258.1, 2011.

He, X., Sonnenborg, T. O., Refsgaard, J. C., Vejen, F., and Jensen, K. H.: Evaluation of the value of radar QPE data and rain gauge data for hydrological modeling, Water Resour. Res., 49, 5989–6005, https://doi.org/10.1002/wrcr.20471, 2013.

Holleman, I.: Bias adjustment and long-term verification of radar-based precipitation estimates, Meteorol. Appl., 14, 195–203, https://doi.org/10.1002/met.22, 2007.

Holleman, I. and Beekhuis, H.: Review of the KNMI clutter removal scheme, Tech. Rep. TR-284, Royal Netherlands Meteorological Institute KNMI, available online at www.knmi.nl/publications/fulltexts, 2005.

Holleman, I., Huuskonen, A., Kurri, M., and Beekhuis, H.: Operational monitoring of weather radar receiving chain using the sun, J. Atmos. Oceanic Technol., 27, 159–166, https://doi.org/10.1175/2009JTECHA1213.1, 2010.

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.

KNMI: Handbook for the Meteorological Observation, Tech. rep., Koninklijk Nederlands Meteorologisch Instituut, De Bilt, Netherlands, Available at http://projects.knmi.nl/hawa/pdf/Handbook_H01_H06.pdf, 2000.

Koistinen, J. and Pohjola, H.: Estimation of Ground-Level Reflectivity Factor in Operational Weather Radar Networks Using VPR-Based Correction Ensembles, J. Appl. Meteor. Clim., 53, 2394–2411, https://doi.org/10.1175/JAMC-D-13-0343.1, 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.

Löwe, R., Thorndahl, S., Mikkelsen, P. S., Rasmussen, M. R., and Madsen, H.: Probabilistic online runoff forecasting for urban catchments using inputs from rain gauges as well as statically and dynamically adjusted weather radar, J. Hydrol., 512, 397–407, https://doi.org/10.1016/j.jhydrol.2014.03.027, 2014.

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.

Madsen, H., Gregersen, I. B., Rosbjerg, D., and Arnbjerg-Nielsen, K.: Regional frequency analysis of short duration rainfall extremes using gridded daily rainfall data as co-variate, Water Sci. Technol., 75, 1971–1981, https://doi.org/10.2166/wst.2017.089, 2017.

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.

Niemi, T. J., Warsta, L., Taka, M., Hickman, B., Pulkkinen, S., Krebs, G., Moisseev, D. N., Koivusalo, H., and Kokkonen, T.: Applicability of open rainfall data to event-scale urban rainfall-runoff modelling, J. Hydrol., 547, 143–155, https://doi.org/10.1016/j.jhydrol.2017.01.056, 2017.

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.

Ntelekos, A. A., Smith, J. A., and Krajewski, W. F.: Climatological Analyses of Thunderstorms and Flash Floods in the Baltimore Metropolitan Region, J. Hydrometeorol., 8, 88–101, https://doi.org/10.1175/JHM558.1, 2007.

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.

Ochoa-Rodriguez, S., Wang, L.-P., Gires, A., Pina, R. D., Reinoso-Rondinel, R., Bruni, G., Ichiba, A., Gaitan, S., Cristiano, E., van Assel, J., Kroll, S., Damian Murlà-Tuyls, D., Tisserand, B., Schertzer, D., Tchiguirinskaia, I., Onof, C., Willems, P., and ten Veldhuis, M.-C.: Impact of spatial and temporal resolution of rainfall inputs on urban hydrodynamic modelling outputs: A multi-catchment investigation, J. Hydrol., 531, 389–407, https://doi.org/10.1016/j.jhydrol.2015.05.035, 2015.

Ogden, F. L. and Julien, P. Y.: Runoff model sensitivity to radar rainfall resolution, J. Hydrol., 158, 1–18, 1994.

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.

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

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.

Rafieeinasab, A., Norouzi, A., Kim, S., Habibi, H., Nazari, B., Seo, D.-J., Lee, H., Cosgrove, B., and Cui, Z.: Toward high-resolution flash flood prediction in large urban areas - Analysis of sensitivity to spatiotemporal resolution of rainfall input and hydrologic modeling, J. Hydrol., 531, 370–388, https://doi.org/10.1016/j.jhydrol.2015.08.045, 2015.

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.

Rico-Ramirez, M. A., Liguori, S., and Schellart, A. N. A.: Quantifying radar-rainfall uncertainties in urban drainage flow modelling, J. Hydrol., 528, 17–28, https://doi.org/10.1016/j.jhydrol.2015.05.057, 2015.

Rodríguez-Iturbe, I. and Mejía, J. M.: On the transformation of point rainfall to areal rainfall, Water Resour. Res., 10, 729–735, https://doi.org/10.1029/WR010i004p00729, 1974.

Rossa, A., Liechti, K., Zappa, M., Bruen, M., Germann, U., Haase, G., Keil, C., and Krahe, P.: The COST 731 Action: a review on uncertainty propagation in advanced hydro-meteorological forecast systems, Atmos. Res., 100, 150–167, https://doi.org/10.1016/j.atmosres.2010.11.016, 2011.

Ruzanski, E., Chandrasekar, V., and Wang, Y. T.: The CASA nowcasting system, J. Atmos. Oceanic Technol., 28, 640–655, https://doi.org/10.1175/2011JTECHA1496.1, 2011.

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.

Saltikoff, E., Haase, G., Delobbe, L., Gaussiat, N., Martet, M., Idziorek, D., Leijnse, H., Novák, P., Lukach, M., and Stephan, K.: OPERA the Radar Project, Atmosphere, 10, 1–13, 2019.

Schilling, W.: Rainfall data for urban hydrology: what do we need?, Atmos. Res., 27, 5–21, https://doi.org/10.1016/0169-8095(91)90003-F, 1991.

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., Seo, D. J., Baeck, M. L., and Hudlow, M. D.: An intercomparison study of NEXRAD precipitation estimates, Water Resour. Res., 32, 2035–2045, https://doi.org/10.1029/96WR00270, 1996.

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.

Stransky, D., Bares, V., and Fatka, P.: The effect of rainfall measurement uncertainties on rainfall-runoff processes modelling, Water Sci. Technol., 55, 103–111, 2007.

Thomsen, R. S. T.: Drift af Spildevandskomitéens RegnmålersystemÅrsnotat 2015, Tech. rep., DMI, Copenhagen, accessed: 2019-12-13, 2016.

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., Nielsen, J. E., and Jensen, D. G.: Urban pluvial flood prediction: a case study evaluating radar rainfall nowcasts and numerical weather prediction models as model inputs, Water Sci. Technol., 74, 2599–2610, https://doi.org/10.2166/wst.2016.474, 2016.

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.

Vejen, F.: Teknisk rapport 06-15, Nyt SVK system, Sammenligning af nedbørmålinger med nye og nuværende system, Tech. rep., DMI, Copenhagen, accessed: 2019-12-13, 2006.

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.

Villarini, G., Smith, J. A., Baeck, M. L., Sturdevant-Rees, P., and Krajewski, W. F.: Radar analyses of extreme rainfall and flooding in urban drainage basins, J. Hydrol., 381, 266–286, https://doi.org/10.1016/j.jhydrol.2009.11.048, 2010.

Wang, Y. and Chandrasekar, V.: Algorithm for Estimation of the Specific Differential Phase, J. Atmos. Oceanic Technol., 26, 2565–2578, https://doi.org/10.1175/2009JTECHA1358.1, 2009.

Wang, Y. T. and Chandrasekar, V.: Quantitative precipitation estimation in the CASA X-band dual-polarization radar network, J. Atmos. Oceanic Technol., 27, 1665–1676, https://doi.org/10.1175/2010JTECHA1419.1, 2010.

Wessels, H. R. A. and Beekhuis, J. H.: Stepwise procedure for suppression of anomalous ground clutter, in: Proc. COST-75, pp. 270–277, Weather Radar Systems, International Seminar, Brussels, Belgium, 1995.

WMO: Guide to Meteorological Instruments and Methods of Observation, WMO-No.8, World Meteorological Organization, Geneva, 7th ed. edn., 2008.

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.: Hydroclimatology of flash flooding in Atlanta, Water Resour. Res., 48, W04524, https://doi.org/10.1029/2011WR011371, 2012.

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.

Yang, L., Smith, J., Baeck, M. L., Smith, B., Tian, F., and Niyogi, D.: Structure and evolution of flash flood producing storms in a small urban watershed, J. Geophys. Res. Atmos., 121, 3139–3152, https://doi.org/10.1002/2015JD024478, 2016.

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.

Zhou, Z., Smith, J. A., Yang, L., Baeck, M. L., Chaney, M., Ten Veldhuis, M.-C., Deng, H., and Liu, S.: The complexities of urban flood response: Flood frequency analyses for the Charlotte metropolitan region, Water Resour. Res., 53, 7401–7425, https://doi.org/10.1002/2016WR019997, 2017.

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.

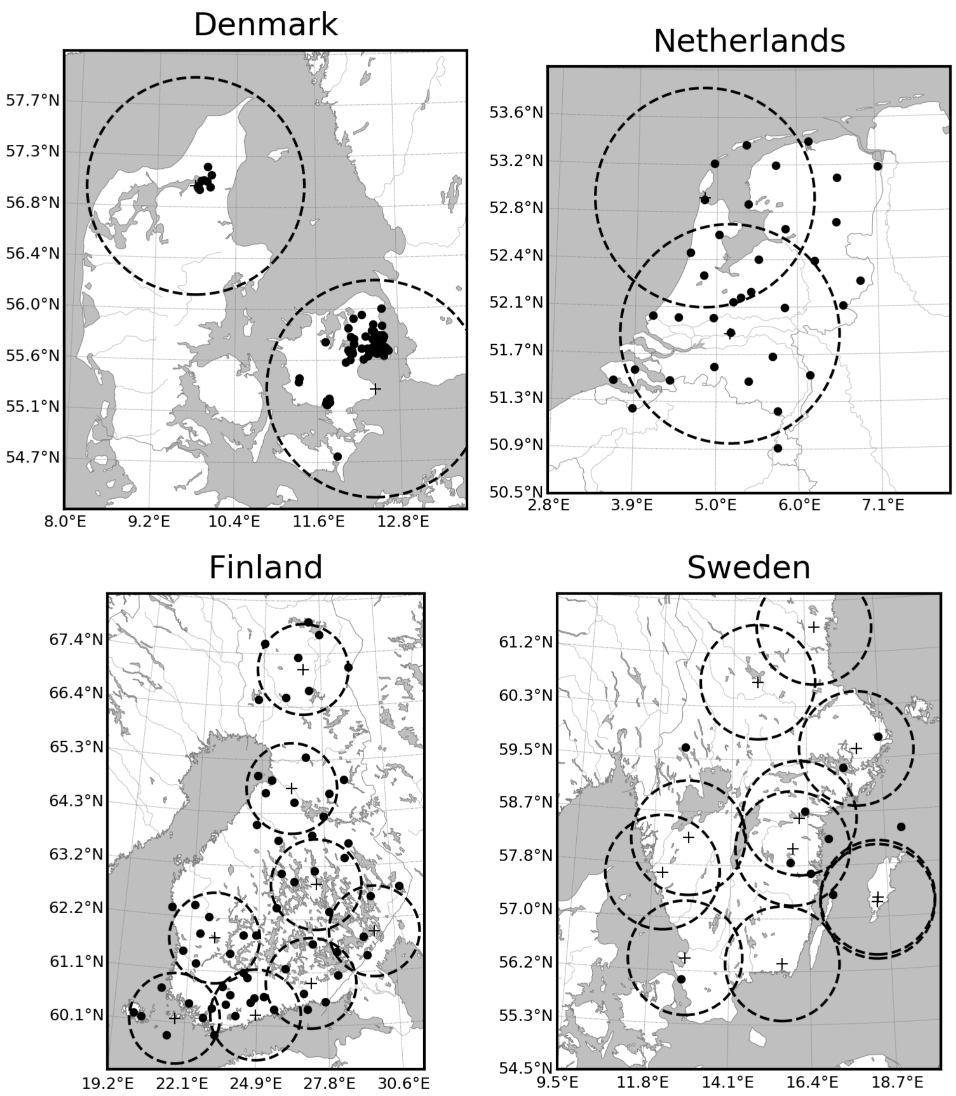

Figure 1. The four considered study areas in Denmark, the Netherlands, Finland and Sweden with the used rain gauges (~~red diamonds~~black dots) and the location of the C-band radars ~~(~~marked by black crosses~~)~~. The dashed lines denote circles of 100 km radius around each radar. Due to maintenance and relocations, not all the radars were operating at the same time.

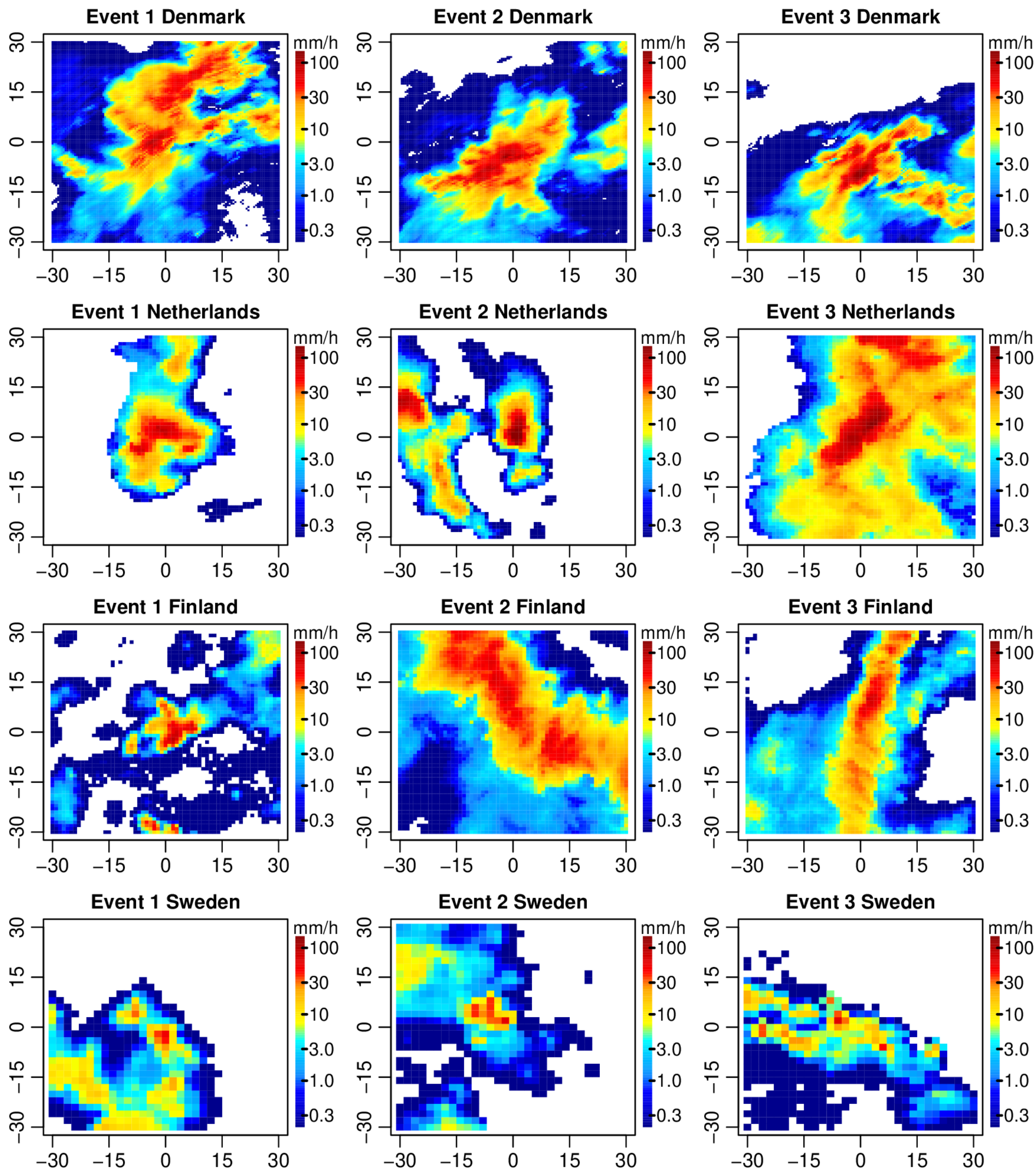

Figure 2. Snapshots of the ~~3 most intense events for each country~~ radar rainfall estimates (in mmh^{-1}) at the time of peak intensity ~~(~~for the 3 most intense events in ~~mm/h)~~each country. Each map is a square of size 60×60 km^2 with the gauge ~~location~~ located in the center of the domain.

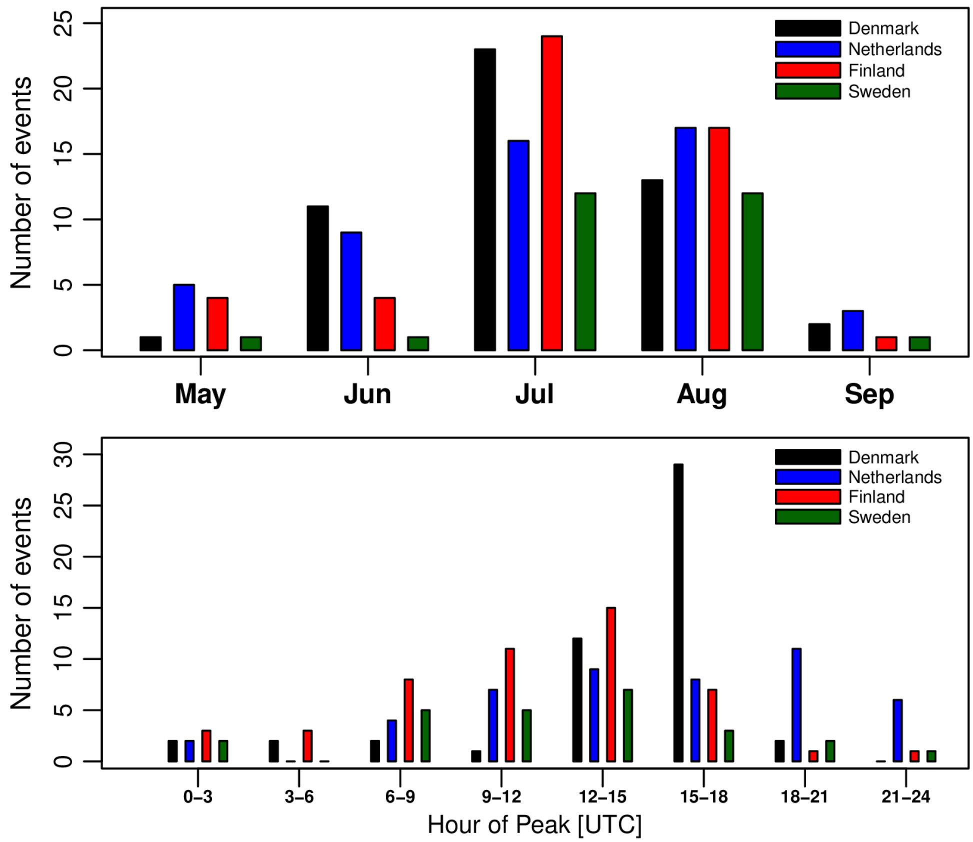

Figure 3. Distribution of the 50 top events over the month (top panel) and hour of the day (bottom panel).

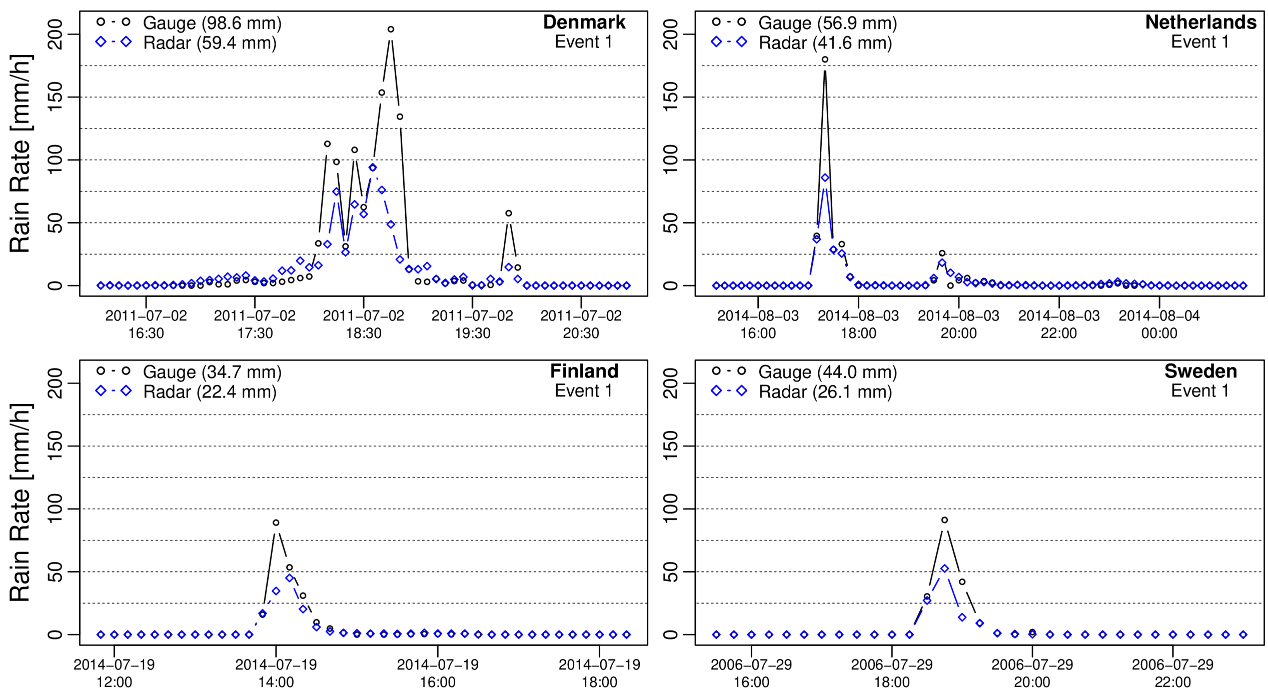

Figure 4. Time series of radar and gauge intensities (in ~~mm/h~~mm h^{-1}) ~~at the highest available temporal resolution~~ for the most intense event ~~of~~ in each country.

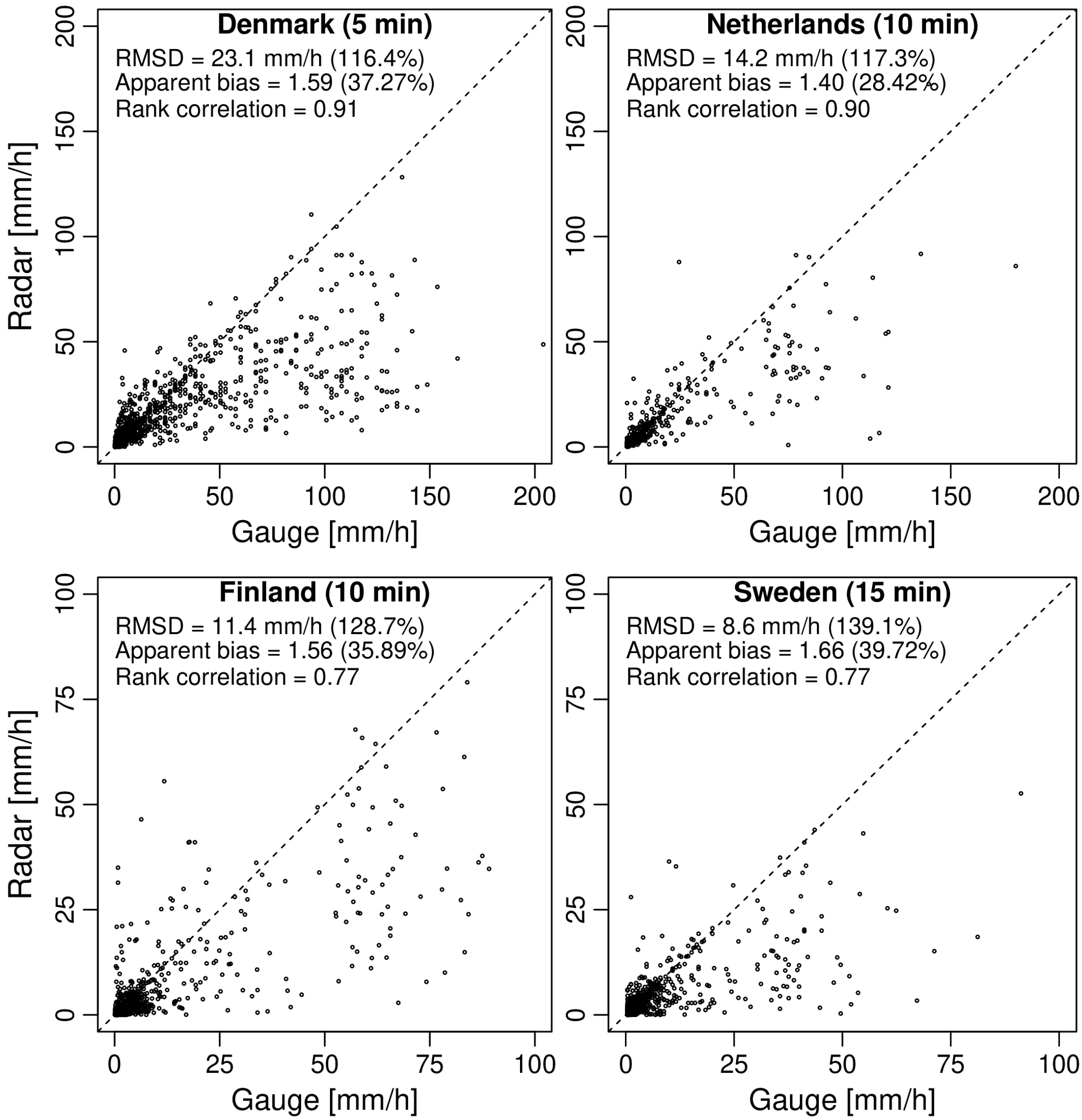

Figure 5. Radar versus gauge intensities (in ~~mm/h~~mmh^{-1}) at the highest available temporal resolution for each country (all 50 events combined). The ~~dotted~~ dashed line represents the diagonal.

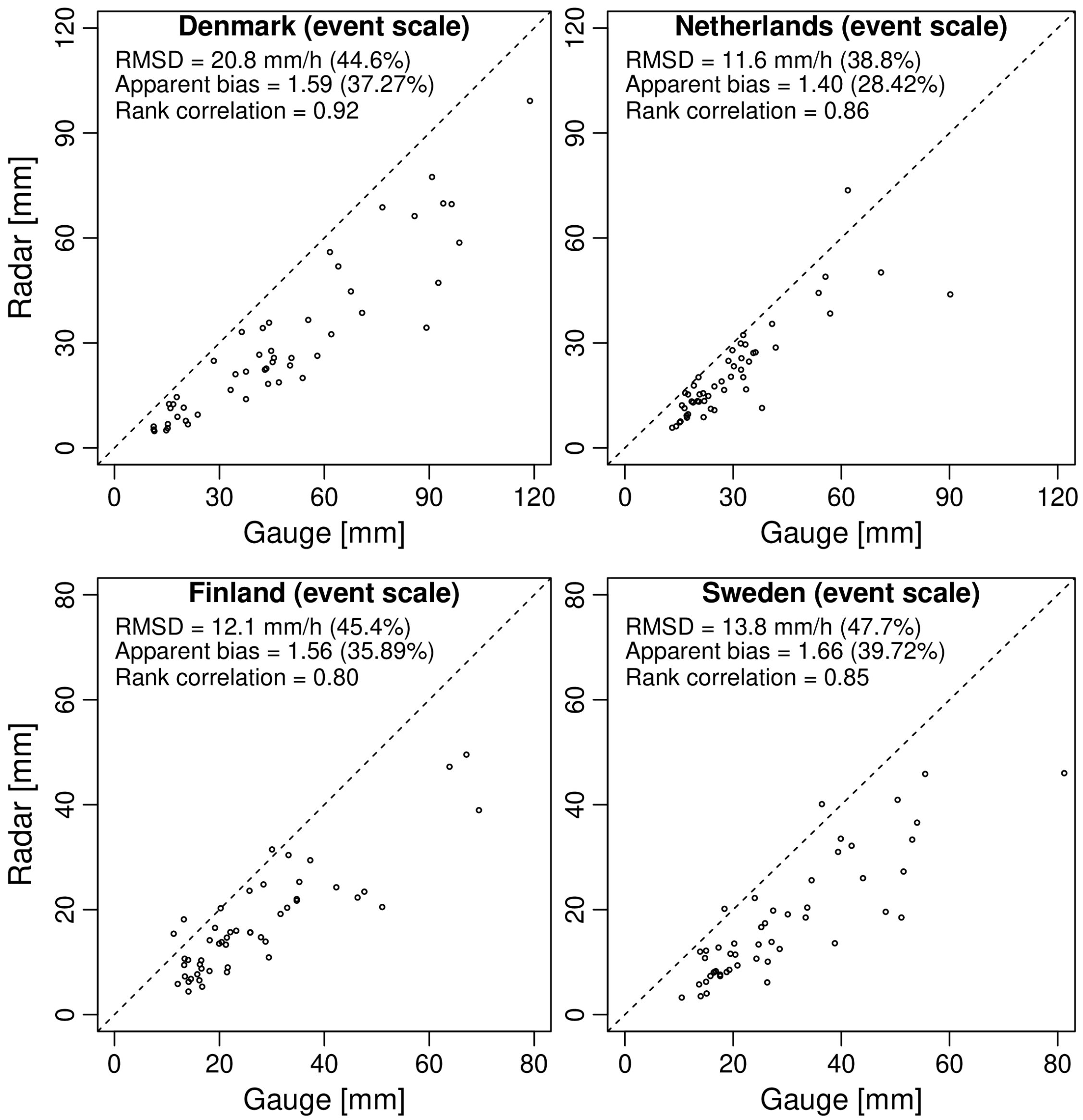

Figure 6. Radar versus gauge accumulations (in mm) at the event scale for each country (i.e., one dot per event). The ~~dotted~~ dashed line represents the diagonal.

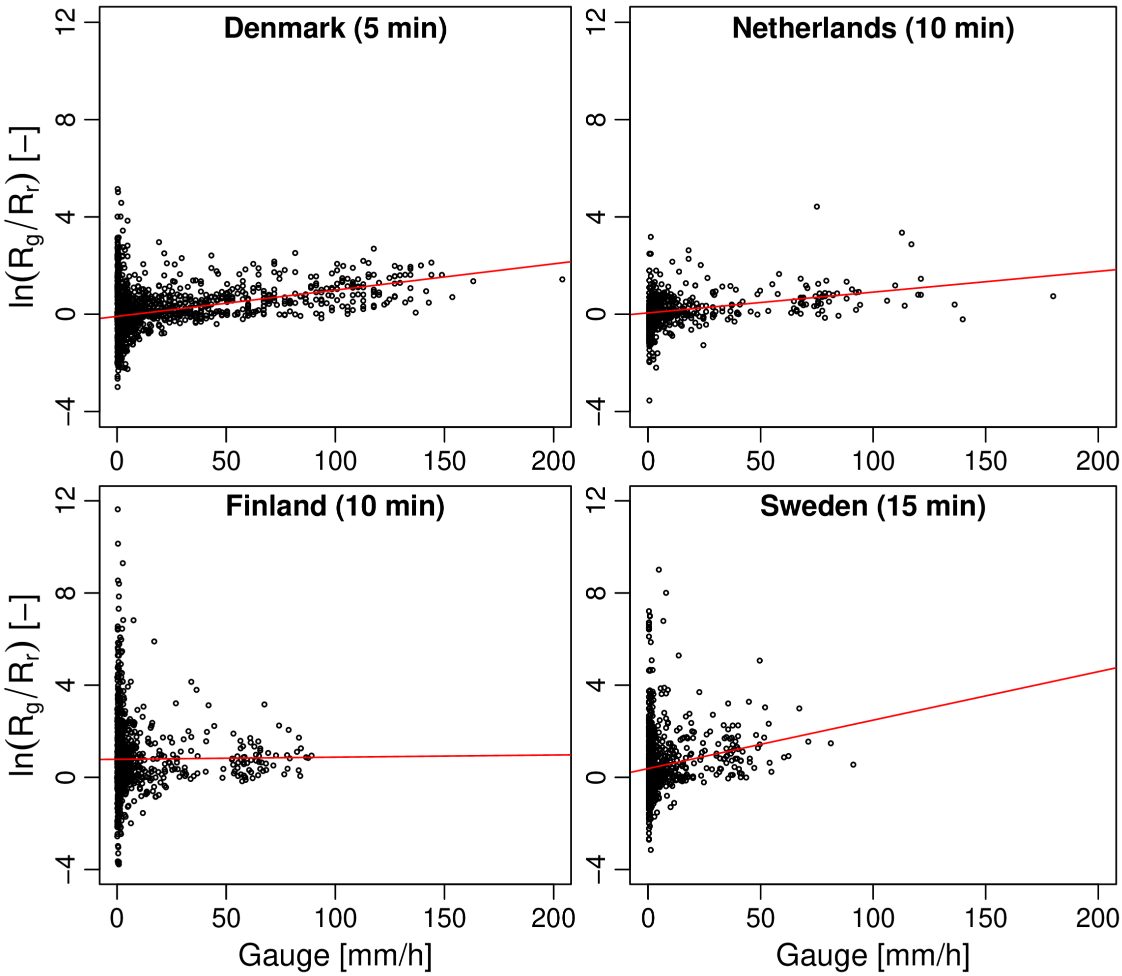

Figure 7. ~~Total rainfall accumulations (in mm), peak rainfall intensities~~ Log ratio of gauge over radar values as a function of rain gauge intensity (in mmh^{-1}) ~~and event duration (in hours)~~ for each country. The ~~boxplots denote the 10%, 25%, 50%, 75% and 90% quantiles of~~ red lines represent the ~~50 events in each country~~fitted linear regression models.

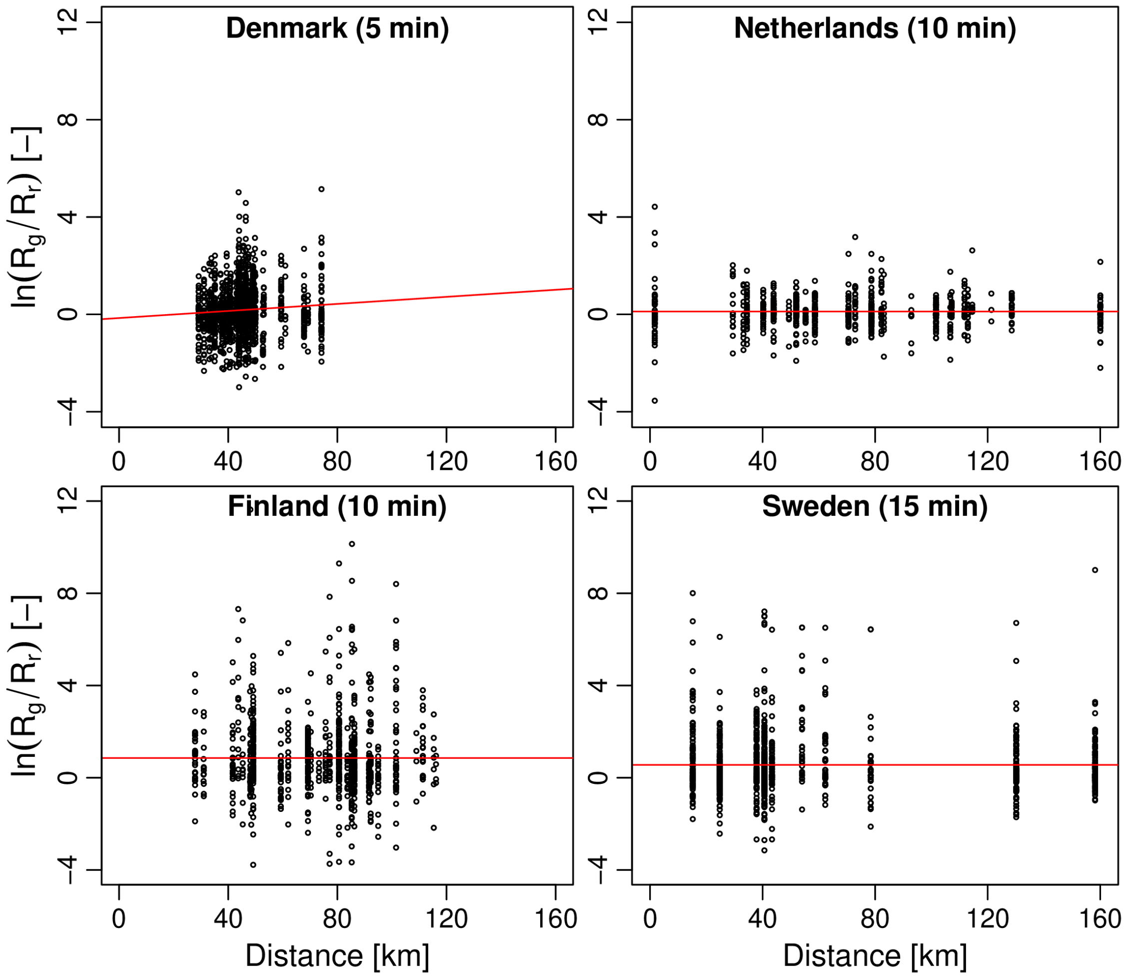

Figure 8. ~~Relative root mean square error and correlation coefficients~~ Log ratio of ~~radar versus rain~~ gauge ~~estimates at different aggregation time scales between 5 min and 2 h (all 50 events combined)~~over radar values as a function of the distance to the nearest radar. The red line represents the fitted linear regression model.

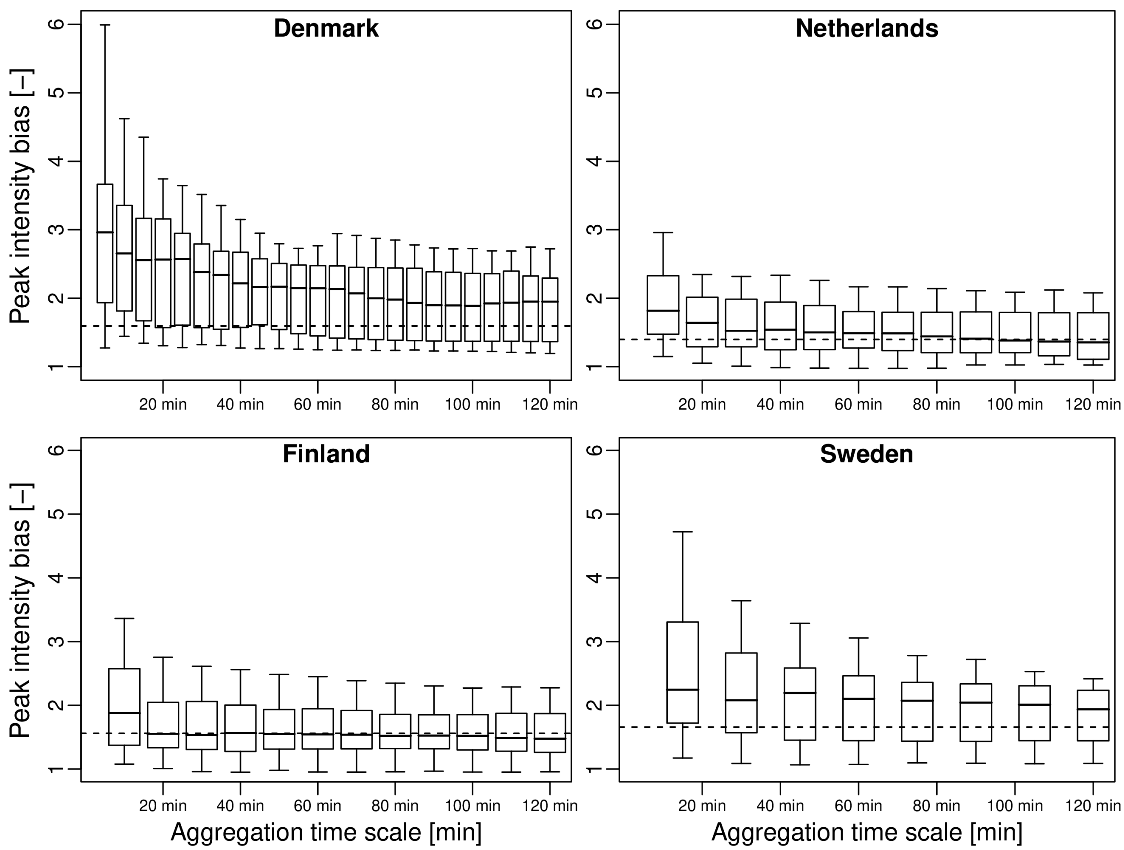

Figure 9. ~~Underestimation~~ Boxplots of peak ~~rainfall~~ intensity ~~by radar compared with gauges (expressed in %)~~ bias versus ~~temporal~~ aggregation time scale. Each boxplot represents the 10%, 25%, 50%, 75% and 90% quantiles for the 50 top events in each country. The horizontal lines ~~represent~~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

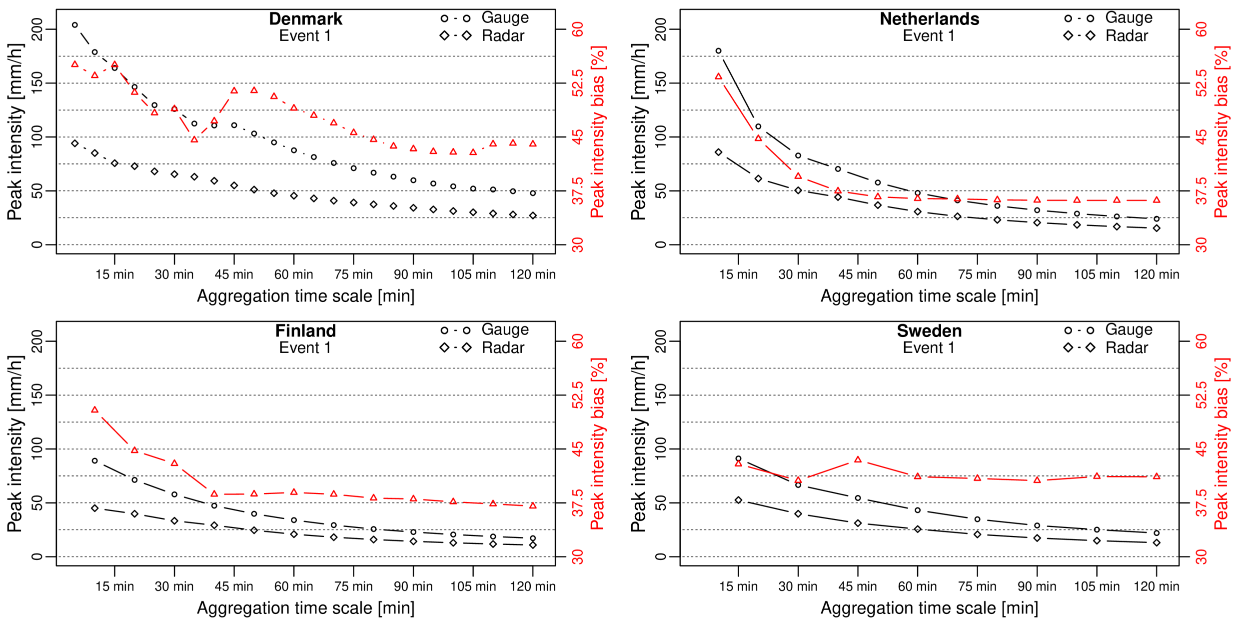

Figure 10. Peak rainfall intensities measured by radar and gauges for the top 1 event in each country. The red triangles show the peak ~~rainfall~~ intensity bias between radar and gauges as a function of the aggregation time scale (axis on the right).

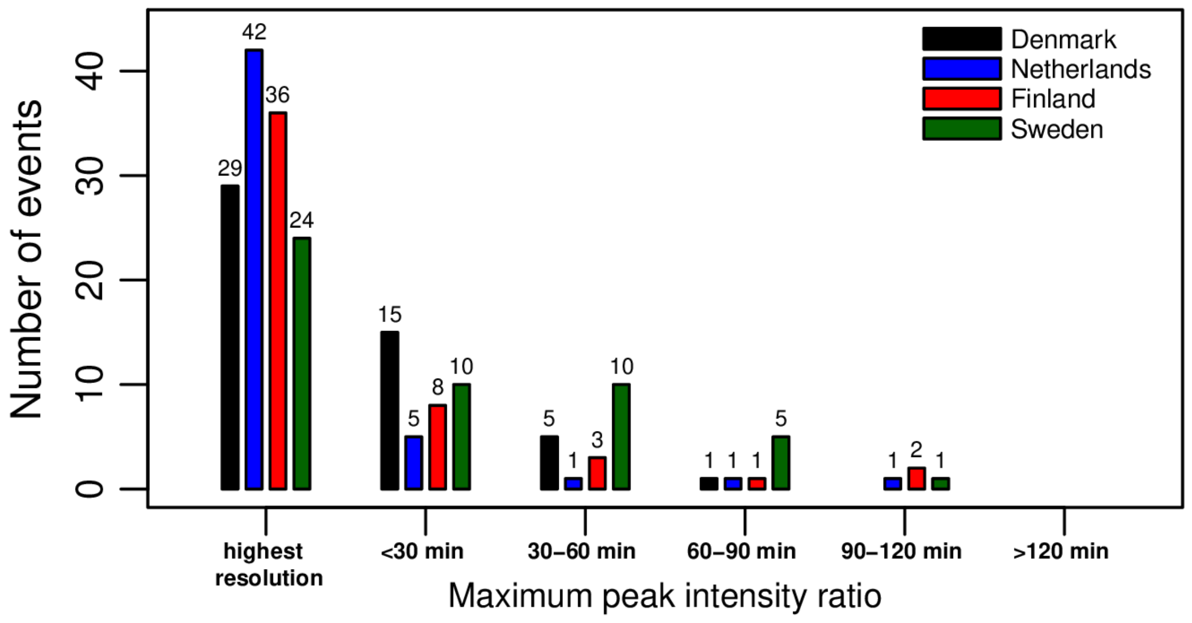

Figure 11. Aggregation time at which the maximum error on peak intensity between gauge and radar occurred.

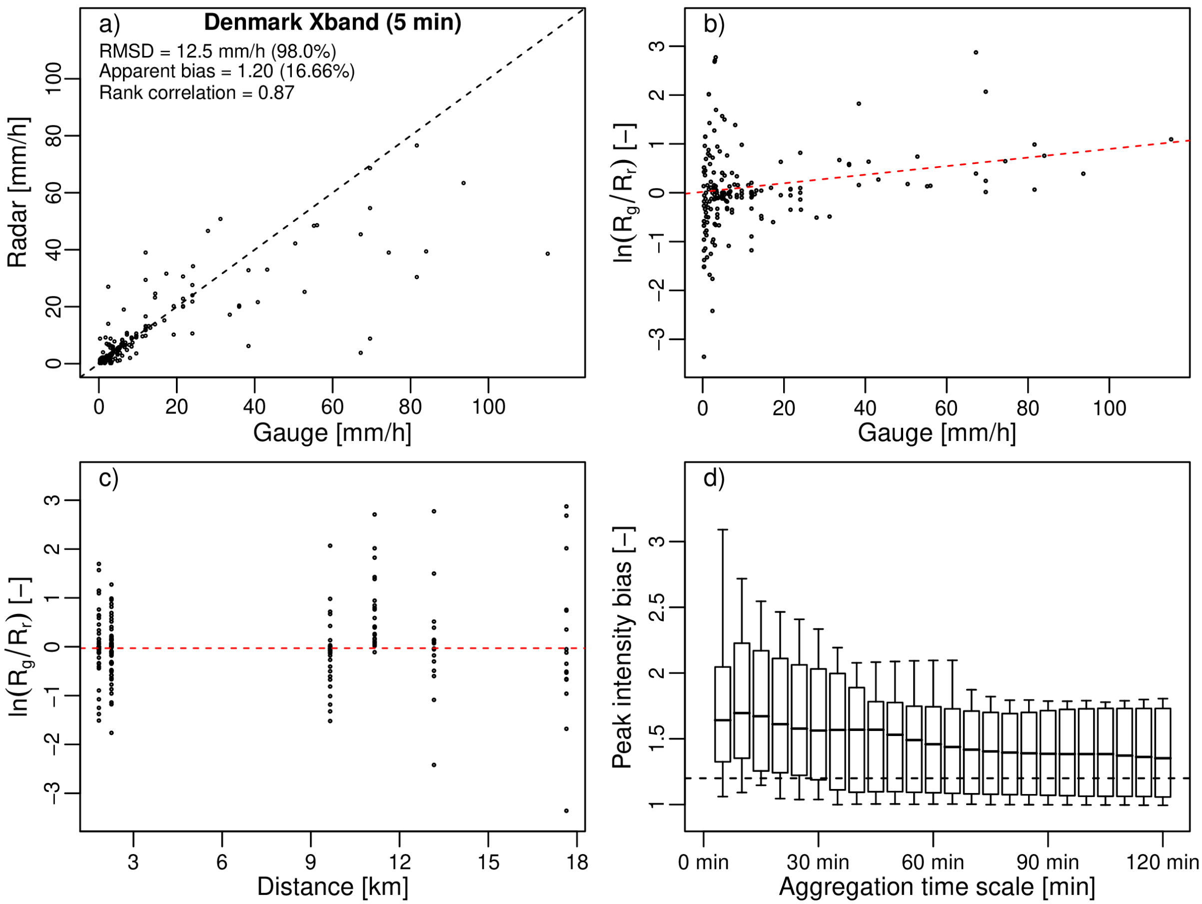

Figure 12. Performance metrics for the Danish X-band radar system (top 10 events).

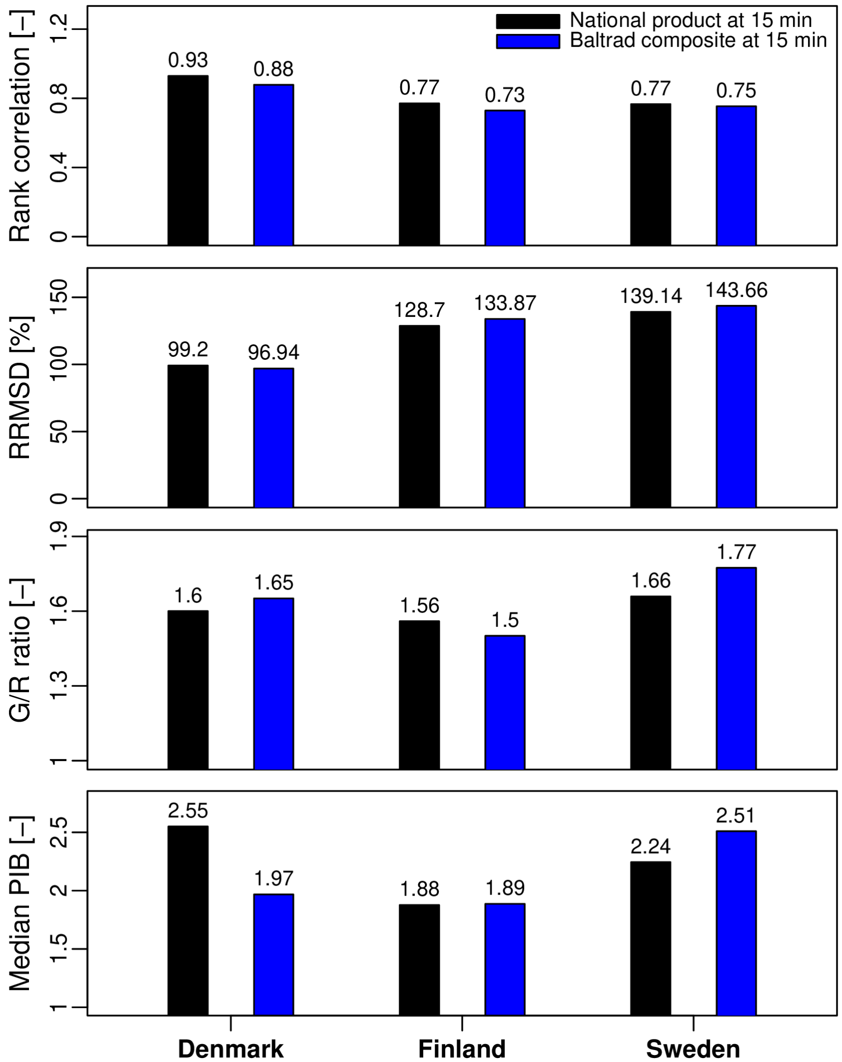

Figure 13. Comparison of relative root mean square error and peak intensity ratios (at 15 min resolution) between the national radar products and the BALTRAD composite.

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

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

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~~ aggregation time scale (~~RRMSE~~all 50 events combined). Average intensity for gauges and radar μ_g and μ_r, ~~multiplicative bias (MB)~~ standard deviations σ_g and ~~area reduction factor (ARF) between a point measurement~~ σ_r, G/R ratio, coefficient of variation, scale parameter σ_ε and ~~a radar pixel (expressed as a percentage)~~"true" underlying model bias β.

Country	~~CC~~ μ_g mmh^{-1}	~~RRMSE~~ μ_r mmh^{-1}	~~MB~~ σ_g mmh^{-1}	~~ARF~~ σ_r mmh^{-1}	G/R [-]	$\frac{CV_g}{CV_r}$ [-]	σ_ε [-]	β [-]
Denmark (500 m, 5 min)	~~0.78~~ 19.8	~~116.4%~~ 12.4	~~37.1%~~ 32.7	~~12.8%~~ 17.6	1.59	1.17	0.93	1.04
Netherlands (1 km, 10 min)	12.1	8.6	23.7	15.5	1.40	1.09	0.89	0.94
Finland (1 km, 10 min)	8.8	5.7	17.2	11.1	1.56	1.00	0.83	~~117.3~~1.11
Sweden (2 km, 15 min)	6.2	3.7	11.4	6.2	1.66	1.11	0.90	1.11

Table 4. Summary statistics for the highest aggregation time scale (all 50 events combined). G/R ratio, G/R ratio corrected for areal reduction factor ARF, model bias β assuming log-normal distribution and relative increase in β with respect to intensity and range.

Country	G/R [-]	G/R corrected for ARF [-]	model bias β [-]	relative increase in β with intensity [(mm/h)$^{-1}$]	relative increase in β with range [km^{-1}]
Denmark (500 m, 5 min)	1.59	1.39	1.04	1.09%	~~29.1%~~ 0.73%
Netherlands (1 km, 10 min)	~~18.6%~~ 1.40	1.14	0.94	0.86%	0
Finland (1 km, 10 min)	~~0.78~~ 1.56	~~128.7%~~ 1.27	~~35.8%~~ 1.11	~~18.6%~~ 0.09%	0
Sweden (2 km, 15 min)	~~0.71~~ 1.66	~~139.1%~~ 1.17	~~39.8%~~ 1.11	~~29.6%~~ 2.12%	0

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~~ 3h	55.4	127.2
21	2007-06-23 09:15	5655	6h5min	38.8	122.4
22	2007-06-23 09:30	5670	~~5h60min~~ 6h	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~~ 2h	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

Table A2. Top 50 events for ~~Finland~~the Netherlands

Event	Starting Time [UTC]	Gauge	Duration	Amount [mm]	Peak [mmh^{-1}]
1	~~2014-07-19 13:50~~ 2014-08-03 17:10	~~101787~~ 380	~~2h30min~~ 6h30min	~~34.7~~ 56.9	~~89.1~~ 180.0
2	~~2014-07-31 09:00~~ 2014-07-28 11:30	~~101103~~ 275	~~1h20min~~ 3h	~~18.1~~ 61.8	~~87.5~~ 139.8
3	~~2014-07-30 15:50~~ 2011-06-28 18:20	~~101289~~ 356	~~19h20min~~ 6h	~~34.8~~ 90.2	~~86.6~~ 136.2
4	~~2014-05-25 16:40~~ 2016-06-23 01:10	~~101555~~ 260	~~29h50min~~ 1h	~~31.6~~ 36.2	~~84.2~~ 121.2
5	~~2014-07-31 11:10~~ 2015-08-30 22:20	~~101690~~ 283	~~2h60min~~ 3h50min	~~51.0~~ 30.2	~~83.9~~ 120.0
6	~~2014-07-18 08:40~~ 2013-08-19 11:20	~~101799~~ 286	~~1h60min~~ 2h10min	~~25.7~~ 29.8	~~83.2~~ 114.0
7	~~2013-08-07 10:10~~ 2015-08-30 19:40	~~100951~~ 356	~~14h60min~~ 6h20min	~~25.9~~ 55.6	~~82.4~~ 112.8
8	~~2014-07-19 09:50~~ 2012-05-20 14:20	~~101194~~ 375	~~50min~~ 4h30min	~~14.6~~ 21.8	~~79.1~~ 109.8
9	~~2014-05-25 09~~2013-07-26 12:50	~~101339~~ 286	~~25h60min~~ 30min	~~48.4~~ 22.0	~~78.6~~ 106.2
10	~~2014-07-31 11:00~~ 2016-09-15 21:20	~~101787~~ 375	~~3h60min~~ 1h30min	~~28.4~~ 18.9	~~78.1~~ 94.2
11	~~2015-07-22 09:00~~ 2011-06-28 19:50	~~101603~~ 273	~~2h30min~~ 11h40min	~~29.4~~ 25.1	~~77.9~~ 93.6
12	~~2014-07-09 14~~2012-08-15 19:40	~~101800~~ 370	~~20min~~ 1h	~~22.1~~ 15.4	~~76.6~~ 92.4
13	~~2014-08-13 21~~2011-08-22 23:40	~~100908~~ 375	~~6h50min~~ 12h	~~28.9~~ 33.4	~~74.2~~ 92.4
14	~~2014-08-09 14:40~~ 2011-08-18 16:30	~~101826~~ 391	~~30min~~ 4h10min	~~16.3~~ 29.4	~~72.8~~ 92.4
15	~~2014-08-11 22:50~~ 2016-06-23 20:20	~~100953~~ 380	~~3h20min~~ 3h30min	~~37.3~~ 27.5	~~71.6~~ 90.6
16	~~2013-08-10 13:50~~ 2015-08-31 14:30	~~100917~~ 270	~~40min~~ 2h20min	~~14.1~~ 32.2	~~69.2~~ 88.2
17	~~2016-07-31 17:20~~ 2009-07-03 14:10	~~101572~~ 391	2h10min	~~21.2~~ 38.0	~~68.3~~ 88.2
18	~~2016-08-06 16:40~~ 2013-08-05 23:00	~~101338~~ 280	~~60min~~ 30min	~~35.2~~ 14.2	~~68.2~~ 84.0
19	~~2016-07-31 09:40~~ 2012-06-21 20:00	~~101555~~ 290	~~11h20min~~ 3h10min	~~27.9~~ 17.2	~~67.5~~ 82.2
20	~~2016-07-03 12:30~~ 2009-07-21 16:50	~~101603~~ 269	~~7h30min~~ 3h	~~67.1~~ 17.2	~~66.9~~ 80.4
21	~~2016-06-30~~ 2016-06-15 10:~~10~~ 50	~~126736~~ 277	~~25h50min~~ 7h30min	~~63.9~~ 34.5	~~66.2~~ 80.4
22	~~2014-08-12 23~~2008-08-07 07:10	~~100955~~ 240	~~7h60min~~ 7h10min	~~20.1~~ 32.9	~~65.6~~ 79.2
23	~~2014-08-11 07:00~~ 2008-07-26 18:10	~~101726~~ 270	~~4h30min~~ 8h10min	~~13.5~~ 26.8	~~65.6~~ 78.6
24	~~2016-07-25~~ 2015-07-05 09:~~00~~ 50	~~101743~~ 270	~~6h20min~~ 6h30min	~~25.9~~ 15.4	~~65.6~~ 78.6
25	~~2014-07-14 11:50~~ 2016-06-23	~~101339~~ 344	~~1h30min~~ 10h10min	~~23.2~~ 32.8	~~65.0~~ 78.6
26	~~2015-08-30 17:10~~ 2014-07-28 02:20	~~100953~~ 257	~~20min~~ 10h20min	~~15.8~~ 71.3	~~65.0~~ 77.4
27	~~2016-07-12 05:10~~ 2009-07-14 12:20	~~101537~~ 286	~~3h10min~~ 3h20min	~~21.4~~ 17.5	~~64.7~~ 77.4
28	~~2014-08-22 12:20~~ 2012-08-05 13:10	~~101805~~ 323	~~1h60min~~ 6h40min	~~16.3~~ 18.5	~~63.6~~ 77.4
29	~~2015-07-08 14:00~~ 2009-05-25 20:50	~~101537~~ 260	~~25h10min~~ 6h30min	~~46.3~~ 23.8	~~62.9~~ 76.8
30	~~2013-06-27 10:20~~ 2012-05-10 14:40	~~101338~~ 375	~~8h30min~~ 3h50min	~~33.2~~ 15.3	~~62.1~~ 76.2
31	~~2014-06-06 13:00~~ 2014-07-10 23:20	~~101690~~ 269	~~6h30min~~ 50min	~~16.7~~ 20.7	~~61.4~~ 75.6
32	~~2013-09-01 06:10~~ 2008-07-06 08:00	~~101272~~ 277	~~9h30min~~ 30min	~~33.0~~ 20.1	~~61.2~~ 75.6
33	~~2016-07-31 06:40~~ 2009-06-09 10:50	~~100974~~ 319	~~3h40min~~ 8h20min	~~21.6~~ 24.8	~~61.0~~ 75.6
34	~~2013-08-15 14:00~~ 2014-07-10 21:10	~~101124~~ 391	~~50min~~ 20min	~~14.0~~ 20.4	~~60.5~~ 75.6
35	~~2014-05-19 18:40~~ 2008-09-11 23:50	~~101537~~ 265	~~4h10min~~ 16h40min	~~21.4~~ 41.8	~~59.6~~ 74.4
36	~~2015-08-08~~ 2011-06-05 16:~~50~~ 10	~~101632~~ 286	~~2h30min~~ 1h30min	~~11.3~~ 19.1	~~58.9~~ 73.8
37	~~2013-08-31 11:30~~ 2015-08-24 15:00	~~100955~~ 269	~~3h20min~~ 3h40min	~~30.0~~ 13.3	~~58.7~~ 70.8
38	~~2016-07-11 14~~2012-05-20 21:30	~~103794~~ 278	~~11h30min~~ 30min	~~14.1~~ 15.8	~~58.4~~ 70.2
39	~~2014-07-14 13:00~~ 2013-07-27 21:40	~~101555~~ 350	2h10min	~~20.2~~ 33.6	~~58.1~~ 70.2
40	~~2016-07-31 06:20~~ 2011-08-03 14:00	~~101632~~ 278	~~6h30min~~ 7h50min	~~16.5~~ 40.8	~~58.1~~ 69.0
41	~~2016-08-04 11:~~2011-08-23 10:40	~~101194~~ 283	~~6h60min~~ 1h30min	~~18.1~~ 16.5	~~58.0~~ 69.0
42	~~2016-07-27 14:50~~ 2008-08-12 23:40	~~101950~~ 257	~~20min~~ 12h20min	~~13.2~~ 23.1	~~57.3~~ 68.4
43	~~2014-08-13 16~~2010-07-14 15:50	~~100967~~ 377	~~3h40min~~ 1h30min	~~12.1~~ 16.7	~~56.8~~ 68.4
44	~~2014-08-11 08:30~~ 2014-07-27 22:00	~~126736~~ 240	~~3h20min~~ 14h20min	~~13.4~~ 53.7	~~56.7~~ 67.8
45	~~2015-07-16 12:20~~ 2009-05-15 05:00	~~101103~~ 273	~~24h30min~~ 16h20min	~~69.5~~ 28.8	~~56.6~~ 67.8
46	~~2016-07-27 04:00~~ 2012-08-04 14:40	~~101805~~ 273	~~5h20min~~ 4h10min	~~16.6~~ 17.5	~~55.5~~ 67.8
47	~~2016-07-14 10:10~~ 2013-07-27 23:50	~~101933~~ 278	~~60min~~ 50min	~~20.4~~ 20.5	~~55.2~~ 67.8
48	~~2014-05-19 13:40~~ 2009-07-03 14:30	~~100967~~ 290	~~20min~~ 4h10min	~~13.3~~ 32.1	~~55.1~~ 66.0
49	~~2014-08-11 23:40~~ 2015-08-14 18:10	~~101603~~ 310	~~12h10min~~ 4h	~~42.4~~ 21.7	~~53.9~~ 66.0
50	~~2013-06-27 11:00~~ 2011-09-06 10:20	~~101150~~ 257	~~5h10min~~ 11h20min	~~19.2~~ 33.1	~~53.2~~ 64.8

Table A3. Top 50 events for ~~the Netherlands~~Finland

Event	Starting Time [UTC]	Gauge	Duration	Amount [mm]	Peak [mmh^{-1}]
1	~~2014-08-03 17:10~~ 2014-07-19 13:50	~~380~~ 101787	~~6h30min~~ 2h30min	~~56.9~~ 34.7	~~180.0~~ 89.1
2	~~2014-07-28 11:30~~ 2014-07-31 09:00	~~275~~ 101103	~~2h60min~~ 1h20min	~~61.8~~ 18.1	~~139.8~~ 87.5
3	~~2011-06-28 18:20~~ 2014-07-30 15:50	~~356~~ 101289	~~5h60min~~ 19h20min	~~90.2~~ 34.8	~~136.2~~ 86.6
4	~~2016-06-23 01:10~~ 2014-05-25 16:40	~~260~~ 101555	~~60min~~ 29h50min	~~36.2~~ 31.6	~~121.2~~ 84.2
5	~~2015-08-30 22:20~~ 2014-07-31 11:10	~~283~~ 101690	~~3h50min~~ 3h00min	~~30.2~~ 51.0	~~120.0~~ 83.9
6	~~2013-08-19 11:20~~ 2014-07-18 08:40	~~286~~ 101799	~~2h10min~~ 2h00min	~~29.8~~ 25.7	~~114.0~~ 83.2
7	~~2015-08-30 19:40~~ 2013-08-07 10:10	~~356~~ 100951	~~6h20min~~ 15h	~~55.6~~ 25.9	~~112.8~~ 82.4
8	~~2012-05-20 14:20~~ 2014-07-19 09:50	~~375~~ 101194	~~4h30min~~ 50min	~~21.8~~ 14.6	~~109.8~~ 79.1
9	~~2013-07-26 12~~2014-05-25 09:50	~~286~~ 101339	~~30min~~ 26h	~~22.0~~ 48.4	~~106.2~~ 78.6
10	~~2016-09-15 21:20~~ 2014-07-31 11:00	~~375~~ 101787	~~1h30min~~ 4h	~~18.9~~ 28.4	~~94.2~~ 78.1
11	~~2011-06-28 19:50~~ 2015-07-22 09:00	~~273~~ 101603	~~11h40min~~ 2h30min	~~25.1~~ 29.4	~~93.6~~ 77.9
12	~~2012-08-15 19~~2014-07-09 14:40	~~370~~ 101800	~~60min~~ 20min	~~15.4~~ 22.1	~~92.4~~ 76.6
13	~~2011-08-22 23~~2014-08-13 21:40	~~375~~ 100908	~~11h60min~~ 6h50min	~~33.4~~ 28.9	~~92.4~~ 74.2
14	~~2011-08-18 16:30~~ 2014-08-09 14:40	~~391~~ 101826	~~4h10min~~ 30min	~~29.4~~ 16.3	~~92.4~~ 72.8
15	~~2016-06-23 20:20~~ 2014-08-11 22:50	~~380~~ 100953	~~3h30min~~ 3h20min	~~27.5~~ 37.3	~~90.6~~ 71.6
16	~~2015-08-31 14:30~~ 2013-08-10 13:50	~~270~~ 100917	~~2h20min~~ 40min	~~32.2~~ 14.1	~~88.2~~ 69.2
17	~~2009-07-03 14:10~~ 2016-07-31 17:20	~~391~~ 101572	2h10min	~~38.0~~ 21.2	~~88.2~~ 68.3
18	~~2013-08-05 23:00~~ 2016-08-06 16:40	~~280~~ 101338	~~30min~~ 1h	~~14.2~~ 35.2	~~84.0~~ 68.2
19	~~2012-06-21 20:00~~ 2016-07-31 09:40	~~290~~ 101555	~~3h10min~~ 11h20min	~~17.2~~ 27.9	~~82.2~~ 67.5
20	~~2009-07-21 16:50~~ 2016-07-03 12:30	~~269~~ 101603	~~2h60min~~ 7h30min	~~17.2~~ 67.1	~~80.4~~ 66.9
21	~~2016-06-15~~ 2016-06-30 10:~~50~~ 10	~~277~~ 126736	~~7h30min~~ 25h50min	~~34.5~~ 63.9	~~80.4~~ 66.2
22	~~2008-08-07 07~~2014-08-12 23:10	~~240~~ 100955	~~7h10min~~ 8h	~~32.9~~ 20.1	~~79.2~~ 65.6
23	~~2008-07-26 18:10~~ 2014-08-11 07:00	~~270~~ 101726	~~8h10min~~ 4h30min	~~26.8~~ 13.5	~~78.6~~ 65.6
24	~~2015-07-05~~ 2016-07-25 09:~~50~~ 00	~~270~~ 101743	~~6h30min~~ 6h20min	~~15.4~~ 25.9	~~78.6~~ 65.6
25	~~2016-06-23~~ 2014-07-14 11:50	~~344~~ 101339	~~10h10min~~ 1h30min	~~32.8~~ 23.2	~~78.6~~ 65.0
26	~~2014-07-28 02:20~~ 2015-08-30 17:10	~~257~~ 100953	~~10h20min~~ 20min	~~71.3~~ 15.8	~~77.4~~ 65.0
27	~~2009-07-14 12:20~~ 2016-07-12 05:10	~~286~~ 101537	~~3h20min~~ 3h10min	~~17.5~~ 21.4	~~77.4~~ 64.7
28	~~2012-08-05 13:10~~ 2014-08-22 12:20	~~323~~ 101805	~~6h40min~~ 2h	~~18.5~~ 16.3	~~77.4~~ 63.6
29	~~2009-05-25 20:50~~ 2015-07-08 14:00	~~260~~ 101537	~~6h30min~~ 25h10min	~~23.8~~ 46.3	~~76.8~~ 62.9
30	~~2012-05-10 14:40~~ 2013-06-27 10:20	~~375~~ 101338	~~3h50min~~ 8h30min	~~15.3~~ 33.2	~~76.2~~ 62.1
31	~~2014-07-10 23:20~~ 2014-06-06 13:00	~~269~~ 101690	~~50min~~ 6h30min	~~20.7~~ 16.7	~~75.6~~ 61.4
32	~~2008-07-06 08:00~~ 2013-09-01 06:10	~~277~~ 101272	~~30min~~ 9h30min	~~20.1~~ 33.0	~~75.6~~ 61.2
33	~~2009-06-09 10:50~~ 2016-07-31 06:40	~~319~~ 100974	~~8h20min~~ 3h40min	~~24.8~~ 21.6	~~75.6~~ 61.0
34	~~2014-07-10 21:10~~ 2013-08-15 14:00	~~391~~ 101124	~~20min~~ 50min	~~20.4~~ 14.0	~~75.6~~ 60.5
35	~~2008-09-11 23:50~~ 2014-05-19 18:40	~~265~~ 101537	~~16h40min~~ 4h10min	~~41.8~~ 21.4	~~74.4~~ 59.6
36	~~2011-06-05~~ 2015-08-08 16:~~10~~ 50	~~286~~ 101632	~~1h30min~~ 2h30min	~~19.1~~ 11.3	~~73.8~~ 58.9
37	~~2015-08-24 15:00~~ 2013-08-31 11:30	~~269~~ 100955	~~3h40min~~ 3h20min	~~13.3~~ 30.0	~~70.8~~ 58.7
38	~~2012-05-20 21~~2016-07-11 14:30	~~278~~ 103794	~~30min~~ 11h30min	~~15.8~~ 14.1	~~70.2~~ 58.4
39	~~2013-07-27 21:40~~ 2014-07-14 13:00	~~350~~ 101555	2h10min	~~33.6~~ 20.2	~~70.2~~ 58.1
40	~~2011-08-03 14:00~~ 2016-07-31 06:20	~~278~~ 101632	~~7h50min~~ 6h30min	~~40.8~~ 16.5	~~69.0~~ 58.1
41	~~2011-08-23 10:40~~ 2016-08-04 11:10	~~283~~ 101194	~~1h30min~~ 7h	~~16.5~~ 18.1	~~69.0~~ 58.0
42	~~2008-08-12 23:40~~ 2016-07-27 14:50	~~257~~ 101950	~~12h20min~~ 20min	~~23.1~~ 13.2	~~68.4~~ 57.3
43	~~2010-07-14 15~~2014-08-13 16:50	~~377~~ 100967	~~1h30min~~ 3h40min	~~16.7~~ 12.1	~~68.4~~ 56.8
44	~~2014-07-27 22:00~~ 2014-08-11 08:30	~~240~~ 126736	~~14h20min~~ 3h20min	~~53.7~~ 13.4	~~67.8~~ 56.7
45	~~2009-05-15 05:00~~ 2015-07-16 12:20	~~273~~ 101103	~~16h20min~~ 24h30min	~~28.8~~ 69.5	~~67.8~~ 56.6
46	~~2012-08-04 14:40~~ 2016-07-27 04:00	~~273~~ 101805	~~4h10min~~ 5h20min	~~17.5~~ 16.6	~~67.8~~ 55.5
47	~~2013-07-27 23:50~~ 2016-07-14 10:10	~~278~~ 101933	~~50min~~ 1h	~~20.5~~ 20.4	~~67.8~~ 55.2
48	~~2009-07-03 14:30~~ 2014-05-19 13:40	~~290~~ 100967	~~4h10min~~ 20min	~~32.1~~ 13.3	~~66.0~~ 55.1
49	~~2015-08-14 18:10~~ 2014-08-11 23:40	~~310~~ 101603	~~3h60min~~ 12h10min	~~21.7~~ 42.4	~~66.0~~ 53.9
50	~~2011-09-06 10:20~~ 2013-06-27 11:00	~~257~~ 101150	~~11h20min~~ 5h10min	~~33.1~~ 19.2	~~64.8~~ 53.2

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~~ 1h	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~~ 3h	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~~ 1h	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~~ 21h	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~~ 3h	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~~ 8h	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~~ 5h	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~~ 4h	24.0	35.6
43	2006-08-03 00:15	89230	~~4h60min~~ 5h	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~~ 1h	14.0	34.8
48	2014-07-28 06:45	89230	1h30min	15.8	34.8
49	2012-06-11 10:15	97280	~~1h60min~~ 2h	16.4	34.8
50	2010-08-09 06:45	76420	~~7h60min~~ 8h	15.0	34.0

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~~ 2h	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