Quantification of the spatial variability of rainfall based on a dense network of rain gauges

Lisbeth Pedersen [a,b,*], Niels Einar Jensen [a], Lasse Engbo Christensen [b], Henrik Madsen [b]

[a] DHI, Gustav Wieds Vej 10, DK-8000 Aarhus, Denmark
[b] Institute for Informatics and Mathematical Modelling, Technical University, Denmark, Building 321, 2800 Lyngby, Denmark

ABSTRACT

The spatial variability of rainfall within a single Local Area Weather Radar (LAWR) pixel of 500×500 m is quantified based on data from two locations. The work was motivated by the need to quantify the variability on this scale in order to provide an estimate of the uncertainty of using a single rain gauge for calibrating the LAWR. A total of nine rain gauges were used, each representing one-ninth of the 500×500 m area. The analysis was carried out based on a dataset obtained using tipping bucket gauges during the summer and fall of 2007 and 2008, and the results were compared with results from an earlier campaign in 2003. The fact that the 2007–2008 dataset was almost four times larger than the original dataset from 2003 motivated this extended study. Two methods were used to describe the variability: the coefficient of variation and the spatial correlation structure of the rainfall field. Despite the small area of 0.25 km^2, accumulated rainfall was found to vary significantly within individual events with durations ranging from 5 min to 13 h. The coefficient of variation was found to range from 1–26% in the 2007–2008 dataset and in some special cases even higher. The 95% prediction interval for a given rainfall depth is estimated and can be used to address the uncertainty of using a single rain gauge to represent the rainfall within a 500×500 m area.

© 2009 Elsevier B.V. All rights reserved.

1. Introduction

Extensive rainfall measurements are required as input for a range of applications ranging from real time online warning systems in hydrology and meteorology to complex models for post analysis of critical events. These types of applications all require a precise and representative measurement of the rainfall. The most common method of precipitation measurement has always been, and still is, the use of gauges. However, gauges cannot provide information on the spatial variability of the rainfall, and this is required for most meteorological and hydrological applications. An obvious solution is to use radar for rainfall measurements, since a single radar can cover a large area, and is able to sample the spatial as well as the temporal properties of rainfall. Weather radars have been used for precipitation measurements since the end of World War II, when a relationship between the radar measurement of the energy backscattered from the hydrometeors in the atmosphere and the rain rate at ground level — the Z–R relationship — was established (Marshall and Palmer, 1948). Despite more than half a century of dedicated research in the field of weather radars and major advancement in the field, radar measurements are still shrouded in distrust — especially by hydrologists. One of the major reasons for this distrust is that the most common approach to evaluating the radar performance is to compare the rainfall estimated by radar with that observed by a single rain gauge. Many consider the rain gauge data to be the "ground truth", despite the fact that it gives no information on the spatial patterns of the rainfall nor does it contain measurement uncertainties. When modeling hydrologic processes in urban areas in particular, the standard assumption that a single gauge is representative of the entire catchment area can lead to large errors, as the individual catchments may have different, and in some cases very rapid, run-off times. Sempere-Torres et al. (1999) demonstrate that even raw radar data provide a better input to a distributed

---

urban drainage model than data from a dense rain gauge network in cases of modeling combined sewer overflows resulting from strong local convective events.

Weather radars operate with pixel sizes from 0.1×0.1 to 2×2 km and provide output every 1–15 min, with 5–10 min being the most common. In reality the output is not a surface measurement, but a volume measurement at a given height, increasing with range. The quality of this measurement is evaluated on the basis of a rain gauge, which is several orders of magnitude smaller and records every time a given volume is collected. In terms of addressing the accuracy of the radar it is, therefore, of great interest to examine the representativeness of the rain gauge within a single radar pixel, since the accuracy of radar precipitation estimation can only be as accurate as a single gauge representativeness of a pixel. Einfalt et al. (2005) show that weather radars provide information on the spatial pattern of the rainfall often missed by rain gauges, and the peak intensities of an event are rarely captured by a gauge network.

Being a mechanical device, the rain gauge also suffers from uncertainties due to hardware and external forces, however, the primary issue is probably the representativeness of the measurement in the spatial domain. The catchment surface of a rain gauge is typically 0.02–0.04 m^2 and for this area the gauge can be assumed to yield a precise measurement, provided it is well maintained and wisely placed. The temporal resolution depends on the gauge type, but most automatic gauges today record every time a given volume of water is collected — often corresponding to a rainfall depth of 0.2 or 0.4 mm.

The spatial domain difference between rain gauges and conventional weather radars is one of the major obstacles to convincing hydrologists that weather radars can provide data, which is equally as good as that of a rain gauge. To bridge this domain span and to act as a supplement and not a substitute for rain gauges or conventional weather radars, a cost-efficient weather radar was developed in 1999 as part of the EP-23475 EU project. This weather radar was to provide sufficient information on rainfall for use in real time online warning systems for flash floods in urban areas. The radar system is called Local Area Weather Radar (LAWR) and was developed and manufactured by DHI. The system is based on a 25 kW marine X-band radar and operates with pixels ranging from 100×100 up to 500×500 m, a temporal resolution of 1 or 5 min and a maximum range of 60 km, of which the inner 20 km range is usable for quantitative precipitation estimation (DHI, 2009). As it is based on X-band technology, the LAWR system is affected by attenuation as outlined in Rahimi et al. (2006) and others, but the LAWR signal processing compensates for this so that extension due to attenuation was only observed in a limited number of situations during the 10 years of operation.

The LAWR is designed as a supplement to standard rain gauges and requires the presence of rain gauges for calibration.

This paper focuses on the spatial variability of the rainfall within a single LAWR pixel and the uncertainties of representing the rainfall of a pixel by a single gauge.

2. Background and past experience

In order to evaluate the uncertainties of calibrating a LAWR with a single rain gauge, a field experiment was conducted south of Aarhus, Denmark during the fall of 2003. Nine high resolution optical rain gauges were evenly distributed within an area of 500×500 m, corresponding to a single pixel from the Aarhus LAWR. The purpose was to test the initial hypothesis: "On an individual event basis, rainfall within an area of 500×500 m is uniform". If a single rain gauge is to be used for calibration of a LAWR, the hypothesis would have to be accepted, otherwise the calibration would vary in the same order as the rainfall within the pixel. An event had to contain a minimum of two gauge registrations less than 60 min apart. The event durations ranged from 5 min to 13 h in the collected dataset. The analysis carried out in this work did not consider different aggregation times, i.e. 5, 10, 15 or 60 min, because the original focal point of the experiment was related to LAWR calibration. However, plans are being made to consider these differences in future analyses.

Surprisingly, the results from the experiment showed high variability in the rainfall depths observed by the nine gauges within independent events. Expressed as coefficient of variation (CV), the variability ranged from 10%–100%, and even if the most extreme of the gauges were omitted from the analysis, the variation was more than 50% in several events (Jensen and Pedersen, 2005).

If the order of variability based on the 2003 dataset is representative, parts of the uncertainties related to radar rainfall estimation are not due to the radar, but are a result of the variability of rainfall influencing the calibration on a very small scale. The properties of rainfall on scales equal to or less than a single pixel are still more or less unknown, whether the pixel be 0.1×0.1 km or 2×2 km. Rainfall within these scales is central for the evaluation of radar performance as well as a range of hydrological applications, in particular urban hydrological applications.

Inter-gauge distances in operational gauge networks used for radar estimation evaluation and for hydrological modeling are often more than 20–30 km, and this yields no information on the spatial structure at small scales (Ciach and Krajewski, 2006). The definition of small-scale and network density varies from source to source, since small-scale density is defined using inter-gauge distances ranging from 0.1 km to more than 10 km (Ciach and Krajewski, 2006, Krajewski et al., 2003) and sources herein. There are many experimental gauge networks around the world, some of which are relatively dense, e.g. EVAC PicoNet (Oklahoma, USA) with 25 gauges spaced 0.6 km apart covering an area of 9 km^2 (Ciach and Krajewski, 2006). Networks with such gauge density are unfortunately quite rare since they require a high number of gauges, frequent maintenance and calibration, as well as time dedicated to data quality control — all of which are expensive.

On the basis of five gauge networks of different densities and in different climate regimes Krajewski et al. (2003) conclude that the small-scale variability of rainfall is significant, but state that their dataset is insufficient for final conclusions. The variability found by Krajewski et al. (2003) is not as large as that found in the rain gauge experiment of 2003. There may be several reasons for this discrepancy, such as local orographic effects, climate zones, rainfall types, gauge types and gauge errors. The dataset from 2003 is limited to three months of sampling during the fall months of September to November (Denmark). During the experiment it was observed that the optical drop-counting gauges vibrated due to the wind, and low temperature caused ice to freeze in the

Table 1
Overview of gauge experiment. The data analyzed are from the years 2003, 2007 and 2008 since the gauges were damaged during the 2006 season. The different gauge types and locations along with calibration information are listed for each year.

	Measuring time frame	Gauge type	Location	Calibrated	In-situ calibration
2003	09.19.2003 11.25.2003	0.01 mm Optical drop-counting gauges (Rosted DigiRain)	Open field	Yes	No
2006	No data	0.2 mm Tipping bucket gauge (Pronamic)	Open field	Yes	No
2007	09.13.2007 11.07.2007	0.2 mm Tipping bucket gauge (Pronamic)	Estuary	Yes	No
2008	06.17.2008 11.13.2008	0.2 mm Tipping bucket gauge (Pronamic)	Estuary	Yes	Yes

funnel of the gauge, causing uncertainties in some events. The large variation in the 2003 dataset gave reason for concern, and a larger dataset is required to reach a more final conclusion. Therefore, the experiment was recommenced, with minor alternations, as part of a 3-year field experiment in 2006.

This paper presents the results from the 3-year measuring campaign conducted from 2006 to 2008 with nine gauges equally spaced within a 500×500 m area. In addition the dataset from 2003 is included for comparison. The aim of the experiment was to validate the results from 2003 by increasing the data foundation and carrying out a more in-depth statistical analysis of the spatial variability. The paper is a further development of the work presented in Pedersen et al. (2008) and it presents new data and methods.

3. Rain gauge setup 2006–2008

In order to validate previous results, the gauge experiment was recommenced in the spring of 2006. The gauges used in the original setup were optical drop-counting gauges with a resolution of 0.01 mm manufactured by Rosted DigiRain. The uncertainties observed during the experiment in 2003 had been confirmed by other internal DHI setups using this gauge type. As a result of this, and because of a desire for more

comparable results with other similar studies. a standard 0.2 mm tipping bucket gauge from Pronamic was chosen for the 2006–2008 period. The gauges were all equipped with a data logger for data storage. For an overview of the different setups see Table 1 and for illustrations see Fig. 1.

In 2006 the gauges were installed at the same location as in 2003 (on land as illustrated in Fig. 2), but unfortunately they were more or less all destroyed by farming equipment as a result of miscommunication with the landowners. After this the landowners would only allow the gauges to be placed on their property from November to February, which in Denmark are winter months dominated by widespread frontal rain with occasional sleet and snow showers. Rainfall events with high intensities and large variability often occur during the summer and early fall, and a new location was therefore needed. The new location was selected based on the following requirements: undisturbed location, homogeneous conditions without terrain differences, and less than 15 km from the Aarhus LAWR. It proved hard to find such a location on land, and therefore the gauges were moved approximately 1 km north-east out into the shallow estuary "Norsminde Fjord". The layout of the experiment is illustrated in Fig. 2 and the inter-gauge distances are listed in Table 2.

The gauges were mounted on poles pounded down into the seabed of the estuary and because it was problematic doing this with a sledge hammer while standing up in a rubber dinghy, not all the poles were level. To alleviate this problem all the gauges were installed with an adjustable fitting, allowing for adjustment of the gauge on site. During the fall of 2007 it was observed that the poles holding the gauges were affected by the current, which may have caused a reduction or increase in the effective volume of the gauge bucket, due to tilting of the gauge bucket. If the gauge installation is tilted along the longitudinal axis of the bucket; the bucket will be tilted upwards or downwards. As result it will tip at a different water volume since the point of gravity has been shifted. The gauge bucket is shown on Fig. 2, marked as B2. As a result of this the gauges were calibrated in-situ throughout the measuring period in 2008 in order to establish the tipping volume as described in Section 3.1.

3.1. Calibration of gauges

All gauges were calibrated prior to deployment each year. A known water volume was poured into the gauge at
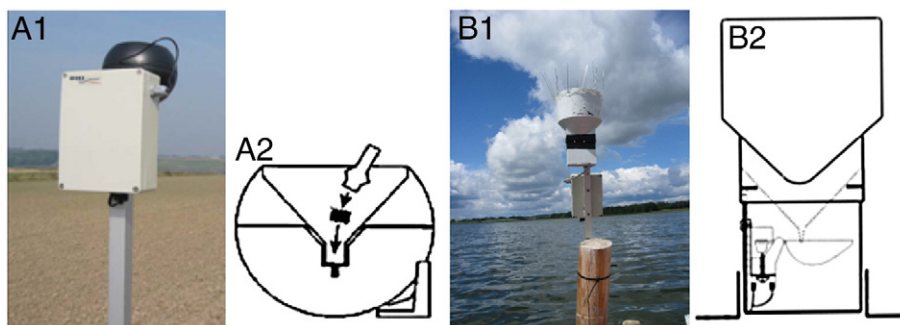

Fig. 1. Gauge types used in field experiment. Left: the optical drop-counting gauge (marked A) based on a Rosted DigiRain gauge (2003) and a sketch of the gauge showing the sponge and funnel. Right: the tipping bucket gauge (marked B) by Pronamic (2006–2008) along with a sketch of the gauge illustrating the location of the tipping bucket. In 2008 the gauges where furthermore equipped with metal spikes on the rim of the gauge to avoid resting birds.
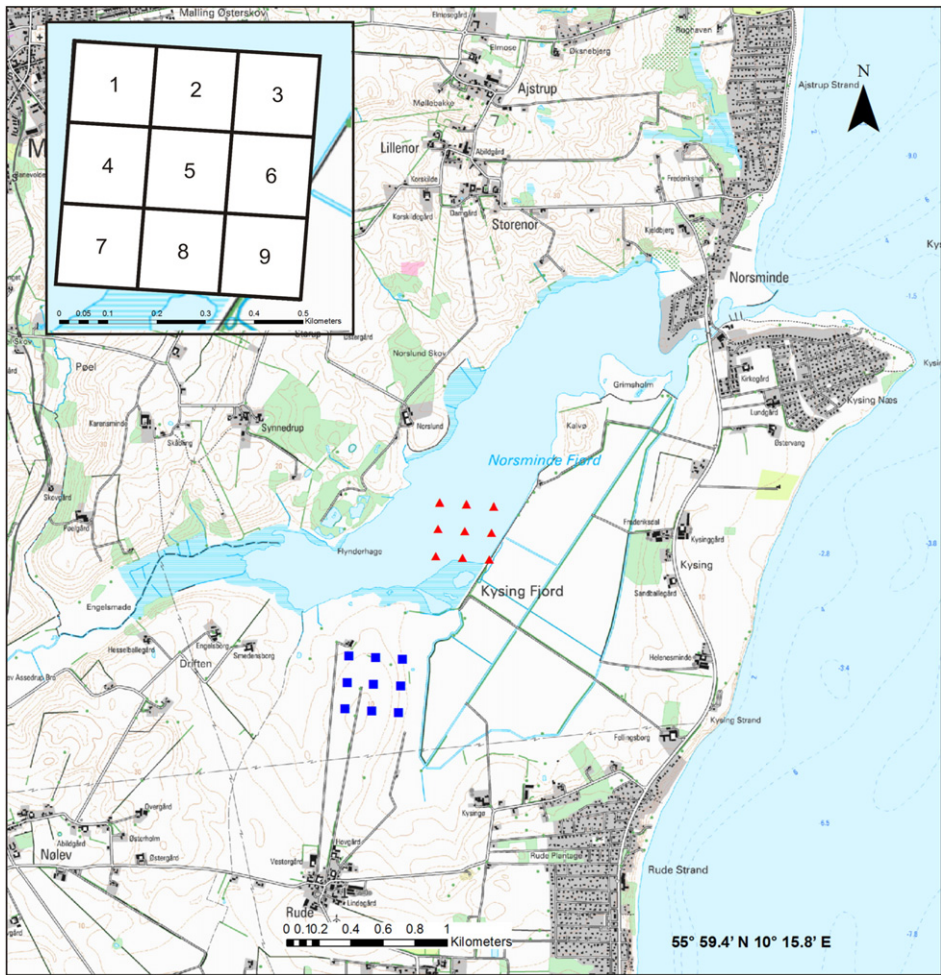
Fig. 2. Location of the gauge sites. The 2003 and 2006 sites are on land marked with squares, while the 2007 and 2008 sites are in the estuary (Norsminde Fjord) marked with triangles. Upper right corner: a close-up of the experiment layout and area represented by each gauge. The tilted orientation of the gauge grid corresponds to the radar grid of the Aarhus LAWR in 2006, which is located 10 km north of the gauge sites.

different speeds to simulate a range of intensities. The observed number of times the buckets tipped was compared with the expected number. The dataset from the optical drop-counting gauges was supplemented with a gauge-specific calibration factor based on the calibration prior to deployment. Unfortunately, the gauges were not recalibrated when dismantled. This would have provided information on any drifting of the gauge.

The calibration prior to installation in 2006 revealed that six of the gauge buckets did not tip at 0.2 mm, but at different volumes corresponding to rainfall depths ranging from 0.1–0.4 mm. All gauge buckets differing more than 2% in the

Table 2
Inter-gauge distances in field experiment.

Possible inter-gauge distances [m]	
I (e.g. 1–2)	167
II (e.g. 1–3)	333
III (e.g. 1–5)	236
IV (e.g. 1–6)	373
V (e.g. 1–9)	472

calibration were replaced and the replacements were tested and accepted. During the calibration it was observed that this type of rain gauge is very sensitive to misalignment, since a slight tilting of the bucket around the longitudinal axis altered the effective tipping volume. This resulted in either an overestimation or an underestimation. If the degree of tilting is unknown or varying, this may corrupt data. Since the gauges were placed on a pole in an unstable environment, an in-situ calibration method was developed to facilitate this problem.

The challenge was to obtain a known volume of water and empty it down into the gauge at a controllable continuous speed in a manner reproducible with a maximum 2% error margin at sea in a small dinghy. The method chosen used a 270 ml plastic wash bottle with a swan neck. The bottle was totally submerged under the water and filled to the brim (no air bubbles), then the spout was screwed on (still submerged) and the bottle was placed upside down in the gauge. The method was tested in the lab in order to determine the degree of error caused by excess water on the bottle and hands, the ability to fill the bottle with exactly the same amount of water each time, and the drip rate of the spout. A total of 28

experiments resulted in a mean result of 69.2 tips with a standard deviation of 0.74 tips. A one sided *t*-test gives 95% confidence bounds of 68.9 and 69.5 and it is therefore concluded that this method can be reproduced within $+/-1$ tip, thus fulfilling the 2% error margin requirement. The water used for the on-site calibration was brackish with a density of approximately 1.02 mg/l and the volume of the bottle would result in 69 tips if the gauge was level. At every visit to the gauge site each gauge was calibrated in-situ. The average number of recorded tips for each gauge per visit was interpolated to obtain a daily calibration factor reflecting changes over time, e.g. drift of the pole. Due to practical time limitations, only two in-situ calibrations were applied at each visit, which is less than desirable for determining the calibration factor. Therefore a weight was used in the estimation of the calibration factor. A factor of 1/3 was assigned to each of the adjacent visits whereas the actual visit was assigned 3/5 and the actual value was the weighted average of the three visits. The calibration factor as a function of time is shown in Fig. 3 for two selected gauges — gauges 3 and 7.

The calibration factor of gauge 7 increased significantly in the final two months, but the factors were consistent with an equal number of registrations on the respective days, so the increase was probably caused by the pole drifting or the adjustable fitting becoming dislodged.

3.2. Data quality control

As with any other field experiments, practical problems can cause data corruption rain gauge clogging due to debris collected in the orifice which was probably one of the most frequent problems. In this experiment a whole range of debris types caused problems. On one occasion, a dead bee was found clogging the funnel and partly dissolved into the gauge bucket, decreasing the effective volume. However, the biggest practical problem was bird droppings resulting in clogging and reduced bucket volume. This was not observed to be a problem when the gauges where situated on land in 2003, but when they were moved to the estuary for the first period of sampling in 2007, the gauges were seriously affected. The reason for the difference is probably a different type of funnel, cf. Fig. 1 and the change in location. The gauges were equipped with a crown of netting, but it became evident that the material used was too flexible because birds rested on the gauges anyway. Next the gauges were equipped with pigeon

spikes glued to the rim as shown in Fig. 1, and this turned out to work well. Unfortunately, this meant that there were few events with nine functioning rain gauges; most events had somewhere between three and nine gauges in working order.

Rain gauge measurements are known to be affected by wind, and a number of correction schemes were developed to correct for the wind-induced loss. Studies carried out under the World Meteorological Organization (WMO) report of wind-induced losses in the range of 4–6% (Sevruk et al., 2009). It was not possible to obtain reliable wind correction parameters for the site in question as neither wind nor temperature data were available. The error was due to potential underestimation of the rainfall since the site was unsheltered (open waters). The underestimation would most likely be relative as all gauges were equally due to the homogeneous site conditions.

Another source of error causing data corruption was lightning. During the summer of 2008 several intense thunderstorms occurred over the site. On at least one occasion six out of the nine gauges were struck by lightning, causing them to stop recording. The data from this event was split into two events in order to use both events in the analysis, since it was a large event with high intensities and large rainfall depths. The time of interruption occurred within a minute on all the affected gauges and the data from the working gauges was cut off at that time, cf. Fig. 4. The full time series from gauges 4, 6, and 9, unaffected by the lightning is illustrated in Fig. 5.

In some cases short circuits in the electric system were observed. Whether this was a result of a previous lightning incident or malfunctioning hardware was not established, but the data affected was omitted from the analysis when detected.

A substantial amount of effort and time was invested in scrutinizing the dataset prior to the analysis. Despite this there are cases where it was not possible to determine whether suspicious data was due to a gauge error or to natural variability in the rainfall. This problem could have been alleviated by using a setup with multiple gauges at each point. On the basis of their observations from various field experiments, Krajewski et al. (2003) recommend a double gauge setup, since they observed difficulties in identifying error-affected data if only a single gauge was available. This was not possible due to financial and practical reasons. The double gauge setup would furthermore have been useful since data were often missing due to malfunctioning gauges as a consequence of the problems accounted for above.
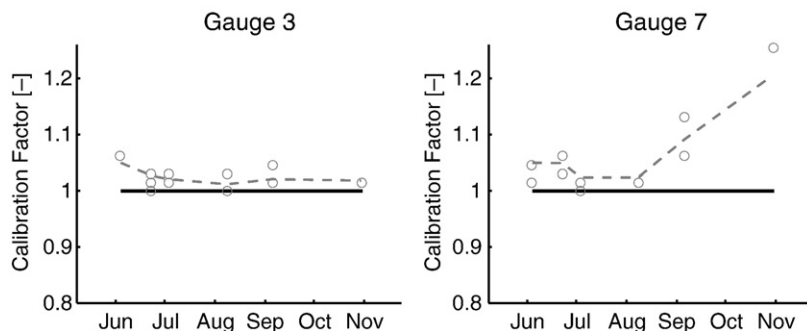

Fig. 3. Calibration coefficients as a function of time for two of the nine gauges — gauges 3 and 7. Circles mark the date of calibration. Gauge 3 is an example of a very stable gauge, while gauge 7 changed properties during the experiment. The rest of the calibration profiles lay between the two shown.
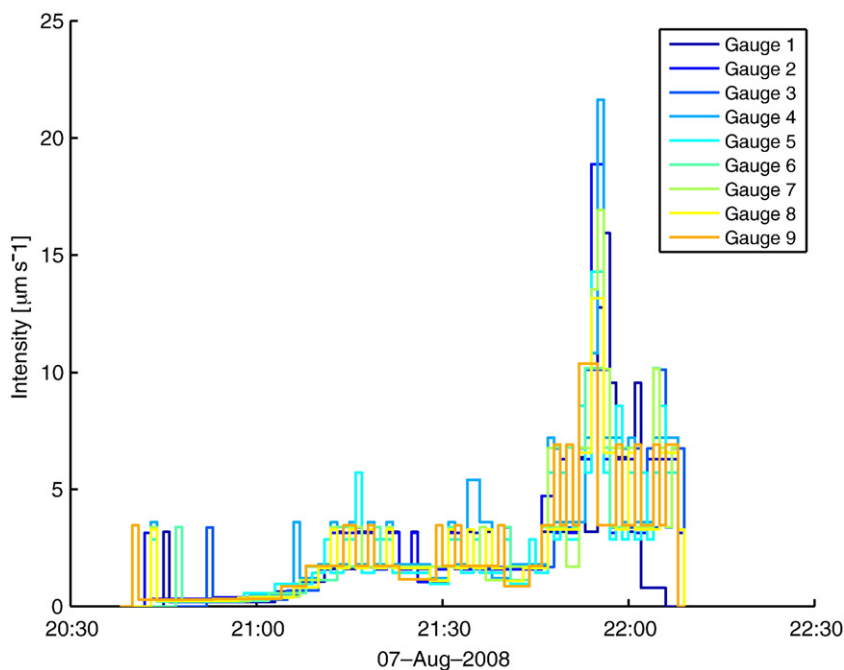
Fig. 4. First part of intense rain event on 7 August 2008 with all nine gauges working. The gauges recorded between 11 and 14 mm of rainfall over a period of 76–87 min (1.4 h) until gauges 1, 2, 3, 5, 7 and 8 were struck by lightning.

4. Data characteristics

The scrutinized data was subdivided into events on the basis of the criteria applied to the Danish Water Pollution Control Committee network of rain gauges in Denmark operated by the Danish Meteorological Institute (Thomsen, 2006). An event was to consist of at least two registrations and the time span between these registrations was to be less than 60 min (Thomsen, 2006). An additional requirement of a minimum rainfall depth of 1 mm was also applied to the data in this analysis.
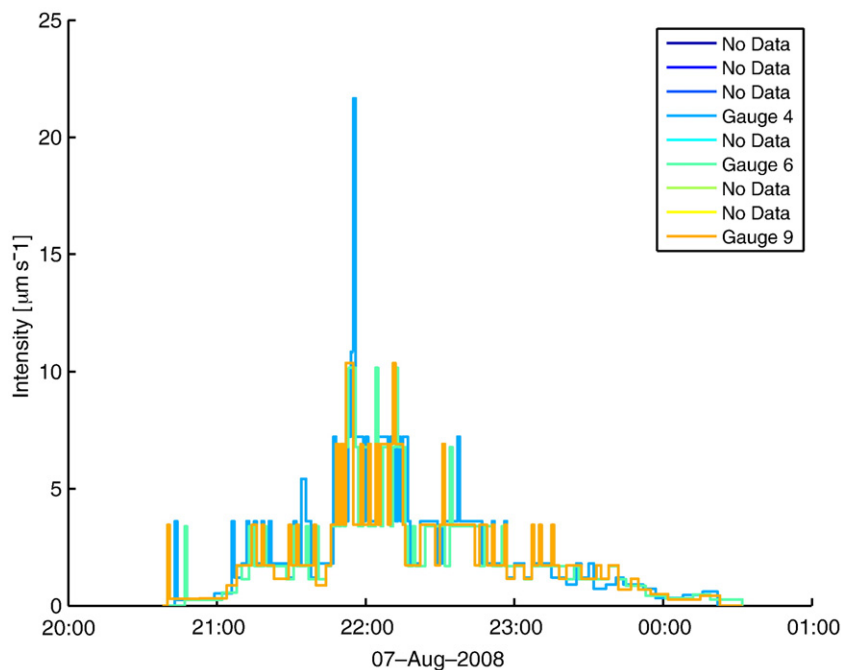

Fig. 5. Full extent of rain event on 7 August 2008 based on gauges 4, 6 and 9, which survived the lightning. The gauges recorded between 38 and 30 mm of rainfall during a 218–224 minute (~3.7 h) interval.

The characteristics of the dataset are listed in Table 3. Unfortunately, all three years had periods, in which one or multiple gauges malfunctioned, leaving ten events in 2003 with all nine gauges functioning, five in 2007 and eight in 2008. The lower half of Table 3 gives the characteristics of events with a minimum of three functioning gauges. By limiting the number of required gauges the dataset becomes significantly larger. Out of the ten events with nine working gauges in 2003, only three events have rain depths larger than 5 mm; there are none in 2007 and five in 2008.

2008 had a large number of events and a very high total rainfall depth of 222 mm, which is due to a large number of intensive convective events in August. The national average precipitation statistics for August 2008 confirm this with a recorded average regional rainfall depth of 146 mm compared to the normal average rainfall of 67 mm for the month of August (DMI, 2008).

The aim of this work is to quantify the variability of rainfall on very small scales by means of descriptive statistical methods focusing on the estimation of the spatial correlation structure. Correlation is known to be an adequate measure of the linear relation between two variables, provided that the data is normally distributed. In hydrology and hydrometeorology correlation is widely used on rainfall data assuming they are normally distributed. This issue motivated Habib et al. (2001) to investigate the matter further, and they state that, among statisticians it is a well known fact that skewed data results in biased correlation results, but this is rarely taken into account in the communities using rainfall data in operational applications. Rainfall data on smaller spatial scales ranging from a few meters to 8 km are typical scales for radar products used in hydrological applications (Habib et al., 2001). On such local scales rainfall data can be considerably skewed and rainfall data should be log-transformed (Habib et al., 2001). The data in this study was therefore first tested to establish which type of distribution it followed.

Due to the very limited number of events where all nine gauges were recording properly the data analysis was carried out on the data-subset corresponding to the lower part of Table 3. The rainfall events used in the analysis contain data from three to nine gauges.

If the data was normally distributed, the variance would be independent of the mean, but, as Fig. 6 clearly illustrates, this was not the case. There is a clear tendency of increasing

Table 3

Characteristics of the data samples from the three seasons of operation − 2003, 2007 and 2008.

	2003	2007	2008
Sample size [events >1 mm by 9 gauges]	10	5	8
Average total rainfall depth (mean of 9 gauges) [mm]	53	14	69
Range rainfall intensity [mm/h]	0.1–2.5	0.5–9.1	1.2–14.3
Mean rainfall intensity (mean of 9 gauges) [mm/h]	0.8	1.9	5.4
Sample size [events >1 mm by min. 3 gauges]	20	19	55
Average total rainfall (mean of 3–9 gauges) [mm]	89	71	222
Range rainfall intensity [mm/h]	0.1–3.4	0.5–11.5	0.3–32.5
Mean rainfall intensity (mean of 3–9 gauges) [mm/h]	0.9	2.4	3.6

variance with increasing mean rainfall depth. The variance is lower for 2007 and 2008 than for 2003, where in some cases the variance is of the same order as the mean value. Common for all three datasets is an indication of a downward trend in variance with an increasing mean rainfall depth. As mentioned earlier, the dataset from 2003 is based on a different type of gauge at a different location and, as illustrated in Fig. 6, this set has a higher variability than those for 2007 and 2008.

Outliers in the 2003, 2007 and 2008 datasets are marked with a square on the graph. These events were examined in detail and were considered dubious, so they were omitted from the datasets.

After log-transforming (natural logarithm) the data, the variance dependency of the mean rainfall depth is not as strong, cf. Fig. 7.

The Anderson–Darling test was chosen for testing the hypothesis of the normality of the log-transformed rainfall data. The p-values obtained from the Anderson–Darling test for each gauge in 2003, 2007 and 2008 are listed in Table 4. There is no evidence to reject the null hypothesis of normality. The Anderson–Darling test requires a minimum of seven samples, and since only five events were observed at gauge 1 in the 2007 dataset, the Lilliefors test is used to test for normality. The strength of this p-value is less than that of the others, partly due to the low number of observations and partly due to the test type. The dataset is hereafter considered as being log-normally distributed.

5. Results

By adding the 2008 dataset using the same gauges and location as in 2007, the basis for the conclusions improved significantly compared to a prior work reported in Jensen and Pedersen (2005) and Pedersen et al.(2008). The 2008 dataset is larger than the 2003 and 2007 datasets combined. The data from 2003 and 2007 was scrutinized more extensively and more data was omitted than in the prior work due to the information gained in the process.

To identify gauges with systematic errors, the non-parametric Kruskal–Wallis test was used to determine whether the median of the observed rainfall depths (log transformed) for each year was constant across the nine gauges. If the median differs significantly, one or more gauge(s) may have systematic errors assuming that the accumulated rainfall within such a small area is uniform over a period of several months. The null hypothesis of equal population medians is accepted with p-values of 0.64 (2003), 0.98 (2007) and 0.95 (2008) and no gauge in any of the three datasets has been identified as containing systematic errors.

The distribution of rainfall depths observed for the three years is shown in Fig. 8. For all three observation years the events with rainfall depths of less than 5 mm dominate the datasets. Only a few events exceeding 15 mm are present in the 2008 datasets and there are none in 2003 and 2007.

When the variability issue was addressed based on 2003 and 2007 data in Pedersen et al. (2008) the smaller variability in the 2007 dataset was attributed to the limited dataset of 2007 and the lack of large events. The rainfall event depths of 2003 vary much more than those observed in 2007 and 2008, cf. Fig. 7, in which the variance is shown as a function of the
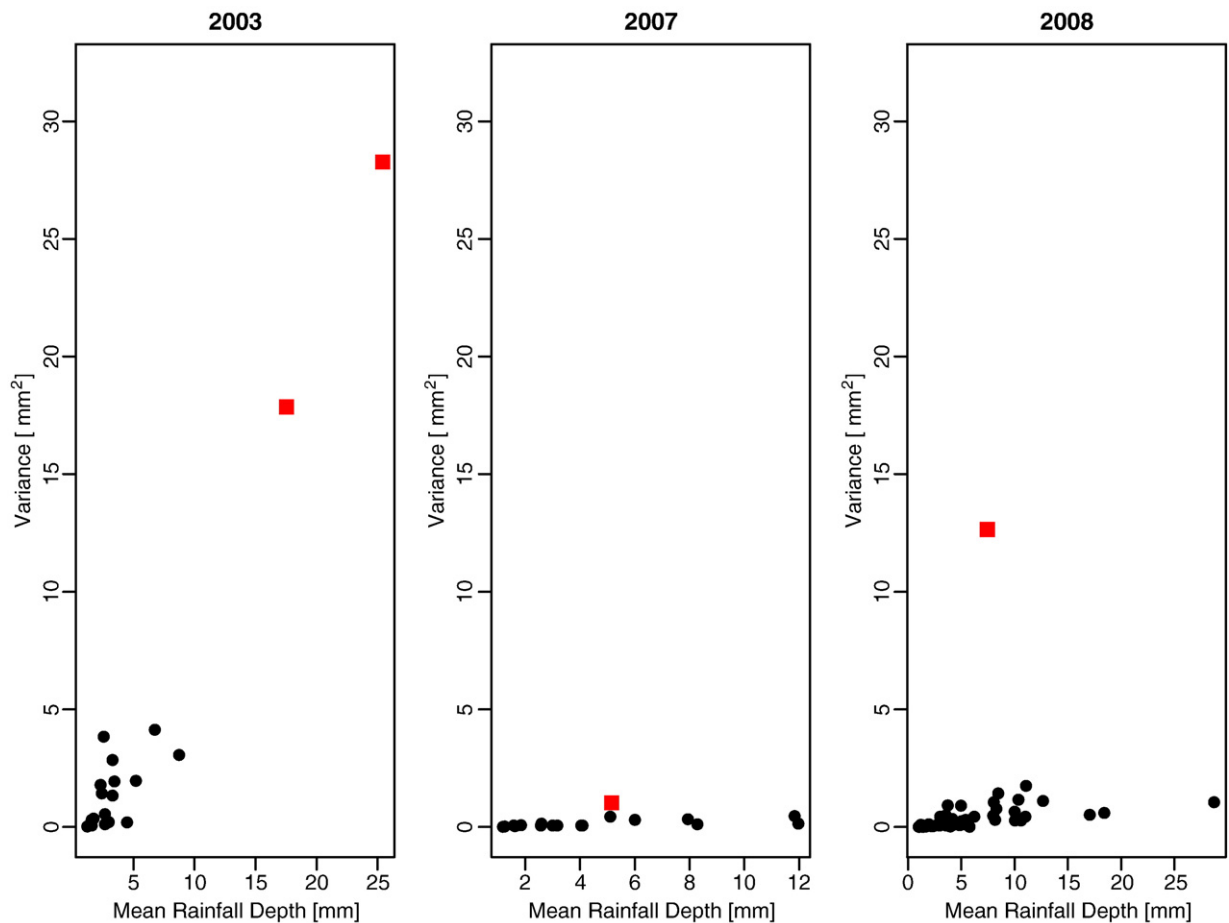
Fig. 6. Variance of the rainfall depth as a function of the mean rainfall depth for each event; note — one graph per year. The events marked with a square are omitted from the dataset. The 2003 data was collected by optical drop-counting gauges, while the 2007 and 2008 data were obtained using 0.2 resolution tipping bucket gauges, cf. Section 3.

mean depth (log-transformed). The dilemma is that the 2003 dataset was obtained using a different type of gauge and from a different location, and it is therefore difficult to directly compare the datasets, despite the identical layout. The variability of the 2008 dataset is larger than that of 2007 using the same location and gauge type, cf. Fig. 6. The 2008 dataset also includes events with large rainfall depths of the same order as those observed in 2003. The large variation in the 2003 dataset cannot be recovered in the 2007 and 2008 datasets. There could be several reasons for this. The mean rainfall intensity ranges from 0.1–5.4 mm/h for 2003 to 0.5–11.5 mm/h for 2007 and 0.3–32.5 mm/h for 2008, indicating that the large
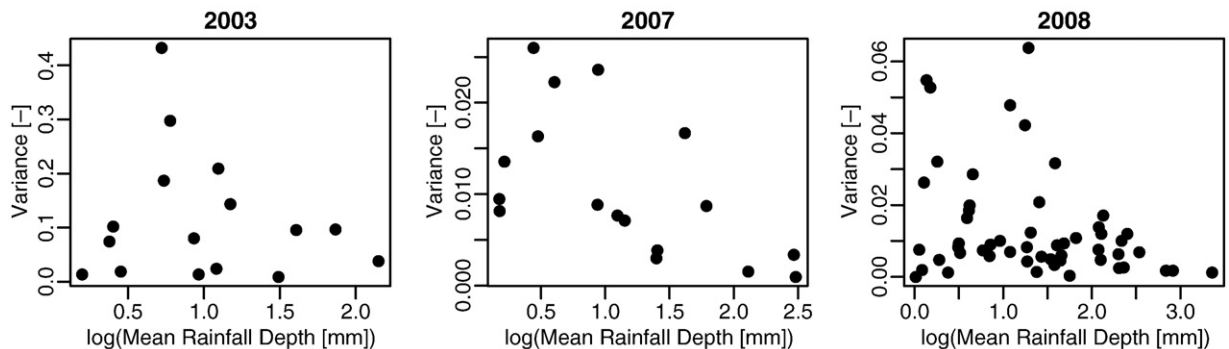

Fig. 7. Variance of the rainfall depth as a function of the mean rainfall depth for each event-based on natural logarithmic transformed data. The 2003 data was collected by optical drop-counting gauges, while the 2007 and 2008 data were obtained using 0.2 resolution tipping bucket gauges, cf. Section 3.

Table 4
P-values from the Anderson–Darling test of normality. *P-value from Lilliefors test due to <7 observations.

	Gauge 1	Gauge 2	Gauge 3	Gauge 4	Gauge 5	Gauge 6	Gauge 7	Gauge 8	Gauge 9
2003	0.13	0.20	0.63	0.38	0.20	0.34	0.14	0.78	0.26
2007	0.06*	0.38	0.32	0.19	0.61	0.58	0.59	0.51	0.59
2008	0.28	0.12	0.09	0.62	0.26	0.34	0.06	0.11	0.44

rainfall depths occurred over a much longer period of time in 2003 than in 2008. The small mean event intensities in 2003 indicate that the large events are not of convective character, but are more frontal, whereas the very high mean event intensities observed in 2008 are of convective character.

5.1. Spatial variability

The objective is to describe the variability of rainfall depths on the basis of events in order to quantify the uncertainty related to using data from a single gauge as input to a hydrological application or to calibrate a radar as defined in Section 4. The first approach is to analyze the variability expressed by the spatial coefficient of variation (CV) of the precipitation field, estimated as the ratio of the standard deviation to the arithmetic mean depth. The estimated CVs are shown in Fig. 9 as a function of mean event depths, and in Fig. 10 as a function of mean event intensities.

It is evident that the spatial variability is notably larger in 2003 than in 2007 and 2008. Fig. 9 indicates correlation between rainfall depths and spatial coefficient of variation, since the CV is decreasing with increasing rainfall depths. The large variability represented with CV above 50% occurs only at depths of less than 5 mm and only in the 2003 dataset.

Generally, two types of rainfall occur in Denmark — widespread frontal rain with long duration and relatively small rainfall volumes, and convective events of short duration, but often releasing large volumes of water. As stated earlier the 2003 dataset has only low mean event intensities in contrast to 2008 and as shown in Fig. 10. Events with a high CV value are events with small event intensity and a low mean event depth.

Based on the CV values for the three datasets, the variability ranges from 1–77%, with a mean value of 14% variability. The CV

values in the 2003 dataset are significantly larger compared to the 2008 dataset, and the most likely cause for this discrepancy is the gauge type. Past experience from other projects with optical drop-counting gauges confirms the suspected instability of this gauge type. Optical drop-counting gauges have been observed to be unstable, e.g. recording continuously in known cases with no precipitation. If the variability is estimated based on 2007 and 2008 data, it ranges from 1–26% with a mean of 10%.

A coefficient of variation of up to 26% is lower than previously reported, but nevertheless it is still quite large because the area of interest is only 0.25 km² and the inter-gauge distances, cf. Table 2, are very short. The upper value of 26% is based on the dataset in which the extreme event of 2008 has been omitted, cf. Fig. 6. If the omitted event is taken into account, the range of the 2007 and 2008 dataset becomes 1–47%.

The event causing the rise in CV values has been omitted due to the extremely high variance, but there is a potential risk of omitting something that could be true. The rainfall depths of this event vary from 12.4 to 2.5 mm, with the largest depths in the north-west corner of the field, cf. Fig. 11.

Unfortunately, gauge 1 malfunctioned during this event, and without this gauge it is very difficult to determine whether this rain event only covered part of the gauge area, or whether there are errors on these gauges. The gauges agree on the duration of the event, which suggests it is valid; furthermore the pattern of the rainfall depths is realistic, with a gradient over the field and no abrupt jumps. A single event is not enough to draw any conclusions, but with rainfall depths potentially varying by up to 47% within an individual event, this may result in uncertainties of this order in output from applications assuming that rainfall is uniform within 500×500 m. The effect is a potential difference of factor five
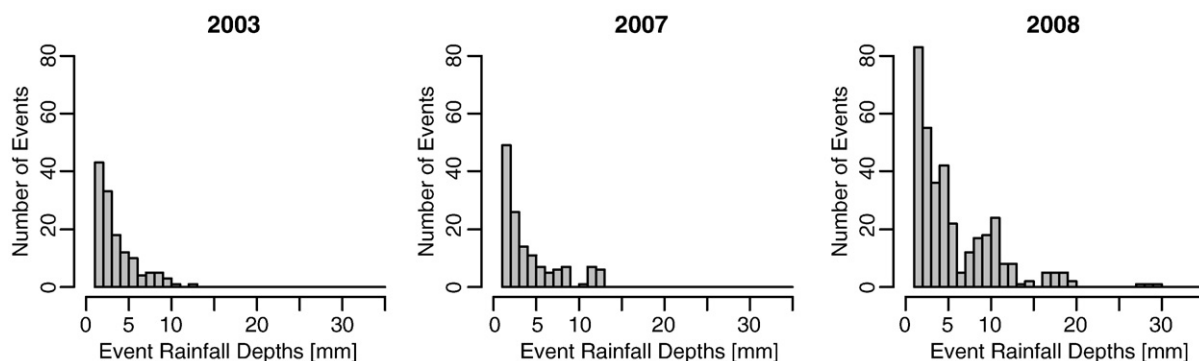

Fig. 8. Histogram of rainfall event depths in mm for all functioning gauges. There is an individual plot for each year.
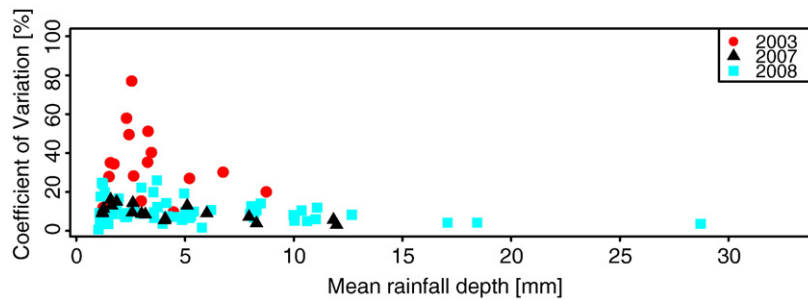
Fig. 9. Coefficient of variation (CV) as a function of the event-based mean rainfall depth in mm for all three datasets. 2003 (circles), 2007 (triangles) and 2008 (squares).

in the radar calibration depending on the location of the gauge used in the pixel by gauge calibration.

5.2. Spatial correlation of rainfall data

So far, the variability has been quantified on the basis of the coefficient of variation of rainfall depths. To add more confidence to the findings, the inter-gauge correlation coefficient is estimated in the following section. The correlation coefficient gives a degree of linear dependency between a pair of variables, and the key prerequisite is that the data is normal distributed. As described in Section 4, the dataset in question can be assumed to be log-normally distributed and is used as such in the following. Despite the very different behavior of the data, as well as the physical circumstances, it cannot be ruled out that the variability observed in the 2003 dataset is correct. However, since other experience with the gauges has later shown strong irregularities, the 2003 dataset has been omitted from the following analysis in order to get a more consistent result. The 2008 dataset covers the range of rainfall depths and intensities observed in 2003.

The traditional approach for estimating the correlation of a pair of rainfall processes observed by two gauges is by using the population correlation coefficients $\rho(X,Y)$, where (X,Y) is the pair of observed rainfall events at two locations. To estimate $\rho(X,Y)$ the Pearson sample correlation coefficient r, derived for N observations is used:

$$r(X,Y) = \frac{\overline{XY} - \overline{X}\,\overline{Y}}{\sqrt{(\overline{X^2} - \overline{X}^2)(\overline{Y^2} - \overline{Y}^2)}} \qquad (2)$$

There are some problems in using the Pearson sample correlation coefficient as an estimation. Habib et al. (2001) state that scatter in the correlation data became more evident when the sample size was reduced and the estimated values became more biased upwards. Furthermore, the Pearson sample correlation coefficient is limited in order to give a local correlation coefficient of two locations not taking the inter-gauge distances into account if a cluster of gauges is considered.

The Pearson sample correlation coefficients are estimated for all possible gauge combinations plotted as a function of the inter-gauge distance in Fig. 12. The correlation coefficients are estimated for 2007 and 2008 separately, cf. Fig. 12. Three correlation coefficients deviate significantly from the rest in the 2007 dataset, and these have been identified as being part of pairs including gauge 1. There are only five events where gauge 1 is functioning, so the foundation for the computation is very weak. Fig. 12 — right-hand side — shows the correlation coefficients without 2007 gauge 1 data. The estimated correlation coefficients are all very close to 1, indicating that over many events the gauges measure the same. It is therefore assumed that they are working equally well and that no gauge is biased. There is an indication of reduced correlation with increasing inter-gauge distance, which in turn indicates that there is variability within the area. Unfortunately, there are only two gauge combinations with the highest inter-gauge distances due to the layout of the experiment.

The correlation in Fig. 12 provides only limited information on the variability of individual events of interest to this study. The correlation was, therefore, determined based on an
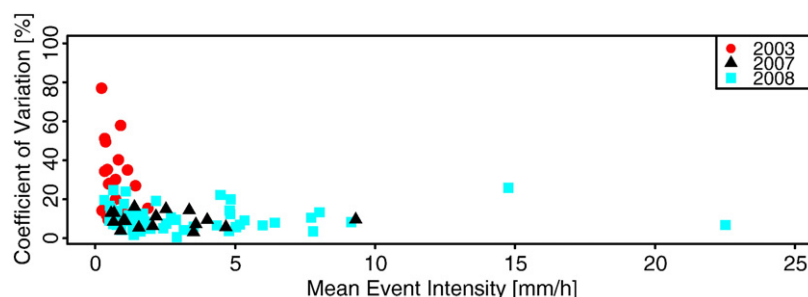

Fig. 10. Coefficient of variation (CV) as a function of the mean event intensity in mm/h for all three datasets. 2003 (circles), 2007 (triangles) and 2008 (squares).

No Data	12.4 mm (23 min)	7.0 mm (30 min)
12.0 mm (23 min)	8.6 mm (28 min)	4.2 mm (25 min)
2.5 mm (25 min)	6.0 mm (27 min)	6.3 mm (29 min)

Fig. 11. Accumulated rainfall observed by the nine rain gauges on 4 August 2008. Numbers in brackets are duration in minutes. This event has been omitted due to the large variance, but the variation could be true because only a part of the area with the gauges was hit by the peak of the event and it cannot be ruled out that the coefficient of variation can be as high as 47%, as for this event.

can be caused by a gradient over the field, which can be explained by a dominant direction of the rainfall fields. The average correlation coefficients are not all zero, which makes the assumption of independent observations weaker. In reality the correlation may be zero, especially as only very few observations contribute to correlation at far ranges. Based on the explanation of the gradient and the sparse dataset at the far ranges, the standard deviations are used to describe the variability in the following.

In order to estimate the standard deviations, the covariance matrices of the two multivariate normally-distributed datasets are determined. The diagonal elements of the covariance matrix are the variance of the gauges, and by exploiting this property it becomes possible to use the variance as an expression of the variability between the gauge-observed rainfall depths. In order to aid interpretation, the standard deviations are shown. The estimated values for each gauge are listed in Table 5.

A small standard deviation is interpreted as a well-functioning gauge, and gauges with large standard deviation values are considered more uncertain. This is confirmed by the general observations in the field and the fact that gauges 1 and 7 were found to be the most unstable and to yield the highest standard deviations. It should be noted that gauge 1 has significantly fewer observations than any of the other gauges, but there is no apparent correlation between the number of observations and the standard deviations. The standard deviations vary from gauge to gauge, cf. Table 5.

To illustrate the uncertainty of using a single rain gauge for the minimum and maximum standard deviation of the rain gauges, the 2008 dataset from Table 5 has been chosen to calculate the 95% prediction intervals for a given rain depth. The 95% prediction interval for a given rainfall depth is defined as $\mu \pm 2\sigma_{(\min \text{ or } \max)}$, and the increase in the band (interval) width is due to the log transformation. The example shown in Fig. 14 uses rainfall data from 2008, in which the narrow interval represents the scenario where the variability interval is predicted based on the gauge with less variability ($\sigma_{\min}$), and the wide interval represents the scenario where the

altered dataset, in which the mean value of the event is subtracted from all observations in the corresponding event. By doing this it is assumed that each event can be considered to be multivariate normally distributed. This dataset will be referred to as multivariate normally distributed. With zero correlation in the multivariate normally distributed dataset only independent white noise should remain, and each gauge can therefore be treated as an independent observation. Thus it is possible to use the standard deviation of the individual gauges as a measure of variability.

The correlation coefficients estimated on the multivariate dataset are shown for each dataset separately in Fig. 13.

The average estimated correlation coefficient for the shorter inter-gauge distances (167 m and 236 m) is very close to 0, with a negative trend at the longer distances for 2007 and 2008, cf. Fig. 13. The increase in negative correlation
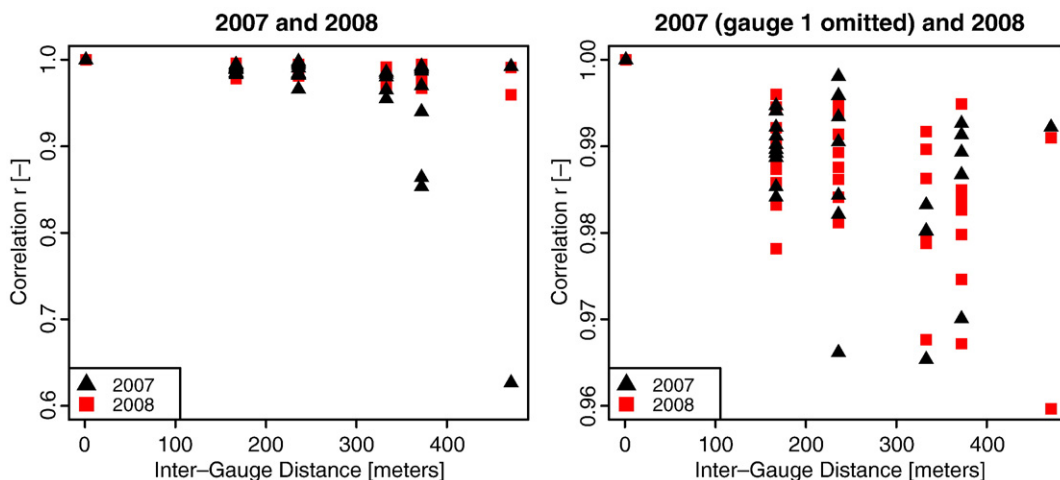
Fig. 12. Pearson's sample correlation coefficients as a function of inter-gauge distances. The left plot shows the correlation coefficients based on all gauge pairs, while gauge 1 has been omitted from the 2007 dataset in the plot on the right since it results in the three significantly smallest correlation coefficients in the left plot.
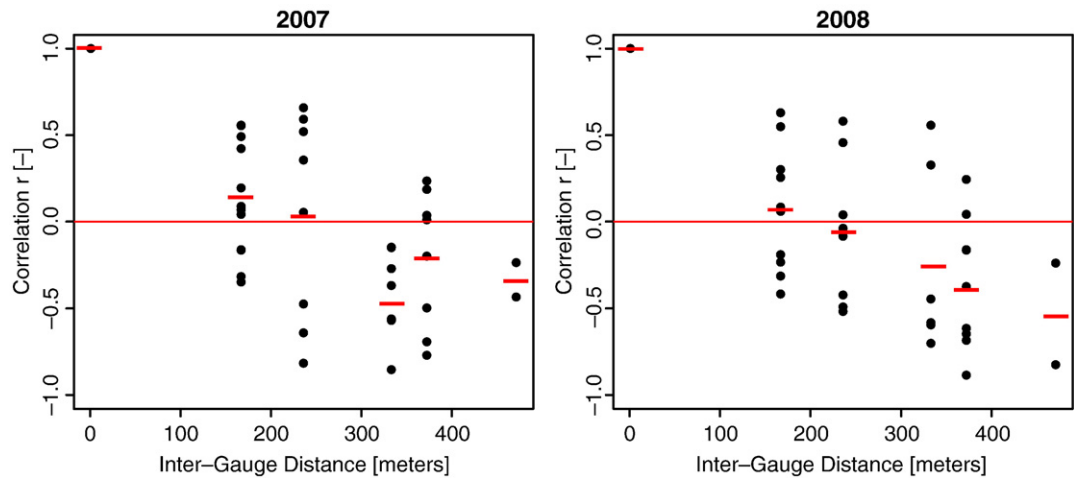
Fig. 13. Correlation coefficients based on the multivariate datasets for 2007 and 2008. The bars are the average for the different inter-gauge distances. Gauge 1 is included in the 2007 dataset.

variability is predicted based on the gauge with most variability (σ_{max}). The prediction intervals are transformed from log-space to normal-space for the plot in Fig. 14. In order to simplify the illustration, the prediction intervals shown in Fig. 14 are estimated based on the mean event depth, but the uncertainty is for the separate observations.

To aid clarity, Fig. 14 contains only 2008 data, but the range of the standard deviations of the two datasets is almost identical. A close-up of the 1–5 mm interval of the mean rainfall depths from Fig. 14 is shown in Fig. 15 to better illustrate the uncertainties related to small rainfall depths.

As illustrated by Figs. 14 and 15, the uncertainty of using a single gauge as a representative measure of the mean rainfall depth over the 500×500 m area is largest for events with depths smaller than 5 mm, since a higher number of the individual gauge depths are outside the 95% prediction interval.

The narrow interval ($\pm 2\sigma_{min}$) is approximately $\pm 12\%$ and the wide interval ($\pm 2\sigma_{max}$) is approximately $\pm 23\%$. Due to the transformation from the logarithmic values, the positive part is wider than the negative and therefore the percentage value is an approximation.

6. Conclusions

The aim of this paper was to quantify the uncertainties of using a single rain gauge to represent the rainfall falling over a 500×500 m area corresponding to a single pixel of a Local Area Weather Radar (LAWR). The motivation for the work was the assumption of uniform rainfall used in connection with

the calibration of weather radars, where the rainfall measured by a single gauge is assumed to be representative for a pixel ranging from 0.25 km^2 to 16 km^2. The same assumption is used in connection with urban drainage modeling, where the rainfall measured by a single gauge is assumed to be uniform over a large catchment.

A field experiment placing nine rain gauges within an area of 500×500 m, each representing one-ninth of the area, was used to address the issue. The gauges were originally placed in an open field, but were later moved to a shallow estuary. It can be concluded that an offshore location is not recommendable for future gauge sites since, although the inaccessibility of the site may be excellent for avoiding vandalism, it also complicates service and maintenance of the gauges to a severe degree. Furthermore, resting birds and their droppings became a major problem, resulting in corrupted data.

2007 resulted in a total of 19 events with a total average rain depth of 71 mm over the 500×500 m area. 2008 resulted in 55 events with a total average rain depth of 222 mm over the 500×500 m area. The data from 2007 and 2008 were pooled with a previous dataset from a similar experiment carried out in 2003, in which there were 20 events having a total average rain depth of 69 mm. The rainfall data was determined to be log-normally distributed; a property facilitating the use of standard statistic methods for estimating the variability.

The variability of accumulated rainfall within the 500×500 m area was estimated in different ways in order to obtain a robust estimate of the variability and thereby the uncertainty of using a single gauge to represent the rainfall

Table 5
Standard deviations for each gauge based on the covariance matrix of the multivariate normal distributed data. The standard deviations are estimated based on all events for the gauges separately.

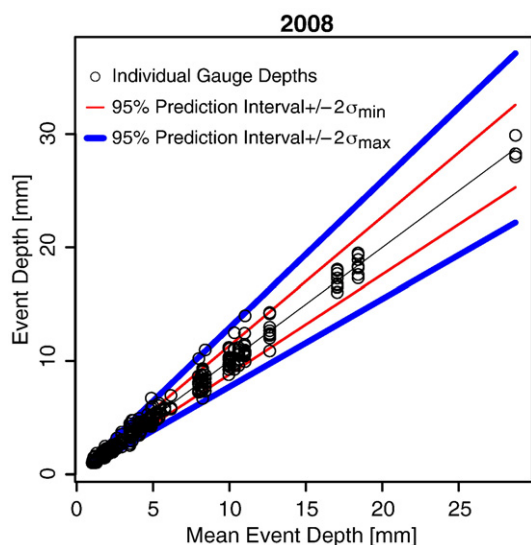|	Gauge 1	Gauge 2	Gauge 3	Gauge 4	Gauge 5	Gauge 6	Gauge 7	Gauge 8	Gauge 9
$\sigma(\log(2007))$ [−]	0.10	0.11	0.06	0.07	0.04	0.09	0.06	0.11	0.11
$\sigma(\log(2008))$ [−]	0.13	0.08	0.06	0.10	0.07	0.10	0.11	0.07	0.09

Fig. 14. Illustration of the 95% prediction intervals for rainfall depths recorded by a rain gauge. The 95% prediction interval is shown as a function of the mean depth of 3–9 working gauges for illustrative purposes — the prediction intervals are estimated based on the individual gauges. The minimum and maximum of $\sigma(\log(2008))$ from Table 5 has been used.

over the area. The first and simplest approach uses the coefficient of variation (CV) as a measure of the variability on an event basis. The CV values decrease with increasing rainfall depths, indicating that the largest variability is in events with an average rainfall depth of less than 5 mm. The CV values ranged from 1–77% for the complete dataset, and from 1–26%
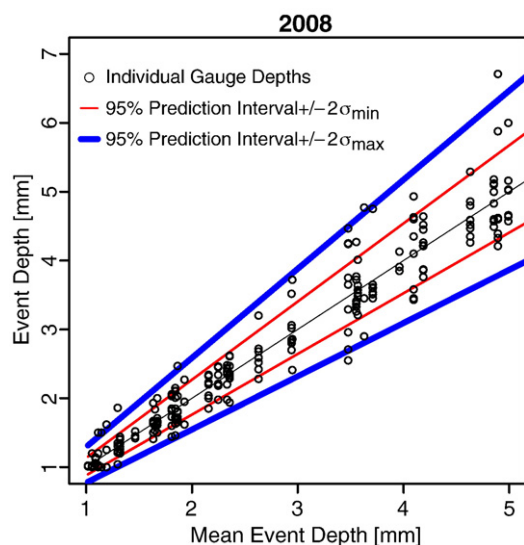

Fig. 15. Close-up of Fig. 14 showing the interval 1–5 mm rainfall depths illustrating the 95% prediction intervals for small rainfall depths recorded by rain gauge. The 95% prediction interval is shown as a function of the mean depth of 3–9 working gauges for illustrative purposes — the prediction intervals are estimated based on the individual gauges. The minimum and maximum of $\sigma(\log(2008))$ from Table 5 has been used.

based on the 2007 and 2008 dataset alone. The large CV values for the 2003 dataset are to be considered, keeping in mind the type of gauge used, since the optical drop-counting gauges were found to be more unreliable than the tipping bucket type. It cannot, however, be concluded that the values are false, but the large CV values were all for events with a depth of less than 5 mm.

The correlation analysis of the data showed a very strong correlation among the gauges, but it decreased somewhat with increasing inter-gauge distance, signifying variability over the area. The correlation analysis only provided an overall estimate of the variability among the gauges, whereas the focus of this work was on the inter-event variability. The data was therefore transformed into multivariate normally-distributed data. The standard deviation of the gauges can thus be used to express the variability as a function of rainfall depth. A 95% prediction interval of the gauges based on the rainfall depth multiplied by $\pm 2\sigma$ is used to give an estimate of the interval within which the observation from a single gauge would be. The standard deviations range from 0.4–0.11 mm for 2007 and 0.6–0.13 for 2008. For the events larger than 15 mm the standard deviations are all within $\pm 2\sigma_{min}$, whereas for the smaller events they are more scattered. However, most of the events with depths less than 5 mm are within $\pm 2\sigma_{max}$.

On the basis of the analysis of the coefficients of variation the conclusion is that the intra-event variability, which ranges from 1–26%, decreases with increasing rainfall depths and is independent of the mean event intensity. This is confirmed by the standard deviations estimated for each gauge separately. The standard deviations are used to defined the 95% prediction interval for a given rainfall depth using a single gauge. The narrow interval of the two estimated 95% prediction intervals ($\pm 2\sigma_{min}$) includes all the large rainfall depth observations, whereas some of the smaller observations fall outside the 95% prediction interval determined by $\pm 2\sigma_{min}$. Whether to use σ_{min} or σ_{max} or an average of these, depends on how much trust one wishes to place in the gauge used for the application in question. A conservative approach would be to use σ_{max} (the wide interval) to define the variability range of input data and thereby the uncertainty that would have to be added to the output of an application, assuming rainfall is uniform within 500×500 m.

References

Ciach, G.J., Krajewski, W.F., 2006. Analysis and modeling of spatial correlation structure in small-scale rainfall in Central Oklahoma. Advances in Water Resources 29, 1450–1463.

DHI, 2009. DHI-Local Area Weather Radar — LAWR Documentation ver. 3.2.

DMI, 2008. Måneden, sæsonen og årets vejr [Online]. . October 15, 2008. — http://www.dmi.dk/dmi/index/danmark/maanedens_vejr_-_oversigt. htm.

Einfalt, T., Jessen, M., Mehlig, B., 2005. Comparison of radar and rain gauge measurements during heavy rainfall. Water Science and Technology 51, 195–201.

Habib, E., Krajewski, W.F., Ciach, G., 2001. Estimation of Rainfall Interstation Correlation. Journal of Hydrometeorology 2, 621–629.

Jensen, N.E., Pedersen, L., 2005. Spatial Variability of Rainfall. Variations Within a Single Radar Pixel. Journal of Atmospheric Research 77, 269–277.

Krajewski, W.F., Ciach, G.J., Habib, E., 2003. An analysis of small-scale rainfall variability in different climate regimes. Hydrological Sciences 2 (48), 151–162.

Marshall, J.S., Palmer, W.McK, 1948. The distribution of raindrops with size. Journal of Meteorology 5, 165–166.

Pedersen, L., Jensen, N.E., Madsen, H., 2008. Estimation of radar calibration uncertainties related to the spatial variability of rainfall within a single radar pixel — statistical analysis of rainfall data from a dense network of rain gauges. Preceedings from Environmental World Congress 2008, Honolulu.

Rahimi, A.R., Holt, A.R., Upton, G.J.G., Krämer, S., Redder, A., Verworn, H.R., 2006. Attenuation calibration of an X-band weather radar using a microwave link. Journal of Atmospheric and Oceanic Technology 23, 395–405.

Sempere-Torres, D., Corral, C., Raso, J., 1999. Use of weather radar for combined sewer overflows monitoring and control. Journal of Environmental Engineering 372–380.

Sevruk, B., Ondrás, M., Chvila, B., 2009. The WMO precipitation measurement intercomparisons. Journal of Atmospheric Research 92, 376–380.

Thomsen, R.S., 2006. Drift af Spildevandskomitéens Regnmålersystem Årsnotat 2006. Danish Meteorological Institute.