Anonymous referee #1
Received and published: 05 Jul 2016

General comments
The goal of the paper is the analysis of the uncertainty of the areal interpolation of precipitation data from a dense gauge network with a high temporal resolution. The topic itself is relevant and currently discussed, in particular, in the context of radar measurements.
From this point of view it is very interesting to analyse such a very dense gauge network with respect to the influence of the spatial variability/measurement errors on the areal interpolation. The paper is well structured and the methods used are (in most cases) clearly presented using a step wise explanation. However, a major problem of the study is the sample size of only 13 events (of 2 years). As these events are highly variable as shown in table 1, the uncertainty of the entire study is rather high. Although there is a pooling method presented that tries to overcome this issue, the underlying assumption might still be problematic (see comment [5: 5]). Furthermore, many descriptive results are presented, but their consequences are rarely discussed. There are also a lot of trivial results in the paper. It should become much clearer, what is a logical consequence of the precipitation structure and what is an actually new result of this study. For example, temporal averaging will always reduce precipitation peaks. Finally, the language could be more precise because several sentences are too fuzzy.

<u>Authors:</u> *We thank the reviewer for the professional and thorough review of this paper. We have attempted to address all the comments and our response is listed below.*

….However, a major problem of the study is the sample size of only 13 events (of 2 years). As these events are highly variable as shown in table 1, the uncertainty of the entire study is rather high.

<u>Reply:</u> *The major technical comment from the reviewer is that our analyses are based on a small sample size of only 13 events. We would like to clarify that the entire ten months of rainfall data from 8 locations were used for the development and calibration of the geostatistical model. Webster and Oliver (2007) suggested around 100 samples to reliably estimate a variogram model. Even in the case of 30 min temporal averaging interval and > 10 mm/hr (where we had the least observations) we had 196 sampling to calculate the variogram which is sufficiently larger than 100. Hence all our variogram models (based on which further results are derived) are stable and reliable. The event separation (ref Table 1) is used only for the analyses presented in Sections 4.2.3 and 4.2.4. We apologise if it was not clear in the manuscript that the entire 10 months of data was used and we will add additional text in the revised manuscript to clarify this.*

Furthermore, to test the stability of variogram estimation we carried out the following experiment. We randomly selected 80% of the data from each intensity class and produced the variograms again to compare these with the variograms presented in Figure 7 of the manuscript. The variograms computed from 80% of the data are presented below in Figure C1.

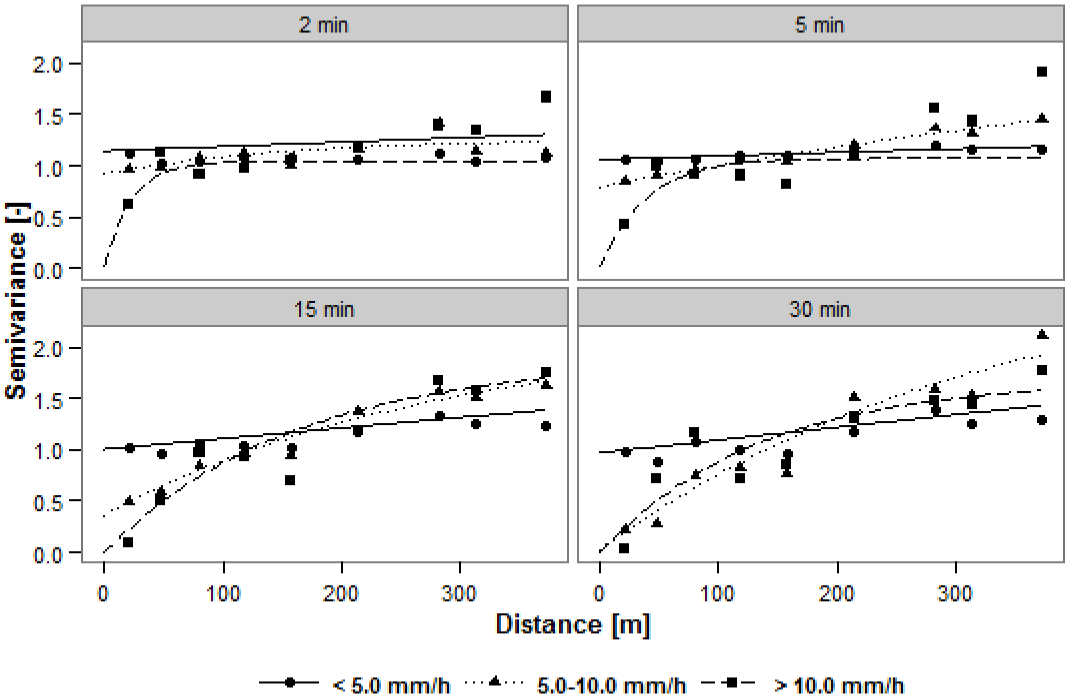

Figure C1: Calculated variograms for each temporal averaging interval and for each range of intensity within a temporal averaging interval using randomly selected 80% of the data from each subclasses

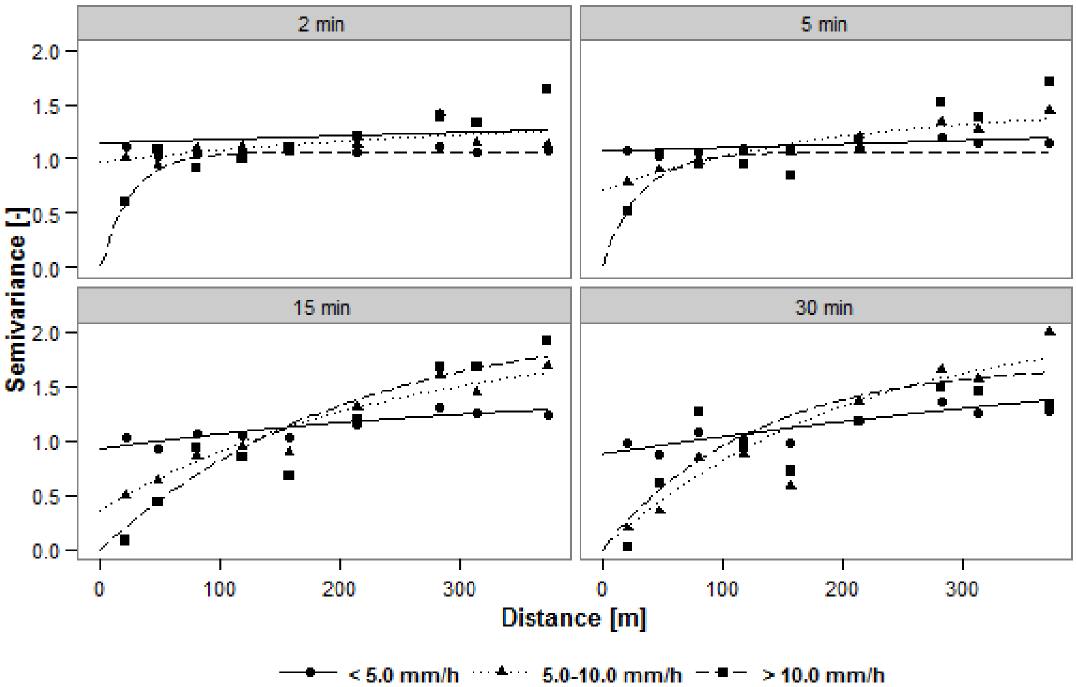

Figure 1: Calculated variograms for each temporal averaging interval and for each range of intensity within a temporal averaging interval

Comparison of Figure C1 with Figure 7 shows that these are very similar. This analysis supports our claim that the variograms shown in Figure 7 are stable and an adequate representation of the rainfall spatial variation for each intensity class and temporal averaging interval.

We acknowledge that the data cover only 10 months i.e. two summer periods in 2012 and 2013. However, for previous studies using such a dense network the duration of data collection is similar (e.g.: 15 months - Ciach and Krajewski, 2006; 16 months - Jaffrain and Berne, 2012). These durations are a reflection of the practical and funding issues to maintain such networks operating accurately for extended periods. The characteristics of our data are comparable with (Ciach and Krajewski, 2006; Fiener and Auerswald, 2008) as these studies also used rainfall data from warm months to investigate the spatial correlation structure.

Although there is a pooling method presented that tries to overcome this issue, the underlying assumption might still be problematic (see comment [5: 5]).

<u>Reply:</u> *The underlying assumption of pooling is discussed in detail in our reply to specific comment 5: 5.*

Furthermore, many descriptive results are presented, but their consequences are rarely discussed. There are also a lot of trivial results in the paper. It should become much clearer, what is a logical consequence of the precipitation structure and what is an actually new result of this study. For example, temporal averaging will always reduce precipitation peaks. Finally, the language could be more precise because several sentences are too fuzzy.

<u>Reply:</u> *We will attempt to deal with the concerns stated by the referee. Superfluous sentences will be removed and some will be rewritten to reduce the attention paid to trivial results, to improve clarity and discuss the consequence of the analysis carried out in the paper. Details of what is proposed are described below.*

Trivial results removed, or reduced
Several sentences will be removed and some will be rewritten to try and reduce the level of trivial results, including the ones mentioned by the reviewer in comments. For more details please refer to the specific comments [5: 11], [5: 12-13] and [10: 16].

Improvement of clarity and language
The following parts will be modified to give a clearer explanation, as indicated in specific comments of the reviewer:
Quality control of rainfall data: See our reply to comment [3:32]
Event separation: See our reply to comment [4: 14]
Stochastic simulation: See our reply to comment [6:29-7:22]
Interpretation of the variograms: See our reply to comment [8: 16]

New results/novelty
Because of the nature of this study more emphasis is given to the methodology. We believe that this methodology is novel for the following reasons:

- *In literature, geostatistics has been used to analyse the spatial correlation structure of rainfall at various spatial scales, but its application to estimate the level of uncertainty in rainfall upscaling has not been fully explored mainly due to its inherent complexity and demanding data requirements. In this study we attempted to address these challenges which include the use of repetitive rainfall measurements (pooling) to increase the spatial samples artificially.*

- *We used the spatial stochastic simulation to address the combination of change of support (from point to catchment) and non-normality for prediction of rainfall and associated uncertainty. To the best of our knowledge this has not been done previously.*

- *We classified intensity classes and derived different geostatistical models (variograms) for each class separately. On top of that we also used different temporal averaging intervals. To the best of our knowledge no previous study attempted to assign geostatistical models for a combination of intensity class and temporal averaging interval.*

In addition we presented and discussed the following new results:

- *Spatial correlation structure of rainfall at a spatial extent of 200m × 400m has been presented. To the best of our knowledge no previous studies have analysed spatial correlation structure of rainfall at such a small spatial extend.*

- *The current study shows that for this spatial scale the use of a single geostatistical model based on a single variogram is not appropriate. Instead, different variograms for different rainfall intensity classes should be used. This is a key finding of the study.*

- *In addition to visual interpretation we also presented quantification of uncertainty in upscaled rainfall predictions in terms of CV for peaks of rainfall intensity. Previous studies (e.g. Villarini et al. 2008) have provided such quantification of uncertainty only using simpler error measures (normalized root mean squared error and normalized mean absolute error) while we use geostatistical approaches that take the effect of spatial correlation into account.*

Specific comments
[Page: Line]

[2: 19] "Since rainfall can vary over space significantly, any method for scaling up the point rainfall measurements adds uncertainty on top of existing measurement error."
The measurement error is not explained so far. It should be introduced as it is a major aspect in the rest of the paper.

<u>Reply:</u> *A short introduction (given below) on measurement error mainly focusing on tipping bucket errors will be added in our revised manuscript.*

"….The uncertainty mainly comes from two sources; uncertainty due to measurement errors and uncertainty associated with spatiotemporal variability of rainfall. The characteristics of measurement errors can vary depending on the rain gauge type. For example, errors associated with tipping bucket rain gauges which is one of the widely used rain gauge type range from errors due to wind, wetting, evaporation, and splashing (Fankhauser, 1998; Sevruk and Hamon, 1984) to errors due to its sampling mechanism (Habib et al., 2001)….."

[Fig 5] For a better comparison, the classes of the histogram should be the same as the classes used later for the variogram.

<u>Reply:</u> *The rainfall intensity class intervals were the same for histograms (Fig. 5) and variograms (Fig. 7).*

[2: 32] ": : :not always: : :" Rainfall intensity values are almost never normally distributed

<u>Reply:</u> *Thank you for pointing this out. This will be corrected in the revised manuscript.*

[3: 32] "During the dynamic calibration: : :" How did you identify the volume error per tip using a dynamic calibration? The entire section is not clear: "Every long set of data: : :if the differences: : :" What is the long set and what kind of differences?

Reply: *This section will be modified in the revised manuscript as given below to give more precise information on how the quality control was carried out using paired setup.*

Quality control procedures were performed prior to statistical analysis, taking advantage of the paired gauge setup. The paired gauge design provides efficient quality control of the rain gauge data records as it helps to identify the instances when one of the gauges fails, and to flag the periods of missing or incorrect data (Ciach and Krajewski, 2006). During the dynamic calibration of all rain gauges in the laboratory before the deployment, it was identified that the highest and lowest values of the calibration factors for the tipping bucket size are 0.196mm and 0.204mm, respectively, resulting in a maximum potential error of ± 4 % due to the observed variation per tip for any two of the rain gauges used in this study. It was therefore decided that this is the maximum acceptable difference in cumulative rainfall that could be accepted between any pair of gauges. Sets of cumulative rainfall data corresponding to specific events from the paired gauges were checked against each other and if the (absolute) difference in cumulative rainfall was greater than 4 %, that complete set was identified as unreliable (e.g. due to partial clogging caused by debris) and the data from both gauges were removed from further analysis.

[3: 8] ": : : to obtain a more normally : : :" After the transformation they are perfectly normally distributed.

Reply: *Thank you for pointing out this out. This will be corrected in the revised manuscript*

[4: 14]: How are the events defined? What are the criteria of the end of the event; if all stations show zero precipitation? Is there a minimum separation time of two events?
Or will a few minutes without rainfall already separate the events? Why does the event need to be at least 20 min?

Reply: *There are two conditions for defining an event: the yield should be more than 10 mm and the event duration should be larger than 20 minutes. If all stations show zero precipitation based on 5 min averaging interval it is considered as the end of an event and there is no minimum separation time between events. Figure 3 and Table 1 show that except for events 6 and 7 there are no events close to each other.*

We chose a 20 min window in order to get a sufficient number of data values when temporal averaging intervals of more than 2 min are used in the analysis. For example, a 10 min event gives two data values when using a 5 min averaging interval and gives only 1 data value for 15 min and 30 min averaging intervals. Hence in order to have at least two data values for all temporal averaging intervals examined, a minimum event duration of 20 minutes is needed. Table 1 shows that the lowest event duration in the collected data was 1.5 hours. Hence all events had at least 45 data values for a 2 minute averaging interval and at least 3 data values for a 30 minutes averaging interval.

Also note that, as we mentioned already, the event separation (ref Table 1) is used only for the analyses presented in Sections 4.2.3 and 4.2.4. Hence these criteria don't leave out any data in the development and calibration of the geostatistical models.

[5: 5] "The underlying assumption: : :" This is your major assumption for the pooling based on the trade-off between independence and number of available time instants for the three defined classes. However, you did not show any analysis to validate the independence assumption. Instead, literature is cited, which shows the dependence of the spatial correlation on the intensity. How can you be sure that the influence is small enough to be disregarded, and that the pooling procedure does not mess up your further analysis?

<u>Reply:</u> We agree with the concerns expressed by the reviewer. This is the reason why we introduced step 2 (6: 1-5) to treat each and every time instant within a subset, by individually calculating their mean and standard deviation. Although variograms are derived only for the whole subset, step 2 (before geostatistical upscaling) and step 9 (after geostatistical upscaling) ensure that the probabilistic model is adjusted for each time instant separately, based on the mean and standard deviation for that particular time instant. Effectively, we assume the same correlogram for time instants of the same subclass, not the same variogram. Although this does not justify the assumption of similar spatial correlation structure with in the pooled classes, it at least relaxes the assumption of the same variogram within subclasses.

The similarity in Figures C1 and 7 also shows that the characteristics of the data from subsets are consistent with those of the entire pooled class. Moreover, several studies (e.g. Dirks et al., 1998; Ly et al., 2011; Tao et al., 2009) used only a single geostatistical model in the form of single variograms/correlograms for the entire range of rainfall intensity. The current study shows that for small time and space scales the use of a single geostatistical model based on a single variogram is not appropriate. This is a key finding of this study. We agree that with narrower intervals the assumption of consistency in spatial variability would be more realistic. But with the available data we had to find a compromise with the number of time instants. We believe that using three intensity subclasses is a workable compromise. Based on Figure C3 (in our reply to comment 8:16), where variograms are produced for narrower subclasses, we conclude that the variograms shown in Figure 7 are good representations of the average spatial variability conditions for corresponding intensity classes.

[5: 11] "As expected, with increasing temporal averaging: : :" This is not only expected, this is obvious for an aggregation process of rainfall.

<u>Reply:</u> We agree with this comment.
"As expected, with increasing temporal averaging the number of time instants t reduces." will be corrected as "Number of time instants t reduces with increasing temporal averaging intervals due to the aggregation process."

[5: 12-13] This is also not surprising due to the skewed intensity distribution of precipitation. It should become clearer in this part [5:11-13], that the results are just natural characteristics of precipitation.

<u>Reply:</u> We agree that the dominance of low intensity rainfall is a natural phenomenon. It will be corrected in the manuscript.
"Figure 5 also shows that the higher the intensity, the smaller the t. There is a large difference between t for lower and higher intensity ranges which shows the dominance of lower intensity (0.1-5.0 mm/h) rainfall over the recording periods" will be corrected as
"The natural characteristic of rainfall data results in the dominance of lower intensity rainfall (0.1-5.0 mm/h) over the recording period."

[5: 14] ": : :only eight: : :" Where does this number come from? In figure 5 these are way more than eight for the 30 min average.

<u>Reply:</u> Thank you for pointing out this. We apologise for the mismatch between text and figure. This histogram shows the number of time instants (t) × number of stations (eight). It will be corrected in the revised manuscript to show the number of time instants (t) to be consistent with the text. Actually there are seven (not eight) time instants where intensity exceeds 10 mm/hr at 30 min temporal averaging interval. The new figure is presented below. This figure will replace Figure 5 in the revised manuscript.

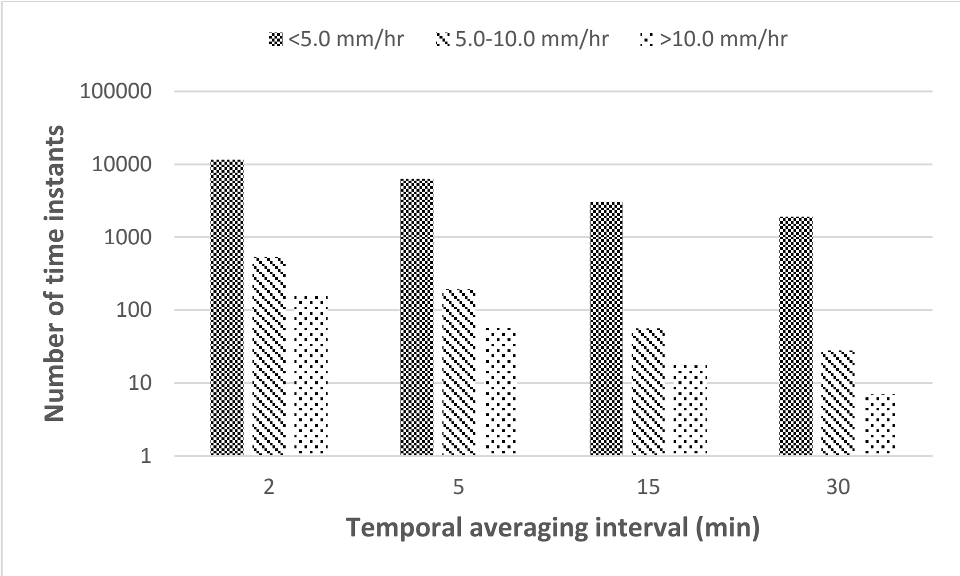

Figure C2: Calculated variograms for 5 min averaging interval and for a narrower range of intensity.

[6: 26] "It is negligible small: : :" Isn't that contradicting to the later parts where the uncertainty of the tipping bucket error is analysed (for example in [12:15]).

<u>Reply:</u> We wanted to argue that any physical micro-scale spatial variation between rain gauges, which theoretically may be one of the causes of the nugget effect, is negligibly small for this case. We did not mean to claim that the nugget itself is negligibly small. We accept that the particular sentence was poorly constructed. We also think that this sentence can be removed as the nugget effect is discussed in detail at a later stage in the manuscript.

[6:29-7:22] In the entire chapter it becomes not clear what kind of stochastic simulation was performed (conditional/unconditional) and which method was used to obtain the 500 simulation results.

Reply: We accept that this section requires more detail. We will modify this section in the revised manuscript as given below.

The output from spatial stochastic simulation is a set of alternative realisations ('possible realities') of rainfall at user-defined grid points. The differences among these realisations are used as a measure of uncertainty. Hence it involves the following two steps
1. Definition of grid cells (25m × 25m in this case)
2. Production of substantial number of possible realisations (500 in this case) for each time instant using stochastic simulation based on the corresponding geo statistical model (variograms in this case).

We used the stochastic simulation conditioned on the available rainfall data at measured stations. The grid size and number of simulations were selected considering the spatial resolution of available measurements and computational demand. It was observed that neither a finer grid nor more simulations improves the results much. Increasing the resolution to 10 m × 10 m only improves the standard deviation of the prediction by less than 5% in most cases while almost doubles the computational time.

[7: 10] Kriging would be possible, if the back-transformation of the individual points was performed before the averaging. (No block kriging, but ordinary kriging of single points
(25 x 25m grid)

<u>Reply:</u> This might work for the prediction but not for the prediction uncertainty of spatially aggregated rainfall, which we also needed to quantify in this study. That is why we turned to spatial stochastic simulation.

[8: 1]": : :is spatial aggregation of each and every simulation: : :" Should it be of "each time step"?

<u>Reply:</u> It should be 'spatial aggregation of each and every simulation (realisation)' (i.e. 500 per time instant). This yields 500 values of the areal mean per time instant. We will modify Section 3.5 to give a clearer explanation of spatial stochastic simulation (refer comment [6:29-7:22]) which we believe that it will help to understand the subsequent steps better.

[8: 16] How can you be sure that this effect is caused by measurement errors? Couldn't the nugget effect be also caused by the pooling technique, that is, by mixing different time steps; or by a high variability in the natural precipitation process? (see Remark [9:
7-21])

<u>Reply:</u> The two other possible reasons for the nugget effect as mentioned by the reviewer are discussed in detail separately below.

1. *Pooling of time instants*
   We do not think pooling is the reason for the nugget effect. We pooled sample variograms from different time instants. But that cannot cause a nugget effect; it just gives back the average nugget of those for all time instants. If none of the variograms from different time instants have a large nugget, then the pooled one also will not have a large nugget.

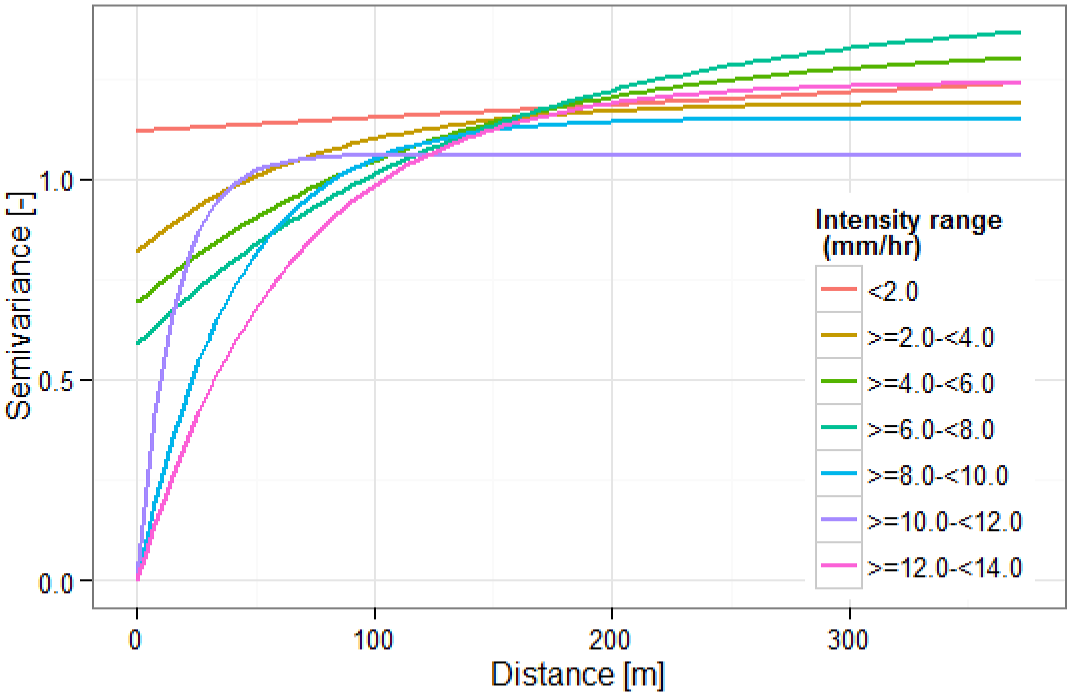

Figure C3: Calculated variograms for 5 min averaging interval and for a narrower range of intensity.

Furthermore, Figure C3 shows the behaviour of variogram models for narrower intensity classes ranging from 0 to 14 mm/hr for the 5 min averaging interval. The highest intensity class is limited to ≥12 - <14 mm/hr as for further narrower ranges (i.e ≥14 - <16 mm/hr and so on) there are not enough sample points to produce a meaningful variogram. The variograms for the narrower classes show that the assumption of similar spatial variability *with in a pooled subset is stronger. Hence the effect of variation caused by pooling is reduced. Comparing this figure with Figure 7 not only shows the behaviour of the nugget effect against the intensity as seen from Figure 7, it also makes clear that there is a clear decreasing trend in the nugget as intensity increases. This also indicates that pooling is not responsible for the pattern of the nugget effect.*

2. *High variability in the natural precipitation process*
   Please refer to our response to the comment [9: 7-21]

[9: 2] Is there a reason why the content of Habib et al. is mentioned explicitly and not of Villarini et al?

<u>Reply:</u> *The main aim of (Habib et al., 2001) was to investigate the sampling error of tipping bucket measurements, whereas (Villarini et al., 2008) was a later study and derived a similar conclusion as a part of their findings. Among the two only (Habib et al., 2001) discuss the sampling errors of tipping bucket rain gauges extensively. Hence their work is explicitly mentioned.*

[9: 7-21] The important point of the interpretation of the variograms is not clearly stated.
The variograms actually show, that for short time periods < 5 min (except for high intensities), there is almost no spatial correlation, that is the field is just random. If the nugget (explained here as tipping bucket error) is almost as high as the sill, there are two options. First, there is just no spatial correlation at the regarded distance, or the spatial correlation of the field cannot be detected by the tipping buckets because of the measurement error. There is also a very weak correlation (even for high aggregations) for the intensities smaller 5.0 mm/h. How can you be sure, that the nugget comes from the tipping bucket error, and does not represent a very high spatial variability of the natural precipitation at very short distances?

<u>Reply:</u> *We accept that the nugget effect could be due to a combination of micro-scale spatial variability and measurement error. Since we cannot quantify the nugget effect caused by measurement error we cannot prove quantitatively that the nugget is only due to the measurement error.*

But with regarding the clear increasing trend of nugget against (a) decreasing temporal averaging interval and (b) increasing intensity range, we believe this trend is as a result of similar trend in sampling related error of tipping bucket (TB error). If it is due to spatial variability of rainfall then such trends expected to be consistent at greater distance too. But the variograms show no consistent trend against intensity range at greater distance. In a previous similar study (Ciach and Krajewski, 2006), where the behaviour of spatial correlation against rainfall intensity was analysed, they also could not find a consistent trend and concluded that such trends are not consistent.

In summary,

1. *Nuggets corresponds to each variogram could be a combined effect of measurement error and spatial variability. We cannot comment on individual contribution as we cannot quantify them.*
2. *The trends in the nugget against a. rainfall intensity and b. temporal averaging interval correspond well with TB error. Hence these trends in the nugget can be attributed to the TB error.*

We will include the above discussion in the revised manuscript.

[10: 16] “Here it can be noted: : :” As the precipitation intensities are never uniformly distributed, this effect is a logic consequence.
<u>Reply:</u> *Thank you pointing this out, this will be corrected in the revised manuscript.*

"Here it can be noted that in this event this effect decreases the event peak AARI from around 50 mm/h to around 20 mm/h as the temporal averaging interval increases from 2 min to 30 min." will be corrected as "As a result of temporal aggregation the event peak gets reduced from around 50 mm/h to around 20 mm/h as the temporal averaging interval increases from 2 min to 30 min"

[10: 30] This chapter should be rewritten, could be shortened and included in the next one. As explained in the last sentence of the paragraph, it is difficult to compare the standard deviation of different absolute values. Figure 10 is rather useless, as the intervals (uncertainty) cannot be read. A table including the standard deviations in addition with the CVs for the single events could help.

Reply: Sections 4.2.3 and 4.2.4 will be combined into one section in the revised manuscript as suggested by the reviewer. We agree that the standard deviation in Figure 10 is too small to be read. Figure 10 will be modified to include CV instead of standard deviation. The modified figure is given below (Figure C4) followed by the discussion that we plan to include (please refer to our response to comment 11: 12-26 regarding the sample size)

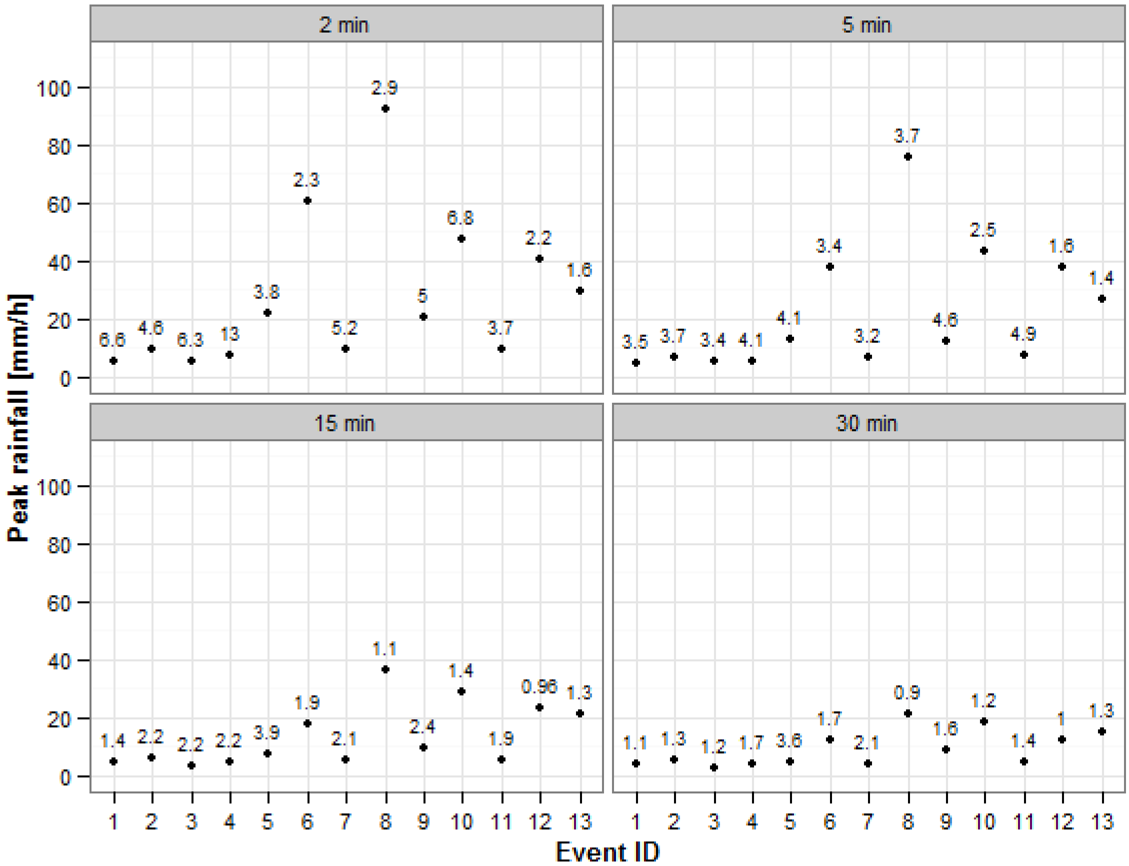

Figure C4 : Predictions of event peaks of AARI (indicated by points) together with labels indicating corresponding CV (%) values.

Rainfall event peaks are of significant interest in urban hydrology as most of the hydraulic structures in urban drainage systems are designed based on peak discharge which is often derived from peak rainfall. Hence it is important to consider the uncertainty in prediction of peaks of AARI. Figure 10 presents predicted peaks of AARI for all 13 events presented in Table 1, together with labels indicating corresponding CV (%) values. Since CV is a normalised measure it gives a clearer picture of behaviour of error against prediction. The peak intensities range from 6 mm/h to 92 mm/h at 2 min averaging interval and this range narrows down to 3 mm/h – 21 mm/h at averaging interval of 30 min as a result of temporal aggregation. Temporal aggregation from 2 min to 30 min also results in the *reduction of CV as expected. The highest CV at 2 min averaging intervals is 13% for event 4 and it gets reduced to 1.7% at 30 min averaging interval. But it can also be noted that events 5, 6, 8 and 11 show their highest CV at 5 min averaging interval and not at 2 min averaging interval. Tracking back these events, they indeed show more spatial variation over 5 min period compared to 2 min period around the peak.*

Although the sample size is not large enough to derive a firm conclusion about the behaviour of CV against the intensity ranges, there is a tendency that CV is higher when the predicted intensity is lower than 10 mm/h. This compliments what is observed from the variograms in Fig. 7, where below 10 mm/h the TB error becomes high. Hence it is expected to see a higher uncertainty characterised by CV at lower rainfall intensity (< 10 mm/h) especially at a temporal averaging interval of 2 min where the TB error is at its highest. But when the temporal averaging interval is 30 min where the TB error is at its lowest, the difference between CV for lower (< 10 mm/h) and higher (> 10 mm/h) intensity becomes smaller. At 30 min averaging interval the mean CV below and above 10 mm/h are 1.7 % and 1.2 % respectively, but they increase to 6.6 % and 3.5 % at 2 min averaging interval. The maximum CV at 2 min averaging interval are 13 % and 6.8 % for lower (< 10 mm/h) and higher (> 10 mm/h) rainfall intensity respectively. This is fairly high considering the required accuracy defined in standard guidelines of urban hydrological modelling practises. For example urban drainage verification guidelines (WaPUG, 2012) in UK set a maximum allowable deviation of 25 % to -15 % in peak runoff demanding more accurate prediction of rainfall which is the main driver of the runoff process.

[11: 12-26] What is the actual goal of that chapter? Has there any kind of significance testing be performed? As there is one very large CV value <10 mm (2 min) out of six, the comparison between the means might be biased. If this value was an outlier, would the result still be that considerable? There seem to be a tendency, but the sample size is very small and thus, there is a lot of uncertainty in this result.

<u>Reply:</u> *We plan to remove Figure 11 from the manuscript and discussions will then be based on revised Figure 10 (please refer our response to comment 10: 30). Although we do not think peak prediction at event 8 is an outlier, we accept the fact that the sample size is too small to derive a firm conclusion. This will also be mentioned clearly in the revised manuscript. We also performed the test without event 8, which results in the same trend. But the average CV for the range < 10 mm/hr is reduced to 5.3 % from 6.6% at 2 min averaging interval.*

We also tried to relax the condition of minimum rainfall yield from 10 mm to 5 mm to derive more events. This gave 19 events instead of 13. Similar analysis was carried out on these events and the results are presented in Figure C5 below,

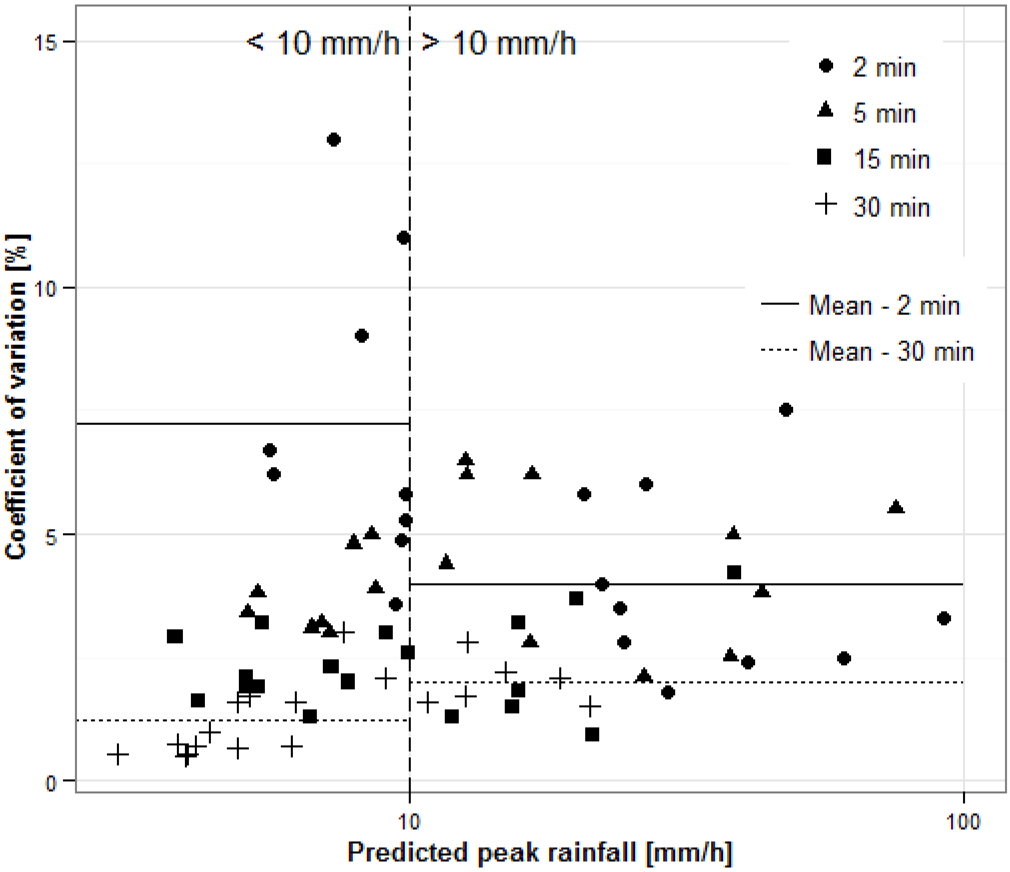

Figure C5: Coefficient of variation plotted against predicted peaks of AARI for the 19 events

The figure shows similar behaviour to Figure 11 where a higher uncertainty (characterised by CV) is seen for lower rainfall intensity (< 10 mm/h) at a temporal averaging interval of 2 min, where the TB error is at its highest. The trend is clearer this time as the highest three CV values at 2 min averaging interval correspond with lower intensity rainfall. The differences between average CV corresponding to lower (<10 mm/hr) and higher (>10 mm/hr) intensity rainfall are 7.2% and 4.0%, respectively. This is mainly due to the dominance of TB error corresponding to lower intensity rainfall at 2 min averaging interval. But when the temporal averaging interval is 30 min and the TB error is minimal this effect is not there anymore. In fact the average CV is slightly higher (by 0.8%) for lower intensity rainfall than for higher intensity rainfall. But the difference is too small to derive a conclusion for the 30 min averaging interval.

[11: 22] "This is fairly high: : :" How did you judge that? 25% to -15 % of the peak runoff is your reference. But what is the expected influence of the uncertainty of the rainfall when estimating the runoff from the precipitation? This is an important question for the judgement as one should know the necessary accuracy of the input precipitation and not the total uncertainty of the runoff.

Reply: *It will need a hydrologic modelling set up for this catchment to find out the exact influence of rainfall and rainfall uncertainty in urban runoff. This is beyond the scope of this study. Hence a simpler approach using Rational Formula to calculate the surface runoff is used here to form an idea of the influence. According to the rational formula (Q=kCIA) runoff has a linear relationship with rainfall intensity given a constant area and a constant runoff coefficient. Hence a 13% deviation in rainfall intensity would result in 13% deviation in surface runoff, which is high given the allowed limit of -15% to 25%, also not to forget other sources of uncertainty due to parameter and model structure. Furthermore, in several recent studies (Gires et al., 2012; Schellart et al., 2012) the effect* *of this small scale (< 1km) variability of rainfall on urban runoff peak has been proven to be significant.*

[11: 26] "Hence a better trade-off : : :" What does this actually mean? How could this be achieved? There is a link missing between the shown problems and the actual application. Later, in the conclusion, it becomes clearer, but in this chapter this sentence is kind of fuzzy.

Reply: *We accept that this sentence is not clear. This sentence will be removed from this section to avoid repetitive discussion of the same.*

[12: 6-10] It should be mentioned, that the result of "peak" intensities are explained here.

Reply: *Thank you for pointing out this. We will do so in the revised manuscript.*

[12: 24] ": : : radar measurements : : : would be much higher: : :" Please, give some references here.

Reply: *The following references will be added in the revised manuscript: (Seo and Krajewski, 2010; Villarini et al., 2008).*

[12: 27] You did not show explicitly the advantage of the paired rain gauges. Unless the improvements are shown in the main chapters, they should not be part of the conclusions.

Reply: *Thank you for this comment. This will be removed from the conclusion in the revised manuscript.*

Technical corrections

[2: 7] : : :where time series of areal : : : "are" needed.
[2: 27] : : : [8,9] : : : citations missing
[4: 9] "That" is because: : :
[8:5/8] Index Error: px should be pi
[9: 21] : : : "look" similar: : :

Reply: *Thank you for pointing out these errors. All will be corrected in the revised manuscript.*

References

Ciach, G. J. and Krajewski, W. F.: Analysis and modeling of spatial correlation structure in small-scale rainfall in Central Oklahoma, Adv. Water Resour., 29, 1450–1463, doi:10.1016/j.advwatres.2005.11.003, 2006.

Dirks, K. N., Hay, J. E., Stow, C. D. and Harris, D.: High-resolution studies of rainfall on Norfolk Island Part II: Interpolation of rainfall data, J. Hydrol., 208(3-4), 187–193, doi:10.1016/S0022-1694(98)00155-3, 1998.

Fankhauser, R.: Influence of systematic errors from tipping bucket rain gauges on recorded rainfall data, Water Sci. Technol., 37(11), 121–129, doi:10.1016/S0273-1223(98)00324-2, 1998.

Fiener, P. and Auerswald, K.: Spatial variability of rainfall on a sub-kilometre scale, , 859(February), 848–859, doi:10.1002/esp, 2008.

Gires, A., Onof, C., Maksimovic, C., Schertzer, D., Tchiguirinskaia, I. and Simoes, N.: Quantifying the impact of small scale unmeasured rainfall variability on urban runoff through multifractal downscaling: A case study, J. Hydrol., 442-443, 117–128, doi:10.1016/j.jhydrol.2012.04.005, 2012.

Habib, E., Krajewski, W. F. and Kruger, A.: Sampling Errors of Tipping-Bucket Rain Gauge

Measurements, J. Hydrol. Eng., 6(2), 159–166, doi:10.1061/(ASCE)1084-0699(2001)6:2(159), 2001.

Jaffrain, J. and Berne, A.: Quantification of the small-scale spatial structure of the raindrop size distribution from a network of disdrometers, J. Appl. Meteorol. Climatol., 51(5), 941–953, doi:10.1175/JAMC-D-11-0136.1, 2012.

Ly, S., Charles, C. and Degré, a.: Geostatistical interpolation of daily rainfall at catchment scale: the use of several variogram models in the Ourthe and Ambleve catchments, Belgium, Hydrol. Earth Syst. Sci., 15, 2259–2274, doi:10.5194/hess-15-2259-2011, 2011.

Schellart, A. N. A., Shepherd, W. J. and Saul, A. J.: Influence of rainfall estimation error and spatial variability on sewer flow prediction at a small urban scale, Adv. Water Resour., 45, 65–75, doi:10.1016/j.advwatres.2011.10.012, 2012.

Seo, B.-C. and Krajewski, W. F.: Scale Dependence of Radar Rainfall Uncertainty: Initial Evaluation of NEXRAD's New Super-Resolution Data for Hydrologic Applications, J. Hydrometeorol., 11(5), 1191–1198, doi:10.1175/2010JHM1265.1, 2010.

Sevruk, B. and Hamon, W. R.: International comparison of national precipitation gauges with a reference pit gauge, instruments and observing methods, Geneva., 1984.

Tao, T. A. O., Chocat, B., Liu, S. and Xin, K.: Uncertainty analysis of interpolation methods in rainfall spatial distribution–a case of small catchment in Lyon, J. Environ. Prot. (Irvine,. Calif)., 1(01), 50, doi:10.4236/jwarp.2009.12018, 2009.

Villarini, G., Mandapaka, P. V., Krajewski, W. F. and Moore, R. J.: Rainfall and sampling uncertainties: A rain gauge perspective, J. Geophys. Res. Atmos., 113(11), 1–12, doi:10.1029/2007JD009214, 2008.

Webster, R. and Oliver, M. a: Geostatistics for environmental scientists, Second edi., John Wiley & Sons, Ltd, West Sussex, England., 2007.