H. Müller (mueller@iww.uni-hannover.de)
Received and published: 12 Aug 2016

Complete author list: Müller, Hannes; Callau Poduje, Ana; Fangmann, Anne; Plötner, Stefan; Shehu, Bora; Uniyal, Bhumika
(all from the Institute of Water Resources Management, Hydrology and Agricultural Hydraulic Engineering, Leibniz Universität Hannover, Hanover, 30167, Germany)

This review results from six reviewers, all interested in the topic of the manuscript. Due to the number of six additional reviewers, the review is organized in major comments, suggestions and technical notes.

Brief summary: The manuscript deals with uncertainties resulting from upscaling of rainfall data. Upscaling hereby includes temporal aggregation as well as the determination of areal rainfall from point measurements. The topic is highly interesting and the investigation can be a good contribution to this field. However, we think that the manuscript can be improved significantly in the methods and the results part.

Authors: We thank the reviewers for the professional and thorough revision of this paper. We have attempted to address all the comments listed below

Major Comments:
Data:

p4 l7-21 The measuring period is quite short with two summer periods in 2012 and 2013. However, for such a dense network this is often the case. The two periods differ clearly and hence it is difficult to draw general conclusions from results.

Reply: We acknowledge that the data cover only 10 months, i.e. two summer periods in 2012 and 2013. However, for such a dense network this is often the case (Eg: 15 months - Ciach and Krajewski, 2006; 16 months - Jaffrain and Berne, 2012) due to practical difficulties and funding issues to maintain such networks for longer periods. The characteristics of our data are comparable with those studied in Ciach and Krajewski (2006) and Fiener and Auerswald (2008) as they also used rainfall data only from warm months to investigate spatial correlation structure of rainfall data.

We would also like to point out that the entire ten months of rainfall data, from 8 locations were used for the development of the geostatistical model in the form of variograms. 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.

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.

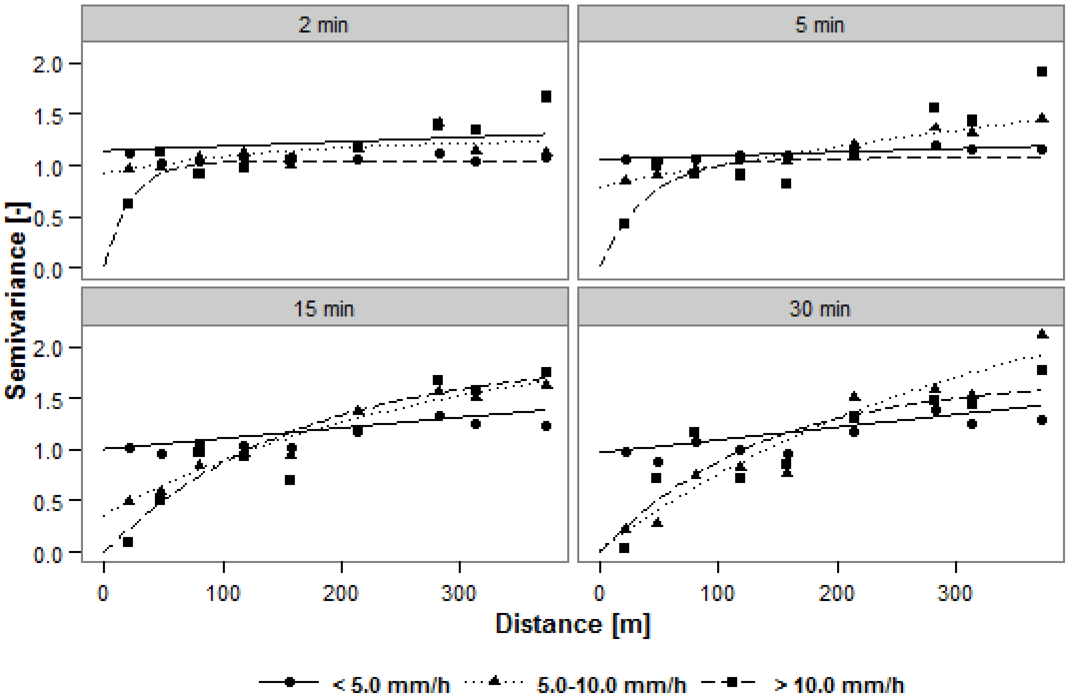

Figure C1: Calculated variograms for each temporal averaging interval and for each range of intensity within a temporal averaging interval using randomly selected 80% of the data from each subclasses

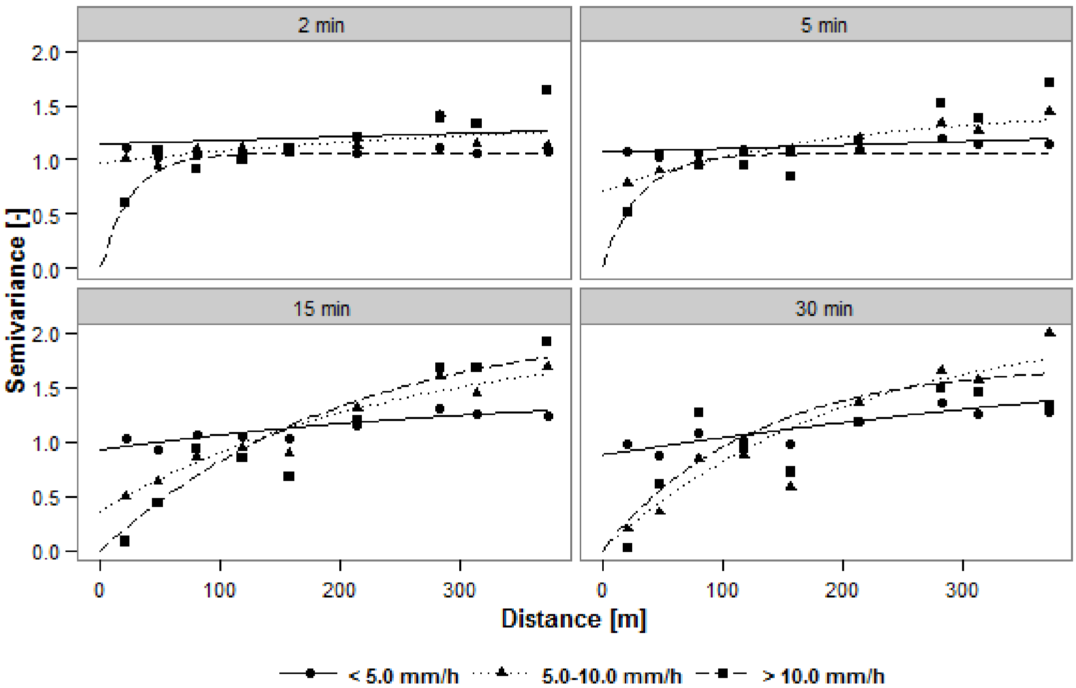

Figure 1: Calculated variograms for each temporal averaging interval and for each range of intensity within a temporal averaging interval

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.

Nevertheless we agree that different climatology of two years might influence the results and we will discuss this in the revised manuscript.

- From the two periods events are selected using a certain threshold. Why are events selected and why is the investigation not carried out for the whole observation periods?
The results shown later are not based on/related to events. Is there a need for the event separation?

Reply: *The event separation (ref Table 1) is used only for the analyses presented in Sections 4.2.3 and 4.2.4. We apologies if this was not clear in the manuscript and we will add some text in the revised manuscript to clarify this position.*

- A threshold of 10 mm network average rainfall depth and a minimum of 20 min rainfall duration were chosen for the event selection. How have these thresholds been chosen? The chosen thresholds can lead to exclusion of convective events with high rainfall amounts at one station, but no rainfall at the other stations. This is also indicated by the durations of the resulting events, ranging from 1.5 h to 11.4 h, which are more typical for stratiform events and not convective ones. Indeed these convective ones are crucial for urban hydrology and the resulting uncertainty in spatial upscaling is very high. Have convective events be excluded from the investigation by the chosen thresholds?

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

As we mentioned in the above response the entire ten months of rainfall data from 8 locations were used for the development of the geostatistical model in the form of variograms. Hence no data are excluded from the investigation. The event separation (ref Table 1) is used only for the analyses presented in Sections 4.2.3 and 4.2.4.

- How is the network average rainfall depth calculated for the event selection? In the introduction several methods are discussed. Is ordinary kriging applied here?

Reply: *The network average rainfall depth is calculated using arithmetic mean of the rainfall depths of 8 stations over the network. Ordinary kriging is not applied here.*

Methodology:
p5 l12 What is pooled – events or single time steps? In the text before, time steps (p5 l1) and events (p5 l6-7) are mentioned. If time steps are pooled (and not events), later for one event different variograms may be used due to different intensities of the single time steps in the event, right?

Reply: *Time steps, not events, were pooled to increase the number of spatial pairs. Please refer to p4 | 26 – p5 | 3 in the manuscript for a detailed explanation. P5 l 6-7 is just a part of an explanation on how the intensity classes are chosen.*

Different variograms corresponding to different intensity classes can be used for a single event as a single event can contain a range of intensity values which fall into different intensity classes.

p7 l14 What is spatial stochastic simulation? All results are based on this method, so an explanation in the text is necessary (not only a reference). Is it applied as a subsequent step to the ordinary kriging or instead of the ordinary kriging? What is the stochastic simulation based on?

Reply: Since spatial stochastic simulation is robust to nonlinear data transformation it was used as an alternative to ordinary (block) kriging. 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.

Results:

p8 l25 The nugget-to-sill ratio is interpreted as measurement error, decreasing with an increasing temporal aggregation. The movement of events is ignored, which could significantly contribute to this ratio. With 2 min time steps, the event has reached one (pair) of the gauges, after 30 min all gauges are influenced by the event. This explanation should be implemented. Is it possible with other measurements (wind velocity,: : :) to exclude / quantify this effect? Also, can the whole nugget effect be described as measurement error from the author's point of view?

Reply: *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.

Since TB error is sampling related, other measurements (wind velocity, etc) cannot help quantifying or reducing this error.

Conclusions:
General comment: Some conclusions are trivial (e.g. the intensity becomes less with increasing averaging interval), and there could be more conclusions out of the investigation.
What is the message to the urban hydrologic modelers? How can this uncertainty be involved in the calibration process/result discussion? Is the uncertainty greater/smaller than other uncertainties in urban hydrological modeling? Is it useful to take this uncertainty into account, if others are higher? What are results of other investigations concerning areal rainfall uncertainties? Is it assumed, that the uncertainty increases with increasing area sizes in the lumped model? What is the recommendation for rain gauges number per square kilometer from this investigation? How sensitive are the results, if the station density/combination of stations is changed in the investigation? The measurement set-up is quite dense. Can general conclusions be drawn to less dense networks (and how)? Can the results be validated with an urban hydrologic model?

Reply: Please refer to our response on specific comment about trivial conclusion p12 l10-13.

Thank you for suggesting more conclusions. Please find below our response.

….What is the message to the urban hydrologic modelers? How can this uncertainty be involved in the calibration process/result discussion? Can the results be validated with an urban hydrologic model?

We believe the summary of our finding (p12|1-18) are all in interest to urban hydrologic modellers. In addition, results from this study can be used for uncertainty analyses of hydrologic and hydrodynamic modelling of similar sized urban catchments as it provides information on uncertainty associated with rainfall estimation. This estimate of uncertainty in combination of estimates of uncertainty due to model structure and model parameter will help to indicate the significance of rainfall uncertainty. This estimate of the relative importance of uncertainty sources can help to avoid false calibration and force fitting of model parameters (Vrugt et al., 2008). We already discussed this briefly in our manuscript (p12|29-35. it is a challenging task to validate these results using hydrological modelling as such validation also needs estimations of other sources of uncertainty (structural and parameter) as well as overall uncertainty in the model output. We will include this in the revised manuscript.

….What are results of other investigations concerning areal rainfall uncertainties?

We think this is something that should be included in the Discussion rather than the Conclusion. We already discussed some other related studies (Ciach and Krajewski, 2006; Krajewski et al., 2003; Villarini et al., 2008) when comparing our results with those of other studies.

….Is the uncertainty greater/smaller than other uncertainties in urban hydrological modeling? Is it useful to take this uncertainty into account, if others are higher? Is it assumed, that the uncertainty increases with increasing area sizes in the lumped model?

Individual uncertainties will be catchment specific, but It is still useful to take this uncertainty into account because only by quantifying it will you know if it is larger or smaller than others. We have not assumed that uncertainty in rainfall increases with increasing area size and since our scope does not cover this we cannot draw any conclusion on this issue.

…What is the recommendation for rain gauges number per square kilometer from this investigation? The measurement set-up is quite dense. Can general conclusions be drawn to less dense networks (and how)?

We think such a hard and fast rule on number of data points cannot be derived for this methodology. Because, like any other geostatistical interpolation method, the efficiency of this method also heavily depends on reliable estimation of the geostatistical model (variogram). Hence it basically comes *down to the question of whether a given rain gauge network can produce a meaningful variogram? As we mentioned in the manuscript, Webster and Oliver (2007) suggested around 100 measurement points to calculate a geostatistical model. But there is no hard and fast rule to define minimum number of bins and the number of samples for each bin to produce a reliable variogram. Further, since pooling of repeated measurements would produce a multiplication of spatial lags, the length of the available data would also play a role in deciding the number of measurement locations.*

….How sensitive are the results, if the station density/combination of stations is changed in the investigation?

Leaving one station out would definitely affect the results. First it will reduce the accuracy of the estimation of the variograms as the number of spatial lags per time instant would come down to 21 from 28. But the further effect of leaving one station out needs to be analysed in detail to see how it effect the uncertainty in the estimation of areal average rainfall intensity. In the manuscript we have not included such sensitivity analyses considering the direct relevance to the main scope of this study and the work load required to perform such an analyses.

Suggestions:
Title: The title doesn't fit to the content of the manuscript. There are no urban hydrological models applied. Also, if no kriging is applied (not sure about that, see major comment p7 l14), it's not an geostatistical upscaling. The title "Estimation of uncertainties from spatial and temporal upscaling on an urban scale" is therefore misleading.

<u>Reply:</u> *Please refer to our response to comment p7 l14 for the explanation on spatial stochastic simulation, which is a geostatistical method. We believe it is quite clear from the methodology that this study uses geostatistical upscaling (e.g. the derivation of variograms, the use of spatial stochastic simulation). Further, we think the spatial extent (0 – 400 m) and the temporal averaging intervals (2 min -30 min) considered in this study are in interest of urban hydrology. Also we think the uncertainty estimation in areal rainfall would be more useful for the hydrology community working on uncertainty. These are the main reasons why the paper was oriented towards the urban hydrology community. Hence we don't think the tittle is misleading.*

Introduction:

p2 l19-25 There exist other methods for the estimation of uncertainties (bootstrapping,: : :), which should be mentioned in this context. Indeed, a focus should put on these methods, their comparisons and a reasonable decision for the applied method should be given at the end.

<u>Reply:</u> *Thank you for the suggestion. We would like to point out that we did not aim to discuss all uncertainty methods that are available for hydrological applications. Rather we wanted to keep the Introduction mainly focused on the methodology that we adapted for this study. Hence we followed the below order in our introduction*

- *Lumped hydrological models need spatial average rainfall over catchments*
- *Focus is on rainfall observations at points*
- *Thus point observations need to be scaled up*
- *Review of existing methods that do this*
- *Disadvantages of these methods*
- *Solution that does not have these disadvantages is to take geostatistical approach*
- *The main challenges with geostatistical approach and how it can be dealt*

And we quoted the most relevant studies wherever necessary. The paper did not aim to compare uncertainty methods but to examine the levels of uncertainty in rainfall intensities.

p2 l8-19 In the introduction a number of interpolation methods are mentioned, which are not used afterwards in the investigation. They could be left out.

Reply: Please refer to our response to specific comment p2 l19-25

Methodology:
p5 l11 How have the thresholds for the pooling been chosen?

Reply: The maximum threshold value was limited to 10mm/hr to have enough time instants for the highest range (i.e. > 10 mm/hr) to produce stable variograms even at 30 min temporal averaging interval. We then decided to divide 0 – 10 mm/hr in to two equal subclasses (i.e. < 5mm/hr and 5-10 mm/hr). This gave us three subclasses which we thought a reasonable number given the length of the data, work load and computational demand.

p6 l2 Methods and results (Fig. 6) are mixed.

Reply: We accept that in Fig 6 part of a result is presented, but we believe that this combined figure helps to explain step 3 clearly and consequently makes it easier for the reader to understand. Further this is not a major result, but just an outcome of one of the steps.

p6 l10 The method of NST could be explained briefly.

Reply: Thank you for the suggestion. Section 3.3 already briefly explains NST with the basic theory and literature where detailed description of NST including the steps involved can be found. We will add some more explanation in the revised manuscript.

p6 l17 Step 4 is not a step, only a description, and can be moved to step 5.

Reply: Step 4 is one of the major steps of this study. It involves the construction of variograms. The reviewer is kindly requested to refer to Section 3.4 in the manuscript for further explanation.

p7 l4 explanation for q is missing

Reply: Thank you for pointing out this. q is the total number of measurement points. We will add this in the revised manuscript.

p8 l9 With the standard deviation and the mean of areal rainfall intensities the restandardisation is carried out. For the former standardisation the standard deviation and the mean of point values have been used (since it is not clear, what has been pooled (see comment p5 l12), we assume time steps). Shouldn't be standard deviation and mean for standardization and re-standardisation be from the same type, so either from area or point values?

Reply: Time steps, not events, were pooled to increase the spatial pairs. Hence mean and standard deviation from each time step is used for standardisation (ref Equation 1) as well as inverse standardisation.

Results:

p8 16-17 The nugget is interpreted as spatial rainfall variability or measurement error.
The network offers the great possibility to have rain gauge pairs with distances of 1 m.
Measurement errors have been excluded before by the paired measured time series.
So the spatial variability can be shown for these small distances, or why should this not be possible?

Reply: *Although paired gauges are used for efficient quality control, it cannot avoid sampling related error of tipping buckets (Habib et al., 2001). For further explanation on why we associate nugget effect with sampling related error of tipping buckets, please refer to our response to specific comment p8|25.*

p8 l28 Since all errors have been excluded under the usage of the paired time series, the word TB error is somehow misleading. "Sampling error" could be more appropriate.

Reply: Thank you for the suggestion. But this sampling error is only associated with tipping bucket type rain gauges. That is why we preferred to call it TB error. Further the same term was used in a previous study (Habib et al., 2001)on sampling errors of tipping bucket type rain gauge which was quoted in our discussion. Hence we wish to use the same term to be consistent with previous studies.

p10 l3 Showing the CV would be more effective then showing the standard deviation in Fig. 8. An increasing of the standard deviation with an increasing intensity is trivial
(which is even stated on p11 l7-10). Also, a logarithmic plot would be useful.

Reply: Thank you for the suggestion. But since the main reviewers did not think that this figure should be changed, we will wait for the editor's decide to decide if any change is required in this this figure.

p10 l31-32 Design on peak rainfall intensity: The intensity AND the duration are important and both are used for the dimensioning of e.g. a sewer system.

Reply: We agree with the reviewer that both peak and duration are important in the design of urban hydraulic structures. We will correct the sentence in the revised manuscript.

p11 l1 Fig. 10 Maybe it would be useful to use violin plots instead of only the standard deviation to show the uncertainty.

Reply: Thank you for the suggestion. But based on one of the assigned reviewer's comment we will replace this figure with Figure C2 below. This plot includes labels of CV values instead of error bars to make it easy to read.

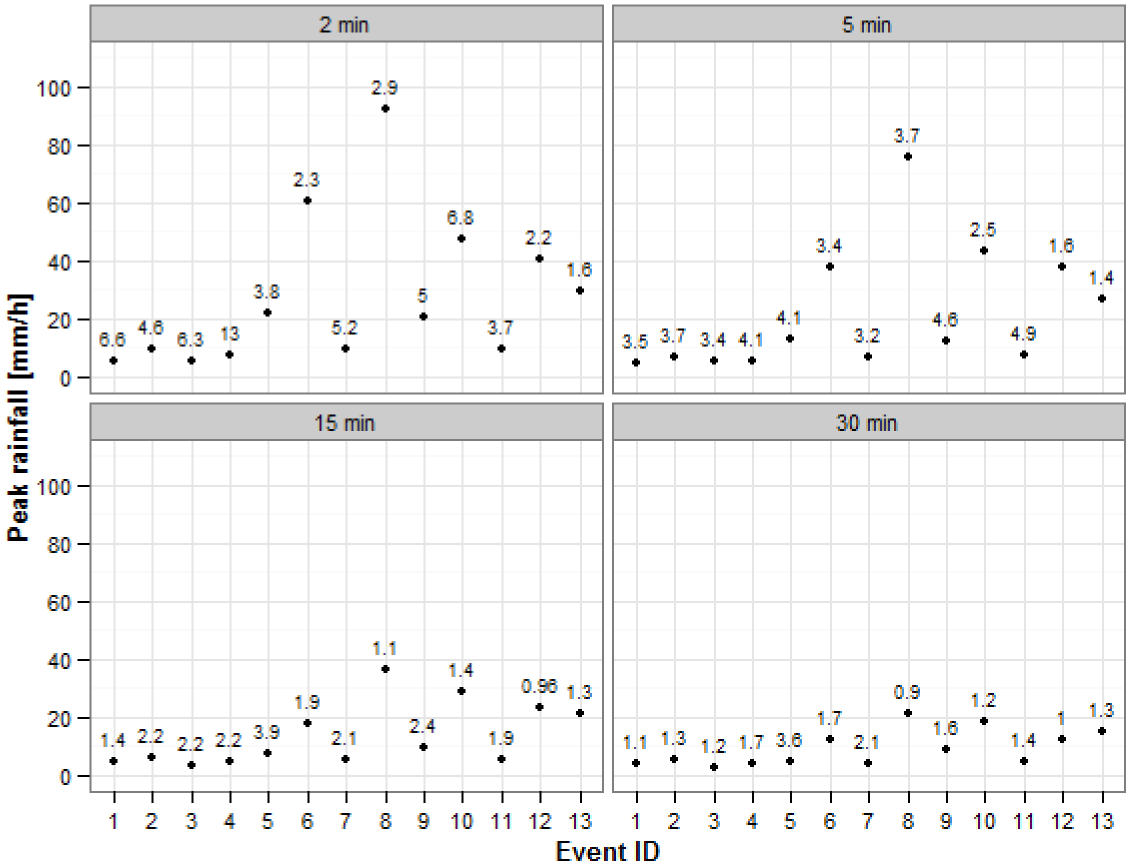

Figure C2: Predictions of event peaks of AARI (indicated by points) together with labels indicating corresponding COV (%) values

p11 l12 Fig. 11 The readability of the figure could be increased by colors and/or drawing only contours, not filling them. Showing the means as functions, not as fixed values, would give better conclusions. The high mean for 2 min, <10 mm/h is caused by only one extreme CV (_13 %) and is not representable.

Reply: Thank you for the suggestion. We wanted to avoid the use of colours in the plots whenever possible. But we agree empty markers would improve the readability and we will use them in the revised manuscript.

Although we don't think peak prediction at event 8 (the extreme value) is an outlier, we accept the fact that the sample size is too small to derive a firm conclusion. This will be mentioned clearly in the revised manuscript. We also tested without event 8. and the trend remains the same, the average CV for the range < 10 mm/hr is reduced to 5.3 % from 6.6% at 2 min averaging interval when test 8 is excluded..

We also tried to relax the condition of minimum rainfall yield from 10 mm to 5 mm to derive more number of events. This gives us 19 such events. Similar analysis is carried out on these events and the results are presented below,

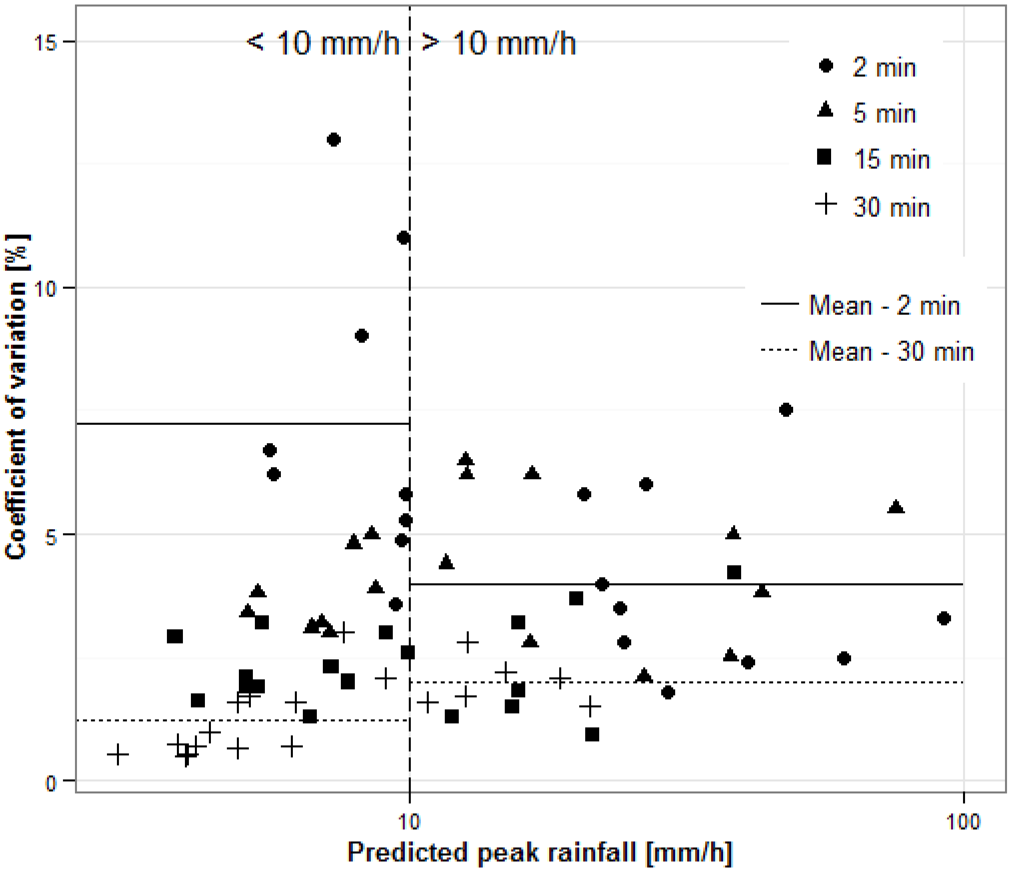

Figure C3: Coefficient of variation plotted against predicted peaks of AARI for the 19 events

It shows the similar behaviour to figure 11 where a higher uncertainty (characterised by CV) is seen for lower rainfall intensity (< 10 mm/h) at 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 belongs to lower intensity rainfall. The difference between average CV corresponds to lower (<10 mm/hr) and higher (>10 mm/hr) intensity rainfall is 7.2% and 4.0% respectively. We believe this is mainly due to the dominance of TB error corresponds to lower intensity rainfall at 2 min averaging interval.

But when the temporal averaging interval is 30 min where 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 higher intensity rainfall. But the difference is too small to derive a conclusion for 30 min averaging interval.

Conclusion:
p12 l10-13 Decreasing peaks due to aggregation in time is trivial and not a conclusion.

Reply: *We agree with the reviewer that decreasing peaks due to aggregation in time is trivial. But our conclusion gives a quantification of this reduction to show its significance. It helps to subsequently discuss the trade-off between temporal resolution and accuracy in rainfall prediction, which is the main aim of that bullet point.*

p12 l32-33 “This information can help to avoid false calibration and force fitting of model parameters” It remains unclear, how the result of the investigation can be used for the avoidance of the before mentioned issues.

Reply: Results from this study can be used for uncertainty analyses of hydrologic and hydrodynamic modelling of similar sized urban catchments as it provides information on uncertainty associated with rainfall estimation. This estimate of uncertainty in combination of estimates of uncertainty due to model structure and model parameter will help to indicate the significance of rainfall uncertainty. This estimate of the relative importance of uncertainty sources can help to avoid false calibration and force fitting of model parameters (Vrugt et al., 2008).

We will include this in the revised manuscript.

Technical notes:

p7 l5 "locationsxl" to "locations xl" and "locationsx0" to "locations x0"

Reply: Thank you for pointing it out. It will be corrected in the revised manuscript.

p8 l5 Eq. (6) "pi" and not "px", also the division by "m" is missing – Since this is a simple equation, it could left out, also Eq. (7)

Reply: Thank you for pointing it out. It will be corrected in the revised manuscript. Although these equations are well-known in general, given the context of the application, we think it would help the reader to understand the step 8 clearly.

p8 l6 Eq. (7) The term under the root has to be squared.
Reply: Thank you for pointing it out. It will be corrected in the revised manuscript.

p8 l25 "nugget-to-still ratio" to "nugget-to-sill ratio" (several times)
Reply: Thank you for pointing it out. It will be corrected in the revised manuscript.

p9 l3 "in their study found" to "found in their study"
Reply: Thank you for pointing it out. It will be corrected in the revised manuscript.

p9 l15 "(2003) in their" to "(2003) found in their"
Reply: Thank you for pointing it out. It will be corrected in the revised manuscript.

p10 l11-13 In Fig. 9 event 10 is shown, not event 11 (regarding to Fig. 10).
Reply: Thank you for pointing it out this error. It will be corrected in the revised manuscript.

p12 l22 "methods to a certain extent" – fuzzy phrase
Reply: //The pooling procedure used in this study makes use of the continuous measurement of rainfall and helps provide a solution to meet the data requirements for geostatistical interpolation methods to a certain extent.//
By the term 'certain extent' we wanted say pooling can only partially solve the problem of scarcity in measurement points as it does not produce any new spatial lags, but only extends the information for existing lags. We agree it was not very clear and will reformulate the sentence in the revised manuscript.

Reference

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.

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

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.

Krajewski, W. F., Ciach, G. J. and Habib, E.: An analysis of small-scale rainfall variability in different climatic regimes, Hydrol. Sci. J., 48(2), 151–162, doi:10.1623/hysj.48.2.151.44694, 2003.

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.