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. We apologise if it was not clear in the manuscript that the entire 10 months of data was used and we have included additional text in the revised manuscript to clarify this. (Revised manuscript [4: 27-29] and [5:9])

Further a discussion on the size of the data used in this study has been added to the revised manuscript (Revised manuscript [10:30 – 11:11])

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 have addressed the concerns stated by the referee. Superfluous sentences have been removed and some rewritten to reduce the attention paid to trivial results improve clarity and discuss the consequence of the analysis carried out in the paper. Details of these modifications are described below.

Trivial results removed, or reduced:

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

Improvement of clarity and language:
The following parts have been modified to give a clearer explanation, as indicated in specific comments:
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 addressed these challenges which include the use of repetitive rainfall measurements (pooling) to increase the number of observations used for variogram estimation.*

- *We used 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 defined intensity classes and derived different geostatistical models (variograms) for each class. 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 extent.*

- *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 on measurement error mainly focusing on tipping bucket errors has been added to our revised manuscript (Revised manuscript [2:22-26]).

[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 has been 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?

<u>Reply:</u> This section has been modified in the revised manuscript to give more precise information on how the quality control was carried out using a paired setup (Revised manuscript [4: 8-18]).

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

<u>Reply:</u> Thank you for pointing out this out. This has been 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?

<u>Reply:</u> 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. Fig. 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. We have included additional text in the revised manuscript to clarify this. (Revised manuscript [4: 27-29] and [5:9])

[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?

Reply: *We agree with the concerns expressed by the reviewer. This is the reason why we introduced step 2 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 between Fig. C1 and Fig. 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 Fig. C1 (in our reply to comment 8:16), where variograms are produced for narrower subclasses, we conclude that the variograms shown in Fig. 7 are good representations of the average spatial variability conditions for corresponding intensity classes.

We discussed this issue in the revised manuscript (Revised manuscript [11:12 – 26])

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

Reply: *We agree with this comment and we revised this sentence in the revised manuscript (Revised manuscript [5:30 -6:3]).*

[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.

Reply: *We agree with this comment and we revised this sentence in the revised manuscript (Revised manuscript [5:30 -6:3])*

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

Reply: *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 has now been 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. This figure has now been revised (Revised manuscript: Fig. 5).*

[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]).

Reply: 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 have now modified the text (Revised manuscript [7:10-11])

[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 have modified this section in the revised manuscript (Revised manuscript Section 3.5).

[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)
Reply: 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. We agree it was not clear from the text. Now this section has been revised (Revised manuscript [7:22-8:6])

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

Reply: 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 have modified Section 3.5 to give a clearer explanation of spatial stochastic simulation (refer comment [6:29-7:22]).

[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])

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

1. *Pooling of time instants:*
   We do not think pooling is the reason for the nugget effect. Note that we did not pair rainfall measurements from one time instant with rainfall measurements from another time instant. We only pooled sample variograms from different time instants. But that cannot cause an additional 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.

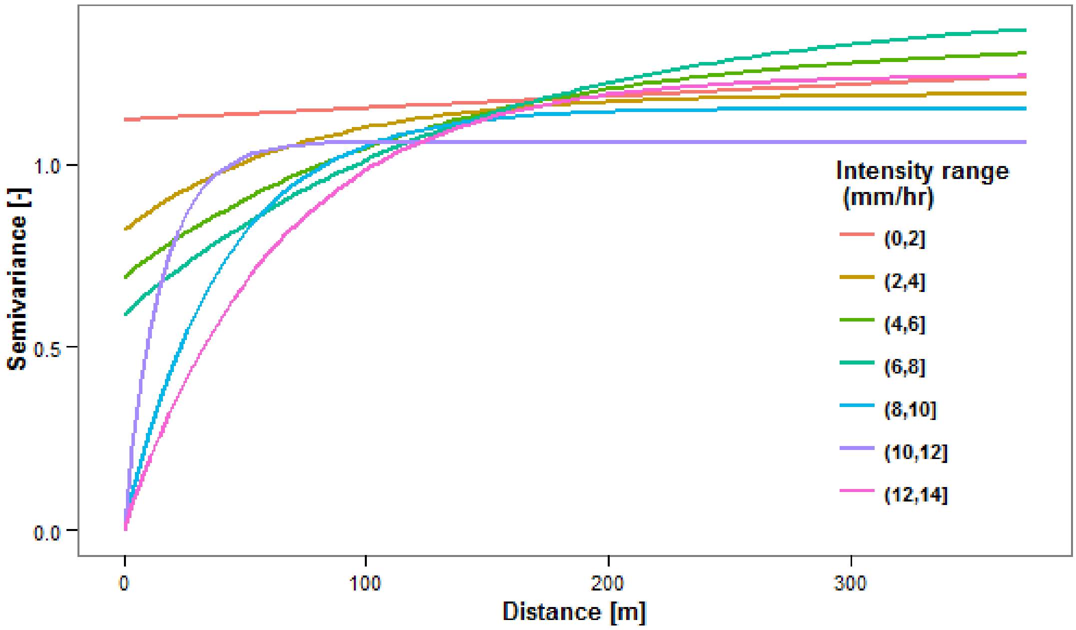


Fig. C1: Calculated variograms for 5 min averaging interval and for a narrower range of intensity.

Furthermore, Fig. C1 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 within a pooled subset is stronger. Hence the effect of variation caused by pooling is reduced. Comparing this figure with Fig. 7 not only shows the behaviour of the nugget effect against the intensity as seen from Fig. 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.

We have added the Fig. C1 in the revised manuscript and discussed the effect of pooling (Revised manuscript [11:12-26)

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?

Reply: 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. We modified this section to address this and to interpret the variograms better (Revised manuscript section 4.1).*

[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. We agree that this is a logical consequence, but this result leads to the discussion of trade-off between timescale temporal resolution and accuracy in rainfall prediction. Hence we decided to keep this, but we have now reformulated this discussion in the revised manuscript to emphasise on the fact that this is a logical consequence of temporal aggregation (Revised manuscript [13:2-6]).*

[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. Fig. 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.

<u>Reply:</u> *Sections 4.2.3 and 4.2.4 have been combined into one section in the revised manuscript as suggested by the reviewer. We agree that the standard deviation in Fig. 10 is too small to be read. Fig. 10 has been modified to include CV instead of standard deviation.(Revised manuscript Fig. 11, Section 4.2.3)*

Please refer to our response to comment 11: 12-26 regarding the sample size.

[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 removed Fig. 11 from the manuscript and discussions are now based on revised Fig. 10 (please refer to our response to comment 10: 30).*

We accept the fact that the sample size is too small to derive a firm conclusion. To comment on prediction uncertainty against intensity with a larger sample size, we modified the Fig. 8 (Revised manuscript Fig. 9) to show CV instead of standard deviation. Now it can be seen from modified Fig. 8 that there is a clear trend of increasing CV with increasing AARI. We also modified the discussions in the sections 4.2.1 and 4.2.3 accordingly.

Further although we do not think peak prediction at event 8 is an outlier, 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.

[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.

<u>Reply:</u> *A 13% uncertainty in rainfall will result in a similar level of uncertainty in runoff prediction for a completely impervious surface according to the well-established rational formula (Viessman and Lewis, 1995) which is still widely used for estimating design discharge in small urban catchments. .*

This uncertainty 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.

We have added some more text in the revised manuscript to clarify this (Revised manuscript [13:32-14:2])

[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 has now been reformulated in the revised manuscript (Revised manuscript 13:2-6)

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

Reply: Thank you for pointing this out. We have done so in the revised manuscript.

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

Reply: The following references have been added:
(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 has been 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 have been 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.

Anonymous referee #2
Received and published: 15 Jul 2016

General
This paper suggests a method to upscale rainfall intensity from point scale to catchment scale. The authors suggest a Kriging-based stochastic method for this upscaling; a method that allows an uncertainty estimation of the areal rainfall. I found the method suggested by the authors very interesting. I think that it will be of interest mainly for the hydro-meteorologist community (dealing with weather radar estimations) rather than for the urban hydrologist community. Main problem in the paper is the short period of observation (two seasons) that is expressed in a low confidence in the presented results.
I found that some key papers dealing with dense rain-gauge networks and rainfall variability in the past were not mentioned and that some trivial aspects discussed in the past are repeated in here. I would suggest the authors to thoroughly revise the paper as follow: make the upscaling method as the main focus of the paper, explain it with much further details and with a much clearer language. Use the data you have from the dense rain-gauge network as a case study to demonstrate how you can upscale rainfall for the catchment / weather radar scale and show the advantages of estimating uncertainties with the method you are suggesting. Please find below my specific comments, following by some general comments.

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

…Main problem in the paper is the short period of observation (two seasons) that is expressed in a low confidence in the presented results…

<u>Reply:</u> Please refer to our detailed response to the reviewer's specific comment on data used in the study [3:19-20]

…I found that some key papers dealing with dense rain-gauge networks and rainfall variability in the past were not mentioned

<u>Reply:</u> Please refer to our detailed response to reviewer's specific comment on literature used in the study [2:19-21]

….and that some trivial aspects discussed in the past are repeated in here.

<u>Reply:</u> We attempted to deal with the concerns stated by the reviewer. Several sentences have been removed and some rewritten to try and reduce the level of trivial results. They are listed below.

[5: 11-13] "As expected, with increasing temporal averaging the number of time instants t reduces. Fig. 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" has been reformulated to emphasise on the fact that these observations are either natural or logical consequence. (Revised manuscript [5:30 – 6:1])

[10: 16] "It is expected that with increasing temporal averaging interval the local minima and maxima of AARI get smoothed out. 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." has been reformulated to emphasise on the fact that this observations is a logical consequence of the aggregation process. (Revised manuscript [13:2-6])

… I would suggest the authors to thoroughly revise the paper as follow: make the upscaling method as the main focus of the paper, explain it with much further details and with a much clearer language. Use the data you have from the dense rain-gauge network as a case study to demonstrate how you can upscale rainfall for the catchment / weather radar scale and show the advantages of estimating uncertainties with the method you are suggesting.

Reply: *Thank you for the suggestion. With all due respect to the reviewer's suggestion, we think that the manuscript is already heavily focused on methodology with a dedicated section which covers around 35% of the manuscript (pagewise). Further, to enable the reader to follow the methodology more easily, we explained it with a step by step procedure for a general case of estimating uncertainty in upscaling of point rainfall data. We tried to keep the methodology as general as possible while also providing enough detail on how each step is applied in the case of the Bradford case study. Furthermore, we believe that introducing the data before the methodology enables the reader to follow the methodology more easily as some of the steps require a pre-introduction to the data to explain why such step is needed. Hence we feel that the structure of the manuscript follows a logical work flow.*

Nevertheless we accept that some part of the manuscript needs more explanation, especially the spatial stochastic simulation methodology. We also agree that the language could be clearer throughout the manuscript. Hence we modified several sections/ parts of sections including the following based on reviewer's specific comments:

Quality control of rainfall data using paired gauge set up: please refer to our response to reviewer's specific comment [4:1-5]

Spatial stochastic simulation: please refer to our response to reviewer's specific comment [7:13-22]

Specific comments

[Page:Lines]

[2:4-6] Please support this statement with a reference.

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

[2:19-21] Since you are dealing with rainfall uncertainty for small domains, taking into consideration the rainfall spatial and temporal variability, I am strongly recommend you to check also the following papers that were published in the recent years, which I think you will find them all relevant to your study:

Peleg N, Ben-Asher M, Morin E. Radar subpixel-scale rainfall variability and uncertainty: lessons learned from observations of a dense rain-gauge network. Hydrol. Earth Syst. Sci. 2013 Jun 14;17(6):2195-208.

Krajewski, W. F., Kruger, A., and Nespor, V.: Experimental and numerical studies of small-scale rainfall measurements and variability, Water Sci. Technol., 37, 131–138, doi: 10.1016/s0273-1223(98)00325-4, 1998.

Pedersen, L., Jensen, N. E., Christensen, L. E., and Madsen, H.: Quantification of the spatial variability of rainfall based on a dense network of rain gauges, Atmos. Res., 95, 441–454, doi:10.1016/j.atmosres.2009.11.007, 2010.

Fiener, P. and Auerswald, K.: Spatial variability of rainfall on a sub-kilometre scale, Earth Surf. roc. Land., 34, 848–859, doi:10.1002/esp.1779, 2009.

Seo, B. C. and Krajewski, W. F.: Investigation of the scale dependent variability of radar-rainfall and rain gauge error covariance, Adv. Water Resour., 34, 152–163, doi:10.1016/j.advwatres.2010.10.006, 2011.

Gebremichael, M. and Krajewski, W. F.: Assessment of the statistical characterization of small-scale rainfall variability from radar: Analysis of TRMM ground validation datasets, J. Appl. Meteorol., 43, 1180–1199, doi:10.1175/1520-0450(2004)043<1180:aotsco>2.0.co;2, 2004.

Ciach, G. J. and Krajewski, W. F.: On the estimation of radar rainfall error variance, Adv. Water Resour., 22, 585–595, doi:10.1016/s0309-1708(98)00043-8, 1999.

Reply: Thank you for suggesting the above papers. We wanted to keep our 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 a case where the input data are rainfall observations at points*
- *Thus point observations need to be scaled up*
- *Review of existing methods that can do this*
- *Disadvantages of these methods*
- *Solution that does not have these disadvantages is to take a geostatistical approach*
- *The main challenges with geostatistical approach and how these can be dealt with*

We quoted the most relevant studies wherever necessary. We accept that there are many other studies like the ones mentioned by the reviewer which can be relevant because of the similar spatial extent of the rainfall data. But most of these publications are based or partly based on radar data (areal rainfall data) and outside the main scope of our study, which is upscaling of point rainfall data and estimating associated uncertainty. Discussion of the literature where both radar data and point rainfall are used together to compare and/or analyse spatial correlation is slightly out of the context and might lead to a very lengthy Introduction and might also confuse the readers even if the spatial extent of interest is the same. Therefore to keep the Introduction to the point and concise and focused on our objectives (upscaling of point rainfall data and associated uncertainty) we have not included literature which are completely/partly based on radar data.

However, from the reviewer's suggestion we found the following papers directly relevant to our study. We thank the reviewer for suggesting these papers. We have discussed these papers in appropriate sections of the revised manuscript. We summarise how these papers are related to our study below.

Pedersen, L., Jensen, N. E., Christensen, L. E., and Madsen, H.: Quantification of the spatial variability of rainfall based on a dense network of rain gauges, Atmos. Res., 95, 441–454, doi:10.1016/j.atmosres.2009.11.007, 2010.

The aim of this paper is to quantify the uncertainties of using a single rain gauge to represent the rainfall over a 500 × 500 m area. A field experiment placing nine 0.2 mm tipping bucket type rain gauges within an area of 500 × 500 m, each representing one-ninth of the area, was used to address the issue. The variability of rainfall is studied and uncertainty in areal rainfall is estimated for different time scales. Although this study uses a simpler approach to estimate the uncertainty, results are still comparable to our study.
(Revised manuscript [12:7-11])

Fiener, P. and Auerswald, K.: Spatial variability of rainfall on a sub-kilometre scale, Earth Surf. roc. Land., 34, 848–859, doi:10.1002/esp.1779, 2009

In this study, a network of 13 tipping-bucket rain gauges was operated on a 1·4 km2 test site in southern Germany for four years to quantify spatial trends in rainfall depth, intensity, erosivity, and predicted runoff. Their data is comparable to ours as they also used summer half-year data for their analyses. Although they did not calculate any uncertainty in areal rainfall estimation, their analyses on spatial trend against temporal averaging interval could be of interest to our study. One of their conclusions suggests that in the longer term there is no difference in rainfall depth within the test site, but in short-time periods or for single events the assumption of spatially uniform rainfall is invalid on the sub-kilometre scale, This complements one of the findings from our study.
(Revised manuscript [10:4-6], [10:34-11:1])

[2:27] Reference typo.

<u>Reply:</u> *The following references have been added in the revised manuscript.*
(Ly et al., 2013; Mair and Fares, 2011)

[2:32] I would even claim that it is rare to find locations where the rainfall is normally distributed.

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

[3:3] word "often" can be deleted.

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

[3:19-20] This is a very short period of observation, and winter rainfall is not represented at all. How does it affect your results? Moreover, in [4:7-12] you mention the large difference between the two years of observation. This imply that the climatology was different between the two years and therefore I would expect that it will somehow influence on the rainfall spatial correlation. With only two years, the variability expected for the spatial rainfall structure cannot be represented and this should be discussed.

<u>Reply:</u> *We acknowledge that the data cover only 10 months, i.e. two summer periods in 2012 and 2013, but however our geo statistical models (based on which further results are produced) are stable. This has now been discussed in detail in our revised manuscript. (Revised manuscript [10: 30-11:11])*

[Fig. 1] The recommendation is to mount rain-gauges elevated at 1.2 m above ground, where here the gauges seem to be placed directly on ground level (roof top). I wonder how this affects rainfall intensity estimations.

<u>Reply:</u> *We understand that different guidelines suggest different elevations when it comes to height of a rain gauge from the surrounding ground level. In our case we followed the standard UK practice (<u>http://www.metoffice.gov.uk/guide/weather/observations-guide/how-we-measure-rainfall</u>) which suggests the rim of the tipping bucket to be around 0.5 m above the surrounding ground level. This clarification has been added to the manuscript (Revised manuscript [3:30-33])*

[3:29] what about time drift? Did you reset the loggers every 4-5 weeks to avoid this problem?

<u>Reply:</u> *Thank you for pointing this out. The data loggers were reset every 4-5 weeks during data collection to avoid any significant time drift. We have included this information in the revised manuscript (Revised manuscript [4:5-6])*

[4:1-5] This is not clear to me. If I got it right, you are comparing paired gauges for each rain event by accumulating the rainfall over the gauges and comparing them and if the difference exceed the 4%

Peleg, N., Marra, F., Fatichi, S., Paschalis, A., Molnar, P. and Burlando, P., 2016. Spatial variability of extreme rainfall at radar subpixel scale.
Journal of Hydrology.

Reply: This section has been modified in the revised manuscript to give more precise information on how the quality control was carried out using paired setup (Revised manuscript [4:8-18]).

[5:10] three rainfall intensity classes were SUBJECTIVELY selected? What was the criterion?

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

We added the above clarification in the revised manuscript (Revised manuscript [5:25-29])

Furthermore, we performed the following test to see if these three classes represent the entire intensity range. We produced 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. Looking at these variogram at Fig. C1, we believe the variograms in Fig. 7 are good representations of the average conditions for corresponding intensity classes.

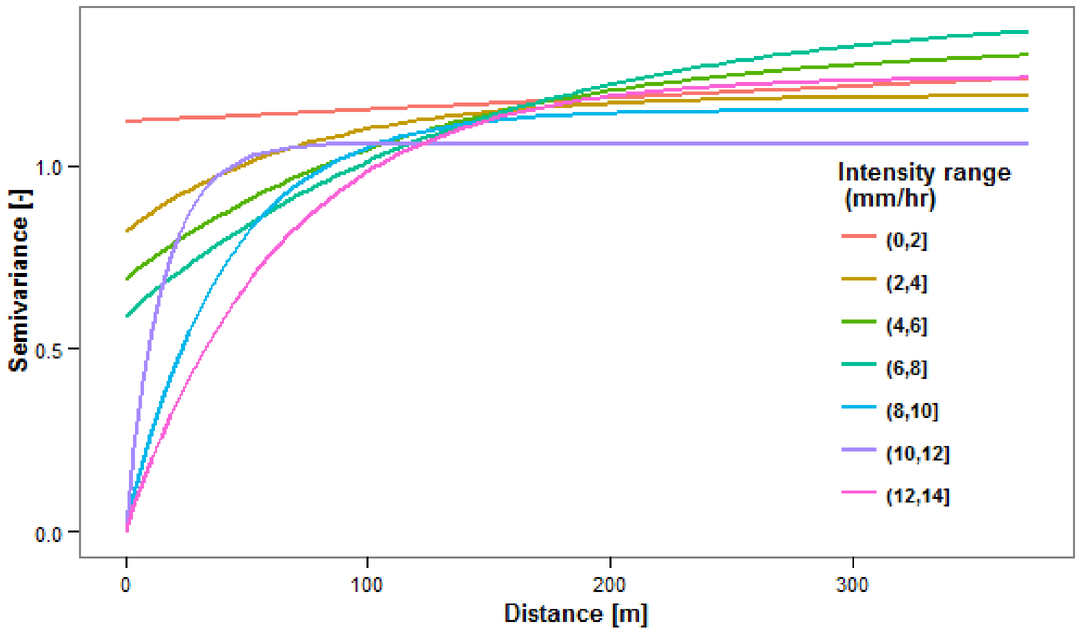


Fig. C1: Calculated variograms for 5 min averaging interval and for a narrower range of intensity

We have added the Fig. C1 in the revised manuscript and included the above discussion (Fig. 9, Revised manuscript [11:12-26])

[6:8] n.d.?

<u>Reply:</u> Thank you for pointing out this. It should be Van der Waerden (1953). We have corrected this in the revised manuscript

[6:26] "It is negligibly small in the case of rainfall intensity data". Not necessarily, Peleg et al. (2013) reported a 0.92 nugget for 1-min time resolution. I am not sure that this can be neglected.

<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 have now modified the text (Revised manuscript 7:10-11)

[7:13-22] spatial stochastic simulation – Please provide more information about how the actual stochastic engine works. How was the variogram reproduced?

Reply: We accept that this section requires more detail. We have modified this section in the revised manuscript (Revised manuscript Section 3.5).

[7:22] I would except that a finer grid would improve your predictions. Especially when very high rainfall intensity is recorded over the domain, as a rapid (exponential) decrease in rainfall intensity from the centre away is expected (for convective rainfall at least).

Reply: We agree that convective rainfall would vary rapidly and therefore having a higher resolution grid might improve the results. But we had only 8 measurement points over the area of 200m × 400m which gives a measurement resolution of 10000 m^2/measurement point. Hence prediction at every 25 m× 25 m (625 m^2) is fine enough for prediction of areal rainfall, also for convective rainfall. Increasing the resolution to 10 m × 10 m only reduces the standard deviation of the prediction by less than 5% in most cases while making the computational time six times higher (a summary on computation power is presented as supplementary material together with the revised manuscript).

[Equation 6 and 7] Equation 6 – doesn't it also need to be divided by m? I think the readers are aware to the statistics of mean and standard deviation thus you can probably delete these equations.

Reply: Thank you for pointing out this error in Equation 6. As per reviewer's suggestion we decided to remove these equations and modified the section accordingly. (Revised manuscript - Section 3.6)

[8:17] "nugget effect . . . at zero lag distance due to measurement error" – are you sure it is just because of a measurement error? Rainfall variability exists between pair gauges, even for a 1 m distance, at least for temporal resolution of 1-5 min. Please check over the paper I have mentioned at comment [2:19-21] above. Your statement is repeated several times again during the text. I would at least discuss the possibility of having the nugget effect as more than a simple representation of rain-gauges measurement error.

<u>Reply:</u> We accept that the nugget effect could be due to a combination of micro-scale spatial variability and measurement error. We modified this section to address this and to interpret the variograms better (Revised manuscript section 4.1).

[9:18-21] I would argue that reason why "the behaviour of spatial correlation against rainfall intensity class is not very distinctive" in your study is due to the short period of data you have used.

<u>Reply:</u> As stated in our response to comment [3:19-20], we had enough data points to develop meaningful and stable variograms based on which the above statement ("the behaviour of spatial

correlation against rainfall intensity class is not very distinctive") is made. 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 substantially larger than 100. Hence, we do not think that the absence of a clear trend in the behaviour of spatial correlation against rainfall intensity class is due to lack of data. Moreover 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.

[Equation 8] CV equation is also commonly known, I suggest you to delete this equation as well.

<u>Reply:</u> *Thank you for your suggestion. We agree that CV is a very common measure, but given the context and for completeness we decided to keep it in.*

[12:23] X-band radar can reach 250 m and 3 min resolution. I think it is good enough for small urban catchments.

<u>Reply:</u> *We agree with the reviewer. What we wanted to argue was that the resolution of most commonly available radar data (1000 m) is not enough for an urban catchment of this spatial extent (< 1000m). In addition the level of uncertainty in radar measurements would be much higher than that of point measurements, especially at a fine averaging interval (< 5 min) which is often of interest in urban hydrology (Seo and Krajewski, 2010; Villarini et al., 2008). We have modified this sentence in the revised manuscript. (Revised manuscript 15:20-23)*

[12:29-31] for a similar climate.

<u>Reply:</u> *Thank you for pointing out this. We have included this in the revised manuscript. (Revised manuscript 15:25-27)*

[General comment 1] I think your method suggested for rainfall upscaling is really interesting and can be very useful to some of the reader. However I, as a reader, would like to have more information, such as: What is the minimum number of rain-gauges required for a given catchment is order to apply your suggested upscaling (e.g. are 3 gauges over 1 km2 are enough?)? What should be the spatial configuration of these rain-gauges over the domain? Another question- if you would leave one of the gauges out of your analysis, how it would affect the results (what is the sensitivity of the network design?)?

<u>Reply:</u> *Thank you for the suggestion. We agree that this additional information would be useful to some readers, but to answer some of these questions we would need a more extensive analyses on the sampling design which we think is a research topic in its own right. Please find below our detailed response:*

What is the minimum number of rain-gauges required for a given catchment is order to apply your suggested upscaling (e.g. are 3 gauges over 1 km2 are enough?)? What should be the spatial configuration of these rain-gauges over the domain?

We think that a simple and generic 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 obvious 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.

We have included the above discussion in the revise manuscript (Revised manuscript 15:12-19)

Another question- if you would leave one of the gauges out of your analysis, how it would affect the results (what is the sensitivity of the network design?)?

Leaving one station out would 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 reduce to 21 from 28. But the further effect of leaving one station out needs to be analysed in detail to see how it affects the uncertainty in the estimation of areal average rainfall intensity. In the manuscript we have not included such sensitivity analysis considering the direct relevance to the main scope of this study, length of the manuscript and the work load required to perform such analysis.

[General comment 2] You stated that the stochastic model require some “computational demands”. Can you give some details? How much time is needed to run the stochastic model per time step? What kind of machine do you need to use? It can be given as supplementary information but some readers might be interested to know.

Reply: *Thank you for this comment. We have provided this information as supplementary material together with revised manuscript.*

[General comment 3] The paper is oriented for the urban hydrology community, but if fact who will benefits the most from your method are hydro-meteorologists that are often looking for different methods to upscale rainfall observation from point scale to weather radar scale. Consider changing the title and addressing this as well. I think that due to the lack of sufficient length of observation, you should focus more on the method and who can benefit from it and less on your results.

Reply: *Thank you for the suggestion. We think the spatial extent (0 – 400 m) and the temporal averaging intervals (2 min - 30 min) considered in this study are in the interest of the urban hydrology community. 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.*

Regarding the comment on lack of data and focusing more on the methodology, we request the reviewer to refer to our responses to the general comment and specific comment [3:19-20].

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.

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.: Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale : a review, Biotechnol. Agron. Soc. Environ., 17(2), 392–406, 2013.

Mair, A. and Fares, A.: Comparison of Rainfall Interpolation Methods in a Mountainous Region of a

Tropical Island, , (April), doi:10.1061/(ASCE)HE.1943-5584.0000330., 2011.

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.

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.

Van der Waerden, B. L. .: Order tests for two-sample problem and their power I-III, Indag. Math., 14,15, 453–458 (14), 303–316(15), 1953.

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

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 review 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, but however our geo statistical models (based on which further results are produced) are stable. This has now been discussed in detail in our revised manuscript. (Revised manuscript [10: 30-11:11])

- 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 Section 4.2. We apologise if this was not clear in the manuscript and we clarified this in the revised manuscript. (Revised manuscript [4: 27-29])

- 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 temporal averaging intervals examined a minimum event duration of 20 minutes was 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 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.

- 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 instants (i.e. sample variograms for time instants with similar rainfall characteristics), 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 (In revised manuscript 5:13-18). P5 l 6-7 is just a part of an explanation on how the intensity classes were 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: We accept that this section requires more detail. We have modified this section in the revised manuscript to answer the reviewer's comment (Revised manuscript Section 3.5).

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. We modified this section to address this and to interpret the variograms better (Revised manuscript section 4.2).*
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 of 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. The 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 included this discussion in the revised manuscript (Revised manuscript [15:25-31])

It is a challenging task to validate these results using hydrological modelling as such validation also needs estimation of other sources of uncertainty (structural and parameter) as well as overall uncertainty in the model output.

….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; Fiener and Auerswald, 2009; Krajewski et al., 2003; Pedersen et al., 2010; 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 be known if it is larger or smaller than other uncertainty sources. 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 that a simple and generic rule on the number of data points cannot be derived from this methodology. It will be case-dependent but our methodology can help solve this problem in any particular case. 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 size of the available data set would also play a role in deciding the number of measurement locations.*

We have included the above discussion in the revise manuscript (Revised manuscript [15:12-19])

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

We did not investigate the sensitivity of results to changes in station density/combination so we cannot answer this question. Since our paper is already quite long we decided not to add this specific issue as well. It could be addressed in subsequent research.

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.

Reply: *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, aggregation of simulations at points to spatial averages). Further, we think the spatial extent (0 – 400 m) and the temporal averaging intervals (2 min -30 min) considered in this study are of interest to urban hydrology. Also, we think uncertainty estimation in areal rainfall would be more useful for the (urban) 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 title 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.

Reply: *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 a case where the input data are rainfall observations at points*
- *Thus point observations need to be scaled up*
- *Review of existing methods that can do this*
- *Disadvantages of these methods*
- *Solution that does not have these disadvantages is to take a geostatistical approach*
- *The main challenges with geostatistical approach and how these can be dealt with*

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.

<u>Reply:</u> Please refer to our response to specific comment p2 l19-25.

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

<u>Reply:</u> The maximum threshold value was limited to 10mm/hr to have enough time instants for the highest range (i.e. > 10 mm/hr) in order to produce stable variograms even at 30 min temporal averaging interval. It was then decided to divide the 0 – 10 mm/hr class to two equal subclasses (i.e. < 5mm/hr and 5-10 mm/hr). This resulted in three subclasses, which is a reasonable number given the size of the data set and computational demand.

We have included the above text in the revised manuscript. (Revised manuscript [5:25-29])

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

<u>Reply:</u> 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.

<u>Reply:</u> 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. Considering the length of the paper and the length of methodology section itself we decided to keep this section as it is.

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

<u>Reply:</u> 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

<u>Reply:</u> Thank you for pointing this out. We have replaced q with n, which is the total number of observation points.

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?

<u>Reply:</u> Time instants (i.e. sample variograms for time instants with similar rainfall characteristics), not events, were pooled to increase the spatial pairs. Hence mean and standard deviation from each time instant is used for standardisation (ref Equation 1) as well as inverse standardisation. Clearly, the mean and standard deviation used for standardisation should also be used for re-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). Regarding the interpretation of variograms, 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 gauges, 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. We modified the Fig. 8 (Revised manuscript Fig. 9) to show CV instead of standard deviation and modified the section 4.2.1 accordingly. We also made the x-axis logarithmic for easier interpretation.

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 have corrected the sentence in the revised manuscript (Revised manuscript 13:12-14)

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 reviewer 1's comment we have replaced this figure with Fig. C1 below and modified the discussion accordingly. This plot includes labels of CV values instead of error bars.

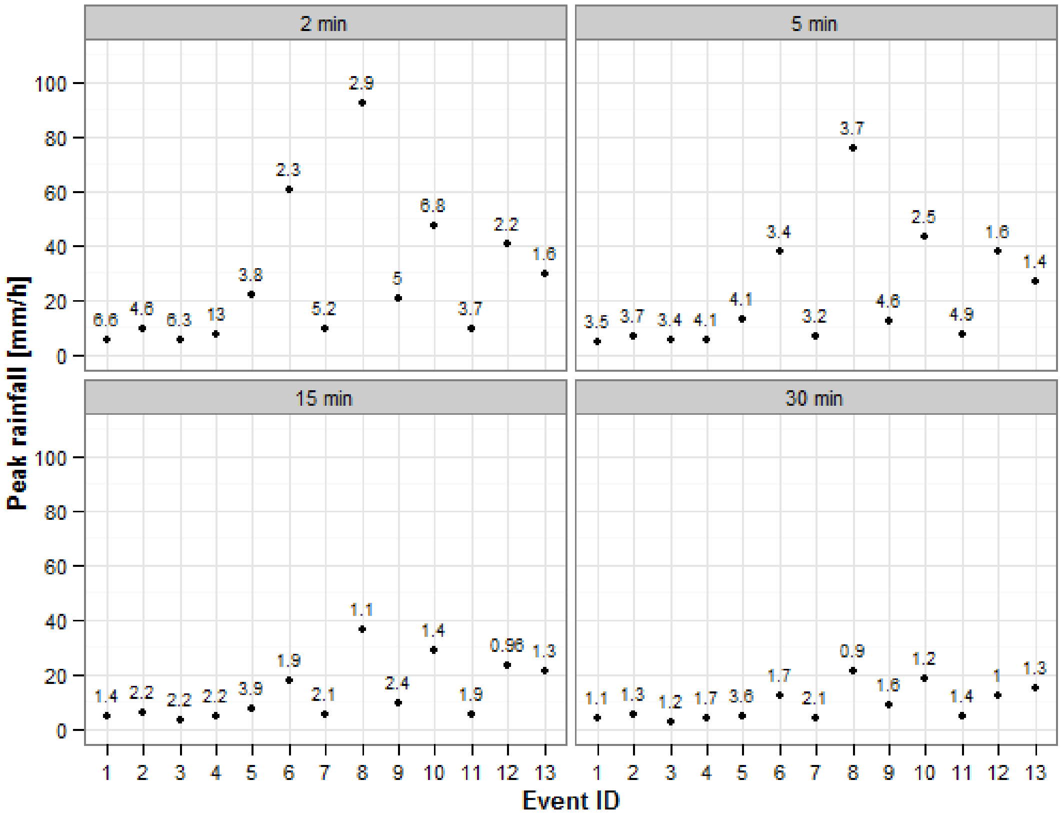


Fig. C1: Predictions of event peaks of AARI (indicated by points) together with labels indicating corresponding CV (%) values

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.

<u>Reply:</u> We removed Fig. 11 from the manuscript and discussions are now based on revised Fig. 10 (please refer to our response to comment p11 l1*).*

We accept the fact that the sample size is too small to derive a firm conclusion. To comment on prediction uncertainty against intensity with a larger sample size, we modified the Fig. 8 (Revised manuscript Fig. 9)to show CV instead of standard deviation (refer to our response to comment p10 l3*). Now it can be seen from modified Fig. 9 that there is a clear trend of increasing CV with increasing AARI. We also modified the discussions in the sections 4.2.1 and 4.2.3 accordingly.*

Further, although we do not think peak prediction at event 8 is an outlier, 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.

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. In any case, we moved this conclusion to the discussion and emphasize more on the fact that reduction in the peaks is obvious. (Revised manuscript 13:3-6)

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, in similar climates, as it provides information on uncertainty associated with rainfall estimation which is arguably the most important input in these models. This estimate of uncertainty in combination with 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 have included this explanation in the revised manuscript (Revised manuscript [15:25-31]).

Technical notes:

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

Reply: Thank you for pointing this out. It has been corrected in the revised manuscript. We also revised the symbols to be consistent throughout the 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 out this error in Equation 6. As per reviewer's suggestion we decided to remove these equations and modified the section accordingly. (Revised manuscript - Section 3.6)

p8 l6 Eq. (7) The term under the root has to be squared.
Reply: Thank you for pointing this out.

p8 l25 "nugget-to-still ratio" to "nugget-to-sill ratio" (several times)
Reply: Corrected in the revised manuscript.

p9 l3 "in their study found" to "found in their study"
Reply: Corrected in the revised manuscript.

p9 l15 "(2003) in their" to "(2003) found in their"
Reply: 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. 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 meant that 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 have reformulated the sentence in the revised manuscript (Revised manuscript 15:10-12)

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, Earth Surf. Process. Landforms, 34(6), 848–859, doi:10.1002/esp.1779, 2009.

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.

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.

Pedersen, L., Jensen, N. E., Christensen, L. E. and Madsen, H.: Quantification of the spatial variability of rainfall based on a dense network of rain gauges, Atmos. Res., 95(4), 441–454, doi:10.1016/j.atmosres.2009.11.007, 2010.

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.

Geostatistical upscaling of rain gauge data to support uncertainty analysis of lumped urban hydrological models

Manoranjan Muthusamy[1], Alma Schellart[1], Simon Tait[1], Gerard B.M. Heuvelink [2]

[1] Department of Civil and Structural Engineering, University of Sheffield, Sheffield, S1 3JD, UK
[2] Soil Geography and Landscape group, Wageningen University, Wageningen, 6700, The Netherlands

Correspondence to: Manoranjan Muthusamy (m.muthusamy@sheffield.ac.uk)

Abstract. In this study we develop a method to estimate the spatially averaged rainfall intensity together with associated level of uncertainty using geostatistical upscaling. Rainfall data collected from a cluster of eight paired rain gauges in a 400 × 200 m^2 urban catchment are used in combination with spatial stochastic simulation to obtain optimal predictions of the spatially averaged rainfall intensity at any point in time within the urban catchment. The uncertainty in the prediction of catchment average rainfall intensity is obtained for multiple combinations of intensity ranges and temporal averaging intervals. ~~Two~~The two main challenges addressed in this study are scarcity of rainfall measurement locations and non-normality of rainfall data, both of which need to be considered when adopting a geostatistical approach. Scarcity of measurement points is dealt with by pooling sample variograms of repeated rainfall measurements with similar~~ spatial~~ characteristics. Normality of rainfall data is achieved through the use of Normal Score Transformation. Geostatistical models in the form of ~~variogram~~variograms are derived for transformed rainfall intensity. ~~Then~~Next spatial stochastic simulation which is robust to nonlinear data transformation is applied to ~~derive predictions and prediction error variances of spatially averaged rainfall intensity. Realisations~~produce realisations of rainfall fields. These realisations in transformed space are first back-transformed and next spatially aggregated to derive a random sample of the spatially averaged rainfall intensity. Results show that the prediction uncertainty comes mainly from two sources: spatial variability of rainfall and measurement error. At smaller temporal averaging intervals both these effects are high, resulting in a relatively high uncertainty in prediction~~, especially for low intensity rainfall.~~. With longer temporal averaging intervals the uncertainty becomes lower due to ~~both better~~stronger spatial correlation of rainfall data and relatively smaller measurement error. Results also show that the measurement error increases with decreasing rainfall intensity resulting in a higher uncertainty at lower intensities. 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, which is arguably the most important input in these models. This will help to better interpret model results and avoid false calibration and force-fitting of model parameters.

Keywords: Geostatistical upscaling, spatial stochastic simulation, areal average rainfall intensity, hydrological modelling, uncertainty

1. Introduction

Being the process driving runoff, rainfall is arguably the most important input parameter in any hydrological modelling study. But it is a challenging task to accurately measure rainfall due to its highly variable nature over time and space, especially in small urban ~~areas~~catchments. Despite recent advances in radar technologies rain gauge measurements are still considered to be the most accurate way of measuring rainfall, especially at short temporal averaging intervals (< 1 hour~~)~~ ~~which are of most interest in urban hydrology studies. However, many of the~~), which are of most interest in urban hydrology studies (Seo and Krajewski, 2010; Villarini et al., 2008). However, many commonly used urban hydrological models (e.g. SWMM, HBV, MIKE11) are lump catchment models (LCM) where time series of areal average rainfall intensity (AARI) ~~is~~are needed as model input. Therefore, point observations of rainfall need to be scaled up using spatial aggregation in order to be fed in to a LCM. There are a number of interpolation methods available for spatial aggregation and ~~are~~ used in ~~these models~~the various LCM to scale up point rainfall data. The simplest ~~interpolation~~ method is to take the arithmetic average (Chow, 1964) of the point observations within the catchment. But this method does not account for the spatial correlation structure of the rainfall and the spatial organisation of the rain gauge locations. Another commonly used method in hydrological modelling is the nearest neighbour interpolation (Chow, 1964; Nalder and Wein, 1998) which leads to Thiessen polygons. In this method the nearest observation is given a weight of one and other observations are given zero weights during interpolation, thereby ignoring spatial variability of rainfall to a certain extent. There are also other methods, with varying complexity levels, including inverse distance weighting (Dirks et al., 1998), polynomial interpolation (Tabios III and Salas, 1985)~~,~~ and moving window regression (Lloyd, 2005). The predictive performance of the above methods are found to be case-dependent and no single method has been shown to be optimal for all catchments and rainfall conditions (Ly et al., 2013). One common drawback with all the above methods is that they do not provide any information on the uncertainty of the predictions of AARI as all the methods are deterministic. ~~Since~~The uncertainty in prediction of AARI mainly comes from two sources; uncertainty due to measurement errors and uncertainty associated with spatial variability of rainfall. The characteristics of measurement errors can vary depending on the rain gauge type. For example, errors associated with commonly used tipping bucket rain gauges 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). In addition to measurement errors and since rainfall can vary over space significantly, any spatial aggregation method for scaling up the point rainfall measurements ~~adds~~incorporates more uncertainty ~~on top of existing measurement error~~ (Villarini et al., 2008). The magnitude of the uncertainty depends on many factors including rain gauge density and location, rainfall variability, catchment size, topography and the spatial interpolation technique used. Quantification of the level of uncertainty is essential for robust interpretation of hydrological model outputs. For instance, the absence of information on uncertainty ~~level~~ can lead to force fitting of hydrological model parameters to compensate for the uncertainty in rainfall input data (Schuurmans and Bierkens, 2006).

Geostatistical methods such as kriging present a solution to this problem by providing a measure of prediction error. ~~In addition to this capability, these statistical methods also take into account the spatial dependence structure of the measured rainfall data [8,9]. Although these features make geostatistical methods better than the~~In addition to this capability, these statistical methods also take into account the spatial dependence structure of the measured rainfall data (Ly et al., 2013; Mair and Fares, 2011). Although these features make geostatistical methods more attractive than deterministic methods, they are rarely used in LCM due to their inherent complexity and heavy data requirements. Since they are statistical methods encompassing multiple parameters the amount of spatial data required for model inference is ~~high~~higher compared to deterministic methods. In addition the underlying assumption of geostatistical approaches typically requires data to be normally distributed (Isaaks and Srivastava, 1989). In general, catchments, especially those at small urban scales, do not contain as many measurement locations as required by geostatistical methods. Furthermore, rainfall intensity data are ~~not always~~almost never normally distributed, especially at smaller averaging intervals (< hour) (Glasbey and Nevison, 1997). ~~But despite~~Despite these challenges geostatistical methods can provide information on uncertainty associated with predicted AARI. This capability can be utilised in uncertainty propagation analysis in hydrological models. In literature, ~~geostatistics has~~geostatistical methods have been used to analyse the spatial correlation structure of rainfall at various spatial scales (Berne et al., 2004; Ciach and Krajewski, 2006; Emmanuel et al., 2012; Jaffrain and Berne, 2012), however its application to support uncertainty analyses of upscaling rainfall data has not been ~~often~~ explored.

In this paper we present a geostatistical approach to derive AARI and the level of uncertainty associated with it from observations obtained from multiple "paired" rain gauges located in a small urban catchment. The proposed approach presents solutions to the above~~ mentioned~~ described challenges of geostatistical methods. ~~Firstly~~First, it uses pooling of sample variograms of rainfall measurements at different times but with similar characteristics to ~~artificially~~ increase the ~~measurement data available~~number of paired observations used to ~~fulfil the requirements of the geostatistical method. Secondly~~fit variogram models. Second, a data transformation method is employed to transform the rainfall data to obtain a ~~more~~ normally distributed data set. The level of uncertainty in the prediction of AARI is then quantified for different combinations of temporal averaging intervals and intensity ranges for the studied urban catchment. We focused on a small urban catchment with a spatial ~~extend~~extent of less than a kilometre given the recent ~~findings~~research on the significance of unmeasured spatial rainfall variability at such spatial scales, especially ~~on~~for urban hydrological and hydrodynamic modelling applications (Gires et al., 2012, 2014; Ochoa-Rodriguez et al., 2015).

2. Data collection

2.1 Location and rain gauge network design

The study area is located in Bradford, a city in West Yorkshire, England. Bradford ~~benefits from~~has a maritime climate, with an average yearly rainfall of 873 mm recorded from 1981-2010 (MetOffice, UK). The rain gauge network, used in this study was located at the premises of Bradford University (Fig. 1) and rainfall data were collected from paired tipping bucket rain gauges placed at eight locations covering an area of 400 × 200 m^2. Data used in this study ~~was~~were collected from April, 2012 to August, 2012 and from April, 2013 to August, 2013. These stations were located on selected roofs of the university buildings, thereby providing controlled, secure and obstruction-free measurement locations. Each station consists of two tipping bucket type rain gauges mounted ~~1m~~1 m apart. On each roof the paired gauges were placed such that the height of the nearest obstruction is less than two times the distance between the gauges and the obstruction~~. An example of this design~~ The rim of each rain gauge was set up around 0.5 m above the surrounding ground level following UK standard practice (MetOffice, UK). An example of the measurement setup (Station 6) is also shown in Fig. 1. A histogram of the inter-station distances of the rain gauge network is presented in Fig. 2. Lag distances covered in this network are distributed between 21 m (St. 4-St. 5) and 399 m (St. 1-St. 3).

All rain gauges are ARG100 tipping bucket type with an orifice diameter of 254 mm and a resolution of 0.2 mm. Dynamic calibration was carried out for each individual gauge before deployment and visual checks were carried out every 4-5 weeks during the measurement period to ~~make sure~~ensure that the instruments were free of dirt and debris. ~~Measurement~~Data loggers were reset every 4-5 weeks during data collection to avoid any significant time drift. Measurements (number of tips) were taken every minute and recorded on TinyTag data loggers mounted in each rain gauge.

Quality control procedures were performed prior to ~~any~~ statistical analysis, taking advantage of the paired gauge setup~~.~~ to detect gross measurement errors. The ~~double~~ 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~~clearly identify periods of missing or incorrect data (Ciach and Krajewski, 2006). During the dynamic calibration of all rain gauges in the laboratory before deployment, it was identified that ~~there could be~~ the highest and lowest values of the calibration factors for the tipping bucket size are 0.196 mm and 0.204 mm. The gauges were recalibrated in the laboratory after the first period of measurement and it was found that the largest change in calibration factor for any gauge was a maximum of ~~±~~ 4 ~~%~~% of the original calibration factor. Therefore a maximum difference of 4% in volume per tip ~~between two rain gauges due~~was assumed to ~~the tipping bucket mechanism.~~be caused by inherent instrument error. It was therefore decided that this is the maximum acceptable difference ~~permitted~~ between ~~paired~~any pair of gauges. ~~Hence every long set~~Sets of cumulative rainfall data ~~from~~corresponding to specific events from the paired gauges were checked against each other and if the (absolute) difference ~~were > ± 4~~ in cumulative rainfall was greater than 4%, that complete set was identified as ~~data affected by some~~

~~type of measurement error (e.g. partial clogging caused by debris), and they were~~unreliable and removed from further analysis.

2.2 Characteristics of the data

The total average network rainfall depth for the summer seasons of 2012 and 2013 are 538 mm and 207 mm, respectively. Figure 3 shows time series of daily rainfall averaged over the network for 2012 and 2103. There is a significant difference in cumulative rainfall between 2012 and 2013. ~~The~~This is because 2012 was the wettest year recorded in 100 years in the UK (MetOffice, UK) and 558 mm of rainfall during 2012 summer was unusually high. An average rainfall of only 360 mm was recorded during April to September over the 1981 - 2010 period at the nearest operational rain gauge station at Bingley, which is around 8 km from the study site with a similar ground elevation (MetOffice, UK).

The data set for 2012 and 2013 contains 13 events yielding more than 10 mm network average rainfall depth each and lasting for more than 20 min. A summary of these events is presented in Table 1. ~~The~~Note that this event separation is only used for the presentation of results in chapter 4. Hence it does not leave out any data from the development and calibration of the geostatistical model as presented in chapter 3. Table 1 shows that the total event duration ranges from 1.5 h to 11.4 h while the event network average rainfall intensity varies from 1.79 mm/h to 7.96 mm/h. Table 1 also includes summary statistics of peaks of events (temporal averaging interval of 5 min) for the eight stations within the network. Although the spatial extent of the area is only 400 × 200 m^2, it is clear that there is a considerable difference in rainfall intensity measurements indicated by the standard deviation and range of peaks observed in the individual events. The maximum standard deviation between peaks of individual events is 9.27 mm/h for event 8, which is around 12.5-% of the mean network peak intensity of 74.4 mm/h. This ~~shows~~ variation provides evidence of the potential importance of analysing ~~spatial~~ uncertainty in the ~~calculation~~estimation of AARI even in such a small urban catchment.

3. Methodology

~~Fig.~~Figure 4 summarises the ~~complete~~ procedure of geostatistical upscaling of the rainfall data adapted in this study in a step-by-step instruction followed by the detail ~~description~~descriptions of each step. This complete procedure was repeated for temporal averaging intervals of 2 min, 5 min, 15 min and 30 min in order to investigate the effect of temporal aggregation on the prediction of AARI. The entire ten months of collected data were used for the development and calibration of the geostatistical model.

3.1 Step 1: Pooling of ~~time steps~~sample variograms

~~This~~The rain gauge network contains eight measurement locations. These eight measurement locations give 28 spatial pairs at a given time instant which ~~is far less~~yields too few spatial lags than ~~what geostatistics~~would normally be used in geostatistical modelling ~~requires~~. For example, ~~(Webster and Oliver, 2007)~~Webster and Oliver (2007) recommends around

100 measurement points to calibrate a geostatistical model. The procedure adapted in this study increases the number of pairs by pooling ~~rainfall measurements~~sample variograms for time instants with similar rainfall characteristics. With n measurement locations and measurements taken at t time instants, the pooling over t time instants creates $t \times ½ \times n \times (n-1)$ spatial pairs. Although this procedure increases the number of spatial pairs by a factor ~~t times~~, the spatial ~~lags~~separation distances for which information is available will be limited to the original configuration of the n measurement locations~~, n~~. The underlying assumption of this pooling procedure is that the spatial variability over the pooled time instants is the same. Therefore it is important to pool sample variograms of rainfall measurements with similar rainfall characteristics. Since the spatial rainfall variability is often intensity-dependent (Ciach and Krajewski, 2006), the characteristics of a less intense rainfall event may not be the same as that of a high intensity rainfall event. Hence to make the assumption of consistency of spatial variability, the range of rainfall intensity over the pooled time instants should be reasonably small. On the other hand, one should also make sure that there are enough time instants within a pooled subset to meet the data requirement to calibrate the geostatistical model. Based on the above two criteria, three rainfall intensity classes were selected. The maximum threshold value was limited to 10mm/hr to have enough time instants for the highest range (i.e. > 10 mm/hr) in order to produce stable variograms even at 30 min temporal averaging interval. It was then decided to divide the 0 – 10 mm/hr class to two equal subclasses (i.e. < 5mm/hr and 5-10 mm/hr). This resulted in three subclasses, which is a reasonable number given the size of the data set and computational demand. The number of time instants (t) within each rainfall intensity class is presented for three temporal averaging intervals in Fig. ~~5. As expected, with increasing temporal averaging intervals the number of time instants t reduces. 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~~5. The natural characteristic of rainfall data results in the dominance of lower intensity rainfall (0.1-5.0 mm/h) ~~rainfall~~ over the recording ~~periods. For~~period. In addition, the number of time instants t obviously reduces with increasing temporal averaging intervals due to the ~~30 min temporal averaging interval~~aggregation process. As a consequence there are only ~~eight~~seven time instants for the intensity range > 10 mm/h~~.~~ at the 30 min temporal averaging interval. This limits the maximum temporal averaging interval to 30 min for our analyses. ~~But for~~For a catchment of this size (400 × 200 m^2) it is very unlikely to have a~~ catchment~~ response time of more than 30 min. Hence, from a hydrological point of view consideration of temporal averaging intervals longer than ~~30 min would not be sensible~~30 min would not be sensible. Note that although there are only seven time instants, the pooling procedure will produce 196 (=7×28) points to calculate and calibrate the geostatistical model for that intensity class.

3.2 Step 2: Standardisation of rainfall intensities

Having chosen the rainfall intensity classes to create pooled time instants, there can still be inconsistency in spatial variability between time instants within a class and therefore assuming a single geostatistical model for the whole subset may not be ~~valid~~realistic. To reduce this effect to a certain extent, all observations within an intensity class were standardised using the mean and standard deviation of each ~~and every~~ time instant ~~given by Eq. (1). Further steps were carried out on the standardised rainfall intensity.~~as follows:

$$\cancel{\tilde{r}_{ij} = \frac{r_{ij} - m_t}{sd_t}}\tilde{r}_{ix} = \frac{r_{ix} - m_i}{sd_i} \quad (1)$$

~~Where~~where i=1… t ~~where t is number of time instances; j,~~ x=1… n ~~where n is number of locations~~; ~~$\tilde{r}_{ij}$~~$\tilde{r}_{ix}$ is standardised rainfall intensity at a time ~~step~~instant i and location ~~j~~x; ~~r_{ij}~~r_{ix} is rainfall intensity at time ~~step~~instant i and location ~~j~~x and; m_i, sd_i are mean and standard deviation of rainfall intensities at time ~~step~~instant i, respectively. Further steps were carried out on the standardised rainfall intensity.

3.3 Step 3: Normal transformation of data

The upper part of Fig. 6 shows the distribution of standardised rainfall intensity for a temporal averaging interval of 5 min derived using Eq. (1). From the figure it is clear that the data are not normally distributed~~do not follow a normal distribution~~. Distributions for other temporal averaging intervals (i.e. 2 min, 15 min and 30 min) show a similar behaviour. But the geostatistical upscaling method to be used is based on the based on the ~~Gaussian distribution which means that it is assumed that the underlying data are from a~~ normal distribution. This requires the rainfall data to be normally distributed prior to the calibration of the geostatistical model. Among the non-parametric transformation, The Normal Score Transformation (NST, also known as normal quantile transformation (Van der Waerden, 1953))~~(Van der Waerden, n.d.)~~ is a widely used method to transform a variable distribution to the Gaussian distribution. It has widely been applied in many hydrological applications (Bogner et al., 2012; Montanari, A., & Brath, 2004; Todini, 2008; Weerts et al., 2011). The concept of NST is to match the p-quantile of the data distribution of the p-quantile of the standard normal distribution. Consider a standardised rainfall intensity $\tilde{r}$ with cumulative distribution $F_{\tilde{R}}(\tilde{r})$~~a data variable z with edf $F_Z(z)$~~. It ~~will be~~ is transformed to a r_N ~~y~~-value with a Gaussian cumulative distribution $F_{R_N}(r_N)$ ~~edf $F_Y(y)$~~ as follows

$$\cancel{y = F_Y^{-1}(F_Z(z))}\ r_N = F_{R_N}^{-1}\left(F_{\tilde{R}}(\tilde{r})\right) \quad (2)$$

Detailed description of NST including the steps involved can be found in Bogner et al. (2012), Van der Waerden (1953) and Weerts et al. (2011). The lower part of Fig. 6 shows the transformed standardised intensity for the temporal averaging interval of 5 min.

3.4 Step 4: Calibration of Geostatistical model

A geostatistical model of ~~rainfall intensity r (in our case normally transformed~~(normalised) rainfall intensity r_N (derived from Section 3.3) at any location x can be written as:

$$\cancel{r(x)}\, r_N(x) = p(x) + \varepsilon(x) \qquad (3)$$

~~Where~~where $p(x)$ is the trend (explanatory part) and $\varepsilon(x)$ is the stochastic residual (unexplanatory part). Considering the availability of ~~rainfall intensity~~ data, small catchment size and scope of this study, it ~~can be~~was assumed that the trend is constant and does not depend on explanatory variables (e.g. topography of the area, wind direction). The stochastic term ε is spatially correlated and ~~can be~~ characterised by a variogram model. A variogram model typically consists of three parameters; nugget, sill and range (Isaaks and Srivastava, 1989). The nugget is the value of the semi-variance at near-zero distance ~~caused by~~. It is often greater than zero because of random measurement error and micro-scale spatial variation. ~~It is negligibly small in the case of rainfall intensity data.~~ The range is the distance beyond which the data are no longer spatially correlated. The sill is the maximum variogram value and equal to the variance of the variable of interest~~.~~

3.5 Step 5: Spatial stochastic simulation

~~During spatial interpolation, the assumption of a constant trend makes this an ordinary kriging system~~ (Isaaks and Srivastava, 1989) ~~which is given by the following set of equations~~

The assumption of a constant trend makes that the spatial interpolation can be solved using an ordinary kriging system (Isaaks and Srivastava, 1989):

$$\cancel{\sum_{k=1}^{q} w_k \gamma_{lk}} \sum_{x=1}^{n} w_x \gamma_{xy} - \mu = \cancel{\gamma_{l0} \quad \forall\, l}\, \gamma_{yz} \quad \forall\, y = 1, \ldots, \cancel{q}\, n \qquad (4)$$

$$\cancel{\sum_{l=1}^{q} w_l =} \sum_{y=1}^{n} w_y = 1 \qquad (5)$$

~~Where l, k are indices; the $w_l, l = 1, \ldots, q$~~where $w_y, y = 1, \ldots, n$ are ordinary kriging weights~~;~~, ~~γ_{lk}~~γ_{xy} is the semivariance between rainfall intensities at locations ~~x_l~~ x and ~~x_k;~~ y ~~γ_{l0}~~γ_{yz} is the semivariance between rainfall intensities at location ~~x_l~~y and prediction location ~~x_0,~~ z, and μ is a Lagrange parameter. Once the ordinary kriging weights~~, which give the minimum prediction error variance,~~ are calculated using Eq. (4) and Eq. (5), point rainfall intensities can be ~~calculated~~predicted using point kriging at any given point~~. But~~ by taking the weighted average of the observed rainfall intensities, using the w_l as weights. In this case we need a change of support from point to block as our intention is to ~~estimate~~predict the average rainfall intensity over the catchment. This is usually done by ~~simply~~ predicting at all points inside the catchment and ~~then~~

integrating these over the catchment~~ (this~~. This procedure is known as block kriging~~).~~, which also has provisions for calculating the prediction error variance of the catchment average. But the procedure of NST as explained in Section 3.4 also involves back-transformation of kriging predictions to the original domain at the end (Step 6). Since this transformation is typically non-linear, the back-transform of the spatial average of the transformed variable that is obtained from block kriging is not the same as the spatial average of the back-transformed variable; we need the latter and not the former. In principle, we could predict at all points within the block, back-transform all and next calculate the spatial average, but standard block kriging software implementations do not support this and neither is it possible to compute the associated prediction error variance. Hence block kriging cannot be applied. The alternative used in this study is to ~~use~~apply a computationally more demanding spatial stochastic simulation approach, which involves generation of a larger number of ~~possible~~ realisations and spatial aggregation of these realisations. ~~-~~Unlike kriging, spatial stochastic simulation ~~do~~does not aim to ~~minimizing a local~~minimize the prediction error variance but ~~focus~~focuses on the reproduction of the statistics such as the histogram ~~or semivariogram~~and variogram model (Goovaerts, 2000). ~~The output from spatial stochastic realisation is a set of alternative realisations of rainfall predictions at user defined grid points. The differences among these realisations are used as a measure of uncertainty.~~The output from spatial stochastic simulation is a set of alternative rainfall realisations ('possible realities'). The mean of a large set of realisations approximates the kriging prediction, while their standard deviation approximates the kriging standard deviation. We used the sequential Gaussian simulation algorithm which involves the following steps (Goovaerts, 2000):

i. ~~In this study,~~Define a prediction grid ~~with cells of 25 m × 25 m was laid over the study area and 500 simulations of~~ (a 25m × 25m regular grid in this case);

ii. Visit a randomly selected grid cell that has not been visited before and predict the transformed rainfall intensity at the grid cell ~~centres were generated~~centre using ~~spatial stochastic simulation~~ordinary kriging, this yields a kriging prediction and a kriging standard deviation;

iii. Use a pseudo-random number generator to sample from a normal distribution mean equal to the kriging prediction and standard deviation equal to the kriging standard deviation and assign this value to the grid cell centre;

iv. Add the simulated value to the conditioning data set, in other words treat the simulated value as if it were another observation;

v. Go back to step ii and repeat the procedure until there are no more unvisited grid cells left.

The five steps above produce a single realisation. This must be repeated as many times as the number of realisations required (500 in this study). It must also be repeated for each time ~~step. This~~instant, which explains that the computational burden can be high. Implementation of these steps with the gstat package in R (Pebesma, 2004) is straightforward.

The grid size and number of simulations (i.e., the sample size) were selected considering the spatial resolution of available measurements and computational demand. It was observed that neither a finer grid nor more ~~simulation improves~~simulations improved the ~~predictions much.~~results significantly. Increasing the resolution to 10 m × 10 m only reduces the standard deviation of the prediction by less than 5% in most cases while making the computational time six times higher (a summary on computation power is presented as supplementary material).

3.6 Step 6-9: Calculation of AARI and associated uncertainty

Once ~~we derived all~~ the realisations ~~from the spatial stochastic simulation~~have been prepared these are back-transformed by applying the ~~next~~ inverse of Eq. (2) to all grid cells (step ~~is to transform them to the original domain using back-transformation (NST^{-1}~~6). Some values derived from spatial stochastic simulation were outside the transformed data range. Hence during back transformation (step 6) of these values linear extrapolation was used. These linear models were derived using a selected number of head and tail portion of normal Q-Q plot. This is one of the simplest and most commonly used solutions for NST back-transformation (Bogner et al., 2012; Weerts et al., 2011). Considering the scope of this study and the relatively small number of data which had to be extrapolated, other extrapolation methods ~~have~~were not ~~been~~ explored. ~~Having~~ After step 6, the back transformed ~~all~~ realisations ~~to the original domain the next step is spatial aggregation~~are spatially aggregated one by one (step 7~~) of each and every simulation which gives us m (=500) simulations~~). This yields as many spatially averages as the number of realisations that had been generated in step 5. This set of ~~the areal~~values is a simple random sample from the probability distribution of the catchment average rainfall. Thus, the sample mean~~. The mean prediction ($\bar{p}$)~~ and standard deviation ~~($s_{\bar{p}}$) were then calculated using Eq. (6) and Eq. (7)~~provide estimates of the mean and standard deviation of the distribution, respectively (~~Step~~step 8).

$$\cancel{\bar{p} = \sum_{i=1}^{m} p_x \qquad (6)}$$

$$\cancel{s_{\bar{p}} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(p_i - \bar{p})} \qquad (7)}$$

~~Where p_x is prediction from the i th simulation.~~ Finally, by doing the inverse standardisation of ~~$\bar{p}$ and $s_{\bar{p}}$~~the mean and standard deviation of the distribution to account for step 2, the AARI and associated uncertainty measure (standard deviation) were derived (Step 9).

4. Results and Discussion

4.1 Calibration of the geostatistical model of rainfall

As explained in Section 3.4, the geostatistical model of transformed rainfall data ~~was~~were calibrated using variograms for three different intensity ranges. This procedure was repeated for temporal averaging intervals of 2 min, 5 min, 15 min and 30 min. Exponential models were fitted to empirical ~~semi-~~variograms. The resulting variograms are presented in Fig. 7.

The variograms illustrate two properties of the collected rainfall measurements; spatial variability of rainfall, and measurement error. One of the main parameters which characterises these properties is the nugget. Theoretically at zero lag distance the variance should be zero. However most of ~~these~~the variograms exhibit a positive nugget effect (generally presented as nugget-to-sill ratio) at zero lag distance. This nugget effect can be due to two reasons; random measurement error~~–~~ and microscale spatial variability of rainfall. Unfortunately we cannot quantify these causes individually using the variograms. But there is a consistent pattern of nugget against both rainfall intensity class and temporal averaging interval which helps to interpret the variograms.

~~From Fig. 7 it is clear that the nugget to sill ratio varies with both rainfall intensity class and temporal averaging interval.~~ Considering the behaviour of nugget-to-~~still~~sill ratio against rainfall intensity class, it can be observed that the smaller the intensity the higher the nugget-to-sill ratio, regardless of temporal averaging interval. For example, at 2 min averaging interval the nugget-to-~~still~~sill ratio increases from zero to almost one (nugget variogram) as the rainfall intensity class changes from > 10 mm/h to < 5 mm/h. The pure nugget variogram at < 5 mm/h means that either there is no spatial correlation at the regarded distance, or the spatial correlation of the field cannot be detected by the measurements because of the measurement error. Looking at the behaviour of nugget-to-~~still~~sill ratio against temporal averaging interval, Fig. 7 shows that the smaller the averaging interval the higher the nugget-to-~~still~~sill ratio, regardless of rainfall intensity class. For example, for rainfall intensity class ~~of~~ 5.0–10.0 mm/h the nugget-to-~~still~~sill ratio ~~decrease~~decreases from almost one to zero as the temporal averaging interval increases from 2 min to 30 min. ~~These~~Overall these observations show that the combined effect of random measurement error and microscale special variability of rainfall characterised by nugget-to-~~still ratio (a) decreases with increasing rainfall intensity class and (b)~~ sill ratio decreases with increasing (a) rainfall intensity class and (b) averaging interval.

~~This behaviour of measurement error against rainfall intensity and averaging interval is mainly attributed to measurement-related error of tipping bucket type rain gauges~~Regarding the behaviour of the nugget-to-sill ratio against averaging interval, it is expected that with the averaging interval the (microscale) spatial correlation of rainfall would increase, which partly explains the observed pattern. The increase in spatial correlation of rainfall intensity with increasing temporal averaging interval agrees with other similar studies (e.g. Ciach and Krajewski, 2006; Fiener and Auerswald, 2008; Krajewski et al., 2003; Villarini et al., 2008). For example, Krajewski et al. (2003) observed in their study on analysis of spatial correlation structure of small-scale rainfall in central Oklahoma a similar behaviour using correlogram functions for different temporal averaging intervals. But commenting on the decreasing trend of the nugget-to-sill ratio against intensity class, it cannot be attributed to improvement in microscale spatial correlation as it is neither natural nor proven. In fact, in Fig. 7 the behaviour of spatial correlation against rainfall intensity class does not show a distinctive trend except at the origin, i.e. the nugget effect. The absence of any consistent trend of spatial variability against intensity class was also observed in Ciach and Krajewski, (2006). Meanwhile this decreasing trend of nugget-to-sill ratio against rainfall intensity corresponds well with measurement errors of tipping bucket type rain gauges caused by its sampling mechanism (hereafter referred as TB error). This is due to the rain gauges' inability to capture small temporal variability of the rainfall time series. The behaviour of TB error against rainfall intensity ~~and temporal averaging interval~~ as seen from Fig. 7 complements results from previous studies (Habib et al., 2001; Villarini et al., 2008). These studies also show that the TB error decreases with temporal averaging interval. Habib et al. (2001) ~~in their study~~ found similar behaviour of TB error with increasing intensity (0-100 mm/h) and also with increasing averaging interval (1 min, 5 min and 15 min~~) and also with increasing intensity (0-100 mm/h~~). Although the bucket size used in their study (0.254 mm) is slightly different from our rain gauge bucket size of 0.2 mm, the characteristic of the TB error against rainfall intensity for different averaging interval is consistent in both cases. In summary, the behaviour of nugget-to-sill ratio of the variograms against temporal averaging interval can be explained by the combined effect of microscale spatial variability of rainfall and TB error, while the behaviour of nugget-to-sill ratio against intensity range can mainly be attributed to the latter.

~~It is clear from Fig. 7 that the spatial correlation of rainfall intensity improves with increasing temporal averaging interval. Decorrelation distance (range) is a good indicator of the improvement in spatial correlation. At lower temporal averaging intervals (≤ 5 min) the variograms for all rainfall intensity classes reach the decorrelation distance very quickly (< 100 m). But at averaging intervals ≥ 15 min, the decorrelation distance has not been reached even at a maximum separation distance, showing the improvement in spatial correlation. High spatial variability of rainfall at shorter temporal averaging interval (≤ 5 min) is an important observation in the context of urban drainage run off modelling, as the time step used in such models is generally around 2 min for small catchments. The increase in spatial correlation of rainfall intensity with increasing temporal averaging interval is an expected observation and agree with other similar studies (Ciach and Krajewski, 2006; Krajewski et al., 2003; Villarini et al., 2008). For example, Krajewski et al. (2003) in their study on analysis of spatial correlation structure in small-scale rainfall in central Oklahoma observed similar behaviour using correlogram functions for different temporal averaging intervals.~~

~~The behaviour of spatial correlation against rainfall intensity class is not very distinctive from Fig. 7. At lower temporal averaging intervals (≤ 5 min) the variograms for different rainfall intensity classes look similar apart from the nugget effect caused by TB error. At averaging intervals ≥ 15 min, variograms for rainfall intensity classes 5.0–10.0 mm/h and > 10 mm/h looks similar, showing a common pattern of spatial correlation.~~

In addition to the nugget-to-sill ratio, another parameter that characterises the variograms is the range, i.e. the distance up to which there is spatial correlation. At lower temporal averaging intervals (≤ 5 min) the variograms for all rainfall intensity classes reach the variogram range very quickly (< 100 m). But at averaging intervals ≥ 15 min, the range has not been reached even at a maximum separation distance, showing the improvement in spatial correlation. High spatial variability of rainfall at shorter temporal averaging interval (≤ 5 min) is an important observation in the context of urban drainage runoff modelling, as the time step used in such models is generally around 2 min for small catchments.

The fact that the data set covers only 10 months of data from two years with varying climatology is something that need to be acknowledged. 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 time periods are reflection of the practical and funding issues to maintain such dense networks operating accurately for extended periods. The characteristics of our data are comparable with (Ciach and Krajewski, 2006; Fiener and Auerswald, 2009) as these studies also used rainfall data from warm months to investigate the spatial correlation structure. Despite the fact that the data cover only 10 months all derived variogram models are stable and reliable. 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 fewest number of observations) we had a total of 196 spatial lags to calculate the variogram. Furthermore, we demonstrated that all derived variogram models are stable and reliable by examining sub-sets of the data. We randomly selected 80% of the data from each intensity class and reproduced the variograms to compare them with the variograms presented in Fig. 7. We had to limit the subclass percentage to 80% to give enough time instants to reproduce variograms for all subclasses. We repeated this procedure a few times. Comparing these variograms with Fig. 7 shows that these variograms are very similar. One set of the variograms computed from 80% of the data are presented as supplementary material. This analysis supports our claim that the variograms shown in Fig. 7 are stable and an adequate representation of the rainfall spatial variation for each intensity class and temporal averaging interval.

One of the assumptions we made during the pooling procedure is that the spatial variability is reasonably consistent within a pooled intensity class. We acknowledge 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 reasonable compromise. Further we also introduced step 2 (section 3.2) that standardises the rainfall for each time instant within a subset. 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. 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 within the pooled classes, it at least relaxes the assumption of the same variogram within subclasses. To compare the behaviour of variogram models for a narrower intensity interval, we produced variograms 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. Narrower intensity classes means that the assumption of similar spatial variability within a pooled subset is more realistic. Comparing Fig. 8 with Fig. 7, we conclude that the variograms shown in Fig. 7 are accurate representations of the average spatial variability conditions for corresponding intensity classes.

4.2 Geostatistical upscaling of rainfall data

Having calculated all variograms, the next step is to apply spatial stochastic simulation for the time instants of interest followed by steps 6 to 9 in Fig. 4 to calculate the AARI together with associated uncertainty. This procedure was carried out for all events presented in Table 1. The following sections present and discuss the predicted AARI and associated uncertainty levels derived from step 9.

4.2.1 Prediction error vs AARI

The scatter plot in Fig. ~~8~~9 shows the coefficient of variation of the prediction ~~standard deviation~~error (CV, refer Eq. (6)) plotted against predicted AARI at 5 min averaging interval for all time instants of all events presented in Table 1. ~~The uncertainty level represented by the standard deviation is due to the combined effect of both spatial variability of rainfall and TB error in the rainfall data. Most of the time the coefficient of variation (CV, refer Eq. (8)) is around 5 % which is not high. But there are also instances where it is more than 10 % with a maximum of 13 % corresponding to a predicted rainfall of 11 mm/h.~~

$$CV = \frac{AARI\ prediction\ error\ standard\ deviation}{Predicted\ AARI} \times 100\% \qquad \cancel{(8)} \quad (6)$$

~~As expected, a large portion of the predicted AARI is within 0-10 mm/h. The corresponding standard deviations are highly scattered, possibly due to the large TB error for rainfall intensities smaller than 10 mm/h as already seen from Fig. 7, where the nugget-to-still ratio is close to one at this intensity range. In this range there are also a few small clusters of points separated from the larger cluster with almost zero standard deviation, even for a predicted AARI of around 8 mm/h. It shows a consistent rainfall over the area at these time instants which results in a very small standard deviation in the predicted AARI.~~

The uncertainty level in the prediction of AARI represented by the CV is due to the combined effect of both spatial variability of rainfall and TB error in the rainfall data. It can be seen here that there is a clear trend of increasing CV with increasing AARI. The CV values are as high as 80% when the AARI is smaller than 1 mm/hr and they get reduced to less than 10% when AARI is larger than 10 mm/hr. In a previous study by Pedersen et al. (2010) using rainfall measurements from similar tipping bucket type rain gauges, they also found that the uncertainty in prediction of mean rainfall depth decreases with increasing mean rainfall depth, but due to the limited information in their results they could not analyse this observation in detail. But here it is clear that this observation corresponds well with what we already observed in variograms in Fig. 7. These variograms show higher nugget-to-sill ratio at lower intensity due to high TB error consequently causing higher uncertainty in the prediction of AARI. At intensity class 0-5 mm/hr the nugget-to-sill ratio was almost one (nugget variogram) and as a result the derived CV values are significantly higher than other two intensity classes. It is interesting to note that, in the range of 1- 10 mm/hr, there are few points that are separated from the larger cluster with almost zero CV. It shows a consistent rainfall measurement over the area at these time instants, which results in a very small CV in the predicted AARI.

The above discussion is based on results from 5 min temporal averaging interval. The following section discusses the effect of temporal averaging interval on prediction error. Further, although CV in Fig. 9 gets as high as 80%, the corresponding AARI is less than 1 mm/hr, thus the prediction error has a very less significance in urban hydrology. Hence we also analysed the prediction error associated with rainfall events' peaks in the last section.

4.2.2 Prediction error vs temporal averaging interval

Having analysed the behaviour of the prediction error CV against predicted AARI, this section presents the effect of temporal averaging interval on the prediction error of AARI. Figure ~~9~~10 shows the kriging predictions with 95 % prediction intervals derived from the prediction standard deviation for temporal averaging intervals of 2 min, 5 min, 15 min and 30 min for event 11. Event 11 ~~represent~~has average conditions ~~roughly~~ in terms of event duration and peak intensity. Prediction errors of other events against the temporal averaging interval follow the same pattern of behaviour.

~~It is expected that with increasing~~ While short time intervals are of greater interest in urban hydrology, they also lead to large uncertainties. Figure 10 shows the smaller the temporal averaging interval ~~the local minima and maxima of AARI get smoothed out. 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. It shows the importance of selecting appropriate temporal scales in the hydrologic modelling of small urban catchments (10–20 ha) where the time steps will be in the range of 2 to 5 min. But on the other hand~~, the larger the ~~temporal averaging interval, the smaller the~~ prediction interval and the ~~smaller~~larger the level of uncertainty. The reason is that with increasing temporal averaging interval while the spatial correlation of rainfall improves the TB error becomes smaller as seen from the variogram in Fig. 7. In fact, when the averaging interval is larger than 15 min the prediction interval width becomes negligible. But ~~as mentioned above,~~ temporal scales of interest in urban hydrology of similar sized catchment can be as ~~small~~low as 2 min where there is still considerable uncertainty. The 95 % prediction interval shows around ± 13 % of error in rainfall intensity corresponding to a prediction of peak rainfall of 47 mm/h at 2 min averaging interval. While temporal aggregation decreases uncertainty, it obviously leads to a significant reduction of the predicted peaks of AARI. For example, the peak of event 11 gets reduced to around 20 mm/h from around 50 mm/h when averaging interval increases from 2 min to 30 min. Hence a careful trade-off between temporal resolution and accuracy in rainfall prediction is needed to decide the most appropriate time step for averaging point rainfall data for urban hydrologic applications.

The decreasing trend of uncertainty in the prediction of AARI with increasing temporal averaging interval agrees with a previous study by Villarini et al. (2008). Although the spatial ~~extend~~extent of their study is much larger (360 km^2), their results also show that the spatial sampling uncertainties tend to decrease with increasing temporal averaging interval due to improvement in measurement accuracy and ~~increased~~improved spatial correlation.

4.2.3 Prediction ~~of~~error Vs peak rainfall intensity

~~Rainfall~~In addition to rainfall event durations, rainfall event peaks are also 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~~11 presents predicted peaks of AARI for all 13 events presented in Table 1, together with labels indicating corresponding ~~95 % prediction intervals indicated by vertical bars~~CV (%) values. 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. ~~In general, Fig. 10 shows that (a) the larger~~

~~the peak value the larger the prediction interval and (b) the larger the temporal averaging interval the smaller the prediction interval~~ as a result of temporal aggregation. As expected, temporal aggregation from 2 min to 30 min also results in the reduction of CV. The highest CV at 2 min averaging intervals is 13% for event 4 and reduces 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.

~~Another interesting observation is that the prediction intervals for smaller peaks (< 20 mm/h) are very small even at an averaging interval of 2 min, where both spatial variability and TB error are at their maximum. But it should be noted that the prediction interval in Fig. 9 and Fig. 10 are derived from standard deviation, a measure which is proportional to the mean. Hence it is important to look at the CV, a normalised measure, (refer Eq. (8)) of these predictions together with their prediction interval to get a clearer picture of uncertainty.~~

~~4.2.4 Prediction vs CV~~

~~CV against all predicted peaks of AARI are presented in Fig. 11. This plot is divided in two segments based on predicted AARI at a threshold value of 10 mm/h to analyse the effect of lower and higher rainfall intensity on CV. Regardless of the range of intensity, CV decreases with averaging interval which is what we expected. Considering CV against intensity range, in general CV is higher when the predicted intensity is lower than 10 mm/h for all temporal averaging intervals. It 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~~As discussed in section 4.2.1, CV decreases with increasing predicted rainfall peaks and this effect is dominant when the averaging interval is at the lowest, i.e. 2 min. This is when the TB error is at its highest. 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~~Even though these values are significantly less than what we observed from Fig. 9 when the rainfall intensity is less than 1 mm/hr, they are still high considering the required accuracy defined in standard guidelines of urban hydrological modelling ~~practises~~practice. For example, the current urban drainage verification ~~guidelines~~guideline (WaPUG, 2012) in the UK ~~set~~defines 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~~. Hence a better trade-off between temporal resolution and accuracy in catchment rainfall prediction is needed when deciding the time step of hydrologic modelling of small urban catchments.~~ in urban areas. A 13% uncertainty in rainfall will result in a similar level of uncertainty in runoff prediction for a completely impervious surface according to the well-established rational formula (Viessman and Lewis, 1995) which is still widely used for estimating design discharge in small urban catchments.

5. Conclusions

Geostatistical methods have 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 presented a method to overcome these challenges and predict AARI together with associated uncertainty using geostatistical upscaling. ~~Rainfall~~We used a spatial stochastic simulation approach to address the combination of change of support (from point to catchment) and non-normality of rainfall observations for prediction of AARI and the associated uncertainty. We addressed the issue of scarcity in measurement points by using repetitive rainfall measurements (pooling) to increase the number of spatial samples used for variogram estimation. The methods were illustrated with rainfall data collected from a cluster of eight paired rain gauges in a 400 × 200 m^2 urban catchment in Bradford, UK~~ were used in~~. The spatial lag ranges from 21 m to 399 m. As far we are aware these are the smallest lag ranges in which spatial variability in rainfall is examined in an urban area using point rainfall measurements. We defined intensity classes and derived different geostatistical models (variograms) for each intensity class separately. We also used different temporal averaging intervals, ranging from 2 to 30 min, which are of interest in urban hydrology. To the best of our knowledge this ~~study. We~~is the first such attempt to assign geostatistical models for a combination of intensity class and temporal averaging interval. Finally, we quantified the level of uncertainty in the prediction of AARI for these different combinations of temporal averaging intervals~~, ranging from 2 to 30 min,~~ and rainfall intensity ranges.

A summary of the significant findings are listed below~~,~~:

- Several studies (e.g. Berne et al., 2004; Gebremichael and Krajewski, 2004; Krajewski et al., 2003) used a single geostatistical model in the form of variogram/correlogram 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 and a distinction between rainfall intensity classes and length of temporal averaging intervals should be made.
- The level of uncertainty in the prediction of AARI using point measurement data essentially comes from two sources; spatial variability of the rainfall and ~~TB~~measurement error. The significance and characteristics of the ~~TB~~measurement error observed here mainly corresponds to sampling related error of tipping bucket type rain gauges (TB error) and may vary for other types of rain gauges.
- TB error decreases with increasing rainfall intensity. As a result of that the prediction error decreases with increasing AARI. At 5 min averaging interval the CV values are as high as 80% when the AARI is smaller than 1 mm/hr and they get reduced to less than 10% when AARI is larger than 10 mm/hr
- At smaller temporal averaging intervals, the effect of both spatial variability and TB error is high, resulting in higher uncertainty levels in the prediction of AARI~~ especially for lower intensity rain (at 2 min temporal averaging interval the average CV is 6.6 % and the maximum CV is 13 %).~~. With increasing temporal averaging interval the uncertainty becomes ~~lower~~smaller as the spatial correlation ~~improves~~increases and the TB error reduces~~ (~~. At 2 min

temporal averaging interval the average CV in the prediction of peak AARI is 6.6 % and the maximum CV is 13 % and they reduced to 1.5 % and 3.6 % respectively at 30 min averaging interval~~ the average CV is 1-2 %)~~.

- ~~With increasing temporal averaging interval while the uncertainty decreases, the predicted peaks of AARI reduces significantly (e.g. peak of event 8 reducing to 21 mm/h from 92 mm/h when averaging interval increases from 2 min to 30 min). Hence a careful trade-off between temporal resolution and accuracy in rainfall prediction is needed to decide the most appropriate time step for averaging point rainfall data for urban hydrologic applications.~~
- TB error at averaging intervals of less than 5 min, especially at low intensity rainfall measurements, is as significant as spatial variability. Hence proper attention to TB error should be given in any application of these measurements, especially in urban hydrology where averaging interval are often as small as 2 min.

~~Although spatial stochastic simulation used in this study needs more computational power than block kriging, it is a robust approach and allows data transformation during spatial interpolation. Such data transformation is important to ensure that rainfall data satisfy the normality assumption of ordinary kriging. 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.~~

~~An urban catchment of this size needs rainfall data at a temporal and spatial resolution which current radar data technology cannot meet. In addition the level of uncertainty in radar measurements would be much higher than that of point measurements especially at a fine averaging interval (< 5 min) which are often of interest in urban hydrology. Hence experimental rain gauge data similar to the one used in this study are crucial for similar studies focused on small urban catchments. Furthermore, the paired gauges setup used in rainfall data collection improves the reliability on rainfall data as it helps to flag the missing or wrong dataset of the measurements and it consequently helps efficient quality control of the data.~~

Although the spatial stochastic simulation method used in this study needs more computational power (a summary on computation power is presented as supplementary material) than block kriging, it is a robust approach and allows data transformation during spatial interpolation and aggregation. Such data transformation is important because rainfall data are not normally distributed for small temporal averaging intervals. The pooling procedure used in this study helps provide a solution to meet the data requirements for geostatistical methods as it extends the available information for variogram estimation. Commenting on the minimum number of measurement points needed to employ this method is difficult, 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 or not a given measurement network can produce a meaningful variogram. As mentioned, Webster and Oliver (2007) advised that around 100 measurement points are needed to adequately estimate a geostatistical model. But there is no single universal rule to define the minimum number of bins and the number of samples for each bin to produce a reliable variogram. Further, since pooling sample variograms of repeated measurements would produce a multiplication of spatial lags, the size of the available data set would also play a role in deciding the minimum number of measurement points.

An urban catchment of this size needs rainfall data at a temporal and spatial resolution which is higher than the resolution of most commonly available radar data (1000 m, 5 min). In addition the level of uncertainty in radar measurements would be much higher than that of point measurements, especially at a small averaging interval (< 5 min), which are often of interest in urban hydrology (Seo and Krajewski, 2010; Villarini et al., 2008). Hence, experimental rain gauge data similar to the ones used in this study are crucial for similar studies focused on small urban catchments.

Results from this study can be used for uncertainty analyses of hydrologic and hydrodynamic modelling of similar sized urban catchments, in similar climates, as it provides information on uncertainty associated with rainfall estimation which is arguably the most important input in these models. ~~This will help to differentiate input uncertainty from total uncertainty thereby helping to understand other sources of uncertainty due to model parameter and model structure. This information can help to avoid false calibration and force fitting of model parameters (Vrugt et al., 2008).~~This estimate of uncertainty in combination with 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). This study can also help to judge optimal temporal averaging interval for rainfall estimation of hydrologic and hydrodynamic modelling especially for small urban catchments.

Acknowledgements

This research ~~work is~~was done as ~~a~~ part of the Marie Curie ITN - Quantifying Uncertainty in Integrated Catchment Studies project (QUICS). This project has received funding from the European Union's Seventh Framework Programme for research, technological development and demonstration under ~~grant agreement~~Grant Agreement no. 607000.

References

Berne, A., Delrieu, G., Creutin, J. D. and Obled, C.: Temporal and spatial resolution of rainfall measurements required for urban hydrology, J. Hydrol., 299(3-4), 166–179, doi:10.1016/j.jhydrol.2004.08.002, 2004.

Bogner, K., Pappenberger, F. and Cloke, H. L.: Technical Note: The normal quantile transformation and its application in a flood forecasting system, Hydrol. Earth Syst. Sci., 16(4), 1085–1094, doi:10.5194/hess-16-1085-2012, 2012.

Chow, V. Te: Handbook of Applied Hydrology. A compendium of water-resources technology, McGraw-Hill, New York., 1964.

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.

Emmanuel, I., Andrieu, H., Leblois, E. and Flahaut, B.: Temporal and spatial variability of rainfall at the urban hydrological scale, J. Hydrol., 430-431, 162–172, doi:10.1016/j.jhydrol.2012.02.013, 2012.

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, Earth Surf. Process. Landforms, 34(6), 848–859, doi:10.1002/esp.1779, 2009.

Gebremichael, M. and Krajewski, W. F.: Assessment of the statistical characterization of small-scale rainfall variability from radar: Analysis of TRMM ground validation datasets, J. Appl. Meteorol., 43(8), 1180–1199, doi:10.1175/1520-0450(2004)043<1180:AOTSCO>2.0.CO;2, 2004.

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.

Gires, A., Tchiguirinskaia, I., Schertzer, D., Schellart, A., Berne, A. and Lovejoy, S.: Influence of small scale rainfall variability on standard comparison tools between radar and rain gauge data, Atmos. Res., 138, 125–138, doi:10.1016/j.atmosres.2013.11.008, 2014.

Glasbey, C. A. and Nevison, I. M.: Modelling Longitudinal and Spatially Correlated Data, edited by T. G. Gregoire, D. R. Brillinger, P. J. Diggle, E. Russek-Cohen, W. G. Warren, and R. D. Wolfinger, pp. 233–242, Springer New York, New York, NY., 1997.

Goovaerts, P.: Estimation or simulation of soil properties? An optimization problem with conflicting criteria, Geoderma, 97(3-4), 165–186, doi:10.1016/S0016-7061(00)00037-9, 2000.

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.

Isaaks, E. H. and Srivastava, R. M.: An Introduction to Applied Geostatistics, Oxford University Press, New York, USA., 1989.

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.

Lloyd, C. D.: Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain, J. Hydrol., 308(1-4), 128–150, doi:10.1016/j.jhydrol.2004.10.026, 2005.

Ly, S., Charles, C. and Degré, A.: Different methods for spatial interpolation of rainfall data for operational hydrology and hydrological modeling at watershed scale : a review, Biotechnol. Agron. Soc. Environ., 17(2), 392–406, 2013.

Mair, A. and Fares, A.: Comparison of Rainfall Interpolation Methods in a Mountainous Region of a Tropical Island, , (April), doi:10.1061/(ASCE)HE.1943-5584.0000330., 2011.

MetOffice UK: MetOffice, UK, [online] Available from: http://www.metoffice.gov.uk/ (Accessed 25 November 2015), n.d.

Montanari, A., & Brath, A.: A stochastic approach for assessing the uncertainty of rainfall-runoff simulations, Water Resour. Res., 40(1), 1–11, doi:10.1029/2003WR002540, 2004.

Nalder, I. a. and Wein, R. W.: Spatial interpolation of climatic Normals: test of a new method in the Canadian boreal forest, Agric. For. Meteorol., 92, 211–225, doi:10.1016/S0168-1923(98)00102-6, 1998.

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

Pebesma, E. J.: Multivariable geostatistics in S: The gstat package, Comput. Geosci., 30(7), 683–691, doi:10.1016/j.cageo.2004.03.012, 2004.

Pedersen, L., Jensen, N. E., Christensen, L. E. and Madsen, H.: Quantification of the spatial variability of rainfall based on a dense network of rain gauges, Atmos. Res., 95(4), 441–454, doi:10.1016/j.atmosres.2009.11.007, 2010.

Schuurmans, J. M. and Bierkens, M. F. P.: Effect of spatial distribution of daily rainfall on interior catchment response of a distributed hydrological model, Hydrol. Earth Syst. Sci., 3(4), 2175–2208, doi:10.5194/hessd-3-2175-2006, 2006.

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.

Tabios III, G. Q. and Salas, J. D.: A comparative analysis of techniques for spatial interpolation of precipitation, J. Am. Water Resour. Assoc., 21(3), 365–380, 1985.

Todini, E.: A model conditional processor to assess predictive uncertainty in flood forecasting, Int. J. River Basin Manag., 6(2), 123–137, doi:10.1080/15715124.2008.9635342, 2008.

Viessman Jr., W. and Lewis, G. L.: Introduction to hydrology, in Introduction to hydrology, p. 311, Harper Collins., 1995.

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.

~~Vrugt, J. a., ter Braak, C. J. F., Clark, M. P., Hyman, J. M. and Robinson, B. a.: Treatment of input uncertainty in hydrologic modeling: Doing hydrology backward with Markov chain Monte Carlo simulation, Water Resour. Res., 44(12), doi:10.1029/2007WR006720, 2008.~~

Van der Waerden, B. L. .: Order tests for two-sample problem and their power I-III, Indag. Math., 14,15, 453–458 (14), 303–316(15), ~~n.d~~1953.

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

Weerts, A. H., Winsemius, H. C. and Verkade, J. S.: Estimation of predictive hydrological uncertainty using quantile regression: examples from the National Flood Forecasting System (England and Wales), Hydrol. Earth Syst. Sci., 15(1), 255–265, doi:10.5194/hess-15-255-2011, 2011.

Table 1: Summary of events which yielded more than 10 mm rainfall and lasted for more than 20 min with summary statistics of event peaks (derived at 5 min temporal averaging interval) from all stations.

Event ID.	Date	Network average duration (h)	Network average intensity (mm/h)	Network average rainfall (mm)	Summary statistics of peaks between different stations (mm/h)			
					Mean	Std. Dev	Max	Min
1	18/04/2012	6.33	2.20	13.9	5.10	0.550	6.02	4.74
2	25/04/2012	6.42	2.55	16.3	7.05	0.751	8.32	5.92
3	09/05/2012	8.92	1.79	16.0	5.10	0.537	5.97	4.74
4	14/06/2012	6.83	1.99	13.6	5.25	0.636	6.04	4.74
5	22/06/2012	11.4	2.39	27.3	12.7	1.72	15.4	9.67
6	06/07/2012	4.42	5.31	23.4	38.5	4.52	42.9	30.5
7	06/07/2012	3.25	3.23	10.5	7.20	0.679	8.46	5.93
8	07/07/2012	1.50	7.84	11.8	74.4	9.27	86.5	61.9
9	19/07/2012	3.08	3.35	10.3	12.7	2.01	14.5	9.74
10	15/08/2012	2.00	7.96	15.9	43.0	3.69	47.8	37.5
11	14/05/2013	7.92	2.14	17.0	8.08	1.20	9.55	6.09
12	23/07/2013	1.75	6.51	11.4	37.7	2.09	42.6	35.7
13	27/07/2013	8.17	4.34	35.5	26.6	1.23	27.5	23.8

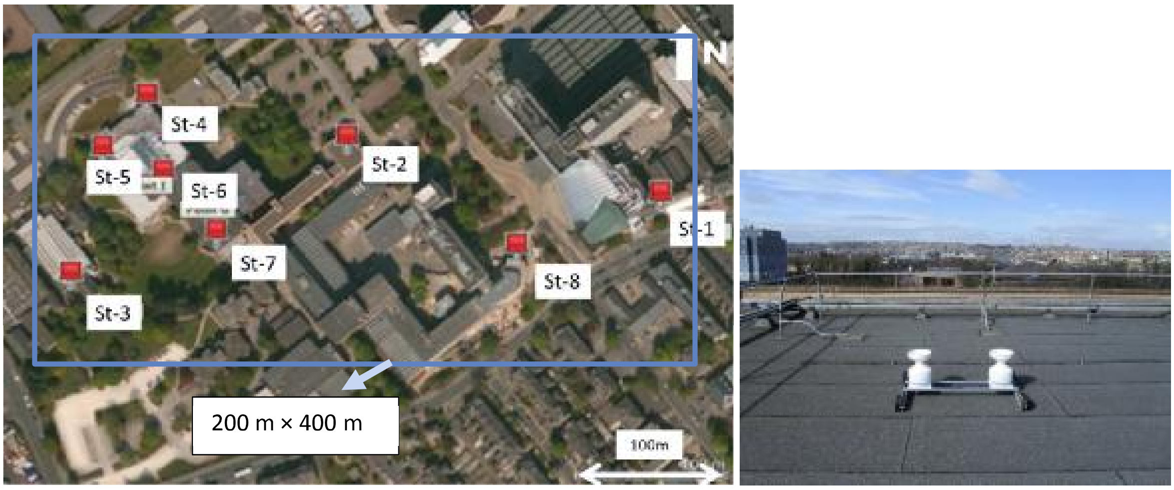


Figure 1: (Left) A aerial view of - rain gauge network covering an area of 200 × 400 m^2 at Bradford University, UK. (Right) A photograph of paired rain gauges at station 6.

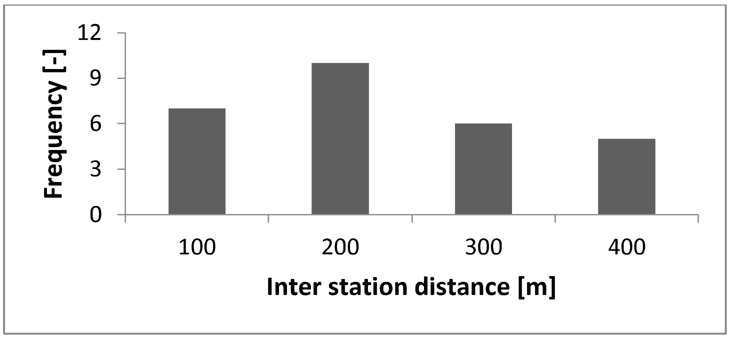


Figure 2: Histogram with class interval width of 100 m showing frequency distribution of inter-station distances (m)

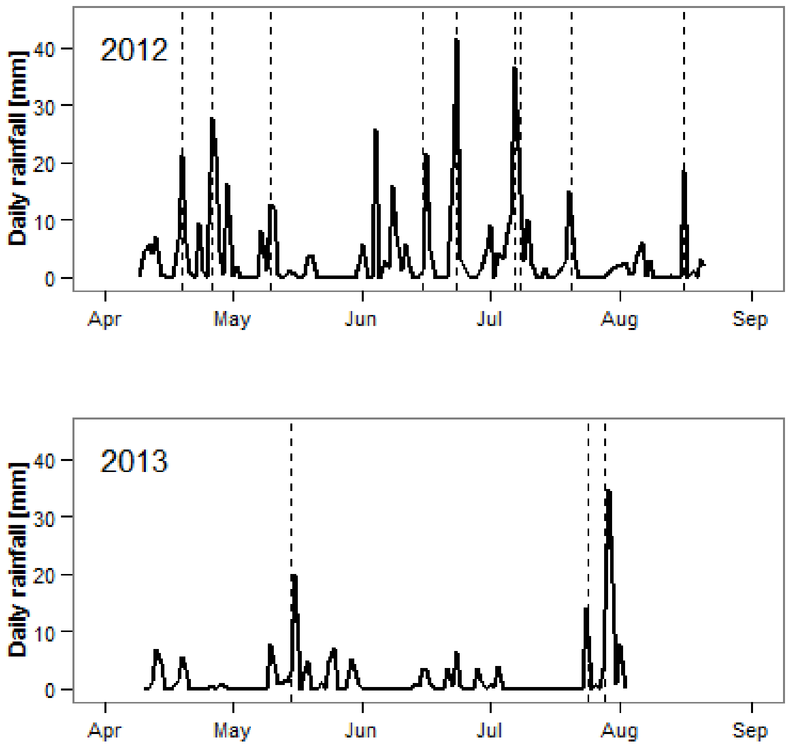
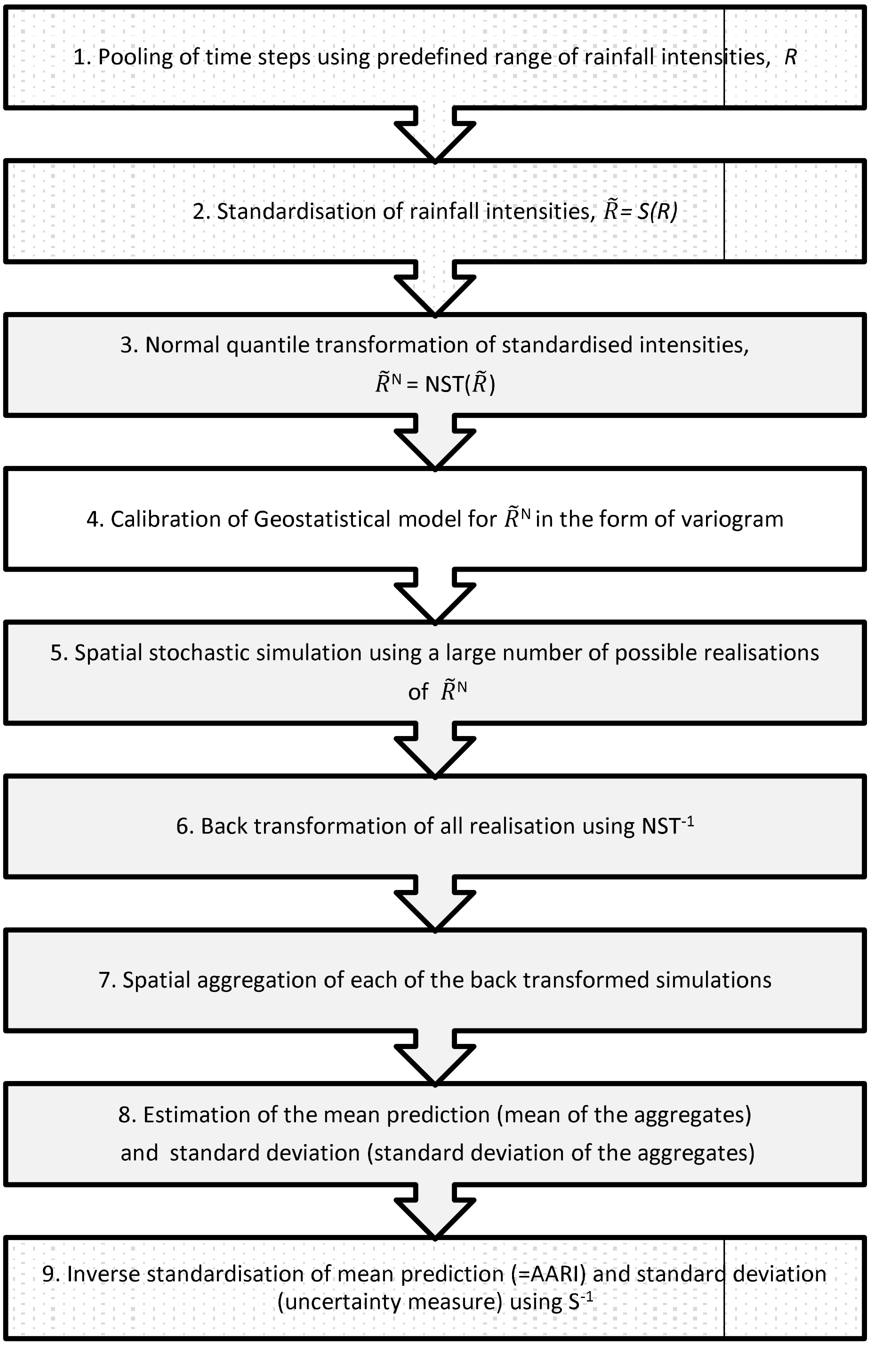


Figure 3 : Time series of network average daily rainfall in the two seasons of 2012 and 2013 with vertical dashed lines indicating the events presented in Table 1


1. Pooling of time steps using predefined range of rainfall intensities, R
2. Standardisation of rainfall intensities, $\tilde{R} = S(R)$
3. Normal quantile transformation of standardised intensities, $\tilde{R}^N$ = NST($\tilde{R}$)
4. Calibration of Geostatistical model for $\tilde{R}^N$ in the form of variogram
5. Spatial stochastic simulation using a large number of possible realisations of $\tilde{R}^N$
6. Back transformation of all realisation using NST^{-1}
7. Spatial aggregation of each of the back transformed simulations
8. Estimation of the mean prediction (mean of the aggregates) and standard deviation (standard deviation of the aggregates)
9. Inverse standardisation of mean prediction (=AARI) and standard deviation (uncertainty measure) using S^{-1}

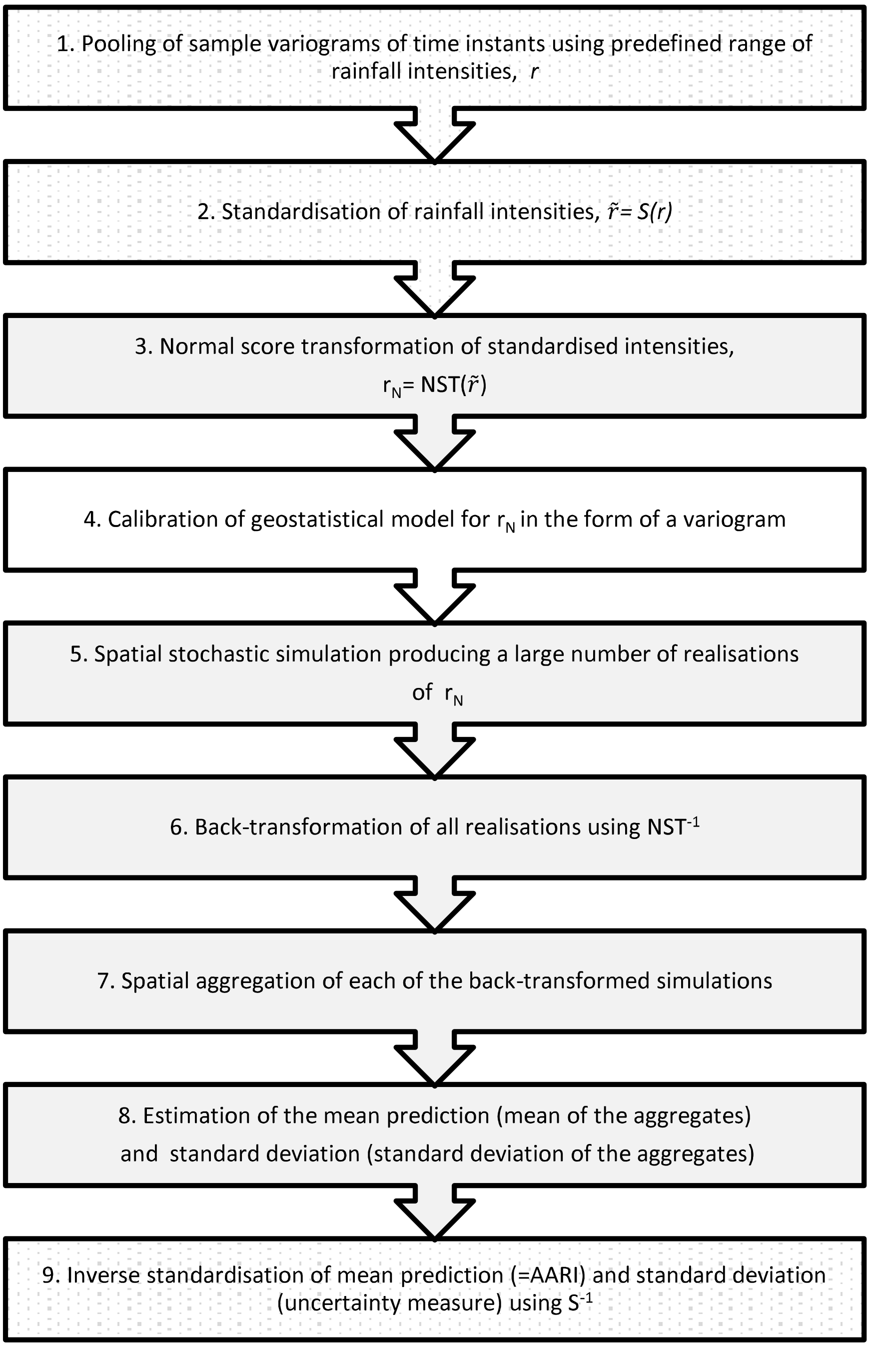


Figure 4: Step by step procedure developed in this study to predict AARI and associated level of uncertainty. Boxes highlighted in dots indicate the steps to resolve the problem of scarcity in measurement locations, grey boxes show the steps introduced to address non-normality of rainfall data.

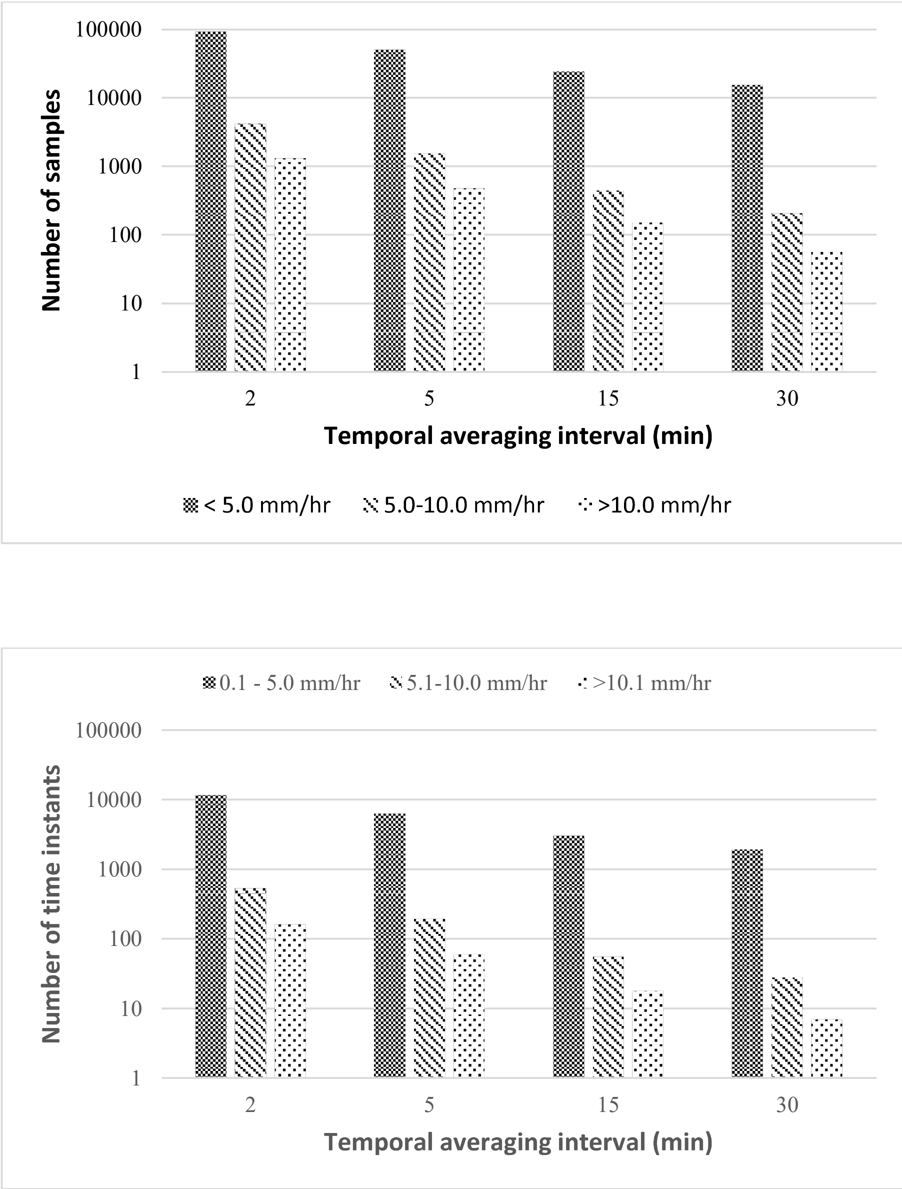


Figure 5 : Number of time instants for each temporal averaging interval and rainfall intensity class combination.

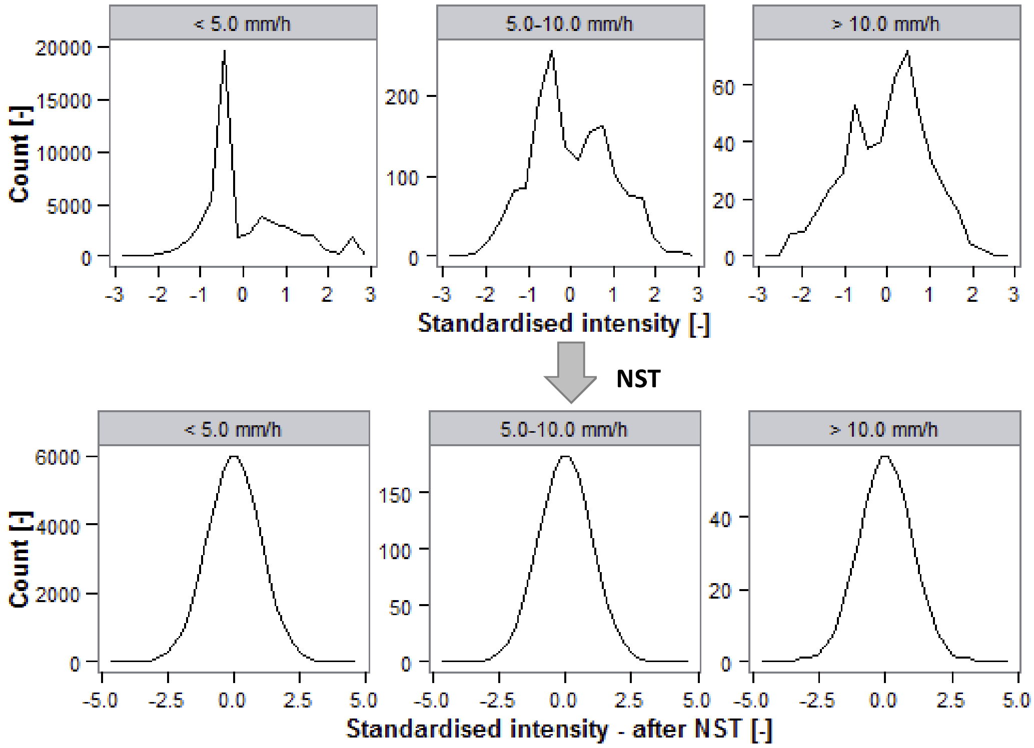


Figure 6: Distribution of standardised rainfall intensity for different rainfall intensity classes at a temporal averaging interval of 5 min before (upper part) and after (lower part) normal score transformation (NST~~)~~).

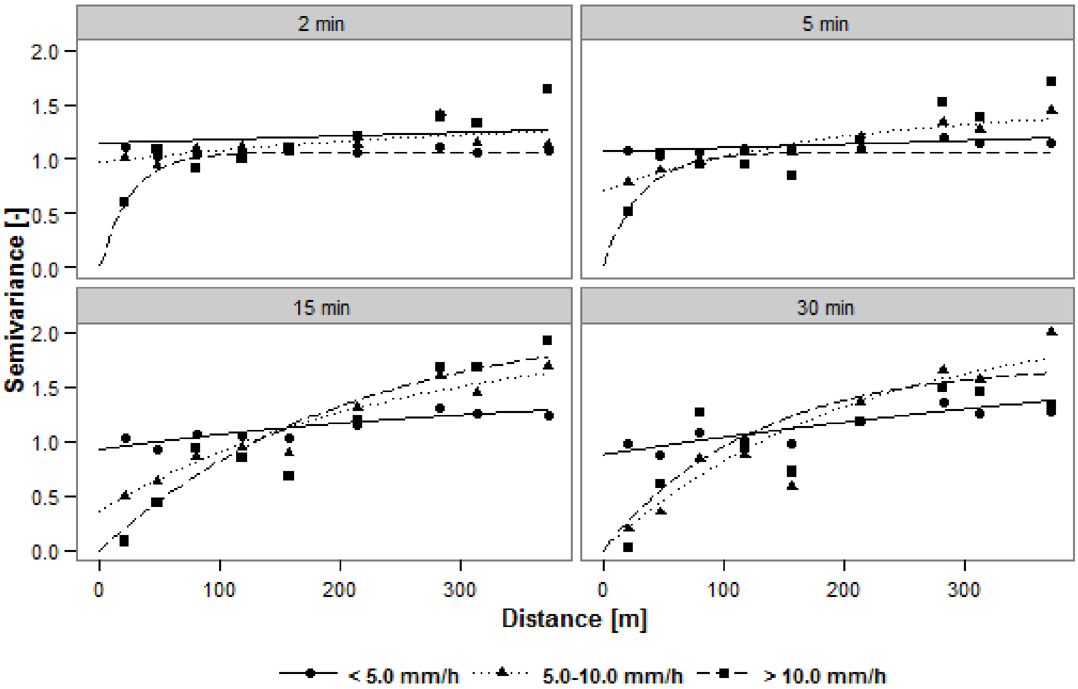


Figure 7: Calculated variograms for each intensity class within each temporal averaging interval ~~and~~

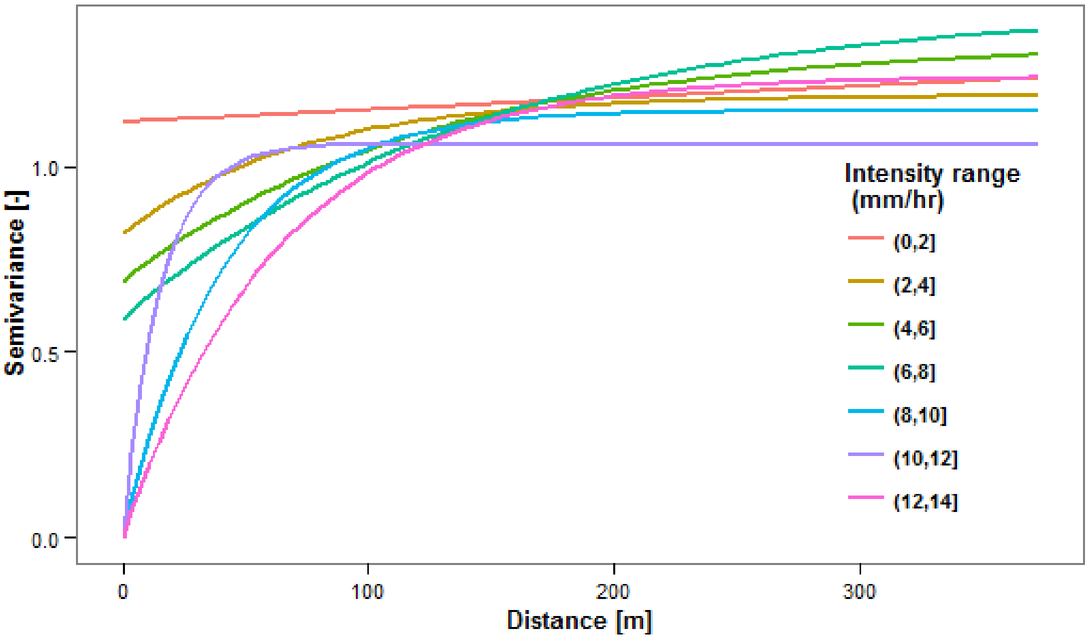
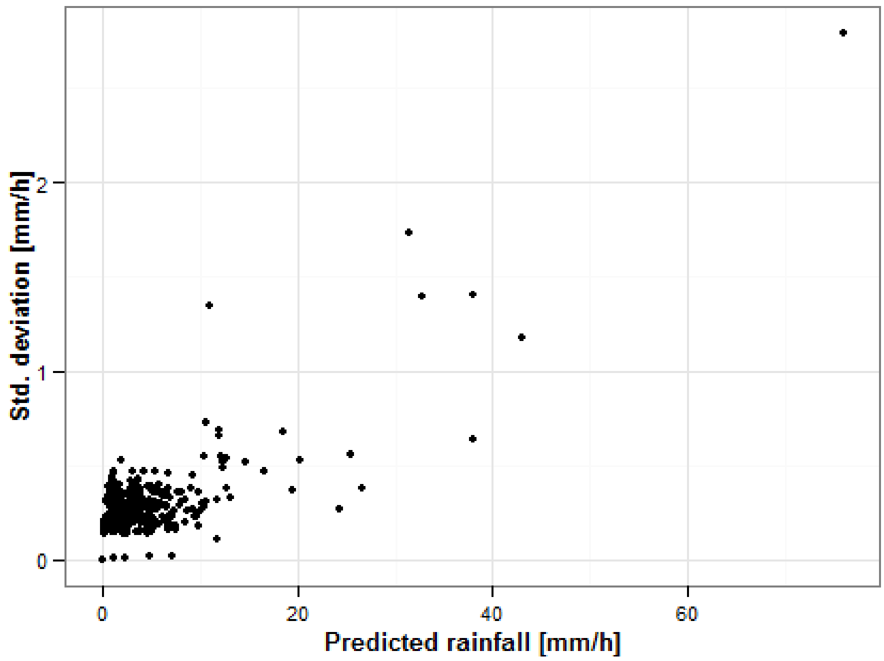


Figure 8: Calculated variograms for a narrower range of intensity classes at 5 min averaging interval.


Std. deviation [mm/h]
Predicted rainfall [mm/h]
0
1
2
0
20
40
60

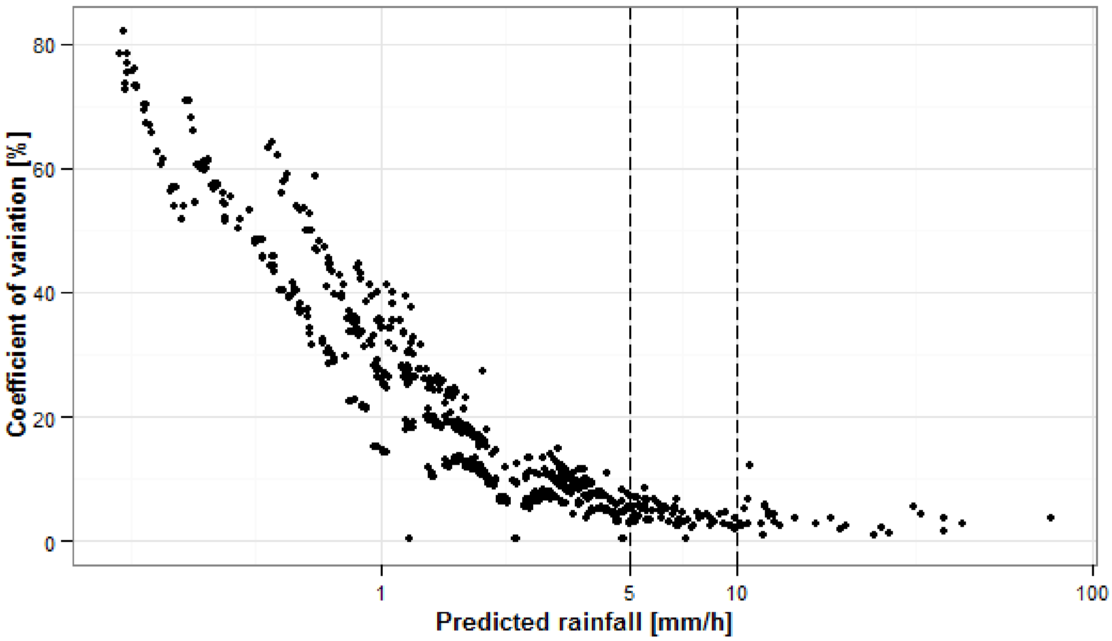


Figure 9~~8~~: AARI ~~standard deviation (~~prediction error~~)~~ CV (%) values against predicted AARI for averaging interval of 5 min.

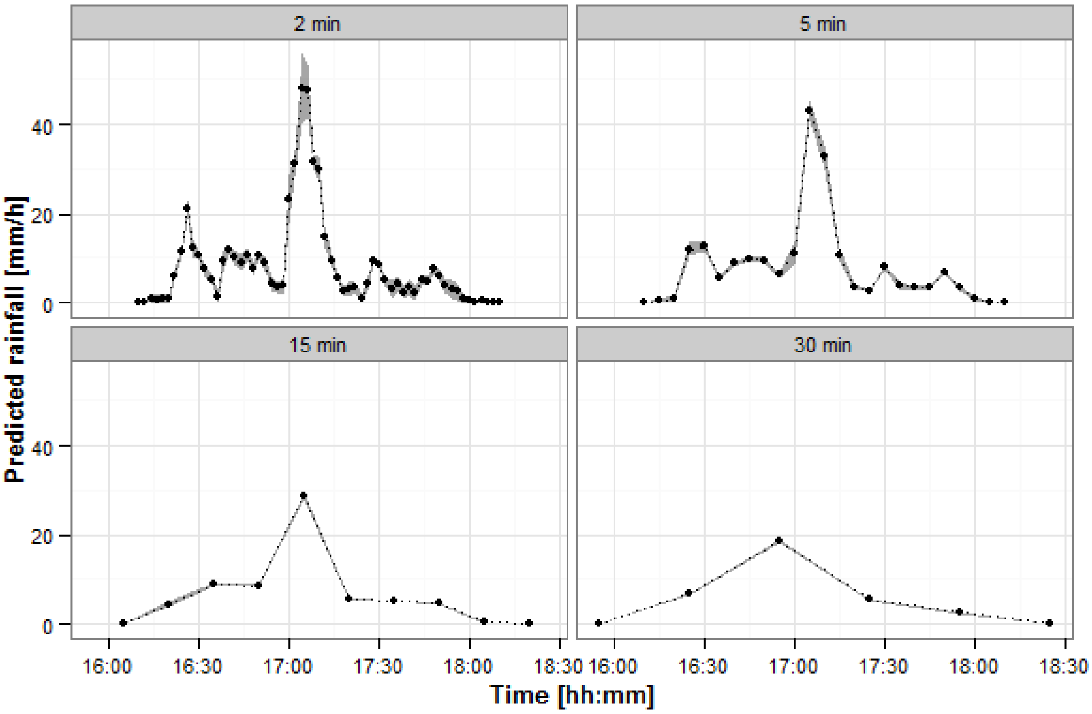
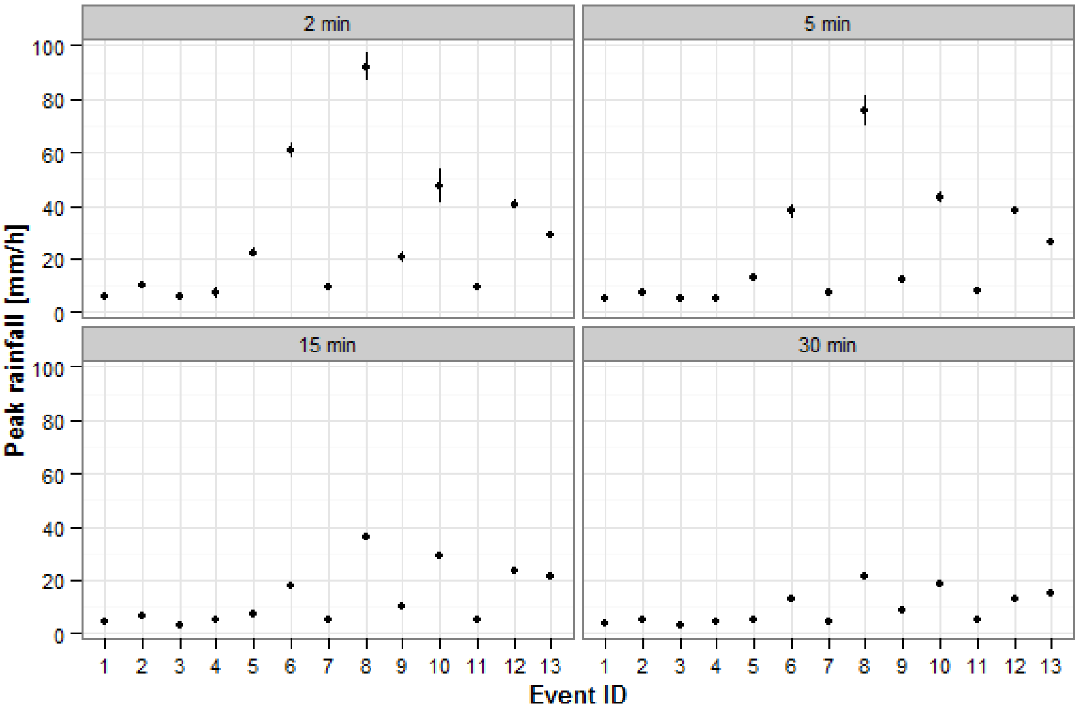


Figure 109 ~~:~~: Predictions of AARI (indicated by points) together with 95 % prediction intervals (indicated by grey ribbon) for rainfall event 11 for different averaging intervals.


2 min
5 min
15 min
30 min
Peak rainfall [mm/h]
Event ID

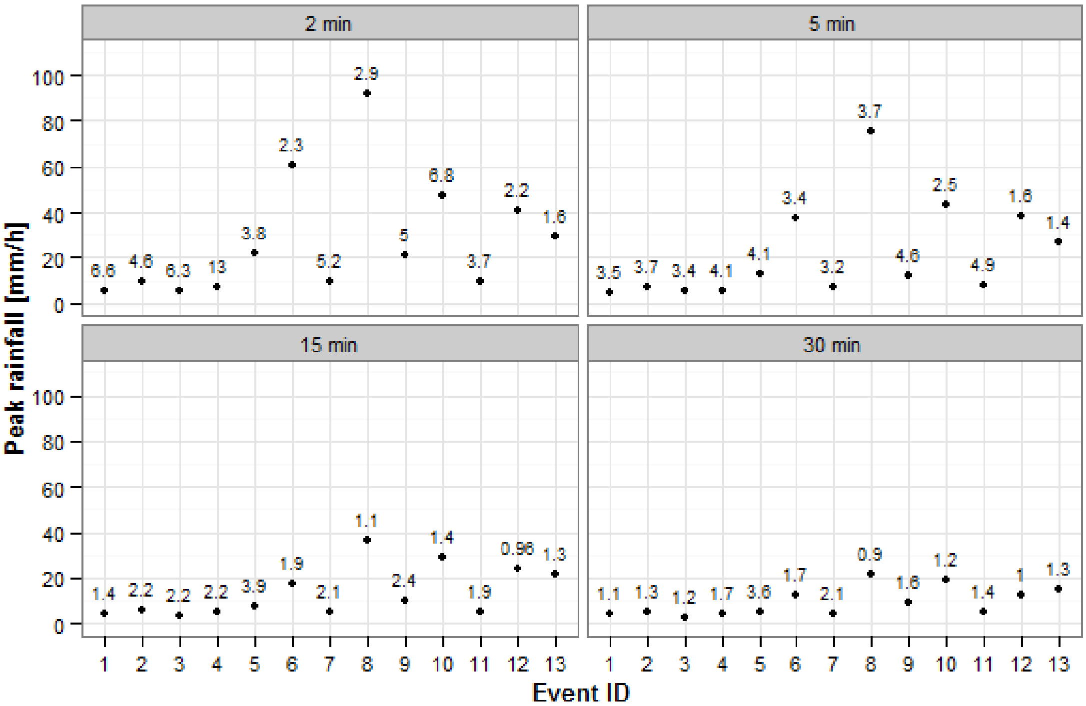
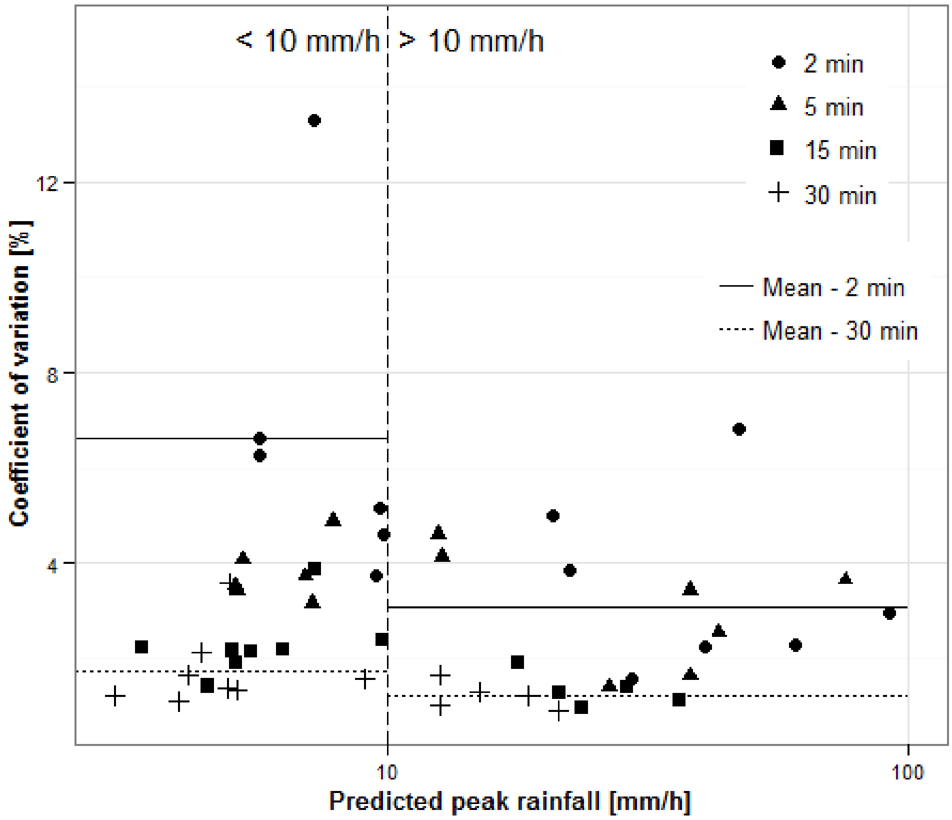


Figure ~~10~~11 : Predictions of event peaks of AARI (indicated by points) ~~and 95 % prediction intervals (indicated by vertical bars)~~together with labels indicating corresponding CV (%) values.



~~Figure 11: Coefficient of variation plotted against predicted peaks of AARI for the events presented in Table 1~~