Author's Response – Reviewer 1

Dear Reviewer,

first of all, we thank you for the useful comments and suggestions. We have considered these carefully in preparing the revised version of the manuscript.
Below we present a point-by-point answer to each comment.

"- Clarify more precisely the added value with regards Licznar et al 2011 a and b"
In the revised version, we underline the novelties of our study.
In the Abstract we have specified that
"We address the following issues partially investigated in literature: 1) the calibration of microcanonical cascade generators in conditions of short time series (say, 2.5-5 yrs.); 2) the identification of the probability distribution of breakdown coefficients through ranking criteria; 3) the variability among the gauges of the monitoring network of the empirical distribution of breakdown coefficients.
In particular, 1) we introduce an overlapping moving window algorithm to determine the histogram of breakdown coefficients, and compare it with the classic non-overlapping moving window algorithm; 2) we compare the 2N-B distribution, which is a mixed distribution composed by two Normal (N) and one Beta (B), with the classic Beta distribution to represent the breakdown coefficients using the Akaike information criterion; 3) we use the cluster analysis to identify patterns of breakdown coefficient histograms among gauges and timescales."

"- With regards to cluster analysis, it not straightforward what "non-obvious" information it brings. Authors should clarify the purpose of this study and why this methodology was tested on the various rain gauges of this rather small area."
We agree with this comment and we address it in more detail at the beginning of section 2.3:
"To our knowledge, until now, the variability of MCM generators among a group of gauges was investigated comparing the value of the parameter of Beta distribution (Molnar and Burlando 2008). Here, we preferred to compare directly the empirical distribution of BDCs instead of the parameters of the theoretical distribution, possibly biased by fitting errors. We have encountered the same problems found in the implementation of statistical tests due to the large sample size. For this, we have used the cluster analysis technique to compare the shape of BDC histograms among the stations of the monitoring network in Warsaw, and with other Polish and German gauges."

"- Results should be discussed much more in-depth with some physical interpretation rather than simply stating output of statistical analysis."
We agree with this suggestion and we add:

- explanation about the better performance of overlapping moving window algorithm (see section 2.2, 4th paragraph);
- explanation about the distribution change from B to 2N-B (see section 3.2, 2nd paragraph);
- plausible explanation of differences in the behavior of gauges R15 and R25 by their location in specific sub-urban environments. For this purpose we prepare the much

detailed plan of gauges location with respect to urban fabric (see Fig. 1) and in addition we discuss the issue of urban heat island (see section 3.4, last paragraph).

"- Why some spatial (some elements are in Rupp et al, 2012) or even spatio-temporal downscaling has not been tested with this data set."
Here we concentrate our attention only on precipitation time series disaggregation.
We do not believe that our set of data, 25 gauges allows for development of spatio-temporal downscaling model. We work on project oriented on spatio-temporal downscaling model development for Warsaw city but we plan to realize it by coupling radar and gauge data.

"- Some methodological aspects should be justified, especially the overlapping methods, statistical significance of the differences in cluster analysis..."
We agree with this comment and in version of manuscript we underline the fact that cluster analysis is used for objective comparison of shapes of empirical BDCs. Moreover in Fig. 16 and 17 we add additional curves obtained for synthetic (average) series generated by outlining MCMs for gauges R15 and R25. In text we comment that differences retrieved by cluster analysis are resulting in significant changes at statistics of generated precipitation series.
Our special attention is focused on the better explanation of the overlapping moving window algorithm. For that purpose, we prepare the new version of Fig. 6. Moreover we test the performance of this algorithm and compare it with the ordinary overlapping non-moving window algorithm for additional 25-year rainfall dataset for gauge R7. Please refer to Figs 11÷14 and their discussion at section 3.3.

Detailed comments

Abstract:
"Last sentence : "The cluster . . .. time scales"; I believe this is overstated with regards to the content of the paper."
This sentence was removed.

1) Introduction
There is a lack of references and "context" statements in the introduction:
"- 1st paragraph : some references about urban hydrology and required time step should be added (ex among many Berne et al 2004, Gires et al. 2013)"
Suggested citations were added, all the Introduction was significantly rewritten.

"- 2nd paragraph: are you considering real time?"
In new version of manuscript in introduction we discuss differences between data necessary for RTC systems and standard evaluation of probabilistic functioning of urban drainage systems (eg. calculation of overtopping frequencies, etc.) . We underlined the fact that for standard engineering calculations/simulations of urban drainage most often only historical rainfall time series are used. And as a result of it, point rainfall disaggregation is interesting for urban hydrology practice.

"- 3rd paragraph: micro-canonical cascades are mentioned, but I think that other types (such as macro-canonical ones) should be mentioned, and you should justify the use of one type rather than the other ones."

We underline the fact that MCM is only one of many alternatives of disaggregation. We do not hide that it is not the best alternative from perspective of statistical properties of synthetic time series. But we underline the simplicity of MCM generators fitting and we decide to use it for study of local rainfall dynamics variability.

"- 4th paragraph : in the context of cascades process for rainfall the "beta model" often refers to a dead or alive cascade process to model the rainfall occurrence (Over and Gupta, 1996). May be it should be briefly clarified that you are not dealing with this here."
Done, we clarify the obvious difference and we add citation of Over and Gupta (1996) work.

2) Data
2.1)
"- p.5 l.16-22: the setting of the gauges seems quite impressive for an urban area !"
Please note, that finding good places for installation of gauges was only possible due to the fact that the MWSSC in Warsaw operates a vast number of local water intakes, water and sewage pumping stations. All these installations due to sanitary standards has to occupy some terrain with green arrears around serving as buffers e.g. for odors spread. In additions all facilities are fenced and guarded for safety reasons.

"- p.6 l.6 -8: the notion of "step error" should be clarified, and the results displayed Fig. 2 commented in the text."
We explain the "step error" in more detail (section 2.1. 2nd paragraph).

"- p.6 l.8 -10 : it is not clear why the data was aggregated to 5 min, especially for heavy rainfall events which are the trickiest to handle. By the way the discrete cascades that are implemented also exhibit some "step-like" features."
We explain it in the revised version of the manuscript. We consider data aggregation to 5-min resolution to remove, or at least reduce, the step error (see section 2.1. 2nd paragraph).

"- p.6 l.16-18 : if the network is compared with other cities, an approx. estimate of the area covered area by these networks should be added."
Done

2.2)
"- For readers not necessarily familiar with the cascade framework, I think a schemewould be quite helpful."
Done, see Fig. 4.

"- p.7 l.1-6: From my understanding the final number of time steps is 128 (n=7), but 1280/128 is different from 5 min which is mentioned in the text. Please check the consistency."
It is corrected. It was our obvious error, for which we apologize.

"- p.7 l.14-17 : I believe the discussion with regards to long lasting debate micro/macro – canonical cascades should be extended. Indeed micro-canonical cascades can limit the appearance of extremes. A discussion on this topic in a framework of urban hydrology applications can be found in Gires et al. 2012, details are also in Schertzer and Lovejoy 1989."

Discussion is added, we do not hide strong limitations of MCMs. Beyond the suggested literature, we mention Lombardo et al. (2012) highlighting the problems with the autocorrelation structure of rainfall series disaggregated by MCMs.

"- p.9 l.1-6 : The procedure seems very artificial... could you justify more precisely it (tests with some simulations, or extend the discussion to previous work using it)."
We explain the randomization procedure more precisely and we illustrate its influence on calculated BDCs distributions in Fig. 5.

"- p.9 l.7-20: the problem is that the suggested method still use the same data... it would be more convincing if justified more carefully (may be with some tests on simulations, see also comments in section 3)."
We test the overlapping moving window algorithm on additional 25 years of rainfall data. Please refer to results in Figs. 11-14 and their discussion.

2.3)
"- The probability distributions of the BDC for each rain gauge could also by compared between themselves using statistical tests to see whether they are statistically significantly different."
We explain our decision to use cluster analysis instead statistical tests in 1st paragraph section 2.3.

- Could you clarify the quantification of the distance between two clusters, and especially criteria for statistical significance of the differences between two clusters (in the following all the rain gauge of Warsaw seem very close...)
We clarify that the distance plotted on dendrograms is only a measure of similarity of patterns if is small or dissimilarity if it is large. We confirm that cluster technique is fully objective in retrieval of similar patterns.

3) Results and discussion
"- I would suggest to divide this section into several sub-sections to clarify the presentation of the results"
Done.

"- Fig. 5 and 6 : improve readability"
Done.

"- p.12 l.8-12 : the explanation is too short. A solution to actually illustrate the relevance of the technique could be to use a long series to estimate the histograms and then select only a portion of your series and re-estimate the histograms with your new method."
Done, please refer to results in Figs. 11-14 and their discussion. We appreciate this suggestion.

"- p 12 l13-16 /p13 l1-9: there are neither justifications for the choice of the distribution nor quantification of the quality of the fitting. Fitting 5 parameters with this histogram seems a lot. It should be added."
We explain why we could not use statistical tests. We propose to use AIC criterion for this. Please refer to Tab. 1 and its discussion.

"- p 12 l 17-19 : please clarify this sentence"
Done.

"- p13 l 11: I think "up" should be changed to "down""
Done.

"- p10 l10-22 : there is only a description of the parameter evolution with scale; a physical interpretation should be added."
Most probably***" p13 l10-22"*** a physical interpretation is supported, please refer to section 3.2 4th paragraph.

"- p13 last paragraph: you should give more details on the disaggregation processes"
For the brevity we refer interested readers to previous studies of Molnar and Burlando 2005, Licznar et al. 2011a and b.

- fig 9 : I did not found it very easy to read. If this is point shared by other referee, may be you should change it.
We decided to leave this figure, however we comment it on text (see section 3.3 6th paragraph).

"- comments on Fig 10 (p14 l5-18): "behaviour . . . is very similar", this is not obvious from the curve, especially since it is plotted in log-log, the differences (overestimation on the whole range) in physical unit must be quite significant, the statement should be much more justified; for the extremes the discrepancies are even greater and should be discussed more in details (potential issues with the model but also the measurements for extreme events ...); why is the validation performed at a single scale whereas you are working in a cascade framework; may be other comparison tool should be implemented."
We agree and we admit discrepancies between observed and synthetic series. Please refer to section 3.3, last paragraph.

"- Fig 11 : again why validation is done at a single scale?"
We concentrate on single smallest timescale of 5-min because this is the most important scale for urban hydrology. We clearly state it in the revised version of the manuscript and we add that the same practice is used by other authors as eg. Molnar and Burlando (2005).

"- p14 l25 – p15 l14 : it should be clarified what knowledge does the rather "heavy" methodology brings, given that the conclusions as mentioned in the text are rather obvious. What is the statistical significance of the observed differences?"
Results could be judge as obvious, however remain of practical importance. Still a common practice in Poland is to use airport gauge data for urban drainage modelling. We discourage this practice and some statistical effects of such practice we illustrate in revised versions of Figs. 16 and 17. Finally we recognize the necessity of future research exploring the effects of poorly assumed parameters of local precipitation models based on airport gauge.

"- p15 l14- p16 l16 : some explanations and physical interpretations should be added."

We add this explanation on the basis of gauges installation (Fig. 1) and discussion of most probable effects of urban heat island, giving references to papers of Peng et al. (2011), Kłysik and Fortuniak (1999) and Fortuniak et al. (2006) (section 3.4 see last paragraph).

Author's Response – Reviewer 2

Dear Reviewer,

first of all, we thank you for the useful comments and suggestions. We have considered these carefully in preparing the revised version of the manuscript.
Below we present a point-by-point answer to each comment.

"1) I can hardly find the novelty of this paper with respect to Licznar et al. (2011a, b) and Rupp et al. (2012). I mean that the vast majority of the concepts presented in the paper under review seems to have been already discussed elsewhere."
In the revised version, we point out clearly the novelties of our study.
In the Abstract we have specified that
"We address the following issues partially investigated in literature: 1) the calibration of microcanonical cascade generators in conditions of short time series (say, 2.5-5 yrs.); 2) the identification of the probability distribution of breakdown coefficients through ranking criteria; 3) the variability among the gauges of the monitoring network of the empirical distribution of breakdown coefficients.
In particular, 1) we introduce an overlapping moving window algorithm to determine the histogram of breakdown coefficients, and compare it with the classic non-overlapping moving window algorithm; 2) we compare the 2N-B distribution, which is a mixed distribution composed by two Normal (N) and one Beta (B), with the classic Beta distribution to represent the breakdown coefficients using the Akaike information criterion; 3) we use the cluster analysis to identify patterns of breakdown coefficient histograms among gauges and timescales."
Similarly we have stressed the novelties also at the end of the Introduction.

"2) The Authors state (p. 5253, lines 14-18): "(: : :) when some local precipitation datasets are accessible, questions and doubts about the representativeness and reliability of data arise. Synthetic time series, generated from precipitation models, could be considered as probable precipitation scenarios to feed hydrodynamic urban drainage system models". Actually, this is a general comment, but it is very important in my view. I believe that the dangers for science become most evident when models (abstracts of more complex real-world problems, generally rendered in mathematical terms) are assumed to be more reliable than observational data (even if uncertain). This is particularly the case for statistical models like the model proposed in the paper under review."
We agree with this comment, for sure our previous statement might suggested that we disregard the good quality observational datasets and we proposed to substitute them completely with synthetic time series. It was not our intention. On the contrary, our interest in developing rainfall time series disaggregation models in Poland originates from the fact of almost complete lack of access to high resolution precipitation time series by engineers. In new version we explain it (see lines 76-95 –section 1, paragraphs 3 and 4).

"3) Please, note that there is ongoing discussion about the inappropriateness of multiplicative random cascade models in providing credible simulations of rainfall time series. For example,

Lombardo et al. (2012) show that the autocorrelation function of the simulated series corresponds to a non-stationary process simply inherent to the model structure (see also Mandelbrot, 1974; Over, 1995; Veneziano and Langousis, 2010). The Authors should investigate whether or not their model is affected by this problem, because the reproduction of the autocorrelations as well as marginal probabilities are major requirements for statistical models."

We would like to thank you for focusing our attention on important communication of Lombardo et al. (2012). In the new version of the manuscript, we do not hide the imperfections of MCMs originating first from their discrete character and microcanonical assumption of rainfall depth conservation across cascades levels. Simultaneously, we would like to stress once more, that the goal of our study was not to present MCMs as simple and effective rainfall disaggregation models for urban hydrology (which was analyzed in numerous previous studies). Our idea was to use the simple framework of MCM generators assessing the precipitation variability among gauges of urban precipitation monitoring network.

"4) The Authors are also encouraged to study the statistical implications of the so-called "overlapping moving window algorithm" for the calculation of the breakdown coefficients (see paper eq. (2)). In other words, the Authors should investigate the joint distributional properties of their simulations when using the classical non-overlapping and their overlapping methods. I guess the dependence of the generated rainfall at a certain time interval with the time intervals preceding and following it may change significantly for the two methods."

We appreciate this comment and made a specific analysis to address it. This analysis was performed using an additional 25-year rainfall dataset available for gauge R7. Results of it are presented graphically in Fig.s 11-14 and their discussion is given in section 3.3.

"5) It is well acknowledged that parsimony is a very important and desirable property in good modelling practice. However, the model proposed is over-parameterized and thus not parsimonious because it uses a somewhat artificial probability distribution for the breakdown coefficients (i.e. 2N-B distribution, which combines two Normal (N) and one Beta (B) distribution). Then, being a complicated discrete-time model, it does not correspond to a continuous time process; but natural processes typically evolve in continuous time.

6) Furthermore, as the Authors have posed the 2N-B distribution as an assumption, rather than deriving it theoretically from other principles. Then, they should apply a goodness-of-fit test to justify their choice of the 2N-B distribution.)"

We do not agree that 2N-B distribution is an over-parameterized model, especially for smallest timescale. For this we have calculated the AIC for both 2N-B and B distributions. Despite 2N-B is penalized for having more parameters, 2N-B model is characterized by lower AIC values, than the Beta distribution model (see Tab. 1). In addition in Fig. 8, we report both the distributions 2N-B and B with data to illustrate the visual comparison.

Concerning the statistical derivation of 2N-B, we will consider it in future investigations.

"7) The reason for fitting a statistical model to data is to make conclusions about some essential characteristics of the population from which the data were drawn. Such conclusions can be sensitive to the accuracy of the fitted model, so it is necessary to check that the model fits well. The main issue concerns the ability of the model to describe variations in the wider population, and this is usually achievable when there are additional sources of data against

which the model can be judged in terms of the reproduction of the autocorrelations as well as marginal probabilities. By contrast, the Authors seem to judge the accuracy of their model in terms of its agreement with the data that were actually used to calibrate it. This limits the value of their results."

Thanks, we will consider it in future investigations.

Author's Response – Reviewer 3

Dear Reviewer,

first of all, we thank you for the useful comments and suggestions. We have considered these carefully in preparing the revised version of the manuscript.
Below we present a point-by-point answer to each comment.

"* It appears that the methodology is not new in comparison with the previous publications by the authors (Licznar et al., 2011a, b; Rupp et al., 2012). In Licznar et al. (2011a, b), similar methodology was applied but to other rain gauges in Poland and Germany; and Rupp et al. (2012) made use of the same rain gauge data of Warsaw??
Can the authors clarify the scientific innovations of the current paper more clearly?
These should better demonstrate that this paper is more than a new case study (applying existing methodology), hence deserves publication in a journal as HESS."
In the revised version, we underline the novelties of our study.
In the Abstract we have specified that
"We address the following issues partially investigated in literature: 1) the calibration of microcanonical cascade generators in conditions of short time series (say, 2.5-5 yrs.); 2) the identification of the probability distribution of breakdown coefficients through ranking criteria; 3) the variability among the gauges of the monitoring network of the empirical distribution of breakdown coefficients.
In particular, 1) we introduce an overlapping moving window algorithm to determine the histogram of breakdown coefficients, and compare it with the classic non-overlapping moving window algorithm; 2) we compare the 2N-B distribution, which is a mixed distribution composed by two Normal (N) and one Beta (B), with the classic Beta distribution to represent the breakdown coefficients using the Akaike information criterion; 3) we use the cluster analysis to identify patterns of breakdown coefficient histograms among gauges and timescales."
Similarly we have stressed the novelties also at the end of the Introduction.

"* The authors propose to use of overlapping moving windows for the calculation of the empirical histograms and calibration of the small-scale rainfall generators (microcanonical cascade generators in this study) for use in urban hydrological applications. This indeed leads to more smooth histograms and parameter calibration results, as shown in the paper. But, it does not meet the shortcoming of local precipitation data shortage (see motivation of the authors on page 5253 line 13) !! The histograms and calibrations are still based on the same short rainfall record. Given that rainfall statistics may strongly be influenced by climate oscillations (e.g. at decadal time scales; see Willems, 2013), precipitation statistics derived from short records may be biased. They may systematically differ from the long-term statistics, which is the main problem of the local precipitation data shortage. This is not solved by the proposed use of overlapping moving windows. I suggest that the authors make this clear in the paper."
We are in special debt due to this comment. In order to address it, we have made an additional analysis using 25-year long data set of rainfall time series available at gauge R7. For results of these experiments please refer to Figs. 11÷14 and their discussion in section 3.3. Experiments results are quite promising in our judgment, since the repeatability of fitted BDCs distributions among 5-year periods and all 25-year period is high. We suppose that a possible explanation of it

is the fact that we work only on subdaily timescales and the variability of extreme precipitation (proved by Willems, 2013) does not influence strongly all the population of BDCs. Still however we recognize the need of additional analysis on this topic, made for better resolution and much longer time periods. Dataset of such quality are rare and completely inaccessible in Poland.

"Additional minor comments (mainly textual suggestions), follow below. The English language needs strong polishing. Some sentences that need to be rephrased are mentioned below, but there are many more."
We have fully revised the text to improve its presentation.

"Page 5255 Line 4: Motivation is given here for studying the spatial variability in the MCM generators. The authors claim "a number of MCM generators varying in space" need to be built, but what about the spatial correlations? These need to be addressed as well."
In the Introduction of the revised manuscript, we recognize the need of spatio-temporal data in urban hydrology. At the same time we underline the fact that in practice most of the currently used urban drainage models is still oriented on point time series. In addition, in practice it is used only a single time series for the whole city in the calculation of overtopping frequencies or combined sewage overflows.

"Page 5260 Lines 13-16: It is unclear why this sampling in 100 points was done? Why not simply considering the full empirical distribution?"
In the application of the cluster analysis, we have considered the empirical BDC distribution described by 100 attributes\components to be used to calculate the distances between each couple of gauges. We have clarified this point in the revised version of the manuscript (see section 2.3).

"Page 5264 Lines 14-15: "this overproduction does not disqualify the use . . . for practical engineering tasks . . .": why not? Did you quantify the impact of the overestimation in rainfall on, for instance, sewer or stormwater reservoir design?? If not, I suggest to weaken that statement. What could be done here is to quantify the bias in return period of the design value, if the biased MCM-based rainfall data would be applied."
We agree with this. Our old statement is not only weaken but completely deleted.

"Figures 7 and 8: I suggest adding the confidence interval limits to the points in the figure. This would allow to check whether the parameter estimation uncertainty can explain the higher value for the parameter a for timescale 8 in comparison with timescale"
Done for former Fig. 7. In case of former Fig. 8, it is impossible as confidence interval limits for parameter p_0 are not calculated.

"Figure 10: highest rainfall intensities show strange behavior (constant exceedance probability above a given precipitation threshold; + discontinuous behavior). Please discuss."
Explained, please refer to the last paragraph of section 3.3.

"Page 5252 Line 4: "resolution": please clarify time and/or space ??"
Done

"Page 5252 Line 4: "probabilistic assessment": What kind of probabilistic assessment? Of the network itself? Or. . .? Clarify to the reader."

Done, we mentioned the calculation of overtopping frequencies or combined sewage overflows and comparing results with recommendations given by Euro code EN-752.

"Page 5255 Line 3: "answer to the question": remove "to"
Removed

"Page 5255 Line 3: "is it sufficient a single . . .": check language"
Removed

"Page 5256 Line 5: "step response error": I don't think it is clear to the reader what this means. Explain more."
Done. For explanation refer to 2nd paragraph of section 2.1.

"Page 5257 Line 22: "we do not however limit": check language"
Page 5258 Line 23: "gauges are however": "are" is missing
Page 5258 Line 26: "making difficult the identification": check language"
Page 5267 Line 15: "cluster analysis technique": remove "technique"
Page 5267 Line 27: "it is enough one single": check language
Page 5268 Lines 3-4: "the quite common practice": sure ??
Page 5268 Line 4: remove "gauges data adoption"
Page 5268 Line 5: remove "fully"
Page 5268 Line 5: "to find more clear": add "more"
Page 5268 Line 6: remove "our observations"
Page 5258 Line 17: "to be equal to 0": add "to"
Page 5258 Line 23: "gauges are however": add "are"
Page 5264 Line 14: change "disqualifies" to "disqualify"
Done.