Review by anonymous Referee #1

The objective of the paper is to compare different versions of a rainfall disaggregation model that aims to produce high resolution times series (10min) from daily time series of precipitation. The versions is applied on a set of 24 recording stations. The main challenge for the author is to reproduce well the autocorrelation that was observed in the measurements.

This issue is obviously of high interest for the specific configuration of hydrological design in rapid response catchments, especially in urban areas. The manuscript however suffers from a number of limitations. The model is very rough and some basic assumptions of it do likely not hold. This make the model likely poorly relevant. On the other hand, other disaggregation models have been proposed in past years and the present work should at least include some in their comparison (it just compares variants of the present model but those are not really convincing). The description of the model / results is often rough and requires improvement. I could not understand what is done with some model variants and with the "resampling" process.

I would thus suggest rejection with the recommendation for a resubmission after clarification / improvements. For the Editor, it is not comfortable to have a numbering of lines which is reset to zero at every page. A unique numbering would make the review more convenient.

I thank the reviewer for the effort and the time spend on this manuscript. The reviewer points out the importance of the study for urban hydrological applications. His/her major concerns are an insufficient model description and missing comparisons of the results with other references. For both concerns detailed examples are provided by the reviewer. A point-by-point reply can be found below. All page and line numbers refer to the original submission.

Specific comments:

1. The objective of the paper is to compare different versions of a rainfall disaggregation model that aims to produce high resolution times series (10min) from daily time series of precipitation. The versions is applied on a set of 24 recording stations.

This comment is copied from the general part of the review for clarification. The disaggregated time series have a final temporal resolution of 5 min, not 10 min. Although the observed time series have a temporal resolution of 5 min (see e.g. P4L20, P8L3-5, P13L15-16,…), it seems the information was missing at other parts of the manuscript. Hence, the information was added in the abstract (P1L9-10):

"In this paper two cascade model modifications are analysed regarding their ability to improve the autocorrelation **in disaggregated time series with 5 minute resolution**."

and in the introduction of the method section (P5L5):

"The overall aim of this investigation is the improvement of the autocorrelation $r_{t1,t2}$ of the generated time series **with a temporal resolution of 5 minutes.**"

2. In the introduction, the authors mention that cascade models underestimate autocorrelation. This is not always true. See the comparative study of Hingray and Ben Haha (2005). They present results obtained for different disaggregation models including the so-called pattern-based microcanonical model. The reproduction of the autocorrelation with this model is almost perfect. This model should likely be included in the present work for comparison.

The reference of Hingray and Ben Haha (2005) was not included in the review of autocorrelation results so far. However, it was not the intention to "hide" results which do not fit to the motivation for the current study. Indeed, it is pointed out that under- and overestimations can occur by the application of a micro-canonical cascade model (P2L29-P3L4) due to several reasons (P3L5-10). Hingray and Ben Haha use for their investigation the time series of only one rain gauge and over a short range of disaggregation levels. Despite the low representation using only one rain gauge, by the few disaggregation levels it can be expected to achieve good results in terms of autocorrelation. The

disaggregation starts with observed hourly values (so on this level the autocorrelation is "perfect", because it results from observations) and ends after four disaggregation steps (start: 1 h, 30 min, 15 min, 7.5 min…even less, since afterwards the time series is aggregated back to 10 min time steps). To keep the "perfect" autocorrelation over four disaggregation steps is much easier than over 7 disaggregation steps (start: 1d, 8 h, 4 h, 2 h, 1 h, 30 min, 15 min, 7.5 min [5 min]) as done in the current study. It can be questioned if the autocorrelation would be as good represented by a disaggregation of daily values down to 5 min values. Also, the results are shown in Hingray and Ben Haha only for lag-1 autocorrelation. As shown in the current study, although a good fit can be achieved for lag-1, strong underestimations can occur for other lags (see e.g. Fig. 6, Method C). Hence, I have implemented the reference into the literature review, but do not attempt to apply the disaggregation model additionally for the sake of comparison (especially in accordance with comment 8 of the reviewer, that the manuscript is already complex).

The following sentences were implemented after P3L4:
"A good representation of the lag-1 autocorrelation was achieved by Hingray and Ben Haha (2005) with two micro-canonical cascade models. However, since only four disaggregation steps were applied (from hourly to 7.5 min time steps) it remains unclear, if the good representation would have been achieved for more disaggregation steps."

3. Other models mentioned in the introduction perform also relatively well for AC reproduction. The best ones should be at least included in the present comparative work.
I double-checked the references cited in the introduction regarding autocorrelation results. As far as I can see, the references not discussed do not include autocorrelation results (e.g. Molnar and Burlando, 2005, Licznar et al. 2011, 2015). The autocorrelation results of Paschalis et al. (2014) are not discussed, since they result from combination of different rainfall generators, thus the error cannot be assigned exactly to the micro-canonical cascade model or the others. However, for the majority of rain gauges and rainfall generator combinations, the lag-1 autocorrelation is also underestimated. Since the reviewer mentioned no reference explicitly, it remains unclear for me which results should be discussed additionally.
A comparison with other rainfall generators is way beyond the scope of the current study, which aims at improving the general model structure. The basic idea behind the preceding cascade model and the resampling approach can be transferred to other micro-canonical cascade models, so this study is considered more as a methodological development than a best-choice model comparison.

4. The author state in their introduction (p.1 – ln 21/22) that "Since time series with 1280 minutes do not exist as observations, these studies [the studies related to the other models] are rather theoretical than practical from an engineering point of view." It seems to be the reason why the author disregarded the related models. This statement does obviously not hold. All the suggested models can be of high practical interest even from an engineering point of view. You just have to push the disaggregation process at the right temporal level (as the author does it actually in the present disaggregation process > disaggregation to 2.5min + reagregation to 5min).
The reviewer is right, a disaggregation could have started also on daily values applying the same disaggregation model until the disaggregated time series have a temporal resolution below the desired resolution and then a kind of transformation can be applied. However, this was not done in the mentioned studies. As shown in Müller and Haberlandt (2015) the kind of transformation has a significant influence on the disaggregation results, e.g. over- or underestimation of the rainfall intensities and especially on extreme rainfall values. Hence, a fair comparison is not possible. However, it was not the intention to disregard these disaggregation models. Hence I added the following subsequent sentence:
„Of course, by the application of an additional transformation process a desired temporal resolution can be achieved, whereby the transformation process affects the characteristics of the disaggregated time series."

5. Relevance of the present model: Variant A : in the first disaggregation step, the daily amount is distributed uniformly on the wet times steps (the wet 8hrs time step can be 1, 2 or 3). This is obviously not realistic at all. The model should relax this strong assumption which obviously cannot be validated from observations (or if it can, this has to be shown)

The reviewer is right, the uniform distribution over the as wet identified 8 h-time steps is a strong assumption. Although the number of wet 8 h-time steps is realistic, the resulting rainfall amounts are maybe less realistic. There are two reasons for this assumption.

First, the number of wet 8 h-time steps tells us something about the genesis of the rainfall event. If it is only one wet time step, it is likely a convective event and hence the whole daily rainfall amount is put into this single 8 h-time step. If the rainfall event lasts longer than 8 h, it is likely a stratiform event, so a long-lasting and less intense event (with often more or less uniform rainfall occurrence). The uniform distribution can be found only on the 8 h-level, on finer temporal resolutions intensities vary due to the b=2-splitting so the resulting final time series does not show any uniform rainfall intensity distribution anymore.

Second, the assumption with the uniform distribution demands only a few parameters. For method A, two parameters are required (P(0/0/1) and P(0/0.5/0.5)). The only other tested approach with a b=3-splitting in the first disaggregation step was introduced by Lisniak et al. (2013), who use for example 8 distribution functions for a splitting with 3 wet 8 h intervals. To keep the cascade model parameter parsimonious, the uniform distribution has been chosen. Also, in previous publications (Müller and Haberlandt, 2015, 2018, Müller-Thomy et al., 2018) with this assumption good representation of rainfall characteristics have been achieved.

6. The amount generated at the 7.5min time step are distributed uniformily on 2.5min time steps and then aggregated back to 5min. The uniform assumption is again really strong. Why don't you do the disaggregation to a finer resolution (3.75 min) and then aggregate back (using the same disaggregation model than the one you used for the previous time step) ?

The idea of the reviewer is to apply a different transformation to achieve a final temporal resolution of 5 min. In a previous publication Müller and Haberlandt (2018) tested transformations at 3 different temporal resolutions, starting with 15 min, 7.5 min and 3.75 min, for the same data set. The transformation starting with 7.5 min led to the overall best results.

An alternative would be to continue with the disaggregation to a very fine temporal resolution (few seconds) and then aggregating the time steps to 5 min, so that there is no need for any transformation. However, since a bounded cascade model is applied, the parameters for this so-called 'fine-graining' process have to be estimated from observations with the same temporal resolution, which are not available for the current study.

7. Variant C : Clarify. I can not understand how it works. The scheme of Figure 2 is very unclear. I do not understand at all.

For a better understanding, the disaggregation level and temporal resolutions have been added to Fig. 2 and colours have been removed. Unfortunately, without a certain comment what is unclear for the reviewer (the position definition, the time step indices, the increase in number of position classes or what exactly in the figure remains unclear) the manuscript can hardly be improved further in this point.

8. Avoiding time steps with too small rainfall intensities. Two approaches are considered to tackle this issue. This makes the paper rather complex. The results obtained with both approaches differ not a lot. I would suggest to keep only one of both (The one that mimic the measurement device would be likely relevant).

The reviewer suggests to leave the MRA approach out to reduce the complexity of the manuscript. However, both approaches represent possible solutions and none of them has been tested before to the authors knowledge. To reduce the complexity, only the MMD approach has been applied for the resampling investigations. I prefer to keep both approaches in the current study, especially since reviewer 2 identifies this issue as the more fundamental one. The comparison of the two approaches

is also useful for other researchers, because they know that the outcome is very similar and don't have to carry out a study on their own.

8. Resampling. What do you do with the resampling step? Please clarify.
The introduction of subsection 3.3 has been extended to clarify, what is done in the resampling algorithm (which is explained more in detail afterwards). The following sentence has been added after P10L7:
„In a resampling process, two elements (here: relative diurnal cycles of the disaggregated time series) are swapped to improve an objective function (here: minimizing the deviation of the autocorrelation function of the disaggregated time series from the observed time series)."

9. What is the archive of observed structure you use ? Please clarify. Give perhaps a graphical scheme for illustration.
For the resampling, there is no archive of observed structures. The structure of the disaggregated time series results solely from the disaggregation process. After the disaggregation, two relative diurnal cycles of the disaggregated time series are swapped with the aim to improve the autocorrelation function (see point 1 of the resampling scheme, P12L2-3).

10. P10 ln 12. Which structure ? what are the volume classes ?
The reviewer refers to the sentence: "The structure of position and volume classes in the disaggregated time series generated by the cascade model should be conserved." Indeed, the sequence of position (e.g. starting, enclosed, ending, isolated, see Fig. 3) and volume classes combination (lower and upper volume class) of a day defines the structure of the disaggregated time series.
Both type of classes are introduced on P7L15-18.

11. P10 ln 16. "Restriction b) is fulfilled by swapping only relative diurnal cycles as time series elements, which does not affect the daily rainfall amount." I do not understand. Clarify.
One advantage of the micro-canonical cascade model is that the daily rainfall amounts are conserved exactly. If for the resampling absolute diurnal cycles are swapped, daily totals will differ from the original coarse time series, which has been disaggregated. To avoid this issue, no absolute values, only relative diurnal cycles are swapped. To achieve a "relative" diurnal cycle, for each 5 min time step its fraction of the total daily rainfall amount is determined. This sequence of fractions are swapped and subsequently multiplied with the rainfall amount of the other day chosen for the swap. Hence, the daily rainfall amount remains constant for each day.

12. Why should you swap structure from one day to another ?
The aim of the resampling is to improve the autocorrelation function (P10L7). Structures are not swapped, only relative diurnal cycles. The structure of the time series is kept.

13. what about the configuration where the rainfall event lasts more than one day ? do you swap 2-days structures ? if no, why not ?
If rainfall lasts for more than one day, this is taken into account by the position classes. For example, in three consecutive days there is {no rain; rain; rain}, the rainfall of the second day (starting position) can be most likely found at the end of the day after the disaggregation process due to the position dependency of the parameters. This is the content of Table 3, 4 and 5 and explained at P14L3-11. To answer the question, the relative diurnal cycle of only one day is swapped, but the long-lasting event should not be destroyed by the swap.

14. Definition of an event : P4. Ln 23. Why should you define "events". The separation of events is always rather subjective and all results would depend on the separation rules. Here, you consider that "An event is hereby defined as a wet period enclosed by at least one time step without rainfall before and after the wet period." What is the time step there ? This definition seems to be not really

relevant if it is 10mn or 1hour. A number of events present intensities interruption. We cannot consider that a break of one or 2 hours makes different events.
The reviewer questions the event definition in the manuscript. Indeed, the event definition is always subjective and from the authors opinion there is no right or wrong for this case. The reviewer argues that 5 min dryness are not long enough for a separation, but also 2 hours are not long enough in the reviewers opinion. A plausible event definition would differ between events, since it depends on the kind of event (for a convective event with a duration of few minutes or hours the separation time would be much shorter than for stratiform events, which last over several days). For this kind of validation much more information than the rainfall time series would be required, e.g. circulation patterns or air pressure time series. But this would be way beyond the scope of the current study. Since there is no "true" separation time" as the reviewer points out herself/hisself, I prefer to keep the 5 min time step for separation as an objective criteria. For clarification, the sentence has been updated to:
„The definition for a single event is according to Dunkerley (2008); having a minimum of one dry **5 min** time step before and after the rainfall occurrence."

Minor comments:

P8 : clarify : is model B non preceding model ? model C : preceding model ?
Method C is the only cascade model modification which is considered as "preceding" cascade model. For clarification purposes, the following sentence has be added after P8L18:
"Since only method C is referred to as preceding cascade model, method A and B can be considered as non-preceding cascade models."

P9 ln 2 : what is the "so-called non preceding" ??
Please see the reply to your former comment.

p7 ln 20 Clarify : how works the bounded cascade
The differentiation into "bounded" and "unbounded" relates to the range of temporal scales used for the estimation of the cascade model parameters. In an unbounded cascade model it is assumed, that the parameters are similar over the applied range of temporal scales due to a mon-fractal scaling behaviour and hence the same parameter set can be applied for all disaggregation steps (e.g. from daily values to hourly time steps). In a bounded cascade model, the parameter differ between the disaggregation levels (e.g. the parameter set applied for the disaggregation from 8 h to 4 h is different than from 10 min to 5 min) to take into account the multi-fractal scaling behaviour, so for each disaggregation step a particular parameter set is applied (which is 'bound' to the certain disaggregation step). The explanation on P7L20:
"For each step of the disaggregation process a particular parameter set is applied, which is 'bound' to the specific transition of temporal resolution."
has been extended to:
„For each step of the disaggregation process a particular parameter set is applied, which is 'bound' to the specific transition of temporal resolution. The need for particular parameter sets for each disaggregation step arise from the wide temporal range (from daily to 5 min time steps) and hence the underlying processes covered causing multi-fractal scaling behaviour."
to clarify this issue.

Equation 2 and equation 4 : what is the sum of probabilities ?
The sum for each line of the equations is '1' as mentioned at P6L1 for Eq. 2:
"The sum of the weights is equal to 1 in each split, so the rainfall amount is conserved exactly throughout the disaggregation process."
For clarification purposes, this information has been repeated for Eq. 4:
"Again, all probabilities (P(0/1), P(1/0) and P(x/(1-x))) sum up to 1. "

How many parameters have to be estimated for each model ? A table is required.
I agree with the reviewer, a new table (Table 2) has been implemented, which lists the parameter amount for each branching type and cascade model variant exactly.

Can you precise what is an event based and a continuous based evaluation ?
For an event-based evaluation first events are separated and then the characteristics of these events are determined. For a continuous-based evaluation the whole time series is considered, event-independent. The two sentences written bold has been added for explanation (at P13L1-5):
„**For an event-based evaluation, first the rainfall events are identified and then the characteristics of these events are determined.** Event-based rainfall characteristics include wet and dry spell duration as well as wet spell amount. An event is hereby defined as a wet period enclosed by at least one time step without rainfall before and after the wet period.
For a continuous-based evaluation, the whole time series is considered, without differentiation into single events. As continuous time series characteristics, the average intensity, the fraction of dry intervals and the autocorrelation are analysed.“

P 11. Ln 13-15. This statement has to be justified
The statement is justified by the citation of Müller and Haberlandt (2018). Additional explanations as e.g. deviations of flooding volume or combined sewer overflow volume would require an additional brief description of the investigated sewage system to plausibilize these values (independent if absolute or relative deviations would be shown). This would move the scope of the paragraph to a wrong direction. Hence, no additional information has been added for justification.

P11. Ln 18-19. “The return period Tn=1.5 years is assumed to be representative for typical return periods for dimensioning purposes in urban hydrology (Tn={1, 2, 5, 10 years},”. This can not be possible T15 can not be representative of other T.
The reviewer points out a sentence, which is indeed misleading. Of course, the absolute values of Tn=1.5 years are different from other return periods. The intention was to express, that an under- or overestimation of observed values for Tn=1.5 yrs is similar for these return periods, since they are close to each other from a statistical point of view. Of course, for a return period of Tn=100 years additional analysis would have to be carried out. The sentence was rephrased to:
“It is assumed that the results regarding under- or overestimation for the return period T_n=1.5 years are representative for typical return periods for dimensioning purposes in urban hydrology (T_n={1, 2, 5, 10 years}, DWA-A 531 (2012)).“

p11. Ln 23 : the amount of diurnal cycle ? what is this ?
The word “amount” was wrong here. It was changed to “number”.

p12. Ln 23 : for the sake of completeness. I do not see why this is completeness there
Without the parameter values it would be only a method description, not exactly repeatable. With the absolute values a repetition is possible, because all information are provided. However, the sentence has been changed to:
“~~For the sake of completeness the~~ The following setup was chosen for the resampling: $T_{a,start}=1*10^{-4}$, dt=0.99, K=500, M=200 and $thr_{O,auto}=1*10^{-9}$.”

p13 : what is partial duration series ??? is it a standard terminology ?
Yes, at least in Germany it is a standard terminology. However, the subsequent sentence includes the information that it is similar to the peak-over-threshold approach:
„Partial duration series are similar to the peak-over-threshold approach, whereby the threshold is defined in order to select 3 extreme rainfall events on average per year.“

p13 : ln 22 : what is “the single out of all n realisations “?

The sentence has been rephrased to:
“with *i* as control variable of all realisations *n*:“

Table 1 : what are AC values (to be given in the table)
The autocorrelation values for lag-1 have been added to Table 1.

Table 2 : can not understand what is presented there
The caption of the table was changed to “Nomenclature of dataset abbreviations used in this study“. The intention of the table is to provide an overview, which combinations of data sets were investigated in the study.

Table 3 and 4 and 5 : which model ?
The first column of each table (“position-independent”) belongs to method A, while all other columns belong to method B (“position-dependent).
The sentence:
„The resulting probabilities are shown in Table 3, 4 and 5.“ (P14L3)
has been extended to:
„The resulting probabilities are shown in Table 3, 4 and 5 (columns with position-dependent entries) in comparison to the position-independent probabilities estimated for method A (first column in each table).

what should be the sum of probabilities ?
The sum of probabilities has to be 1 as defined on P7L1-2: “The sum of the weights is equal to 1 in each split, so the rainfall amount is conserved exactly throughout the disaggregation process.“ For example, the sum of the probabilities for one wet interval, position-independent (Tab. 3, first column, first entry) P=40 %, for two wet intervals, position-independent (Tab. 4) P=35 % and for three wet intervals, position-independent (Table 5) P=25 % is ∑=100 %. Small deviations from 100 % can occur, since the mean for 24 stations is shown and hence rounding errors can occur.

Table 4 and 5 : they have the same caption !!!
The reviewer is right, for Table 5 it should have been “…three wet 8 hour interval…” instead of “…two wet 8 hour interval…”. This has been corrected.

Table 6 / 7 : what is the variability between stations ?
In Table 6 and 7 the relative errors for rainfall characteristics and extreme values are presented. These error variability differs between the investigated methods. However, I think an information about the standard deviation of the relative error or something similar is not useful for the reader and hence it was not added to the tables..

<u>Reference</u>:
Hingray, B., Ben Haha, M. (2005). Statistical performances of various deterministic and stochastic models for rainfall series disaggregation. Atm. Res. 77:152-175.
Müller-Thomy, H., Wallner, M., Förster, K. (2018): Rainfall disaggregation for hydrological modeling: Is there a need for spatial consistence?, Hydrology and Earth System Sciences, 22, 5259-5280