R2_0

This paper investigates the use of downscaling model to forecast green roofs performance in the context of climate change. It uses a downscaling approach based on multiplicative cascades. The topic is interesting and relevant for the community. However, I would not recommend to publish this paper in its current state and suggest major revisions. Indeed, it requires significant clarifications on the downscaling model. Indeed, its presentation is hard to follow and should be more detailed.

Thank you for your valuable comments. The original orientation of the paper was on developing downscaling model in order to apply them to GI models. That was the reason why the methodology was not fully detailed on the MRC model development. However, in the light of your comments, the authors agree that the methodology could be further detailed: i) by careful edits in the method section, and ii) by providing supplementary material such as model details. Please also note that one of the reasons why the python codes where not shared in the first version of the manuscript is because we plan to release it as a python package and making it available and stable does require more work than a direct code sharing.

R2_1

It notably seems that different distributions of weights are used according to the cascade step suggesting they are not scale invariant.

Indeed, the distribution is not scale invariant: the probability to get a weight equal to zero depends on time-scale (and possibly depth and temperature depending on the model). The non-zero weights follow a truncated normal distribution in which the sigma parameter depends on time-scale. In order to improve the manuscript, two main aspects will be developed: i) clarification in the method section, ii) details of the models (functions and structure) together with the python codes or a pseudo code corresponding to the models.

R2_3

The calibration process of the numerous parameters (up to 19!) needs to be explained.

The calibration procedure will be clarified. Please note that this procedure is a methodology for model development. For a wider use of the model, another methodology would be more appropriate: a more formal approach could be applied. It consists in different steps: see word file attached. The number of parameters, despite appearing high, is in fact still quite low compared to other micro-canonical cascades from 12 to 36 in total (Bürger et al., 2019) up to from 6 to 224 per disaggregation step (Müller-Thomy, 2020). The reason for that large number of parameters is that often a parameter set has to be estimated for each cascade steps. Similarly to Bürger et al., (2019), our models include timescale as a dependency, therefore there is a single (bigger) parameter set instead of a parameter set per cascade step.

The main idea of our calibration method is to first (**step 1**) calibrate for each time-scale, with moving window of depth and temperature. (**Step 2**) The timescale dependency is added by calibrating the parameter of step 1 depending on time-scale. The timescale dependency prevents for having a number of parameters at each cascade level which would lower the robustness of the model and reduce the number of parameters.

Taking the example of the MCDTS, without the time scale dependency, given 8 cascade step and 5 parameter per cascade step, there would be a total of 40 parameters. With the timescale dependency there is a total of 18 parameters. The model is them more flexible since it allows to use variable time-scale input data. The robustness of the model also improved using this procedure since for small time-scale the parameters are often noisy. See details below:

A1	Fit the proportion of zero-weight depending on time-scale to a function by non-linear least square.
A2	Given a time-scale Fit the proportion of zero-weight depending on depth to a function by non-linear least square. Fit the parameters depending on time-scale to a function by non-linear least square.
A3	Given a time-scale, given a window of temperature, fit the proportion of zero-weight depending on depth to a function by non-linear least square. Given a time-scale, fit the parameter depending on temperature to a gaussian function. Fit the parameters depending on time-scale to a function by non-linear least square.
B	Fit the distribution of non-zero weight to a truncated normal distribution on [0,0.5] with mu = 0.5 by fitting to the standard deviation of the sample.
C	Fit a function to the proportion of high weight on the side of the highest neighbour.

Model	Calibration steps
MC	A1, B
MCS	A1, B, C
MCD	A2, B
MCDS	A2, B, C
MCDT	A3, B
MCDTS	A3, B, C

R2_4

Please also clarify that what is called "observed data" for the various figures is actually simulations with observed rainfall. Am I correct?

Exactly, it is simulation based on observed fine resolution time-series. It will be clarified.

R2_5

- l. 40-44: It should clearly be stated that canonical cascades ensure conservation on average only while micro-canonical ones ensure exact conservation of intensity at each step.

It will be clarified.

R2_6

- l. 54: should MC be MRC? In general, the use of numerous abbreviations does not really help the reader. I would suggest limiting their use to words really often used in the paper.

Yes, MC refer to the first model developed in this study. The abbreviation will be reviewed and a table for abbreviation will be provided according to the suggestions of the first reviewer.

Abbrev.	Meaning	Change made	Reason
GI	Green infrastructure	-	-
MRC	Multiplicative Random Cascade	-	-
IDF curves	Intensity Duration Frequency curves	-	-
NVE	Norwegian Water Resources and Energy Directorate	-	-
MET	Norwegian Meteorological institute	-	-
S	Temporal coherence indicator at time-step i and time-scale $2j$	$S_{i,2j}$	Avoid confusion
d	Depth at time-step i and time-scale $2j$	$d_{i,2j}$	Avoid confusion

w	minimum weight at time-step step i and time-scale $2j$ from aggregation of time-step $\{2i, 2i+1\}$ at time-scale j	$w_{i,2j}$	Avoid confusion
S			-
MC	**MRC** model with only time**S**cale dependence	MRC_S	-
MCS	**MRC** model with time**S**cale dependence and **S**tochastic 2-**E**lement **P**ermutation	MRC_{S-SEP}	-
MCD	**MRC** model with time**S**cale and depth/**I**ntensity dependence	MRC_{SI}	-
$MCDS$	**MRC** model with time**S**cale, depth/**I**ntensity dependence and **S**tochastic 2-**E**lement **P**ermutation	MRC_{SI-SEP}	-
$MCDT$	**MRC** model with time**S**cale, depth/**I**ntensity and **T**emperature dependence	MRC_{SIT}	-
$MCDTS$	**MRC** model with time**S**cale, depth/**I**ntensity, **T**emperature dependence, and **S**tochastic 2-**E**lement **P**ermutation	$MRC_{SIT-SEP}$	-
PET	Potential EvapoTranspiration	-	-
AET	Actual EvapoTranspiration	-	-
SMEF	Soil Moisture Evaluation Function	Removed	Used once
E-Green roof	Extensive green roof	-	-
D-Green roof	Detention based extensive green roof	-	-
WC_i	Water content in the roof at time i	Not in table	Equation variable
P_i	Precipitation depth at time i	Not in table	Equation variable
Q_i	Discharge released by the roof at time i	Not in table	Equation variable
T_mean	Mean daily temperature	Not in table	Equation variable
C	Calibrated factor accounting for Crop factor and maximum storage capacity	Not in table	Equation variable
S_K	Smoothing factor	Not in table	Equation variable
K	Conductivity slope	Not in table	Equation variable
WC_K	Starting delay	Not in table	Equation variable
DREAM	DiffeRential Evolution Adaptative Metropolis	Not in table	Used once
RCP8.5	Representative Concentration Pathway scenario with an 8.5 W/m2 radiative forcing in 2100	-	-
NSE	Nash Sutcliffe Efficiency	-	-
VM	Variational Method	-	-

Figure 1: Review example of the different abbreviations

R2_7

- Section 2.2.1: I think there is a need to be more specific, notably for the reader not specialist. Index i and j should be consistent between equations 1 – 2 and Fig. 1. Please also clarify the range of possible values (if “i” refers to a time step then it belongs to 1… 2^n where n is the cascade step and j * 2^n = total duration?). Eq. 2: S is said to measure a proportion while it has only 3 possible values. Please clarify.

We will clarify and be more consistent in the naming. "i" will refer to a time step and j a time-scale in minute. The first step of the cascade n=0 allow to go from2j=1440 minute to j = 720 minutes. You can see below a formal version of figure 1. The improved version will include those aspect together to a better readability for non expert reader.

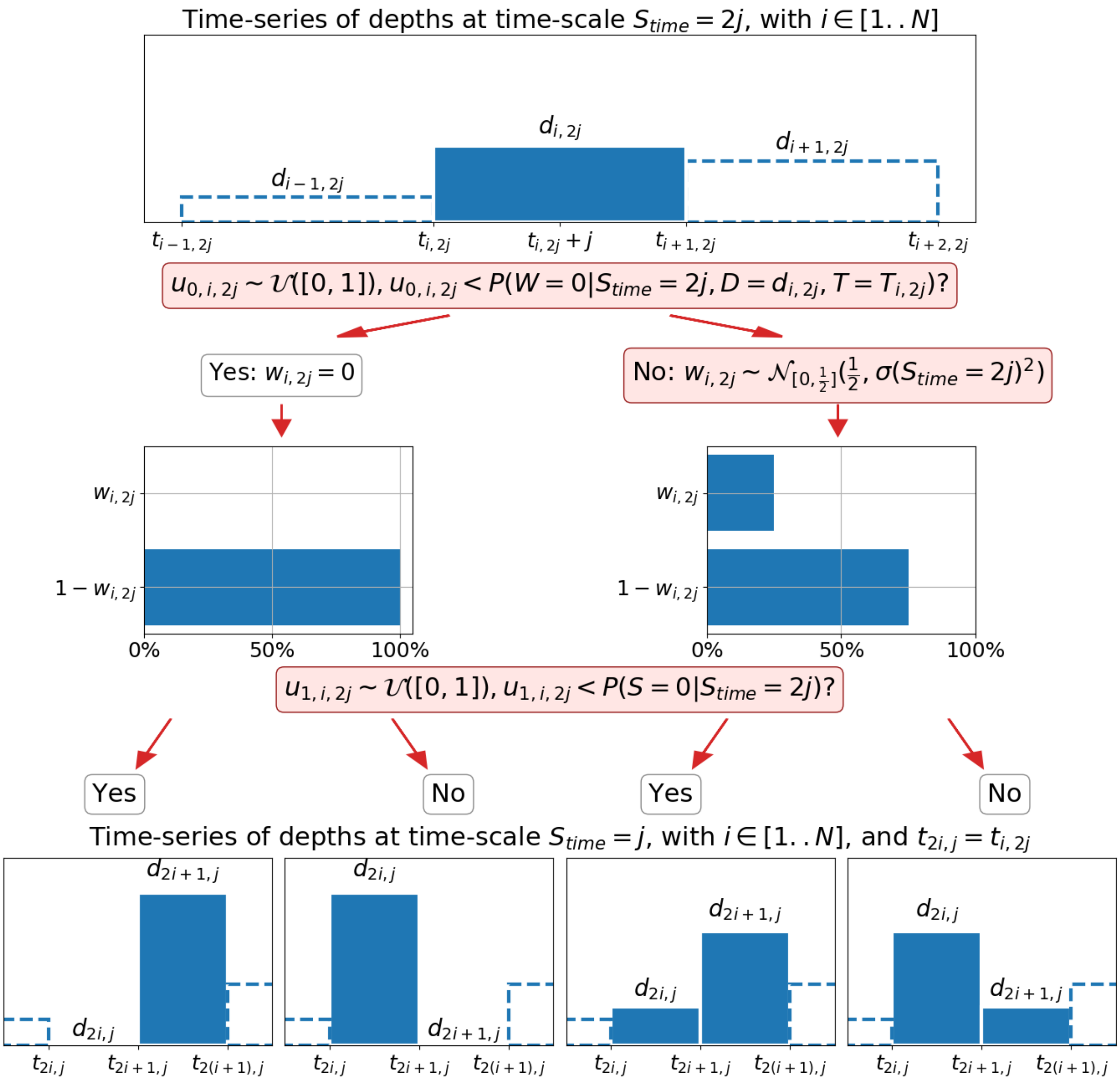

Figure 2: Formal version of figure 1 in the first version of the manuscript.

R2_8

- Section 2.2.2: Please clarify how the fitting of the models was done. Is the probability distribution used the same at all cascade steps (only for P(W=0) if I understand well table 2)? Was a scaling break identified in the data? What would be the consequences of such break? l. 101: "all included 5", may be say all included the use of eq. 5 to help the reader. Please explain how the depth or temperature dependency was included. It would also be needed to clarify in Table 2 to what refer the parameters mentioned.

About the fitting description, please see reply to R2_3. The distributions and functions depend on the time-scale, it is therefore different at each cascade step. In practice the parameters for the distribution vary more at small time-scale than (approx. lower than 45 min time-scale) as it can be seen in the manuscript on figure 3.a. The consequence of such a break is that it is relevant to add time-scale dependency since *a priori*, given 2 different time-scales the distribution is not the same.

It will be clarified both in the text and together with supplementary material and details on the model (R2_1). Table 2 was corrected due to wrong version leading to confusion (see below).

Model	$P(W=0)$				$CDF(W,W\neq 0)$				$P(S_W=high)$				Number of parameters
	R	D	T	N	R	D	T	N	R	D	T	N	
1-MC	x			x									8
2-MCS	x			x				x			x		13
3-MCD	x	x		x									14
4-MCDS	x	x		x				x			x		19
5-MCDT	x	x	x	x									13
6-MCDTS	x	x	x	x				x			x		18

Model	$P(W=0)$				$CDF(W,W\neq 0)$				$P(S_W=high)$				Number of parameters
	R	D	T	N	R	D	T	N	R	D	T	N	
MRC_S	x				x								8
MRC_{S-SEP}	x				x				x			x	13
MRC_{SI}	x	x			x								14
MRC_{SI-SEP}	x	x			x				x			x	19
MRC_{SIT}	x	x	x		x								13
$MRC_{SIT-SEP}$	x	x	x		x				x			x	18

Figure 3: Current table with wrong column alignment (top) table with alignment correction (bottom)

The table 2 will be reviewed to clarify that all process involves a time-scale dependency.

The detail of the different models (including functions) will be provided in appendix.

R2_9

- l. 144-145: why limiting to lag-1?

In the manuscript the authors made the choice of applying a lag-1 autocorrelation at each step of the cascade. The main reason is that, at fine resolution, the autocorrelation might be influenced by the rain gauge resolution. The computational time is also shorter. Informally, the lag-1 at 90 min includes information relative to lag 2 or lag 3 at 45 min resolution, the same principle can be applied for all timesteps.

Since the autocorrelation is often computed for other lags (Müller-Thomy, 2020), we will further investigate other lag times and if relevant include it in the main text, or in an appendix.

R2_10

- l. 152-153: how do you define "small", "major" and "extreme" events?

We defined "small", "major" and "extreme" with common threshold for roofs and locations. In order to qualify different operating mode of the roofs and different climates while being common for all locations in order to facilitate the comparison, we had to set a compromise: 1 L/s/ha, 10 L/s/ha and 100 L/s/ha. Since one of the indicators is exceedance frequency, a common frequency could not be chosen as threshold.

We will clarify the choice in the method section.

R2_11

- Section 3.2: How to you interpret physically the differences of behaviour in Fig. 3.a? Is the shape of Fig. 3.b the same for other time scales? How was the fitting of the model done from this analysis?

In this context the shape in figure 3.a. cannot be linked to a physical behaviour but to the properties of the datasets. Those properties in the datasets can be linked both to a data collection problematic or to a physical property. In general, this curve is linked to the probability to have a long continuous evenly distributed event or a shorter event. The shape for figure 3b is the same for other timesteps, t=48 was chosen as an example since the effect is visual a support the explanation. The conceptualization of the Temperature dependent models was based on this analysis. cf. R2_3 for fitting. A figure will be provided in appendix (cf. R1_18: fig 3a for all locations and fig 3 (b) for other locations and another time-scale).

R2_12

- l. 212-213 and comments on Fig. 4. c and d. Why is the discharge considered to be only slightly underestimated while the observations do not fall in the 5-95% percentile?

The discharge looks are qualified as slightly underestimated because the distance between the distributions is small. The log axis was used here to track visually the magnitude to this underestimation. The reason for the use of log axis is that the exceedance frequency we are interested in are rarely occurring. That is also the reason why a Kolmogorov Smirnov test is not relevant here: we are interested in reliable metrics for rare occurrences.

In practice, those survival distributions are used to estimate the time above threshold. And from a practitioner point of view, the authors think that accounting for natural variability with a window of time is more relevant than looking at a single point estimate from a full time series since the duration associated to discharge exceedance frequency can vary from year to year, especially because of the rare occurrence of extreme events.

The figure 5 shows in 3 different thresholds that, while accounting for natural variability on the flow duration curve, the results from this method are close enough to inform on the magnitude of runoff occurrence. We will provide statistical distance (and possibly provide a Q-Q plot in appendix) to support this analysis. Showing the range of this 3 year-window from both observed and simulated data was initially excluded to favour readability of figure 4.

We will rework the paragraph to introduce this definition of good estimate earlier in order to clarify those aspects. We will also consider a way to add this information into the graphs 4.

R2_13

- l. 216-218 and fig 4.e: the discrepancies between models and observations for the lag-1 autocorrelation should be discussed more.

The MC led to poor lag-1 autocorrelation depending on time-scale because the depth is not taken into account when splitting the data. therefore, a high depth will be split in 2 in the same way as a small depth. It directly influences the lag-1 autocorrelation. About the MCDS and MCDTS, it is possible to improve the lag-1 autocorrelation. However, since the Rainfall continuity indicator does not take into account the depth of neighbouring approach, it was not possible to improve further the quality of the autocorrelation. See (Müller-Thomy, 2020) for other possibilities in improving autocorrelation (excluded to not increase further the number of parameter)

R2_14

- l. 235-238: I have trouble to find the figures mentioned in Fig. 5.

The results are indeed not easy to read on this because of the large amount of data and variable. The current example refers to Bergen **[1st column]**, major events (i.e., 10 L/s/ha) **[2nd**

row], and D-green roof **[right part of the subplot]**. Moreover, this example is not the easier to read. Therefore, we will: *i)* carefully review the explanation linked to figure 5, 7 A1 and A2; *ii)* we will choose a clearer example. E.g., Bergen [1st column], small event [1st row] E-green roof [middle of the subplot] predicted between 30 to 45 days per year above threshold with simulation based on observed data. With MCS we predicted 18 to 20 day / year: it is a bad estimate (according to definition l. 235). With the MCDS we predicted 35 to 37 day/year: it is a good estimate. We will also consider changing the scale to a log scale in order to make the results more easily readable (see comment below).

R2_15

- Fig. 5: last row (extreme events). May be the vertical scale could be split to enable a zoom on the lower part which concentrate most of the information which is not visible now.

This figure was originally made in order to include the results from one location for each of the climate investigated and each of the events defined which made it challenging to include all. We will try both your suggestion and the use of a log axis (since the magnitude of the estimates matters especially).

R2_16

- Fig. 6: "observed data" is not visible in the graphs

It will be fixed.

R2_17

- Section 3.5: without explaining everything, I believe that some details the variational approach are needed for the non-specialist reader. How do authors interpret the fact that the differences between the two approaches are much more pronounced extensive roofs than the detention-based ones?

More details will be provided. Section 2.4 will be renamed "Evaluation of the downscaling models", and the use of the variational method will be added. It consists, given an IDF curve, in using as an estimate the constant duration rainfall leading to the worst-case scenarios. In our case, in terms of peak discharge.

The authors are not sure what the reviewer means by the 2 approaches: current climate vs future climate or downscaling based vs VM? For both cases, the difference between the approaches can be explained by the properties of the roofs (which is also the reason why they were selected). The D-green roof has a higher detention capacity which results in a very narrow distribution. However, as it can be seen with 10-year RP future, once this capacity is overcome (i.e., the layer saturated), it leads to much higher runoff and therefore a more spread distribution of performance. The E-green roof has a simple setup, it led more easily to high runoff and therefore it has a large range of performance even for lower return period events. However, increasing the return period shifts the range but does not increase the range as much as the E-green roof.

References:

Bürger, G., Pfister, A., & Bronstert, A. (2019). Temperature-driven rise in extreme sub-hourly rainfall. *Journal of Climate*. https://doi.org/10.1175/JCLI-D-19-0136.1

Müller-Thomy, H. (2020). Temporal rainfall disaggregation using a micro-canonical cascade model: possibilities to improve the autocorrelation. *Hydrology and Earth System Sciences*, *24*(1), 169–188.