Response to the editor:

We confess that we were extremely surprised that the third referee, who only participated in the last round, seems to have ignored the previous rounds. In particular, he totally ignored our detailed responses on the novelty of the paper and the corresponding changes in the manuscript. Moreover, no dissatisfaction from the first two referees was communicated to us. As this was the only main weakness in the paper content claimed by the referee, we do not understand why a major revision was decided. Indeed, the second main weakness claim was limited to the quality of the writing.

Despite these objections, we have carefully analysed, commented on and taken into account the referee's numerous comments, which focused mainly on the presentation of the manuscript.

Response to 3rd Anonymous Referee's assessment/recommendation/comments/suggestions:

General assessment:

This is a rather low quality manuscript. Its two main weaknesses are 1) lack of novelty and 2) poor writing/structure. The first relates to the fact that UM cascade have been used to analyze/simulate rainfall time series for a long time, including extremes. Therefore, the contribution of the present paper remains unclear. In terms of writing, I would say that the paper is lengthy and overly complicated. The text contains many digressions on other, irrelevant issues that are outside the scope of the paper and distract the reader from the essential.

Also, there is a lot of jargon, which makes the research sound more complicated than it really is.

At the same time, lots of important practical details about the actual implementation of the cascade are missing, which makes it impossible to reproduce the work.

Response: We are grateful to the referee for reviewing our manuscript in detail and providing numerous comments and suggestions. However, the referee seems to have ignored the previous rounds. In particular, he totally ignored our detailed responses on the novelty of the paper and the corresponding changes in the manuscript. Following questions of the two first referees, we sharpened out that we addressed three kinds of knowledge gaps:

- a general discrepancy between standard procedures for defining reference precipitation and the strong multiscale intermittency of precipitation.
- missing procedure to adapt multifractal precipitation modelling to given partial statistical references.
- missing procedure to assess the accuracy of the method.

with the corresponding challenges:

- to tackle multiscale intermittency head-on, based on extreme non-Gaussian statistics and scaling behaviour over two subranges of time scales, due to the finite size of the earth. This requires a given adaptation of the multifractal modelling procedure.
- to define a renormalizing procedure for the multifractal model to make the simulations fit with these partial statistical references.
- to define multiscale metrics to assess distance between (closeness of) two time series (observed and simulated) across time scales.

In short, one should not confuse the UM cascades used by earlier studies to simulate rainfall (without any reference constraints), and the UM cascades being used here to simulate reference rainfall scenarios (with the constraint of certain durations, return periods and intensities) that could be used to optimise the design of certain urban storm-water management infrastructure.

The second weakness claim is limited to our writing quality, not to the content of the paper. Unfortunately, this is a somewhat subjective point-of-view and is partly due to our text edits based on suggestions from other referees. However, we have reviewed the whole text, in particular to limit the use of multifractal jargon to what is absolutely essential for this paper.

We respectfully disagree with the referee's comment on the work being impossible to reproduce, since it is fully based on the discrete UM cascades that have been explained in great detail in several earlier studies, some of them cited here.

In this document we provide our detailed response to the referee's comments/suggestions and also mention how we have addressed some of these queries in the revised version of this manuscript.

<u>Recommendation:</u> *major revisions.*

Response: We consider that the term "major" is not supported by the above responses, in particular those addressed to the first two referees we have recalled. Most of the current comments and corresponding suggestions, several of which correspond to the literature review based Table and Figure in the introductory part, simply call for some text edits! It is worth to note that 7 out of the 10 "major" comments where rather focused on the presentation of the results, not the on the results themselves.

<u>Major Comments:</u>

<u>Comment a:</u> *The abstract does not state/summarize the most important results. It is too long and misrepresents the scope of the paper.*

<u>Suggestion a:</u> *rewrite the entire abstract. Describe what this paper is about, highlighting the novelty and contribution. Be concise and clearly mention the main results.*

Reply: This revised abstract, although a bit long, was actually a result of modifications considering comments from referees #1 and #2. It lists the 3 research gaps and the corresponding 3 contributions that we recalled above, they highlight the important novelty of the paper. Therefore, we respectfully disagree with the referee.

<u>Comment b:</u> *The Introduction is too broad and contains irrelevant information. This paper is about the simulation of rainfall time series using discrete UM cascades. Therefore, I do not see any need to dwell on spatial models and space-time models. The part about the 8* *characteristics of rainfall fields on lines 53-67 is not necessary for the understanding of the paper, and so are Table 1 and Fig 1. Most of these aspects never come back in the results part of the paper.*

<u>Suggestion b:</u> *tighten the scope of the introduction/methods by focusing on time series only. Instead of wandering off topic, include more relevant background information about the current weaknesses/strengths of rainfall time series models, including their ability to reproduce extremes and IDFs. Highlight what the knowledge gap(s) is/are and how the methods proposed in this paper address it/them.*

> **Reply:** The discussion about space-time cascades was added in response to comments from referee #2. We partly agree with referee#3 that the last 2 characteristics - high parameter parsimony and low computational complexity – seem at first glance not essential for the understanding, but are nevertheless of prime importance for data analysis and stochastic simulations, and thus for the choice of methods.
>
> Anyway, the other characteristics (except space-time complexity) are in line with the referee's suggestion. The recommendation to include information about the weaknesses of current rainfall time series models in simulating extremes was in fact done earlier: the necessity to go well beyond Gaussian statistics has already been explained in L64.

<u>Comment c:</u> *Figure 1 and Table 1 are not necessary for the comprehension of the paper. The computational complexity never comes back and none of the other methods are implemented/used.*

<u>Suggestion c:</u> *remove/shorten them or consider adding other methods to compare against.*

> **Reply:** We partly agree with the referee, and as mentioned in our response to comment b, we mention parameter parsimony and computational complexity in Fig.1 and Table. 1 only as bare facts, e.g. 5 parameters over time periods ranging from a few minutes/hours to years. As the referee points out below, any attempt to compare with other methods easily becomes complex and/or partial. On the contrary, we briefly recall that for each scaling range UM enable to work with the minimal number of parameters (2) that is theoretically needed to obtain multiscaling, i.e. a nonlinear scaling function (Schertzer and Lovejoy, 1987, 1997).

Comment d: *Fig 1 and Table 1 are deeply misleading. Within a given category, many different implementations/flavors have been proposed. The complexity and number of parameters vary a lot depending on which publication you consider.*

Suggestion d: *to make the comparison fair, you should refer to specific papers (e.g., authors +year + name of method) or give a range of values for multiple publications.*

> **Reply:** This issue is resolved in the revised manuscript (see Table. 1, Fig. 1) as mentioned in our response to comments b and c, i.e. there can be no more room for misunderstanding or uncertainty.

Comment e: *The random cascade model implemented in this paper uses 4 parameters and not 2 as claimed in Table 1. Therefore, it is not objectively more parsimonious than many of the other methods mentioned in the introduction. According to your own definition in Fig 1, the model would not be labeled as highly parsimonious.*

Suggestion e: *do not label models as highly parsimonious, etc. Focus on explaining the differences in approach, and how much of the original complexity can be reproduced with a given set of parameters. Depending on the application, different characteristics will be important, such as extremes, mean, variance, autocorrelation, intermittency etc. Clearly explain which characteristics are the most important to you.*

Note: *actually, the number of parameters is 5, because you also need to count the scale break (which needs to be estimated from the data).*

> **Reply:** The issue of parameter parsimony is resolved in the revised version as mentioned in our response to comments b, c and d, in particular we clarified, thanks to referee's remarks, that the minimal number of parameters is 2 per scaling sub-ranges.

Comment f: *There is crucial information missing about how the cascade models are implemented, and how the time series are generated. Because of this, the research is impossible to reproduce.*

Suggestion f: *restructure section 4. Consider creating more sub-sections in 4.1 to explain the different parts, from the simulation itself (using the Lévy random variables) to the renormalization. Provide a step-by-step description and mention the software packages/tools used. If possible, provide documented example codes.*

Reply: As mentioned in our response to the referee's general assessment, the simulation procedure is fully based on the discrete UM cascades that have been explained in great detail in several earlier studies, some of which have been cited here. Furthermore, Fig. 5 already shows the step-by-step algorithm used by the Python-based simulation code. Moreover, in line with the referee's demand to have a paper focused on the novelty, we do not feel we can give in more details in the present paper. However, the interested reader has all the means to reproduce the simulations.

Comment g: *The Results/Discussion part is too short and too shallow. The outcomes need to be discussed in more details. The scores are not enough to understand/interpret the results.*

Suggestion g: *extend the Discussion part. Include more diagnostic plots and critically discuss the pros/cons. If possible, compare the outcomes to what is possible to achieve with another of the mentioned simulation techniques (not UM based).*

Reply: Our idea of defining the metrics was to make a quantitative, robust yet quick comparison of the simulations with observed datasets, and they seem quite adequate considering the objective of this manuscript. It should be noted that these metrics (MCM, SCM, CCM) are defined across scales, unlike the usual scores (such as RCM) which are limited to the estimation of a given scale. It is also worth noting that discussion of results in this paper is not restricted to section 5. Section 2 and 3 discuss some data analysis results, whereas section 4 discusses simulation results.

Comment h: *Be more critical with respect to obtained results. While reading the paper, I got the impression that the authors were very quick at praising the UM cascade model and how amazing it is. However, UM cascade also come with limitations and the whole approach relies on some pretty strong assumptions which need to be discussed.*

Suggestion h: *objectively report on what the method can/cannot do and critically discuss the assumptions it relies on.*

Reply: One main limitation in this paper is that of discrete UM cascades, they use integer scale-ratios which can be considered to be a non-physical assumption. We have already mentioned this in L232, L399. The method proposed here can only do what it was developed for i.e. simulating realistic reference rainfall scenarios to design storm-water management infrastructure. Simulating rainfall in real time and/or forecasting rain is not the goal of this method. Furthermore, it cannot be used directly to simulate additional related variables such as temperature that could be relevant in the design of urban storm-water management devices including green roofs.

Comment i: *explicitly state what you actually mean by seasonality. Different characteristics of the precipitation process may have different seasonal patterns. For example, the wet/dry spell lengths, the average precipitation amounts or the extremes. In addition, you don't actually need the UM framework to assess seasonality.*

Suggestion i: *clearly define what seasonality means in the context of this paper and use traditional metrics such as the coefficient of variation (or related) to quantify the observed/simulated seasonality. Check whether the UM cascade can reproduce these quantities.*

Reply: As mentioned in L366 - 374**,** we use the time gap between the maximum and minimum monthly average of cumulative precipitation as an indicator of seasonality. The *classical* UM framework does not address seasonality because it assumes a form of statistical stationarity. However, this framework can be generalised to include a given type of seasonality (Tchiguirinskaia et al. 2002). To keep the present paper as focused as possible, we only wanted to take into account a question of the referee #1 on possible biases of UM simulation vs. empirical data due to the difference of periodicity. This is why we use this simple indicator, which just assesses whether the time gap between the maximum and minimum monthly rainfall is similar for both observed and simulated rainfall. With respect to the traditional coefficient of variation, it had the advantage not to be limited to quasi-Gaussian/second order statistics. Again, to keep the paper as focused as possible, we do not feel we have to elaborate more.

Comment j: *lots of self-referencing: Out of the 71 references, at least 27 refer to work done by people in the same group as the co-authors. This is a lot and could be qualified as excessive self-citation.*

Suggestion j: *check whether all these self-references are really needed.*

Reply: We have removed a few references that weren't too relevant in the revised manuscript.

Minor Comments:

Comment 1: *Throughout the paper: avoid using too many parentheses at the end of your sentences. This gets annoying very quickly and is bad writing. Just add a new sentence or consider combining the two parts using a comma.*

Reply: We have reduced the usage of parenthesis in the revised version.

Comment 2: *The TM and DTM methods have already been explained in great detail in other studies. You could save space by not repeating the theory and referring to the relevant papers.*

Reply: Although we agree, these parts are already not that space consuming. While the description of TM method is around 5-6 lines, that concerning basic DTM is around 5 lines.

Comment 3: *ll.46-47: "[…] however they do make some non-physical simplifying assumptions […]": Which ones?*

Reply: L78-80 already discusses such a simplification in Radar-based bead models. Cell clusters and Modified turning band models both make Gaussian assumptions. We have added this later sentence in the revised manuscript.

Comment 4: *ll.53-67: when mentioning the 8 properties, you should better distinguish actual properties (as seen in observations) from model properties.*

Reply: Since the UM model parameters correspond to data statistical estimators (as mentioned in L55) such a distinction is rather limited to the fact that the latter has uncertainties.

Comment 5: *l.60: "[…] for instance fields are not presumed to be additive": Please explain what you mean by this. Are you referring to additive errors?*

Reply: Not at all, but to the fact that the underlying processes are presumably not additive, e.g. like a Gaussian or a Lévy process, but multiplicative. The former are linear, while the latter are strongly nonlinear. Therefore, we use the UM cascade models where the Levy distribution is used only for simulating the generator which is then exponentiated to obtain rainfall. Although this is already explained a bit in L239, we have made this clearer in the revised manuscript.

Comment 6: *ll.63-64: "[…] extreme rainfall values are more frequent than usual resulting into strongly non-Gaussian statistics.": Nonsense. By definition, extreme values are less frequent than usual ones. Just say that the distributions are positively skewed, with long right tails.*

Reply: We meant extremes occur more frequently in fat-tailed distributions than in Gaussian distributions. So the comparison was obviously between the occurrence of extreme events in Gaussian and non-Gaussian distributions, not between extremes and usual events! We have made this clearer in the revised manuscript.

Comment 7: *l.145 Equation 2: Why not give the general expression with the H?*

Reply: Ok, we have given the generalized expression in the revised manuscript.

Comment 8: *l.151: "Larger the sample size, better will": Gibberish*

Reply: Unfortunately, it is unclear what the referee thinks is Gibberish here as the entire sentence is "Larger the sample size, better will be the estimate of spectral slope". If the issue is with how to get a larger sample, then we added the example that "Spectral slope obtained from a time series that is split into a number of smaller samples is more reliable than that obtained from the whole time series".

Comment 9: *ll.184-185: "Generally, this could be due to two different issues: […]": The most plausible issue should also be mentioned here: that the data are not really multifractal. In other words: the assumption itself should be questioned (on top of how the parameters are estimated).*

Reply: With finite samples, we can only estimate how much the observed field *could reasonably be* multifractal, but not how much it *is really* multifractal. This is achieved by assessing how closely the empirical statistical moments follow a scaling law for each moment order over a given range of resolutions. We think Figure. 3 already made this point very clear.

Comment 10: *ll.207-208: "These low values of MCI justify the aforementioned selection procedure": Well, maybe. But without context, this number does not mean much. What is an acceptable value?*

Reply: As shown in Eqs. 6, 7 the MCI here is totally dependent on α, C_1. Since $0 \leq \alpha \leq 2$ and $0 \leq C_1 \leq 1$ (due to the assumption of a single sample), this implies that the maximum and minimum value of γ_s are close to 1, 0 respectively. Therefore, it is rather straightforward to see that the maximum value of MCI is around 1 due to which the MCI values obtained in the text (0.03, 0.03, 0.04) are low and can be considered acceptable. We have added this explanation in the revised manuscript.

Comment 11: ll.221-222: "[…] a property respected even by the Navier-Stokes equation used by state-of-the-art NWP models for operational forecasting": Please add a reference here.

Reply: Ok.

Comment 12: l.223: "[…] can be considered as a bridge between purely statistical and purely physical models": Nonsense. A bridge is what you use to cross from one side to another. Here, you just have a method that combines the properties of both worlds. But that does not make it a bridge and does not tell us how to go from the physical to the statistical world.

Reply: These cascade models are based on Richardson's idea of energy transfer embodied in his 1922 Poem "Big whorls have little whorls Which feed on their velocity, And little whorls have lesser whorls And so on to viscosity." So the ideology of cascade models is firmly rooted in the so called physical world, while generating fields that have the right statistical properties. Therefore, these cascade models take us from the physical world to the statistical world due to which we see no issue in calling them a bridge between these two worlds. The importance of this type of bridge has gained recognition from the Nobel Committee for Physics, (Schertzer and Nicolis, 2022).

Comment 13: l.236: "[…] and suitable amplitude [...]": make this part more explicit by stating exactly how the Lévy variable is simulated. See major comment (f).

Reply: As mentioned in our response to comment (f) there are several studies that have already explained such simulation procedures in great detail, and we have already cited them in L237, L243.

Comment 14: l.351: "[…] it can be seen that they are somewhat similar to another.": Too vague. Provide the absolute and relative differences. Actually, one could make the point that since the UM parameters for different sites with different rainfall properties are very similar, they do not really offer a great physical interpretation. Otherwise, one would be able to see the differences just by looking at the parameter values. Is this because small differences in parameter values can have large differences in terms of patterns? Please elaborate.

Reply: Similarity of the parameter values confirms that rainfall at the three different locations have some common properties, e.g. intermittency**.** At the same time, small differences in parameter values can result in significant changes in the probability of occurrence of events exceeding a given threshold, therefore possible location dependent processes, for instance, different levels of intermittency. We have added this explanation in the revised manuscript.

Comment 15: ll.360-364: "[…] thereby confirming that the simulations have reasonably realistic seasonality features.": I don't think that you have presented enough evidence to conclude this. The simple, subjective comparison with some ratios close to 0 is very sketchy and some more in-depth analyses and diagnostic plots are necessary to convince me of the realism of seasonal features in the simulations.

Reply: As mentioned in our response to comment i, although this was in no way the primary objective of this manuscript. Indeed, we only answered to a question of referee #1 on possible biases related to periodicity. We suggested a very simple metric to have a first look to it. Let us underline that the maximum time gap in months will give a metric close to 1, while a value close to 0 suggests that the observed time gap and simulated time gap between maximum and minimum monthly rainfall is very similar.

Comment 16: l.374: […] "~~physically~~ statistically realistic reference rainfall ensembles"

Reply: We feel that the simulations being physically and statistically realistic go hand in hand. The reason is that the UM Framework and its parameters, unlike those of simpler are physically meaningful (as already explained in L221 & L352), consequently they help to produce rainfall scenarios with the right statistics and probably the right physics.

Comment 17: ll.377-378: "[…] seems to be the most reliable comparison metric.": Reliable is a strange word in this context. Did you mean robust? Or adequate?

Reply: We feel the word reliable is rather adequate here, given the fact that traditional metrics like the RCM seem too dependent on dataset sizes, therefore being unreliable for quantitatively comparing simulations with observations.

Comment 18: ll.388-391: This is a strange way to conclude a paper. This paragraph would fit better in the Introduction, to justify the UM cascade model.

Reply: Ok, this line has been removed.