Referee 2:

The manuscript proposes a new approach for rainfall disaggregation, based on the framework of Multiplicative Random Cascade (MRC) yet able to overcome its limitations, namely possible over-parametrization or limited accuracy in presence of intermittency (more specifically, for a large amount of zero values). The Author investigates the effectiveness of a disaggregation scheme based on the notions of scaling and Equal Volume Areas (EVA), with maximum three parameters that can be inferred directly from the coarse-scale data. Investigation is performed for 2D (spatial) rainfall fields. The proposed approach allows for intermittency simulation without explicitly coding zero values and their distribution and structure; on the other hand, the simulated rainfall field needs to be re-sampled in grid cells from variable to constant size (i.e., re-gridding).

General comment:

The manuscript is well written and organized, and the topic is surely of interest for the hydrologic community. Hence, I suggest to consider the manuscript for publication in HESS. Before this, more details are necessary to allow for reader understanding and repeatability of the work (that is also the application of the framework in different contests). The most important problem is that the theoretical properties of the simulated process, that are the marginal probability distribution function and the joint distribution, are not explicitly presented and discussed; these are fundamental to understand e.g. if the model is based on the same scaling property ruling the theoretical behavior of the traditional MRC. I have some additional concerns that are listed below. I hope they will be helpful for manuscript improvement.

<u>Response:</u> Thank you for taking the time to review this paper. I will add more details in Section 2.4 and 2.5 in order to better explain the mechanisms through which the EVA cascade generates intermittency and how model parameters can be estimated from the coarse-scale data. However, I don't think that it is possible (and that I would be capable) to derive meaningful, analytical expressions for characterizing the marginal probability distribution function of rainfall rates and dry/wet areas produced by the EVA cascade. Such theoretical derivations are beyond the scope of this paper and would depend a lot on the assumptions you make about the input field (e.g., distribution, stationarity, spatial structure, etc…). I understand the reviewer's interest for these issues but don't think that this is important for understanding the basic idea. So instead of making this very theoretical, I would like to stick to a more empirical approach and present the EVA cascade as a simple, alternative and convenient way to numerically transform a coarse-scale rainfall field into a high-resolution output. The application of this new cascade

model to a large number of rainfall fields shows promising performance, especially in the presence of intermittency.

Specific and technical comments:

• I'm curious about the reasons why 2D simulation is discussed instead of "starting" from the simpler 1D case, i.e. for temporal disaggregation, which is important for several application problems. The Author should at least discuss the applicability to the framework to temporal disaggregation.

Response: Thank you for this comment. Indeed, the same technique could be used to downscale time series based on the notion of inter-amount times (http://10.1175/JHM-D-15-0078.1). However, in this paper, I wanted to focus on designing a new spatial downscaling method, which is more challenging and interesting than time series downscaling because of the higher intermittency of 2D rainfall fields. This makes it easier to outline the advantages of using an adaptive sampling approach. But of course, the same approach could be used to downscale rainfall time series as well. Actually, in a previous paper (https://doi.org/10.1175/JHM-D-16-0221.1), I already outlined the superior scaling properties of inter-amount times in the presence of intermittency (without explicitly using them for downscaling). For completeness, I will add a few sentences about the downscaling of time series in the Discussion part. But this would have to be done within the context of another paper.

• It should be important also to mention one fundamental problem characterizing MRC model, i.e. the stationarity of the disaggregated process. The Author mentions the paper by Lombardo et al. (2017); yet in a previous paper (Lombardo et al., 2012) the authors demonstrated why MRC does not generate stationary processes. The Author should discuss the stationarity issue for the model proposed here and generally mention the possible limitation of the framework for temporal and spatial disaggregation.

Response: I will add a few lines about the stationarity issue in MRCs in the discussion part of the paper, together with a reference to Lombardo et al. 2012. But honestly, I don't see a lot of value in investigating stationarity issues within the context of this paper. The new EVA cascade proposed in this paper is just a convenient mathematical way to transform a coarse resolution input to a fine resolution output. Yes, there is a whole discipline in mathematics that deals with studying the theoretical properties of MRCs under idealized conditions. But rainfall is a complicated, highly intermittent and non-stationary process which means that these idealized conditions are very unlikely to be satisfied in reality. So instead of theorizing too much, I think there is more value in assessing the performance and practical advantages of the EVA cascade on real data, by comparing the EVA cascade with classical MRCs and discussing the pros and cons.

• A figure could help to better explain the splitting rule.

Response: Thanks for the suggestion. I will add a new figure during revision.

• Line 17, page 5. How can we reproduce the stochastic properties of the process of alternation of wet and dry cells? If we do not explicitly model this process, how does the structure of this process depend on the model formulation?

Response: The process by which intermittency is created during the cascade is explained at the end of Section 2.4, equations (9) and (10). The particularity of the EVA cascade is that some grid cells will get ``stuck'' during the downscaling, meaning that their area will converge to a fixed lower limit (different from zero) while the rainfall volume inside these grid cells will tend to zero. These are the areas which, after the final regridding, will correspond to dry areas. The spatial structure of the dry/wet areas and their frequency at a give scale is therefore controlled by the splitting rule or, equivalently, by the values of the three cascade generator parameters a, b and c. This was already explained in the text but may not have been formulated very clearly. I will add additional details during revision to make sure that this part is better explained, for example by splitting Section 2.4 in two parts and adding a new subsection for explaining how the cascade generator parameters influence the intermittency pattern.

• Lines 12-17, page 7. Is it possible to gain the same or similar advantage by using a different generator?

Response: Sure, many other similar distributions could be used with various degrees of interpretation for the parameters. I chose the logit-normal model because it was the easiest and most convenient that I could think of. I will add some text during the revision to mention some alternatives (e.g., the beta distribution).

• Section 2.5. Since it is not clear which are the theoretical properties of the simulated random field (see general comment), it is difficult to follow this section.

Response: Some additional details will be added to Section 2.4 during revision. This should make it easier to follow Section 2.5

• Lines 21-22. Is it possible to quantify this additional uncertainty?

Response: Yes, in theory, if we know the spatial autocorrelation structure of the rainfall field we could approximate the uncertainty due to linear interpolation. I will add some text to explain this. However, the quantification of this uncertainty and its incorporation into the estimation process goes beyond the scope of this paper.

• Lines 29-30. Based on this, it seems that the theoretical behavior of the simulated process is determined based on empirical reasons. Is this correct?

Response: I'm not sure to fully understand this comment. To clarify: the theoretical properties of the simulated rainfall fields are fully determined by the coarse-scale input data and the cascade generator model. The cascade generator model is fixed: it's a logit-normal distribution with mu=0 and variable standard deviation sigma. The parametric relation between sigma, the rainfall intensity R and area A in Equation 8 is based on empirical evidence from a large number of studies. It assumes that on average, spatial intermittency increases as we move to smaller scales and lower intensities. The parameters a, b and c needed to define the changes with A and R are inferred from the coarse scale data, depending on the properties of the considered rainfall field. The only empirical parts are therefore: 1) the choice of a logit-normal distribution for the cascade generator and 2) the assumption that the dependence of sigma on A and R can be approximated by Eq 8.

• Line 10, page 9. "Smoother than the observed one, being based on interpolation" . . . Thus, can this method be considered a disaggregation model? Instead of linear interpolation the Author could have considered kriging, which preserves the variogram.

Response: Yes, the reviewer is right: bilinear interpolation with rescaling is not a disaggregation technique because it does not conserve the total rainfall amount in each coarse-scale grid cell. It only preserves the average intensity over the whole field, similarly to a canonical cascade. I will change the corresponding sentence during revision to clarify this point. Anyway, I don't think this is a big issue as bilinear interpolation is not the main focus of the paper and is only used as a first-order benchmark against which the improvements of the microcanonical cascades can be assessed. The suggestion to use kriging is interesting. I had thought of this initially and indeed, there are some situations in which kriging works better than bilinear interpolation. But there wasn't a big improvement in performance and any simple micro-canonical cascade would do better than kriging. In addition, there are many downsides to using kriging in downscaling applications. The first is that kriging is a linear estimator whereas bilinear interpolation is not. The second is that kriging is slower than bilinear interpolation. The third is that kriging makes pretty strong assumptions about the data. In particular, the mean and the variance of the rainfall process are not allowed to change with the location. This is often not true in practice, especially in the presence of intermittency. For more information about this, please have a look at my 2014 paper in JHM: "Non-stationarity in intermittent rainfall: the dry drift" (http://10.1175/JHM-D-13-095.1). Lastly, it must be said that kriging requires the estimation and fitting of a variogram model from the data. This creates all sorts of numerical issues and complications, with many exceptions and special cases. If the spatial structure is complicated and the variogram model do not fit the sample variogram very well, chances are that kriging will not be as good as anticipated. By contrast, bilinear interpolation is a non-parametric and local method that makes no assumptions about the data and is therefore hard to beat in practice. For all these reasons, I think it is better to stick with bilinear interpolation as the benchmark rather than kriging.

• Lines 13-18, page 9. This configuration of the MRC is indeed not very common in the literature; how does this configuration affect the reliability of the disaggregation model Can we expect better results (as depicted in figures 7, 8 and 11) by using the most common approach in the literature or one based on a larger number of parameters. Models can be compared also in case of a different complexity; furthermore, the parameters of disaggregation models could also be estimated by considering different additional properties of the generated process (e.g. the expected value of the number of wet cells in the spatial case or the dry spell average length in the case of time disaggregation).

<u>Response:</u> Sure, one could easily improve the performance of the classical MRC by introducing additional parameters and performing a separation between dry and wet components. There is little doubt that such a state-of-the-art model with 6-7 parameters would outperform the simple EVA cascade proposed in this paper. At the same time, I don’t think that such comparisons are really fair and helpful at this stage. Future model developments directed at improving the EVA cascade with the help of additional parameters and more flexible generator models should be part of a follow-up study. Performance is obviously important. But the primary goal of this paper is to introduce a new conceptual way of approaching the downscaling problem in the presence of intermittency. The comparisons with the MRC are done to demonstrate potential and gain insight into the pros and cons of each approach. Optimization is not the primary concern here and can always be done later.

• The calibration procedure is not totally clear to me; additional efforts are required to explain it in theory and practice.

<u>Response:</u> I will add more details about this during the revision. The model parameters are estimated in a very natural way, by aggregating the coarse-scale field to larger scales and studying the splitting behavior of larger grid cells.

• I’m not sure figure 6 is really useful. Are the numerical differences between EVA and traditional approach W values (as depicted in figure 6) really significant from a practical point of view?

Response: Yes, of course! They are the main reason why, for similar grid cell areas and intensities, the EVA model tends to have a generator with lower variance. This is crucial for understanding why the EVA method produces slightly smoother fields and lower, more realistic rainfall extremes.

• Given the large amount of observed rainfall fields, it could be of interest to understand how estimated model parameters depend on large scale event characteristics, so that a general parametrization valid for all the events can be found. This means calibrating the model based on observed small scale observations when available to apply the model to other events.

Response: Thank you for the suggestion. I have already looked at this but did not find any strong link between large-scale properties such as intermittency, variance or decorrelation range and the parameter values of the generator. Also, this will depend a lot on which sensor you used for measuring the rainfall field and would not be easily transferable to other regions or datasets. One promising idea though could be to model the bias affecting the estimated model parameters and compensate for it during the parameterization step. This can significantly improve the performance for larger downscaling ratios. I will add a short paragraph about this in the discussion.

• Is it possible to compare simulated and observed fields in terms of the spatial structure of variability (spatial correlation) of the wet/dry alternation process? See also previous comment on this.

Response: Yes, sure! I will add a figure similar to Fig10 but for the dry/wet transitions together with some explanations.

• The Author uses R2 as a metric for model performance with respect to observation; why not using a different metric, not based on the normality assumption?

Response: Actually, the coefficient of determination (R2) does not require any normality assumption. The normality assumption is only needed for the interpretation of the R2 within the framework of a linear regression model (e.g., in terms of explained variance). But this is not the goal here and I just use the R2 as an easy way to measure the relative goodness of fit and support the visual assessment of the scatterplots. This is not critical and any other related metric could be used for this without affecting the conclusions.