Dear Reviewer#1,

Thank you for your detailed comments about our manuscript. All your suggestions have been considered, and we propose the following changes to address the questions you raised in your review.

In the following point-by-point responses, RC denotes a referee comment (in black), AR denotes the author response (in blue), and PM denotes the proposed modifications (in green).

Hoping that the proposed improvements will fulfill your expectations,

Best regards,

Lionel Benoit, Mathieu Vrac and Gregoire Mariethoz.

RC: Rainy images are considered those with > 10% rainy pixels (what rain intensity threshold was applied to set a pixel as rainy?). I wonder if the selection of 10% as a threshold is important and may affect the conclusions? For example, if there is a distinct rain type with a small coverage area (<10%) we totally miss it. I understand a threshold needs to be set, but it should be clarified that there is no high sensitivity to this definition.

AR: The threshold of 10% is indeed pretty high, and leads to classify around one third of rainy images in the dry type. To deal with this ‘dry bias’, the typing method we use to classify radar images actually incorporates a second step (Benoit et al, 2018). It aims at assigning to images with less than 10% rain coverage the type of the closest (along the time axis) classified image. Therefore we do not define a distinct rain type for images with a low rain coverage (<10%), but rather consider that this ‘pseudo-type’ is due to a transient behavior of rainfall at the onset and at the end of rain storms.

This point was overlooked in the early version of our manuscript, and we will therefore address it in details in the revised manuscript. In particular, a paragraph will be added in the main text to discuss how we deal with images with rain coverage <10%. And an appendix will be added to show that the performance of the proposed rain type simulation method is not degraded when images with less than 10% rain coverage are considered as dry.

PM:

New paragraph in section 2.1:

“*Using only radar images with more than 10% wet pixels to define rain types ensures a reliable classification, but at the cost of a dry bias (in the present dataset, 32% of the images measuring some rain have a rain fraction under 10%, and are therefore assigned to the dry type). To deal with images with less than 10% wet pixels, Benoit et al. (2018b) proposed to classify images with a small rain fraction (i.e. 0% < rain fraction < 10<%) in a second step by assigning them the type of the closest classified image (i.e. nearest neighbor interpolation in time). This post-processing scheme is not directly transferable to the context of simulation because no information about the previously misclassified images is available in simulation outputs. Two options can be considered to alleviate this problem. First, the rain type model defined in Sect 3 can be calibrated on the final classification (i.e. including images with low rain coverage), which results in simulations that preserve the actual rain proportion. However, using a classification that includes the beginning and the end of rain storms leads to less clear relationships between climate covariates and rain type occurrence, which may degrade simulation results. Hence the second option, which we follow in this paper that consists in (1) calibrating and running the rain type model for rain types defined only from radar images with more than 10% rain coverage, and next (2) re-adjusting the wet/dry balance by post-processing. The dry bias is corrected* *assuming the ratio R = Ns/Nl between the number Ns of images with small rain coverage and the number Nl of images with large rain coverage as constant in observations and simulations. Subsequently, the number of epochs for which rain is simulated is increased by propagating the closest rain type to the RxNl dry epochs located at the beginning and at the end of rain storms. Appendix A shows that such post-processing performs well to mitigate the dry bias originating from the use of a 10% rain coverage threshold to define a wet image. However, since the present study focuses on climate - rain type relationships, which are better defined when considering only the first step of the classification, the aforementioned post-processing is not applied in the remainder of this paper. Hence, one should keep in mind that in the following the dry type also includes epochs with a small rain coverage (under 10%), and that a post-processing is required if the end-use is an application that involves the stochastic simulation of actual rain fields.*"

New appendix:

"*Appendix A: Cross-validation when accounting for small rain coverage.*

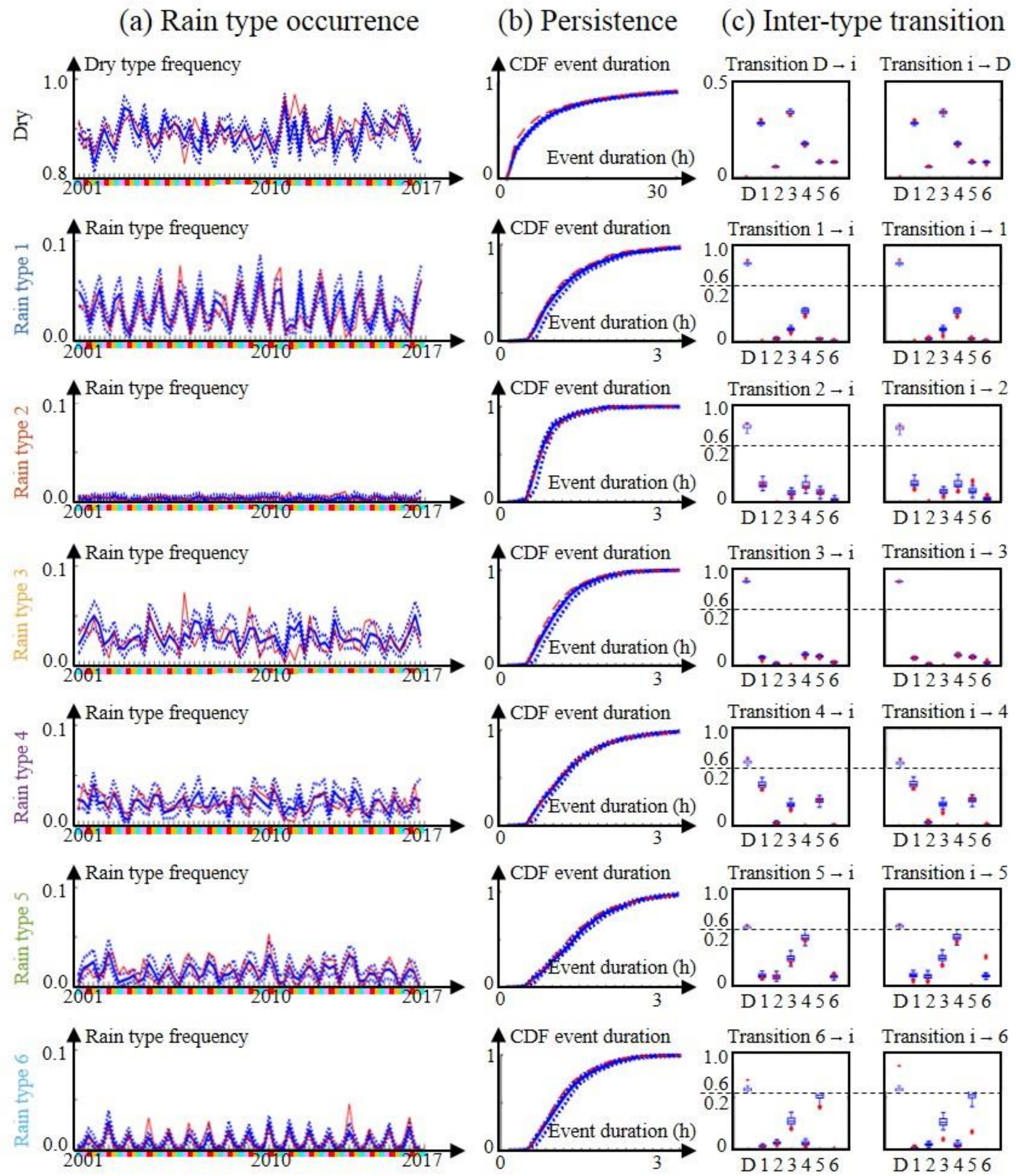


Figure A1. Results of the cross-validation experiment when images with low rain coverage (rain fraction between 5% and 10%) are regarded as wet. (a) Seasonality of rain (and dry) type occurrence (Seasons are DJF (light blue), MAM (pink), JJA (red) and SON (yellow)), (b) rain type persistence, and (c) probability of transition between rain types. Observations are in red and are obtained by assigning the type of the closest classified image to epochs with rain fraction between 5% and 10%. Simulations are in blue and are obtained by propagating the closest rain type to the beginning and to the end of each rain event. In simulations, continuous lines represent the median of the simulated ensembles (50 realizations), and dashed lines represent the Q10 and Q90 quantiles."

RC: Radar data are not gauge-adjusted – this may affect the indices, especially the intensity indices. For example, the "Mean rain intensity over all rainy pixels" index is obviously affected by biases in the radar data that could be corrected with gauge adjustment. It can be claimed that such biases will be consistent and therefore will not affect classification, but biases in radar data could change along the years and thus the inter-annual frequency of rain types may be affected. I suggest addressing this point in the paper.

AR: This is true that changing radar biases could negatively affect the results of this study. Fortunately, the use of uniformly processed radar images minimizes the risk of drift in radar biases. In practice, we do not find any long-term trend in observed rain type frequencies. In addition, the results of the leave-one-year-out cross-validation procedure do not show any disagreement between climate-driven simulated rain types and observed rain types, which reflects the absence of outliers in observed rain types.
The above will be addressed in the first paragraph of Section 2.1 when we present the radar dataset.

PM: "*Using raw radar images can lead to biases in estimated rain intensities, but the impact of such biases on the classification are deemed negligible since the adopted approach focuses on rainfall space-time behavior rather than rainfall intensity. A more troublesome source of errors would be the change of radar biases along time, which could alter the inter-annual frequency of rain types. To alleviate this problem, uniformly reprocessed radar images are used as basis for the classification, which ensures a consistent data-cube throughout the period of interest. In practice, no adverse trend is noted in the observed rain type distribution (Fig. 3).*"

RC: Rain type duration: please explain how is it computed? is it simply derived from consecutive maps series with the same type and does a single map with a different type end this series?

AR: This is correct. The following description of rain type duration will be added in the revised manuscript (Sect. 2.1).

PM: "*Here rain type duration is defined as the duration (i.e. length along the time axis) of a segment of rain type time series with constant type. Each curve in Fig.3a therefore corresponds to the probability that a rain event of a given type does not exceed the duration given in abscissa.*"

RC: Long and short dry duration: is the 24h long/dry duration necessarily defined over days (i.e, midnight to midnight) or is it 24h with an arbitrary starting point? if the latter, how dry periods are split between "long" and "short" intervals?

AR: It is 24h with an arbitrary starting point. A dry period is split in as many 'long' intervals as possible, and the remaining dry time is distributed into two 'short dry' intervals at the beginning and at the end of the 'long dry' period. For example if there in case of a 49h dry period, it is split into: 30min short dry + 2 x 24h long dry + 30min short dry.

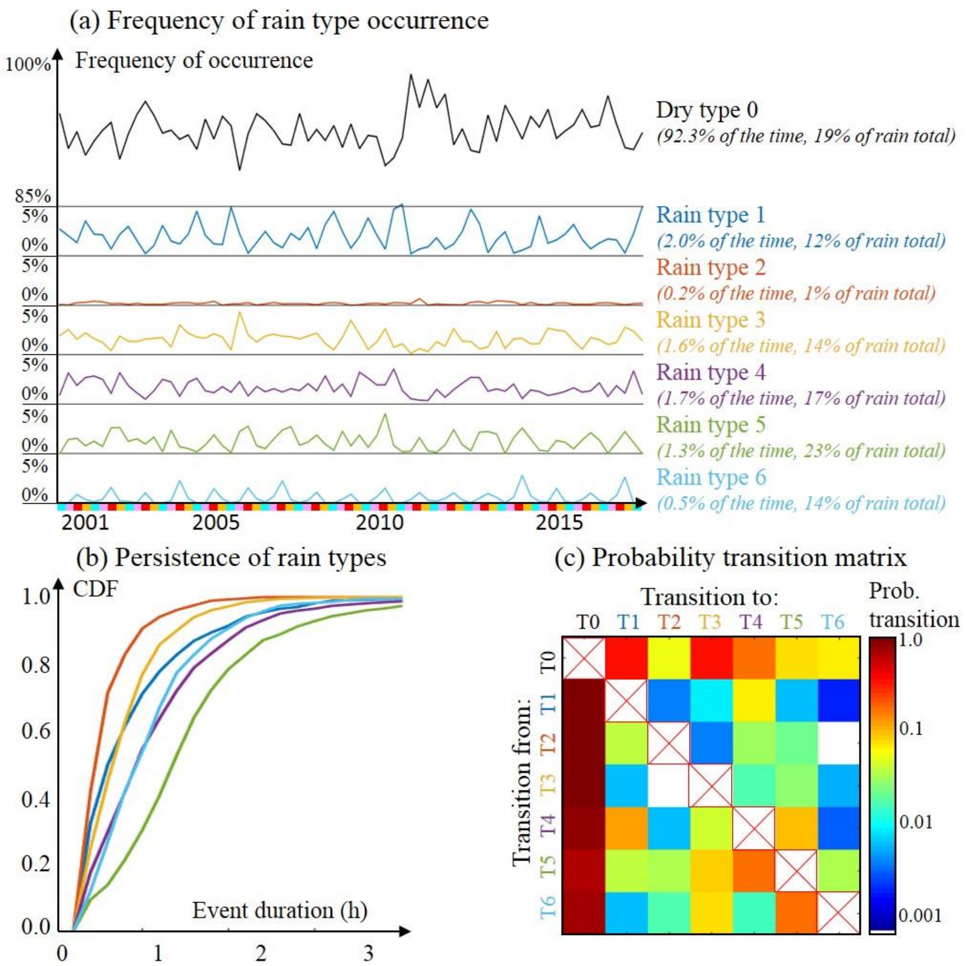
The following clarification will be added in the description of the parametric model (please note that following comments of reviewer#2, the description of the parametric model will be moved in appendix).

PM: "*In practice, if a dry period exceeds 24h, it is split into as many 'long dry' spells as possible, and the remaining time is distributed into two 'short dry' spells of equal length at the beginning and at the end of the overall dry period.*"

RC: It would be helpful to indicate the percent of rain amount out of total amount for each rain type, as well as for "dry" time-steps (i.e., with <10% rainy pixels).

AR: This will be indicated in the figure that investigates the main features of rain type time series.

PM:



"*Figure 3. Main features of a rain type time series (2000-2017) observed over central Germany. (a) Frequency of rain type occurrence computed at a seasonal basis (Seasons are DJF (light blue), MAM (pink), JJA (red) and SON (yellow)). (b) CDF of event duration stratified by rain type. (c) Empirical matrix of transition probability between rain types.*"

RC: Figure 2A: please indicate months so seasonality can be better realized.

AR: Ok, this will be implemented.

PM: Seasonality will be indicated with a color code in all figures where relevant, see Fig 3a above as an example.

RC: It would be helpful to mention statistical significance for the comparisons presented in Figure 3.

AR: Following the comments of Reviewer#2 the parametric model will be moved to appendix and the comparison between the two models will be removed. Fig 3 will therefore disappear.

RC: Eq. 1: I think that it should be written as: P(St=j|St-1=i,x)=…. if not, please explain what is the source of index j on the right side of the equal sign.

AR: This is indeed a better formulation.

PM: The equation will be corrected accordingly.

RC: Page 9, Lines 9-10: can you explain why different data are used for determining the Sigma and Mue matrices?

AR: Our initial description was misleading, in fact the same data are used for determining the Sigma and Mue matrices. We will reformulate the description of these matrices as follow.

PM: "*More precisely, Σ is the empirical covariance matrix and μ_{ij} is the empirical mean of the covariates for the time steps where the transition from i to j occurs.*"

RC: Fig. 6: can you explain the systematic negative bias in rain occurrence frequency for the two models?

AR: In our opinion the systematic negative bias is significant only for the parametric model. For the non-parametric model, the occurrence of dry and rain types is properly simulated. Regarding the dry bias in the parametric model, we believe that it is due to the difficulty of conditioning to covariates in the framework of semi-Markov models. The discussion about the performance of the parametric model will be modified as follows (in Appendix C).

PM: "*Figure C2 shows the results of the same cross-validation experiment than in Sect. 4.1, but applied to the parametric model described above. Results show that this model generates a strong dry bias (ratio simulated/observed rain frequency = 0.61) and does not properly capture the inter-annual variability of rain occurrence (correlation between observed and simulated time series = 0.4). This can be explained by the fact that the relationships between the meteorological covariates and the presence of rain are probably more complex than the linear relationship assumed in the non-homogeneous Markov chain formulation of the parametric model, and by the fact that semi-Markov models do not allow for an easy conditioning to continuous-time covariates. These hypotheses are reinforced by the fact that simulations driven by daily mean temperature only do not generate a dry bias (not shown here).*"

References: Benoit, L., Vrac, M. and Mariethoz, G.: Dealing with non-stationarity in sub-daily stochastic rainfall models, Hydrology and Earth System Sciences, 22, 5919-5933, doi:10.5194/hess-22-5919-2018, 2018.