

## Authors answer

*We thank the editor for taking into consideration our paper for publication and the referees for the useful commentary provided. Any suggested corrections have been carefully examined and the corrected paper is presented with the following answers to the referees comments.*

*The main improvements brought to the paper are:*

- 1) Figure 1 has been modified to better show the rationale of the algorithm.*
- 2) A more detailed illustration of the algorithm for the multivariate case has been provided in Section 2.2.*
- 3) The minimum moving sum has been added as an indicator of the long-term dependence structure (Figure 7). The commentary to this result can be found in Section 4.4.*
- 4) Section 3.1 “Imposing a trend” has been simplified and a commentary about the application of the technique has been added in section 4.2.*

*Line numbering refers to the revised manuscript attached. The second file attached contains the tracking of the corrections.*

*We hope that the revised paper will be evaluated positively and we keep ourselves available for any further suggested improvements.*

*Best regards.*

*F.Oriani, J.Staubhaar, P.Renard and G.Mariethoz.*

### Anonymous Referee #1

General comments The discussion paper demonstrates a simple and robust data generation method that has not been widely applied in hydrology. Its application to daily rainfall generation therefore adds considerable value to stochastic hydrology and highlights the ability of non-parametric approaches for data generation. The methods applied are valid but some details are left out and it would be difficult for the reader to replicate the analysis. The discussion and conclusions reflect the analysis and results obtained. Specific comments Section 2.2 of the paper describes the Direct Sampling (DS) method and uses Figure 1 to illustrate the method. It is not clear exactly how SG is obtained. How different is ST from the historic record?

*We understand that section 2.2 was unclear and we hope to have substantially improved it in the revision. The definition of SG has been elaborated in section 2.1 (lines 73-76) of the revised paper.*

Is the value of  $t$  (in  $x(t)$ ) get randomly obtained from a uniform distribution ( $\text{Random}[0,1] \times \text{length of simulated time series}$ )?

*Correct, the simulation order is randomly generated from a uniform distribution. This is now clarified in section 2.2, lines 103-104.*

Figure 1 is not very informative and might be better if it illustrates a single or two iterations in chronological order.

*We agree, figure 1 has been modified in order to better explain the copy-paste rule at the base of the algorithm.*

Section 2.2 does not inform how auxiliary variables are used as part of the DS method. It seems that the search for  $Z(y_i)$  continues until the thresholds for all the auxiliary variables are met but this is not stated in Section 2.2.

*We agree, the procedure has now been clarified in section 2.2, lines 147-150.*

It may be possible to get rid of patches (Section 3.2) by imposing a condition that the  $Z(y_i)$  selected should not result in a patch in addition to its meeting the set threshold of dissimilarity.

*We think the suggested improvement can be a valid optional feature in cases where a total absence of patching is critical, and it will be considered in further work.*

*Nevertheless, for the application shown in the paper and the proposed setup, we do not see a real need for it since the observed patching is very low. We believe that in this case, forcing the algorithm to totally get rid of the patches is not going to bring an effective improvement to the simulation and may reduce the performance by over-conditioning. As shown in the results, the patching obtained is negligible when using the proposed setup together with an appropriate training dataset (i.e. sufficiently long with respect to the simulated time-series and with a low amount of gaps).*

*Moreover, as far as we have seen in the results of the tests conducted until now, the algorithm is not naturally prone to patching. A considerable patching is generally due to an inadequate parameterization or too limited/fragmented training dataset and results in a bad overall performance of the algorithm. The user is warned about these issues (Section 3 lines 208-211, 272-275). Therefore, in the most part if not all the cases in which a considerable patching occurs, a more efficient solution would be to find an opportune setup or training dataset instead of pushing the algorithm not to generate the patching itself.*

In addition to the 10-years MS comparisons presented in Figure 6, the minimum run sums for various lengths (up to say 10 or 20 years) could be used to assess how well DS replicates the long-term dependence characteristics of the rainfall.

*We agree, the minimum moving average (the moving sum divided by the window length, which improves the visibility) with various windows up to 60 years has been computed for the stationary simulations (see fig.7) and a commentary has been added at lines 420-430.*

Suggestion changes to sentence structure etc. Page 3214 line 13 . . . . . reproduced adequately, reducing the . . . . . Page 3214 line 23-24 . . . . Solutions to deal with this limitation . . . . . Page 3215 line 12 . . . . . completely capture a complex . . . . . Page 320 line 2 . . . .event and acceptance threshold. . . . Page 3220 line 20-21 and other locations: should it be datum or data? Page 3222 line 7 Table 1 presents the dataset . . . . . Page 3222 lines 14-15 Mariethoz and Renard (2010) show how direct sampling can be used for data reconstruction Page 3222 line 3 and page 3239: why is (\*) included?

*The training image includes the target and the auxiliary variables. To clarify this point, “\*” as been changed to “6” in Section 3 line 219 as well as Table 1.*

Page 3228 line 16: — discussed in the following section. Page 3242 replace ‘dotted line’ with ‘blue dots’

*We agree with the suggestions, the revised paper has been changed accordingly.*

## **Anonymous Referee #2**

General comments:

The manuscript proposes the Direct Sampling (DS) technique to simulate daily rainfall data as an alternative to the parametric models. As this method resamples the data from the training image based on certain criteria, it cannot simulate values larger than the ones in the training image. Based on this one can say that this method is inferior to the other non-parametric methods such as Harrold et al. (2003b) and Mehrotra and Sharma (2007).

*We agree with the referee, this limitation is put in evidence at line 374 and in the conclusions, line 486. On the other hand, the advantage of the DS with respect to the parametric techniques is the faithful reproduction of the time-dependence structure and distribution at higher scales, where also extremes higher than the reference are generated. Ongoing tests and a detailed comparison between the DS and the mentioned family of techniques will be the subject of a future publication.*

Apart from this, the model adequately preserves the statistical characteristics of the historical data used in the simulation. The section on non-stationary simulation (Section 4.6) is not clear, confusing and not relevant to manuscript.

*We think that the simplicity in which even a complex non-stationarity can be reproduced is a valuable and essential aspect of the algorithm and should be illustrated for time-series simulation, therefore we did not remove this part of the manuscript.*

*We agree about the lack of clarity: the explanation of the methodology has been simplified in Section 3.1 and the relevance of the application is now put in evidence at lines 467-470 .*

I cannot understand why PACF was used to assess the correlation in the data. ACF should have been used in its place.

*ACF and PACF are algebraically linked by the Yule-Walker equations (see for example [1] p.64) and contain the same information. Since here the aim was to investigate how efficiently each time-lag dependence is reproduced by the algorithm, the PACF has been chosen since it shows the linear dependence for each time-lag independently, which is not the case for the ACF. This is clarified at lines 335-340.*

The manuscript should be revised before it can be published in HESS.

Specific comments:

The word "global" appears at a number of places and I cannot understand what it really means. Please explain.

*We agree, the term is ambiguous, it has been changed to "marginal" referring to the probability distribution (lines 294,417,496).*

PACF is not relevant and there is no need to calculate the correlations for lags up to 10 or 20.

*Since the algorithm operates in a non-parametric way and imposes a variable time-dependence, the eventuality of modifying the persistence of the signal cannot be excluded a priori. That is why the daily PACF is calculated up to the 20th lag, just to show that no artifacts are introduced. This has been clarified at lines 440-443.*

*At the monthly scale a more complex dependence structure justifies the computation until the upper lags.*

Technical corrections:

Page 3214, Line 13: Replace "exhaustively" with "well" or "satisfactorily" Page 3214, Line 23: Change "overtake" to "overcome" Page 3216, Lines 12-15: Sect 3 is missing. Sect 3 describes the application of the method. Page 3219, Line 17: Changed "informed" to "covered" Page 3219, Line 24: Change "respect" to "preserve" Page 3220, Line 5: Change "informed" to "selected" Page 3222, Line 5: Change "showing and extreme" to "showing an extreme" Page 3223, Line 4: Change "respect more strictly" to "preserve" Page 3223, Line 15: Change "showed" to "shown" Page 3224, Line 18: What is the statistics mentioned here? Page 3226, Line 16: Change "Another used validation criterion" to "Another validation criterion used " Page 3226, Line 17: Change "transformed in a" to "transformed into a" Page 3226, Line 19: Change "region" to "spell"

*We agree with the suggestions, the revised paper has been changed accordingly*

Page 3227, Lines 1 -24: PACF is not appropriate here. ACF should be used to assess the correlation with the well-known confidence limits. Delete lines 15 – 24.

*We agree on the fact that the detailed explanation about the confidence limits is unnecessary, it has been deleted accordingly. The motivation for using PACF instead of ACF is explained above. Besides, the confidence limits are still valid for PACF since based on the autocorrelation of an IID  $\sim N(0, \sigma^2)$ , for which the two functions are statistically zero valued and equivalent (see [1] p.65).*

Page 3229, Line 8: What is meant by "border"? Do the authors mean the start and end of the time series.

*Correct, the term has been changed accordingly.*

Page 3228, Line 16: Insert "section" after "following"

Page 3230, Lines 12-13: Not clear. Page 3230, Line 17: Change "respected" to "preserved"

*We agree with the suggestions, the revised paper has been changed accordingly.*

Page 3231, Lines 8-28: These can be deleted. There is no need to calculate the correlations up to 10 or 20 lags. Lag one correlation coefficient is adequate.

*We agree on the fact that we do not expect a significant autocorrelation in the reference for lags greater than 1. The reason for computing that is explained above.*

Besides, the model only cater for lag one correlation by considering the sum of current and previous day rainfall (2MS) as a covariate.

*We do not agree, the 2MS is used to respect more accurately the lag-one autocorrelation, since we know a priori that it is the most important short-term dependence for daily rainfall. But it is not the sole lag the algorithm takes into account. As explained in Section 2.2 lines 136-143, higher order dependences are variably taken into account by the data event of the target variable, which changes size during the simulation. This concept has been clarified at lines 228-230.*

Page 3232, Line 1: What is the non-stationarity imposed?

*The non-stationarity is the one found in the TI. The sentence has been rewritten (line 456) to clarify this point.*

Page 3232, Line 7: What is meant by "global" statistic? This word has been used at several places Page 3233, Line 21: Change "Goundwater" to "Groundwater"

*We agree with the suggestions, the revised paper has been changed accordingly.*

*[1] Box, G. E. and Jenkins, G. M.: Time series analysis, control, and forecasting, Revised Edition, San Francisco, CA: Holden Day, 1976.*