

Manuscript: "Virtual Rain: A Unified Toolkit for High-Resolution Rainfall Simulation and Disaggregation"

Authors Reply

(in black Reviewers' comments; in blue Authors' comments; in red Authors' actions)

Reviewer 1

General Comments

The manuscript introduces Virtual Rain, a two-step toolkit that: a) generates synthetic daily rainfall time series, and b) disaggregates them to finer temporal resolution (defined by the user). The framework is implemented through a set of Python and R routines designed to facilitate practical application. The toolkit is also available through an interactive web-based platform, facilitating its use by a broad range of users. The Authors illustrate and evaluate Virtual Rain performance through a real-world case study. Based on the above, I recommend that the manuscript is accepted for publication after minor revisions. After careful reading of the manuscript, I recommend that the Authors provide some clarifications, as follows:

Authors' comment: We sincerely thank the Reviewer for the careful reading of the manuscript and for the positive evaluation of our work. We will carefully consider all the comments and suggestions provided and revise the manuscript accordingly. Detailed responses to each point are reported below, together with the corresponding modifications that will be introduced in the revised version of the manuscript.

Comment 1

Routine M1.2 is used to identify the probability distribution that best describes rainfall intensities in each season. If I understand correctly, the routine provides as output the estimated parameters for all candidate distributions. I think it will be helpful for the user to get as output, also, the name of the best distribution model.

Authors' comment: We thank the Reviewer for this valuable suggestion. The Reviewer is correct that Routine M1.2 returns the estimated parameters for all candidate probability distributions, allowing users to inspect the fitted models. We intentionally chose not to automatically report a single "best" distribution because the objective of the routine is to identify a single marginal distribution family to be adopted consistently across all seasons, while estimating its parameters separately for each season. When the same distribution provides the best fit for every season, the selection is straightforward. However, it is common for different distributions to provide the best seasonal fit in different seasons, meaning that no unique distribution is optimal across the entire dataset. In these situations, Routine M1.2 provides a summary table containing the Relative Mean Absolute Error (RMAE) values for each candidate distribution and season, allowing users to evaluate the overall performance of the competing distributions and select the most appropriate common marginal distribution for all seasons according to the specific objectives of the application. This design preserves the flexibility of the framework while ensuring consistency in the selected marginal model.

Authors' action: The manuscript will be revised to clarify that Routine M1.2 provides both the estimated parameters for all candidate distributions and a summary table of seasonal RMAE values. These outputs are intended to support an informed user-driven selection of a single marginal distribution family to be used consistently across all seasons, with parameters estimated independently for each season. When different distributions provide the best seasonal fit, the final choice is intentionally left to the user based on the overall RMAE summary.

Comment 2

Maybe I miss something here, but I think it would be helpful for the reader to have a clarifying note on the specific type of outputs produced by resM1.2 and resM1. 3 (lines 183-184), which are used as inputs for routine M1.4 (e.g. parameters of the best fitted distribution model, autocorrelation function).

Authors' comment: We thank the Reviewer for this helpful observation. We agree that explicitly describing the outputs of routines resM1.2 and resM1.3 improves the readability of the workflow. Routine resM1.2 provides the estimated parameters for all candidate marginal distributions, together with the observed dry-day frequency (p_0) and the RMAE summary used to support the selection of a common marginal distribution across seasons. Routine resM1.3 provides both the empirical and fitted autocorrelation structure coefficients. These outputs are then used as inputs by Routine M1.4 to generate multiple daily rainfall time series.

Authors' action: The manuscript will be revised to explicitly describe the outputs of routines resM1.2 and resM1.3 before introducing Routine M1.4. Specifically, it will be clarified that resM1.2 provides the estimated parameters for all candidate marginal distributions together with the observed dry-day frequency (p_0), whereas resM1.3 provides the empirical and fitted ACS coefficients, which are subsequently used by Routine M1.4 for rainfall generation.

Comment 3

In line 240, is there a minimum threshold for the length of annual maxima below which the model cannot be used?

Authors' comment: We thank the Reviewer for this relevant observation. The routine does not impose a minimum threshold on the length of the annual maxima series, and therefore it can technically be applied to datasets of any length. However, we acknowledge that short records may substantially reduce the robustness of the GEV parameter estimation and, consequently, the reliability of the estimated α -coefficients. Since these coefficients are subsequently used to calibrate the MRC disaggregation model, uncertainty in the GEV fit may propagate to the disaggregation performance. Therefore, although no minimum record length is enforced, longer annual maximum series are recommended to obtain more reliable calibration results. As an empirical indication, we applied the methodology to rainfall time series spanning only 20 years. Although the results naturally exhibited greater variability than would be expected from longer records, the overall calibration and disaggregation performance remained satisfactory.

Authors' action: The manuscript will be revised to clarify that the routine does not require a minimum number of annual maxima to operate. A note will be added to emphasize that short annual maximum records may reduce the robustness of the GEV fitting and the estimation of the scaling-based α -coefficients potentially affecting the calibration and performance of the MRC disaggregation model.

Comment 4

In line 245, the Authors state that the α -coefficients are associated with return periods ranging from 2 to 999 yr. Is there a particular limitation or practical reason that this range is used?

Authors' comment: We thank the Reviewer for this relevant question. The selected range of return periods (2–999 years) is not imposed by any methodological limitation of the GEV model but rather reflects a practical design choice adopted during the development of the software. The selected return periods are used to constrain the matching between the theoretical and empirical IDF curves during the calibration procedure. In practice, if the input rainfall time series spans n years, it is generally more robust to focus on return periods up to approximately $n/2$, as estimates associated with longer return periods become increasingly uncertain. Therefore, the upper limit of 999 years should be regarded simply as an operational maximum that provides flexibility to the user; in practical applications, such high return periods would rarely, if ever, be selected.

Authors' action: The manuscript will be revised to clarify the adopted return period range (2–999 years).

Comment 5

In line 249, how the α -coefficients relate to the inputs of routine M2.2 in lines 254–256?

Authors' comment: We thank the Reviewer for this helpful observation. We agree that the relationship between routines M2.1 and M2.2 should be made more explicit. The scaling-based α -coefficients estimated by Routine M2.1 constitute one of the main inputs to Routine M2.2, where they are used for setting up the

empirical IDF used for the calibration of the MRC disaggregation model. In addition, Routine M2.2 also uses the mean annual rainfall intensity estimated by M2.1 together with the observed rainfall time series.

Authors' action: The manuscript will be revised to explicitly state that the output of `GEV_Fitting()` (resM2.1) provides the scaling-based α -coefficients, the estimated mean annual rainfall intensity, and the observed rainfall time series, all of which are subsequently used as inputs to Routine M2.2.

Comment 6

In line 260, please state the specific output/s used.

Authors' comment: We thank the Reviewer for this helpful suggestion. We agree that the specific outputs used by the routine should be explicitly identified. As described above, the routine uses the output of `GEV_Fitting()` (resM2.1), which includes the scaling-based α -coefficients, the estimated mean annual rainfall intensity, and the observed rainfall time series.

Authors' action: The manuscript will be revised to explicitly specify the outputs of `GEV_Fitting()` used by the subsequent routine.

Comment 7

In line 261, please provide a typical range of admissible values.

Authors' comment: We thank the Reviewer for this helpful suggestion. The calibration routine requires admissible ranges for the three MRC parameters (C_β , C_{LN} , and D). These ranges are not arbitrary but are based on values commonly adopted in the literature for MRC model calibration. To facilitate reproducibility and ease of use, these literature-based bounds are implemented as default values both in the software routines and in the web-based platform, so users are not required to define them unless they wish to customize the calibration.

Authors' action: The manuscript will be revised to specify the default admissible parameter ranges adopted during calibration.

Comment 8

In line 283, the Authors mention that there is an option for the user to determine whether the routine will preserve daily rainfall totals. For disaggregation purposes, is it technically acceptable that the simulated daily rainfall totals are not preserved at finer temporal resolutions? Please clarify.

Authors' comment: We thank the Reviewer for this important observation. We agree that, in classical rainfall disaggregation, preserving the daily rainfall totals is generally the preferred and physically consistent option, as it ensures coherence between the daily and sub-daily time scales. Indeed, the original MRC model disaggregates daily rainfall while preserving daily rainfall totals, maintaining scale-invariant properties and reproducing the prescribed IDF scaling exponent through a stochastic MRC approach. For this reason, mass preservation is implemented as the default and recommended setting in both the Virtual Rain routines and the web platform. The option to disable daily total preservation was intentionally introduced to provide additional flexibility for advanced users. Some research applications may require the generation of sub-daily rainfall sequences that reproduce the statistical scaling properties of the MRC model without enforcing exact conservation of the daily totals. This option is therefore intended for specific methodological investigations rather than for standard rainfall disaggregation.

Authors' action: The manuscript will be revised to clarify that preserving daily rainfall totals is the default and recommended option for rainfall disaggregation, consistent with the original MRC formulation. A sentence will be added explaining that the alternative option, which does not enforce mass preservation, is provided solely to support advanced applications requiring greater flexibility in the generation of sub-hourly rainfall time series and is not intended as the standard operational setting.

Comment 9

In Figure 6 (upper panel), which type of marginal distribution is eventually used, as the reproduction of the ACS is highly affected by its selection. Also, is it technically acceptable that ACS preservation depends on the type of marginal distribution selected? Also, please note that the B autocorrelation function corresponding to Burr XII distribution, does not match the empirical one. Please provide some explanation.

Authors' comment: We thank the Reviewer for this important observation. Routine `ACS_Fitting()` is independent of the marginal distribution selected in Routine M1.2. It starts from the empirical Spearman autocorrelation structure of the binary occurrence process $B(t)$ and the continuous intensity process $X(t)$, and estimates the parameters of the candidate parametric ACS models by minimizing the discrepancy between empirical and theoretical autocorrelation values. Therefore, the upper panel of Figure 6 does not correspond to a specific marginal distribution, and the fitting of the ACS is not affected by the marginal model adopted for rainfall intensities.

The user can subsequently choose whether to use the fitted parametric ACS or directly adopt the empirical ACS in the simulation routines. The latter option avoids any loss of information introduced by the parametric approximation and may be preferable when none of the candidate ACS models adequately reproduces the observed dependence structure.

We also agree that the Burr XII ACS curve for the occurrence process $B(t)$ does not closely match the empirical ACS. This result is plausible because the empirical ACS exhibits a slower decay than that reproduced by the fitted Burr XII ACS. The discrepancy therefore reflects a limitation of the selected parametric form in representing the observed persistence, rather than an effect of the marginal distribution.

Authors' action: The manuscript will be revised to clarify that `ACS_Fitting()` operates independently of the marginal distribution selected in Routine M1.2. We will also specify that users may choose between the fitted parametric ACS and the empirical ACS when generating the daily rainfall series. In addition, the discussion of Figure 6 will be expanded to explain that the mismatch observed for the Burr XII fit of $B(t)$ is due to the slower decay of the empirical ACS, which is not fully captured by the selected parametric model.

Comment 10

The model tends to overestimate the observed rainfall statistics (see Figure 7). Is this fact related to some BIAS issue? Please clarify.

Authors' comment: We thank the Reviewer for this observation. The slight overestimation observed in Figure 7 should not be interpreted as evidence of a systematic bias of the model. As Virtual Rain is based on a stochastic rainfall generator, a certain degree of variability around the observed statistics is expected across simulated realizations. In this specific case, the observed overestimation is primarily attributable to the ability of the selected marginal distribution to reproduce the empirical rainfall intensity distribution. This issue is exacerbated by the relatively limited amount of information available for calibration: the case study consists of 20 years of daily observations, of which nearly 70% correspond to dry days. Consequently, the effective sample of non-zero rainfall events available to estimate the marginal distribution is relatively small, making an exact reproduction of the observed statistics inherently more challenging.

Authors' action: The manuscript will be revised to clarify that the differences observed in Figure 7 do not indicate a systematic model bias but rather reflect the combined effects of the stochastic nature of the rainfall generator and the limited number of non-zero observations available for calibrating the marginal distribution. A brief discussion will be added to explain that this limitation is particularly relevant for relatively short daily rainfall records characterized by a high proportion of dry days.

Comment 11

Also, it seems that in August the monthly simulated rainfall tends to be larger than September and almost equal to October (see Figure 9). Any particular reason for this? Please clarify.

Authors' comment: We thank the Reviewer for this insightful observation. The slight discrepancy observed between the simulated monthly rainfall totals in August, September, and October is primarily attributable to the adopted seasonal aggregation rather than to a deficiency of the rainfall generator. In the present case

study, the model was configured using three seasonal groups instead of a monthly (12-season) parameterization. This choice was made because the available rainfall record spans only approximately 20 years, making monthly calibration less robust due to the limited number of observations per month. Consequently, some intra-seasonal differences, such as those between August and September, are partially smoothed by the seasonal parameterization. This represents an intentional trade-off between preserving seasonal variability and ensuring robust parameter estimation from the available data.

Authors' action: The manuscript will be revised to clarify that the slight discrepancies in the simulated monthly rainfall pattern are a consequence of adopting a three-season parameterization. A sentence will be added explaining that this configuration was intentionally selected as a compromise between model parsimony and the robustness of parameter estimation when relatively short rainfall records are available, whereas a monthly (12-season) parameterization would require substantially more data to obtain reliable calibration.

Reviewer 2

General Comments

The manuscript presents Virtual Rain, a toolkit linking the CoSMoS-2s daily generator to an IDF-calibrated multifractal cascade, distributed as Python and R routines with a web interface and demonstrated on one Arno Basin gauge over 2001 to 2020. The motivation is sound: sub-hourly records are scarce, daily records and IDF curves are not. My first concern is what is new. Module 1 is CoSMoS-2s (Papalexiou, 2018, 2022), already implemented in Python by the same group (Cappelli et al., 2024). Module 2 is the cascade of Cappelli et al. (2025a), previously applied to this basin (Cappelli et al., 2025b). The coupling of the two is the framework of Cappelli et al. (2026). The manuscript says as much at lines 33 to 34, 322 to 326 and 440 to 442. What remains is the implementation and the platform, which can be a worthwhile contribution but should be claimed as such and substantiated. I have not seen Cappelli et al. (2026) and cannot judge the overlap directly. My second concern is the evaluation. Calibration and assessment use the same record; Module 2 rests on one realization; a 20-year observation is compared with a 1000-year simulation; the two modules' errors are not separated; the source of n is not reported; and the main text and Appendix disagree numerically. Most of this calls for new computation rather than rewriting.

Authors' comment: We sincerely thank the Reviewer for the thorough and constructive assessment of our manuscript. We appreciate the recognition of the motivation underlying Virtual Rain and the insightful comments regarding both the novelty of the contribution and the evaluation strategy.

Regarding the first point, we agree that the manuscript did not sufficiently distinguish the methodological developments from the software contribution. We will therefore substantially revise the Introduction, Discussion, and Conclusions to explicitly identify the manuscript as a software paper. In particular, we will clarify that Virtual Rain does not introduce new stochastic rainfall methodologies but provides the first integrated software implementation of the methodological framework proposed by Cappelli et al. (2026), which combines the CoSMoS-2s daily rainfall generator (Papalexiou, 2022) with the MRC rainfall disaggregation methodology developed in Cappelli et al. (2025a, 2025b). We will also explicitly distinguish Virtual Rain from PyCoSMoS (Cappelli et al., 2024), emphasizing that the latter implements the earlier CoSMoS-1s methodology, whereas Virtual Rain implements the more advanced two-state CoSMoS-2s framework together with the complete MRC disaggregation workflow.

Regarding the second point, we agree that the distinction between methodological validation and software verification should be made clearer throughout the manuscript. The purpose of the presented case study is not to provide a new validation of the CoSMoS-2s or MRC methodologies—which have already been developed and validated in the cited methodological papers—but rather to demonstrate the implementation and application of the complete Virtual Rain workflow. Accordingly, we will revise the manuscript to explicitly distinguish calibration targets, software verification, and methodological validation, and clarify the rationale underlying the adopted evaluation strategy.

Following the Reviewer's suggestions, we will also introduced numerous clarifications throughout the manuscript, including: (i) the source and role of the imposed scaling exponent n ; (ii) the relationship between the outputs of the different software modules; (iii) the rationale for the adopted seasonal grouping; (iv) the interpretation of the statistical comparisons presented in the case study; (v) the propagation of

uncertainty from Module 1 to Module 2; (vi) the intended use of empirical and parametric autocorrelation structures; and (vii) the limitations associated with relatively short rainfall records. In addition, a dedicated Discussion section will be introduced to explicitly interpret the results and discuss the scope and limitations of the software.

We believe that these revisions substantially strengthen the manuscript and more clearly position Virtual Rain as an operational, reproducible, and user-friendly software implementation of previously established stochastic rainfall methodologies. While some of the additional analyses suggested by the Reviewer would undoubtedly provide valuable methodological insight, we believe they are beyond the intended scope of the present software paper and constitute important directions for future methodological investigations.

Comment 1

State what is new, and verify the software. Please say whether this is a software paper, technical note or application paper, and align title, abstract and conclusion accordingly. A short table identifying what is new relative to Cappelli et al. (2024, 2025a, 2025b, 2026) would help, together with a direct statement on how M1.1 to M1.6 relate to PyCoSMoS. If the contribution is the implementation, it needs testing rather than description: agreement with the original CoSMoS-2s and MRC codes, agreement between platform and routines under identical inputs and seeds, fixed-seed reproducibility, a volume-conservation test, runtime, and a versioned archive with licence and dependencies.

Authors' comment: We thank the Reviewer for this valuable comment. We agree that the novelty and scope of the manuscript should be stated more explicitly. The present work is intended as a software paper, whose primary contribution is not the development of new stochastic rainfall methodologies but the implementation and integration of previously published methods through an open, reproducible, and user-friendly software environment.

The methodological framework implemented in Virtual Rain was introduced by Cappelli et al. (2026), which integrates two complementary components: (i) the CoSMoS-2s stochastic rainfall generator originally proposed by Papalexiou (2022) for daily rainfall simulation, and (ii) the Multifractal Random Cascade (MRC) rainfall disaggregation methodology developed in Cappelli et al. (2025a) and subsequently refined in Cappelli et al. (2025b) for sub-hourly rainfall generation. While these studies established and validated the underlying methodologies, they did not provide a publicly available software implementation tailored for a common user.

Accordingly, Module 1 of Virtual Rain implements the complete CoSMoS-2s workflow, including data preprocessing, automatic configuration, calibration, stochastic simulation, and diagnostic evaluation. This implementation should not be regarded as an extension of PyCoSMoS (Cappelli et al., 2024), which implements the earlier CoSMoS-1s framework. In CoSMoS-1s, rainfall is modeled as a single intermittent process, whereas CoSMoS-2s explicitly separates rainfall occurrence and rainfall intensity into two coupled stochastic processes. This two-state formulation provides a more realistic representation of wet and dry spell dynamics and improves the reproduction of rainfall statistics across multiple temporal scales.

Similarly, Module 2 implements the complete MRC workflow, including GEV fitting, estimation of scaling-based IDF coefficients, calibration of the MRC parameters, and stochastic temporal disaggregation. Virtual Rain therefore represents the first publicly available software integrating both the complete CoSMoS-2s daily rainfall simulation workflow and the complete MRC sub-hourly rainfall disaggregation workflow within a unified framework implemented in Python and R and accessible through an interactive web-based platform.

Accordingly, the objective of the present manuscript is to document the software architecture, implementation, functionality, and usability of Virtual Rain rather than to re-validate the underlying stochastic methodologies, whose methodological development and validation have already been presented in Papalexiou (2022) and Cappelli et al. (2025a, 2025b, 2026).

Regarding software verification, the focus of this work is therefore on verifying the correctness and consistency of the software implementation. The standalone Python and R routines and the web platform implement the same computational workflow and parameterization, ensuring consistency across all software components. Reproducibility is guaranteed through user-controlled random seeds available in both modules, while the MRC implementation preserves daily rainfall totals by default, consistently with the original methodology. Furthermore, the case study demonstrates that the implemented software reproduces the statistical behavior expected from the original CoSMoS-2s and MRC formulations. We agree that additional software-engineering benchmarks, such as runtime analysis and formal equivalence tests between the

standalone routines and the web platform, would further enrich the software documentation and will be considered in future releases.

Authors' action: The manuscript will be revised to more clearly define the scope and contribution of Virtual Rain as a software paper. Specifically:

- the Introduction and Conclusions will explicitly state that the objective of the manuscript is the implementation and dissemination of previously validated methodologies rather than their methodological development or re-validation;
- the relationship between PyCoSMoS (CoSMoS-1s) and Virtual Rain (CoSMoS-2s) will be clarified, emphasizing that Virtual Rain implements the two-state CoSMoS-2s framework rather than extending the CoSMoS-1s implementation;
- the roles of Module 1 and Module 2 will be explicitly described, clarifying that Module 1 implements the complete CoSMoS-2s workflow, whereas Module 2 implements the complete MRC rainfall disaggregation workflow;
- the software verification strategy will be clarified by specifying that the Python and R routines and the web platform implement the same computational workflow, that reproducibility is ensured through fixed random seeds, and that daily rainfall totals are preserved by default during MRC disaggregation, consistently with the original methodology.

Comment 2

Report where n came from. The origin of $n = 0.251$ is not given. If it was estimated from the 15-minute record, that record supplied scale information, and the statement at lines 415 to 416 that it was not used for calibration would need qualifying; if it came from a regional IDF product, please cite it. Agreement between imposed and simulated n (lines 427 to 429) shows the calibration works as intended, not that the model predicts well, and the wording should reflect that. Please also mark n as an imposed input in Table 3. Clarify and more directly demonstrate the data-sparse pathway.

Authors' comment: We thank the Reviewer for this important observation. The Reviewer is correct that, in the present case study, the imposed benchmark scaling exponent $n = 0.251$ was estimated from the observed 15-minute rainfall series. Therefore, although the 15-minute observations were not directly used to calibrate the MRC model parameters (C_β , C_{LN} , and D), they were used to estimate the target scaling exponent governing the IDF relationship. We agree that the original wording could be misleading, and we will revise the manuscript to explicitly clarify this point.

We also agree that the agreement between the imposed and simulated values of n demonstrates that the calibration procedure successfully reproduces the prescribed scaling behavior, rather than representing an independent validation of the model predictive capability. The text will be modified to reflect this distinction more clearly.

Finally, Table 3 will be updated to explicitly identify n as an imposed calibration input, rather than as a simulated output.

Authors' action: The manuscript will be revised as follows:

- the description of the case study will explicitly state that the imposed scaling exponent ($n = 0.251$) was estimated from the observed 15-minute rainfall series;
- the sentence describing the benchmark will be revised to clarify that the 15-minute observations were used only to estimate the target scaling exponent, whereas the calibration of the MRC parameters (C_β , C_{ln} , and D) was performed independently through the MRC calibration procedure;
- the discussion of the results will be revised to clarify that the agreement between the imposed and simulated scaling exponent demonstrates the correct implementation of the calibration procedure rather than the predictive skill of the model;
- Table 3 will be updated to explicitly identify the scaling exponent n as an imposed calibration input.

Comment 3

The workflow estimates $a(T)$ from daily annual maxima and the imposed exponent n , while M2.2 also permits users to supply externally derived a -coefficients. The present case could represent a data-sparse application if n was obtained independently of the 15-minute record, but this cannot be assessed because the source of n is

not reported. Please clarify this point. A complementary application using n and/or $a(T)$ from a published regional IDF product, with the 15-minute record reserved strictly for evaluation, would provide a more direct demonstration of the claimed use case.

Authors' comment: We thank the Reviewer for this insightful observation. We agree that, in the present case study, the scaling exponent n was estimated from the observed 15-minute rainfall series, and we will clarify this point in the revised manuscript.

We would also like to emphasize that the possibility of supplying externally derived $a(T)$ coefficients, instead of estimating them directly from the available rainfall series, was intentionally incorporated into Virtual Rain to increase the flexibility of the software. While this option certainly addresses situations where only short or incomplete rainfall records are available, its rationale extends beyond data scarcity. In particular, it allows users to adopt $a(T)$ coefficients derived from regional or national IDF products, previous studies, or expert knowledge, thereby supporting applications in ungauged locations or regions with limited observations. Furthermore, it enables users to investigate plausible future scenarios by modifying the input $a(T)$ coefficients to account for anticipated changes in rainfall extremes, for example those associated with climate change, without requiring modifications to the underlying MRC methodology. We believe that this flexibility represents one of the strengths of the proposed software.

Regarding the suggested complementary application, we agree that using externally derived n and/or $a(T)$ values (e.g., from regional IDF products) while reserving sub-daily observations exclusively for evaluation would constitute an interesting additional case study. However, demonstrating this application lies beyond the scope of the present software paper, whose primary objective is to document the implementation and functionality of the Virtual Rain framework rather than to exhaustively evaluate all possible application scenarios.

Authors' action: The manuscript will be revised to explicitly state that, in the presented case study, the imposed scaling exponent n was estimated from the observed 15-minute rainfall series. In addition, the description of Routine M2.2 will be expanded to clarify that the option of supplying externally derived $a(T)$ coefficients were intentionally introduced to provide greater flexibility. Specifically, it allows users to (i) overcome limitations associated with short or unavailable sub-daily rainfall records, (ii) directly incorporate regional or published IDF information, and (iii) investigate plausible future rainfall scenarios by prescribing modified $a(T)$ coefficients that reflect expected changes in rainfall extremes, such as those associated with climate change. This clarification better highlights the intended scope and versatility of the Virtual Rain framework.

Comment 4

Add independent validation. Seasonal grouping, marginal distributions, p_0 , ACS form and AR order are all chosen from the full record, and Figures 7 to 9 then compare against statistics from that same record. Some agreement follows by construction. A blocked design would be appropriate, for example calibration on 2001 to 2010 and verification on 2011 to 2020. Please also separate, in the text, calibration targets from software checks and from independent tests.

Authors' comment: We thank the Reviewer for this thoughtful suggestion. We agree that a blocked calibration–verification design is appropriate when assessing the predictive performance of statistical or machine-learning models. However, the objective of CoSMoS-2s and the MRC model is fundamentally different. Both are stochastic simulation models, whose purpose is not to predict future rainfall observations but to generate synthetic rainfall realizations that reproduce the statistical properties of the observed process. Consequently, the model parameters (seasonal grouping, marginal distributions, dry-day frequency, autocorrelation structure, and autoregressive order) are estimated from the available record to characterize the underlying stochastic process. The subsequent comparison between observed and simulated statistics is therefore intended to verify the ability of the calibrated stochastic model to reproduce the target statistical characteristics, rather than to assess predictive performance on an independent validation dataset.

Moreover, the results presented in this manuscript intentionally follow the workflow proposed in Cappelli et al. (2026), which constitutes the methodological foundation of Virtual Rain. Since the purpose of the present manuscript is to document the software implementation and guide users through the complete workflow, we believe it is important to preserve the same sequence of calibration, simulation, and statistical evaluation that users will obtain when applying the software. Introducing an independent calibration–

validation experiment would shift the focus from software implementation toward methodological assessment, whereas the latter has already been addressed in Papalexiou (2022) and Cappelli et al. (2025a, 2025b, 2026).

We nevertheless agree that the distinction between calibration targets, software verification, and methodological validation should be stated more explicitly in the manuscript.

Authors' action: The manuscript will be revised to clarify that the comparisons reported in Figures 7–9 are not intended to evaluate predictive performance, but rather to verify that the implemented stochastic simulation workflow correctly reproduces the target statistical characteristics of the observed rainfall process after calibration. Additional text will be included to more clearly distinguish (i) the calibration of the stochastic models, (ii) the verification of the software implementation through the reproduction of the imposed statistical properties, and (iii) the methodological validation of the CoSMoS-2s and MRC models, which has already been presented in Papalexiou (2022) and Cappelli et al. (2025a, 2025b, 2026).

Comment 5

Assess Module 2 with an ensemble at matched length. Line 411 and Figure A.12 indicate a single realization, which cannot show how stable the tail metrics are and sits oddly beside the 50-member ensemble used for Module 1. Table 3 also sets 20 years of observation against a 1000-year simulation, so record length and model bias are confounded in the maxima rows. Please use an ensemble matched to the observed length, report each observed statistic as a rank within it, and show convergence. Long-record return levels can be reported separately with their uncertainty.

Authors' comment: We thank the Reviewer for this thoughtful suggestion. We agree that ensemble-based analyses are appropriate for quantifying the variability of stochastic simulations. However, the objectives of Modules 1 and 2 are substantially different, and this motivated the different presentation adopted in the manuscript.

For Module 1 (CoSMoS-2s), multiple realizations are generated because the user must select among several competing model configurations (seasonal grouping, marginal distribution, autocorrelation structure, and autoregressive order). These choices represent multiple sources of uncertainty, and generating an ensemble of realizations provides the user with a robust basis for evaluating the quality and stability of the selected configuration. The 50-member ensemble is therefore an integral part of the CoSMoS-2s configuration workflow proposed by Cappelli et al. (2026).

For Module 2 (MRC), the situation is fundamentally different. Once the daily rainfall series and the scaling exponent n (estimated from sub-daily observations or obtained from regional IDF relationships) are available, the MRC model performs the stochastic temporal disaggregation using the calibrated parameter set. The objective of this section is therefore to demonstrate the implementation and use of the software rather than to re-evaluate the statistical performance of the MRC methodology.

The methodological validation of the MRC model, including extensive ensemble analyses and assessment of its statistical performance, has already been presented in Cappelli et al. (2025a, 2025b). Reproducing these analyses within the present manuscript would substantially increase the computational burden while shifting the focus from software implementation to methodological validation, which is beyond the scope of this software paper.

Authors' action: The manuscript will be revised to better explain the different roles of the two modules. In particular, we will clarify that the 50-member ensemble adopted in Module 1 is required to support the automatic configuration and evaluation of competing CoSMoS-2s model structures, whereas Module 2 illustrates the implementation of an already calibrated MRC methodology. Additional text will be included to explicitly state that the statistical validation of the MRC model has already been reported in Cappelli et al. (2025a, 2025b), and that the purpose of the present manuscript is to document its software implementation and practical application rather than to repeat its methodological validation.

Comment 6

Run Module 2 on the observed daily series. The GEV behind $a(T)$ is fitted to the simulated 1000-year series (Figures 10b and 11), and Figure 7 suggests that series overestimates the daily upper tail, so Module 1 bias can propagate into the calibration and into Table 3. The attribution of the extreme-value differences to record

length at lines 424 to 425 has not been tested against this. The toolkit already supports the experiment (lines 47 to 49); it is inexpensive and would clarify how Table 3 should be read.

Authors' comment: We thank the Reviewer for this valuable suggestion. We agree that applying Module 2 directly to the observed daily rainfall series represents an appropriate workflow when reliable daily observations are available. Indeed, this capability is one of the key features of Virtual Rain and is explicitly described in the Introduction (lines 47–49), where we state that Module 2 can be applied independently of Module 1 whenever observed daily rainfall data are available.

The objective of the presented case study, however, was to illustrate the complete integrated workflow implemented in Virtual Rain, from stochastic daily rainfall generation (Module 1) to sub-hourly rainfall disaggregation (Module 2). For this reason, Module 2 was intentionally applied to the synthetic daily rainfall series generated by Module 1, thereby demonstrating the end-to-end functionality of the software.

We also agree that directly applying Module 2 to the observed daily rainfall series would isolate the performance of the disaggregation model from the uncertainty introduced by the stochastic rainfall generator. This analysis has already been performed and discussed in Cappelli et al. (2026b), where the MRC methodology was evaluated using observed daily rainfall as input. Since the primary objective of the present manuscript is to document the software implementation rather than to repeat the methodological validation of the MRC model, we would prefer not to include this additional experiment here.

Regarding the Reviewer's observation on the attribution of the differences in extreme rainfall statistics, we agree that the original wording was too restrictive. The comparison presented in this manuscript does not allow the effect of record length to be disentangled from the uncertainty introduced by the stochastic daily rainfall generator. We will therefore revise the Discussion to explicitly refer readers to Cappelli et al. (2026b), where this aspect is examined in detail using observed daily rainfall as input.

Authors' action: The manuscript will be revised to clarify that the case study was intentionally designed to demonstrate the complete Virtual Rain workflow, combining Modules 1 and 2. Additional text will be included to emphasize that Module 2 can also be applied directly to observed daily rainfall series whenever reliable daily data are available. Furthermore, the discussion of the extreme rainfall statistics will be revised to acknowledge that the observed differences may result from both the different record lengths and the uncertainty propagated from the stochastic daily rainfall simulation. A citation to Cappelli et al. (2026b) will be added, explicitly indicating that the application of the MRC model to observed daily rainfall series and the corresponding assessment of extreme-value reproduction are presented in that methodological study.

Comment 7

Add benchmarks. The toolkit is compared only with the data used to calibrate it, so its added value over simpler procedures cannot be judged. Please add one daily-generation baseline and one established disaggregation baseline under the same periods, record length and metrics. A Markov occurrence model with a fitted intensity distribution, and a method-of-fragments or resampling scheme, are reasonable examples; other choices would serve. Where a baseline needs sub-daily data, please say so.

Authors' comment: We thank the Reviewer for this valuable suggestion. We agree that benchmarking the implemented methodologies against simpler stochastic rainfall generators and alternative disaggregation approaches would provide additional insight into their relative performance.

However, the primary objective of the present manuscript is to introduce Virtual Rain as a software platform, providing an operational implementation of the CoSMoS-2s and MRC methodologies through reproducible Python and R routines and an interactive web-based interface. Accordingly, the emphasis is placed on documenting the software architecture, workflow, and functionality, enabling users to readily apply these methodologies rather than on conducting an extensive methodological comparison with alternative models. We therefore believe that including benchmark experiments against other stochastic rainfall generators and disaggregation techniques would substantially broaden the scope of the manuscript and shift its focus from software implementation to methodological evaluation. Nevertheless, we agree that such comparative analyses would be valuable and represent a natural direction for future work, particularly to assess the relative strengths and limitations of different stochastic rainfall simulation and disaggregation approaches within a common software framework.

Authors' action: The manuscript will be revised to more clearly emphasize that Virtual Rain is presented as a software paper, whose primary contribution lies in the implementation and dissemination of established stochastic rainfall methodologies within an integrated software environment. To better define this scope, additional text will be included in both the Introduction and the Conclusions. We also acknowledge that a comprehensive benchmarking against alternative daily rainfall generators and temporal disaggregation methods would provide valuable additional insight and represents a worthwhile direction for future development and evaluation of Virtual Rain.

Comment 8

Distinguish fitted dependence information from diagnostic behaviour.

The case study adopts empirical ACS inputs with $p=3$, and the Appendix states that the number of empirical ACS lags equals the AR order. The lag-1 to lag-3 dependence in the binary and continuous component models is therefore directly informed by the observed record. Although Figure 8 shows the ACF of the combined rainfall series rather than the fitted component ACSs themselves, agreement at these short lags is strongly constrained by calibration and should not be presented as fully independent validation. Longer lags provide a more independent diagnostic of whether the fitted dependence structure propagates beyond the calibration range. Please clarify this distinction and explain how the empirical ACS option was selected relative to the parametric alternatives.

Authors' comment: We thank the Reviewer for this insightful observation. We agree that, in the presented case study, the agreement observed at the first three lags should not be interpreted as an independent validation of the dependence structure. Since the empirical ACS option was selected and the autoregressive order was set to $p = 3$, the lag-1 to lag-3 dependence is directly informed by the observed autocorrelation structure used during calibration.

The purpose of Figure 8 is therefore not to provide an independent validation of the short-lag dependence, but rather to illustrate that the implemented simulation workflow correctly reproduces the dependence structure prescribed during model configuration. The longer lags, which are not directly constrained by the empirical ACS used for calibration, provide a more informative diagnostic of how the simulated dependence propagates beyond the fitted range.

We also clarify that the choice between the empirical and parametric ACS is left to the user. The empirical ACS was intentionally selected in this case study because it provided the closest representation of the observed dependence structure. The parametric ACS option remains available within Virtual Rain for applications where a smoother analytical representation is preferred or where extrapolation beyond the empirical lag range is required.

Authors' action: The manuscript will be revised to explicitly distinguish calibration information from diagnostic evaluation. Specifically, we will clarify that the agreement observed over the first p lags results from the use of the empirical ACS during model calibration and should therefore not be interpreted as an independent validation of the model's ability to reproduce the temporal dependence structure. The discussion of Figure 8 will be expanded to emphasize that the longer lags provide a more informative diagnostic of the model behavior beyond the calibration range. In addition, the description of Routine M1.3 will be updated to explain that users may choose either the empirical or the fitted parametric ACS, and that the empirical option was adopted in the present case study because it provided the most faithful representation of the observed autocorrelation structure.

Comment 9

Quantify uncertainty and examine identifiability. With roughly 20 annual maxima extrapolated to 100 and 200 years, uncertainty in the GEV parameters and derived coefficients is likely considerable and should be quantified by bootstrap or profile likelihood, with reported precision matched to the information available. The MRC objective function is never defined, so the value of $8.19 \text{ mm}^2/\text{h}^2$ cannot be interpreted and the description of it as low (lines 404 to 406) has no reference scale. Please give its expression, weighting and optimizer, and report parameter profiles, correlations, sensitivity to n and to the chosen return periods, and multiple optimization starts. The Appendix run cannot substitute, since it differs in several inputs at once.

Authors' comment: We thank the Reviewer for this detailed and technically important comment. We agree that, with approximately 20 annual maxima, extrapolation to long return periods is inevitably affected by considerable uncertainty. The resulting $a(T)$ coefficients should therefore be interpreted as calibration inputs conditioned on the available observations rather than as precise estimates of extreme rainfall quantiles. We will revise the manuscript to make this limitation more explicit.

We also agree that the MRC calibration objective and optimization procedure should be clearly documented. Rather than repeating the full methodological derivation, we will explicitly refer readers to Section 2.3 of Cappelli et al. (2025b), where the calibration procedure, including the objective function, weighting strategy, and optimization approach, is described in detail. The purpose of the present manuscript is to document the software implementation of that established methodology.

We further acknowledge that bootstrap confidence intervals, profile-likelihood analysis, parameter-identifiability assessment, sensitivity to n and to the selected return periods, and multi-start optimization would provide valuable additional insight. However, these analyses fall outside the scope of the present software paper and represent relevant directions for future methodological investigation.

Finally, we agree that describing the objective-function value of $8.19\text{mm}^2\text{h}^{-2}$ as “low” may be misleading, since no universal reference scale exists. We will therefore revise the corresponding wording and will report this value only as a calibration diagnostic for the specific case study.

Authors' action: The manuscript will be revised to:

- explicitly acknowledge the uncertainty associated with fitting a GEV distribution to a relatively short annual-maximum series and extrapolating to long return periods;
- remove the qualitative characterization of the objective-function value as “low”;
- report the value $8.19\text{mm}^2\text{h}^{-2}$ only as a case-specific calibration diagnostic;
- add an explicit reference to Section 2.3 of Cappelli et al. (2025b), where the MRC calibration procedure, objective function, weighting strategy, and optimization approach are described in detail;
- identify uncertainty quantification, parameter-identifiability analysis, sensitivity analysis, and multi-start optimization as possible directions for future work.

Comment 10

Support the seasonal partition. Lines 332 to 333 present Figure 4c as supporting three groups starting in April, but the panel shows a November maximum and a July minimum, with April on the rising limb, and no dispersion is shown around means based on 20 values each. Lines 368 to 372 then explain the seasonal discrepancies by the grouping without a control run. Please state the criterion, add dispersion, document the automated grouping option in Figure A.3, and compare one or two alternative settings.

Authors' comment: We thank the Reviewer for this thoughtful observation. We agree that including information on the variability of the monthly rainfall statistics would provide additional context. However, the purpose of Figure 4c is not to statistically justify the selection of three seasonal groups, but rather to provide users with a visual summary of the mean seasonal cycle to support the choice of the seasonal partition.

The seasonal grouping implemented in Virtual Rain is not selected automatically. Instead, users choose both the number of seasonal groups and the starting month, using the information provided in Figures 4c and A.3 together with their knowledge of the rainfall regime. The rationale is to group months exhibiting broadly similar rainfall characteristics while maintaining a sufficient sample size for robust estimation of the marginal distributions and autocorrelation structure.

We agree that aggregating similar—but not identical—months inevitably introduces some discrepancies in the monthly statistics, since each month retains its own variability. These discrepancies represent the expected consequence of adopting a parsimonious seasonal representation. A more accurate reproduction of the monthly seasonal cycle could be achieved by adopting a monthly (12-season) parameterization; however, this would substantially increase the number of parameters to estimate, reduce the amount of data available for each calibration period, and increase the computational cost. The adopted seasonal grouping therefore represents a deliberate trade-off between model parsimony, robustness of parameter estimation, and computational efficiency, particularly when relatively short rainfall records are available.

We will acknowledge that comparing alternative seasonal partitioning schemes would be informative. However, such an analysis would constitute a methodological sensitivity study, whereas the objective of the present manuscript is to document the implementation and use of the software.

Authors' action: The manuscript will be revised to clarify that Figures 4c and A.3 are intended to support user-defined seasonal grouping rather than to automatically determine the optimal partition. Additional text will be added to emphasize that the adopted grouping represents a compromise between faithfully reproducing monthly variability and obtaining robust parameter estimates from relatively short rainfall records. We will also explicitly state that a monthly (12-season) parameterization may improve the representation of the seasonal cycle but at the expense of increased model complexity, reduced sample size per season, and higher computational cost.

Comment 11

Add a Discussion, and match claims to evidence. The manuscript goes straight from the case study to the conclusion, so the results are not interpreted, not compared with the group's earlier work in this basin, and no limitation is stated. Worth discussing: the autocorrelation deviations in Table 3, which grow with lag; an outer-scale parameter of about 12 days alongside a cascade applied independently to each day; discrepancies concentrated near the seasonal-group boundaries; and Module 1 bias entering Module 2. If robustness and transferability are retained as claims, a multi-station evaluation is needed; otherwise these should be removed and the study described as a single-site demonstration. Please also give the station-selection criterion, missing-data fraction, quality control, and gauge resolution if applicable.

Authors' comment: We thank the Reviewer for this comprehensive and constructive comment. We agree that the manuscript would benefit from a clearer discussion of the results and of the limitations of the proposed software.

Accordingly, we will introduce a dedicated Discussion section to better interpret the case-study results, relate them to the underlying methodologies, and explicitly discuss the main limitations of the current implementation. In particular, we will discuss the deviations in the autocorrelation structure at longer lags, the implications of the adopted seasonal grouping, and the propagation of uncertainty from Module 1 to Module 2 within the integrated Virtual Rain workflow.

Regarding robustness and transferability, we agree that these properties cannot be inferred from a single-site application. The objective of the present manuscript is to demonstrate the implementation and application of the software through a representative case study, rather than to provide a comprehensive multi-site evaluation. Accordingly, we will revise the manuscript to avoid overly broad claims regarding the robustness or transferability of the approach. Instead, we will explicitly present the study as a single-site demonstration of the software workflow, highlighting its implementation while acknowledging that broader validation across multiple sites is beyond the scope of the present work.

With respect to the requested comparison with previous studies, we note that the methodologies implemented in Virtual Rain are those previously developed by Papalexiou (2022) and Cappelli et al. (2025a, 2025b, 2026). The present manuscript focuses on their software implementation rather than on a new methodological assessment. Nevertheless, the Discussion will explicitly place the presented results within the context of these previous methodological studies.

Finally, we agree that additional information describing the observational dataset improves the reproducibility of the case study. The manuscript will therefore be expanded to include the station-selection criterion together with information on missing data, quality-control procedures, and rain-gauge temporal resolution.

Authors' action: The manuscript will be revised as follows:

- a new Discussion section will be added, interpreting the main results and explicitly discussing the limitations of the proposed software implementation;
- the discussion will be expanded to address the behavior of the autocorrelation structure at increasing lags, the implications of seasonal grouping, and the propagation of uncertainty from Module 1 to Module 2;
- claims regarding robustness and transferability will be revised to clearly indicate that the presented application is a single-site demonstration intended to illustrate the software workflow rather than a comprehensive methodological evaluation;

- references to the previous methodological studies (Papalexiou, 2022; Cappelli et al., 2025a, 2025b, 2026) will be strengthened to distinguish the methodological developments from their software implementation;
- the description of the case-study dataset will be expanded to include the station-selection criterion, the proportion of missing data, the adopted quality-control procedure, and the temporal resolution of the rain-gauge observations.