Authors' Response

We thank the Editor and the Reviewers for taking time carefully reviewing our manuscript and for the constructive comments. All the reviewer comments have been considered and incorporated in the revised manuscript. Here, we first summarise the main changes made in the revised manuscript, and we list, point by point, our replies to reviewers' comments.

1. Summary of main changes in the revised manuscript

- Incorporated a comprehensive baseline model analysis (Section 4.2.1).
- Presented calibrated model parameters using full 69-year records (Table 2).
- Demonstrated the baseline model's capacity to preserve key statistical properties across various timescales (5-min, 1-h, 6-h and 24-h timescales) (Figures 3-6).
- Highlighted the current model's limitations at the monthly scale (Figure 7).
- Provided formulae for selected BL model statistical properties in Appendix A (Appendix A).
- Extended the description of the model calibration, including the explanations of the numerical challenges, our proposed numerical method and the factors affecting the model calibration quality (Section 2.2 and Supplement S1).
- Conducted a sensitivity analysis of BL model parameters and provided results as Supplement of the manuscript (Supplement S1).
- Suggested potential improvements in the future versions of the pyBL, such as incorporating the shuffling components method from Kim and Onof (2020) to better address large-scale rainfall variability caused by storm dependence (Sections 4.2.1 and 5).

2. Referee #1

1. The readers will benefit from seeing all formulations of the BL model - either in the section on the model structure (Section 2.1) or in the supplementary material if you do not wish to lengthen the paper. The lack of equations in a paper that describes a model is somewhat unexpected.

   R/ Thank you for the comment. Indeed, it would be helpful to have the formulations of the BL model. We have added them to the appendix of the revised manuscript.

2. In the case study presented, you address the issue of sample size. As part of the discussion of model calibration, I would present this information in advance to the reader.

   R/ Thank you for your comment. We have addressed this by incorporating the information regarding sample size in the model calibration section.

3. I suggest adding a short section to provide readers with a more comprehensive understanding of the sensitivity of the model parameters. You may list all model parameters in a table (this would be very useful for readers to gain a better understanding of the model engine) and present the local or global sensitivity.

   R/ Thank you for your comments. We will address both comments together as they are closely related.

   A further discussion on model calibration will indeed provide readers with a better understanding of the model and the sensitivity of the parameters. To our understanding, three main factors affect the sensitivity of model parameters, ranging from local to global sensitivities: (1) the capacity of the numerical solver to determine optimal parameters, (2) the estimation of observed rainfall properties, and (3) the sample size.

Before discussing these three factors, as suggested, we will first provide a summary of the model parameters and the calibration results obtained from 69 years of 5-min rainfall records at Bochum (the reference used in the submitted manuscript). This information will be presented in a new table (as Table R1 below), which we will add to the revised manuscript.

In terms of the sensitivity of model parameters, we begin with the local sensitivity introduced by the numerical solver. Determining optimal parameters for a Bartlett-Lewis (BL) model is always a numerical challenge due to its complexity. Several strategies have been proposed. For example, Onof and Wang (2020) introduced a 2-stage solver, employing simulated annealing for heuristic searching followed by the Nelder-Mead algorithm to efficiently refine optimal parameters. In our work, we utilised a basin-hopping algorithm to reduce the likelihood of being trapped in local optima and help identify optimal parameters. As noted by Baioletti et al. (2024), basin-hopping outperforms algorithms like Differential Evolution and Particle Swarm Optimization in terms of computational efficiency and solution accuracy. Our numerical solver runs basin-hopping iteratively 20 times for each model calibration. The first iteration starts with a randomly assigned initial guess, while subsequent iterations use the solution from the previous basin-hopping iteration to refine the optimal solution.

To demonstrate the impact of the numerical solver, we conducted an experiment using 69 years of rainfall records from Bochum. As shown in Fig. R1, when a fixed random seed (related to the initial guess) is used, the solver consistently results in the same parameters. When varying random seeds are used, the solver produces nearly identical parameters in most months, except for July and September, where greater variability in some parameters is observed. However, when these parameters are used to compute rainfall properties such as skewness at 5-minute and 1-day time scales (shown in Fig. R2), the variability in skewness estimates is minimal. Comparing these results with those derived from the bootstrapping method (Sect. 2.4 of the original manuscript) shows that the variability in rainfall properties from bootstrapping is consistently larger than that caused by the numerical solver. Even in July, where parameter variation from varying random seeds exceeds that from bootstrapping, the resulting variability in rainfall properties remains smaller. This suggests that the sensitivity of model parameters, and the derived rainfall properties, is largely driven by the estimation of observed rainfall properties rather than the numerical solver.

Finally, we conducted another experiment to examine the impact of sample size, using the bootstrapping method to derive model parameters from 5- and 69-year rainfall records. As shown in Fig. R3, sample size has a significant impact on parameter variability (or sensitivity). When the sample size is small, the variability in model parameters is much greater than when using the full records. Furthermore, this sensitivity propagates into rainfall properties, as can be seen when comparing Figures 2 and 4. We conclude that sample size has the largest impact on the sensitivity of model parameters compared to the other two factors discussed.

A detailed discussion of this experiment is provided in Supplement S1.

Table R1. Model parameters calibrated using full data records

Month	λ $[h^{-1}]$	ι [mm]	α [-]	α/ν $[h^{-1}]$	κ [-]	ϕ [-]
Jan	0.013	0.223	0.780	4.407	0.769	0.026
Feb	0.012	0.203	0.982	4.120	1.001	0.033
Mar	0.015	0.216	0.975	6.209	0.572	0.027
Apr	0.012	0.312	0.712	5.723	0.408	0.020
May	0.015	0.521	0.642	6.856	0.388	0.038
Jun	0.013	1.183	0.464	7.619	0.132	0.022
Jul	0.018	1.440	0.612	6.480	0.106	0.032
Aug	0.012	1.834	0.437	4.907	0.076	0.021
Sep	0.013	1.141	0.480	5.496	0.161	0.030
Oct	0.011	0.303	1.060	5.859	0.508	0.022
Nov	0.042	2.028	1.993	0.477	1.798	20.00
Dec	0.012	0.248	0.725	4.559	0.686	0.023
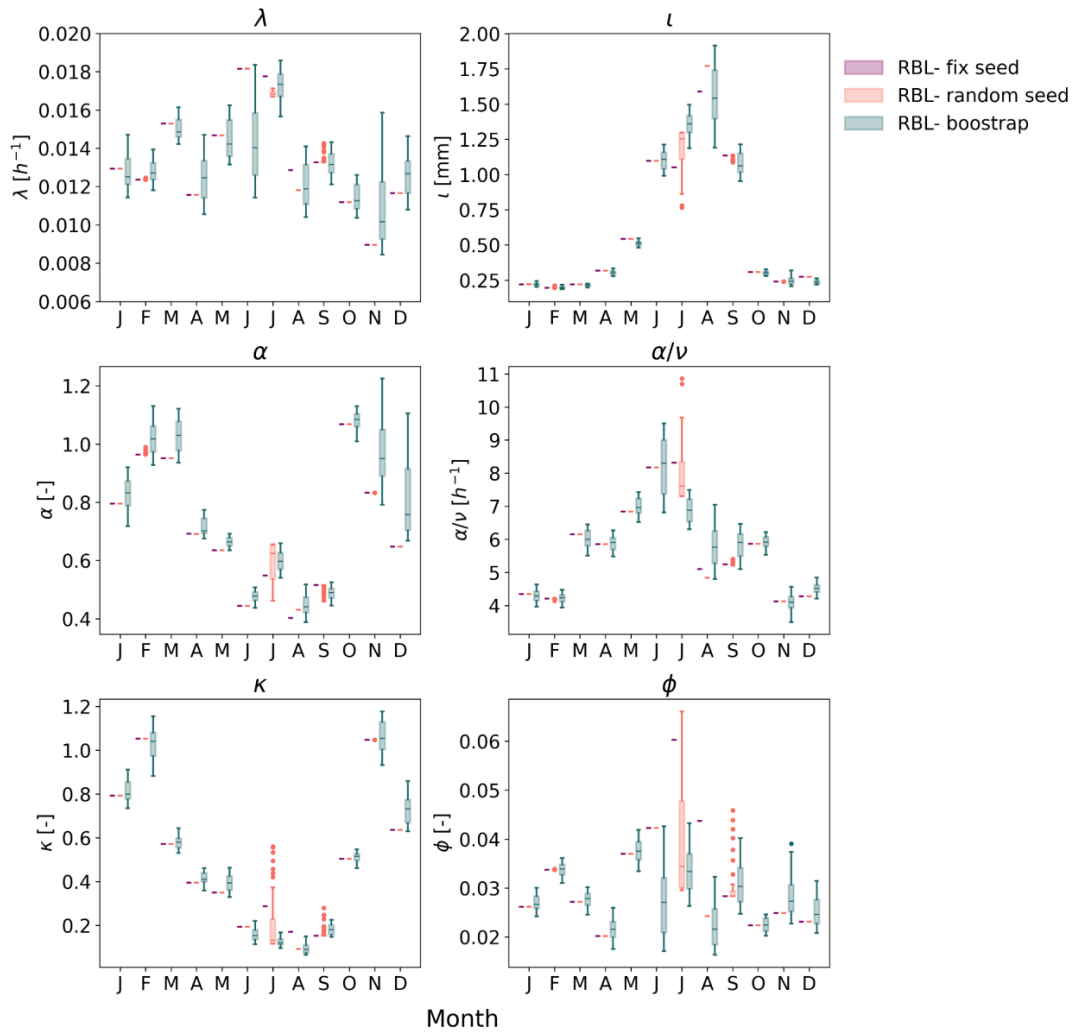
Figure R1. Variability of model parameters calibrated under three different scenarios: (1) fixed random seeds, (2) varying random seeds for initial guesses, and (3) the bootstrapping method.
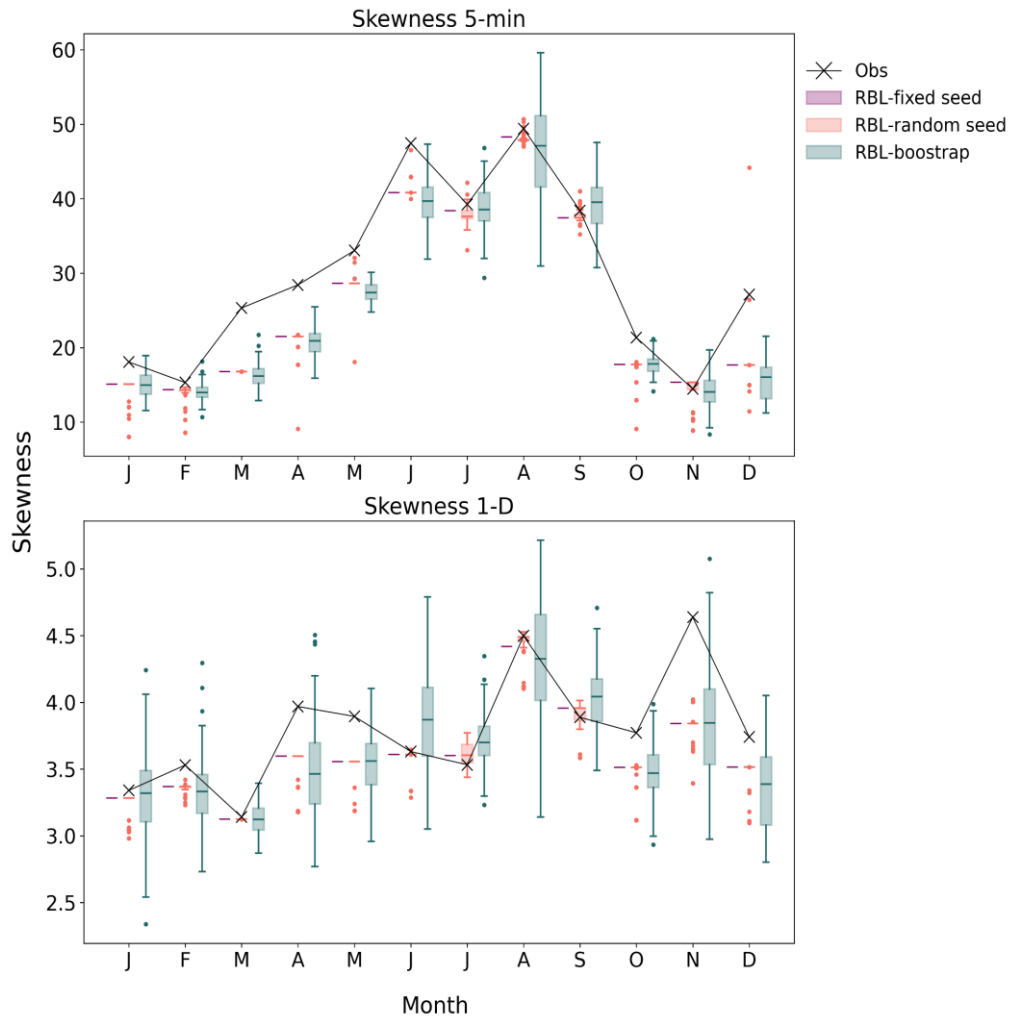
Figure R2. The corresponding 5-min and 1-day skewness estimates computed using the model parameters shown in Fig. R1.
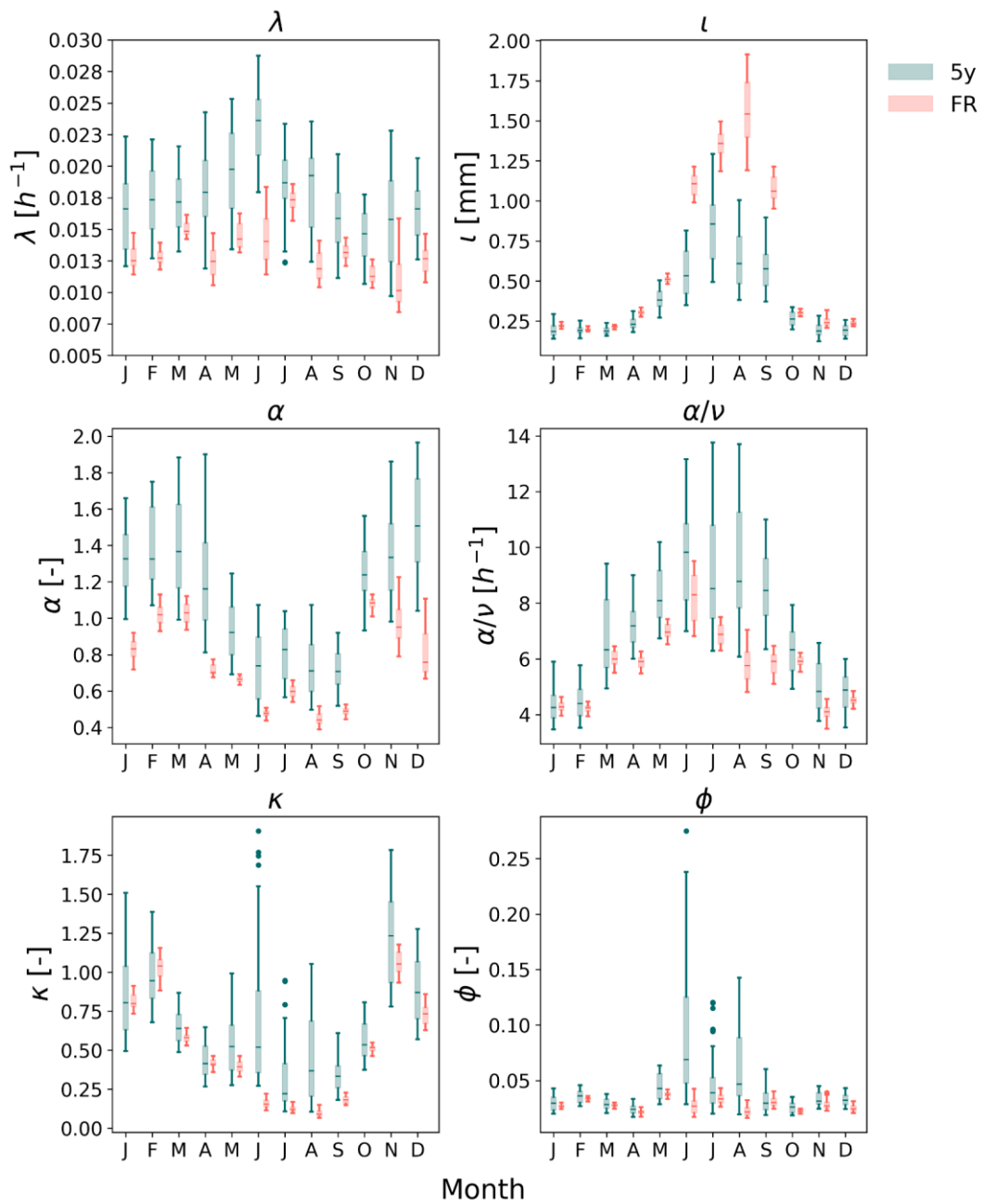
Figure R3. Boxplots of monthly model parameters (storm arrival rate (λ), cell intensity ratio (ι), cell duration shape (α), cell duration scale (ν), cell arrival rate (κ), and storm duration rate (φ)) calibrated using different record lengths (light blue boxes: 5 years, green boxes: full records). Each member of 100 bootstrapping iterations is calculated, and their interquartile ranges (IQR) are presented.
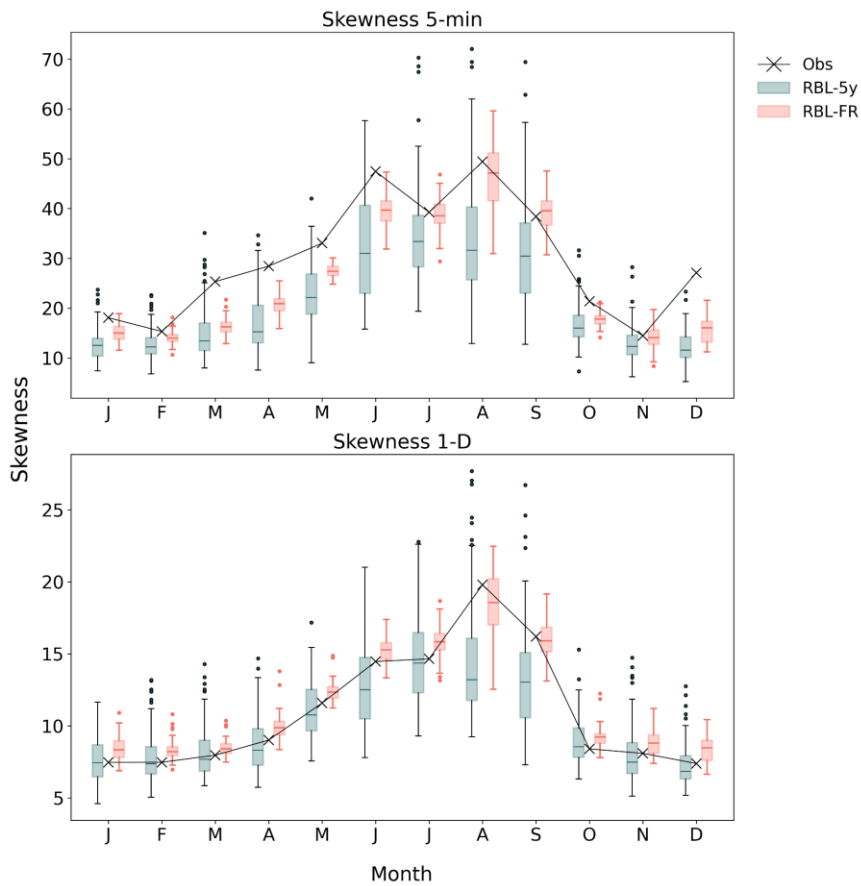
Figure R4. The corresponding 5-min and 1-day skewness estimates computed using the model parameters shown in Fig. R3.

4. It would be useful to begin by presenting several box plots demonstrating the model's ability to reproduce the interannual variability, monthly and daily rainfall statistics before presenting the results of the extreme rainfall (e.g., Figure 3). Currently, it appears that the model is only calibrated to simulate extreme events correctly.

R/ Thank you for the comment. Indeed, including information about interannual variability, and monthly and daily rainfall statistics is helpful to better understand the model's ability to reproduce 'standard' rainfall properties. It is also helpful to clarify that the BL model doesn't require extreme rainfall properties for model calibration. A new section discussing these findings, along with figures illustrating the standard statistics, has been added to the manuscript (Section 4.2.1).

Here, we follow the method used by Wang et al. (2006) and Kim, D. and Onof (2020) to calculate interannual variability, where observed monthly variance of the mean daily rainfall and the corresponding quantiles of monthly variance derived from 100 sample time series at each calendar month across study years (in this case, 69 years of rainfall records from Bochum) were used. As illustrated in Fig. R5, for all calendar months, the observed variances are well reproduced by the sampled ones over most variance range. However, it is also observed that the observed maximum variances for each calendar month tend to be overestimated by the sampled variances.

Apart from interannual variability, Figures R6-R9 show the standard rainfall properties derived from the calibrated BL model at selected timescales. As seen, the BL model can well reproduce all selected rainfall properties at daily and sub-daily timescales. However, for rainfall properties at the 1-month (1-M) timescale, only rainfall mean can be well reproduced. Failing in preserving monthly properties lies in the fact that these properties are not considered during model

calibration in this version of the BL model. This highlights the limitation of the current implementation. We understand the importance of a rainfall model to be able to reflect monthly rainfall variation, thus it may be addressed in the future version of the BL model. Candidate methods like adding the shuffling components proposed by Kim, D. and Onof (2020) can help involve the consideration of monthly rainfall variability, enhancing the ability of the BL model to reproduce this variation accurately.
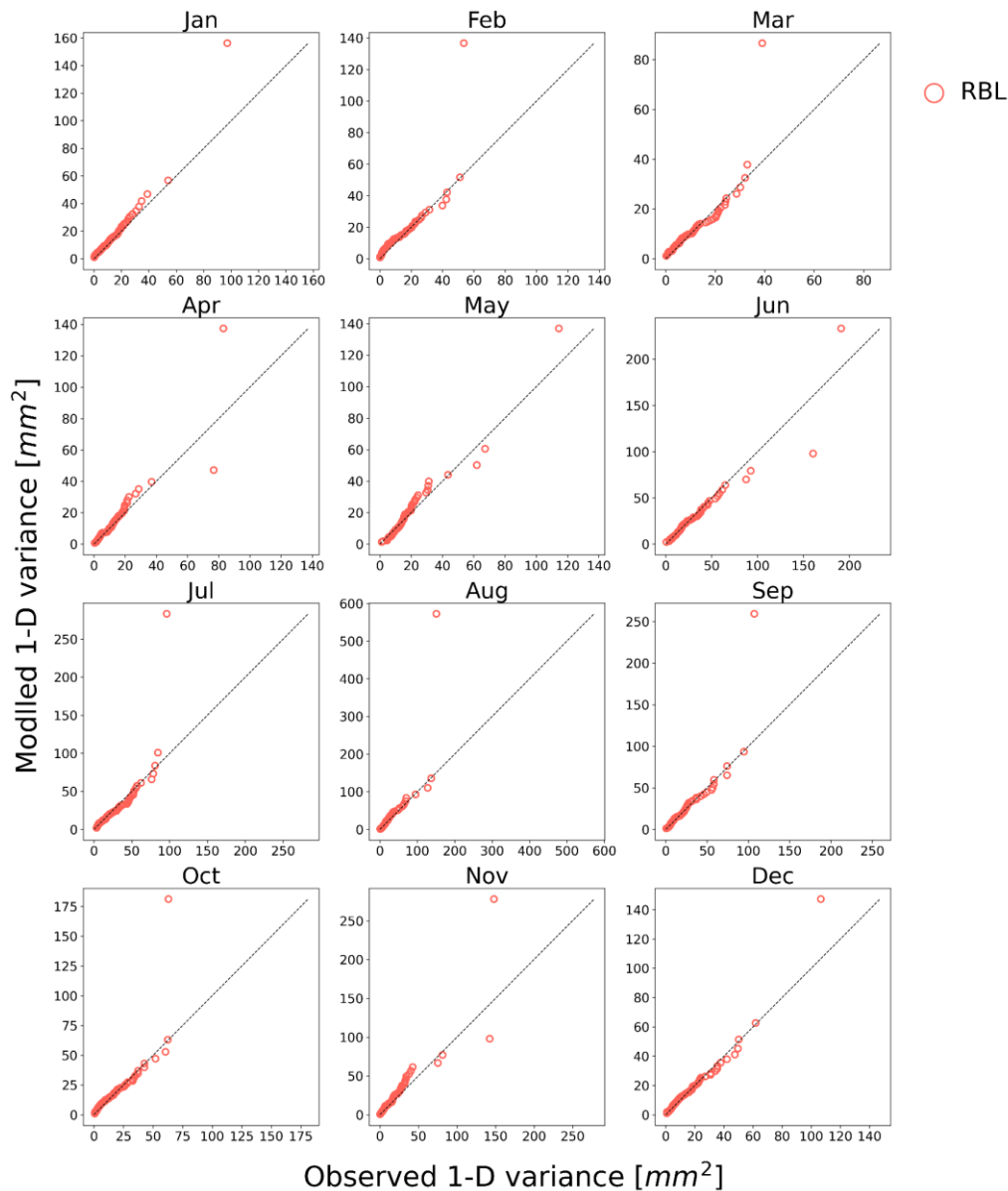
Figure R5. Relationship between observed and modelled monthly variances of daily rainfall amount. Red dots indicate the monthly variance of each year (69 samples for each month) with the diagonal line for reference (black dash).
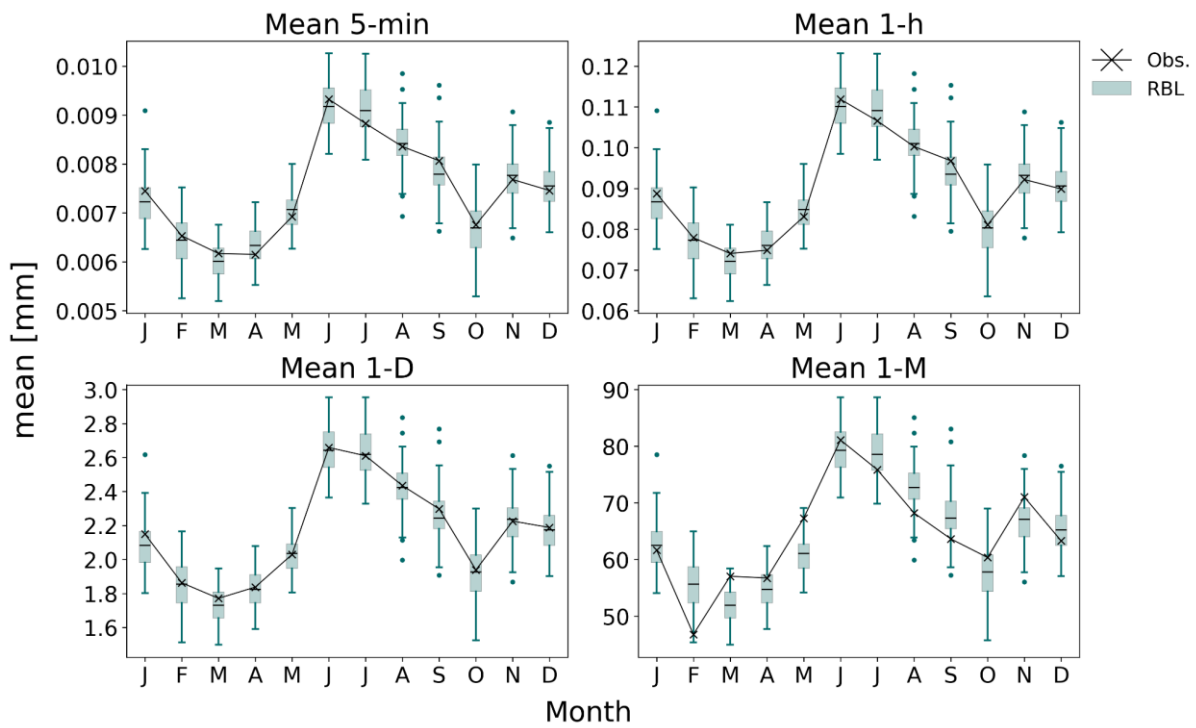
Figure R6. Mean by month at Bochum: comparison between RBL (boxes) and observation (crosses)
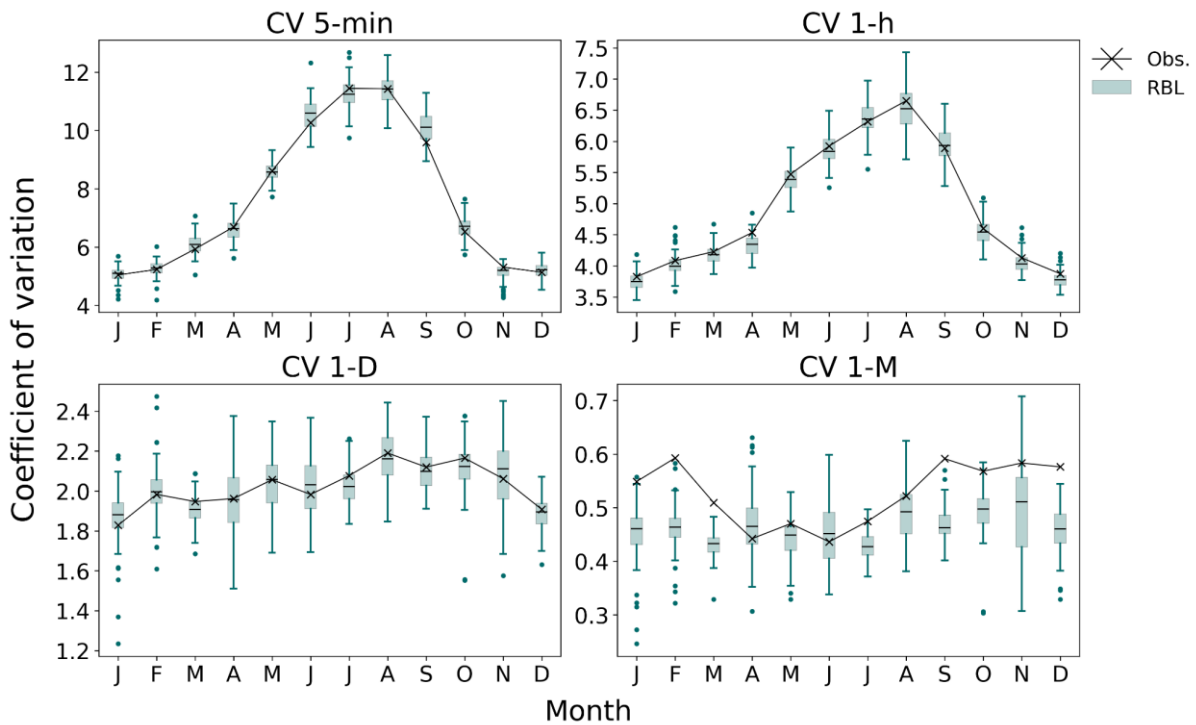
Figure R7. Coefficient of variation (CV) by month at Bochum: comparison between RBL (boxes) and observation (crosses)
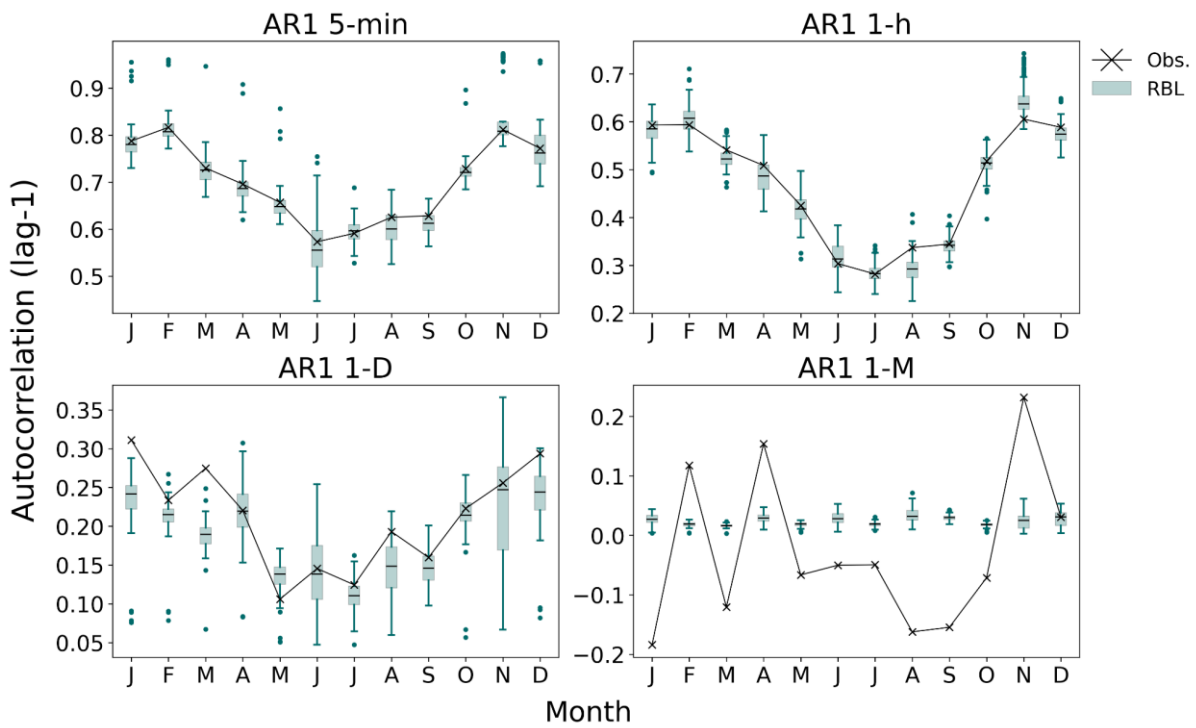
Figure R8. Lag-1 autocorrelation (AR1) by month at Bochum: comparison between RBL (boxes) and observation (crosses)
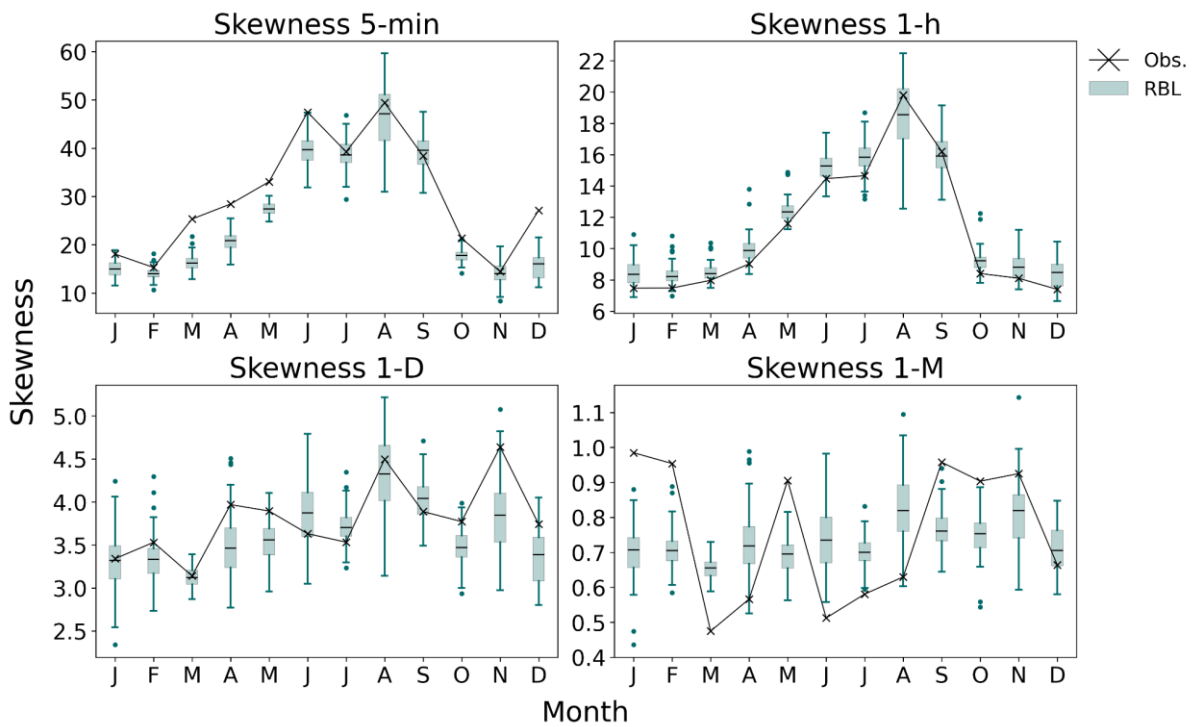
Figure R9. Skewness by month at Bochum: comparison between RBL (boxes) and observation (crosses)

5. The font size in Figure 2 is too small.

R/ Thank you for the comment. We will adjust the figure in the manuscript as Fig. R10 shows.
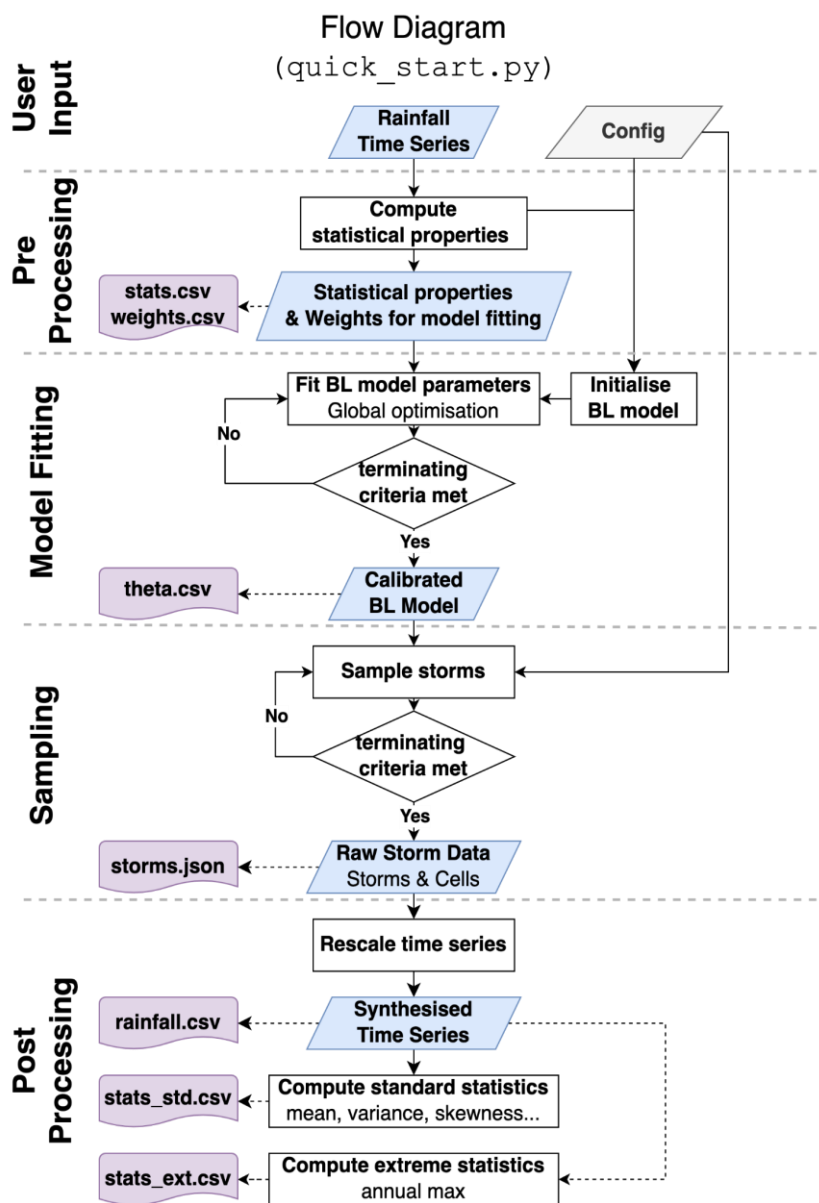
Figure R10. Workflow for generating synthetic rainfall time series using historical records with the pyBL package.

3. Referee #2

1. Line 110: As a reviewer with a personal interest in the practical application of this model, I have applied it across various fields. Based on that experience, although Equation 2 has a solid theoretical foundation (as cited in Kaczmarska et al., 2014, which states that statistics with greater interannual variability should be given less weight, and vice versa), it has shown problems such as underestimation of extreme values in real-world applications. The most significant reason, I speculate, is that interannual variability, as mentioned by Marani (2003) and Kim and Onof (2020), is a large-scale variability that the Poisson cluster rainfall model cannot replicate. This large-scale variability is related to extreme values that pose real-world problems. For example, if a time series shows high interannual variability in 1-hour variance, the year that contributed to this high variability is likely to contain extreme values. Therefore, I believe it would be more appropriate to apply greater weight to statistics with large interannual variability. Additionally, the magnitude of each MMM in this equation varies significantly. Thus, the weight factor should be adjusted to account for these relative differences, which could introduce confusion. Therefore, I recommend adopting a method of determining the weight factor based on the application field of the generated rainfall, as suggested by Kim and Olivera (2012). Moreover, I suggest using a normalized form of the function, such as Sigma(w_i \times (1 - f_k / f'_k)), instead of Equation 2. At the very least, users should have the option to choose such a method.

   R/ Thank you for the comment. We note the suggestion of having the possibility of choosing a different objective function, which we agree would be a good idea, so that will be included as an option in the software.

   The reviewer correctly mentions the issue of underestimation of large-scale variance which, indeed, has potential impacts upon the reproduction of extremes. However, the solution proposed by the reviewer conflicts with the theoretical result obtained by Jesus and Chandler (2011) which shows that statistics with lower variability should have more weight and that the weight should indeed be the inverse of the variance of the corresponding statistic. While we appreciate the reviewer's intention to obtain parameters that will achieve greater variability by giving more weight to statistics with greater variability, the problem is that the objective function only includes the mean value of the corresponding statistic. It therefore has no information indicating that this statistic has greater variability.

   Depending on the application, the underestimation of large-scale variability will or will not be a concern. If it is, then the hypothesis of storm independence would have to be revised, and the best way of doing that would be to use the added shuffling components of the model by Kim and Onof (2020). We include the coding of these components as our next task in terms of further research.

2. Section 2.4: The Bartlett-Lewis model is likely to produce different parameters corresponding to different local minima with each calibration attempt. However, there is no way to discern whether the variability of the parameters derived from the method presented here is due to parameter calibration or sampling. To demonstrate the validity of the method proposed in this section, you must show that calibration consistently produces the same parameters for the same rainfall statistics.

   R/ Thank you for your comment. The issue raised by the Reviewer can be addressed from two perspectives: (1) identifying globally optimal parameters and (2) ensuring the consistency of parameter estimation during each model calibration.

   Regarding the identification of optimal parameters, it is indeed challenging to verify whether the parameters found in each calibration are globally optimal. To address this, we employed a numerical strategy based on the basin-hopping algorithm, which reduces the likelihood of being

trapped in local optima and helps determine optimal parameters. According to a recent study by Baioletti et al. (2024), basin-hopping outperforms many other numerical algorithms, such as Differential Evolution and Particle Swarm Optimization, in terms of computational efficiency and solution accuracy. Specifically, in the default setting, our numerical solver runs basin-hopping iteratively 20 times for each model calibration (i.e., for a given set of rainfall statistics). The first iteration begins with a randomly assigned initial guess, and subsequent iterations use the solution from the previous basin-hopping as input to refine the optimal solution.

For the consistency of parameter estimation, we demonstrate that when the same random seed is used, the proposed numerical solver consistently produces the same parameters. However, when a different random seed is used, the solver may indeed yield different parameters. This variation is however due to the complexity of the Bartlett-Lewis model, where the parameters are inter-correlated, allowing different sets of parameters to produce similar rainfall statistics.

To further elaborate on the consistency issue, we conducted an experiment on the sensitivity of model calibration using pyBL under three scenarios: (a) fixed random seeds, (b) varying random seeds for initial guesses, and (c) the proposed bootstrapping method (see Sect. 2.4 of the original manuscript).

Figure R11 shows the results of model calibration using 69 years of rainfall records from Bochum (as used in the submitted manuscript). As can be seen, with fixed random seeds, the numerical solver consistently produces the same parameters. When varying random seeds, the solver yields nearly identical parameters in most months, except for July and September, where greater variability in some parameters is observed. However, when these parameters are used to compute rainfall properties, such as skewness at 5-minute and 1-day time scales (shown in Fig. R12), the variability in skewness estimates is minimal. This confirms the consistency of the solver and supports the statement that different parameter sets can produce similar rainfall statistics.

We also included results from model calibration using the proposed bootstrapping method. As shown in Fig. R12, the variability of rainfall properties derived from bootstrapping is consistently larger than that obtained from the other two scenarios. Interestingly, even for July, where the variation in some parameters from varying random seeds exceeds that from bootstrapping, the variability in rainfall properties is still smaller with varying random seeds. This suggests that, while the consistency of the numerical solver is important, the uncertainty in model calibration is largely driven by the estimation of observed rainfall properties rather than the solver's consistency.
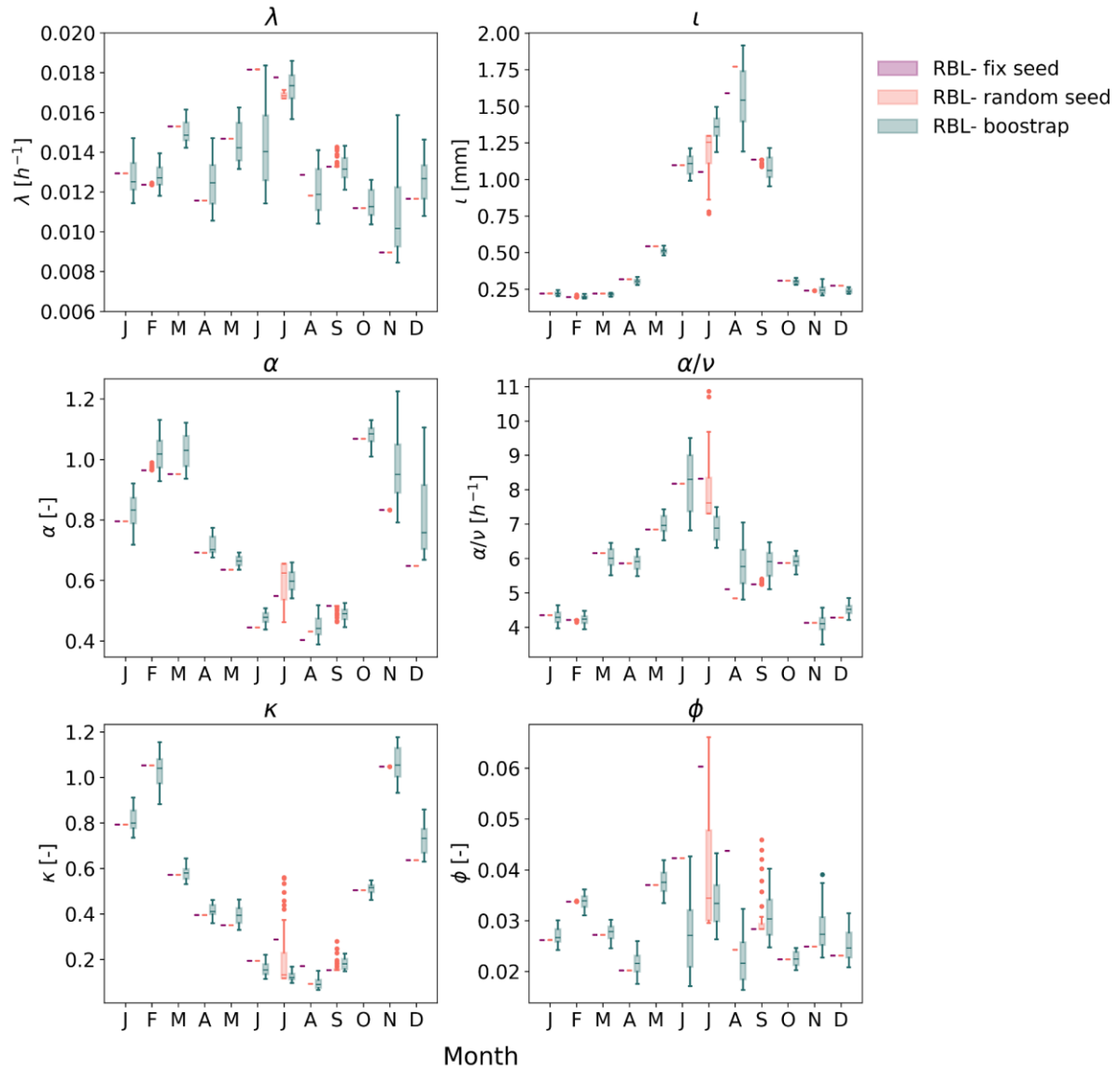
Figure R11. Variability of model parameters calibrated under three different scenarios: (1) fixed random seeds, (2) varying random seeds for initial guesses, and (3) the bootstrapping method.
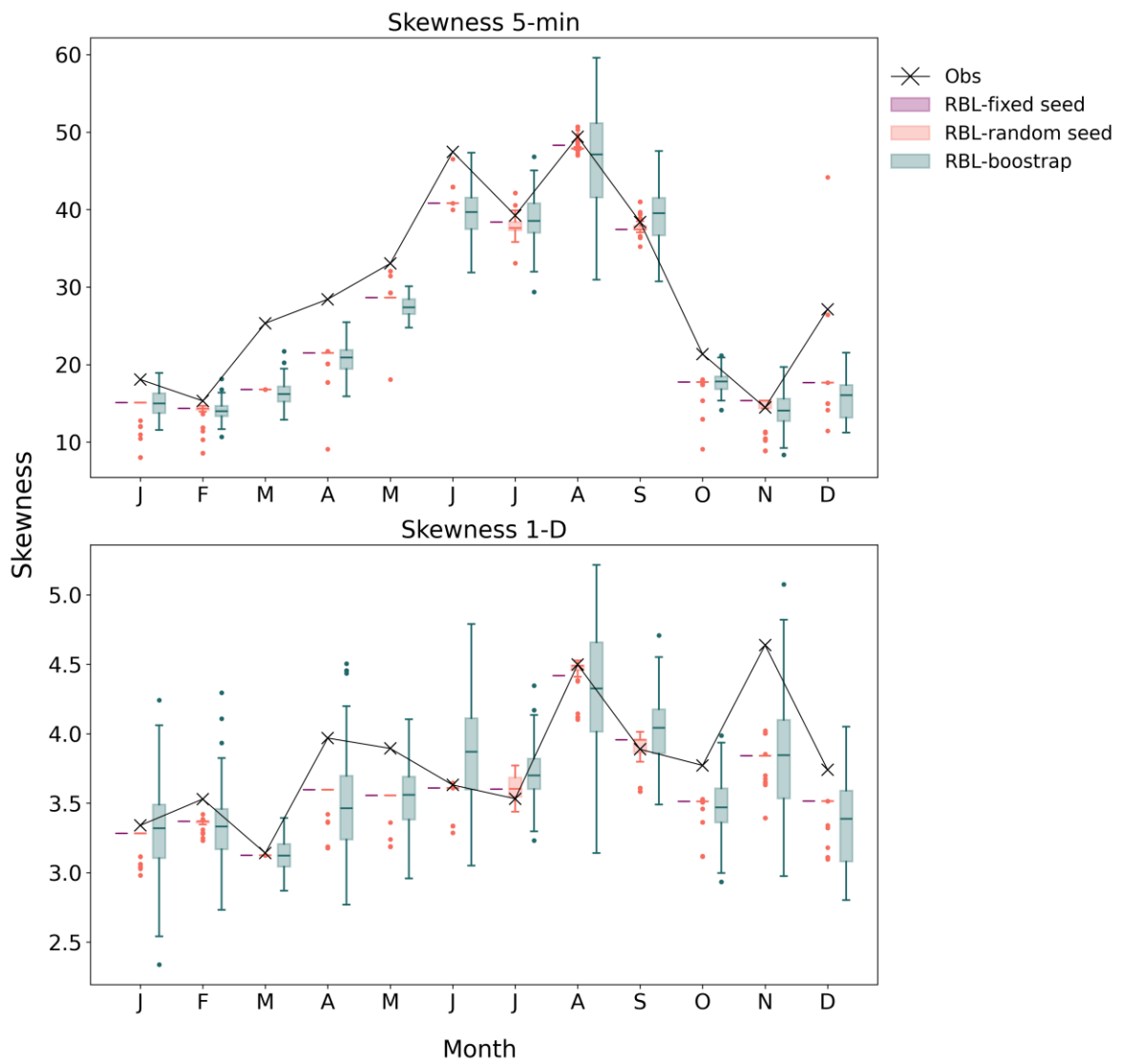
Figure R12. The corresponding 5-min and 1-day skewness estimates computed using the model parameters shown in Fig. R11.

References

Baioletti, M., Santucci, V., & Tomassini, M. (2024). A performance analysis of Basin hopping compared to established metaheuristics for global optimization. Journal of Global Optimization, 1-30.

Jesus, J. and Chandler, R. E.: Estimating functions and the generalized method of moments, Interface Focus, 1, 871-885, https://doi.org/10.1098/rsfs.2011.0057, 2011.

Kim, D. and Onof, C.: A stochastic rainfall model that can reproduce important rainfall properties across the timescales from several minutes to a decade, J. Hydrol., 589, 125150, 2020.

Wang, J., Anderson, B. T., & Salvucci, G. D. (2006). Stochastic modeling of daily summertime rainfall over the southwestern United States. Part I: Interannual variability. Journal of Hydrometeorology, 7(4), 739-754.