https://doi.org/10.5194/egusphere-2024-1918
Preprint. Discussion started: 22 July 2024

Modelling rainfall with a Bartlett-Lewis process: pyBL (v1.0.0), a Python software package and an application with short records

Chi-Ling Wei[1,*], Pei-Chun Chen[1,*], Chien-Yu Tseng[1,*], Ting-Yu Dai[1,3], Yun-Ting Ho[1], Ching-Chun Chou[1], Christian Onof[2], and Li-Pen Wang[1,2]

[1]National Taiwan University, Taipei 10617
[2]Imperial College London, London SW7 2AZ
[3]University of Texas at Austin, TX 78705
[*]These authors contributed equally to this work.

Correspondence: Li-Pen Wang (lpwang@ntu.edu.tw)

Abstract. The Bartlett-Lewis (BL) model is a stochastic framework for representing rainfall based upon Poisson cluster point process theory. This model has been used for over 30 years in the stochastic modelling of daily and hourly rainfall time series. Historically, the BL model was known to underestimate sub-daily rainfall extremes, but recent advancements have addressed this issue, making it a viable alternative to traditional rainfall frequency analysis methods, such as those based on annual maxima time series. Despite its potential, calibrating the BL model is a not a trivial task. The model's formulation is complex, and calibrating it involves a nonlinear optimisation process that can be numerically unstable, which has limited its broader application. To promote the use of the BL model and demonstrate its capabilities in modeling sub-hourly rainfall –both standard and extreme statistics– we have developed an open-source Python package called pyBL. This paper details the design of the BL model and summarises the key features of the pyBL package. It includes a brief explanation of how to use the package in selected user scenarios. In addition, we report upon scientific experiments that resemble real-world situations to showcase pyBL's ability to model sub-hourly rainfall extremes with short records and its flexibility in utilising records of various timescales and lengths.

1 Introduction

Stochastic rainfall modeling is an increasingly popular technique used by the water and weather risk industries. It can be employed to synthesise sufficiently long rainfall time series to support hydrological applications, such as runoff and flood modeling (Koutsoyiannis et al., 2003; Gires et al., 2012; Kim et al., 2017a; Park et al., 2019), or weather-related risk analysis, such as the quantification of the impact of climate change (Onof and Arnbjerg-Nielsen, 2009; Cross et al., 2019; Kim and Onof, 2020; Papalexiou, 2022; Ebers et al., 2023).

The Bartlett-Lewis (BL) rectangular pulse model is a type of stochastic model that represents rainfall using a Poisson cluster point process to define the arrival of rectangular pulses representing short duration constant intensity contributions to the cumulative rainfall. The model parameters are identified with standard statistical properties of rainfall data, such as mean, coefficient of variation, skewness, and autocovariance of the time-series of rainfall depths at various important scales, as well

https://doi.org/10.5194/egusphere-2024-1918
Preprint. Discussion started: 22 July 2024

the proportion of dry periods at those scales. Since the basic model type was published in 1987 (Rodriguez-Iturbe et al., 1987), several model variants have been developed (Rodriguez-Iturbe et al., 1988; Onof and Wheater, 1993; Onof et al., 2000; Kaczmarska et al., 2014; Onof and Wheater, 1994). In early versions, despite these models' ability to capture rainfall variability over a range of scales, two issues limiting their hydrological applicability have been commonly raised in the literature. First, BL models tend to underestimate hourly and sub-hourly extremes while overestimating daily extremes (Verhoest et al., 2010). Second, as identified by Marani (2003), BL models fail to reproduce rainfall variability for scales equal to or coarser than a few days. These limitations have been reported in many studies and pose significant challenges to the broader application of the BL model (see Verhoest et al. (2010) and references therein).

Recent advancements have addressed these issues, enhancing BL models' ability to preserve extreme statistics of rainfall at multiple timescales simultaneously (Cross et al., 2018; Onof and Wang, 2020; Kim and Onof, 2020). For example, Onof and Wang (2020) re-derived the analytical expressions for the rainfall depth moments in BL models and discovered that the parameter space is wider than was assumed in past studies. This relaxation of the parameter solution domain effectively improves BL models' capacity to preserve sub-hourly rainfall extremes. Furthermore, Kim and Onof (2020) extended Onof and Wang (2020)'s model by reorganising the temporal sequence of storms with a double shuffling algorithm.This enhances the model's ability to reproduce rainfall variability for scales ranging from a few days to a decade. In addition, preliminary studies suggest that the BL models are less sensitive to observational data length compared to existing rainfall frequency analysis methods that rely on, for example, annual maxima time series (Wang et al., 2020). Thus, it offers an alternative approach for modelling rainfall extremes when long datasets are not available.

In this work, we introduce an open-source Python package named pyBL. This package is implemented based on the state-of-the-art BL model developed in Onof and Wang (2020) since this version of BL model is capable not only of reproducing standard statistical properties but also of preserving extreme value statistics of rainfall across various timescales, from sub-hourly to daily. There are three main modules in the proposed pyBL package. These are the statistical properties calculation module, the model fitting (i.e. calibration) module, and the sampling (i.e. simulation) module. The statistical properties calculation module processes the input rainfall data and calculates its standard statistical properties at chosen timescales. The model fitting module calibrates the model parameters based upon the re-derived BL equations given in Onof and Wang (2020). To ensure efficient calibration and prevent the optimisation process from being trapped in local optima, a numerical solver employing Dual Annealing optimization together with Nelder-Mead local minimisation techniques is implemented. Finally, the sampling module generates stochastic rainfall time series at a specified timescale and for any required data length, based upon the fitted BL model.

The design of the BL is highly modularised, and the standard Comma Separated Values (CSV) format is used for file exchange between modules. Users can easily incorporate specific modules into their existing applications. In addition, a team comprising researchers from National Taiwan University and Imperial College London will consistently implement new advancements in BL models in the package, ensuring that users have access to the latest developments.

This paper is organised as follows. In Sect. 2, we provide detailed explanations of the formulation of the Bartlett-Lewis (BL) model. This includes a presentation of the model structure, as well as an overview of model calibration and sampling

https://doi.org/10.5194/egusphere-2024-1918
Preprint. Discussion started: 22 July 2024

processes. In addition, inspired by the bootstrapping method, we propose a novel approach to the estimation of model parameter uncertainty. Section 3 focuses on introducing the pyBL package. Specifically, we explain the workflow for using the package and summarise the pre-requisite Python packages needed to install pyBL. In Sect. 4, we use a case-study to demonstrate and evaluate the BL model's ability to generate realistic rainfall time-series. Two scenarios resembling data settings commonly found in many countries are designed and tested, showcasing the BL model's ability to produce rainfall extremes with short records. Finally, Section 5 summarises the key findings from this work and discusses potential further developments and applications of the proposed package.

2 Formulation of Bartlett-Lewis Rectangular model

This section will first explain the general structure of the BL model and the key adjustment in the most recent version of the model. Then, the processes of model calibration and of rainfall time series sampling will be detailed. Finally, a method, inspired by the well-known bootstrapping, is proposed here to estimate the model uncertainty.

2.1 Model structure

The model is constructed by using a point process to represent the arrival of rainfall cells. This process is a Poisson-cluster process which allows for the model to represent the observed clustering of such cells within longer rainfall events that are usually referred to as 'storms'. As seen in Fig. 1), the storms therefore arrive as a Poisson process at rate λ. The clustering mechanism is that of the Bartlett-Lewis process. it involves the generation of a second Poisson process of rate β starting at the storm inception. and of random duration, which we choose as exponentially distributed with parameter γ.

The rainfall is then added to this point process: each cell is represented by a random rectangular pulse. This means that the rainfall intensity produced by each cell is random (with a distribution characterised by its first three non-centred moments μ_x, μ_{x2} and μ_{x3}) but constant over the random duration of the cell. The latter is chosen as exponentially distributed with parameter η. Various cells will overlap, thereby producing a hyetograph that has a noisiness comparable to that of observed rainfall.

This describes the basic structure of a Bartlett-Lewis Rectangular Pulse model, which in its original version (Rodriguez-Iturbe et al., 1987), had all these distribution parameters define constant model parameters. It was however noted by Rodriguez-Iturbe et al. (1988), Onof and Wheater (1993) and others that a model in which the distribution parameters characterising the temporal structure of storms were allowed to vary from storm to storm would be preferable. This was achieved by randomising distribution parameter η: this becomes a random variable that is fixed for each storm but varies between storms and is Gamma distributed with shape parameter α and rate parameter ν, i.e. scale parameter $1/\nu$. To ensure that all the temporal statistical features of storms scale in the same way, the cell arrival rate and the storm duration parameter are chosen as: $\beta = \kappa\eta$; $\gamma = \phi\eta$. This defines a Randomised Bartlett-Lewis model which has been widely applied (e.g. Khaliq and Cunnane (1996); Verhoest et al. (1997); Kim et al. (2017b); Kim et al. (2017a); Kossieris et al. (2018)).

A more recent version of the Randomised Bartlett-Lewis model extends the randomisation to all the properties characterising the structure of the storm, i.e. also to the cell intensity parameters (Kaczmarska et al., 2014). So μ_x (and also μ_{x2} and μ_{x3}

https://doi.org/10.5194/egusphere-2024-1918
Preprint. Discussion started: 22 July 2024

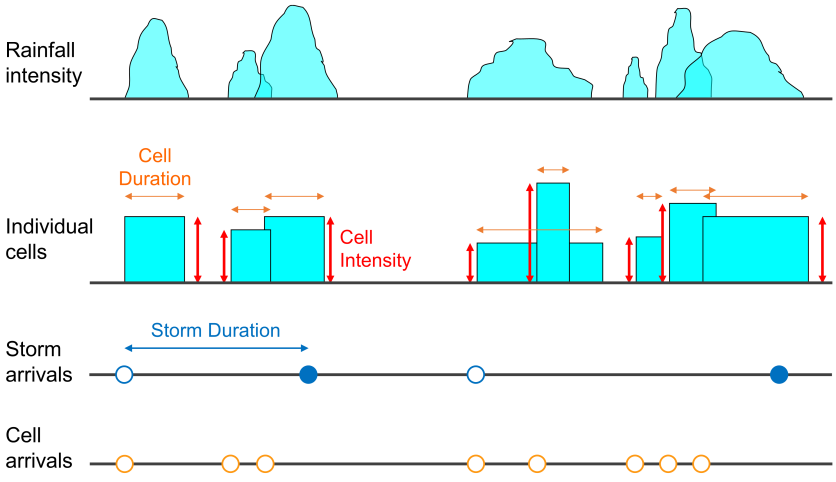

Figure 1. Conceptualisation of the Bartlett-Lewis Rectangular Pulse model (adapted from Figure 1 in Onof and Wang (2020)).

if another distribution than the exponential is chosen for the cell intensity). is now a random variable that takes on a fixed value throughout a storm but which varies between storms proportionally to η. This defines a new model parameter ι such that $\mu_x = \iota\eta$. It is this Randomised Bartlett-Lewis model (hereafter 'BL'), as further developed by Onof and Wang (2020), that is coded up in pyBL.

2.2 Model calibration

The BL model is a continuous-time model, i.e. it defines a continuous-time stochastic process $\{Y(t)\}_{t\in\mathbb{R}}$ where $Y(t)$ is the rainfall intensity at time t resulting from the superposition of the contributions of all the cells that are active at time t. Since rainfall data are generally available in discrete time, i.e. as a time-series, the BL model can only be calibrated by using the model's properties for rainfall aggregated to discrete time-scales (e.g. h hours). These are properties of the discrete random variable defined at time-step i by:

$$Y_i^{(h)} = \int_{(i-1)h}^{ih} Y(t)\,dt \tag{1}$$

It is not possible to obtain an analytical expression for the probability density function of $Y_i^{(h)}$ so that maximum likelihood estimation is not an option. What can be obtained are analytical expressions of the moments of the rainfall depth distribution (they are tractable up to order three) of this variable (Onof and Wang, 2020) in terms of the model parameters and the time-scale. Further, the probability of a dry interval at any time-scale h, i.e. $P(Y_i^{(h)}) = 0$ can also be estimated (ibid.). With these expressions, a Generalised Method of Moments (Onof et al., 2000) is used to obtain parameters that produce values of these

https://doi.org/10.5194/egusphere-2024-1918
Preprint. Discussion started: 22 July 2024

various properties that are as close as possible to their estimates from observed time-series of rainfall depths. This defines an optimisation problem which is the minimisation of the sum of the squares of the differences between analytical expressions of statistical model properties (the moments of the rainfall depth distribution or the proportion of dry periods at various time-scales) as a function of model parameters, and estimates of these properties from an observed time-series:

$$\sum_{\mathcal{M}\in\Omega} \omega(\mathcal{M}) \left\{ \mathcal{M} - \hat{\mathcal{M}} \right\}^2 \tag{2}$$

This minimisation is subject to certain feasibility constraints on the model parameters (Onof and Wang, 2020). In this expression, Ω is a set of statistical model properties, $\omega(\mathcal{M})$ a weight assigned to property $\mathcal{M}$ (whose analytical expression is a function of the model parameters and the time-scale) in the objective function, and $\hat{\mathcal{M}}$ is the estimate of that same property from the observed time-series. The choice of weights is discussed in Kaczmarska et al. (2014).

Ultimately, the choice of which statistics to include in Ω will depend upon which properties are deemed most important to reproduce, given the application for which the rainfall model is used. If the application does not obviously guide this choice, then Kaczmarska et al. (2014) recommend using the mean 1-h rainfall depth, and the coefficient of variation as well as the autocorrelation lag-1 and coefficient of skewness of rainfall depths at time-scales of 1-, 6- and 24-hours (and also at sub-hourly scales if the data are available at such scales). Inspired by the method proposed by Efstratiadis et al. (2002), the optimisation method first involves the Simulated Annealing algorithm. This is used to identify a promising region of parameter space. In the second phase, the method implements a downhill simplex Nelder-Mead algorithm in that region to identify the parameter set that minimises the objective function in equation (2).

2.3 Sampling

The sampling process of the BL model is fairly straightforward. It follows the concept of the BL model explained in Sect. 2.1. It involves two Poisson processes –one embedded in the other one– to model storm and rain cells, respectively. For each parameter set (usually corresponding to parameters for a given calendar month), the model first samples the number of storms based upon the specified sampling period (e.g. 10 years). For each storm event, the model then samples its arrival time and duration of activity (i.e. the time during which rain cells cam arrive), as well as the parameters associated to the distributions used to sample the properties of the embedded rain cells. Based upon the storm duration, the number of embedded rain cells is determined; and the arrival rates, durations and intensities of these rain cells are sampled. It is worth mentioning that, to ensure the consistency of the starting time between a given storm and the corresponding cells, the starting time of the first cell has to align with that of the storm.

2.4 Modelling uncertainty

The uncertainty ranges from the above sampling process represent the *sampling* uncertainty, i.e. that arising from the variability between various samples of given size (i.e. length of the simulated time-series). This is different from *model parameter* uncertainty, which is the uncertainty in the estimation of optimal model parameters. The sampling uncertainty can be decreased by

https://doi.org/10.5194/egusphere-2024-1918
Preprint. Discussion started: 22 July 2024

extending the length of the simulation: it converges asymptotically to *0* as this length goes to infinity. The model parameter uncertainty on the other hand is a feature of the calibration process.

To model the parameter uncertainty of a BL model is no trivial task due to the complexity of its model structure. Here, a method, inspired by the bootstrap method (or bootstrapping), is proposed. Assuming the full record length is N-years, one can randomly sample N years of data with replacement (that means data from any given year may be picked more than once) N_b times. Each N-year data sample is then used for BL model calibration, such that a total of N_b sets of BL parameters are obtained and used for sampling the corresponding time series with specified lengths. Based upon this bootstrapping process, the distribution of model parameters can be obtained and thus model parameter uncertainty can be quantified.

The proposed method can be further extended if one wants to model the uncertainty resulting from the available data length. Instead of sampling N-year data, one can randomly sample N_s years of data with replacement (where $N_s \leq N$), and proceed with the same process to model the corresponding uncertainty.

3 The pyBL package

As suggested by its name, the pyBL package is developed using the Python language. Python was chosen due to its open-source environment, extensive support libraries, and low learning curve, which facilitates the expansion of the pyBL user community. Here, we outline the complete workflow of running pyBL, starting from computing statistical properties from input rainfall records and fitting model parameters to sampling rainfall time series at a given timescale and length. In addition, we provide instructions for the usage of the package. The source code, example scripts and test data for pyBL v1.1. can be downloaded at https://doi.org/10.5281/zenodo.12605935, and please stay tuned with the further development at pyBL's github repository: https://github.com/NTU-CompHydroMet-Lab/pyBL.

3.1 Workflow

The workflow for using pyBL to build a BL model, sample rainfall time series, and calculate the associated statistics is illustrated in Figure 2. As shown, it comprises five main steps. These are:

1. **User Input**
   Users must provide two input files. The first file includes rainfall time series records, either as a 1D array of rainfall intensity data or a 2D array including rainfall intensity data and associated timestamps. The second file is the model configuration file (Config), which allows users to control the entire modelling process.

2. **Pre Processing**
   This step calculates the required statistical properties from the input rainfall records and estimates the associated 'weights' for each property needed for BL model fitting (as mentioned in Sect. 2.2). Users can choose to export the calculation results to a CSV (Comma Separated Value) file for future use.

https://doi.org/10.5194/egusphere-2024-1918
Preprint. Discussion started: 22 July 2024

3. **Model Fitting**

   This step derives the BL model parameters based on the statistical properties and weights. As suggested by Onof and Wang (2020), a two-stage numerical minimisation strategy is used as the default solver. This strategy reduces the chance of the solution being trapped in a local optimum. The combination of dual-annealing and basin-hopping methods is implemented in this version of the package. The fitting process terminates when it reaches the error threshold or exceeds the iteration limit. Users can choose to export the fitted parameters to a CSV file for future use.

4. **Sampling**

   This step uses the fitted BL model to sample rainfall time series that preserves the statistical properties observed in the input records. The sampling process terminates when the required number of valid storms and cells are obtained. Users can choose to export the sampled storm and cell data to a JSON file for reference.

5. **Post Processing**

   This step calculates standard and extreme statistics from the sampled time series to support model evaluation. Users can choose to export these statistical results to CSV files. In addition, users can choose to convert the raw storm and cell data into rainfall time series at a specified temporal resolution (e.g., 5-min or 1-h) and export it to a CSV file for subsequent hydrological applications.

A Python notebook script, named `quick_start.ipynb`, is provided, accompanying with the package download (https://doi.org/10.5281/zenodo.12605935). This script enables users to run the entire workflow detailed above to stochastically generate rainfall time series with the test data. Notably, the package is designed to be highly modular, meaning that each of the above steps can be executed independently if the corresponding input is provided. For more examples, please refer to pyBL's GitHub repository (https://github.com/NTU-CompHydroMet-Lab).

3.2 External libraries and package installation

The implementation of pyBL depends on a number of external libraries. A list of these dependencies is summarised in Table 1. Amongst these libraries, Numpy and Pandas are used mainly for computing statistical properties from the input rainfall records. SciPy is used for BL model fitting. A robust numerical solver built on SciPy optimisers is used to obtain parameters for the pyBL model. Numba accelerates calculations using compiled C/C++ code, parallelisation, and CUDA kernels. Finally, matplotlib is an optional library for visualisation.

To install the pyBL package, it is recommended to use `pip`, which automatically resolves all dependencies and installs the pyBL package, simplifying the installation process for users.

https://doi.org/10.5194/egusphere-2024-1918
Preprint. Discussion started: 22 July 2024

Table 1. Summary of external libraries used by the pyBL package.

Library	Website	Reference	Description
Numpy	https://numpy.org/	Harris et al. (2020)	Mathmatic and Datastructures
Pandas	https://pandas.pydata.org/	Wes McKinney (2010)	
SciPy	https://www.scipy.org/	Virtanen et al. (2020)	
numba	https://numba.pydata.org/	...	Performance Optimization
matplotlib	https://matplotlib.org/	...	Visualization

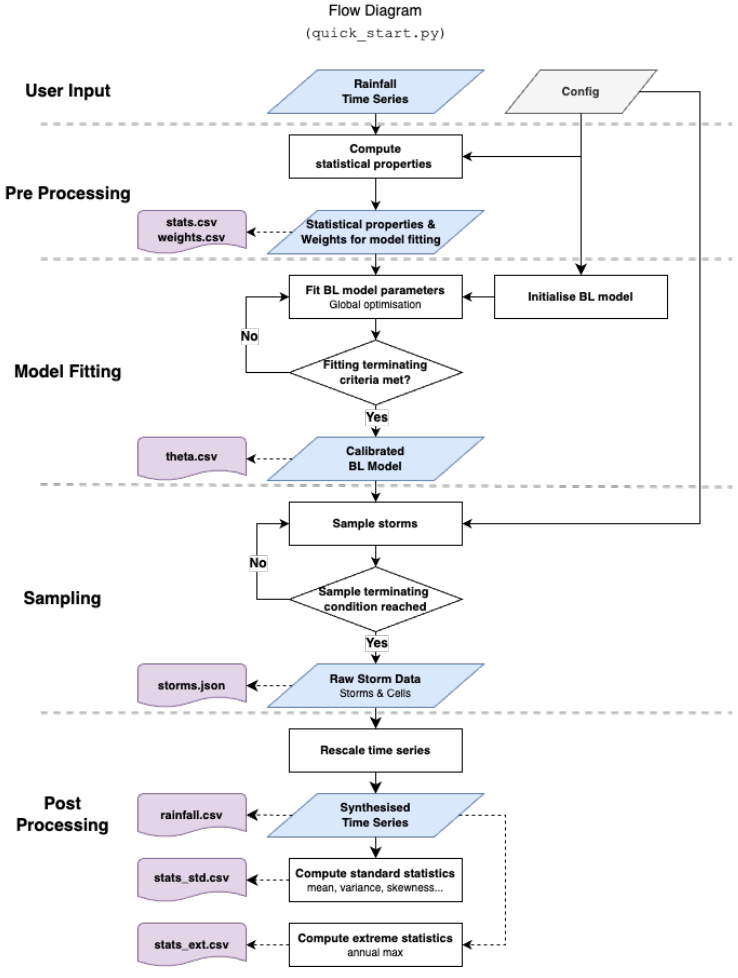

Figure 2. Workflow for generating synthetic rainfall time series using historical records with the pyBL package.

https://doi.org/10.5194/egusphere-2024-1918
Preprint. Discussion started: 22 July 2024

4 Case study

In this section, we conduct experiments based on two scenarios that resemble real-world settings in many countries. In both scenarios, the extreme value performance of the BL model is compared with that of a conventional rainfall frequency analysis approach based on the annual maxima series and the generalised extreme value distribution (hereafter, AM analysis). These two scenarios will demonstrate that the BL model can not only serve as an alternative to conventional frequency analysis methods but also provides the flexibility to combine rainfall records at different temporal resolutions and recording periods.

4.1 Experimental Design

4.1.1 Data sets

Rain gauge data at a 5-min time-scale from a rain gauge in Bochum, Germany, is used to demonstrate the application of the pyBL package in this paper. The Bochum rainfall records used here span 69 years, from January 1931 to December 1999.

4.1.2 Scenarios

Two scenarios are designed here to resemble two real-world settings that can be found in many countries or regions in the world. These two scenarios will be compared with a baseline, which represents a near ideal setting where long (over 30 years) sub-hourly rainfall data are available.

These scenarios are:

- Baseline (BS) resembles a near ideal setting where long (over 30 years) sub-hourly rainfall data is available. Five-minute records over full recording periods from two gauges are used. This baseline scenario is to demonstrate that the BL model can be used as an alternative to the conventional AM analysis in modelling rainfall extremes.
- Scenario 1 (SC1) resembles a widely-seen setting in many regions where long-term rainfall records (30+ years) are not available. In this scenario, short records with different lengths (5, 10, 15 and 20 years) are used. The uncertainty ranges resulting from the BL model at various data lengths are compared with those from the traditional AM analysis. This scenario will enable us to showcase the greater ability of the BL model to preserve extreme statistics with short records, when compared with the traditional AM analysis.
- Scenario 2 (SC2) resembles another relatively realistic setting seen in some countries. That is, sub-hourly rainfall records are available only in a shorter period (e.g. for 5-10 years), whilst hourly or coarser rainfall records are available for a longer period (i.e 20+ years). In this scenario, we would like to demonstrate that the BL model provides a flexible framework enabling the combination of rainfall records at different timescales with different data lengths.

https://doi.org/10.5194/egusphere-2024-1918
Preprint. Discussion started: 22 July 2024

4.1.3 Evaluation methods and metrics

The focus of the evaluation lies in the impact of modelling sub-hourly extreme statistics with short records. In scenario 1 (SC1), rainfall records at all timescales under consideration (i.e. 5-min, 1-, 6- and 24-h) are randomly sampled with replacement for N years (with $N = 5, 10, 15$ and 20), using the bootstrapping-inspired method described in Sect. 2.4 (where 100 bootstrapping iterations, i.e. $N_b = 100$, are conducted). In scenario 2 (SC2), we assume that full records are available for hourly or coarser timescales (i.e. 1-, 6- and 24-h), and the statistical properties obtained from these records are combined with those derived from the N-year short records at the 5-min timescale for the simulation. Here, the last and the earliest N-year 5-min records from the original dataset are used (with $N = 5$ and 20, respectively). The setup allows us to demonstrate how variations and uncertainties in short sub-hourly rainfall records affect the modeling process and impact the resulting extreme statistics.

In both scenarios, we compare the return levels at T_r return periods (with $T_r = 20, 50$ and 100) with those derived from the base scenario (BS). In addition, for SC1, two widely-used, non-dimensional, normalised evaluation metrics to quantify the estimation error of quantiles at T_r return periods. The first metric is to quantify the multiplicative bias between the estimated quantile resulting from the i-th bootstrapping iteration x_{est}^{i,T_r} and the corresponding reference $x_{ref}^{T_r}$, which is termed:

$$\mathrm{B}^{i,T_r} = \frac{x_{est}^{i,T_r}}{x_{ref}^{T_r}} \tag{3}$$

where bias (B^{i,T_r}) ranges from 0 to $+\infty$, with 1 indicating the perfect match. The second metric assesses the (relative) error between the estimated quantile and the corresponding reference at given return periods. This is done using the fractional standard error (FSE), which is termed:

$$\mathrm{FSE}^{T_r} = \frac{\sqrt{\frac{1}{N_b}\sum_{i=1}^{N_b}(x_{est}^{i,T_r} - x_{ref}^{T_r})^2}}{x_{ref}^{T_r}} \tag{4}$$

where FSE ranges from 0 to $+\infty$, with 0 representing no error.

4.2 Results and discussion

4.2.1 Modelling rainfall with short records (SC1)

Here, we first focus on assessing the hourly and sub-hourly rainfall extremes with various data lengths. In Fig. 3, return levels at 5-min (left column of plots) and 1-h (right column of plots) timescales obtained from the BL model (green boxplots, denoted RBL) and the AM analysis (yellow boxplots, denoted AM (GEV)) are given, respectively. From top to bottom rows, the return level estimates for 20-, 50- and 100-year return periods are given. Each plot presents the estimates resulting from rainfall records with 5-, 10-, 15-, 20-year data lengths and full records (FR, i.e., 69 years).

As seen in the plots, the median estimates of the return levels from the BL model generally align with those from the AM analysis. Nonetheless, the uncertainty ranges are significantly different, with those from the BL model being much smaller

https://doi.org/10.5194/egusphere-2024-1918
Preprint. Discussion started: 22 July 2024

than those from the AM analysis. The difference is particularly evident as the records are short and the targeted return periods are high. Moreover, the BL model results in far fewer outliers compared to the AM analysis. Notably, for the 100-year 5-min return levels, the uncertainty range from the BL model fitted with 20-year records is similar to that from the AM analysis fitted with full records (69 years). Similar observations can be made for the 100-year 1-h return levels (see Fig.3 (f)).

The quality of the return level estimation and the corresponding uncertainty can be further quantified using the bias (B) and the FSE measures. Figures 4 (a) and (c) compare the (multiplicative) biases of the 5-min and 1-h return levels, resulting from the BL model and the AM analysis against the reference return levels (i.e. those estimated from the full records). The full quantile intervals are used to represent uncertainty ranges.

From the median estimates (blue solid lines), it is evident that the BL model tends to slightly underestimate the return levels, particularly when the rainfall records are shorter than 15-20 years. This underestimation is consistent across both 5-min and 1-h timescales and for all return periods examined. However, the uncertainty ranges are adequate to cover the unbiased line (B = 1.0, dark dashed lines). The median estimates from the AM analysis (yellow solid lines) exhibit a different behavior. While the AM analysis appears to provide more unbiased estimates at relatively low return periods ($T_R = 20$) compared to the BL model, a significant overestimation is observed at higher return periods, especially with records shorter than 15 years. In addition, the AM analysis results in much larger uncertainty ranges than the BL model when records are shorter than 20 years, with the size of these ranges increasing drastically as return periods become higher. Unlike the AM analysis, the uncertainty ranges resulting from the BL model remain relatively stable across all return periods examined.

This difference in uncertainty ranges is further highlighted in the FSE estimates (see Fig. 4 (b) and (d)). The FSE estimates from the BL model remain consistently similar as return periods increase, whereas those from the AM analysis increase significantly, particularly when records are shorter than 15-20 years.

To summarise results of the SC1 experiment, the BL model shows itself able to preserve extreme rainfall statistics at 5-min and 1-h timescales even though extreme rainfall records are not used for model calibration. Moreover, its estimation uncertainty of the extreme statistics is significantly less sensitive to record lengths, as compared to the traditional AM analysis. This robustness can be attributed to the fact that the BL model works with standard statistics calculated from the entire rainfall records, rather than just the annual maximum data, which represents a small subset of the records. Consequently, the BL model proves to be a robust alternative to the AM analysis, particularly when only short records are available.

4.2.2 Modelling rainfall with short 'sub-hourly' records

Following the SC1 with short record analysis, we now shift our focus to a SC2 setting where hourly (or coarser-resolution) rainfall data is available for 69 years, whilst sub-hourly rainfall data is available for only 5 years (the most recent 5-year period: 1995-1999). Here, we compare the 5-min rainfall extremes derived from the traditional AM analysis and the BL model. The associated uncertainty is calculated using the bootstrapping method detailed in Sect. 2.4, with the full quantile intervals representing the estimation uncertainty ranges.

As illustrated in Fig. 5, we first observe that the dark dashed line, representing median estimates of 5-min extremes from the AM analysis (denoted AM (GEV)-5y, 5min), is nearly horizontal (with negligible increase) after the 10-year return period.

https://doi.org/10.5194/egusphere-2024-1918
Preprint. Discussion started: 22 July 2024

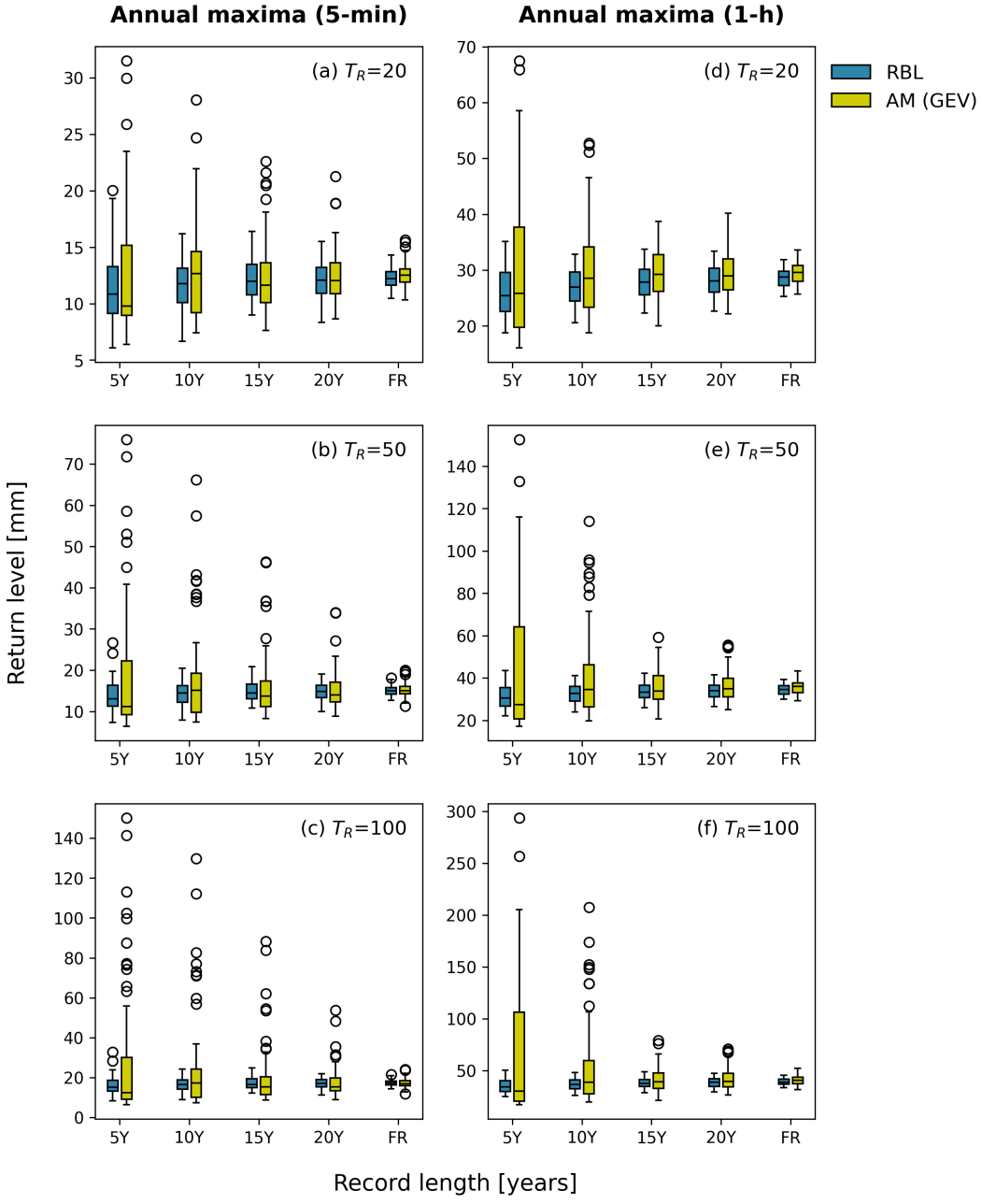

Figure 3. Boxplots of 5-min and 1-h return levels from RBL (blue boxes) and AM analysis (yellow boxes) for different record lengths (5, 10, 15, 20 years, and full records) at 20-, 50-, and 100-year return periods (top to bottom).

https://doi.org/10.5194/egusphere-2024-1918
Preprint. Discussion started: 22 July 2024

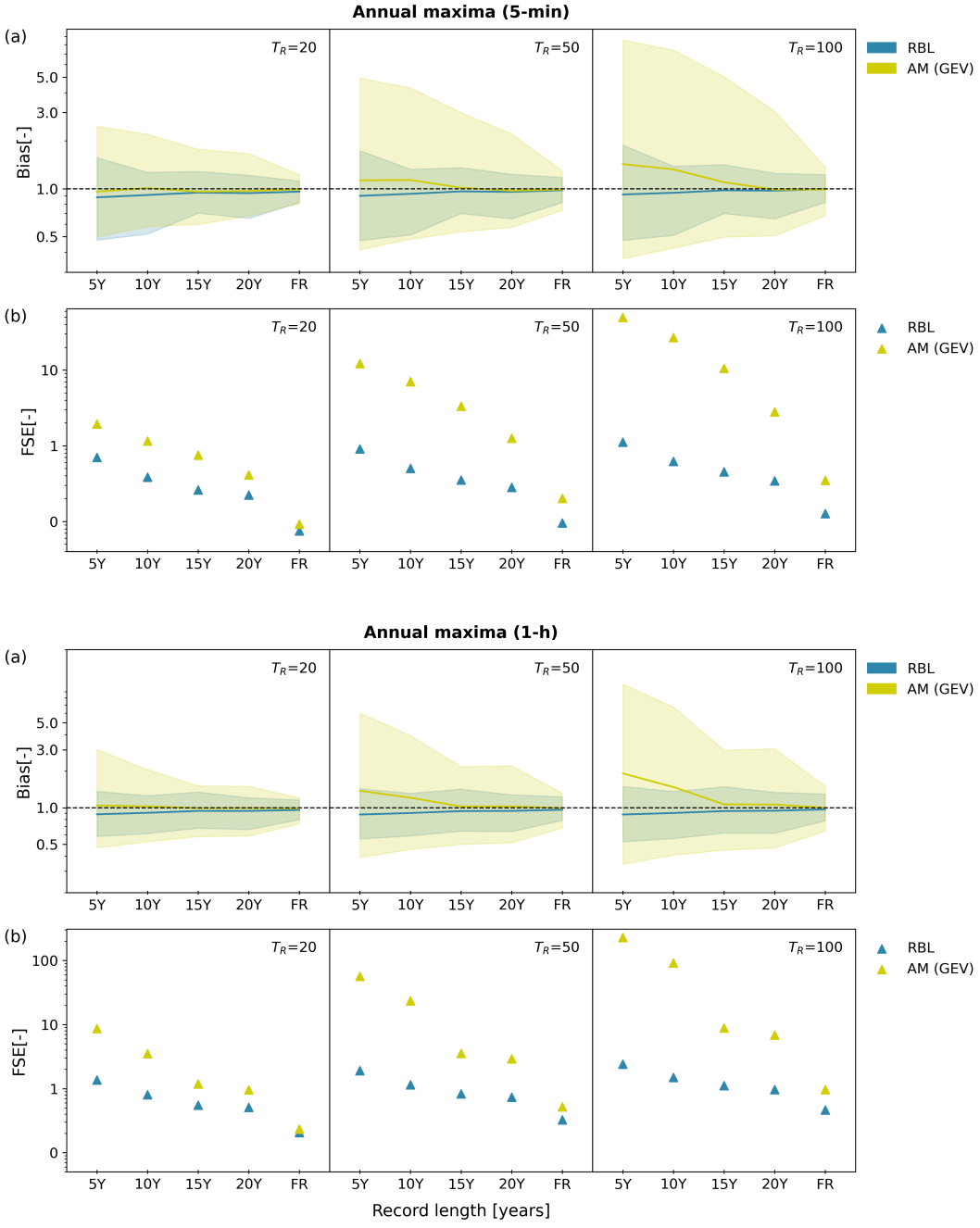

Figure 4. Comparisons of 20-, 50- and 100-year return levels at 5-min (a-b) and 1-h (c-d) timescales derived from the BL model (denoted RBL) and AM analysis (denoted AM (GEV)) using 5-, 10-, 15- and 20-year short records, as well as full records (69 years). Multiplicative biases and Fractional standard errors (FSEs) of each member of 100 bootstrapping iterations are calculated against the reference values, and the full quantile intervals are presented.

https://doi.org/10.5194/egusphere-2024-1918
Preprint. Discussion started: 22 July 2024

In addition, the associated uncertainty range grows exponentially after the same return period. This is likely caused by fitting the GEV distribution with a very small dataset, which is numerically challenging. The BL model effectively addresses this issue and results in more reasonable estimates of 5-min extremes (see yellow line, denoted RBL-5y, 5-min). Moreover, the uncertainly interval is significantly reduced compared to that from the AM analysis. This result is consistent with that in SC1.

We then further calibrate the BL model using the aforementioned SC2 setting, i.e., using 5-year 5-min data and the 69-year 1-h data (see blue line, denoted RBL-69y, 1-h + 5y, 5-min). This combination results in similar median estimates to those from the RBL model with 5 years of 5-min data only but leads to a further reduction in the uncertainty range. This highlights the capacity of the BL model to integrate data at different timescales and lengths, adding value to short sub-hourly rainfall records.

The benefit of integrating data at different timescales and lengths can be further explored from another perspective. It has been observed in the literature that the impact of climate change on rainfall patterns varies across different timescales. Specifically, many studies have noted that increases in temperature lead to more pronounced variations in rainfall extremes at finer timescales (e.g., sub-hourly or hourly) compared to coarser timescales (e.g., multiple hourly or daily) (Chan et al., 2016; Fowler et al., 2021; Huang et al., 2022; Cannon et al., 2024). In other words, the level of statistical non-stationarity is different for rainfall at different timescales –generally higher for sub-hourly rainfall and lower for hourly or coarser-scale rainfall. Given that the data used to construct a BL model is assumed to be statistically stationary, it makes sense to calibrate BL models using data at different timescales and lengths to better comply with the stationary assumption and to more accurately represent underlying rainfall features.

To further explore this multi-time-scale perspective, we continue with the SC2 setting but introduce some minor changes. Specifically, we combine 1-h full records with 5-min data from different periods –the earliest or the most recent 5/20 years– to reflect the impact on fine-scale rainfall extremes, caused by the non-stationarity in the sub-hourly rainfall time series. The reason for choosing these two periods lies in the variations in 5-min rainfall extremes observed between them. As illustrated in Fig. 6, the 69-year 5-min annual maxima from 1931 to 1999 is presented. As highlighted in the plot, there is a notable difference in annual maxima between the earliest (1931-1935, blue shading) and the most recent (1995-1999, yellow shading) 5 years, where the average difference is over 2 mm. This difference is, however, largely reduced if we extend the average from 5 to 20 years.

We find that this 'dynamics' observed for the 5-min rainfall extremes, effectively propagate through the BL modelling. As shown in Fig. 7 (upper left), the BL model calibrated with 69-year 1-h data and the earliest 5-year 5-min data (see light blue boxes) results in much higher 5-min extremes compared to those calibrated with 69-year 1-h data and the most recent 5-year 5-min data (see blue boxes). This relative difference observed in the sampled 5-min return levels aligns with that presented in Fig. 6, where the average of the 5-min annual maxima over the earliest 5 years is much higher than that over the most recent 5 years. Please note that, in this experiment, all results presented come from the BL modelling, so the variations in uncertainty ranges caused by different models is not our main evaluation focus. Thus, the bootstrapping method is not conducted here for model uncertainty estimation. The uncertainty ranges presented here come from the sampling process.

We also observe that the difference in annual maxima is largely reduced at the 1-h timescale (see Fig. 7 (lower left)), as well as when the available 5-min records increase from 5 to 20 years (see Fig. 7 (upper right)). The former, together with the result

https://doi.org/10.5194/egusphere-2024-1918
Preprint. Discussion started: 22 July 2024

presented in Fig. 7 (upper left), demonstrates that the BL model not only reflects the variations in sub-hourly rainfall extremes but also effectively maintains the stationarity in hourly rainfall extremes. The latter however showcases that the variations in sub-hourly rainfall extremes may be smoothed out when a longer time period of data is used.

To summarise the results of the SC2 experiment, we demonstrate the flexibility of the BL model in working with rainfall data at different timescales and lengths, highlighting the corresponding benefits. Specifically, the BL model can effectively reduce the estimation uncertainty of sub-hourly rainfall extreme calculations by integrating long hourly data with short sub-hourly records. In addition, it offers a straightforward approach to account for the varying impacts of climate dynamics on rainfall properties across different timescales.

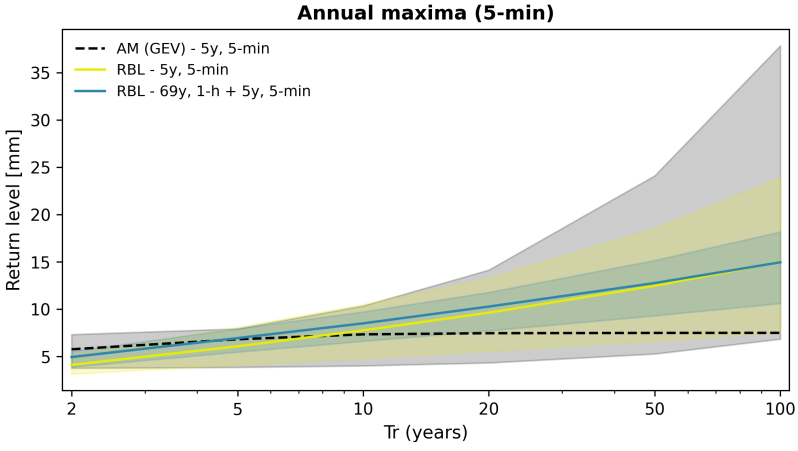

Figure 5. Comparison of 5-min return levels for 2- to 100-year return periods, with estimation uncertainty. Traditional AM analysis (dark dashed line, grey shading), BL model using 5 years of 5-min data (yellow solid line, shading), and BL model calibrated with 5 years of 5-min data and 69 years of 1-h records (blue solid line, shading).

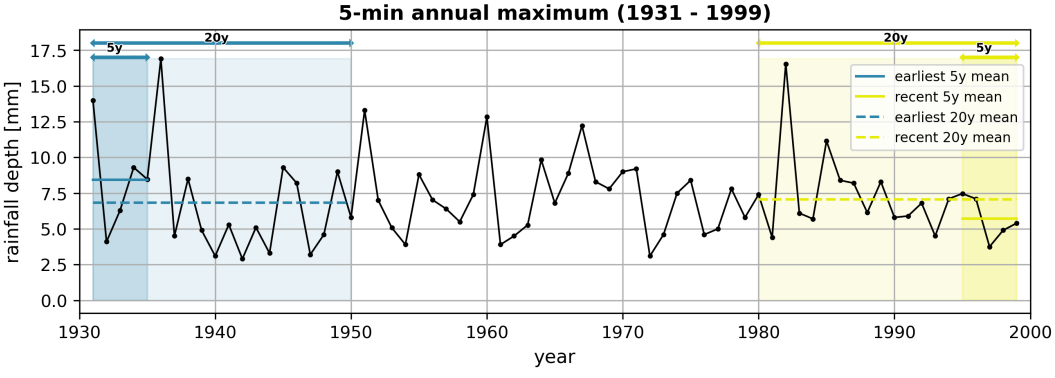

Figure 6. Five-minute annual maximum rainfall records at Bochum (1931-1999) are shown (dark solid line). Average values for the earliest 5 and 20 years (blue lines and shading) and most recent 5 and 20 years (yellow lines and shading) highlight significant differences in extreme statistics over these periods.

https://doi.org/10.5194/egusphere-2024-1918
Preprint. Discussion started: 22 July 2024

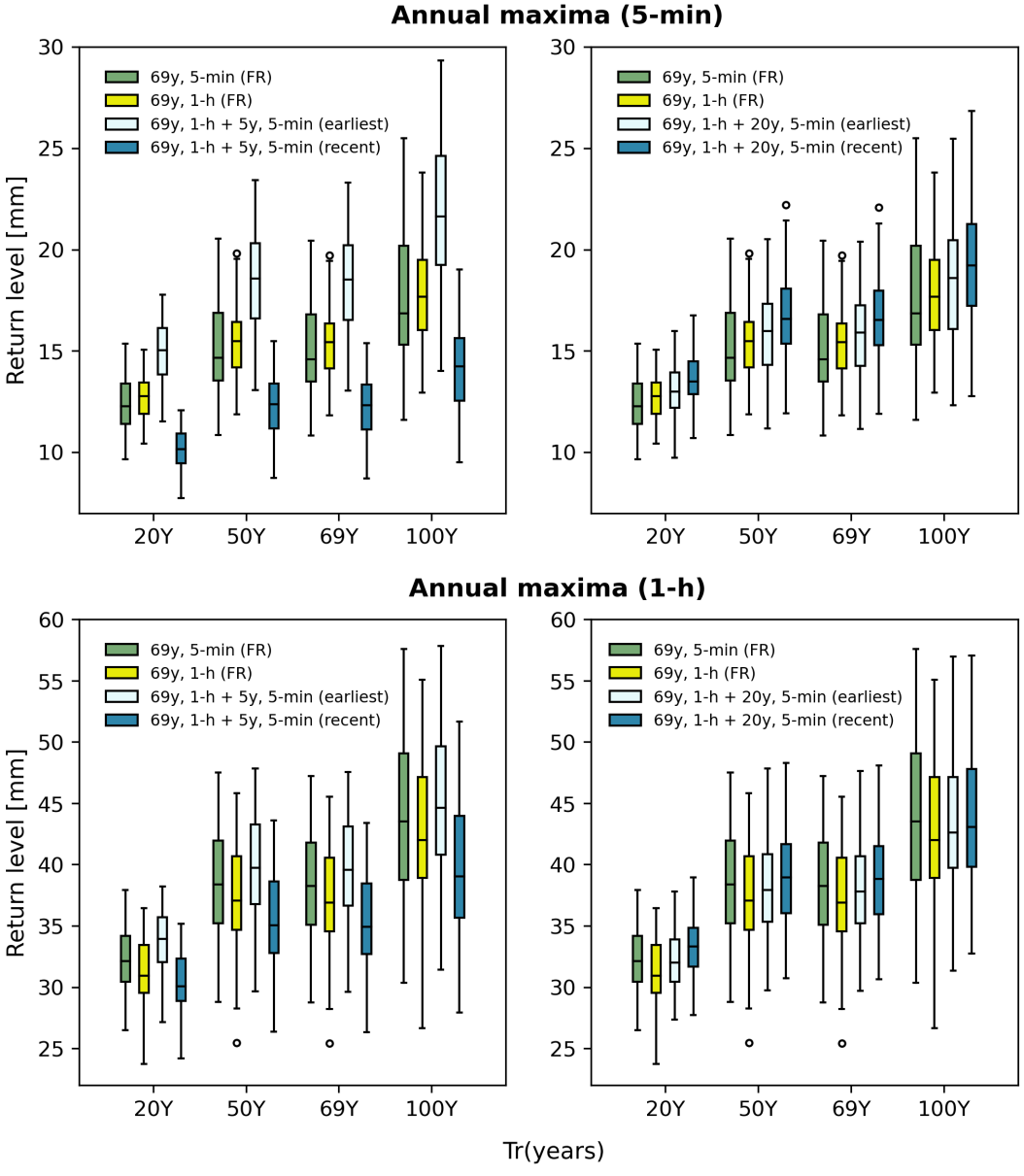

Figure 7. Boxplots of 5-min and 1-h return levels from the BL model calibrated with various data settings, including 69-year 5-min data (full records, green boxes: 69y, 5-min (FR)), 69-year 1-h data without 5-min data (yellow boxes: 69y, 1-h (FR)), 69-year 1-h data with the earliest 5/20 years of 5-min data (light blue boxes: 69y, 1-h + 5/20y, 5-min (earliest)), and 69-year 1-h data with the most recent 5/20 years of 5-min data (blue boxes: 69y, 1-h + 5/20y, 5-min (recent)). Return levels for 20-, 50-, 69- (full records), and 100-year periods are shown.

https://doi.org/10.5194/egusphere-2024-1918
Preprint. Discussion started: 22 July 2024

5 Conclusions

This work introduces an open-source Python package named pyBL for generating rainfall time series using randomised Bartlett-Lewis rectangular pulse models (BL models). Historically, BL models have been effective in producing rainfall time series with realistic statistical properties across various timescales. However, they have also been known to underestimate rainfall extremes at sub-daily timescales. Recent advancements have addressed this issue, enabling the BL model to preserve both standard and extreme statistics at sub-hourly and hourly timescales while maintaining its ability to generate realistic rainfall features (Kaczmarska et al., 2014; Onof and Wang, 2020).

Implementing the BL model is a challenging task due to its complex formulation and the nonlinear optimisation required to derive its parameters. To overcome these challenges and promote the widespread use of the BL model, we developed pyBL. This paper provides explanations of its structure and installation instructions. In addition, we explored a potential application of the BL model with two scenarios that mimic real-world situations where only short sub-hourly records are available.

In the first scenario (SC1), we demonstrated that the BL model can produce robust sub-hourly and hourly rainfall extremes with short records. Compared to conventional annual maximum analysis, the BL model achieves similar consistency in estimating sub-hourly rainfall extremes with only half the record length (or even shorter).

In the second scenario (SC2), we showcased the BL model's flexibility in integrating rainfall records at different timescales and lengths. We demonstrated that the estimation uncertainty of sub-hourly rainfall extremes, when using only short sub-hourly records, can be significantly reduced by incorporating long hourly records. In addition, the BL model provides a straightforward method to account for the varying impacts of climate dynamics on rainfall properties across different timescales.

These findings suggest that the BL model is a viable alternative to traditional annual maximum analysis, especially for short records. This ability to work with short records and integrate data of different lengths can help in regions that have only recently started collecting high-resolution rainfall records make better use of their data. In addition, it opens the door to other applications. For example, recent developments by Islam et al. (2022) and Islam et al. (2023) highlight the potential of applying the BL model to satellite-derived IMERG (Integrated Multi-satellite Retrievals for Global Precipitation Measurement mission) rainfall products.

The pyBL package developed in this work will not only help countries overcome the barrier of short records but also accelerate the exploration of various applications. By providing a robust and flexible tool for rainfall time series generation, pyBL can facilitate a more accurate and comprehensive analysis of rainfall extremes, which is crucial for water resource management, urban planning, and climate impact studies. The package's ability to integrate different timescales and lengths of data will particularly benefit regions with limited historical rainfall data, enabling them to make informed decisions based on more reliable rainfall statistics.

Code and data availability. The pyBL package 1.0.0, test dataset and the script to run the rainfall modelling with the Bartlett-Lewis process are available at https://doi.org/10.5281/zenodo.12605935 (Wei et al., 2024)

https://doi.org/10.5194/egusphere-2024-1918
Preprint. Discussion started: 22 July 2024

Author contributions. Li-Pen Wang (LW) conceptualised the research idea. Chi-Ling Wei (CW), Ting-Yu Dai (TD) and LW designed the package structure. CW, TD and Yung-Ting Ho (YH) led the development of the the package with support from Pei-Chun Chen (PC), Chien-Yu Tseng (CT) and Ching-Chun Chou (CC). PC and CT led the experimental design and result evaluation with support from LW and CW. Christian Onof (CO) provided advises on model validation and applications. LW and CO prepared the manuscript with contributions from CW, PC and CT

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. The authors express their gratitude for the financial support received from two National Science and Technology Council research projects (NSTC 111-2221-E-002 -060 -MY2 and 113-2124-M-002-014-).

https://doi.org/10.5194/egusphere-2024-1918
Preprint. Discussion started: 22 July 2024

References

Cannon, A. J., Jeong, D.-I., and Yau, K.-H.: Updated Observations Provide Stronger Evidence for Increases in Sub-hourly to Hourly Extreme Rainfall in Canada, Journal of Climate, https://doi.org/10.1175/JCLI-D-23-0501.1, 2024.

Chan, S., Kendon, E., Roberts, N., Fowler, H., and Blenkinsop, S.: The characteristics of summer sub-hourly rainfall over the southern UK in a high-resolution convective permitting model, Environmental Research Letters, 11, 094 024, 2016.

Cross, D., Onof, C., Winter, H., and Bernardara, P.: Censored rainfall modelling for estimation of fine-scale extremes, Hydrology and Earth System Sciences, 22, 727–756, https://doi.org/10.5194/hess-22-727-2018, 2018.

Cross, D., Onof, C., and Winter, H.: Ensemble simulation of future rainfall extremes with temperature dependent censored simulation, Advances in Meteorology, 2019.

Ebers, N., Schröter, K., and Müller-Thomy, H.: Estimation of future rainfall extreme values by temperature-dependent disaggregation of climate model data, EGUsphere, 2023, 1–25, 2023.

Efstratiadis, A., Koutsoyiannis, D., and Polytechniou, H.: An evolutionary annealing-simplex algorithm for global optimisation of water resource systems, in: Hydroinformatics 2002: Proceedings of the Fifth International Conference on Hydroinformatics, International Water Association, 2002.

Fowler, H. J., Ali, H., Allan, R. P., Ban, N., Barbero, R., Berg, P., Blenkinsop, S., Cabi, N. S., Chan, S., Dale, M., et al.: Towards advancing scientific knowledge of climate change impacts on short-duration rainfall extremes, Philosophical Transactions of the Royal Society A, 379, 20190 542, 2021.

Gires, A., Onof, C., Maksimovic, C., Schertzer, D., Tchiguirinskaia, I., and Simoes, N.: Quantifying the impact of small scale unmeasured rainfall variability on urban runoff through multifractal downscaling: A case study, Journal of Hydrology, 442, 117–128, 2012.

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., Gérard-Marchant, P., Sheppard, K., Reddy, T., Weckesser, W., Abbasi, H., Gohlke, C., and Oliphant, T. E.: Array programming with NumPy, Nature, 585, 357–362, https://doi.org/10.1038/s41586-020-2649-2, 2020.

Huang, J., Fatichi, S., Mascaro, G., Manoli, G., and Peleg, N.: Intensification of sub-daily rainfall extremes in a low-rise urban area, Urban Climate, 42, 101 124, https://doi.org/https://doi.org/10.1016/j.uclim.2022.101124, 2022.

Islam, M. A., Yu, B., and Cartwright, N.: Coupling of satellite-derived precipitation products with Bartlett-Lewis model to estimate intensity-frequency-duration curves for remote areas, Journal of Hydrology, 609, 127 743, 2022.

Islam, M. A., Yu, B., and Cartwright, N.: Bartlett–Lewis Model Calibrated with Satellite-Derived Precipitation Data to Estimate Daily Peak 15 Min Rainfall Intensity, Atmosphere, 14, 985, 2023.

Kaczmarska, J., Isham, V., and Onof, C.: Point process models for fine-resolution rainfall, Hydrological Sciences Journal, 59, 1972–1991, https://doi.org/10.1080/02626667.2014.925558, 2014.

Khaliq, M. and Cunnane, C.: Modelling point rainfall occurrences with the modified Bartlett-Lewis rectangular pulses model, Journal of Hydrology, 180, 109 – 138, https://doi.org/https://doi.org/10.1016/0022-1694(95)02894-3, 1996.

Kim, D. and Onof, C.: A stochastic rainfall model that can reproduce important rainfall properties across the timescales from several minutes to a decade, Journal of Hydrology, 589, 125 150, https://doi.org/https://doi.org/10.1016/j.jhydrol.2020.125150, 2020.

https://doi.org/10.5194/egusphere-2024-1918
Preprint. Discussion started: 22 July 2024

Kim, D., Cho, H., Onof, C., and Choi, M.: Let-It-Rain: a web application for stochastic point rainfall generation at ungaged basins and its applicability in runoff and flood modeling, Stochastic Environmental Research and Risk Assessment, 31, 1023–1043, https://doi.org/10.1007/s00477-016-1234-6, 2017a.

Kim, J.-G., Kwon, H.-H., and Kim, D.: A hierarchical Bayesian approach to the modified Bartlett-Lewis rectangular pulse model for a joint estimation of model parameters across stations, Journal of Hydrology, 544, 210–223, https://doi.org//10.1016/j.jhydrol.2016.11.031, 2017b.

Kossieris, P., Makropoulos, C., Onof, C., and Koutsoyiannis, D.: A rainfall disaggregation scheme for sub-hourly time scales: Coupling a Bartlett-Lewis based model with adjusting procedures, Journal of Hydrology, 556, 980 – 992, https://doi.org/10.1016/j.jhydrol.2016.07.015, 2018.

Koutsoyiannis, D., Onof, C., and Wheater, H. S.: Multivariate rainfall disaggregation at a fine timescale, Water Resources Research, 39, https://doi.org/10.1029/2002WR001600, 2003.

Marani, M.: On the correlation structure of continuous and discrete point rainfall, Water Resources Research, 39, 1128, https://doi.org/10.1029/2002WR001456, 2003.

Onof, C. and Arnbjerg-Nielsen, K.: Quantification of anticipated future changes in high resolution design rainfall for urban areas, Atmospheric Research, 92, 350 – 363, https://doi.org/10.1016/j.atmosres.2009.01.014, 7th International Workshop on Precipitation in Urban Areas, 2009.

Onof, C. and Wang, L.-P.: Modelling rainfall with a Bartlett–Lewis process: new developments, Hydrology and Earth System Sciences, 24, 2791–2815, https://doi.org/10.5194/hess-24-2791-2020, 2020.

Onof, C. and Wheater, H. S.: Modelling of British rainfall using a random parameter Bartlett-Lewis Rectangular Pulse Model, Journal of Hydrology, 149, 67 – 95, https://doi.org/10.1016/0022-1694(93)90100-N, conventry Groundwater Investigation: Sources and Movement of Chlorinated Hydrocarbon Solvents, 1993.

Onof, C. and Wheater, H. S.: Improvements to the modelling of British rainfall using a modified Random Parameter Bartlett-Lewis Rectangular Pulse Model, Journal of Hydrology, 157, 177 – 195, https://doi.org/10.1016/0022-1694(94)90104-X, 1994.

Onof, C., Chandler, R. E., Kakou, A., Northrop, P., Wheater, H. S., and Isham, V.: Rainfall modelling using Poisson-cluster processes: a review of developments, Stochastic Environmental Research and Risk Assessment, 14, 384–411, https://doi.org/10.1007/s004770000043, 2000.

Papalexiou, S. M.: Rainfall generation revisited: Introducing CoSMoS-2s and advancing copula-based intermittent time series modeling, Water Resources Research, 58, e2021WR031 641, 2022.

Park, J., Onof, C., and Kim, D.: A hybrid stochastic rainfall model that reproduces some important rainfall characteristics at hourly to yearly timescales, Hydrology and Earth System Sciences, 23, 989–1014, https://doi.org/10.5194/hess-23-989-2019, 2019.

Rodriguez-Iturbe, I., Cox, D. R., and Isham, V.: Some models for rainfall based on stochastic point processes, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 410, 269–288, https://doi.org/10.1098/rspa.1987.0039, 1987.

Rodriguez-Iturbe, I., Cox, D. R., and Isham, V.: A point process model for rainfall: further developments, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 417, 283–298, https://doi.org/10.1098/rspa.1988.0061, 1988.

Verhoest, N., Troch, P. A., and Troch, F. P. D.: On the applicability of Bartlett–Lewis rectangular pulses models in the modeling of design storms at a point, Journal of Hydrology, 202, 108 – 120, https://doi.org/10.1016/S0022-1694(97)00060-7, 1997.

Verhoest, N. E. C., Vandenberghe, S., Cabus, P., Onof, C., Meca-Figueras, T., and Jameleddine, S.: Are stochastic point rainfall models able to preserve extreme flood statistics?, Hydrological Processes, 24, 3439–3445, https://doi.org/10.1002/hyp.7867, 2010.

https://doi.org/10.5194/egusphere-2024-1918
Preprint. Discussion started: 22 July 2024

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C. J., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P., and SciPy 1.0 Contributors: SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, Nature Methods, 17, 261–272, https://doi.org/10.1038/s41592-019-0686-2, 2020.

Wang, L.-P., Marra, F., and Onof, C.: Modelling sub-hourly rainfall extremes with short records – a comparison of MEV, Simplified MEV and point process methods, EGU General Assembly 2020, https://doi.org/10.5194/egusphere-egu2020-6061, an optional note, 2020.

Wei, C.-L., Chen, P.-C., Dai, T.-Y., Wang, L.-P., Tseng, C.-Y., and Chou, C.-C.: NTU-CompHydroMet-Lab/pyBL: v1.0.0, https://doi.org/10.5281/zenodo.12605935, 2024.

Wes McKinney: Data Structures for Statistical Computing in Python, in: Proceedings of the 9th Python in Science Conference, edited by Stéfan van der Walt and Jarrod Millman, pp. 56 – 61, https://doi.org/10.25080/Majora-92bf1922-00a, 2010.