

## Modelling annual maximum daily rainfall with the STORAGE (STOchastic RAInfall GEnerator) model

Andrea Petroselli <sup>a</sup>, Davide Luciano De Luca <sup>b</sup>, Dariusz Młyński <sup>c,\*</sup> and Andrzej Wałęga <sup>c</sup>

<sup>a</sup> Department of Economy, Engineering, Society and Business Organization (DEIM), Tuscia University, Viterbo, Italy

<sup>b</sup> Department of Informatics, Modelling, Electronics and System Engineering, University of Calabria, Arcavacata di Rende, Italy

<sup>c</sup> Department of Engineering Sanitary and Water Management, University of Agriculture in Krakow, Krakow, Poland

\*Corresponding author. E-mail: [dariusz.mlynski@urk.edu.pl](mailto:dariusz.mlynski@urk.edu.pl)

 AP, 0000-0003-4943-0928; DLD, 0000-0001-5255-7343; DM, 0000-0002-8925-2591; AW, 0000-0001-6839-4745

### ABSTRACT

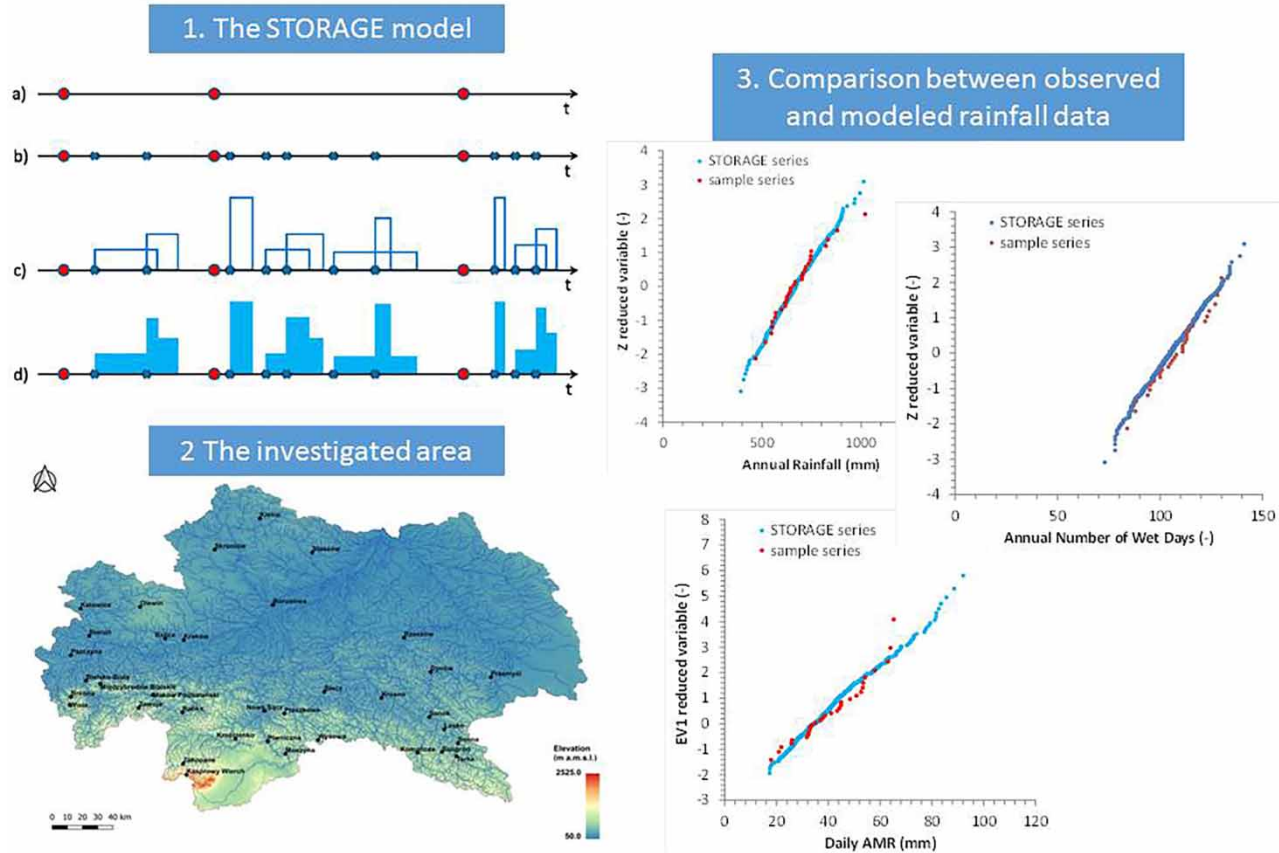
In this work, the capability of STORAGE (STOchastic RAInfall GEnerator) model for generating long and continuous rainfall series for the upper Vistula basin (southern Poland) is tested. Specifically, in the selected area, only parameters of depth–duration–frequency curves are known for sub-daily rainfall heights (which are usually estimated in an indirect way by using Lambor's equations from daily data), while continuous daily series with a sufficient sample size are available. Attention is focused on modelling the sample frequency distributions of daily annual maximum rainfall. The obtained results are promising for further elaborations, concerning the use of STORAGE synthetic continuous rainfall data as input for a continuous rainfall-runoff approach, to be preferred with respect to classical event-based modelling.

**Key words:** annual maximum daily rainfall, continuous approach, STORAGE model

### HIGHLIGHTS

- STORAGE model as an effective tool in rainfall modelling.
- Possibility of modelling rainfall with different characteristics.
- Promising results in the upper Vistula basin, paving the way for a continuous approach.

## GRAPHICAL ABSTRACT



## 1. INTRODUCTION

Hydro-meteorological data availability plays a crucial role in planning and managing the optimum use of water resources. In recent years, high maintenance costs and the need of trained staff have led to a consistent decrease in the number of observations in hydro-meteorological networks (Turnipseed & Sauer 2010). Hence, the modelling of hydrological or meteorological factors assumes an increasingly important role, for instance in the calculation of design discharges and flood prone areas in scarcely gauged catchments, for which the knowledge of design rainfall is mandatory (Sabzevari 2017). Rainfall temporal distribution, frequency, and duration are the basic information necessary to correctly estimate the runoff amount in such catchments (Wałęga *et al.* 2016; Petroselli *et al.* 2020). This is particularly important when using hydrological and hydrodynamic models, which are sensitive to changes in the input parameter values (Mannina *et al.* 2006; Sabzevari *et al.* 2009).

When sufficiently long observed rainfall information is available, statistical methods based on theoretical distribution functions are commonly used to assess the design rainfall. In particular, concerning annual maximum time series, commonly adopted functions are generalized extreme value (GEV), Log-Normal (LN), Weibull, Gumbel, generalized exponential distribution, generalized logistic distribution (Fitzgerald 2005; Boni *et al.* 2006; Villarini *et al.* 2011; Kaźmierczak & Kotowski 2012, 2015; Młyński *et al.* 2019a, 2019b).

Another solution is taking advantage of radar observations that also consider the spatial variability of the rainfall field, in order to reconstruct the sequences of rainfall characteristics, and then using them to determine the depth–duration–frequency (DDF) curves (Wright *et al.* 2013). Since there is a lack of radar meteorological data availability in some areas of the world, there is often the need to use the more widely available satellite-based products (Siddique-E-Akbor *et al.* 2014). Generally, the satellite and radar products underestimate extreme rainfall in comparison to gauge-based observations (Bharti *et al.* 2016; Chen *et al.* 2020). This is mainly due to the different spatial resolutions, i.e. satellite or radar outputs are average estimates on areal pixels,

while raingauge data refer to measures in specific points. Spatial resolution refers to the size of one pixel on the land. Landsat data, for example, has a 30 m resolution which is a medium-resolution image. It could cover bigger area alone, but this resolution level is not enough to distinguish individual objects like houses or cars. Currently, the following resolution levels are provided by commercial satellites: low (over 60 m per pixel), medium (from 30 to 10 m per pixel) and high to very high (from 5 m to 30 cm per pixel). For instance, [Derin \*et al.\* \(2019\)](#) investigated the ability of six satellite-based products to estimate extreme rainfall in nine mountainous locations with a dense raingauge network, determining a constant underestimation of extreme rainfall values. Consequently, satellite or radar estimates should require downscaling or bias correction methods ([Will-ems \*et al.\* 2012](#); [Maraun 2016](#)) for hydrological analyses. Other methods using satellite products to indirectly estimate rainfall from soil moisture can be used like SM2RAIN ([Brocca \*et al.\* 2018](#); [Dari \*et al.\* 2020](#)).

However, the evaluation of the design rainfall as input for the calculation of the design discharge (also named as event-based approach) can be affected by several relevant assumptions: the hyetograph shape, the critical rainfall duration, the pre-event soil moisture condition and the implicit hypothesis, assuming that the design storm and the design hydrograph have the same return period. So, in order to address the major drawbacks of event-based approaches, continuous rainfall-runoff models have been recently developed ([Grimaldi \*et al.\* 2012, 2021](#); [Biondi & De Luca 2015](#)), but they require long and continuous rainfall series at high resolutions (which are not so long or available in many areas of the world) as input.

In this context, synthetic rainfall generators (SRGs) are able to produce long and continuous rainfall series at the desired time resolution. Their calibration is carried out using the usually scarce information derived from data at sub-daily resolution (which should correspond to long high-resolution continuous data, but, in many cases, it only refers to the knowledge of annual maxima and DDF curves) and the more available and longer continuous daily time series.

For instance, [Iwański & Kuchar \(2003\)](#) described a SRG using daily series of meteorological data for a single geographical site with a known climatic profile. [Pierzga & Jarosińska \(2014\)](#) presented a SRG that used annual maximum daily rainfall data that later were disaggregated into hourly values, achieving to generate many synthetic hyetographs. [Agilan & Umamahesh \(2019\)](#) proposed a modified version of a K-nearest neighbour weather generator that incorporates nonstationarity in the extreme rainfall series. [Zhao \*et al.\* \(2019\)](#) described a daily, spatially distributed, SRG based on a first-order Markov chain model to simulate three types of rainfall, i.e. convective, frontal and tropical storms.

Among the SRGs, the Neyman–Scott Rectangular Pulse (NSRP) ([Rodriguez-Iturbe \*et al.\* 1987](#); [Cowpertwait 1991](#); [Burton \*et al.\* 2010](#)) is a commonly used model, which is suitable for generating continuous rainfall series at several time scales, from fine (e.g. 1 or 5 min) to coarse ones (e.g. average annual), and capable of preserving the statistical properties of the observed rainfall data ([Fowler \*et al.\* 2000](#)). In a stationary context, [De Luca & Galasso \(2019\)](#) calibrated a NSRP model by using information from extreme value distributions. Their results were promising, because they found acceptable relationships among extreme value distributions and rainfall statistical properties of continuous series, and this constitutes an interesting alternative to calibration when only annual maximum data are available. In a climate change context, [De Luca \*et al.\* \(2020a, 2020b\)](#) investigated the modelling of climate changes with a NSRP model, while [Wasko & Sharma \(2017\)](#) proposed a novel method for simulating continuous rainfall sequences for future warming climate using one-dimensional NSRP model.

Recently, [De Luca & Petroselli \(2021\)](#) developed a modified NSRP model named as STORAGE (STOchastic RAInfall GEnerator). Its main advantages with respect to a classic NSRP model are:

1. The model calibration is carried out using summary statistics from annual maxima series (from sub-hourly/hourly to daily scales), annual and seasonal cumulative rainfall heights, and annual number of wet days, which are usually longer than continuous observed high-resolution series (mainly adopted for SRGs calibration but typically very short or absent in many locations) and
2. The seasonality is modelled by using series of goniometric functions. This approach makes STORAGE more parsimonious with respect to the use of monthly or seasonal sets for parameters.

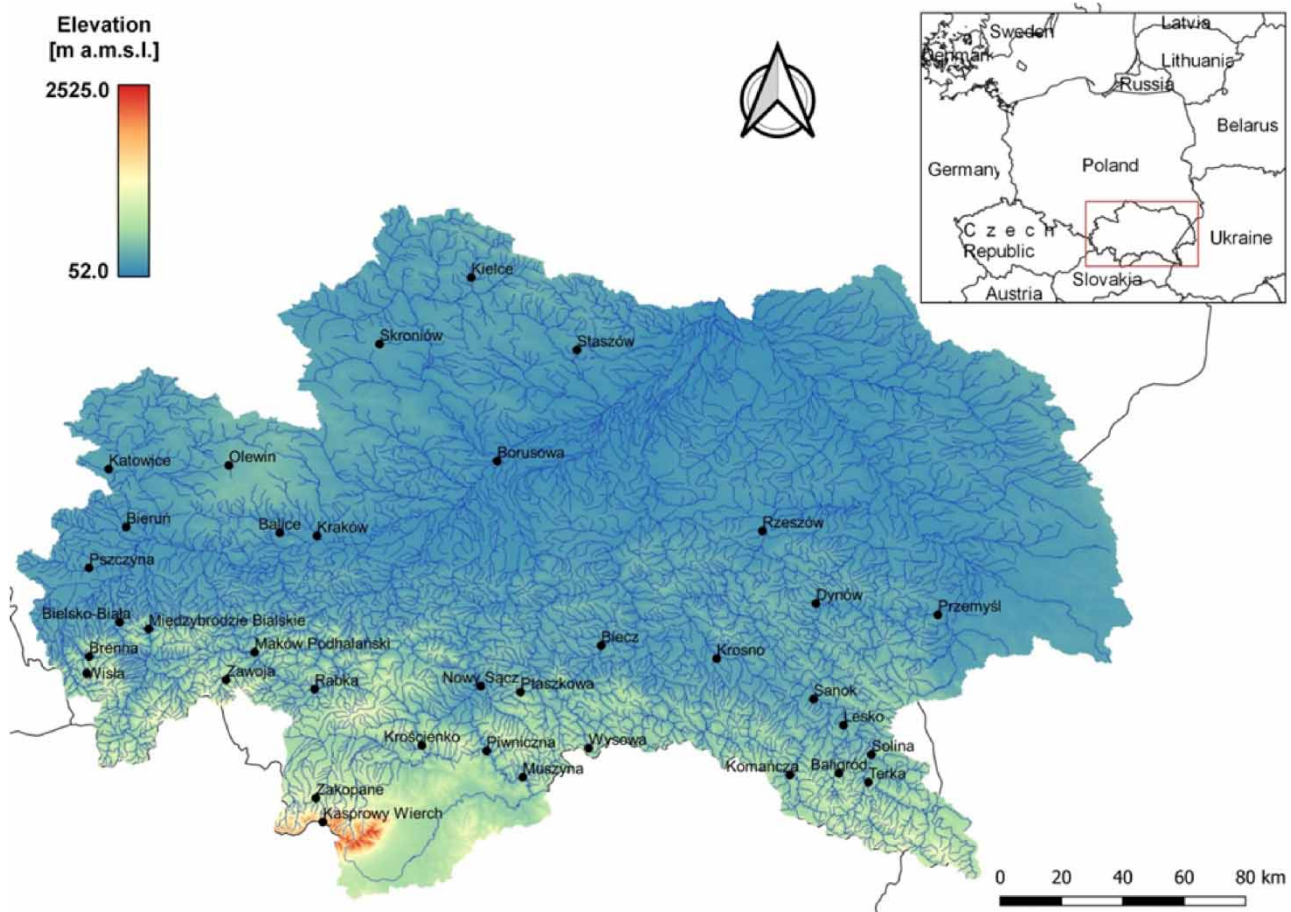
In the present work, we investigated the performances of the STORAGE model for the raingauge network of the upper Vistula basin, which is located in Poland. In the selected study area, only parameters of DDF curves are known for sub-daily rainfall heights, which are usually estimated in an indirect way by using Lambor's equations from daily data ([Młyński \*et al.\* 2020](#)), while continuous daily series with a sufficient sample size are available. In particular, the goal of the present work is to assess the capability of STORAGE in reconstructing the annual maxima of daily rainfall, which can constitute a starting point for further elaborations, related to the use of STORAGE synthetic high-resolution rainfall data as input for a continuous approach aiming in evaluating the design discharge in scarcely gauged basins.

## 2. STUDY AREA

The study was conducted for 36 rainfall stations that are located in the upper Vistula basin, Poland (Figure 1). This area covers about 25% of the entire Vistula basin and about 15% of the whole Poland. The region is divided into three main physiographic units: Carpathian's, uplands and flats. The water resources of the upper Vistula basin are the highest in the entire Poland. The water cycle in the mountain part is very dynamic, due to steep relief and low permeability of the ground, what causes that floods events develop rapidly here (Pociask-Karteczka 2016). The elevation of rainfall stations ranges from 171 to 1,991 m a.s.l. The used data were continuous daily rainfall series, which were recorded in the period 1989–2018, and the parameters of the DDF curves, which were estimated in an indirect way by using Lambor's equations from daily data (Młyński *et al.* 2020). The data were obtained from the public database of the Institute of Meteorology and Water Management, National Research Institute in Warsaw, Poland. Table A1 in Supplementary Material, Appendix 1 presents names, coordinates (ETRS89/Poland CS92, authority ID: EPSG:2180), elevation of analysed stations, and the mean annual precipitation (MAP), the mean annual number of wet days (MANWD) and the mean values for the seasons DJF (December–January–February), MAM (March–April–May), JJA (June–July–August) and SON (September–October–November). In this work, a wet day is defined when the daily rainfall height is at least equal to 1 mm.

## 3. METHODOLOGY

For each investigated raingauge, the study involved the following steps: (1) extracting the correspondent observed series of daily annual maximum rainfall (AMR, also indicated with  $P_{max}$  in the following) and determining its statistical properties; (2) analysing the trends in the observed  $P_{max}$  time series, in order to justify the use of STORAGE model in a stationary context and (3) generating the  $P_{max}$  using the STORAGE model and analysing the uncertainty due to the sample size.



**Figure 1** | Location of analysed rainfall stations in the upper Vistula basin.

Then, focusing on some specific raingauges as examples, STORAGE performances (in terms of comparison among observed and generated data for annual maxima, cumulative annual and seasonal rainfall data, plus annual number of wet days) are shown and discussed.

### 3.1. Statistical analysis

The statistical analysis of  $P_{max}$  observed time series regarded the estimation of minimum (*min*), mean, maximum (*max*), standard deviation (*s*) and coefficient of variation ( $C_v$ ) of the sample, together with the modelling of these series with the LN distribution (Kottegoda & Rosso 2008), which is commonly adopted for annual maxima data in Poland. However, Extreme Value Type 1 (EV1) distribution (Kottegoda & Rosso 2008) is also suitable for  $P_{max}$  time series in Poland, as shown in Section 4 for some rain gauge data series.

### 3.2. Trend analysis

Trend analysis for  $P_{max}$  observed time series was conducted using the modified Mann–Kendall (MK) test. The test considers the relative magnitudes of sample data rather than the data values themselves, and it does not require the data to follow any specific probability distribution. The null hypothesis  $H_0$  of test indicates a lack of monotonic trend for the time series. The alternative hypothesis  $H_1$  claims its existence. The  $S$  statistic of the modified Mann–Kendall test is expressed as (Wałęga *et al.* 2016; Ali *et al.* 2019; Agbo *et al.* 2021) follows:

$$S = \sum_{k=1}^{n-1} \sum_{j=k+1}^n \operatorname{sgn}(x_j - x_k) \quad (1)$$

$$\operatorname{sgn}(x_j - x_k) = \begin{cases} 1 & \text{for } (x_j - x_k) > 0 \\ 0 & \text{for } (x_j - x_k) = 0 \\ -1 & \text{for } (x_j - x_k) < 0 \end{cases} \quad (2)$$

where  $n$  is the number of elements of the time series,  $x_j$  is the observation at time  $j$  and  $x_k$  is the observation at time  $k$ .

The  $S$  is approximately normally distributed with the mean  $E[S]$  equal to 0 and the variance  $\operatorname{Var}(S)$  is computed as follows:

$$\operatorname{Var}(S) = \frac{1}{18} \cdot (n \cdot (n-1) \cdot (2 \cdot n + 5)) \quad (3)$$

The standardized test statistic for the modified Mann–Kendall test  $Z(S)$  was calculated based on the following equation:

$$Z(S) = \frac{S - \operatorname{sgn}(S)}{\operatorname{Var}(S)^{1/2}} \quad (4)$$

Positive values of  $Z$  statistic indicate increasing trends, while negative values of  $Z$  show decreasing trends. Moreover, once chosen a specific significance level and calculated the theoretical quantile  $Z_{1-(\alpha/2)}$ , the null hypothesis of no trend is rejected when  $|Z_S| > Z_{1-(\alpha/2)}$ . For the frequently adopted value of  $\alpha = 0.05$ , the null hypothesis of no trend is rejected when  $|Z_S| > 1.96$ .

The main assumption of the modified Mann–Kendall test is the lack of autocorrelation in analysed time series. Because the autocorrelation leads to underestimation of the variance  $S$ , the following adjustment for variance corrections was included only for data where the significant autocorrelation was detected (Sridhar & Raviraj 2017; Młyński *et al.* 2019a, 2019b):

$$\operatorname{Var}^*(S) = \operatorname{Var}(S) \cdot \frac{n}{n_s^*} \quad (5)$$

where  $n/n_s^*$  is a correction factor, which is calculated as follows:

$$\frac{n}{n_s^*} = 1 + \frac{2}{n(n-1)(n-2)} \cdot \sum_{k=1}^{n-1} (n-k)(n-k-1)(n-k-2)\rho_k \quad (6)$$

where  $n$  is the number of elements of the time series,  $n_s^*$  is the effective number of observations to account for the autocorrelation in the data and  $\rho_k$  is the autocorrelation function for the ranks of the observations.

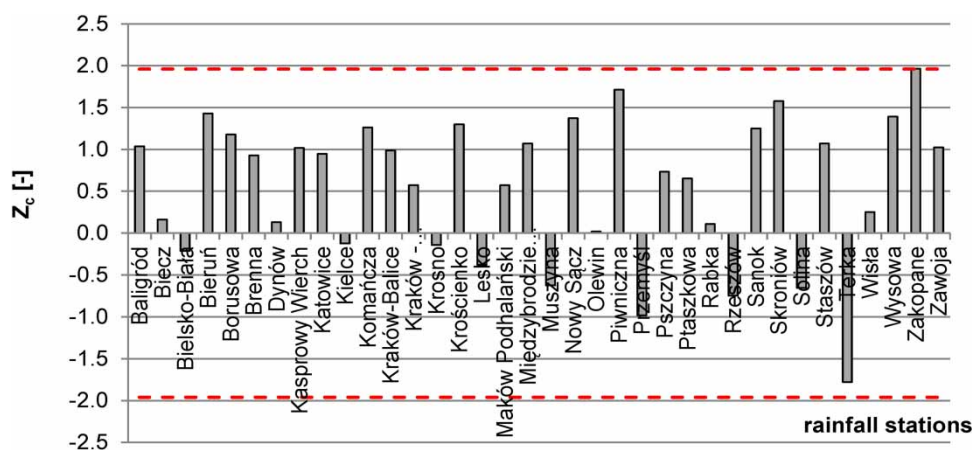
It should be remarked that a more comprehensive trend investigation should regard fine and coarse time scales and not only the extremes, but if high-resolution data are scarce or absent, then trend/no trend hypothesis for specific fine time scales can only be assumed on the basis of the obtained MK test results for daily data. Moreover, since the present study aims at evaluating the design discharge, the extreme analysis is sufficient for investigation purposes.

### 3.3. Generation of daily AMR using the NSRP model and analysis of uncertainty due to the sample size

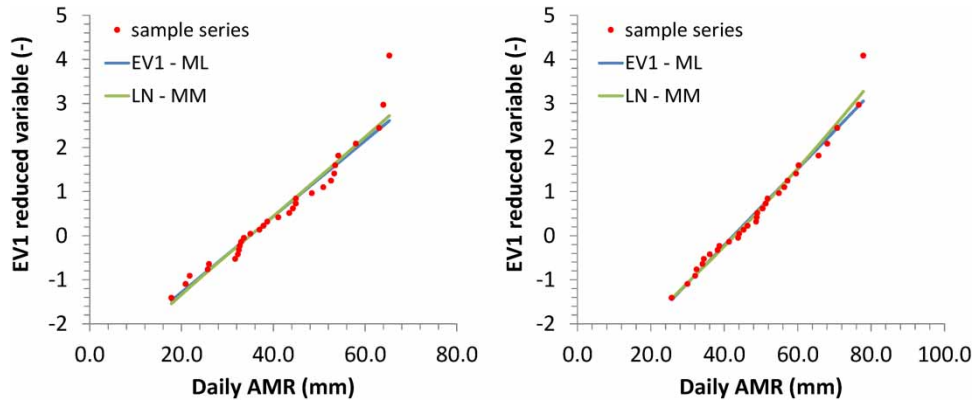
Authors adopted a modified version of the NSRP (Rodriguez-Iturbe *et al.* 1987; Cowpertwait 1991; Burton *et al.* 2010) model, named STORAGE – STOchastic RAINfall GEnerator (De Luca *et al.* 2020a, 2020b; De Luca & Petroselli 2021). The NSRP model usually presents a set of parameters, which represent mean values for (1) inter-arrival time between the origins of two consecutive storms; (2) the number of rain cells, also indicated as bursts or pulses, inside a specific storm; (3) waiting time between a specific burst origin and the origin of the associated storm; (4) intensity and (5) duration of the bursts. Such parameters are estimated by minimizing an objective function (OF), which is defined as the sum of residuals between the considered summary statistics of the observed data at chosen time resolutions and their theoretical expressions. The summary statistics are typically referred to high-resolution continuous time series, which unfortunately are very short or absent in many locations. Moreover, the main limitation of the basic version of NSRP is that the model fitting to a particular rainfall time series involves highly nonlinear equations, regarding analytical and sample moments, and then the parameters estimation needs nontrivial solutions. Even when the nonlinear sophisticated methods are used, the fitting process is a trial-and-error process consisting in trying different initial parameter values. The process is finished when physically reasonable parameter final values are found that provide a good fit (Morrissey 2009). For this reason, STORAGE was developed, being characterized by two main innovations as respect to the basic versions of the NSRP model. First, the parametric calibration is carried out using data series (e.g. annual maxima rainfall, annual and monthly cumulative rainfall and annual number of wet days), which are usually longer than observed high-resolution series. Second, the seasonality is modelled using series of goniometric functions. This approach makes STORAGE strongly parsimonious as respect to basic NSRP models.

In the present work, STORAGE was applied, for all the considered raingauge stations, by employing as input data, the continuous rainfall daily data, recorded in the period 1989–2018, and the parameters of the DDF curves. MAP, MANWD and mean values of seasonal rainfall (DJF, MAM, JJA and SON) were calculated from daily data for each raingauge (see Table A1 in Supplementary Material, Appendix): they constitute (together with parameters of DDF curves) the input information for STORAGE calibration (De Luca & Petroselli 2021).

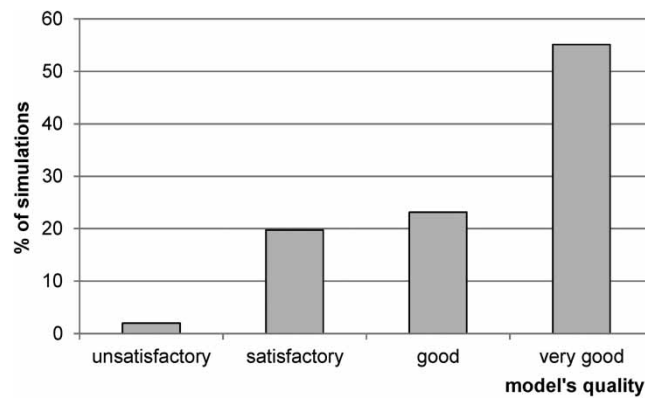
Then, a 500-year continuous rainfall series with a resolution of 60 min was generated for each raingauge. It must be highlighted that both stationary and non-stationary time series can be generated with STORAGE, depending on the observed/



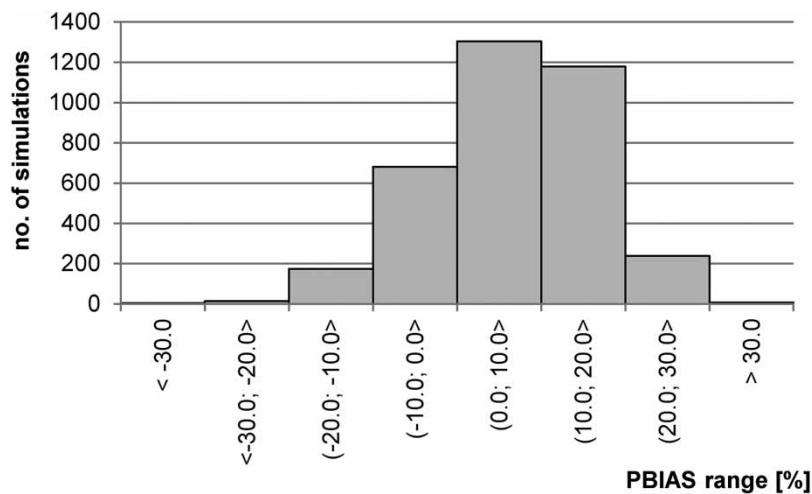
**Figure 2** | Results of the Z statistics for the Mann-Kendall test.



**Figure 3** | EV1-ML and LN-MM fitting of sample daily AMR distribution for Kraków-Balice (left) and Solina (right).



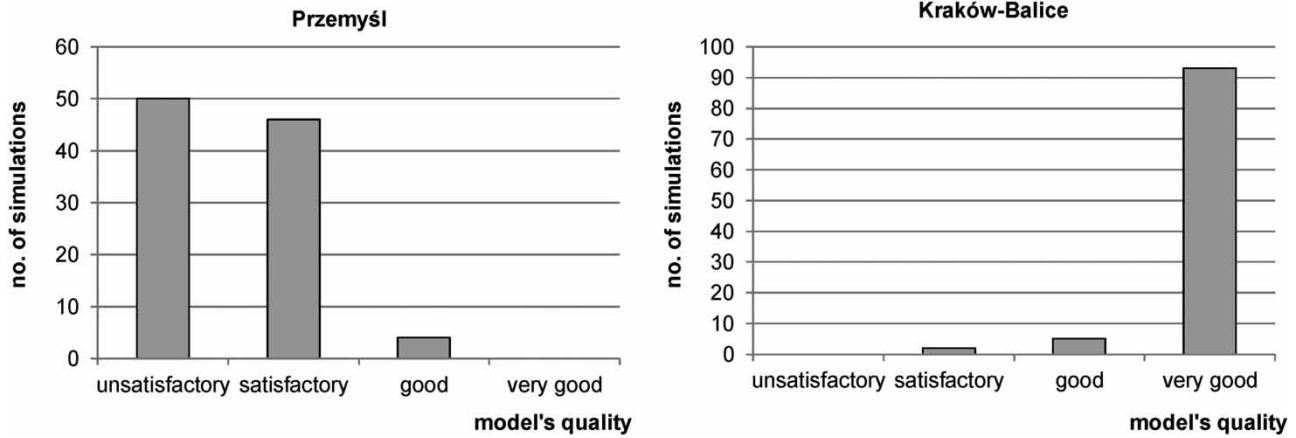
**Figure 4** | Summary of the quality of the STORAGE model for the entire upper Vistula basin.



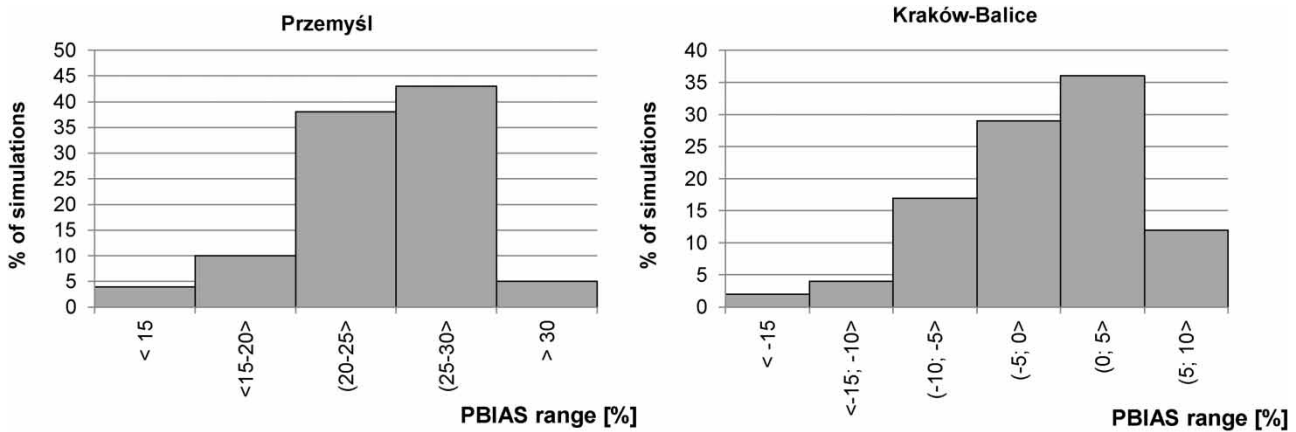
**Figure 5** | The values of PBIAS for the upper Vistula basin.

predicted trend. In this work, the stationary version was adopted based on results from the modified Mann-Kendall test (MMK) test. For further details related to the STORAGE application, see De Luca & Petroselli (2021).

Next, for each raingauge, STORAGE daily AMR series were extracted and statistically examined, also considering uncertainty analysis due to the sample size. In detail, from the 500-year series of annual maxima, 100 samples, each one with 30



**Figure 6** | Summary of the quality of the STORAGE model for the station with the worst (Przemyśl) and best (Kraków-Balice) results.



**Figure 7** | The values of PBIAS for the station with the worst (Przemyśl) and best (Kraków-Balice) results.

elements, were randomly sampled. After that, the comparison between the observed 30-year time series and each 30-year series generated using the STORAGE model was made by calculating the mean value of *PBIAS* (Kim & Kim 2016; Młyński 2020):

$$PBIAS = \frac{P_T - P_{T_{STORAGE}}}{P_T} \cdot 100[\%] \tag{7}$$

where  $P_T$  is the daily design rainfall, for the assumed return period, which is calculated from observed time series, and  $P_{T_{STORAGE}}$  is the daily design rainfall, for the assumed return period, which is calculated from the STORAGE model.

Both  $P_T$  and  $P_{T_{STORAGE}}$  are evaluated by using LN distribution. The mean of *PBIAS* (indicated as *mPBIAS* in the following) was estimated by considering the obtained values from specific return periods (2, 5, 10, 20, 50 and 100 years).

**Table 1** | Annual and seasonal summary statistics for Kraków-Balice and Solina raingauges

Raingauge	MAP (mm)	MANWD (-)	DJF (mm)	MAM (mm)	JJA (mm)	SON (mm)
Kraków-Balice	673.2	108.0	100.3	163.9	251.9	157.1
Solina	940.3	129.6	140.1	235.4	343.9	220.9



The performance rating for mean values  $mPBIAS$  can be expressed as in the following:  $|mPBIAS| < 10$  very good;  $10 \leq |mPBIAS| < 15$  good;  $15 \leq |mPBIAS| < 25$  satisfactory;  $|mPBIAS| \geq 25$  unsatisfactory (Ajmal *et al.* 2020).

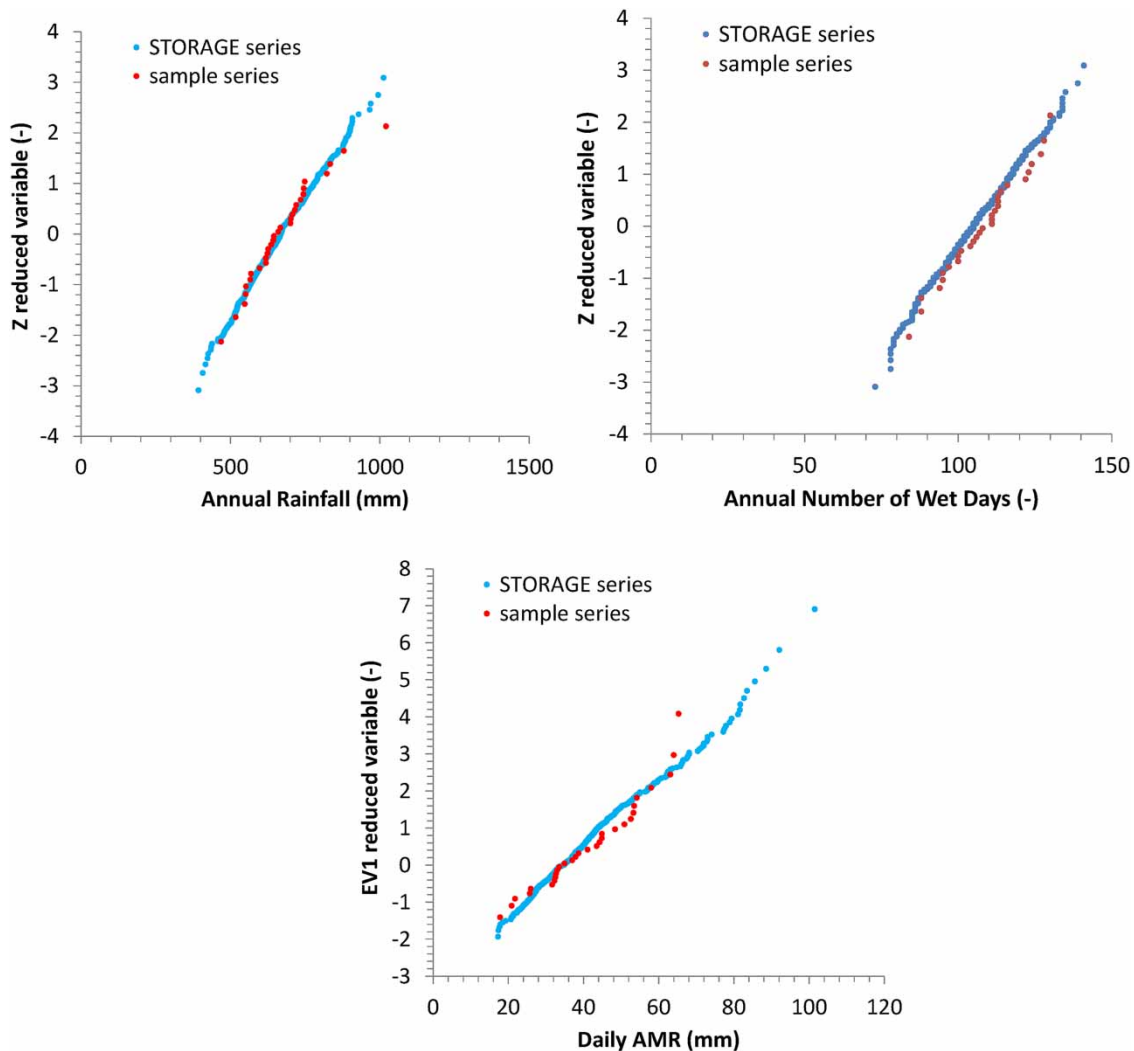
## 4. RESULTS

### 4.1. Initial analysis of the observed daily AMR

The preliminary analysis included the calculation of summary statistics of the observed time series: minimum (*min*), mean, maximum (*max*), standard deviation (*s*) and coefficient of variation ( $C_v$ ). The results are presented in Table A2 in Supplementary Material, Appendix. Analysing the coefficient of variation, it was concluded that 78% of all stations were characterized by medium dynamics of changes in AMR, which was described by  $C_v$  values in range from 0.2 to 0.4. Only for the eight stations: Bielsko-Biała, Brenna, Kielce Krościenko, Maków Podhalański, Międzybrodzie Bielskie, Skroniów, Wisła and Zawoja, the values of  $C_v$  are greater than 0.4, which means high dynamics of changes in their AMR.

### 4.2. Analysis of trend in the observed daily AMR

As previously mentioned, trend investigation was carried out for the observed daily AMR series using the modified Mann-Kendall test.



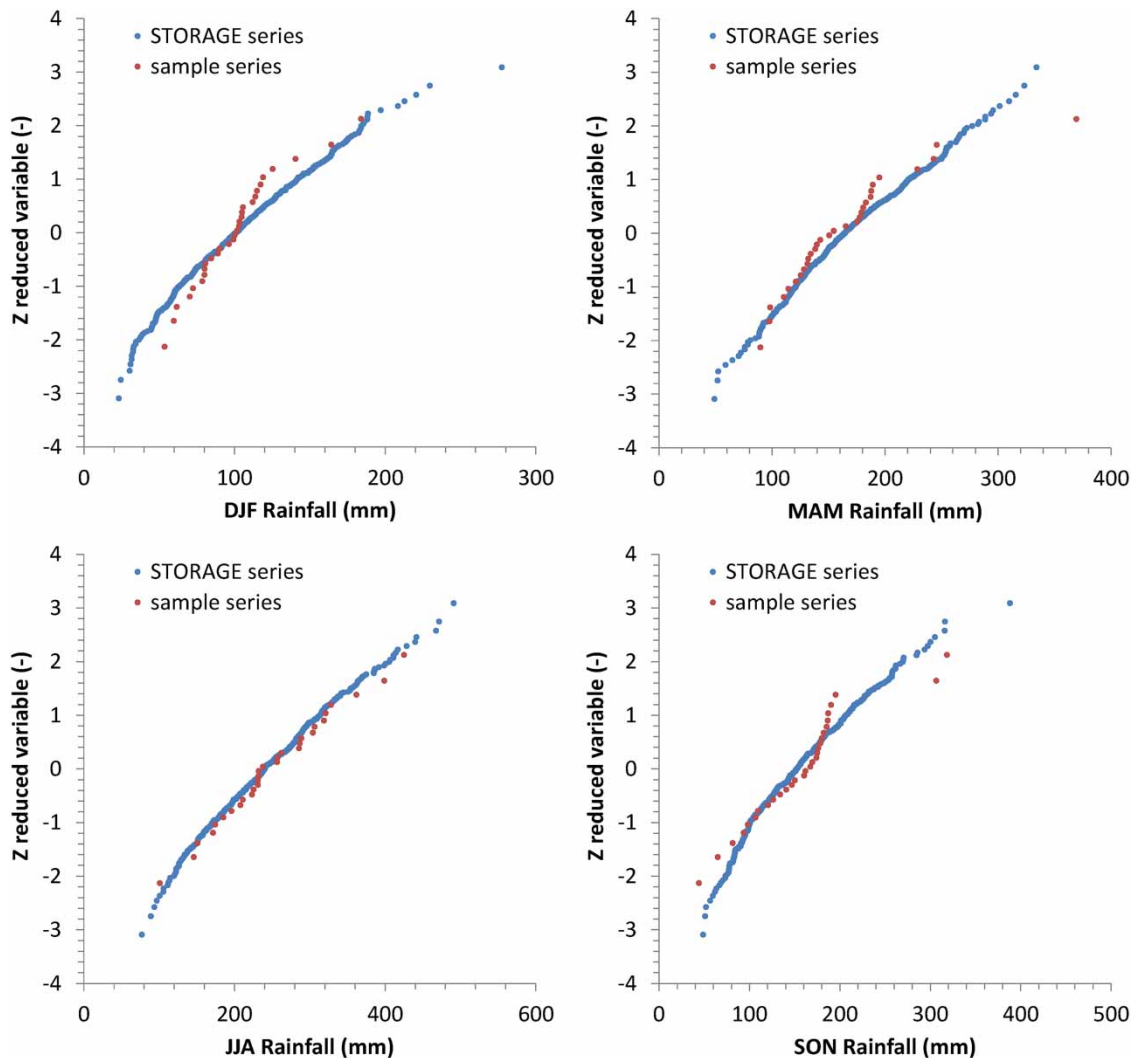
**Figure 8** | Kraków-Balice raingauge: STORAGE performances in terms of reproduction of sample frequency distribution for annual rainfall, annual number of wet days and daily AMR.

From the obtained results (Figure 2), it should be emphasized that there was no case where the  $|Z|$  statistic is higher than 1.96, and then the hypothesis of stationary processes cannot be rejected. Similar results were obtained by Młyński *et al.* (2018), where significant trends in the annual daily maximum rainfall for the upper Vistula basin have not been found. The study conducted by Kubiak-Wójcicka (2020) has shown a slight increase of rainfall for the Vistula basin, mostly in particular seasons. Seasonal trends were analysed also by Pińskwar *et al.* (2019). The analysis of rainfall extremes showed that maximum daily rainfall mostly increased for the summer half-year and that the increases during summer half-year are more numerous than in winter half-year. However, many changes were weak and statistically insignificant. Finally, Niedźwiedź *et al.* (2014) published similar results for southern Poland, which indicated growing but not significant trends in daily rainfall.

Concerning  $P_{max}$  modelling, under the hypothesis of stationarity, both EV1 and LN distributions (Kottegoda & Rosso 2008) usually provide comparable performances. As examples, the EV1 and LN fittings are illustrated in Figure 3 for Kraków-Balice and Solina raingauges using EV1 probabilistic plots. In detail, EV1 distribution is estimated using the maximum-likelihood (ML) method, while the method of moments (MM) is adopted for LN function. The observed time series for Kraków-Balice and Solina raingauges are in-depth analysed in Section 4.4, where examples of the STORAGE application are discussed.

### 4.3. STORAGE uncertainty analysis due to the sample size

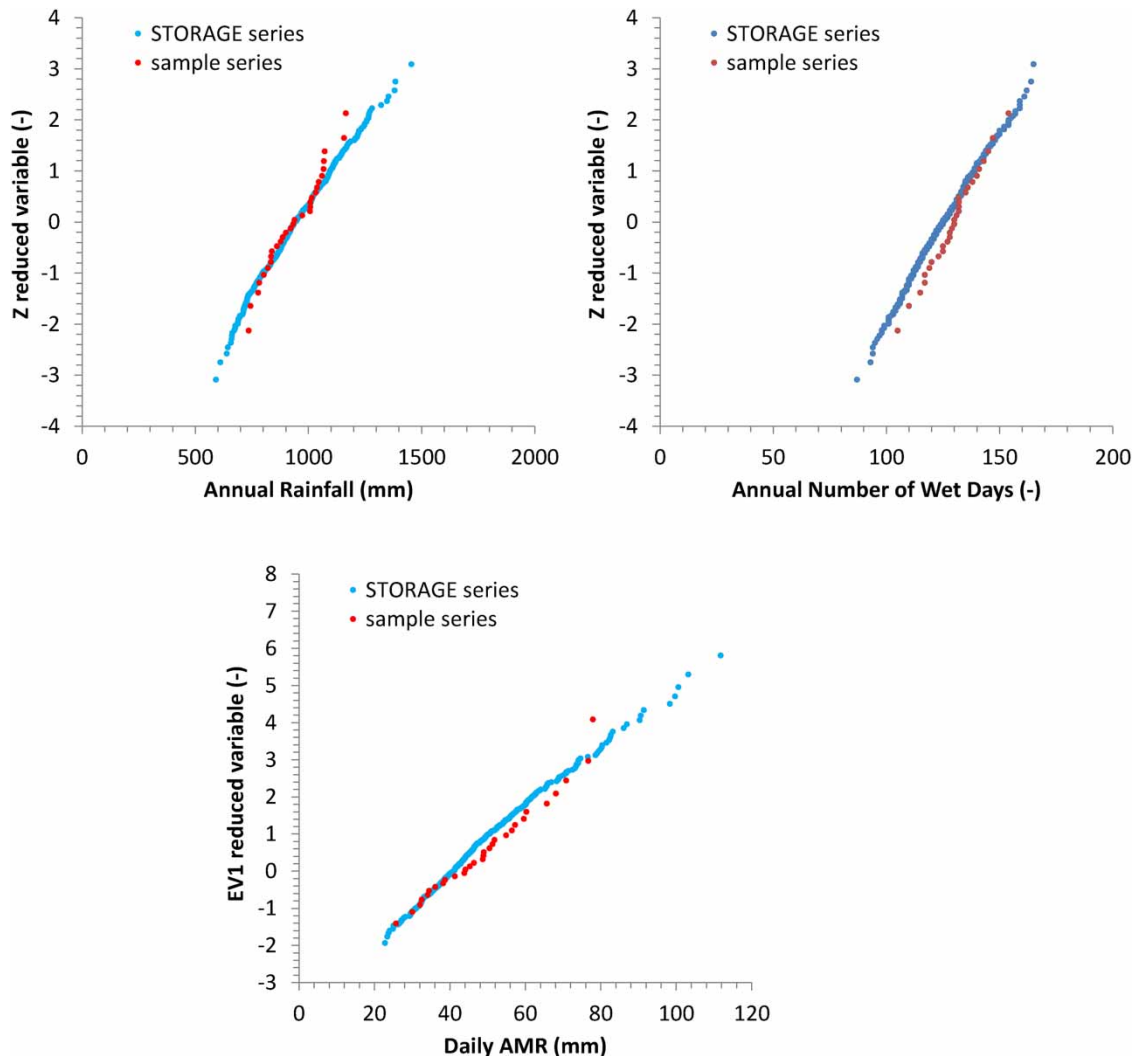
As previously mentioned in Section 3.3, the uncertainty analysis of the STORAGE model consisted in selecting randomly 30-element series from the 500-year synthetic AMR series, for each raingauge station, and comparing with the observed annual



**Figure 9** | Kraków-Balice raingauge: STORAGE performances in terms of reproduction of sample frequency distribution for seasonal rainfall.

maxima. Next, the values of *PBIAS* were determined. The operation was repeated 100 times. The results are presented in Figures 4 and 5 and in Table A3 in Supplementary Material, Appendix.

Based on the obtained results, it was concluded that the STORAGE model generated the daily AMR series usually with good or very good quality. The best results were obtained for Bieruń, Borusowa, Brenna, Kraków-Balice and Międzybrodzie Bialskie rainfall stations, where the *PBIAS* quality was very good in more than 90 out of 100 simulations. It should be emphasized that in 75% of all analysed rainfall stations, the unsatisfactory quality of model was not observed. Only for the Przemyśl station, there was 50 out of 100 unsatisfactory cases. Generally, analysing the model's quality for the entire upper Vistula basin (Figure 4), it was concluded that only 2% of all simulations were unsatisfactory, 20% was satisfactory, 23% was good and 55% was very good, i.e. 98 and 78% of simulations were at least satisfactory and good, respectively. Figure 6 presents the summary of quality for the station with best (Kraków-Balice) and worst results (Przemyśl). Figure 7 presents the values of *PBIAS* for the station with best and worst results. Figure 6 shows that the very good quality of the model was not obtained for the Przemyśl station. The most common *PBIAS* values were in range from 25 to 30%. Moreover, for the Przemyśl station, the values of *PBIAS* were only positive. For the Kraków-Balice station, the unsatisfactory quality of model was not observed and the most common *PBIAS* values were in the range from 0 to 5%. In this case, negative values of *PBIAS* were also observed.



**Figure 10** | Solina rain gauge: STORAGE performances in terms of reproduction of sample frequency distribution for annual rainfall, annual number of wet days and daily AMR.

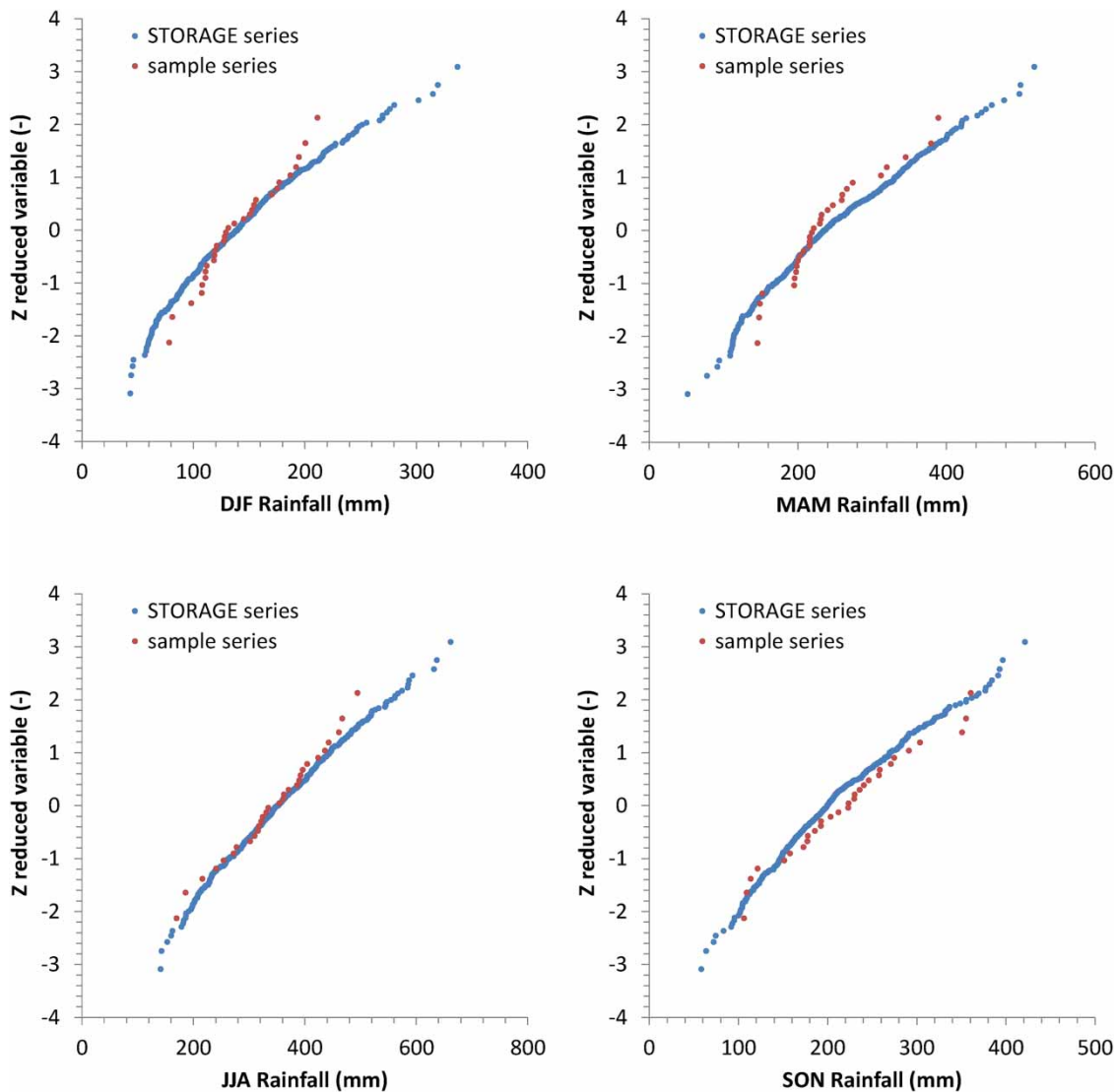
#### 4.4. Examples of STORAGE generation for two observed time series

As examples of in-depth analyses of STORAGE application, we focused on continuous daily series for Kraków-Balice and Solina raingauges, for which the summary statistics concerning annual and seasonal rainfall are reported in Table 1. These selected time series present good performances in terms of mPBIAS and are located in different areas of the investigated Vistula basin (Figure 1).

In detail, from the generated 500-year synthetic series with a resolution of 60 min, we carried out model validation by analysing the reproduction of sample frequency distributions of daily AMR, annual and seasonal rainfall, and annual number of wet days (Figures 8–11).

Concerning annual and seasonal frequency distributions, STORAGE provided for both raingauges (on Gaussian probabilistic plots), a very good fitting for annual rainfall and a slight underestimation for annual number of wet days, while a good/acceptable fitting is obtained for seasonal rainfall heights. Sample daily AMR distributions are fitted in good/acceptable way as well (on EV1 probabilistic plots).

These specific results for Kraków-Balice and Solina raingauges clearly constitute a first promising step for further investigations of rainfall fields in Poland employing the STORAGE model, which will specifically focus on the high resolutions and



**Figure 11** | Solina raingauge: STORAGE performances in terms of reproduction of sample frequency distribution for seasonal rainfall.

on the use of STORAGE synthetic rainfall data for a continuous approach aiming at evaluating the design discharge in ungauged or scarcely gauged basins.

## 5. CONCLUSION

This work aimed at modelling, for 36 raingauge stations located in the upper Vistula basin (Poland), the daily annual maximum rainfall (AMR) time series, extracted from synthetic continuous rainfall data which were generated using the STORAGE model. The present analysis constitutes a preliminary step of a more comprehensive study, focused on the use of the synthetic high-resolution rainfall series as input for a continuous rainfall-runoff modelling, aiming at evaluating the design discharge in ungauged or scarcely gauged basins.

For all the investigated raingauge stations, a lack of significant trends for the observed data series was assessed; hence, a stationary version of STORAGE was adopted for the successive steps. Moreover, the uncertainty analysis, due to sample size, for STORAGE output highlighted that the generated values of daily AMR are usually characterized by good or very good quality (in terms of mPBIAS), with respect to the observed data. Good or very good qualities have been confirmed from the in-depth analysis of STORAGE performances for two raingauges, chosen as examples: Kraków-Balice and Solina stations. The obtained reconstruction of sample frequency distributions, in terms of daily AMR, annual and seasonal rainfall heights and annual number of wet days, allows to consider STORAGE as a promising tool for modelling Polish rainfall fields. Further steps of the more comprehensive study of interest (i.e. design discharge evaluation for ungauged or scarcely gauged basins in Poland) are encouraged.

## AUTHORS CONTRIBUTIONS

A.P., D.L.D.L., D.M. and A.W. conceptualized and investigated the study; A.P. and D.L.D.L. did the data curation, formulated the study methodology and provided software and validation; D.M. provided resources; D.L.D.L. and A.W. provided supervision; D.M. did the visualization; writing of the original draft was done by D.M. and A.W., and writing of review & editing was done by A.P. and D.L.D.L.

## CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

## DATA AVAILABILITY STATEMENT

Data cannot be made publicly available; readers should contact the corresponding author for details.

## REFERENCES

- Agbo, E. P., Ekpo, C. M. & Edet, C. O. 2021 Analysis of the effects of meteorological parameters on radio refractivity, equivalent potential temperature and field strength via Mann-Kendall test. *Theoretical and Applied Climatology* **143**, 1437–1456.
- Agilan, V. & Umamahesh, N. V. 2019 Rainfall generator for nonstationary extreme rainfall condition. *Journal of Hydrologic Engineering* **24** (9), 1–13.
- Ajmal, M., Waseem, M., Kim, D. & Kim, T.-W. 2020 A pragmatic slope-adjusted curve number model to reduce uncertainty in predicting flood runoff from steep watersheds. *Water* **12**, 1–15.
- Ali, R., Kuriqi, A., Abubaker, S. & Kisi, O. 2019 Long-term trends and seasonality detection of the observed flow in Yangtze River using Mann-Kendall and Sen's innovative trend method. *Water* **11**, 1–17.
- Bharti, V., Singh, C., Ettema, J. & Turkington, T. A. R. 2016 Spatiotemporal characteristics of extreme rainfall events over the Northwest Himalaya using satellite data: spatiotemporal characteristics of extreme rainfall events. *International Journal of Climatology* **36**, 3949–3962.
- Biondi, D. & De Luca, D. L. 2015 Process-based design flood estimation in ungauged basins by conditioning model parameters on regional hydrologic signatures. *Natural Hazards* **79** (2), 1015–1038.
- Boni, G., Parodi, A. & Rudari, R. 2006 Extreme rainfall events: learning from raingauge time series. *Journal of Hydrology* **327**, 304–314.
- Brocca, L., Tarpanelli, A., Filippucci, P., Dorigo, W., Zaussinger, F., Gruber, A. & Fernández-Prieto, D. 2018 How much water is used for irrigation? A new approach exploiting coarse resolution satellite soil moisture products. *International Journal of Applied Earth Observation and Geoinformation* **73**, 752–766.
- Burton, A., Fowler, H. J., Blenkinsop, S. & Kilsby, C. G. 2010 Downscaling transient climate change using a Neyman-Scott rectangular pulses stochastic rainfall model. *Journal of Hydrology* **381** (1–2), 18–32.

- Chen, S., Liu, B., Tan, X. & Wu, Y. 2020 Inter-comparison of spatiotemporal features of precipitation extremes within six daily precipitation products. *Climate Dynamics* **54**, 1057–1076.
- Cowpertwait, P. S. P. 1991 Further developments of the Neyman-Scott clustered point process for modeling rainfall. *Water Resources Research* **27**, 1431–1438.
- Dari, J., Brocca, L., Quintana-Seguí, P., Escorihuela, M. J., Stefan, V. & Morbidelli, R. 2020 Exploiting high-resolution remote sensing soil moisture to estimate irrigation water amounts over a Mediterranean region. *Remote Sensing* **12** (16), 1–22.
- De Luca, D. L. & Galasso, L. 2019 Calibration of NSRP models from extreme value distributions. *Hydrology* **6**, 1–20.
- De Luca, D. L. & Petroselli, A. 2021 STORAGE (STOchastic RAInfall GEnerator): a user-friendly software for generating long and high-resolution rainfall time series. *Hydrology* **8**, 1–32.
- De Luca, D. L., Petroselli, A. & Galasso, L. 2020a Modelling climate changes with stationary models: is it possible or is it a paradox? In: *Numerical Computations: Theory and Algorithms* (Sergeyev, Y. & Kvasov D, D., eds). Springer, Cham.
- De Luca, D. L., Petroselli, A. & Galasso, L. 2020b A transient stochastic rainfall generator for climate changes analysis at hydrological scales in Central Italy. *Atmosphere* **11** (12), 1–21.
- Derin, Y., Nikolopoulos, E. & Anagnostou, E. N. 2019 Chapter seven – estimating extreme precipitation using multiple satellite-based precipitation products. In: *Extreme Hydroclimatic Events and Multivariate Hazards in a Changing Environment* (Maggioni, V. & Massari, C., eds). Elsevier, Cambridge, pp. 163–190.
- Fitzgerald, D. L. 2005 Analysis of extreme rainfall using the log-logistic distribution. *Stochastic Environmental Research and Risk Assessment* **19**, 249–257.
- Fowler, H. J., Kilsby, C. G. & O'Connell, P. E. 2000 A stochastic rainfall model for the assessment of regional water resource systems under changed climatic conditions. *Hydrology and Earth System Sciences* **4**, 261–280.
- Grimaldi, S., Petroselli, A. & Serinaldi, F. 2012 Design hydrograph estimation in small and ungauged watersheds: continuous simulation method versus event-based approach. *Hydrological Process* **26**, 3124–3134.
- Grimaldi, S., Nardi, F., Piscopia, R., Petroselli, A. & Apollonio, C. 2021 Continuous hydrologic modelling for design simulation in small and ungauged basins: a step forward and some tests for its practical use. *Journal of Hydrology* **595**. <http://doi.org/10.1016/j.jhydrol.2020.125664>.
- Iwański, S. & Kuchar, L. 2003 Spatial generation of daily meteorological data. *Acta Scientiarum Polonorum Formatio Circumiectus* **2** (1), 113–121.
- Kaźmierczak, B. & Kotowski, A. 2012 A depth–duration–frequency rainfall model for dimensioning and modelling of Wrocław drainage systems. *Environment Protection Engineering* **38**, 127–138.
- Kaźmierczak, B. & Kotowski, A. 2015 The suitability assessment of a generalized exponential distribution for the description of maximum precipitation amounts. *Journal of Hydrology* **525**, 345–351.
- Kim, S. & Kim, H. 2016 A new metric of absolute percentage error for intermittent demand forecast. *International Journal of Forecasting* **32**, 669–679.
- Kottegoda, N. T. & Rosso, R. 2008 *Applied Statistics for Civil and Environmental Engineers*. Wiley-Blackwell, pp. 736.
- Kubiak-Wójcicka, K. 2020 Variability of air temperature, precipitation and outflows in the Vistula Basin (Poland). *Resources* **9**, 1–26.
- Mannina, G., Freni, G., Viviani, G., Saegrov, S. & Hafskjold, L. S. 2006 Integrated urban water modeling with uncertainty analysis. *Water Science and Technology* **54** (6–7), 379–386.
- Maraun, D. 2016 Bias correcting climate change simulations – a critical review. *Current Climate Change Reports* **2** (4), 211–220.
- Młyński, D. 2020 Analysis of problems related to the calculation of flood frequency using rainfall-runoff models: a case study in Poland. *Sustainability* **12** (17), 1–18.
- Młyński, D., Cebulska, M. & Wałęga, A. 2018 Trends, variability, and seasonality of maximum annual daily precipitation in the Upper Vistula Basin, Poland. *Atmosphere* **2018** (9), 1–14.
- Młyński, D., Wałęga, A., Bugajski, P., Operacz, A. & Kurek, K. 2019a Verification of empirical formulas for calculating mean low flow in reflect to affecting on disposable water resources. *Acta Scientiarum Polonorum Formatio Circumiectus* **18** (2), 83–92.
- Młyński, D., Wałęga, A., Petroselli, A., Tauro, F. & Cebulska, M. 2019b Estimating maximum daily precipitation in the Upper Vistula Basin, Poland. *Atmosphere* **10**, 1–17.
- Młyński, D., Wałęga, A., Ozga-Zielinski, B., Ciupak, M. & Petroselli, A. 2020 New approach for determining the quantiles of maximum annual flows in ungauged catchments using the EBA4SUB model. *Journal of Hydrology* **589**, 125198.
- Morrissey, M. L. 2009 Superposition of the Neyman-Scott rectangular pulses model and the Poisson white noise model for the representation of tropical rain rates. *Journal of Hydrometeorology* **10** (2), 395–412.
- Niedźwiedz, T., Łupikasza, E., Pińskwar, I., Kundzewicz, Z. W., Stoffel, M. & Malarzewski, L. 2014 Climatological background of floods at the northern foothills of the Tatra Mountains. *Theoretical and Applied Climatology* **119**, 273–284.
- Petroselli, A., Asgharina, S., Sabzevari, T. & Saghafian, B. 2020 Comparison of design peak flow estimation methods for ungauged basins in Iran. *Hydrological Sciences Journal* **65** (1), 127–137.
- Pierzga, K. & Jarosińska, E. 2014 Annual maximum daily rainfall generator. In: *Hydrology in Environment Protection and Development* (Magnuszewski, A., ed.). PAN, Warsaw, pp. 57–66.
- Pińskwar, I., Choryński, A., Graczyk, D. & Kundzewicz, Z. W. 2019 Observed changes in extreme precipitation in Poland: 1991–2015 versus 1961–1990. *Theoretical and Applied Climatology* **135**, 773–787.

- Pociask-Karteczka, J. 2016 The Upper Vistula Basin – a geographical overview. In: *Flood Risk in the Upper Vistula Basin* (Kundzewicz, Z., Stoffel, M., Niedźwiedz, T. & Wyzga, B., eds). Springer, Cham.
- Rodriguez-Iturbe, I., Cox, D. R. & Isham, V. 1987 Some models for rainfall based on stochastic point processes. *Proceedings of the Royal Society A: Mathematical, Physical* **410**, 269–288.
- Sabzevari, T. 2017 [Runoff prediction in ungauged catchments using the gamma dimensionless time-area method](#). *Arabian Journal of Geosciences* **10** (6), 1–11.
- Sabzevari, T., Ardakanian, R., Shamsaee, A. & Talebi, A. 2009 Estimation of flood hydrograph in no statistical watersheds using HEC-HMS model and GIS (Case study: Kasilian watershed). *Journal of Water Process Engineering* **4**, 1–11.
- Siddique-E-Akbor, A. H. M., Hossain, F., Sikder, S., Shum, C. K., Tseng, S., Yi, Y., Turk, F. J. & Limaye, A. 2014 [Satellite precipitation data-driven hydrological modeling for water resources management in the Ganges, Brahmaputra, and Meghna Basins](#). *Earth Interactions* **18**, 1–25.
- Sridhar, S. I. & Raviraj, A. 2017 [Statistical trend analysis of rainfall in Amaravathi river basin using Mann-Kendall test](#). *Current World Environment* **12** (1), 89–96.
- Turnipseed, D. P. & Sauer, V. B. 2010 *U.S. Geological Survey Techniques and Methods*, Vol. A8. USGS, Ch. Discharge Measurements at Gaging Stations, 87.
- Villarini, G., Smith, J. A., Baeck, M. L., Vitolo, R., Stephenson, D. B. & Krajewski, W. F. 2011 [On the frequency of heavy rainfall for the Midwest of the United States](#). *Journal of Hydrology* **400**, 103–120.
- Wałęga, A., Młyński, D., Bogdał, A. & Kowalik, T. 2016 [Analysis of the course and frequency of high water stages in selected catchments of the Upper Vistula Basin in the South of Poland](#). *Water* **8**, 1–15.
- Wasko, C. & Sharma, A. 2017 [Continuous rainfall generation for a warmer climate using observed temperature sensitivities](#). *Journal of Hydrology* **544**, 575–590.
- Willems, P., Arnbjerg-Nielsen, K., Olsson, J. & Nguyen, V. T. V. 2012 [Climate change impact assessment on urban rainfall extremes and urban drainage: methods and shortcomings](#). *Atmospheric Research* **103**, 106–118.
- Wright, D. B., Smith, J. A., Villarini, G. & Baeck, M. L. 2013 [Estimating the frequency of extreme rainfall using weather radar and stochastic storm transposition](#). *Journal of Hydrology* **488**, 150–165.
- Zhao, Y., Nearing, M. A. & Guertin, D. P. 2019 [A daily spatially explicit stochastic rainfall generator for a semi-arid climate](#). *Journal of Hydrology* **574**, 181–192.

First received 20 September 2021; accepted in revised form 21 March 2022. Available online 6 April 2022