­A hydrological calibration procedure for stochastic rainfall model

# Abstract

# Introduction

Stochastic rainfall modelling involves the generation of sequences of rainfall at a scale of interest (e.g., sub-daily, daily, monthly, annual, multi-annual) that is statistically similar to observed rainfall timeseries, typically measured at rain gauges. Simulated rainfall is a primary input to hydrological model, for simulating streamflow. The simulated streamflow is then used to assess risks such as floods and droughts. Stochastic rainfall models (SRMs) are commonly fitted to observed rainfall data. However, when simulated rainfall data are used as input for hydrological models, it is possible that the simulated rainfall will not be translated to realisations of streamflow that is statistically similar to the observed one due to, for example, limited understanding of the rainfall-runoff process or the incapability of SRMs to capture important rainfall attributes. This paper introduces a new hydrological calibration procedure that allows SRMs to help capture streamflow statistics that cannot be achieved with common calibration approach (fitting with observed rainfall statistics) when SRMs are used to generate rainfall input for continuous hydrological simulation.

In catchment hydrology, the term continuous simulation is the simulation of the wet and dry condition of a catchment by estimating the loss in rainfall and generating streamflow at daily, hourly, or sub-hourly time scales. Generally, the process of continuous simulation requires a rainfall-runoff model with sequences of meteorological inputs such as rainfall, evapotranspiration or temperature depending on the specification of the model ([Beven, 2012](#_ENREF_5)). Continuous simulation for catchment process has an advantage over the popular event-based approach by eliminating the assumptions of initial and continuing losses. Continuous simulation also allows the generation of long sequences of streamflow from which important flood or drought statistics can be extracted for the establishment of appropriate mitigation strategies, early warning, or long-term projection ([Linsley and Crawford, 1974](#_ENREF_26), [Boughton and Hill, 1997](#_ENREF_8), [Blazkova and Beven, 2002](#_ENREF_6), [Lamb, 2005](#_ENREF_24), [Viviroli et al., 2009](#_ENREF_39)).

The type of rainfall-runoff model component is selected depending on different purposes of the hydrological assessment. Based on the spatial resolution, for example, rainfall runoff models can be categorised into lumped, semi-distributed and distributed models. There have been a numerous developments and reviews of lumped rainfall runoff models ([Burnash et al., 1973](#_ENREF_9), [Perrin et al., 2003a](#_ENREF_29), [Boughton, 2004](#_ENREF_7), [Croke et al., 2006](#_ENREF_13), [Chiew, 2010](#_ENREF_10)) as well as semi-distributed and distributed rainfall-runoff model [ref]([Knightes, 2017](#_ENREF_23)). Different rainfall-runoff models have their own strengths and weaknesses, however by definition, they are all simplifications of the catchment characteristics and processes such as soil properties, topography, vegetation, water and energy balance ([Beven, 2012](#_ENREF_5), [Gupta et al., 2005](#_ENREF_20)). Therefore, uncertainties and assumptions accompanied with rainfall-runoff models should be accounted to ensure the reliability of the predictions from continuous simulation.

In terms of rainfall time series, a primary input to rainfall-runoff model, can be (1) the observed rainfall data collected at a rain-gauge or a network of rain-gauges or (2) generated from a rainfall model. There has been a substantial amount of study on rainfall model development to capture both the temporal pattern including daily ([Richardson and Wright, 1984](#_ENREF_33), [Semenov and Barrow, 1997](#_ENREF_34), [Sharma and Lall, 1999](#_ENREF_35), [Srikanthan and McMahon, 2001](#_ENREF_36)), sub-daily ([Gupta and Waymire, 1993](#_ENREF_21), [Khaliq and Cunnane, 1996](#_ENREF_22), [Cowpertwait, 2006](#_ENREF_11)), year to year variation ([Thyer and Kuzera, 1999](#_ENREF_38), [Srikanthan and Pegram, 2009](#_ENREF_37)), as well as the spatial pattern ([Wilks, 1998](#_ENREF_40), [Qian et al., 2002](#_ENREF_31), [Leonard et al., 2008](#_ENREF_25), [Rasmussen, 2013](#_ENREF_32), [Baxevani and Lennartsson, 2015](#_ENREF_2), [Bennett et al., 2018](#_ENREF_3), [Evin et al., 2018](#_ENREF_16)) of the observed rainfall. However, one of many challenges to the continuous simulation approach is to assess the impact of uncertainties in rainfall data on rainfall-runoff models response. Poor understanding of the effect of rainfall variability or uncertainties in rainfall data on the output of rainfall-runoff models will hamper the capability of continuous simulation in providing reliable hydrological assessment ([Michaud and Sorooshian, 1994](#_ENREF_27), [Faurès et al., 1995](#_ENREF_17), [Andréassian et al., 2001](#_ENREF_1), [Cristiano et al., 2017](#_ENREF_12)). The sensitivity of rainfall-runoff models to the spatial variability of rainfall input has become an attention to researchers. [Wilson et al. (1979)](#_ENREF_41) found that poor spatial representation of rainfall input causes significant errors to the estimated hydrograph through an experiment on a small-sized catchment (68.6 km2). Similarly, [Michaud and Sorooshian (1994)](#_ENREF_27) examined the effect of rainfall sampling errors on peak-flow estimations on a 150 km2 semi-arid catchment and concluded that poor spatial representation of rainfall (inadequate rain-gauges network in point sampling) accounts for 58% underestimation in observed peak-flow. However, [Nicótina et al. (2008)](#_ENREF_28) pointed out mixed conclusions in the effect of rainfall spatial variability on rainfall-runoff models response from multiple mesoscale catchments, where some catchments tend to dampen the effect of spatial distributions while others tend to amplify it and concluded that the sensitivity of runoff response to the spatial distribution of rainfall is also affected by catchment characteristics such as hillslope and routing time distribution.

Sources of simulated streamflow deficiencies have also been reported to be caused by the deficiencies of simulated rainfall in different time periods. In particular, poor simulated streamflow within a month could be the result of rainfall deficiencies in the concurrent month, the preceding month, or a contiguous block of month ([Bennett et al., 2019](#_ENREF_4)). It was also found that in some cases, “good” simulated rainfall can create “poor” streamflow estimations while “poor” simulated rainfall can create “good” streamflow estimations ([Bennett et al., 2019](#_ENREF_4)). The case of “good” rainfall - “poor” streamflow has also been reported in ([Gao et al., 2020](#_ENREF_18)), in which a single site SRM was developed and was shown to capture well a range of rainfall statistics including the wet and dry spell distributions, the lower and the upper tails. However, when the model is used to generate rainfall input for a rainfall-runoff model, the streamflow estimation was shown to have relatively large underestimation in the high flow ranges (from the 90th to the 99th percentiles, with increasing magnitude of underestimation). It is evident that a SRM should not only be able to preserve observed rainfall attributes but also be able to capture observed streamflow characteristics for practical hydrological application. Hitherto, studies on SRMs development generally considered observed rainfall statistics as benchmarks without any consideration on streamflow statistics. Hence, the aim of this paper is to develop a new approach to SRMs development that considers both rainfall and streamflow statistics.

The three key objectives of this paper are:

1. Present the hydrological calibration procedure for SRMs.
2. Demonstrate scenarios where simulated rainfall categorised as “good” translated to simulated runoff categorised as “bad”. A range of catchments within Australia are examined with the use of a SRM and lumped conceptual rainfall runoff model.
3. Demonstrate the improvements that the hydrological calibration procedure provided to the SRM through examining different rainfall and runoff statistics.

This paper is structured as follows. Sect. 2 introduces the hydrological calibration procedure for SRMs. Sect. 3 presents the case studies that are used to demonstrate the hydrological calibration procedure. Results, discussion, and conclusion are in sect. 4, sect. 5, sect. 6 respectively.

# Hydrological calibration of SRMs

## Overview

A typical calibration procedure for SRMs involves computing some observed rainfall statistics which are then used as model parameter to generate simulated rainfall. Fig. 1 illustrates the general process of SRMs development.

Diagram

Description automatically generated

Figure 1. Schematic of SRMs development with observed rainfall data.

For example, the popular WGEN models ([Richardson and Wright, 1984](#_ENREF_33)) has a daily rainfall generator component that requires 4 rainfall statistics: the probability of dry-wet event (PDW), the probability of wet-wet event (PWW), the shape (α) and rate (β) parameters of the gamma distribution. The first two probabilistic parameters PDW and PWW control the wet/dry pattern of the rainfall using a 1st order Markov chain while parameters α and β controls the amount of rainfall occurs on wet days. These parameters can be computed directly from the observed rainfall and used for simulating rainfall estimates.

Calibrating with observed rainfall data allows SRMs to preserve the observed rainfall statistics. The simulated rainfall time series can be used as input for hydrological models to produce streamflow time series for hydrological assessment. However, it is not necessarily given that simulated rainfall time series will translate to streamflow time series that preserve the properties of observed streamflow data. Therefore, another approach is to calibrate model parameters with streamflow statistics. Fig. 2 illustrates a schematic of the hydrological calibration procedure for SRMs. This approach utilises an optimization process based on targeted streamflow attributes to identify best suited set of SRM parameters. The use of optimization loop for SRM has also been used in previous studies to perturb rainfall data for scenarios-neutral climate impact assessment ([Guo et al., 2018](#_ENREF_19), [Culley et al., 2019](#_ENREF_14))

Diagram

Description automatically generated

Figure 2. Schematic of the hydrological calibration for SRMs with observed runoff

In comparison to the typical procedure of SRMs development, the hydrological calibration of SRMs requires an additional rainfall-runoff model component to simulate runoff estimates and observed streamflow data. The proposed steps for the hydrological calibration procedure are as follows.

1. Generating a sequence of simulated rainfall from the SRM with a pre-defined set of parameters.
2. Generating a sequence of simulated runoff using the rainfall runoff model with the simulated rainfall as input.
3. Computing a pair of statistics from the simulated runoff and the observed runoff (e.g., the flow duration curve).
4. Computing an objective function with the simulated and observed runoff statistics (e.g., the sum of squared errors)
5. Finding the SRM parameters set that optimizes the objective function.

From Fig. 2, the simulated runoff statistic is compared against the observed runoff statistic. This comparison could affect the performance of the hydrological calibration procedure due to potential errors in the measurement of the observed runoff. To overcome this issue, the observed runoff time series is replaced by a virtual observed runoff time series. The virtual observed runoff is generated by inputting observed rainfall to the same rainfall-runoff model in previous step. Fig. 3 illustrates the hydrological calibration procedure with virtual observed runoff.

Diagram

Description automatically generated

Figure 3. Schematic of the hydrological calibration for SRMs with virtual observed runoff.

To further illustrate the hydrological calibration procedure, an example with the WGEN model ([Richardson and Wright, 1984](#_ENREF_33)) and the GR4J rainfall-runoff model ([Perrin et al., 2003b](#_ENREF_30)) that are used for this study is presented in the following sub-sections.

## Step 1 – rainfall simulation

The WGEN mentioned above is used to generate simulated rainfall time series. To initiate the procedure, the 4 monthly model parameters (i.e., 4 parameters for each of the 12 month) are computed from the observed rainfall data. These values will be used as the starting point for the optimization process.

The occurrence parameters PDW and P­WW are computed using the following equations:

Where nDW is the number of wet days given a dry occurred previously, nDD is the number of dry days given a dry day occurred previously, nWW is the number of wet days given a wet day occurred previously and nWD is the number of dry days given a wet day previously.

The parameters α and β is computed by simply using the method of moment (MoM):

Where μ is the mean of wet days amount and ν is the variance of wet days amount.

In total, there will be 48 parameters computed as initial values for the optimization. The bounds of the parameters are tabulated in Table 1

Table 1. Bounds of the 4 WGEN model parameters

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Parameter | PDW | P­WW | α | β |
| Lower bound | 0 | 0 |  |  |
| Upper bound | 1 | 1 | Inf | Inf |

## Step 2 – streamflow simulation

The GR4J model is used to generate both simulated streamflow and virtual observed streamflow. The model is calibrated with observed runoff data which yields 4 parameters. These 4 parameters are fixed throughout the whole process.

## Step 3 – Runoff statistic

From the simulated and virtual observed runoff, the targeted statistic such as the flow duration curve is computed.

## Step 4 – Objective function

## Step 5 - Optimization

The Shuffle Complex Evolution (SCE) optimization algorithm ([Duan et al., 1992](#_ENREF_15)) is used to find the best set of rainfall model parameters. Given that the WGEN model has in total 48 parameters, the search space will become relatively large which could affect the efficiency of the optimization. Therefore, to reduce the size of the parameter space, the optimization process can be performed on a monthly basis. In this way, the parameters set for each optimization loop is reduced to 4 parameters only. Then the optimization can be run 12 time to obtain the parameters set for each month.

# Case study

The hydrologic reference stations database (http://www.bom.gov.au/water/hrs/) is used for the selection of case studies in this project. The database provides streamflow records from 467 stations across Australia that are known to be of a high quality. Thirteen stations have been selected based on their distribution on contrasting hydro-climate regions as potential case studies for this project

Map

Description automatically generated

Figure 4. \*An example of potential figure for the paper

Table 2. Identified stations for case study

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| ID | Lat | Long | Jurisdiction | Area (km2) | Record length |
| Cotter River at Gingera | -35.59 | 148.82 | ACT | 130 | 1963-2019 |
| Corang River at Hockeys | -35.15 | 150.03 | NSW | 165.6 | 1950-2019 |
| Apsley River at Apsley Falls | -31.05 | 151.77 | NSW | 851.9 | 1960-2019 |
| Bremer River at Walloon | -27.6 | 152.69 | QLD | 628.1 | 19611003 |
| Barambah Creek at Litzows | -26.3 | 152.04 | QLD | 646.6 | 19641002 |
| Daintree River at Bairds | -16.18 | 145.28 | QLD | 907.3 | 19680927 |
| Elliot River at Guthalungra | -19.94 | 147.84 | QLD | 279.6 | 19730314 |
| Goulburn River at Dohertys | -37.33 | 146.13 | VIC | 700.2 | 19671214 |
| Wannon River at Dunkeld | -37.63 | 142.34 | VIC | 384.9 | 19701015 |
| Happy Valley Creek at Rosewhite | -36.58 | 146.82 | VIC | 138 | 19610630 |
| Leven River at Bannons Bridge | -41.25 | 146.09 | TAS | 499.3 | 19630620 |
| Scott Creek at Scott Bottom | -35.1 | 138.67 | SA | 29 | 19690329 |
| Harvey River at Dingo Road | -33.09 | 116.04 | WA | 148 | 19700320 |

* Example of applying the procedure with WGEN and GR4J (Section 2.1 to 2.6)

# Results

# Discussion

# Limitations and opportunities

# Conclusions

* Limitation and future opportunities:
* Rainfall runoff models
* Objective function
* Feasibility – runtime
* Other Rainfall model (extend WGEN to capture seasonal pattern - harmonic function)

*Data availability.* All the data used in this study can be requested by contacting the corresponding author Thien Nguyen at truonghuythien.nguyen@adelaide.edu.au.

*Author contributions.*

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

*Acknowledgements.*

# Reference

ANDRÉASSIAN, V., PERRIN, C., MICHEL, C., USART-SANCHEZ, I. & LAVABRE, J. 2001. Impact of imperfect rainfall knowledge on the efficiency and the parameters of watershed models. *Journal of Hydrology,* 250**,** 206-223.

BAXEVANI, A. & LENNARTSSON, J. 2015. A spatiotemporal precipitation generator based on a censored latent Gaussian field: SPATIOTEMPORAL STOCHASTIC GENERATOR. *Water resources research,* 51**,** 4338-4358.

BENNETT, B., THYER, M., LEONARD, M., LAMBERT, M. & BATES, B. 2018. A comprehensive and systematic evaluation framework for a parsimonious daily rainfall field model. *Journal of Hydrology,* 556**,** 1123-1138.

BENNETT, B., THYER, M., LEONARD, M., LAMBERT, M. & BATES, B. 2019. A virtual hydrological framework for evaluation of stochastic rainfall models. *Hydrology and Earth System Sciences,* 23**,** 4783-4801.

BEVEN, K. J. 2012. *Rainfall-runoff modelling : the primer,* Chichester, West Sussex ;, Wiley-Blackwell.

BLAZKOVA, S. & BEVEN, K. 2002. Flood frequency estimation by continuous simulation for a catchment treated as ungauged (with uncertainty). *Water resources research,* 38**,** 14-1-14-14.

BOUGHTON, W. 2004. The Australian water balance model. *Environmental modelling & software : with environment data news,* 19**,** 943-956.

BOUGHTON, W. & HILL, P. 1997. A Design Flood Estimation Procedure Using Data Generation And A Daily Water Balance Model. Cooperative Research Centre For Catchment Hydrology.

BURNASH, R. J. C., FERRAL, R. L. & MCGUIRE, R. A. 1973. *A Generalized Streamflow Simulation System: Conceptual Modeling for Digital Computers*, U.S. Department of Commerce, National Weather Service, and State of California, Department of Water Resources.

CHIEW, F. H. S. 2010. Lumped Conceptual Rainfall-Runoff Models and Simple Water Balance Methods: Overview and Applications in Ungauged and Data Limited Regions. *Geography compass,* 4**,** 206-225.

COWPERTWAIT, P. S. P. 2006. A spatial–temporal point process model of rainfall for the Thames catchment, UK. *Journal of hydrology (Amsterdam),* 330**,** 586-595.

CRISTIANO, E., TEN VELDHUIS, M.-C. & VAN DE GIESEN, N. 2017. Spatial and temporal variability of rainfall and their effects on hydrological response in urban areas – a review. *Hydrology and Earth System Sciences,* 21**,** 3859-3878.

CROKE, B. F. W., ANDREWS, F. & JAKEMAN, A. J. 2006. IHACRES Classic Plus: A redesign of the IHACRES rainfall-runoff model. *Environmental Modelling and Software,* 21**,** 426-427.

CULLEY, S., BENNETT, B., WESTRA, S. & MAIER, H. R. 2019. Generating realistic perturbed hydrometeorological time series to inform scenario-neutral climate impact assessments. *Journal of Hydrology,* 576**,** 111-122.

DUAN, Q., SOROOSHIAN, S. & GUPTA, V. 1992. Effective and efficient global optimization for conceptual rainfall‐runoff models. *Water resources research,* 28**,** 1015-1031.

EVIN, G., FAVRE, A.-C. & HINGRAY, B. 2018. Stochastic generation of multi-site daily precipitation focusing on extreme events. *Hydrology and Earth System Sciences,* 22**,** 655-672.

FAURÈS, J.-M., GOODRICH, D. C., WOOLHISER, D. A. & SOROOSHIAN, S. 1995. Impact of small-scale spatial rainfall variability on runoff modeling. *Journal of Hydrology,* 173**,** 309-326.

GAO, C., BOOIJ, M. J. & XU, Y.-P. 2020. Development and hydrometeorological evaluation of a new stochastic daily rainfall model: Coupling Markov chain with rainfall event model. *Journal of hydrology (Amsterdam),* 589**,** 125337.

GUO, D., WESTRA, S. & MAIER, H. R. 2018. An inverse approach to perturb historical rainfall data for scenario-neutral climate impact studies. *Journal of Hydrology,* 556**,** 877-890.

GUPTA, H. V., BEVEN, K. J. & WAGENER, T. 2005. Model Calibration and Uncertainty Estimation. *Encyclopedia of Hydrological Sciences.*

GUPTA, V. K. & WAYMIRE, E. C. 1993. A Statistical Analysis of Mesoscale Rainfall as a Random Cascade. *Journal of Applied Meteorology,* 32**,** 251-267.

KHALIQ, M. N. & CUNNANE, C. 1996. Modelling point rainfall occurrences with the modified Bartlett-Lewis rectangular pulses model. *Journal of Hydrology,* 180**,** 109-138.

KNIGHTES, C. 2017. An overview of rainfall-runoff model types. USA: The U.S. Environmental Protection Agency.

LAMB, R. O. B. 2005. Rainfall-Runoff Modeling for Flood Frequency Estimation. Chichester, UK: John Wiley & Sons, Ltd.

LEONARD, M., LAMBERT, M. F., METCALFE, A. V. & COWPERTWAIT, P. S. P. 2008. A space-time Neyman–Scott rainfall model with defined storm extent. *Water resources research,* 44**,** W09402-n/a.

LINSLEY, R. & CRAWFORD, N. 1974. Continuous simulation models in urban hydrology. *Geophysical research letters,* 1**,** 59-62.

MICHAUD, J. D. & SOROOSHIAN, S. 1994. Effect of rainfall-sampling errors on simulations of desert flash floods. *Water Resources Research,* 30**,** 2765-2775.

NICÓTINA, L., CELEGON, E. A., RINALDO, A. & MARANI, M. 2008. On the impact of rainfall patterns on the hydrologic response. *Water Resources Research,* 44.

PERRIN, C., MICHEL, C. & ANDRÉASSIAN, V. 2003a. Improvement of a parsimonious model for streamflow simulation. *Journal of Hydrology,* 279**,** 275-289.

PERRIN, C., MICHEL, C. & ANDRÉASSIAN, V. 2003b. Improvement of a parsimonious model for streamflow simulation. *Journal of hydrology (Amsterdam),* 279**,** 275-289.

QIAN, B., CORTE‐REAL, J. & XU, H. 2002. Multisite stochastic weather models for impact studies. *International journal of climatology,* 22**,** 1377-1397.

RASMUSSEN, P. F. 2013. Multisite precipitation generation using a latent autoregressive model. *Water Resources Research,* 49**,** 1845-1857.

RICHARDSON, C. W. & WRIGHT, D. A. 1984. WGEN: a model for generating daily weather variables. U.S Department of Agriculture. *Agriculture Research Service,* ARS-8**,** 83.

SEMENOV, M. A. & BARROW, E. M. 1997. Use of a Stochastic Weather Generator in the Development of Climate Change Scenarios. *Climatic Change,* 35**,** 397-414.

SHARMA, A. & LALL, U. 1999. A nonparametric approach for daily rainfall simulation. *Mathematics and Computers in Simulation,* 48**,** 361-371.

SRIKANTHAN, R. & MCMAHON, T. A. 2001. Stochastic generation of annual, monthly and daily climate data: A review. *Hydrology and earth system sciences,* 5**,** 653-670.

SRIKANTHAN, R. & PEGRAM, G. G. S. 2009. A nested multisite daily rainfall stochastic generation model. *Journal of hydrology (Amsterdam),* 371**,** 142-153.

THYER, M. & KUZERA, G. 1999. Modelling long-term persistence in rainfall time series: Sydney rainfall case study. *International Conference on Water Resources & Environment Research.* Brisbane, Qld.: Institution of Engineers, Australia.

VIVIROLI, D., MITTELBACH, H., GURTZ, J. & WEINGARTNER, R. 2009. Continuous simulation for flood estimation in ungauged mesoscale catchments of Switzerland – Part II: Parameter regionalisation and flood estimation results. *Journal of Hydrology,* 377**,** 208-225.

WILKS, D. S. 1998. Multisite generalization of a daily stochastic precipitation generation model. *Journal of hydrology (Amsterdam),* 210**,** 178-191.

WILSON, C. B., VALDES, J. B. & RODRIGUEZ-ITURBE, I. 1979. On the influence of the spatial distribution of rainfall on storm runoff. *Water Resources Research,* 15**,** 321-328.