Heliyon 9 (2023) e15740

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon

Research article

Rainfall-runoff modelling based on global climate model and tropical rainfall measuring mission (GCM -TRMM): A case study in Hulu Terengganu catchment, Malaysia

Wan Norsyuhada Che Wan Zanial [a], Marlinda Abdul Malek [b], Mohd Nadzri Md Reba [c], Nuratiah Zaini [a], Ali Najah Ahmed [a,*], Mohsen Sherif [d,e], Ahmed Elshafie [f]

[a] *Civil Engineering Department, College of Engineering, Universiti Tenaga Nasional, 43000 Kajang, Selangor, Malaysia*
[b] *Cataclysmic Management and Sustainable Development Research Group (CAMSDE), Department of Civil Engineering, Kulliyyah of Engineering, International Islamic University Malaysia, 53100, Selangor, Malaysia*
[c] *Geoscience and Digital Earth Centre (INSTeG), Research Institute for Sustainable Environment (RISE), Faculty of Built Environment & Surveying, Universiti Teknologi Malaysia, 81310, Johor Bahru, Johor, Malaysia*
[d] *Civil and Environmental Engineering Department, College of Engineering, United Arab Emirates University, P.O. Box 15551, Al Ain, United Arab Emirates*
[e] *National Water and Energy Center, United Arab Emirates University, P.O. Box 15551, Al Ain, United Arab Emirates*
[f] *Department of Civil Engineering, Faculty of Engineering, University of Malaya (UM), 50603 Kuala Lumpur, Malaysia*

ARTICLE INFO

Keywords:
Hydropower plant
GCM
SDSM
TRMM
Inflow

ABSTRACT

The hydropower Plant in Terengganu is one of the major hydroelectric dams currently operated in Malaysia. For better operating and scheduling, accurate modelling of natural inflow is vital for a hydroelectric dam. The rainfall-runoff model is among the most reliable models in predicting the inflow based on the rainfall events. Such a model's reliability depends entirely on the reliability and consistency of the rainfall events assessed. However, due to the hydropower plant's remote location, the cost associated with maintaining the installed rainfall stations became a burden. Therefore, the study aims to create a continuous set of rainfall data before, during, and after the construction of a hydropower plant and simulate a rainfall-runoff model for the area. It also examines the reliability of alternative methods by combining rainfall data from two sources: the general circulation model and tropical rainfall measuring mission. Rainfall data from ground stations and generated data using inverse distance weighted method will be compared. The statistical downscaling model will obtain regional rainfall from the general circulation model. The data will be divided into three stages to evaluate the accuracy of the models in capturing inflow changes. The results revealed that rainfall data from TRMM is more correlated to ground station data with R2 = 0.606, while SDSM data has R2 = 0.592. The proposed inflow model based on GCM-TRMM data showed higher precision compared to the model using ground station data. The proposed model consistently predicted inflow during three stages with R2 values ranging from 0.75 to 0.93.

* Corresponding author.
E-mail addresses: syuhadazanial@gmail.com (W.N. Che Wan Zanial), marlinda@iium.edu.my (M.A. Malek), nadzri@utm.my (M.N. Md Reba), Nur_Atiah@uniten.edu.my (N. Zaini), Mahfoodh@uniten.edu.my (A.N. Ahmed), msherif@uaeu.ac.ae (M. Sherif), elshafie@um.edu.my (A. Elshafie).

https://doi.org/10.1016/j.heliyon.2023.e15740
Received 17 January 2023; Received in revised form 19 April 2023; Accepted 20 April 2023
Available online 23 April 2023

1. Introduction

Most of the hydrological studies mainly used ground station data. Rainfall is one of the most challenging elements of the hydrological cycle to forecast [1]. In most locations, rain gauge stations are accessible as precipitation estimation gadgets. Although the results are generally acceptable, constraints such as no coverage at certain remote regions or with one or two rain gauges at a single enormous catchment area are common in developing countries such as Malaysia. General Circulation Models or GCM is one of example for numerical model that signify physical methods in the environment, sea, cry sphere and land surface, are among the innovative tools as of now available to simulate responses of worldwide climate framework. Due to its conjunction with nested regional model, only a more straightforward models used to give all-inclusive or found the middle value of assessments of atmosphere reaction and have the potential in sway investigation which to give geologically and truly predictable appraisals of worldwide environmental change [2].

Environmental change can affect water assets through changes in the hydrological cycle. Rainfall is one of the principal factors, caused from environmental change. To appraise future environmental change due to nonstop increment of ozone harming substance fixation in the air, Global Climate Models (GCMs) are utilized. Direct yield from GCMs cannot be used for hydrological evaluation because of its coarse spatial goals. Therefore, downscaling is used to turn the coarse spatial objectives of GCMs into an appropriate plan that can involve the production of station data for unique territories by using GCM environment yield factors [2–4].

Tukimat et al. [5] utilized the model of an integrated Statistical Downscaling Model and Geographic Information System (SDSM-GIS) to analyse the long-term precision of predicted ungauged station rainfall. It is appropriate to assess the precision of projected rainfall mapping at ungauged extent. The outcome of this study showed that the SDSM-GIS model has a high capacity for ungauged catchment to produce long-term rainfall patterns. In the simulated results, the SDSM technique was found to be effective in providing long-term climate trends at measured stations with a lower percentage of mean absolute error and a higher R^2 value of about 1.0.

Tahir et al. [6] contemplated the importance of precipitation by downscaling system utilizing SDSM at Limbang stream bowl. The examination explores model capabilities at tropical area, and anticipated the atmosphere inconstancy at nearby scale under a few emanation situations of RCP2.6, RCP4.5, and RCP8.5. The outcomes demonstrated that under RCP2.6 situation, there will be an expansion of 8.13%, while 14.7% ascent in RCP4.5 situation during time of year 2071–2100. A sudden increment of about 40.6% was seen under the solid situation of RCP8.5. In this way, it is reasoned that future example of precipitation at Limbang waterway catchment under all situations is consistently expanding because of environmental change. Discoveries of this examination may help, policymakers and individual experts for better arranging of water the board and seepage framework during change atmosphere in the future.

Singh et al. [7] analyzed the potential relevance of SDSM in day by day downscaling of Tmax, Tmin and precipitation in a piece of Sutlej bowl, to investigate feasibility of yields on third era of Canadian Coupled Global Climate Model (CGCM3) and Hadley Center Coupled Model, variant 3 (HadCM3) in downscaling of Tmax, Tmin and precipitation using SDSM. This examination likewise explored future changes on Tmax, Tmin and precipitation under various spread circumstances (A1B and A2 of CGCM3 and A2 and B2 of HadCM3) for the 21st century.

Tukimat & Harun [8] utilized SDSM model to create atmosphere designs with temperature, precipitation, wet and dry length for 30 years (year 2040–2069). Results indicated that future atmosphere design will even now be interrelated to the verifiable record, yet with more noteworthy extents.

Hassan & Harun [9] investigated the flexibility of SDSM for downscaling temperature and precipitation. During calibration and validation, the SDSM model showed a decent reproduction of monthly for precipitation and temperature.

Wilby et al. [10] defined a decision support tool to evaluate neighbourhood environmental change impacts utilizing a solid measurable downscaling practice. Genuine Downscaling Model (SDSM) underpins fast improvement of different, ease, single-site conditions of consistently surface environment factors under present and future local air driving. In addition, the development performed subordinate assignments of marker variable pre-screening, model arrangement, focal symptomatic testing, quantifiable evaluations and illustrating of condition information.

In previous study, statistical downscaling method (SDSM) was utilized to determine bigger scale rainfall data to a more sufficient scale through induction of cross-scale relationship with irregular or potentially deterministic capacities [11]. Measurable downscaling is used to accomplish data from environmental change at finer grids by improving direct statistical connections between massive scale climatic dissemination of rainfall and its neighbourhood factors. It creates quantitative relationships between massive scale barometrical factors (predictors) and nearby surface factors (predictands). In this study, programming coded in Visual Basic 6.0 of SDSM 4.2 was utilized [12].

Mohd Zad et al. [13], analyzed the exhibition of precipitation estimation from TRMM 3B42–V7 utilizing precipitation measure information in Malaysia, explicitly from the Pahang stream bowl and using a lot of execution marker. The findings indicate that the district's elevation affects the scoring displays. Root Mean Squared Error (RMSE) is lower at a higher height and mid-elevation for the most part. In the identification and volumetric correspondence between TRMM and downpour measures, TRMM demonstrates a mild show. The purpose of satellite recognitions is to be equipped for parameter considerations and the reinstitution of water-driven state components that better reflect the hydrological characteristics of a catchment. Streamflow is at last assessed by the model itself. It is entirely expected to fuse satellite perceptions of flood immersion zone, and water arranges data from radar altimetry or construed from water surface region and high-goals geography information, into hydrodynamic or flood directing model.

The processing and incorporation of the use of heterogeneous data from space and ground-based information sources has been used in past studies. This has resulted in high precision in forecasting. The flood monitoring and forecasting system has facilitated major social security changes and reduced economic damage caused by floods [14].

The study by Brakenridge et al. [15] found that a comparison of gauging station versus modelled discharge often showed that a small positive model bias with satellite-observed annual runoff errors observed is also positive and could be increased by eliminating bias from rating curves.

The parameters of rainfall-runoff models in such catchment cannot be obtained by merely calibrating runoff data for catchment with a lack of technical knowledge on its hydrological regime, limited ground data available, inadequate runoff data are reported available or very few ground rain gauges installed in a huge catchment; therefore, other alternative methods need to be acquired. Model parameters involved calibration are usually transposed from identical measured catchments in order to solve problems with incomplete inaccessible rainfall data or ungauged catchments. Physically dependent model parameters are typically inferred near the ungauged catchment of interest from other rainfall stations. In ungauged catchments, the key problem is the absence of local ground rainfall and runoff data to be used in data calibration.

In this study, the study area is located in a rural forested area of Hulu Terengganu. Based on data exploration conducted, minimal numbers of ground rain gauges were set-up in this area before and during the construction of the Hydropower plant. Due to this reason, this study has adopted Inverse Distance Weighted (IDW) technique in providing continuous Ground Rainfall Data before, during and after the construction of hydropower plant and the use of National Centers for Environmental Prediction (NCEP) data through GCM downscaled by SDSM and weather satellite data to produce a continuous rainfall in this study area is most desirable.

Inverse Distance Weightage (IDW) is one of the most well-known interpolation methods, which is a simple and intuitive deterministic interpolation. IDW is easy to be implemented and available in almost any GIS software, so it is applied frequently in various disciplines [16]. Interpolation of the Inverse Distance Weighted (IDW) is a careful methodology that upholds the condition that a point's rough worth is impacted more by well-established realities close by than by those farther away.

This study aims to produce a continuous set of rainfall data at the study area before, during, and after the construction of the Hydropower plant and to simulate a rainfall-runoff model at the study area.

2. Methodology

2.1. Study area

The hydropower plant in Terengganu is one of the major hydropower plants in Malaysia. It also serves as a multipurpose hydropower plant. This hydropower plant station is located in the district of Kuala Berang in Terengganu. The best of Malaysia, Hydropower plant is one of the primary sustainable power sources accessible.

This hydropower plant commands a catchment area of 2600 km^2 for before, during and after Hydropower plant construction (based on catchment and watershed delineation from DEM conducted). This hydropower plant is contributed by five (5) major river tributaries, namely Sg. Terengganu Sg. Cacing, Sg. Petang, Sg.Tembat and Sg. Petuang. In this study, rainfall is the input parameter used to develop the proposed hydrodynamic model at the chosen study area. Based on its strategic location and proximity to the hydropower

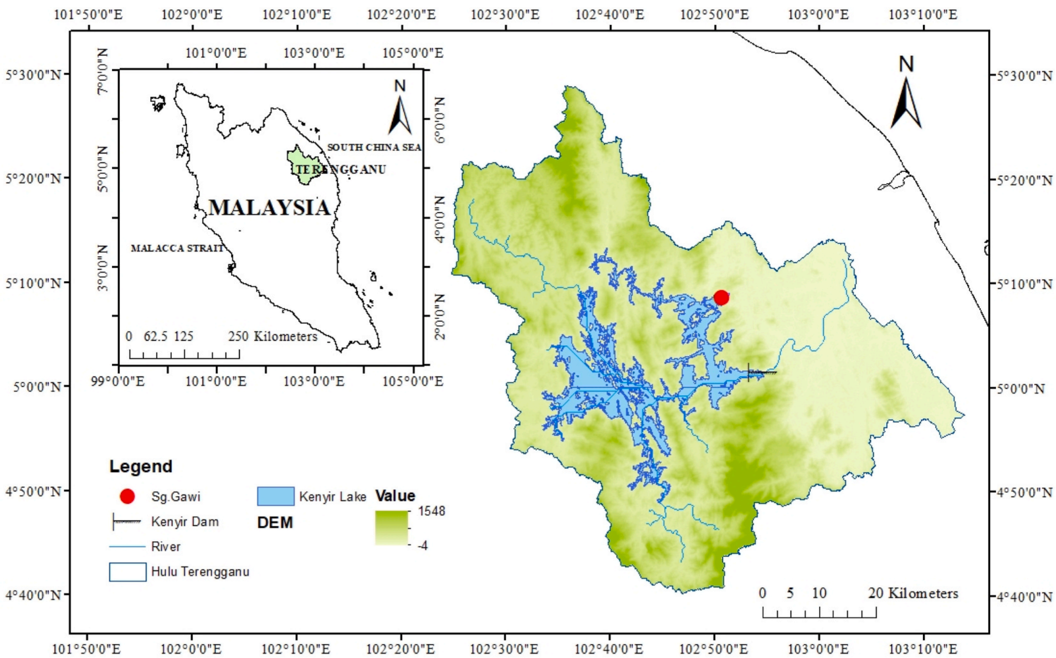

Fig. 1. Location of Study Area (this map was created with ArcGIS 10.8 software).

plant, Sg. Gawi (Station No: 5128001) ground rainfall station monitored by Department of Irrigation and Drainage Malaysia (DID), situated at Latitude of 5.143°N and Longitude of 102.84°E was chosen as the main ground rainfall station used in this study as shown in Fig. 1.

2.2. Data acquisition

Rainfall data used in this study are from the year 1971 until 2017. These data are divided into three (3) stages, namely before the construction of hydropower plant (1971–1979), during the construction of hydropower plant (1980–1986), and after the construction of hydropower plant (1987–2017). The reason analyses were conducted in these substages is to investigate the trend of river flow during these three different stages in order to detect any possibilities of impacts from the operation of the hydropower plant. The robust historical data used in this analysis would boost the outcomes of the rainfall-runoff model. After the construction of the hydropower plant, data were divided into two (2) parts, namely the year 1987–1997 and the year 1998–2017. This is because TRMM weather satellite data are available only after the year 1997.

Rainfall data used in this study are divided into two (2); namely, Ground Rainfall and Atmospheric Rainfall from General Circulation Model (GCM), downscaled using Statistical Downscaling Model (SDSM) and TRMM Satellite Rainfall.

The Tropical Rainfall Measuring Mission (TRMM) 3B42–V7 dataset was obtained from Giovanni Earth Data website, https://www.earthdata.nasa.gov/technology/giovanni. The ground rainfall data was obtained from DID Ground Rain Gauge stations.

In this study, satellite estimates were carried out at regular resolution to balance the time resolution of the data from the ground rain gauge, at daily intervals and 0.25° × 0.25° (approximately 27.8 km × 27.8 km) spatial resolution. Data were extracted within latitudes and longitude boundaries of Terengganu at 19 years duration of 1998–2017.

Simulated inflow data are used to validate rainfall-runoff simulation from General Circulation Model (GCM), downscaled using Statistical Downscaling Model (SDSM) and TRMM Satellite Rainfall (GCM-TRMM). The inflow data was simulated because there is no monitoring station within the surrounding area of the hydropower plant, therefore the data from rainfall stations which is in the same catchment was used to simulate the inflow of the river.

2.3. Rainfall data infilling methods

As shown in Fig. 2, most of the Ground Rain gauges were found to be located towards the east of Hulu Terengganu, away from the upstream of hydropower plant. Due to this reason and data availability, this study has adopted the Inverse Distance Weighted (IDW) technique in providing continuous Ground Rainfall Data before, during, and after the hydropower plant's construction. Equation (1) shows the general equation for IDW method.

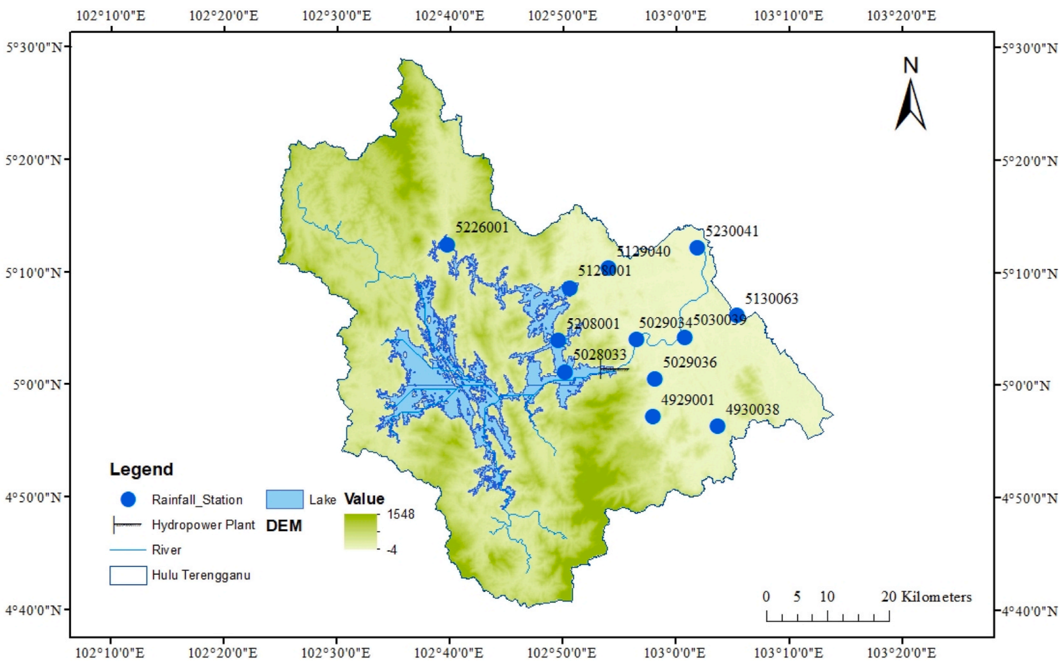

Fig. 2. Location of Ground Rain gauges used for IDW technique (this map was created with ArcGIS 10.8 software).

$$z_0 = \frac{\sum_{i=1}^{s} z_i \frac{1}{d_i^k}}{\sum_{i=1}^{s} \frac{1}{d_i^k}} \tag{1}$$

where z_0 at 0 point is the simulated value, z_i at the known point, i is the z value, d_i is the distance between point i, and point 0, s is the number of known points used in the calculation, and k is the stated power.

Power k governs the degree of local control. A power of 1.0 implies a constant rate of change between points in value (linear interpolation). A power of 2.0 or higher means that near a known point and levels away from it, the rate of change in values is higher. An important characteristic of IDW interpolation is that all predicted values are within the range of maximum and minimum values of the known points.

2.4. Statistical downscaling model (SDSM) rainfall data

In order to obtain the necessary refined rainfall values from General Circulation Model (GCM), Statistical Downscaling Model (SDSM) was utilized. In SDSM, seven (7) key capacities were performed to be specific Quality Control, Transform Variables, Screen Variables, Calibrate Model, Weather Generator, Scenario Generator and Compare Results. In the quality control exercise conducted in this study, all gross data errors and missing data will be identified. This is importance to ensure the quality of input data used.

The input data will then be transformed into logarithm, power, inverse, lag, binomial, and others depending on its suitability to input data used. Next, the predictors selection will be conducted based on statistical analyses such as monthly and partial correlations between predictand and predictor presented in the Screen Variables. The principal reason for Screen Variables activity is to help the client in choosing the fitting downscaling indicator factors. This is one of the most testing stages being developed of SDSM, since the selection of indicators exceptionally decides the character of downscaled atmosphere situation.

Screening potential helpful indicator predictand connections for model adjustment is an extremely significant stage being

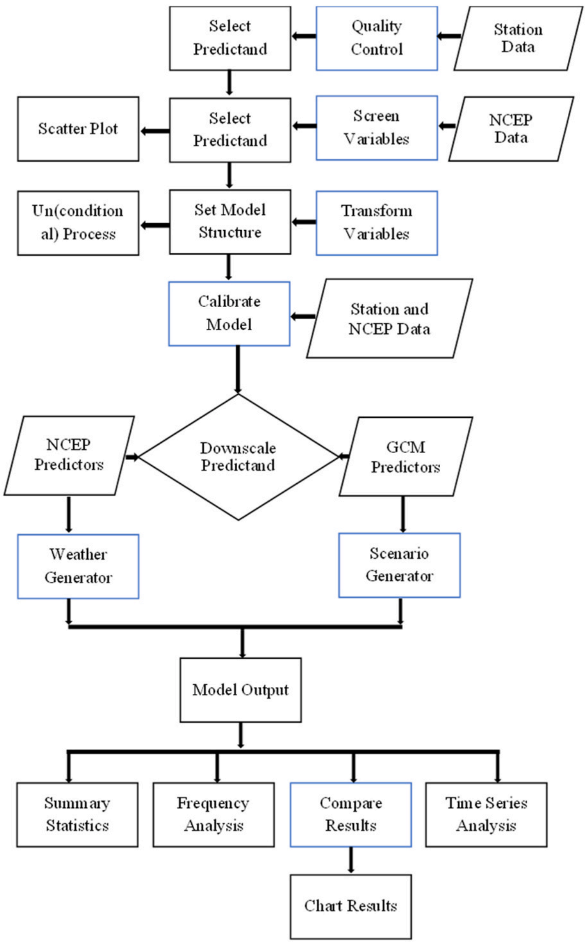

Fig. 3. Shows the flow chart of operating SDSM Version 4.2 analysis.

developed of SDSM. Choice of fitting indicator factors exceptionally decides the accomplishment of SDSM and the character of downscaled atmosphere situation [12]. Predictors were selected based on a combination of the correlation matrix, partial correlation and p-value.

After predictor selection, the predictor-predictand equation will then be tested at calibration and validation stages. The Calibrate Model activity clients a client indicated predictand alongside a lot of indicator factors. It figures the parameters of various relapse conditions utilizing streamlining calculation at either double simplex of conventional least squares. In the interim, Weather Generator operation generates ensembles of synthetic daily weather series given observed (or NCEP re–analysis) atmospheric predictor variables.

The technique empowers the confirmation of aligned models utilizing free information and the amalgamation of artificial time arrangement for current atmosphere conditions. Three choices for graphical investigation were given by SDSM 4.2 through Frequency Analysis, Compare Results, and Time Series Analysis screens. The Scenario Generator operation produces ensembles of synthetic daily weather series given atmospheric predictor variables supplied by a climate model (either for present or future climate experiments), rather than observed predictors. The input files for both the Weather Generator and Scenario Generator options will not be the same length as those used to obtain the model weights during the calibration phase.

There are two (2) kinds of sub-models in SDSM, namely unconditional and conditional used according to the requirement of predictands. For an independent variable such as temperature, the unconditional sub-model is used. For a conditional dependent variable such as precipitation, the conditional is used [10]. Projection for future temperature at regional scale are required to determine possible changes in evapotranspiration (ET) due to climate change. In this study, SDSM version 4.2 was used to downscale the precipitation (rainfall) in the study area over the years 1971–1997. Fig. 3 shows the flow chart of operating SDSM Version 4.2 analysis.

SDSM is a hybrid of multiple linear regression (MLR) and the stochastic weather generator (SWG). MLR establishes a statistical/empirical relationship between NCEP, large-scale variables, and local scale variables, and produces some regression parameters. These calibrated parameters, along with NCEP and GCM predictors, are then used by SWG to simulate up to 100 daily time series in order to create a better correlation with the observed time series [17].

In SDSM, the generation of climate parameters on the station scale is linearly conditioned by observed large-scale environmental predictors ($j = 1, 2, \ldots, n$). The downscaled method is either unconditional (as with the occurrence of rainy days) or conditional (as with wet days) (as with rainfall amounts). As shown in Equation (2), the frequency of wet day (W_i) on day I is linearly dependent on the vector n predictors of X_{ij}.

$$W_i = \alpha_0 + \sum_{j=1}^{n} \alpha_j X_{ij} \tag{2}$$

Under the requirement of $0 \le W_i \le 1$. Precipitation occurs at the point when the unvarying random number $r \le W_i$. The wet-day edge (mm) can fluctuate between areas, depending on the rainfall or the accuracy of the estimate. It is also critical to say that daily rainfall aggregates have not been established for many days, downplaying the rainfall frequencies, thus aligning the model.

In order to establish tirelessness of wetland droughts, indicator variables X_{ij} may be simultaneous or slowed. The precipitation of Pi is decreased at the point where a rainy day is restored, using Equation (3):

$$P_i^k = \beta_0 + \sum_{j=1}^{n} \beta_j X_{ij} + \varepsilon_i \tag{3}$$

where k (normally 0.25) is utilized to change every day. Nevertheless, other changes example, logarithm or inverse normal) may likewise be applied to P_i. Direct linear relationships are formed between the predictand U_i and the selected NCEP/NCAR predictors X_{ij} on individual sites such as Equation (4) in the case of unconditional processes such as daily temperature:

$$Ui = \gamma_0 + \sum_{j=1}^{n} \gamma_j X_{ij} + \varepsilon_i \tag{4}$$

The chosen NCEP/National Center for Atmospheric Research (NCAR) predictors on day i are where Ui is temperature on day i and X_{ij}. Regression coefficients evaluated for each month using the least-square relapse are alpha, βj and γj, and εi is model fault. These parameters are generated using stochastic arrangements of sequentially autonomous Gaussian numbers and are included in the regular schedule of deterministic segments [7].

2.5. TRMM Satellite Rainfall Data

The Tropical Rainfall Measuring Mission (TRMM) 3B42–V7 dataset was obtained from Giovanni Earth Data website. In this study, satellite estimates were carried out at regular resolution to balance the time resolution of the data from the ground rain gauge., at daily interval with $0.25^\circ \times 0.25^\circ$ (approximately 27.8 km × 27.8 km) spatial resolution. Data were extracted within latitudes and longitude boundaries of Sg. Gawi station at 19 years duration of 1998–2017.

2.6. Validation and calibration of SDSM rainfall and TRMM Satellite Rainfall

In this study, Ground Rainfall data from DID were used to validate SDSM Rainfall and TRMM Satellite Rainfall data. Rainfall from Statistical Downscaling Model (SDSM) rainfall were used from year 1971–1997 at 26 years duration.

The accuracy of TRMM Satellite data was validated at daily interval by comparing with the selected ground rain gauges at Sg. Gawi. The comparisons between TRMM Satellite Rainfall Data and Ground Rainfall Data were conducted with point-to-pixel approach. This method is chosen to prevent any potential errors and uncertainties during the field rain gauge interpolation cycle. Over the years 1998–2017, Gawi ground rain gages as well as compared with their respective grid points. Root Mean Square Error (RMSE) and Correlation Coefficient Squared (R^2), as presented in the Statistical Analysis subsection, are statistical tests used to validate the rainfall results.

2.7. Hydrological model development on rainfall – runoff

In this study, the Ground Rainfall and Atmospheric Rainfall namely, SDSM Rainfall and TRMM Satellite Rainfall are converted into runoff using Hydrological Procedure No. 27, Estimation of Design Flood Hydrograph based on Clark method for Rural Catchment in Peninsular Malaysia by Department of Irrigation and Drainage (DID), Malaysia.

a) Areal Reduction Factor

Precipitation is typically not equitably transmitted over an area for a storm case, the amount of precipitation decreases with good ways from the middle of the storm. In Peninsular Malaysia, enormous varieties can occur in short separations in precipitation sum, especially when storms ruled. HP1 was adopted in this analysis as Area Reduction Factors of Hydrological Procedure No.1 (1982).

b) Temporal Distribution

Numerous trend and peak discharge can be attained from temporal pattern. Temporal patterns of storm events have substantial effect on the computed peak values of river flow. DID Hydrological Procedure No.1 [18] defined a temporary distribution of the cumulative annual rainstorms over 1⁄2, 3, 6, 24 and 72 h in length. For this reason, nine (9) ground rainfall stations were selected, located in different parts of Peninsular Malaysia. The average time distributions over record years were determined. Temporary distributions of the Peninsular East and West Coasts. In this study, 24 h temporal pattern were used.

c) Rainfall – Runoff Relationships

The method used in Hydrological Procedure No. 11 (HP11) [19] was adopted in this study to determine the relationship between rainfall runoffs. The estimated cumulative volume of storm rainfall at a given flood event and direct runoff derived from the flood hydrograph is used to monitor the rainfall-runoff relationship.

In HP11, for catchments in Peninsular Malaysia, at year 1970–2000, 177 storms from 37 catchments out of 40 catchments were used to develop rainfall runoff relationship. Equation (5) and Equation (6) were then fitted to the observed data to represent larger floods analyzed. Equation (5) and Equation (6) were derived for catchments in Peninsular Malaysia. Where P is in mm of total rainfall, and Q is in mm of direct runoff.

$$Q = 0.33\,P, P < 75 \text{ mm} \tag{5}$$

$$Q = \frac{P^2}{P + 52}, P > 75 \text{ mm} \tag{6}$$

d) Time Distribution of Runoff

Unit Hydrograph is one of the methods to distribute runoff volume with time. Synthetic unit hydrograph (SUH) method was used to describe the entire unit hydrograph (UH) with only a few parameters for a gauged catchment. Hydrograph parameters are related to the properties of the catchment from which the parameters are derived. For this analysis the Clark Unit Hydrograph is used. Clark [20], stated that the interpretation of stream is portrayed when region bend. The time area twist shows the catchment area as a limited amount of time from the earliest starting point of actual precipitation, in addition to overflowing before the catchment outlet. Real precipitation is precipitation that are not lost through penetration or held ashore surface for example direct overflow. Clark [20], utilized straightforward direct repository where capacity is identified with inflow so as to portray reduction as shown in Equation 7, S is storage of catchment, R is coefficient of catchment storage and O is outflow from the catchment.

$$S = RO \tag{7}$$

e) Clark Parameter Determination

Estimation of Time of Concentration (Tc) and Catchment Storage Coefficient (R) for ungauged catchment used Equations (8) and (9). The catchment features of the study area are represented by these equations. In this study, numerous linear regression program named Cuanalo & Webster [21] was utilized to decide scientific connections of Tc and R with catchment qualities, for example, territory, slope and length of the standard for hydropower plant.

$$Tc = 2.32\ A^{-0.1188} L^{0.9573} S^{-0.5074} \tag{8}$$

$$R = 2.976\ A^{-0.1943} L^{0.9995} S^{-0.4588} \tag{9}$$

where A is the area of catchment in km^2, L is the main stream length in km and S is the main stream slope weighted in m/km.

f) Design Baseflow

To settle on the hydrograph of the total structure, a baseflow is needed. Before a big storm, verifiable baseflow characteristics can hardly be expected. Baseflows were captured under very dry and mildly humid catchment conditions. A better fit condition was induced for general use, as shown in Equation (10). Where the baseflow in m^3/s is Q_B and A in km^2 is the catchment area.

$$Q_B = 0.11 A^{0.85889} \tag{10}$$

g) Validation and Calibration of Generated River Flow

In this study, statistical analyses were conducted to evaluate the performance of generated River Flow from Atmospheric Rainfall (GCM-TRMM) compared to River flow station generated from Ground Rainfall.

2.8. Statistical analysis

In computational metrics, quantitative statistical tests using metrics such as Root Mean Square Error (RMSE), Determination Coefficient (R^2), and Relative Error percentage (RE%) are commonly accepted [1,22]. Equation (11), Equation (12), and Equation (13) shows the RMSE, R^2, and RE used to validate both rainfall and river flow in this study.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - X_i)^2} \tag{11}$$

$$R^2 = 1 - \frac{\sum(\widehat{x}_i - x_i)^2}{\sum(x_i - mean\widehat{x}_i)^2} \tag{12}$$

$$RE\% = \frac{(x_i - \widehat{x}_i)}{x_i} * 100 \tag{13}$$

where,
x_i is Ground Rainfall/River flow station generated from Ground Rainfall.
$\widehat{x}_i$ is Atmospheric Rainfall/generated River Flow from Atmospheric Rainfall
n is the number of observations.

3. Results and discussion

The ground rainfall data in this study obtained from 12 stations covering the study area of hydropower plant catchment from 1983 until 2017. Since this study covers the duration before constructing the dam (from 1971 to 1979), the inverse distance weighted (IDW) method was adopted to simulate the rainfall for the period from 1971 to 1983. To interpolate the rainfall during the first duration stage using IDW, the rainfall data from the 12 stations used for this purpose. The generated rainfall data and the measured rainfall data can be seen in Fig. 4. The generate rainfall depth shows similar fluctuations compared with the measured rainfall data except for the last

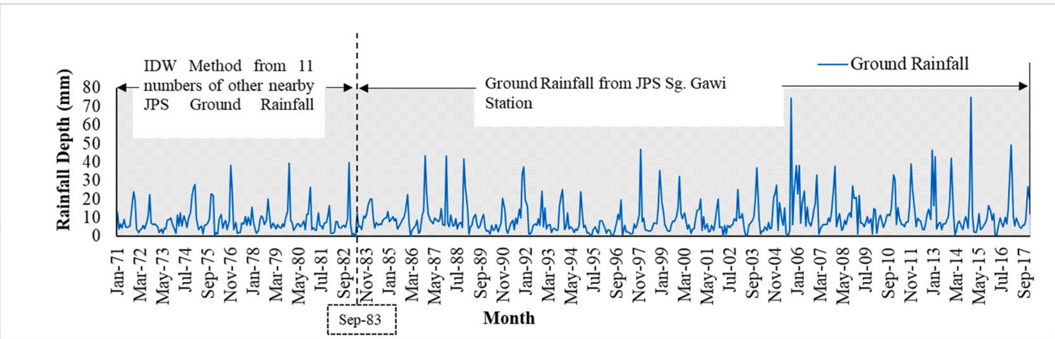

Fig. 4. Ground Rainfall data used in the study.

recent ten years where extreme rainfall events occurred which are considered as rare extreme events that happened at the study area [23].

After obtaining the required rainfall data for the entire duration of the three stages from ground stations and IDW method, the next step in to downscale the rainfall data from the global circulation model (GCM) for the same duration. Downscaling steps were carried out using the statistical downscaling model (SDSM) to acquire 26 years of rainfall data from 1971 to 1997. It can be observed in Fig. 5 that the rainfall depths for both grounds and downscaled data similar display similar patterns during each month, and the latter model capable of detecting the peak events of rainfall with an adequate level of precision.

For rainfall from 1998 to 2017, the estimated data obtained from the tropical rainfall measuring mission (TRMM) satellite. Fig. 6 shows that the monthly data of the rainfall from TRMM satellite and the ground stations. It can be shown that the calculation of TRMM rainfall displays a close range of values to the measured ground station rainfall and achieves encouraging precision.

The results in Table 1 reveal that the rainfall data obtained from TRMM is more correlated to the measured rainfall data from the ground stations where the value of R^2 are range from 0.5 to 0.8, while R^2 are range from 0.5 to 0.7 for SDSM rainfall data. The proposed inflow model based on GCM-TRMM rainfall data exhibits a high level of precision compared to the developed inflow model using rainfall data from ground stations.

Fig. 7 shows the rainfall for the entire duration during the three stages from both the ground rainfall, a combined rainfall data from the IDW model, and measured rainfall. In contrast, the atmospheric rainfall data is a combination data obtained from SDSM and TRMM satellite. The next step will be using the two sets of data to develop the rainfall-runoff model to predict the inflow during these three stages.

Finally, the two inflow models that were built can be seen in Fig. 8. The first inflow model is developed using the rainfall data from ground stations, while the second inflow model is developed by integrating the rainfall data from two sources GCM-TRMM. It can be seen that the proposed inflow model based on rainfall data from GCM-TRMM capable of capturing the changes of the developed inflow model using the rainfall data from ground stations. In addition to that, a significant increase in the predicted inflow rate from the developed model during stage three after constructing the dam compared with the first stage period. Such increases have been reported by previously conducted studies [24,25], indicating the robustness of the proposed model in predicting the pattern of the inflow during the entire duration of the three stages.

Summary of statistical analysis using R^2 and RMSE to validate the proposed inflow model using GCM-TRMM rainfall data is tabulated in Table 2. Based on these statistical indicators, it can be assumed that the model proposed in this study is can be used as a reliable predictive model where R^2 values are ranging from 0.749 to 0.933.

The lowest performance was found to be during the first stage duration, which is expected since, during this period, the obtained rainfall data was from SDSM model; despite that, the accuracy for this period still acceptable where R^2 is equal to 0.749. It is also noteworthy that during the following steps, the proposed model performs with a reliable degree of precision. Even after constructing the dam, the model able to capture the changes in the inflow rate with reasonable precision for the entire duration of the third stage. To conclude, this study's finding proved the inflow model's superiority when the used rainfall data obtained from satellite and can be used as an alternative to the measured rainfall data from the ground stations.

For more in-depth understanding for the performance of the proposed river inflow model, Relative Error distribution for the whole duration has been calculated and shown in Fig. 9. It could be observed from Fig. 9 that the proposed model could be successfully provide reliable accuracy for the inflow compared to the actual ones. It can be depicted that the maximum relative error obtained using the proposed model is over-estimate the actual value by 20% or under-estimate the actual value by −25%.

By considering that 20% or less relative error in estimating the river inflow is acceptable level of accuracy, it could be noticed that the model successfully provides accurate inflow estimation for 53 events out of 56 which is considered as remarkable performance.

Therefore, it can be concluded that GCM-TRMM rainfall data is a reliable source for rainfall-runoff modelling, as it exhibits a high degree of correlation compared to generated data from Machine Learning (as reported by Ref. [23]). The accuracy of rainfall data and runoff modelling is crucial for hydropower generation, providing valuable information on how flood waves move through a point over time. This study emphasizes the importance of accurate rainfall data not only for hydropower generation but also for decision-making in agriculture, infrastructure development, and water supply.

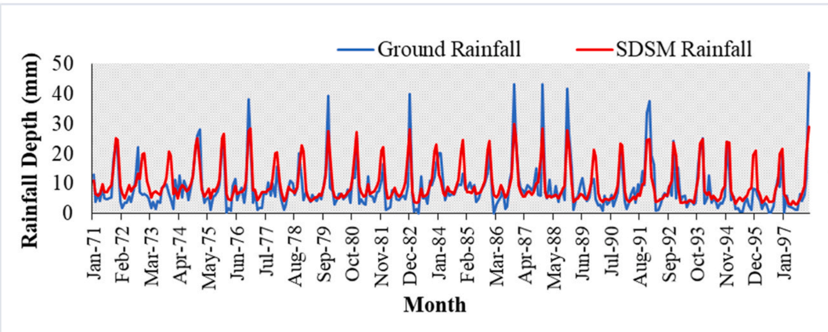

Fig. 5. Graphical comparison between SDSM Rainfall and Ground Rainfall.

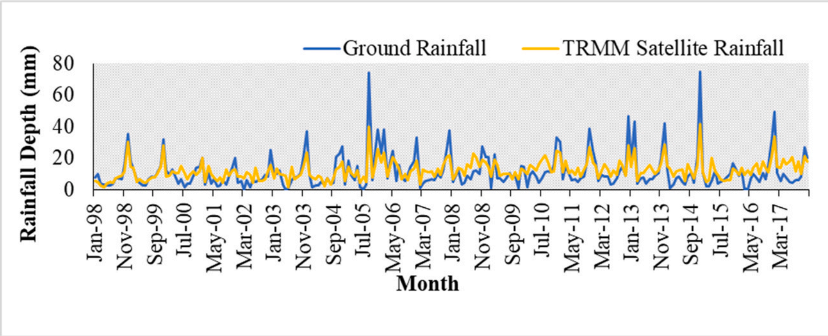

Fig. 6. Graphical comparison between ground rainfall and TRMM satellite rainfall.

Table 1
Statistical test (R^2 and RMSE) between ground rainfall and atmospheric rainfall.

In-Situ Rainfall	Atmospheric Rainfall	Year	R^2	RMSE
Ground Rainfall	SDSM Rainfall (1971–1997)	1971–1975	0.6	14.2
	Overall Study (Before, During and After hydropower plant Construction)	1976–1980	0.6	13.9
		1981–1985	0.5	13.8
		1986–1990	0.6	18.8
		1991–1995	0.5	12.4
		1996–1997	0.7	14.9
	TRMM Satellite Rainfall (1998–2017)	1998–2002	0.5	13.4
	After hydropower plant Construction	2003–2007	0.8	19.9
		2008–2012	0.5	19.3
		2013–2017	0.6	20.8

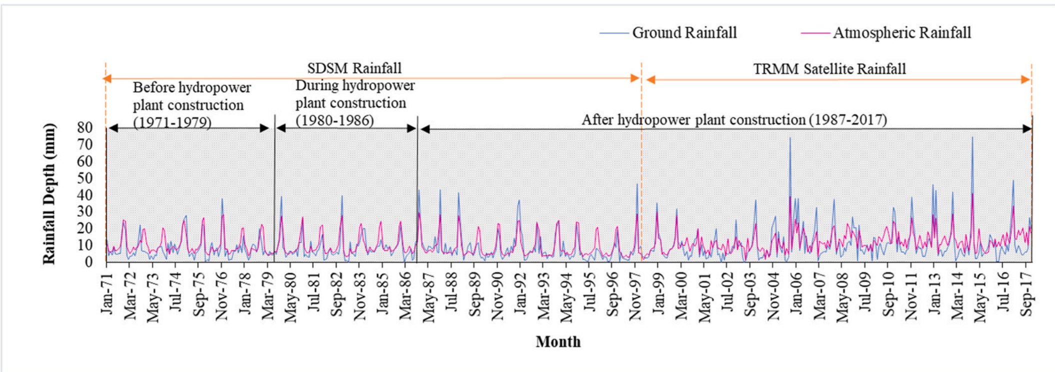

Fig. 7. Graphical Comparison between Ground Rainfall and Atmospheric Rainfall (GCM-TRMM) for the overall study period (1971–2017).

4. Conclusions

This study used rainfall data from the general circulation model and tropical rainfall measurement mission (GCM-TRMM) to forecast inflow for the period from 1971 to 2017. To obtain the regional rainfall data from GCM, a statistical downscaling model (SDSM) was put into place. The suggested model's validity was tested using intensive distance weighted (IDW) data and rainfall data from 12 ground stations deployed in the catchment of hydropower plants. The entire duration is divided into three stages presenting the development that occurred in the catchment (before constructing hydropower plant, during and after the construction). The study's findings reveal that the proposed model (GCM-TRMM) can be used as a reliable predictive model where R^2 values are ranging from 0.749 to 0.933. The rainfall-runoff model was developed using the Clark method in this study; more development could be achieved if a more sophisticated model could capture the non-linearity associated with a rainfall-runoff relationship, such as data-driven models. Future research could focus on the employing the Machine Learning (ML) models to solve the non-linearity feature in the rainfall-runoff model, especially those associated with image processing to such as Convolution neural network.

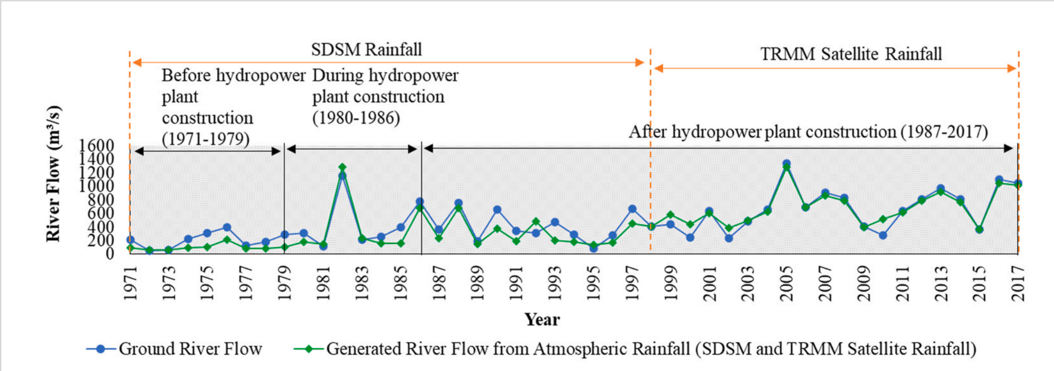

Fig. 8. Comparison between the two inflow models.

Table 2
The performances of the proposed inflow model.

Stages	Year	R^2	RMSE (m^3/s)
Before hydropower plant construction	1971–1979	0.749	129.2
During hydropower plant construction	1980–1986	0.933	114.4
After hydropower plant construction	1987–2017	0.848	123.5

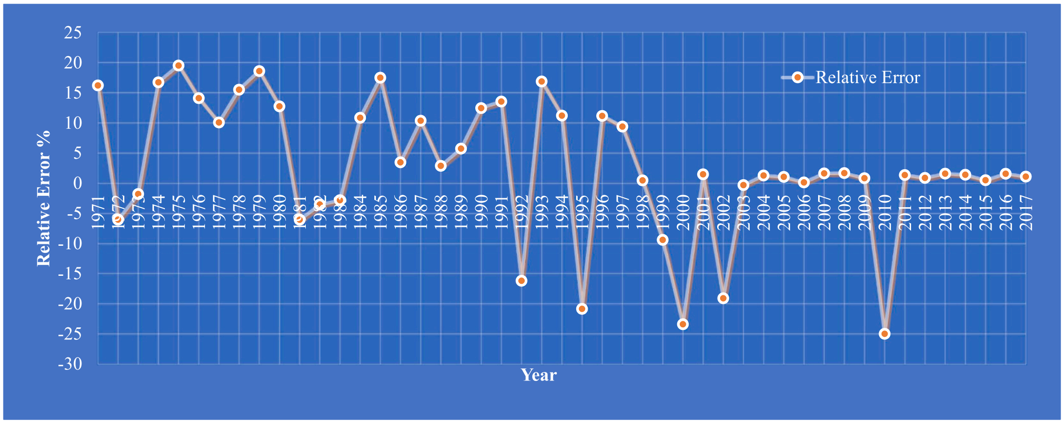

Fig. 9. Relative error distribution for the river flow.

Author contributions

Wan Norsyuhada Che Wan Zanial : Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Contributed reagents, materials, analysis tools or data; Wrote the paper.
Marlinda Abdul Malek: Conceived and designed the experiments, Contributed reagents, materials, analysis tools or data; Wrote the paper.
Mohd Nadzri Md Reba: Conceived and designed the experiments, Contributed reagents, materials, analysis tools or data.
Nuratiah Zaini: Analyzed and interpreted the data; Performed the experiments; Wrote the paper.
Ali Najah Ahmed: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.
Mohsen Sherif: Contributed reagents, materials, analysis tools or data; Wrote the paper.
Ahmed Elshafie: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.

Conflicts of interest

The authors declare no conflict of interest.

Availability of data and materials

The datasets generated during the current study are not publicly available which is considered as copyright between the funded agency and the university but are available from the corresponding author on reasonable request.

Funding

This MSc study is financially supported by BOLD Scholarship, College of Graduate School. Universiti Tenaga Nasional, Malaysia.

Acknowledgement

The authors would also like to acknowledge technical support from Geoscience and Digital Earth Centre (INSTeG), Research Institute for Sustainable Environment (RISE), Universiti Teknologi Malaysia, and access to data from Department of Irrigation and Drainage Malaysia (DID) and Department of Survey and Mapping Malaysia (JUPEM).

References

[1] N.D. Bennett, et al., Characterising performance of environmental models, Environ. Model. Software 40 (2013) 1–20, https://doi.org/10.1016/j.envsoft.2012.09.011.
[2] R.L. Wilby, C.W. Dawson, The Statistical DownScaling Model : insights from one decade, Int. J. Climatol. 33 (2013) 1707–1719, https://doi.org/10.1002/joc.3544.
[3] M.Z. Hashmi, A.Y. Shamseldin, B.W. Melville, Statistical downscaling of precipitation: state-of-the-art and application of bayesian multi-model approach for uncertainty assessment, Hydrol. Earth Syst. Sci. Discuss. 6 (5) (2009) 6535–6579, https://doi.org/10.5194/hessd-6-6535-2009.
[4] R.L. Wilby, T.M.L. Wigley, Downscaling general circulation model output: a review of methods and limitations, Prog. Phys. Geogr. 21 (4) (1997) 530–548, https://doi.org/10.1177/030913339702100403.
[5] N.N.A. Tukimat, N.A. Ahmad Syukri, M.A. Malek, Projection the long-term ungauged rainfall using integrated statistical downscaling model and geographic information system (SDSM-GIS) model, Heliyon 5 (9) (2019), https://doi.org/10.1016/j.heliyon.2019.e02456 e02456, Sep.
[6] K.W. Tahir, T. Hashim, A.M. Yusof, Statistical Downscaling of Rainfall under Transitional Climate in Limbang River Basin by Using SDSM Statistical Downscaling of Rainfall under Transitional Climate in Limbang River Basin by Using SDSM, 2018.
[7] D. Singh, S.K. Jain, R.D. Gupta, Statistical downscaling and projection of future temperature and precipitation change in middle catchment of Sutlej River Basin, India, J. Earth Syst. Sci. 124 (4) (2015) 843–860, https://doi.org/10.1007/s12040-015-0575-8.
[8] N.N.A. Tukimat, S. Harun, Climate change impact on rainfall and temperature in Muda irrigation area using multicorrelation matrix and downscaling method, J. Water Clim. Chang. 6 (3) (2015) 647–660, https://doi.org/10.2166/wcc.2015.015.
[9] Z. Hassan, S. Harun, Statistical Downscaling for Climate Change Scenarios of Rainfall and Temperature, 2011, https://doi.org/10.13140/RG.2.1.2336.9446.
[10] R.L. Wilby, C.W. Dawson, E.M. Barrow, SDSM - a decision support tool for the assessment of regional climate change impacts, Environ. Model. Software 17 (2002) 145–157, https://doi.org/10.1016/s1364-8152(01)00060-3.
[11] M. Widmann, C.S. Bretherton, E.P. Salathé Jr., Statistical precipitation downscaling over the northwestern United States using numerically simulated precipitation as a predictor, J. Clim. 16 (5) (2003) 799–816.
[12] R.L. Wilby, C.W. Dawson, SDSM 4.2— a decision support tool for th e assessment of regional climate change impacts, User Manual," Dep. Geogr. Lancaster Univ. UK (2007) 1–94, no. August.
[13] S.N.M. Zad, Z. Zulkafli, F.M. Muharram, Satellite rainfall (TRMM 3B42-V7) performance assessment and adjustment over Pahang river basin, Malaysia, Rem. Sens. 10 (3) (2018) 1–24, https://doi.org/10.3390/rs10030388.
[14] G. Merkuryeva, Y. Merkuryev, B. V Sokolov, S. Potryasaev, V.A. Zelentsov, A. Lektauers, Advanced river flood monitoring , modelling and forecasting, J. Comput. Sci. 10 (2015) 77–85.
[15] Brakenridge, et al., Calibration of satellite measurements of river discharge using a global hydrology model, J. Hydrol. 475 (2012) 123–136, https://doi.org/10.1016/j.jhydrol.2012.09.035.
[16] J. Li, A.D. Heap, A Review of Spatial Interpolation Methods for Environmental Scientists, 2008.
[17] M. Hussain, K.W. Yusof, M.R. Mustafa, N.R. Afshar, Application of statistical downscaling model (SDSM) for long term prediction of rainfall in Sarawak, Malaysia, November, Water Resour. Manag. VIII 1 (2015) 269–278, https://doi.org/10.2495/wrm150231.
[18] JPS, Hydrological Procedure NO . 1 (REVISED AND UPDATED 2015) Estimation of Design Rainstorm in Peninsular, 2015 vol. 1, no. 1.
[19] JPS, *Hydrological Procedure 11*, Design Flood Hydrograph Estimation For Rural Catchments In Malaysia, 2018.
[20] C.O. Clark, Storage and the unit hydrograph, Proc. Am. Soc. Civ. Eng. 69 (9) (1945) 1333–1360.
[21] H.E.C. de laCuanalo, R. Webster, A comparative study of numerical classification and ordination of soil profiles in a locality near Oxford. Part I. Analysis of 85 sites, J. Soil Sci. 21 (1970) 340–352.
[22] M. Amirabadizadeh, A.H. Ghazali, Y.F. Huang, A. Wayayok, Downscaling daily precipitation and temperatures over the Langat River Basin in Malaysia : a comparison of two statistical downscaling approaches, December, Int. J. Water Resour. Environ. Eng. 8 (2016) 120–136, https://doi.org/10.5897/IJWREE2016.0585.
[23] W.M. Ridwan, M. Sapitang, A. Aziz, K.F. Kushiar, A.N. Ahmed, A. El-Shafie, Rainfall forecasting model using machine learning methods: case study Terengganu, Malaysia, Ain Shams Eng. J. (2020), https://doi.org/10.1016/j.asej.2020.09.011.
[24] M. Sapitang, W.M. Ridwan, K. Faizal Kushiar, A. Najah Ahmed, A. El-Shafie, Machine learning application in reservoir water level forecasting for sustainable hydropower generation strategy, Sustainability 12 (15) (2020), https://doi.org/10.3390/su12156121, 6121, Jul.
[25] M. Vafakhah, S. Mohammad Hasani Loor, H. Pourghasemi, A. Katebikord, Comparing performance of random forest and adaptive neuro-fuzzy inference system data mining models for flood susceptibility mapping, Arabian J. Geosci. 13 (11) (2020) 1–16, https://doi.org/10.1007/s12517-020-05363-1. Jun.