# Multi-spatial scale hybrid rainfall-runoff modelling - A case study of Godavari river basin

Syed Abdul Mateen[1] and Regonda S.K[2]

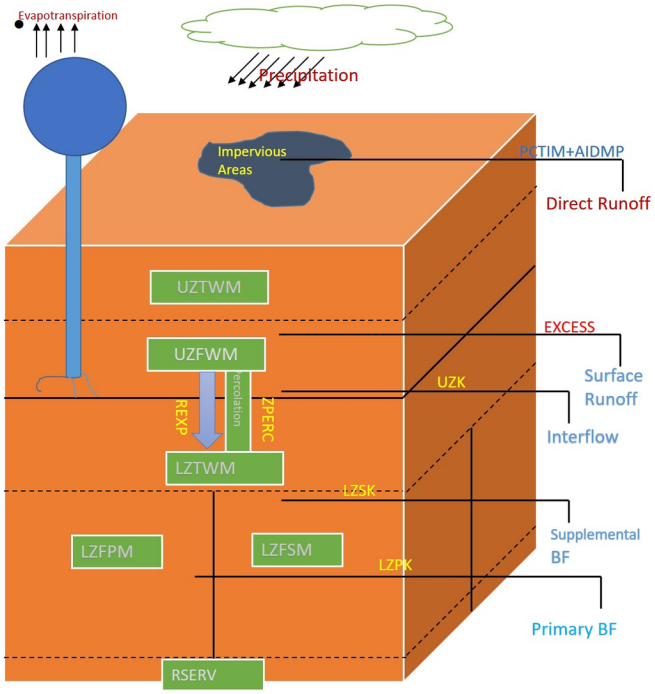
[1]IIT Bombay
[2]IITH

November 16, 2022

**Abstract**

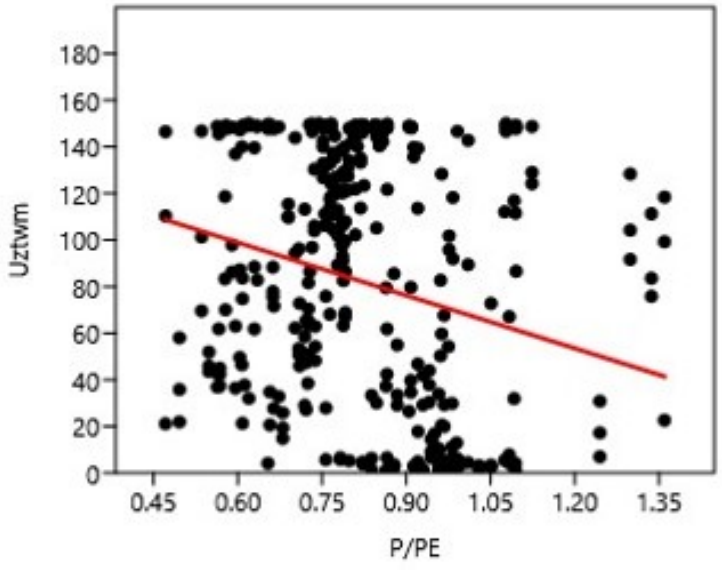
Transformation of rainfall to runoff is a complex hydrological phenomenon involving various interconnected processes. Besides, the distribution of rainfall and basin characteristics are not uniform across time and space leading to a poor understanding of the process. Hydrologists have been using various hydrological models to understand transformation of rainfall into runoff. Conceptual models developed in the 1960s represent various individual components of hydrological cycle via interconnected conceptual elements, thus model various aspects of the hydrological cycle. On the other hand, data-driven models such as Artificial Neural Networks (ANNs) are widely regarded as universal approximators due to their ability to model many complex problems. Very few studies reported the application of a widely used conceptual model, Sacramento Soil Moisture Accounting model (SAC-SMA), in the Indian river basins context. Considering that the hydrological cycle is very complex and may never be fully understood in detail, conceptual models like Sacramento Soil Moisture Accounting model (SAC-SMA) can be integrated with data-driven models which can take care of poorly described and understood aspects of hydrological modelling. In this study, a hybrid rainfall-runoff model was developed and applied over the Godavari river basin in India at multiple spatial scales for capturing the spatial variations in model inputs and catchment charateristics.The hybrid model by virtue of the semi-distributed configuaration and addition of ANN component led to improved simulations of streamflow in comparison to the standalone SAC-SMA model.

Posted on 16 Nov 2022 — CC-BY 4.0 — https://doi.org/10.1002/essoar.10511283.1 — This a preprint and has not been peer reviewed. Data may be preliminary.

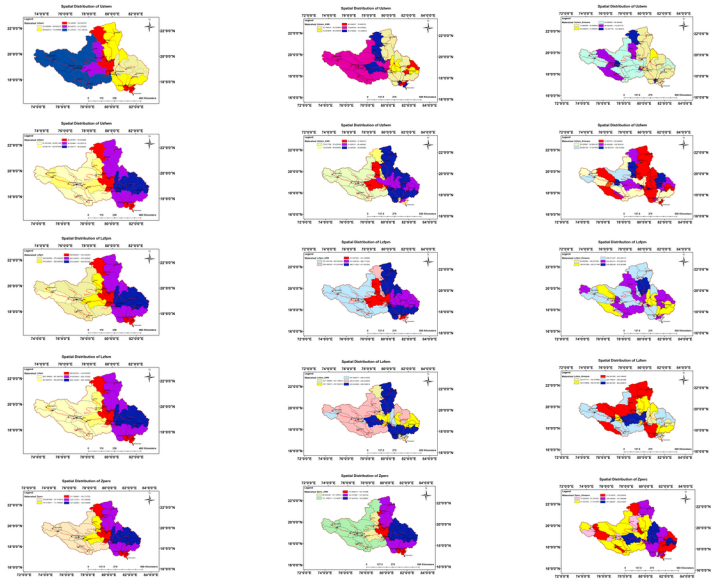



9

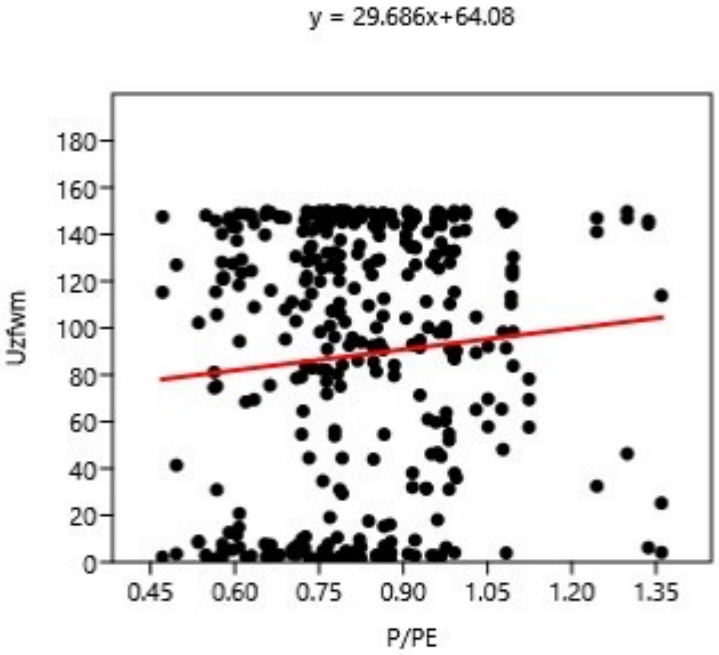
y = -75.672x+144.32



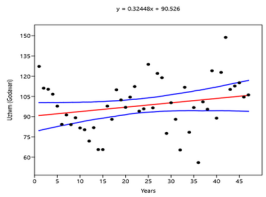

Posted on 16 Nov 2022 — CC-BY 4.0 — https://doi.org/10.1002/essoar.10511283.1 — This a preprint and has not been peer reviewed. Data may be preliminary.



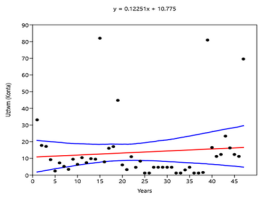
(a) (b) (c)



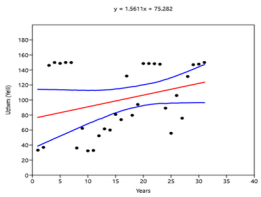

Posted on 16 Nov 2022 — CC-BY 4.0 — https://doi.org/10.1002/essoar.10511283.1 — This a preprint and has not been peer reviewed. Data may be preliminary.




(a)




(b)




(c)

Posted on 16 Nov 2022 — CC-BY 4.0 — https://doi.org/10.1002/essoar.10511283.1 — This a preprint and has not been peer reviewed. Data may be preliminary.

# Multi-spatial scale hybrid rainfall-runoff modelling - A case study of Godavari river basin

**S.A Mateen, S.K Regonda**

[1]Department of Civil Engineering,IIT Hyderabad,TS,India
[2]Department of Civil Engineering,IIT Hyderabad,TS,India

**Key Points:**

- In this study, a hybrid model framework was developed for rainfall-runoff modelling for Godavari basin
- The hybrid model consisted of the well validated SAC-SMA model and ANN component.
- The ANN's with their structural flexibilty, improved the runoff estimates from the lumped SAC-SMA model.
- The SAC-SMA model was used in five different configurations and the spatially distributed information was introduced gradually.

Corresponding author: S.A Mateen, mateen1634@gmail.com

Corresponding author: S.K Regonda, satishr@ce.iith.ac.in

**Abstract**

Transformation of rainfall to runoff is a complex hydrological phenomenon involving various interconnected processes. Besides, the distribution of rainfall and basin characteristics are not uniform across time and space leading to a poor understanding of the process. Hydrologists have been using various hydrological models to understand transformation of rainfall into runoff. Conceptual models developed in the 1960s represent various individual components of hydrological cycle via interconnected conceptual elements, thus model various aspects of the hydrological cycle. On the other hand, data-driven models such as Artificial Neural Networks (ANNs) are widely regarded as universal approximators due to their ability to model many complex problems.

Very few studies reported the application of a widely used conceptual model, Sacramento Soil Moisture Accounting model (SAC-SMA), in the Indian river basins context. Considering that the hydrological cycle is very complex and may never be fully understood in detail, conceptual models like Sacramento Soil Moisture Accounting model (SAC-SMA) can be integrated with data-driven models which can take care of poorly described and understood aspects of hydrological modelling. In this study, a hybrid rainfall-runoff model was developed and applied over the Godavari river basin in India at multiple spatial scales for capturing the spatial variations in model inputs and catchment charateristics.The hybrid model by virtue of the semi-distributed configuaration and addition of ANN component led to improved simulations of streamflow in comparison to the standalone SAC-SMA model.

**Keywords**: ANN, Conceptual model, Godavari River Basin, Flood Forecasting, Hydrologic Cycle, NWS, SAC-SMA, Scaling

## 1 Introduction

All rainfall-runoff models represent some sort of simplification of the hydrological cycle to a varying degree. In hydrological sciences, it's challenging to set a benchmark for model development which has led to an ever-growing list of hydrological models being developed to simulate the same mechanism, i.e., rainfall to runoff transformation. The rainfall-runoff modelling of today traces its origins to the availability of computing power in the 1960s (Todini, 2011; Singh & Woolhiser, 2002). Most of these models were of the conceptual type, meaning they represented various individual components of the watershed by interconnected conceptual elements such as plane, converging and diverging sections and routing channels. Along with the conceptual models, data driven models which primarily rely on observational data began to be used for hydrologic studies. The Unit hydrograph is viewed as the first data driven model in hydrology (Sherman, 1932). Similar empirical models relating flow observations to catchment descriptors were developed and fall in the category of data driven models. Data driven approach introduced the loss of "physicality" (no consideration for physical laws) which increased further with development of ANN's, fuzzy logic and genetic algorithms (Todini, 2007).

One of the conceptual models used for flood forecasting in the United States is the Sacramento Soil Moisture Accounting Model (SAC-SMA), developed by the staff of National Weather Service (NWS), River Forecast Center (RFC) in Sacramento, California. SAC-SMA is a 16 parameter hydrological model which treats the watershed (or basin) as a soil block and divides it into two zones, i.e., a thin upper zone and a thicker lower zone. The upper zone is typically top soil layer and corresponds to the interception storage, whereas the lower zone represents the bulk of the soil moisture. The depth up-to which plant roots can penetrate and extract soil moisture limits the depths of the upper and lower zones (Armstrong, 1978). Within each zone, the soil moisture storage is composed of "Tension Water" and "Free Water". Tension water is that portion of soil moisture that is bonded with soil particles and is used by evapotranspiration. Free wa-

ter is that water which moves under the influence of gravity and is depleted by percolation, evapotranspiration, interflow, surface runoff and refilling of tension water. The SAC-SMA model, by design, first fills the tension water storage in the upper zone because the moisture content of soil must be raised to a certain level to initiate the downward movement of the wetting front; the wetting front corresponds to soil moisture profile in the ground (Burnash & Ferral, 1996; Bodman & Colman, 1944). Filling of upper tension water storage is followed by filling of upper free water storage, which controls interflow, generates surface runoff and allows for percolation of water to lower zones. Thus, the moisture content in the upper zone filled first and then will meet requirements of the lower zone. The ability of the model to perform well depends on simulating extended periods of varying flows, comprising of dry periods with low flows and wet periods with high flows which necessitate the division of lower zone free water storage into two, primary which is slow draining and simulates the year-round base flow and supplementary which is quick releasing (Burnash & Ferral, 1996; Anderson, 2002). Thus,the SAC-SMA is a continuous water balance model simulates changes in soil moisture based on the wetting and drying cycles of soil which results due to precipitation and subsequent depletion of accumulated soil moisture by evapotranspiration, production of runoff and gravity draining (Shamir et al., 2006). The model conceptualisation is shown in Figure 1. For more details on the SAC-SMA model the readers are referred to the companion review paper.

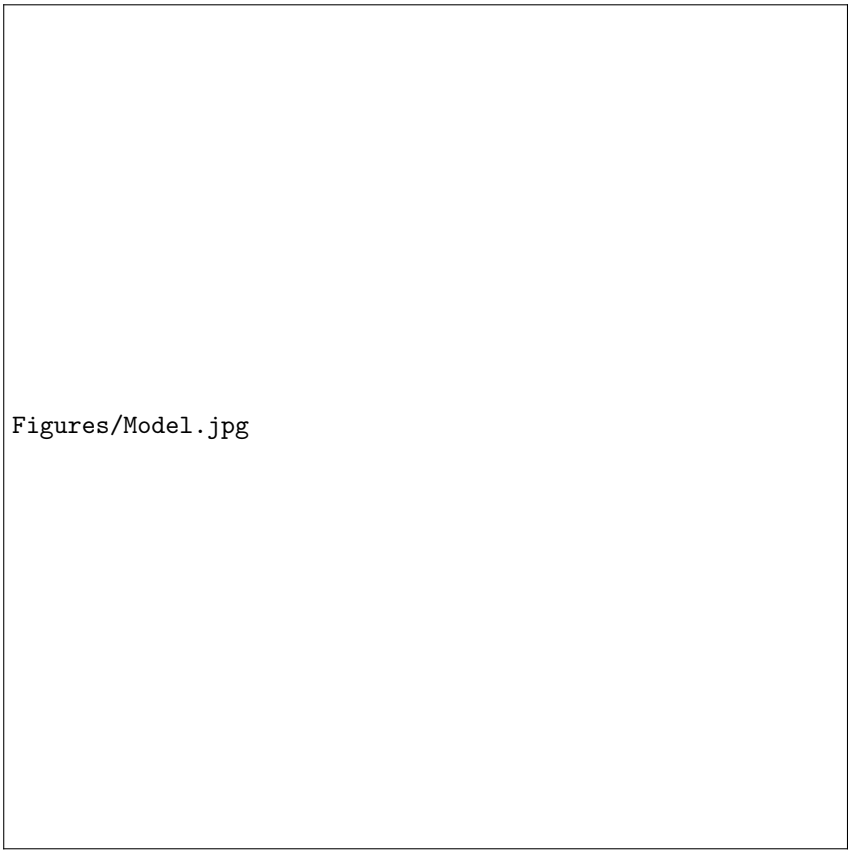



**Figure 1.** Conceptualisation of the SAC-SMA model

Among the data-driven models, ANN's are widely used to model many complex problems and thus regarded as universal approximators. ANN's work by imitating the

human brain and its neural structure. ANN's receive the input data at nodes which pass on the information to the other nodes depending on layers and structure. The transfer of the information is dependent on the connection weights and some bias is also added to compensate for the uncertainty effects. The components of an artificial neuron are shown in Figure2. By applying a non-linear transformation at each node ANN's are able to model varying engineering problems which is the reason for the widespread adoption (on Application of Artificial Neural Networks in Hydrology, 2000b; Tayfur, 2014). The ability

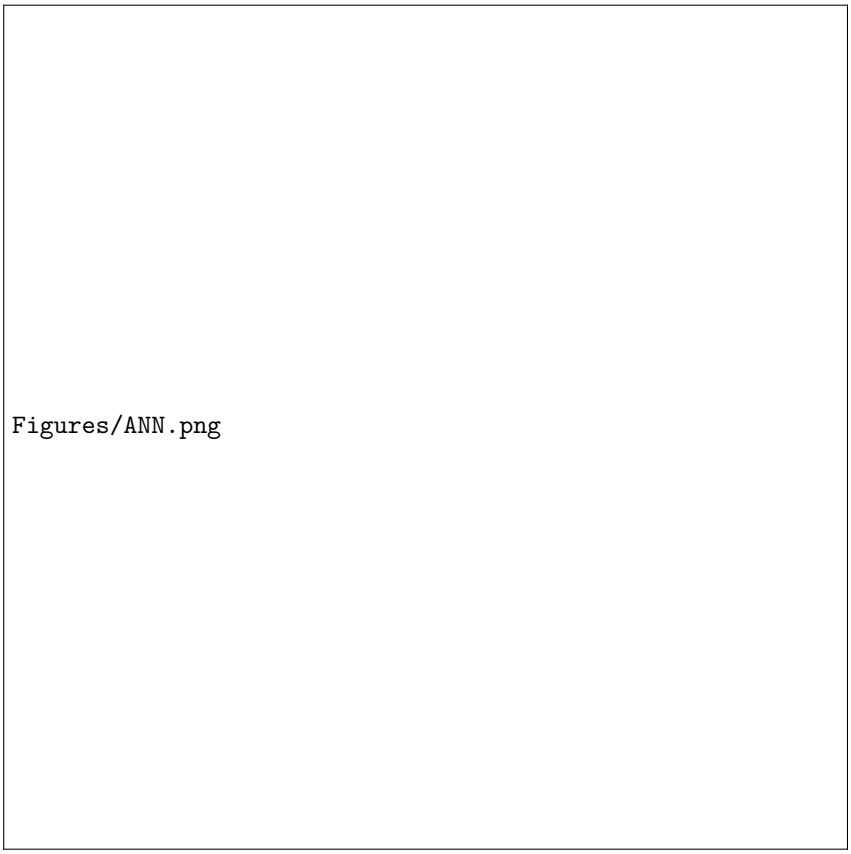



**Figure 2.** Components of an artificial neuron. I1, I2, I3 are the inputs to the neuron with W1, W2, W3 being the connection weights. B1 is the bias term. The net information is transformed by the activation function and carried forward in proportion to the connection weights.

of ANN's to learn from the input data and adjust the weights and biases make them attractive for hydrological modelling. Without giving importance to the underlying physical considerations, ANN's can capture the non-linearity inherent in the rainfall-runoff process (Singh & Woolhiser, 2002). The usage of ANN's in hydrologic applications started in 1990's with Daniell (1991) reporting some of the first applications. Halff et al. (1993) constructed the earliest rainfall runoff model using ANN's with the rainfall as inputs and streamflow as outputs at Bellvue, Washington. Similar modelling approaches were considered for other basins and also by including more inputs such as temperature, snowmelt etc. Various researchers reported improved performance with the use of ANN over standalone conceptual models. Daliakopoulos and Tsanis (2016) compiled results from twenty-two studies comparing performances of conceptual models with ANN for rainfall-runoff applications. It is significant to mention that a whopping twenty-one studies reported per-formance of ANN's superior or equivalent to conceptual models. Only one study by Gaume and Gosset (2003) found the conceptual model GR4J outper-forming

the ANN's and the authors were in favour of a limited complexity conceptual model than ANN's. They also compared SAC-SMA and ANN's on a standalone basis. SAC-SMA was calibrated using Genetic Algorithm optimisation scheme while as for ANN's, trial and error approach was adopted to determine the hidden layer neurons for im-proved performance. Although both models performed well, overall, ANN's results were about 10 per cent better than SAC-SMA, and it outmatched SAC-SMA's performance in high flows as well. However, SAC-SMA performed better to simulate low flows than ANN's. Use of ANN's in the hydrologic application is further detailed in the second part of the ASCE Task Committee on Application of Artificial Neural Networks in Hydrology. As the applications of ANN's grew it was realised their usage is not without drawbacks such as the complete disregard for the physical processes happening in the catchment. Also, the results from the ANN are heavily dependent on the quality and length of input training data. Thus extrapolation of results can become a issue if the problem domain is outside the scope of the training data. Selection of optimal training data range is mostly based on trial and error rather than having a sound scientific basis (on Application of Artificial Neural Networks in Hydrology, 2000a).

With each modeller promoting his own approach and a plethora of hydrological models available, there arises a need to reconcile the various lines of thinking. In this regard, Koren et al. (2014) integrated a physically-based model (NOAH-LSM) with a conceptual model (SAC-SMA) to include the effect of freeze and thaw on runoff generation. The soil moisture changes in the SAC-SMA storages due to rainfall were transferred to a heat transfer model which divided the total water content into liquid and frozen water portions based on the simulated soil temperature profile. The updated soil moisture states are computed back into the tension and free water storages. The modified version of SAC-SMA is referred to as SAC-HT (SAC Heat transfer) and does not need calibration for frozen ground parameters. Julien and Halgren (2014) explored a hybrid approach to hydrologic modelling by combining the SAC-SMA model with the TREX distributed surface hydrology model. The hybrid model was composed of an additional TREX surface layer in addition to SAC-SMA soil zones. The models are linked by infiltration from TREX as input to the upper zone of SAC-SMA and sub-surface flow as a point source to TREX surface flow. The hybrid model was tested on the California Gultch watershed near Leadville, Colorado. The simulations were once obtained with only TREX model and other time with the hybrid model. The results show that by introducing the soil moisture component alongside the TREX model, the hydrograph peaks are reduced in overestimation. At stations with low infiltration rates of soil, the hybrid model does not improve the performance due to reduced importance of soil moisture. Chen and Adams (2006) integrated three conceptual models (the Xinanjiang model, the Soil Moisture Accounting and Routing (SMAR) model, and the Tank model) with ANN's based on the premise that ANN will take care of non-linear transformation of rainfall to runoff. The results from the hybrid models were compared to a lumped, semi-distributed and semi-distributed with linear regression configuration. It was consistently observed that the hybrid model outperformed the above-mentioned configurations since the ANN's were able to map the non-linear relationship effectively. Corzo et al. (2009) expanded onto this work by formulating two schemes in relation to integrating conceptual models with data-driven models. The semi-distributed HBV model over the Meuse river basin was used with a linear regression model, artificial neural networks and M5 model trees. In the first scheme, out of 15 sub-basins in the Meuse river basin on which the semi-distributed HBV model was run, some sub-basins were replaced by these data-driven models. The replacement was done based on error contribution, runoff response and basin area with the effects were seen in per cent of area replaced, average discharge contribution, RMSE reduction. Replacement with an ANN model worked best and was proceeded further. The addition of ANN helped improve the simulation performance with low flows simulated better. In the second scheme rather than traditional routing techniques, ANN's were used for routing. Any error in the model simulation due to traditional routing techniques was corrected by ANN's to some extent resulting in hybrid model outperforming the HBV model.

In this study we attempted to integrate a conceptual (SAC-SMA) with a data-driven model (ANN) leading to the hybrid model setup shown in Figure 3. As the name suggests, hybrid models employ two or more modelling approaches, thus combining the strengths and masking the limitations of each approach. Considering that the hydrological cycle is very complex and may never be fully understood, conceptual models like SAC-SMA can be integrated with data-driven models which can take care of poorly described and understood aspects of hydrological modelling. This is so since data driven models don't require information of the physical domain.

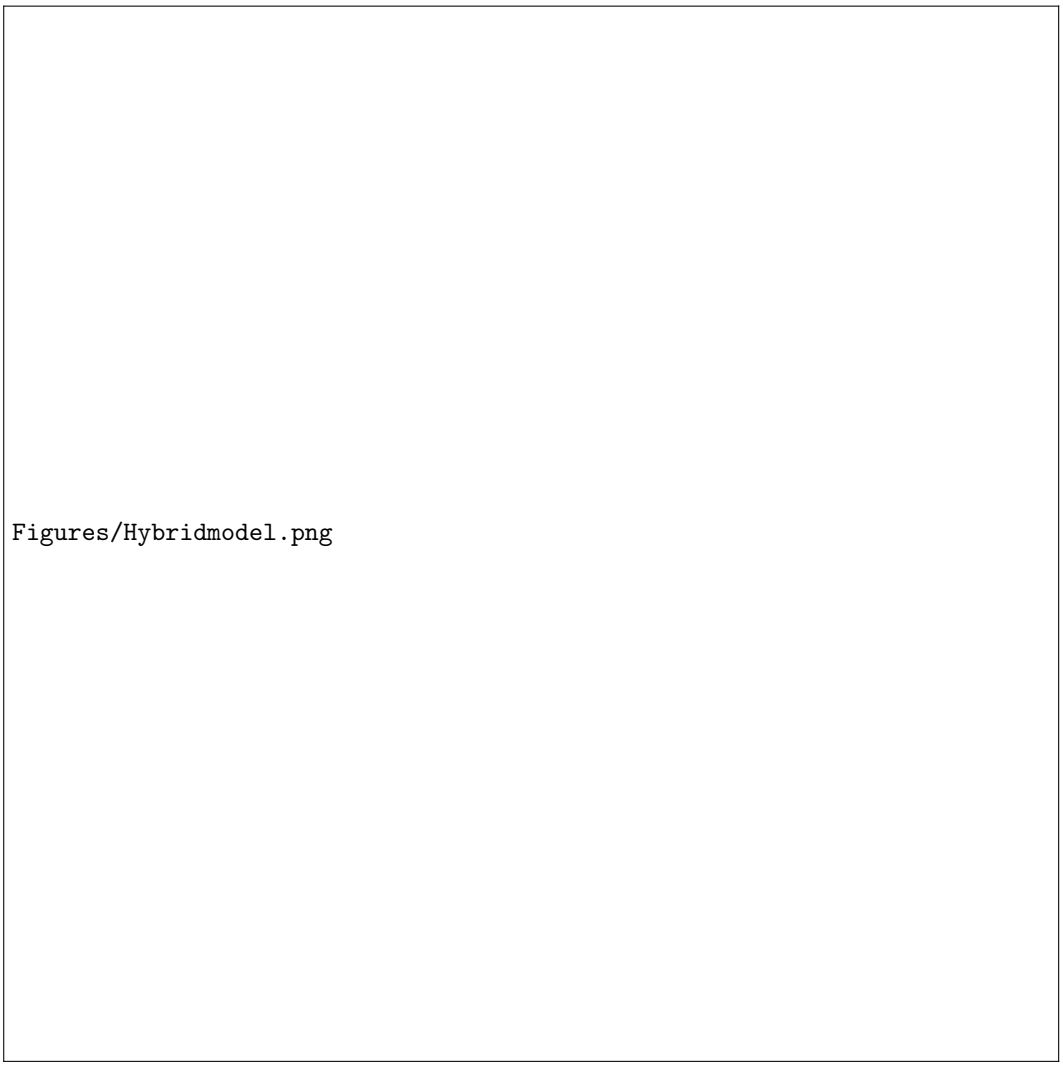



**Figure 3.** Hybrid model conceptualisation

The paper is organised as follows: Firstly we introduce the various modelling approaches in first section while building towards the development and the need for a hybrid modelling approach; the second section details the methodology adopted with a focus on comparisons between lumped SAC-SMA and hybrid model ; the third section lays out the results and its analysis with the fourth section summarising the important observations from the study.

# 2 Materials and Methods

## 2.1 Overview

A hybrid modelling approach was developed to estimate streamflows for the Godavari River Basin. As part of this approach, a conceptual model SAC-SMA was integrated with ANN's for Cases 2 through 5. The addition of ANN can be seen as a post-processing technique similar to bias correction techniques. Thus the hybrid model, instead of using the principle of superposition for the routed runoff from all the 49 sub-catchments in the formation of the total runoff output at the outlet of the entire catchment uses ANN for non-linear mapping.The overall methodology adopted is enumerated below and show in Figure 4:

1. SAC-SMA was run in lumped mode and streamflows are estimated at the basin outlet, Pollavaram, which are then compared with observed streamflows. This is considered as Case 1 scenario.
2. SAC-SMA was run in the semi-distributed mode with model Inputs (Precipitation and , Potential Evapotranspiration) at sub-basin level and parameters estimated from lumped mode. The Godavari basin was delineated into 49 sub-basins. This is considered as Case 2 scenario
3. In the third case we again ran SAC-SMA in the semi-distributed mode but with mixed parameters. Thus for sub-basins above Godavari Arch bridge (GR bridge) station, parameters obtained by calibration at GR bridge were used while as for remaining downstream stations lumped parameters were used.
4. SAC-SMA was run in the semi-distributed mode with model inputs (Precipitation and , Potential Evapotranspiration) at sub-basin level and distributed parameters that were obtained by parameter transfer techniques. This is considered as Case 4 scenario.
5. SAC-SMA was run in the semi-distributed mode with model inputs (Precipitation and , Potential Evapotranspiration) at the basin level and distributed parameters that were obtained from Case 4. This is Case 5 scenario.

## 2.2 Study Area

The Godavari river is the third largest river in India and is of immense significance in peninsular India. The Godavari river originates in the Western Ghats at an elevation of 1067 m and traverses a distance of about 1,465 Km till it drains into the Bay of Bengal. The basin is situated between latitude $16^0$ $16'$ North and $22^0$ $36'$ North and longitude $73^0$ $26'$ East and $83^0$ $07'$ East covering states of Maharashtra, Telangana, Andhra Pradesh,Madhya Pradesh, Chhattisgarh, Orissa, Karnataka and the Union Territory of Puducherry. The river is joined by various tributaries namely Pravara, Manjira and Maner along the right bank, and Purna, Pranhita, Indravathi and Sabari along the left bank. Pranhita and Indravathi contribute the most in terms of discharge volume into the Godavari river (Commission et al., 2006; Bhawan & Puram, 2014).

Based on the availability and quality of observed streamflow data, Pollavaram was chosen as the outlet of the basin despite the existence of streamflow measurements at Sir Arthur Cotton Barrage (Godavari barrage) 42 Km downstream. Pollavaram dam project has immense significance in terms of irrigation potential and tackling drought problems in Telangana and Andhra Pradesh. The project has been embroiled in controversies due to objections by other states regarding water utilisation and environmental groups (Gujja et al., 2006). The drainage area of the Godavari basin up to Pollavarm is 310375.3 $Km^2$. Figure 5 shows the Godavari river denoted by red colour and the elevation profile of the basin.

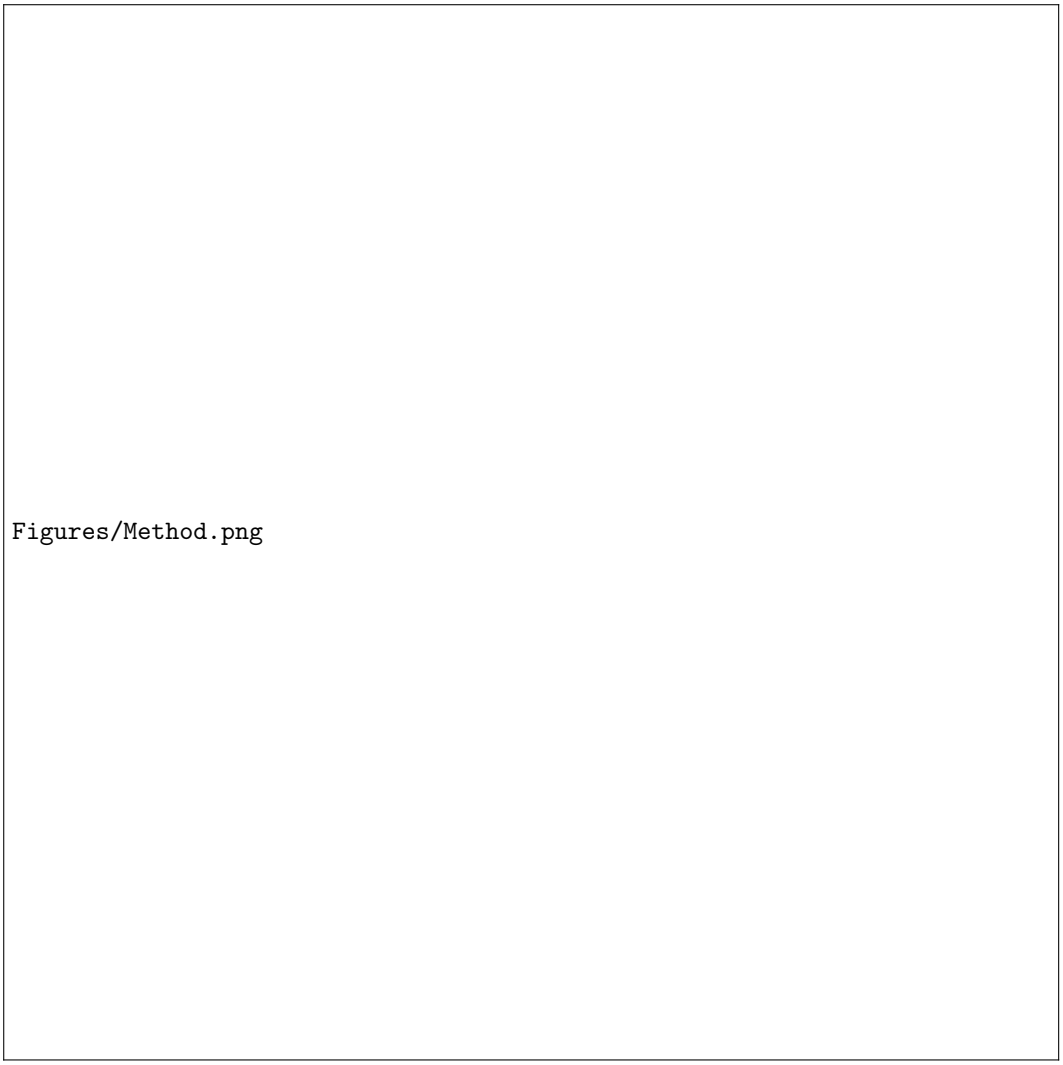



**Figure 4.** Overview of the methodology adopted

### 2.3 SAC-SMA model inputs

SAC-SMA requires a time-series of precipitation, potential evapotranspiration and observed streamflow as input. The modelling time steps depend on the temporal resolution of the input data and the nature of the problem. Significant number of studies of SAC-SMA is are on a hourly, six-hour and daily time-scale (Peck, 1976). The mean areal precipitation time-series can be prepared by station weighting schemes such as Thiessen polygons or inverse squared distance weighting methods (Anderson, 2002; Zhang, 2003). Daily Potential Evapotranspiration(PET) can be estimated from semi-emprical equations such as Penman requiring extensive meteorological data (air temperature, solar radiation, relative humidity and wind speed) (Penman, 1948). In countries such as India due to paucity of data simpler temperature based methods can be adopted such as Hamon's method for PET estimation (Hamon, 1960). Observed streamflow data is used to calibrate the model as it is the response of the basin and contains the combined effects of all the processes occurring in the basin.

Table 1 details about the inputs that are used in the SAC-SMA model while as Table 2 provides the input statistics. Figure 6 presents the time-series plots of the inputs.

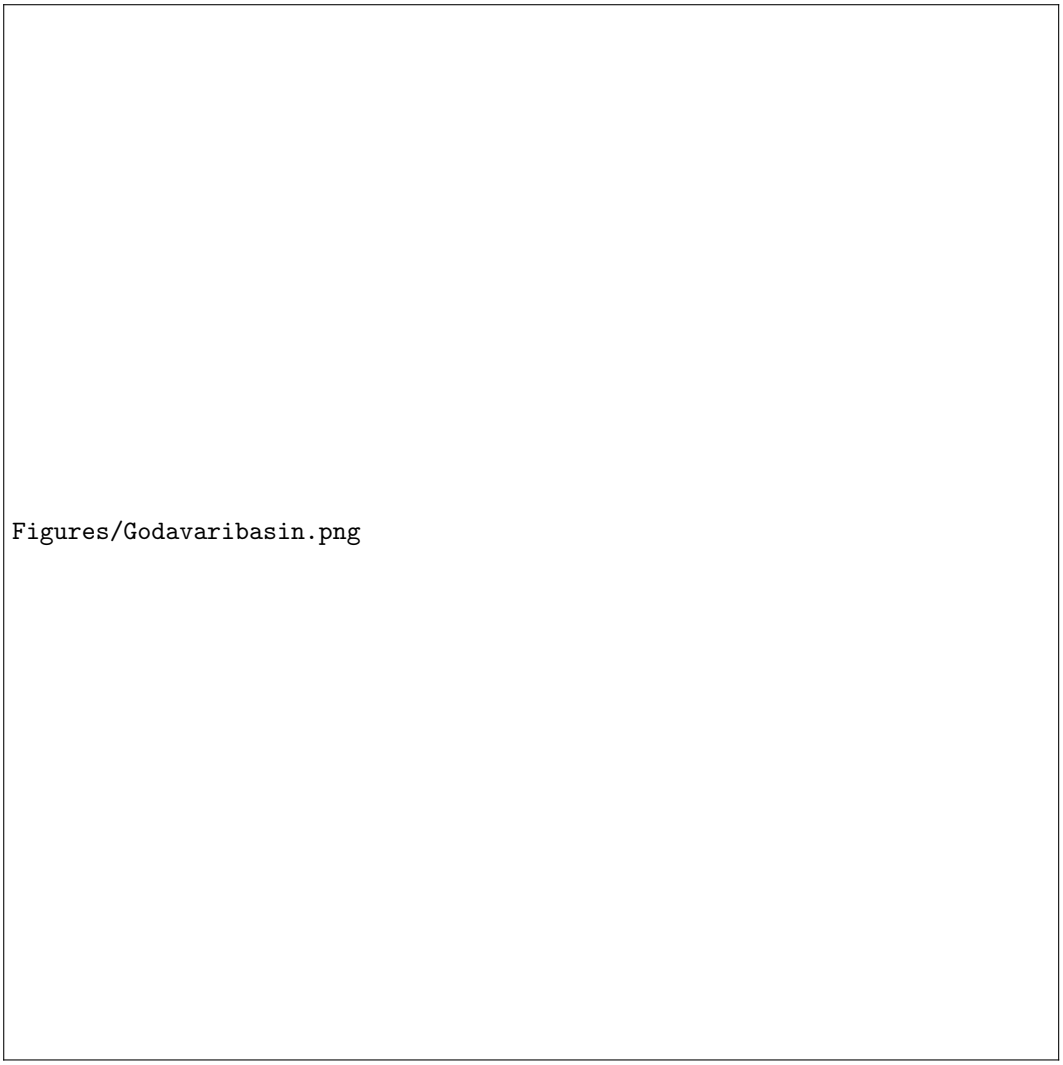



**Figure 5.** Godavari river basin.Various discharge stations are also shown in the figure.

Precipitation shows the most variability amongst all the inputs, whereas PET which was derived based on mean temperature shows the least variability. Around 85 per cent of rainfall in Godavari basin happens during the southwest monsoon season of June to September. The annual rainfall amount varies from 600 to 3000 mm. Cyclonic depressions originating in the Bay of Bengal result in extreme rainfall in the basin. The mountainous region of Western Ghats intercepts the southwestern monsoon winds producing heavy rainfall over the thin mountainous belt. This results in arid conditions in the area on the leeward side of the Western Ghats. The mean annual surface temperature in the Godavari basin is around $26^0$C. The western arid regions of the Godavari River Basin barring the Western Ghats are comparatively hotter as compared to the eastern areas of Godavari. In the western parts of the Godavari region, the mean daily maximum temperature generally exceeds $30^0$C while as its slightly less than $30^0$C over the eastern areas. The east part of the Godavari being closer to sea shows a comparatively low annual range of temperature (Commission et al., 2006).

**Table 1.** Model Inputs (Godavari Basin)

| Inputs | Scale | Source | Span |
|---|---|---|---|
| Precipitation | Daily | IMD Gridded ($0.25^0$*$0.25^0$) | $1^{st}$ Jan 1966 to $31^{st}$ Dec 2014 |
| Temperature (PET) | Daily | IMD Gridded ($1^0$*$1^0$) | $1^{st}$ Jan 1966 to $31^{st}$ Dec 2014 |
| Observed streamflow at Pollavaram | Daily | Central Water Commision | $1^{st}$ Jan 1966 to $31^{st}$ Dec 2014 |

IMD- Indian Meteorological Department

**Table 2.** Input statistics

| Input(mm/day) | Mean | Median | Standard Deviation | Maximum |
|---|---|---|---|---|
| Precipitation | 3 | 0.35 | 5.83 | 69.25 |
| PET | 3.83 | 3.7 | 1.38 | 8 |
| Observed streamflow at Pollavaram | 0.77 | 0.13 | 1.58 | 17.16 |

### 2.4 SAC-SMA in lumped configuration (Case 1)

SAC-SMA is a lumped conceptual model. Both, precipitation and temperature were spatially averaged over the Godavari River Basin from the gridded precipitation and temperature data obtained from the Indian Metrological Department (IMD). Potential Evapotranspiration was estimated from the mean temperature by Hamon's method (Hamon, 1960). Hamon's formula is given as

$$PET = 13.97 * Daylightfraction * Saturatedvapourpressure \quad (1)$$

Saturated vapour pressure is determined with the help of following relation:

$$Saturatedvapourpressure = 4.95 * e^{(0.062*Meantemperature)}/100 \quad (2)$$

The daylight fraction is the total sunshine hours (calculated in the units of 12 h) for a particular day. The daylight hours at the centroid ($79^0$ 00' E, $19^0$ 30' N) of the Godavari basin for the year 2012 were obtained from Astronomical Applications Department of U.S. Naval Observatory. Although the daylight hours may vary across the years and the basin itself, the difference was observed to be small and hence ignored. Similar performance has been observed among both less data intensive simpler methods and data intensive Penman equations, therefore,simpler PET estimation methods such as Hamon's method are used in this study (Oudin et al., 2005).

The SAC-SMA model ran in the lumped configuration supplying the input for the entire Godavari basin considering it as a single entity, as shown in Figure 5.The model was implemented in the R programming environment using the package "Hydromad" (Andrews & Guillaume, 2018). The version used for this study is 0.9-22. The modelling framework consists of the SAC-SMA model and a routing module which transforms the effective rainfall into runoff as shown in Figure 7. The routing module includes a unit hydrograph transfer function, defined by Autoregressive Moving Average Models with exogenous variables (ARMAX) (Jakeman et al., 1990). A warmup period of one year was chosen for stable parameter estimates. The model was calibrated automatically using the "Dynamically dimensioned search" (DDS) algorithm. DDS algorithm was introduced for automatic calibration of watershed simulation models. It scales the search to find reliable and accurate solutions within the maximum number of user-specified function. The adjustment from global to local search is achieved by dynamically and probabilistically modifying the set of decision variables or parameters from their best value known as reducing the dimensions (Tolson & Shoemaker, 2007). The daily streamflow data from 01-

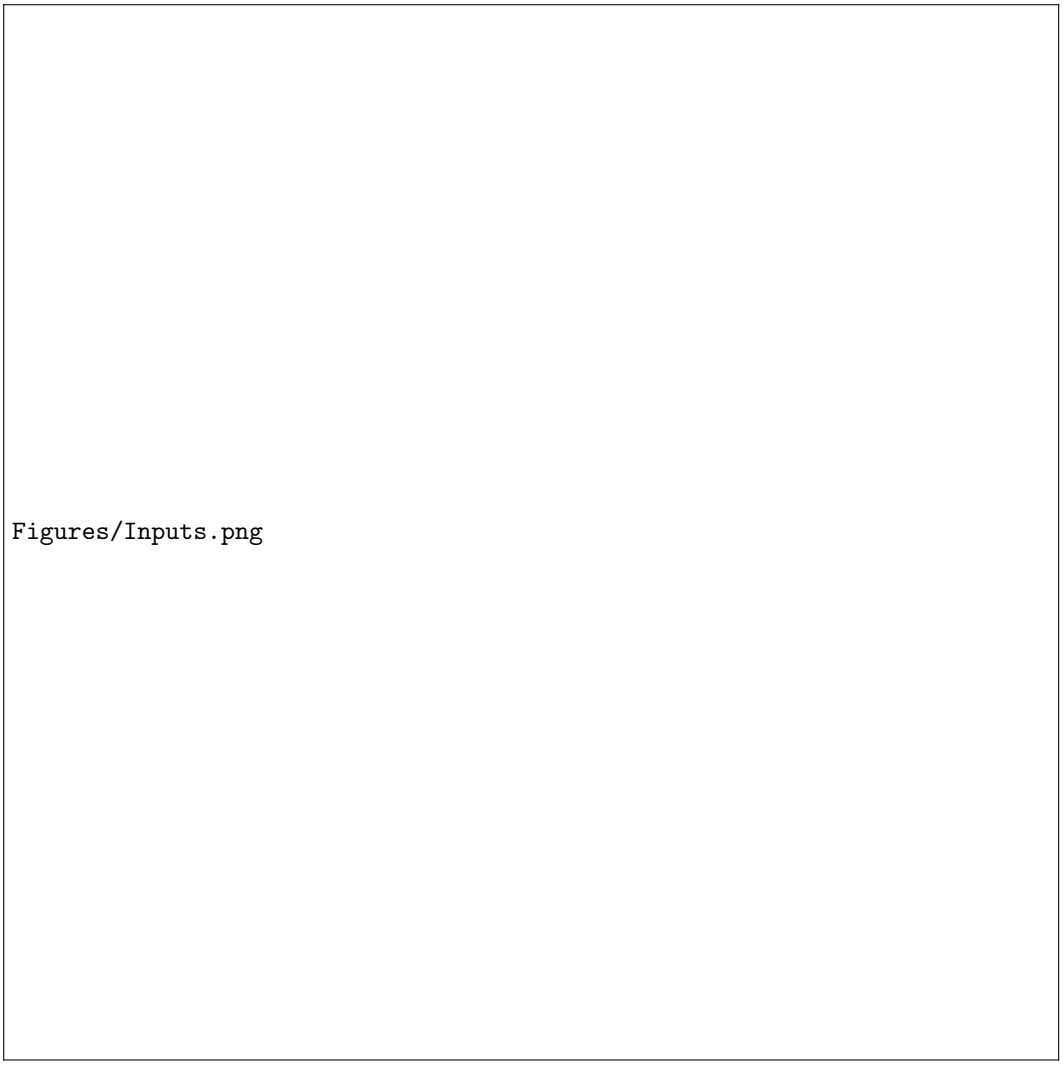



**Figure 6.** Model Inputs (Godavari Basin)

01-1966 to 31-12-1994 and 01-01-1995 to 31-12-2014 was used as calibration and validation periods, respectively. The calibration period consists of at least eight years of extremes of dryness and wetness, which is considered essential for the good calibration of the SAC-SMA model (Anderson, 2002).

### 2.5 SAC-SMA in semi-distributed configuration

The need for a semi-distributed approach comes from the fact that both the precipitation and PET are spatially varying across the Godavari basin, and we need to account for it for an improved understanding of the watershed behaviour. The high-resolution data available from satellites and radars can thus be incorporated in the SAC-SMA model leading to improved simulations. In the interior locations of the basin, streamflow estimates can be made which can be mainly used in the construction of hydraulic structures and water resources management.

Automatic delineation of the Godavari basin using the ARC-SWAT plugin in ARC-GIS was performed to divide it into 49 sub-basins. Three tiles of the 30 m SRTM DEM

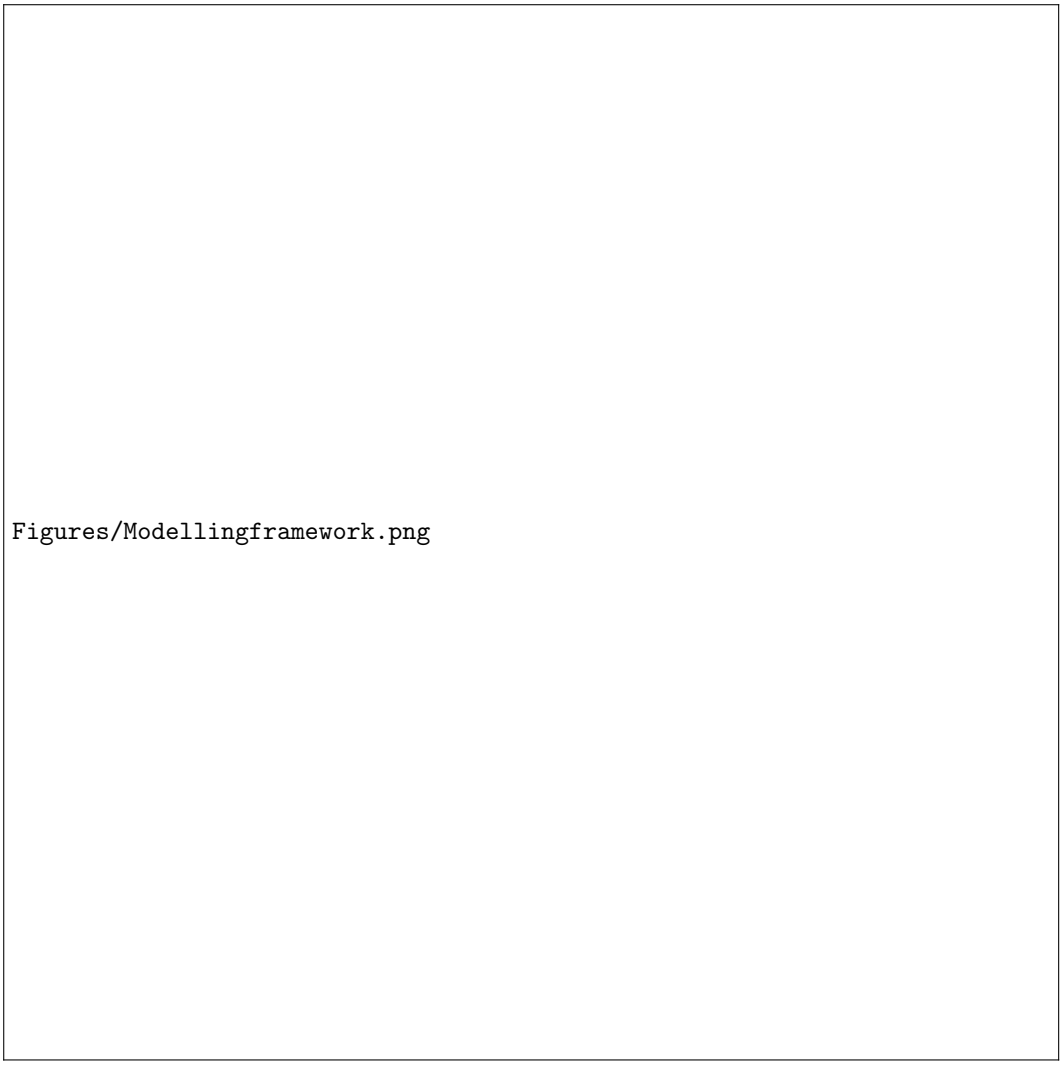



**Figure 7.** Modelling framework

were mosaicked and then clipped to the shapefile of the Godavari basin and input as the source DEM in ARC-SWAT. The stream network was defined based on the DEM, allowing the software to determine the flow direction based on the eight-direction (D8) flow model. The D8 flow model calculates flow direction based on the direction of steepest descent with flow travelling into the eight adjacent cells (Jenson & Domingue, 1988). The critical stream area threshold to create stream network was set as 5000 Ha (50 $Km^2$). It was necessary to incorporate information on the discharge stations into the delineation process so as to calibrate the model effectively . From the "Integrated hydrological data book for non-classified river basins", Table 4 gives the details of the historical observation sites located in the Godavari basin (Commission et al., 2006). The coordinates of these 48 stations were input to the ARC-SWAT software as outlets for the sub-basins thus effectively forcing it to create the sub-basins around these stations. However, only 44 sub-basins had the sub-basin outlet matching with the on-ground stations since it was not possible to create all the sub-basins based on the input stations coordinates. Pollavaram outlet was selected as the drainage point for the whole basin, and the basin was delineated into 49 sub-basins as shown in Figure 8. The station names are also mentioned for the input stations.

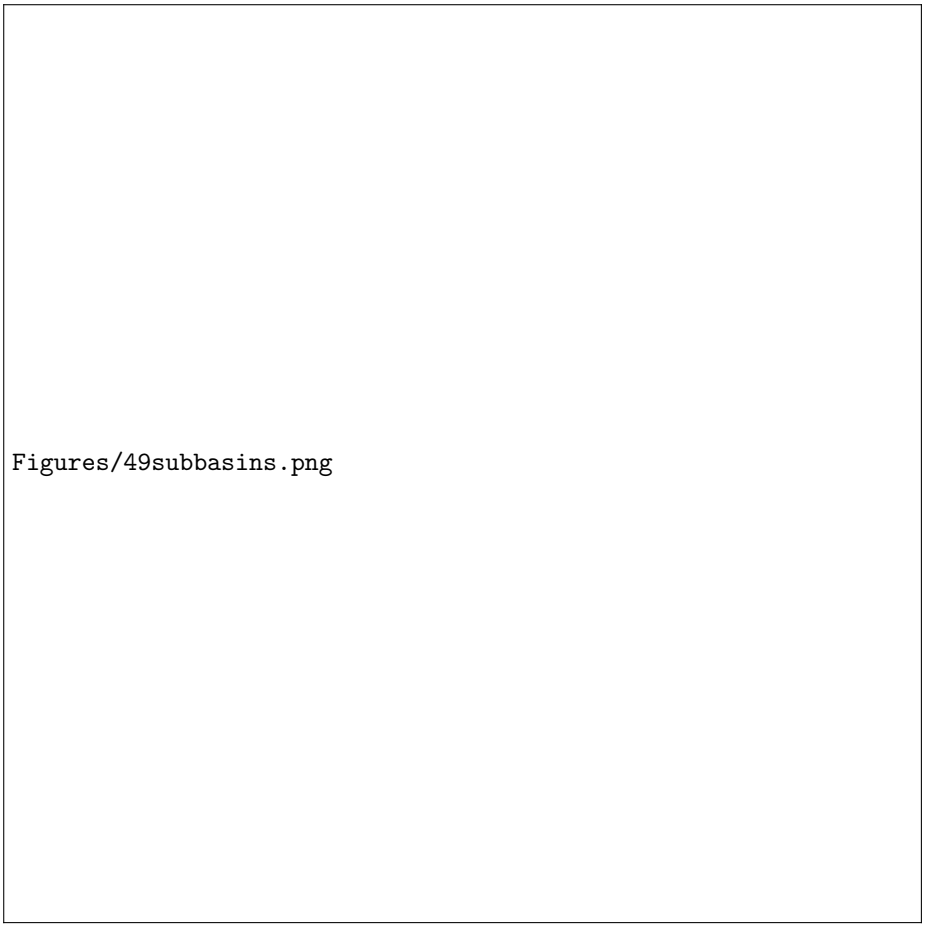



**Figure 8.** Delineation into 49 Sub-basins

**Table 3.** Details of the ANN setup

| | |
|---|---|
| Optimum Number of Hidden Layers | 2 |
| Hidden Nodes in First Layer | 2 |
| Hidden Nodes in Second Layer | 4 |
| Transfer Function | Sigmoid |

### *2.5.1 Spatially distributed inputs and lumped parameters (Case 2)*

This case pertains to spatially varying inputs, i.e. Precipitation and PET. Spatially averaged precipitation and PET were obtained for all the 49 sub-basins for the period spanning 01-01-1966 to 31-12-2014. Using the parameters from the lumped configuration, the model was run in the lumped mode for each of the 49 sub-basins. However, overall, the model was semi-distributed, and we were taking the spatial variations into account. The streamflow outputs from each of the 49 sub-basins were thus obtained and passed as inputs to the ANN as shown in Figure 9. The details of the ANN setup are mentioned in Table 3.

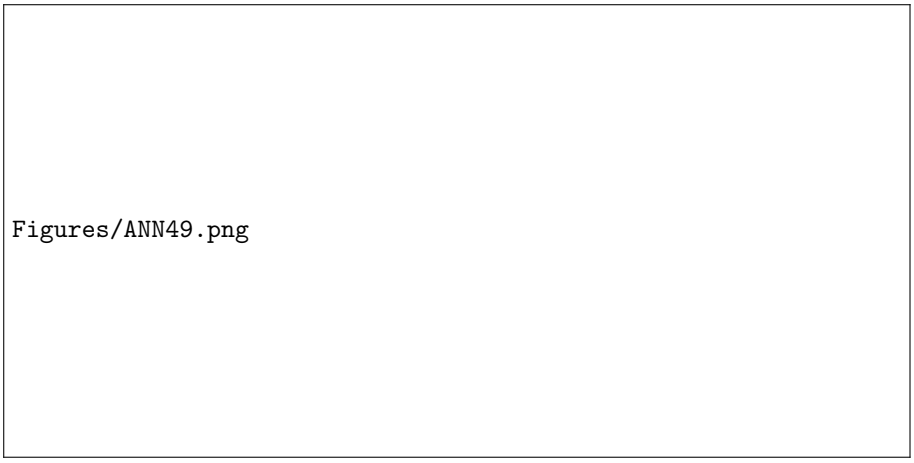



**Figure 9.** ANN setup in the hybrid model

### *2.5.2 Spatially distributed inputs and mixed parameters (Case 3)*

In this case, five sub-basins upstream of Godavari Arch (GR) bridge station on the Godavari river were merged and will be referred to as G.R. Bridge from here on as shown in Figure 10. The streamflow data at the G.R. bridge was available from 01-07-1976 to 31-12-2014. At the same time, the extracted IMD spatially averaged precipitation and PET for G.R. bridge was available from 01-01-1966 to 31-12-2014. For calibrating the lumped SAC-SMA model over the G.R. bridge sub-basin, observed streamflow corresponding to 01-01-1977 to 31-12-1995 was selected since the period comprises a sufficient number of high and low flow events for meaningful calibration of the model. This case differs from the previous case in the aspect that a mix of parameters has been used here. Parameters obtained by calibration at G.R. bridge were used along with lumped parameters over the rest of the 43 sub-basins. Hence spatial variability of the parameters has been accounted for partially. Again the streamflow outputs from the G.R. bridge and rest of sub-basins served as input to the ANN.

### *2.5.3 Fully semi-distributed modelling with spatially varying parameters (Case 4)*

Before proceeding with the fully semi-distributed configuration, we needed to determine the parameters for all 49 sub-basins. The parameter transfer was done in such a way for Godavari, Yelli and Konta basins so as to absorb as much information in the calibrated parameters so that the parameter transfer would be meaningful and effective. One of the most reliable signs suggesting that the transferred parameters represent the spatial and temporal variability in the Godavari basin would be the difference in parameter values among the eastern and western regions of Godavari basins as will be discussed in results section.

*2.5.3.1 Parameter transfer -* As seen in Figure 8, only basins with with no inlets and only one outlet could be calibrated. However, the streamflow data for many of these basin outlets was unavailable or not of sufficient length for effective calibration. Also, the basins which had stream order greater than one were effectively ruled out in terms of calibration since the contribution to the streamflow at the basin outlet was coming from upstream basins in addition to the basin under consideration. Thus we had to resort to parameter transfer techniques based on various factors which are listed below:

1. Climatic index, i.e., factors, e.g. ratio of Annual Precipitation to PET

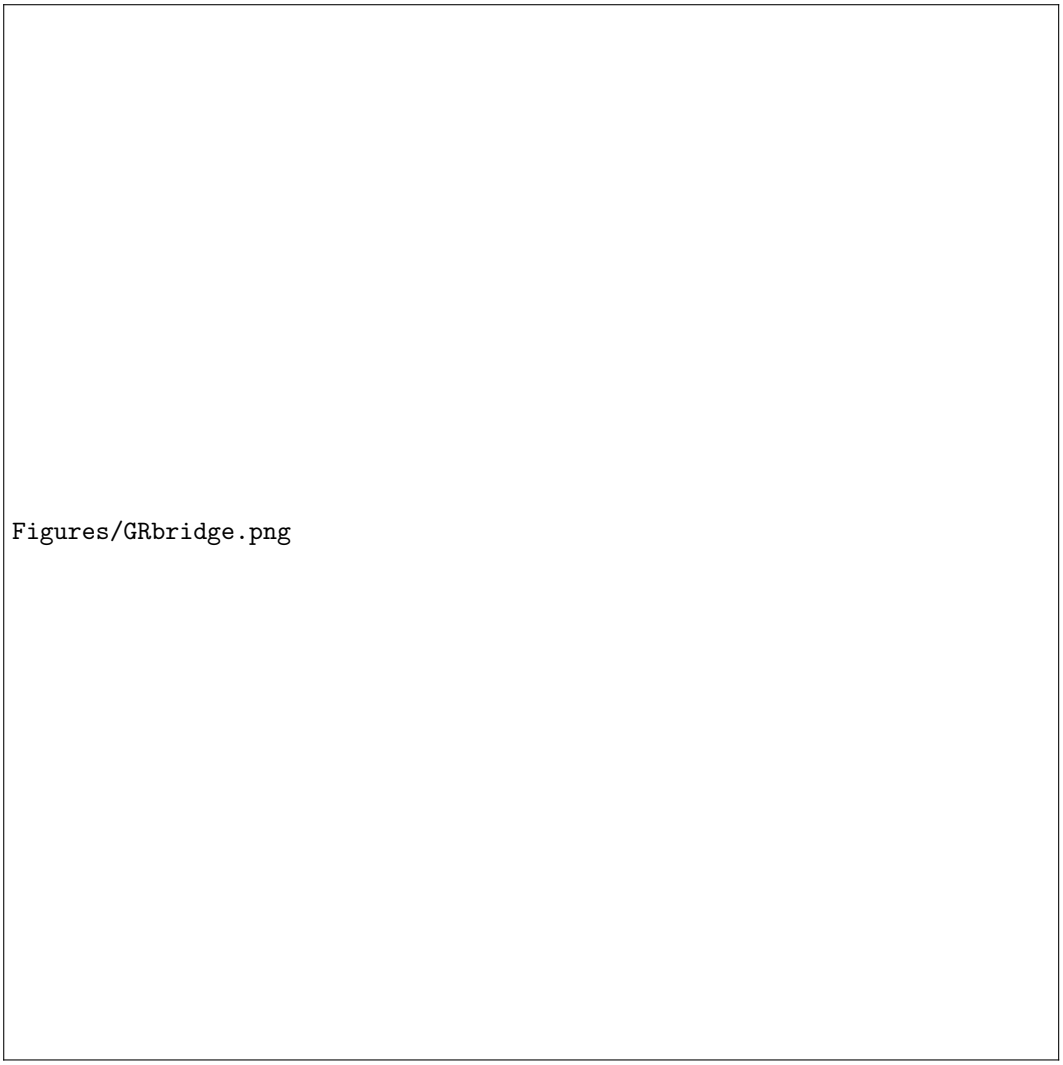



**Figure 10.** Merging of sub-basins leading to creation of G.R Bridge basin. Streamflow at G.R. Bridge outlet was used for model calibration.

2. Topographic factors e.g. elevation, terrain
3. Basin characteristics, e.g. stream density, area, land use

It is a reasonable assumption that the calibrated parameters reflect combined effects of all or a few of the factors that are mentioned above.

The climate index signifies the "wetness" of the basin (Koren et al., 2006). Figure 11 shows the spatial distribution of climate index over the 49 sub-basins. Since the precipitation and PET were readily available and of very high resolution , it was decided to use climate index solely for parameter transfer. This was because the climate index also influences the land use as will be seen later. As was already mentioned, most of the western Godavari basin lies on the leeward side of the Western Ghats and receives scanty rainfall as seen by a lower value of P/PE. Also, comparatively its hotter compared to the eastern parts of the basin. The eastern parts are frequented by cyclonic depressions originating in the Bay of Bengal result in extreme rainfall (*PMP Atlas for Godavari River Basin*, 2014). In Figure 11 a clear divide can be seen in terms of the climate index be-

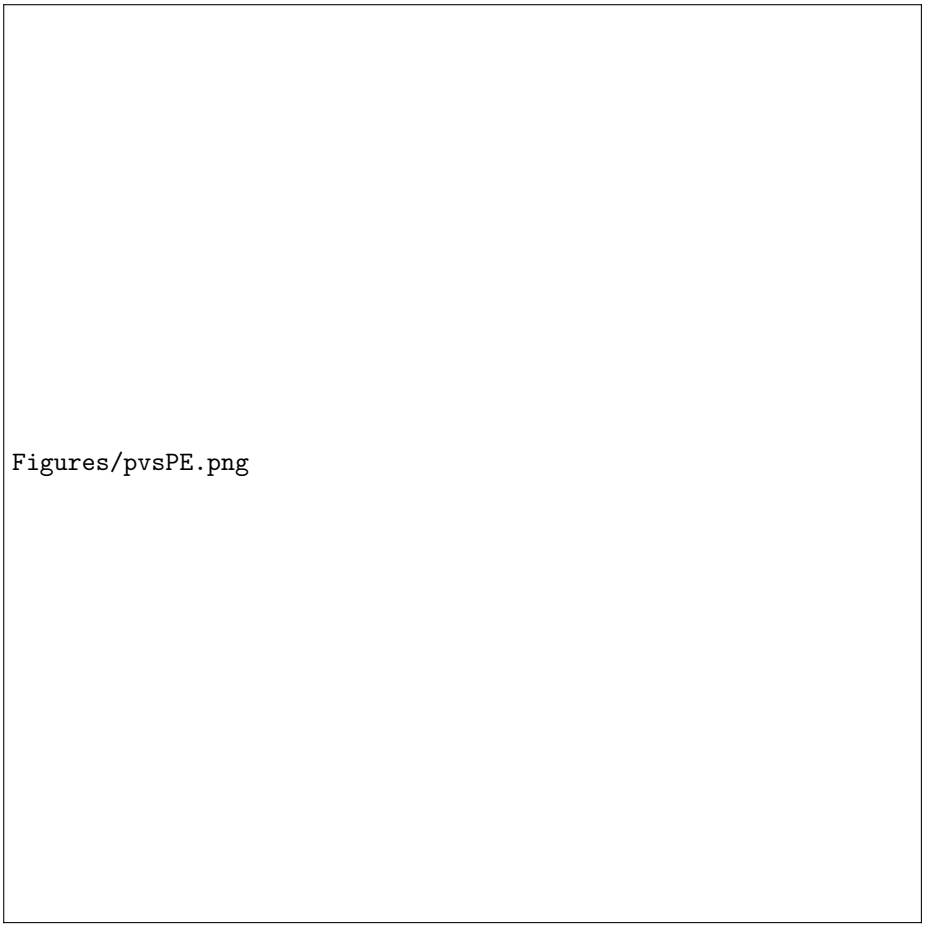



**Figure 11.** Spatial distribution of climate index (P/PE)

tween western and eastern parts of Godavari basin. It is assumed that this clear pattern must be may most likely reflected in the parameters corresponding to these regions. In this context, all the upstream sub-basins in the western part of the basin up to Yelli outlet were combined and treated as one single hydrologic unit, referred as Yelli sub-basin. Similarly upstream sub-basins in the eastern part of the basin up to Konta outlet were merged treating the entire region as one unit and referred as Konta region (or basin). The new sub-basins are shown in Figure 12. Differences in parameters emerging from these two basins assist in understanding the watershed behaviour under different hydroclimatic conditions and consequently in parameter transfer. Both Yelli and Konta outlet stations had daily streamflow records available for sufficient years for calibration. Yelli has a period of record from 01-06-1978 to 31-12-2011, i.e., approximately 33 years, while as Konta streamflow record spans from 01-01-1966 to 31-12-2014, i.e., approximately 48 years.

The decadal Land use land cover (LULC) maps for Godavari basin shown in Figures 13 and 14 were created by clipping the Indian LULC maps with the Godavari shapefile in ARC-GIS. Here only the maps correponding to years 1970 and 2000 have been shown. The maps for India were created by Moulds et al. (2018) using the Historic Land Dynamics Assessment (HILDA) land change model fed with district-level environmental and socio-economic data. The model allocates a particular pixel to a specific land use pattern by considering the environmental and socio-economic value of the land use under consideration. Table 4 gives the percentage of various land-use patterns from 1970 to 2000

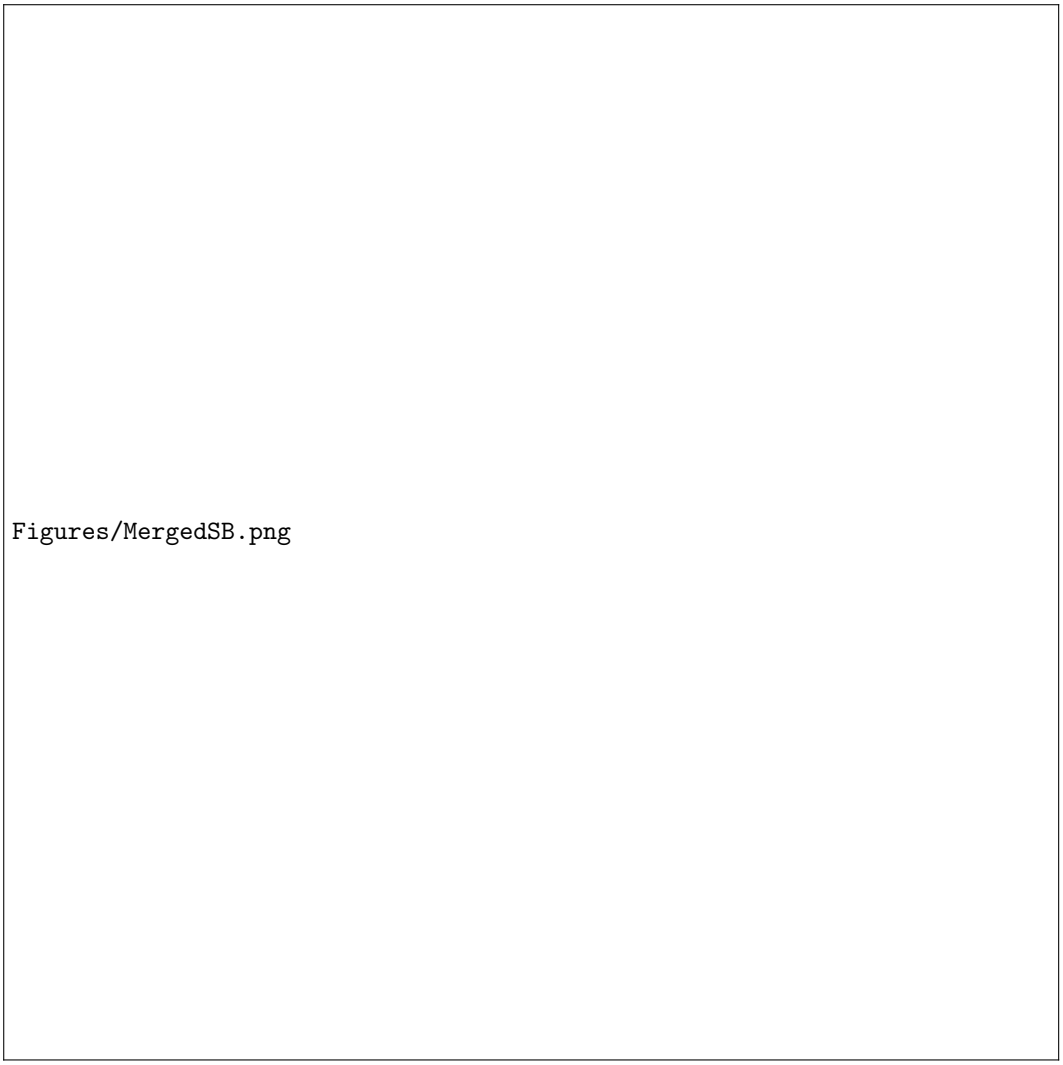



**Figure 12.** Yelli and Konta regions (shown in yellow)

for the Godavari basin and Yelli and Konta sub-basins. It can be noted that cropland and urban areas are increasing in the all three basins at the expense of forest area.The LULC pattern is somewhat similar to the climate index map with a stark contrast between western (Yelli) and eastern (Konta) regions of Godavari basin. This sharp divide in climate index and LULC pattern gives further credence that the parameters obtained for these two basins will be quite different from each other in magnitudes. The Yelli basin is dominated by agricultural activities where as Konta basin is less influenced by human activities. In the Yelli basin obstructions to natural flows has been created to facilitate agricultural activities and is one of the reasons for low flows in Godavari river in this basin.

The elevation and basin slope determine the amount of precipitation which gets converted into streamflow at the basin outlet. Table 5 gives the mean elevation, basin slope and Q/P ratio for Godavari, Yelli and Konta basins. Again it can be seen that the Q/P ratio for Yelli is far lower than Konta, where almost 50 per cent of rainfall gets converted into streamflow due to a steeper slope in comparison to Yelli. The Godavari basin lies in the middle since in a way it averages out both the basins.

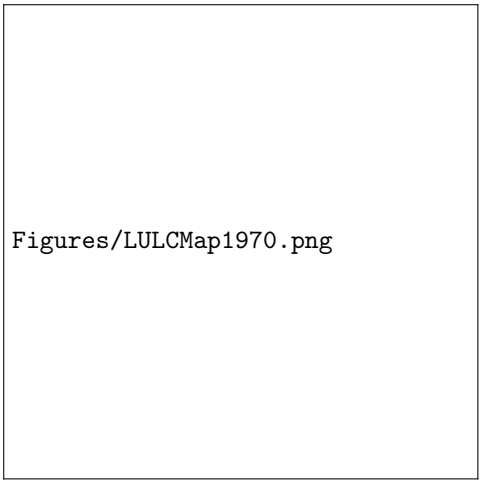



**Figure 13.** LULC Map for year 1970

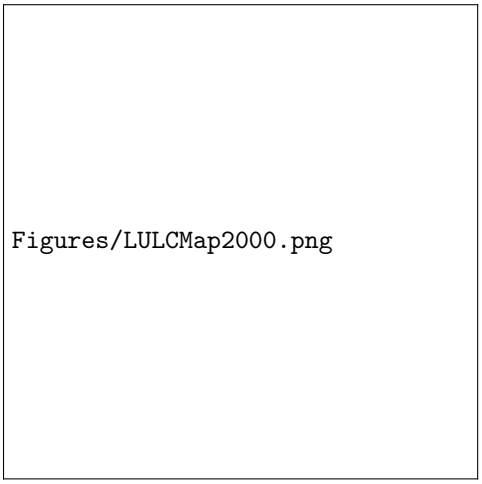



**Figure 14.** LULC Map for year 2000

**Table 4.** Percentage of various LULC classes

| Year | Cropland | | | Forest | | | Urban | | | Water bodies | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Year | G | Y | K | G | Y | K | G | Y | K | G | Y | K |
| 1970 | 57.66 | 87.63 | 29 | 29.34 | 3.74 | 50 | 0.51 | 0.46 | 0.2 | 3.6 | 0.62 | 3.34 |
| 1980 | 57.86 | 87.58 | 32.7 | 29.52 | 3.71 | 48.26 | 0.72 | 0.64 | 0.24 | 3.6 | 0.62 | 3.34 |
| 1990 | 58.05 | 87.625 | 36.73 | 29.37 | 3.62 | 44.79 | 0.76 | 0.79 | 0.48 | 3.6 | 0.62 | 3.34 |
| 2000 | 58.33 | 87.61 | 39.62 | 29.33 | 3.68 | 42.71 | 1.15 | 0.84 | 0.61 | 3.6 | 0.62 | 3.34 |

G- Godavari basin, Y- Yelli sub-basin, K- Konta sub-basin

**Table 5.** Mean Elevation, Basin slope and its effect on Q/P ratio

| Mean Elevation (m) | | | Basin slope | | | Q/P | | |
|---|---|---|---|---|---|---|---|---|
| G | Y | K | G | Y | K | G | Y | K |
| 838.5 | 624.7 | 648.76 | 7.65 | 5.98 | 12.86 | 0.25 | 0.1 | 0.49 |

G- Godavari basin, Y- Yelli sub-basin, K- Konta sub-basin

Knowing the factors influencing the SAC-SMA parameters, the next step was to create a sort of "parameter bank " which would be used for parameter transfer. Accounting for spatial and temporal variability of the parameters is important since that would influence the parameter transfer. This was achieved in the following ways:

1. Calibrating the SAC-SMA model on a three year moving window (blocks of 3-years and each block will have a overlap to 2-years with the previous block) for Godavari, Yelli and Konta basins for their respective period of record. Three model runs were performed for each basin, and 46 sets of calibrated parameters were obtained per run for Godavari and Konta while as 30 sets were obtained for Yelli. The calibrated parameters were related to the P/PE ratio for the corresponding three years moving average window for all three basins.
2. By including the effects of calibration period since the period of record of observed streamflow differs across basin outlets and influences calibration. Thus for the Godavari, Konta and Yelli, the data set that is available for calibration will increase every year, therefore, can be treated as updation of parameters every year.

*2.5.3.2 Methods of parameter transfer-* The parameter transfer was accomplished using the methods described below:

1. **Linear Regression (L.R.)** - It was considered as a baseline method and was expected to perform the worst since a linear relationship was assumed between the parameters and the climate index (P/PE) when it is not the case as shown in Figures 15 and 16.
2. **ANN's**- Knowing that ANN's can capture the non-linear relationships between the parameters and climate index, it made them ideal for this case. The ANN network was trained on the "parameter bank" and parameters were estimated for the 49 sub-basins. The ANN model setup showing the estimated parameters as output nodes and climate index as input node is depicted in Figure 17.
3. **K-Means algorithm**- The K Means algorithm is also suited for parameter transfer problems. The known parameters were grouped into 49 clusters based on climate index values. The algorithm allotted the parameters based on these clusters to the 49 sub-basins. This method captured most variability among the parameters.

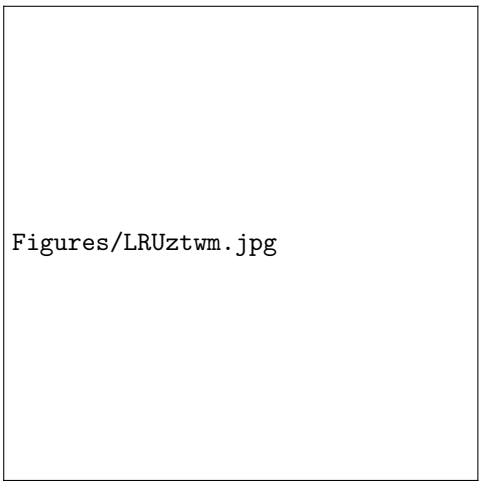



**Figure 15.** Linear regression method for parameter (Uztwm) estimation

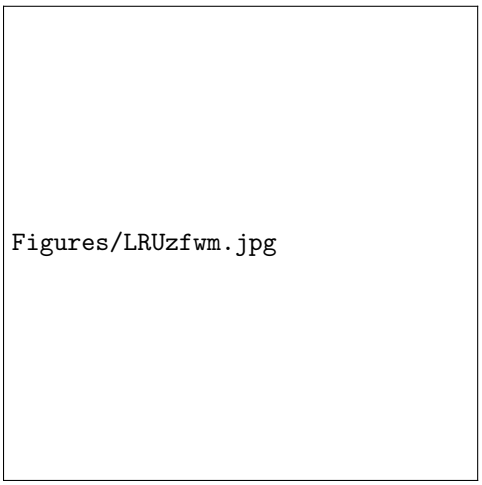



**Figure 16.** Linear regression method for parameter (Uzfwm) estimation

### 2.6 Semi-distributed with lumped inputs with spatially varying parameters (Case 5)

This configuration can be seen as the opposite of the second case where the inputs were spatially distributed. In this framework, lumped inputs as used in original SAC-SMA configuration were combined with the 'best' performing distributed parameters from Case 4.

## 3 Results and Discussion

### 3.1 SAC-SMA in lumped configuration (Case 1)

SAC-SMA has 16 parameters out of which only 13 parameters were optimised using automatic calibration as given in Table 6. The remaining three parameters were left at their defaults as per recommendations of Peck (1976). The visual comparison of performance for the calibration and validation periods is shown in Figure 18 with the over-

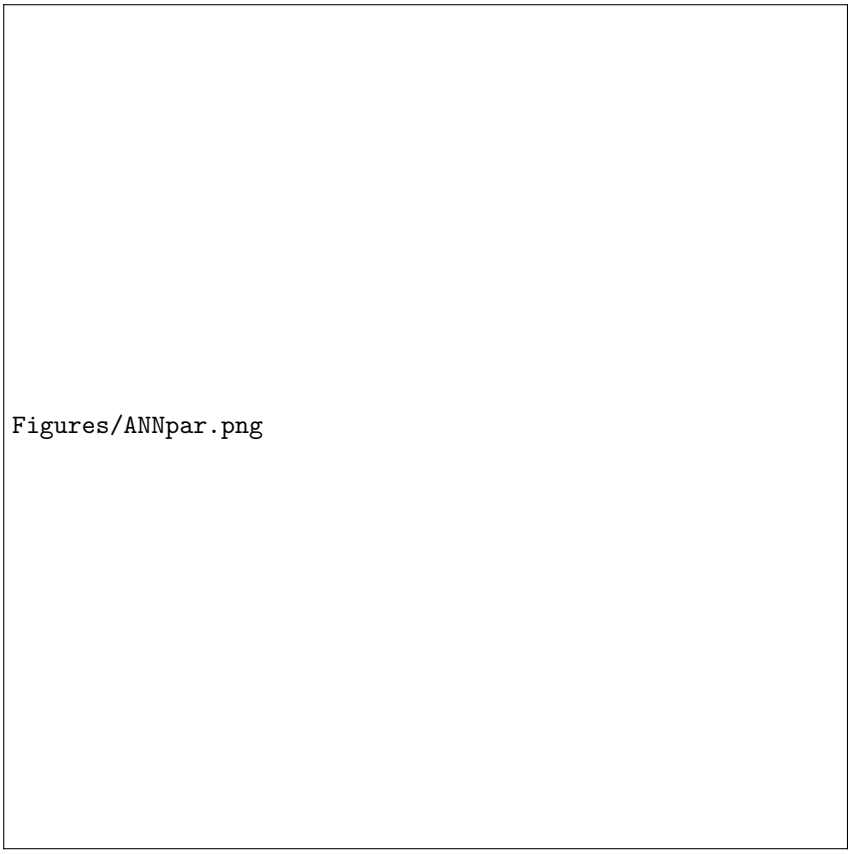



**Figure 17.** ANN model setup for parameter estimation

all verification metrics given in Table 7 . Both, visual verification of hydrographs and numerical verification metrics indicate good agreement between the simulated and observed streamflows and emphasize the simpler lumped modeling. In lumped mode, as the values are averaged over the larger area spatial variations in model inputs consequently in model output decreased, therefore importance of these variations as-well. Nevertheless, improvements in simulations can be pursued by the hybrid model.

**Table 6.** Optimised SAC-SMA parameters

| Optimised | | | | | | | | | | | | | Not optimised | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Uztwm | Uzfwm | Uzk | Pctim | Adimp | Zperc | Rexp | Lztwm | Lzfsm | Lzfpm | Lzsk | Lzpk | Pfree | Riva | Side | Rserv |
| 122.06 | 149.68 | 0.49 | 0.07 | 0.39 | 7.97 | 4.96 | 312.5 | 543.43 | 678.88 | 0.01 | 0.08 | 0.04 | 0 | 0 | 0.3 |

Like many other hydrological models, SAC-SMA has its own limitation in the estimation of high and low flows. Probability plot, i.e., quantiles of streamflow data plotted against quantiles of the standardized normal distribution, is developed separately for observed and simulated streamflows in Figure 19. Deviation from the straight line suggests that streamflows do not follow Gaussian distribution and both observed and simulated streamflows differ in their distribution at high- and low- flows, of which differences in distribution of low flows are relatively high. This is so since the model is structurally tuned to capture high flows (Onyutha, 2019). The objective function used here "R Squared " also gives more weightage to high flows than low flows. This emphasizes the need to

**Figure 18.** Visual comparison of model performance for calibartion and validation periods

analyze and assess the high and low flows separately as shown in Figure 20 . There is no rationale in selecting 4 and 1/4 times mean criteria and is entirely arbitrary. The daily flows which were common between observed and simulated flows were selected since comparison could be made of equal length data. Lower flows dominate in the Godavari basin except in monsoon months leading to relatively larger sample size for the low flows. The verification metrics for high and low flows are tabulated in Table 8. The better simulation of the high flows is one of the reasons for the use of SAC-SMA as a flood forecasting model. However, in Indian conditions, many of the rivers carry low flows , and the model performance needs to improve in this aspect. Thus the need for a hybrid model. Overall the SAC-SMA model by itself has simulated the flows well and any improvements in model performance will come from improving the simulation of high and low flows. Beyond some point, it will not be worth the time and effort to improve the model performance further due to imperfections in model structures, irregularities in inputs and uncertainties in parameters.

### 3.2 Spatially distributed inputs and lumped parameters (Case 2)

Based on the verification metrics as given in Table 7, including the spatially distributed input information leads to better simulations.The visual comparisons of observed and simulated hydrographs are in Figure 21. Although the improvements may appear marginal, they are significant when we look at high and low flows (Table 8). We can see that the improvement in model performance is mainly result better simulation of high flows. In both Cases 1 and 2, the model overestimates the high flows albeit more in Case 2. However, the lumped SAC-SMA model (Case 1) underestimates the low flows in comparison to the Case 2 configuration of hybrid model as depicted in Figure 22. The negative values for the NSE indicate, at least for low flows, the predictive power of Case 2 configuration of hybrid model decreases with respective to the lumped SAC-SMA model (Case 1). A constant value such as mean of daily low flows can be more in agreement and better than the models used.

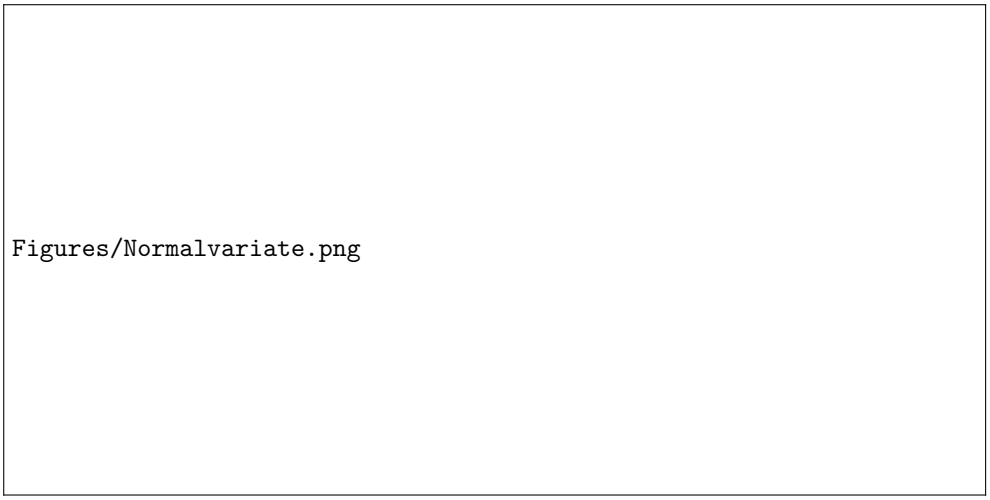



**Figure 19.** Streamflow plotted against standard normal variate

**Figure 20.** Criteria for high and low flows

### 3.3 Spatially distributed inputs and mixed parameters (Case 3)

With the parameters also spatially varying, the overall results are not significantly different from Case 2 as given in Table 7. However, the model performance is slighly different when low- and high- flows are looked at separately (Table 8). Incorporating spatial variability in hydrologic information mainly led to improved simulations of high flows with degradation in the performance of low flows in line with the results of Boyle et al. (2001).

### 3.4 Fully semi-distributed modelling with spatially varying parameters (Case 4)

#### *3.4.1 Evolved parameters*

One of the constituents of the "parameter bank" created were the evolved parameters. The mean of 125 parameter sets (47 for the Godavari, Konta and 31 for Yelli) is

**Figure 21.** Visual comparison of observed and simulated hydrographs for various scenarios

**Table 7.** Verfication metrics (Daily scale) for all scenarios

| Case | MAE | RMSE | PBIAS | NSE |
|---|---|---|---|---|
| 1 | 0.24 | 0.55 | 8 | 0.88 |
| 2 | 0.21 | 0.49 | 6.6 | 0.90 |
| 3 | 0.21 | 0.49 | 6.6 | 0.90 |
| 4 | 0.25 | 0.61 | 5 | 0.85 |
| 5 | 0.24 | 0.56 | 7.4 | 0.87 |

depicted in the bar plot in Figure 23 . It can be observed that Uztwm and Lztwm are highest for Yelli in comparison to the Konta basin. This is due to presence of significant irrigation and storage structures present in the Yelli sub-basin which influences the initial loss and detention storage and subsequently the Uztwm parameter. Due to mean annual temperature in Yelli being 4-5 degrees higher than Konta basin, the evapotranspiration losses as represented by the Lztwm parameter acquire a higher value. The highest value of Lzfpm and Lzpk for Yelli indicates that most of the flow contribution is from base flow component, which is understood since the basin is primarily semi-arid. Uzfwm and Uzk parameters are highest for Konta signalling a significant contribution from surface runoff and that the streamflow may be particularly flashy due to steeper slope as was noted from Q/P ratio in Table 5. Similarly, Lzsk and Lzfsm are also highest for Konta indicating a quick release of baseflow and conversion of most of the rainfall to runoff at outlet amongst all the basins considered. This is facilitated by the maximum value of Zperc also for Konta, which increases the supply of water to the lower zone by percolation.

So far, it has been seen that the SAC-SMA model struggles to simulate the extreme flows well. This was attributed to the parameters not being "activated" fully and also the choice of the objective function. The "activation" of the parameters can be explained with the help of the evolved parameters from the point of view of stability. Figure 24 shows the stability of the evolved parameter (Uztwm) for the calibration period. Although results here have been shown for Uztwm, similar behaviour was observed for most of the parameters. The blue curves are the 95 % confidence bounds and the red line being the

**Table 8.** Verfication metrics (High and low flows) for all scenarios

| Case (High Flows) | MAE | RMSE | PBIAS | NSE |
|---|---|---|---|---|
| 1 | 1.26 | 1.68 | 0.3 | 0.56 |
| 2 | 1.13 | 1.52 | 1.5 | 0.65 |
| 3 | 1.09 | 1.49 | 0.8 | 0.66 |
| 4 | 1.43 | 2 | -2.6 | 0.48 |
| 5 | 1.28 | 1.71 | 0.2 | 0.54 |
| Case (Low Flows) | MAE | RMSE | PBIAS | NSE |
| 1 | 0.03 | 0.04 | -2.8 | -0.02 |
| 2 | 0.04 | 0.05 | 39.9 | -0.43 |
| 3 | 0.04 | 0.05 | 39.9 | -0.44 |
| 4 | 0.04 | 0.05 | 27.5 | -0.24 |
| 5 | 0.04 | 0.05 | 34 | -0.62 |

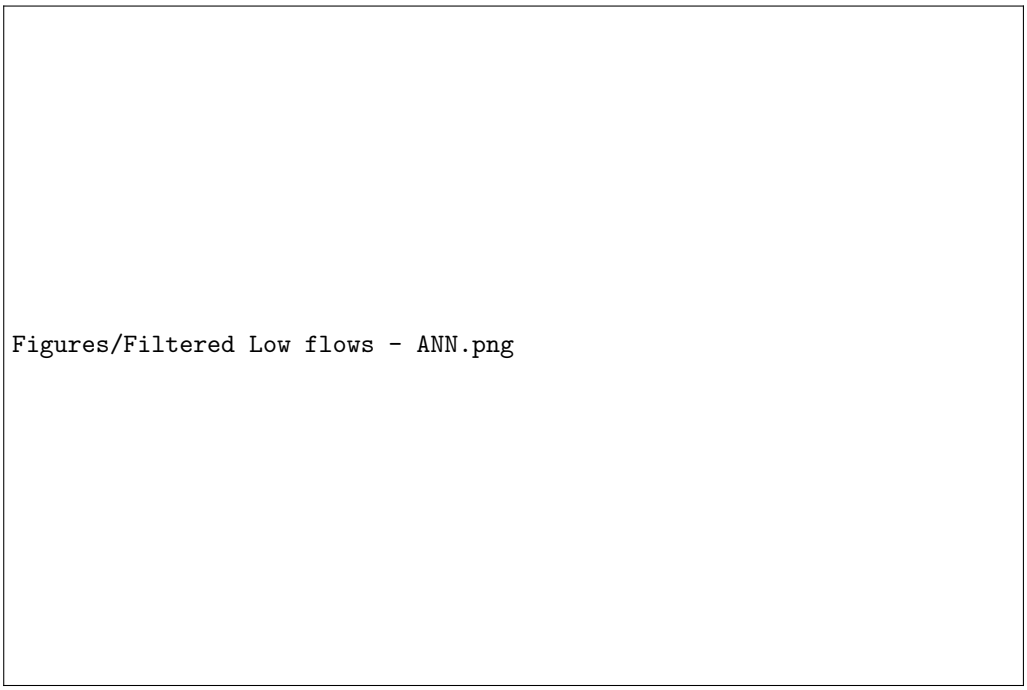



**Figure 22.** Scatter plot for comparison of low flows between cases 1 and 2

best fit line. It can be seen that the maximum number of points lie within the 95% confidence bounds for Konta and Godavari , and only a few points for the Yelli. This indicates that the parameters for Yelli are not well determined and may have higher uncertainty. It is known that the repeated cycles of extremes of dryness and wetness are required to determine the parameters with less uncertainty which may not be the case with Yelli parameters given that Yelli is a semi-arid basin. For example there will less events generating surface runoff in Yelli basin given the dominance of baseflow as was seen earlier. Thus the value of UZFWM will generally contain much uncertainty. The determination of parameters is well accomplished when there is sufficient runoff generated. In very dry regions, there is not enough runoff to be confident in any of model parameter values.

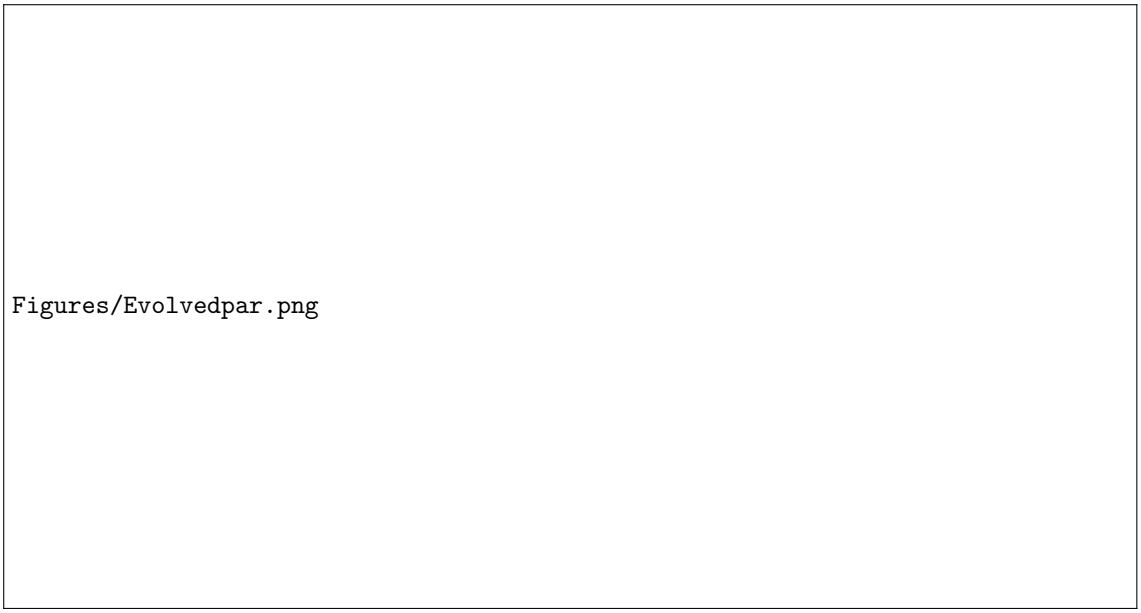



**Figure 23.** Mean of Evolved parameters for Godavari, Yelli and Konta

**Figure 24.** Stability of Evolved parameter (Uztwm) for (a) Godavari (b) Konta (c) Yelli

Figure 25 shows the bar plot of the standard deviation of the evolved parameters. Relatively high variance in capacity (Uzfwm, Uztwm) and withdrawal parameters (Lzpk, Uzk, Lzsk) for Yelli basin does reinforce the arguments made above. It is the reason for poor model performance in the Yelli basin. On the other hand, the Konta basin, which has the highest Q/P ratio also shows the relatively high variance in capacity parameters (Lzfpm, Lzfsm) leading to the high flows being over or under-estimated . The Godavari basin, which falls in the middle in terms of Q/P ratio and encompasses both the Yelli and Konta basins has the least variance in most of the parameters.

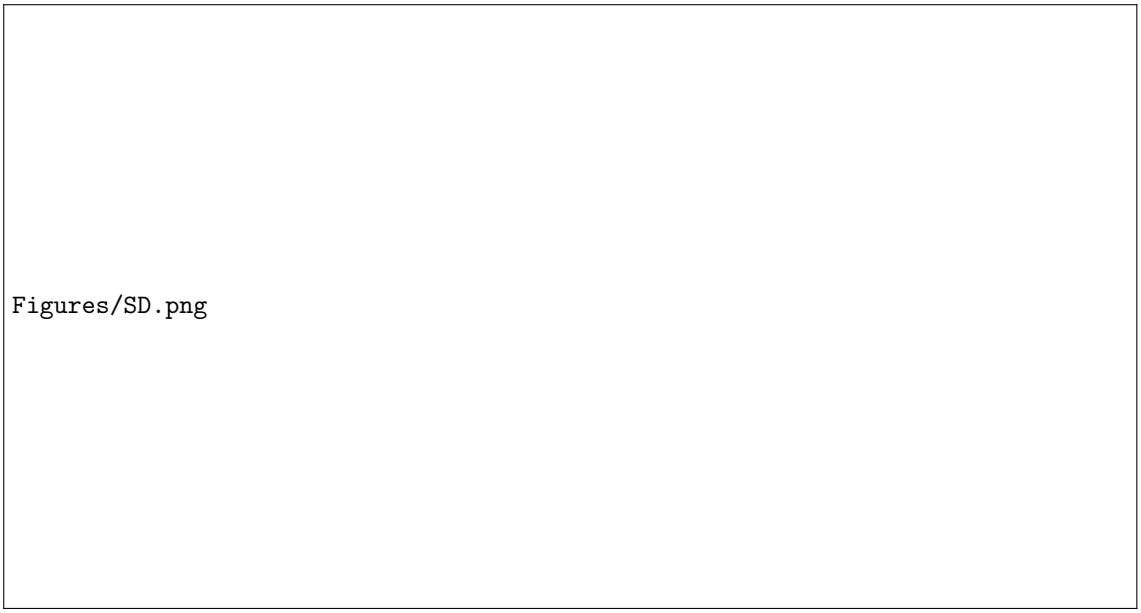



**Figure 25.** SD of evolved parameters. It can be seen that both Konta and Yelli which represent extreme cases in terms of Q/P ratio have higher variance than Godavari basin

### *3.4.2 Parameter Transfer*

Now knowing the reasons for the decreased model performance in some basins, we now focus on the results of parameter transfer.

1. **Linear Regression derived parameters**- The baseline linear regression method captured a small range of parameter variations as seen in Figure 26. Also, the higher values of Uztwm and Lztwm for Western regions of Godavari are due to a higher temperature, intensive agricultural activities . A relatively higher value of Uzfwm and Uzk over the eastern portions of Godavari indicates a significant contribution from surface runoff to total runoff. The base flow dominates the streamflow contribution in the western Godavari as implied by the higher value of the product of Lzpk and Lzfpm. The percolation parameter Zperc is showing its maximum values towards the eastern parts of Godavari to dissipate the water from upper to lower zones quickly. Thus we see the method is able to replicate some of the parameter trends seen for the evolved parameters which are one of the components of the "parameter bank".
2. **ANN derived parameters**- ANN improves upon the range captured by the linear regression method, as shown in Figure 26. For the Uztwm parameter derived by ANN, the spatial distribution represented a range from 32-112 mm, which is a significant improvement over the baseline linear regression method. Thus the ANN was able to capture the extreme values of the parameters over the basins. Although the overall results are similar to those obtained by linear regression, more spatial variation of the parameters was captured.
3. **K Means derived parameters**- The parameters applied from this method were directly sourced from the "parameter bank" and hence reflected the maximum variations in the values as was the case for the "parameter bank". As illustrated in Figure 26, the range of values captured by this method for Uztwm is 12-144 mm. Thus with this method, the spatial distribution of the parameters improves further.

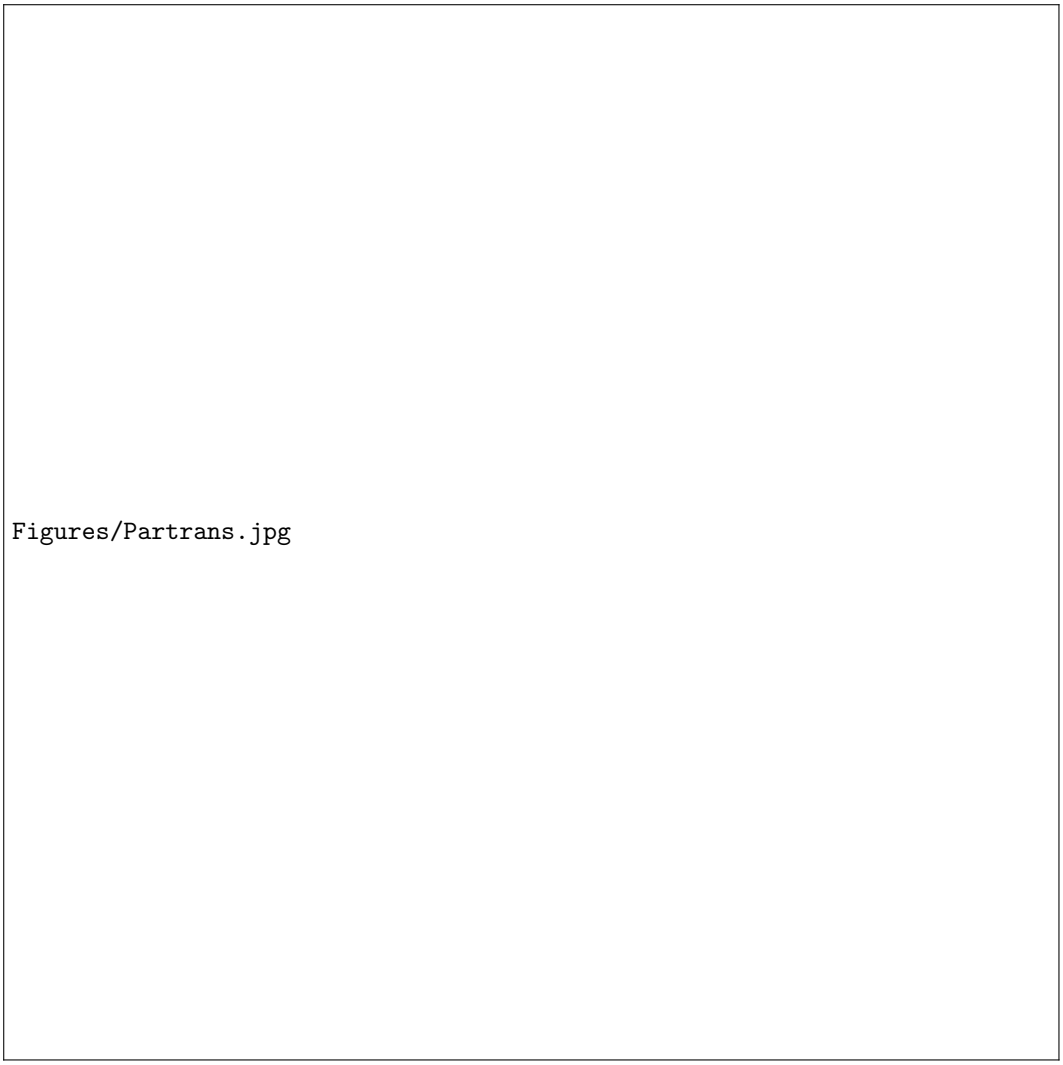



**Figure 26.** Spatial distribution of (a) Linear regression (b) ANN (c) K-Means derived parameters. Only Uztwm, Uzfwm, Lzfpm, Lzfsm and Zperc parameters have been shown

### *3.4.3 Hybrid model with derived parameters*

We can postulate that the methods which capture most of the parameter values and the underlying spatial distribution will outperform other methods. This can be verified in terms of how well the hybrid model reproduces overall streamflow at Pollavaram.

1. **Hybrid model with LR parameters**- A significant drop in model performance is seen in comparison to all the methods discussed so far (Table 9). This is expected since the parameters have been derived in the absence of any streamflow data for many of the sub-basins and an assumption of linearity between parameters and climate index.
2. **Hybrid model with ANN parameters**- The use of ANN derived parameters produced an improvement in model performance with the results comparable to the SAC-SMA lumped model as shown in Table 9. The ANN was able to capture the underlying non-linearity between parameters and climate index. The ANN de-

rived parameters yielded the best results and were selected for the next model configuration (Case 5) .

3. Hybrid model with K-Means parameters- The hybrid model run with parameters derived by the K-Means method closely followed the hybrid model with ANN parameters in terms of model performance (Table 9). As was our initial guess, the methods which captured the inherent spatial variations in the parameters did outperform the baseline (L.R) simulated parameters.

### 3.5 Semi-distributed with lumped inputs with spatially varying parameters (Case 5)

The distributed parameters derived from ANN's in the previous configuration (Case 4) were used along with lumped inputs. As seen from Figure 21 and Table 7, the model simulations did improve in comparison to the fully semi-distributed configuration . It seems the model performance is more sensitive to the spatial variation of inputs than the parameters. Based on this reasoning, we would have expected then that the fully semi-distributed configuration would outperform other hybrid model configurations. However, it is not the case since going fully semi-distributed also leads to an increase in complexity due to a multifold increase in parameters and underlying uncertainty in parameters. This may have been the reason for a deterioration in performance in case of fully semi-distributed configuration.

### 3.6 Selection of the best model framework

Of all the model configurations (Case 1 to 5), the configuration with spatially distributed inputs and lumped parameters (Case 2) was the best performing edging out the model configuration (Case 3) based on improved simulations of low flows, utilisation of existing parameter sets obtained from lumped calibration, taking spatial variations of inputs into account and less intensive (Compuataionally and time required to setup).

**Table 9.** Verification metrics (Daily flows) for hybrid model with the use of derived parameters

| Method | MAE | RMSE | PBIAS | NSE |
|---|---|---|---|---|
| Linear Regression | 0.28 | 0.84 | 9 | 0.71 |
| ANN | 0.25 | 0.61 | 5 | 0.85 |
| K-Means | 0.24 | 0.64 | 8.9 | 0.83 |

## 4 Conclusions

In this study, a hybrid model framework was developed for rainfall-runoff modelling for Godavari basin. The hybrid model consisted of the well-validated SAC-SMA model and ANN component. The ANN's, with their structural flexibility, improved the runoff estimates from the lumped SAC-SMA model. It was seen that increasing the model complexity does not necessarily improve the accuracy of the simulations as fully semi-distributed hybrid model perform as well as the standalone SAC-SMA model . It was a result of the increase in the total number of parameters, the challenges in the calibration of the semi-distributed form of the SAC-SMA and uncertainty associated with parameters . The uncertainty was higher for calibrated parameters derived from basins with low flows as that they were poorly determined.

For the fully semi-distributed configuration, parameter transfer methods (ANN and K-Means) accounting for non-linearity, variations and based on similar hydro-climatic conditions performed the best. Overall the hybrid configuration with distributed inputs and lumped parameters (Case 2) was found the most suitable in terms of simulating the flows and the effort required.The improved results of the Case 2 configuration indicate that the model is more sensitive to the spatial variation of rainfall and PET than the parameters. This may be a basin specific conclusion since we could not test this on other basins.

**Acknowledgments**

The authors acknowledge Ranga Reddi, Raghu Ram and Mohan, CWC Kishna and Godavari Basin Organization (KGBO) for providing streamflow data and other basin related information.The authors are grateful to PhD students Vimal .C. Sharma and M. Azharudin for valuable comments during this work and T. Prasad for providing code for lumped rainfall calculations.

## References

Anderson, E. A. (2002). Calibration of conceptual hydrologic models for use in river forecasting. *Office of Hydrologic Development, US National Weather Service, Silver Spring, MD*.

Andrews, F., & Guillaume, J. (2018). *Hydromad: hydrological model assessment and development.*

Armstrong, B. L. (1978). Derivation of initial soil moisture accounting parameters from soil properties for the national weather service river forecast system.

Bhawan, S., & Puram, R. (2014). Watershed atlas of india.

Bodman, G., & Colman, E. A. (1944). Moisture and energy conditions during downward entry of water into soils 1. *Soil science society of america journal*, *8*(C), 116–122.

Boyle, D. P., Gupta, H. V., Sorooshian, S., Koren, V., Zhang, Z., & Smith, M. (2001). Toward improved streamflow forecasts: Value of semidistributed modeling. *Water resources research*, *37*(11), 2749–2759.

Burnash, R., & Ferral, L. (1996). Conceptualization of the sacramento soil moisture accounting model. *NWSRFS Users Manual, Part II*, *3*.

Chen, J., & Adams, B. J. (2006). Integration of artificial neural networks with conceptual models in rainfall-runoff modeling. *Journal of Hydrology*, *318*(1-4), 232–249.

Commission, C. W., et al. (2006). Integrated hydrological data book (non-classified river basins). *Hydrological Data Directorate, Information Systems Organization Water Planning and Projects Wing, Central Water Commission, New Delhi, India*, *383*.

Corzo, G., Solomatine, D., De Wit, M., Werner, M., Uhlenbrook, S., Price, R., et al. (2009). Combining semi-distributed process-based and data-driven models in flow simulation: a case study of the meuse river basin. *Hydrology & Earth System Sciences Discussions*, *6*(1).

Daliakopoulos, I. N., & Tsanis, I. K. (2016). Comparison of an artificial neural network and a conceptual rainfall–runoff model in the simulation of ephemeral streamflow. *Hydrological Sciences Journal*, *61*(15), 2763–2774.

Daniell, T. (1991). Neural networks. applications in hydrology and water resources engineering. In *National conference publication- institute of engineers. australia.*

Gaume, E., & Gosset, R. (2003). Over-parameterisation, a major obstacle to the use of artificial neural networks in hydrology?

Gujja, B., Ramakrishna, S., Goud, V., et al. (2006). *Perspectives on polavaram, a*

*major irrigation project on godavari*. Academic Foundation.

Halff, A. H., Halff, H. M., & Azmoodeh, M. (1993). Predicting runoff from rainfall using neural networks. In *Engineering hydrology* (pp. 760–765).

Hamon, W. R. (1960). *Estimating potential evapotranspiration* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Jakeman, A., Littlewood, I., & Whitehead, P. (1990). Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. *Journal of hydrology*, *117*(1-4), 275–300.

Jenson, S. K., & Domingue, J. O. (1988). Extracting topographic structure from digital elevation data for geographic information system analysis. *Photogrammetric engineering and remote sensing*, *54*(11), 1593–1600.

Julien, P. Y., & Halgren, J. (2014). Hybrid hydrologic modeling. *Handbook of engineering hydrology, modeling climate changes and variability*, 331–352.

Koren, V., Moreda, F., Reed, S., Smith, M., & Zhang, Z. (2006). Evaluation of a grid-based distributed hydrological model over a large area. *IAHS Publ*, *303*, 47–56.

Koren, V., Smith, M., & Cui, Z. (2014). Physically-based modifications to the sacramento soil moisture accounting model. part a: Modeling the effects of frozen ground on the runoff generation process. *Journal of Hydrology*, *519*, 3475–3491.

Moulds, S., Buytaert, W., & Mijic, A. (2018). A spatio-temporal land use and land cover reconstruction for india from 1960–2010. *Scientific data*, *5*, 180159.

on Application of Artificial Neural Networks in Hydrology, A. T. C. (2000a). Artificial neural networks in hydrology. ii: Hydrologic applications. *Journal of Hydrologic Engineering*, *5*(2), 124–137.

on Application of Artificial Neural Networks in Hydrology, A. T. C. (2000b). Artificial neural networks in hydrology. i: Preliminary concepts. *Journal of Hydrologic Engineering*, *5*(2), 115–123.

Onyutha, C. (2019). Hydrological model supported by a step-wise calibration against sub-flows and validation of extreme flow events. *Water*, *11*(2), 244.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall–runoff model?: Part 2—towards a simple and efficient potential evapotranspiration model for rainfall–runoff modelling. *Journal of hydrology*, *303*(1-4), 290–306.

Peck, E. L. (1976). *Catchment modeling and initial parameter estimation for the national weather service river forecast system* (Vol. 31). Office of Hydrology, National Weather Service.

Penman, H. L. (1948). Natural evaporation from open water, bare soil and grass. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, *193*(1032), 120–145.

*Pmp atlas for godavari river basin* (Vol. 1). (2014).

Shamir, E., Carpenter, T. M., Fickenscher, P., & Georgakakos, K. P. (2006). Evaluation of the national weather service operational hydrologic model and forecasts for the american river basin. *Journal of Hydrologic Engineering*, *11*(5), 392–407.

Sherman, L. K. (1932). Streamflow from rainfall by the unit-graph method. *Eng. News Record*, *108*, 501–505.

Singh, V. P., & Woolhiser, D. A. (2002). Mathematical modeling of watershed hydrology. *Journal of hydrologic engineering*, *7*(4), 270–292.

Tayfur, G. (2014). *Soft computing in water resources engineering: Artificial neural networks, fuzzy logic and genetic algorithms*. WIT Press.

Todini, E. (2007). Hydrological catchment modelling: past, present and future.

Todini, E. (2011). History and perspectives of hydrological catchment modelling. *Hydrology Research*, *42*(2-3), 73–85.

Tolson, B. A., & Shoemaker, C. A. (2007). Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research*, *43*(1).

Zhang, Z. (2003). Hydrologic model calibration in the national weather service. *Calibration of watershed models*, 133.

**Figure 1.**

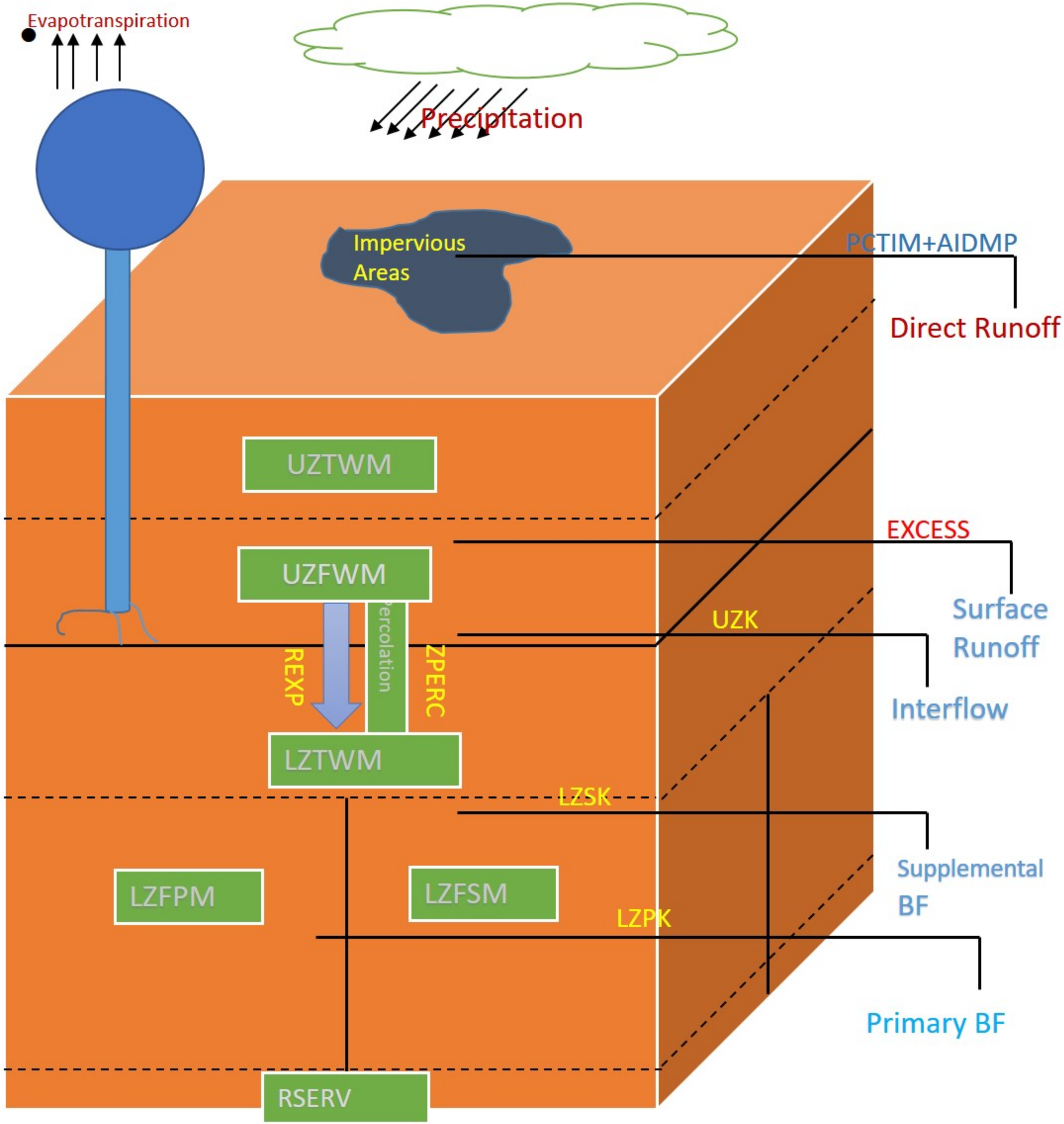
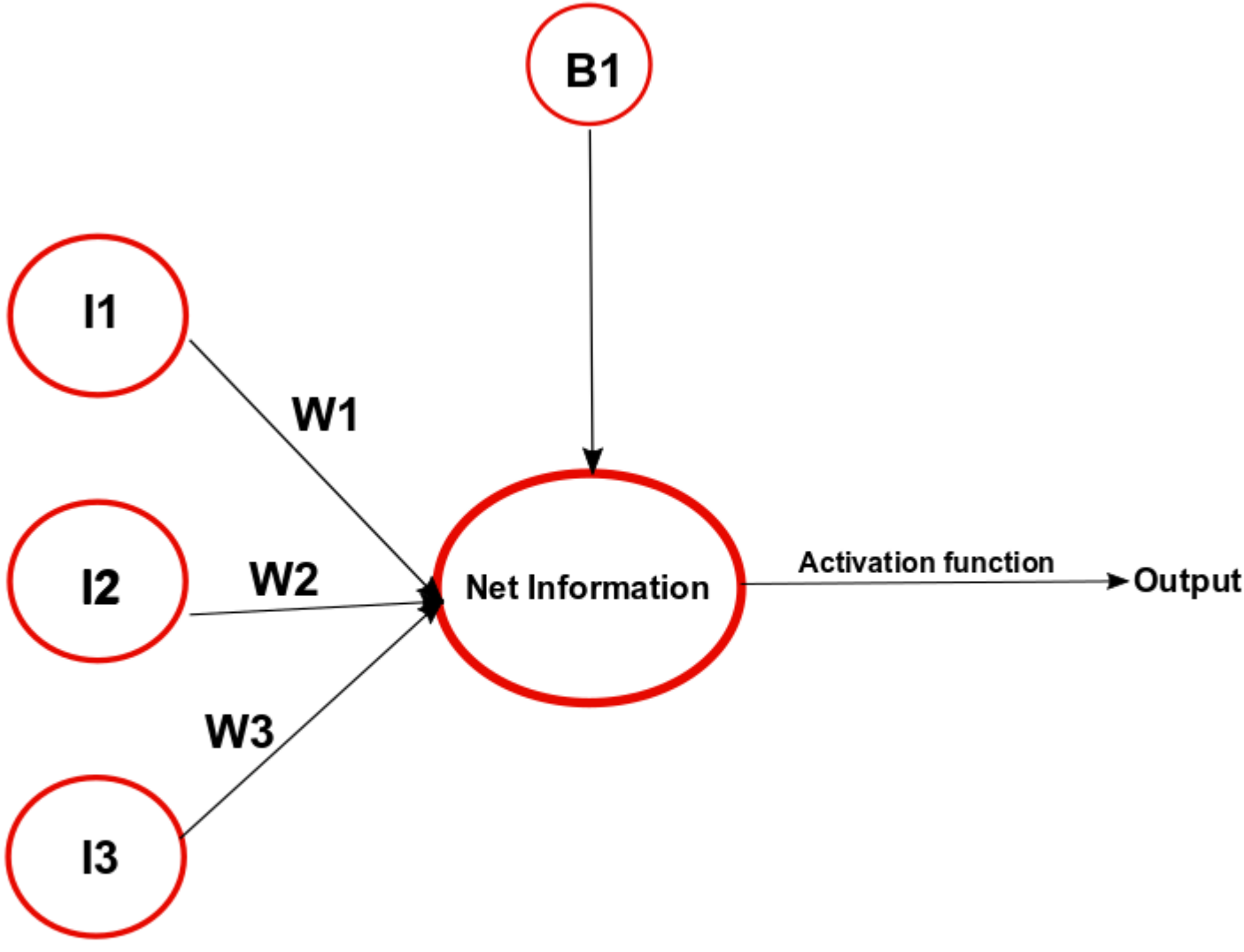

Evapotranspiration
Precipitation
Impervious Areas
PCTIM+AIDMP
Direct Runoff
UZTWM
EXCESS
UZFWM
Surface Runoff
UZK
REXP
Percolation
ZPERC
Interflow
LZTWM
LZSK
Supplemental BF
LZFPM
LZFSM
LZPK
Primary BF
RSERV

**Figure 2.**




$$\text{Net Information} = \sum_{i=1}^{i=3} I_i W_i - B_1$$

**Figure 3.**

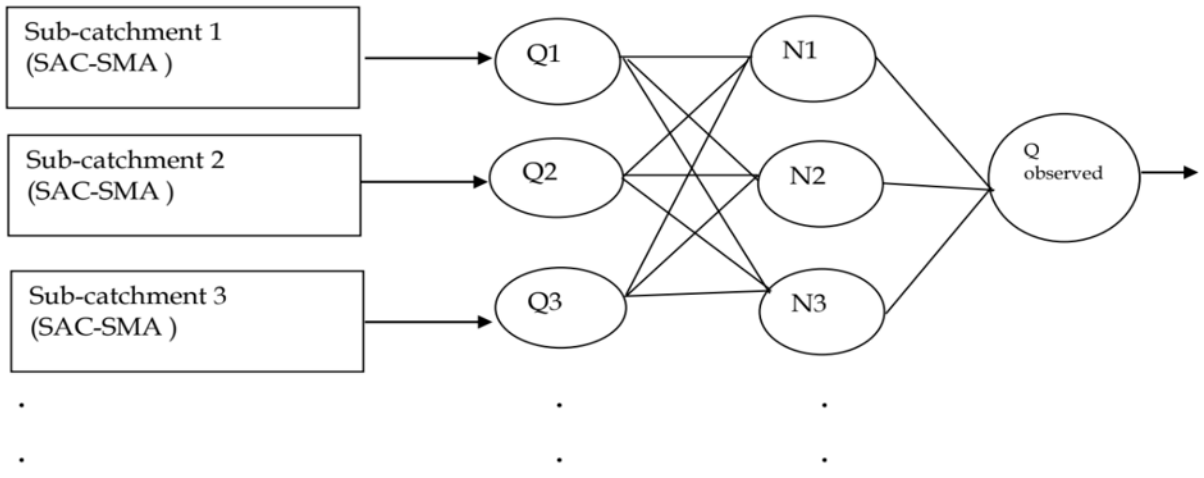

Sub-catchment 1
(SAC-SMA )
Sub-catchment 2
(SAC-SMA )
Sub-catchment 3
(SAC-SMA )
Q1
Q2
Q3
N1
N2
N3
Q
observed

**Figure 4.**

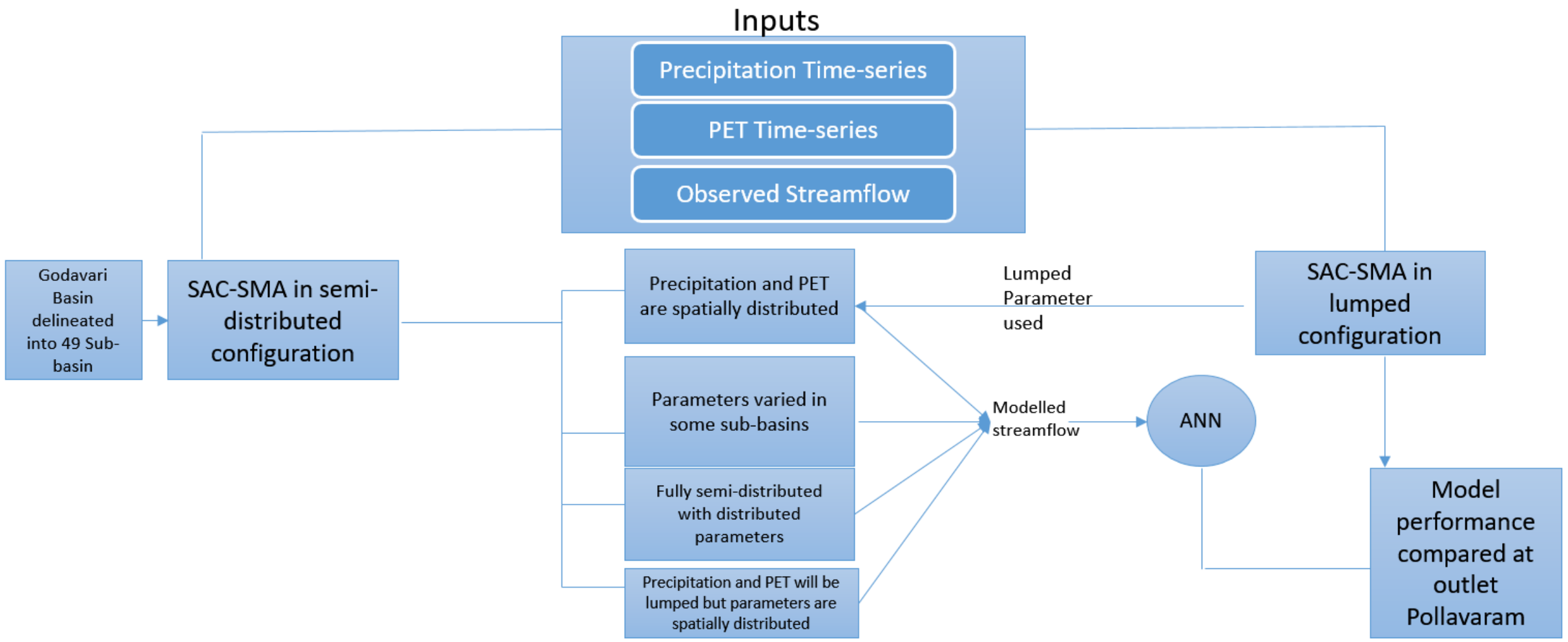

Inputs
Precipitation Time-series
PET Time-series
Observed Streamflow
Godavari Basin delineated into 49 Sub-basin
SAC-SMA in semi-distributed configuration
Precipitation and PET are spatially distributed
Parameters varied in some sub-basins
Fully semi-distributed with distributed parameters
Precipitation and PET will be lumped but parameters are spatially distributed
Lumped Parameter used
SAC-SMA in lumped configuration
Modelled streamflow
ANN
Model performance compared at outlet Pollavaram

**Figure 5.**

# Godavari Basin

72°0'0"E 74°0'0"E 76°0'0"E 78°0'0"E 80°0'0"E 82°0'0"E 84°0'0"E

22°0'0"N 20°0'0"N 18°0'0"N 16°0'0"N

N

Kumhari
Satrapur
Hivra
Bamini
P.G Bridge
Sirpur
Ashti
Tumnar
Mancherial
Somanpalli
Perur
Sardarput
Sukhma
Koida
Konta
Bhadrachalam
Koperagaon
Toka
Dhalegaon
Phalla
Yelli
G R Bridge
Basar
Saigaon
Raipally
Godavari
Pollavaram

Legend

SourceDEM

Value

High : 1677

Low : -15

0 137.5 275 550 Kilometers

**Figure 6.**

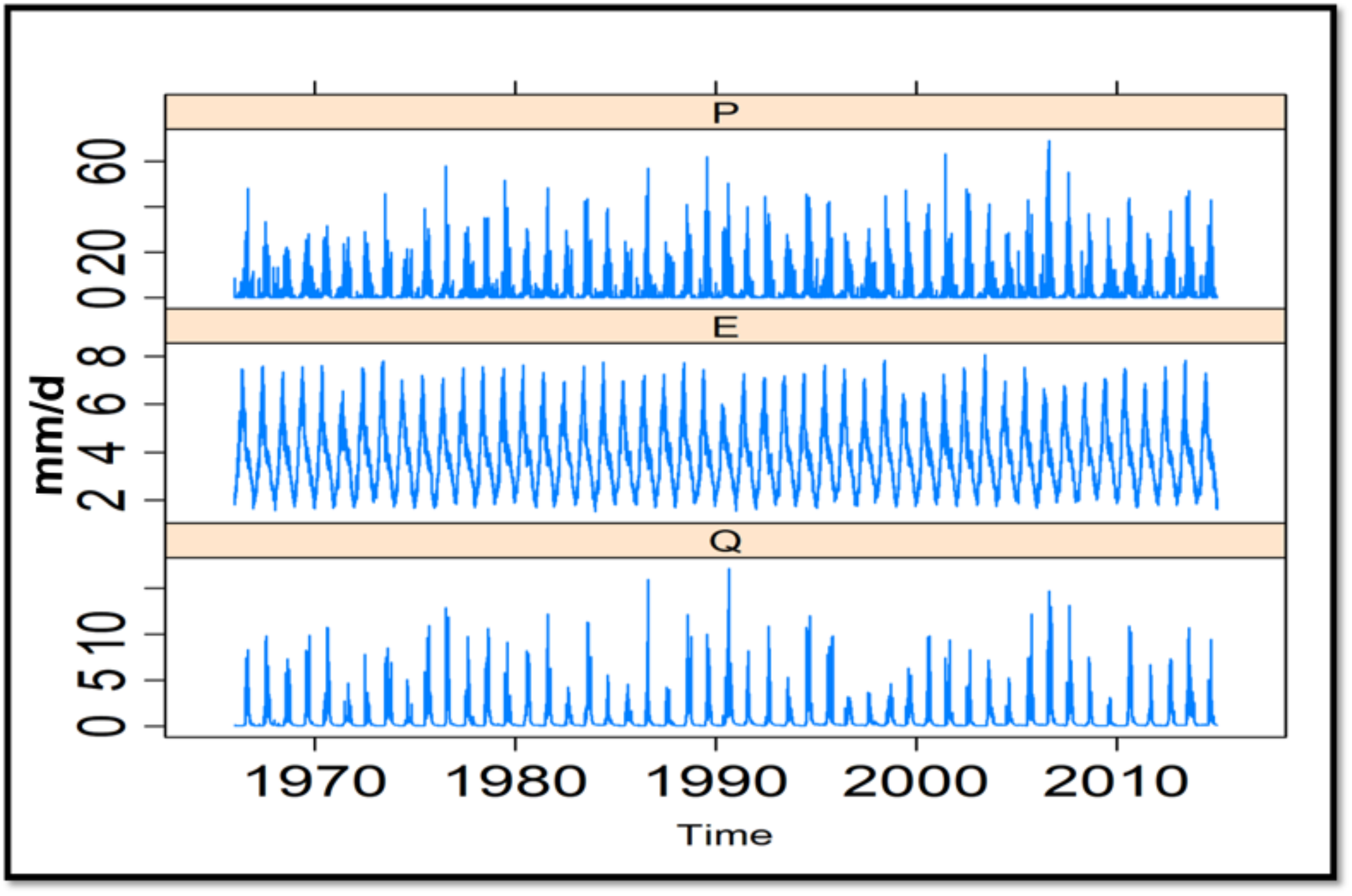

P
E
Q
mm/d
0 20 60
2 4 6 8
0 5 10
1970
1980
1990
2000
2010
Time

**Figure 7.**

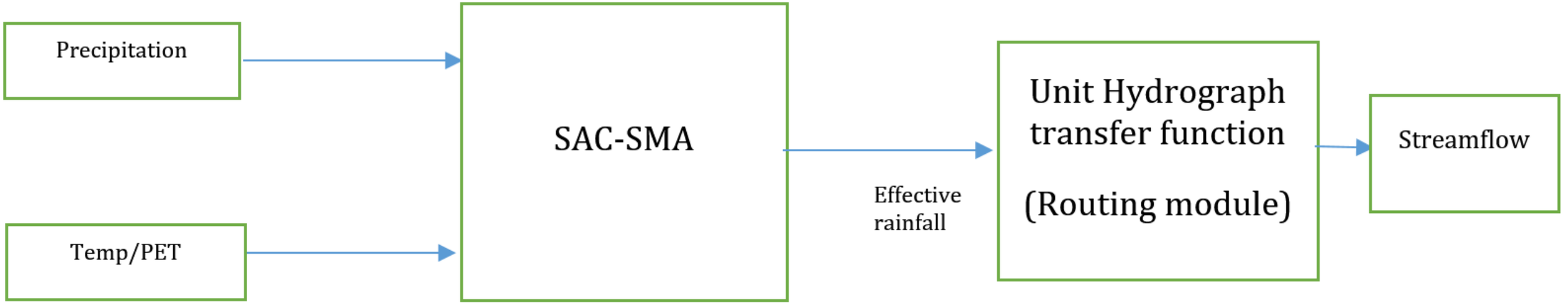

Precipitation
Temp/PET
SAC-SMA
Effective rainfall
Unit Hydrograph transfer function
(Routing module)
Streamflow

**Figure 8.**

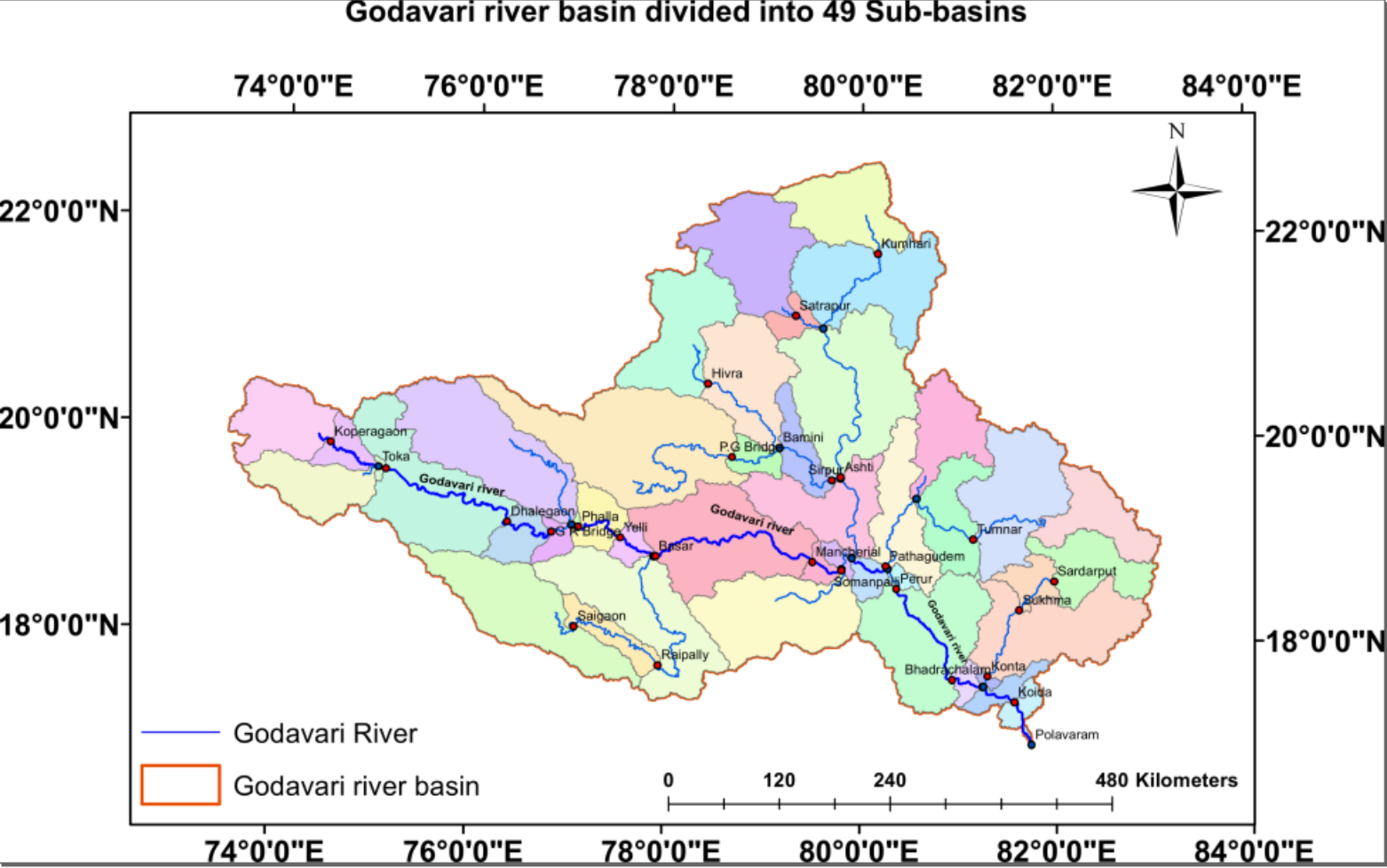

Godavari river basin divided into 49 Sub-basins
74°0'0"E
76°0'0"E
78°0'0"E
80°0'0"E
82°0'0"E
84°0'0"E
22°0'0"N
20°0'0"N
18°0'0"N
N
Kumhari
Satrapur
Hivra
Bamini
P.G Bridge
Sirpur Ashti
Koperagaon
Toka
Godavari river
Dhalegaon
Phalla
G.K Bridge
Yelli
Basar
Godavari river
Tumnar
Mancherial
Pathagudem
Somanpalli
Perur
Sardarput
Sukhma
Godavari river
Saigaon
Rajpally
Bhadrachalam
Konta
Kolda
Polavaram
Godavari River
Godavari river basin
0
120
240
480 Kilometers

**Figure 9.**

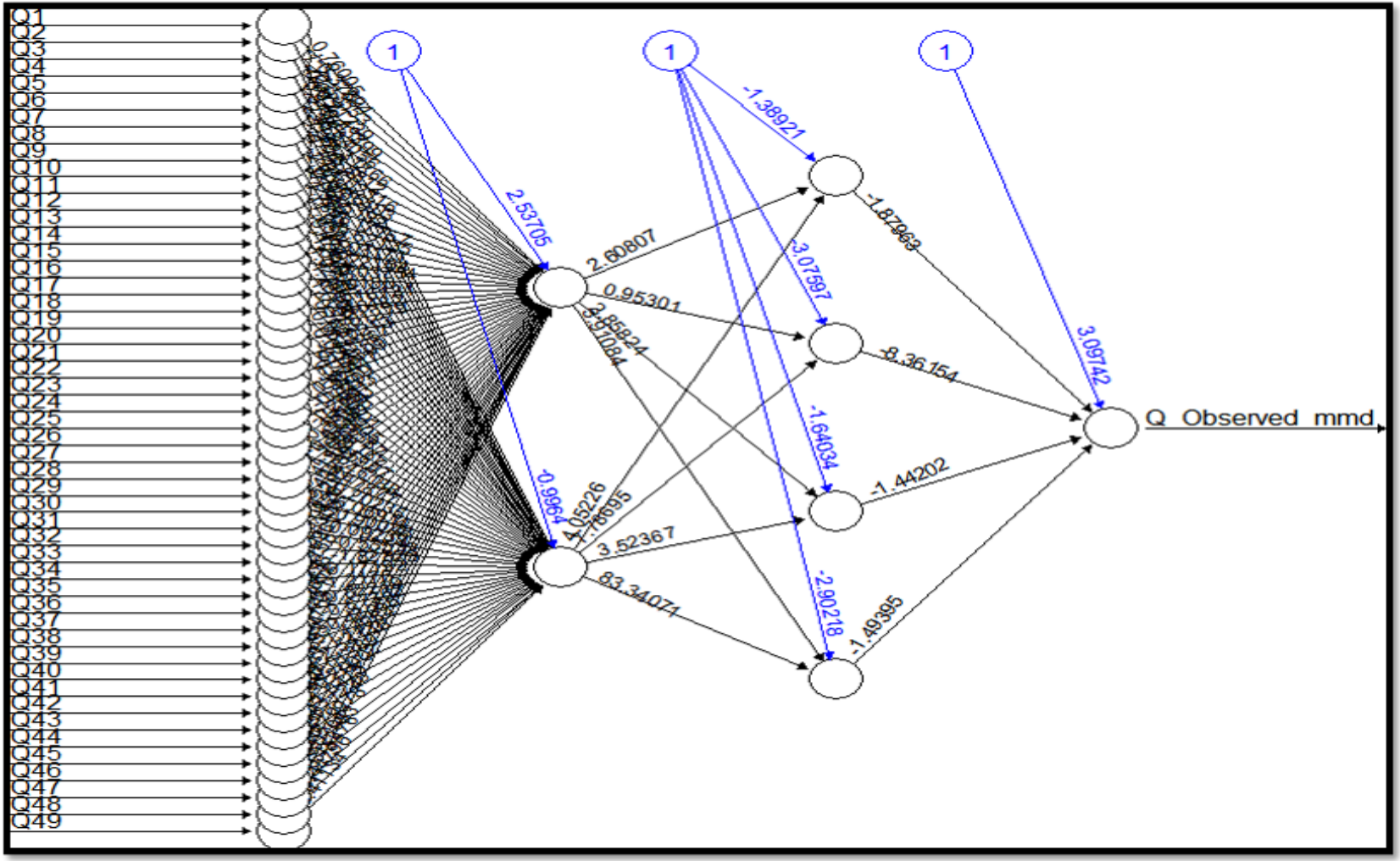
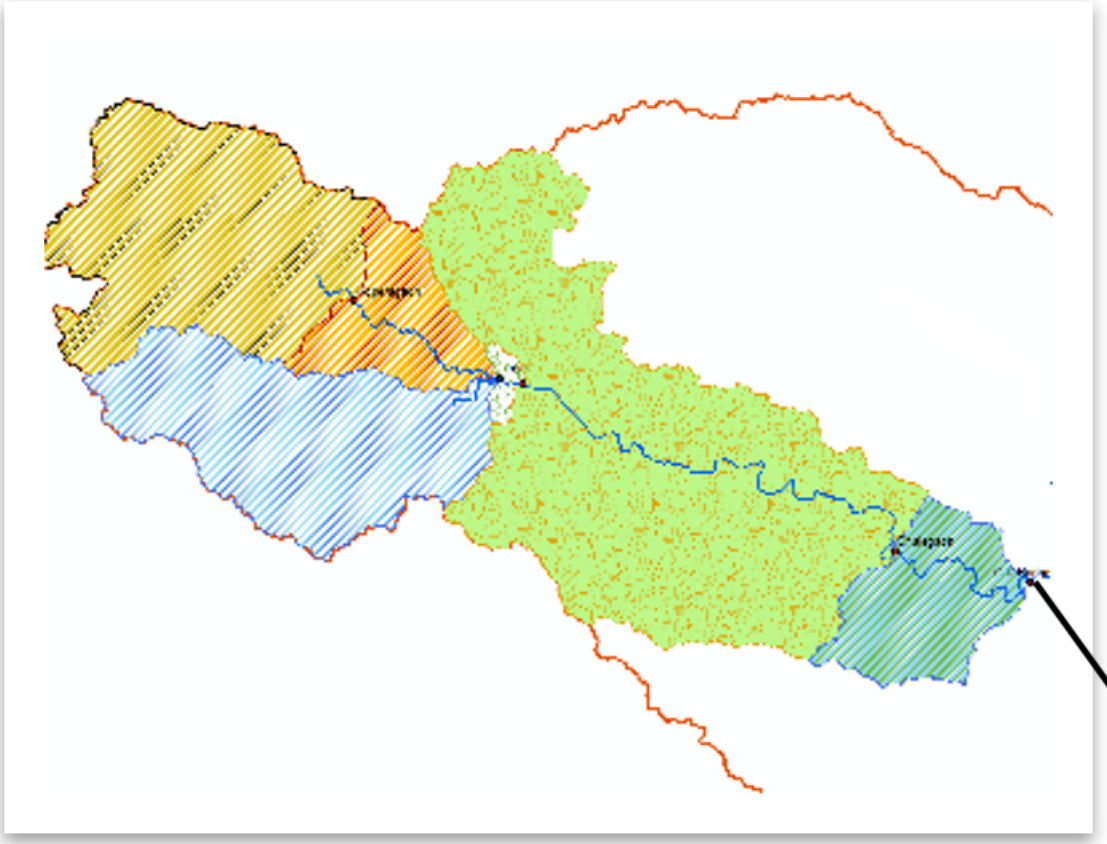

Q1
Q2
Q3
Q4
Q5
Q6
Q7
Q8
Q9
Q10
Q11
Q12
Q13
Q14
Q15
Q16
Q17
Q18
Q19
Q20
Q21
Q22
Q23
Q24
Q25
Q26
Q27
Q28
Q29
Q30
Q31
Q32
Q33
Q34
Q35
Q36
Q37
Q38
Q39
Q40
Q41
Q42
Q43
Q44
Q45
Q46
Q47
Q48
Q49
1
1
1
2.53705
-0.9964
2.60807
0.95301
3.85824
3.91084
4.05226
4.76695
3.52367
83.34071
-1.38921
-3.07597
-1.64034
-2.90218
-1.87969
-8.36154
-1.44202
-1.49395
3.09742
Q_Observed_mmd

**Figure 10.**



GR Bridge

**Figure 11.**

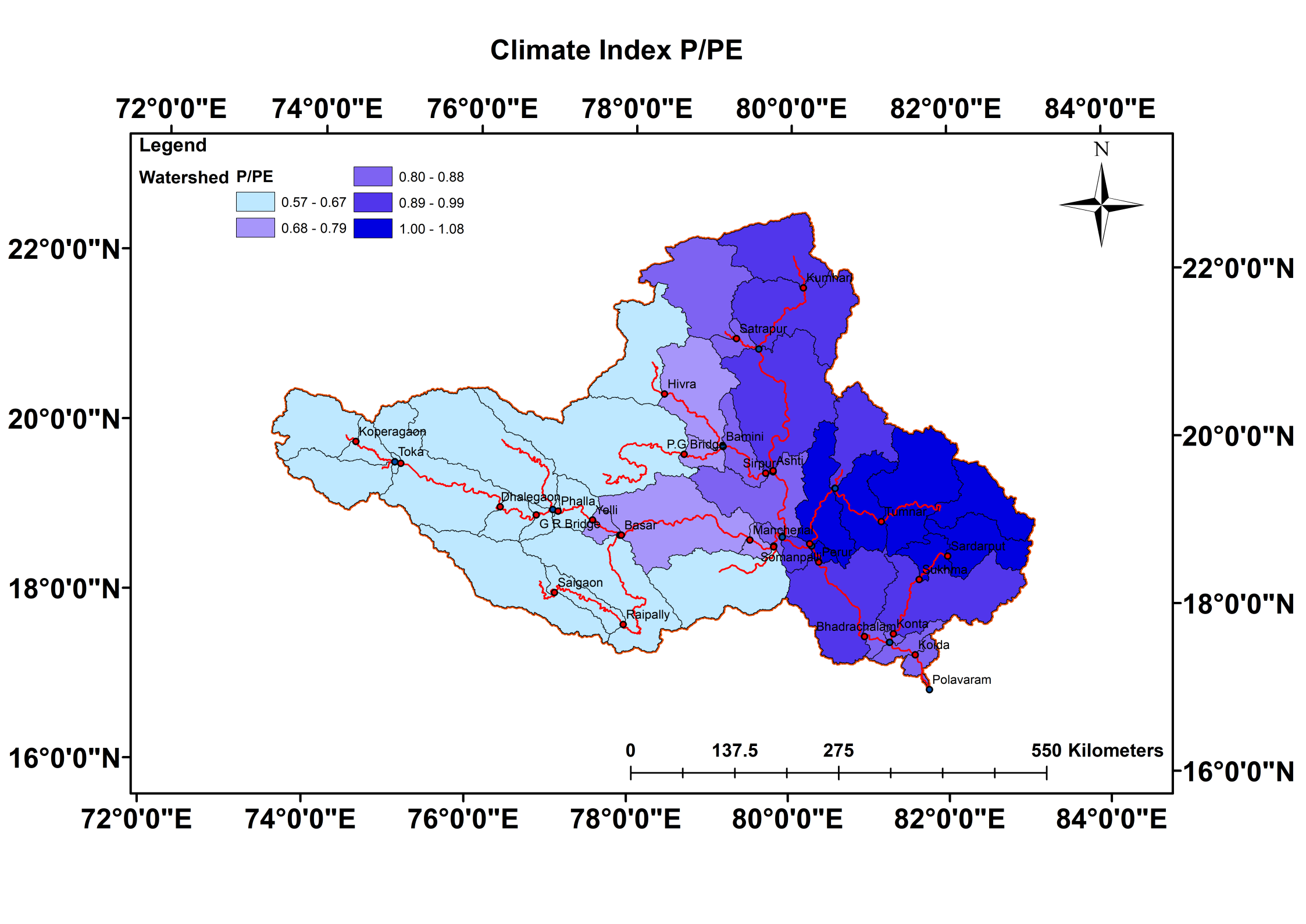

Climate Index P/PE
72°0'0"E
74°0'0"E
76°0'0"E
78°0'0"E
80°0'0"E
82°0'0"E
84°0'0"E
22°0'0"N
20°0'0"N
18°0'0"N
16°0'0"N
Legend
Watershed P/PE
0.57 - 0.67
0.68 - 0.79
0.80 - 0.88
0.89 - 0.99
1.00 - 1.08
N
Kumnar
Satrapur
Hivra
Bamini
P.G.Bridge
Sirpur
Ashti
Koperagaon
Toka
Phalegaon
Phalla
Yelli
G R Bridge
Basar
Tumnar
Mancherial
Sardarput
Perur
Somanpalli
Sukhma
Saigaon
Raipally
Bhadrachalam
Konta
Kolda
Polavaram
0
137.5
275
550 Kilometers

**Figure 12.**

Koperagaon
Toka
Dhalegaon
G R Bridge
Phalla
Yelli
Kutrelam
Satrapur
Hivra
Bamini
P G Bhadi
Simli
Ashti
Koperagaon
Toka
Dhalegaon
Phalla
G R Bridge
Yeth
Basar
Manchikal
Somanpalli
Tomnar
Perur
Sardarput
Sukhma
Konta
Kpida
Polavaram
Rajaram
Rakpally
Bhadrachalam
Sardarput
Sukhma
Konta
Konta

**Figure 13.**

# LULC Map of Godavari Basin (1970)

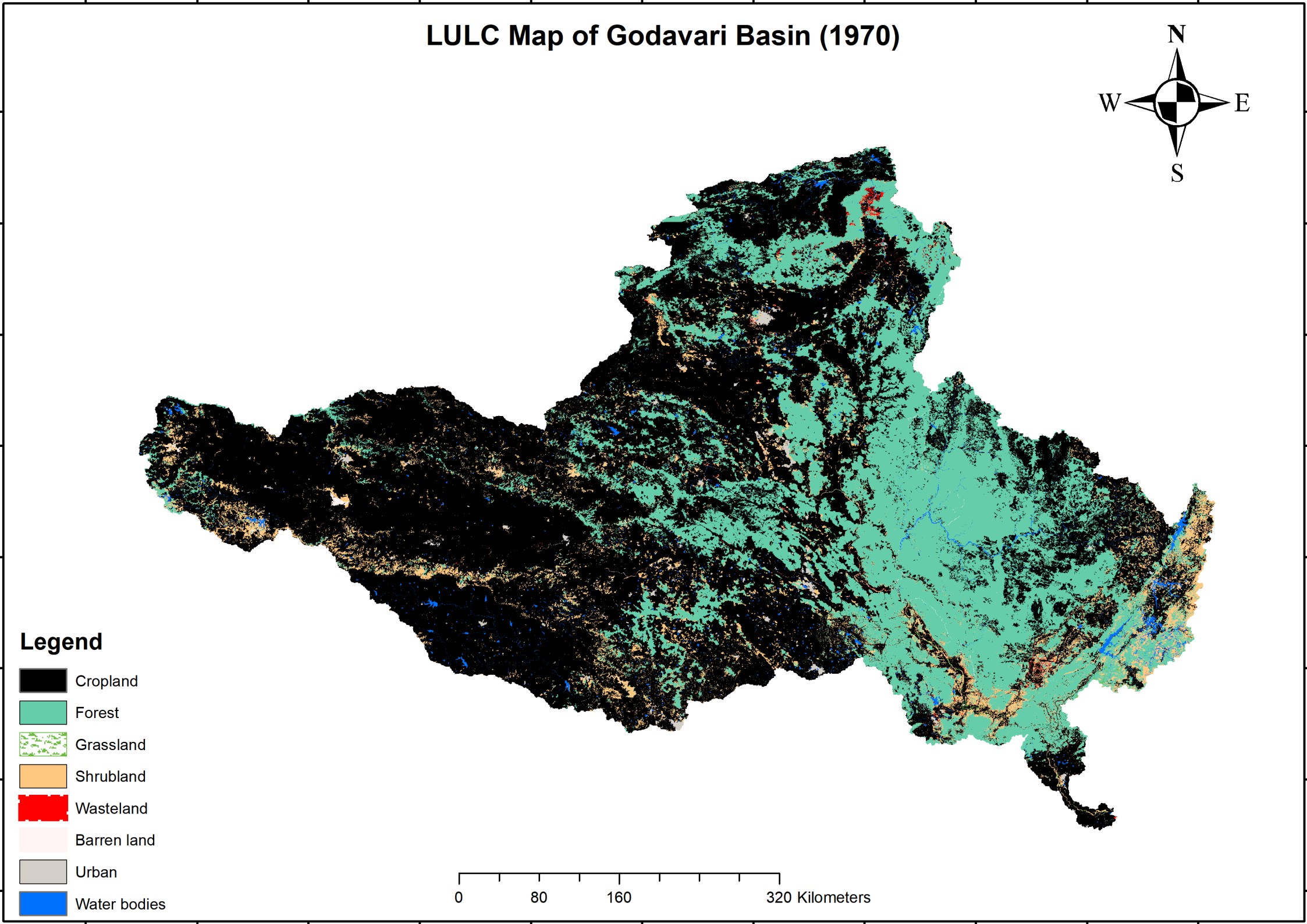
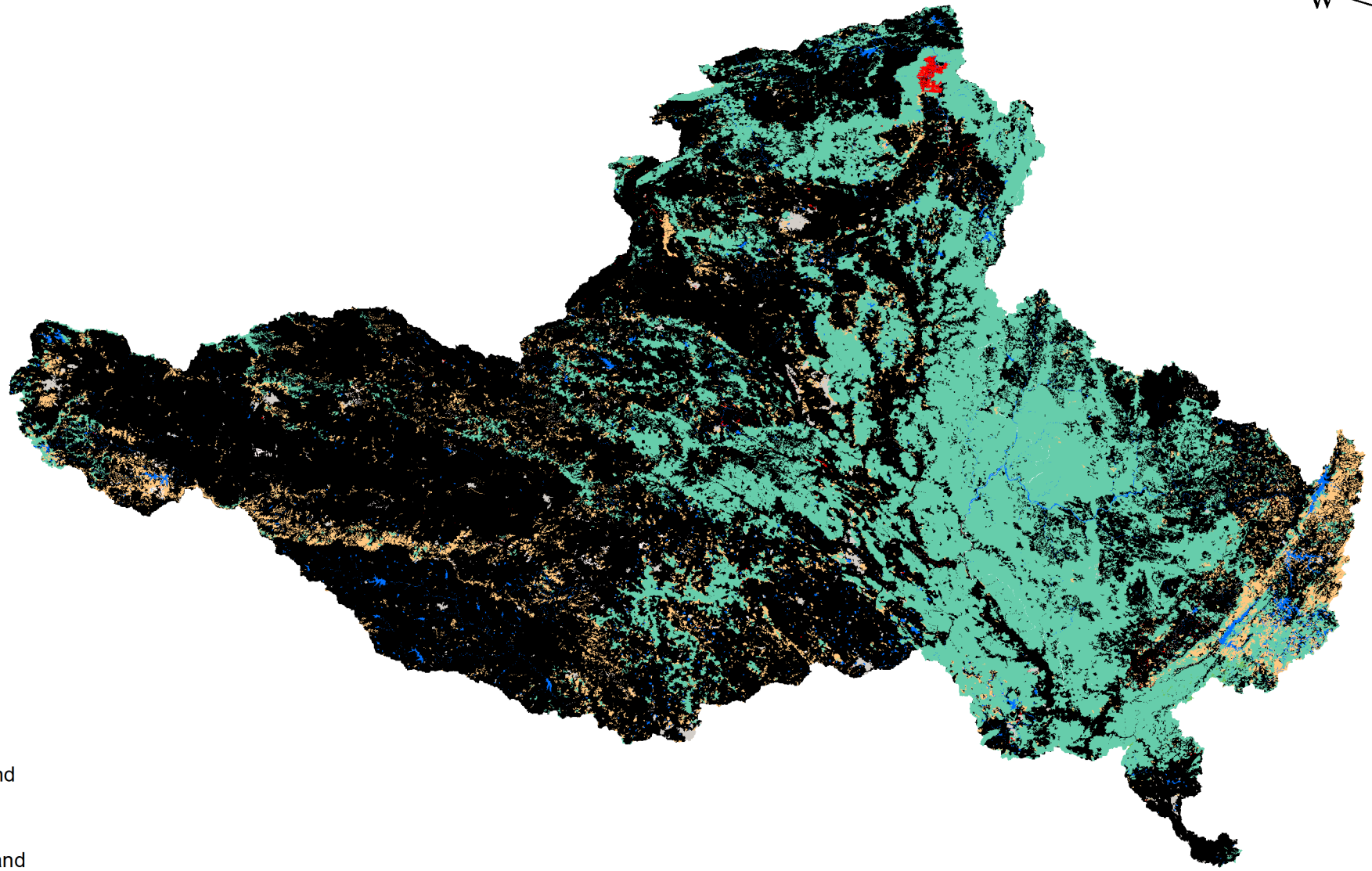


**Figure 14.**

# LULC Map of Godavari Basin (2000)



N
W
E
S

## Legend

- Cropland
- Forest
- Grassland
- Shrubland
- Wasteland
- Barren land
- Urban
- Water bodies

0 87.5 175 350 Kilometers

**Figure 15.**

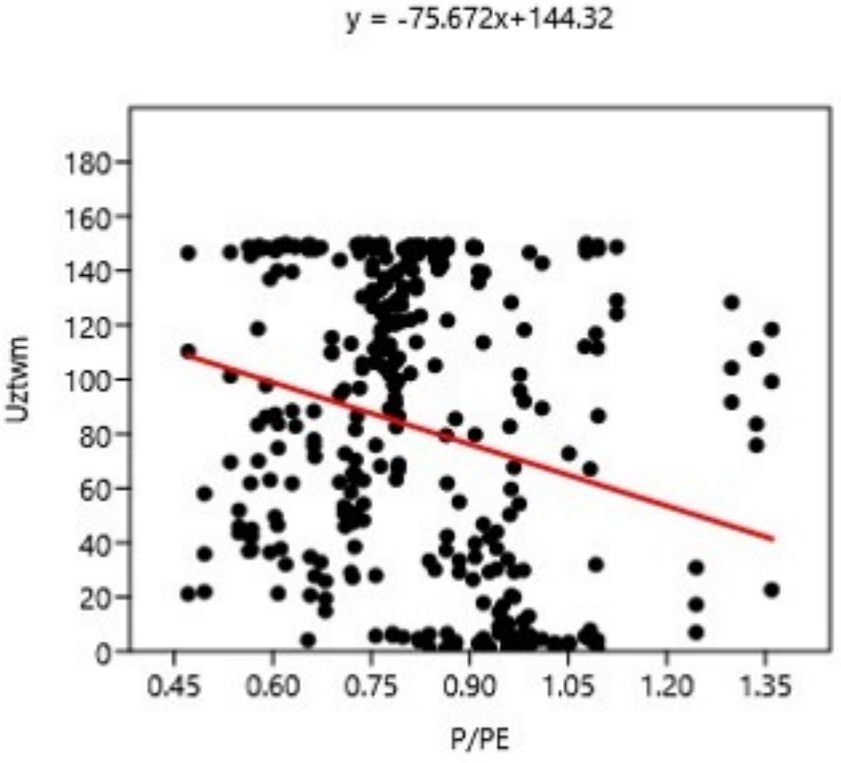

y = -75.672x+144.32
Uztwm
180
160
140
120
100
80
60
40
20
0
0.45
0.60
0.75
0.90
1.05
1.20
1.35
P/PE

**Figure 16.**

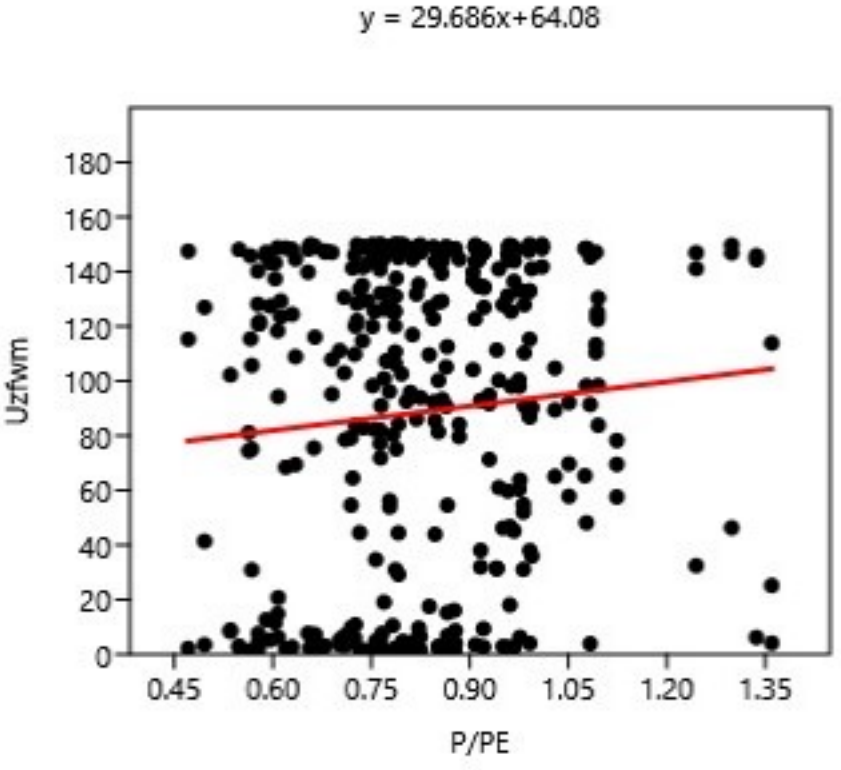

y = 29.686x+64.08
Uzfwm
180
160
140
120
100
80
60
40
20
0
0.45
0.60
0.75
0.90
1.05
1.20
1.35
P/PE

**Figure 17.**

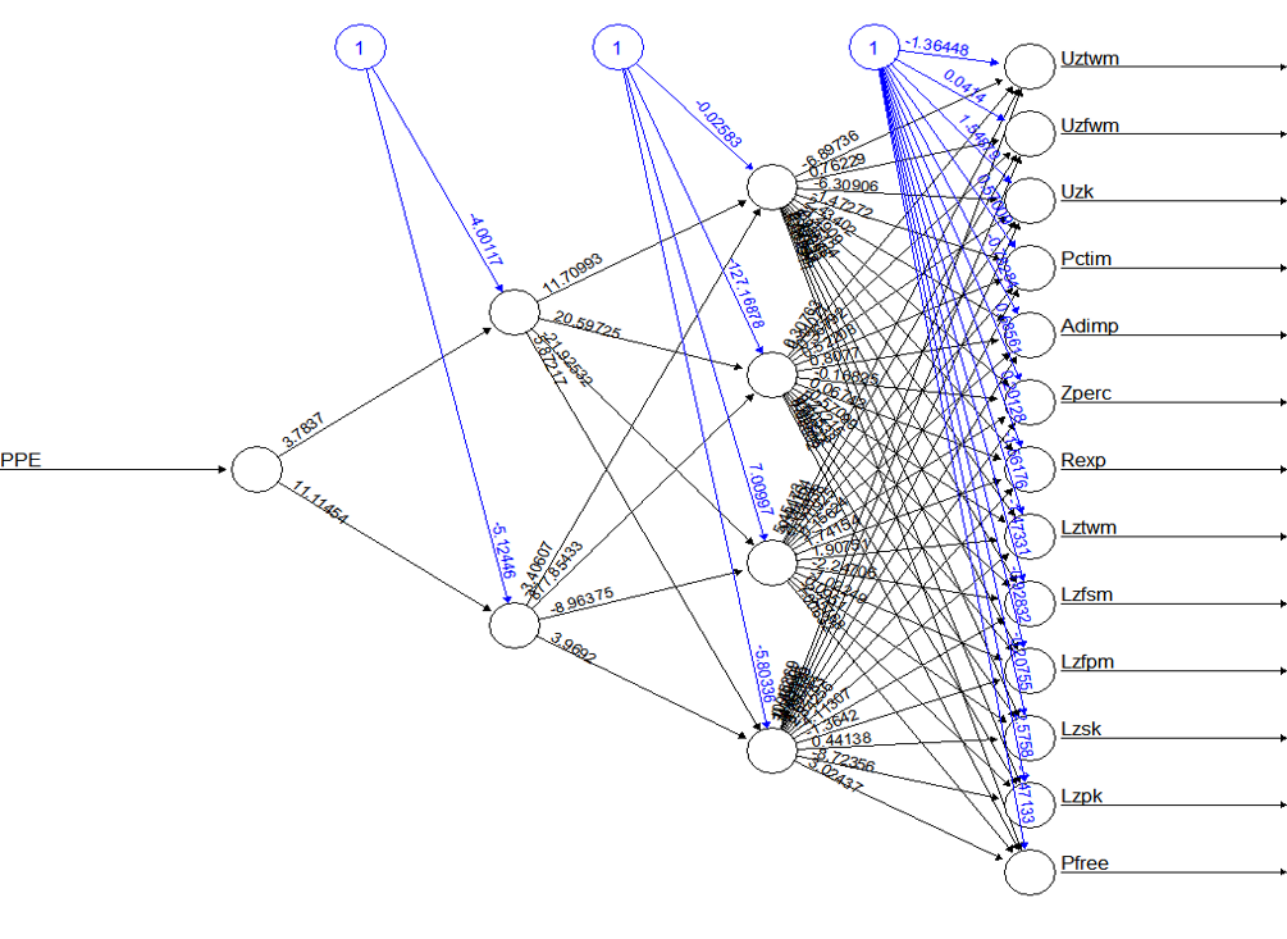

PPE
1
1
1
3.7837
11.11454
-4.00117
-5.12446
11.70993
20.59725
-2.92532
-3.87217
3.40607
-17.85433
-8.96375
3.9692
-0.02583
-121.16878
7.00997
-5.80336
-1.36448
0.0414
-8.89736
0.76229
0.30906
0.8077
0.16625
0.44138
Uztwm
Uzfwm
Uzk
Pctim
Adimp
Zperc
Rexp
Lztwm
Lzfsm
Lzfpm
Lzsk
Lzpk
Pfree

**Figure 18.**

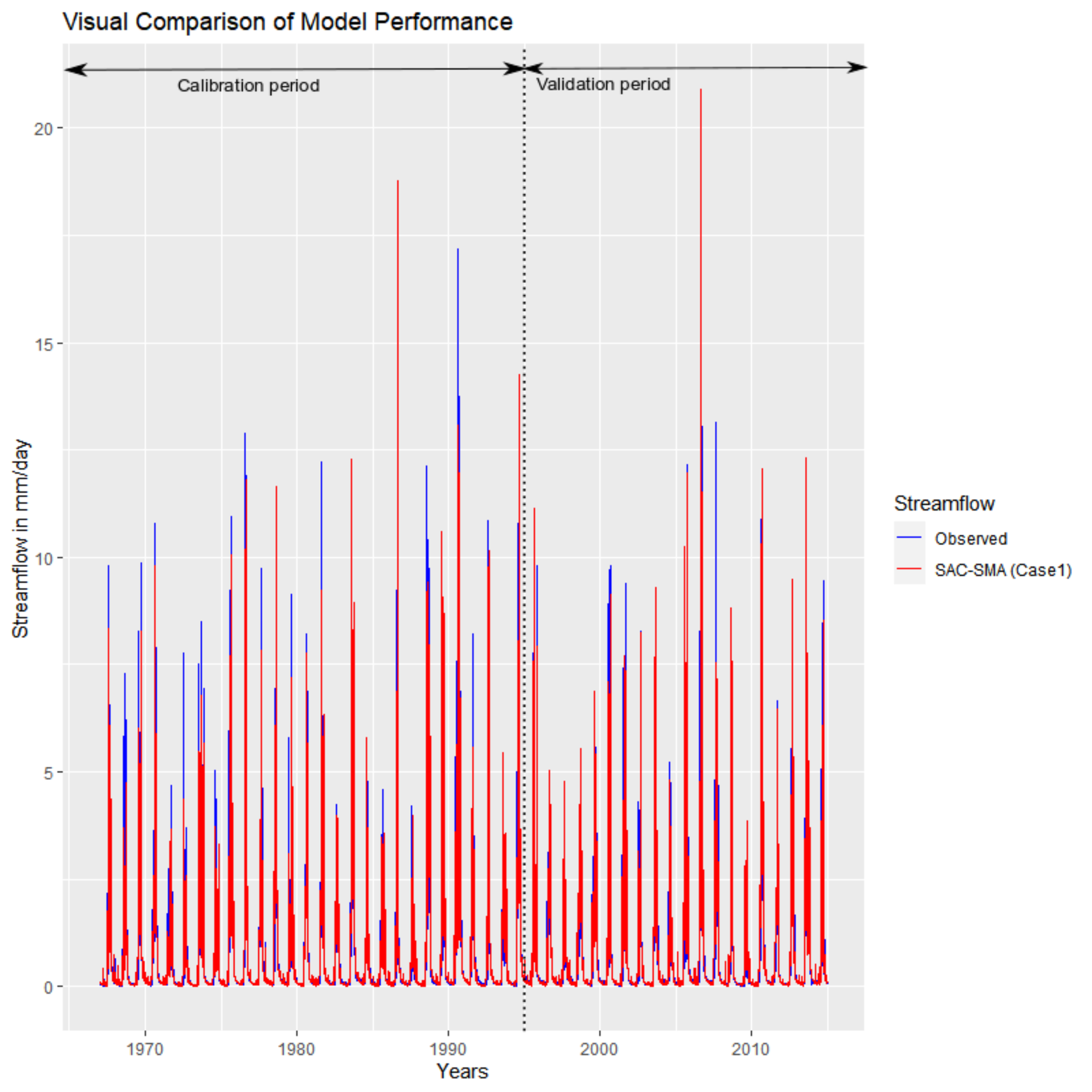

Visual Comparison of Model Performance
Calibration period
Validation period
Streamflow in mm/day
20
15
10
5
0
1970
1980
1990
2000
2010
Years
Streamflow
Observed
SAC-SMA (Case1)

**Figure 19.**

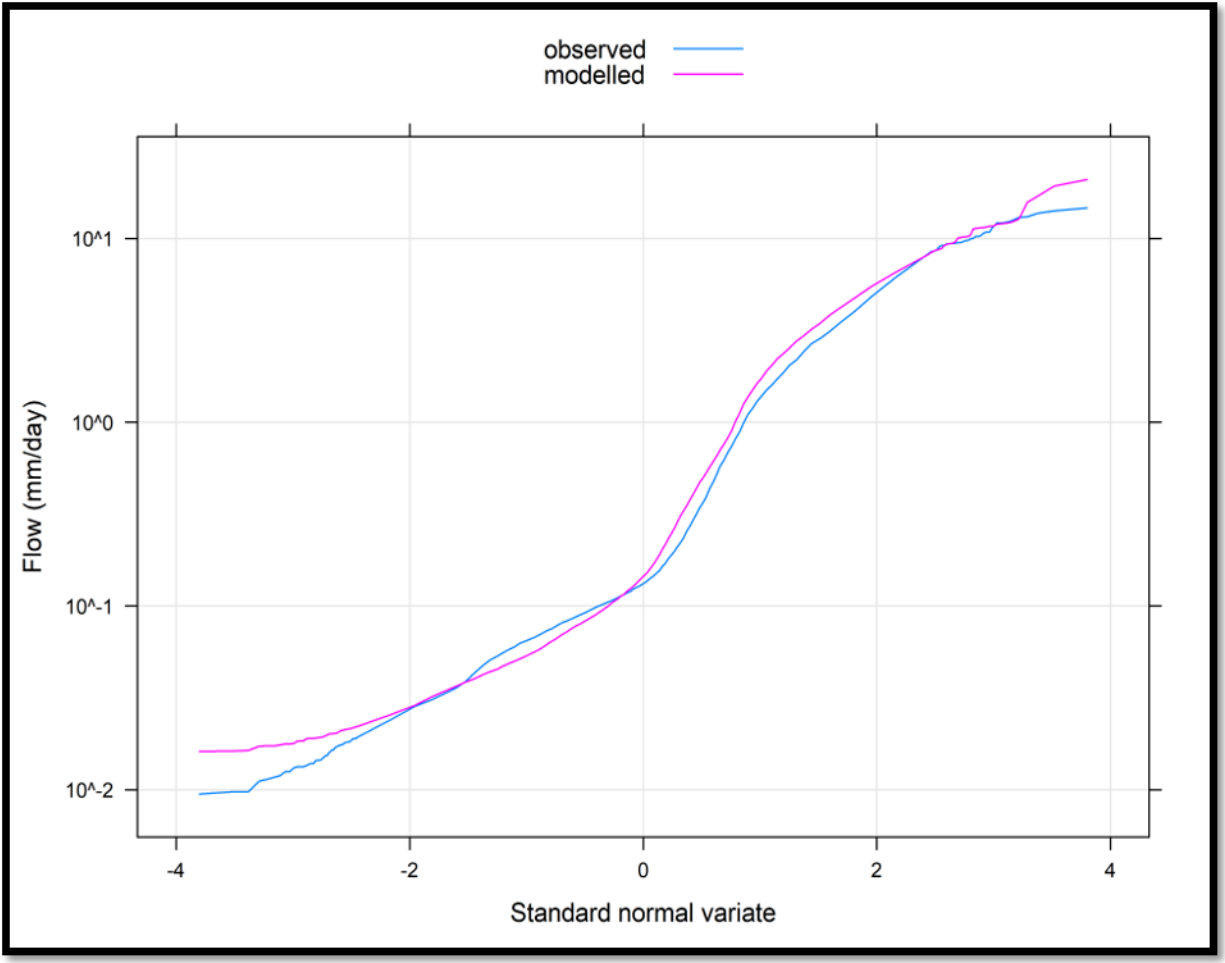

observed
modelled
10^1
10^0
10^-1
10^-2
Flow (mm/day)
-4
-2
0
2
4
Standard normal variate

**Figure 20.**

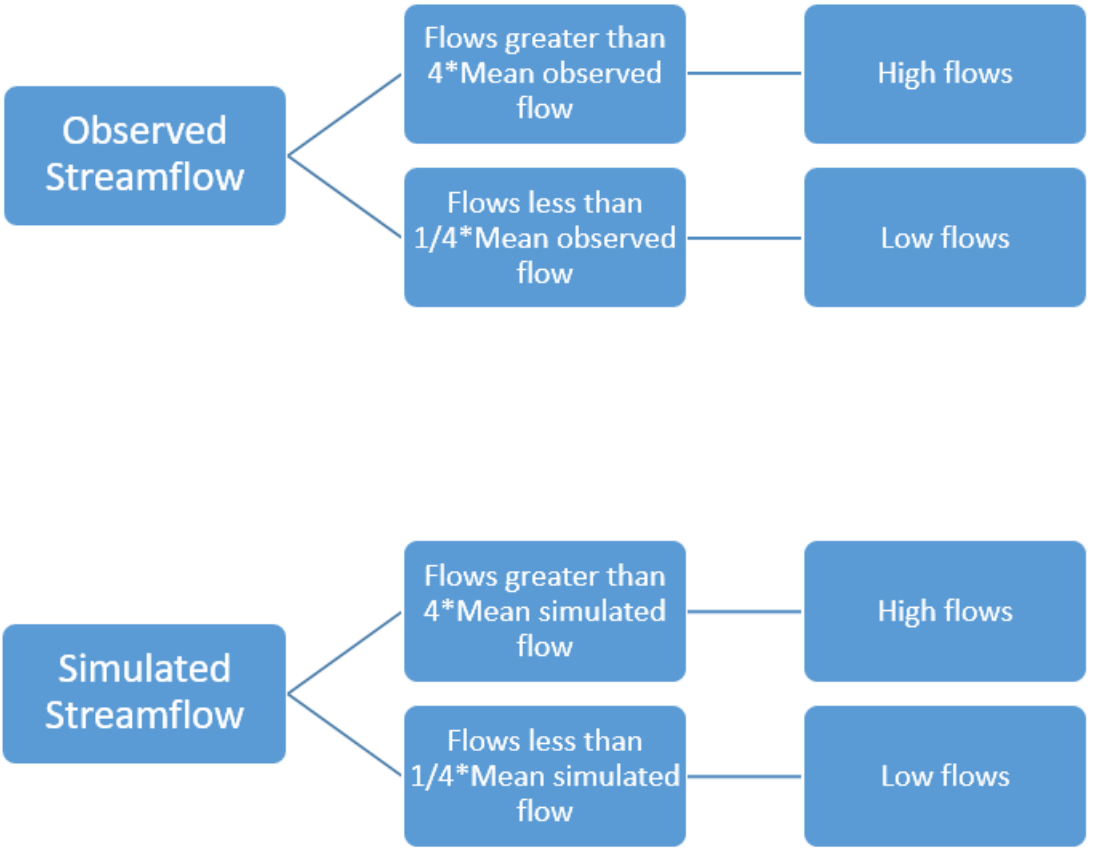
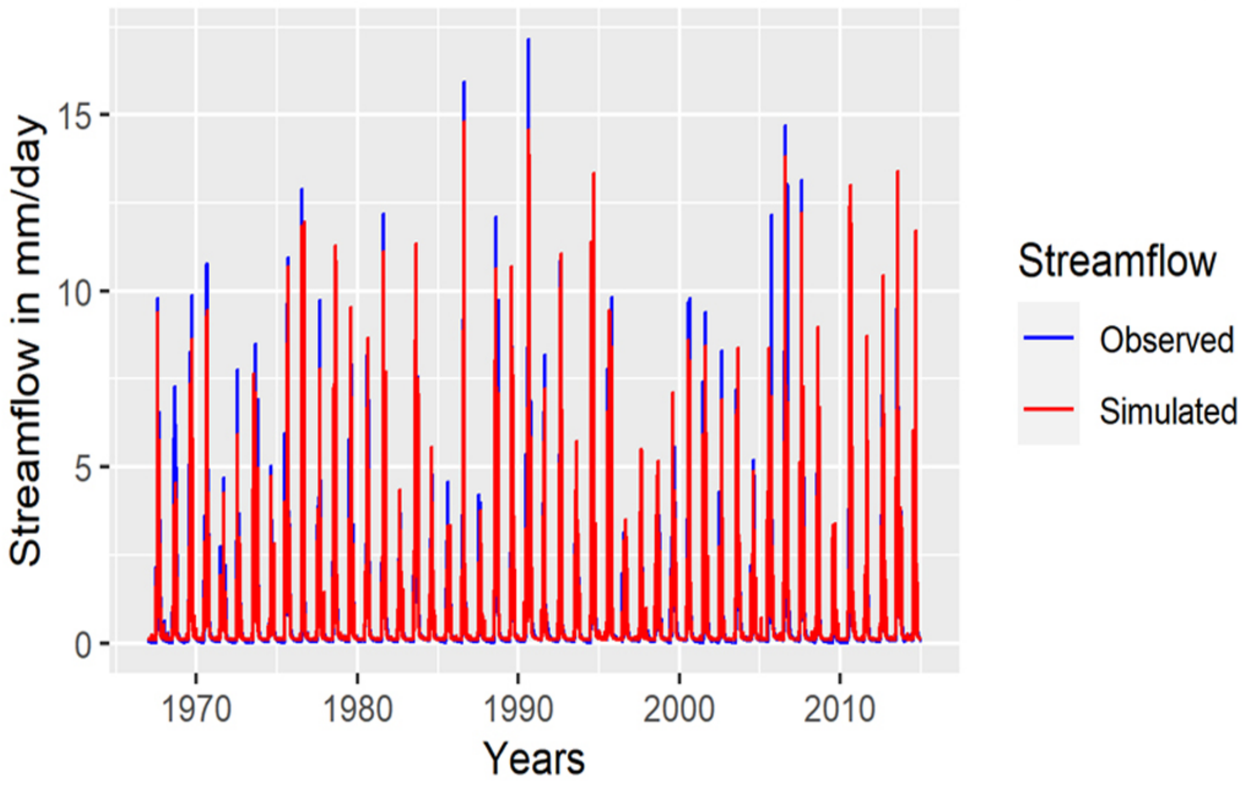
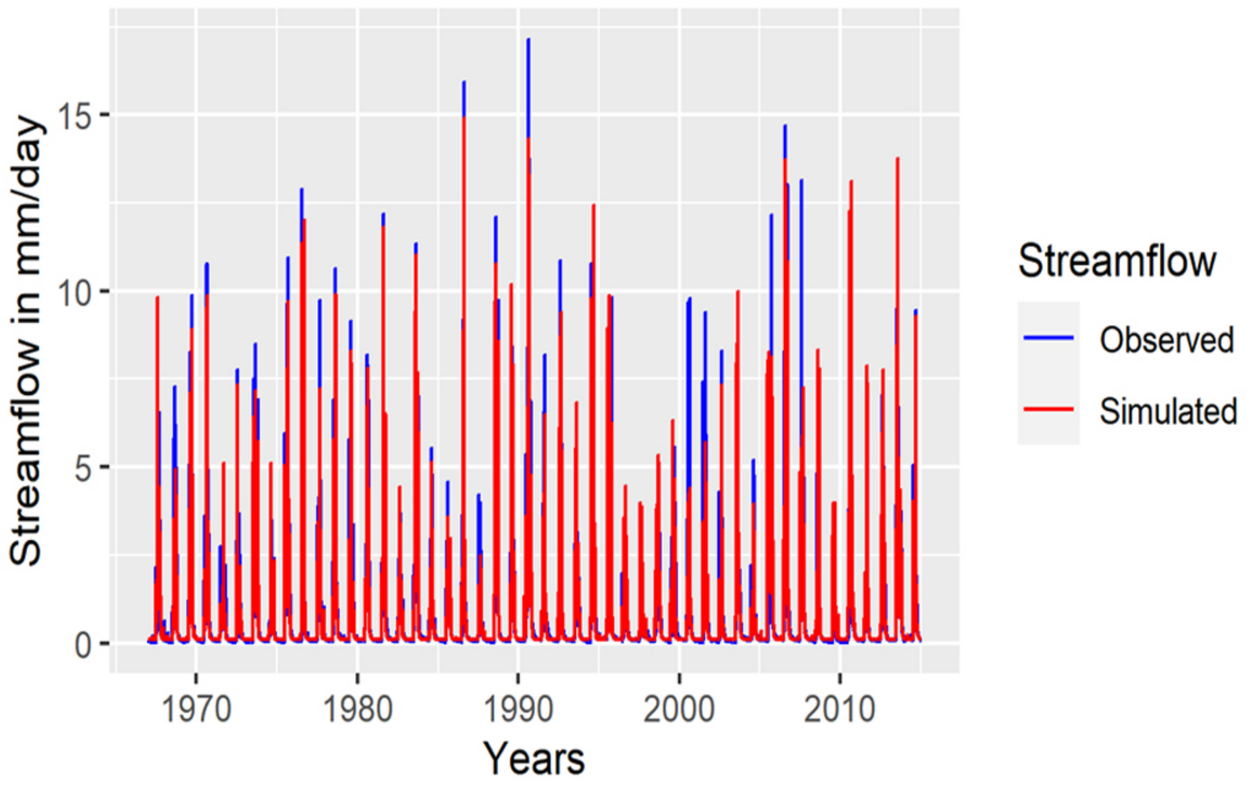
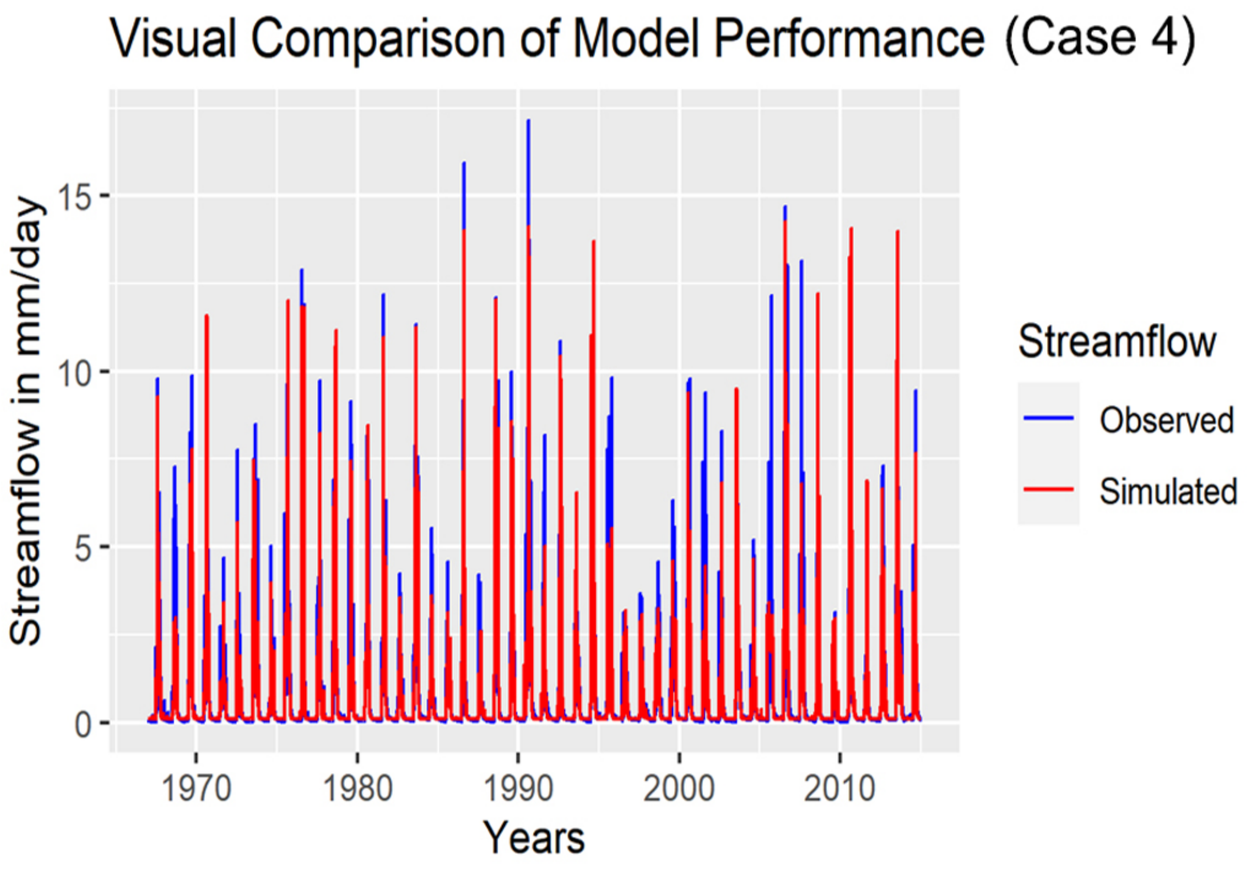

Observed Streamflow
Flows greater than 4*Mean observed flow
High flows
Flows less than 1/4*Mean observed flow
Low flows
Simulated Streamflow
Flows greater than 4*Mean simulated flow
High flows
Flows less than 1/4*Mean simulated flow
Low flows

**Figure 21.**


Visual Comparison of Model Performance (Case 2)
Streamflow in mm/day
0
5
10
15
1970
1980
1990
2000
2010
Years
Streamflow
Observed
Simulated


Visual Comparison of Model Performance (Case 3)
Streamflow in mm/day
0
5
10
15
1970
1980
1990
2000
2010
Years
Streamflow
Observed
Simulated


Visual Comparison of Model Performance (Case 4)
Streamflow in mm/day
0
5
10
15
1970
1980
1990
2000
2010
Years
Streamflow
Observed
Simulated

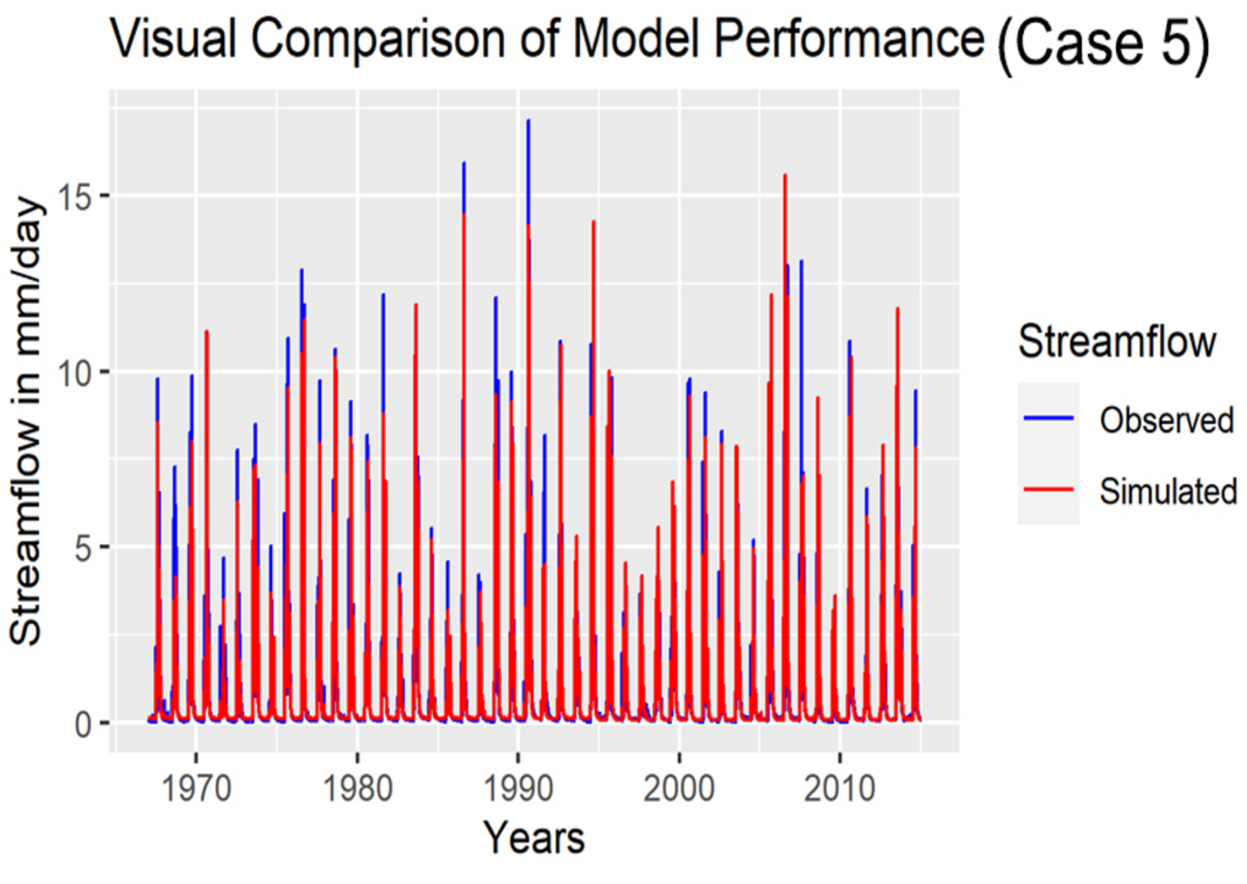

Visual Comparison of Model Performance (Case 5)
Streamflow in mm/day
0
5
10
15
1970
1980
1990
2000
2010
Years
Streamflow
Observed
Simulated

**Figure 22.**

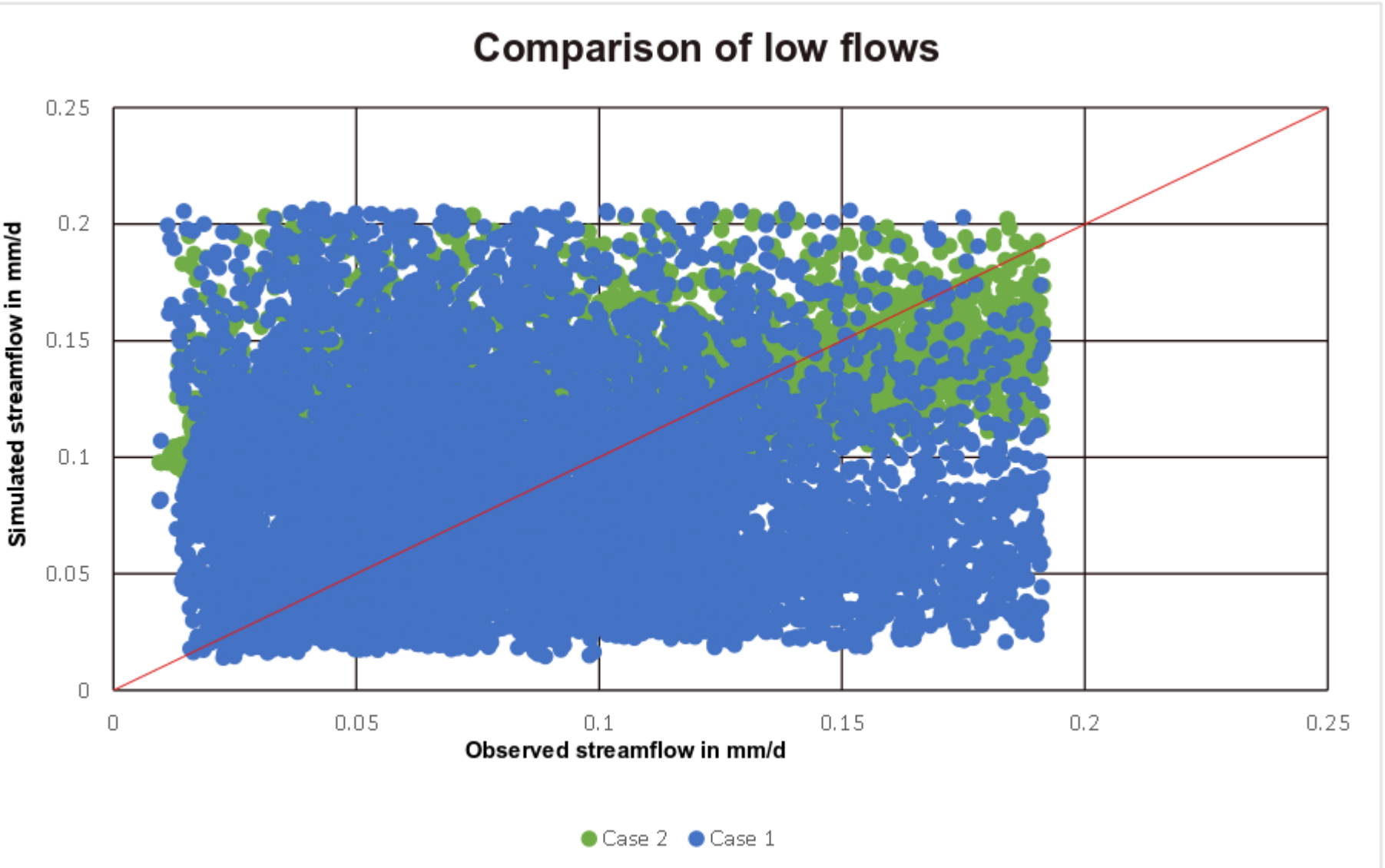

Comparison of low flows
Simulated streamflow in mm/d
0.25
0.2
0.15
0.1
0.05
0
0
0.05
0.1
0.15
0.2
0.25
Observed streamflow in mm/d
Case 2
Case 1

**Figure 23.**

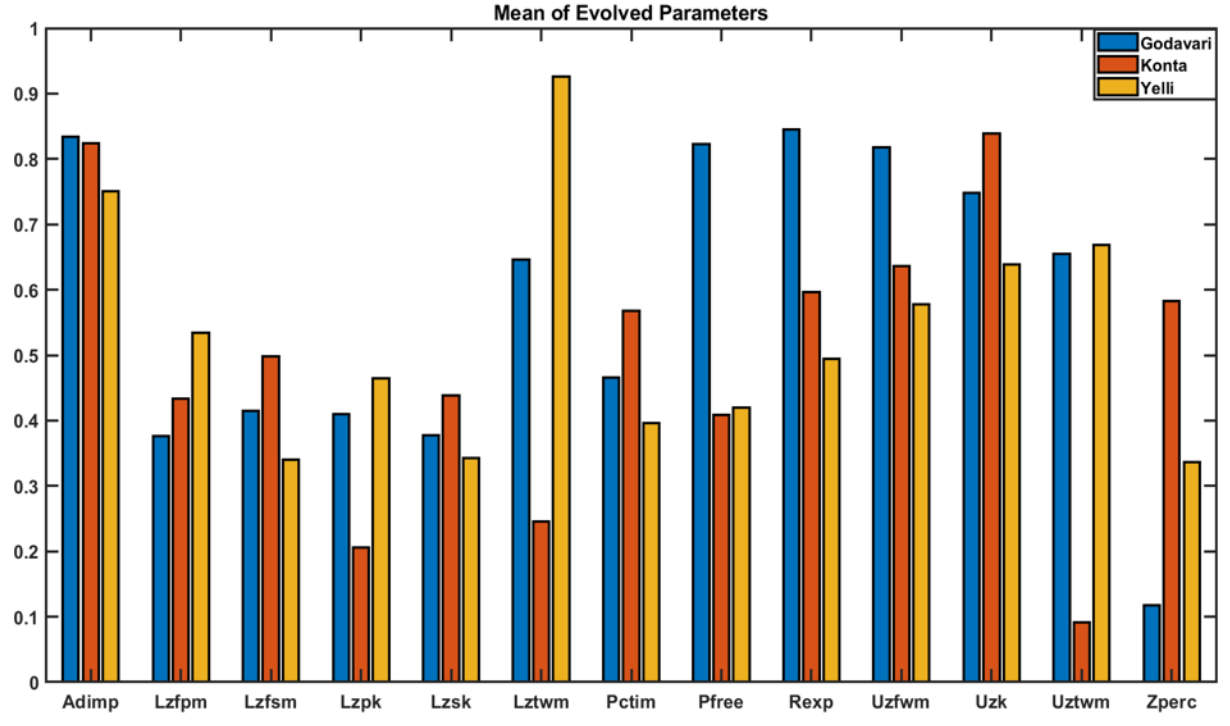
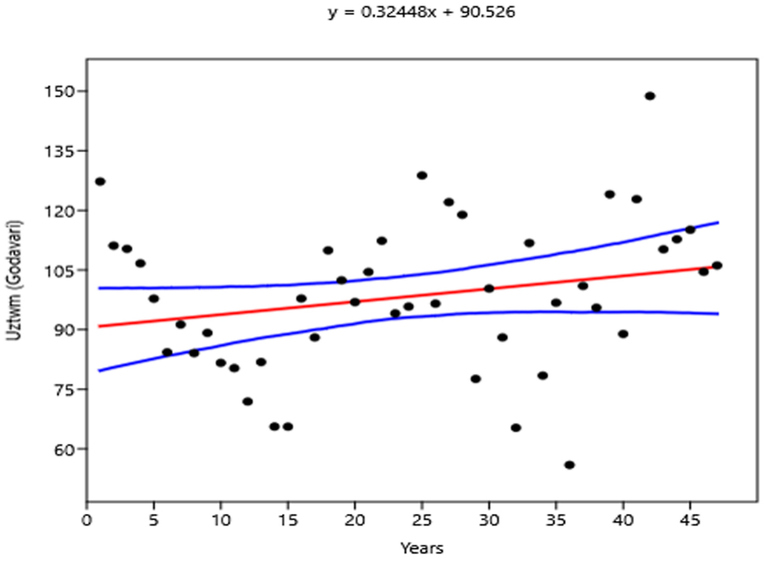
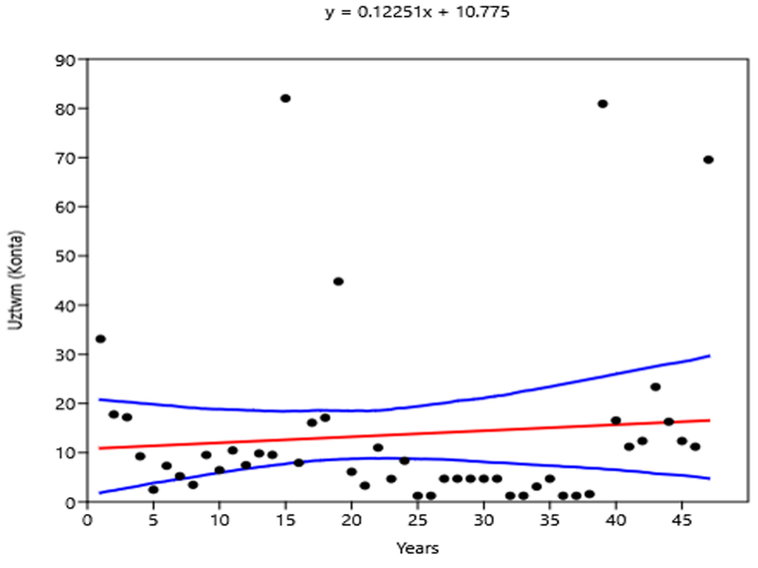

Mean of Evolved Parameters
Godavari
Konta
Yelli
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
Adimp
Lzfpm
Lzfsm
Lzpk
Lzsk
Lztwm
Pctim
Pfree
Rexp
Uzfwm
Uzk
Uztwm
Zperc

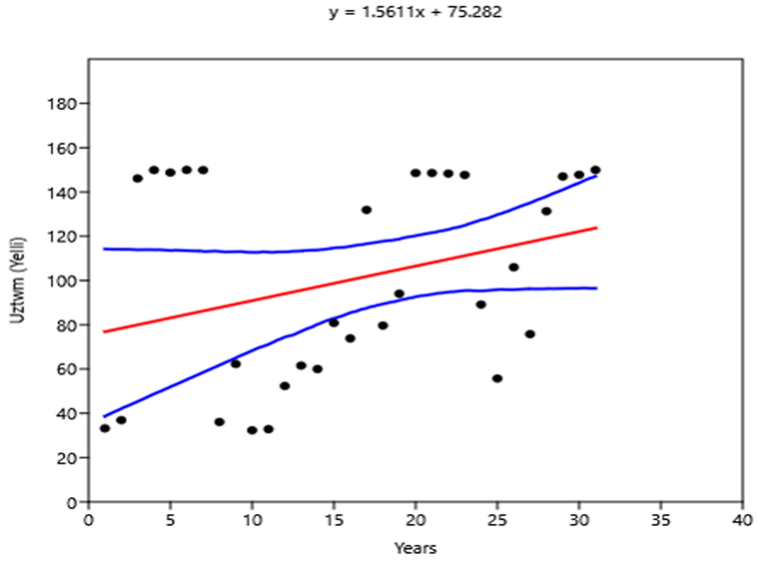
**Figure 24.**


y = 0.32448x + 90.526
Uztwm (Godavari)
Years

(a)


y = 0.12251x + 10.775
Uztwm (Konta)
Years

(b)


y = 1.5611x + 75.282
Uztwm (Yeli)
Years

(c)

**Figure 25.**

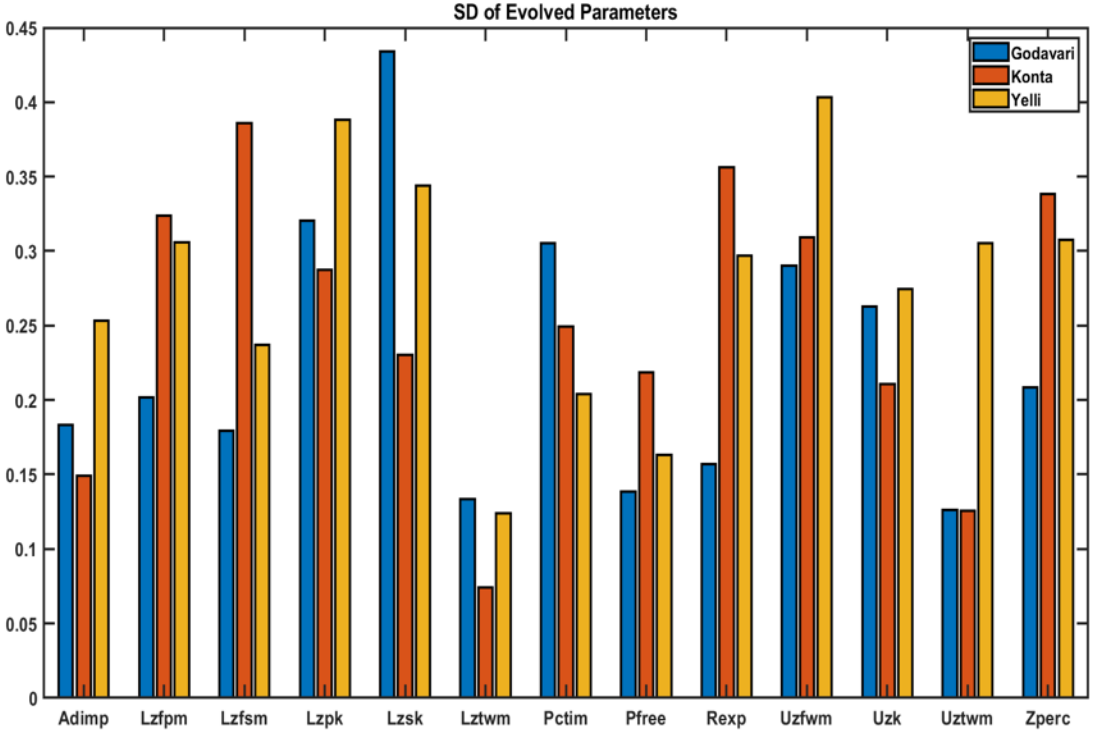
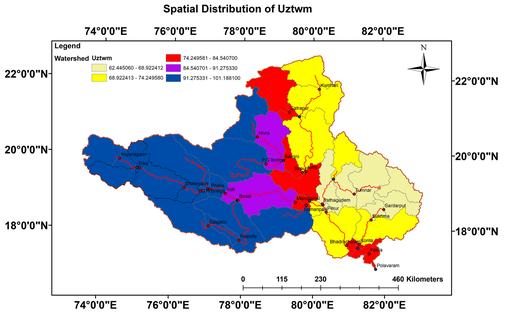
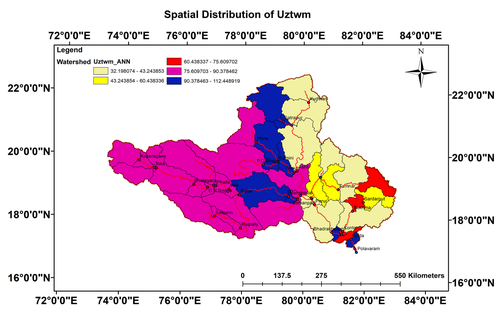
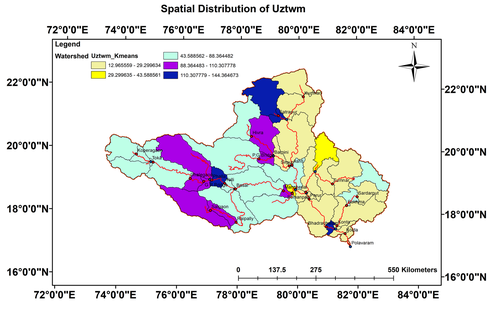
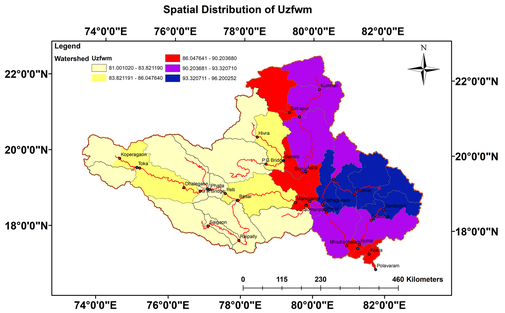

SD of Evolved Parameters
Godavari
Konta
Yelli
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Adimp
Lzfpm
Lzfsm
Lzpk
Lzsk
Lztwm
Pctim
Pfree
Rexp
Uzfwm
Uzk
Uztwm
Zperc

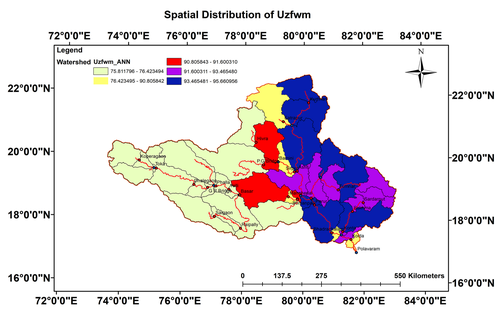
**Figure 26.**

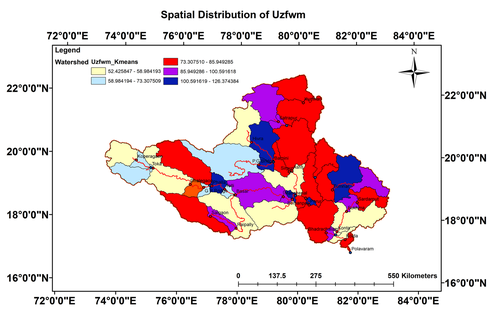

Spatial Distribution of Uztwm
Legend
Watershed Uztwm
74.249581 - 84.540700
62.445060 - 68.922413
84.540701 - 91.275330
68.922413 - 74.249580
91.275331 - 101.188100
74°0'0"E
76°0'0"E
78°0'0"E
80°0'0"E
82°0'0"E
22°0'0"N
20°0'0"N
18°0'0"N
N
0 115 230 460 Kilometers

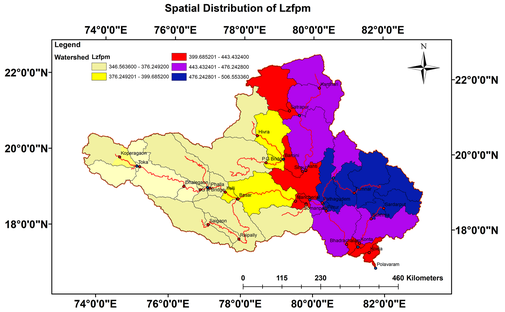

Spatial Distribution of Uztwm
Legend
Watershed Uztwm_ANN
60.435337 - 75.609702
32.199074 - 43.243853
75.609703 - 90.378462
43.243854 - 60.438336
90.378463 - 112.448919
72°0'0"E
74°0'0"E
76°0'0"E
78°0'0"E
80°0'0"E
82°0'0"E
84°0'0"E
22°0'0"N
20°0'0"N
18°0'0"N
16°0'0"N
N
0 137.5 275 550 Kilometers

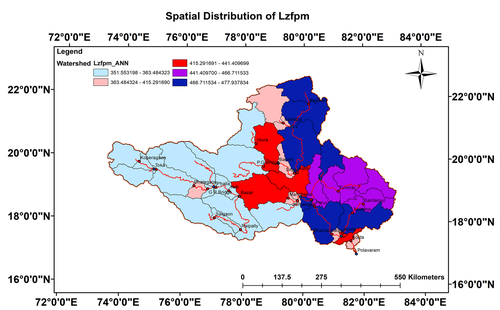

Spatial Distribution of Uztwm
Legend
Watershed Uztwm_Kmeans
43.588562 - 88.364482
12.965559 - 29.299634
88.364483 - 110.307778
29.299635 - 43.588561
110.307779 - 144.364673
72°0'0"E
74°0'0"E
76°0'0"E
78°0'0"E
80°0'0"E
82°0'0"E
84°0'0"E
22°0'0"N
20°0'0"N
18°0'0"N
16°0'0"N
N
0 137.5 275 550 Kilometers

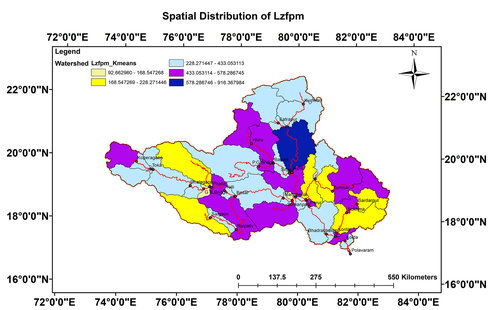

Spatial Distribution of Uzfwm
Legend
Watershed Uzfwm
86.047641 - 90.203680
81.001020 - 83.821190
90.203681 - 93.320710
83.821191 - 86.047640
93.320711 - 96.200252
74°0'0"E
76°0'0"E
78°0'0"E
80°0'0"E
82°0'0"E
22°0'0"N
20°0'0"N
18°0'0"N
N
0 115 230 460 Kilometers

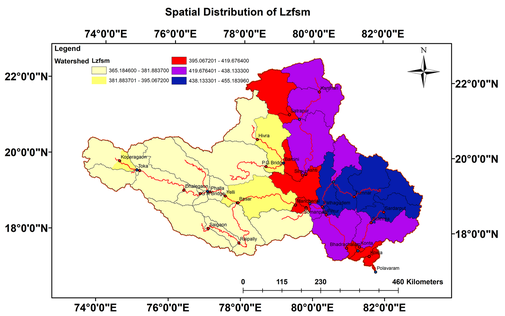

Spatial Distribution of Uzfwm
Legend
Watershed Uzfwm_ANN
90.805643 - 91.600310
75.811796 - 76.423494
91.600311 - 93.465480
76.423495 - 90.805642
93.465481 - 95.660956
72°0'0"E
74°0'0"E
76°0'0"E
78°0'0"E
80°0'0"E
82°0'0"E
84°0'0"E
22°0'0"N
20°0'0"N
18°0'0"N
16°0'0"N
N
0 137.5 275 550 Kilometers

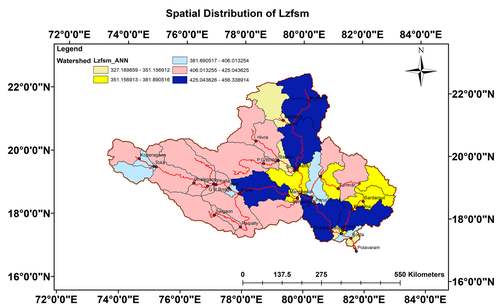

Spatial Distribution of Uzfwm
Legend
Watershed Uzfwm_Kmeans
73.307510 - 85.949285
52.425647 - 58.984193
85.949286 - 100.591618
58.984194 - 73.307509
100.591619 - 126.374384
72°0'0"E
74°0'0"E
76°0'0"E
78°0'0"E
80°0'0"E
82°0'0"E
84°0'0"E
22°0'0"N
20°0'0"N
18°0'0"N
16°0'0"N
N
0 137.5 275 550 Kilometers

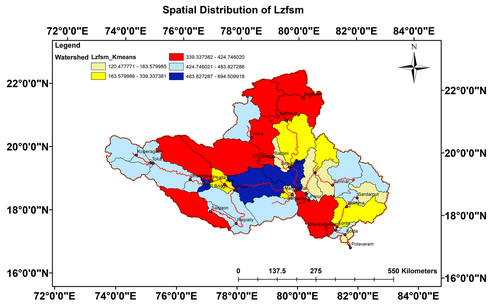

Spatial Distribution of Lzfpm
Legend
Watershed Lzfpm
399.665201 - 443.432400
346.563600 - 378.249200
443.432401 - 476.242800
378.249201 - 399.665200
476.242801 - 506.553360
74°0'0"E
76°0'0"E
78°0'0"E
80°0'0"E
82°0'0"E
22°0'0"N
20°0'0"N
18°0'0"N
N
0 115 230 460 Kilometers

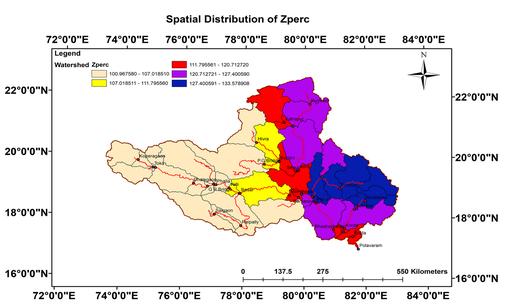

Spatial Distribution of Lzfpm
Legend
Watershed Lzfpm_ANN
415.291691 - 441.409699
381.553198 - 383.484323
441.409700 - 466.711533
383.484324 - 415.291690
466.711534 - 477.937934
72°0'0"E
74°0'0"E
76°0'0"E
78°0'0"E
80°0'0"E
82°0'0"E
84°0'0"E
22°0'0"N
20°0'0"N
18°0'0"N
16°0'0"N
N
0 137.5 275 550 Kilometers

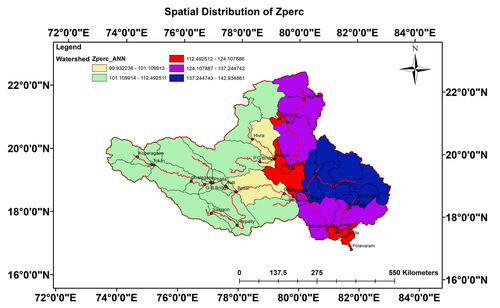

Spatial Distribution of Lzfpm
Legend
Watershed Lzfpm_Kmeans
228.271447 - 433.053113
92.882960 - 166.547268
433.053114 - 578.286745
166.547269 - 228.271446
578.286746 - 916.387984
72°0'0"E
74°0'0"E
76°0'0"E
78°0'0"E
80°0'0"E
82°0'0"E
84°0'0"E
22°0'0"N
20°0'0"N
18°0'0"N
16°0'0"N
N
0 137.5 275 550 Kilometers

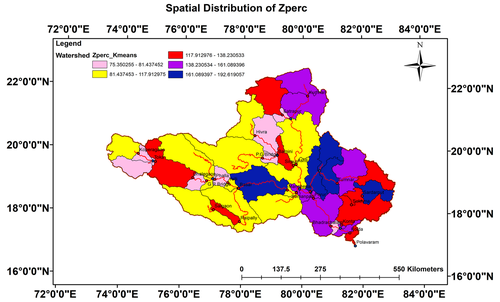

Spatial Distribution of Lzfsm
Legend
Watershed Lzfsm
395.067201 - 419.676400
365.184600 - 381.883700
419.676401 - 438.133300
381.883701 - 395.067200
438.133301 - 455.183960
74°0'0"E
76°0'0"E
78°0'0"E
80°0'0"E
82°0'0"E
22°0'0"N
20°0'0"N
18°0'0"N
N
0 115 230 460 Kilometers


Spatial Distribution of Lzfsm
Legend
Watershed Lzfsm_ANN
381.890517 - 406.013254
327.189659 - 351.156912
406.013255 - 425.043625
351.156913 - 381.890516
425.043626 - 456.338914
72°0'0"E
74°0'0"E
76°0'0"E
78°0'0"E
80°0'0"E
82°0'0"E
84°0'0"E
22°0'0"N
20°0'0"N
18°0'0"N
16°0'0"N
N
0 137.5 275 550 Kilometers


Spatial Distribution of Lzfsm
Legend
Watershed Lzfsm_Kmeans
339.337382 - 424.746020
120.477771 - 163.579966
424.746021 - 483.827286
163.579967 - 339.337381
483.827287 - 694.509918
72°0'0"E
74°0'0"E
76°0'0"E
78°0'0"E
80°0'0"E
82°0'0"E
84°0'0"E
22°0'0"N
20°0'0"N
18°0'0"N
16°0'0"N
N
0 137.5 275 550 Kilometers


Spatial Distribution of Zperc
Legend
Watershed Zperc
111.795661 - 120.712720
100.967580 - 107.018510
120.712721 - 127.400590
107.018511 - 111.795660
127.400591 - 133.578908
72°0'0"E
74°0'0"E
76°0'0"E
78°0'0"E
80°0'0"E
82°0'0"E
84°0'0"E
22°0'0"N
20°0'0"N
18°0'0"N
16°0'0"N
N
0 137.5 275 550 Kilometers


Spatial Distribution of Zperc
Legend
Watershed Zperc_ANN
112.402512 - 124.107886
99.932238 - 101.109913
124.107887 - 137.244742
101.109914 - 112.492511
137.244743 - 142.934861
72°0'0"E
74°0'0"E
76°0'0"E
78°0'0"E
80°0'0"E
82°0'0"E
84°0'0"E
22°0'0"N
20°0'0"N
18°0'0"N
16°0'0"N
N
0 137.5 275 550 Kilometers


Spatial Distribution of Zperc
Legend
Watershed Zperc_Kmeans
117.912976 - 138.230533
75.350255 - 81.437452
138.230534 - 161.089396
81.437453 - 117.912975
161.089397 - 192.619057
72°0'0"E
74°0'0"E
76°0'0"E
78°0'0"E
80°0'0"E
82°0'0"E
84°0'0"E
22°0'0"N
20°0'0"N
18°0'0"N
16°0'0"N
N
0 137.5 275 550 Kilometers

(a) (b) (c)