Chapter 3

RESEARCH METHODOLOGY

3.1 Introduction

The present work aims to estimate runoff based on available rainfall data. The present chapter discuss the details of Rainfall Prognostic Model and prediction of discharge using Hydrological models (HEC-HMS, Soil and Water Assessment Tool) and soft computing technique (Random Forest, M5P models). The study also aimed to estimate model errors in the computational runoff, the time-dependent and posterior fire-breathing network is utilized and the prophetic multilayer network is directly considering the relationship between infiltration and soil moisture through a collection of parsimonious computational parameters to estimate the infiltration rate. This study also aimed to estimate the velocity of runoff on steep slopes by utilizing well-ordered selective genetic algorithm. This study compares the forecast discharge of (HEC-HMS,SWAT) vs Soft computing techniques (Random Forest vs M5P).

3.2 Rainfall Prognostic Model-Based Artificial Framework

For modeling the rainfall-runoff process, models have been developed that are based on conceptual representations of the physical processes of the water flow lumped over the entire catchment area (lumped conceptual type of models). The parameters of such models cannot, in general, be obtained directly from measurable quantities of catchment characteristics, and hence model calibration is needed. To calibrate a model, values of the model parameters are selected so that the model simulates the hydrological behavior of the catchment as closely as possible. The process of model calibration is normally done either manually or by using computer-based automatic procedures. In manual calibration, a trial-and-error parameter adjustment is made. In this case, the goodness-of-fit of the calibrated model is based on a visual judgment by comparing the simulated and the observed hydrographs. However, since there is no generally accepted objective measure of comparison, and because of the subjective judgment involved, it is difficult to assess explicitly the confidence of the model simulations. The development of such relationships is an example of a regionalization

methodology and is a useful goal for a number of reasons. For example, the construction of a hydrologic structure such as a bridge or dam may require a prediction to be made of the hydrologic response of a catchment at an ungauged point

Rainfall prediction is the integration of science and technology to estimate the atmosphere's environment. In order to effectively utilize water supplies, crop production, and pre-planning water systems, the rainfalls should be calculated in advance. Various strategies for runoff can avoid precipitation**,** and accurate input of rainfall in time and space is crucial to the response of rainfall-runoff modeling. Recently conceptual rainfall-runoff models use basic spatially lumped storage-flow relationships to compensate for the spatial complexities of hydrological processes within a basin. These rough approximations introduce model errors that are hard to describe. Also, it is imperative to estimate uncertainties in measured data (e.g., discharge), model inputs (e.g., rainfall, parameters), and the model itself. Failure to consider any sources of uncertainty would result in an underestimation of the prediction ability. While formulating an effective model error and parameterization, the parameters (e.g. variance of random noise) must be calculated from the entire available data. The hybrid AI framework is designed by utilizing the novel rainfall prognostic model-based artificial framework which use the posterior fire-breathing network to estimate model errors. Also, the model errors are estimated by prophetic multilayer network which relies on soil run off levels. Furthermore, to tackle the numerical challenge of measuring surface water from steep slopes, the system used a well-ordered selective genetic algorithm to determine the velocity of runoff in distinct bend areas. Further to determine the runoff level in the state-of-the-art methods, many limitations lead to a high infiltration rate. In such modeling methods, runoff rates from metropolitan centers are believed to be a flat rate value that is essentially disconnected from the underlying soil moisture. Since soils vary from undeveloped areas to urban areas due to the mixture of compaction and synthetic materials in the soil, results in different infiltration rates than uncompacted soils. As a result, runoff in urban areas is growing due to impermeable surfaces; soils also impact runoff rates in urban green areas (such as parks and sports fields). Furthermore, to protect the surface water from steep slopes face velocity measurements on different scales and a watershed application evaluates the velocity field. Many works are

involved in describing the fields of velocity in a particular case and some works that tackle numerical problems. The low water depths runoff steep slopes which contribute to the physical representation of the friction at the bottom for specific water depth to height ratios of roughness. Nonetheless, the capability can be very difficult in mountain environments like heterogeneous and high low slope values as well as intermittent rain inputs. Therefore, the paper proposed a novel Rainfall prognostic artificial model framework, the framework relies on the models of artificial neural networks. The architecture of the proposed framework is shown in figure 3.1.

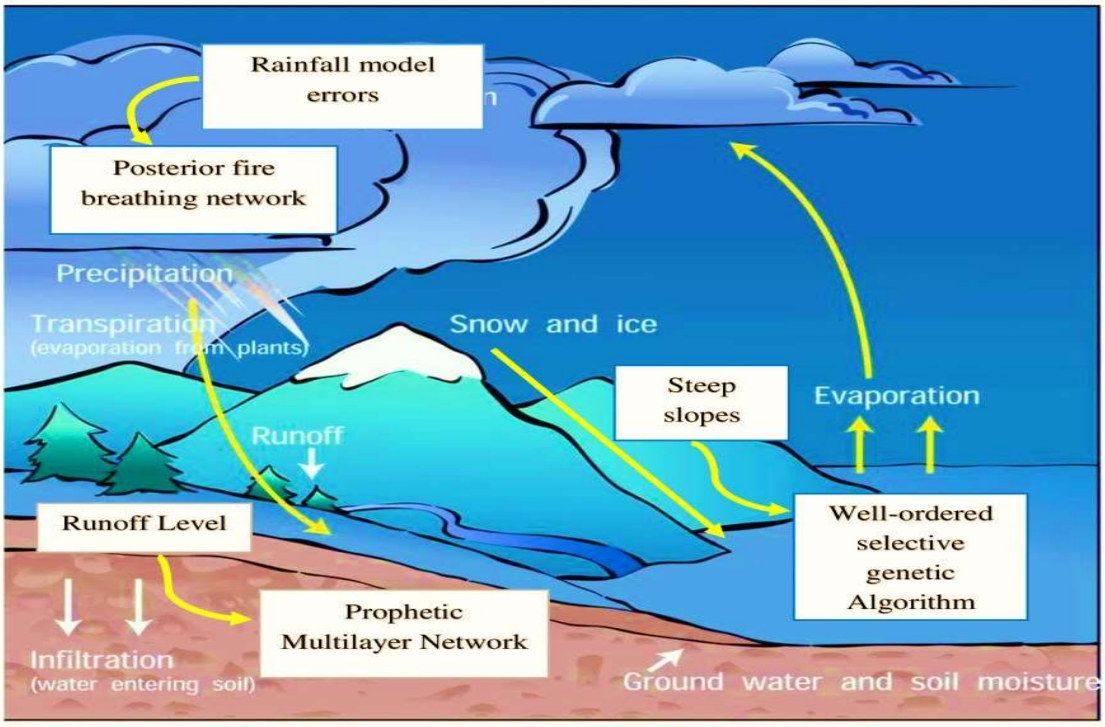

Figure 3.1 Architecture of Rainfall Prognostic Model-based Artificial Framework

Initially, the framework introduced and applied a novel Posterior fire-breathing network to estimate model errors of a conceptual rainfall-runoff model and to perform uncertainty analysis. The solution accounts for model errors by adding time-dependent, random noise into the internal storage of the hydrologic model. The model also effectively quantifies variability in the performance and the parameters arising from variance in the input data used to calibrate. Further, the level of runoff is simulated using the Prophetic Multilayer Network technique which is employed to model infiltration. The technique is also

found to be related to the input as the exactness of the ANN model was higher for soil types with higher proportions in the training data. The fraction of precipitation converted into either runoff or infiltration depends on the amount of soil moisture, and the dynamics of soil moisture vary according to the model's rural or urban portion. In addition to overcoming runoff at low water depths on steep slopes, the framework proposed a well-ordered selective genetic algorithm (WSGA) for velocity measurements and predictions in different bend sections and comparing their test results with each other and with actual data. Four experimental test cases have been used to explore and evaluate the numerical model and to understand its behavior better. The overall rainfall modeling is enlightened in the following sections.

3.2.1 Posterior Fire-breathing Network

The network calculates a conceptual rainfall-runoff process's model errors and analyzes uncertainty. The model error assessment assumes that the results of the discharge reflect exactly the hydrological behavior of the catchment. However, time series are successively extracted from stage measurements and conversions of stage ratings and therefore may contain highly uncertain values. Thus, by incorporating time-dependent, spontaneous noise into the internal storage of the hydrological model, the solution accounts for model errors and thus efficiently quantifies uncertainty in model output and parameters resulting from variation in the input data used for calibration. Adoption of posterior fire breathing network in driving a rainfall-runoff model reduce the error by estimating the run off at the catchment outlet with different rainfall inputs from different parts of the catchment. Initially, the noise parameter of the model errors was calculated and revised sequentially through the approximation of the Gamma density to its posterior. With the approach suggested in the article, Gamma approximations were obtained in these graphs, while Gaussian approximations have the same mean and variance as the Gamma approximations. These observations suggest that the posterior is originally asymmetric and non-Gaussian, but as more data is assimilated, it tends to become more Gaussian.

The objective is to recursively estimate posteriors for the model error precision including for the Hymod states and parameters using the posterior fire-breathing network. Each time a major discharge discovery appears accessible it is incorporated by

a) Updating the posterior of τ (precision of model errors),
b) Updating the states and parameters of Hymod.

It is realized that the model error accuracy posterior is modified separately from the state and parameter posterior. The alternative of introducing the precision as an external unknown to the state vector, i.e., the state-augmentation method, is never considered here. Information on mathematics to update posteriors is given in the following two sections.

3.2.1.1 Sequential updating of the posterior of τ

The posterior precision or inverse variance of the model errors is calculated sequentially in the system. It is accomplished by preserving a Gamma density approximation of the posterior: as new data become available the approximation is modified in real-time to represent new knowledge on model errors. Thus, the Gamma approximations are used for their corresponding calculations of the accuracy (or variance) of model errors a positive sum. In Bayesian analysis, the Gamma density is a popular approach for modeling precision variables, partially even with its mathematical properties; e.g., a Gamma prior is conjugated to a Linear-Gaussian probability, leading to posterior closed-form expressions. In the study, the focus is on dealing with differential equations. This is because it's not possible to access the posterior analytically. However, there's an effective approach for estimating it, as indicated in the following discussion. The factors in an approximate but effective recursive approach for determining the posterior of τ that is completely Bayesian and extends point (posterior mode) estimation of the model noise.

Mathematically, the posterior of τ at time t is written recursively in terms of the posterior at the time $t-1$ and the likelihood at the time t. For model noise on a selected flow or state variable x, as shown in equation 1;

$$p(\tau \mid D_t) \alpha p(\tau \mid D_{t-1}) s \int_{x_t} \int_{\mu_t} p(\mu_t \mid D_{t-1}) N(x_t \mid \mu_t, \tau^{-1}) p(D_t \mid x_t) d\mu_t dx_t \qquad \text{eq (1)}$$

Where, D_t is the current discharge measurement, and $D_t = (D_{t-1}, D_t)$ what holds all up to and including the date t of discharge. The prior of τ at the time t is inscribed as the double integral represents the likelihood τ at the time t: $p(\mu_t \mid D_{t-1})$ is the predictive distribution

of the chosen flow or state variable before adding noise, $N\left(x_t \mid \mu_t, \tau^{-1}\right)$ is the model noise term $x_t \square N\left(\mu_t, \tau^{-1}\right)$, and $p\left(D_t \mid x_t\right)$ is a likelihood at time t of the true flow or state (after adding noise).

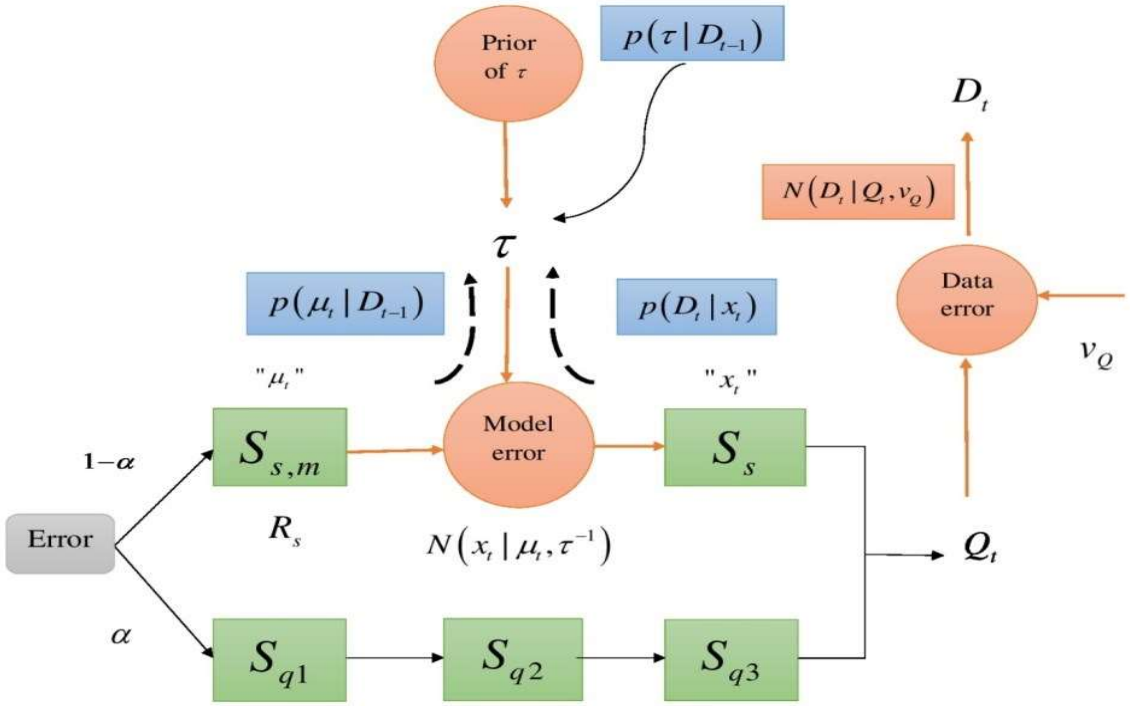

Figure 3.2 Schematic diagram of the deterministic (black color) and stochastic (red color) components of the modified Hymod model.

Rectangles in Figure 3.2 denote reservoirs for which a water balance is determined. Model error is applied to the slow-reservoir state as an example. The changes in areas that make up the posterior τ are seen in blue**.** Equation (1) can be understood by reference to the model error depicted in Figure 3.2, which shows the case of model error added to the slow-reservoir state S_s. The case μ_t corresponds to the modeled state $S_{s,m}$ ("m" indicates modeled) before adding model error and x_t corresponds to the true state S_s after adding model error. To update the posterior for τ at time t, famine to combine three sources of information (arrows in Figure 3.2):

(i) the prior for τ (i.e., the posterior for τ at the time $t-1$),

(ii) the predictive distribution of S_s, $p(\mu_t \mid D_{t-1})$ and

(iii) Probabilistic information on *Ss* provided by the current discharge observation $p(D_t \mid x_t)$

As Figure 3.2 shows, the latter two are combined $(x_t \mid \mu_t, \tau - 1)$ to provide a probabilistic information stream τ; which constitutes the likelihood term for τ and it corresponds to the double integral in Equation (1).To make the double integral in Equation (1) tractable, both the predictive distribution $p(\mu_t \mid D_{t-1})$ and likelihood $p(D_t \mid x_t)$ are approximated by Gaussian distributions. The predictive distribution $p(\mu_t \mid D_{t-1})$ is approximated by $N(\mu_t \mid \mu_\mu, v_\mu)$ where μ_μ *and* v_μ are mean and variance, respectively, of an ensemble of predicted values μ_t, which is the selected flow or state variable before adding model noise. Likewise, likelihood $p(D_t \mid x_t)$ is approximated by $N(x_t \mid \mu_x, v_x)$, where μ_x *and* v_x are estimated by linearizing via linear regression the relation between predicted Q_t and x_t $Q_t = \mu_{Q,pred} + \psi(x_t - \mu_{x,pred})$ *or* $x_t = \frac{Q_t - \mu_{Q,pred}}{\psi} + \mu_{x,pred}$. The last equation is then used to translate mean D_t and variance v_Q for the likelihood of Q_t into mean and variance of x_t:

$$\mu_x = \frac{D_t - \mu_{Q,pred}}{\psi} + \mu_{x,pred} \qquad \text{eq (2)}$$

$$V_x = \frac{v_Q}{\psi^2} \qquad \text{eq (3)}$$

Where $\mu_{Q,pred}$ *and* $\mu_{x,pred}$ (ensemble) means of are predicted Q_t *and* x_t, respectively, and ψ is the linear regression coefficient between the two ensembles. With these Gaussian approximations, can rewrite Equation (1) in the following form:

$$p(\tau \mid D_t) \alpha p(\tau \mid D_{t-1}) \int_{x_t} \int_{\mu_t} N(\mu_t \mid \mu_\mu, v_\mu) N(x_t \mid \mu_t, \tau^{-1}) N(x_t \mid \mu_x, v_x) d\mu_t dx_t \qquad \text{eq (4)}$$

As all terms under the integral are now Gaussian, the double integral can be easily computed (see appendix A). While the yields a closed-form formula for the posterior, the resulting density does not correspond to a known parametric density. To have an interactive recursive estimation algorithm, i.e., one that makes fast posterior sampling and eliminates the need to repurpose data and likelihoods from specified time measures, it recommends estimating the posterior by a new Gamma density. Under such estimation, the posterior will increase in complexity every time step, for each time steps an additional possibility term is introduced that transforms the posterior into a non-standard dimension. Therefore, the mathematical expression for the posterior increases in complexity even as the posterior is one-dimensional. The Gamma approximation avoids and preserves a reliable estimation of the posterior, and therefore enables for recursive updating rather than going back to different results, while still being simple to analyze the form.

Hence, at each time step, the posterior τ is represented by a Gamma distribution: $p(\tau \mid D_{t-1}) \approx Ga(\tau \mid \alpha_{t-1}, \beta_{t-1})$ *and* $p(\tau \mid D_t) \approx Ga(\tau \mid \alpha_t, \beta_t)$.The distributor allows us to derive closed-form recursive update equations for α_t and β_t

$$\beta_t = \alpha_{t-1} - \tau^2 \frac{d^2 \log f(\tau)}{d\tau^2} \qquad \text{eq (5)}$$

$$\beta_t = \beta_{t-1} - \frac{d \log f(\tau)}{d\tau} - (\alpha_t - \alpha_{t-1})\tau^{-1} \qquad \text{eq (6)}$$

$$\tau = \frac{\alpha_t - 0.5}{\beta_t} \qquad \text{eq (7)}$$

Where expressions for log f and its derivatives are given in the appendix. Since α_t *and* β_t equations 5 and 6 depend on τ, update τ in equation 7, and recomputed α_t *and* β_t again with equations 5 and 6. The update in equation 7 sets τ to a value that lies in between the mean $\left(\frac{\alpha_t}{\beta_t}\right)$ and the mode $\left(\frac{\alpha_t - 1}{\beta_t}\right)$ of the new Gamma density approximation. At each time *t*, these three equations are iterated 10 times to yield a Gamma approximation of the posterior of τ for that time step. Iterate 10 times but typically observe convergence after 2-5

iterations. The iteration could be made more sophisticated by monitoring convergence of the iterations and using a stopping criterion, but the computational cost of 10 iterations is small. As new data come in, these update equations result in an automatic tuning of the amount of model noise τ so that the resulting probabilistic prediction is neither too wide nor too narrow. The use of gamma approximation calculates the time between independent events that occur at a constant average rate. Using this approximation, analysts specify the number of events, such as modeling the time until the nth accident occurs.

3.2.1.2 Sequential updating of the Hymod state and parameter posteriors

The states and parameter posteriors are expressed by ensembles, which are produced in the prediction phase illustrated above. The ensemble Kalman Filter (EnKF) is used to modify the actual flow observation of an ensemble participant which is ideal for systems that are continuously changing. Ensemble kalman filter have the advantage that are light on memory which don't need to keep any history other than the previous state, and very fast, making them well suited for real-time problems and embedded systems than particle filter.

$$\left(S_t,\theta \mid D_t\right)_i = \left(S_t,\theta \mid D_{t-1}\right)_i + K_t\left(D_t - Q_t\right)_i \quad i = 1,2,.....N \qquad \text{eq (8)}$$

Where D_t represents randomly perturbed observed discharge (with perturbation variance v_Q). On the right-hand side of Equation 8, predicted values for states S_t and discharge Q_t are generated by the sampling process described in the predict step. Finally, K_t Equation 8 is the Kalman gain vector and is written as:

$$K_t = V_{(S_t,\theta),Q_t}\left(v_{Q,pred} + v_Q\right)^{-1} \qquad \text{eq (9)}$$

Where $V_{(S_t,\theta),Q_t}$ is a vector of ensemble cross-covariance between states-parameters $\left(S_t,\theta\right)$ and predicted discharge Q_t *and* $v_{Q,pred}$ is the ensemble variance of predicted discharge Q_t.

$$\left[S_t\right] > 0 \qquad \text{eq (10)}$$

$$Ib < [\theta] < ub \qquad \text{eq (11)}$$

$$SMc < C_{\max} / (\beta + 1) \qquad \text{eq (12)}$$

An upgrade to EnKF Values for states and parameters outside their physical range may result in Equation 8. With the purpose, of preceding the EnKF update, the accompanying restrictions are extended to states and the model parameters with better optimization in the gamma distribution. Further, to approximate the penetration and runoff level the system integrates a prophetic multilayer network which is concise in the following section.

3.3 Prophetic Multilayer Network

The paper explores whether linking soil moisture and urban runoff will facilitate model simulations that overtake fixed percentage runoff. A modern approach to modeling infiltration and runoff levels across urban surfaces is implemented for the purpose, which directly considers the relationship between infiltration and soil moisture through a collection of parsimonious computational parameters. The models of infiltration are incorporated within a framework of uniform rainfall-runoff. The prophetic multilayer network allows for probability-based predictions and the classification of items into multiple labels. The advantages of a prophetic multilayer network are capable to learn non-linear models and models in real-time.

Prophetic Multilayer Network is trained by a back-propagation algorithm, which is fast and easy to program also it does not require additional parameters to tune apart from the numbers of input and does not require prior knowledge about the network. Back-propagation ANN is used to model the relationship between experimentally obtained scattering coefficient data and the degree of ground unevenness, and the soil moisture content of the field. As a system of linked layers of artificial neurons, the BP network is structured, input, hidden, and output layers are shown in Figure 3.3 as a simple representation of such a network. When data is provided to a neural group through the input layers, the input layer neurons spread the weighted data and randomly draw bias via the hidden layers. The output at each node of the hidden layer is then calculated consuming a transmission purpose.

Afterward, the generated signals are transferred either to the next hidden layer or the output layer.

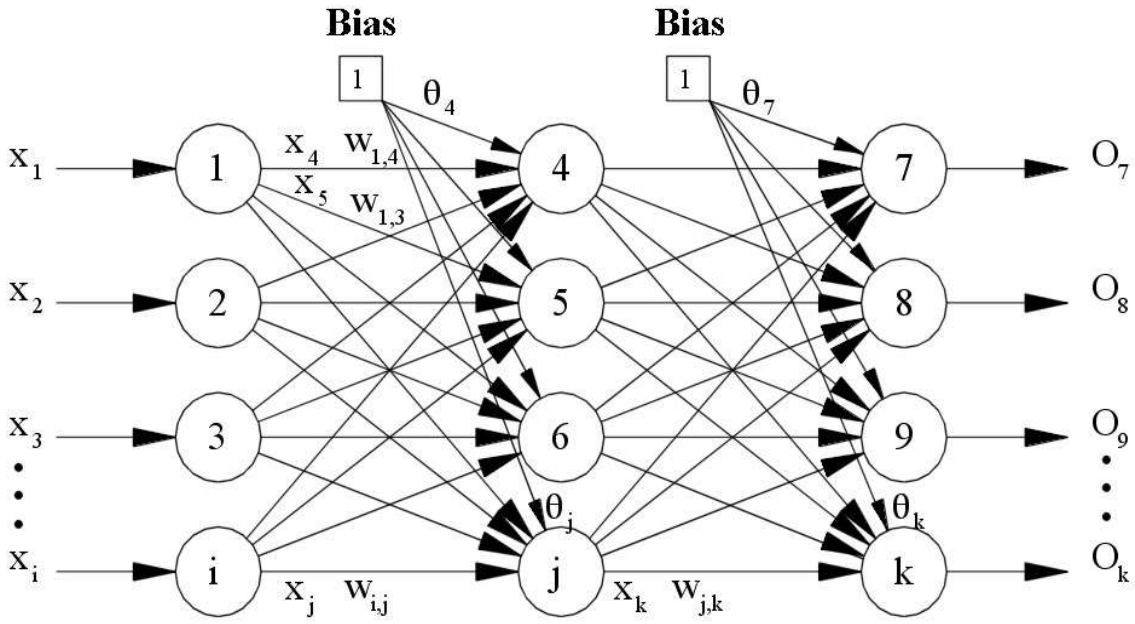

Figure 3.3 Network Layer

The prophetic multilayer network incorporates a deterministic, continuous-time; lumped, computational rainfall-runoff model intended to simulate catchment runoff like the urban land-cover (URMOD) model for sustainable development. A lumped model doesn't break down catchments into different sub-catchments but takes into consideration a single catchment with only one outlet. The hydrological implications of urbanization are directly accounted for by dividing a catchment into a rural and an urban portion, with specific infiltration (and therefore runoff) and routing features allocated to each segment. URMOD has nine parameters requiring calibration of the observed precipitation, possible evaporation, and river flow.

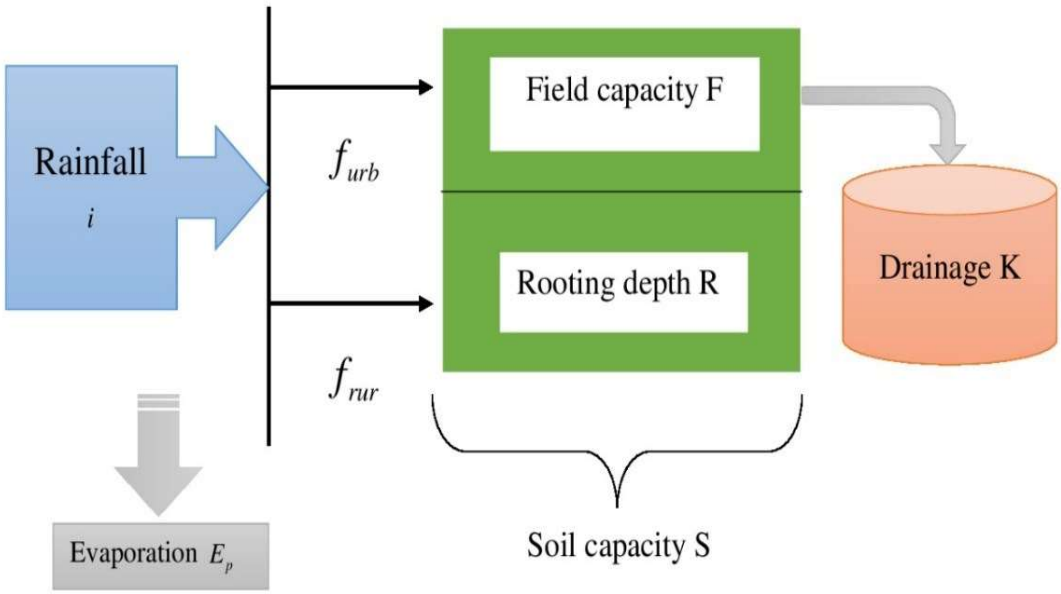

Figure 3.4 Visual representation of the URMOD model.

Figure 3.4 shows the processes of infiltration and runoff with the conceptual soil column and three sectors, sector 2 with field capacity but exceeding the rooting depth (R < m < F) and sector 3 with soil moisture below rooting depth (m < R) where soil moisture regulates infiltration and runoff generation.

3.3.1 Infiltration and runoff line of attack

A soil-column-based approach is used to model the surface runoff and infiltration. The water which does not enter into the column of the soil is transformed into a direct runoff. The proportion of precipitation that turns into either runoff or absorption depends on the amount of soil moisture, and soil moisture characteristics differ according to the rural or urban component of the model. In the rural region, the temporal increase in soil moisture is induced by three processes: (i) infiltration, (ii) runoff, and (iii) evaporation. When the urban land cover is present then two inputs will influence the infiltration through the catchment; infiltration from rural areas and industrial areas respectively. The cumulative infiltration from the two land-cover classifications is expressed as a weighted average infiltration (f) as shown in equation 13:

$$f = i(1-u) f_{rur} + iuf_{urb} \qquad \text{eq (13)}$$

U is the fraction of the total catchment area covered by urban land cover where there is rainfall f_{rur} represents rural infiltration and is defined in equation (14), and they signify infiltration in urban areas f_{urb}. The rural infiltration indicated in Equation (14) is based on the PDM model

$$f_{rur} = \left(1 - \frac{m}{s}\right)^{\frac{1}{2}} \qquad \text{eq (14)}$$

Where, soil moisture content (mm) is m, and soil column capacity (mm) is S, thus $0 \le m/S \le 1$. When the column of the soil reaches saturation $\left(m/S \approx 1\right)$, the infiltration is poor and the rest of the rain is turned into a direct runoff. It is assumed that the conceptual soil column has three separate zones reflecting soil moisture levels, all of which are balanced parameters, controlled by field ability (F), and rooting depth (R) Drainage and soil column evaporation depend on the degree of soil moisture shown in Figure 3.4.

Section 1, near the soil surface, is defined as if soil moisture is above the capacity of the field (m > F). In this case, the possible rate of evaporation (Ep) is assumed, and drainage to deeper storage depends on moisture content (m) and a calibrated drainage coefficient (k) so drainage from the column occurs at a rate of k (m−F). Then the equation for the region is shown in Equations 15.

$$\frac{dm}{dt} = \underbrace{i\left(1-u\right)\overbrace{\left(1-\frac{m}{S}\right)^{\frac{1}{2}}}^{f_{rur}} + iuf_{urb}}_{Infiltration} - \underbrace{k\left(m-F\right)}_{Drainage} - \underbrace{E_p}_{Evaporation} \qquad \text{eq (15)}$$

Sector 2 is where the soil moisture does not exceed the potential of the field but exceeds the rooting depth (R < m < F) the evaporation is again at the highest rate (Ep) but no drainage takes place. Then the equation for the region is shown in Equations 16.

$$\frac{dm}{dt} = \underbrace{i\left(1-u\right)\overbrace{\left(1-\frac{m}{S}\right)^{\frac{1}{2}}}^{f_{rur}} + iuf_{urb}}_{Infiltration} - \underbrace{E_p}_{Evaporation} \qquad \text{eq (16)}$$

Sector 3 when soil moisture is below the rooting depth (m < R), there is again no drainage, and evaporation reduces linearly with depth $E = E_p\left(m/R\right)$ until it reaches E = 0 for m = 0. Three Different differential equations explain the distribution of soil moisture in each of the three different regions. The definition of infiltration does not alter in any of these equations and is defined by Equation 14. Therefore, the zone equation is shown in Equations (17).

$$\frac{dm}{dt} = \underbrace{i\left(1-u\right)\overbrace{\left(1-\frac{m}{S}\right)^{\frac{1}{2}}}^{f_{rur}} + iuf_{urb}}_{Infiltration} - \underbrace{E_p\frac{m}{R}}_{Evaporation} \qquad \text{eq (17)}$$

A leveling concept for the amount of evaporation occurring in urban areas is applied. Although there is no agreement on the importance of evaporation in urban areas in modeling studies, importance is agreed to be lower than the total in rural areas and greater than no evaporation. The first urban extension assumes a fixed percentage of infiltrated rainfall, thus de-coupling the hydrological cycle on urban surfaces with soil moisture. The second extension presumes the generation of runoff and infiltration in urban areas are dependent on a scaling term denoted γ, the urban surface infiltration depends on the soil moisture content of the rural areas but is decreased by a factor $\left(1-\gamma\right)$. It thus links urban infiltration directly to soil humidity levels.

3.3.2 Urban enlargement 1: Fixed percentage runoff

Urban enlargement 1 assumes a fixed proportion of runoff from the metropolitan region, and typically runoff is independent of soil moisture. The fixed percentage runoff will be denoted ω. Comparing ω the runoff generated for the rural areas is obvious that ω a threshold would be defined since there are high levels of soil moisture that the amount of runoff generated from rural areas exceeds urban runoff rates, which is considered counter-intuitive. The threshold for soil moisture $\frac{m}{s}$ at which the shift occurs is derived as in equation 18;

$$\underbrace{1-\left(1-\frac{m}{S}\right)^{\frac{1}{2}}}_{rural} > \underset{urban}{\omega} \Rightarrow \frac{m}{S} > 1-(1-\omega)^2 \qquad \text{eq (18)}$$

For the soil moisture levels above and below the point, therefore, penetration from urban areas must be addressed. When the amount of soil moisture reaches this threshold then urban area infiltration can return to actions as in rural areas. Therefore, the infiltration into urban areas is described as equation 19;

$$f_{urb} = \begin{cases} 1-\omega & : 0 \le m/S \le 1-(1-\omega)^2 \\ \left(1-\frac{m}{S}\right)^{\frac{1}{2}} & : 1-(1-\omega)^2 < m/S \le 1 \end{cases} \qquad \text{eq (19)}$$

By replacing Equations 14 and 19 with Equation 13, total infiltration can be described as urban and rural.

$$f = \begin{cases} (1-u)\left(1-\frac{m}{S}\right)^{\frac{1}{2}} + iu(1-\omega) & : m/S \le 1-(1-\omega)^2 \\ \left(1-\frac{m}{S}\right)^{\frac{1}{2}} & : m/S > 1-(1-\omega)^2 \end{cases} \qquad \text{eq (20)}$$

While ω would traditionally be set as a fixed value such as 70 percent or 100 percent, Evaluate the specific amount valuation that depends entirely on urbanization is categorized as a calibrated value for the research.

3.3.3 Extension 2 to urban infiltration: Multiplicative impacts on urban areas

Urban extension 2 believes that urban infiltration relies on soil moisture, comparable to rural infiltration but reduced by a multiplicative factor $(1-\gamma)$. Extension 2 thus prevents the explicit implementation of a threshold as provided in extension 1. Defines the functional form as equation 21:

$$f_{urb} = (1-\gamma)\left(1-\frac{m}{S}\right)^{\frac{1}{2}} \qquad \text{eq (21)}$$

The calibration parameter $\gamma \in [0,1]$ is implemented to account for the variability of infiltration across various urban catchments so that a large one indicates that the urban area is largely impervious and more runoff is generated, while a smaller value indicates that the urban area has more perennial surfaces and less runoff is produced. If γ is zero then penetration is the same for the impervious area as the rural area, whereas if γ is one area then the area will be fully sealed and no intrusion will occur. In substitution of Equation 14 and Equation 21 into Equation 13, the complete infiltration is derived as:

$$f = \underbrace{i(1-u)\left(1-\frac{m}{S}\right)^{\frac{1}{2}}}_{rural} + \underbrace{iu(1-\gamma)\left(1-\frac{m}{S}\right)^{\frac{i}{2}}}_{urban} = i(1-u\gamma)\left(1-\frac{m}{S}\right)^{\frac{1}{2}} \qquad \text{eq (22)}$$

Equation 22 defined for infiltration can then be replaced by the three soil moisture equations. Thus, the infiltration rate in the soil moisture and runoff level is determined successfully in the urban and rural areas. Then the steep slopes in the low depth of hills affect the surface water is estimated by the velocity with a novel genetic algorithm is discussed below.

3.4 Well-ordered selective genetic algorithm

The well-ordered selective Algorithm evaluates the importance of a mathematical model to determine the velocity profile in a watershed with rain-induced runoff, using only a displacement rule customized to every disturbance and upwelling regime. More broadly, for the representation of the velocity field, describe the speed field in each situation and trust the numerical model.

The two-dimensional deep ocean models are transformed from the Naiver-Stokes equations with hydrostatic pressure and standardized vertical velocity statements where the flow is deep and the slope fluctuations are not increasing, certain assumptions are accessible which is written as

$$\begin{cases} \dfrac{\partial h}{\partial t} + \dfrac{\partial hu}{\partial x} + \dfrac{\partial h\upsilon}{\partial y} = R - I \\ \dfrac{\partial hu}{\partial t} + \dfrac{\partial\left(hu^2 + \dfrac{gh^2}{2}\right)}{\partial y} = gh\left(-\dfrac{\partial z}{\partial y} - S_{fx}\right) \\ \dfrac{\partial h\upsilon}{\partial t} + \dfrac{\partial hu\upsilon}{\partial x} + \dfrac{\partial\left(h\upsilon^2 + \dfrac{gh^2}{2}\right)}{\partial y} = gh\left(-\dfrac{\partial z}{\partial y} - S_{fy}\right) \end{cases} \qquad \text{eq (23)}$$

Where t is the time, h is the height of the water, v the velocity of flow in the y-direction, u the velocity of flow in the x-direction R the strength of the rain, g is the gravity constant, z is the bottom elevation, and S_{fx} *and* S_{fy} the friction slope respectively in the x and y directions.

3.4.1 Numerical resolution

An unstructured mesh is used to discretize the domain in space. The model enables better adaptation than structured mesh to a basin's complex topography by better representing its topography, especially the river and gully shape in the river system. Also, these meshes lie in the ability to deal with arbitrary complex geometries and the ease with which local refinements of the grid are to be implemented. At each node of the mesh, a control volume is constructed creating a boundary between each neighboring node of the mesh, as shown in Figure 3.5.

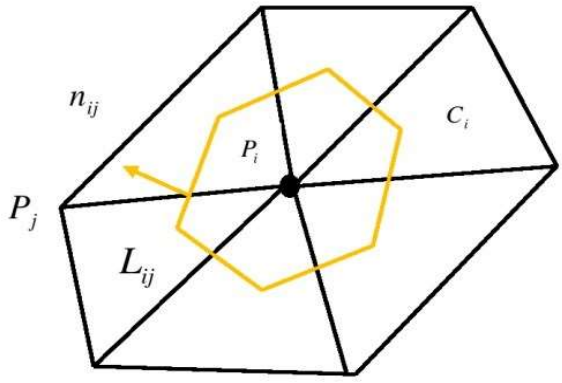

Figure 3.5 Representation of the formalism used for the 2D unstructured mesh resolution.

The Shallow Water Eq. (23) can be solved with a finite volume scheme which writes:

$$U_i^{t+1} = U_i^t - \sum_{j \in V_i} \left(\frac{\Delta t}{\Delta x_{ij}} F_{ij}^t + \frac{\Delta t}{\Delta x_{ij}} S_{ij}^t \right) \qquad \text{eq (24)}$$

where $U_i^t = \left(h_i^t, h_i^t u_i^t, \upsilon_i^t \right)$ the states at a time t in the node P_i, V_i all the nodes surrounding P_i, Δ_t the time step, $\Delta x_{ij} = \frac{C_i}{L_{ij}}$ the space step with C_i the area of the cell constructed around the node P_i *and* L_{ij} the length of the boundary edge between the cells C_i *and* C_j, F_{ij}^t is the numerical value at the interface between the cells C_i *and* C_j along n_{ij}, the normal vector to the boundary edge and $S_{ij}^t = \left(R_i - I_i, s_{ijx}^t, s_{ijy}^t \right)$ are the source terms of mass and momentum along *x* and *y* directions. The key challenge in solving these equations is how to treat the terms slope source s_{ijx}^t *and* s_{ijy}^t to respect the water depth and hydrostatic balance positivity and, more specifically, for wet/dry transitions and steep slopes. The Shallow Water Equations are solved by the scheme (24).

A fully wet state when the water depth is higher than the bottom difference between two cells, and a partially wet regime otherwise, are distinguished at the interface of two cells. Consider the non-dimensional number $\beta = \frac{h}{\Delta x \partial_x z}$, if β is inferior to 1 the regime is partially wet and fully wet otherwise. The exact error made on the source term can be estimated. Then the experimental test cases are to be carried out.

3.4.2 Experimental test cases

Four test cases are selected at different scales and degrees of complexity. The first two laboratory test cases are selected with simple geometry. The first laboratory case is used to evaluate the effect of the resolution method and the other one to assess the effect of Friction

Law on local velocities. Then on a small plot, the model is applied to a real case to confirm the results of the two previous test cases. Finally, the model is tested over a watershed of around 1 km^2 on which the discharge at the outlet can be repeated.

The first experiment is a flume with a constant slope on which a constant raindrop is used to measure the error due to the numerical scheme. The second test case is a furrow-shaped sinusoidal mold reflecting the agricultural land. In comparison with measurements, the conclusions of these simulated results are then represented using two laboratory experiments obtained on the This test case. Eventually, the model was applied to a specific catchment in the Laval where only outlet discharge measurements are available.

3.4.2.1 Flume experiment

A steady stream, with an intensity of 25 mm / h or 50 mm / h depending on the chosen scenario, is applied with a stream simulator over the domain during 600 s. With a time, phase of 0.1s the temporal evolution of the discharge at the outlet of the channel is observed. The temporal evolution of the discharge is achieved by measuring the water that flows from the canal. Once the steady-state is achieved, the water depth and the velocities are also determined along the canal every 60 cm. These are calculated using a system of salt tracing. For the three slope values, the dataset is available: 2%, 5%, and 25%. All combinations of slope and rain intensity are used except for the case with a slope of 5 percent and a rain intensity of 50 mm / has no data on the event. This result implies that have convergence in mesh size because whatever the slope and the positive water height defined, a mesh size is adapted in the case where depth is large enough to reduce the error.

3.4.2.2 Sinusoidal moulds experiment

The second test case is a laboratory test case describing a real area with several flood-ratio values. Runoff of water depths from 1 to 3 mm, of high speeds is observed at the inlet. The flow is then led to furrows where the depths of the water can exceed 2 cm. The test case is representative of the hydraulic change that can be seen on an actual plot with a hydraulic leap between the runoff and the flow in the furrows. The number of furrows ranges from one to three depending on the slope option and the discharge from the inlet. The data on the test

case are calculated in a steady state so that they do not change in time. The water depths and the bottom elevation are determined using a laser scanner with a spatial resolution of 0.5 mm and the experimental water depth uncertainty is 1 mm. concerning the velocities, they are measured using the LSPIV process, with a 5 mm spatial resolution and a 0.1 m / s experimental uncertainty. The location of the hydraulic jump and the flow in the furrows are in good agreement with the experiments. Indeed, the maximum water depths, as well as the location of the hydraulic jump entrance in the furrows are very well reproduced. However, the water level is a little overestimated in the farthest furrows, especially

3.4.2.3 Plot experiment

The plot is ten meters long, and four meters wide. For the domain, a uniform rain is applied using a rain simulator. Following many rain events that eroded the soil, the mobility of sediments in the soil was considered marginal and the soil is assumed to be fixed for the time of interest. The soil is composed mainly of sand with a ruggedness scale that will be called spatially uniform. Measurements of the velocities were taken at 62 separate locations on the map. As in the first experimental scenario, the measuring technique is salt-tracing. For plot experiment case, there is no water level calculation is required as compared to the previous test case. Some anomalies were found on the topography at the edges, which can be attributed primarily to measurement errors.

3.4.2.4 Watershed application

The model is then put to the test by measuring the effects of the experiment to Observed discharges on a particular catchment from field observations. The climate is a system of mountains influenced by the Mediterranean climate, reflected in particular by heavy summer storms and an annual average of 900 mm of rainfall. Its total area is 0.86 km2 and 58 percent of the mean slope. The soil consists primarily of black marls, and 68 percent of the surface is bare earth. The data are valid for several rainfall events at the catchment source, with a time stage of 60 s. Only the rainfall is measured every 60 s for each case. The flow rate is measured over a specified fixed section by measuring the water depth by laser. The basin size and steep slopes make the hydrological response fast. The choice of these events is

motivated by the fact that the soil's initial state is very different from one to another, but also because they are the two events that have had the greatest impact on the transport of sediments in one year.

The model highlights well-known difficulties in solving the Shallow Water Equations for steep slopes and low depths of water. This offers an empirical criterion for determining, in a given situation, the discretization of the term slope source to prevent the public from erosion. The work flow of proposed method given in figure 3.6.

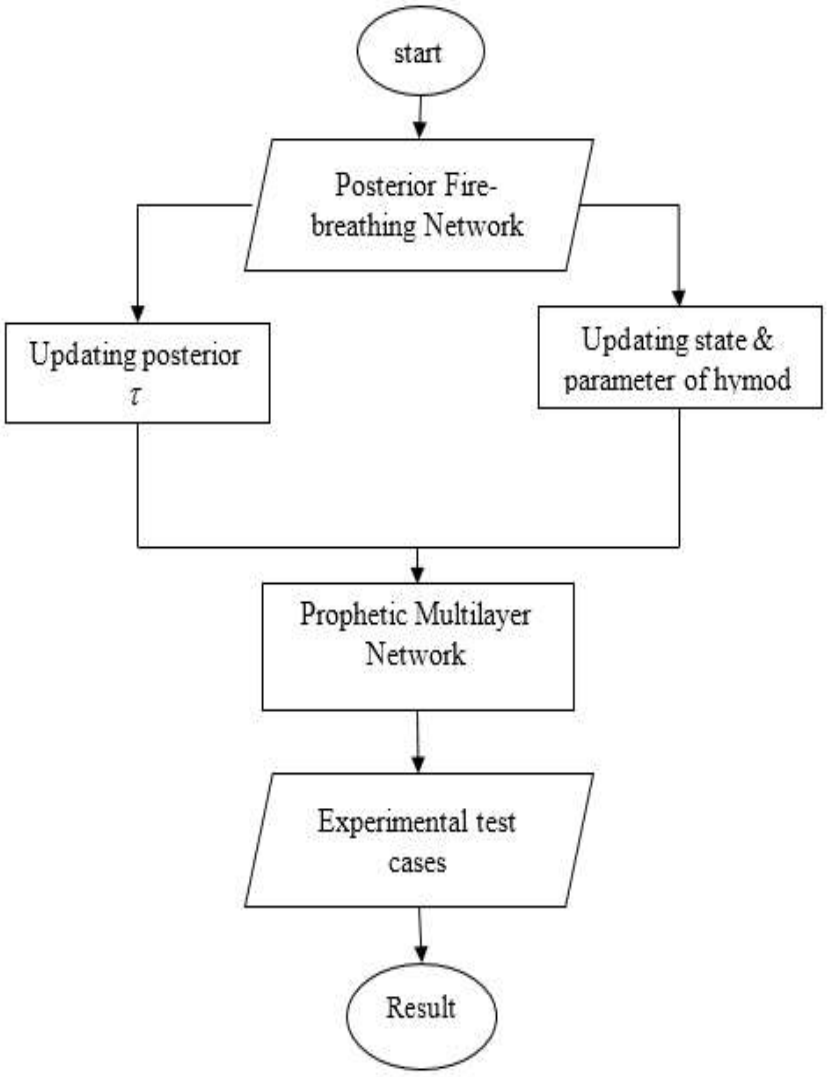

Figure 3.6: Flow chart of the proposed framework

Thus, a greater prediction of precipitation is given by the proposed rainfall model-based artificial framework. The method analyses the errors of the model and decreases the uncertainty for the fine discharge of rainfall. The infiltration regulation also comes with the runoff level estimation in the soil moisture. Formerly the model analyzes the velocity from the steep slopes to inhibit the surface water. Thus, the model estimates the predicted total precipitation.

3.5 Model descriptions

In this study, there are two Hydrological models namely HEC-HMS and SWAT and two soft computing techniques namely M5P and Random Forest (RF) models were applied to simulate discharge in the M.H. Halli Station.

3.5.1 Hydrological Model

3.5.1.1 HEC-HMS Model

HEC-HMS is a hydrologic model package developed by the United State Army Corps of Engineers-Hydrologic Engineering Centre (HEC). It is a semi-physically based and conceptual semi-distributed model designed to simulate continuous and event-based rainfall-runoff processes in a wide spatial scale range, from large river basin flood hydrology to small urban and natural catchment runoff. The software package includes runoff transform, losses, channel routing, base flow, canopy, surface, rainfall-runoff simulation and parameter estimation. HEC-HMS hydrological model uses different packages to represent each component of the river runoff process, including models that compute runoff volume, models of base flow, and models of direct runoff. Each model run combines a meteorological model, basin model and control specifications with run options to obtain results (Choudhari et al., 2014).

The basin model, which describes the different elements of the hydrologic system. It consists of different methods like infiltration loss and transforms method. The infiltration method is used to compute runoff volume and estimates losses resulting from infiltration and evapotranspiration during rainfall event. In this study the soil moisture accounting method has been used as the loss method which simulates the water movement and its storage on

plants, soil surface, and soil profile and groundwater layers. The transform method is used to transform the excess precipitation at the watershed into a hydrograph at the outlet. In this study the transformation of precipitation into surface runoff was accomplished by SCS unit hydrograph. The Soil Conservation Service (SCS) unit hydrograph method defines a curvilinear unit hydrograph by first setting the percentage of the unit runoff that occurs before the peak flow (NRCS, 2007). A triangular unit hydrograph can then be fit to the curvilinear unit hydrograph so that the total time base of the unit hydrograph can be calculated. The SCS unit hydrograph method requires only one parameter for each sub-basin: lag time between rainfall and runoff in the sub-basin. The program computes Tc (time of concentration) and Qp (peak flow) to rescale the SCS-CN dimensionless unit hydrograph. This is then used to compute the direct runoff hydrograph for the sub-basin.

3.5.1.2 Soil and Water Assessment Tool (SWAT)

SWAT is a semi-distributed hydrologic model that can simulate surface runoff, evapotranspiration, sediment yields, the dissemination of nutrients and pesticides, plant growth, groundwater movement, and estimate reservoir storage over the long term (Grusson et al., 2015). In this study ArcSWAT 2012 (Arnold et al., 2012) has been used for runoff simulation. After a good model design, the model can effectively mimic vast and complicated watersheds with various management techniques and best management practices. A proper setup includes the high quality of the input data and rearranging them in SWAT model in a proper manner. So it is important that while creating the sub-basins in the SWAT model the study area should thoroughly be investigated i.e. in a physical sense the distribution of the drainage network and shape of the basin upto its outlet along with the land use pattern and soil type.

While generating the sub-basins in the SWAT model a care should be taken that the number of the generated sub-basins neither be very much nor very less because it affects the runoff generation and separation process as well as the simulation of computation time and also creates the difficulties while writing the output results. While creating the HRU's it is most important to understand the HRU Thresholds and HRU's definition first. The SWAT model facilitates the four HRU definitions and two threshold criteria, in this way as per the

thresholds selected i.e., Percentage or Area, the HRU's definition can be chosen depending upon the study area, therefore a proper investigation of the study area is essential prior to selecting the HRU's definition. In the SWAT model, the HRU's definition may be Dominant of Land use, Soils and slopes, Dominant of HRU's, Target no. of HRU's and Multiple HRU's. In this study we have used Multiple HRU's and percentage of Thresholds. According to the standards outlined by Her et al., (2015) the values for land-use % over sub-basin area, soil class percentage over land use area, and slope class percentage over soil area have been determined. For the SWAT model setup of this study area, we have taken Land use and Soil data from https://swat.tamu.edu/data/india-dataset/. Since this data set is easily available and ready to use otherwise finding out the HSG (Hydrologic Soil Group) is a tedious job which needs an in-depth field investigation of the soil type and multiple infiltration test for each and every soil type found in the study area. The Digital Elevation Model (DEM) of 30m resolution from SRTM has been used for preparing the SWAT model setup. The Hargreaves method has been used as the PET method since this method requires only the temperature as an input for estimating the PET. For reaches channel routing, the Muskingum method has been selected with its default parameters value.

The calibration of model parameters for any hydrological model is an essential part only if the first simulation results are unsatisfactory or expecting more improvements. In this study, we found that the results were satisfactory which are discussed in the results part. Such type of performance from the model can only be obtained if the model setup has been done in a proper manner as we have explained earlier. Therefore, it was decided that for this setup of the SWAT model and for this study area no calibration of the parameters is required. Which shows that why the understanding of the physical meaning of all the input required to the model are necessary while setting up the model.

3.5.2 Soft Computing Technique

3.5.2.1 Random Forest

RF was first developed by Breiman (2001), who used a decision tree classifier to generate the model (Breiman and Cutler, 2004). A method for regression and classification called RF is a grouping of different learning algorithms that can be used in conjunction. The random

forest methodology relies on the creation of a forest of random trees, which are produced by randomly dispersing the information at each decision tree node. In an effort to create a more accurate model, RF uses an ensemble approach to aggregate the resilience of many trees (Goeschel, 2016).

Several studies (Rudianskait et al., 2015; Yaseen and colleagues 2019) have examined the use of RF in engineering and proved its feasibility for prediction methodologies. The "out-of-bag" data that is not picked for the bootstrapping procedure is referred to as "out-of-the-bag" data. The huge number of trees in the RF technique ensures that over-fitting does not occur, and random variables of the suitable type are chosen to ensure precise categorization. The number of trees, the minimum gain, and the maximum tree depth are all characteristics that must be optimized in RF.

For each bootstrap sample of $b = 1, 2, 3, \ldots, B$, *an* RF model is fitted to calculate the output of $\hat{f}_{rf}^{B}(x)$ in input x as shown in equation *25* (Hastie et al., 2009).The following formula computes the output of the RF model by the creation of a random forest tree Tb on bootstrapped data:

$$\hat{f}_{rf}^{B}(x) = \frac{1}{B}\sum_{b=1}^{B} T_b(x). \qquad \text{Eq (25)}$$

The fundamental disadvantage of the RF model is that it is a hyper-parameter method.

3.5.2.2 M5P model

Quinlan (1992) introduced the M5 tree as a regression decision tree learner (Quinlan, 1992). Regression analysis functions are assigned to the terminal nodes of this tree approach, and a multivariate linear regression analysis is fitted to each subspace. The M5 tree scheme, rather than discrete groups, deals with continuous class problems and can handle high-dimensional problems. It shows the piecewise information for each linear model that was used to estimate nonlinear relationships in the data set.

Information about the M5 model tree's splitting criteria is gathered using error calculations at each node. To evaluate the error, the standard deviation of the class values that arrive at a node is used. As a consequence of checking each property at that node, it is

determined which attribute for splitting at that node maximizes the projected error reduction. The following formula is used to calculate the standard deviation reduction (SDR) is shown in equation 26.

$$\mathcal{SDR} = \mathcal{sd}(\mathcal{K}) - \sum \frac{|\mathcal{K}_i|}{|\mathcal{K}|} \mathcal{sd}(\mathcal{K}_i) \qquad \text{eq (26)}$$

Where $\mathcal{K}$= set of instances; $\mathcal{K}_i$ = the subset of illustrations that have the i^{th} product of the possible set; and $\mathcal{sd}$ is the standard deviation.

The next chapter provides the Data and Model setup for this study.