LONG TERM HYDROLOGICAL SIMULATION MODELLING BASED ON PHYSICAL CHARACTERISTICS OF WATERSHED

Dilip G. Durbude[1], B. B. Jadia[2], R. S. Sontakke[3]

[1]*Associate Professor, Engineering Faculty, WALMI, Aurangabad, (India)*

[2]*Professor and Head, IWDM Faculty, WALMI, Aurangabad, (India)*

[3]*Joint Director and S. E., WALMI, Aurangabad, (India)*

ABSTRACT

The physical characteristics of watershed play a vital role in generating runoff and significantly affect the hydrological behavior of the watershed. The future performance of this hydrological behavior of watershed can be predicted by conceptually modellin. But, in the conceptual rainfall-runoff modelling, finding the value of conceptual model parameters is a challenging task particularly in ungauged basins or basins where very less measurements are available. Hence, it has been the endeavor of many hydrologists to quantify and relate geomorphological characteristics of ungauged watersheds to their hydrologic response characteristics. Very recently developed Jain et al.(2012) Modified Long Term Hydrologic Simulation Advance Soil Moisture Accounting (MLTHS ASMA) 15-parameters model performed better than the existing LTHS models to simulate total stream flow, but for its pragmatic application, it is required to correlate model parameters with some measurable physical characteristics of the watersheds. Therefore, in the present study, the measurable physical characteristics of the seventeen watersheds lying in various agro-climatic zones of India are correlated with model parameters using step-wise backward elimination procedure via p-value of F-statistic of multiple regression analysis. In the majority cases of the watershed under study, the model parameters exhibited a significant relationship with physical characteristics of the watersheds.

***Keywords: Advance Soil Moisture Accounting, Geomorphological Characteristics, Hydrologic Modeling, Runoff, Watershed**.*

I. INTRODUCTION

The conceptual rainfall-runoff models generally involves certain parameters relating to watershed characteristics such as size, shape, orientation, topography, geology, geomorphology, land use and soil characteristics etc. play very important role in generating runoff and affect significantly the hydrological response of a watershed. Finding the value of these parameters is a challenging task particularly in ungauged basins or basins where very less measurements are available. Hence, it has been the endeavor of many hydrologists to quantify and relate geomorphological parameters of these watersheds to their hydrologic response characteristics (Chandra, 1993). As such, Horton (1945) pioneered the hydro-geomorphologic analysis of watershed and provided a rational and systematic base, rather a framework of outlines of geomorphological characteristics to relate them to various hydrological properties of the watershed. Strahler's (1952) modification of this technique has generally been adopted for use in hydrologic study. Potter (1953) and Benson (1962) related peak discharge to watershed area, a topographical factor, and a rainfall frequency factor. Boyd (1978) developed a conceptual model using watershed geomorphological properties. Using a probabilistic framework, Rodrguez-Iturbe and Valdes (1979) and Gupta et al. (1980) presented a geomorphological instantaneous unit hydrograph with the exponential probability distribution for the time of travel of water drops which is essentially equivalent to using a linear reservoir. Rosso (1984) derived the Nash IUH parameters as functions of Horton's ratios. Hydro-geomorphological analysis was carried out in a number of Indian watersheds to compute the runoff in water resources development and management projects (Roohani and Gupta, 1988; Karnieli et al. (1994); Chalam et al., 1996; Chaudhary and Sharma, 1998; Hsieh and Wang (1999); Kumar et al., 2001; Ali and Singh, 2002; Durbude and Kumar, 2002; Durbude (2004); Singh et al., 2003; Suresh et al., 2004; Durbude, 2005; Durbude and Chandramohan, 2007; Dabral and Pandey, 2007; etc.).

A general fact is that the model containing too many parameters for simulation of limited components of hydrological processes exhibit difficulty in field applications. Usability of a model can be enhanced if its parameters can be related to measurable catchment characteristics. The physical characteristics of the watersheds which are measurable entities and influencing the runoff characteristics of watershed can be correlated with the model parameters by means of some techniques such as regression analysis. In regression analysis, investigations are made to relate dependent variable (Y) for example model parameters to independent predictors (Xs) such as physical characteristics of watershed. It can be used for modeling causal relationships between model parameters and physical characteristics of watershed.

Therefore, in the present study, parameters of the recently developed long term hydrological simulation Jain el al. (2012) MLTHS ASMA model is correlated with the measurable physical characteristics of the watershed by using the multiple regression technique.

II. MATERIAL AND METHODS

2.1 Existing MLTHS ASMA Model

The existing MLTHS ASMA model proposed by Jain et al. (2012) is primarily based on the physical concepts that describe water movement through a watershed; the total runoff of the catchment is quantified by incorporating sub-modules for direct surface runoff, lateral flow, and base flow. The accounting for soil moisture and ground water store is considered on daily basis. The initial soil moisture store level is used to calculate the space available for water retention, which is updated on daily basis using evapotranspiration, drainage from soil moisture store and level of soil moisture. Direct surface runoff is computed using modified formulation of the Soil Conservation Service Curve Number (SCS-CN) method given by Michel et al. (2005) and Durbude et al. (2011). The sub-surface drainage flow is modeled using the formulations based on concepts given by Yuan et al. (2001) for computation of the sub-surface drainage flow (Jain et al., 2012). Jain et al. (2012) MLTHS ASMA model uses the ASMA procedure both for surface and sub-surface flow components by formulating sub-surface drainage flow component based on the modification in SCS-CN method through theoretical analogy. As this model operates on daily time step, it requires daily rainfall as input and the observed runoff is used only to calibrate parameters of the model and its validation.

The mathematical formulations for computation of the surface flow and sub-surface flow components and losses (such as evapotranspiration and deep percolation) involved in the MLTHS ASMA (Jain et al., 2012) model are again reproduced here, as follows.

2.1.1 Surface Flow Components

The surface flow, which occurs only when the rainfall rate is greater than the rate of infiltration. It is modeled using the ASMA procedure proposed by Durbude et al. (2011), in which soil moisture store at time t is computed by using equation 1 as (Michel et al., 2005);

$$V_t = V_{0(t)} + P_t - RO_t \tag{1}$$

where $V_{0(t)}$ is initial soil moisture store level (mm) at time 't', P_t is accumulated rainfall at time 't' along a storm (mm), RO_t is direct runoff at time 't' along a storm (mm), and V_t is soil moisture store level at time 't', i.e. when the accumulated rainfall is equal to P_t (mm).

The direct surface runoff (RO), a component of surface flow can be computed based on AM (Durbude et al., 2011) as;

If $V_{0(t)} \le S_{a(t)} - P_t$**, then** $RO_t = 0$ **(2)**

If $S_{a(t)} - P_t < V_{0(t)} < S_{a(t)}$**, then**

$$RO_t = \frac{(P_t + V_{0(t)} - S_{a(t)})^2}{P_t + V_{0(t)} - S_{a(t)} + SM_t} \tag{3}$$

If $S_{a(t)} \le V_{0(t)} \le S_{a(t)} + SM_t$**, then**

$$RO_t = P_t\left[1 - \frac{(SM_t + S_{a(t)} - V_{0(t)})^2}{SM_t^{\,2} + (SM_t + S_{a(t)} - V_{0(t)})P_t}\right] \tag{4}$$

In this model, it was assumed that the current space available for water retention (S_t) is constant for the first 5 days of simulation. The value of S_t for the first day of simulation (S_t=S_0) is computed from Eq. (5) by using the initial value of CN (i.e. CN_0 to be determine by optimization). Afterwards, S_t is modified based on antecedent moisture (AM) into modified water retention (SM_t) to avoid the sudden variation in the daily curve number that may affect the performance of model (Geetha et al., 2007; Durbude et al., 2011) as follows;

$$SM_t = \frac{(S_t)^2}{(AM_t + S_t)} \tag{5}$$

Where AM can be computed using the following expression.

$$AM_t = \delta\sqrt{P_{5(t)}} \tag{6}$$

Here $P_{5(t)}$ is the 5 days antecedent rainfall at time 't' and δ is the coefficient of antecedent moisture to be determined by optimization.

The next parameter S_a (Eqs.7-9) is set as a fraction of S (as per the Michel et al., 2005), as follows;

$$S_{a(t)} = \alpha S_t \tag{7}$$

Where α is a parameter (fraction) of S_a, which is treated as calibration parameter and obtained through optimization.

Similarly, the initial soil moisture store level (V_0) can be computed as follows (Durbude et al., 2011);

If $V_{00(t)} \le S_{a(t)} - P_{5(t)}$**, then**

$$\mathbf{V_{0(t)} = V_{00(t)} + \beta\,(P_5)_t} \tag{8}$$

If $S_{a(t)} - \mathbf{P_{5(t)}} < \mathbf{V_{00(t)}} < S_{a(t)}$ **, then**

$$\mathbf{V_{0(t)} = V_{00(t)} + \beta} \left[P_{5(t)} - \frac{(P_{5(t)} + V_{00(t)} - S_{a(t)})^2}{P_{5(t)} + V_{00(t)} - S_{a(t)} + SM_t} \right] \tag{9}$$

If $S_{a(t)} \leq V_{00(t)} \leq S_{a(t)} + SM_t$ **, then**

$$\mathbf{V_{0(t)} = V_{00(t)} + \beta P_{5(t)}} \left[\frac{(SM_t + S_{a(t)} - V_{00(t)})^2}{SM_t^{\,2} + (SM_t + S_{a(t)} - V_{00(t)}) P_{5(t)}} \right] \tag{10}$$

Where β is a model parameter obtained through optimization, and V_{00} is pre-antecedent moisture level defied as (Durbude et al., 2011)

$$\mathbf{V_{00(t)} = \gamma\, SM_t} \tag{11}$$

Here γ ranges from 0.0 to 1.0 and can be obtained by optimization.

2.1.2 Routing of Direct Runoff

The direct surface runoff RO_t (Eqs. 7-9) is routed using a single linear reservoir to produce the surface runoff (SRO_t) at the outlet of the basin after the number of days exceeds 5 (Nash, 1957) to account for catchment induced storage effects as follows;

$$\mathbf{SRO_t = C_0 \,.\, RO_t + C_1 \,.\, RO_{(t-1)} + C_2 \,.\, SRO_{(t-1)}} \tag{12}$$

Where

$$C_0 = \frac{(1/K)}{2 + (1/K)} \tag{13}$$

$$\mathbf{C_1 = C_0} \tag{14}$$

$$C_2 = \frac{2 - (1/K)}{2 + (1/K)} \tag{15}$$

K is the storage coefficient determined from optimization. In linear reservoir routing, the amount of attenuation is a function of Δt/K. Values of Δt/K greater than 2 can lead to negative attenuation, therefore due care was taken during optimization to restrict value of Δt/K to not to exceed 2.

2.1.3 Evapotranspiration

Evapotranspiration (ET) is the amount of water that goes back or lost to the atmosphere. It is the combination of evaporation from the soil surface and transpiration from the vegetation and can be obtained by the summation of daily evaporation from the water bodies and transpiration from the soil zone in the watershed. Since, in the evapotranspiration process, the transpiration process is more dominant than evaporation process, hence, the evapotranspiration is assumed equivalent with transpiration and expressed as follows:

$$\mathbf{ET_t = P_1 \,.\, (\theta_t - \theta_w)} \tag{16}$$

Where P_1=coefficient of transpiration from soil zone, θ_t = soil moisture at time‘t’, θ_w = wilting point of the soil.

2.1.4 Sub-Surface Flow Components

Sub-surface flow occurs beneath the ground surface, when infiltrated rainfall meets an underground zone of low transmission and travels above the zone to the soil surface downhill, and appears as a seep or spring. In this model, the expression for sub-surface drainage flow (DR_t) at time't' is derived based on theoretical analogy (Yuan et al., 2001) as follows;

$$DR_t = \frac{(P_t - (S_{a(t)} - V_{0(t)}) - SRO_t - I_{d(t)})^2}{(P_t - (S_{a(t)} - V_{0(t)}) - SRO_t - I_{d(t)} + S_{d(t)})} \quad \textbf{(17)}$$

The Eq. (22) is valid for ($P_t - S_{a(t)} + V_{0(t)} - SRO_t) \geq I_{d(t)}$; $DR_t = 0$, otherwise. Here,

$$\mathbf{I_{d(t)} = \lambda_d \, . \, S_{d(t)}} \quad \textbf{(18}$$

$$\mathbf{S_{d(t)}} = \frac{25400}{CN_{d(t)}} - 254 \quad \textbf{(19)}$$

I_d is initial abstraction in saturated zone at time't'; λ_d is coefficient of initial abstraction in saturated zone (I_d); $S_{d(t)}$ is potential maximum retention in saturated zone at time 't'; and $CN_{d(t)}$ is curve number for sub-surface (drainage) flow at time 't'.

The sub-surface drainage flow is further partitioned into two components: (i) sub-surface drainage flow in lateral direction as lateral flow and (ii) sub-surface drainage flow in vertical direction as percolation into ground water zone.

2.1.4.1 Lateral Flow

Fraction of sub-surface drainage flow moving in lateral direction (Putty and Prasad, 1994, 2000) is given as:

$$\mathbf{THR_t = P_3 \, . \, DR_t} \quad \textbf{(20)}$$

Where THR_t is lateral flow at time 't' and P_3 is unsaturated soil zone runoff coefficient.

The remaining portion of sub-surface drainage flow moving in the vertical direction to meets the ground water store (GWS) due to the permeability of the soil is considered as percolation and modeled as follows (Mishra et al., 2005, Putty and Prasad, 1994, 2000):

$$\mathbf{PR_t = (1 - P_3) \, . \, DR_t} \quad \textbf{(21)}$$

Where PR_t = percolation at time 't'.

The saturated store GWS is considered as a non-linear reservoir from which the outflow occurs at an exponential rate in the form of deep seepage as follows:

$$\mathbf{DSP_t = (\psi_t - \psi_f)^{Eg}} \quad \textbf{(22)}$$

Where DSP_t = deep seepage at time 't', ψ_t = ground water at time 't', ψ_f = field capacity of the soil in the ground water zone and E_g = exponent of ground water zone

The deep seepage which travels in lateral as well as vertical direction through GWS is further bifurcated into active ground water flow (base flow) and inactive ground water flow (deep percolation) into the aquifers.

2.1.4.2 Base Flow

The base flow (BF_t) or delayed flow, which is an active ground water flow, is modeled as outflow from a non-linear storage as follows:

$$\mathbf{BF_t = P_4 . \, DSP_t} \quad \textbf{(23)}$$

Where P_4=ground water zone runoff coefficient.

The inactive sub-surface flow into aquifers is termed as deep percolation and occurs from the saturated ground water zone in vertical direction, and is considered as a loss from the saturated store which is modeled as:

$$\mathbf{DPR_t = (1 - P_4) \, . \, DSP_t} \quad \textbf{(24)}$$

Where DPR_t = deep percolation at any time 't' and P_4 = ground water zone runoff coefficient. Here, it is worth emphasizing that the proposed model considers deep seepage which is partitioned into two components, base flow and deep percolation.

2.1.5 Total Stream Flow

The total stream flow (TRO_t) at time't', is the sum of the surface runoff, lateral flow, and base flow (Eqs. 4, 17, 25, and 28).

$TRO_t = RO_t + THR_t + BF_t$ if $t \leq 5$ days (25)

$TRO_t = SRO_t + THR_t + BF_t$ if $t > 5$ days (26)

The daily water balance can be maintained by daily water retention storage or soil moisture budgeting from both the SMS and GWS by defining the lower and upper limits of wilting point and field capacity of the soil**.** The current space available for retention of water S_t in unsaturated zone and $S_{d(t)}$ in saturated zone is upgraded on daily basis by taking into account the changes in SMS and GWS as:

$$S_{t+1} = S_t + \theta_t - \theta_{(t+1)} + \psi_t - \psi_{(t+1)} \quad \textbf{(27)}$$

$$S_{d(t+1)} = S_{d(t)} + \theta_t - \theta_{(t+1)} + \psi_t - \psi_{(t+1)} \quad \textbf{(28)}$$

were, $S_{(t+1)}$ is the next day's potential maximum retention (mm); $S_{d(t+1)}$ is the next day's potential maximum retention (mm) in saturated zone; $\theta_{(t+1)}$ is the next day's soil moisture (mm); $\psi_{(t+1)}$ is the next day's ground water (mm). The soil moisture store (SMS) and ground water store (GWS) are upgraded on daily basis as:

$$\theta_{(t+1)} = \theta_t + V_t - V_{0(t)} - ET_t - DR_t \quad \textbf{(29)}$$

$$\psi_{(t+1)} = \psi_t + PR_t - BF_{(t)} - DPR_t \quad \textbf{(30)}$$

2.2 Multiple Regression Technique

In the multiple regression, the multiple correlation coefficient (R) is Pearson's product moment correlation between the predicted values (Y') and the observed values (Y). Just as coefficient of determination (r^2) is the proportion of the total variance (s^2) of Y that can be explained by the linear regression of Y on X, R^2 is the proportion of the variance explained by multiple regressions. The significance of R can be tested by the F-statistic of the analysis of variance for the regression. The basic idea is to select the most significant regression equation, which corresponds to the minimum p-value of F-test. In regression analysis, selecting variables is very important. As a matter of fact, the first problem that has to be solved in practice is to determine which variables should be included in the model. Obviously, the goodness of regression model depends on the selection of variables. How to select variables that can yield the best regression equation? Aitkin (1974) defined a class of "adequate" regression equations, characterized by a lower bound on the multiple correlation coefficients. Here "adequate" means that each member of the class is not significantly poorer than the complete equation. As Aitkin pointed out, this does not solve the problem of finding the "best" equation for prediction. Besides, Spjotvoll (1972) constructed a multiple comparison method, which usually gives a set of many equations none of which is significantly better than any other. Many criteria have been presented by Draper and Smith (1981) for the selection of the "best" regression equation, but none of these has been considered as the best one. Using different criteria, one gets different (the "best") regression equations. Among these criteria are residual mean square (s^2), adjusted multiple correlation coefficient (R), Cp-statistic (Mallows, 1964) and so on. In order to develop a good model based on these criteria, it is necessary to select the best subset.

2.2.1 Best Subset Selection

A problem arises frequently in multiple regression analysis how to predict the value of a dependent variable when there are a number of variables available to choose as independent variables. Though the high speed of modern algorithms is available to perform the multiple linear regression calculations, it is tempting to select a subset instead of just using all the variables in the model. It is always better to make predictions with models that do not include irrelevant variables. Dropping independent variables that have small (non-zero) coefficients will improve the predictions as it will reduce the mean square error (MSE). Hence, there is a need for selecting subset of the independent parameters to correlate with the dependant parameters. There are several methods for selecting a subset of predictors that produce the "best" regression. Many statisticians discourage general use of these methods because they can detract from the real-world importance of predictors in a model. Examples of predictor selection methods are step-up selection, step-down selection, stepwise regression, and best subset selection. The fact that there is no predominant method indicates that none of them are broadly satisfactory (Draper and Smith, 1998).

2.2.2 Algorithms for Subset Selection

Selecting subsets to improve MSE is a difficult computational problem for large number of independent variables. The most common procedure for more than 20 independent variables is to use heuristics to select "good" subsets rather than to look for the best subset for a given criterion. The heuristics most often used and available in statistics software are step-wise procedures. There are three common procedures: forward selection, backward elimination, and step-wise regression (Draper and Smith, 1998). In forward selection procedure, the variables are kept on adding one at a time to construct what we hope is a reasonably good subset (Draper and Smith, 1998). Starting with constant term only in subset, compute the reduction in the sum of squares of the residuals (SSR) obtained by including each variable that is not presently in *S*. For the variable, say, *i* that give the largest reduction in SSR compute as:

$$F_i = Max_{i \notin S} \frac{SSR(S) - SSR(S \cup \{i\})}{\sigma^2 (S \cup \{i\})} \quad \textbf{(31)}$$

If $Fi > Fin$, where Fin is a threshold (typically between 2 and 4) add i to S. Repeat until no variables can be added.

The backward elimination started with all variables in *S*. Compute the increase in the sum of squares of the residuals (SSR) obtained by excluding each variable that is presently in S (Draper and Smith, 1998). For the variable, say, i that give the smallest increase in SSR compute as:

$$F_i = Min_{i \notin S} \frac{\sigma^2 (S)}{SSR(S - \{i\}) - SSR(S)} \quad \textbf{(32)}$$

If $Fi < Fout$, where $Fout$ is a threshold (typically between 2 and 4) then drop i from S. Repeat until no variable can be dropped. Backward Elimination has the advantage that all variables are included in *S* at some stage. This addresses a problem of forward selection that will never select a variable that is better than a previously selected variable that is strongly correlated with it. The disadvantage is that the full model with all variables is required at the start and this can be time-consuming and numerically unstable.

The step-wise regression procedure is like forward selection except that at each step we consider dropping variables as in backward elimination. Convergence is guaranteed if the thresholds Fout and Fin satisfy: Fout < Fin. It is possible, however, for a variable to enter *S* and then leave S at a subsequent step and even rejoin S at a yet later step. As stated above these methods pick one best subset. There are straightforward variations of the methods that do identify several close to best choices for different sizes of independent variable subsets.

None of the above methods guarantees that they yield the best subset for any criterion such as adjusted R^2. These are reasonable methods for situations with large number of independent variables. Hence, in the present study, stepwise multiple regressions with p-value of F-statistic were followed to select the best subset of various combinations of measurable characteristics of study watersheds by using EXCEL 2007: Multiple regression and statistical software, namely SYSTAT.

2.2.3 Stepwise Multiple Regression

Stepwise regression was introduced by Efroymson (1960). This method is an automated procedure used to select the most statistically significant variables from a large pool of explanatory variables. The method does not take into account industrial knowledge about the process, and therefore, other variables of interest may be later added to the model, if necessary. If properly used, the stepwise regression option in EXCEL 2007 and SYSTAT (or other stat packages) puts more power and information than does the ordinary multiple regression option, and it is especially useful for shifting through large number of potential independent variables and/or fine-tuning a model by poking variables in and/or out. If improperly used, it may converge on a poor model while giving a false sense of security. The stepwise regression option either begins with no variables in the model or proceeds forward (adding one variable at a time) or start with all potential variables in the model and proceed backward (removing one variable at a time). At each step, the SYSTAT program performs various calculations via for each variable currently in the model, it computes the t-statistic for its estimated coefficient, squares it, and reports this as its "F-to-remove" statistic; for each variable not in the model, it computes the t-statistic that its coefficient would have if it were the next variable added, squares it, and reports this as its "F-to-enter" statistic. At the next step, the program automatically enters the variable with the highest F-to-enter statistic or removes the variable with the lowest F-to-remove statistic in accordance with certain control parameters that have been specified. Under the forward method, at each step, it enters the variable with the largest F-to-enter statistic, provided that this is greater than the threshold value for F-to-enter. When there are no variables left to enter whose F-to-enter statistics are above the threshold, it checks to see whether the F-to-remove statistics of any variables added previously have fallen below the F-to-remove threshold. If so, it removes the worst of them, and then tries to continue. It finally stops when no variables either in or out of the model have F-statistics on the wrong side of their respective thresholds. The backward method is similar in spirit, except it starts with all variables in the model and successively removes the variable with the smallest F-to-remove statistic, provided that this is less than the threshold value for F-to-remove. Whenever a variable is entered, its new F-to-remove statistic is initially the same as its old F-to-enter statistic, but the F-to-enter and F-to-remove statistics of the other variables will generally all change. Similarly, when a variable is removed, its new F-to-enter statistic is initially the same as its old F-to-remove statistic. Until the F-to-enter and F-to-remove statistics of the other variables are recomputed, it is impossible to tell what the next variable to enter or remove will be. Hence, this process is myopic, looking only one step forward or backward at any point. There is no guarantee that the best model that can be constructed from the available variables (or even a good model) will be found by this one-step-ahead search procedure. Hence, when the procedure terminates, one should study the sequence of variables added and deleted, think about whether the variables that were included or excluded make sense. For example, the variable with the lowest F-to-remove or highest F-to-enter may have just missed the threshold value, in which case one may wish to tweak the F-values and see what happens. Sometimes adding a variable with a marginal F-to-enter statistic, or removing one with a marginal F-to-remove statistic, can cause the F-to-enter statistics of other variables not in the model to go up and/or the F-to-remove statistics of other variables in the model to go down, triggering a new chain of entries or removals leading to a very different model. The selection of stepwise forward or backward multiple regression method depends on the set of independent variables. If a very large set of potential independent variables is available from which one has to extract a few, i.e. one is on fishing expedition, one should generally go forward. On the other hand, if one has a modest-sized set of potential variables from which one wishes to eliminate a few, i.e. one is fine-tuning some prior selection of variables, one should generally go backward. As noted above, after SYSTAT completes a forward run based on the F-to-enter threshold, it takes a backward look based on the F-to-remove threshold, and vice versa. Hence, both thresholds come into play regardless of which method are using, and the F-to-enter threshold must be greater than or equal to the F-to-remove threshold (to prevent cycling). Usually the two thresholds are set to the same value. Keeping in mind that the *F*-statistics are squares of corresponding t-statistics, an F-statistic equal to 4 would correspond to a *t*-statistic equal to 2, which is the usual rule-of-thumb value for "significance at the 5% level." (4 is the default value for both thresholds.). It is always better using a somewhat smaller threshold value than 4 for the automatic phase of the search-for example 3.5 or 3. Since the automatic stepwise algorithm is myopic, it is usually OK to let it enter a few too many variables in the model, and then one can weed out the marginal ones later on by hand. However, beware of using too low an F-threshold if the number of variables is large compared to the number of observations or if there is a problem with multicollinearity in data.

III. RESULTS AND DISCUSSION

The optimized values of parameters of MLTHS ASMA (Jain et al., 2012) model applied to the daily data of rainfall and stream flow of the 17 watersheds varying in size/shape, physical properties and situated in different agro-climatic zones of India are presented in Table 1.

Table 1. Optimized Values of Parameters of MLTHS ASMA (Jain et al. 2012) Model

Sr. No.	Name of Watershed	Model Parameters														
		CN_0	δ	α	β	γ	K	P_1	P_3	P_4	θ_w	ψ_f	E_s	ψ_0	CNd_0	λ_d
1.	Hemavati	30.5	2.33	0.41	0.14	0.44	1.52	0.01	0.11	0.99	40.3	300.0	0.32	300.3	64.7	0.01
2.	Hridaynagar	27.5	0.01	0.54	0.01	0.52	1.23	0.04	0.00	0.59	299.7	496.2	0.26	253.6	60.7	0.01
3.	Mohegaon	37.0	0.05	0.46	0.01	0.50	5.00	0.03	0.60	0.16	119.7	395.3	0.32	469.4	58.3	0.07
4.	Manot	35.3	6.63	0.85	0.00	0.95	3.02	0.02	0.60	0.48	298.9	413.8	0.36	377.1	56.6	0.57
5.	Amachi	17.4	0.10	0.40	0.02	0.42	2.00	0.01	0.20	0.44	80.0	310.0	0.34	242.9	32.3	0.02
6.	Anthrolli	28.6	4.92	0.43	0.01	0.46	1.69	0.01	0.39	0.10	220.0	259.8	0.54	358.8	78.2	0.13
7.	Attigundi	20.0	0.02	0.45	0.001	0.58	5.98	0.05	0.24	0.98	162.5	373.9	0.38	199.4	44.8	0.26
8.	Barchi	38.1	0.01	0.53	0.01	0.54	2.00	0.01	0.35	0.50	148.9	279.2	0.38	551.5	55.4	0.09
9.	Khanapur	11.3	0.14	0.46	0.02	0.57	3.18	0.01	0.60	0.99	179.9	405.0	0.50	198.3	18.3	0.15
10.	Hirehalla	14.6	0.06	0.48	0.10	0.50	2.07	0.09	0.80	0.06	130.3	300.0	0.49	350.0	35.8	0.03
11.	Sagar	21.4	0.21	0.50	0.01	0.54	0.64	0.01	0.59	0.54	45.5	400.0	0.90	191.0	50.8	0.09
12.	Sorab	52.8	0.13	0.53	0.01	0.61	1.63	0.01	0.50	0.84	105.0	302.3	0.40	397.0	51.9	0.05
13.	Dasanakatte	15.6	0.11	0.59	0.24	0.61	0.76	0.00	0.40	0.99	123.6	399.9	0.38	394.4	20.4	0.02
14.	Haladi	25.6	0.02	0.56	0.23	0.60	1.66	0.01	0.47	0.99	103.8	400.0	0.35	389.1	37.1	0.05
15.	Jadkal	30.8	0.98	0.47	0.34	0.53	0.53	0.05	0.48	0.99	160.0	400.0	0.39	405.7	42.2	0.06
16.	Kokkarne	15.8	4.99	0.57	0.18	0.56	2.06	0.00	0.55	0.99	127.4	407.9	0.30	220.2	25.9	0.01
17.	Halkal	26.5	0.01	0.65	0.30	0.66	1.34	0.01	0.47	0.99	240.0	262.9	0.38	303.4	33.5	0.01

As stated earlier, the geomorphological characteristics of the watersheds that significantly affect the runoff are geographical area, length, slope, shape, land use, and soil characteristics of watershed. The various geomorphological characteristics of study watersheds were extracted from the Survey of India (SOI) toposheets. The toposheets were scanned and projected into Universal Traverse Mercator (UTM) projection system into zone (43-in which the study area lies), using Everest (India, 1956) Ellipsoid and Everest (India, Nepal) datum using image processing utilities of Integrated Land and Water Information System (ILWIS) software. The rectified toposheets were further used for the delineation of different features in the study watersheds like contour lines and drainage networks etc. The base map of the watershed boundary at 1:50,000 scale was prepared using the location of various contour and drainage lines. Different thematic maps, viz., contour map and drainage map were prepared using the base map. Digital Elevation Models (DEMs) were created using the contour maps, which were further used for the assessment of relief aspects. The mathematical expressions given by various researchers such as Horton (1945), Miller (1953), and Schum (1956) were used to compute geomorphological characteristics of watershed as shown in Table 2.

Table 2. Geomorphological Characteristics of Study Watersheds

Sr. No.	Name of Watershed	Geomorphological characteristics of the watershed								
		Area (A)	Perimeter (P)	Length (Lb)	Hydrologic length (Lm)	Form factor (Rf)	Circulatory ratio (Rc)	Elongation ratio (Re)	Total relief (H)	Vegetation (V)
		(Km^2)	(Km)	(Km)	(Km)				(m)	(%)
1	Hirehalla	1296.00	162.23	55.49	51.31	0.42	0.62	0.73	150	6
2	Hridaynagar	3370.00	402.42	182.92	215.20	0.10	0.26	0.36	228	65
3	Amachi	87.00	33.74	11.28	11.51	0.68	0.96	0.93	224	70
4	Barchi	4661.00	326.97	148.62	174.85	0.21	0.55	0.52	391	58
5	Mohegaon	14.50	20.81	8.18	8.60	0.22	0.42	0.53	254	94
6	Anthroli	503.00	98.21	35.47	24.23	0.40	0.66	0.71	246	57
7	Manot	5032.00	503.03	228.65	269.00	0.10	0.25	0.35	660	35
8	Sorab	96.00	45.30	15.81	24.61	0.38	0.59	0.70	266	60
9	Khanapur	320.00	143.74	30.83	48.08	0.34	0.19	0.65	146	63
10	Sagar	75.00	33.56	11.84	10.92	0.54	0.84	0.83	108	55
11	Attigundi	4.51	8.81	3.47	2.34	0.37	0.73	0.69	188	85
12	Hemavati	600.00	127.35	57.89	55.13	0.18	0.46	0.48	350	12
13	Kokkarne	343.00	116.94	34.35	53.17	0.29	0.32	0.61	1147	82
14	Halkal	108.00	48.23	18.39	17.64	0.32	0.58	0.64	1101	92
15	Dasanakatte	135.00	57.95	19.92	28.56	0.34	0.50	0.66	869	92
16	Jadkal	90.00	39.45	13.12	18.75	0.52	0.73	0.82	1142	85
17	Haladi	505.00	105.07	34.79	42.75	0.05	0.57	0.73	968	87

The multiple regression analysis was carried out to develop the relation between the MLTHS ASMA model parameters and the geomorphological characteristics of the watershed. For this purpose, the regression matrix was prepared to have an idea about the poorly correlated geomorphological characteristics of study watersheds with model parameters as shown in Table 3. This analysis helps to take decision for carrying out multiple linear regression analysis. The multiple regressions were performed using the Data Analysis Add-in facilities of EXCEL 2007. The regression matrix (Table 3) was used for choosing the best subset of the watershed characteristics to correlate with model parameters. Since EXCEL 2007 has limited facilities and required several trials to select the

Table 3. Regression Matrix between MLTHS ASMA Model Parameters and Geomorphological Characteristics of Study Watersheds

Sr. No.	Model Parameters	Regression coefficient (r^2)								
		A	P	Lb	Lm	Rf	Rc	Re	H	V
1	CNd_0	0.178	0.146	0.196	0.145	0.090	0.001	0.170	0.089	0.175
2	α	0.181	0.261	0.244	0.264	0.307	0.366	0.293	0.218	0.003
3	β	0.160	0.154	0.149	0.138	0.000	0.022	0.057	0.628	0.168
4	γ	0.171	0.221	0.212	0.238	0.300	0.310	0.280	0.131	0.000
5	δ	0.014	0.024	0.025	0.022	0.079	0.027	0.016	0.001	0.024
6	P_3	0.001	0.003	0.014	0.010	0.034	0.004	0.075	0.005	0.012
7	P_4	0.173	0.093	0.101	0.068	0.048	0.071	0.003	0.281	0.191
8	K	0.032	0.079	0.060	0.060	0.124	0.166	0.073	0.127	0.060
9	P_1	0.041	0.029	0.026	0.017	0.011	0.014	0.001	0.043	0.120
10	θ_w	0.262	0.385	0.368	0.370	0.142	0.285	0.291	0.028	0.006
11	ψ_f	0.222	0.329	0.283	0.334	0.221	0.324	0.179	0.085	0.001
12	ψ_0	0.094	0.038	0.050	0.045	0.159	0.032	0.084	0.072	0.000
13	E_g	0.147	0.205	0.190	0.207	0.486	0.483	0.418	0.236	0.000
14	CN_0	0.053	0.011	0.036	0.033	0.062	0.001	0.088	0.010	0.001
15	λ_d	0.172	0.163	0.173	0.173	0.028	0.023	0.079	0.021	0.017

best subset of watersheds characteristics, the multiple regressions using stepwise backward elimination procedure based on p-value of F- statistics (Zhang and Wang, 1997) is performed in SYSTAT 10. Here, the p-value is the probability (prob(F)) of obtaining a test statistic at least an extreme as the one that was actually observed, assuming that the null hypothesis is true. Generally, one rejects the null hypothesis if the p-value is smaller than or equal to the significance level (α). If the level is 0.05, then results that are only 30% likely or less are deemed extraordinary, given that the null hypothesis is true. The calculated p-value exceeds 0.05, so the observation is consistent with the null hypothesis. Likewise, if prob(F)<0.05, then the model is considered significantly better that would be expected by chance and reject the null hypothesis of no linear relationship of model parameters to the measurable physical characteristics of the watersheds. Some of the statisticians also considered the model is highly significant if p-value is less than or equal to 0.001.

The various combination of p-value -to enter and p- value-to remove and/or F-value-to enter and F-value to remove were tried in SYSTAT 10 to choose the correct combination to develop the regression equations. The regression statistics along with analysis of variance (ANOVA) for correct combination of physical characteristics of watersheds to estimate the various model parameters are computed. As seen from the ANOVA, the multiple correlation coefficient (multiple R) for most of the model parameters such as CN_0, α, β, P_3, P_4, θ_w,ψ_f,ψ_0, E_g, and CNd_0, are more than 0.60, which indicates that there exists a good correlation between model parameters and measurable physical characteristics of the watersheds. A very good correlation (multiple R=0.95) is found between initial ground water content (ψ_0) and physical characteristics of the watershed such as A, P, Lb, Lm, Rf, Rc, and Re. From the p-value of F-statistics, it is found that the regression equation developed for β, θ_w, ψ_0, and E_g parameters are highly significant (level of significance or p- value is more than 0.001), while some parameters such as δ, K, P_1, and λ_d are very poorly significant (or insignificant) at 95% confidence interval (level of significance or p-value is less than 0.05). From the ANOVA, the regression equations for various model parameters were formulated as shown in Table 4.

Table 4. Regression Equations Showing Relationship between LTHS ASMA II Model Parameters and Geomorphological Characteristics of Watersheds

Model Parameter	Multi-linear Regression Equation	Multiple (R^2)
CN_0	CN_0=31.991-0.317P+0.702Lb	0.39
α	α=0.64-0.424Rc+0.0001H	0.52
β	β =0.022-0.001L+0.0003H	0.76*
γ	γ =0.751-0.37Rc	0.31
δ	δ =0.023Lb+10.38Rf-2.75	0.24**
P_3	P_3=0.45+0.0003A+0.005P-0.02Lb	0.56
P_4	P_4=1.38-0.0001A-0.018Lb+0.017Lm-1.125Rf+0.0001H	0.74
K	K=1.27-0.011H+0.147V	0.40**
P_1	P_1=0.02+0.001P-0.002Lb-0.0002Lm+0.17Rc-0.18Re	0.24**
θ_w	θ_w=10.968Lb-8.248Lm+275.567Rf-481.685Rc+319.221Re+2.709V-133.414	0.85*
ψ_f	θ_f=180.578+0.835Lm-330.33Rc+444.463Re	0.54
ψ_0	ψ_0=0.26A-15.69P+48.91Lb-17.0Lm-320.07Rf-3144.86Rc+4051.43Re-207.56	0.91*
λ_d	λ_d =0.056+0.001Lm	0.17**
CNd_0	CNd_0=36.79+1.51Lb-1.19Lm	0.52
E_g	E_g=0.242+0.601Rc-0.0001H	0.65*

Note: * highly significant at 95% confidence interval

** not significant at 95% confidence interval

As seen from Table 4, the multi-linear regression equations developed for some of the model parameters of MLTHS ASMA model are highly significant, while there exists a significant relationship for most of the parameters. Hence, the multi-linear regression equations developed for these model parameters (except for those parameters for which regression equations are found insignificant) may be used for parameter estimation using the measurable physical characteristics of watersheds. Thus, many parameters of MLTHS ASMA model could be estimated from catchment characteristics and could potentially be used for field application when sufficient data for better calibration of parameters of the model do not exist.

IV. CONCLUSIONS

The relationship between model parameters and measurable geomorphologocal characteristics of the watersheds using multiple regression analysis is developed in this paper. The step-wise regression with backward elimination on p-value of F-statistics is followed to develop regression equations. In most of the cases, a significant relationship is found between the model parameters and geomorphological characteristics of study watersheds at 95% confidence interval. The multi-linear equations thus developed for various parameters of MLTHS ASMA model can be used to estimate the total runoff from the unguaged watershed in various agro-climtic zones of India.

REFERENCES

[1] M. K. Jain, D. G. Durbude and S. K. Mishra, SCS-CN based modified long term hydrologic simulation model for daily stream flow simulation, Journal of Hydrologic Engineering (ASCE), 17(11), 2012, 1204–1220.

[2] S. Chandra, *Geomorphology of Sabarmati basin up to Dharoi*, Rep. National Institute of Hydrology, Roorkee, TR-138, 1993, 3-5.

[3] R. E. Hortan, Erosional development of streams and their drainage basins-hydrophysical approach to quantitative morphology, Bull. Goel. Soc. of Ame., 56, 1945, 275-370.

[4] A. N. Strahler, Hypsometric (area-altitude) analysis of erosional topography, Bull. Geol. Soc. Ame., 63, 1952, 1117-1142.

[5] W. D. Potter, Rainfall and topographic features that affects runoff." Trans. Ame. Geophys. Un., 34, 1953, 67-73.

[6] M. A. Benson, Flood peaks related to hydrological factors in the south-west, USGS Prof. Paper 450 E, 1962, 161-163

[7] M. J. Boyd, A storage-routing model relating drainage basin hydrology and geomorphology, Water Resour. Res., 14 (15), 1978, 921-928.

[8] I. Rodrguez-Iturbe, and J. B. Valdes, The geomorphologic structure of hydrology response, Water Resour. Res. 15(6), 1979, 1409-1420.

[9] V. K. Gupta, E. Waymire and C. T. Wang, A representation of an instantaneous unit hydrograph from geomorphology, Water Resour. Res., 16 (5), 1980, 855–862.

[10] R. Rosso, Nash model relation to Horton order ratios, Water Resour. Res., 20(7), 1984, 914-920.

[11] M. S. Roohani, and R. P. Gupta, Quantitative hydrogeomorphic investigations in the Chenab catchment, Himalayas, J. Hydro. (IAH), 11(4), 1988, 25-43.

[12] A. M. Karnieli, M. H. Diskin, and L. J. Lane, CELMOD 5-A semi-distributed cell model for conversion of rainfall into runoff in semi-arid watersheds, J. Hydro., 157 (1-4), 1994, 61–85.

[13] B. N. S. Chalam, M. Krishnaveni, and M. Karmegam, Correlation of runoff with geomorphic parameters, J. Appl. Hydro., 9 (3-4), 1996, 24-31.

[14] R. S. Chaudhary and P. D. Sharma, Erosion hazard assessment and treatment prioritization of Giri River catchment, North Western Himalayas, Indian J. Soil Cons., 26(1), 1998, 6-11.

[15] L. S. Hsieh, and R. Y. Wang, A semi-distributed parallel-type linear reservoir rainfall– runoff model and its application in Taiwan, Hydrol. Process., 13(8), 1999, 1247–1268.

[16] R. Kumar, A. K. Lohani, S. Kumar, C. Chatterjee, and R. K. Nema, GIS based morphometric analysis of Ajay river basin up to Sarthgauging site of South Bihar, J. Appl. Hydrol. (AHI), 14(4), 2001, 45-54.

[17] S. Ali, and R. Singh, Morphological and hydrological investigation in Hirakud catchment for watershed management planning, J. Soil Water Cons. (India), 1(4), 2002, 246-256.

[18] Dilip G. Durbude, and C. P. Kumar, Application of GIS for the determination of morphological characteristics of a watershed, Proc. National Conf. on Modern Trend in Water Resources Development and Environmental Management, Vellore, India, Vellore Institute of Technology, Vellore, March 2002, 11-17.

[19] Dilip G. Durbude, Remote sensing application for hydro-geomorphological analysis of a watershed, Proc. National Conf. on Hydraulic. Water Resour. (HDRO 2004), ISH & VNIT, Nagpur, December, VNIT, Nagpur, India, 2004, 42-49.

[20] R. K. Singh, C. M. Bhatt, and V. H. Prasad, Morphological study of a watershed using remote sensing and GIS techniques, Hydro. J. (IAH), 26(1&2), 2003, 55-66.

[21] M. Suresh, S. Sudhakar, K. N. Tiwari and V. M. Chowdary, Prioritization of watersheds using morphometric parameters and assessment of Surface Water Potential using Remote Sensing, J. Indian Soc. Rem. Sens., 32(3), 2004, 249-259.

[22] Dilip G. Durbude, Remote sensing and GIS approach for land use, soil texture and morphological characteristics of a watershed, J. Appl. Hydrol. (AHI), 18(3), 2005, 42-50.

[23] Dilip G. Durbude and T. Chandramohan, Morphometric analysis of a forested watershed using GIS technique, In:Forest Hydroloogy, (Ed.) Venkatesh, B., Purandara, B. K. and Ramasastri, K. S., Capital Publishing, New Delhi, 2007, 130-136.

[24] P. P. Dabral and A. Pandey, Morphometric analysis and prioritization of a eastern Himalayan river basin using satellite data and GIS, Asian J. Geoinformatics, 7(3), 2007, 3-14.

[25] C. Michel, A. Vazken and C. Perrin, Soil conservation service curve number method: how to mend a wrong soil moisture accounting procedure, Water Resour. Res., 41:W02011, 2005, 1–6.

[26] Dilip G. Durbude, M. K. Jain and S. K. Mishra, Long term hydrologic simulation using SCS-CN based improved soil moisture accounting procedure, Hydrol. Process., 125, 2011, 561-579.

[27] Y. Yuan, J. K. Mitchell, M. C. Hirschi and R. A. C. Cooke, Modified SCS curve number method for predicting subsurface drainage flow, Trans. ASAE, 44(6): 2001, 1673-1682.

[28] K. Geetha, S. K. Mishra, T. I. Eldho, A. K. Rastogi and R. P. Pandey, Modification to SCS-CN method for long-term hydrologic simulation, J. Irrig. Drain. Engg., 133(5), 2007, 475-486.

[29] S. K. Mishra, K. Geetha, A. K. Rastogi and R. P. Pandey, Long-term hydrologic simulation using storage and source area concepts, Hydrol. Process., 19, 2005, 2845-2861.

[30] M. R. Y. Putty and R. Prasad, Understanding runoff processes using a watershed model–a case study in the Western Ghats in South India, J. Hydro., 228, 2000, 215-227.

[31] M. R. Y. Putty, and R. Prasad, Stream flow generation in Western Ghats, In: 6th National Sympo., Hydrology, Shillong, 1994.

[32] M. A. Aitkin, A general linear model approach to data analysis, Symposium on Statistical Computing, Sydney, August 1974.

[33] E. Spjotvoll, On the optimality of some multiple comparison procedure, The Annals of Math. Statistics, 43(2), 1972, 398-411.

[34] N. R., Draperand, H. Smith, *Applied regression analysis, 2d Edition, New York: John Wiley & Sons, Inc.*, 1981.

[35] C. L. Mallows, Choosing variables in a linear regression: a graphical aid, Presented at the Central Regional Meeting of the Institute of Mathematical Statistics, Manhattan, Kansa, 1964.

[36] N. R. Draper and H. Smith, *Applied regression analysis (3rd edition), New York: John Wiley and Sons, Inc.*, 1998.

[37] M. A. Efroymson, *Multiple regression analysis, In Ralston, A. and Wilf, HS, editors, Mathematical Methods for Digital Computers. Wiley*, 1960.

[38] Jin, Zhang, and X. Wang, Selecting the best regression equation via the p-value of F-test, Metrika, **46**: 1997, 33-40.

[39] V. C. Miller, A quantitative geomorphic study of drainage basin characteristics in the Clinch mountain area, Virginia and Tennessee, Proj. NR 389-402, Tech Rep 3, Columbia Univ., Deptt. Geol., ONR, New York, 1953.

[40] S. A. Schum, Evaluation of drainage systems and slopes in badlands at Perth Amboy, New Jersey, Geol. Soc. Ame. Bull., 67, 1956, 597-646.