Journal Of Flood Engineering And Science Research

Volume No. 9

Issue No. 2

May - August 2025

ENRICHED PUBLICATIONS PVT. LTD

S-9, IInd FLOOR, MLU POCKET,
MANISH ABHINAV PLAZA-II, ABOVE FEDERAL BANK,
PLOT NO-5, SECTOR-5, DWARKA, NEW DELHI, INDIA-110075,
PHONE: - + (91)-(11)-47026006

Journal Of Flood Engineering And Science Research

Aims and Scope

The Journal of Flood Engineering and Science Research aims at opening an easy access to the valuable source of information and providing an excellent publication channel for the global community of researchers /scholars/phd guides and phd pursuing students and others in the civil engineering subjects and especially geotechnology and its applications.

Typical subjects covered by this journal include:

- Computational and Theoretical Geomechnics
- Foundations
- Tunneling
- Earth Structures
- Site Characterization
- Soil-Structure Interactions

Journal Of Flood Engineering And Science Research

Journal Of Flood Engineering And Science Research

(Volume No. 9, Issue No. 2, May - August 2025)

Contents

Assessment Of Adequate Probability Distribution For Extreme Wind Speed Analysis Under Missing Data Scenario

N. Vivekanandan[1], Dr. C. Ramesh[2], R.S. Jagtap[3]

[1]Assistant Research Officer,

[2]Scientist-C,

[3]Joint Director

Central Water and Power Research Station, Pune 411024

Telephone No.: +91 20 24103367; E-mail: vivek.cwprs@gmail.com

A B S T R A C T

Assessment of extreme wind speed is of utmost importance in estimating the design value of the wind load-effect for any structural design. This can be carried out by Extreme Value of Analysis (EVA) by fitting of probability distributions to the recorded wind speed data. This paper illustrates the adoption of Gumbel (EV1), Frechet (EV2), Log Normal (LN2) and Log Pearson Type-3 distributions in EVA of wind speed for Delhi with two different scenarios viz., data series with imputation (Scenario-1) and data series without imputation (Scenario-2). Method of Moments (MoM) and maximum likelihood method are used for determination of parameters of EV1, EV2, LN2 and LP3 distributions. In addition to above, method of least squares and Order Statistics Approach (OSA) are used for determination of parameters of EV1 and EV2. The adequacy of fitting of probability distributions is evaluated by Goodness-of-Fit tests viz., Anderson-Darling and Kolmogorov-Smirnov and diagnostic test using D-index. The study suggests the EV1 (OSA) is better suited probability distribution for estimation of Extreme Wind Speed (EWS) for Scenario-1 whereas LN2 (MoM) for Scenario-2. By considering the design-life of the structure, the study recommends the estimated EWS using EV1 (OSA) obtained from Scenario-1 for design purposes.

Keywords: Anderson-Darling, D-index, Extreme Value Analysis, Kolmogorov-Smirnov, Probability Distribution, Wind Speed

1. Introduction

In hydrometeorological context, it is well recognized fact that, whatsoever extreme be the design-loading, more severe conditions are likely to occur in nature. Generally, structures are designed to withstand extreme wind loads over the intended economic lifetime. As the maximum wind loads experienced by a structure is an important factor in design, estimation of the occurrence of Extreme Wind Speed (EWS) for a particular return period is carried out by fitting of probability distribution to the recorded wind speed data (Simiua et al., 2001).

Out of a number of probability distributions that are adopted in frequency analysis, Gumbel (EV1), Frechet (EV2), 2-parameter Log Normal (LN2) and Log Pearson Type-3 (LP3) are extensively used

Extreme Value Analysis (EVA) of wind speed. Based on the applicability, standard parameter estimation procedures viz., Method of Moments (MoM), Maximum Likelihood Method (MLM), Method of Least Squares (MLS) and Order Statistics Approach (OSA) are used for determination of parameters (Celik, 2004). In the recent past, number of studies has been carried out by researchers adopting probability distributions for Extreme Value Analysis (EVA) of wind speed. Kunz et al. (2010) compared the Gamma and Generalized Pareto (GPA) distributions for estimation of EWS and concluded that GPA provides better estimates than Gamma. Morgan et al. (2011) applied Extreme Value, Gamma and Normal family of probability distributions for estimation of EWS using the 10-minute wind speed observations recorded at various stations around North America. They have found that the LN2 distribution yielded the best estimate of EWS, but still exhibited large errors. El-Shanshoury and Ramadan (2012) applied EV1 distribution to estimate EWS for Dabaa area in the north-western coast of Egypt. Lee et al. (2012) applied Gumbel and Weibull probability distributions for estimation of EWS using the Korea wind map. They have observed that the Gumbel distribution gives better results than the Weibull. Daneshfaraz et al. (2013) carried out the wind speed frequency analysis adopting LN2, truncated extreme value, truncated logistic and Weibull probability distributions and found that the truncated extreme value is the most appropriate distribution for Iran. Lawan et al. (2015) evaluated the suitability of five different probability distributions through GoF tests and found that the Gamma and LN2 distributions are better suited for modelling wind speed data of Miri, Malaysia. Generally, when different probability distributions are used for EVA, a common problem that arises is how to determine which model fits best for a given set of data. This can be evaluated by quantitative assessment using Goodness-of-Fit (GoF) and diagnostic tests. GoF tests such as Anderson-Darling (A2) and Kolmogorov-Smirnov (KS) are applied for checking the adequacy of fitting of probability distributions to the recorded wind speed data (Palutikof et al., 1999). A diagnostic test of D-index is used for the selection of a suitable probability distribution for estimation of EWS. Thus, there exist research efforts in assessing EWS for aiding design parameter of interest and present work is an effort in this direction. This paper illustrates the procedures involved in assessing the suitable probability distribution for estimation of EWS though GoF and diagnostic tests with illustrative example.

2. Methodology

The study is to assess the adequacy of Probability Density Function (PDF) for EVA of wind speed under scenario of missing value. In this context, various steps followed for data processing, validation and analysis are: (i) prepare two different data series from the recorded hourly maximum wind speed with imputation and without imputation; (ii) select the PDFs for EVA (say, EV1, EV2, LN2 and LP3); (ii) select parameter estimation methods (say, MoM, MLM, MLS and OSA wherever applicable; (iii)

Probability Density Function (PDF) and quantile estimator (XT) of these distributions are presented in Table 1.

Table 1 PDF and quantile estimator of probability distributions

Distribution	PDF	Quantile estimator
EV1	$f(X;\alpha,\beta)=\dfrac{e^{-(X-\alpha)/\beta}e^{-e^{-(X-\alpha)/\beta}}}{\beta}$, β>0	$X_T=\alpha+Y_T\beta$
EV2	$f(X;\beta,\gamma)=\dfrac{\gamma}{\beta}\left(\dfrac{\beta}{X}\right)^{\gamma+1}e^{-\left(\dfrac{X}{\beta}\right)^{-\gamma}}$, β>0	$X_T=\beta e^{(Y_T/\gamma)}$
LN2	$f(X;\mu_Y,\sigma_Y)=\dfrac{1}{\sqrt{2\pi}\sigma_Y X}exp\left(-\dfrac{(ln(X)-\mu_Y)^2}{2\sigma_Y^2}\right)$,-∞<X<∞,σ>0	$X_T=e^{\mu_Y+K_P\sigma_Y}$
LP3	$f(X;\alpha,\beta,\gamma)=\dfrac{1}{\beta X\Gamma\gamma}\left(\dfrac{ln(X)-\alpha}{\beta}\right)^{\gamma-1}e^{-\left(\dfrac{ln(X)-\alpha}{\beta}\right)}$,β,γ>0	$x_T=Exp((\alpha+\beta\gamma)+K_P\beta\sqrt{\alpha})$

In Table 1, the symbols μ_y and σ_y represents the mean and standard deviation of the log-transformed series of recorded data. α, β and γ denotes the location, scale and shape parameters of the distributions respectively. For EV1 and EV2 distributions, the reduced variate (Y_T)corresponding to the return period (T) is defined by $Y_T=-ln(-ln(1-(1/T)))$ while in the mathematical representation of LN2 and LP3, K_P denotes the frequency factor corresponding to the probability of exceedance. The Coefficient of Skewne(C_S) is C_S=0.0 for LN2 whereas $SC$is based on the log transformed series of the recorded data for LP3 (Bobee and Askhar, 1991; Bivona et al., 2003).

2.1 Goodness-of-Fit tests

Generally, A^2 statistic is applied for checking the adequacy of fitting of EV1 and EV2 distributions. The procedures involved in application of A^2 statistic for LN2 and LP3 are more complex though the utility of the test statistic is extended for checking the quantitative assessment. In view of the above, KS test is widely applied for the purpose of quantitative assessment. Theoretical descriptions of GoF tests statistic are as follows:

A^2 statistic is defined by:

$$A^2=(-N)-(1/N)\sum_{i=1}^{N}\{(2i-1)Ln(Z_i)+(2N+1-2i)Ln(1-Z_i)\} \quad \ldots (1)$$

Here, $Z_i=F(X_i)$ for i=1,2,3,…,N with $X_1<X_2<\ldots<X_N$, $F(X_i)$ s the Cumulative Distribution Function (CDF) of i^{th} sample (Xi) and N is the sample size.

KS statistic is defined bv:

$$KS=\underset{i=1}{\overset{N}{Max}}(F_e(X_i)-F_D(X_i)) \quad \ldots (2)$$

Here, $F_e(X_i)$ is the empirical CDF of X_i and $F_D(X_i)$ is the derived CDF of X_i by PDFs. For the present work, Weibull plotting position formula is used for computation of empirical CDF. The theoretical value A^2 and KS statistic for different sample size (N) at 5% significance level is available in the technical note on "Goodness-of-Fit Tests for Statistical Distributions" by Charles Annis (2009).

Test criteria: If the computed values of GoF tests statistic given by the distribution is less than that of theoretical values at the desired significance level (either at 5% or 1%), the distribution would be taken as acceptable for EVA of wind speed at that level.

2.2 Diagnostic test

Sometimes the GoF test results would not offer a conclusive inference thus posing a problem for the user in selecting a suitable PDF for their application. In such cases, a diagnostic test in adoption to GoF is applied for making inference. The selection of a suitable probability distribution for estimation of EWS is performed through D-index test (USWRC, 1981), which is defined as below:

$$\text{D-index} = \left(1/\overline{X}\right)\sum_{i=1}^{6}\left|X_i - X_i^*\right| \quad \ldots (3)$$

Here, $\overline{x}$ is the average value of the recorded data whereas X_i (i= 1 to 6) and X^*_i are the six highest recorded and corresponding estimated values by different PDFs. The distribution having the least D-index is considered as better suited distribution.

3. Application

EVA of wind speed data was carried out to estimate EWS for different return periods adopting four different PDFs viz., EV1, EV2, LN2 and LP3. MoM, MLM, MLS and OSA were used for determination of parameters of EV1 and EV2 distributions whereas MoM and MLM for LN2 and LP3 distributions. Hourly wind speed data (with missing values) for the period 1969 to 2012 was used. The series of Annual Maximum Wind Speed (AMWS) was extracted from the hourly data and used for EVA. From the scrutiny of the wind speed data, it was observed that the data for the period of twelve years (1974, 1979 to 1981, 1983 to 1988, 1990 and 2004) are missing. So, the data for the missing years were imputed by the series maximum value as per Atomic Energy Regulatory Board (AERB) guidelines and the entire data set is used for EVA. To study the effect of imputation, two different scenarios viz., (i) data series with imputation as per AERB (2008) guidelines (Scenario-1) and (ii) data series without imputation (Scenario-2) were considered while carrying out EVA of wind speed. Table 2 gives the descriptive statistics of the series of AMWS.

Table 2 Descriptive statistics of the series of AMWS

Data series of AMWS	Statistical parameters			
	Average (km/hr)	SD (km/ hr)	C_S	C_K
Scenario-1	66.6	15.4	-0.007	-1.584
Scenario-2	59.9	12.4	0.574	-0.657

SD: Standard Deviation; C_S: Coefficient of Skewness; C_K: Coefficient of Kurtosis

4. Results and Discussions

Based on the parameter estimation procedures of EV1, EV2, LN2 and LP3 distributions (Rao and Hameed, 2000), computer codes were developed in FORTRAN language and used for EVA of wind speed. These programs compute the distribution parameters, estimates of EWS with standard error for different return periods and also perform GoF tests statistic and D-index. For Scenarios 1 and 2, the estimated EWS (XT) with Standard Error (SE) computed from EV1, EV2, LN2 and LP3 probability distributions are presented in Tables 3 to 6.

Table 3 Estimated EWS with Standard Error (SE) using EV1 and EV2 distributions (Scenario-

Return period (year)	Estimated EWS with SE (km/hr)															
	EV1								EV2							
	MoM		MLM		MLS		OSA		MoM		MLM		MLS		OSA	
	EWS	SE	EWS	SE	EWS	SE	EWS	SE	EWS	SE	EWS	SE	EWS	SE	EWS	SE
2	64.1	2.1	64.1	2.1	64.2	2.4	65.1	1.6	62.3	1.6	62.4	1.6	62.4	10	63.3	64.9
5	77.7	3.6	77.6	3.5	80	4	75.2	2.5	77	3.2	80.3	3.3	79.7	17	74.5	77.6
10	86.7	4.8	86.5	4.7	90.4	5.5	81.9	3.3	88.5	4.9	94.9	5.2	93.8	23	83.1	87.6
20	95.4	6.1	95	6	100.4	7	88.3	4.2	101.3	6.9	111.4	7.6	109.6	29.1	92.2	98.5
50	106.6	7.8	106	7.7	113.3	8.9	96.7	5.2	120.5	10.5	137	12	134.2	37.1	106	114.7
100	115	9	114	8.9	123	10.4	102.9	6	137.3	14	160.1	16.3	156.1	43.2	117	128.6
200	123.4	10.3	123	10.2	132.7	11.8	109.1	6.9	156.3	18.2	186.9	21.8	181.5	49.3	129	144.1
500	134.4	12	134	11.9	145.4	13.9	117.3	7.9	185.5	25.3	229.2	31.4	221.4	57.5	147	167.5
1000	142.8	13.3	142	13.1	155	15.4	123.5	8.8	211.1	32.1	267.4	40.8	257.3	63.7	163	187.6
2000	151.1	14.6	150	14.4	164.7	16.8	129.7	9.7	240.2	42.5	312.1	55.1	299	69.9	180	211.8
5000	162.1	16.4	161	16.1	177.4	18.8	137.9	10.7	285	53.8	382.6	72.3	364.7	78.1	206	244.3
10000	170.5	17.6	169	17.4	187	20.3	144.1	11.5	324.3	66.6	446.5	91.5	423.8	84.3	227	273.8

Table 4 Estimated EWS with Standard Error (SE) using LN2 and LP3 distributions (Scenario-1)

Return period (year)	Estimated EWS with SE (km/hr)							
	LN2				LP3			
	MoM		MLM		MoM		MLM	
	EWS	SE	EWS	SE	EWS	SE	EWS	SE
2	64.8	2.3	64.8	2.3	65.3	4.8	56.9	4.5
5	79.2	3.3	79.1	3.2	79.4	6.5	71.7	6.3
10	88	4.3	87.7	4.2	87.6	8.2	80.5	8.2
20	96	5.3	95.6	5.2	94.7	10.1	88.3	10.2
50	105.9	6.7	105	6.6	103.2	12.5	97.7	12.9
100	113	7.8	112	7.7	109.1	14.4	104.3	15
200	120	8.9	119	8.7	114.7	16.3	110.7	17.1
500	129	10.4	128	10.2	121.7	18.8	118.7	19.9
1000	135.7	11.6	135	11.4	126.8	20.7	124.6	22.1
2000	142.3	12.9	141	12.6	131.7	22.5	130.3	24.2
5000	150.6	13.5	150	12.4	137.1	24.9	136.6	26.6
10000	157.7	13.6	156	13.4	141	27	142	29

It was observed from the results of two scenarios that the estimated EWS using EV2 (MLM) is consistently higher than the corresponding values of EV1, LN2 and LP3 distributions. From Table 4, it was noted that there is no significant difference between the estimated EWS while MoM and MLM is considered for determination of parameters of LN2 and LP3 distributions. Also, from Table 6, it was noted that the estimated EWS using LP3 (MoM) is relatively higher than the corresponding values of LN2 (MoM and MLM) and LP3 (MLM). In general, the estimates for a given probability distribution in Scenario-1 are higher than Scenario-2 are in expected line as Scenario-1 represents a conservative approach. The EVA results obtained from Scenario-1 and Scenario-2 were used to develop the probability plots and presented in Figures 1 and 2 respectively.

Table 5 Estimated EWS with Standard Error (SE) using EV1 and EV2 distributions (Scenario-

Return period (year)	Estimated EWS with SE (km/hr)															
	EV1								EV2							
	MoM		MLM		MLS		OSA		MoM		MLM		MLS		OSA	
	EWS	SE	EWS	SE	EWS	SE	EWS	SE	EWS	SE	EWS	SE	EWS	SE	EWS	SE
2	57.9	1.9	57.8	2	57.9	2.1	57.9	2	57.1	1.7	56.8	2	56.8	2	56.9	2
5	68.8	3.4	68.9	3.1	69.5	3.5	69.1	3.2	67.1	4.7	69.3	4	68.7	4	69	4
10	76.1	4.5	76.3	4	77.2	4.7	76.4	4.3	74.6	7.7	79.1	6.1	78	6	78.4	6
20	83.1	5.7	83.3	5	84.6	6	83.5	5.3	82.7	11.2	89.7	8.7	88.1	8.4	88.6	8.6
50	92.1	7.3	92.5	6.2	94.1	7.7	92.7	6.7	94.4	17.1	105.7	13	103	12.7	104	12.8
100	98.9	8.4	99.3	7.3	101.2	9	99.5	7.8	104.2	22.6	119.5	17.2	115.9	16.7	117	16.9
200	105.6	9.7	106	8.3	108.4	10.2	106.4	8.8	115	29.2	135	22.4	130.3	21.6	132	21.8
500	114.5	11.3	115	9.6	117.8	11.9	115.4	10.3	131	39.9	158.6	31	152	29.7	154	30
1000	121.2	12.5	122	10.6	124.9	13.2	122.2	11.3	144.6	49.7	179.2	39	170.9	37.2	173	37.7
2000	127.9	13.7	129	11.6	132	14.5	129	12.7	159.5	72.8	202.4	61.7	192	58.6	195	59.5
5000	136.8	15.3	138	12.9	141.4	16.2	138	13.9	181.7	80	237.7	64.6	224.1	60.9	228	62
10000	143.5	16.5	145	13.9	148.5	17.4	144.8	15	200.5	97	268.5	79.5	251.8	74.6	257	76

Table 6 Estimated EWS with Standard Error (SE) using LN2 and LP3 distributions (Scenario-2)

Return period (year)	Estimated EWS with SE (km/hr)							
	LN2				LP3			
	MoM		MLM		MoM		MLM	
	EWS	SE	EWS	SE	EWS	SE	EWS	SE
2	58.7	2.1	58.7	2.1	58.1	4.6	58.5	7.2
5	69.7	2.9	69.4	2.8	69.7	4.8	69.6	8.8
10	76.3	3.7	75.8	3.6	76.5	8.8	76.3	14.3
20	82.2	4.5	81.5	4.3	83.3	10.9	82.3	17.9
50	89.4	5.6	88.4	5.4	91.9	14	89.8	22.6
100	94.5	6.5	93.4	6.2	98.3	16.6	95.2	26.4
200	99.5	7.4	98.1	7.1	104.7	19.2	100.4	30.4
500	105.8	8.6	104	8.2	113.2	23	107.2	35.7
1000	110.5	9.5	109	9.1	119.8	25.9	112.3	39.9
2000	115.2	10.4	113	9.9	126.4	29.1	117.3	44.4
5000	120.3	11.5	118	11	134.1	32.9	123.9	49.9
10000	125	12	123	12.1	142.3	37.1	128.9	55.2

4.1 Analysis based on GoF tests

The adequacy of fitting four different PDFs for EVA of wind speed was performed by adopting GoF tests such as A2 and KS, as described in Section 2.1. The GoF tests results for the Scenarios 1 and 2 adopted in EVA of wind speed are presented in Table 7.

Table 7 Computed and theoretical values of GoF tests statistic adopting Ev1, EV2, LN2 and LP3 distributions for Scenario-1 and Scenario-2

GoF tests	Computed values of GoF tests statistic												Theoretical value at	
	EV1				EV2				LN2		LP3			
	MoM	MLM	MLS	OSA	MoM	MLM	MLS	OSA	MoM	MLM	MoM	MLM	5%	1%
Scenario-1														
A^2	3.047	2.685	1.678	9.423	2.678	1.638	1.715	8.06	1.736	1.951	1.718	1.944	0.757	1.04
KS	0.186	0.188	0.157	0.229	0.177	0.142	0.147	0.22	0.186	0.185	0.175	0.187	0.198	0.24
Scenario-2														
A^2	0.414	0.452	0.342	0.392	0.517	0.335	0.38	0.36	0.621	0.502	0.518	0.441	0.757	1.04
KS	0.075	0.078	0.064	0.072	0.088	0.092	0.091	0.09	0.106	0.123	0.106	0.116	0.231	0.28

From the computed results of GoF tests (A^2 and KS), the following observations were made from the study:

i) For Scenario-1, A^2 test didn't support the selection of EV1, EV2, LN2 and LP3 distributions for EVA of wind speed.

ii) KS test supported the use of EV1, EV2, LN2 and LP3 distributions for EVA of wind speed adopting Scenario-1.

iii) For Scenario-2, A^2 and KS tests suggested the EV1, EV2, LN2 and LP3 distributions were found to be acceptable for EVA of wind speed.

In Scenario-1, 12 out 44 AMWS missing values were replaced with historical maximum AMWS so as to get conservative estimates for design parameter. Such arrangement however may affect the series characteristics of randomness, independence and homogeneity. The reason for rejection of A^2 test for Scenario-1 and acceptance for Scenario-2 can be attributed to the above mentioned factor. On the other side, KS test accepts GoF in both scenarios and was unable to detect the difference of two scenarios indicates its inconsistency as compared with A^2 test.

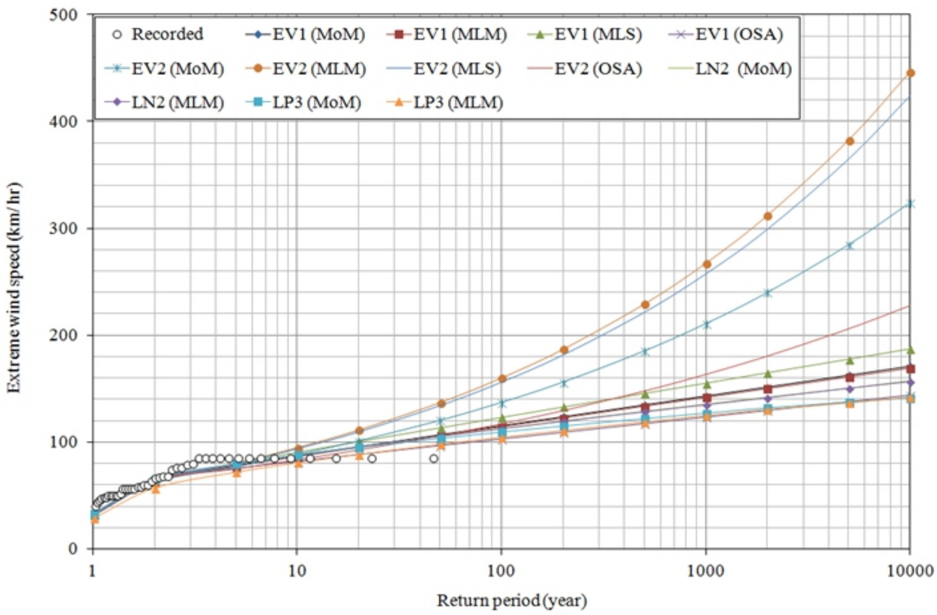
4.2 Analysis based on diagnostic test

A diagnostic test of D-index, as given in Section 2.2, was used for assessing the adequate probability distribution for EVA of wind speed adopting Scenarios 1 and 2. The D-index values computed for EV1, EV2, LN2 and LP3 distributions by different methods are given in Table 8.

Table 8 D-index values computed by EV1, EV2, LN2 and LP3 distributions

Data series	Computed values of GoF tests statistic											
	EV1				EV2				LN2		LP3	
	MoM	MLM	MLS	OSA	MoM	MLM	MLS	OSA	MoM	MLM	MoM	MLM
Scenario-1	0.877	0.853	1.283	0.51	1.433	2.358	2.199	0.81	0.749	0.688	0.544	0.551
Scenario-2	0.391	0.404	0.342	0.374	0.528	0.656	0.599	0.62	0.332	0.37	0.351	0.341

From Table 8, it was noted that the D-index value of EV1 (OSA) is minimum when compared with the corresponding values of other distributions for Scenario-1 whereas LN2 (MoM) having minimum for Scenario-2.



iii) For Scenario-2, A^2 and KS tests suggested the EV1, EV2, LN2 and LP3 distributions were found to be acceptable for EVA of wind speed.

In Scenario-1, 12 out 44 AMWS missing values were replaced with historical maximum AMWS so as to get conservative estimates for design parameter. Such arrangement however may affect the series characteristics of randomness, independence and homogeneity. The reason for rejection of A^2 test for Scenario-1 and acceptance for Scenario-2 can be attributed to the above mentioned factor. On the other side, KS test accepts GoF in both scenarios and was unable to detect the difference of two scenarios indicates its inconsistency as compared with A^2 test.

4.2 Analysis based on diagnostic test

A diagnostic test of D-index, as given in Section 2.2, was used for assessing the adequate probability distribution for EVA of wind speed adopting Scenarios 1 and 2. The D-index values computed for EV1, EV2, LN2 and LP3 distributions by different methods are given in Table 8.

Table 8 D-index values computed by EV1, EV2, LN2 and LP3 distributions

Data	Computed values of GoF tests statistic											
series	EV1				EV2				LN2		LP3	
	MoM	MLM	MLS	OSA	MoM	MLM	MLS	OSA	MoM	MLM	MoM	MLM
Scenario-1	0.877	0.853	1.283	0.51	1.433	2.358	2.199	0.81	0.749	0.688	0.544	0.551
Scenario-2	0.391	0.404	0.342	0.374	0.528	0.656	0.599	0.62	0.332	0.37	0.351	0.341

From Table 8, it was noted that the D-index value of EV1 (OSA) is minimum when compared with the corresponding values of other distributions for Scenario-1 whereas LN2 (MoM) having minimum for Scenario-2.

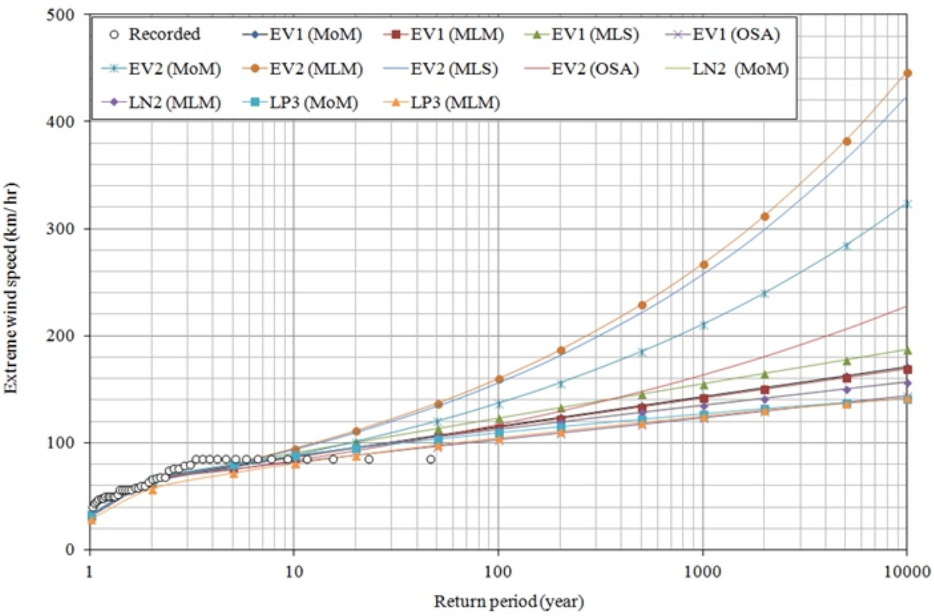


Figure 1 Recorded and estimated EWS using four probability distributions for Delhi (Scenario-1)

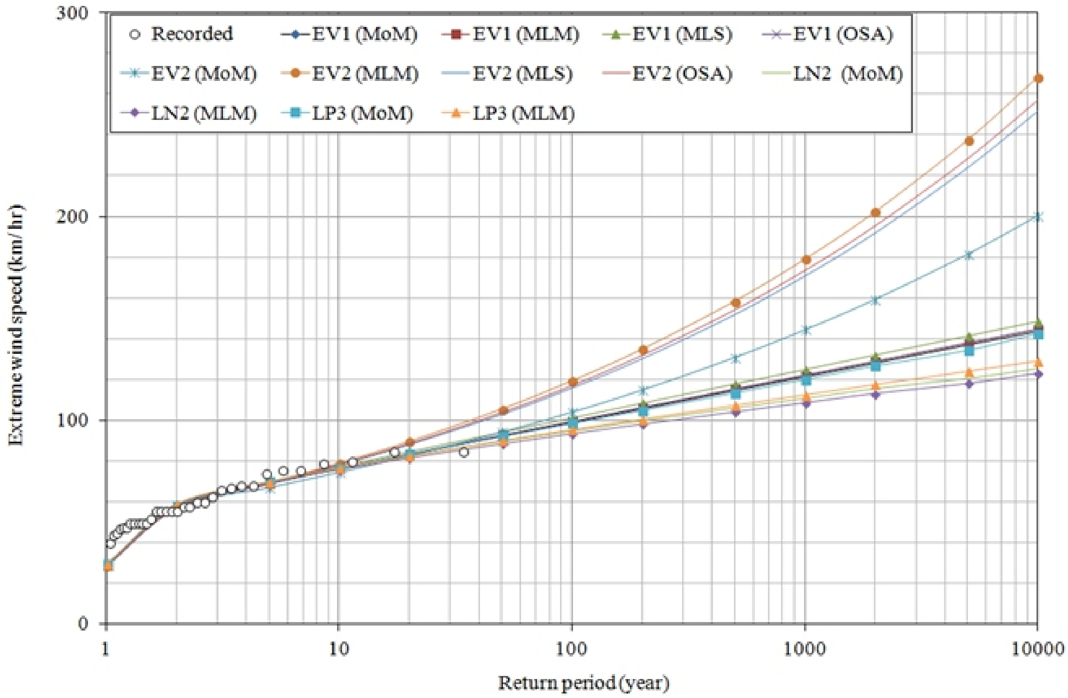


Figure 2 Recorded and estimated EWS using four probability distributions for Delhi (Scenario-2)

4.3 Selection of probability distribution

Based on the findings obtained through GoF and diagnostic tests results, the study suggested the EV1 (OSA) is the most appropriate distribution for EVA of wind speed using the data series with imputation (Scenario-1) whereas LN2 (MoM) for the data series without imputation (Scenario-2). The Mean and Mean+1 (where Mean denotes the estimated EWS and the Standard Error) values for 100-yr, 1000-yr and 10000-yr computed by the selected probability distribution for Scenario-1 and Scenario-2 are given in Table 9.

Data series	Probability distribution	Parameter estimation method	Estimated extreme values (km/hr)					
			100-year		1000-year		10000-year	
			Mean	Mean+1σ	Mean	Mean+1σ	Mean	Mean+1σ
Scenario-1	EV1	OSA	102.9	108.9	123.5	132.3	144.1	155.6
Scenario-2	LN2	MoM	94.5	101	110.5	120	125	137

5. Conclusions

The paper presents the study carried out for estimation of EWS by adopting EVA and also assessing the adequacy of probability distributions using a computer aided procedure with applicable parameter estimation methods. The following conclusions are drawn from the study:

I) It was found that the estimated EWS by EV2 (MLM) distribution are higher than the corresponding values of other three distributions in EVA of wind speed adopting Scenarios 1 and 2.

ii) Suitability of probability distribution is evaluated by GoF (using A2 and KS) and diagnostic (using D-index) tests.

a) The A^2 test results suggest the EV1, EV2, LN2 and LP3 distributions are not acceptable for estimation of EWS using the data series of Scenario-1.

b) The A^2 test results showed that the four probability distributions are acceptable for EVA of wind speed using the data series of Scenario-2.

c) For Scenarios 1 and 2, it was observed that the KS test results suggest the use of all the probability distributions considered for EVA of wind speed.

iii) On the other hand, the qualitative test, i.e., the trend lines of the fitted curves using estimated EWS values, the study presented that the EV1 distribution (using OSA) is better suited amongst four distributions studied for estimation of EWS at Delhi using the data series of Scenario-1 whereas LN2 (MoM) for Scenario-2.

iv) The estimated EWS obtained from Scenario-1 with EV1 (OSA) is higher than the values of Scenario-2 with LN2 (MoM). However, by considering the design-life of the structure, the 10000-year return period Mean+1 value of 155.6 km/ hr obtained from Scenario-1 is recommended for the design of civil structures.

Acknowledgements

The authors are grateful to Shri S. Govindan, Director, Central Water and Power Research Station, Pune, for encouragement during conduct of the studies and also according permission to publish this paper. The authors are thankful to S/Shri N. Madhusudana Rao, Chief Engineer, S.K. Sharma and K. Giridhar, Additional Chief Engineers, M/s Nuclear Power Corporation of India Limited, Mumbai for supply of wind speed data used in the study and also for useful discussions.

References

1) AERB (2008), Extreme values of meteorological parameters, Atomic Energy Regulatory Board safety guide AERB/ NF/ SG/ S-3.

2) Bivona, S., Burlon, R. and Leone, C. (2003), Hourly wind speed analysis in Sicily, Renewable Energy, 28 (9): 1371-1385.

3) Bobee, B. and Askhar, F. (1991), The Gamma family and derived distributions applied in hydrology, Water Resources Publications.

4) Celik, A.N. (2004), On the distributional parameters used in assessment of the suitability of wind speed probability density functions, Energy Conversion and Management, 45 (11 & 12): 1735-1747.

5) Charles Annis, P.E. (2009), Goodness-of-Fit tests for statistical distributions, [http://www.statisticalengineering.com/goodness.html].

6) Daneshfaraz, R., Nemati, S., Asadi, H. and Menazadeh, M. (2013), Comparison of four distributions for frequency analysis of wind speed: A case study, Journal of Civil Engineering and Urbanism, 3 (1): 6-11.

7) El-Shanshoury, I, and Ramadan, A.A. (2012), Estimation of extreme value analysis of wind speed in the North-Western Coast of Egypt, Arab Journal of Nuclear Science and Applications, 45 (4): 265-274.

8) Kunz, M., Mohr, S., Rauthe, M., Lux, R. and Kottmeier Ch. (2010), Assessment of extreme wind speeds from regional climate models – Part 1: Estimation of return values and their evaluation, Natural Hazards Earth System Sciences, 10 (4): 907-922.

9) Lawan, S.M., Abidin, W.A.W.Z., Chai, W.Y., Baharun, A. and Masri, T. (2015); Statistical modelling of long-term wind speed data, American Journal of Computer Science and Information Technology, 3 (1): 1-6.

10) Lee, B.H., Ahn, D.J., Kim, H.G. and Ha, Y.C. (2012), An estimation of the extreme wind speed using the Korea wind map, Renewable Energy, 42 (1): 4–10.

11) Morgan, E.C., Lackner, M., Vogel, R.M. and Baise, L.G. (2011), Probability distributions for offshore wind speeds, Energy Conversion and Management, 52 (1): 15–26.

12) Palutikof, J.P., Brabson, B.B., Lister, D.H. and Adcock, S.T. (1999), A review of methods to calculate extreme wind speeds, Meteorological Applications, 6 (2): 119-132

13) Rao, A.R. and Hameed, K.H. (2000), Flood frequency analysis, CRC Publications, Washington D.C., New York.

14) Simiua, E., Heckertb, I, N.A., Fillibenb, J.J. and Johnson, S.K. (2001), Extreme wind load estimates based on the Gumbel distribution of dynamic pressures: an assessment, Structural Safety, 23 (3): 221-229.

15) USWRC (1981), Guidelines for determining flood flow frequency, United States Water Resources Council, Bulletin No. 17B.

Groundwater Pollution Source Identification Through Inverse Modelling

Jyoti Chaubey1, Deepak Kashyap2

[1]*Research Scholar, Department of Civil Engineering, IIT Roorkee,Roorkee 247667, India*

Email: jyotidce@iitr.ernet.in

[2]*Professor, Department of Civil Engineering, IIT Roorkee, Roorkee 247667, India*

Email: dkashfce@iitr.ernet.in

ABSTRACT

The present study involves development and illustration of an inverse problem model of groundwater pollution source identification. The model is based upon linked simulation optimization approach and invokes a finite difference based simulator and a Sequential Unconstrained Minimization Technique (SUMT) based optimizer. The model arrives at such pollution source characteristics that provide the closest match between observed and the simulated concentration field. The estimated pollution source characteristics include number of source, their location and strength. Finding the optimum number of sources and their location is an advantage of this model, since most of the source identification problem involves a priori known potential location. The model has been illustrated for a hypothetical study area, with known boundary conditions and flow/transport parameters. Two dimensional steady state flow and transport has been considered. Pollutant source is assumed to release a conservative pollutant. Identified results indicated that the proposed methodology can be used to solve the inverse problem of groundwater pollution source identification.

Keywords: Groundwater pollution, source characteristics identification, simulation, optimization.

1. Introduction

Groundwater source identification is an important problem in groundwater risk assessment studies and groundwater management problems. It is a type of an inverse problem in context of groundwater solute transport modelling. In this inverse problem source characteristics are obtained for a given set of concentration and head fields and flow and transport parameters.

Several methods have been reported to solve the source identification inverse problem. These methods can be broadly classified as Deterministic direct methods (Skaggs and Kabala 1994, 1995, 1998; Sidauruk et. al. 1997; Ball et al. 1997; Liu and Ball 1999; Atmadja and Bagtzoglou 2001), Probabilistic and Geostatistical simulation approaches (Bagtzoglou et al. 1991, 1992; Snodgrass and Kitanidis 1997; Michalak and Kitanidis 2010; Woodbury and Ulyrich 1993, 1996, 1998) and Indirect approaches such as Optimization approaches (Gorelick 1982, 1983; Mahar and Dutta 2000, 2001; Aral et al. 2001; Singh et al. 2004; Singh and Datta 2006; Datta et al. 2009; Ayvaz 2010; Jha and Dutta 2013, 2015; Chaubey and Kashyap 2014). Detailed review of studies related to source identification problem has been performed by Atmadja and Bagtzoglou (2001) and Amirabdollahian and Datta (2013).

Optimization is one of the most common solution approaches for source identification problem. Optimization approach consists of the integration of simulation with the optimization model. Depending upon the incorporation of Simulation model into the optimization model it can be termed as embedded method i.e. incorporating simulation model as constraints (Mahar and Dutta 2000, 2001), Kernel method i.e. incorporating simulation model as as concentration response matrix (Gorelick 1982, 1983) and Linked Simulation optimization approach i.e. linking simulation model externally with an optimization model (Datta et al. 2009; Aral et al. 2001; Singh et al. 2004; Singh and Datta 2006; Ayvaz 2010;Jha and Dutta 2013, 2015;Chaubey and Kashyap 2014).

Simulation model solves the governing flow and solute transport equation for a given initial and boundary conditions, flow and transport parameters. It is integrated with the optimization model which aims at selecting that set of sources characteristics (location, magnitude and release history) which results in simulated concentrations closest with the local groundwater solute concentration data.

Present study is based on the linked simulation optimization approach for groundwater source identification. Source characteristics usually involve source location, strength and release history. In most of the studies related to groundwater pollution source identification, source number and location is known a priori. An attempt has been made by Ayvaz (2010) in solving the inverse problem with unknown number of sources and their location by considering a hypothetical example.

Present study is an extended work of Chaubey and Kashyap (2014). In latter work source locations were assigned a priori and then the source strength was determined. Source characteristics in the present context comprise number of sources, their location and their fluxes. The model invokes a finite difference based simulator and an Exhaustive Search based optimizer.

2. Methodology

Simulation model based on governing 2D steady state groundwater solute transport equation is used. It can be given as (Bear, 1972)

$$\frac{\partial}{\partial x}\left(D_{xx}\frac{\partial C}{\partial x}\right) + \frac{\partial}{\partial x}\left(D_{yy}\frac{\partial C}{\partial y}\right) - \frac{\partial}{\partial x}(uC) - \frac{\partial}{\partial y}(vC) + \frac{W}{b\varphi} = 0 \qquad 1$$

where, D_{xx} and D_{yy} are hydrodynamic dispersion coefficients, u and v are seepage velocities along x and y direction respectively, W is the pollution source flux, b is aquifer thickness, φ is effective porosity of the aquifer, C is concentration of solute dissolved in groundwater. The velocities in Eqn. 1 are given by,

$$u = \frac{\frac{\partial h}{\partial x} T_{xx}}{b * \varphi} \text{ and } v = \frac{\frac{\partial h}{\partial y} T_{yy}}{b * \varphi} \qquad 2$$

where, h is nodal head, T_{xx} and T_{yy} are transmissivities in x and y direction respectively.

Hydrodynamic dispersion coefficients are taken as $D_{xx} = \alpha_L |u|$ and $D_{yy} = \alpha_T |v|$ where, α_L is longitudinal dispersivity and α_T is transverse dispersivity.

The necessary head fields for computing velocity vector are simulated by solving following two dimensional, steady state groundwater flow equation (Bear, 1972):

$$\frac{\partial}{\partial x}\left(T_{xx}\frac{\partial h}{\partial x}\right) + \frac{\partial}{\partial y}\left(T_{yy}\frac{\partial h}{\partial y}\right) - W' = 0 \qquad 3$$

where, x and y are cartesian coordinates in the directions of principal permeability and *W'* is net vertical abstraction.

Iterating Alternative Direction Implicit Explicit (IADIE) finite difference scheme has been applied for solving Eqn. 3 to generate head field and corresponding velocities calculated are used in Eqn. 1. Eqn. 1 is solved by using IADIE finite difference scheme to generate concentration fields.

This simulation model is linked externally with the optimizer, to arrive at closest match between simulated and observed concentration distribution.

Optimization problem formulation

The objective of the optimization problem is

$$\text{Minimize } Z = \sum_{i \in I}\left(C_i^{obs} - C_i^{simu}\right)^2$$

$$\text{such that } C = f(X, Y, W)$$

$$\text{subject to}$$

$$W_j \geq 0 \quad j=1, 2\ldots, n$$

$$X_L \leq X \leq X_U$$

$$Y_L \leq Y \leq Y_U$$

where C_i^{obs} is observed concentration at node i , C_i^{simu} is simulated concentration at node i, W_j s point source flux, $i \in I$ where I is the set of nodes at which observed concentration data are available, X $(=\{x_1,x_2 \ldots ,x_n\})$ is the x coordinate vector of point sources, Y $(=\{y_1,y_2,\ldots.,y_n\})$ is the y coordinate vector of point sources, n is the number of unknown sources and C is the simulated concentration vector which is a function of the unknown sources location and fluxes. Thus Decision variables are *X, Y, and W* . n is also a decision variable but it is not used as an explicit variable.

Sequential Unconstrained Minimization Technique (SUMT) technique was employed to solve the above posed optimization problem.

For each n, an initial set of x,y and W was given to the optimizer. This was passed to the simulator and corresponding concentration field was simulated which was matched with the observed concentration field and the residual was calculated. This residual was used to calculate the objective function value, which was utilized by optimization algorithm to improve the candidate solution. The process continued till an optimal solution was obtained. The process was repeated for different n values. n for which least objective function value was obtained was chosen as optimal n and corresponding x, y and W as optimal solution to the source identification problem.

3. Model illustration

The model developed has been illustrated by a hypothetical two dimensional study area, shown in Figure-1. Observed concentration field was needed for demonstrating the model applicability. It was generated by simulation model for assumed source flux values and known source location. This generated concentration field was then used as an input to the linked simulation optimization model, masking the assumed source flux values and known source location.

3.1 Study area and database

Study area of dimension 549 m×732 m was taken and is shown in Figure 1. The north and south boundaries were considered as impermeable boundaries while east and west boundaries were taken as constant head boundaries. Two pollution sources (S1 and S2) were there in the study area, the location of which is shown in the Figure-2. A pond was located in the area, which was considered free from contamination. The flow and transport parameters used in the simulation model are given in Table 1.

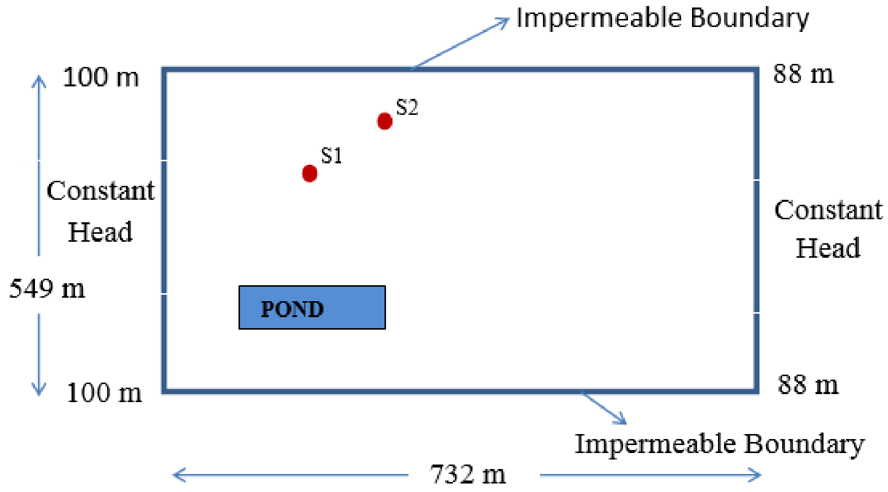


Figure1Study Area

Table 1 Simulation model Parameters

Parameters	Value
T_{xx} (m^2/d)	26.35
T_{yy} (m^2/d)	26.35
Grid spacing in x direction, Äx (m)	91.5
Grid spacing in y direction, Äy (m)	91.5
Longitudinal dispersivity, $á_L$ (m)	40
Transverse dispersivity,$á_T$ (m)	9.6
Aquifer Thickness, b (m)	30.5
Effective porosity, ?	0.2
Pond Recharge (m/d)	0.011
Source flux (S1 and S2) ($kg/m^2/d$)	31.7

3.2 Observed Concentration field generation

The study area was divided into grids of size 91.5×91.5 m each, shown in Figure 2. Initial concentration of pollutant is assumed to be zero in the groundwater. Pollution sources were assumed to release conservative pollutant at a uniform rate throughout the activity period. The corresponding steady state concentration field was generated by solving the solute-transport equation numerically, employing the parameters and the assumed source flux listed above. Locations of the sources are given in Table 2.

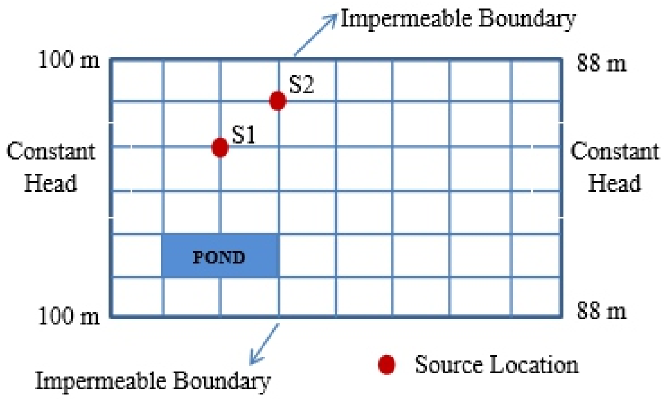


Figure 2Finite Difference Grid for the study area

Table 2 Source Locations

Source	S1	S2
x (m)	274.5	183
y(m)	549	457.5

3.3 Model Runs

The model was run by assuming a particular number of source (n let) and the decision variables varying accordingly (3n). Search for number of sources started from n=2, on the basis of suspected number of sources.

The boundary constraints (Lower and Upper limits) for the location coordinates of both the sources were

$$0 \leq x \leq 732$$

$$0 \leq y \leq 549$$

i.e. the entire study domain was searched.

4. Results

Results obtained for the source identification problem are given in Table 3. For n=2, the objective function value was found to be 0.0911, representing a close enough match between observed and simulated concentration filed. Since the objective function value was very less for n=2, we did not go for n=3. The value chosen for n was 2 and corresponding optimal source location and flux chosen as the solution to the source identification problem.

Table 3 Identified results for different number of pollution sources

Number of Sources	Actual source flux ($kg/m^2/d$)	Actual source location (m)	Identified Location (m)	Identified source flux ($kg/m^2/d$)	Objective Function
n=2	31.7	(274.5,549)	(260.8,554.4)	31.4686	0.0911
	31.7	-183,457.50	(140.7,453.1)	31.4686	

5. Conclusion

From the results it is evident that the model is generally able to identify the source characteristics, i.e. the location of the pollutant sources and source flux. There was some deviation in the second source flux location from the actual source location, which could be because of the numerical errors or due to inadequate convergence during optimization.

Reference

Amirabdollahian, M. and Datta B. (2013). Identification of Contaminant Source Characteristics and Monitoring Network Design in Groundwater Aquifers: An Overview. Journal of Environmental Protection, 4, 26-41.

Atmadja, J. and Bagtzoglou, A.C. (2001).Pollution source identification in heterogeneous porous media. Water Resources Research, 37 (8), 2113-2125.

Atmadja, J. and Bagtzoglou, A.C. (2001).State of Art Report on Mathematical Methods for Groundwater Pollution Source Identification. Environmental Forensics, 2, 205-214.

Ayvaz, M.T. (2010). A linked simulation–optimization model for solving the unknown groundwater pollution source identification problems. Journal of Contaminant Hydrology, 117, 46–59.

Bagtzoglou, A.C., Dougherty, D.E., Tompson, A.F.B. (1992). Application of particle methods to reliable identification of groundwater pollution sources. Water Resource Management, 6(1), 15–23.

Bagtzoglou, A.C., Tompson, A.F.B., Dougherty, D.E, (1991). Probabilistic simulation for reliable solute source identification in heterogeneous porous media.Water resources engineering risk assessment, 189–201.

Ball, W.P., Liu, C., Xia, G. and Young, D.F. (1997). A diffusion-based interpretation of tetrachloroethene and trichloroethene concentration profiles in a groundwater aquitard. Water Resources Research, 33 (12), 2741-2757.

Bear, J. (1979). Hydraulics of groundwater.Mc-Graw Hill, New York, 569 pp.

Chaubey, J. and Kashyap, D. (2014).Identification of groundwater pollution source through inverse modelling. International Journal of Scientific Engineering and Technology, Special Issue, 138-142.

Datta, B., Chakrabarty, D., Dhar, A. (2009).Simultaneous identification of unknown groundwater pollution sources and estimation of aquifer parameters. Journal of Hydrology, 376(1–2), 48–57.

Gorelick, S. M. (1982).Optimal Dynamic Management of Groundwater Pollutant Sources. Water Resource Research, 18 (1), 71–76.

Gorelick, S. M., Evans, B. E. and Remson, I. (1983).Identifying sources of groundwater pollution: An optimization approach. Water Resource Research, 19 (3), 779–790.

Jha, M. and Datta, B. (2013).Three-Dimensional Groundwater Contamination Source Identification Using Adaptive Simulated Annealing.Journal of Hydrologic Engineering, 18(3), 307-317.

Jha, M. and Datta, B. (2015).Application of Unknown Groundwater Pollution Source Release History Estimation Methodology to Distributed Sources Incorporating Surface-Groundwater Interactions. Environmental Forensics, 16(2), 143-162.

Liu, C. and Ball, W.P. (1999).Application of inverse methods to contaminant source identification from aquitard diffusion profiles at Dover AFB, Delaware. Water Resources Research, 35 (7), 1975-1985.

Mahar, P.S. and Dutta, B. (2000).Identification of Pollution Sources in Transient Groundwater Systems. Water Resources Management, 14, 209–227.

Mahar, P. S. and Datta, B. (2001).Optimal Identification of Groundwater pollution Sources and Parameter Estimation, Water Resources Planning and Management, 127 (1), 20-29.

Michalak, A.M. and Kitanidis, P.K. (2010).Application of geostatistical inverse modeling to contaminant source identification at Dover AFB, Delaware. Journal of Hydraulic Research, 42, 9-18.

Sidauruk, P., Cheng, A.H.D. and Ouazar, D. (1997).Groundwater Contaminant Source and Transport Parameter Identification by Correlation Coefficient Optimization. Groundwater, 36 (2), 208-214.

Singh, R.,Datta, B., and Jain, A. (2004).Identification of unknown groundwater pollution sources using artificial neural networks. Journal of Water Resource Planning and Management, 130 (6), 506–514.

Singh, R., and Datta, B. (2006).Identification of groundwater pollution sources using ga-based linked simulation optimization model. Journal of Hydrologic Engineering, 11 (2), 101–109.

Skaggs, T.H. and Kabala, Z.J. (1994).Recovering the release history of a groundwater contaminant. Water Resources Research, 30 (1), 71-79.

Skaggs, T.H. and Kabala, Z.J. (1995).Recovering the history of a groundwater contaminant plume: Method of quasi-reversibility. Water Resources Research, 31 (11), 2669-2673.

Skaggs, T.H. and Kabala, Z.J. (1998).Limitations in recovering the history of a groundwater contaminant plume. Journal of Contaminant Hydrology, 33, 347–359.

Snodgrass, F.M. and Kitanidis, P.K. (1997).A geostatistical approach to contaminant source identification. Water Resources Research, 33 (4), 537-546.

Woodbury, A.D. and Ulrych, T.J. (1993).Minimum Relative Entropy: Forward Probabilistic Modeling. Water Resources Research, 29 (8), 2847-2860.

Woodbury, A.D. and Ulrych, T.J. (1996).Minimum relative entropy inversion: Theory and application to recovering the release history of a groundwater contaminant. Water Resources Research, 32 (9), 2671-2681.

Woodbury, A., Sudicky, E., Ulrych, T.J. and Ludwig, R. (1998).Three-dimensional plume source reconstruction

Distributed Surface Water Flow Model For Gopalkheda Sub-Catchment, Maharashtra, India

V. D. Loliyana[1], P. L. Patel[2]

[1]Research Scholar, Centre of Excellence (CoE) on 'Water Resources and Flood Management', Department of Civil Engineering, Sardar Vallabhbhai National Institute of Technology (SVNIT), Surat – 395007, India.

[2]Professor, Department of Civil Engineering, Sardar Vallabhbhai National Institute of Technology (SVNIT), Surat – 395007, India

Email: viraj_nishi@yahoo.com, plpatel@ced.svnit.in, premlalpatel1966@gmail.com

A B S T R A C T

A distributed hydrologic model reflects its predictive capability of surface water flow, thus, making it a very useful tool for effective water management and decision making for hydrological extremes. The objective of present study is aimed to assess the performance of a distributed physics based hydrologic model, MIKE SHE to describe the overland flow processes in Gopalkheda sub-catchment (area of 8647 km2), part of upper Tapi basin, Maharashtra, India. The distributed model is discretized at 500m resolution scale, and calibrated using observed stream flow data for years 1991-2000. The Manning's roughness coefficient for overland flow are selected from existing land use-land cover of the sub-catchment by using their appropriate values from literature. The performance of calibrated model has been validated using independent data for years 2001-2010 and computing statistical performances indices, like correlation of coefficient (r), Nash-Sutcliffe's model efficiency (EF) and root mean square error (RMSE). The estimated values of r, EF and RMSE for aforesaid independent are found to be 0.78, 0.66 and 106.73 m3/s respectively. The performance of the model has been found to be satisfactory, and can be improved further while considering the effects of sub-surface flow and ground water flow explicitly in the model.

Keywords: Distributed physics based approach; MIKE SHE; Calibration; Gopalkheda sub-catchment

1. Introduction

The physics based hydrological models describe rainfall-runoff relationship of a catchment using mass and momentum equations. Presently, rainfall-runoff models are being developed through distributed approach at sub-basin scales, and investigators are extensively using the same due to inherent limitations of lumped models. The lumped hydrological models are not able to provide the precise results over the catchment scale as the physiographic and climatic characteristics are averaged area over the whole catchment. On other hand, distributed hydrological models require extensive sets of surface, subsurface, ground water and climatic data for applying the same in the catchment.

A spatially distributed physically based model comprises an integrated description of both surface and groundwater processes; and requires a large amount of data to perform reliable simulations (Henriksen et al. 2003). While applying a model of this type for large area, it is required to have a rigorous methodology for assessing the model parameters from field database of the catchment. Model calibration is another challenge for rainfall runoff modeling of large watersheds in heterogeneous mountain environment. Many researchers developed and calibrated hydrologic model under lumped manner for rainfall-runoff processes in recent past (Madsen 2003; Henriksen et al. 2003; Rahim et al. 2012; Qin et al. 2013). Madsen (2003) formulated the consistent structure of a distributed and integrated hydrologic model MIKE SHE including three aspects, namely (i) model parameterization and choice of calibration parameters (ii) specification of calibration criteria and (iii) choice of optimization algorithm, using automatic calibration with multiple objectives. The MIKE SHE model was optimized for Danish Karup catchment (440 km2), Denmark. The parameters were selected for calibration for different components of MIKE SHE model, viz. saturated zone (hydraulic conductivity, storage coefficient, and specific yield), unsaturated zone (saturated hydraulic conductivity, conductivity curve parameters, and retention rate parameters), surface zone (Manning's M, detention storage), river system (river bed resistance, leakage coefficient for river aquifer) and drainage system (time constant for drainage routing) and evapotranspiration (evapotranspiration parameters). They estimated upper and lower limits of 12 model parameters by automatic calibration with shuffle complex evolution (SCE) algorithm. The model was calibrated for 1971 – 1974 storm events and validated for 1975 – 1977 storm events with comparison of observed and simulated data based on bias and RMSE performance statistics. Further, Dai et al. (2010) evaluated the performance of MIKE SHE model using Bi-criterion analysis for a Santee experimental forest, South Carolina, USA (area of 160 ha). Bi-criteria approach used for constructing uncertainty bounds of model parameters to simulate the hydrological processes, i.e., stream flow and ground water flow. Rahim et al. (2012) estimated total water balance of Paya Indah wetlands watershed, Malasiya using fully distributed MIKE SHE modeling system. Their model results revealed that the overall water balance is predominantly controlled by climate variables. The estimated values of ET and overland flow (in terms of rainfall) were 65% and 58%; 12.38% and 12.30% during calibration and validation stages respectively. Qin et al. (2013) implicated sustainable water management using integrated hydrologic MIKE SHE for North China Plain (area 1,40,000 km2). The developed model was calibrated for a period of 2000-2005 and validated for period of 2006-2008 for observations of groundwater heads and stream flows on monthly and annual scales. They successfully simulated the hydrologic processes, i.e., evapotranspiration (ET), groundwater flow and surface water flow of the catchment. They also suggested from the obtained results, to demonstrate an effective approach to improve surface and subsurface water use efficiency in the North China Plain by presenting changes in cropping and water saving technologies.

In present study, an attempt has been made to model the rainfall-runoff processes of Gopalkheda sub-catchment of Purna River Basin, Maharashtra, India using MIKE SHE and relevant inputs (in terms of physiographic and climate) for GIS tool. The calibration of the model parameters has been undertaken using the data of year 1991-2000, and validated the same for the data of years 2001-2010 on daily as well as monthly time scale.

2. Study area

The present study area includes Gopalkheda subcatchment (Purna River up to Gopalkheda gauging station) of Tapi basin measuring an area of 8647 km2, draining through Madhya Pradesh and Maharashtra states. The Purna River originates from Gawilgarh steep mountains of eastern Satpura range (Betul district of Madhya Pradesh); and meets the Tapi River just upstream of Hatnur dam. The other four tributaries drain into the Purna River (length ≈ 229.66 km) up to Gopalkheda gauging station, see Fig. 1. The south-west monsoon sets in by the middle of June and withdraws by mid-October having an average rainfall of 750-800 mm. The minimum and maximum temperatures vary from 10^0C to 15^0C and from 38^0C to 48^0C, respectively in the Gopalkheda sub-catchment. The main topographical features of the study area is included in Table 1. There are eleven raingauge stations, one discharge-sediment gauging station in Gopalkheda catchment. Fig. 1 shows the location of raingauges, discharge-sediment gauging station in the Gopalkheda catchment. Brief description of the data being used in present study are included in Table 2.

Table 1 Physiography of the Gopalkheda sub-catchment

Features	Description
Latitude	$21^0\ 38^{'}$ N
Area	8647.00 km^2
Perimeter	513.50 km
Mean altitude	393.81 m
Maximum altitude	1162 m
Minimum altitude (at outlet of catchment)	240 m
Length of main river	229.66 km
Average slope of main river	0.00414

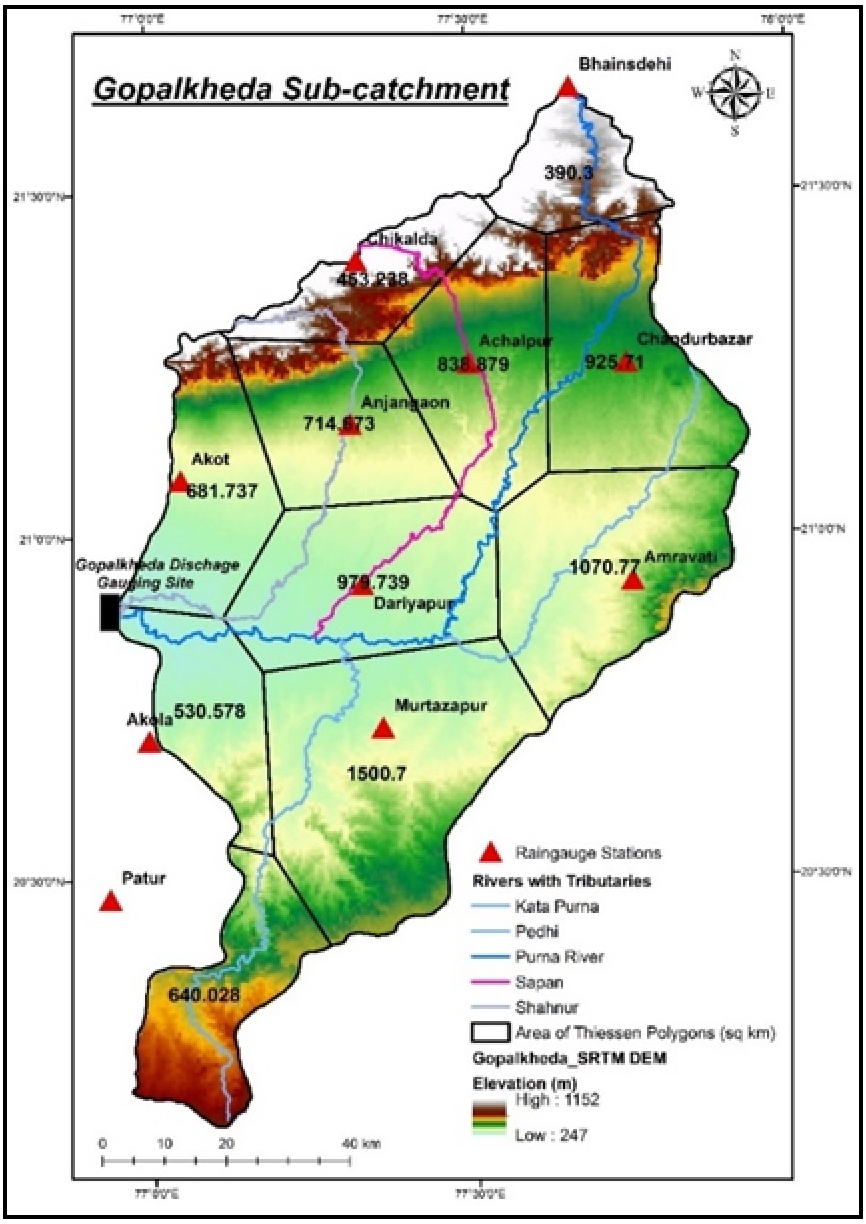


Figure 1 Index map of Gopalkheda sub-catchment

Table 2 Data inputs for model development

Sr. No.	Type of Data	Description of Data	Remarks
1	Satellite data	(1) SRTM DEM (500 m X 500 m)	(1) For catchment delineation
		(2) IRS P6 LISS III – Imagery (2009)	(2) For land use/cover classification
2	Topographic data	(1) Topographic sheets	(1) For digitizing river network
3	Hydrologic data	(1) Rainfall data	For calibration: 1991-2000
		(2) Discharge data	For validation: 2001-2010

3. Description of MIKE SHE model

MIKE SHE (Système Hydrologique Europeén) is a fully distributed, physics based, deterministic hydrological model, capable of both continuous and single event analysis (DHI, 2012). The schematic of hydrologic processes included in MIKE SHE are shown in Fig. 2. The process starts with precipitation input, a fraction of which is intercepted by vegetation before it reaches the surface. The intercepted precipitation is either stored on the plant material or later evaporated back into the atmosphere or depending on soil conditions detained on the soil surface where it can undergo surface runoff or infiltration. As infiltration continues, the unsaturated zone will be saturated, and after all surface storage areas are filled up, overland flow will begin from one cell to the next based on topographic features. As the process goes on, moisture from the unsaturated zone is transferred to the saturated zone at a rate dependent upon soil parameters within the vertical soil profile. The saturated zone flow can either continue down as deep percolation or flow laterally as subsurface drainage, or interflow. The lateral flow would eventually contribute to inflow to surface lakes and streams, and also contribute to groundwater in the form of recharge to the saturated zone. MIKE SHE simulates aforesaid processes, and give final output as total water balance of the basin. In present study, an attempt has been made to simulate the surface flow (overland flow) process only.

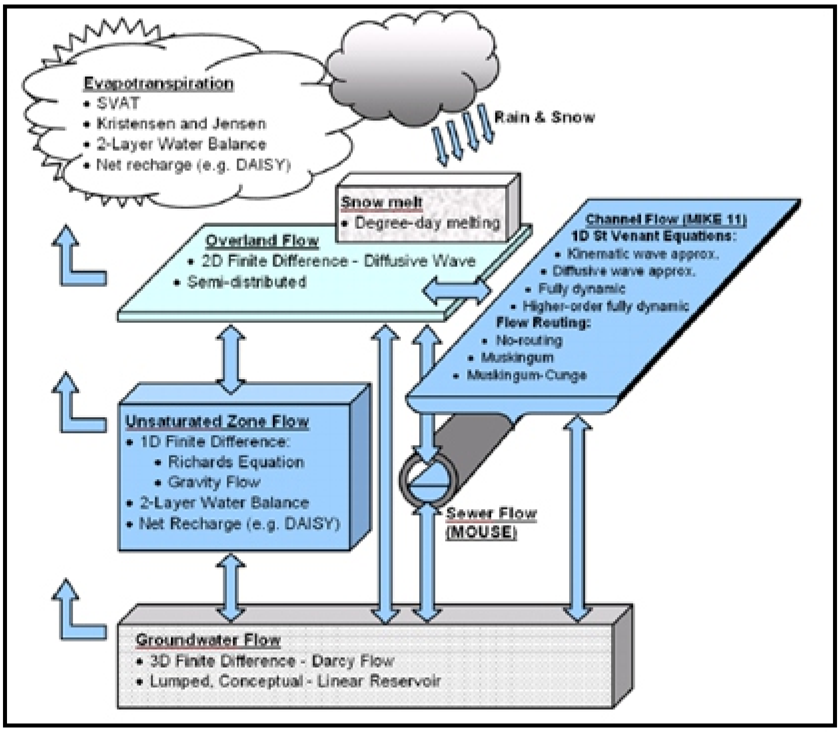


Figure 2 Mathematical schematic diagram of MIKE SHE model (Refsgaard and Storm, 1995)

The prominent parameters in the overland flow analysis of MIKE SHE are included in Table 3.

Table 3 MIKE SHE model parameters (DHI, 2012)

Parameters	Description
Manning's M	Roughness coefficient, which is inversely proportional to Manning's n
Net Rainfall Fraction	Fraction of rainfall that is available for infiltration and overland flow
Detention Storage	Limit the amount of water can flow over ground surface

4. Model development

Rainfall-runoff model has been developed for Gopalkheda sub-catchment using MIKE SHE model. The SRTM DEM data was used to get flow direction in the catchment area using spatial hydrologic tool in ArcGIS 10.0. The model domain discretized as 500 m x 500 m spatial resoulution. Land use/ cover were classified using supervised classification for the study area using ERDAS Imagine 10 package. The satellite imagery of study area were classified into twelve classes as shown in Fig. 3. Rainfall data for 20 years (1991-2010) were used as inputs to simulate the runoff at the outlet of the catchment. Missing data analysis was also carried out for rainfall data by inverse distance method (Subramanya, 2011), to estimate the rainfall of missing period at relevant gauging station using Eq. (1).

$$P_i = \frac{\sum_{i-1}^{n} \frac{R_i}{{d_i}^2}}{\sum_{i-1}^{n} \frac{1}{{d_i}^2}} \qquad (1)$$

where, P_i = missing rainfall value at missing raingauge station,

R_i = known rainfall value at surrounding raingauge stations,

d_i = distance between missing raingauge station and surrounding raingauge station

The 2D overland flow has been simulated using finite difference method (using successive over relaxation solver) by solving St. Venant's equation under diffusive wave conditions. The MIKE SHE parameters (see Table 3), depending upon land use-land cover, soil characteristics of the sub-catchment, were estimated according to hydrological or physical interpretation, and refereeing relevant literature on the same. The calibration and validation of model parameters are described in succeeding paragraphs. The two calibration parameters, net rainfall fraction and detention storage were calibrated by taking several trial simulations for the period of 1991-2000. These parameters have been validated during the time period of 2001-2010 for the same sub-catchment.

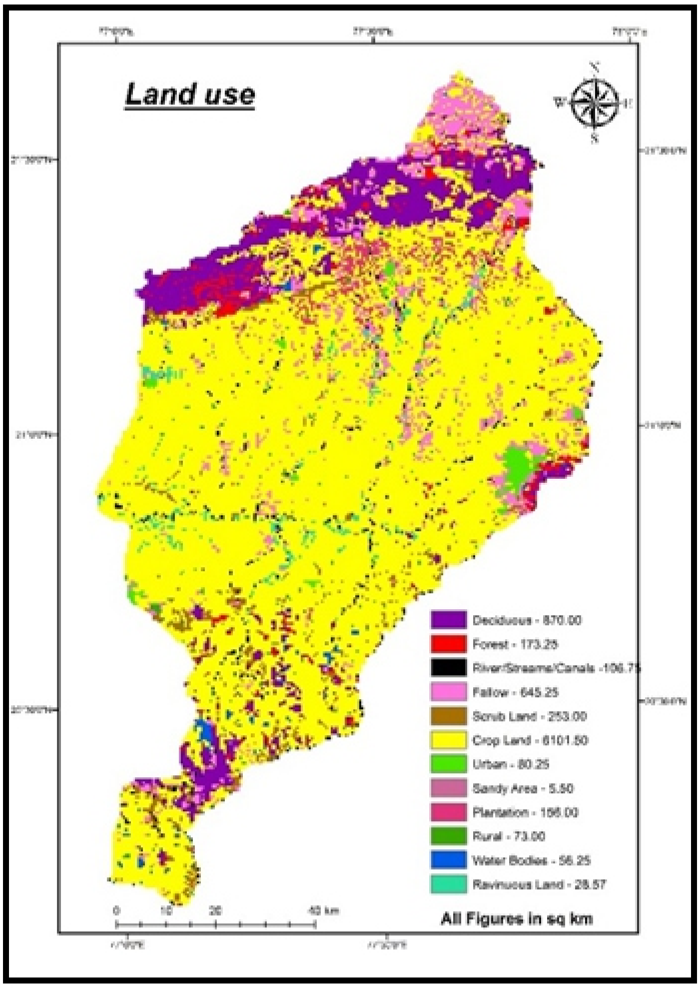


Figure 3 Land use-land cover of Gopalkheda sub-catchment

5. Calibration

For calibration of the model, the values of the model parameters are selected such that model simulates the observed hydrological parameters in catchment as closely as possible. The two parameters, viz., net rainfall fraction and detention storage, are calibrated while third parameter, i.e., Manning's M, has been selected from the literature depending upon land use-land cover in the catchment. The calibration of rainfall-runoff SHE model is based on the entire drainage area contributing to the Gopalkheda gauging station, and is accomplished by adjusting aforesaid two parameters. The calibration of the model has been accomplished by optimizing two objective functions: (i) agreement between the average simulated and observed runoff volume from the catchment (i.e. minimizing the water balance error, %WBL) and (ii) overall agreement of the shape of the hydrograph, by minimizing the overall root mean square error (RMSE) for the year 1991-2000 on daily time scale. Table 4 shows the Manning's M values for different land use classes taken from the literature (Chow et al., 1988; Engman, 1986; Kothyari et al. 2010; Vieux, 2001; Rawls and Brakensiek, 1989).

Table 4 Manning's M for different land use classes

Land use Classes	Manning’s M
Deciduous	6.67
Forest	6.67
Rivers/Stream/Canals	35.71
Fallow	25
Scrub Land	18.18
Crop Land	28.57
Urban	6.67
Sandy Area	25
Plantation	25
Rural	6.67
Water Bodies	35.71
Gullied/ Ravinuous Land	28.57

Table 5 MIKE SHE calibrated parameters

SHE Model Parameters	Initial Values	Calibrated Values	%WBL error	RMSE (m3/s)
Net Rainfall Fraction	0.2	0.63	7.12%	101.37
Detention Storage (mm)	0.5	5.6		

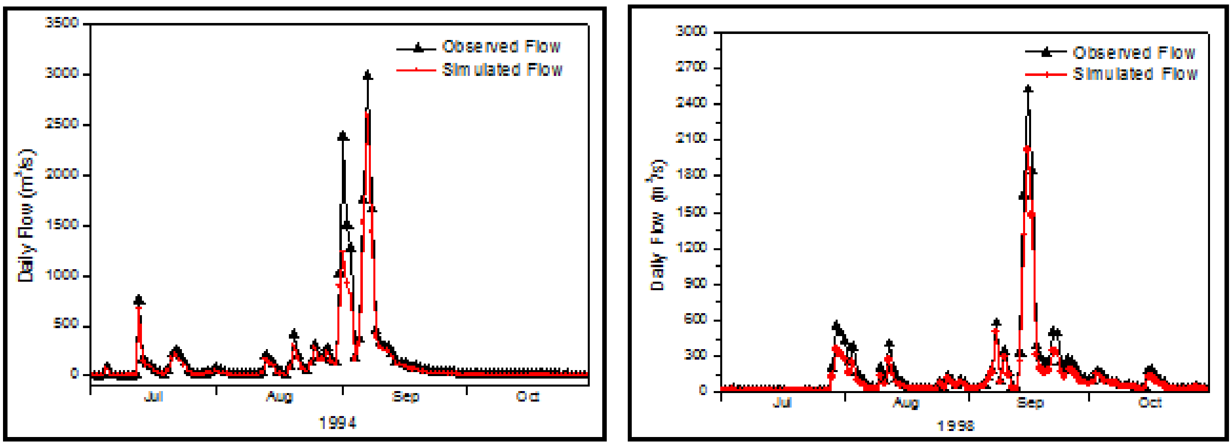


Figure 4 Comparison of observed and simulated runoff under calibration period (a) 1994 (b) 1998

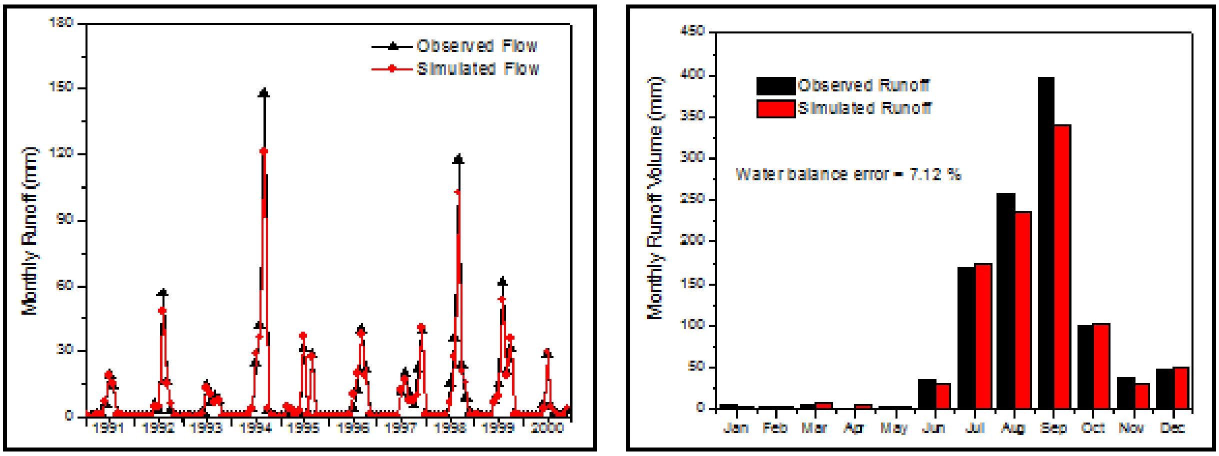


Figure 5 Comparison of observed and simulated runoff under calibration (a) Monthly Runoff (b) Monthly Runoff Volume

The statistical parameters during the calibration stage, i.e. % water balance and RMSE, are included in Table 5. From Table 5, it is revealed that calibrated values of rainfall fraction and detention storage are 0.63 and 5.6 mm respectively. The performance of model under calibration for years 1994 and 1998 at daily time scale are shown in Fig. 4 (a) and (b) respectively. The model invariably under predicts the large peak, may be due to non-consideration of subsurface and base flow in the model results. Also, the simulated results, at calibration stage, are compared with observed discharge at Gopalkheda gauging site on monthly scale, see Fig. 5 (a). Invariably, both simulated and observed discharges are in close agreement. The month wise simulated vs observed % water balance are shown in Fig. 5(b). The observed runoff volume during the monsoon, are invariably higher than simulated runoff because the simulated runoff does not account explicitly the base flow and sub-surface flow from the system.

6. Validation of Calibrated Model

The calibrated model for the parameters described in the preceding section are validated using independent observed data at Gopalkheda gauging site for period 2001-2010. The performance of the model has been assessed using statistical performance indices, like Root mean square error (RMSE), Coefficient of correlation (r) and Nash-Sutcliffe efficiency (NSE), see Table 6. The quantitative performance of the model based on aforesaid performance indices is included in Table 7. The statistical performance indices reported for the model validation, i.e. r = 0.78 and NSE = 0.66 for daily time scale and r = 0.90 and NSE = 0.84 for monthly time scale, are within satisfactory range. Also, the hydrographs plotted on daily and monthly scales are shown in Figs. 6 and 7 (a) respectively. Invariably, the performance of the model is in line with those reported for calibration stage, i.e. the calibration model under predicts the peak discharges for both daily and monthly time scales. Also, the simulated % water balance are perpetually lower from observed runoff volume, see Fig. 7(b).

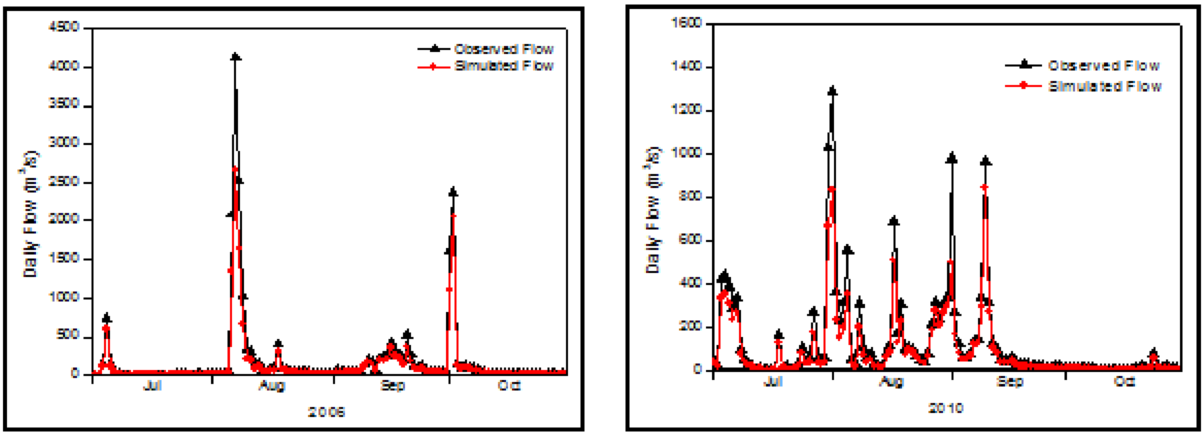


Figure 6 Comparison of observed and simulated runoff under validation period (a) 2006 (b) 2010

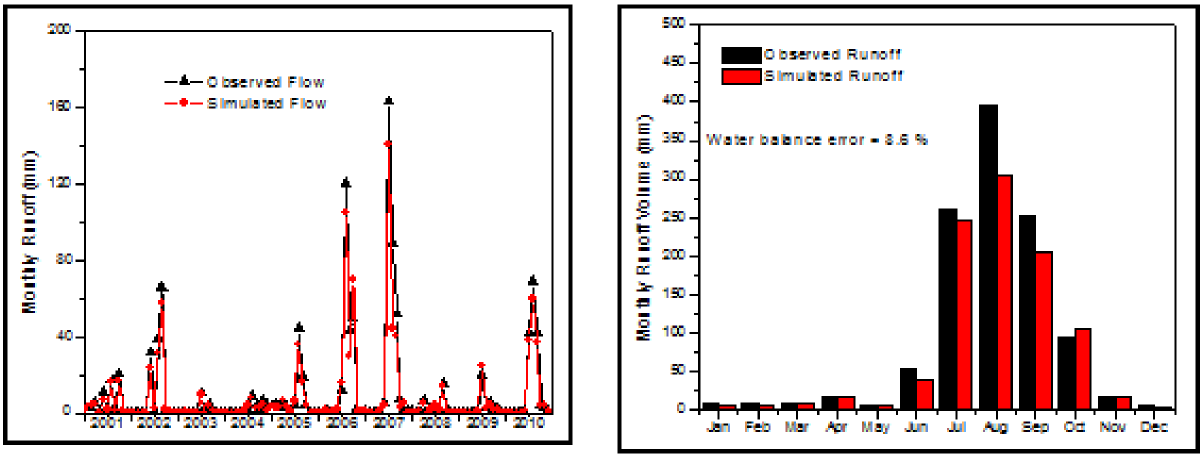


Figure 7 Comparison of observed and simulated runoff under validation (a) Monthly Runoff (b) Monthly Runoff Volume

Table 6 Description of performance indicators (Dai et al. 2010)

Sr. No.	Statistical Indices	Description	Equation	Acceptable Value
1	Root Mean Square Error (RMSE)	Measure of scatter of residuals (flow)	$RMSE = \left[\frac{\sum_{i=1}^{n}\left(Q_{sim_i} - Q_{obs_i}\right)^2}{n}\right]^{1/2}$	0
2	Coefficient of Correlation (r)	Measure of correlation between observed and simulated flows	$r = \left\{\frac{\sum_{i=1}^{n}\left(Q_{obs_i}-\bar{Q}_{obs}\right)\left(Q_{sim_i}-\bar{Q}_{sim}\right)}{\sqrt{\left[\sum_{i=1}^{n}\left(Q_{obs_i}-\bar{Q}_{obs}\right)^2\right]}\sqrt{\left[\sum_{i=1}^{n}\left(Q_{sim_i}-\bar{Q}_{sim}\right)^2\right]}}\right\}$	1
3	Nash-Sutcliffe Efficiency (NSE)	Ratio of mean square error to the variance of observed flow subtracted from unity	$NSE = 1 - \frac{\sum_{i=1}^{n}\left(Q_{sim_i} - Q_{obs_i}\right)^2}{\sum_{i=1}^{n}\left(Q_{obs_i} - \bar{Q}_{obs}\right)^2}$	1

Table 7 Performance of developed model under calibration and validation

Performance Indicators	Calibration (Jan 01, 1991 – Dec 31, 2000)				Validation (Jan 01, 2001 – Dec 31, 2010)			
	Daily Flow	Unit	Monthly Runoff	Unit	Daily Flow	Unit	Monthly Runoff	Unit
RMSE	101.37	m^3/s	11.4	mm	106.73	m^3/s	12.67	mm
r	0.82	-	0.93	-	0.78	-	0.9	-
NSE	0.67	-	0.86	-	0.66	-	0.84	-

7. Conclusions

For Purna River basin, regional (Gopalkhea sub-catchment = 8647 km2) model has been developed using distributed MIKE SHE. The model has been calibrated against ten years (1991-2000) hydrologic data and validated for the data of another ten years (2001-2010).

(a) The statistical performance indices have been used to assess the performance of model while comparing the observed and simulated runoff at Gopalkheda gauging station. There are 7.12 % and 8.6% of water balance errors for calibration and validation stages respectively, which indicates that developed model gives acceptable results for modelling the overland flow.

(b) The calibrated parameters of the model, i.e. rainfall fraction and detention storage are 0.63 and 5.6 mm respectively. The Manning's resistance parameters, M, has been selected based on land use-land cover classification of the catchment.

(c) The simulated discharges invariably under predicts the peak discharges may be due to non-consideration of subsurface and base flow parameters explicitly in the modelling of runoff.

(d) The performance of calibrated model has also been quantified in terms of statistical performance indices for independent data sets. The NSE and r values for independent data sets are 0.84 and 0.90 respectively for monthly time scale, which indicates reasonably good agreement of calibrated model with observed data sets.

Acknowledgement

Authors are thankful to MHRD-NPIU-TEQIP-II for providing the funding through a Centre of Excellence (COE) Project on 'Water Resources and Flood Management' at SVNIT Surat under which present investigation was undertaken.

Reference

Chow, V.T., Maidment, D. R., and Mays, L. W. (1988). Applied Hydrology. Mcgraw Hill, New York.

Dai, Z., Li, C., Trettin, C., Sun, G., Amatya, D., and Li, H. (2010). Bi-criteria evaluation of the MIKE SHE model for a forested watershed on the South Carolina coastal plain. Hydrologic Earth System Sciences, 14, 1033-1046.

DHI. (2012). MIKE SHE reference manual. MIKE by DHI, Denmark.

Engman, E. T. (1986). Roughness coefficients for routing surface runoff. Journal of Irrigation Drainage Engineering, 112(1), 39-53.

Henriksen, H. J., Troldborg, L., Nyegaard, P., Sonneborg, T. O., Refsgaard, J. C., and Madsen, B. (2003). Methodology for construction, calibration and validation of a national hydrological model for Denmark. Journal of Hydrology, 280, pp. 52-71.

Kothyari, U. C., Raamsankaran, Raaj, Satish Kumar, D., Ghosh, S. K., and Mendiratta, N. (2010). Geospatial based automated watershed modeling in Garhwal Himalaya. Journal of Hydroinformatics, 12(4), 502-520.

Madsen, H. (2003). Parameter estimation in distributed hydrological catchment modelling using automatic calibration with multiple objectives. Advances in Water Resources, 26(2), 205-216.

Qin, H., Cao, G., Kristensen, M., Refsgaard, J. C., Rasmussen, M. O., He, X., Liu, J., Shu, Y., and Zheng, C. (2013). Integrated hydrological modeling of the Norh China Plain and implications for sustainable water management. Hydrologic Earth System Sciences, 17, 3759-3778.

Rahim, B-e. E. A., Yusoff, I., Jafri, A. M., Othman, Z., and Ghani, A. A. (2012). Application of MIKE SHE modelling system to set up a detailed water balance computation. Water and Environment Journal, 26, 490-503.

Rawls, W. J., and Brakensiek, D. L. (1989). Estimation of soil water retention and hydraulic properties. In: Morel-Seytoux (ed.), Unsaturated flow in hydrologic modeling: Theory and practice. Kluwer Academic Publisher, Boston, 275-300.

Subramanya, K. (2011). Engineering Hydrology. Tata McGraw-Hill Publishing Company Limited, New Delhi.

Vieux, B. E. (2001). Distributed hydrologic modeling using GIS. Kluwer Academic Publishers, Dordrecht, The Netherlands.

Assessment Of Regional Flood Frequency Model For The Upper Tapi Basin, India

Resmi S R[1], P L Patel[2]

[1]M. Tech Scholar, Center of Excellence on 'Water Resources and Flood Management', SardarVallabhbhai National Institute of TechnologySurat, Surat-395007, India

[2]Professor, Department of Civil Engineering, SardarVallabhbhai National Institute of TechnologySurat, Surat-395007, India

Email: iiresmii@gmail.com, plpatel@ced.svnit.ac.in

A B S T R A C T

In present study, annual maximum discharge data of five stream gauging stations of Upper Tapi basin, viz. Lakhpuri, Gopalkheda, Yerli, Dedtalai and Burhanpur, are used to develop regional flood frequency model of the study area. The data were checked for time independency using the lag-one autocorrelation coefficient; time homogeneity using the cumulative deviations test;and outliersusing the G-B outlier test. Screening of the data has also been undertaken using the Discordancy measure (Di) test,and the regional homogeneity has been measured in terms of Hosking and Wallis' L-moment based heterogeneity measure (H), whilecarrying out 500 simulations using the four-parameter Kappa distribution. The data of all the five stations have been found to constitute a homogenous region. The comparative regional flood frequency analysis has been carried out for the region using the L-moment based three-parameter frequency distributions, like generalized extreme value (GEV), lognormal (LN-III),Pearson type-III (P-III),generalized logistic (GLO) andgeneralized Pareto (GP) distributions. The parameters of the distributions were calculated using the regional L-moment ratios as given by Hosking and Wallis (1997). Based on the L-moment ratio diagram, and the L-moment goodness-of-fit statistics, GEV has been identified as the most robust distribution for the study area. The parameters of the GEV distribution as obtained from the analysis arelocation parameter, =0.6079, scale parameter, =0.4721 and shape parameter, k =0.2062. The regional flood frequency relationship developed would be helpfulin estimating the flood quantiles corresponding to desired return periodsin the upper Tapi basin, India.

Keywords:Regional flood frequency analysis, L-moments, Regional homogeneity, Kappa distribution,Goodness-of-fit, Regional flood frequency relationship, GEV

1. Introduction

Information on flood magnitudes and their associated probability of occurrence is often required for the design of hydraulic structures like dams, spillways, culverts, road and railway bridges, flood plain zoning, urban drainage system, economic evaluation of flood protection projects etc. The most widely used methods for flood estimation are rational method, U.S. Soil Conservation method and regional flood frequency methods (Pilgrim and Cordey 1992). The availability of data and the design criteria adopted for the hydraulic structure plays a very important role in the choice of flood estimation method.

Reliable flood estimates can directly be derived if river flow records are available at or near the site of interest. However, such flood derivation becomes really difficult asand when the reliable flood records are not available at the proposed project site. In India, frequency based flood estimation approaches are being used for the design of new hydraulic structures. The primary objective of regional flood frequency method is to develop flood formulae for the region which gives fairly good flood estimates for the sites included in it.

The L-moments are a set of statistics used to describe the shape of probability distributions (Hosking 1990). They are linear combinations of L-statistics, which are analogous to conventional moments like standard deviation, skewness and kurtosis, known as L-scale, L-skewness and L-kurtosis respectively (L-mean is same as that of the conventional mean). The L-moment ratios are also known as standardized L-moments. A theoretical distribution has a set of population L-moments. Sample L-moments can be calculated for a sample, and can be used as estimators of the population L-moments.L-moments are the modifications of the probability weighted moments (PWM's) introduced by [illegible]

$$\boldsymbol{\beta_r = E[x\{F(x)\}^r] = \int_0^1 x(F)F^r dF} \tag{1}$$

where, $F = F(x)$ is the cumulative distribution function (CDF) for x, $x(F)$ is the inverse CDF of x evaluated at the probability F, and $r = 0, 1, 2, \ldots$, is a non-negative integer. When $r = 0$, β_0 is equal to the mean of the distribution $\mu = E[x]$. Hosking (1990) established a relationship between the r^{th} L-moment, $\lambda r, and$ the r^{th} PWM for any probability distribution as

$$\boldsymbol{\lambda_{r+1} = \sum_{k=0}^{r} \beta_k (-1)^{r-k} \binom{r}{k} \binom{r+k}{k}} \tag{2}$$

The L-moment ratios as defined by Hosking (1990) are:
L-coefficient of variation, L-CV = $\boldsymbol{\lambda_2/\lambda_1}$ **(3)**
L-coefficient of skewness, L-skew = $\boldsymbol{\lambda_3/\lambda_2}$ **(4)**
L-coefficient of skewness, L-skew = $\boldsymbol{\lambda_4/\lambda_2}$ **(5)**

Burn et al. (2000) developed a new technique to delineate homogenous basins for a region. The methodology begins with the k-means clustering algorithm coupled with the weighted dissimilarity measure and Euclidean distance to define the initial regions. These regions are then tested for homogeneity using the Hosking and Wallis Heterogeneity measure (H). The findings of the study was finally applied and were found to be very effective in Indian conditions. Kumar et al. (2003) developed regional flood frequency relationship for Middle Ganga Plains Subzone 1(f) of India using a GEV index-flood procedure while incorporating the method of L-moments. The L-moment ratio diagram and ZDIST statistic showed that GEV was the best fit distribution of the region. Based on the results they developed regional frequency relationship for gauged as well as ungauged catchments. Hussainet al.

discordancy measure was calculated to screen the data; heterogeneity measure was obtained by conducting simulation experiments using the 4-parameter Kappa distribution and the best distribution was chosen using the L-moment ratio diagram and the ZDIST statistics. Saf (2010) assessed the reliability of the probability distribution assigned to the region without taking into consideration the effect of discordancy. However, flood frequency studies undertaken further showed that the results were much better when the data was screened using the discordancy measure. Hussain (2011) performed an L-moment based regional flood frequency analysis to the annual maximum peak flows observed at seven stations located on the main stream of the Indus River, Pakistan. The consistency tests like run test, lag-one autocorrelation analysis, G-B outlier test and Discordancy measure test were performed to make sure that the data was fit for theanalyses. The robust distribution for the regions were finally obtained by performing accuracy checks as described by Hosking and Wallis (1997).It could be observed that most of these studies did not give ample focus on the time homogeneity and time independency of individual stations which are the major requirements for performing flood frequency analysis.

In present study, a regional approach for flood frequency analysis at five stream-gauging stations,viz., Lakhpuri,Gopalkheda, Yerli, Dedtalai and Burhanpur, of Upper Tapi basin has been undertaken. The fundamental idea is as follows. By making use of the annual maximum discharge data at each stations, a frequency distribution model is developed for estimating flood peaks for different return periods using an appropriate probability distribution and the method of L-moments. Before performing flood frequency analysis, data were checked for time homogeneity, time independency, outliers and discordancy. Finally, the developed regional flood frequency model has been used to obtain fairly reasonable flood estimate of the region as a whole.

2. Studyarea

The Tapi River basin situated in the northern part of Deccan plateau covers an area of 65,145 km^2 which constitutes 2% of India's total geographic area. The basin lies in the states of Maharashtra (51,504 km^2), Madhya Pradesh (9,804 km^2) and Gujarat (3,837 km^2). The basin has two well defined physiographic regions, (i) hilly regions which is covered with forest areas taking up 25% of the basin area and (ii) plain areas which are characterized by the presence of black soil.

The river basin is divided into three major sub-basins:

- **Upper Tapi Basin:** It covers an area of 29,430 km^2 of the total basin area, up to Hathnur reservoir which is just downstream of Purna River (one of the major tributaries of Tapi River).

- **Middle Tapi Basin:** It starts from Hathnur dam and ends at Gidhade gauge-discharge site with a total area of 25,320 km^2.

- **Lower Tapi Basin:** It constitutes an area of 10,395 km^2 and ends at the Arabian Sea.

The study area covered in present work includes Upper Tapi basin (Fig. 1) with its outlet at Hathnur dam. It lies between 20 15' 00" N to 22 45' 00" N latitudes and 75 45' 00" E to 78 30' 00" E longitudes. The basin has varying altitude from 752 m near Multai to about 207 m near Hathnur dam. The basin covers two sub-catchments, viz., Purna and Burhanpursub-catchments. There are five gauge-discharge stations, i.e., Dedtalai and Burhanpur in Burhanpur sub-catchment; and Lakhpuri, Gopalkheda and Yerli in Purna sub-catchment, see Fig.1.

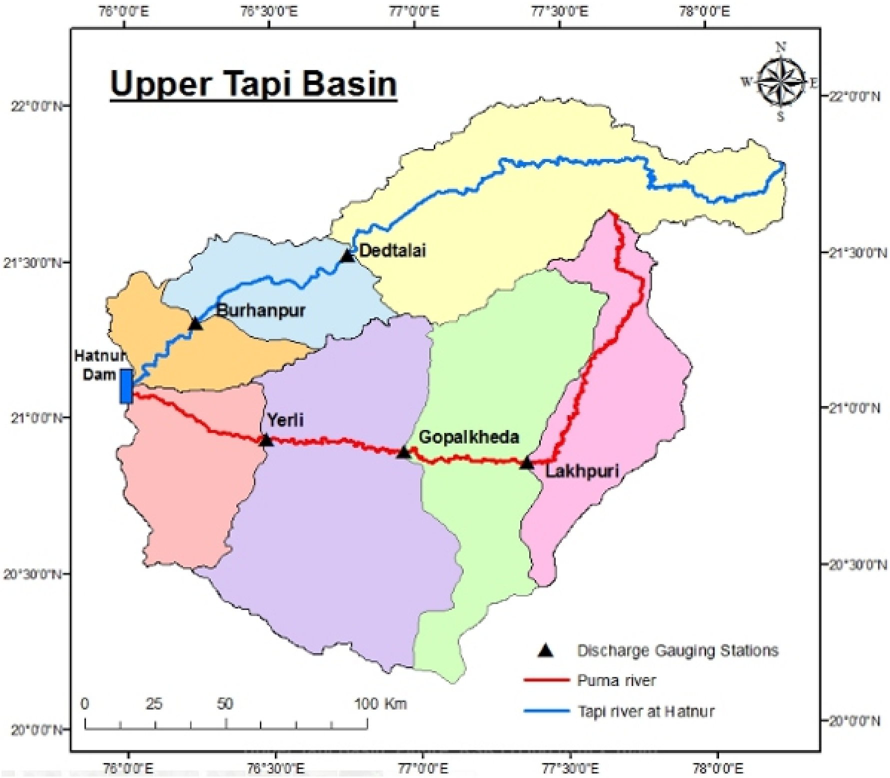


Fig.1Index map of Upper Tapi Basin

3. Data Availability

The daily discharge data for Lakhpuri, Gopalkheda, Yerli, Dedtalai and Burhanpur gauging stations were available for period 1978-2011 from Central Water Commission (CWC), the government agency responsible for gauging the rivers in India. Annual maximum series of discharge were prepared from the data available for aforesaid gauging stations. Locations of each station and the corresponding drainage area, in km^2, were also collected from CWC. The statistical parameters of discharge data at five gauging stations along with locations and drainage areas, in km^2, are presented in Table 1.

Table 1 Statistical parameters of the gauge-discharge stations

Site	Period	$\bar{Q}$ (m^3/s)	*SD*, σ (m^3/s)	*CV*	*CS*	A (km^2)	Latitude	Longitude
Lakhpuri	1978-2011	907.8	630.3	0.69	1.28	3564	20° 50' 44"	77° 21' 38"
Gopalkheda	1978-2011	1425.5	1059.7	0.74	0.82	8724	20° 52' 27"	76° 59' 29"
Yerli	1978-2011	2692	2331.8	0.87	1.76	15960	20° 56' 09"	76° 28' 33"
Dedtalai	1978-2011	4528.2	3568.6	0.79	2.34	6786	21° 30' 47"	76° 45' 26"
Burhanpur	1978-2011	7997.3	7448.8	0.93	2	9018	20° 17' 54"	76° 14' 10"

$\bar{Q}$ - Mean Annual Peak Flood, SD – Standard Deviation, CV – Coefficient of Variation, CS – Coefficient of Skewness, A – Catchment Area

4. Data Analysis And Results

Five three-parameter distributions, viz., Generalised extreme value (GEV), lognormal (LN-III), Generalised logistic (GLO), Pearson Type-III (P-III) and Generalised Pareto (GPA) distributions, were selected as candidate distributions to develop regional flood frequency relationship for the Upper Tapi River Basin, India. The parameters of these distributions were obtained using the L-moments' approach as described in Hosking and Wallis (1997). In order to ascertain suitability of available data in estimating floods, statistical tests for time independency, time homogeneity and outliers were carried out. Screening of the data was also undertaken using the Discordancy measure (Di) test. All the five stations are considered as a single entity while measuring the homogeneity using the Hosking and Wallis (1997) heterogeneity measure (H). These statistical tests, along with the development of regional flood frequency model are discussed in succeeding paragraphs.

4.1 Test for independency (lag-one autocorrelation)

The time independency of annual maximum discharge series for all five gauging sites were undertaken by computing lag-one autocorrelation (r1) using Eq. (6). The computed values of r1for Lakhpuri, Gopalkheda, Yerli, Dedtalai and Burhanpur gauging sites are -0.100, -0.013, -0.017, 0.004 and 0.027 respectively. Also, upper and lower threshold limits of lag-one autocorrelation at 5% significance level, for 34 years of data were computed from Eq. (7), and have been found to be 0.306 and -0.366 respectively.

$$r_1 = \frac{\frac{1}{n}\sum_{t=1}^{n-1}(X_t - \bar{X})(X_{t+1} - \bar{X})}{\frac{1}{n-1}\sum_{t=1}^{n}(X_t - \bar{X})^2} \qquad (6)$$

where, $\bar{X} = \frac{1}{n}\sum_{t=1}^{n} X_t$,is the overall mean of sample data points, (X_t), with n observations.

$$r_1(5\%) = \frac{-1 \pm 1.96\sqrt{n-2}}{(n-1)} \qquad (7)$$

As calculated values of lag-one auto correlation coefficients for all stations lie within threshold limits (0.306 to -0.366), it is inferred that annual maximum discharge series at all stations in the study area come from a population that is time-independent.

4.2. Test for time homogeneity (cumulative deviations test)

To ascertain whether observed discharge data at all discharge-gauging stations come from populations that had not undergone any serious change over time, homogeneity test was conducted. In the present study, a non-parametric cumulative deviations test is adopted. The test statistics are calculated as per Eq. (8)and compared with critical value at 5% level of significance.

$$\frac{Q}{\sqrt{n}} = \frac{\max\limits_{0\le k\le n}\left|\frac{\sum_{t=1}^{k}(X_t-\bar{X})}{\sigma}\right|}{\sqrt{n}} \qquad (8)$$

Here, is the sample standard deviation. The critical value of test statistic for a sample of size 34 at 5% level of significance is 1.124 (Machiwal and Jha 2012). Values of $\frac{Q}{\sqrt{n}}$ (estimated from Eq. (8) ranges from 0.532 to 1.036 for all the five sites in the basin, and are less than the critical value. Hence, it is concluded that data come from the homogenous populations.

4.3 Test for outliers (G-B test)

The Grubbs and Beck (G-B) test has been used for the detection of outliers in the data series. The G-B test statistics (Kn) at 10% level of significance for a sample size 34 was calculated using Eq. (9) and the corresponding higher and lower threshold values for the log transformed series were calculated using Eqs.(10) and(11) respectively for all the stations. The result indicated that neither high outliers nor low outliers are present in the discharge data of all the stations, i.e. all the discharge values fall within the threshold bandsofrespective stations.

$$K_n = -3.62201 + 6.2846n^{\frac{1}{4}} - 2.49835n^{\frac{1}{2}} + 0.491436n^{\frac{3}{4}} - 0.037911 \qquad (9)$$

$$y_H = \bar{y} + K_n S_y \qquad (10)$$

$$y_L = \bar{y} - K_n S_y \qquad (11)$$

4.4. Screening of data using discordancy measure test

The discordancy measure (D_i) of i[th] site is an important parameter being used to ascertain the suitability of the site in performing regional flood frequency analyses. The discordancy measure, Di, for site i is defined as:

$$D_i = \frac{1}{3}N(u_i - \bar{u})^T {A_s}^{-1}(u_i - \bar{u}) \qquad (12)$$

Here, N is the number of sites in the region; ui= [$t^{(i)}$ $t_3^{(i)}$ $t_4^{(i)}$] T is a vector containing L-moment ratios of site I; superscript T denotes the transpose of the vector;

$$\bar{u} = N^{-1} \sum_{i=1}^{N} u_i \qquad (13)$$

A_s is the matrix of sums of squares and cross products and is defined as:

$$A_s = \sum_{i=1}^{N} (u_i - \bar{u})(u_i - \bar{u})^T \qquad (14)$$

The site is said to be discordant, if the value of D_i exceeds the critical value of discordancy statistics at 10% significance level, as given elsewhere (Hosking and Wallis 1997).The value of critical discordancy measure corresponding to 5 sites is 1.6480. The individual discordancy measures ranges from 0.501 to 1.276 which is less than the critical value. Hence, it is concluded that all the sites are non-discordant, and fit for regional flood frequency analysis.

4.5. Regional Homogeneity Test

To assess whether the given region has to be divided into number of homogenous sub-regions, regional homogeneity test, as described by Hosking and Wallis (1993),is undertaken. The test statistics, H, compares the inter-site variations in sample L-moments for the group of sites with what would be expected of a homogeneous region. The inter-site variation of L-moment ratio is measured as the standard deviation (V) of the at-site L-CV's weighted proportionally to the record length at each site.

Considering the whole set of five sites as a single region at first instance, the regional average L-CV is estimated using Eq. (15) as:

$$t^R = \sum_{i=1}^{N} n_i t_k^{(i)} \Big/ \sum_{i=1}^{N} n_i \qquad (15)$$

$t^{(I)}$, $t3^{(i)}$ and $t4^{(i)}$ are the values of L-CV, L-skewness and L-kurtosis for the sample of size n_i at i^{th} site. and N is the number of sites in the region. The estimated values of L-CV, L-skewness and L-kurtosis for the selected region are 0.4120, 0.3089 and 0.2062 respectively.

The inter-site variation of L-CV is measured as the standard deviation (V) of the at-site L-CV's (t(i)), weighted proportionally to the record length at each site, and expressed as

$$V = \left[\frac{\sum_{i=1}^{N} n_i \left(t^{(i)} - t^R\right)^2}{\sum_{i=1}^{N} n_i} \right]^{1/2} \qquad (16)$$

Using four-parameter Kappa distribution (Hosking and Wallis, 1997), 500 data regions, having same record length as selected region, were generated. The inter-site variation of L-CV (V) of each generated region is computed and the mean (μ_v) and standard deviation (σ_v) of the computed inter-site variation is obtained. Finally the heterogeneity measure of the selected group of sites is computed as:

$$\boldsymbol{H} = \frac{\boldsymbol{V} - \boldsymbol{\mu_v}}{\boldsymbol{\sigma_v}} \quad (17)$$

Hosking and Wallis (1993) suggested that a region may be regarded as 'acceptably homogenous, if H < 1;'possibly homogenous', if 1< H < 2;'definitely heterogeneous', if H > 2. The value of heterogeneity measure, H is -0.6743 suggesting that the region is acceptably homogenous. Thus, a homogenous region containing five sites, Lakhpuri, Gopalkheda, Yerli, Dedtalai and Burhanpur is defined for the Upper Tapi Basin and regional flood estimates are derived for the same.

4.6Selection of regional frequency distribution

Five three-parameter distributions i.e., generalized extreme value (GEV), log normal (LN-III), Pearson type-III (P-III), generalized logistic (GLO) and generalized Pare to (GPA) were selected as candidate distributions to fit the region. The choice of the best distribution is done by comparing the moments of the candidate distributions with that of the regional average L-moment ratios of the observed data, i.e. by observing how well the L-skewness and L-kurtosis of the fitted distribution match with the regional average L-skewness and L-kurtosis of the observed data (Hosking 1991). The Z^{DIST} statisticscriteria has been used to assess the suitability of the distribution.

The goodness of fit measure, as proposed by Hosking and Wallis (1997),Z^{DIST}, measures the difference between the L-kurtosis of the fitted distribution ($\tau 4^{DIST}$) and that of the regional average L-kurtosis ($t4^{R}$). The Z^{DIST} statistics is defined as:

$$\boldsymbol{Z^{DIST}} = (\boldsymbol{\tau_4^{DIST}} - \boldsymbol{t_4^R} + \boldsymbol{B_4})/\boldsymbol{\sigma_4} \quad (18)$$

$$\boldsymbol{B_4} = \boldsymbol{N_{sim}^{-1}} \sum_{m=1}^{N_{sim}} (\boldsymbol{t_4^{[m]}} - \boldsymbol{t_4^R}) \quad (19)$$

$$\boldsymbol{\sigma_4} = \left((\boldsymbol{N_{sim}} - \mathbf{1})^{-1} \left(\sum_{m=1}^{N_{sim}} \left(\boldsymbol{t_4^{[m]}} - \boldsymbol{t_4^R} \right)^2 - \boldsymbol{N_{sim}} \boldsymbol{B_4^2} \right) \right)^{1/2} \quad (20)$$

Here, B4 and 4 are the bias and standard deviation of regional average L-kurtosis,t4[m], for the m^{th} simulation. Both B4 and4 are obtained from the N_{sim} simulations that are conducted while testing the homogeneity of the region.

The fit is said to be adequate, if Z^{DIST} is sufficiently close to zero. The fit is said to be reasonable when $|Z^{DIST}| < 1.645$ at 10% significance level. If more than one distributions are found to pass the test, the one with the minimum value of $|Z^{DIST}|$ is taken as the most suitable distribution. For the region, all the distributions have been found to pass the goodness-of-fit criterion as their Z^{DIST} values are well below $|Z^{DIST}| = 1.645$ at 10% significance level (see Table 2). However, $| Z^{DIST} |$ for GEV is 0.079 which is sufficiently close to zero when compared with the other candidate distributions. Hence, GEV is recommended as the robust distribution for regional flood frequency analysis of Upper Tapi Basin.

Table 2 $| Z^{DIST} |$ values for different distributions

$\|Z^{DIST}\|_{GEV}$	$\|Z^{DIST}\|_{LN-III}$	$\|Z^{DIST}\|_{GLO}$	$\|Z^{DIST}\|_{P-III}$	$\|Z^{DIST}\|_{GPA}$
0.079	0.446	0.692	1.36	1.62

4.7. Development of regional flood frequency relationship for gauged catchments of Upper Tapi basin

Generalized extreme value (GEV) distribution has been identified as the most robust distribution for the study area. The regional frequency relationships have been developed using the identified distribution with its relevant parameters. The form of regional frequency relationship for GEV distribution corresponding to a return period T is given as:

$$GEV: \frac{Q_T}{\bar{Q}} = \xi + \frac{\alpha}{k}\left[1 - \left\{-ln\left(1 - \frac{1}{T}\right)\right\}^k\right] \quad (21)$$

Here, QT is flood event of T-years return period. The values of regional parameters of the GEV distribution for Upper Tapi basin arξ =0.607 α =0.4721 and k =-0.2062. Substituting the values in Eq. (21), the regional flood frequency relationship, for the gauged catchments of Upper Tapi basin is expressed as:

$$\frac{Q_T}{\bar{Q}} = 0.6709 - 2.2895\left[1 - \left\{-ln\left(1 - \frac{1}{T}\right)\right\}^{-0.2062}\right] \quad (22)$$

Using Eq. (22), dimensionless growth factors can be computed for various return periods, see Table 3 and Fig.3. The flood quantiles can be obtained by multiplying the mean annual peak flood of the catchment $\bar{Q}$ by the corresponding value of the growth factors for GEV distribution.

Table 3 Values of growth factors $(\frac{Q_T}{\bar{Q}})$ of GEV distribution for Upper Tapi basin

Distribution	Return Period (years)								
	2	5	10	25	30	50	100	150	200
	Growth factors								
GEV	0.851	1.501	2.023	2.809	2.982	3.5	4.293	4.811	5.205

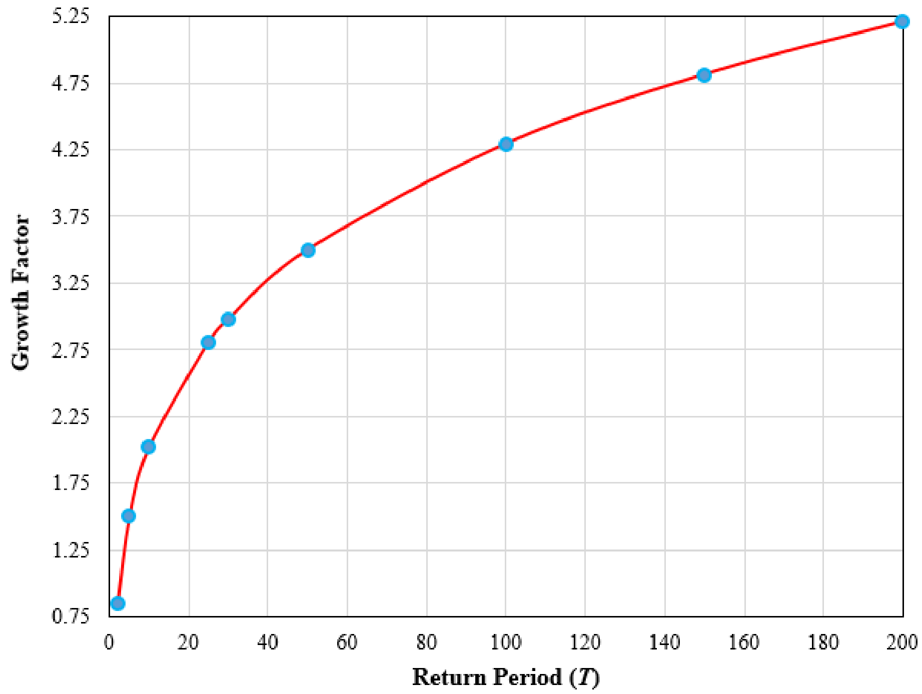


Fig. 3Growth factors computed for GEV distribution

5. Conclusions

The upper catchment of Tapi basin has been selected for regional flood frequency analysis of the region. The following key conclusions can be derived from foregoing investigation:

- Screening of the data for annual maximum discharge of Upper Tapi basin, using the Discordancy measure (D_i) test and G-B outlier test, reveal that data of all the five stations are suitable for performing regional flood frequency analysis. The Hosking and Wallis homogeneity test, i.e., heterogeneity measure, H, shows that the data of five sites constitute a single homogenous region.

- Probability distributions, GEV, LN-III, GLO, P-III and GPA, have been selected as candidate distributions for the region. Regional parameters of the distributions have been estimated using the L-moments approach. Based on $|Z^{DIST}|$-statistic criteria, GEV distribution has been identified as the robust distribution for the study area.

- Regional flood frequency relationship (growth factor) has been developed incorporating the regional parameters of GEV distribution (Eq. (22)). The mean annual peak discharge of the catchment can be multiplied by the corresponding values of growth factors, to obtain flood magnitudes of various return periods.

- The regional flood frequency relationships can be modified for obtaining more accurate flood frequency estimates, if data for more stream gauging stations made available in the region.

References

Burn, D. H., and Goel, N. K. (2000).The formation of groups for regional flood frequency analysis.Hydrological Sciences Journal, 45(1), 97-112.

Greenwood, J. A., Landwehr, J. M., Matalas, N. C. and Wallis, J. R. (1979). Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form. Water Resources Research, 15, 1049–1054.

Hosking, J. R. M. (1990). L-moments: Analysis and estimation of distributions using linear combinations of order statistics. J. Royal Stat. Soc, Series B,52, 105–124.

Hosking, J. R. M. (1991). Approximations for use in constructing L-moment ratio diagrams.IBM Research Division, TJ Watson Research Center.

Hosking, J. R. M., & Wallis, J. R. (1993).Some statistics useful in regional frequency analysis.Water Resources Research, 29(2), 271-281.

Hosking, J. R. M., and Wallis, J. R. (1997).Regional Frequency Analysis: An Approach Based on L-Moments. Cambridge University Press.

Hussain, Z., and Pasha, G. R. (2009).Regional flood frequency analysis of the seven sites of Punjab, Pakistan, using L-moments.Water Resources Management, 23(10), 1917-1933.

Hussain, Z. (2011). Application of the regional flood frequency analysis to the upper and lower basins of the Indus River, Pakistan.Water resources management, 25(11), 2797-2822.

Kumar, R., Chatterjee, C., Kumar, S., Lohani, A. K., and Singh, R. D. (2003).Development of regional flood frequency relationships using L-moments for Middle Ganga Plains Subzone 1 (f) of India.Water Resources Management, 17(4), 243-257.

Machiwal, D., and Jha, M.K. (2012).Hydrologic Time Series Analysis: Theory and practice.Springer.

Pilgrim, D. H., Cordey, I., & Maidment, D. R. (1992).Flood runoff.Handbook of hydrology,9-1.

Saf, B. (2010). Assessment of the effects of discordant sites on regional flood frequency analysis.Journal of hydrology, 380(3), 362-375.

Improved Vegetation Parameter And Hydrological Modelling In A Snow Covered Basin

Diwan Mohaideen M M[1], Abdul Hakeem K[2], Raju P V[2], Mudgal B V[3], Varija K[1]

[1]National Institute of Technology Karnataka, Surathkal, Mangalore, 575 025, India

[2]National Remote Sensing Centre, ISRO, Balanagar, Hyderabad, 500 037, India

[3]Centre for Water Resources, Anna University, Chennai, 600 025, India

Email: diwan1990@gmail.com

ABSTRACT

The total runoff generated from high altitude region like Himalaya is ensuing from both rainfall and snowmelt. In this study, Variable Infiltration Capacity (VIC), a macro-scale grid-based hydrological model, was used to simulate the rainfall-snowmelt-runoff of Sutlej basin, located in western Himalayan region. The major input parameters to the model are soil physical and hydraulic properties, land cover patterns, vegetation phenology in terms of LAI albedo, height, root distribution and topography. Meteorological data- min/max temperature, precipitation, wind speed are forcing parameters for energy balance/exchange simulations. Temporal variation of LAI was used to characterize different landscapes in the study area. MODIS 8-day LAI time-series data was used to sub-group agricultural dominant areas into major crop groups and corresponding monthly vegetation phenology in terms of LAI, albedo, height, root distribution were arrived. This exercise enabled improved definition of vegetation parameterization for the study area, incorporating the region specific conditions. The model parameterization was setup at 3min (~5.5km) grid resolution. The model computed discharge hydrograph was compared with that field observed and model parameters were optimized till best fit between observed and modeled is obtained. 1999-2002, 2002-2004 periods were used for calibration and validation respectively. Calibration period daily time-step Nash-Sutcliffe Efficiency was 0.5, while it was 0.44 during validation period. Higher NSE values were obtained at monthly time-step during both calibration and validation. For better simulation of snowmelt process, a long term spin-up is preferred in order to build up the snow cover for high altitude grids. However in the present study, only two years of simulation was used for model spin-up. The model generates the outputs of various components of the energy fluxes, snow water equivalent and snow cover fraction. This need to be validated with measured and estimated sources to assess the complete model performance.

Keywords: VIC, Leaf Area Index, Land cover, Hydrological Modeling, Evapotranspiration, Snowmelt.

1. Introduction

The precipitation in India comprises of both rainfall as well snowfall. Snowfall is predominant in northern parts of the country. The Himalayas, the largest mountain range in the world, are snow covered throughout the year. The melt water from glaciers of the Himalayan region covers eight countries across Asia. All the major South Asian rivers originate in the Himalayas and their upper catchments are covered with snow and glaciers. In snow covered area, snow melt runoff that is predominant during summer has to be managed properly, otherwise will lead to inadequate fresh water supply in mountainous region and downstream flooding. In order to address these issues, the total runoff estimation from the basin is needed, which can be done through hydrological modeling. Hydrologic models are primarily used for hydrologic prediction and for understanding hydrologic processes in a watershed. In the recent past, researchers have developed models known as soil–vegetation–atmosphere transfer schemes (SVATS). These models consider the role of vegetation in the estimation of evapotranspiration (Aggarwal et al. 2013). The Variable Infiltration Capacity (VIC) macro scale land surface hydrological model was also developed as a SVATS for general circulation models (GCMs) (Liang et al. 1994). Compared to other SVATS, VIC has various distinguished features such as the sub grid variability in soil moisture storage capacity as a spatial probability distribution (Liang et al. 1996a,1996b), sub grid variability of land cover/land use (LULC), the inclusion of topography that allows for orographic precipitation and temperature lapse rates resulting in more realistic hydrology in mountainous regions and drainage from a lower soil moisture zone (base flow) as a nonlinear recession (Zhao et al.1980).

2. Study area

The Sutlej river basin bounded by latitudes 31° 13' N to 32°23' N and longitudes 76° 22' E to 78° 42' E is located in western Himalayan region. The total geographical area of Sutlej river up to Bhakra dam is about 56,980 km2, of which about 37,153 km2 lies in Tibet (Bhakra Beas Management Board). The remaining area about 19,827 km2 lies in the Indian Territory. The elevation of the catchment varies from about 400 m to 7,000 m. The average annual precipitation of the basin is about 1,400 mm and the average maximum and minimum temperature are observed about 20°C and -2°C respectively (APHRODITE water resource Project). The major crops are maize, rice, wheat and potato. The primary land cover in the basin is bare ground (60%), grassland (25%), evergreen forest (5%), scrubland (5%), and snow cover (5%), which is derived from MODIS and Bhuvan LULC data for 2009-10. The study area is shown in Figure 1.

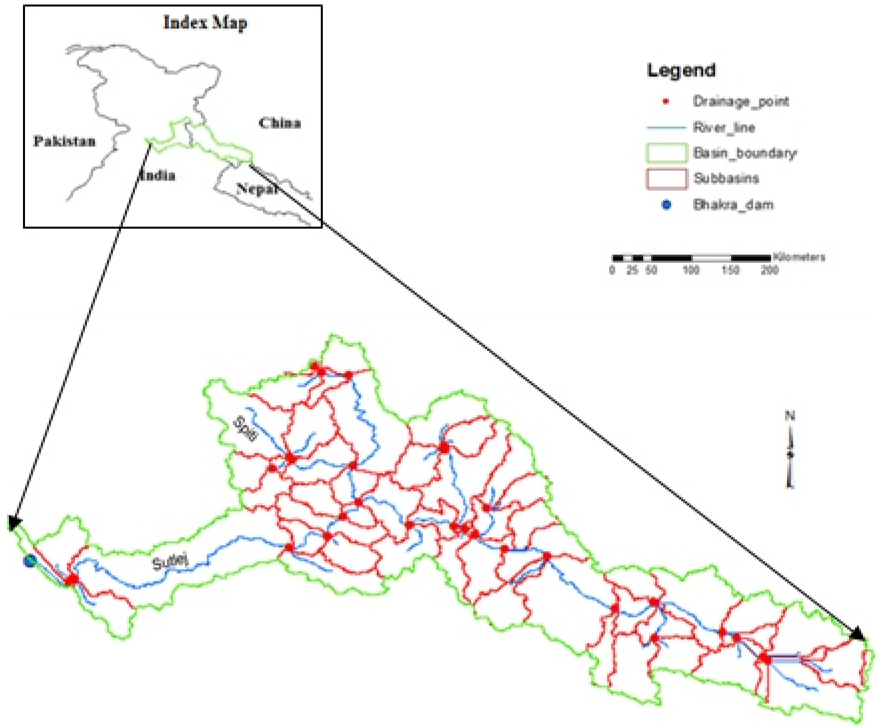


Figure 1 Study area

3. Methodology

3.1 Watershed delineation and basin grids

The watershed boundary and stream network of the study area were delineated using the ASTER Digital Elevation Model (DEM). The model inputs are given in grid wise. It computes the fluxes for all grid cells individually and route the runoff to the outlet of basin. The size of the grid was taken as 3' x 3' for this study. These grids are numbered with descending order along the latitude and ascending order along the longitude. Totally 2,222 grids were identified to model simulation for this basin.

3.2 Input parameterization

The inputs required for the model setup are land surface vegetation classes, soil characteristics, and forcing parameters. Since the study area extends beyond India, both NBSS&LUP and FAO soil map are used for the preparation of the soil parameter. The soil parameters are derived based on the USDA soil texture type. The depth of the three soil layer is considered as 0.15 m, 0.35 m and 1 m respectively. The hydraulic properties of the each soil type is taken from the literature and the area weighted average of those properties for each grid cell are calculated based on the textural class. The hydraulic properties of different soil texture classes used in preparation of soil parameter are given in the Table 1.

Table 1 Hydraulic properties of the soil

Soil / property	Clay	Sandy-clay	Loam	Sandy clay loam	Sandy Loam	Loamy Sand	Sand
K_{sat} (cm/hr)	3.18	1.19	1.97	2.4	5.24	10.87	38
Bulk density (g/cm^3)	1.39	1.57	1.49	1.6	1.57	1.52	1.5
Bubbling (cm)	37.3	29.17	11.15	28.08	14.66	8.69	7.3
Field capacity (cm^3/cm^3)	0.4	0.34	0.29	0.27	0.21	0.15	0.1
Wilting point (cm^3/cm^3)	0.27	0.23	0.14	0.17	0.09	0.06	0
Quartz content	0.25	0.5	0.41	0.61	0.69	0.85	0.9
Slope of retention curve 'b'	12.28	1.19	5.3	8.66	4.84	3.99	4.1
Residual moisture (cm^3/cm^3)	0.09	0.109	0.027	0.068	0.041	0.035	0

Source: (www.hydro.washington.edu/Lettenmaier/Models/VIC/Overview)

The Land Cover data and MODIS LAI data were used for vegetation related parameter definition. For the Indian region of the basin, the Land Use Land Cover (LULC) image from Bhuvan thematic services of National Remote Sensing Centre (NRSC) was used, for year 2009 - 10 (Figure 2a). Vegetation parameter to the model is obtained by compiling the fraction of each vegetation classes and their root distribution. Vegetation library defines the dynamic vegetation parameter such as LAI, albedo, minimum stomatal resistance, architectural resistance, roughness length, displacement length for each vegetation classes in the study area. Land cover data from MODIS (Figure 2b) was recoded based on the LULC from NRSC and combined land cover data is arrived (Figure 2c). An agricultural class mask is created using this combined Land Cover data. Temporal variation of LAI was used to characterize different landscapes in the study area. MODIS 8-day LAI time-series data was used to sub-group agricultural dominant areas into major crop groups and corresponding monthly vegetation phenology in terms of LAI, albedo, height, root distribution were arrived. Unsupervised classification was done using agricultural mask and stacked LAI data for June 2009 - May 2010 for generating LAI profile for different agricultural classes.

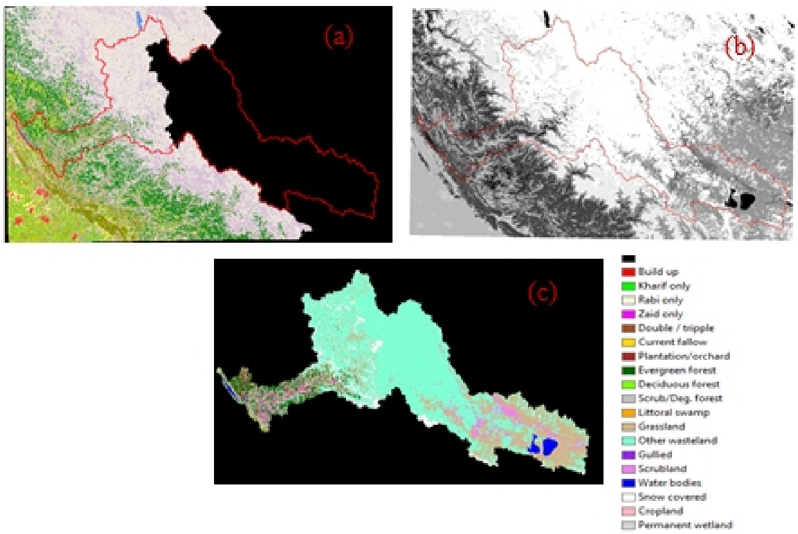


Figure 2 (a) LULC of Indian region, (b) MODIS Land Cover superimposed with catchment boundary, (c) Combined Land Cover from MODIS and LULC/NRSC

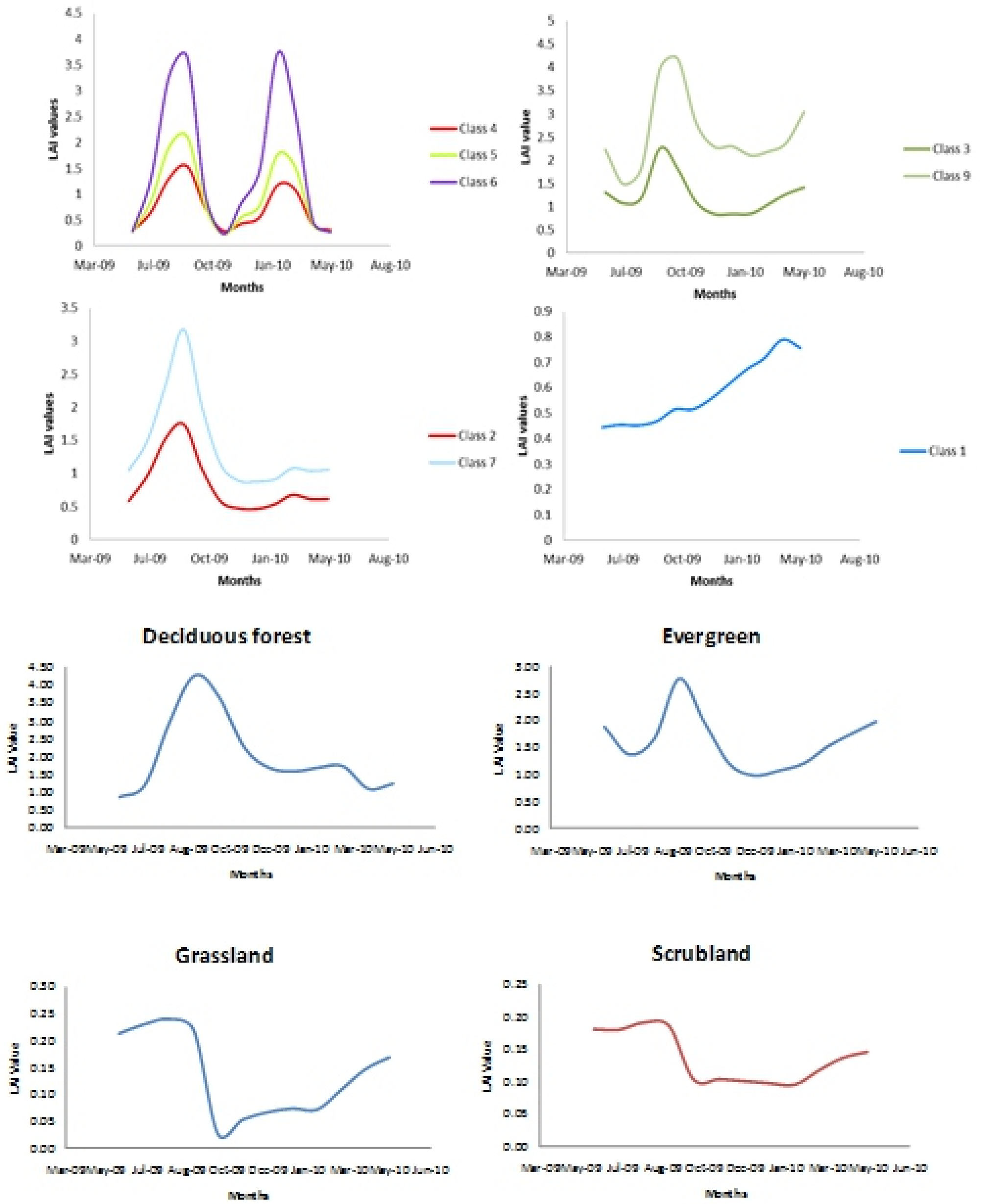


Figure 3 Temporal LAI profile of the agricultural and non-agricultural classes

The unsupervised classification of temporal LAI data resulted in 10 classes, which was grouped based on the temporal variation of the LAI profile of the each class and these classes were taken into the model simulation (Figure 3) (LAI ranges from 0.23 – 5.24). The major cropping patterns in this area were identified as maize – wheat and rice – wheat. These classes can be grouped as single vegetation class. The LAI profile to the classes 2 & 7 were identified as rice. The LAI values of the other non-agricultural classes such as evergreen forest, deciduous forest, grassland, and scrubland are arrived from the representative samples of each class (LAI ranges from 0.02 – 6.8). The above practices enabled the improved definition of vegetation parameterization for the study area, incorporating the region specific conditions. Meteorological data- min/max temperature, precipitation (in mm per step) and wind speed (m/s) are forcing parameters for energy balance/exchange simulations. The precipitation data was taken from APHRODITE dataset (June 1997 – May 2004). Temperature and wind speed data was used from the historical datasets available in the VIC website. ASTER DEM was used to define terrain topography and runoff routing parameterization. The DEM of the entire Himalayan region is classified into 10 classes at the interval of 820 m. The elevation band of the study area was extracted from the classified DEM. The elevation band to model simulation contains area fraction of each classes, mean elevation and precipitation factor for each grid cell. Precipitation factor is the fraction of cell precipitation that falls on each elevation band. Elevation bands representing topography improves simulations of elevation-dependent components within a grid cell.

The simulated runoff is routed through the river network using a simple routing model (Lohmann et al 1996, 1998a, 1998b). Model uses Unit Hydrograph (UH) approach and linearized Saint-Venant equation for within the grids and the channel routing respectively. It assumes all runoff exits a cell in a single flow direction. Model requires the parameters such as fraction of each grid cell that flows into the basin being routed, flow direction, flow velocity and diffusion (assumed to be constant), grid size and UH file, which gives the grid cell impulse response function.

4. Results

4.1 Calibration and validation

The model simulates the energy fluxes as an output for the each grid cell. The model output grid-wise runoff was routed through catchment hydraulic network to obtain runoff hydrograph at basin outlet. The model was calibrated using a set of initial values and optimized till the best fit between observed and modeled is obtained by trial and error. 1999 – 2002 period was used for calibration. The calibration parameters and their calibrated values are given in the Table 2. The comparison between simulated and observed daily, monthly stream flow and their scatter plots in calibration are shown in Figure 4. The model was validated for the water years 2002 - 2004. The comparison between the simulated and observed stream flow for daily, monthly stream flow and their scatter plots in validation are shown in Figure 5. Nash-Sutcliffe Simulation Efficiency coefficient (Nash and Sutcliffe, 1970) was used to determine the agreement. The values of NSE during calibration and validation are listed in the Table 3.

Table 2 Calibration parameter and values

	Calibrated Values				
Avg K_{sat} (cm/hr)	>10	>7	>5	>3	<3
b_{inf}	0.2	0.225	0.275	0.3	0.325
D_s	0.2	0.4	0.4	0.4	0.6
W_s	0.7	0.6	0.6	0.6	0.5
Snow surface roughness	0.0025				
Max. temperature	0.5				
Min. temperature	-0.5				

K_{sat} – Saturated hydraulic conductivity

D_S - Fraction of maximum velocity of the base flow

W_s - Fraction of maximum soil moisture

b_{inf} - Describe the infiltration curve in the model

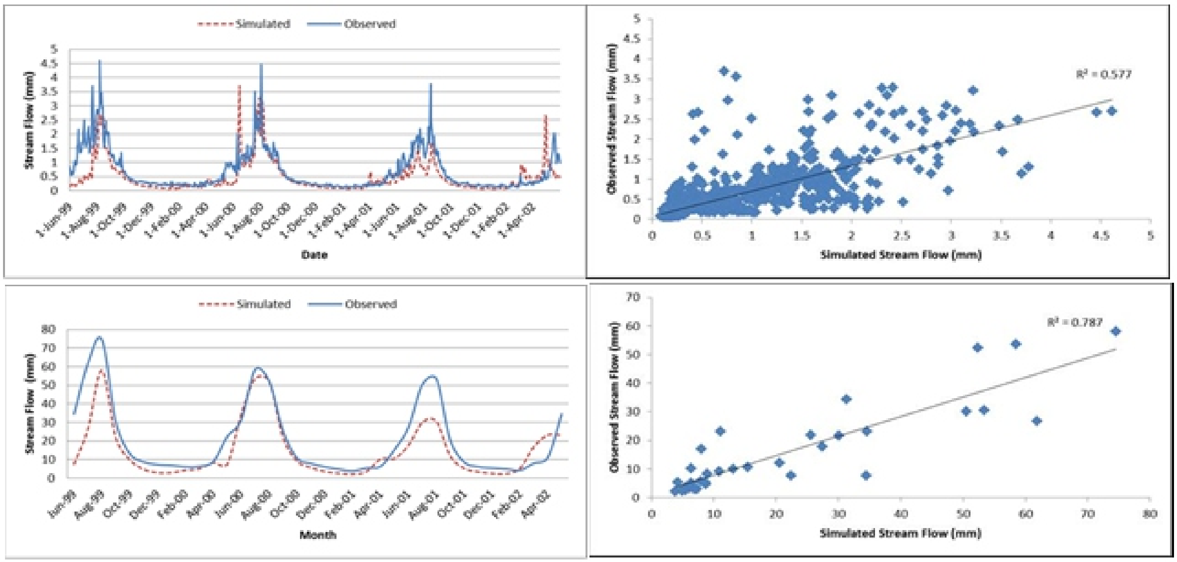


Figure 4 Comparison between simulated and observed flow and scatter plots in calibration

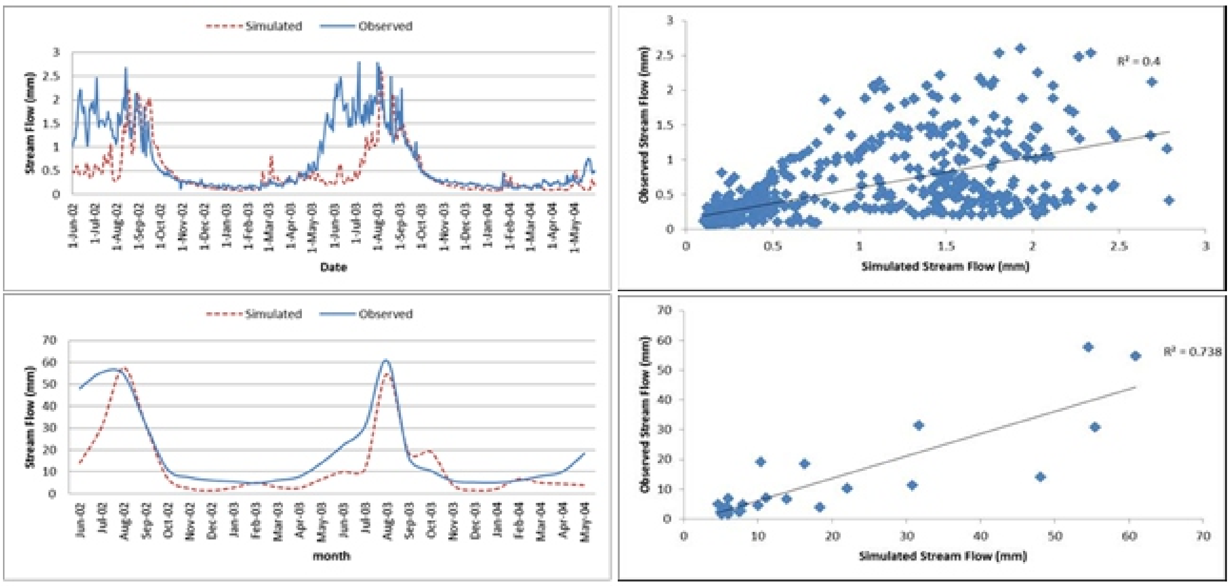


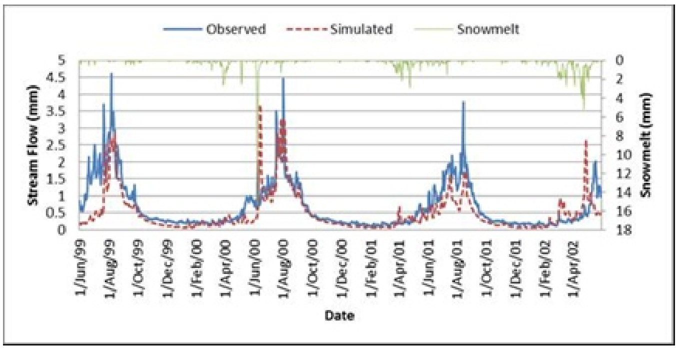
Figure 5 Comparison between simulated and observed flow and scatter plots in validation

Table 3 Nash-Sutcliffe Efficiency

	June 1999 - May 2002	June 2002 - May 2004
Time Step	**Calibration**	**Validation**
Daily	0.5	0.44
Monthly	0.69	0.57

4.2 Discussion

The results show that the model generally underestimated the runoff during both calibration and validation phases. The under estimation was observed more during the base flow period. The gridded precipitation data from APHRODITE is used as one of the forcing parameter in the model. The density of ground observed rain gauge data used in preparation of this gridded data is observed to be poor for this study area. The results show that the observed runoff is 55% of the precipitation which is used for the simulation. It was found that the simulated flow was about 30% less than the observed flow during calibration and validation. The difference between simulated and observed runoff is found to be more during the south-west monsoon period of 2002 and 2003. This has resulted in low NSE value during the validation phase of model. Higher NSE values were obtained at monthly time-step during both calibration and validation. The main objective of the study is to estimate total runoff from the rainfall and snowmelt. Snowmelt generally occurs during the summer months of March to June. The comparison of the model simulated snowmelt with the daily observed and simulated stream flow is shown in Figure 6. The NSE value for the snowmelt period of March – June in the calibration as well as validation needs improvement.



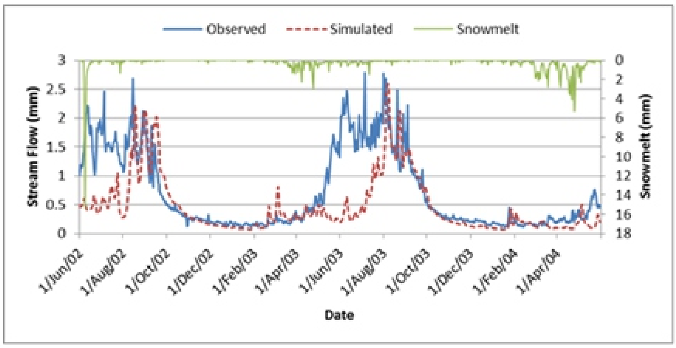


Figure 6 Comparison of simulated snowmelt with daily observed and simulated stream flow in calibration and validation

5. Conclusions

Vegetation parameter and vegetation library information are one of the important input to VIC hydrological model. Among the various parameters, LAI is the major parameter defining the vegetation growth, which affects the simulation of evapotranspiration in the hydrological model. Satellite derived LAI data is found to be the best source of temporal information on LAI for different vegetation classes in a highly inaccessible terrain like Himalayas. In this study, an attempt is made to define vegetation parameter through MODIS LAI time series data. This can improve the definition of the vegetation parameterization of the model simulation. Since the model is highly sensitive to the forcing data, especially the precipitation data, a better source of this data may improve the model performance. For better simulation of snow melt process in the VIC model, a long term spin-up is preferred in order to build up the snow cover for high altitude grids. However in the present study, only two years of simulation was used for model spin-up. This has resulted in poor simulation of snow melt process. The vegetation library file defines the properties of different vegetation types in month wise time step. However for the crops, the phenology changes within the month. This needs to be implemented by modifying the source code of VIC model. The model generates the outputs of various components of the energy fluxes, snow water equivalent and snow cover fraction. This need to be validated with measured and estimated sources to assess the complete model performance.

Acknowledgement

Data support and computational facility was provided by Water Resources Group, National Remote Sensing Centre, Hyderabad. The first author is thankful to Mr. Saksham Joshi (WRM&AD, NRSC) for his constant support during this study.

References

Aggarwal, S. P., Vaibhav Garg, Prasun, K. G., Bhaskar, R. N., Praveen, K. T., and Roy, P. S. (2013). Run-off potential assessment over Indian landmass: a macro-scale hydrological modeling approach. Current Science, Vol. 104, No. 7, pp. 950 – 958.

Asian Precipitation - Highly Resolved Observational Data Integration Towards Evaluation of the Water Resources (APHRODITE), last accessed on 9th July 2015, http://www.chikyu.ac.jp/precip/index.html.

Bhakra Beas Management Board, Last accessed on 9th July 2015 http://bbmb.gov.in /english/ menu2.asp.

Liang, X., Lettenmaier, D. P., Wood, E. F., and Burges, S.J. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. Journal of Geophysical Research, Vol. 99, No. D7, pp. 14 415 – 14 428.

Liang, X., Lettenmaier, D.P., and Wood, E.F. (1996a). A one-dimensional statistical dynamic representation of subgrid spatial variability of precipitation in the two-layer variable infiltration capacity model. Journal of Geophysical Research, Vol. 101, No. D16, pp. 21 403 – 21 422.

Liang, X., Wood, E.F., and Lettenmaier, D.P. (1996b). Surface soil moisture parameterization of the VIC-2L model: evaluation and modifications. Global and Planetary Change, Vol. 13, No. 1, pp. 195–206.

Lohmann, D., R. Nolte-Holube, and E. Raschke, (1996). A large-scale horizontal routing model to be coupled to land surface parametrization schemes, Tellus, 48(A), 708-721.

Lohmann, D., E. Raschke, B. Nijssen and D. P. Lettenmaier, (1998a). Regional scale hydrology: I. Formulation of the VIC-2L model coupled to a routing model, Hydrol. Sci. J.,43(1), 131-141.

Lohmann, D., E. Raschke, B. Nijssen, and D. P. Lettenmaier, (1998b). Regional Scale Hydrology: II. Application of the VIC-2L Model to the Weser River, Germany, Hydrological Sciences Journal, 43(1), 143-157.
Nash, J. E. and J. V. Sutcliffe (1970). River flow forecasting through conceptual models part I — A discussion of principles, Journal of Hydrology 10 (3), 282–290.
Variable Infiltration Capacity (VIC) Macroscale Hydrologic Model, 'University of Washington, last accessed on 9th July 2014, http://www.hydro.washington.edu/Lettenmaier/Models/VIC
Zhao, R.J., Zhang, Y. L., and Frang, L. R. (1980). The Xianjiang model, Hydrological Forecasting Proceeding Oxford Symposium, IASH 129, 351 – 356.

Instructions for Authors

Essentials for Publishing in this Journal

1 Submitted articles should not have been previously published or be currently under consideration for publication elsewhere.

2 Conference papers may only be submitted if the paper has been completely re-written (taken to mean more than 50%) and the author has cleared any necessary permission with the copyright owner if it has been previously copyrighted.

3 All our articles are refereed through a double-blind process.

4 All authors must declare they have read and agreed to the content of the submitted article and must sign a declaration correspond to the originality of the article.

Submission Process

All articles for this journal must be submitted using our online submissions system. http://enrichedpub.com/ . Please use the Submit Your Article link in the Author Service area.

Manuscript Guidelines

The instructions to authors about the article preparation for publication in the Manuscripts are submitted online, through the e-Ur (Electronic editing) system, developed by **Enriched Publications Pvt. Ltd**. The article should contain the abstract with keywords, introduction, body, conclusion, references and the summary in English language (without heading and subheading enumeration). The article length should not exceed 16 pages of A4 paper format.

Title

The title should be informative. It is in both Journal's and author's best interest to use terms suitable. For indexing and word search. If there are no such terms in the title, the author is strongly advised to add a subtitle. The title should be given in English as well. The titles precede the abstract and the summary in an appropriate language.

Letterhead Title

The letterhead title is given at a top of each page for easier identification of article copies in an Electronic form in particular. It contains the author's surname and first name initial .article title, journal title and collation (year, volume, and issue, first and last page). The journal and article titles can be given in a shortened form.

Author's Name

Full name(s) of author(s) should be used. It is advisable to give the middle initial. Names are given in their original form.

Contact Details

The postal address or the e-mail address of the author (usually of the first one if there are more Authors) is given in the footnote at the bottom of the first page.

Type of Articles

Classification of articles is a duty of the editorial staff and is of special importance. Referees and the members of the editorial staff, or section editors, can propose a category, but the editor-in-chief has the sole responsibility for their classification. Journal articles are classified as follows:

Scientific articles:

1. Original scientific paper (giving the previously unpublished results of the author's own research based on management methods).

2. Survey paper (giving an original, detailed and critical view of a research problem or an area to which the author has made a contribution visible through his self-citation);

3. Short or preliminary communication (original management paper of full format but of a smaller extent or of a preliminary character);

4. Scientific critique or forum (discussion on a particular scientific topic, based exclusively on management argumentation) and commentaries. Exceptionally, in particular areas, a scientific paper in the Journal can be in a form of a monograph or a critical edition of scientific data (historical, archival, lexicographic, bibliographic, data survey, etc.) which were unknown or hardly accessible for scientific research.

Professional articles:

1. Professional paper (contribution offering experience useful for improvement of professional practice but not necessarily based on scientific methods);
2. Informative contribution (editorial, commentary, etc.);
3. Review (of a book, software, case study, scientific event, etc.)

Language

The article should be in English. The grammar and style of the article should be of good quality. The systematized text should be without abbreviations (except standard ones). All measurements must be in SI units. The sequence of formulae is denoted in Arabic numerals in parentheses on the right-hand side.

Abstract and Summary

An abstract is a concise informative presentation of the article content for fast and accurate Evaluation of its relevance. It is both in the Editorial Office's and the author's best interest for an abstract to contain terms often used for indexing and article search. The abstract describes the purpose of the study and the methods, outlines the findings and state the conclusions. A 100- to 250-Word abstract should be placed between the title and the keywords with the body text to follow. Besides an abstract are advised to have a summary in English, at the end of the article, after the Reference list. The summary should be structured and long up to 1/10 of the article length (it is more extensive than the abstract).

Keywords

Keywords are terms or phrases showing adequately the article content for indexing and search purposes. They should be allocated heaving in mind widely accepted international sources (index, dictionary or thesaurus), such as the Web of Science keyword list for science in general. The higher their usage frequency is the better. Up to 10 keywords immediately follow the abstract and the summary, in respective languages.

Acknowledgements

The name and the number of the project or programmed within which the article was realized is given in a separate note at the bottom of the first page together with the name of the institution which financially supported the project or programmed.

Tables and Illustrations

All the captions should be in the original language as well as in English, together with the texts in illustrations if possible. Tables are typed in the same style as the text and are denoted by numerals at the top. Photographs and drawings, placed appropriately in the text, should be clear, precise and suitable for reproduction. Drawings should be created in Word or Corel.

Citation in the Text

Citation in the text must be uniform. When citing references in the text, use the reference number set in square brackets from the Reference list at the end of the article.

Footnotes

Footnotes are given at the bottom of the page with the text they refer to. They can contain less relevant details, additional explanations or used sources (e.g. scientific material, manuals). They cannot replace the cited literature.
The article should be accompanied with a cover letter with the information about the author(s): surname, middle initial, first name, and citizen personal number, rank, title, e-mail address, and affiliation address, home address including municipality, phone number in the office and at home (or a mobile phone number). The cover letter should state the type of the article and tell which illustrations are original and which are not.

Notes: