https://doi.org/10.15623/ijret.2018.0704010 Received: 12-02-2018, Accepted: 26-03-2018, Published: 16-04-2018

A NEW HYDROLOGICAL APPROACH TO SOURCE EVALUATION FOR PUBLIC PRIVATE PARTICIPATION PROJECTS FOR URBAN DRINKING WATER

Vasudev Shirahatti[1]

[1]*Consulting Engineer, Black & Veatch India Private Limited, Mumbai, India*

Abstract

Application of 2- site model for generating synthetic stream flows is discussed. Historical simulations based on insufficient field data may often be not reliable due to small sample bias. In the case of public Private Participation (P-P-P) projects rigorous testing of reservoir yields is critically important. The 2-site model applied to a field situation wherein main river having 29-years of daily flow data is treated as key river and its main tributary having only 14 -years concurrent data is treated as follower. Leveraging the lag-zero cross correlation that must exist between two adjacent catchments falling under the same hydrometeorological conditions, record of follower river was extended through application of innovative 2-site model. Study basin lies in tropical wet-dry climate of peninsular India as such high skewness and serial lag properties characterise historical time series. These higher model moments have great economic significance in reservoir sizing for sustaining commercial yields and this calls for extreme care in modelling. In the model high sample skewness was treated by Wilson-Hilfertytransformation in the generating equations for generating longer synthetic time series of 100 items as may be necessary for reservoir simulation purposes. Preserving all historical statistics including higher order moments like skew coefficient in generated samples is of priority in commercial applications. Paper discusses novel evaluation technique devised for evaluating model generated samples

***Keywords**: Historical simulations, Public Private Participation Projects, synthetic simulations, commercial yield, reservoir simulation, 2 -site model, peninsular Indian rivers, skewness coefficient, serial correlation coefficient, cross correlation coefficient, model evaluation criteria*

---***---

1. INTRODUCTORY REMARKS

Effective water resource development and management are basic to sustained growth and poverty alleviation. Broad based water resource interventions meant for industrial use, power use or for drinking water provide national, regional and local benefits from which all people including poor can benefit.

After independence, Government of India as a policy measure mandated state and central government bodies to own, develop and deliver water resources to populace. Under this policy be it creating large irrigation facilities or development of power projects or urban water supply it was the prerogative of government departments to act exclusively as project developers with government finances in concert with regulatory bodies. This striking feature of India's developmental policy has undergone a sea change since ushering of economic liberalization of the decade of 90's and from that time on private participation in core infrastructure projects including water has become more of a norm rather than exception.

Water projects in private sector come with their own peculiar problems. While ownership and regulatory controls remain with the government, water is yet to be accepted as a commodity in India which can be exploited for maximum financial returns. For example, urban drinking water project developed in a Public Private Participation (P-P-P) mode must necessarily accommodate cross subsidies drinking water delivery to poorer sections of urban populace by charging more to industrial users or to urban rich. When sharing the source with other governmental departments P-P-P developers often must operate under peculiar source constraints because of which many an urban water supply project in the P-P-P mode has remained non-starter in India.

Normally source earmarked for projects in private sector will be either in the form of a reservoir or a virgin river site. For ensuring optimum use of allocated source a primary objective in the project design is commercial evaluation of sustainable yield. Such evaluations include collection of reliable and long-term hydro-meteorological data collection followed by up to date hydrological analyses. For drinking water, the mandated reliability of delivering specified quantity of water is 95%.Generally, concession agreements drawn up for P-P-P projects specify penalties for non-delivery. Thus, project design must ensure robust design to ensuring sustainable profitability of the venture. Hydrologist's primary role in design of a water resources project istoensure efficient quantification of scarcewater resources. However, in the context of P-P-P projects

https://doi.org/10.15623/ijret.2018.0704010 Received: 12-02-2018, Accepted: 26-03-2018, Published: 16-04-2018

hydrologist’s responsibility increases two-fold: a) to rigorously testing reservoir yields for ensuring commercial success and b) designing strategies for reservoir operations for mitigating shortfalls and consequent penalties.

2. PRESENT STUDY

It is the prerogative of the Government to allocate water source. When furnishing his bid in response to the Request for Proposal (RFP), regulatory norms mandate that the Proponent must commit nominal source yield through historical (deterministic) reservoir simulations. This is understandable because projected source yield must be based on a common approach to assist in bid analyses. Good and reliable historical data base is hence a prerequisite for yield projections. Historical time series of 50 years is normally preferredby hydrologists for reservoir simulations but in most cases field data available will be far less. Reservoir outcomes based on smaller data sets run the risk of small sample bias. Although historical data could be extended through rainfall -run off correlations or other watershed models, deterministic simulations are beset with a major handicap in that they are based on but onepast run off trace. In all commercial applications where water must be competitively pricedrigorous tests are a prerequisite to confirm reliability of historical yields and / orgenerate allied information on risks like vulnerability etc. Often sources earmarked come with existing user’srights. Such situations call for rule curves to be simulated taking into accountlikely future runoff variations. It is here that stochastic modelling comes to the rescue of hydrologists. Operation Hydrology is generation of hundreds of synthetic traces that have the same probability distribution as observed records. Synthetic traces have the principal advantage in that they are surrogates of historical time series and it is easily recognised that when hundreds of synthesized time series of longer length say of 100 years (same as the useful life of a reservoir) are routed through a candidate reservoir its performance is rigorously tested and estimates of frequency and severity of failures are generated. Operation hydrology has come a long way since its inception[1]and can be of immense benefit to P-P-P projects provided hydrologist can improvise text book formulations for meeting a field situation.It is now realized that stochastic modelling provides a pragmatic platform for testing yields and verify performance of reservoirs be it for drinking water or hydro generation or for any other commercial use. Hydrological analyses form a crucial part of front end engineering in P-P-P projects and hydrologist’s work typically comprises of generating synthetic traces and routing generated traces through the proposed storage. This paper discusses the first part ie; applying operation hydrology as input for reservoir simulation purposes.

3. DECRIPTION OF STUDY BASINS

The city of Mumbai at present abstracts about 2030 Ml /d (450 Mg/d) of raw water from Vaitarna and Tansa river basins by gravity transfer to treatment works. Municipal Corporation of Greater Mumbai (MCGM) have initiated planning for future sources by gravity transfer in the backdrop of future demand projections for the metropolis. In this context government authorities have earmarked Pinjalriver basin as a potential source which is situated about 50 km away and lying just north of the Vaitarna basin. On the basis of historical simulations government agencies have projected a combined total of yield of 2100 Ml/d. It is necessary to build two new impoundages at two separate locations for harnessing river waters. Study river basin (18.5 ^{0} N and 74.5^{0} E) are shown in Fig-1. Hydrometeorological description of this entire coastal belt is tropical wet and dry climate. Pinjal is the main river having catchment area of 327 sq.km at the proposed dam site near village of Andhari and joins Vaitarna river after traversing 85 km further downstream. Gargai its main tributary intercepting a catchment of 109 sq.km up to its proposed dam site near village of Ogade. Flowing further down Gargai confluences with Pinjal downstream of Andhari. As required for a bankable report prepared in 2006hydrometeorological data was procured from the Maharashtra Central Designs Organization (CDO), Nashik which included 29 years of daily flow data going from 1976 to 2004 for Pinjal and similar concurrent 14 years for Gargai from 1992 to 2005

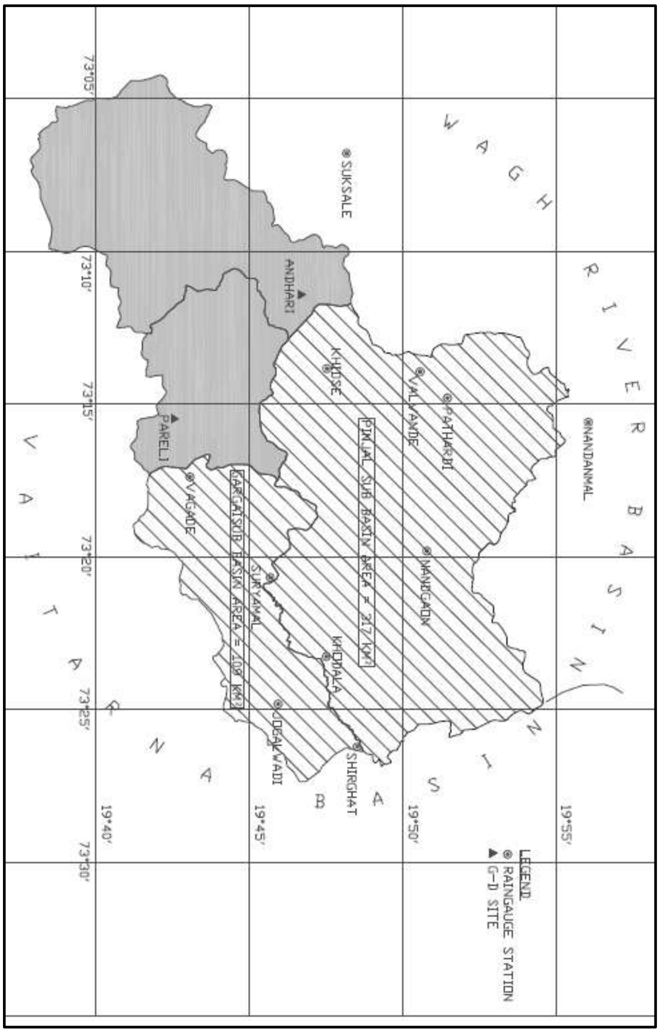

Fig -1: Pinjal Basin

https://doi.org/10.15623/ijret.2018.0704010 Received: 12-02-2018, Accepted: 26-03-2018, Published: 16-04-2018

As a part of comprehensive hydrological analyses, available daily discharge data was processed and concurrent time series files of 10 -day flows in both Pinjal and Gargai rivers were prepared for use in yield analyses. Monthly inflows were aggregated using this processed data along with averages of respective records and statistical parameters obtained based on method of moments which are presented in Table-1 below. Average annual runoff depths of both catchments vary between 2600 to 2700 mm and compare with the IMD long term isohyet of 106.25 inches (2750 mm) indicated for this entire hydrometeorological segment.As one expects of typical peninsular wet and dry climate, estimated skew and serial lag coefficients of both rivers are high. Diagnostic tests confirm that data skewness lieoutside non-zero threshold @ 5 % significance for the respective sample sizes.

Table-1: Historical data and Statistics of Pinjal and Gargai rivers

	Pinjal river basin (CA=317 Sq.km)			Gargai river basin (CA=109 Sq.km)		
	Years of gauged run off data: **1976-2004**			Years of gauged data: **1992-2005**		
	Volume in million cubic meters, Mm3			Volume in million cubic meters, Mm3		
Month	Mean ($\bar{x}$)	Maximum	Minimum	Mean ($\bar{y}$)	Maximum	Minimum
June	68.82	918.89	0	27.49	159.91	0.31
July	340.8	838.43	117.32	82.36	150.15	19.78
August	283.6	772.37	69.73	95.27	381.58	25.23
September	121.4	539.51	17.07	46.9	144.61	12.68
October	30.72	66.18	2.05	10.83	31.98	0
Novenber	3.4	26.3	0	2.12	15.25	0
December	0.53	6.77	0	0	3.16	0
January	0	0	0	0	0	0
February	0	0	0	0	0	0
March	0	0	0	0	0	0
April	0	0	0	0	0	0
May	0	0	0	0	0	0
Seasonal inflow as % of annual total	93	99	68	93	996	84
Mean Annual Inflow x bar/y bar	**869.23**			**264.43**		
Stan. Deviation (s)	342.9			146.2		
Skew Coef , g	0.846			1.338		
Lag-1 Serial Corr. Coef. r_1	0.308			0.849		
Lag-0 (cross) corr. Coef. r_0			0.452			

4. MODEL SELECTION

Results summarised above indicate that a) inflows are predominantly seasonal. Historical data suggests that about 93 % of annual runoff is seasonal occurring during 4 monsoon months of June, July, August and September, b) discharges post monsoon (from October onwards)dwindle rapidly and both rivers run dry come many years mostly in the month of August resulting in heavy flood flows for about a week. A general trend is that computed annual totalsare higher if more number of stormshave occurred in a year and conversely, much smaller in their absence. Because of high runoff events sample skewness is high for both rivers categorizing both data as skewed samples and d) persistence of higher or lower runoffs to occur sequentially in two to three runs. High value (0.85) of sample lag -1 serial correlation coefficient (SCC-1) of the 14-year Gargai data sample is a result of this. In practical designs and particularly so in case of P-P-P projects, these hydrometric characteristics of peninsular tropical wet and dry climate require extreme care in sizing reservoirs. Hence for selecting a model for generating synthetic traces of run off following aspects were duly deliberated upon:

a) Skewness is an important statistic that impact size of live storage in design calculations. For runoff simulation choice is to model sample skewness directly or transform data through normal transformation techniques. For small skewness normal transformation approach might work but in practical situations whereskewness is significant direct modelling approach is the more preferred route keeping in view of economic benefits of storing spills. Gamma probability distribution is known to preserve skewness and hence selected for simulation.

https://doi.org/10.15623/ijret.2018.0704010 Received: 12-02-2018, Accepted: 26-03-2018, Published: 16-04-2018

b) Higher value of serial correlation coefficient (SCC-1) of the historical data portendunder supplies to reservoir in sequential years over the life of reservoir. In operation hydrology either a single season lag(lag-1) or multi-season lag(multi- lags) models are usedfor flow generation. Both Pinjal and its tributary Gargai are small river basins and it is reasonable to assume that most of infiltrated ground water flows into the river within the same hydrologic year and some part may get carried over to the following year. Hencesingle lag model(Markov) is considered appropriate for the present study and this is supportive of use of gamma distribution.

c) For yield testing, seasonality of rivers is leveraged to split generated annual flows into two components. Major component of annual run off @95 % is the seasonal input and balance 5 % assumed as base flow component. Status of reservoir at end of monsoon season (instead of each monsoon month) and another at the end of hydrological year would suffice to evaluate adequacy of reservoir capacity in delivering yield.

d) Using 14-year long historical time series in conventional univariate Markov model is not considered trustworthy for making crucial decisions about yield.

In the light of foregoing stochastic model chosen for flow generation is the well-known2-site generation model conceptualized in standard textbooks [2] . Inbi-variate models short term data of one catchment is extended by spatially correlating it with its adjacent basin having a longer data base through lag-zero correlation. Applying to present problem, Gargai time series which is only 14 years long is extended by using concurrent run off data of 29 – year long Pinjal records through cross correlation that must exist @ zero-lag. Lag-zero cross correlation, $r_{0\,xy}$ is defined by:

$$r_{0x\,y} = \sum_{i=1}^{i=n} \{(x\,i - \bar{x})\,(yi - \bar{y})\} \,/\, \sqrt{[\{\sum_{i=1}^{i=n} (xi - \bar{x})^2 \sum_{i=1}^{i=n} (yi - \bar{y})^2\}]} \text{ ----(1)}$$

where, n=14 and x_i is Pinjal annual runoff and i=1 is the year 1991 and similarly y_i is the recorded Gargai runoff of the same year and $\bar{x}$ and $\bar{y}$ are average of records of Pinjal and Gargai respectively.Since Pinjal and Gargai are two adjacent basins separated by a low relief ridge it is reasonable to assume spatial correlation to exist between two catchments. Based on 14 pairs of annual flows of Pinjal – Gargai r_{0xy} computed is 0.452

2-station model used here is that discussed by Kottegoda & Yevjevich[3]and is essentially a simple autoregressive (AR-1) type or Markov model, convenient for field engineers for manual computing in a recursive manner. Generating equations of the 2-station model in the standardized annual flows format are written as:

$$\xi_{x,\,t} = r_{1x}\,\xi_{x,\,t-1} + \lambda_{1x,\,t}\,\sqrt{(1-r_{1x}^{\;2})} \text{ ------- (2)}$$

$$\xi_{y,\,t} = r_{1y}\,\xi_{y,\,t-1} + \alpha\,\lambda_{1x,\,t} + \beta\,\lambda_{2y,\,t} \text{ ------- (3)}$$

where, $\xi_{x,\,t}$ and $\xi_{y,\,t}$ are the standardized components of flows at Pinjal (X-series) and Gargai (Y-series) sites respectively, $\lambda_{1\,x,\,t}$ and$\lambda_{2\,y,\,t}$ are time independent stochastic components distributed as gamma at the same time t. First terms in both equations are the deterministic components in which r_{1x} and r_{1y} are the first serial correlation coefficients of Pinjal and Gargai records respectively. Main river Pinjal which has longer data base is treated as the key site and tributary Gargai having smaller concurrent data is treated as the subordinate site. X –series is generated through conventional univariate AR-1 model and concurrent Y – series by bi-variate technique through spatial correlation that exists between the main and subordinate rivers.

Treatment of skewness of data series is crucial in this generation scheme. This is achieved by applying Wilson-Hilferty transformation (WH) a technique discussed by Haan[4]. A step by step process of working this technique to field problems is discussed in the monograph of Jackson &Fiering [5] . First skewness of random components is obtained from:

$$\zeta_{\epsilon\,x} = \{1 - r_{1x}^{\;3} / (1 - r_{1x}^{\;2})^{1.5}\}\, g_x \text{ ------------- (4)}$$

$$\zeta_{\epsilon\,y} = \{1 - r_{1y}^{\;3} / (1 - r_{1y}^{\;2})^{1.5}\}\, g_y \text{ ------------- (5)}$$

where, g_x, g_y, r_{1x} and r_{1y} are pairs of historical skewness and serial lag-1 parameters respectively. By substituting sample values $\zeta_{\epsilon\,x}$ works out as 3.52 and $\zeta_{\epsilon\,y}$ 0.95. Impact of higher sample skewness and serial correlations areclear in higher values of transformed skewness parameters. In the case of Y-series the multiplier is 2.6 times while that for the X-series is 1.12. Random variates distributed as gamma viz, $\lambda_{1x,\,t}$ and $\lambda_{2y,t}$usedinEq's (2) and (3) are then obtained through WH transformation by incorporating skewness of random components:

$$\lambda_1\, x,\, t = 2/\zeta_{\epsilon\,x}\,(1 + \zeta_{\epsilon\,x}\, t_{t1}/6 - \zeta_{\epsilon\,x}^{\;2}/36)^3 - 2/\zeta_{\epsilon\,x} \text{-------(6)}$$

$$\lambda_2 y,\, t = 2/\zeta_{\epsilon\,y}\,(1 + \zeta_{\epsilon\,y}\, t_{t2}/6 - \zeta_{\epsilon\,y}^{\;2}/36)^3 - 2/\zeta_{\epsilon\,y} \text{--------- (7)}$$

Where t_{t1} andt_{t2} are two standard random normal numbers at the same time t.

Weighting parameters α and β used in Eq-3 are constituted by the time invariant catchment retention (lag-1 serial correlation coefficient) characteristics and are calculated from:

$$\alpha = r_{0x\,y}\,\{(1 - r_{1x}\, r_{1y}) / (1 - r_{1x}^{\;2})^{1/2}\} \text{ --------------------- (8)}$$

and

$$\beta^2 = \{(1 - r_{1x}^{\;2})(1 - r_{1y}^{\;2}) - r_{0\,x\,y}^{\;2}(1 - r_{1x}\, r_{1\,y})^2\} / (1 - r_{1x}^{\;2}) \text{ -------- (9)}$$

https://doi.org/10.15623/ijret.2018.0704010 Received: 12-02-2018, Accepted: 26-03-2018, Published: 16-04-2018

where, r_{1x}, r_{1y} are lag-1 serial correlation coefficients which are catchment specific and considered time invariant and spatial correlation coefficient r_0 which is indicative of spatial linkage between Gargai and Pinjal run offs. By substituting values in Eq-8 and 9 these are obtained as 0.35 and 0.39 respectively. Thus it can be seen that Gargai flows are synthesized by taking advantage of spatial correlation with its adjacent basin of Pinjalthrough parameters α and βwhich are dependent on lag zero correlation parameter r_0 and additionally on serial correlation coefficients of both catchments.

Finally using equations (2) and (3) dependant variables of annual flows of X and Y –series are obtained from:

$$X = s_x \xi_{x,t} + \bar{x} \text{------------- (10)}$$

$$Y = s_y \xi_{y,t} + \bar{y} \text{------------- (11)}$$

where, s_x and s_y are the standard deviations of the original x and y –series respectively, $\bar{x}$ and $\bar{y}$ are sample means.

5. METHOD OF GENERATION

Individual sets of 150 annual flows in the standardized form were first generated by following a step by step process via equations 1 and 2 with the help of two sets of virgin normal random deviates. Subsequently corresponding dependant series were generated from equations 10 and 11. In all, three such sequences were generated viz, one series (X-series) for Pinjal and two series (Y_1 and Y_2) for Gargai. Y_1- series is the output of bivariate (two-station) model and Y_2- series is from single-site AR-1 process. This facilitates two comparisons, one between two conventional models (X & Y-2) to determine effect of dissimilar data bases, and second between Y-1 & Y-2 to critically assess practical usefulness of the former. In all 30,000 pseudo - random normal deviates were extracted from MS excel spreadsheet generated series of100000 numbers and arranged in two parallel sequences of 15,000 deviates each. First 30 values of dependent variables in each series were removed to eliminate priming bias and the balance 120 were utilized for compiling one series of maximum length of 100. A feature of both Y_1 (bivariate) and Y_2 (univariate) series is the occurrence of several non-zero values either sporadically or in a sequence of up to 4/5 numbers. All negative values and such other positive value/s less than athreshold value of 10 % of the mean of the historical data were ruled out while compiling sequences. Model estimators like the mean ($\hat{\mu}$) standard deviation ($\hat{\sigma}$), skew coefficient ($\widehat{\Upsilon}$) and serial lag coefficient ($\hat{\rho}_1$) of each(three) concurrently generated 100 sequences along with the spatial correlation coefficients$\hat{\rho}_{0\ X\text{-}Y1}$ and $\hat{\rho}_{0\ X\text{-}Y2}$ for each pair of X &Y1- series in the bivariate case and X&Y2 in the univariate case respectively were computed by method of moments and tabulated in Table-2 under respective heads (presented at the end because of size).To be sure,$\hat{\rho}_{0\ X\text{-}Y2}$ is of no relevance here and is purely of academic interest to demonstrate expected spatial non-correlation between two random concurrent time series. For this study, representative value of any estimator viz; mean ($\hat{\mu}$), standard deviation ($\hat{\sigma}$), skew coefficient ($\widehat{\Upsilon}$) or lag-1 serial correlation coefficient ($\hat{\rho}_1$) is the averaged value of that parameter over generated N=100outcomesand this value is designated with double accent $like\,\bar{\bar{\mu}}$.Thus, in all 14 estimators (4 for Pinjal X-series and 5 each for the Y_1 and Y_2 series (Gargai) have been generated and employed in analyses. A 15th estimator called the cross correlation between the random components of X & Y1 series is also added to this table (column 15)

Bias (error) and root mean square error (RMSE) are shown in Table-2 are used for assessment of inter – model and intra-model efficiency. Bias is the algebraic difference between representative estimator and its counterpart historical statistic and while RMSE is the square root value of sum of square of bias added to estimator variance $\hat{\sigma}^2$. Additionally, %bias (PBIAS)which the ratio of bias normalised by counterpart historical statistic was also employed for assessing model efficiency. Lower the magnitude of PBIAS, more accurate is the model. This study was carried out in excel spread sheet format and specific advantages to field hydrologists are that voluminous computations (involving 50,000 numbers) could be carried out without compromising on accuracy or cost and further as computations unfold results one can observe play of underlying random process. This is of benefit to hydrologists because it enables first hand appreciation of random numbers as they unfold so desirable while dealing with practical problems.

Another issue that is worth noting here is the number and length of generated samples. In practical applications there appears to be no consensus as to number of synthetic time series (N) to be generated and what should be the length of each time series (n) to firm uprepresentative estimators. Intuitively, time series of length of 100 years (n=100) is adequate in reservoir simulations because it relates to years of useful life. As far number of simulations (N) are concerned few research applications Guimaraes et al. [6] have indicated that large number of simulations running into thousands will indeed be necessary to exponentially converge on reservoir volumes for ensuring a very level of reliability viz;99.5 %. However, a reliability of 95 % is the mandated minimum for drinking water purposes by regulatory authorities hence as a practical measure a finite N=100 reservoir simulations would suffice for yield determinations each series being of length n=100.

6. DISCUSSION OF RESULTS

Credibility of results in stochastic modelling is of the essence. From a practitioner's view point foremost question that must be satisfactorily answered is that how close the generated series which is nothing but a sequence of random numbers, is to historical time series. Non- causal models replace historical time series by time series of dependant random numbers with no correlation tophysical catchment processes. In contrast, in deterministic (causal)modelling precipitation is the independent variable and resulting

https://doi.org/10.15623/ijret.2018.0704010 Received: 12-02-2018, Accepted: 26-03-2018, Published: 16-04-2018

catchment runoff is the dependant variable which are linked through simple/ complex correlations and available historical run off extended (regressed) by applying correlation equations. For testing goodness of such a mix of measured data with simulated data significant amount of literature exists that deal with evaluation criteria. For instance, Moriasi et al [7] have developed evaluation parameters like percentage bias (PBIAS)and ratio of RMSE to historical standard deviation designated as RSR along with acceptance limits. But in case of non-causal models' practical guidelines for model acceptance have not been articulated except in cases involving pseudo- normal distributions where goodness-of- fit criteria can be applied. Neither do international / Indian codes exist that prescribe uniform standards for accepting random outcomes. This is a major concern in case of P-P-P projects where primary concern is testing rigorously of reservoiryields. Hence new protocols have to be set up on case by case basis which must involve some degree of intuition and common sense. In most basin simulation studies for testing generated series, it is considered acceptable if two first order moments like mean and standard deviations come close to their historical counterparts. However skew and serial correlation are equally important in reservoir yield fixation given their high economic significance in the context of P-P-P projects. Since both are higher order moments large convergence defects must be expected and, in such cases, comparative assessment of model values is the practical way. In this light protocol selected for this study is: a) use intra-model PBIAS of $\bar{\bar{\mu}}$ and $\bar{\bar{\sigma}}$ to judge closeness of all estimators including skewness and serial correlation. In the case of primary estimator mean ($\bar{\bar{\mu}}$), error margin is set @ 5% for accepting closeness to historical mean $\bar{x}$. Similarly apply PBIAS to standard deviation ($\bar{\bar{\sigma}}$)to judge model precision. Intuitively, a lower dispersion and a lower PBIAS means greater precision. b) RMSE of each of the 14 estimators normalized by its range i.e. difference between the maximum and minimum of generated value of that estimator designated as normal RMSE (NRMSE) is taken as a measure of efficiency or performance of the model. Since RMSE is sensitive to outliers' lower values indicate less residual variance and hence better performance. Additionally, inter-model efficiency is also compared through ratio ($\bar{\bar{\sigma}}$ / $\bar{\bar{\mu}}$) termed here as model coefficient of variation (MCV)

6.1 Mean and Standard Deviation

Results presented in Table-2 show all 14 estimators viz $\bar{\bar{\mu}}$,$\bar{\bar{\sigma}}$, $\bar{\bar{\gamma}}$, and $\bar{\bar{\rho}}_1$ and$\bar{\bar{\rho}}_0$from conventional univariate time series (X- series and Y2- series) and bi-variate (Y1 -series) are negatively biased (under estimates). PBIAS of mean taken as a measure of closeness to historical statistic, varies between a negligible0.15 % - X series to 0.85 % Y-1(2 – site model) and 5.75% for Y-2. Similarly, PBIAS of standard deviation taken as a measure of precision varies from a low of 1.16 % for the X-1 to 10.36 and 8.03 for Y-1 and Y-2 respectively. Clearly the mean is preserved well in all three expansions but not to the same extent in case of standard deviation. Between the two conventionally generated series X-1 and Y-2, the former is decidedly far superior to Y-2 in terms of closeness of both primary parameters viz; mean and standard deviation. Perhaps larger data sample of Pinjal (x) has a role in this. Likewise, for smaller ~preserves historical mean both in terms of closeness and precision than its conventional counterpart. Referring to Table-2, normalized RMSE of both$\bar{\bar{\mu}}$ and $\bar{\bar{\sigma}}$ for all three expansions fall in narrow range 0.17 ∼ 0.25, X-series being the best performer withY-1 clearly scoring over Y-2 in terms of residual variance. Further confirmation of this trend can be had if one compares model coefficient of variation (MCV) of three generated series. X -series is the lowest (39%)followed by two -site model which looks a more efficient model than its counterpart single-site model. Overall trend confirms that given a good data base (about 29 items) conventional auto regressive or Markov models produce excellent outcomes in terms of primary parameters like mean and standard deviation. But If short data samples are used clearly innovative 2-site model produces markedly better results than the conventional model

6.2 Skewness

Skewness property is an important characteristic of peninsular Indian river catchments falling in wet and dry climate. Skew coefficients $\hat{\Upsilon}$of all sequences are the one set that show highest bias and dispersion. Cochran-Snedecortest for normality conducted on longer time series (n=100) demonstrates that all three X and Y 1 & Y 2 - series retain skewness @ 90 % level of significance. Both PBIAS and NRMSE of$\bar{\bar{\gamma}}_x$is the lowest @ 7.09 and 0.2respectively,while for Y-1 and Y-2 extensions PBIAS are many times higher than that of X-series with *Y-2 scoring over Y-1*. If one compares the RMSE here again the same trend is maintained and in fact NRMSE are nearly equal for X and Y-2,both univariate time series but in case of Y-1 (bi-variate case)it is significantly higher. This reversal of trend is a surprising result difficult to explain considering that in both generating schemes sample skewness are treated in the same way viz; through Wilson-Hilfertytransformation(WH) (Eq-6&7).One possible explanation could be that both random variates viz, $\lambda_{1x,\,t}$ and $\lambda_{2y,\,t}$ are generated using two parallel streams of random normal variables (Eq-6&7) but when these two combine as in bi-variate model (Eq-3), the stochastic component could either get moderated or accentuated depending on whether concurrent random variables have the same algebraic sign or opposite signs resulting in $\hat{\gamma}$ being either high or low in a series of 100 items. It would be interesting to see how this effect plays out if n is increased to say 500. This effect is absent in univariate time series X and Y2 which are generated through Eq-2,4 and 6.

6.3 Serial Correlation

Persistence represented by $\hat{\rho}_1$ is a crucial feature in judging goodness of synthetic time series. All generated $\bar{\bar{\rho}}_1$ show downward bias. PBIAS of $\bar{\bar{\rho}}_1$ is the lowest forY-1-series followed by X and Y-2 in that order. Unlike in the case of skewness coefficient, $\bar{\bar{\rho}}_1$ of Y-1is much lower that of its

https://doi.org/10.15623/ijret.2018.0704010 Received: 12-02-2018, Accepted: 26-03-2018, Published: 16-04-2018

univariate counterpart and closer to that of X-series. Additionally, NRMSE of the Y-1 series is also smaller than both X and Y-2 cases. In this light one can with some caution, assess that important persistence characteristic (lag-1)is excellently reproduced in the bi-variate model. This is an encouraging result. In non-periodic sequences, SCC's are expected to decline smoothly as lags increase. However in all three correlograms shown in Fig-2a, b, c, such a declining trend is not much in evidence. Nevertheless, it should be borne in mind that initial steep fall in SCC is very much in affirmation of underlying AR-1 process.

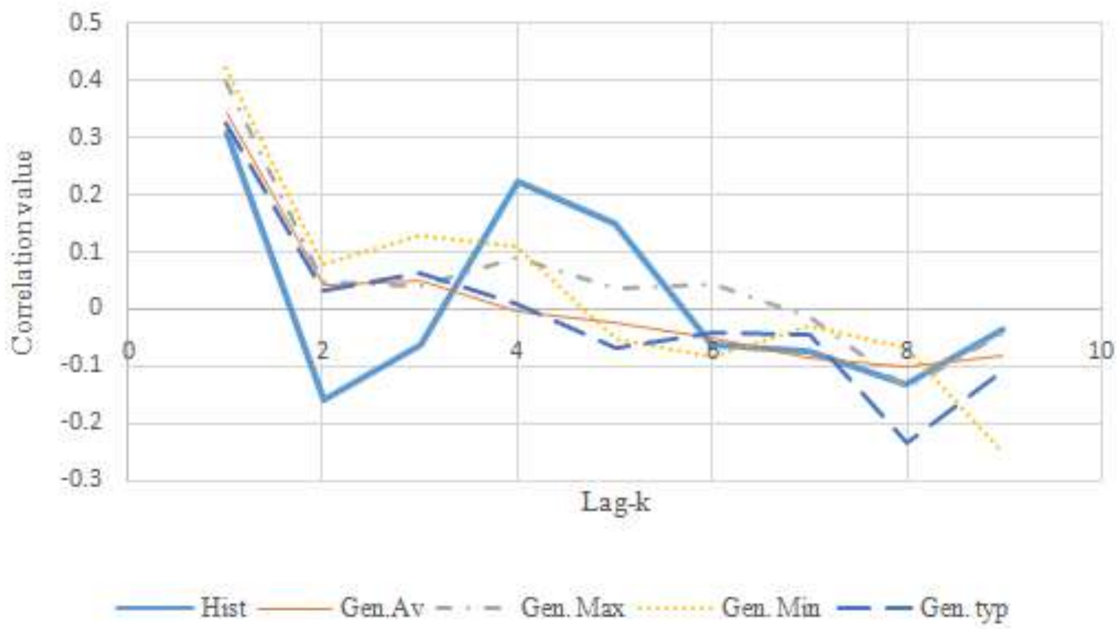

Fig 2a: Correlograms of X-Series

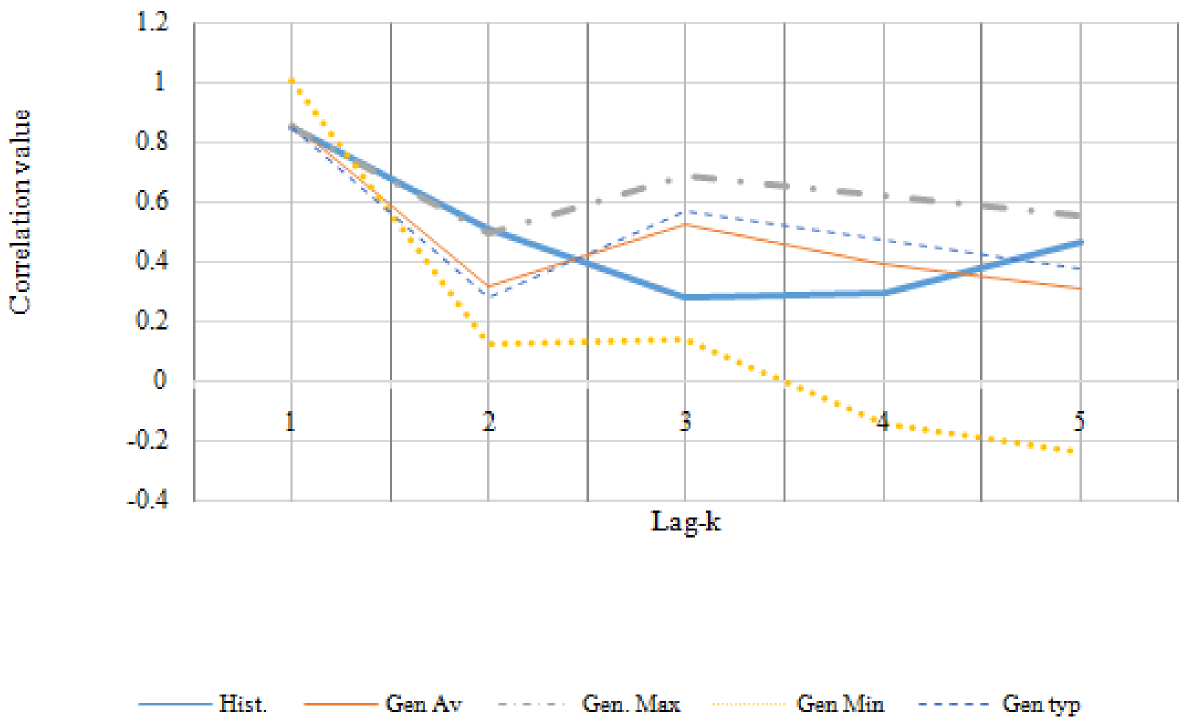

Fig 2b: Correlograms of Y-1 Bi-variate Series

https://doi.org/10.15623/ijret.2018.0704010 Received: 12-02-2018, Accepted: 26-03-2018, Published: 16-04-2018

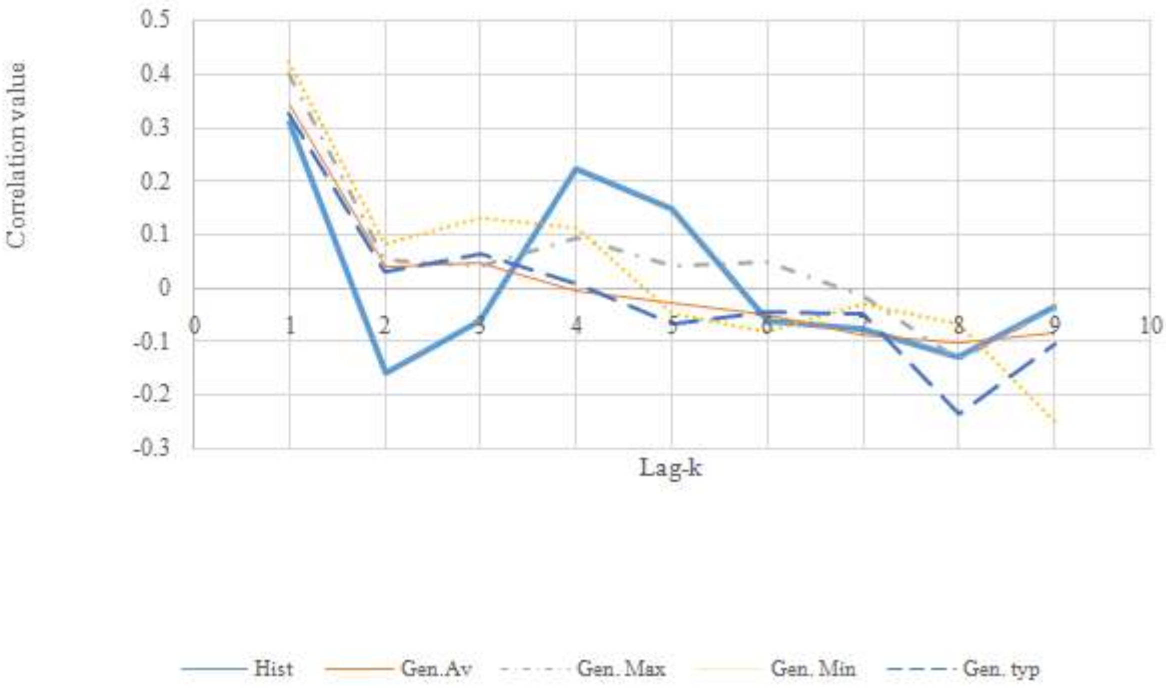

Fig 2c: Serial Correlograms of Y-2 Univariate Series

6.4 Cross Correlation

The most important property of the bi-variate model viz; lag-zero cross correlation ($\bar{\bar{\rho}}_{0\ X\text{-}Y1}$) is excellently reproduced with both PBIAS and normalized RMSE showing lower values viz; 4.87% and 0.2 % respectively. Expectedly, large bias and dispersion of $\bar{\bar{\rho}}_{0\ X\text{-}Y2}$confirmabsence of spatial relationship between two univariate random time series X and Y-2.

According to Kottegoda and Yevjevich(1977) who have pioneered research on multi-sitemodelsexistence of linear correlation between two streams of concurrent run off in the bi-variate model is owing to strength of underlying lag-zero cross correlation between stochastic components of both time series. If this insight is applied to this study it is seen that cross correlation between the random components of X and Y-2 series is higher, 0.67 as compared to $\bar{\bar{\rho}}_0\ xy_1$.In the generating equation (Eq-2&3)dependant variables are products of two components: first one is determinant component represented by the first term in both equations and second component is the random component represented by second term in equation -2 and second and third terms in equation-3. While the determinant components are driven by time invariant catchment retention property, inequation-3the random component is product of two terms(Eq-6&7), one is the time independent component λ_1x, t of the X-series and the second term is the time independent component of Y- series λ_2y, t, which is spatially independent of λ_1x, t. These terms incorporate factors α and β which in themselves are dependent on serial correlation coefficients r_{1x} and r_{1y}. Physical interpretation of moderating effect that α & βmay be having is not clear. Although in this specific instance two-site model is distinctly superior in performance to conventional Markov model and as such provides greater conviction for application, from practitioner's perspective the physical basis for spatial correlations is not clearly understood. Whether historical value of r_0 = 0.452 which is based on limited (14) pairs of concurrent annual run offs is sufficient and strong basis for choosing bi-variate scheme for generating synthetic traces in similar field situations would do well to awaitconfirmatory research studies

7. CONCLUSION

Main conclusions of the study are:

- When historical data series is adequate(29 years), synthesized longer time series (Pinjal - X series) generated through conventional Markov AR-1 model show that all four estimators are close to historical statistics and the generated X-series can be used convincingly in reservoir simulations. However it is necessary to treat skewness of the historical sample through Wilson- Hilferty (WH)transformation to achieve good outcomes.
- Generated Gargai time series (Y1 & Y2-series) based on 14 years of field runoff data, produce mixed results. When generated in bi-variate mode by treating high sample skewness through WH transformation, both primary estimators viz;mean, standard deviation and additionally the serial lag coefficient (SCC-1) are better preserved compared to corresponding estimators generated from conventional Markov modelling. However, the other third order moment viz; skew coefficients in both generated Gargai time series are

https://doi.org/10.15623/ijret.2018.0704010 Received: 12-02-2018, Accepted: 26-03-2018, Published: 16-04-2018

preserved rather poorly with the bivariate model found more wanting.

- Outcomes of Pinjal(X-series) and Gargai (Y1 and Y2-series)point to influence of length of sample sizes. Overall, bi-variate model used in this study seems an effective model whendealing with highly skewed and serially correlated smaller size data samples.Although additional research studies would be required to further substantiate this result, limited success of the innovative two-site modelis encouraging

REFERENCES

[1] Fiering M.B, (1967): *Streamflow Synthesis*, Macmillan, United Kingdom

[2] Kottegoda, N.T (1980): *Stochastic Water Resources Technology, Chapter IV, Para 4.8.1,* The Macmillan Press, London, United Kingdom.

[3] Kottegoda, N.T &Yevjevich, V: (1977): *Preservation of Correlations in Generated Hydrologic Samples through Two-Station Models*, Journal of Hydrology,33 (1977), pp 99-1

[4] Charles T. Haan, (2002): *Statistical Methods in Hydrology*, *Chapter 15*, Iowa State Press, Ames, Iowa 50010

[5] Fiering, Myron B & Barbara, B Jackson, (1971): *Synthetic Stream Flows*, *Chapter 4, Equations 41 to 44*, Water Resources Monograph Series, American Geophysical Union, Washington D.C

[6] Guemaraes, Rita Cabral, Santos, Emidio Gill (2011): *Principles of Stochastic Generation of Hydrologic Time Series for Reservoir Planning and Design - A Case Study*, Journal of Hydrologic Engineering (ASCE), Vol. 116, # 11, pp 891-898, Nov.2011

[7] D. N. Moriasi et.al (2007): *Model Evaluation Guidelines for Systematic Quantification* of *Accuracy in Watershed Simulations* Transactions of the American Society of Agricultural and Biological Engineers (ASABE), March 2007

BIOGRAPHY

Author of this paper Vasudev Shirahatti is a hydraulics and hydrological engineer having more than 35 years of experience in consulting engineering profession. He has worked on several water resources projects aided by international organizations like the World Bank and Asian Development Bank. His area of speciality is application of Operation Hydrology for determination of reliability of source yields.

Table-2: Estimators of Generated series (100 sequences each of 100 items)

	X-series				Y1 (bivariate) series					Y 2 (univariate) series					
Historical Parameters	$\bar{x}$ =869.23 (Mm3) (1)	s_x=342.9 (Mm3) (2)	g_x=0.85 (3)	r_{1x}=0.31 (4)	$\bar{y}$ =264.4 (Mm3) (5)	s_y=146.2 (Mm3) (6)	g_y=1.34 (7)	r_{1y}=0.85 (8)	r_{0xy} =0.452 (9)	$\bar{y}$=264.4 (Mm3) (10)	$s_{y=}$14 6.2 (Mm3) (11)	g_y= 1.34 (12)	r_{1y}= 0.85 (13)	$r_{0\ x\text{-}y}$ =0.452 (14)	Cross correlation-random components (15)
Generated Estimators	$\hat{\mu}$	$\hat{\sigma}$	$\widehat{\Upsilon}$	$\hat{\rho}_1$	$\hat{\mu}$	$\hat{\sigma}$	$\widehat{\Upsilon}$	$\hat{\rho}_1$	$\hat{\rho}_0$ X-Y1	$\hat{\mu}$	$\hat{\sigma}$	$\widehat{\Upsilon}$	$\hat{\rho}_1$	$\hat{\rho}_0$ X-Y2	$\widehat{\rho}_0$ X-Y1
Mean	867.91	339.02	0.79	0.29	262.18	131.05	0.62	0.80	0.43	249.23	134.16	1.00	0.74	0.00	0.67
Max. value	956.18	407.71	1.73	0.49	412.23	206.15	1.86	0.94	0.68	408.31	212.19	2.79	0.93	0.12	0.87
Min. value	742.40	259.37	0.16	0.01	157.90	71.61	-0.29	0.37	0.13	150.76	71.53	0.20	0.34	0.01	0.43
Representative value $\overline{\overline{()}}$	867.91	339.02	0.79	0.29	262.18	131.05	0.62	0.80	0.43	249.23	134.16	1.00	0.74	0.00	0.67

https://doi.org/10.15623/ijret.2018.0704010 Received: 12-02-2018, Accepted: 26-03-2018, Published: 16-04-2018

Std. Deviation	40.28	29.27	0.31	0.09	42.70	25.73	0.42	0.09	0.11	50.01	33.56	0.44	0.14	0.29	0.09
Skewness	-0.22	-0.10	-0.52	-0.31	0.76	0.36	0.31	-2.35	-	0.63	0.29	0.87	1.04	-	
Bias	-1.33	- 3.96	-0.06	-0.02	-2.25	-15.14	-0.72	-0.05	-0.02	-15.20	-12.03	-0.34	-0.11	-0.45	
Root MSE	40.30	29.53	0.31	0.09	42.76	29.86	0.83	0.10	0.11	52.27	35.65	0.56	0.18	0.46	
Coefficient of Variation of Model, % (MCV)	39.06				49.98					53.83					
PBIAS (%)	-0.15	-1.15	-7.09	-6.49	-0.85	-10.36	-53.81	-5.89	-4.87	-5.75	-8.03	-25.26	-12.85	-0.45	
NRMSE	0.19	0.20	0.20	0.19	0.17	0.22	0.39	0.18	0.20	0.20	0.25	0.21	0.30	4.88	