# SIMULATION OF REGIONAL DAILY FLOW AT UNGAGED SITES USING INTEGRATED GIS-SPATIAL INTERPOLATION (GIS-SI) TECHNIQUE

**JU YOUNG LEE [1,2] and GANESHI KRISHINAMURSH[1]**

[1] School of Environmental and Water Resource Engineering, Department of Civil Engineering, Texas A&M University, College Station, TX 77840

[2] KIST-Gangneung Branch, Seoul 130-650, Korea

**Abstract:** The Brazos River is one of the longest rivers contained entirely in the state of Texas, flowing over 700 miles from northwest Texas to the Gulf of Mexico. Today, the Brazos River Authority and Texas Commission on Environmental Quality interest in drought protection plan, waterpower project, and allowing the appropriation of water system-wide and water right within the Brazos River Basin to meet water needs of customers like farmers and local civilians in the future. Especially, this purpose of this paper primarily intended to provide the data for the engineering guidelines and make easily geological mapping tool. In the Brazos River basin, many stream-flow gage station sites are not working, and they can not provide stream-flow datasets enough for development of the Probable Maximum Flood (PMF) for use in the evaluation of proposed and existing dams and other impounding structures. Integrated GIS-Spatial Interpolation (GIS-SI) tool are composed of two parts; (1) extended GIS technique (new making interface for hydrological regionalization parameters plus classical GIS mapping skills), (2) Spatial Interpolation technique using weighting factors from kriging method. They are obtained from the relationship among location and elevation of geological watershed and existing stream-flow datasets. GIS-SI technique is easily used to compute parameters which get drainage areas, mean daily/monthly/annual precipitation, and weighted values. Also, they are independent variables of multiple linear regressions for simulation at ungaged stream-flow sites. In this study, GIS-SI technique is applied to the Brazos river basin in Texas. By assuming the ungaged flow at the sites of Palo Pinto, Bryan and Needville, the simulated daily/monthly/annual time series are compared with observed time series. The simulated daily/monthly/annual time series are highly correlated with and well fitted to the observed times series.

**Keywords**: GIS-SI, Brazos River Basin, ungaged stream-flow sites, multiple linear regression, PMF, Kriging

## 1. INTRODUCTION

Recently, the nature flow regime is the long term unregulated pattern of flow magnitude, duration, seasonality, and frequency. This pattern and daily flow information, in the ecosystem management aspects, determine the functions, integrity and biodiversity and, in the water resource management aspects, the optimal water planning-management, production of hydroelectric generation. Hence, the hydrological estimation and getting daily flow information at ungaged sites have been studying in the water resource research field. Simulations of continuous daily flow times series can be produced using semi-distributed deterministic rainfall-runoff models or multiple linear regression model and by spatial interpolation algorithm using available observed flow and precipitation data and their associated daily Flow Duration Curves (FDCs). The continuous daily flow data does not exist all the time within basin or gage systems do not set up in some of Brazos river basin of USA. From missing gage sites, computed flow data can estimate flow characteristics. It is used to estimate amount of water to maintain ecosystem conservation. For simulation of regional daily flows at ungaged sites, multiple linear regression, flow duration curves, and spatial interpolation algorithm (Huges and Smakhtin, 1996) are used for computing. These processes belong to the function of integrated GIS-SI technique, which is convenient to use.

## 2. MATERIALS AND METHODS

### 2.1 Brazos River Basin

The Brazos River begins in eastern Stonewall County, Texas, and extends along a winding trail over 1448 km to the Gulf of Mexico (Figure 1). A number of tributary rivers (e.g., Clear Fork, Bosque, Lampasas, Leon, Little River, and Navasota) contribute additional water to the flow of the Brazos. The Brazos River watershed reaches into eastern New Mexico and extends in a southeasterly arc through Texas to the Gulf of Mexico near Freeport. According to USEPA 1999, its coordinates are 28.90N and 95.37W, average depth and average salinity are 3.0 m and 28 ppt. In Texas, the Brazos River Basin is about 46,000 square miles of total drainage area. However, only about 36,000 square miles contributes to stream flow, about 1/6 of the total area of the State, and encompasses 65 counties containing a population of over 2 million. It is the longest river in Texas and freshwater discharge to the Gulf is approximately 6,784 million $m^3$ /year (Hendrickson, 2001). At its terminal end, the river mixes with tidal water from the Gulf of Mexico forming the Brazos River estuary. The river has been dammed in several places to form reservoirs for flood control, municipal use important man-made lakes on the river are the Possum Kingdom and Whitney reservoir.

### 2.2 Computation of the parameter values of multiple linear regression using GIS-SI

GIS-SI is the most recent GIS for simulation of regional daily flow at ungaged sites. It has a very user friendly interface, and because of this it is accessible even to users who are not familiar with ArcGIS and Spatial Interpolation technique. GIS-SI operates within ArcGIS 8 and computes the parameter values from digital elevation model (DEM) and weather data at gage stations; drainage area, river routing parameters and mean daily/monthly/annual precipitation at ungaged sites.

Figure 2 shows that extracting hydrologic parameter values from digital elevation model (30m) has been accomplished by extended GIS technique in GIS-SI. The computed hydrologic parameters are independent variables for multiple regression models within Spatial Interpolation algorithm. Projected 30m Digital Elevation (DEM) is the first input for extended GIS technique in GIS-SI. The flow direction grid which stores either of the 8 numbers (e.g., 1,2,4,8,16, 32, 64 or 128) is calculated based on

8 point-pour algorithm and Filling Sink/Burn-In process. This creates a raster of flow direction from each cell to its steepest down slope neighbor. The flow accumulation creates a raster which stores in each cell the number of cell flowing to it based on the flow direction raster. Based on the outlet raster and flow direction raster, delineated watershed raster is made. It will have a field Grid-Code which associates it with outlet and stream link. The final step makes basin/subbasin shape files.

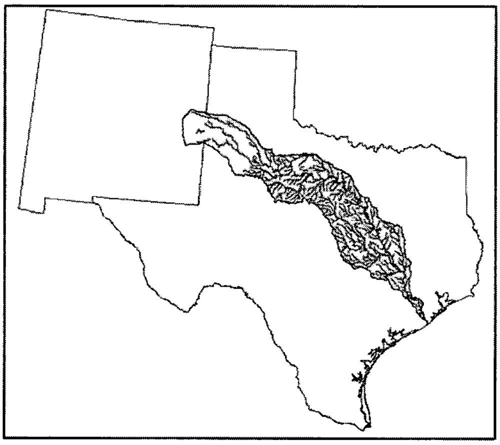


**Fig. 1 Brazos river basin**

### 2.3 Simulating the long term mean daily flow values at ungaged sites using spatial interpolation technique.

Hughes and Smakhtin (1996), Smakhtin et al, (1997), Smakhtin (1999), and Smakthin and Masse (2000) provide this technique. Spatial interpolation technique assumes that flows occurring simultaneously at sites, which are reasonably close to each other and hydrological similar, correspond to similar percentage points on their respective flow duration curves(FDCs) which is the curve that shows the relation between time excess probability and discharge corresponding to the probability. The long term mean daily flow values at ungaged sites as destination sites are generated by available FDCs at gaged sites as source sites.

#### 2.3.1. Generation of a flow duration curves (FDCs) at gaged sites (source sites)

A regional FDCs at source sites reflects regional flow variability and is relatively close to regional flow variability at the ungaged sites (destination sites). A regional FDCs and discharge at 17 point time excess probability values (0.01, 0.1, 1, 5, 10, 20, 30, 40, 50, 60, 70, 80, 90, 99, 99.9, and 99.99 %) are computed at gaged sites.

#### 2.3.2. Generation of a flow duration curves (FDCs) and a continuous flow hydrograph at ungaged sites (destination sites)

The next step is to calculate the regional FDCs based on the normalization. Conversion of FDCs into generation of a continuous flow hydrograph at ungaged sites is computed by Spatial Interpolation Technique. Hughes and Smakhtin (1996) assumed that source and destination flow close proximity to each other on their respective FDCs. Figure 3 shows the procedure how to generate a continuous flow hydrograph at ungaged site using spatial algorithm technique.

In figure 3, step 1 is an observed hydrograph, FDC in step 2 is calculated from observed hydrograph, and step 3 shows that source FDC find the discharge value associated with the same % point on the destination site FDC, step 4 shows that FDC at destination site create the interpolated hydrograph in destination site.

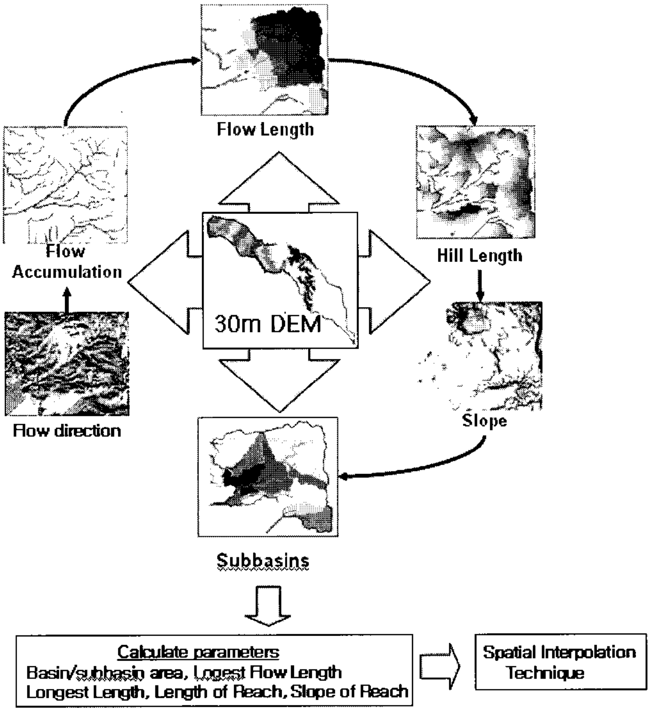



**Fig. 2 Procedure for computing hydrological parameters**

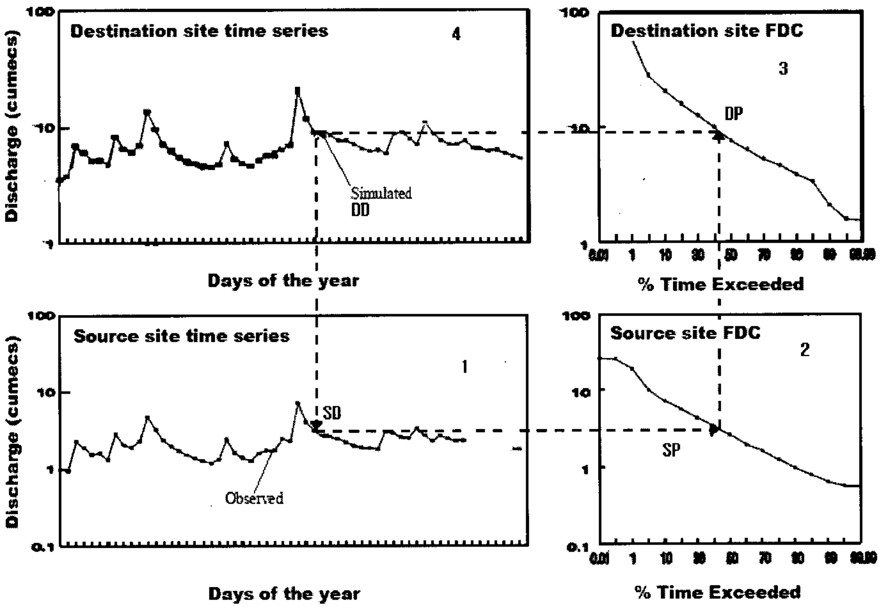



**Fig. 3 Long term daily flow generation procedure at ungaged site using Hughes & Smakthins method (1996)**

### 2.3.3 Generation of long term daily flow using GIS-SI

Measured sites are chosen from geologically closing ungaged sites within basin/subbasin using GIS technique. GIS technique provides different weighed factors to each source sites according to geological aspects (slope, distance and elevation).

The generation of a continuous mean daily flow is computed by following equation 1;

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2$$
$$InQ_{mn} = -\beta_0 + \beta_1 InA + \beta_2 In[MAP/MMP/MDP] \quad (1)$$

where $Q_{mn}$ is continuous mean daily flow(cfs), A is drainage area (mi$^2$), MAP/MMP/MDP are Mean annual/month/daily precipitation (in) from 1960 to 2004, $\beta_{0,1,2}$ are parameters(weighting factors) for linear log fitting. The dataset which is drainage area and MAP/MMP/MDP are then input to GIS-SI technique for calculating discharge-area and discharge-precipitation relationships and formula as equation 1. In the Brazos River, The mean daily flow (cfs) is well fitted linear log relationship among mean daily flow ($Q_{mn}$) and area (A). Especially, Parameters ($\beta_{0,1,2}$ ) for linear log fitting are averaged 0.043, 0.895 and 0.453 and $R^2$ is 93.54 in the Brazos River Basin, Texas.

Equation 2 is used that $Q_{mn}$ at ungaged sites in large basin is computed by weighting the area of different precipitation. Meanwhile, $Q_{mn}$ at ungaged sites within small subbasin is computed by weighting the area of same precipitation;

$$Q_{mn} = \sum_g w_g \frac{Q^g_{mn}}{A_g} A_s \quad (2)$$

$Q_{mn}$ flow at ungaged site
$A_s$ upstream drainage area
$g$ gage site

$Q^g_{mn}$ gage flow
$A_g$ gage drainage area
$w_g$ weights

To obtain weighting factors, spatial interpolation technique is important method. Especially, kriging technique is more precise than other method. Kriging defines that it assigns all of weights from measured data at gage station sites to predict ungaged data at ungaged stations. Kriging method depends on mathematical and statistical aspects. Its basic concept is autocorrelation and distance. Spatial unmeasured data is simulated by their distance and measured MAP/MMP/MDP data are computed as autocorrelation response to a function of distance. The equation 2 is derived from the empirical semivariance values of measured data are fitted as semivariogram as measured gage stations in Brazos River Basin.

Figure 4 shows how to make the simulation of long-term mean daily flow ($Q_{mn}$) at ungaged site using GIS-SI. The process is following like that.

In the first step, it is preparation step for collecting data. Also, it is obtained from extended GIS technique (new making interface for hydrological regionalization parameters plus classical GIS mapping skills). Drainage area and hydrologic parameters are calculated from 30m DEM using extended GIS within GIS-SI like figure 2. For grid data of MDP/MMP/MAP at gaged sites, they are obtained from satellite rainfall data by converting process.

In the second step, Spatial Interpolation technique chooses kriging method and mathematically and geostatistically analyzes the relationship between an ungaged site and

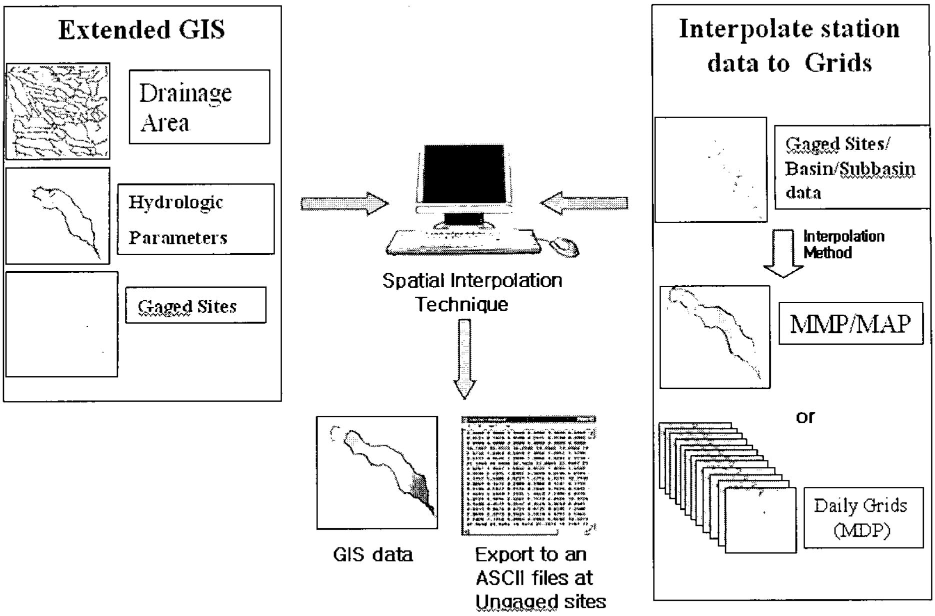



**Fig. 4 The processing procedure for simulation regional daily flow at ungaged site using GIS-SI**

simultaneously within basin or subbasin. After that, long term mean daily flow at ungaged site is computed and exported to ASCII files. For validation, the Nash-Sutcliffe coefficient of efficiency (Nash and Sutcliffe, 1970) is. NS coefficient of efficiency has been widely used to evaluate the comparison of hydrologic simulated and observed value. The coefficient of determination ($R^2$) describes that the proportion of the total variance in observed data that can be explained by the simulation.

$$deviation = \frac{\sum Q_{sim} - \sum Q_{obs}}{\sum Q_{obs}} * 100$$

$$efficiency = 1 - \frac{\sum (Q_{obs} - Q_{sim})^2}{\sum (Q_{obs} - \overline{Q}_{obs})^2} \quad (3)$$

$$R^2 = \left| \frac{\sum_{i=1}^{N} (Q_{obs} - \overline{Q_{obs}})(Q_{sim} - \overline{Q_{sim}})}{\left[\sum_{i=1}^{N} (Q_{obs} - \overline{Q_{obs}})^2\right]^{0.5} \left[\sum_{i=1}^{N} (Q_{sim} - \overline{Q_{sim}})^{0.5}\right]} \right|^2 \quad (4)$$

$Q_{obs}$ : Observed Q

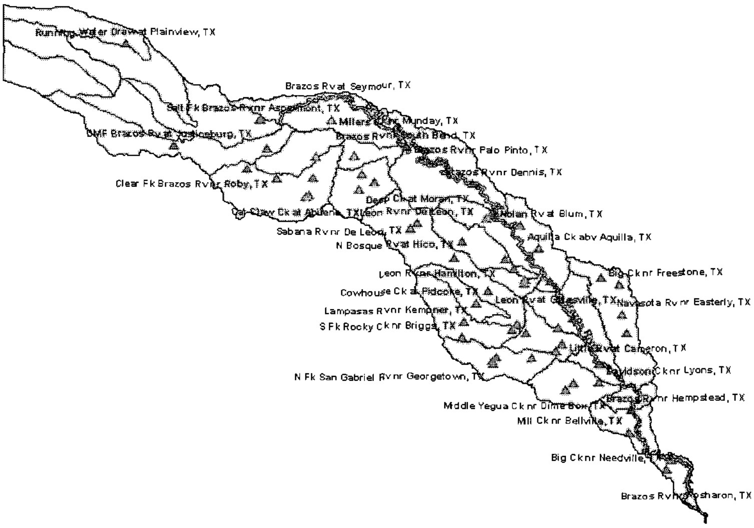
$\overline{Q_{obs}}$ : Observed mean Q

$Q_{sim}$ : Simulated Q

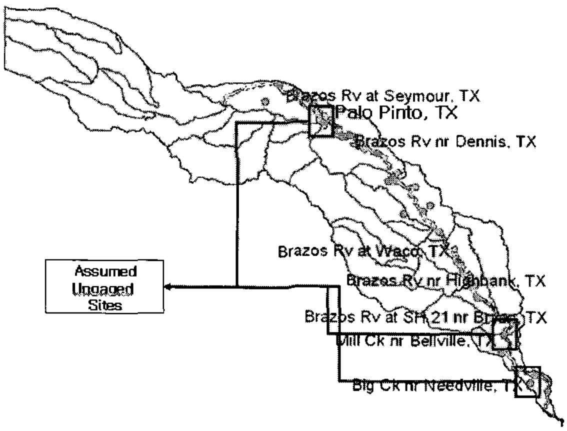
$\overline{Q_{sim}}$ : Simulated Mean Q

## 3. RESULTS AND DISCUSSION

In the figure 5, The Brazos river length is 840 miles and the objective station are at Seymour, Dennis, Palo pinto, Waco, Highbank, Bryan, Bellville, and Needville in Texas. Also, the site Palo pinto, Bryan and Needville are assumed to be ungaged site for the purpose of comparison between simulated and real data. Hence, Palo pinto, Bryan and Needville are simulated by near weighted neighbor stations. Table 1 and Figure 6 show the value of independent variables (Mean annual precipitation (MAP), Drainage area) and the information of geological parameters. In table 1, the drainage area and mean annual precipitation (MAP) are obtained from USGS. Meanwhile, table 2 and 3 show




(a)



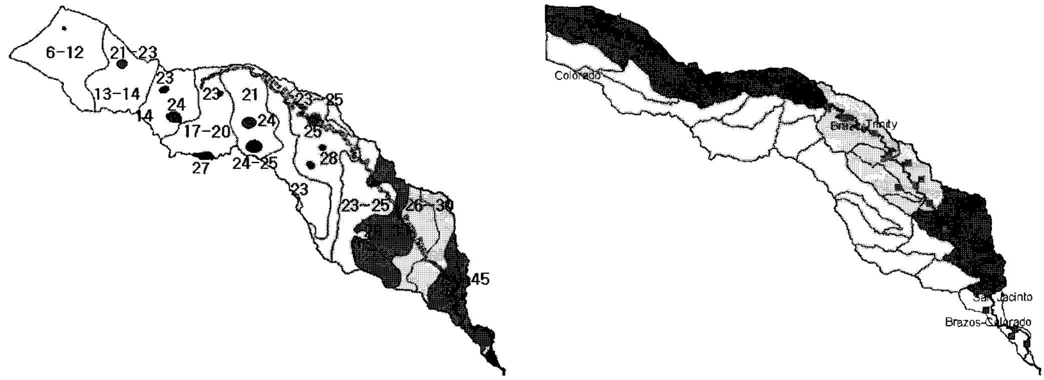

(b)

**Fig. 5 The gaged station (a), the objective stations at ungaged sites in the Brazos river basin (b)**

**Table 1. Characteristics of gaged sites**

| Site Name | Drainage Area(sqmi) | Lat. | Long. | Duration of data | Qmn (cfs) | MAP (in) |
|---|---|---|---|---|---|---|
| Seymour | 20,457 | 33 | 98.6 | 1960~04 | 281.23 | 21.21 |
| Palo Pinto* | 23,811 | 32.8 | 98.3 | 1960~04 | 789.14 | 24.32 |
| Dennis | 24,123 | 32.6 | 98.2 | 1960~04 | 1011.5 | 25.0 |
| Waco | 26,734 | 32.2 | 97.7 | 1960~04 | 1438.5 | 24.68 |
| Highbank | 29,378 | 32 | 97.1 | 1960~04 | 2033.98 | 26.68 |
| Bryan* | 35,325 | 31.4 | 96.4 | 1960~04 | 6912.15 | 27.43 |
| Bellville | 43,260 | 30.1 | 96.2 | 1960~04 | 6205.3 | 36.2 |
| Needville* | 50,631 | 29.6 | 95.7 | 1960~04 | 8831.2 | 42.2 |

* The objective station sites (Needville, Bryan, and Palo Pinto)




(a) (b)

**Fig. 6 Mean annual precipitation (in) (a), major subbasin in Brazos river basin (b)**

**Table 2. Simulated value from GIS-SI and observed value for drainage area ($mi^2$)**

| Palo Pinto($mi^2$) | | | Bryan($mi^2$) | | | Needville($mi^2$) | | |
|---|---|---|---|---|---|---|---|---|
| Obs.* | Sim.* | RE (%)* | Obs. | Sim. | RE (%) | Obs. | Sim. | RE (%) |
| 23,811 | 23,124 | 2.9 | 35,325 | 33,645 | 4.8 | 50,631 | 49,123 | 3 |

* Obs. (Observation), Sim. (Simulation), RE (Relative Error)

**Table 3. Simulated value from GIS-SI and observed value for long term mean annual flow (cfs)**

| Palo Pinto(cfs) | | | Bryan(cfs) | | | Needville(cfs) | | |
|---|---|---|---|---|---|---|---|---|
| Obs.* | Sim.* | RE (%)* | Obs. | Sim. | RE (%) | Obs. | Sim. | RE (%) |
| 789.14 | 770.24 | 2.4 | 6912.15 | 6676.43 | 3.4 | 8831.2 | 8490.9 | 3.8 |

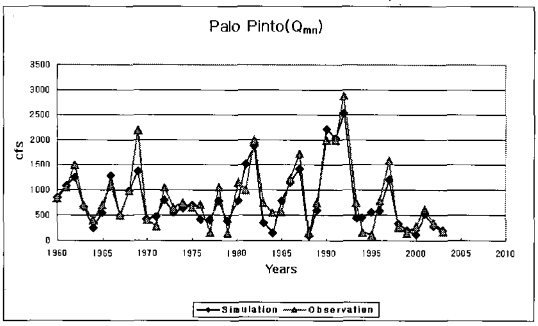
* Obs. (Observation), Sim. (Simulation), RE (Relative Error)

**Table 4. Simulated value from GIS-SI and observed value for validation**

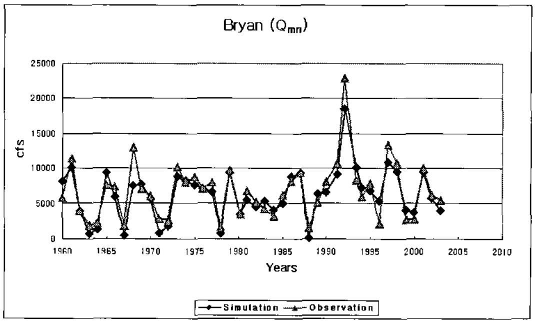
| Palo Pinto | | Bryan | | Needville | |
|---|---|---|---|---|---|
| CE* | *$R^2$ | CE | $R^2$ | CE | $R^2$ |
| 0.87 | 0.98 | 0.77 | 0.95 | 0.83 | 0.97 |

* CE: Coefficient of efficient, $R^2$: Coefficient of determination

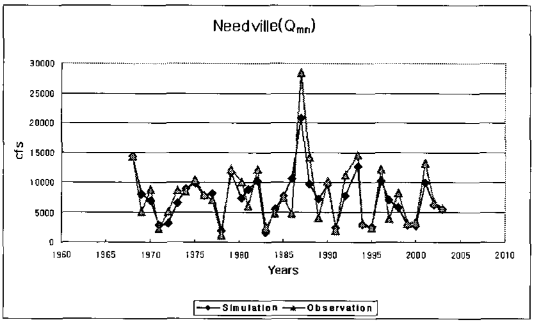
comparison of observed value with obtaining values from GIS-SI. The two independent variables and long term mean daily flow at Palo Pinto, Bryan, and Needville are obtained from GIS-SI. Also the result of computing long term mean annual flow at three sites and the low relative error mean that it is fairly closed. Actually, this paper should show long term mean daily flow as well as long term mean annual flow, but long term mean annual flow are only mentioned except long term mean daily flow. Figure 7 show comparison between observed and computed hydrograph. Though two hydrograph are similar, they are significantly different. In case of observed hydrograph, it is wigglier than simulated hydrograph because




(a) Palo Pinto site




(b) Bryan site




(c) Needville site

**Fig. 7 Comparison between observed and computed hydrograph from 1960 to 2004**

simulated hydrograph are obtained from spatial interpolation algorithm in GIS-SI. Also, in the high flow region of hydrograph, the simulated value underestimates in 3 gaged sites.

The NS coefficients of efficiency (NSCE) are 0.87, 0.77, and 0.83 at Palo Pinto, Bryan, and Needville. Smakhtin *et al* (1997) shows if NSCE is higher than 0.6, the simulation result can be regarded satisfactory. The values of coefficient of determination are 0.98, 0.95 and 0.97 at Palo Pinto, Bryan, and Needville. $R^2$ is the criteria of consonance between the real value and the simulated value. It ranges from 0 (poor model) to 1 (perfect model). The $R^2$ shows that the simulated value is satisfactory to observed value. The NSCE and $R^2$ are criteria for validation test. The NSCE is improved over for model evaluation purposes because it is sensitive to differences in the observed and simulated means and variances as equation 3 and 4. Therefore, NSCE is sensitive to extreme values as is $R^2$.

This purpose of the GIS-SI is computing conveniently and exactly the long term means daily/monthly/annual hydrographs at ungaged sites. Most of all, its results are more exact than manual computational method like labor intensive work. GIS-SI are composed of two categories; one is extended GIS technique which is performing to get independent variables which are drainage area, hydrologic parameters, MAP/MMP/MDP, and determination of weight- ing factors to gaged sites near ungaged sites. The other is spatial interpolation technique to simulate hydrographs at ungaged sites using historical data at gaged sites. Through this paper, the simulated hydrographs are highly correlated with and fitted to the observed hydrographs and GIS-SI is a very convenient tool for analysis of regionalization.

## 4. CONCLUSION AND SUMMARY

The purpose of this paper primarily intended to provide the generated data for engineering guidelines more precise and make easily geological mapping tool. In the Brazos River basin, many stream-flow gage station sites are not working, and they can not provide stream-flow datasets enough for development of the Probable Maximum Flood (PMF) for use in the evaluation of proposed and existing dams and other impounding structures. Also, many citizen and farmers interest in drought protection plan, waterpower project, and allowing the appropriation of water system-wide and water right within the Brazos River Basin to meet water needs of customers in the future. Then, GIS-SI technique is a convenient tool for all users. Therefore it is the role of the planning and supporting document (guidelines) for water planners

## REFERENCE

Hughes, D. A. and Smakhtin, V. Y. (1996) Daily Flow data Time Series Patching or Extension: a Spatial Interpolation Approach Based on Flow Duration Curve, *Hydrol. Sci. J.* 41(6), pp 851-871.

Hendrickson. K.E., Jr. (2001). Handbook of Texas Online. Brazos River, the Texas State Historical Association, 1997-2001

Nash, J.E. and Sutcliffe, J.V. (1970) River flows forecasting through Conceptual Models Part 1-a Discussion of Principles, *J. Hydrol.* 10, pp 282-290.

Smakhtin, V. Y., Hughes, D.A., and Creuse-Naudin, E. (1997). Regionalization of daily flow characteristics in part of the Eastern Cape, South Africa. *Hydrol. Sci. J.* 46(6): 919-936.

Smakhtin, V. Y. and Masse, B. (2000). Continuous daily hydrograph simulation using duration curves of precipitation index. *Hydrological Processes*, 14: 1083-1100.

Smakhtin, V.Y. (1999). Generation of natural daily flow time-series in regulated rivers using a non linear spatial interpolation technique. Regulated Rivers: Research & Management, 15:311-323.

Smakhtin, V. Y. (2000). Estimating daily flow duration curves from monthly streamflow data. Water SA, 26(1): 13-18.

USEPA, 1999. Ecological condition of estuaries in the Gulf of Mexico. EPA 620-R-98-004, U.S.Environmental Protection Agency, Office of Research and Development. National Health and Environmental Effects Research Laboratory, Gulf Ecology Division, Gulf Breeze, Florida. 80pp.

---

To whom correspondence should be addressed.

(E-mail: JUYOUNGLEE@tamu.edu)