Stochastic River Flow Modelling and Forecasting of Upper Indus Basin

Hamza Khan[1] and Syed Ahmad Hassan[2,*]

[1]*Mathematical Sciences Research Centre, Federal Urdu University of Arts, Sciences and Technology, Karachi-75300, Pakistan*

[2]*Institute of Industrial Electronics Engineering (IIEE), PCSIR, St-22/c, Block # 6, Gulshan-e-Iqbal, Karachi-75300, Pakistan*

Abstract: Upper Indus Basin (UIB) region has faced seasonal and sometimes unpredictable disastrous flow in their tributaries and contributing one of the world's largest Indus River System. As these streams emerged from high mountains of Hindukush, Karakorum and Himalaya ranges, and formed as a lifeline for the local population of which 90% is accommodate by Indus River system source. A little change in the regional climate may cause floods and outburst flows in the river and affects the lives, regional ecosystem and long part of the Karakoram highway. On the other hand, the shortage of water in Pakistan can create an alarming condition in future because a huge amount of eastern glaciers is shrinking. According to UN-ICC 2011 report Pakistan is in top of the four risky countries which adversely affected by climate change and especially worst hit by bi-catastrophes in 2010.During summer, intensifications of temperature may dissimilar for different locations/altitudes but it affects the glaciated areas. Moreover, the summer river flow and precipitation in previous winter and spring seasons has significant correlation shows their influence in the UIB region. Consequently, it may also be responsible for fluctuation in the seasonal/regular flows of UIB Rivers. To study these variations this paper analyses two types of data, mean monthly and 4-times moving average of monthly for long-term forecast. They belong to two different rivers Ghizer-Gilgit at Gilgit and Ghizer-Gilgit-Hunza at Alam stations. Both types of data illustrate a strong seasonal cycle. Therefore, seasonal autoregressive integrated moving average (SARIMA) models of time series method have been used. The five selected SARIMA models explore 90% and more river flow forecast. Moreover, the result with 4-times moving average is more accurate than simple data.

Keywords: Upper Indus Basin, SARIMA, Time series, Monsoon, Stochastic.

1. INTRODUCTION

Agro dominated economy of Pakistan is based on one of the world largest integrated irrigation network of Indus River system. They originate from the world highest Gallicized mountain ranges of Hindukush, Karakorum and western Himalayan regions. This is supposed to be the largest area outside the Polar and Greenland region [1]. The total catchments area of upper Indus Basin (UIB) is about 206000 km^2, out of which 22000 km^2 comprises of perennial snow/ice and glaciers [2]. Water from this area is lifeline for local as well as downstream population and accommodates approximately 90% of Indus River system source. Sometimes, a little change in the local climatic condition may cause floods and outburst flows (responsible for abrupt rise in the river in fraction of hours) and become hazardous for the whole region.

There are more than six main tributaries in the UIB region (like Astor and Shyok Ghizer-Gilgit) whose flow is largely depends on the timing and volume of local precipitation. When usual seasonal temperature increases, the melting of remote glaciers and snow covers also rises and they form flood in UIB almost every year [3]. However, in some years disastrous flood arrive and affect the activities of local population and long part of the Karakoram highway. On the other hand, the shortage of water in Pakistan can create an alarming condition in future because a huge amount of eastern glacier is shrinking. It is surprising in view of report of the UN International Climate change conference held in Durban 2011that "Pakistan has been putted on top of four risky countries which are adversely affected by climate change and worst hit by bi-catastrophes caused by extreme weather events like in 2010".

Moreover, during summer (June to August) 2011 and 2013 floods, Pakistan has faced dangerous situation when half of lower Sind submerged under floodwaters and in the later year, Karachi was experiencing heaviest rains of the decade. It may happen due to the combined effect of the climatically influenced dynamical changing weathers on mountainous area and unexpected monsoonal rainfall.

Most of the time snow covers dynamics and hydrological regime of UIB regions contributing complex activities to the river flow network regularly [2]. Consequently, the area has faced fluctuated rivers flow

*Address correspondence to this author at the Institute of Industrial Electronics Engineering (IIEE), PCSIR, St-22/c, Block # 6, Gulshan-e-Iqbal, Karachi-75300, Pakistan; Tel: +9221-99244218 (Ext. 112); E-mail: syed_a_hassan@yahoo.com
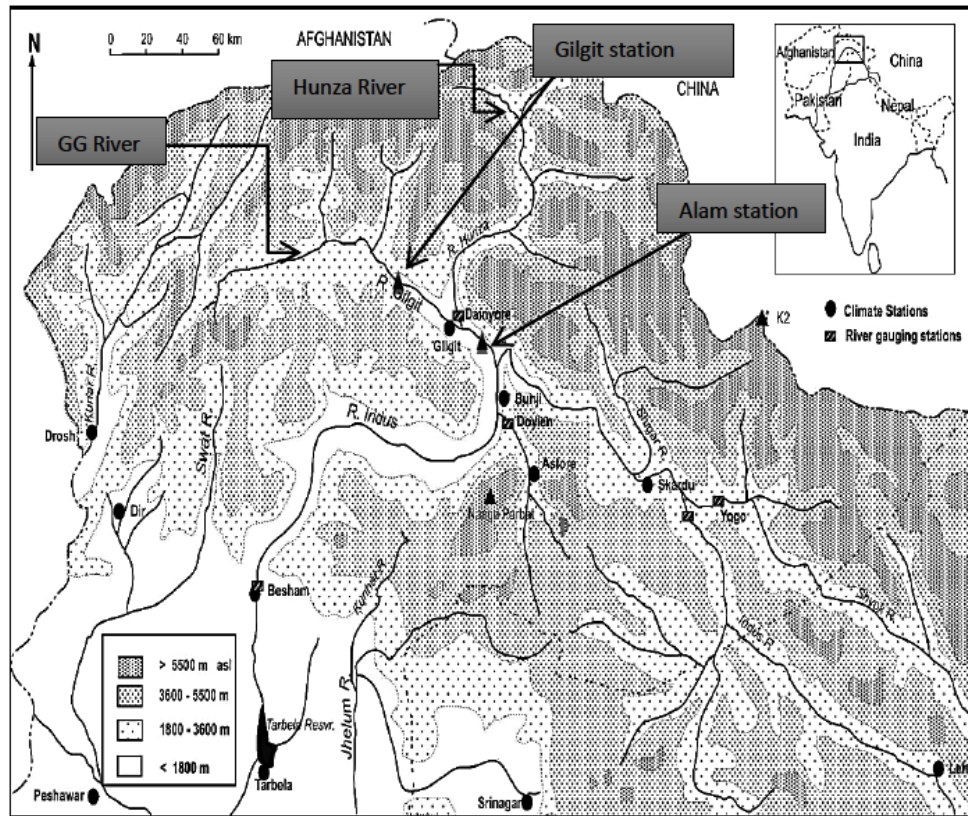
Figure 1: Study area of UIB (Archer & Fowler, 2004).

in their tributaries and the main Indus River. It can be proved that climate change has potential impact on water resources of UIB in both freezing and melting forms [4]. Naturally, the situation calls for a better forecasting of a potential flood, much earlier before its formation. Forecasts of temperature, wind-speed and precipitation, it being correlated to solar activity, expressed by sunspot numbers [3, 5] a few days ahead contribute short-term predictions only. The available temperature and precipitation data of nearby glaciated areas make river flow forecast significant in UIB region. However, due to rough terrain areas and high hazardous rocky Mountains, installation of observatories to measure local climatic parameters is difficult. Moreover, due to the many short and long term variations in the network, complex hydrology with incomplete information about their process, the real time assessments of river flow forecast is considerably uncertain [6]. The investigation of a single time series (univariate analysis) is also important in hydrological modelling [7]. Now this paper utilizes the concept of time series model of auto-regressive integrated moving average (ARIMA) methods. Briefly, ARIMA stochastic model comprises of autoregressive (AR) process, consider the memory of previous events; an integrated (I) process, which describes stabilizing or making the

data stationary, helps to easier forecast; and a moving average (MA) of the forecast errors, support to accurate forecast with conditioned to large historical data. ARIMA has three model parameters, p, d and q for AR, I and MA processes respectively. All these are joined and interrelate with each other and recomposed into ARIMA (p, d, q) model. In order to estimate river flow (flood), it is particularly important to optimize the seasonal forecasting of UIB stream flows. In view of the above, this paper presents modelling and forecasting of rivers flow of UIB using seasonal autoregressive integrated moving average (SARIMA) time series models [8] for the long term forecast. As this model provide good results due to their efficient seasonal adoptability and self-regressed. Seasonal forecasting is a better way to save water, reduce poverty, food security improvement, health, environmental security and management of power generation [9]. The method could provide significant benefits for the management of national power strategies by providing an early indication of surplus or shortfall in hydropower which would require balancing with thermal power source [10]. We will study the impact of climate variables and the method for seasonal forecasting, considering UIB as a case study and the local climates as parameters in our future communication. The data available from

1985-to-2008 of mean monthly Ghizer-Gilgit (GG) River flow at Gilgit station and while 1969-to-2008 of Ghizer-Gilgit-Hunza (GGH) River flow at Alam station (Figure **1**). This paper organizes as follows. In the next section, we present the Time series method. In Section 3, we show the main empirical results and discussion. Finally, we draw some conclusions in Section 4.

2. MATERIAL AND METHODS

2.1. Time Series Method

The hydrological time series data are mostly autocorrelated [11]. Their supported mathematical modelling of linear time series method of hydrological data is most appropriate [12]. Archer [13] also recommends this method of modelling for the UIB region. The most traditional time series models are autoregressive (AR) and moving average (MA) methods [14, 15]. However, some times it is necessary to consider mixed autoregressive and moving average models, ARMA (data assumed to be stationary) and its extended form autoregressive integrated moving average ARIMA (data assumed to be non-stationary) [14, 15, 16]. Some modification to basic ARIMA model involving 'seasonal differencing' allows the model to include seasonal component that is SARIMA (seasonal autoregressive integrated moving average) model, which is of much interest for this study. A shortcoming of these methods is that the relationship between the parameters cannot be permanently represented linearly [17]. The SARIMA time series modeling method, utilized Box & Jinkens [16] scheme. It has three main steps Identification, Estimation & Testing, adequacy of proposed model and Forecasting. Autocorrelation function (ACF) and partial autocorrelation function (PACF) plots are used to identify the most appropriate formation and order of the SARIMA $(p, d, q) \times (P, D, Q)_s$ model [8, 16]. Where, "p", "d" and "q" represent simple autoregressive, differencing and moving average components respectively whereas, "P", "D" and "Q" of all three belongs the same representation for seasonal order "s". The ACF plot shows the existence of the autoregressive and their significant initial lag values recognize the order "q" of the moving average components. Although, the significant initial lag values of simple PACF plot helps to identify the order "p" of autoregressive and behavior of the initial lags indicates the existence of the moving average components [18]. The integrated component appears when we need to stationary the data series using differencing of order "d". Moreover, the ACF plot also indicates the seasonal variations by periodic change of their values [16].

The river flow time series plot (Figure **2**) has a fluctuation cycle and seasonal variations. Both GG & GGH river flow data series shows some autoregressive and moving average components including appropriate significant seasonal variation as depicted in Figures **3**, **4** & **5**. They identify and suggested different order of SARIMA $(p, d, q) \times (P, D, Q)_s$ models described as;

$$\theta(B)\Delta^h\Theta(B^s)\Delta_s^H \varkappa_t = \varphi(B)\Phi(B^s)\xi_t \tag{1}$$

Where, B is the backward shift operator defined as $B\varkappa_t = \varkappa_{t-1}$ and $\Delta = 1 - B$ is the differencing operator. The other terms denote polynomial functions, defined as

$$\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_p B^p),$$

$$\varphi(B) = (1 - \varphi_1 B - \varphi_2 B^2 - \ldots - \varphi_q B^q),$$

$$\Theta(B^s) = (1 - \Theta_1 B^s - \Theta_2 B^{2s} - \ldots - \Theta_p B^{ps}),$$

$$\Phi(B^s) = (1 - \Phi_1 B^s - \Phi_2 B^{2s} - \ldots - \Phi_q B^{qs}),$$

Whereas, θ, φ, Θ and Φ are the parameters of AR, MA, seasonal AR, and seasonal MA of degree p, q, ps and qs respectively; ξ_t is a purely random process with zero mean, constant variance, and no serial correlation (white noise). h and H are the non-seasonal and seasonal degrees of differencing respectively.
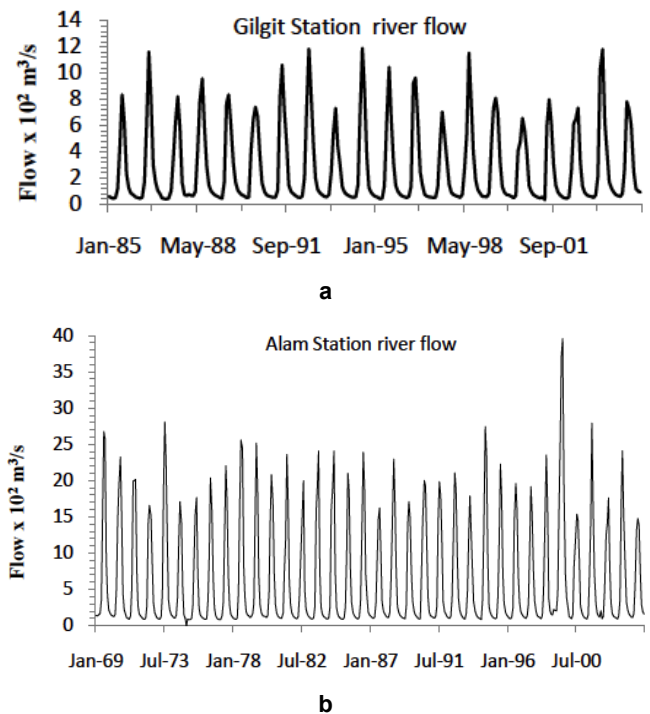

a

b

Figure 2: Mean monthly flow (**a**) GG River at Gilgit Station during Jan.1985 to Dec. 2004 (**b**) GGH River Alam station during Jan.1969 to Dec. 2004.
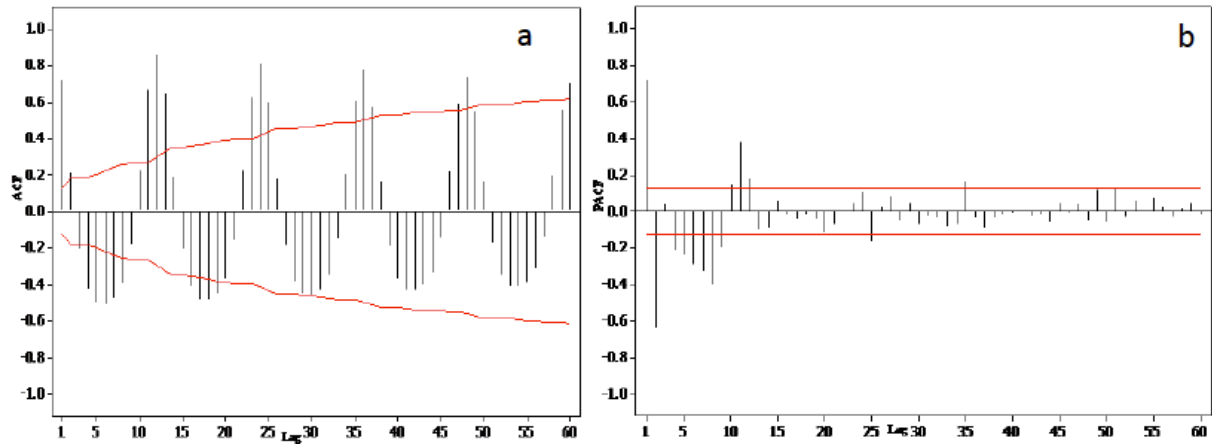
Figure 3: Correlograms of GG River flow at Gilgit station for simple data during 1969-2004. (**a**) ACF (**b**) PACF.
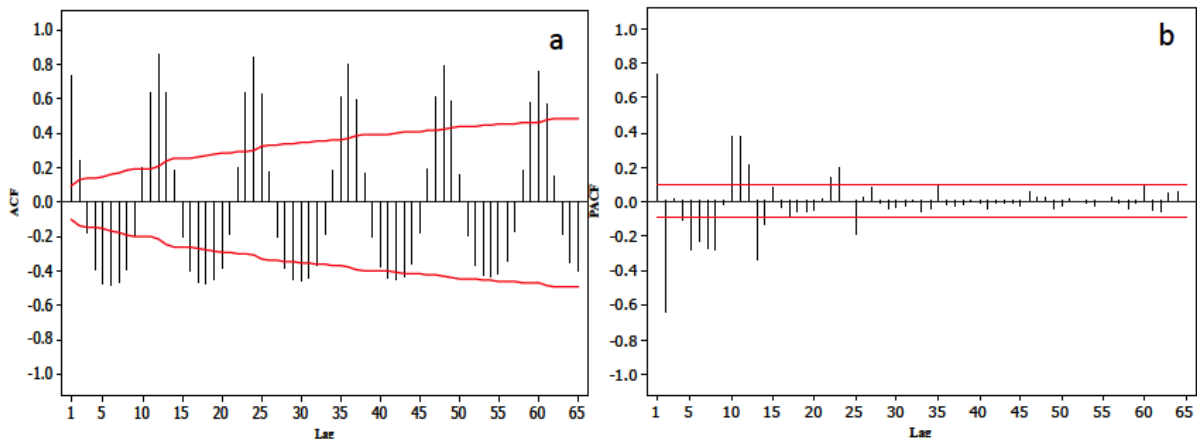

Figure 4: Correlograms of GGH River flow at Alam station during 1985-2004. (**a**) ACF (**b**) PACF.
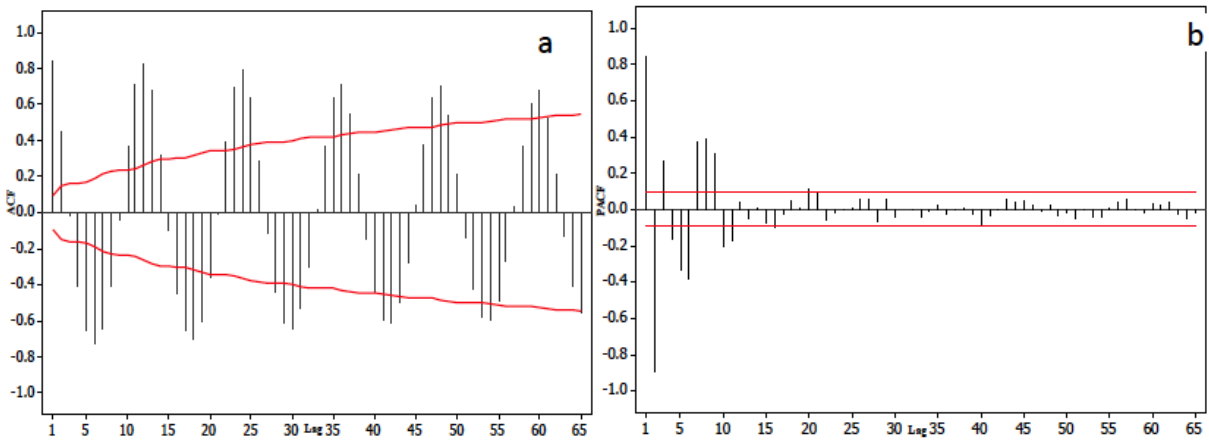

Figure 5: Correlograms of 4 months moving average GGH River flow at Alam station during 1969-2004. (**a**) ACF (**b**) PACF.

The observed data of UIB rivers shows significant PACF plot values (Figures **3b**, **4b**, & **5b**) and exponential decay behavior in the preliminary lags with seasonal in the long run of ACF (Figures **3a**, **4a**, & **5a**) correlograms. Based on these criteria we have selected some models. The estimation of most optimized parameter of the models is achieving using E-Views software [19]. Moreover, the testing of the suggested models based on the prediction parameters statistical tests of R-square, Durban Watson, Akaike Criterion [20] and Schwarz Criterion [21] as shown in Table **1**.

Table 1:   Tests Statistics Values for Seasonal Time Series Models of GG River Flow at Gilgit Station & GGH River Flow at Alam Station

Station	Data Type	S. No.	Model	R-Sqr.	DW	AIC
GG River at Gilgit Station	Simple	1	SARIMA $(2, 0, 1) \times (1, 0, 0)_{12}$	0.83	1.999	12.53
		2	SARIMA $(2, 0, 2) \times (1, 0, 0)_{12}$	0.84	1.995	12.49
GGH River at Alam Station	Simple	3	SARIMA $(2, 0, 1) \times (1, 0, 0)_{12}$	0.87	2.02	14.04
		4	SARIMA $(2, 0, 2) \times (1, 0, 0)_{12}$	0.87	1.998	14.05
	4-times M.A.	5	SARIMA $(2, 0, 1) \times (1, 0, 1)_{12}$	0.96	1.98	11.75

For GG River flow at Gilgit station two SARIMA models has selected. For these models, 240 data points (Jan. 1985 to Dec. 2004) have taken for modeling while 48 data points (Jan. 2005 to Dec. 2008) will be utilized for forecasting purposes.

The first and second selected model are SARIMA $(2, 0, 1) \times (1, 0, 0)_{12}$ and SARIMA $(2, 0, 2) \times (1, 0, 0)_{12}$ as shown in Eqn. 2 and 3 respectively;

$$(1 - \theta_{g,1}^1 B^1 - \theta_{g,2}^1 B^2)(1 - \Theta_{g,1}^1 B^{12})G_t^1 = (1 - \varphi_{g,1}^1 B^1)E_t^1 \quad (2)$$

$$(1 - \theta_{g,1}^2 B^1 - \theta_{g,2}^2 B^2)(1 - \Theta_{g,1}^2 B^{12})G_t^2 = (1 - \varphi_{g,1}^2 B^1 - \varphi_{g,2}^2 B^2)E_t^2 \quad (3)$$

Where, G_t^1 and G_t^2 represents GG River flow at Gilgit station and their superscripts shows the first and second model.

In the next, we consider GGH River flow at Alam station with two types of data one is monthly where 432 (Jan. 1969 to Dec. 2004) data points have been taken for modeling while 48 (Jan. 2004 to Dec. 2008) data points for forecasting purpose. Other is 4-times moving average of monthly data with 429 data points (March 1969 to Nov. 2004) have taken for modeling while 45 data points (Mar. 2005 to Nov. 2008) for forecasting purpose.

The two models for Alam station monthly data are SARIMA $(2, 0, 1) \times (1, 0, 0)_{12}$ and SARIMA $(2, 0, 2) \times (1, 0, 0)_{12}$ are represented as in Eqn. 4 and 5 respectively;

$$(1 - \theta_{a,1}^1 B^1 - \theta_{a,2}^1 B^2)(1 - \Theta_{a,1}^1 B^{12})A_t^1 = (1 - \varphi_{a,1}^1 B^1)E_t^1 \quad (4)$$

$$(1 - \theta_{a,1}^2 B^1 - \theta_{a,2}^2 B^2)(1 - \Theta_{a,1}^2 B^{12})A_t^2 = (1 - \varphi_{a,1}^2 B^1 - \varphi_{2,2}^2 B^2)E_t^2 \quad (5)$$

Where, A_t^1 and A_t^2 represents the first and second model of GGH River flow at Alam station respectively.

While, the model selected for 4-times moving average data of GGH River at Alam station is SARIMA $(2, 0, 1) \times (1, 0, 1)_{12}$ as shown in Eqn. 6;

$$(1 - \theta_{a,1} B^1 - \theta_{a,2} B^2)(1 - \Theta_{a,1} B^{12})M_t = (1 - \varphi_{a,1} B^1)(1 - \Phi_{a,1} B^{12})E_t \quad (6)$$

Where, M_t represent GGH River flow 4-times moving average at Alam station.

To check the adequacy of the proposed model employed ACF and PACF analysis of the prediction residuals. There is no significant spike of ACF, PACF occur, and no outliers are found in the residual data of all five models. The adequacy of the models has sustained. Finally, the forecasting procedure is performed in the next section.

3. RESULT & OUTLOOK

The forecasted result for simple data of GG River flow at Gilgit station is achieved (Table **1**) through SARIMA $(2, 0, 1) \times (1, 0, 0)_{12}$ model with Correlation = 0.903, SSE = 3071764 and MSE = 63995 (Figure **6** & Table **2**). While the result of their second SARIMA $(2, 0, 2) \times (1, 0, 0)_{12}$ model with Correlation = 0.902, SSE = 3069275 and MSE = 63933 is achieved (Figure **6**).

Moreover, the first SARIMA $(2, 0, 1) \times (1, 0, 0)_{12}$ model for simple data of GGH River flow at Alam station (Table **2**) provide Correlation = 0.929, SSE = 6666412 and MSE = 138884 forecast results (Figure **7**). The second SARIMA $(2, 0, 2) \times (1, 0, 0)_{12}$ model of this river gives Correlation = 0.931, SSE = 6501579 and MSE = 135450 forecast results (Figure **7**). It is observed from the above analysis that for GG River flow at Gilgit station SARIMA $(2, 0, 1) \times (1, 0, 0)_{12}$ and for GGH River flow at Alam station SARIMA $(2, 0, 2) \times (1, 0, 0)_{12}$ model shows good forecast results.

The best forecast result for 4-times moving average data of GGH River flow at Alam station is achieved with first selected SARIMA $(2, 0, 1) \times (1, 0, 1)_{12}$ model with Correlation = 0.965, SSE = 723242 and MSE = 16072.06 (Figure **8** & Table **2**).
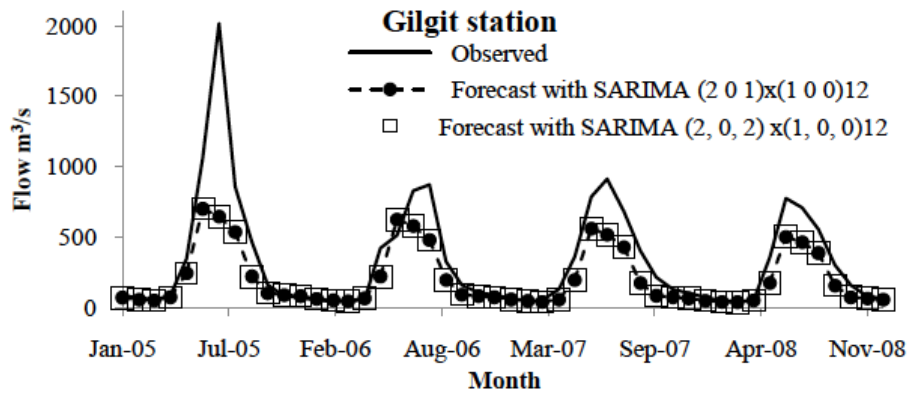
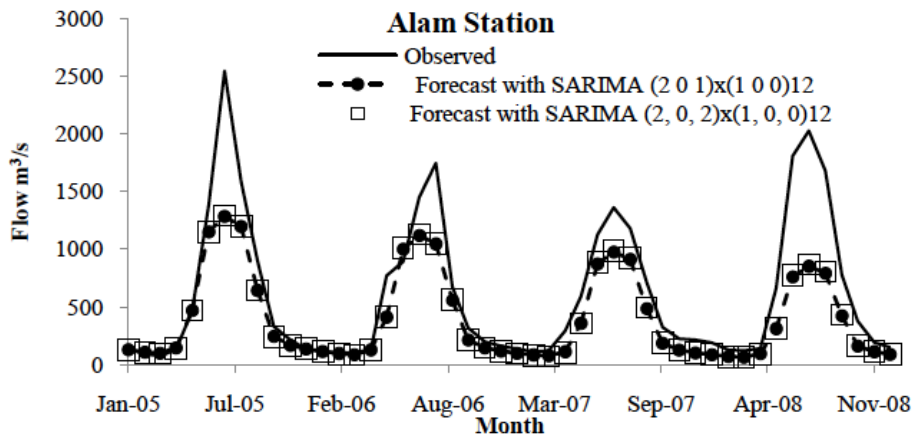**Figure 6:** SARIMA models forecast in comparison to observed GG River flow at Gilgit station for mean monthly data (Jan., 2005-Dec., 2008).

Table 2:   **Forecasting Results of GG and GGH River Flows at Gilgit and Alam Station**

Station	Data Type	S. No.	Model	Correlation	SSE	MSE
GG River at Gilgit Station	Simple	1	SARIMA $(2, 0, 1) \times (1, 0, 0)_{12}$	0.903	3071764	63995
		2	SARIMA $(2, 0, 2) \times (1, 0, 0)_{12}$	0.902	3069275	63933
GGH River at Alam Station	Simple	3	SARIMA $(2, 0, 1) \times (1, 0, 0)_{12}$	0.929	6666412	138884
		4	SARIMA $(2, 0, 2) \times (1, 0, 0)_{12}$	0.931	6501579	135450
	4-times M.A.	5	SARIMA $(2, 0, 1) \times (1, 0, 1)_{12}$	0.965	723242	16072

Figure 7: SARIMA models forecast in comparison to observed GGH River flow at Alam station for simple mean monthly data (Jan., 2005-Dec., 2008).

4. CONCLUSION

Total five seasonal time series models are selected for better forecasting purpose. Moreover, models for GG River flow at Gilgit station provide 90% whereas, models for GGH River flows at Alam Station provide more than 90% forecasting results for simple mean monthly data respectively (Tables **1** & **2**). For the long term forecasts and water management purposes the best result for 4-times moving average data of GGH

River flows at Alam station is achieved. Efficient forecasting models provide vast benefits in agricultural and power generation sectors of Pakistan that always depend on water resources. This type of forecasting is very essential for the current and future water management of Pakistan. A better management scheme can help to overcome the catastrophic conditions. Forecasting of water resources using time series models in such areas can help us to understand the existing scenario of weather change.
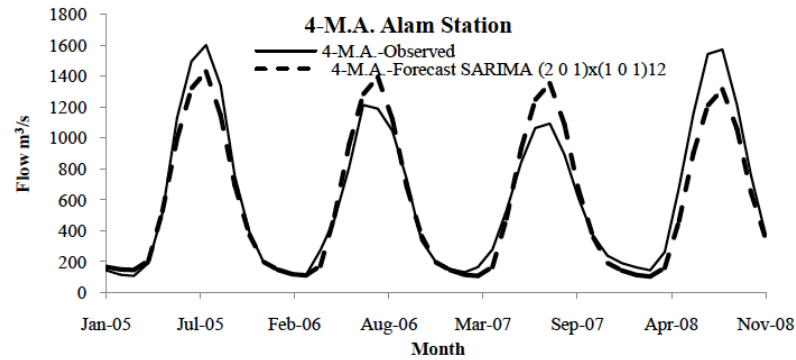
Figure 8: SARIMA model forecast in comparison to observed 4-months moving average (4-M.A.) GGH River flow at Alam station for simple mean monthly data (March., 2005-Nov., 2008).

ACKNOWLEDGEMENT

REFERENCES

[1]   Hewitt K. Tributary glacier surges: An exceptional concentration at panmah glacier, karakoramhimalaya. J Glaciol 2007; 53: 181-188.
http://dx.doi.org/10.3189/172756507782202829

[2]   Tahir AA, Chevallier P, Arnaud Y, Ahmad B. Snow cover dynamics and hydrological regime of the Hunza River basin, Karakoram Range, Northern Pakistan. Hydrol Earth Syst Sci. Discuss 2011; 8: 2821-2860.
http://dx.doi.org/10.5194/hessd-8-2821-2011

[3]   Konecny F. 2 A daily streamflow model based on a jump-diffusion process. New uncertainty concepts in hydrology and water resources 2007; 225.

[4]   Fowler HJ, Archer DR. Conflicting signals of climatic change in the upper Indus basin. J Climate 2005; 19: 4276-4293.
http://dx.doi.org/10.1175/JCLI3860.1

[5]   Anthes RA. Meterology, 7/e, Upper Saddle River, NJ: Prentice-Hall, 1997; pp. 31-50

[6]   Minarik B, Martinka K. Reduction of uncertainities in the flood estimation in the Czechoslovak section of Danube river. In Flood Hydrology, Springer Netherlands, 1987; pp. 187-195.
http://dx.doi.org/10.1007/978-94-009-3957-8_15

[7]   Browning Keith A, Robert J. Gurney. Global energy and water cycles. Cambridge University Press 1999.

[8]   Hassan SA, Ansari MRK. Nonlinear analysis of seasonality and stochasticity of the Indus River. Hydrol Sci J 2010; 55: 250-265.
http://dx.doi.org/10.1080/02626660903546167

[9]   Eastham J, Kirby M, Mainuddin M, Thomas M. Simple water use accounting of the Indus Basin. Challenge Program on Water & Food 2008.

[10] Archer DR, Fowler HJ. Using meteorological data to forecast seasonal runoff on the River Jhelum, Pakistan. Journal of Hydrology 2008; 361: 10-23.
http://dx.doi.org/10.1016/j.jhydrol.2008.07.017

[11] Salas JD. Analysis and modeling of hydrologic time series. Handbook of hydrology 1993; pp. 1-72.

[12] Chatfield C. The analysis of time series: an introduction. CRC press 2013.

[13] Archer D. Contrasting hydrological regimes in the upper Indus Basin. Journal of Hydrology 2003; 274: 198-210.
http://dx.doi.org/10.1016/S0022-1694(02)00414-6

[14] Tong H. Non-linear time series: a dynamical system approach. Oxford University Press 1990.

[15] Pandit SM, Shien-Ming W. Time series and system analysis with pplications. New York: Wiley 1983.

[16] Box GE, Jenkins GM. Time Series Analysis: Forecasting and Control. Holden-Day, San Francisco, California, USA 1994.

[17] Solomatine DP, Velickov S, Wust JC. Predicting water levels and currents in the North Sea using chaos theory and neural networks. In proceedings of the congress-international association for hydraulic research 2001; pp. 353-359.

[18] Makridakis S, Wheelwright SC, Hyndman RJ. Forecasting methods and applications. John Wiley & Sons 2008.

[19] EViews (7.0v) EViews software available from: http://www.eviews.com. Quantitative Micro Software (QMS), Irvine, California, USA 2010.

[20] Akaike H. A new look at the statistical model identification. IEEE Transaction on Automatic Control 1974; 19: 716-723.
http://dx.doi.org/10.1109/TAC.1974.1100705

[21] Schwartz G. Estimating the dimensions of a model. Ann. Statist 1978; 6: 461-464.
http://dx.doi.org/10.1214/aos/1176344136