RESEARCH ARTICLE   OPEN ACCESS

Projection of future Temperature and Precipitation for Jhelum river basin in India using Multiple Linear Regression.

Mehnaza Akhter
Deptt.of Civil Engineering, National Institute of Technology Srinagar, J&K, India.

ABSTRACT
In this paper, downscaling models are developed using a Multiple Linear Regression (MLR) for obtaining projections of mean monthly temperature and precipitation for Jhelum river basin. Precipitation and temperature data are the most frequently used forcing terms in hydrological models. However, the available General Circulation Models (GCMs), which are widely used nowadays to simulate future climate scenarios, do not provide those variables to the need of the models. The purpose of this study is therefore, to apply a statistical downscaling method and assess its strength in reproducing current climate and project future climate. Regression based downscaling technique was usedtodownscaletheCGCM3, HadCM3 and Echam5 GCMpredictionsoftheA1B scenario for the Jhelum river basin located in India. The Multiple Linear Regression (MLR) model shows an increasing trend in temperature in the study area until the end of the 21st century. The average annual temperature showed an increase of 2.37°, 1.50°C and 2.02°C respectively for CGCM3, HadCM3 and Echam5 models over 21st century under A1B scenario. The total annual precipitation decreased by 30.27%, 30.58°C and 36.53% respectively for CGCM3, HadCM3 and Echam5 models over 21st century in A1B scenario using MLR technique. The performance of the linear multiple regression models was evaluated based on several statistical performance indicators.
Keywords: Downscaling, Precipitation, Regression, Climate change, Global Climate Models, Scenario.

I. INTRODUCTION

The issue of climate change is one of the hot topics getting the attention of almost every media outlet since over the last few decades. The discussions are more or less supported by the outputs from Global Climate Models (GCMs) under different emission scenarios that are usually used in impact assessments. General Circulation Models (GCMs) are the most powerful tools available to simulate evolving and future changes in the climate system. GCMs are able to simulate reliably the most important mean features of the global climate at planetary scales. Global circulation models (GCMs) are numerical models that represent the large-scale physical processes of the earth-atmosphere-ocean system and have been designed to simulate the past, present, and future climate [4,8&9].They play a crucial role in generating future projections of climate change using different emission scenarios (Hashmi et al., 2009). However, GCMs are available at a coarse grid resolution of 2° to 4°. Consequently, products of GCMs cannot be used directly for climate impact assessment on a local scale. This has led researches to undertake the development of suitable downscaling methodologies to transfer the GCM information to local scale information. The basic assumption of downscaling is that the large scale atmospheric characteristics highly influence the local scale weather, but that in general the reverse effects from local scales upon global scales are negligible

and thus can be disregarded. There are more than 200 GCMs available which have been developed by different agencies. Input data requirement for the these GCMs are generally same but the output results vary and sometimes with slight variation in input parameters (which may be due to different data collection agencies) the results are contradictory giving confusing future climate scenarios. Despite recent improvements in modeling of the climate dynamics with complex and large-scale models, use of GCMs is still limited in evaluating regional details of climatic changes. For generating future climate scenarios on regional basis there are downscaling models called RCMs which use output of GCMs. However, RCMs do not give basin level scenarios.

1.3 Downscaling

Downscaling is the general name for a procedure to take information known at large scales to make prediction at local scales. It can be achieved in two ways such as dynamic downscaling and statistical downscaling (Christensen et al. 2007). Statistical downscaling is a two step process consisting of a) the development of statistical relationships between local climate variables (e.g., surface air temperature and precipitation)and large-scale predictors(e.g., pressure fields),and b)the application of such relationships to the output of Global climate model (GCM) experiments to simulate local climate characteristics in the future.

Dynamical downscaling is a technique which involves extraction of small scale regional information from large scale GCM output. For the derivation of small scale regional information dynamical downscaling models use a limited area, high resolution model which is driven by boundary conditions. Domain area for Regional Climate Model (RCM) is generally 106 km^2 having a spatial resolution of 20 km to 60 km. Various methods have been employed to derive relationships in statistical downscaling to forecast different climate information in different parts of the world. Such methods include canonical correlation analysis, multiple linear regressions, artificial neural networks and support vector machine. The rest of this section summarizes some previous studies on statistical downscaling methods.

Wilby and Wigley (1997) analyzed the correlation between the predictor and predictand, and selected the predictors that explained the high variance in the precipitation occurrence. These selected predictors were then used to downscale the GCM to get daily precipitation.

Ojha, etal., (2010) used multiple linear regression (MLR) and artificial neural networks (ANN) models for downscaling of precipitation for lake catchment in arid region in India. The results of downscaling models show that precipitation is projected to increase in future for A2 and A1B scenarios, whereas it is least for B1 and COMMIT scenarios using predictors.

Aksornsingchai and Srinilta (2011) have explored three statistical techniques to predict both rainfall and temperature using multiple linear regression (MLR), support vector machine with polynomial kernel and support vector machine with radial basis function. Five predictors namely; temperature, pressure, precipitation, evaporator and net short wave radiation were selected as predictor.

Zulkarnain Bin Hassan(2011) described the application of statistical downscaling method (SDSM) to downscale rainfall and temperature at north of Peninsular Malaysia. The stations in Bukit Merah at Kerian Perak and Ipoh Perak, Malaysia have been selected as a study site to test the methodology for rainfall and temperature respectively. The SDSM model estimates that there was increased total average annual rainfall at station in Bukit Merah and annual average temperature for station in Ipoh, which indicated that the area of station will be wet and warm in the future. The temperature is suspected to rise to 3°C for that area. However, several months such as May, Nov and Dec shows that the drought may happen in the future.

Fiseha,etal.,(2012). Two statistical downscaling techniques, namely regression based downscaling and the stochastic weather generator, were used to downscale the HadCM3, GCM predictions of the A2 and B2 scenarios for the Upper Tiber River basin located in central Italy. The Statistical Downscaling Model (SDSM) based downscaling shows an increasing trend in both minimum and maximum temperature as Well as precipitation in the study area until the end of the 2080s. Long Ashton Research Station Weather Generator (LARS-WG) shows an agreement with SDSM for temperature, however, the precipitation shows a decreasing trend with a pronounced decrease of summer season that goes up to 60% in the time window of the 2080s as compared to the current (1961-1990) climate.

In the present study Multiple Linear Regression (MLR) technique was employed for downscaling of monthly mean temperature and precipitation data of 4 national stations for base time (1979–2009) and then the future scenarios generated up to 2100. Observed as well as Predictors data were calibrated and tested on individual/multiple basis through linear regression. Future scenario was generated based onCGCM3, HadCM3 and ECHAM5monthly data forA1B story line (IPCC 2007).The downscaled data has been tested, and it has shown a relatively strong relationship with the observed data for all three GCMs. The predictors as obtained from Global Climate model (GCMs) were: mslpas (mean sea level pressure), tempas (mean temperature at 2m), humas (specific humidity at 2m), relative humidity (rhum), zonal velocity component (u), meridional velocity component (v).

II. STUDY AREA AND DATA USED
2.1 Study Area

River Jhelum flows through India and Pakistan having a total length of about 725 kilometers. River Jhelum rises from Verinag spring situated at the foot of the Pir Panjal in the south-eastern part of Kashmir valley in India. It flows through Srinagar city, the capital of Jammu and Kashmir and the Wular lake before entering Pakistan. It ends in a confluence with the river Chenab. Srinagar city which is the largest urban centre in the valley is settled on both the sides of river Jhelum and is experiencing a fast spatial growth. The river Jhelum and its associated streams that drain the bordering mountain slopes together constitute the drainage network of the study area. They include the fairly developed systems like Sind, Rambiara, Vishaw and Lidder rivers as well as tiny rivulets such as the Sandran, Bringi and Arapat Kol. The study area chosen spatially lies between 33° 21' 54" N to 34 ° 27 ' 52" N latitude and 74° 24' 8" E to 75° 35' 36" E longitude with a total area of 8600.78 km^2. Fig. 1 shows the catchment map of study area.
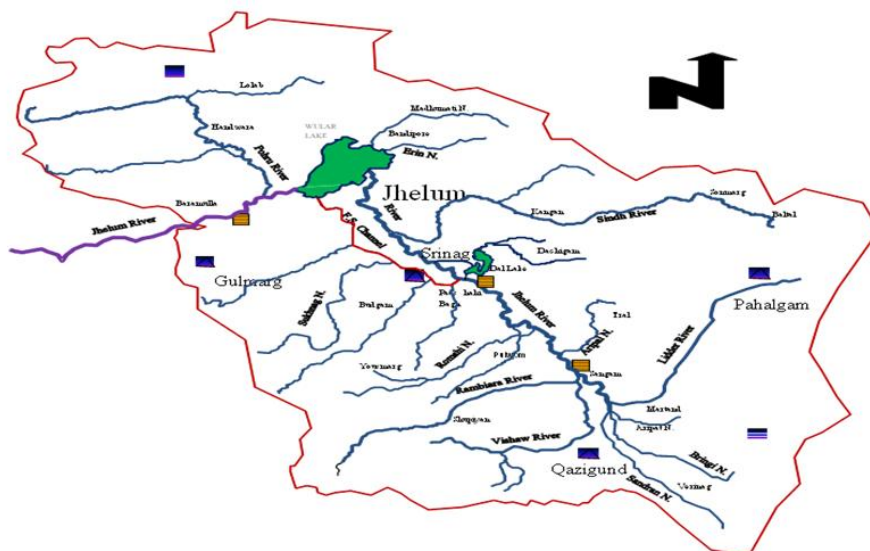
Fig.1. Catchment map of Jhelum river basin

2.2 The Datasets

The historical monthly precipitation and temperature data for the study area were obtained from the National Climate Data Centre (NCDC) website and Indian Meteorological Department (IMD), Rambagh Srinagar. The meteorological stations (Table1), where continuous datasets are produced and which represent the Jhelum basin, are selected. The GCM predictors data for the baseline period and the projection period were downloaded for selected GCM model and selected future scenario from Canadian Climate Data and Scenarios(CCDS) website http://ccds-dscc.ec.gc.ca/ and www.cccsn.ec.gc.ca/.Table 2 shows the list of models used each driven by the SRE scenarios, to capture the possible range and trend of changes

Table 1: The latitude and longitude of the meteorological stations in the Jhelum river basin

S.no	Station name	Latitude(°N)	Longitude(°E)	Elevation (m a.s.l)
1	Srinagar	34.09	74.79	560
2	Qazigund	33.63	75.15	1670
3	Pahalgam	34.01	75.19	2740
4	Gulmarg	34.15	74.25	2690

Following three models are selected for the study; CGCM3, HadCM3 and Echam5. GCM data was downloaded from the CCDS portal.

Table 2. Selected GCM models and their attributes

Model	Centre Name	GCM Resolutions (Long.° vs. Lat.°)
CGCM3	Canadian Global climate model,version3	3.75×3.75
HadCM3	Hadley centre coupled model,version3	3.75×2.5
Echam5	'EC' being short for 'ECMWF' ECMWF stands for European Centre for Medium-range Weather Forecasting Ham stands for Hamburg the place of development of its parameterisation package.	2.8×2.8

The input variables include the following GCM predictors:
1. Tempas (mean temperature at 2m)
2. Mslpas (mean sea level pressure),
3. Humas (specific humidity at 2m),
4. Relative humidity(rhum),
5. Zonal velocity component(u),
6. Meridional velocity component(v)

The GCM predictors of tempas, u, v and rhum were found to be most appropriate dependent variables for temperature and those of tempas, humas, mslpas and u for precipitation after analysis and p-test.

III. METHODOLOGY

Statistical downscaling involves developing quantitative relationships between large-scale atmospheric variables (predictors) and local surface variables (predictands).The most common form has the predictand as a function of the predictor(s), but other types of relationships, such as between predictors and the statistical distribution parameters of the predictands or between predictors and frequencies of extremes of the predictand have also been used. Shortly, the large-scale output of a GCM simulation is fed into a statistical model (as MLR) to estimate the corresponding local and regional climate characteristics. A Multiple Linear Regression Model is a linear Model that describes how a y-variable relates to two or more x-variables. A linear model is one that is linear in the beta coefficients, meaning that each beta coefficient simply multiplies to an x-variable.

Multiple linear regression analysis was performed using MS-Excel Data Analysis Tool Pack and Mini Tab 15 statistical software. The best fit is calculated for the given dataset. The regression model used for prediction is given in Equation below:

$$y = a_0 + a_1 x_{1+...} a_p x_p + \varepsilon$$

Where, y is the dependent (or response) variable, x is independent (or predictor) variable and ε is the error term.

The MLR software performs the task of statistical downscaling of weather data through six discrete processes;

The steps for downscaling can be summarized as follows (Wilby and Dawson, 2004):

1. Screening of downscaling predictors
2. Model Calibration
3. Generation of current weather data using observed data for validation.
4. Statistical analysis of results
5. Generation of future weather using GCM-derived predictors
6. Graphical representation of outputs

In the present study the multiple linear regression technique was employed to relate the GCM predictors with the predictands such as the locally observed precipitation and temperature at four meteorological observatories namely Srinagar, Pahalgam, Qazigund, and Gulmarg of Kashmir valley India. Minitab 15.0 statistical software was also used for regression analysis. The large scale GCM predictors were related to locally observed precipitation and temperature.

The various steps that were undertaken to develop the MLR model are given as under:

1. The data was first normalized using MATLAB 7.10.0 software.
2. The normalized data was transferred to Minitab work sheet.
3. Using the Statistical tool box, regression was conducted on the data.
4. The regression analysis was done and regression equations were obtained for each month from January-December for temperature and precipitation.
5. The MLR model was validated by plotting the temperature vs. time and precipitation vs. time curve for the period 2010-2013 between predicted and observed values.
6. The Analysis of Variance was done and R^2, Mean Square Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Deviation (MAD) were computed.

IV. RESULTSAND DISCUSSIONS:

The study pertained to the projection of future temperature and precipitation over 21st century in Jhelum river basin using Multiple Linear Regression technique. The MLR model was applied on the spatially averaged mean monthly temperature and monthly rainfall of Jhelum river basin. The historical precipitation data for the period 1979 to 2009 was used for calibrating the regression model, whereas the data for the period 2010-2013 was used for validating the model. GCM predictors were taken as input parameters for projecting temperature and precipitation over 21st century for Jhelum river basin. Multiple Linear Regression was done on the normalized data using MS Excel and Minitab 15 statistical software.

From Fig.2 it is clear that the observed and predicted values of temperature varied in the same direction throughout the validation period for all three GCMs. The downscaled monthly mean temperature shows an increasing trend in all months except for October and November for the period 2001–2100 for the CGCM3 and HadCM3 under A1B scenario and for August, October and November for the period 2001–2100 for A1B scenario. In general, the annual mean temperature shows an increasing trend (Figure 6).The average annual temperature will increase by 2.37°,1.50°C and 2.02°C respectively forCGCM3, HadCM3 and Echam5 models over 21st century in A1B scenario. Figure 4.1-4.3 shows the monthly projections of temperature in Jhelum river basin over 21st century by using CGCM3, HadCM3 and Echam5 model. The maximum increase in temperature over coming century was predicted by CGCM3. Also maximum rise in temperature was found in the months of March, May and July by all the three models using MLR. CGCM3 predicted pronounced increase in temperature in the month of May also. The rainfall amounts generally show a decrease trend throughout the year especially in the month of January and August and a small increase for

the month of October for the period 2001–2100 for the CGCM3 model for A1B scenario. For HadCM3 A1B scenario precipitation decreases in all months except for April and June for the period 2001–2100.For Echam5 A1B scenario precipitation decreases in all months except for the month of October for the period 2001–2100.FromFig.7 the total annual precipitation will decrease by 36.53%, 30.58°C and 36.27% respectively for CGCM3,

HadCM3 and Echam5 models over 21st century in A1B scenario.(Fig 5.1-5.3) shows the monthly projections of precipitation in Jhelum river basin over 21st century by using CGCM3, HadCM3 and Echam5 model. The maximum decrease in precipitation over coming century was predicted by CGCM3 and Echam5.Also maximum decrease in precipitation was found in the months of January, March, May and August by all the three models using MLR.
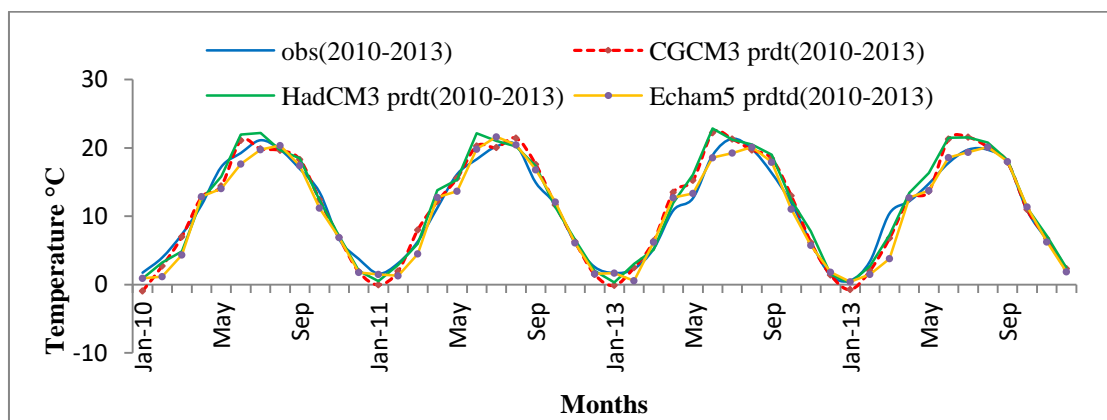
Fig.2 Validation of mean monthly temperature of Jhelum river basin for the period 2010-2013 using MLR.

From the statistical parameters of temperature validation using MLR model, value of MSE ranges from 0.0448-4.3685 for CGCM3 model,0.3943-12.6593 for HadCM3 model and 0.0766-14.2766 for Echam5 model. Similarly the value of RMSE for CGCM3 ranges from 0.2117-2.0901, for HadCM3 ranges from 0.6248-3.5580 and

for Echam5ranges from0.2767-3.7784. The maximum value of MAD is for Echam5 and least for CGCM3. Thus MSE and RMSE works out to be least for CGCM3 and is the best climate model selected. Thus Echam5 is the least accurate model and CGCM3 is the most accurate model selected.
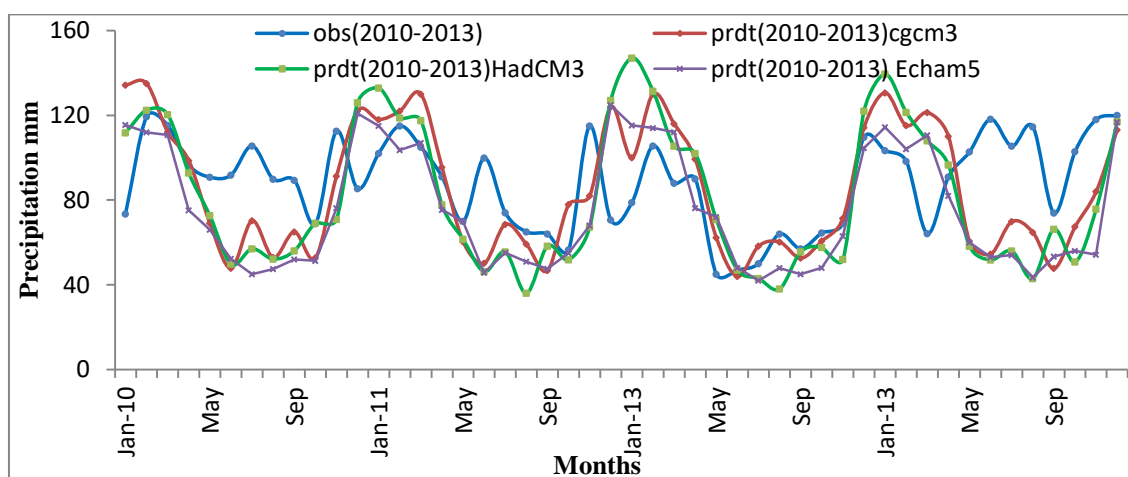

Fig.3 Validation of monthly total precipitation of Jhelum river basin for the period 2010-2013 using MLR.

Regression statistics of MLR model was done for validation period 2010-2013 for precipitation and the values of RMSE, MSE and MAD were found for all the three GCMs i.e., CGCM3, HadCM3 and Echam5.From Figure 3 it is clear that the observed and predicted values of precipitation varied in the same direction throughout

the validation period but showed less accuracy than temperature. Furthermore, the future monthly total precipitation of the Jhelum basin for the period 2001-2100 were predicted by MLR model and are shown in Fig. 5.1-5.3
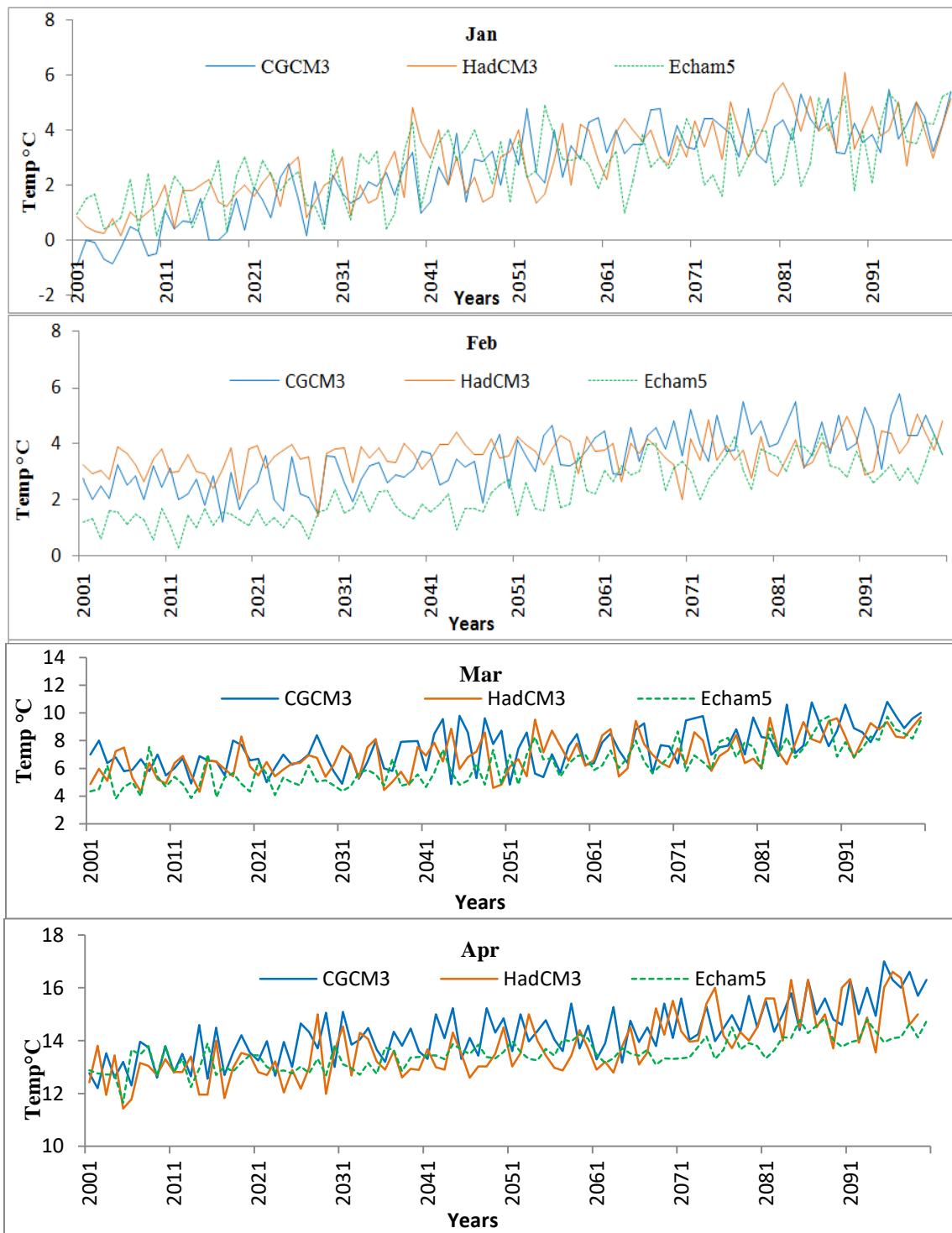
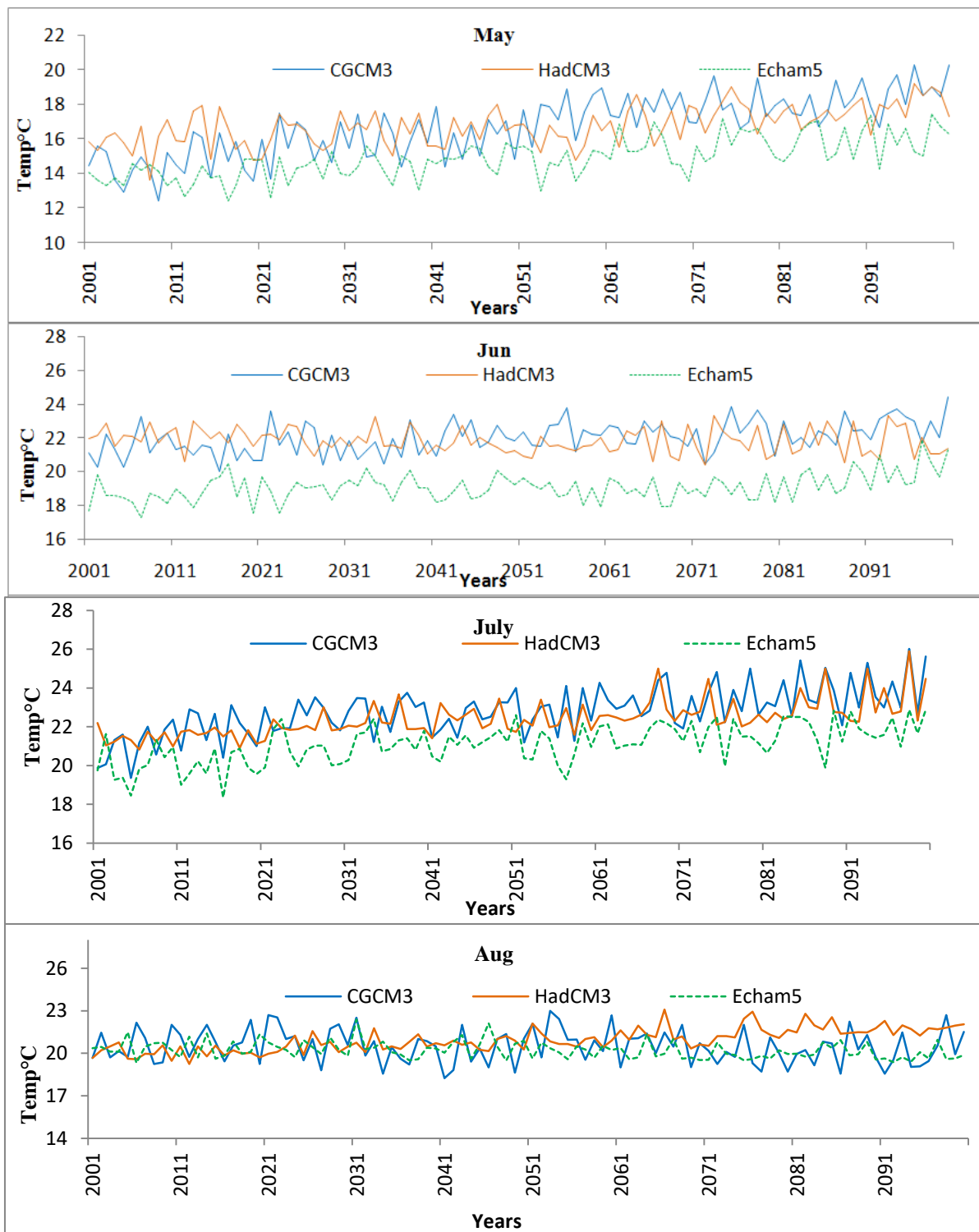**Fig.4.1.** Variation of mean monthly temperature of Jhelum river basin from January-April during 2001-2100

Mehnaza Akhter. Int. Journal of Engineering Research and Application
ISSN : 2248-9622, Vol. 7, Issue 6, (Part -4) June 2017, pp.89-101

www.ijera.com

Fig.4.2 Variation of mean monthly temperature of Jhelum river basin from May-August during2001-2100
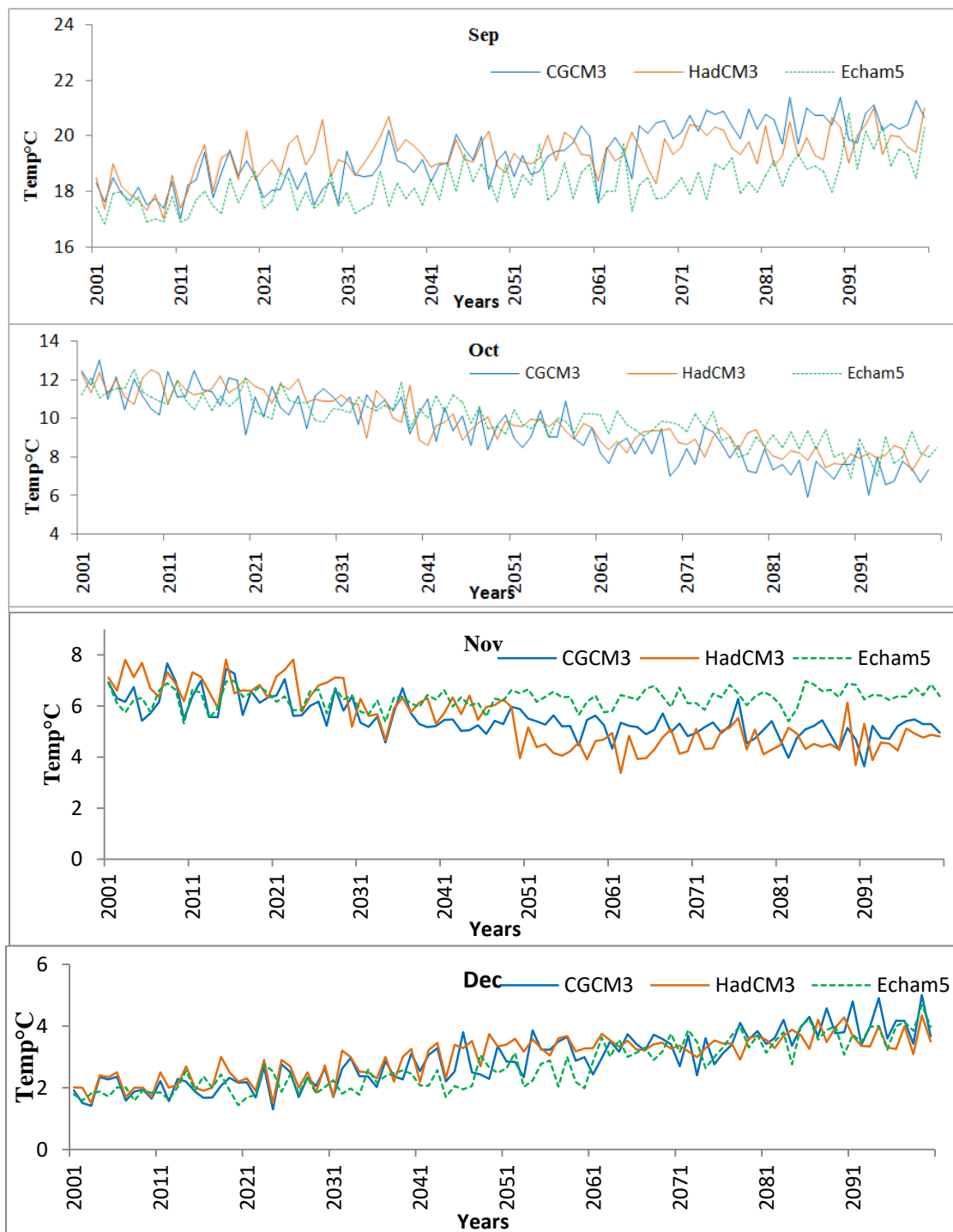
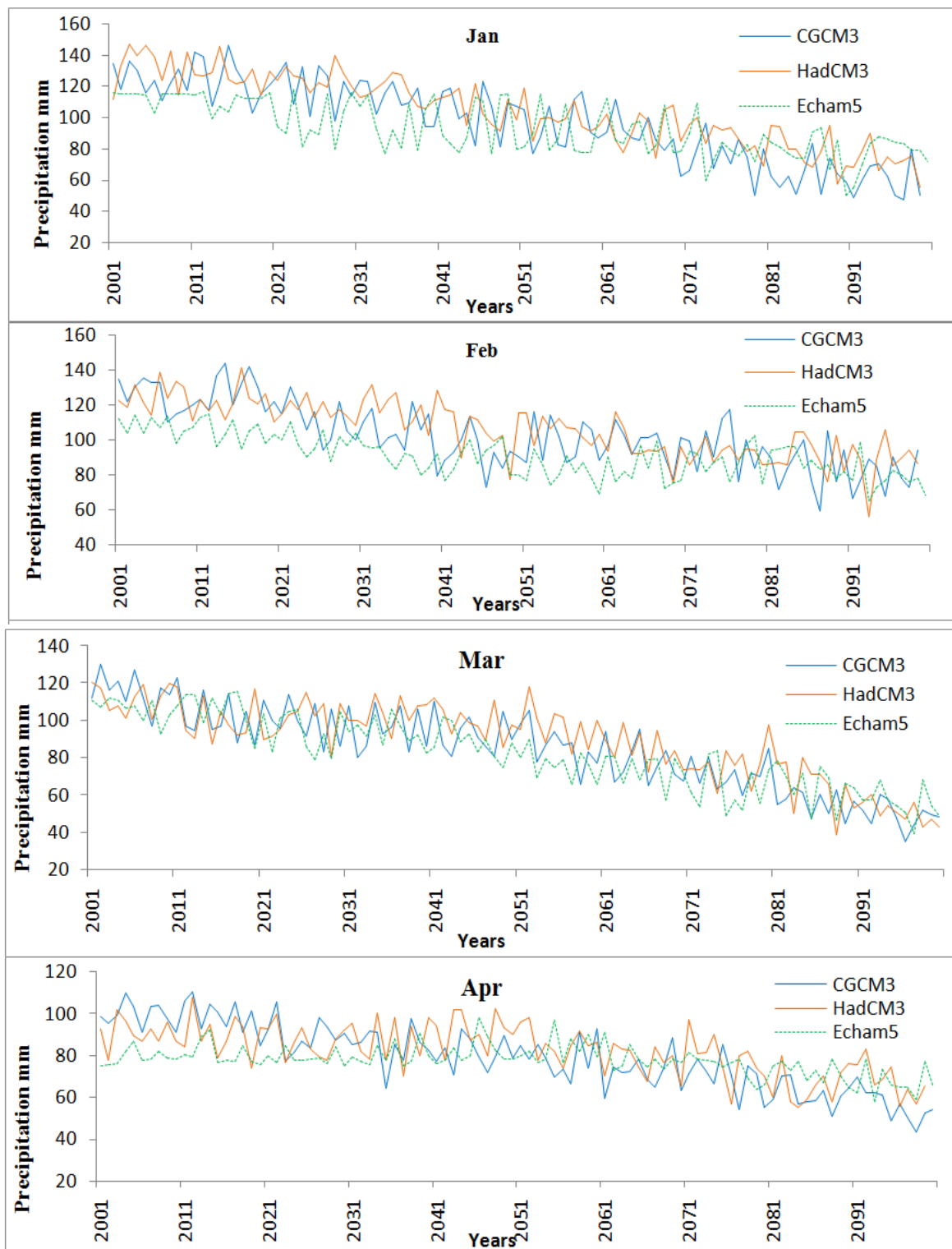**Fig.4.3** Variation of mean monthly temperature of Jhelum river basin from September-December during 2001-2100

Fig. 5.1 Variation of monthly precipitation of Jhelum river basin from January -April during 2001-2100
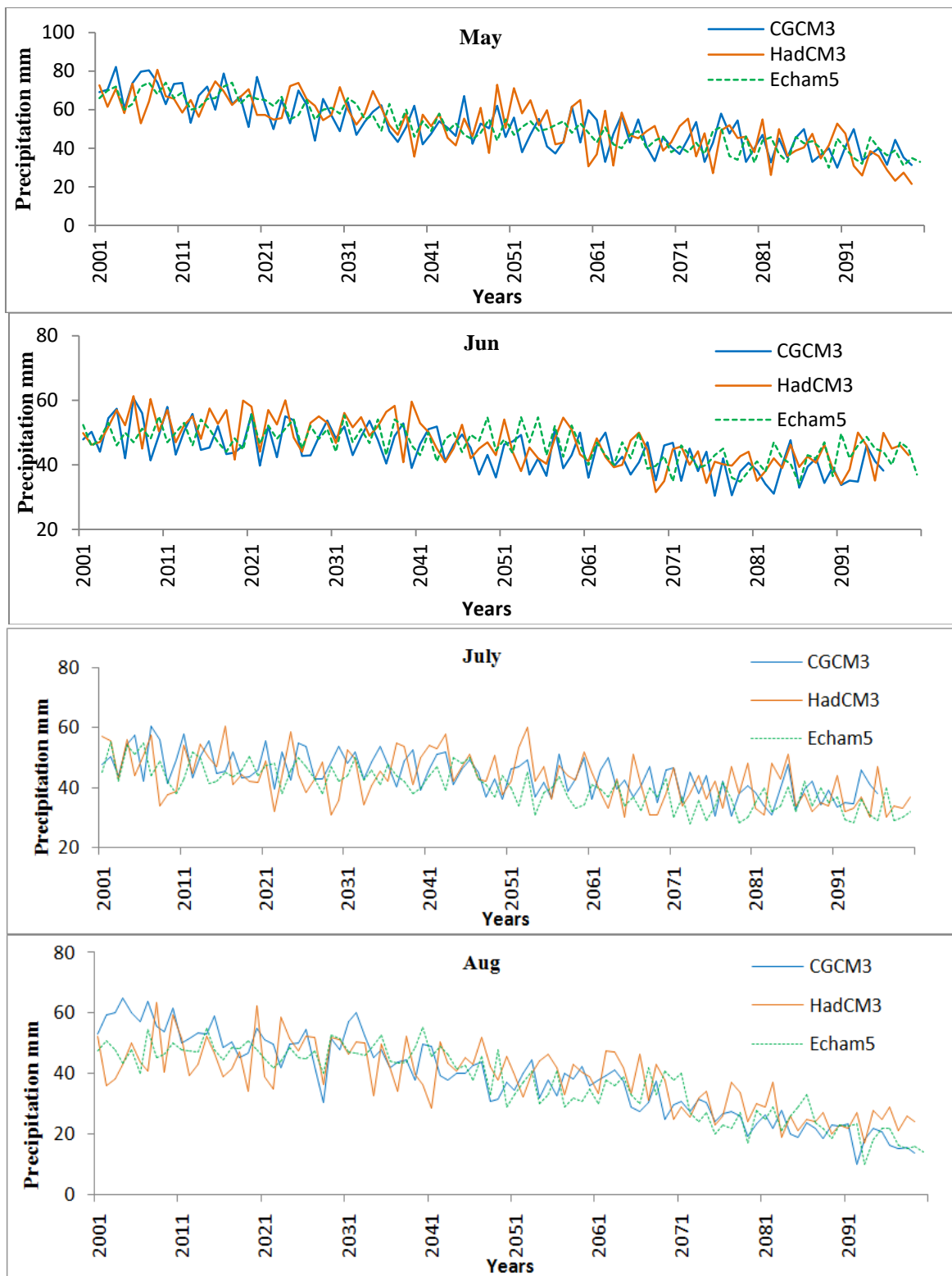
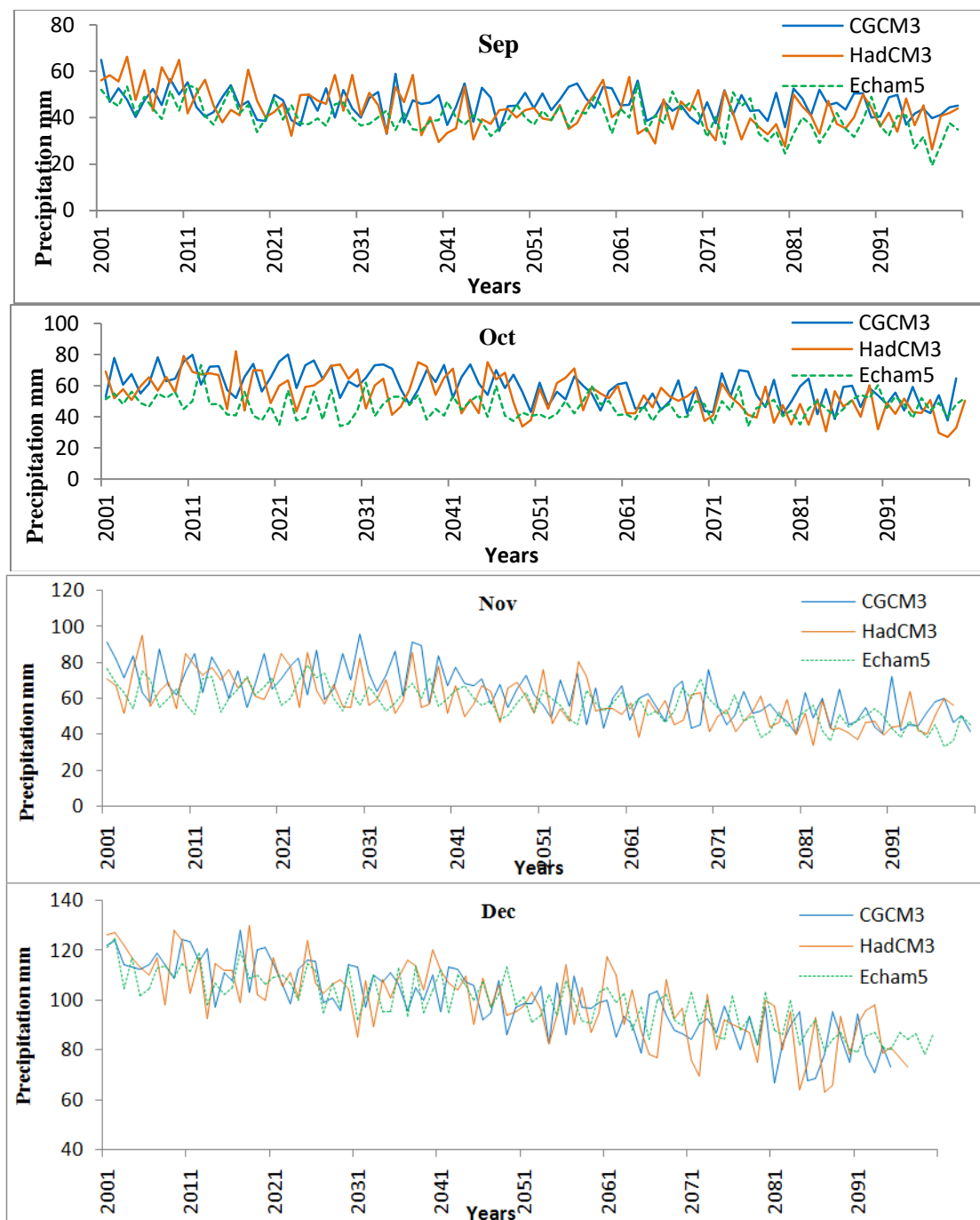**Fig.5.2** Variation of monthly precipitation of Jhelum river basin from May-August during 2001-2100

Fig.5.3 Variation of monthly precipitation of Jhelum river basin from September-December during 2001-2100
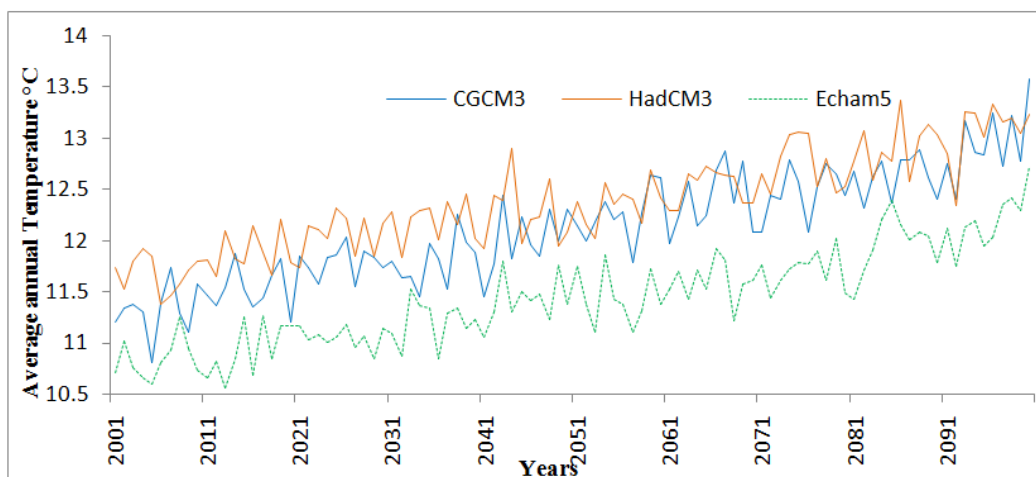
Fig. 6. Variation of MLR predicted average annual temperature of Jhelum river basin during 21st century
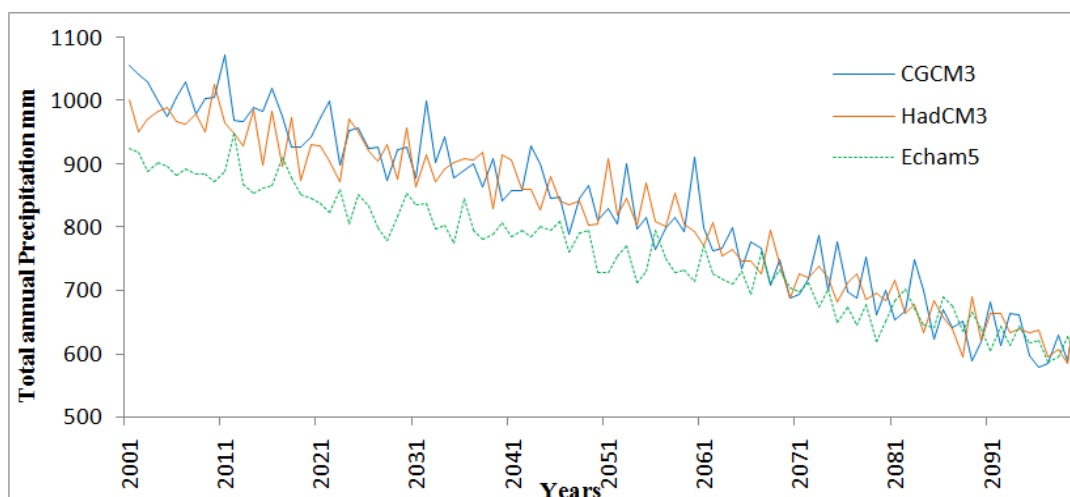

Fig.7. Variation of MLR predicted total annual Precipitation of Jhelum river basin during 21st century

V. CONCLUSION

1) As three GCMs viz., CGCM3, HadCM3 and Echam5 were used in MLR analysis. It was found that the value of RMSE, MSE and MAD were least and R^2 was highest for CGCM3. Thus CGCM3 performed better than other two GCMs.

2) The downscaled monthly mean temperature showed an increasing trend in all months except for October and November for the period 2001–2100 using CGCM3 and HadCM3 under A1B Scenario and for August, October and November using Echam5 for the period 2001–2100 for A1B scenario.

3) The average annual temperature showed an increase of 2.37°,1.50°C and 2.02°C respectively forCGCM3, HadCM3 and Echam5 models over 21st century in A1B scenario.

4) The rainfall amounts showed a decreasing trend throughout the year especially in the month of January and August and a small increase for the month of October for the period 2001–2100 using CGCM3 model for A1B scenario. ForHadCM3 A1Bscenario precipitation decreases in all months

except for April and June for the period 2001–2100. For Echam5 A1B scenario precipitation decreases in all months except for the month of October for the period 2001–2100.

5) The total annual precipitation decreased by 30.27%, 30.58°C and 36.53% respectively forCGCM3, HadCM3 and Echam5 models over 21st century in A1B scenario using MLR technique.

REFERENCES

[1]. Aksornsingchai, P., Srinilta, C. (2011). Statistical Downscaling for Rainfall and Temperature Prediction in Thailand. In, International Multi Conference of Engineers and Computer Scientists Hong Kong.

[2]. Christensen JH, Hewitson B, Busuioc A, Chen A, Gao X, Held I, JonesR, Kolli RK, Kwon WT, Laprise R, Magana Rueda V, Mearns L, Menendez CG, R´ais¨ anen J, Rinke A, Sarr A, Whetton P (2007)Regional climate projections. Climate Change 2007: The Physical Science Basis, Contribution of

Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press: Cambridge.

[3]. Fiseha, B. M., Melesse, A. M., Romano, E., Volpi, E., and Fiori, A. (2012). "Statistical downscaling of precipitation and temperature for the Upper Tiber Basin in Central Italy." Int. J. Water Sci., 1(3), 1–14.

[4]. Grotch SL, MacCracken MC. The use of general circulation models to predict regional climate change. J Clim 1991; 4: 286-303.

[5]. Hashmi, M.Z., Shamseldin, A.Y., Melville, B.W., 2009. Statistical downscaling of precipitation: state-of-the-art and application of Bayesian multi-model approach for uncertainty assessment. Hydrol. Earth Syst. Sci. Discuss. 6, 6535e6579.

[6]. IPCC (2007) IPCC fourth assessment report: climate change 2007 (AR4).Online available at www.ipcc.ch/publications_and_data Accessed28 Oct 2010.

[7]. Ojha C. S. P., Manish Kumar Goyal and A. J. Adeloye "Downscaling of Precipitation for Lake Catchment in Arid Region in India using Linear Multiple Regression and Neural Networks". The open hydrology Journal,2010,4,122-136.

[8]. Robock A, Turco RP, Harwell MA, et al. Use of general circulation model output in the creation of climate change scenarios for impact analysis. Clim Change 1993; 23(4): 293-335.

[9]. Sailor DJ, Hu T, Li X, Rosen JN. A neural network approach tolocal downscaling of GCM output for assessing wind power implications of climate change. Renewable Energy 2000; 19: 359-78.

[10]. Wilby RL, Dawson CW (2007) User manual for SDSM 4.2

[11]. Wilby RL, Wigley TML. 1997. Downscaling general circulation model output: a review of methods and limitations. Progress in Physical Geography **21**: 530–548.

[12]. Zulkarnain Bin Hassan, Sobri Bin Harun(2011). Statistical Downscaling for Climate Change Scenarios of Rainfall and Temperature in Malaysia In, UK-Malaysia-Ireland Engineering Science Conference (UMIES 2011),Kuala Lampur,12-14 July 2011.