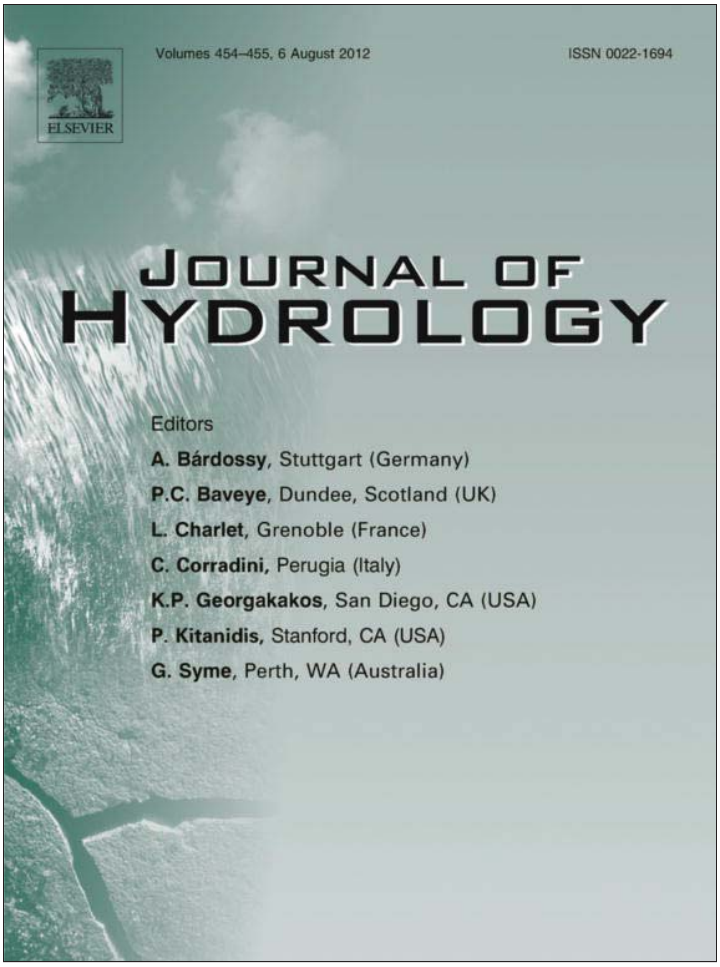
Volumes 454–455, 6 August 2012
ISSN 0022-1694
ELSEVIER
JOURNAL OF
HYDROLOGY
Editors
A. Bárdossy, Stuttgart (Germany)
P.C. Baveye, Dundee, Scotland (UK)
L. Charlet, Grenoble (France)
C. Corradini, Perugia (Italy)
K.P. Georgakakos, San Diego, CA (USA)
P. Kitanidis, Stanford, CA (USA)
G. Syme, Perth, WA (Australia)

Journal of Hydrology 454–455 (2012) 26–41

Contents lists available at SciVerse ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Prediction of monthly rainfall on homogeneous monsoon regions of India based on large scale circulation patterns using Genetic Programming

Satishkumar S. Kashid [a], Rajib Maity [b,*]

[a] Department of Civil Engineering, Walchand Institute of Technology, Solapur 413 006, Maharashtra, India
[b] Department of Civil Engineering, Indian Institute of Technology Kharagpur, Kharagpur 721 302, West Bengal, India

ARTICLE INFO

Article history:
Received 31 May 2011
Received in revised form 20 February 2012
Accepted 14 May 2012
Available online 29 May 2012
This manuscript was handled by Philippe Baveye, Editor-in-Chief, with the assistance of K.P. Sudheer, Associate Editor

Keywords:
El Niño-Southern Oscillation (ENSO)
Equatorial Indian Ocean Oscillation (EQUINOO)
Genetic Programming (GP)
Indian Summer Monsoon Rainfall (ISMR)

SUMMARY

Prediction of Indian Summer Monsoon Rainfall (ISMR) is of vital importance for Indian economy, and it has been remained a great challenge for hydro-meteorologists due to inherent complexities in the climatic systems. The Large-scale atmospheric circulation patterns from tropical Pacific Ocean (ENSO) and those from tropical Indian Ocean (EQUINOO) are established to influence the Indian Summer Monsoon Rainfall. The information of these two large scale atmospheric circulation patterns in terms of their indices is used to model the complex relationship between Indian Summer Monsoon Rainfall and the ENSO as well as EQUINOO indices. However, extracting the signal from such large-scale indices for modeling such complex systems is significantly difficult. Rainfall predictions have been done for 'All India' as one unit, as well as for five 'homogeneous monsoon regions of India', defined by Indian Institute of Tropical Meteorology. Recent 'Artificial Intelligence' tool 'Genetic Programming' (GP) has been employed for modeling such problem. The Genetic Programming approach is found to capture the complex relationship between the monthly Indian Summer Monsoon Rainfall and large scale atmospheric circulation pattern indices – ENSO and EQUINOO. Research findings of this study indicate that GP-derived monthly rainfall forecasting models, that use large-scale atmospheric circulation information are successful in prediction of All India Summer Monsoon Rainfall with correlation coefficient as good as 0.866, which may appears attractive for such a complex system. A separate analysis is carried out for All India Summer Monsoon rainfall for India as one unit, and five homogeneous monsoon regions, based on ENSO and EQUINOO indices of months of March, April and May only, performed at end of month of May. In this case, All India Summer Monsoon Rainfall could be predicted with 0.70 as correlation coefficient with somewhat lesser Correlation Coefficient (C.C.) values for different 'homogeneous monsoon regions'.

1. Introduction

It is scientifically and mathematically challenging to use climate signals for the prediction of basin-scale hydrologic variables, because the climatic systems are very complex and physics of many systems is still not very clearly understood. The difficulties in modeling such complex systems are considerably reduced by the recent Artificial Intelligence tools like Artificial Neural Networks (ANNs); Genetic Algorithm (GA) based evolutionary optimizer and Genetic Programming (GP). Hence such AI tools are tried nowadays for modeling complex systems like basin-scale stream-flow forecasting using the information of large-scale atmospheric circulation phenomena.

Indian Summer Monsoon Rainfall is always found to vary annually leading to profound impacts on agriculture based Indian economy. Generally meteorological forecasts are generated for three timescales, viz. short-range (1–2 days ahead), medium-range (3–10 days ahead) and long-range forecasts for monthly and seasonal scales. In India, India Meteorological Department (IMD) generates the short and long-range predictions, whereas the National Centre for Medium Range Weather Forecasting (NCMRWF), New Delhi is responsible for the medium-range predictions.

Prediction of ISMR is having a long history. It started with the work of Sir Henry Blanford in 1886, which was entirely based on Himalayan snowfall. John Eliot used extra-Indian factors, viz. Pressure of Mauritus, Zanzibar and Seychelles in the monsoon forecast of 1896. Sir Gilbert Walker proposed statistical association for monsoon forecast. He systematically examined the relationship between Indian monsoon rainfall and global circulation parameters. He selected 28 predictors to issue forecast based on regression equation during the year 1906 (Jagannathan, 1960; Rao and Rama Moorthy, 1960; Rao, 1965). Most of the Walker's predictors (except Himalayan snow accumulation) were signs of different facets of Southern Oscillation (Shukla and Paolino, 1983). Savur (1931)

* Corresponding author. Tel.: +91 80 2293 2673; fax: +91 80 2360 0404.
E-mail addresses: rajib@civil.iitkgp.ernet.in, rajibmaity@gmail.com (R. Maity).

0022-1694/$ - see front matter © 2012 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.jhydrol.2012.05.033

showed that 7 out of the 28 parameters had lost their significance in due course of time. Since then, extensive research work has been carried out on empirical seasonal forecasting of Indian Summer Monsoon Rainfall. Some of the noteworthy studies can be listed as following: Banerjee et al. (1978), Kung and Sharif (1982), Bhalme et al. (1986), Gowariker et al. (1989, 1991), Parthasarathy et al. (1988, 1991, 1995), Krishna Kumar et al. (1995, 1997), Rajeevan et al. (2004), etc. In spite of many efforts in the long range prediction of all-India summer monsoon rainfall (AISMR), it is felt that achieved success is not adequate and there is much scope to investigate new predictors and new methodologies of ISMR prediction.

Even though forecast for 'All India Summer Monsoon Rainfall' is available before every monsoon nowadays, still such forecasts have limited use due to significant variation in monsoon rainfall over the country in the same season. Hence it is felt that the rainfall forecasts can be more useful if issued at regional or sub-divisional scale. Hence this work attempts to develop models for issuing medium range forecasts of monthly monsoon rainfall from June through October, at all India level, as well as for five homogeneous monsoon regions of India. It is felt that such regional forecasts will have better utility than total ISMR due to availability of the information at smaller spatio-temporal scale.

Simultaneous variations of climatic conditions and hydrologic variables over widely separated regions on the surface of earth have long been discovered and noted by the meteorologists, world over. Such recurrent patterns are commonly referred to as "hydroclimatic teleconnection". It is established that the natural variation of hydrologic variables is linked with these large-scale atmospheric circulation pattern through hydroclimatic teleconnection (Dracup and Kahya, 1994; Eltahir, 1996; Jain and Lall, 2001; Douglas et al., 2001; Ashok et al., 2001, 2004; Marcella and Eltahir, 2008; Maity and Nagesh Kumar, 2008). Indian hydrometeorology is prominently influenced by two large-scale atmospheric circulation patterns. The first is El Niño-Southern Oscillation (ENSO) from tropical Pacific Ocean and second is the Equatorial Indian Ocean Oscillation (EQUINOO) from Indian Ocean. (Gadgil et al., 2004).

El Niño-Southern Oscillation (ENSO), which is a large-scale circulation pattern from tropical Pacific Ocean, is established to influence Indian Summer Monsoon Rainfall. Another large scale circulation pattern from Indian Ocean viz. Indian Ocean Dipole Mode (IOD) also influences the Indian Summer Monsoon rainfall (Saji et al., 1999).

Equatorial Indian Ocean Oscillation (EQUINOO) is the atmospheric component of the IOD mode (Gadgil et al., 2003, 2004). Gadgil et al. (2003) had shown that the Indian Summer Monsoon Rainfall is not only associated with ENSO, but also with EQUINOO. They suggest that one can scientifically predict the Indian Summer Monsoon Rainfall by knowing the prior EQUINOO status. Equatorial zonal wind index (EQWIN) is considered as an index of EQUINOO, which is defined as negative of the anomaly of the zonal component of surface wind in the equatorial Indian Ocean Region (60°E–90°E, 2.5°S–2.5°N). Weakening of ENSO-ISMR relationship is indicated by few researchers (Krishna Kumar et al., 1999). It is also established that ENSO-ISMR is modified by the influence of Indian Ocean Dipole (IOD) mode. Consideration of these two indices is found to give better results as compared to the analyses by researchers using just ENSO index (Gadgil et al., 2004). Thus, in this study, apart from ENSO index, EQUINOO index from Equatorial Indian Ocean is considered simultaneously, which is supposed to take care the temporal change in relationship between ENSO and ISMR.

Since nearly 80% of Indian Summer Monsoon Rainfall is due to the southwest monsoon, interaction between various oceans due to ENSO and EQUINOO regulates the amount and distribution of the rainfall over the sub continent. Such association is more prominent for the large aerial scale. It is also prominent for longer temporal scale (seasonal) or smaller temporal scale (monthly).

The search for a new methodology for predicting the All India Summer Monsoon Rainfall has been continued for a long time. The search is active for both, the long term as well as short term forecasts of Indian Summer Monsoon Rainfall. In recent years, an efficient Artificial Intelligence tool Genetic Programming has been used for modeling complex systems. Hence the Genetic Programming approach has been used to predict monthly Indian Summer Monsoon Rainfall over India in this study.

2. Objectives of the work

This work intends to develop models for medium range (1 month ahead) forecasts of monthly Indian Summer Monsoon rainfall for 'All India', as well as for five homogeneous monsoon regions of India, by using ENSO and EQUINOO indices as large scale atmospheric circulation information, with the help of Artificial Intelligence tool Genetic Programming. The study also deals with development of models for prediction of one time ISMR forecast on end of May, for All India Summer Monsoon Rainfall and five homogeneous monsoon regions of India. The study intends to compare both the forecasts and discuss usefulness of both approaches. The results of both the analyses can be compared to derive suitable conclusions indicating importance and utility of both the approaches.

3. Data

Sea surface temperature (SST) anomaly from the Niño 3.4 region (120°W–170°W, 5°S–5°N) is used as the 'ENSO index' in this study. Monthly sea surface temperature data from Niño 3.4 region for the period, January 1950 to December 2006, data are obtained from the website of the National Weather Service, Climate Prediction Centre of National Oceanic and Atmospheric Administration (NOAA) (http://www.cpc.noaa.gov/data/indices/). EQWIN, the negative of zonal wind anomaly over equatorial Indian Ocean region (60–90°E, 2.5°S–2.5°N) is used as 'EQUINOO index' (Gadgil et al., 2004). Monthly surface wind data for the period January 1950–December 2010, are obtained from the National Centre for Environmental Prediction (http://www.cdc.noaa.gov/Datasets).

Monthly rainfall data over entire India as well as over homogeneous monsoon regions of India rainfall data used for this study were collected by India Meteorological Department for a period 1871–2010. The data from 1950 through 2010 were only used, as the ENSO and EQUINOO data were available 1950 onwards. Monthly rainfall data from data from 1950 through 1975 were used for the training purpose. The data from 1976 through 1990 were used for the validation purpose. The data from 1991 through 2010 were used as testing the GP models. The analysis consists of rainfall depths all over India for so called Indian Summer Monsoon season (June to September) plus the month of October. The reason behind including the October rainfall in the analysis is to encompass the complete physical processes of Indian Summer Monsoon. October rainfall values are also found reasonable as compared to the monsoon months of year viz. June through September. It is assumed that, no temporal changes have been observed in the summer monsoon rainfall during training, validation and testing periods. Evidence shows that though 'All-India' summer monsoon rainfall is not showing any long-term trend, rainfall over subdivisional scale has been showing increasing/decreasing extremes due global warming and climate change. However, such changes in the extremes can be more effective for daily scale. This study considers monthly rainfall at homogeneous monsoon regions, which consists of several adjoining subdivisions, i.e., a larger

Table 1
Details of area covered by homogeneous monsoon regions of India for regional rainfall data sets. Sources: ftp://www.tropmet.res.in/pub/data/rain/iitm-regionrf.txt.

No.	Region	No. of Sub. Div.	Area (square km)
1	All India	30	2,880,324
2	Central North-East India	5	573,006
3	North-East India	4	267,444
4	North-West India	6	634,272
5	Peninsular India	6	442,908
6	West Central India	9	962,694

spatial extent. Validity of the stationarity assumption for this study is discussed later in the context of the model performance.

The methodology adopted by Indian Institute of Tropical Meteorology for preparation of regional rainfall data series is described as following. The monthly (January–December) area weighted rainfall series for each of the 30 meteorological subdivisions have been prepared by assigning the district area as the weight for each rain-gauge station in that subdivision. Similarly assigning the subdivision area as the weight to each of the subdivisions in the region, area weighted monthly rainfall series are prepared for homogeneous regions of India as well as for all India. The details of area covered by homogeneous monsoon regions used by Indian Institute of tropical Meteorology (IITM) are listed in the Table 1. The 'homogeneous monsoon regions of India' as discussed earlier can be visualized in map of India illustrated in Fig. 1.

4. Methodology

Artificial Intelligence tools like Artificial Neural Networks, Support Vector Machines as well as Neuro-fuzzy Systems and data driven models have become popular tools for predictions in Water Resources related research problems (Drecourt, 1999; Giustolisi and Savic, 2006; Chau, 2006; Li et al., 2006; Lin et al., 2006; Shiri and Kisi, 2011; Muttil and Chau, 2007; Partal and Kisi, 2007). Many researchers have used Genetic Programming for modeling complex systems. (Babovic, 2000; Aytek and Kisi, 2008; Harris et al., 2003; Babovic and Keijzer, 2000; Baptist et al., 2007; Keijzer and Babovic, 2002; Kisi and Shiri, 2011). Applications of Genetic Programming are reported the field of Hydraulics and Fluid Mechanics (Babovic and Abbott, 1997a; 1997b; Giustolisi, 2004). Genetic Programming is widely used in recent years for data mining, rainfall runoff modeling and other hydrologic predictions (Babovic, 2005; Babovic and Keijzer, 2002; Liong et al., 2002).

The problem of predicting meteorological events such as rainfall over a region is much more complex than any other general scientific problem. This is because of the extreme instability of the atmosphere. The systems responsible for the events that we are trying to predict, such as clouds or monsoon depressions (in which thousands of clouds are embedded) are the culmination of the instabilities. These involve nonlinear interaction between different spatial scales from kilometers (as in a single cloud) to hundreds of kilometers (as in a monsoon depression or a hurricane). Due to such inherent complexity in the climate systems, it is mathematically challenging to use climate signals for the prediction of rainfall. The difficulties in modeling such complex systems can be considerably reduced by using the modern Artificial Intelligence (AI) tools like Artificial Neural Networks (ANNs), Genetic Algorithm (GA) and Genetic Programming (GP). Hence such AI tools are tried nowadays for modeling complex systems.

The Artificial Intelligence tool Genetic Algorithm (GA) has given a rise to two new fields of research where (global) optimization is

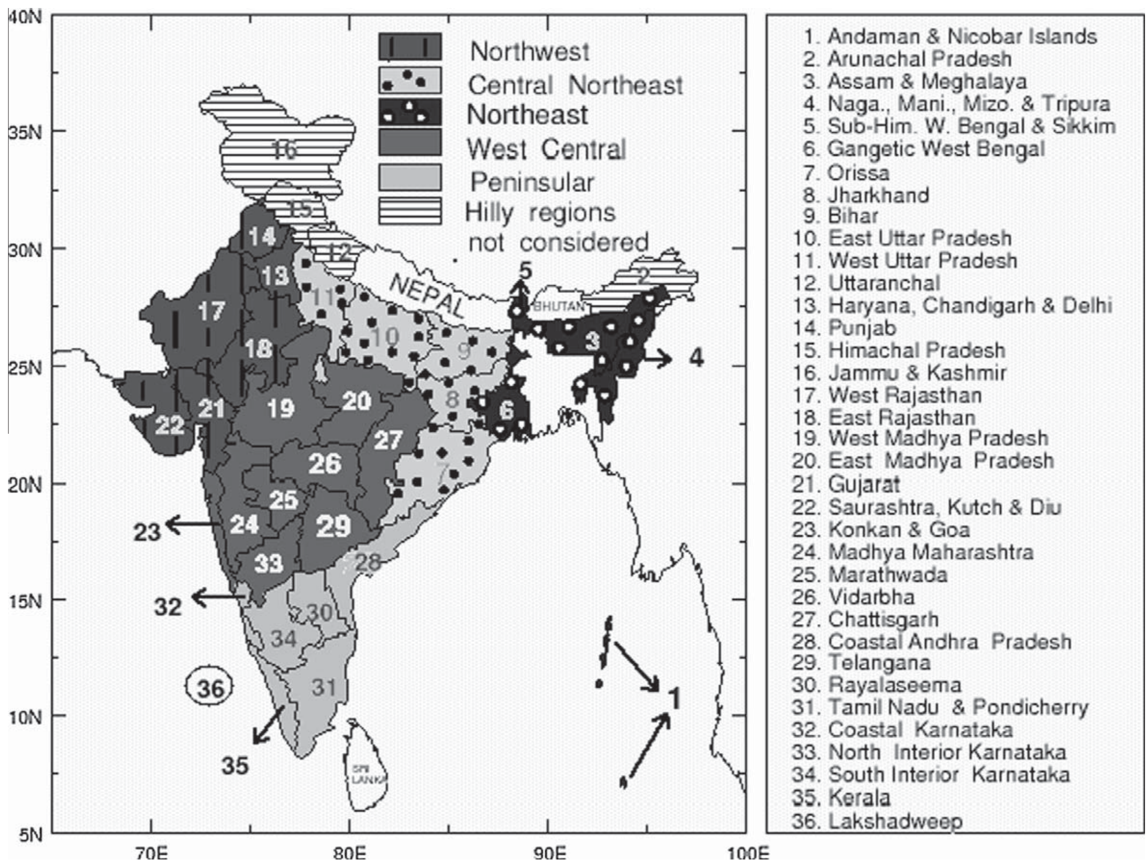

Fig. 1. Homogeneous monsoon regions of India. *Source*: Modified from Indian Institute of Tropical Meteorology.

of crucial importance: 'Genetic Based Machine Learning' (GBML) and 'Genetic Programming' (GP). GP is a member of evolutionary computing family. Further, GP can also be viewed as a GA applied to a population of computer programs (Sette and Boullart, 2001; Saks and Maringer, 2010). GA usually operates on (coded) strings of numbers, whereas GP operates on computer programs. Genetic Programming can solve much more complicated problems. Genetic Programming can also be applied to a greater diversity of problems (Koza, 1992).

It might be interesting to compare GP with the traditional approaches like ANN. ANNs do have many attractive features, but they suffer from some limitations. The difficulty in choosing the optimal network architecture and time-consuming effort involved thereof is one of the key issues. In regression also, the model structure is decided in advance and the model coefficients are determined by the regression method. On the other hand, GP has the unique feature that it does not assume any functional form of the solution. GP can optimize both the structure of the model and its parameters.

Genetic Programming evolves a computer program, relating the output and input variables. Hence it has the advantage of providing inherent functional relationship explicitly over techniques like ANN. The specialty of GP approach lies with its ability to select input variables that contribute beneficially to the model and to disregard those that do not. Hence Genetic Programming is used for modeling regional rainfall prediction in this study.

4.1. Genetic Programming

The theory behind Genetic Programming is almost same as that behind genetic algorithms. The same Darwinian concept of survival-of-the-fittest is applied through genetic operators, but with a small difference. The structures that are manipulated are quite different from the coded strings of genetic algorithms. Fig. 2 depicts how one can visualize a simple genetic program. It can be seen from the 'Tree' structure of a model that there now exists, a clean hierarchical structure, instead of a flat, one dimensional string. The structure is made up of simple functions that can be easily encoded using a high-level language. Tree manipulation routines exist in several high-level computer programming languages.

One of the great challenges in computer science is to get a computer to do what needs to be done, without telling it how to do it. Genetic Programming accepts this challenge by providing a methodology for automatically creating a working computer program,

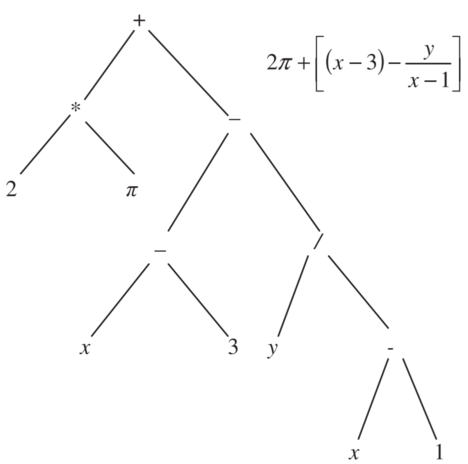

Fig. 2. A simple 'Tree' structure to represent a model.

from statement of the problem. Genetic Programming breeds a population of computer programs to solve a problem. It iteratively transforms a population of computer programs into a new generation of programs by applying analogs of naturally occurring genetic operations. It genetically breeds a population of computer programs using the Darwinian principles of natural selection and biologically inspired operations. The operations include reproduction, crossover, mutation, and architecture altering operations patterned after gene duplication and gene deletion in nature (Koza, 1992).

The tool Genetic Programming has been successfully applied recently by few researchers, for water resources problems. Drecourt (1999) applied neural networks and Genetic Programming for rainfall runoff modeling. Whigham and Crapper (2001) also used Genetic Programming for Rainfall–Runoff modeling. Muttil and Liong (2001) applied Genetic Programming for the problem of improving runoff forecasting by input variable selection. Streamflow rate prediction over a semi-arid coastal watershed in USA was attempted by Drunpob et al. (2005). Makkeasorn et al. (2008) compared Genetic Programming and neural network models for short-term streamflow forecasting with global climate change implications.

Five major preparatory steps in application of GP (Koza, 1992) can be summarized as following: (i) selection of the set of terminals, (ii) selection of the set of primitive functions, (iii) deciding the fitness measure, (iv) deciding the parameters for controlling the run, and (v) defining the method for designating a results and the criterion for terminating a run. A flow chart showing the basic steps involved in GP is shown in Fig. 3. The choice of input variables is generally based on priory knowledge of casual variables and physical insight into the problem being studied. If the relationship to be modeled is not well understood, then analytical techniques can be used.

GP evolves a function that relates the input information to the output information, which is of the form:

$$Y^m = f(X^n) \tag{1}$$

where X^n is an n-dimensional input vector and Y^m is an m-dimensional output vector. In the proposed study, the input vector consists of Historical Average Rainfall for particular month, ENSO indices of three previous monthly time steps and EQWIN indices of three previous monthly time steps. The output vector consists of monthly rainfall for the particular month over the region.

The implementation of GP in this work is done through software tool Discipulus (Francone, 1998) that is based on an extension of the originally envisaged GP called Linear Genetic Programming (LGP). It evolves sequences of instructions from an imperative programming language or machine language. The LGP expresses instructions in a line-by-line mode. The term 'linear' in Linear Genetic Programming refers to the structure of the (imperative) program representation. It does not stand for functional genetic programs that are restricted to a linear list of nodes only. Genetic programs normally represent highly non-linear solutions in this meaning (Brameier and Banzhaf, 2004).

4.2. Control parameters and input impact of different variables

Values of control parameters of Genetic Programming can be selected initially and thereafter varied in trials till the best fitness measures are produced. The fitness criterion is the mean squared error between the actual and the predicted values. The statistical error criteria of Correlation Coefficient (C.C.) between observed and predicted rainfall and Root Mean Square Error (RMSE) have been used in this study to compare the GP predictions with the

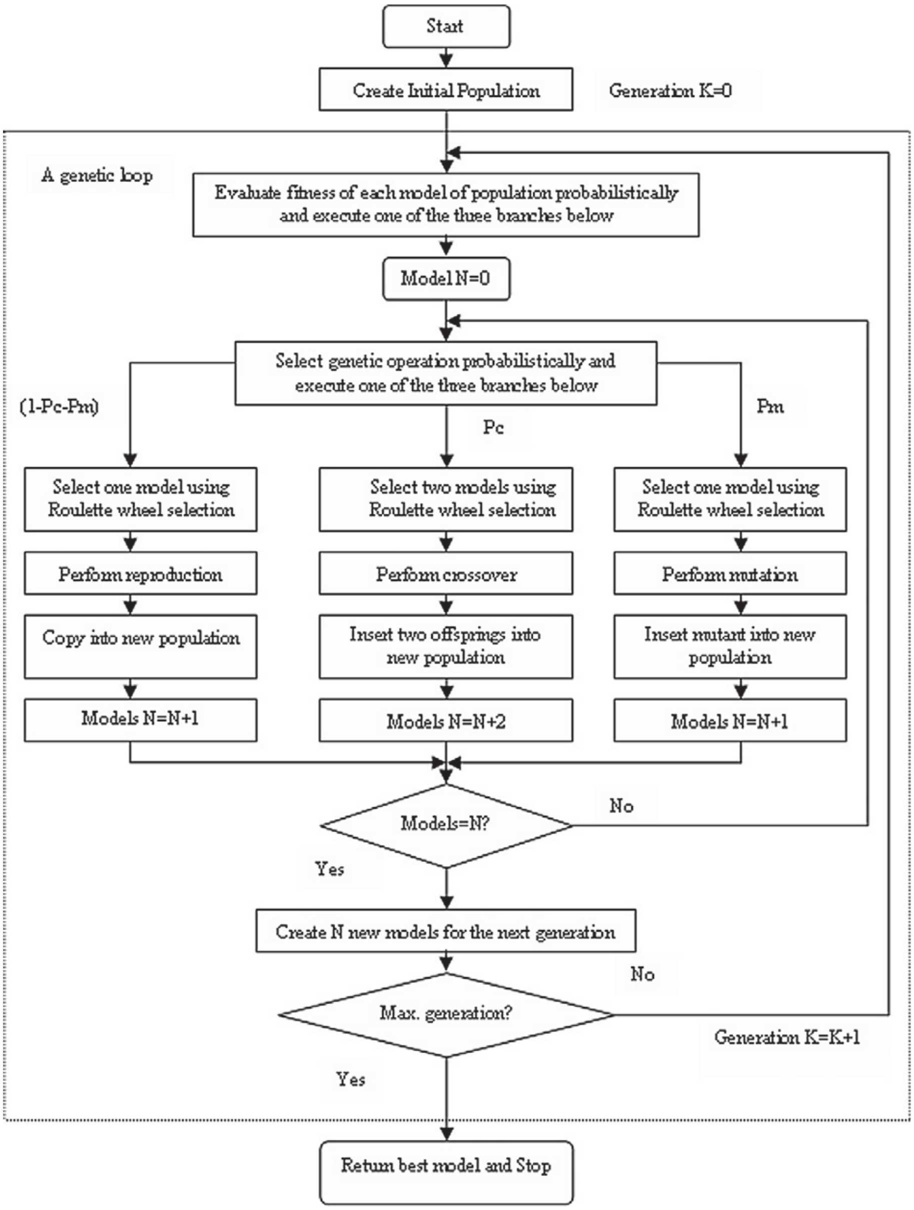

Fig. 3. Flowchart for Genetic Programming. *Source*: Hong and Bramidimarri (2003).

actual observations. Four basic arithmetic operators (+, _, ∗, /), trigonometric functions and some basic mathematical functions like sqrt (x) and power are utilized. A typical choice of the initial GP control parameters adopted in the studied is as follows: Population size: 500, number of generations: 300, mutation frequency: 90%, crossover frequency: 50%. The reproduction rate in a run is left over after the application of the crossover and mutation operators. The reproduction rate may be calculated (in percentage) as follows:

$$\text{Reproduction rate} = 100 - \text{mutation rate} - (\text{crossover rate} * [1 - \text{mutation rate}])$$

The GP tool calculates 'input impact' of every input variable. 'Input impact' calculated by the tool is based on the percentage of the best thirty programs from the project that contained the referenced input variable. For instance, input impact = 0.66 for certain input variable indicates that particular variable appears in 20 out of best 30 programs, evolved by GP tool. Input impact happens to be a measure of sensitiveness of particular variable as input.

Main inputs in this study are monthly values of ENSO indices and EQUINOO indices of few previous months. Trials were taken by including ENSO and EQUINOO indices of last 5 months i.e. $(t-1)$ through $(t-5)$ to decide number of monthly time steps for ENSO and EQUINOO indices in analysis. However it was observed that 'Input Impacts' were significant for up to EN $(t-3)$ and EQ $(t-3)$ time step only. Hence it was decided to use ENSO and EQUINOO indices of time steps $(t-1)$, $(t-2)$ and $(t-3)$ only as inputs. This combination could give the best results.

4.3. Genetic programming approach for monthly rainfall forecasting

Genetic Programming models are developed to predict Indian Summer Monsoon Rainfall. India Meteorological Department (IMD) has divided the country into five 'homogeneous monsoon

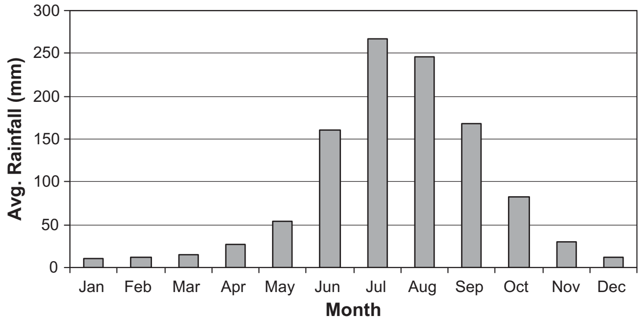

Fig. 4. Historical average of Indian rainfall.

regions'. Hence total six separate analyses are carried out for developing six separate models of ISMR predictions. The first analysis deals with 'All India Summer Monsoon Rainfall' with India as one unit. The other five analyses deal with five 'homogeneous monsoon regions' of India. The monthly data is designated with respect to the central date of that month. Monthly rainfall data from January 1, 1950 to December 31, 2010 was used for this study. Monthly rainfall data are available since January 1871 for All India Rainfall as well as five homogeneous monsoon regions. However, the ENSO indices and EQUINOO indices are available beyond year 1950 only. Hence the period of analysis is limited to 1950 through 2010. The plot of historical monthly Indian Rainfall is depicted in Fig. 4. It can be observed from Fig. 4, that rainfall is significant in the months of June, July, August, September and October. ENSO and EQUINOO indices for three immediate previous months are considered for monthly rainfall prediction.

4.4. Monthly ISMR prediction

Two separate analyses are carried out in this work. The first analysis uses monthly ENSO and EQUINOO indices of three previous time steps for prediction of rainfall. For example, for prediction of August rainfall, ENSO and EQUINOO indices of July, June and May are given as input, with long term avg. rainfall of month August as one long term input. This system of input thus needs input of monthly ENSO and EQUINOO indices of previous months for prediction for the starting month. This can be treated as medium range forecast with 1 month lead time which has its own importance for planning of agricultural activities and judicious management of available water in reservoirs, depending upon likely inflows into reservoirs in near future. Thus the monthly rainfall is modeled as a function of

(i) Historical average monthly rainfall of the particular month.
(ii) ENSO indices of three previous monthly time steps (three values).
(iii) EQUINOO indices of three previous monthly time steps (three values).

Thus,

$$R_t = f\{HR_t,\ (EN_{t-1}, EN_{t-2}, EN_{t-3}), (EQ_{t-1}, EQ_{t-2}, EQ_{t-3})\} \quad (2)$$

For example

$$R_{June} = f\{HR_{June}, (EN_{May}, EN_{April}, EN_{March}), (EQ_{May}, EQ_{April}, EQ_{March})\} \quad (3)$$

Thus the total summer monsoon rainfall is sum of rainfall from June through September, calculated as following.

$$R_{\text{monsoon}} = R_{June} + R_{July} + R_{August} + R_{September} \quad (4)$$

where R_t stands for predicted (computed by using GP model) rainfall of particular month, HR stands for Historical average of rainfall in particular month, *EN* stands for ENSO index, *EQ* stands for EQUINOO Index. The optimum number of lags to be considered for each input variables is decided based on the 'input impacts' of that input variable during model calibration.

The second analysis is altogether different from the first analysis. This analysis aims to forecast monthly rainfall in coming monsoon months June, July, August and September depending upon the climatic signals in form of ENSO and EQUINOO indices of 3 months viz. March, April and May before onset of Indian summer monsoon. This approach thus models rainfall of September also as a function of large scale indices observed over March, April and May and one long term input in form of long term average rainfall of particular month. Thus the monthly rainfall is modeled as a function of

(i) Historical average monthly rainfall of the particular month.
(ii) ENSO indices of three previous monthly time steps (three values).
(iii) EQUINOO indices of three previous monthly time steps (three values).

Thus,

$$R_{June} = f\{HR_{June},\ (EN_{May}, EN_{April}, EN_{March}), (EQ_{May}, EQ_{April}, EQ_{March})\} \quad (5)$$

$$R_{July} = f\{HR_{July},\ (EN_{May}, EN_{April}, EN_{March}), (EQ_{May}, EQ_{April}, EQ_{March})\} \quad (6)$$

$$R_{August} = f\{HR_{August},\ (EN_{May}, EN_{April}, EN_{March}), (EQ_{May}, EQ_{April}, EQ_{March})\} \quad (7)$$

$$R_{Sept} = f\{HR_{Sept}\ (EN_{May}, EN_{April}, EN_{March}), (EQ_{May}, EQ_{April}, EQ_{March})\} \quad (8)$$

$$R_{monsoon} = R_{June} + R_{July} + R_{August} + R_{September} \quad (9)$$

where R_t stands for predicted (computed using GP model) rainfall of particular month, *HR* stands for Historical average of rainfall in particular month, *EN* stands for ENSO index, *EQ* stands for EQUINOO Index. The optimum number of lags to be considered for each input variable is decided based on the 'input impacts' of that input variable during model calibration.

5. Results and discussions

As stated earlier, two separate analyses are carried out in this work. The first analysis uses real time monthly ENSO and EQUINOO indices of three previous monthly time steps for prediction of rainfall, which is nothing but medium range forecast with 2 weeks lead time by using monthly data of ENSO and EQUINOO in real time. The second analysis is one time analysis at the end of month of May, for computing monthly and hence monsoon rainfall in months of June through September at the end of May. It is based on ENSO and EQUINOO indices of March, April and May of the same year. The results of both the analyses are discussed in the following subsections.

5.1. ISMR analysis using ENSO and EQUINOO

Long term average (1950–2010) computed and used in this analysis for All India summer monsoon rainfall (June to September)

Table 2
Average monthly rainfall (mm) for all India and homogeneous monsoon regions during monsoon months June through September.

No.	Region	June (mm)	July (mm)	August (mm)	September (mm)	Total monsoon (June–September) (mm)
1	All India	160.5	267.6	245.3	167.3	840.4
2	Central Northeast India	158.7	308.2	304.4	208.4	979.8
3	West Central India	162.8	292.7	279.7	176.5	911.8
4	Northeast India	367.4	408.6	342.3	279.6	1394.8
5	Northwest India	67.0	187.1	160.4	81.3	495.8
6	Peninsular India	164.0	191.6	158.1	150.1	663.9

is 840.4 mm. All India monthly averages of rainfall for June, July, August and September are 160.0 mm, 267.6 mm, 245.3 mm, and 167.3 mm respectively for the aforesaid period. October rainfall also has sizeable value of 82.3 mm. Monthly distribution of All India Rainfall over the period of 12 months of year (January–December) is shown in Fig. 4. Average Monthly Rainfall (mm) for all India and homogeneous monsoon regions during monsoon months June through September is given in Table 2.

The South west monsoon normally touches Indian continent on the first day of June and returns in November. Hence rainfall in October also can be considered as a part of Monsoon rainfall activity. Hence Monthly rainfall dada of June through October has been analyzed for developing rainfall forecasting models. As only June through September rainfall is treated as 'monsoon rainfall' by India Meteorological Department, June through September rainfall values are summed while reporting 'monsoon rainfall' in this work.

Monthly rainfall values during months June through October have been computed by Genetic Programming models. Monthly rainfall anomalies of observed and computed rainfall with reference to long term average (1950–2010) are also computed and presented through plots. The monthly rainfall values over training period, validation period and testing period for All India Summer Monsoon Rainfall can be visualized in Figs. 5–7 respectively.

To assess the prediction performance, the Pearson's Correlation Coefficient, Root Mean Square Error (RMSE) as well as Index of agreement 'd1' (Willmott et al., 2011) are computed.

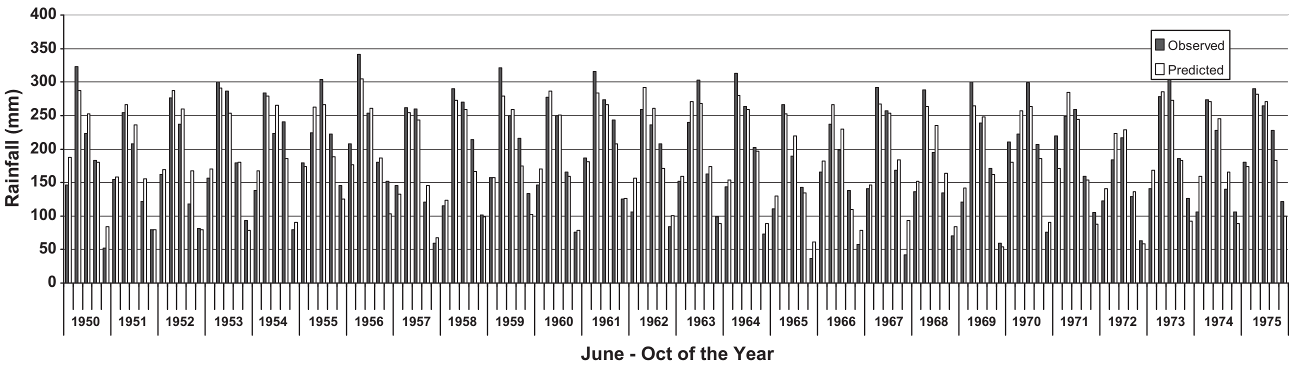

Fig. 5. All India – monthly rainfall June–October (training).

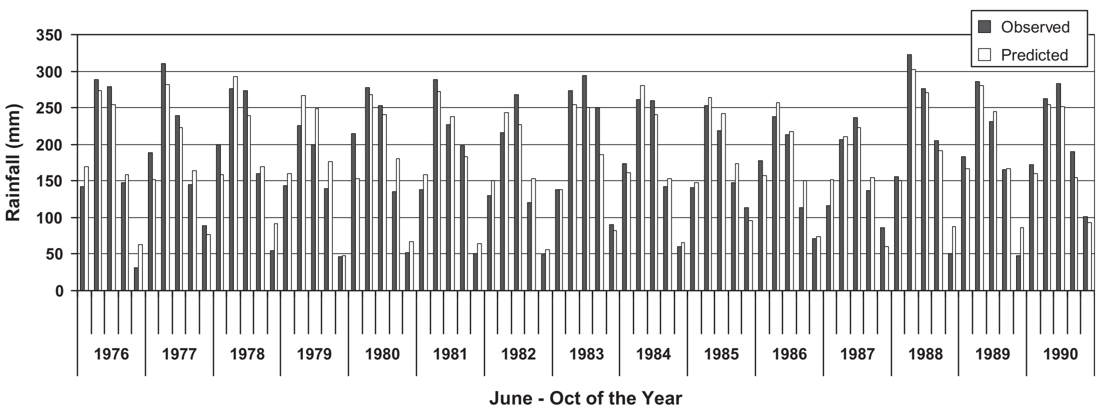

Fig. 6. All India – monthly rainfall June–October (validation).

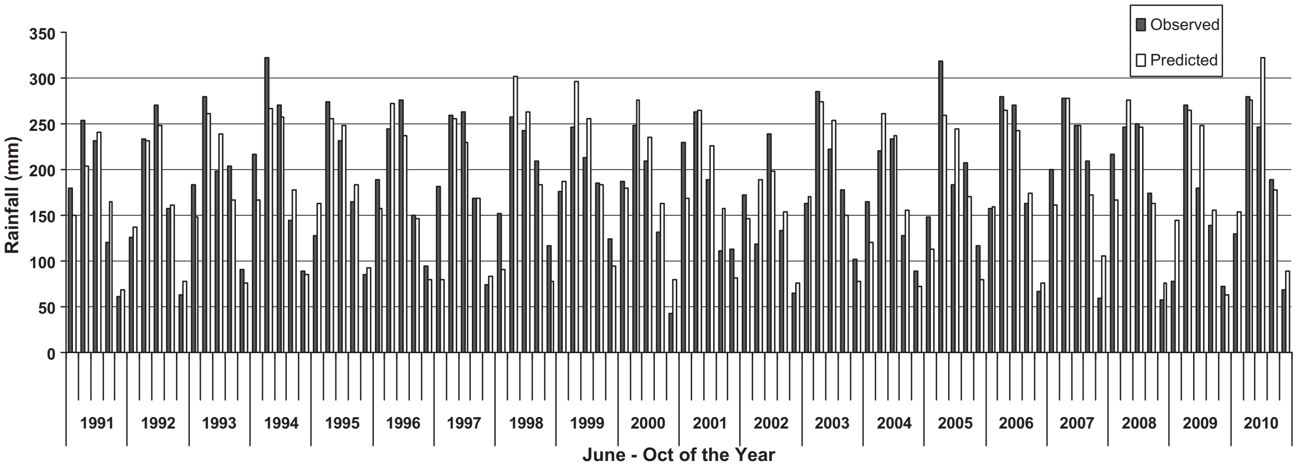

Fig. 7. All India – monthly rainfall June–October (testing).

Table 3
Correlation Coefficients, Index of agreement and RMSE for training, validation and testing (analyses with real time data of ENSO and EQUINOO).

No.	Region	Statistical parameter	Training	Validation	Testing
1	All India	C.C.	0.9385	0.9381	0.8683
		d1	0.8225	0.7754	0.7680
		RMSE (mm)	26.7	26.0	33.7
2	Central Northeast India	C.C.	0.9317	0.9160	0.8402
		d1	0.8090	0.7191	0.7329
		RMSE (mm)	38.9	48.3	55.4
3	West Central India	C.C.	0.9017	0.9088	0.8337
		d1	0.7806	0.7206	0.0.7243
		RMSE (mm)	44.5	41.9	51.5
4	North East India	C.C.	0.8803	0.8701	0.8075
		d1	0.7280	0.6146	0.6625
		RMSE (mm)	56.6	58.0	66.1
5	North West India	C.C.	0.8379	0.8899	0.7204
		d1	0.7358	0.6747	0.6939
		RMSE (mm)	46.0	38.9	49.2
6	Peninsular India	C.C.	0.7797	0.4135	0.3225
		d1	0.4524	0.2061	0.1767
		RMSE (mm)	39.8	42.1	48.6

Note: C.C.: Pearson's product moment Correlation Coefficient.
d1: Index of agreement.
RMSE: Root Mean Square Error.

Table 4
Input impacts of input variables in Genetic Programming models for all India monthly (June–October) rainfall analysis.

No.	Variable	Input impact
1	Historical Avg. of Monthly Rainfall	1.0
2	EN ($t-1$)	1.0
3	EN ($t-2$)	0.7
4	EN ($t-3$)	0.43
5	EQ ($t-1$)	1.0
6	EQ ($t-2$)	0.63
7	EQ ($t-2$)	0.53

Index of agreement 'd1' measuring model performance compares model predictions (Pi; $i = 1, 2, \ldots, n$) with pair-wise-matched observations (Oi; $i = 1, 2, \ldots, n$) that are judged to be reliable. The units of P and O should be the same. The set of model-prediction errors usually is composed of the ($Pi - Oi$) values, with most dimensioned measures of model performance being based on the central tendency of this set. The Index of agreement '$d1$' is given by the following equation:

$$d1 = 1 - \frac{\sum_{i=1}^{n} |P_i - O_i|}{\sum_{i=1}^{n} (|P_i - \overline{O}| + |O_i - \overline{O}|)} \tag{10}$$

where $\overline{O}$ is the estimated mean of the observations, Oi ($i = 1, 2, \ldots, n$). The values of the Pearson correlation coefficients, Index of agreement 'd1', Root Mean Square Error (RMSE) for training, validation and testing of GP models, for these analyses with real time monthly data of ENSO and EQUINOO is given in Table 3. The input impacts of different input variables on the output computed by Genetic Programming are listed in Table 4. The reasonable values of impact factors up to ($t-3$) time steps support the decision of including ENSO and EQUINOO indices of last three monthly time steps in the analysis. Monthly rainfall anomalies for All India Monsoon Rainfall are computed by using GP models.

From the GP predicted monthly values of rainfall, all India total monsoon rainfall is computed as a sum of computed rainfall values during June through September, for the concerned year. Hence monsoon rainfall anomalies are then computed for testing period, to study the consolidated effect over the rainfall of All India Summer Monsoon Rainfall. In the same way, the monthly rainfall values computed for Central Northeast India, West Central India, Northeast India, Northwest India and Peninsular India can be visualized in Figs. 8–12 respectively.

5.2. ISMR prediction at end of May based on ENSO and EQUINOO indices of March–April–May only

Monthly rainfall values during months June through September are computed by Genetic Programming models. The statement of the Pearson correlation coefficients for training, validation and testing for these analyses for this analysis is given in Table 5. The input impacts of different input variables on the output computed by Genetic Programming are listed in Table 6. The reasonable values of impact factors up to ($t-3$) time steps again support the

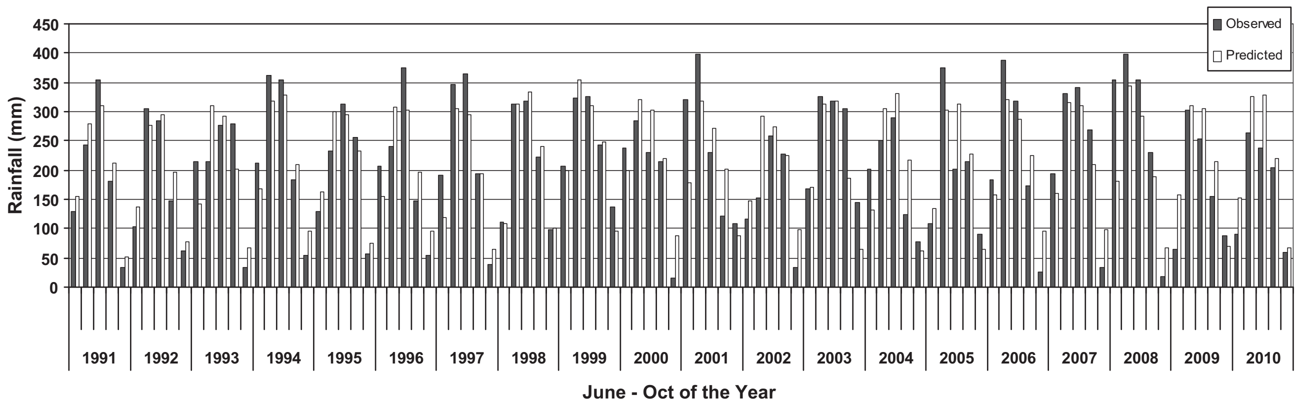

Fig. 8. Central Northeast India – monthly rainfall June–October (testing).

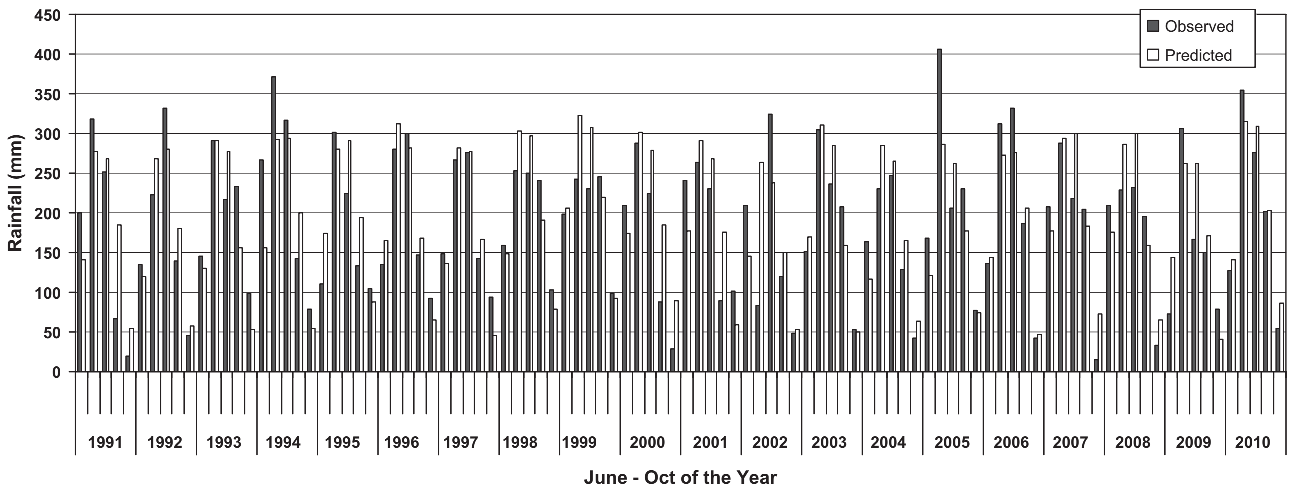

Fig. 9. West Central India monthly rainfall June–October (testing).

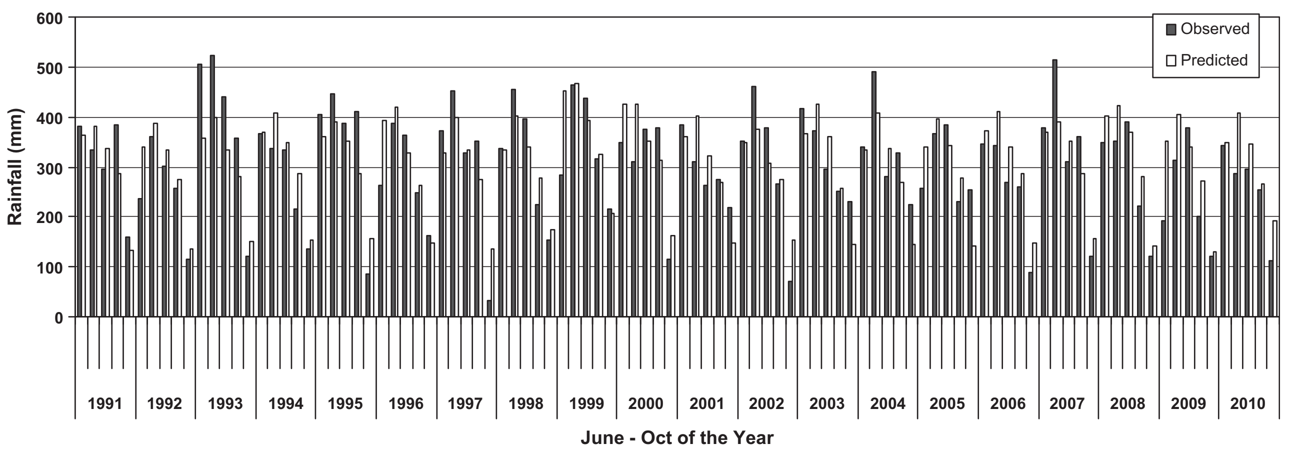

Fig. 10. Northeast India – monthly rainfall June–October (testing).

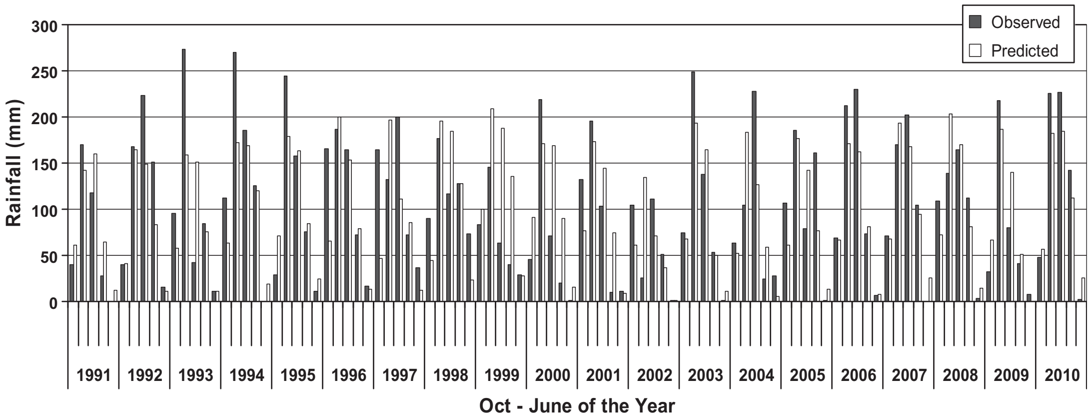

Fig. 11. Northwest India – monthly rainfall June–October (testing).

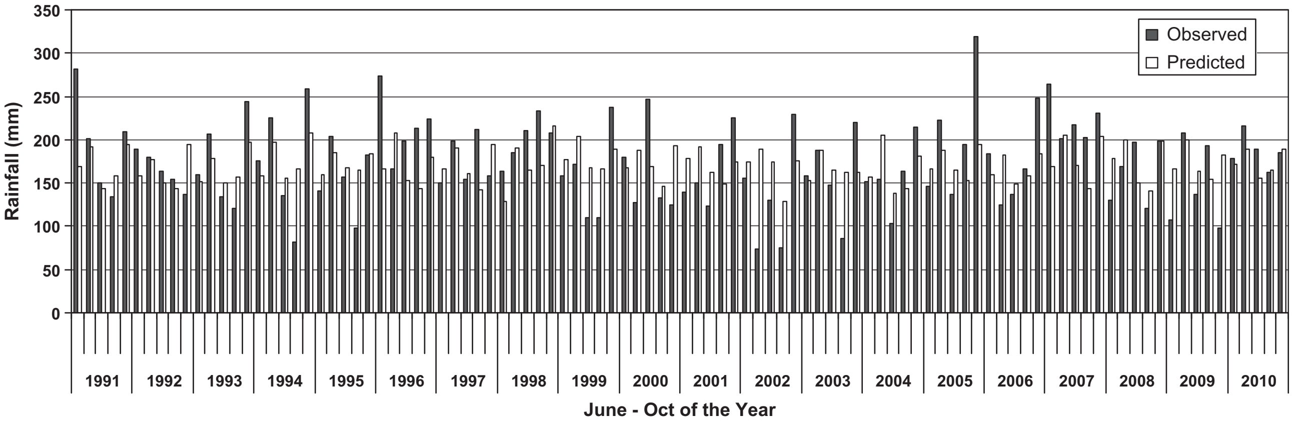

Fig. 12. Peninsular India – monthly rainfall June–October (testing).

decision of including ENSO and EQUINOO indices of last three monthly time steps.

Table 5
Correlation Coefficients (C.C.s), Index of Agreement (d1) and Root Mean Square Error (RMSE) for training, validation and testing (analyses with only March–April–May Indices of ENSO and EQUINOO).

No.	Region	Statistical Parameter	Training	Validation	Testing
1	All India	C.C.	0.8877	0.8832	0.7085
		d1	0.7680	0.7567	0.5311
		RMSE (mm)	32.3	32.0	51.2
2	Central Northeast India	C.C.	0.8955	0.8167	0.6693
		d1	0.7711	0.6119	0.5358
		RMSE (mm)	45.4	66.4	86.6
3	West Central India	C.C.	0.8479	0.8331	0.6442
		d1	0.7263	0.6353	0.5289
		RMSE (mm)	50.0	49.4	70.5
4	North East India	C.C.	0.7443	0.6921	0.4950
		d1	0.5699	0.4358	0.3376
		RMSE (mm)	78.6	87.4	112.2
5	North West India	C.C.	0.7550	0.8031	0.3962
		d1	0.6119	0.6211	0.3345
		RMSE (mm)	48.3	41.3	58.0
6	Peninsular India	C.C.	0.6964	0.4494	0.0361
		d1	0.3246	0.2056	0.1069
		RMSE (mm)	41.4	41.4	54.1

From the computed monthly values of rainfall, the monthly rainfall values were computed for All India, Central Northeast India, West Central India, Northeast India, Northwest India and Peninsular India.

5.3. Discussions

The effects of the large scale atmospheric circulation patterns over Pacific Ocean on ISMR are also modified by the circulation pattern over tropical Indian Ocean. The joint influence of these two patterns is established in earlier studies (Gadgil et al., 2004; Maity and Nagesh Kumar, 2006). Three monthly time lags of both ENSO and EQUINOO are considered in this analysis. The earlier study by Maity and Nagesh Kumar (2006) uses two monthly lags

Table 6
Input impacts of input variables in Genetic Programming models for all India monthly (June–October) rainfall analysis based on ENSO and EQUINOO indices of months March–April–May only.

No.	Variable	Input impact
1	Historical Avg. of Monthly Rainfall	1.0
2	EN (May)	0.83
3	EN (April)	0.80
4	EN (March)	0.77
5	EQ (May)	0.59
6	EQ (April)	0.77
7	EQ (March)	0.67

for ENSO and one monthly lag for EQUINOO. However the reason behind using three monthly lags in this analysis is such that the Genetic Programming methodology has its own mechanism to assess the impact of different input variables on model outputs. Hence the importance of the inputs for getting output can be judged. Reasonable values of 'input impacts' were observed up to three previous time steps in analysis and hence it was decided to use lag of three monthly time steps to encompass the effects all the physical processes on Pacific Ocean and Indian Ocean over a reasonable period of previous time steps. The input impacts of all input variables for the analysis based on real time data are reported in table in Table 4. The input impacts for May end with input ENSO and EQUINOO data up to May only are reported in Table 6.

It is interesting to see that the highest Correlation Coefficient in testing of GP model for monthly rainfall has been observed for 'All India Rainfall', treating India as a single unit. The value 0.868 of Correlation Coefficient is simply alluring. The Correlation Coefficients obtained for Central Northeast India (0.84), West Central India (0.83), and North-East India (0.80) are quite good. The North West India region gives C.C. of 0.72 which happens to be reasonable but the C.C. for Peninsular India is far less than other four regions. This large difference in C.C. of Peninsular India region and other regions indicates that ENSO and EQUINOO indices are not able to capture the total climatic mechanism behind summer monsoon rainfall over peninsular region of India. It may be noted that ENSO and EQUINOO do not well capture rainfall mechanism during summer monsoon over peninsular India. It might be due to the fact that it is not the chief rainy season for peninsular India. Again, rainfall in peninsular India due to North-East monsoon cannot be predicted on basis of ENSO and EQUINOO indices as these indices are almost irrelevant in connection with rainfall in peninsular India due to North-East monsoon. The Pearson product moment Correlation Coefficients for training, validation and testing for All India and five homogeneous monsoon regions are tabulated in Table 3.

Monthly as well as seasonal rainfall anomalies in form of percentage of mean long term average rainfall are calculated for observed and rainfall All India Summer Monsoon Rainfall. Total monsoon rainfall anomalies are computed for 'All India' as well as for five homogeneous monsoon regions.

For 'All India Summer Monsoon Rainfall' analysis, it can be observed that during the testing period of 1991–2010, out of 80 monthly rainfall prediction cases, prediction error less than 10% of mean for 44 number of months and prediction error was between 10% and 20% for 26 number of months. The percentage departure of predicted monthly rainfall values from actual observed values was calculated. It was found to be 16.79% of long term monthly average rainfall of the particular month. The prediction was successfully done for four important years of below normal rainfall 1991, 1992, 1995, 2001, 2002, 2004 and 2009. Above normal rainfall in 1994, 1998, 2006, 2007 were also rightly indicated with reasonable accuracy except year 1994. Monsoon rainfall was predicted with less than 10% of error (10% of long term average) for 15 years out of 20 years of analysis. The monsoon rainfall above/below normal, was predicted correctly in 11 out of 20 cases. The results of these analyses are summarized in Table 7.

Similar analysis was performed for five homogeneous Indian monsoon regions also. The results of these analyses are summarized in Table 8. The results indicate that for All India, Central Northeast India, West Central India and Northwest India regions, the results are excellent. The results for Northwest India are also encouraging. But the results for Peninsular India are not satisfactory. It may be due to the reason that only ENSO and EQUINOO indices are not sufficient to capture the mechanism behind rainfall over Peninsular India.

'May end analyses' also show similar pattern of results. For 'All India' analysis the Correlation Coefficient reached up to 0.708 for monthly predictions. It was followed by 0.66 for Central Northeast India, 0.64 for west central India, 0.495 for Northeast India and 0.396 for Northwest India. Just like the real time analysis the Correlation Coefficient for Peninsular India was far less.

For all India analysis, out of 20 years of prediction, the monsoon rainfall was predicted with less than 10% error (10% of long term average) for 16 years of analysis. The values of above/below normal were identified correctly in 12 cases. More important is that the below normal rains in 1991, 1992, 2000, 2002 and 2004, 2009 were rightly indicated. Above normal rains during 1997, 1998 and 2006, 2007 were also indicated with good accuracy. The results of total monsoon analyses at May end for All India, Central Northeast India, West Central India and Northwest India regions are summarized in Table 9. It can be observed that the results are reasonably good for All India, Central Northeast India and west central India regions. Results for Northeast India and Northwest India are not so appreciating. No indication can be expected for peninsular region based on ENSO and EQUINOO data up to May end.

Table 7
All India analysis of monthly Indian Summer Monsoon Rainfall (real time analysis).

All India monthly monsoon rainfall analysis June–September	
Number of years for which monsoon was tested	20
Number of months (June–September)	80
Prediction error less than 10% of mean for number of months	44
Prediction error between 10% and 20% of mean for number of months	26
Error more than 20% of mean for number of months Number of months	10

Table 8
Prediction performance of monsoon, in terms of percentage error, identification of trends above/below avg. and Correlation Coefficients while testing GP models (For real time analysis).

Zone	No. of monsoons predicted	No. of times prediction error less than 10% of mean rainfall	No. of times correct trend of monsoon indicated	C.C. in testing
All India	20	16	13	0.8683
Central Northeast India	20	13	15	0.8402
West Central India	20	12	12	0.8337
Northeast India	20	13	10	0.8075
Northwest India	20	7	10	0.7204
Peninsular India	20	7	6	0.3225

Note: C.C.: Pearson's product moment Correlation Coefficient.

Table 9
Prediction performance of monsoon, in terms of percentage error, identification of trends above/below avg. and Correlation Coefficients while testing GP models (for May end analysis).

Zone	No. of monsoons predicted	No. of times prediction error less than 10% of mean rainfall	No. of times correct trend of monsoon indicated	C.C. in testing
All India	20	16	12	0.708
Central Northeast India	20	10	10	0.669
West Central India	20	12	11	0.644
Northeast India	20	11	9	0.495
Northwest India	20	6	8	0.396
Peninsular India	20	5	7	0.036

Note: C.C.: Pearson's product moment Correlation Coefficient.

It was mentioned earlier that the proposed approach is based on the assumption of no temporal changes in the summer monsoon rainfall during training, validation and testing period. As per earlier evidence, 'All-India' summer monsoon rainfall is not showing any long-term trend. Thus for 'All India' analyses this can be a safe assumption. Observed increasing/decreasing extremes due global warming and climate change at subdivissional scale can be more effective for daily scale. This study considers monthly rainfall at homogeneous monsoon regions, which consists of several adjoining subdivisions, i.e., a larger spatial extent. Thus, for monthly total rainfall for homogeneous monsoon regions can safely be assumed to be stationary. If there were any temporal change, the model can still be useful with a recommendation that it is necessary to have some period over which the time series can safely be assumed to be stationary. In this study, it is observed that the model is performing equally well for the training, validation and testing periods, which indicates the stationary assumption is not very crude.

5.4. Comparison of two analyses

Comparison between Correlation Coefficients for analysis with continuous data and analysis with only March–April–May Indices of ENSO and EQUINOO was done to derive some conclusions regarding advantage of real time analysis over may end analyses. It can be observed that C.C. values in May end analysis are smaller than corresponding values for real time analysis, which is quite obvious. The comparison of Correlation Coefficients for both analyses in terms of their Correlation Coefficients in testing for All India as well as five homogeneous zones is presented in Table 10. The same can be visualized graphically in Fig. 13.

Essentially the models developed by the first method give better results than the second method. But it can be understood that the long range forecast of rainfall over four monsoon months before onset of monsoon has its own importance as it is one time total forecast of a seasonal rainfall. Such forecast though at lesser accuracy can help in taking crucial decisions of crop planning over the year, deciding reserves of reservoir water at end of May for June and assessing chances of filling in of conservation reservoirs in coming season.

Table 10
Comparison between C.C.s for analysis with continuous data and analysis with only March–April–May Indices of ENSO and EQUINOO.

No.	Region	Testing C.C. for continuous data	Testing C.C. for data up to May end only
1	All India	0.8683	0.7085
2	Central Northeast India	0.8402	0.6693
3	West Central India	0.8337	0.6442
4	North East India	0.8075	0.4950
5	North West India	0.7204	0.3962
6	Peninsular India	0.3225	0.0361

Note: C.C.: Pearson's product moment Correlation Coefficient.

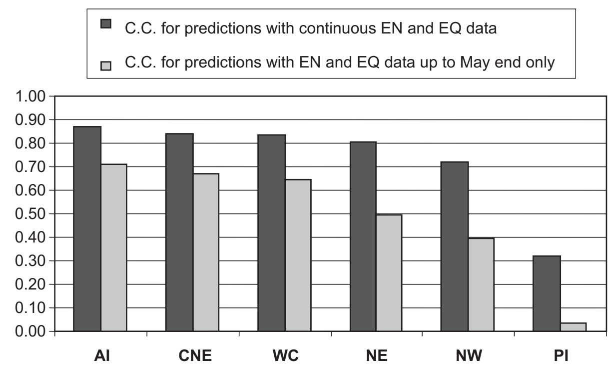

Fig. 13. Comparison of Correlation Coefficients for analysis with continuous ENSO and EQUINOO and with ENSO and EQUINOO data up to May end only. AI: All India, CNE: Central Northeast India, WC: West Central India, NE: North West India, Peninsular India.

6. Conclusions

Established research works indicate an association between the large-scale circulation pattern and hydrologic variables of large spatial and temporal scale. In this study, All India Summer Monsoon Rainfall as well as regional summer monsoon rainfall in India is investigated for possible influence of the large-scale circulation patterns on it. The ENSO and EQUINOO information is used as the large-scale input, which is established to be important for Indian hydroclimatology. Genetic Programming, which is a genetic algorithm based approach, is used to capture the complex relationship between inputs and outputs. Combinations of historical average rainfall of the particular month, and large-scale circulation pattern indices of ENSO and EQUINOO were explored for the monthly ISMR prediction.

It can be concluded that the influence of ENSO and EQUINOO on regional Indian Summer Monsoon Rainfall varies from region to region. The highest correlation was observed for Central North-East India and West Central India followed by North East and North West regions of India. Considerably less correlation was observed for peninsular India, which covers Tamilnadu, Pondicherry, Coastal Andhra Pradesh, Rayalseema and South Interior Karnataka regions of India. This can be attributed to the climate systems other than ENSO and EQUINOO, which cause winter rainfall in months of November and December in south part of Peninsular India. It can be observed that the years of positive as well as negative rainfall anomalies reasonably match in validation as well as testing data years also.

The GP based method proposed in this paper is demonstrated in the contest of Indian Summer Monsoon rainfall and is dependent on ENSO and EQUINOO indices. The performance of the model is found to be alluring. The proposed method is general in many aspects and can therefore be applied for similar studies and of teleconnected hydroclimatic variables. However, this study does not consider the temporal change in relationship between climate indices and rainfall, if any. Though such temporal change in relationship is still under investigation by researchers, inability to consider any possible change in relationship should be considered as a drawback of the proposed approach. In this context, it is also important to mention that the effect of climate change is also expected to modify the climate indices itself, which are being considered in the model over time and, thus, the signal, if already exists in the climate indices, is considered. Still, as stated before, consideration of the effect of change in relationship (which might also be caused by climate change effects) should be under scope of future study.

Appendix A

Appendix-I

****

/* Program for Prediction of Monthly Indian Summer Monsoon Rainfall, for Input Variable Combination, AvgMonRain(t), EN (t-1) through (t-3) and EQ (t-1) through (t-3) weekly time Steps i.e. using large-scale coupled atmospheric-oceanic circulation patterns */

```
#include <stdio.h>
#include<conio.h>
#include <stdlib.h>
#include <float.h>
#include <math.h>
#include <string.h>
```

/* Note: Input output variables in called function are as following:

v[0]= AvgMonRain, v[1]=$EN_{(t-1)}$, v[2]=$EN_{(t-2)}$,v[3]=$EN_{(t-3)}$, v[4]= $EQ_{(t-1)}$,v[5]= $EQ_{(t-2)}$,

v[6]= $EQ_{(t-3)}$, f[0] =$Rain_{(t)}$.

f[1], f[2], f[3], f[4] ,f[5], f[6], f[7] are temporary variables used by program */

```
#define TRUNC(x)(((x)>=0) ? floor(x) : ceil(x))
#define C_FPREM (_finite(f[0]/f[1]) ? f[0]-(TRUNC(f[0]/f[1])*f[1]) : f[0]/f[1])
#define C_F2XM1 (((fabs(f[0])<=1) && (!_isnan(f[0]))) ? (pow(2,f[0])-1) :
((!_finite(f[0]) && !_isnan(f[0]) && (f[0]<0)) ? -1: f[0]))

float DiscipulusCFunction(float v[])
{
 long double f[8];
 long double tmp = 0;
 int cflag = 0;

 f[0]=f[1]=f[2]=f[3]=f[4]=f[5]=f[6]=f[7]=0;

 L0:   f[0]+=v[6];
 L1:   f[0]*=v[4];
 L2:   f[2]-=f[0];
 L3:   f[0]=fabs(f[0]);
 L4:   tmp=f[2]; f[2]=f[0]; f[0]=tmp;
 L5:   f[0]=fabs(f[0]);
 L6:   f[0]-=-1.924433708190918f;
 L7:   f[0]=cos(f[0]);
 L8:   f[0]+=f[0];
 L9:   f[0]-=v[1];
 L10:  f[0]=fabs(f[0]);
 L11:  f[0]*=v[2];
```

```
  L12:  f[0]+=1.252994060516357f;
  L13:  f[0]*=v[4];
  L14:  f[0]/=-0.9636838436126709f;
  L15:  f[0]=cos(f[0]);
  L16:  f[0]/=-0.09100413322448731f;
  L17:  cflag=(f[0] < f[2]);
  L18:  f[0]-=-1.364777803421021f;
  L19:  f[0]-=-1.259177207946777f;
  L20:  f[0]-=-1.924433708190918f;
  L21:  if (cflag) f[0] = f[2];
  L22:  f[0]-=v[6];
  L23:  f[0]+=f[0];
  L24:  f[0]/=-1.427085638046265f;
  L25:  f[0]*=v[4];
  L26:  tmp=f[3]; f[3]=f[0]; f[0]=tmp;
  L27:  f[0]=cos(f[0]);
  L28:  f[0]*=v[1];
  L29:  f[0]+=v[5];
  L30:  f[0]-=v[0];
  L31:  f[0]+=0.7790718078613281f;
  L32:  f[0]*=-1.641227006912231f;
  L33:  f[0]=cos(f[0]);
  L34:  f[0]-=v[1];
  L35:  f[0]/=-0.09100413322448731f;
  L36:  f[0]+=f[3];
  L37:  f[0]/=-0.09100413322448731f;
  L38:  f[0]+=v[0];
  L39:

  if (!_finite(f[0])) f[0]=0;

  return f[0];
}

/*GP generated code ends here*/

  void main(int argc, float argv[])
  {
     FILE *pinputfile;
     FILE *poutputfile;
     char string[1000];
     char *inputs = new char[100];
     char strnull[1];
     strnull[0]=0x00;
      printf("Enter number of Inputs:");
      gets(inputs);
```

```
  int inp = atoi (inputs);  /*Converts string to integer*/
 pinputfile = fopen ("ismr.txt", "r");

if (pinputfile==NULL)
{
  perror("Error opening file input.txt");
  getch();
}
else
{
  poutputfile=fopen("x.txt", "w");
  int icount=0;

  while (fgets (string, 1000, pinputfile)) /*Read string from file*/
      {
        {
           float fargs[7];
           sscanf(string, "%f\t%f\t%f\t%f\t%f\t%f\t%f\t%f",
                   &fargs[0], &fargs[1],&fargs[2], &fargs[3], &fargs[4],&fargs[5],
               &fargs[6]);

               ;/*Read formatted input from a string*/

        for (int icount =inp; icount < 7; icount++)
        {
         fargs[icount] =0;
        }

        puts (string);/*Writes line to file*/
        char output [100];

        float fout = DiscipulusCFunction(fargs);
        sprintf (output, "Prediction: %f", fout);/*Writes formatted output to a string*/
        puts (output);/*Writes a line to file*/

        char outputstring [100];
        char outputfilestring[1000];

        strncpy(outputstring, strnull, 100); /*Copies a given number of characters of one
                                                     string to another*/
        strncpy(outputfilestring, strnull, 1000);

      for (int icnt =0; icnt < inp; icnt++)
      {
        sprintf(outputstring, "%f", fargs[icnt]);/*Writes formatted output to a string*/
        strcat(outputfilestring, outputstring);/*Appends one string to another*/




              strcat(outputfilestring, "\t");/*Appends one string to another*/

            }
            sprintf(outputstring," %f\r\n", fout);/*Writes formatted output to a string*/
            strcat(outputfilestring, outputstring);/*Appends one string to another*/
            fputs(outputfilestring, poutputfile);/*Wries a string to a file*/
      }
     }

      fclose(poutputfile);
      fclose(pinputfile);
    }
   }
```

References

Ashok, K., Guan, Z., Yamagata, T., 2001. Impact of Indian Ocean dipole on the relationship between the Indian monsoon rainfall and ENSO. Geophys. Res. Lett. 28, 4499–4502.

Ashok, K., Guan, Z., Saji, N., Yamagata, T., 2004. Individual and combined effect of ENSO and Indian Ocean Dipole on the Indian summer monsoon. J. Clim. 17 (16), 3141–3155. http://dx.doi.org/10.1175/1520-0442(2004) 017<3141:IACIOE>2.0.CO; 2.

Aytek, A., Kisi, O., 2008. A genetic programming approach to suspended sediment modeling. J. Hydrol. 351, 288–298.

Babovic, V., 2000. Translating data to knowledge: a case study in sediment transport. Comput.-Aided Civil Infrastruct. Eng. 15 (1), 1–7.

Babovic, V., 2005. Data mining in hydrology. Hydrol. Process. 19, 1–7.

Babovic, V., Abbott, M.B., 1997a. The evolution of equation from hydraulic data, Part I: Theory. J. Hydraul. Res. 35 (3), 397–410.

Babovic, V., Abbott, M.B., 1997b. The evolution of equation from hydraulic data, Part II: Applications. J. Hydraul. Res. 35 (3), 411–430.

Babovic, V., Keijzer, M., 2000. Genetic programming as a model induction engine. J. Hydroinform. 2 (1), 35–60.

Babovic, V., Keijzer, M., 2002. Rainfall runoff modeling based on genetic programming. Nord. Hydrol. 33, 331–343.

Banerjee, A.K., Sen, P.N., Raman, C.R.V., 1978. On foreshadowing southwest monsoon rainfall over India with mid-tropospheric circulation anomaly of April. Indian J. Meteorol. Hydrol. Geophys. 29, 425–431.

Baptist, M., Babovic, V., Rodrigues, J., Keijzer, M., Uittenbogaard, R., Mynett, A., Verwey, A., 2007. On inducing equations for vegetation resistance. J. Hydraul. Res. 45 (4), 435–450.

Bhalme, H.N., Jadhav, S.K., Mooley, D.A., Ramana Murty, B.H.V., 1986. Forecasting of monsoon performance over India. J. Climatol. 6, 347–354.

Brameier, M., Banzhaf, W., 2004. Evolving teams of predictors with linear genetic programming. Genetic Programming and Evolvable Machines 2 (4), 381–407.

Chau, K.W., 2006. Particle swarm optimization training algorithm for ANNs in stage prediction of Shing Mun River. J. Hydrol. 329 (3–4), 363–367.

Douglas, W.W., Wasimi, S.A., Islam, S., 2001. The El Niño southern oscillation and long-range forecasting of flows in Ganges. Int. J. Climatol. 21, 77–87.

Dracup, J.A., Kahya, E., 1994. The relationship between U.S. streamflow and La Niña Events. Water Resour. Res. 30 (7), 2133–2141.

Drecourt, J.P., 1999. Application of neural networks and genetic programming to rainfall runoff modeling. Danish Hydraulic Institute (Hydro-Informatics Technologies HIT). 3rd DHI Software Conference \& DHI Software Courses. <http://www.cs.bham.ac.uk/~wbl/biblio/gp-html/drecourt_1999uANNGPrrm.html>.

Drunpob, A., Chang, N.B., Beaman, M., 2005. Stream flow rate prediction using genetic programming model in a semi-arid coastal watershed. Proceedings of EWRI 2005, ASCE.

Eltahir, E.A.B., 1996. El Niño and the natural variability in the flow of the Nile River. Water Resour. Res. 32 (1), 131–137.

Francone, F.D., 1998. Discipulus Owner's Manual. Machine Learning Technologies, Inc., Littleton, Colorado.

Gadgil, S., Vinayachandran, P.N., Francis, P.A., 2003. Droughts of the Indian summer monsoon: role of clouds over the Indian Ocean. Curr. Sci. 85 (2), 1713–1719.

Gadgil, S., Vinayachandran, P.N., Francis, P.A., Gadgil, S., 2004. Extremes of the Indian summer monsoon rainfall: ENSO and equatorial Indian Ocean Oscillation. Geograph. Res. Lett. 31 (L 12213), doi:10.1029/2004GLO19733.

Giustolisi, O., 2004. Using genetic programming to determine Chèzy resistance coefficient in corrugated channels. J. Hydroinform., IWA-IAHR Publishing, UK 6 (3), 157–173.

Giustolisi, O., Savic, D.A., 2006. A symbolic data-driven technique based on evolutionary polynomial regression. J. Hydroinform., IWA-IAHR Publishing, UK 8 (3), 207–222.

Gowariker, V., Thapliyal, V., Sarkar, R.P., Mandal, G.S., Sikka, D.R., 1989. Parametric and power regression models – new approach to long range forecasting. Masaum 40, 115–122.

Gowariker, V., Thapliyal, V., Kulshrestha, S.M., Mandal, G.S., Sen Roy, N., Sikka, D.R., 1991. A power regression models for long range forecast of southwest monsoon rainfall over India. Masaum 42, 125–130.

Harris, E., Falconer, R., Babovic, V., 2003. Velocity predictions in compound channels with vegetated floodplains using genetic programming. J. River Basin Manage. 1 (2).

Hong, Y.S., Bramidimarri, R., 2003. Evolutionary self-organising modelling of a municipal wastewater treatment plant. Water Res. 37, 1199–1212. http://dx.doi.org/10.1016/S0043-1354(02)00493-1.

Jagannathan, P., 1960. Seasonal Forecasting in India: A Review. India Meteorological Department, Pune, p. 67.

Jain, S., Lall, U., 2001. Floods in a changing climate: does the past represent the future? Water Resour. Res. 37 (12), 3193–3205.

Keijzer, M., Babovic, V., 2002. Declarative and preferential bias in GP-based scientific discovery. Gen. Programm. Evolut. Mach. 3 (1), 41–79.

Kisi, O., Shiri, J., 2011. Precipitation forecasting using wavelet genetic programming and wavelet-neuro-fuzzy conjunction models. Water Resour. Manage.. http://dx.doi.org/10.1007/s11269-011-9849-3.

Koza, J.R., 1992. Genetic Programming: On the Programming of Computers by Means of Natural Selection. MIT Press, Cambridge, MA, USA.

Krishna Kumar, K., Soman, M.K., Rupa Kumar, K., 1995. Seasonal forecasting of Indian summer monsoon: a review. Weather 12, 449–467.

Krishna Kumar, K., Rupa Kumar, K., Pant, G.B., 1997. Pre-monsoon maximum and minimum temperature over India in relation to the summer monsoon rainfall. Int. J. Climatol. 17, 1115–1127.

Krishna Kumar, K., Rajagopalan, B., Cane, M.A., 1999. On the weakening relationship between the Indian Monsoon and ENSO. Science 284 (5423), 2156–2159, DOI: 10.1126/science.284.5423.2156.

Kung, E.C., Sharif, T.A., 1982. Long-range forecasting of the Indian summer monsoon onset and rainfall with upper air conditions. J. Meteorol. Soc. Jpn. 60, 672–681.

Li, X.Y., Chau, K.W., Cheng, C.T., Li, Y.S., 2006. A Web-based flood forecasting system for Shuangpai region. Adv. Eng. Softw. 37 (3), 146–158.

Lin, J.Y., Cheng, C.T., Chau, K.W., 2006. Using support vector machines for long-term discharge prediction. Hydrol. Sci. J. 51 (4), 599–612.

Liong, S.Y., Gautam, T.R., Khu, S.T., Babovic, V., Keijzer, M., Muttil, N., 2002. Genetic programming: a new paradigm in rainfall-runoff modeling. J. Am. Water Resour. Assoc. 38 (3), 705–718.

Maity, R., Nagesh Kumar, D., 2006. Hydroclimatic association of monthly summer monsoon rainfall over India with large-scale atmospheric circulation from tropical Pacific Ocean and Indian Ocean region. Atmosph. Sci. Lett., R. Meteorol. Soc. (RMetS), Wiley InterScience 7, 101–107. http://dx.doi.org/10.1002/asl.141.

Maity, R., Nagesh Kumar, D., 2008. Probabilistic prediction of hydroclimatic variables with nonparametric quantification of uncertainty. J. Geophys. Res. – Atmosph. 113, D14105. http://dx.doi.org/10.1029/2008JD009856.

Makkeasorn, A., Chang, N.B., Zhou, X., 2008. Short-term streamflow forecasting with global climate change implications – a comparative study between genetic programming and neural network models. J. Hydrol. 352, 336–354.

Marcella, M.P., Eltahir, E.A.B., 2008. The hydroclimatology of Kuwait: explaining the variability of rainfall at seasonal and interannual time scales. J. Hydrometeor. 9, 1095–1105.

Muttil, N., Chau, K.W., 2007. Machine learning paradigms for selecting ecologically significant input variables. Eng. Appl. Artif. Intell. 20 (6), 735–744.

Muttil, N., Liong, S.Y., 2001. Improving Runoff forecasting by input variable selection in Genetic Programming. ASCE World Water Congress, Orlando, Florida, USA 111 (2001), 76, doi:10.1061/40569(2001)76.

Partal, T., Kisi, O., 2007. Wavelet and neuro fuzzy conjunction model for precipitation forecasting. J. Hydrol. 342, 199–212.

Parthasarathy, B., Diaz, H.F., Eischeid, J.K., 1988. Prediction of All India summer monsoon rainfall with regional and large-scale parameters. J. Geophys. Res. 93 (5), 5341–5350.

Parthasarathy, B., Rupa Kumar, K., Deshpande, V.R., 1991. Indian summer monsoon rainfall and 200mb meridional wind index: application for long range prediction. Int. J. Climatol. 11, 165–176.

Parthasarathy, B., Munot, A.A., Kothawale, D.R., 1995. Monthly and Seasonal Rainfall Series for All-India Homogeneous Regions and Meteorological Subdivisions: 1871–1994. IITM Research, Report No. RR-065.

Rajeevan, M., Pai, D.S., Dikshit, S.K., Kelkar, R.R., 2004. IMD's new operational models for long-range forecast of southwest monsoon rainfall over India and their verification for 2003. Curr. Sci. 86 (3), 22–431.

Rao, K.N., 1965. Seasonal forecasting – India. Proc. of Symp. on 'Research and Development Aspects of Long Range Forecasting'. WMO-IUGG Tech. Note No. 66 WMO-No. 162-TP-79. World Meteorological Organisation Geneva, pp. 17–30.

Rao, K.N., Rama Moorthy, K.S., 1960. Seasonal (Monsoon) rainfall forecasting in India. Proc. of Symp. on 'Monsoon of the World', held at New Delhi February 1958, Published by India Meteorological Department, New Delhi, pp. 237–250.

Saji, N.H., Goswami, B.N., Vinayachandran, P.N., Yamagata, T., 1999. A dipole mode in the tropical Indian Ocean. Nature 401, 360–363.

Saks, P., Maringer, D., 2010. Evolutionary money management. In: Brabazon, A., O'Neill, M., Dietmar, G., Maringer, D.G. (Eds.), Natural Computing in Computational Finance, vol. 3. Springer-Verlag, Berlin, Heidelberg, ISBN-978-3-642-13949-9.

Savur, S.R., 1931. The seasonal Forecasting Formulae Used in the India Meteorological Department. Scientific Notes, vol. 4, no. 37. Published by India Meteorological Department, New Delhi.

Sette, S., Boullart, L., 2001. Genetic programming: principles and applications. Eng. Appl. Artif. Intell. 14, 727–736.

Shiri, J., Kisi, O., 2011. Application of artificial intelligence to estimate daily pan evaporation using available and estimated climatic data in the Khozestan Province (Southwestern Iran). ASCE J. Irrig. Drain. Eng. 137 (7), 412–425.

Shukla, J., Paolino, D.A., 1983. The Southern Oscillation and long range forecasting of the summer monsoon over India. Mon. Weather Rev. 111, 1830–1837.

Web reference. <ftp://www.tropmet.res.in/pub/data/rain/iitm-regionrf.txt>.

Whigham, P.A., Crapper, P.F., 2001. Modeling rainfall–runoff using genetic programming. Math. Comput. Model. 33, 707–721, Canberra, Australia.

Willmott, C. J., Robeson S.M., and Matsuura K., 2011. Short Communication - a refined index of model performance. Int. J. Climatol., in press, http://dx.doi.org/10.1002/joc.2419.