Comparative Analysis of the Predictability of Linear & Non-linear Methods for Seasonal Streamflow Forecasting: A Case Study of New South Wales (NSW)

A Thesis Submitted in Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Rijwana Ishat Esha

Department of Civil and Construction Engineering
Faculty of Science, Engineering and Technology
Swinburne University of Technology
Hawthorn, VIC 3122
Australia

December 2020

Abstract

High interannual variability of streamflow resulting from the extensive topographic variation and climatic inconsistency cause immense difficulties to the water users and planners of Australia. New South Wales, which is situated in the south-eastern part of Australia, is the most populous state and is one of the major contributors of Australia's agricultural income. The inter-annual variation of streamflow hampers the agricultural production and proper allocation of water of the state largely. Therefore, prediction of streamflow over a large time period will enable the water allocators and agricultural producers to take the low-risk decision at an earlier stage of the crop year which will ultimately enhance the economic growth of the country. Since streamflow is largely dependent on rainfall, it appears to be a more complex phenomenon compared to rainfall. Thereby, long-lead forecasting of streamflow rather than rainfall will be more beneficial to the irrigators. To date, many researchers have attempted to predict future streamflow and rainfall using oceanic and atmospheric indices with the help of both statistical and dynamic approaches. While most of the past studies were concentrated on revealing the relationship between streamflow of single concurrent or lagged climate indices, this study makes an effort to explore the combined impact of large-scale climate drivers to forecast seasonal streamflow of New South Wales (NSW) region. To accomplish the aim of this study, several oceanic and atmospheric climate indices are selected considering their influence on the streamflow of NSW which includes but not limited to four major climate drivers of this region PDO (Pacific Decadal Oscillation), IPO (Inter Decadal Pacific Oscillation), IOD (Indian Ocean Dipole) and the ENSO (El Nino Southern Oscillation) indices. Many past research works demonstrated that different regions of NSW are influenced by different climate modes which lead the present study to divide NSW into four regions with a view to identifying the regional variation of the impacts of various climate drivers. At first single lagged correlation analysis is performed to identify the individual interactions of indices with spring streamflow till nine lagged months which is, later on, exploited as the basis for selecting input variables for developing Multiple Linear Regression (MLR) models to examine the extent of the combined impact of the selected climate drivers on forecasting spring streamflow several months ahead. As many researchers have claimed that a non-linear approach may better capture the relationship between climate variables and seasonal streamflow, Multiple Non-Linear Regression (MNLR) Analysis is conducted to explore the underlying non-linear relationship between seasonal streamflow and climate indices. Finally, for further improvement, an Artificial Intelligence (AI) based method, Gene Expression Programming (GEP) is introduced to evaluate the potential of this method for forecasting seasonal streamflow of NSW. Performances of the developed models are assessed using standard statistical measures such as RRSE (Root Relative Squared Error), RAE (Relative Absolute Error), RMSE (Root Mean Square Error), MAE (Mean Absolute Error) and Pearson correlation (r) values. A comparative analysis is performed among the applied methods where GEP method has outperformed the other two methods. The highest predictabilities of the GEP based models are evident from the Pearson correlation (r) values ranging between 0.57 and 0.97, which are mostly about twice the values achieved by MLR and MNLR models. The developed GEP models are able to predict spring streamflow up to 5 months in advance with significantly high correlation values. The current study showed better performances while compared to the previous research studies in this field. This research concludes that GEP models can be used to predict seasonal streamflow of NSW incorporating large-scale multiple climate indices as predictors. In future, a similar concept will be applied to other regions for other seasons to explore the spatial and seasonal variation of influences different climate indices on seasonal streamflow.

Acknowledgements

First and foremost, I wish to express my greatest gratitude to the Almighty Allah (SWT), for all his blessings on me during the course of life and for bestowing me with enough courage and strength to pursue my studies.

I would like to express my sincere appreciation and gratitude to my principal supervisor, Associate Professor Dr. Monzur Alam Imteaz, for his continuous support and invaluable guidance during my study. His insightful feedback pushed me to sharpen my thinking and brought my work to a higher level. I want to thank him, especially for his constant motivation to continue my study amid the toughest phase of my personal life. I would also like to thank my internal supervisory committee member, Dr Shirley. Gato-Trinidad.

I convey my sincere gratitude towards my PhD progress review panel members Dr. Arul Arulrajah, Dr. Abdullah Al Mahmud and Dr. Pathmanathan Rajeev for their constructive criticism, appreciation and suggestions during the progress review presentations. I want to extend my sincere appreciation to Dr. Iqbal Hossain and Dr. Fatemeh Mekanik for their invaluable support and guidance during this research.

I also acknowledge Swinburne University of Technology for supporting and funding this research project by providing me with the Tuition Fee Scholarship. I wish to thank all my friends and colleagues who made my stay at the Swinburne University of Technology an enjoyable experience. Additionally, I would like to extend my appreciation towards the Bureau of Meteorology, Melbourne, VIC, Australia, for their support with the required data throughout the duration of this research.

Last but not least, I wish to express my sincere gratitude to my parents, Late Md Enamul Haque and Shirin Haque for their never-ending blessings and for inspiring me to be ambitious to pursue higher studies. My heartiest thanks to my husband Md Farhat Alam for always being supremely supportive and encouraging for developing my career. I want to express my love towards my newborn daughter Fareesha Alam Zerlin who is my biggest inspiration to move on. I want to especially thank my aunt Sara Shelley for her contribution to support me for my studies. Thanks to my entire family whose love, patience and encouragement have given me power, enthusiasm, and motivation not only during my PhD study but throughout my life.

Declaration

I, Rijwana Ishat Esha, declare that this Doctor of Philosophy (PhD) thesis has been composed solely by myself, except where explicitly stated otherwise by reference or acknowledgement and that this work has not been previously, and is not concurrently, submitted in whole or in part, for any other degree or professional qualification except as specified. As a sole author of this PhD thesis, I retain all the copyright and ownership rights of my research work including its use in any future works (such as articles or books) all or part of my work.

To the best of my knowledge, this PhD thesis does not contain any material previously published or written by any other person except the help that I received in my research work from any person and the contributions of the respective authors are gratefully acknowledged in the publications.

I hereby declare that I am the sole author of this thesis. This is an original copy of the thesis, including any required final revisions, as accepted by my examiners.

Rijwana Ishat Esha

Rijwana Ishat Esha
September 2020

Dedication

I want to dedicate this thesis to my parents, my husband and my daughter, whom I love and owe my life to.

List of Publication

Journals

- Esha, R.I. and Imteaz, M.A., 2019. Assessing the predictability of MLR models for long-term streamflow using lagged climate indices as predictors: a case study of NSW (Australia). Hydrology Research, 50(1), pp.262-281**(Q2, SCOPUS INDEXED)**

- Esha, R. and Imteaz, M.A., 2020. Pioneer use of gene expression programming for predicting seasonal streamflow in Australia using large scale climate drivers. *Ecohydrology*, p.e2242. **(Q1, SCOPUS INDEXED)**

- Hossain, I., Esha, R. and Alam Imteaz, M., 2018. An attempt to use non-linear regression modelling technique in long-term seasonal rainfall forecasting for Australian Capital Territory. Geosciences, 8(8), p.282. **(Q2, SCOPUS INDEXED)**

Conference Proceedings

- Esha, RI & Imteaz, MA 2017, 'MLR Analysis for Forecasting Seasonal Streamflow Using Large Scale Climate Drivers: A Case Study for NSW', *Proceedings of 65 th ISERD International Conference, Mecca, Saudi Arabia, 23rd -24th January 2017,* ISBN: 978-93-86291-92-9
- Esha, RI & Imteaz, MA 2017, Seasonal streamflow prediction using large scale climate drivers for NSW region. MODSIM2017, 22nd International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, December 2017, ISBN: 978-0-9872143-7-9.
- Esha, RI & Imteaz, MA 2017, Non-linear multiple regression analysis for predicting seasonal streamflow using large scale climate mode. MODSIM2017, 22nd International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, December 2017, ISBN: 978-0-9872143-7-9.

- Esha, R.I., Imteaz, M.A. and Nazari, A., 2019, October. Assessing Gene Expression Programming as a technique for seasonal streamflow prediction: A case study of NSW. In IOP Conference Series: Earth and Environmental Science (Vol. 351, No. 1, p. 012004). IOP Publishing.

Table of Contents

List of Figures

List of Tables

List of Acronym

Acronyms	Description
AI	Artificial Intelligence
ANN	Artificial Neural Networks
ANFIS	Adaptive Network-Base Fuzzy Inference System
ARMA	Auto-Regressive Moving Average
GA	Genetic Algorithms
GEP	Genetic Expression Programming
GP	Genetic Programming
BoM	Bureau of Meteorology
BJP	Bayesian Joint Probability
CWNSW	Central West New South Wales
CSIRO	Commonwealth Scientific and Industrial Research Organization
DMI	Dipole Mode Index
EMI	ENSO Modoki Index
ENSO	El Nino Southern Oscillation
IOD	Indian Ocean Dipole
IPO	Interdecadal Pacific Oscillation
KNMI	Royal Netherlands Meteorological Institute
MAE	Mean Absolute Error
MDB	Murray Darling Basin
MLR	Multiple Linear Regression
MNLR	Multiple Non-Linear Regression
NNSW	Northern New South Wales
NSE	Nash–Sutcliffe model efficiency coefficient
NSW	New South Wales
PDO	Pacific Decadal Oscillation
POAMA	Predictive Ocean Atmosphere Model for Australia
RAE	Relative Absolute Error

Acronyms	Description
RRSE	Root Relative Squared Error
SAM	Southern Annular Mode
SLP	Sea Level Pressure
SST	Sea Surface Temperatures
SNSW	Southern New South Wales
SOI	Southern Oscillation Index
SAM	Southern Annular Mode
SST	Sea Surface Temperature
SSTA	Sea Surface Temperature anomaly
WNSW	Western New South Wales

Chapter 1
Introduction

1.1 Background

Australia is considered as the driest inhabited continent of the world with 70% of its land arid or semi-arid land. The geographic location and extensive topographic variation present high climatic inconsistency in Australia which results in even higher inter-annual streamflow variability across the country, which is almost twice of the rivers in any other part of the world (McMahon et al. 1992). As a consequence, it presents many difficulties to the irrigators, agricultural producers, water managers and planners to allocate irrigation water and environmental flows, manage and operate reservoir, municipal supply water, estimate future hydroelectricity supply etc. There is a significant regional variation of streamflow and run-off throughout Australia, and in recent years the streamflow occurrences are more visible in the eastern part compared to the rest of the parts of the country. Therefore, prediction of rainfall and streamflow over long timescales can help in low-risk decisions making for water resource management (White et al. 2004; Abawi et al. 2005).

Australia is surrounded by Pacific, Indian and Southern Ocean, is greatly influenced by the climatic anomalies originated from the ocean. It has been accepted by the hydrologists that there exists a strong correlation between the streamflow and the large-scale atmospheric circulation patterns. The impacts of each climate index, including the Sea Level Pressure (SLP) and Sea Surface Temperature (SST) anomalies have spatial and seasonal variation. Over the years, researchers have studied the relationship between Australian rainfall, streamflow and climate indices. Dutta et al. (2006) indicated the necessity for exploring the skills of forecasting streamflow and rainfall with different lead times exploiting various climate indicators. It was mentioned by him that streamflow forecast is more significant compared to rainfall forecast as it can be predicted with longer lead times. Thereby, streamflow forecast enables the water users to make the decision earlier in the year, which ultimately increases the potential of financial benefits.

Two main sources of streamflow are initial catchment and future climate (for instance, oceanic and atmospheric climate indices) conditions (Robertson et al. 2009). Initial catchment conditions can be indicated by antecedent streamflow, antecedent rainfall, soil moisture or groundwater levels. Chiew et al. (1998) stated that serial correlations (persistence) is very high in Australia and the reason for this is the delaying responses of the rainfall-runoff process which ultimately gives streamflow data memory of several months. While comparing to initial catchment condition, remote climate drivers have better predictability of streamflow as the climate indices fluctuate at very low frequencies which can impact the streamflow easily. Moreover, developing streamflow forecasting models incorporating initial catchment condition is more complicated. Thereby, this study will aim to focus on the potential of lagged climate modes to explain future seasonal streamflow.

The El Niño Southern Oscillation (ENSO) phenomenon, which results from the large-scale interactions between the ocean and atmospheric circulation processes in the equatorial Pacific Ocean, has direct influences on the climate variability over many parts of the world (e.g. Ropelewski and Halpert 1987; Kiladis and Diaz 1989; Nicholls et al. 1991). El Nino and La Nina events are responsible for the different climatic conditions around the Pacific including eastern Australia (Stone and Auliciems 1992; Nazemosadat and Cordery 1997; Hoerling et al. 2001; CPTEC 2006; Chiew 2006). Several studies revealed the influences of ENSO on streamflow throughout Australia (Piechota et al. 1998; Chiew et al. 1998, 2003; Dettinger and Diaz 2000; Dutta et al. 2006). Chiew et al. (1998) and Piechota et al. (1998) found that ENSO based (Southern Oscillation Index (SOI) and SST) streamflow predictions in northeast Australia are better than the forecasts from climatology.

The El Nin˜o–Southern Oscillation Modoki (EMI) events have significant influences on the climate of many parts of the world including Japan, New Zealand, western coast of United States (Ashok et al. 2007), Australia (Taschetto and England 2009), South China (Feng and Li 2011). According to Taschetto and England (2009), EMI significantly decreases northern and north-western rainfall while traditional ENSO indices decrease south-eastern and north-eastern rainfall in Australia.

Though the dominant source of inter-annual variability in Australian rainfall and streamflow is believed to be ENSO phenomenon, some recent evidence shows that Eastern Australia is also influenced by Indian Ocean Dipole (IOD) as well as interdecadal modulation of ENSO as a result of the low-frequency variability in the

Pacific Ocean, which is referred as Pacific Decadal Oscillation or PDO (Westra and Sharma 2008). Cai et al. (2011) and Risbey et al. (2009) found that IOD has an impact on austral winter (June to October) in the southern part of Australia whereas ENSO has a strong influence on austral spring rainfall as a result of the strong covariation of ENSO and IOD. According to Meyers et al. (2007), IOD and ENSO can sometimes occur together in such a way that strengthens each other. Many researchers (e.g., Power et al. 1999; Kiem et al. 2003; Kiem and Franks 2004) have demonstrated the influence of the Interdecadal Pacific Oscillation (IPO) to be significant on rainfall variation on a decadal to multi-decadal timescale. King et al. (2013) suggested that the IPO plays a significant role in the frequency of major floods during the 1950s, 1970s and 2010–2011.

The climate of southeast Australia is influenced by four major climate drivers originating in the Pacific Ocean, the Indian Ocean and the Southern Ocean ENSO, IPO (PDO), Southern Annular Mode (SAM) and IOD (Duc et al. 2017). The correlations between ENSO phenomenon and seasonal rainfall in central NSW are found to be the strongest during spring (McBride and Nicholls 1983). Robertson and Wang (2009) investigated on 12 climatic predictors with a view to selecting the best predictors for forecasting seasonal streamflow in Murrumbidgee catchment of NSW and found that the greatest predictability occurred between September and December while the best indicators were found to be the anomalies of Pacific Ocean that are related to ENSO. These findings were similar to the previous findings (McBride and Nicholls 1983) that evidenced the strongest correlations between seasonal rainfall of NSW and ENSO during spring.

Whiting et al. (2003) studied the rainfall in Sydney and demonstrated the existence of a greater correlation of annual rainfall in Sydney with the PDO index than with SOI. A recent attempt (Duc et al. 2017) was made using the Bayesian Model Averaging (BMA) method to analyse the impact of the four major climate drivers on the rainfall of NSW as well as to compare their relative contributions in the forecast model.

A combination of correlation and wavelet-based methods was applied to identify the principal sources of variation in reservoir inflows of Sydney (Westra and Sharma 2008). The study found ENSO, PDO and IOD to be influential and statistically significant correlations (±0.4) were obtained which varied seasonally, although correlations were comparatively lower for spring (Westra and Sharma 2008). The best skills for three months' streamflow forecast were obtained from April to June and October to January for the catchments in Queensland and NSW respectively using dynamic models which are comparable to the outcomes that were obtained using a statistical BJP approach (Robertson and Wang 2009; Wang et al. 2009). Duc et al. (2017) showed in their study that IPO alone does not have any significant impact on the rainfall of NSW, but its combination with ENSO can make a significant impact on rainfall. Trends in extreme rainfall in NSW was investigated by Evan et al. (2017).

It is evident from the study of Kirono et al. (2010) that statistically, significant lag relationships exist between atmospheric, oceanic variables (thermocline, SOI and NINO4) and winter, summer and spring runoff in the northern part of Moree of northern NSW, which is better than the relation with antecedent runoff. Here, again the lag relationships were assessed by using only simple linear correlation where each predictor was considered separately; thus, no combined impact of the predictors was analysed.

The study of Chiew et al. (2003) explained that spring rainfall and runoff had high correlation (0.3 to 0.5) against winter SOI throughout eastern Australia except for NSW (east of Great Dividing Range), although this region showed high correlation for summer rainfall and runoff against spring SOI. Wang et al. (2011) explained that the Predictive Ocean Atmosphere Model for Australia (POAMA) forecast skill shows significant improvements only for monthly forecasts (not for three monthly forecasts), while compared to the historical ensemble. The extensive literature review reveals that climate indices have strong potential as predictors of futuristic streamflow.

1.2 Statement of Problem

To date, most of the research works were focused on the identification of suitable predictor variables for forecasting rainfall and streamflow on daily or monthly scales while very few of those established the seasonal relationship in different parts of Australia. However, a strong concurrent relationship between climate indices and streamflow does not imply that there is also a lagged relationship existed. Thus, the relationship between lagged climate mode and streamflow needs to be investigated individually. Majority of the previous studies investigated the concurrent relation of single climatic variable with daily, monthly or seasonal streamflow. Even though some studies considered the lagged climate modes, none of those included the combined impact of different climate anomalies on streamflow of eastern Australia. Some recent attempts were made by Abbot and Marohasi (2012), Mekanik et al. (2012) and Rasel et al. (2016) to forecast seasonal rainfall of Queensland, Victoria and South Australia respectively, exploiting linear and non-linear statistical techniques and considering the combined influence of different climate modes. Nonetheless, no research work explored the impact of multiple large-scale indices on seasonal streamflow of NSW. Therefore, further research in this study region is essential to forecast seasonal streamflow of NSW using large-scale climate drivers for the following reasons.

Firstly, it is important to find out the time extent of the multiple climate indices till which they are influential on streamflow of different seasons and different regions.

Secondly, NSW being the highest contributor to Australia's agricultural production, it is of utmost importance to the water stakeholders and irrigators to be served with reliable streamflow forecast. NSW, located in the south-eastern part of Australia, is stricken by frequent droughts, especially in the western and north-eastern region of the state. Agricultural production is the major part of the economy of NSW contributes about AU$15 billion annually. Climate variability has a severe impact on the yield of planted crops like wheat, rice etc. Although efforts have been made to forecast streamflow and rainfall of this region, none of the current practices provides reliable seasonal streamflow forecast, which can enable the water users to take risk-free managing decisions at the early stage of the crop period (Khan et al. 2005).

The study intends to provide deterministic forecast as it can play more important roles in solving water management problems by enabling the water stakeholders to take more accurate decisions knowing the predicted amount of future streamflow, compared to the probabilistic approaches which have been attempted by many researchers till date (Piechota et al. 1998; Ruiz et al. 2007; Robertson and Wang 2009; Wang and Robertson 2011; Duc et al. 2017). Communication of the concept of the probabilistic forecast remains a challenge, whereas end-user confidence is essential for the adaptation of a forecast model for decision making.

Furthermore, the performance of existing seasonal streamflow forecast models is not very satisfactory while compared to the performances of the daily or monthly streamflow forecast models. In addition, though many hydrologists established the existence of strong correlations between streamflow and large-scale climate drivers, the nature of the relationship remained a difficult question to deal with. Thus, the uncertainty of the underlying relationships between atmospheric variables and streamflow and the complexity of the atmospheric processes make it difficult to develop a successful streamflow forecast model with reliable accuracy. Hence it is required to deploy new data-driven techniques like Gene Expression Programming (GEP) to capture the complex relationships between streamflow and climate variables which may result in developing reliable streamflow predictor models with enough accuracy.

1.3 Aims and Objectives

The main aim of this research study is to develop a successful seasonal streamflow predictor model using large-scale climate drivers. To serve the purpose of the study basic linear (MLR) and non-linear(MNLR) techniques have been applied at the first stages which were followed by the main method of the study Gene Expression Programming (GEP). To the best of author's knowledge, this is the very first endeavor to forecast seasonal streamflow applying GEP in Australia in conjunction with the use of lagged climate indices.

- To investigate the relative relation of concurrent and lagged single climate variables with seasonal streamflow of NSW.
- To identify the combination of influential multiple climate indices that can be used for developing MLR models.
- To compare the influences of single lagged indices with that of the combined multiple lagged indices.
- To identify the best predictor variables for each region and the maximum significant lagged months, which can be implemented for reliable seasonal streamflow forecast of NSW.
- To explore the non-linear relationship between seasonal streamflow and climate indices by developing Multiple Non-Linear Regression (MNLR) models.
- To use an advanced Artificial Intelligence (AI) based data-driven technique ((GEP) to capture the underlying sophisticated relationship of climate variables and seasonal streamflow.

- To compare the outputs of linear models with that of non-linear models and suggest the best possible modelling approach that can be exploited to forecast seasonal streamflow of NSW.
- To compare the model outputs with the previous research outcomes to ensure the validity of each developed model.

1.4 Research Scopes

In the current research study, the whole NSW is divided into four distinct regions, while in total, twelve streamflow stations are selected. For each of these twelve stations, monthly streamflow and climate indices data are collected which undergo a pre-processing based on the requirement of the applied techniques. Climate anomalies which are influential on the streamflow of selected four regions are identified through concurrent and lagged correlation analyses. Thus, the spatial variation of influential climate indices on streamflow of different regions is explored.

Single concurrent and lagged correlation analyses are carried out to select the combination sets of multiple climate indices which were later on used as the input data sets for the MLR, MNLR and GEP models. For each of these techniques, calibration and validation tests are performed. After completing the analysis of each technique, the best model for each region is selected. The outcomes of different techniques are compared, and the best streamflow forecasting technique is proposed. Performances of all the techniques are compared with the outcomes of previous research studies in this field.

1.5 Significance of the Research

The present research work is expected to have a great extent of the impact on seasonal streamflow forecasting by replacing the shortcomings of the previous researches on this topic with the successfully developed models.

- The current work will provide a reliable streamflow forecast model which will enable the water users to make low-risk plans for the droughts and high flood seasons.

- Different models will be developed for different regions of NSW; thereby it will provide a forecast model which can be reliable irrespective of spatial variation.
- A large number of climate variables with different lagged months will be investigated to identify the best predictors for each region. Therefore, the developed models will be able to suggest the earliest possible time for successful streamflow forecasting, which eventually will help the irrigators and water planners to take their timely decisions to ensure proper water allocation.

1.6 Outline of the Thesis

The current research has the following outline:

- A detailed literature review is stated and discussed in Chapter 2.
- The study area and the details of used data are discussed in Chapter 3, along with the model verification processes.
- Chapter 4 analyzed the relationship between streamflow and climate indices.
- In Chapter 5, MLR modelling methodology and results are discussed in detail.
- MNLR and GEP methods and outcomes of the models are discussed in Chapter 6 and Chapter 7, respectively.
- Performances of all the applied modelling techniques are compared in Chapter 8, which includes a comparison with the previous research works in this field as well.
- The report is concluded with Chapter 9, which provides the summary and conclusion of the study along with recommendations for future work in this research area.
- The list of references and the appendix are also added at the end of this thesis.

Chapter 2
Literature Review

2.1 Background

Australia is believed to receive the lowest precipitation and runoff comparing to all other continents. On an average 85 percent of the rainwater that falls in Australia, gets evaporated or is directly used by the trees and plants or end up in lakes, wetlands or the ocean. Only less than one-fifth of the total rainfall over Australia ends up in its rivers. Though Australia is considered as the driest continent, there is no short of water. However, the problem lies in the location of available water in connection with the large population centres, i.e. where maximum water demand exists. The pressure has been highly increased on water resource availability since the recent eastern Australian severe drought (2002-2007), (Murphy and Timbal 2008). According to Nicholls (2006) over the latter half of the 20th century, a huge decline in rainfall has been observed in eastern Australia. This has been a matter of great concern to the scientist and researchers considering the large population and economic importance of this region.

The land in Australia is mostly arid and desert. The climate here varies from the tropical climate in the north, arid and dry interior and the southern part has the fringe of the Southern Ocean mid-latitude storm track (Risbey 2011). Since, surrounded by three oceans Pacific Ocean, Indian Ocean and the Southern Ocean, the climatic phenomenon in Australia is dependent on climatic drivers in these oceans. Rainfall variability, either inter-seasonal or inter-annual, is influenced by El Nino Southern Oscillation (ENSO) phenomenon in the Pacific Ocean, Indian Ocean Dipole (IOD) phenomenon in the Indian Ocean, Madden – Julian Oscillation (MJO) in the northern regions and Southern Annular Mode (SAM) on the southern regions. SAM has been studied for the variations in the storm track, atmospheric blocking and the subtropical ridge. (Risbey et al., 2009).

The extreme inter-annual variability of rainfall and streamflow make it very hard for the water management authorities to manage and ensure an adequate amount of supply of water to the water users. Highly inconsistent rainfall pattern results in various extreme events such as flash floods, bush fires, droughts etc. (Gallant et al. 2017; Steffen 2015). The variability of rainfall and flow in conjunction with climate variability may happen on different time scales, from annual to multi-decadal and maybe even longer (Verdon-Kidd and Kiem 2009b). This variation of global climate is expected to be related to the variation of Sea Surface Temperatures (SSTs) and Sea Level Pressures (SLPs) of the oceans all over the world. Various climate drivers are generated from the oceans due to the fluctuation of SST and SLP. Thus, these climate drivers play a predominant role in the variation of rainfall and streamflow availability across the world. Among the most influential climate indices on Australia's streamflow and rainfall ENSO, IOD, IPO and SAM are included along with some other indices.

The two main sources of seasonal streamflow forecast are initial catchment and future climate conditions (Robertson et al. 2009). Initial catchment conditions can be indicated by antecedent streamflow, antecedent rainfall, soil moisture or groundwater levels. Chiew et al. (1998) stated that serial correlations (persistence) is very high in Australia and the reason for this, is the delaying responses of the rainfall-runoff process which ultimately gives streamflow data memory of several months.

As rainfall is the source of soil moisture and streamflow, there is a strong relation between rainfall and streamflow Therefore, it can be hypothesized that the climate drivers that have an influence on rainfall, may have impacts on streamflow as well, thus can play a vital role for streamflow forecasting. So, previous research studies focused on rainfall forecasting must take into account while forecasting streamflow in order to find out the most influencing climate variables. Nevertheless, rainfall events end in runoff into the rivers only after satisfying the requirements of the hydrological processes, e.g. evapotranspiration, infiltration, interception, surface storage etc.

Dutta et al. (2006) indicated the necessity for exploring the skills of forecasting streamflow and rainfall with different lead times exploiting various climate indicators. He also mentioned that streamflow forecast is more significant compared to rainfall forecast as it can be forecasted with longer lead times. Thereby, streamflow forecast enables the water users to make the decision earlier in the year, which ultimately increases the potential of financial benefits.

A successful streamflow predictor model with sufficient accuracy is of great importance to the water stakeholders, which will enable them to take low-risk decisions at the early stage of the crop periods. Significant concurrent and lagged correlations between large-scale climate drivers and seasonal rainfall and streamflow have been found in many studies (Chiew et al. 1998; Drosdowsky and Chambers 2001). Robertson et al. (2009)

explored that climate indices can improve streamflow predictions throughout the year with the greatest improvement.

In the following sections, a brief overview of the past studies on rainfall and streamflow have been discussed considering rainfall being the major source of streamflow in Australia. Then a review of influential climate indices has been presented. Later on, a discussion on the influential climate variables on streamflow of different parts of the world and specific dominant climate indices on Australian streamflow have been carried out. Finally, a critical review of the available forecasting technique is discussed and also the research gap is explored, which is followed by stating a critical summary of this literature review to carry on the detailed analysis.

2.1.1 Short-term streamflow forecasting

After experiencing the severe drought in the past decade and the recent extreme climatic events in Australia, it is now more important than ever to develop more accurate streamflow forecast model. Though long-term streamflow forecasting has got more significance to the water-stakeholders compared to short-term forecasting due to its long lead time which gives sufficient time to take preparation for the extreme events, short-term forecasting is also important, especially in case of the events that have very short lead-time available for the warning system. For instance, a flash flood is such an event having a very short lead time. The flash flood that occurred in June 1990 at Shadyside in Ohio, USA cased huge damages and fatalities which resulted from the onset rainfall within 1 hour (Einfalt et al. 1990; Seo and Smith 1992). Therefore, to forecast events like a flash flood which has very short lead time, short- term forecasting is very important (Seo and Smith 1992; Einfalt et al. 1990; Huff and Vogel 1981; Saffle and Greene 1978).

Researchers have been studying short-term streamflow forecasting (up to 10 days ahead) by using many means including rainfall forecasting and many other modern methods with a view to improving the forecasting skill by quantifying forecast uncertainty. It is the biggest challenge for the hydrologists and water researchers to find out a way to meet the ever-increasing water demand. To solve this problem, the Bureau of Meteorology (BoM) and Commonwealth Scientific and Industrial Research Organization (CSIRO) have come together to develop a new system of Short-term Water information Forecasting Tools (SWIFT), (Short-term Water Forecasting and

Prediction 2012). Previously the hydrological model used by BoM was proving event-based forecast whereas the new tool (SWIFT) is capable of providing a continuous forecast.

A new approach, wavelet-genetic programming was attempted by Karimi et al. (2016) to forecast short-term and long-term streamflow where this new method outperformed all other methods which included autoregressive moving average (ARMA) method, neuro-fuzzy system and artificial neural network. Artificial Neural Network (ANN) was applied in the Winnipeg River system in Northwest Ontario, Canada for making short-term streamflow prediction using quarter monthly time intervals while ANN provided better performance than a conventional model (Zealand et al. 1999). Bureau of meteorology provides short-term streamflow forecast with a lead time up to 7 days applying a deterministic method which uses the rainfall-runoff model called GR4H as the core component of it (BoM 2000). This rainfall-runoff model uses the estimated rainfall, model parameters and catchment condition to generate the streamflow forecast on a daily basis; however, the model is unable to provide any likelihood of the event to occur.

2.1.2 Long-term streamflow forecasting

Long-term rainfall and streamflow forecasting have more significance to the water stakeholders compared to short-term forecasting as it allows them to take considerable time to make proper water management plans and make low-risk decisions. According to the Bureau of Meteorology (BoM, 2009), long-term forecasting plays a very important role while taking preventive measures or making plans for extreme events like droughts, bushfires or high flood events. Thus, over the years, long-term rainfall and streamflow forecasting has got more attention to the researchers and hydrologists. The forecast period for long term forecasting may vary from one month to several months ahead, or even it may become a seasonal forecast if consists of the months of the corresponding season.

So many research studies had been carried out with a view to making successful rainfall and streamflow forecasting with longer lead time using either probabilistic or categorical forecast method (Duc et al. 2017; Wang and Robertson 2011; Risbey et al. 2009, Robertson and Wang 2009; Piechota et al. 1998). However, for solving water resources management problems, deterministic forecasts are of more importance than probabilistic forecasts which have been attempted by many researchers till date (Piechota et al. 1998; Ruiz et al. 2007; Robertson and Wang 2009; Wang and Robertson 2011; Duc et al. 2017).

Furthermore, the Bayesian Joint Probability (BJP) method used by the Australian Bureau of Meteorology (http://www.bom.gov.au/water/ssf/index.shtml) to provide futuristic streamflow is again a probabilistic method. Bureau of Meteorology provides seasonal streamflow forecast based on statistical (BJP) and dynamic models. While the BJP models use the climate indicators, initial catchment conditions and historical streamflow and rainfall data to generate out, the dynamic model uses climate predictions from global climate models to hydrological models. Though these forecast models developed by BoM play an important role to water allocations, cropping and water market strategies, operating a diversified water supply scheme and overall water management, the forecast skill contains uncertainties due to a range of factors. In addition, communication of the concept of the probabilistic forecast remains a challenge, whereas end-user confidence is very important for the adaptation of a forecast model for decision making.

2.1.3 Climate indices

2.1.3.1 El Nino Southern Oscillation (ENSO)

El Nino Southern Oscillation (ENSO) which develops off the western coast of South America is a band of sea surface temperature when unexpectedly warm or cold temperature exists for a long period of time. Southern Oscillation refers to the fluctuation of the sea surface temperature of the tropical eastern Pacific Ocean and fluctuation of air surface pressure in the western Pacific Ocean (Bamston et al. 1997). El Nino and La Nina refer to the warm and cool phase of tropical eastern Pacific Ocean, respectively. The ocean-atmosphericashok variation is known as the El Nino Southern Oscillation. ENSO is considered to be responsible for extreme weather conditions such as floods and droughts that are happening in many regions of the world.

ENSO phenomenon has two components- sea surface temperature and atmospheric pressure, which are intensely correlated and can be represented by two types of indicators, the SLP indicator and the SST indicator (Duc et al. 2017).

SLP indicator refers to the bimodal variation in the sea level barometric pressure, and it is referred as the Troup SOI, which measures the difference between sea-level atmospheric pressures at Papeete (Tahiti) and Darwin (Troup 1965). SLP consists

of two phases, the warm SST anomalies, El Nino where the surface pressure is high and the cold SST anomalies, La Nina where the surface pressure is low in the equatorial Pacific Ocean. BoM provides the following formula to calculate SOI:

$$SOI = 10 \times \frac{P_{diff} - P_{diff(av)}}{SD_{P_{diff}}} \qquad (2\text{-}1)$$

Where,

P_{diff}=(average Tahiti mean SLP for the month) –(average Darwin mean SLP for the month)

P_{diff} = long term average of P_{diff} for the month in question

$SD_{P_{diff}}$ = long term standard deviation of P_{diff} for the studied month

Various SST anomalies are also available which have been derived using different areas of the equatorial Pacific Ocean (Kiem and Franks 2001). Generally, the SST anomalies are monitored in 3 geographic regions of the equatorial Pacific and defined as NINO3 (5°S – 5°N, 150°– 90°W), NINO3.4 (5°S – 5°N, 170° – 120°W) and NINO4 (5°S – 5°N, 160° – 150°W) (Risbey et al. 2009). Hanley et al. (2003) compared the response of pressure based and SST based anomalies to the ENSO extreme event and found that NINO3.4 and Niño-4 indices are equally sensitive to El Niño events whereas SOI is less sensitive to La Niña events than others are. The reason for originating ENSO warm or phase is yet to be discovered. However, it is explored that ENSO events have impacts on the climate of the areas outside the tropical Pacific regions. EL Nino phase indicates the SST anomalies which are greater than or equal to 0.5^0C existing in the Nino 3.4 region and partially in Nino 3 and Nino 4 regions. On the other hand, La Nina Phase has anomalies which are less than or equal to -0.5^0C.

Thermocline, which is one of the important drivers of ENSO, is defined as the regions which separates the warm and well-mixed surface water from cool and deep ocean water. Water temperature below thermocline is considered as less than or equal to 15^0C, whereas the temperature above thermocline is 25^0C. Atmospheric convection is the process of rising the warm air generated by the warm sea surface temperature of the western Pacific, causing heat and moisture in the atmosphere. The drier air moves to the east before falling over the cooler eastern tropical Pacific. This phenomenon of air rising in the west and falling over the east is known as Walker Circulation.

ENSO phenomenon has got three distinct phases: the two opposite phases, "El Niño" and "La Niña," which resulted from certain changes both in the ocean and in the atmosphere and thus ENSO is called a coupled phenomenon; and the "Neutral" phase which is in the middle of the continuum.

El Niño phase occurs when ocean surface has a temperature above-average SST, specifically in the central and eastern tropical Pacific Ocean which is associated with the deepening of thermocline in this region (Figure 1). The trade winds which usually blow from east to west ("easterly winds"), weakens or in some cases start to blow toward the reverse direction (from west to east) and become "westerly winds". This causes reduced rainfall over Indonesia and increased rainfall over the tropical Pacific Ocean. During this period sea surface temperature around, northern Australia remains cooler than the normal phase, and convection moves away Australia eastward towards the central tropical Pacific Ocean resulting in reduced rainfall over Australia. Eastern Australia is mostly affected by this event, while the impacts on southwest western Australia and coastal

NSW may vary from event to event, and western Tasmania is less affected.

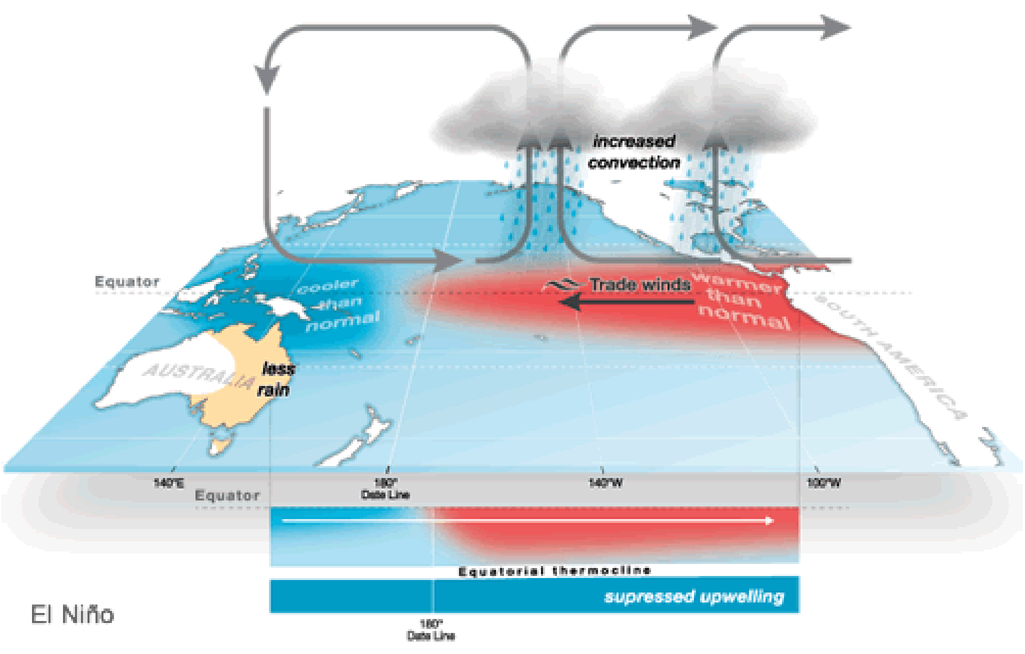


Figure 1. El Niño Phase of ENSO Phenomenon (Source: Record Breaking La Niña Events, BoM, 2012)

On the contrary, La Niña occurs when the ocean surface has below-average temperature, especially in the central and eastern tropical ocean, which resulted from the movement of thermocline towards the surface (Figure 2). The normal easterly winds become even stronger; thus, the warmer water gets confined to the western tropical Pacific. This results in increased rainfall tend over Indonesia while regions of the central tropical Pacific Ocean experience reduced rainfall. Regions around the north of Australia has warmer SST, and the water circulation also intensifies. During this phase, northern and eastern Australia experience increased rainfall and increase humidity s observed in the inland of Australia. La Niña is more influential than El Niño on some parts of northern and central Australia.

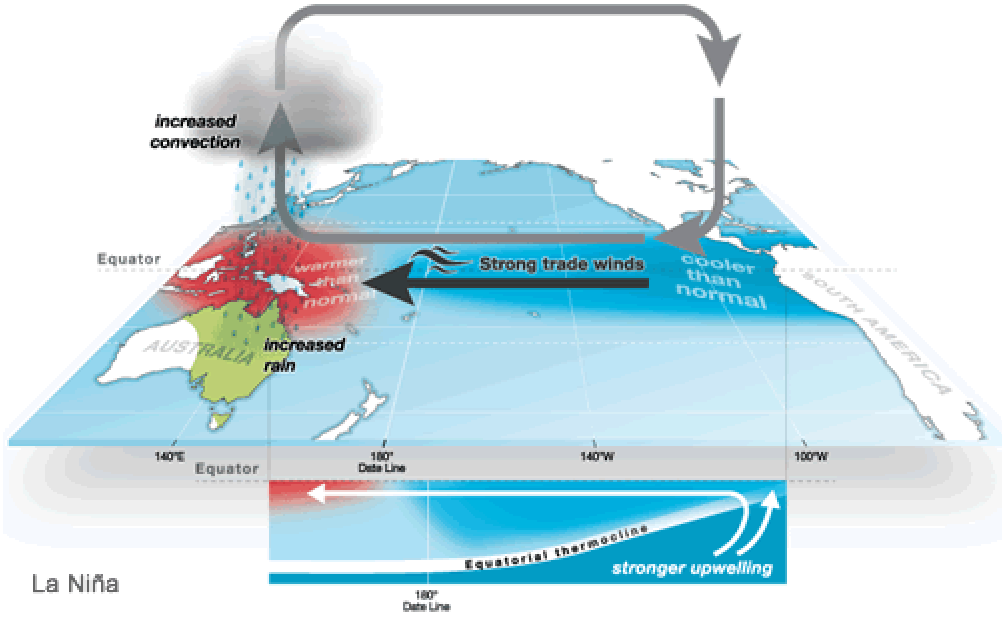


Figure 2. La Niña Phase of ENSO Phenomenon (Source: Record Breaking La Niña Events, BoM, 2012)

During the Neutral phase, the SST of the tropical Pacific is close to the average (Figure 3). The trade winds blow from east to west which brings the warm moist air warmer surface water towards the western Pacific Ocean. Sometimes the ocean looks like to be in El Niño or La Niña state, but the atmosphere does not behave accordingly.

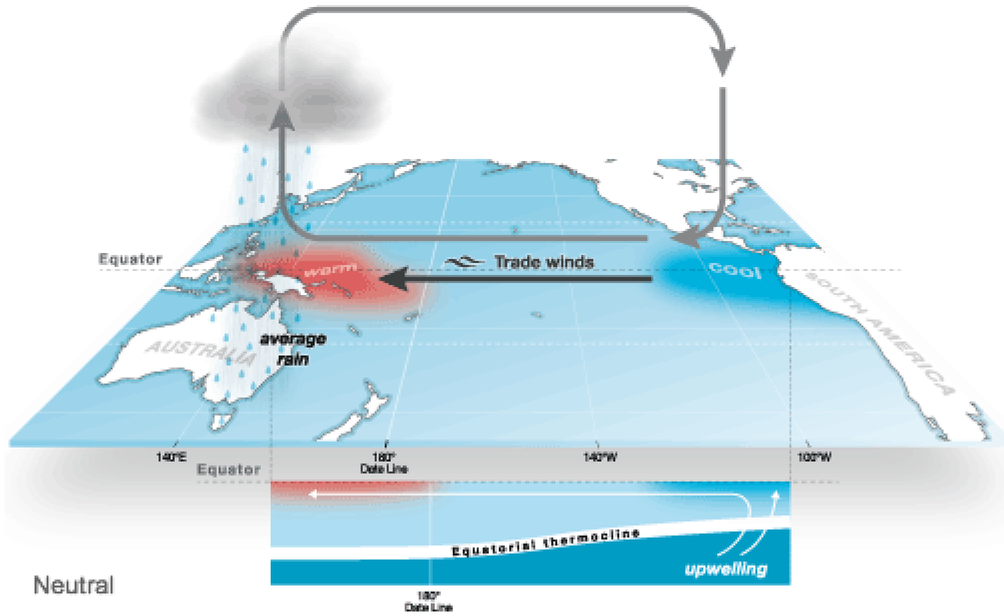


Figure 3. Neutral Phase of ENSO Phenomenon (Source: Record Breaking La Niña Events, BoM, 2012)

The El Nin˜o Modoki is an ocean-atmosphere coupled process (Figure 4), which results in unique tripolar sea level pressure pattern during the evolution, similar to the Southern Oscillation phenomenon of El Nino (Ashok et al. 2007). Therefore, this phenomenon is named as El Nin˜o–Southern Oscillation (ENSO) Modoki and expressed by the following equation

$$EMI = SSTX - (0.5*SSTY) - (0.5*SSTZ) \quad (2\text{-}2)$$

Where, X=165°E–140°W, 10°S–10°N, Y= 110°W–70°W, 15°S–5°N, Z=125°E–145°E, 10°S–20°N (Ashok et al. 2007).

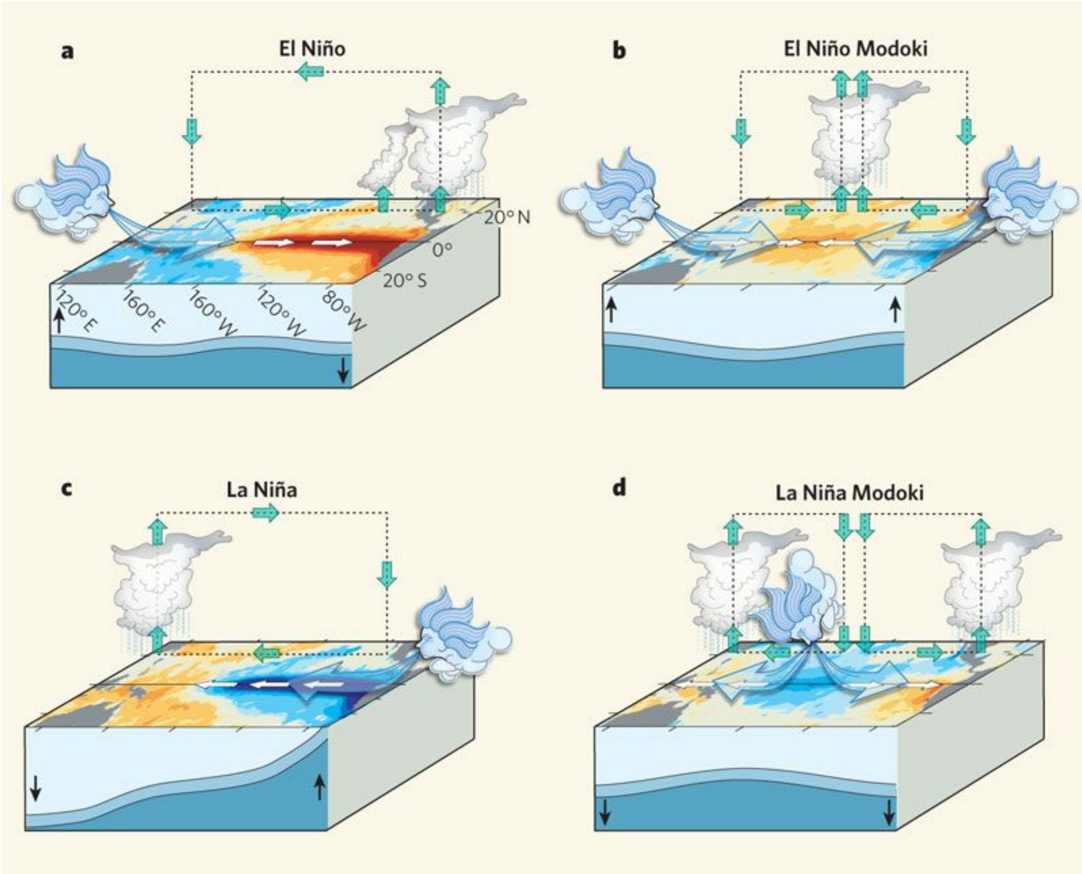


Figure 4. (a) Usual El Niño event; (b) Usual El Niño Modoki Event; (c) Opposite phase(La Niña) of El Niño; (d) Opposite phase (La Niña Modoki) of El Niño Modoki (Source: Ashok and Yamagata 2009)

The ENSO Modoki events have significant influences on the climate of many parts of the world including Japan, New Zealand, western coast of United States (Ashok et al. 2007), Australia (Taschetto and England 2009), South China (Feng and Li 2011).

2.1.3.2 Indian Ocean Dipole (IOD)

The IOD represents the coupled oceanic-atmospheric variability in the tropical Indian Ocean which is classified by SST anomalies of reverse sign in the east and west (Saji et al. 1999; Webster et al. 1999). It is the difference in SST between two poles where a western pole is in the western Indian Ocean (Arabian Sea), and an eastern pole is in the eastern Indian Ocean south of Indonesia. The Dipole Mode Index (DMI) which is a measure of the IOD is defined as the difference in SST anomaly between the tropical western Indian Ocean (10ºS–10ºN, 50º–70ºE) the tropical south-eastern Indian Ocean (10ºS–equator, 90º–110ºE).

The IOD is considered to be a similar phenomenon to El Niño but occurs in the equatorial Indian Ocean and thus often called "Indian Niño". Some researchers (Saji et al. 1999; Ashok et al. 2003; Meyers et al. 2007) have arguments regarding the independent extent of IOD from ENSO because of the extension of the Walker circulation to the west and associated Indonesian throughflow (the flow of warm tropical ocean water from the Pacific into the Indian Ocean). Meyers et al. (2007) developed an index of IOD using a lagged empirical orthogonal function (EOF) approach considering the variation in ENSO in defining IOD.

Like ENSO, IOD has three different phases: positive (Figure 5), "negative" (Figure 6) and "neutral" (Figure 7). A positive IOD happens when the waters near the Horn of Africa are warmer than average leading to enhanced rainfall there, while cooler waters develop off Indonesia resulting in less rainfall and high temperatures in Australia. Thus, a positive IOD refers to a wetter west and drier east. On the contrary, negative IOD is a phase when sea surface temperatures in the western Indian Ocean is cooler compared to the east, and the trade winds become more westerly. It causes more rainfall in southern Australia. It is believed that the positive IOD is related to the El Niño phase while the negative IOD is associated with the La Niña phase. The impacts of El Niño and La Niña are most extreme when IOD and ENSO are in phase, whereas the impacts of El Niño and La Niña are weakened when they are out of phase.

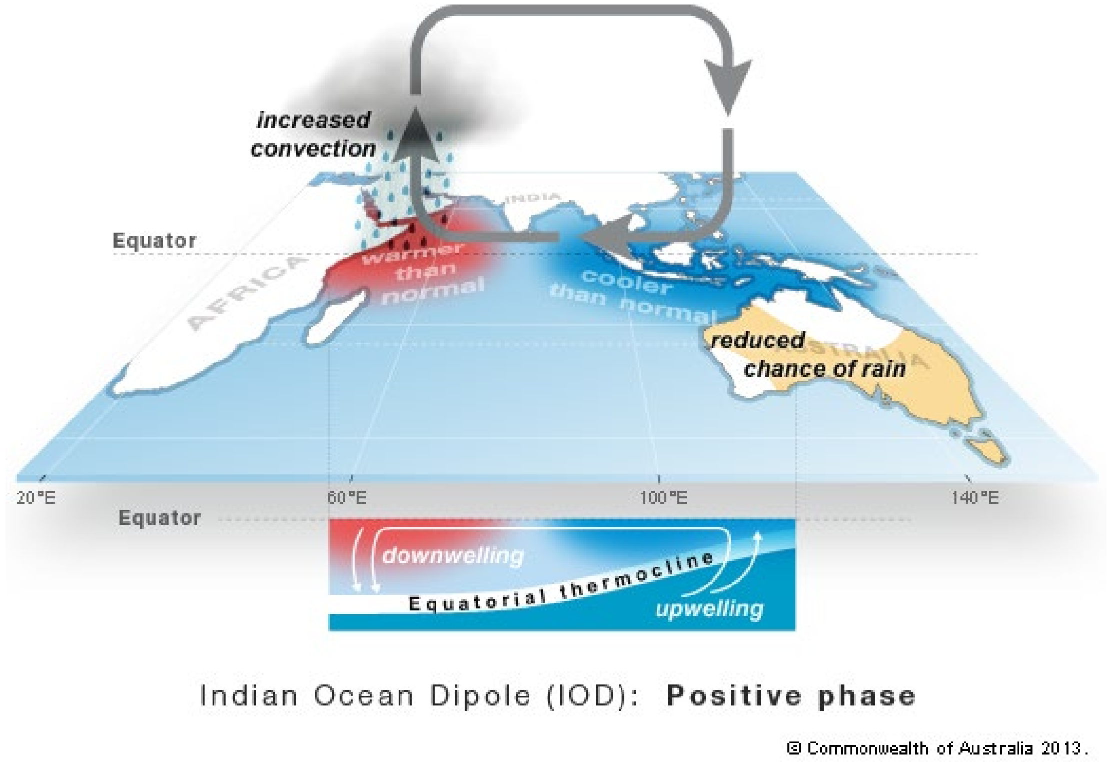


Figure 5. Positive Phase of IOD (Source: Bureau of Meteorology)

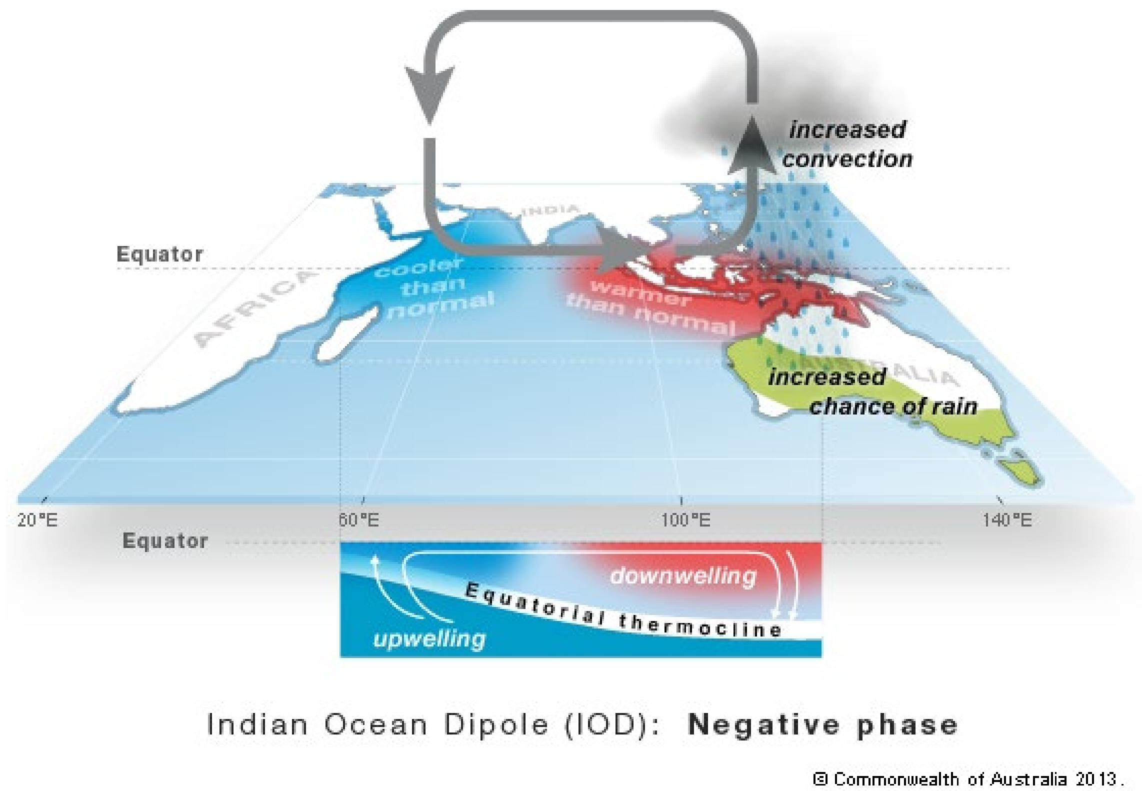


Figure 6 Negative Phase of IOD (Source: Bureau of Meteorology)

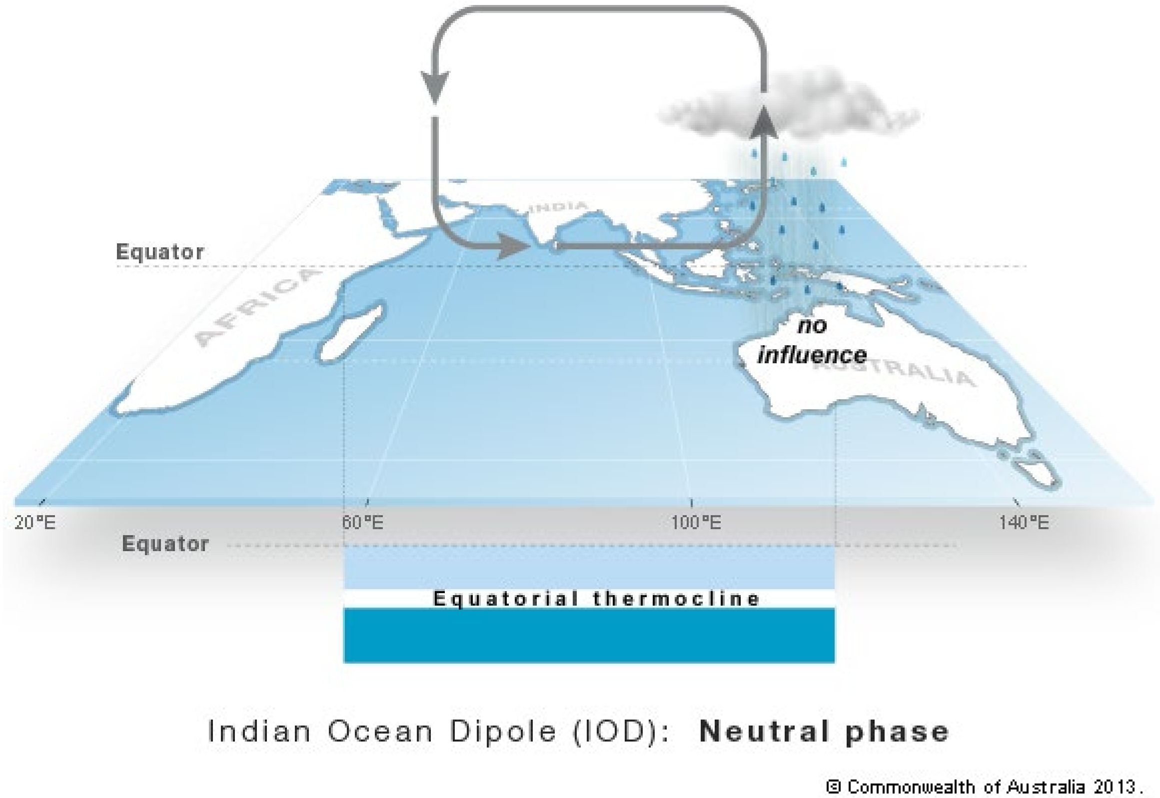


Figure 7. Neutral Phase of IOD (Source: Bureau of Meteorology)

Yamagata et al. (2004) has summarized the dynamics of IOD, which provides a better understanding of the mechanism. According to the researcher, IOD is developed by ocean-atmospheric interactions, which includes the influences of the thermocline. Though studies on the depth of thermocline are very limited, the existing researches suggested that thermocline has huge vertical displacements beneath both poles of the dipole which is correlated to the local SST anomalies (Meyers 1996; Rao et al. 2002; Feng and Meyers 2003). It was also revealed that the depth of thermocline is driven by the remote winds from Indian and Pacific Oceans (Wijffels and Meyers 2004).

2.1.3.3 Interdecadal Pacific Oscillation (IPO)

The term "Interdecadal Pacific Oscillation (IPO)" was first introduced by Power et al. (1999) while he was explaining a trans-Pacific Sea Surface Temperature Anomaly (SSTA) that signifies a Pacific-wide indication of the phenomenon, referred as PDO by Mantua et al. (1997). IPO causes a relatively rapid transformation in the SSTA pattern across the Pacific Ocean, which remains active for two to three decades. Unlike the predominant influence of PDO in the northern Pacific region, IPO creates a link to the Indian Ocean region and ocean-atmospheric coupling by a Pacific basin-wide, bi-hemispheric climatic pattern. The spatial pattern of a PDO variability is characterized by a 'horseshoe' shape. Determined by positive (warm) and negative (cold) phases, during a positive phase higher than average SST is observed from the western coast of North America down to the equator, forming a horseshoe shape, surrounding cooler water in central and north-western Pacific. Whereas in the negative phase, this pattern is reversed (As shown in Figure 8). The warmer water in the horseshoe shape region become cooler than average, and the central and the north-western Pacific have higher than average SSTs (Zhang and Delworth, 2015).

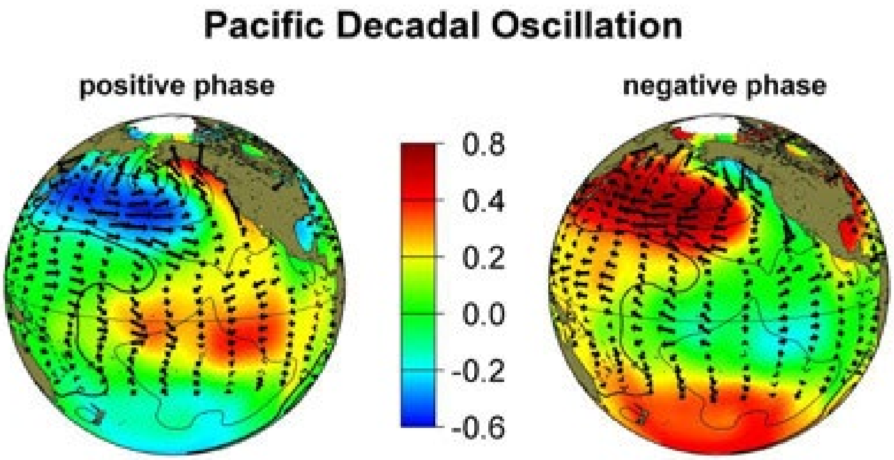


Figure 8. Typical wintertime sea surface temperature (colors), sea level pressure (contours), and surface wind stress (arrows) anomaly patterns during positive and negative phases of the Pacific Decadal Oscillation (PDO). Temperature anomalies (colors) are in degrees Celsius.
(Source: Hare and Mantua, University of Washington)

The IPO is described as the Pacific ENSO-like pattern of SST, which is found in the analysis of near-global inter-decadal SST (Folland et al. 1999). IPO has a cycle of 15-30 years and characterized by two phases, namely positive and negative (Salinger et al. 2001; Henley et al. 2015). While IPO is defined for the whole Pacific Basin, PDO is defined for the North Pacific, poleward of 20°N. IPO and PDO are found to be highly correlated while their phase changes are directly related to the increased and decreased frequency of warm and cool phases of ENSO, respectively. Negative and positive phases of IPO/PDO creates SSTA like La Niña and El Niño phases over the tropical Pacific, respectively. During the positive phase of IPO/PDO increased rainfall is observed in the northeast of the South Pacific Convergence Zone (SPCZ) while the southwest of SPCZ experiences reduced rainfall than normal. Mean Sea Level Pressure remains higher than normal to the west of the dateline while it is lower than normal to the east of the dateline. Southerly flow anomaly is generated due to these pressure difference created during the positive phase of IPO/PDO.

IPO's influences have been observed from Pacific to the Indian Ocean, southern Asia and Madagascar (D'Arrigo et al. 2006; Crueger et al. 2009; D'Arrigo and Ummenhofer 2015), to Australia (Power et al. 1999; Arblaster et al. 2002), and even Antarctica (Palmer et al. 2015; Vance et al. 2015; Meehl et al. 2016b). Mantua et al. (1997) found IPO/PDO to be responsible for the multi-decadal step changes in climate. According to Mantua and Hare (2002), the two distinct features that make IPO/PDO different from ENSO are the persistence range of IPO/PDO (15-30 years) and dominant climatic influence of PDO over the north-Pacific region.

2.2 Impacts of Large Scale Climate Drivers on Streamflow all over the World

As rainfall is the primary source of streamflow in Australia, future rainfall and climate conditions can be exploited to estimate the magnitude of future streamflows. According to Fawcett and Stone (2010), there exists relatively low skill for operational predictions of seasonal rainfall in many parts of Australia; thereby, it may give little additional information about future streamflows.

It has been accepted by the hydrologists that there exists a strong correlation between the streamflow and the large-scale atmospheric circulation patterns. Again, large-scale atmospheric circulation patterns are influenced by an ocean-atmospheric phenomenon, mostly ENSO.

The ENSO phenomenon, which results from the large-scale interactions between the ocean and atmospheric circulation processes in the equatorial Pacific Ocean, has direct influences on the climate variability over many parts of the world (e.g. Ropelewski and Halpert 1987; Kiladis and Diaz 1989). The ENSO phenomenon has three phases- El

Nino (warm), La Nina (cold) and Neutral. The warm oceanic phase (El Nino) is associated with the high air surface pressure in the western Pacific, and the cold oceanic phase (La Nina) is associated with the low air surface pressure in the western Pacific (Verdon et al. 2004). ENSO's warm phase (El Nino) conditions refer to SST anomalies equal to or greater than 0.5°C in the Nino 3.4 region including portions of Nino 3 and Nino 4 regions whereas cool phase (La Nina) conditions are related to anomalies less than or equal to –0.5°C.

The teleconnection between ENSO and climate is used for long lead weather forecasting in many countries all over the world (Chiew et al. 2003). El Nino and La Nina events are responsible for the different climatic conditions around the Pacific including eastern Australia (Stone and Auliciems 1992; Nazemosadat and Cordery 1997; CPTEC 2006; Chiew 2006; Hoerling et al. 2001). Fan et al. (2000) found the predictability to be related to geographical location. Generally, the regions within 30°S-30°N band are considered as highly influenced by ENSO and thus can be predicted more precisely (Frederiksen et al. 2001). It may cause flooding on the west coast of equatorial South America resulting from the heavy precipitation across the ocean (NOAA-2014). El Nino conditions are strongly related to the reduction in the pressure difference, which causes severe drought in parts of the western Pacific, such as Australia.

The IOD influences the climate of the tropical Indian Ocean such as East Africa, India and Indonesia (Webster et al. 1999; Saji and Yamagata, 2003; Yamagata et al. 2004). The western and southern part of Australia is influenced by IOD (Ashok et al. 2003). IOD events can be predicted several months ahead, and this predictability can be exploited to improve water and agricultural decision-making. According to Meyers et al. (2007), IOD and ENSO can sometimes occur together in such a way that strengthens each other.

Many researchers (e.g., Power et al., 1999; Kiem et al., 2003; Kiem and Franks, 2004) have demonstrated the influence of the IPO to be significant on rainfall variation on a decadal to multi-decadal timescale. The IPO is described as the Pacific ENSO-like pattern of SST, which is found in the analysis of near-global inter-decadal SST (Folland et al. 1999). The IPO has a cycle of 15-30 years and characterized by two phases, namely, positive and negative phases (Henley et al. 2015; Salinger et al. 2001). The

IPO is defined for the whole Pacific Basin while PDO is defined for the North Pacific, poleward of 20°N. IPO has some independent impacts on the South Pacific Convergence Zone compared to ENSO. Climate patterns around the world are also influenced by the PDO and IPO (e.g. Kiem et al. 2003; Verdon et al. 2004). From at least the 15th century, the IPO/PDO has been known as a dominant climate mode in the Pacific sector. Therefore, the climate in the future is likely to continue to be influenced by the IPO/PDO (Kiem et al. 2009). IPO/PDO phenomena are linked to the annual fluctuations in maximum temperature, reported water volume, wheat production and overall climate variability in Australia. According to Mantua and Hare (2002), the two features that differentiate PDO/IPO from ENSO are the persistence period (15-30 year) and dominant region (north pacific sector) of PDO/IPO.

2.3 Impacts of Large-Scale Climate Drivers on Streamflow in Australia

The climate of southeast Australia is influenced by four major climate drivers originating in the Pacific Ocean, the Indian Ocean and the Southern Ocean ENSO, IPO (PDO), SAM and IOD (Duc et al. 2017). ENSO and IPO (or PDO) are atmospheric-ocean phenomena results from the SST and SLP anomalies in the Pacific. Similarly, IOD is the atmospheric-ocean phenomenon in the Indian Ocean. Southern Annular Mode (SAM) which is the change of the anomalies in SST and geopotential height in the South Pacific and the Southern Ocean, has an influence on climate processes. ENSO has been considered to be the most dominant climate anomalies for forecasting rainfall and streamflow time series (Ropelewski and Halpert 1987; Chiew et al. 1998; Piechota et al. 1998; Sharma et al. 2000; Sharma 2000; Cai et al. 2001; Drosdowsky and Chambers 2001; Piechota et al. 2001;

Chiew et al. 2003; White et al. 2004; Dutta et al. .2006). Figure 9 shows the influence of different climate variables around Australia.

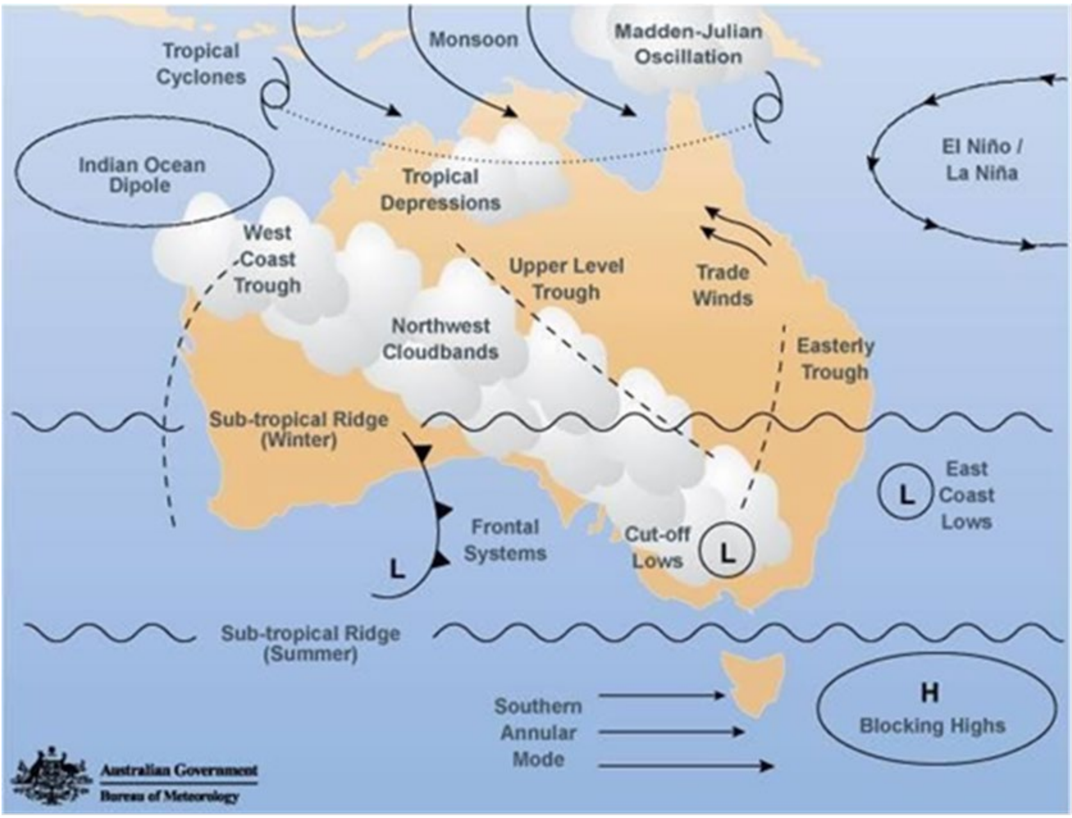


Figure 9. Influence of different climate variables around Australia (Source: Bureau of Meteorology, 2010)

Conditional probability forecast was applied by Simpson et al. (1993) and Allan et al. (1995) using SST anomalies of ENSO in order to forecast annual streamflow at the confluence of River Murray and the Darling Rivers and the natural flows into the Hume Reservoir. They found an opposite relationship between SSTs and the annual flows, which is associated with Australian droughts.

It is revealed from the past studies that the two approaches- dynamic and statistical are widely used in practice for forecasting rainfall and streamflow (Goddard et al. 2001). While predicting ENSO and other climate indices, many researchers have found simple statistical models to outperform sophisticated dynamic models (e.g., Halide and Ridd 2008; Quan et al. 2006). Again, the implementation and operation of dynamic models are more expensive than statistical climate forecasting models (Anderson et al. 1999). On the other hand, the success of the statistical prediction systems relies on not only the availability of long data records but also the stationary relationship between the variable.

Direct use of climate indices to forecast streamflow gives better results than when they are used to forecast rainfall as climate tends to have a stronger relationship with streamflow than rainfall (Wooldridge et al. 2001). Several studies revealed the influences of ENSO on streamflow throughout Australia (Chiew et al. 1998, 2003; Dettinger and Diaz 2000; Dutta et al. 2006; Piechota et al. 1998). Chiew et al. (1998) and Piechota et al. (1998) found that ENSO based (SOI and SST) streamflow predictions in northeast Australia are better than the forecasts from climatology.

Piechota et al. (1998) developed a Seasonal Streamflow Forecast Model (SSFM) that used an optimal linear combination of four statistical models with linear discriminant analysis (LDA). An extension of this research (Piechota et al. 2001) in five Australian catchments showed that when the degree of persistence is less noticeable, SST and SOI could be more useful for streamflow forecasts with longer lead times. The same research also used the ENSO indices (SOI and MEI) and serial correlations of rainfall and streamflow, and the result revealed that lag correlation between rainfall and streamflow versus ENSO indices is statistically significant at $\alpha < 0.01$ in spring and summer in most parts of Australia.

In addition to ENSO, Australian rainfall and streamflow are also influenced by low-frequency variability in the Pacific Ocean, which is referred to as PDO (Mantua et al. 1997; Zhang et al. 1997; Mantua and Hare, 2002). Besides, the modulation impact of IPO, which is a closely related index to PDO, is associated with the Australian rainfall, streamflow and flood risk (Kiem et al., 2003 and Verdon et al. 2004). ENSO-hydroclimate relationship is found to be stronger during the negative phase of IPO than its positive phase (Chiew and Leahy 2003), and thus this relationship can be used to forecast rainfall and streamflow. It was evaluated by using the Linear Error in probability score (LEPS) that the combined model of NINO3– thermocline and NINO3-SST outperformed the skill of the individual models which is more evident in spring and summer with 65% of all stations having increased forecast skill (Jose Eric Ruiz et al. 2007). Strong relationships between the SOI, IPO indices, seasonal rainfall and total streamflow volumes were found in the study of Verdon et al. (2004). Risbey et al. (2009) in their study of the impacts of remote climate drivers on rainfall variability in Australia established the relation between ENSO and rainfall in Australia in weakly modulated by IPO. Researchers (Power et al. 1999; Kiem et al. 2003; Verdon et al. 2004) have found that the decadal and annual-scale fluctuations in maximum temperature, rainfall, water volume transport and wheat crop yield general climate variability in Australia are linked to the IPO/PDO phenomenon. Follan et al. (2002) found IPO/PDO that IPO/PDO influences eastern Australian climate during the austral spring, summer and autumn. The higher rainfall and streamflow occurred across most of eastern Australia during the mid-1940 to mid-1970 was the effect of IPO/PDO (Kiem et al. 2003; Verdon et al. 2004). IPO/PDO also has an influence on the transformation of the magnitude and frequency of ENSO which impact eastern Australian Climate(Power et al. 1999; Kiem et al. 2003; Verdon et al. 2004). The effect of ENSO on Australian rainfall is weakened during IPO/PDO warm phase whereas the effect of ENSO is enhanced on Australian rainfall when the IPO/PDO is in cool phase(Power et al. 1999). It is experienced that the wet events get wetter and more frequent during the cool (negative) phase of IPO/PDO compared to the positive phase of IPO/PDO which results in increased rainfall in the Murray Darling Basin (MDB) of southeast Australia (Kiem et al. 2003; Verdon et al. 2004) whereas the wet events are not that strong when the IPO/PDO is in warm (positive) phase, and this enhances the risk of drought in the MDB regions and many other parts of eastern Australia (Verdon-Kidd and Kiem 2009a). Studies have explored that there was a consistent relationship between ENSO, IPO/PDO and rainfall/streamflow in the northeast Queensland for at least the last 400 years (Lough 2007).

Single site seasonal streamflow forecasting approaches were introduced by Piechota et al. (2001), Sharma (2000), and Chiew and Siridardena (2005) while an extension of a nonparametric approach to forecasting streamflow at multiple sites was carried out by Mehrotra et al. (2006). Westra et al. (2008) proposed an approach for dealing with spatially correlated streamflows by independent component analysis. However, none of these studies includes non- concurrent data analysis.

ENSO Modoki Index (EMI) also has a strong influence on the rainfall of northern and north-western region of Australia. According to Taschetto and England (2009), EMI significantly decreases northern and north-western rainfall while traditional ENSO indices decrease south-eastern and north-eastern rainfall in Australia. Moreover, EMI anomalies mostly influence March to May rainfall. During this period, EMI contributes more to drying the northern Australia region.

Wang and Robertson (2011) applied Bayesian joint probability modelling and identified NINO3.4 to have a stronger impact on forecasting eastern Australian rainfall than SOI, which is contradictory to the findings of Chiew et al. (1998) who proved significant relationship of eastern and central northern Australian rainfall with SOI but could not find any significant correlation with SST anomalies. It is evident from the study of Kirono et al. (2010) that the atmospheric variables (NINO4, thermocline and SOI) having a significant lag relationship for winter, spring and summer runoff are better predictors of runoff than antecedent runoff in the northern parts of southeast Australia. Here the lag relationships were assessed by using only simple linear correlation where each predictor was considered separately. Thus no combined impact of the predictors was analyzed. A Bayesian Joint Probability model was introduced by Robertson et al. (2009) to determine the best predictors for seasonal streamflow forecast and was applied to two catchments of eastern Australia. Assessment with serial correlation, Kruskal-Wallis tests and LEPS skill score proved persistence to be a stronger predictor of streamflow than average SOI where strongest relationships were found in/eastern Australia during late spring and early summer. Cai et al. (2012) demonstrated the devastating southeast Queensland (SEQ) flood and the associated extreme rainfall in January 2011 to be the effect of the transition to a negative phase of the PDO-IPO. King et al. (2013) suggested that the IPO plays a significant role in the frequency of major floods during the 1950s, 1970s and 2010–2011.

An attempt (Wang et al. 2011) was made to explore the skills of two conceptual rainfall-runoff models (MWB model and SIMHYD model) to forecast streamflow at monthly and three monthly scales for two catchments (North Queensland and Murrumbidgee region NSW) in East Australia using simulated catchment initial condition, historical rainfall and SOI. The performance of the models demonstrated that in southeast Australia, SOI is a better predictor for July-September and October-December streamflow while SST is a better predictor of January-March and April-June streamflow.

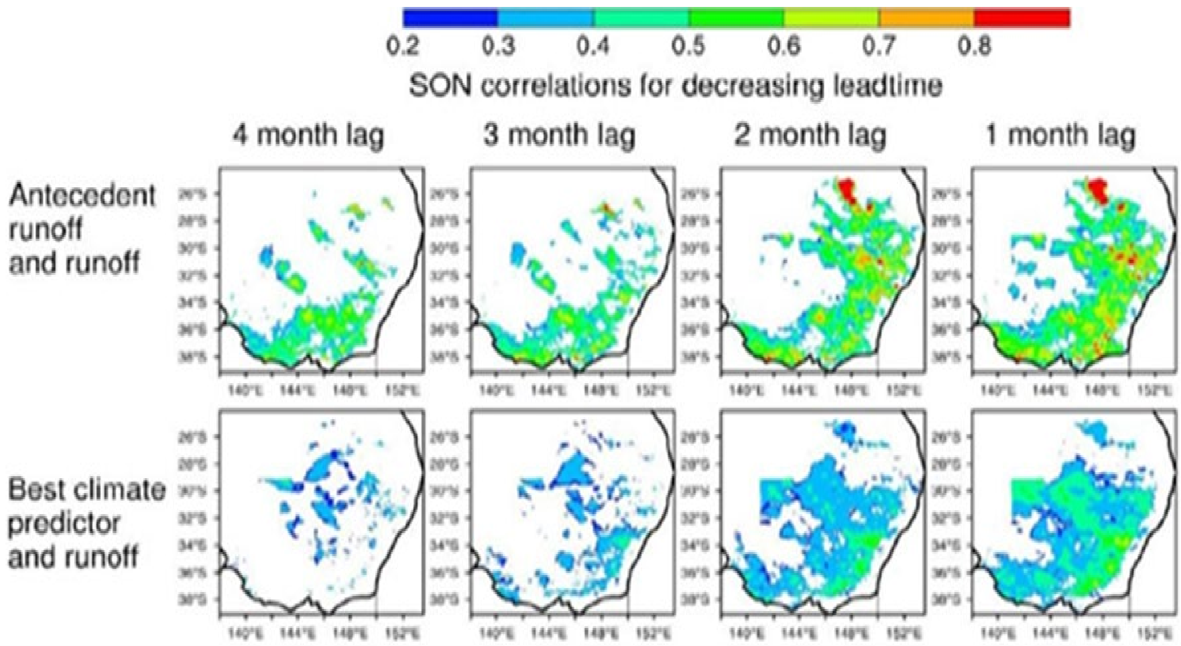


Figure 10. Correlation coefficient between antecedent runoff (top panel) and best atmospheric- oceanic predictor (bottom panel) and runoff (Source: Kirono et al. 2010)

Though the dominant source of inter-annual variability in Australian rainfall and streamflow is believed to be ENSO phenomenon, some recent evidence shows that Eastern Australia is also influenced by IOD as well as interdecadal modulation of ENSO as a result of the PDO (Westra and Sharma 2008). Cai et al. (2011) investigated the teleconnection between ENSO and IOD and their impact on Australian rainfall. He found that IOD has an impact on austral winter in the southern part of Australia, whereas ENSO has a strong influence on austral spring rainfall as a result of the strong covariation of ENSO and IOD. In the southwest and southeast, Australia, IOD is found to be influential on the rainfall from June to October (Risbey et al. 2009). The influence of ENSO and IOD cover the whole continent during this period. Australian rainfall especially the western, southern and central part of the country largely influenced by IOD (Saji et al. 1999; Saji and Yamgata 2003; Ashok et al. 2003; Nicholls 1989; Verdon and Franks 2005).

Nicholls (2010) analyzed the rainfall trends with the high-quality dataset and identified enhanced rainfall pattern over most of the east Australian regions (from 1950 to 1990), which is similar to the findings of Suppiah (2004) who found increasing rain in most parts of south-eastern Australia from 1910. On the contrary, Gallant et al. (2007), in their study of trends in rainfall for the six Australian regions explored that there has been a significant decrease in annual total rainfall in the coastal-east (55 mm per decade) and southeast regions (20 mm per decade. Nicholls (2010) mentioned the recent decline to be the influence of SAM trend from 1958 to 2007.

Wang et al. (2011) took a dynamic forecasting approach based on conceptual rainfall-runoff modelling for analyzing monthly and three-monthly streamflow in east Australia. It was explained in the study that the POAMA forecast skills show significant improvements only for monthly forecasts (not for three monthly forecasts), while compared to the historical ensemble. Gudmundsson et al. (2015) showed a comparative analysis of runoff and streamflow all over the world where it is clearly evident that there is a significant regional variation of streamflow and runoff throughout Australia and in the recent years the streamflow occurrences are more visible in the eastern part compared to the rest of the parts of the country (Figure 11).

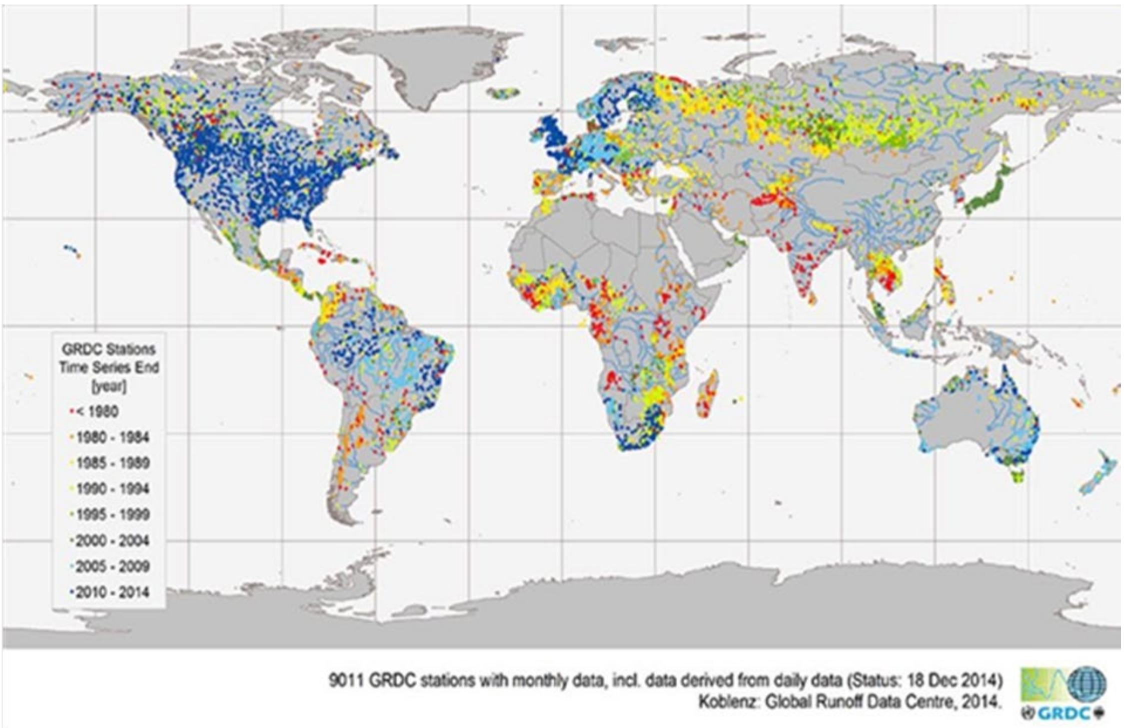


Figure 11. An In-Situ observation of runoff and streamflow (Gudmundsson et al. 2015)

2.4 Streamflow Forecasting in NSW Using Climate Drivers

NSW, located in the south-eastern part of Australia, is stricken by frequent droughts, especially in the western and north-eastern region of the state. Agricultural production is the major part of the economy of NSW contributes about $15 billion annually. Inter-annual variability of rainfall and streamflow impose a threat to the agricultural productivity and production of the state. Climate variability has a serious impact on the yield of planted crops like wheat, rice etc. Advanced skilful knowledge of climate variability can be utilized by farmers to have good production during high rainfall and avoid losses during drought.

Over the years researchers have been trying to determine a systematic trend of rainfall as well as streamflow, which may arise from climate change or global warming. Whiting et al. (2003) studied the rainfall in Sydney and found that there exists a greater correlation of annual rainfall in Sydney with the PDO index than with SOI. A recent attempt (Duc et al. 2017) was made using the Bayesian Model Averaging (BMA) method to analyze the combined impact of the four major climate drivers on the rainfall of NSW as well as to compare their relative contributions in the model. The correlations between ENSO phenomenon and seasonal rainfall in central NSW are found to be the strongest during spring (McBride and Nicholls 1983). Another study found a strong correlation between PDO and the Australian summer monsoon as positive PDO is responsible for below-normal monsoon rainfall (Latif et al. 1997). Robertson and Wang (2009) investigated on 12 climatic predictors with a view to selecting the best predictors for forecasting seasonal streamflow in Murrumbidgee catchment of NSW and found that the greatest prediction accuracy occurred between September and December with only a small increment for the remainder of the year while the best indicators were found to be the anomalies of Pacific Ocean that are related to ENSO. These findings were similar to the previous findings (McBride and Nicholls 1983) that found the strongest correlations between seasonal rainfall of NSW and ENSO during spring.

A combination of correlation and wavelet-based methods was applied to identify the principal sources of variation in reservoir inflows of Sydney (Westra and Sharma, 2009). The study found ENSO, PDO and IOD to be influential and obtained statistically significant correlations (±.4) which varied depending on the seasons, although correlations were lower for spring. (Westra and Sharma, 2009). One of the previous studies by the same authors showed that rainfall in the eastern coastal fringe (east of Great Dividing Range) varied with a period of 13 years, which is longer than the period of 2-8 years associated with ENSO and shorter than the period of 24 years associated with PDO.

The best skills for 3 months' streamflow forecast were obtained from April to June and October to January for the catchments in Queensland and NSW, respectively, which are comparable to the outcomes that were obtained using a statistical BJP approach (Robertson and Wang, 2008; Wang et al. 2009). IPO has a direct influence on the changes in the correlation coefficients between the rainfall patterns of NSW and SOI as it changes from positive correlation to none or negative over time. In accordance with this, Powel et al. (1999) stated that the correlation between SOI and rainfall is weak when IPO is in a positive phase and vice versa. Duc et al. (2017) showed in their study that IPO alone does not have any significant impact on the rainfall of NSW, but its combination with ENSO can make a significant impact on rainfall.

Duc et al. (2017) observed the different rainfall trends from 2000 to present at sites such as Coonabarabran and Yamba from those at sites such as Sydney and Wagga Wagga and suggested that additional climate drivers besides PDO and ENSO could be influencing the rainfall across Australia. He also mentioned that the interaction between ENSO and PDO could be more complicated than it had been considered in the previous studies.

The eastern Australian seaboard has different climate patterns compared to the interior part of the country where climatic influences on each side of the Great Dividing Range are also different. The study of Chiew and McMahon (2003) indicates that there exists a clear El Nino-streamflow teleconnections across most of Australia except in NSW catchments east of Great Dividing Range, which is stronger than the El Nino- Rainfall teleconnection.

Another climate driver that has an influence on the climate of Australia is MJO (Madden-Julian Oscillation), which is an eastward –propagating wavelike atmospheric disturbance in equatorial latitudes between the Indian and Pacific Oceans. Risbey et al. (2009) found it to operate on a shorter time scale with phase changes influential over days or weeks rather than seasons or years. However, the influence of MJO is limited to monsoonal tropical north and to a lesser extent in localized central and southern

Australia (Duc et al. 2017; Wheeler et al. 2009). As it has limited impact on southeast Australia, it is not included in the current study for forecasting streamflow.

The study of Owens et al. (2013) compared the predictability of persistence and average SOI to forecast streamflow and found that NSW including ACT to show the largest percentage of forecasting skill as this region had streamflow records from large catchments and rivers. The study of Chiew et al. (2003) explained that spring rainfall and runoff had a high correlation (0.3 to 0.5) against winter SOI throughout eastern Australia except for NSW east of Great Dividing Range, although this region showed high correlation for summer rainfall and runoff against winter SOI while the correlation was even higher against MEI for northern NSW.

The analysis (Kiem and franks 2001) on several method-index combinations to determine the best method that had the strongest relation with rainfall and runoff in the Williams River catchment of NSW explored Multivariate Index of ENSO (MEI) to provide the best classification. The same study applied "rise rule" and found strong relationships with runoff (rainfall) up to nine months in advance of the summer and autumn period. Wang et al. (2011) found lower rainfall variability in the catchment of NSW than the catchment of Queensland and obtained the best skills after the wetter winter period (October onward) when the catchment was wet, and the influence of initial catchment states was increased on subsequent streamflow. Wang and Robertson (2011) explained that they could not find any evidence of impacts of lagged Tasman Sea Index (TSI) to forecast seasonal rainfall in south-eastern Australia in any season.

2.5 Summary from the Extensive Literature Review

South East Australia contributes to more than half of Australia's agricultural income. Therefore, forecasting streamflow with a long lead time will be useful to the agricultural producers, irrigators and other water users which are even more important than rainfall forecasting as streamflow is the ultimate outcome of rainfall. Although efforts have been made to forecast streamflow and rainfall, none of the current practices provides reliable seasonal streamflow forecast, which can enable the water users to take risk-free managing decisions at the early stage of the crop period (Khan et al. 2005).

To date, most of the studies on streamflow forecasting are focused on selective parts of the country and predictor variables due to unavailability of reliable long data records of streamflow measurements (e.g. Piechota et al. 1998; Kiem and Franks 2004; Ruiz et al., 2007). None of the studies considered the combined impact of the climate indices to predict seasonal streamflow of NSW. A summary of the previous research works on streamflow forecasting of Australia is briefly presented in Table 1.

Table 1. Summary of methods used for forecasting streamflow of NSW

Reference Study	**Study Region**	**Forecast period**	**Method Applied**	**Predictor used**	**Results for NSW region**
Wang et al.(2011)	East Australia (Goodradigbee River at Wee Jasper)	3 monthly streamflow	Dynamic models (SIMHYD and MWB)	Initial catchment condition, historical rainfall, SOI	Best skills during OCT-JAN
Roertsonand Wang (2008); Wang et al. (2009)	South-Eastern Australia (Murrumbidgee)	Seasonal streamflow	Statistical BJP approach	IODMI,ENSO,SAM, EMI, Thermocline, Catchment rainfall etc.	Best predictability during Spring; Best predictors Pacific Ocean anomalies
Piechota et al. (2001)	5 Australian Catchments (1 station in NSW)	Seasonal streamflow	Linear Discriminant Analysis (LDA)	Persistence, SOI, SST	Persistence best but SOI and SST effective in absence of persistence
Piechota et al. (1998)	10 Eastern Australian catchment(5 stations in NSW)	Seasonal streamflow	Linear Discriminant Analysis (LDA)	SOI, SST	SOI better for July-Dec;SST better for (Jan-June)
Ruiz et al. (2007)	48 stations in Northern and Eastern Australia	6 months, 12 months streamflow	Linear Error in Probability Space(LEPS)	Nino, SST, thermocline	(Nino + Thermocline) model superior
Owen et al. (2013)	320 Australian Stations	Upto 4 months lagged streamflow	Australian Rainman streamflow software	Persistence, SOI	Persistence better than SOI;Strongest relation during spring
Kirono et al. (2010)	Southeast Australia (Moree)	Seasonal streamflow (upto 12 months lagged)	Single Lagged Correlation Analysis	12 atmospheric-oceanic predictors, antecedent runoff , Serial correlation	Atmospheric-oceanic predictors for spring, summer, winter
Chiew and McMahon (2007)	284 catchments of Australia	Several months lagged	Teleconnection (Harmonic Analysis)	ENSO,SOI, Serial correlation	Spring streamflow form winter SOI

Chiew et al. (2009)	45 statios in Australia (Coastal and Murray Darling Region of NSW)	Seasonal stremaflow	Single Lag correlation analysis	Persistence, SOI, SST	Spring runoff against 1 and 3 months lagged SOI and SST

It is expected that the different regions of NSW will have variations in the streamflow patterns due to the different influence of various climate drivers and their interactions with each other. Therefore, the aim of this study is to develop a seasonal streamflow forecasting method considering the combined impacts of ENSO based SST anomalies Nino3.4, EMI, IPO, PDO, and DMI (IOD) that have a major influence on spring streamflow in New South Wales.

To accomplish this objective NSW has been divided into four different regions, and for each region, various streamflow stations are chosen to identify the best predictor models for forecasting spring streamflow several months ahead in these regions. The locations are chosen considering the consumptive water use, where streamflow predictions were potential of some value and the longer availability of streamflow data.

In the present study, both linear (MLR) and non-linear (MNLR and GEP) models are developed to capture all possible relations between seasonal streamflow and large-scale climate variables. As the linear regression serves as a more simple, direct and consistent method than the other statistical methods, in the present study, this method is applied as the benchmark method. Later, the MNLR technique was used to explore the non-linear relationship between climate indices and seasonal streamflow. Finally, advanced GEP technique is applied with a view to getting better predictor models. This study is conducted with a view to filling up the existing hydrological research gap that prevails in Australia.

Chapter 3
Study Area and Data Collection

3.1 Background of the Study Area

Australia, the smallest continent in the map, has a huge contribution to agricultural production. Approximately there is 85,681 farm business which produces almost 93% of Australia's daily domestic food supply (National Farmer Federation 2017). Farmers are more dependable on water from streamflow for the agricultural purpose. Australia has extensive topographic variations due to which there are high climatic variabilities consigning hard to predict the streamflow in Australia. So, the farmers had to go under many problems in managing the water supply for agriculture. During 2002-2003, drought in Australia brought a reduction in Gross Domestic Production (GDP) reduction by 1.6% (Horridge et al. 2005). However, for the management of proactive risk management like drought management, the seasonal forecasts play a vital role.

NSW, which is situated in the east coast of Australia covering a land area of 880,0642 km^2 is the most populous state of Australia with a population of 7.5 million two-third of which live in Greater Sydney Area. The state is bordered on the north by Queensland, on the west by South Australia, on the south by Victoria and on the east by the Tasman Sea. The two most important features of NSW are the Great Dividing Range (GDB) and Murray Darling Basin (MDB) which accounts for nearly 40% of the value of agricultural production in Australia and 65% of the irrigated land (Abbot et al. 2015).

NSW possesses almost 61% of the water resources plan area of MDB (Department of Industry, NSW Government) while all of Australia's irrigated rice is produced by Murrumbidgee and NSW Murray irrigation regions (Murray-Darling Basin Authority). According to ABARE, 80.92% parts of the state is agricultural land which contributed 23 per cent of the total gross value of agricultural production in Australia in the year 2015-16.

Geographically NSW can be categorized into four different regions. The coastal regions, which are in the east of the state adjacent to the Tasman Sea, have a rainfall variation of around 800 millimetres to 3000 millimetres. Rainfall is moderate (600mm-1500 mm) and evenly distributed throughout the year in the highlands which is a part of the Great Dividing Range. The main agricultural region of NSW is the western inland slopes, which have a less dense population than coastal areas. This area receives high rainfall (600 mm) throughout the year. The western arid or semi-arid plains, which cover almost two-thirds of the state, experience an average rainfall of 150 mm to 500 mm in almost all the time of the year across the whole region.

3.2 Data Used in this Study

3.2.1 Streamflow Data

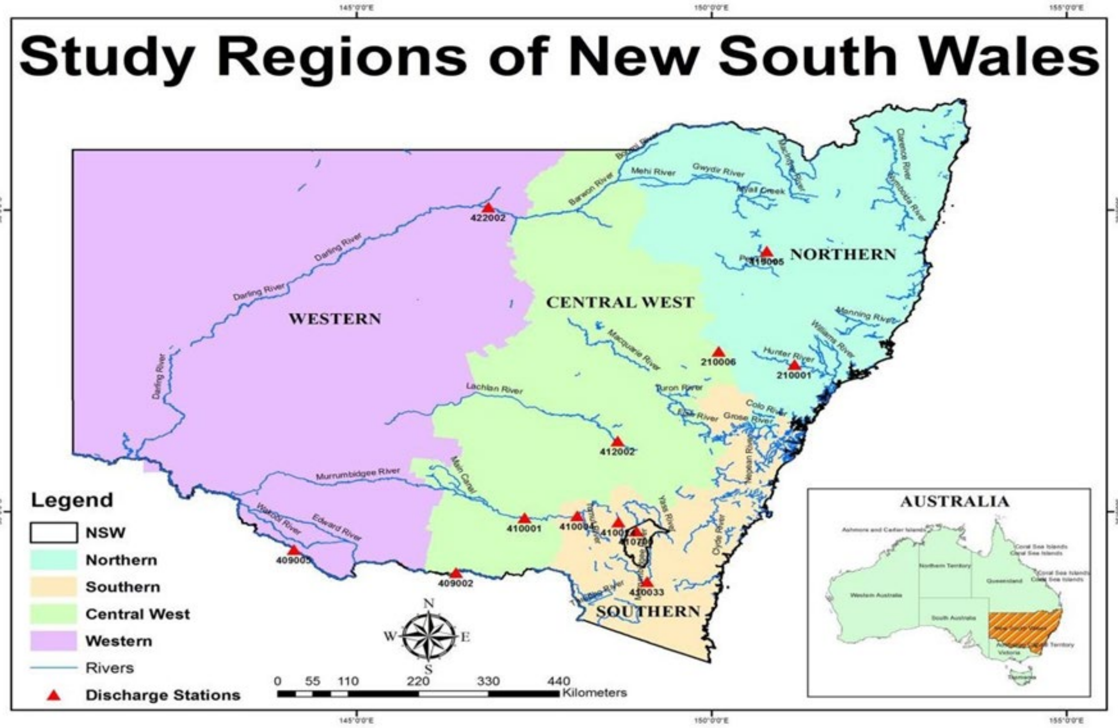


Figure 12. Locations of the discharge stations in the four study regions of NSW

Economically NSW is the most important state of Australia as it contributes most of Australia's agricultural production. Agriculture is spread throughout the eastern two-thirds of the state. Considering the geographical location, regional climatic variation and the agricultural importance, NSW is further divided into four regions for the current study- Northern NSW (NNSW), Southern NSW (SNSW), Central West NSW (CWNSW) and Western NSW (WNSW) (Figure 12). To explore the spatial variation of influences of different climatic variables for each region three, four, three and two stations are selected respectively based on their long data records and fewer missing values (Table 2).

Table 2. Overview of the Selected Discharge Stations

Study Region	Station Number	Latitude	Longitude	River Name	Station Name
Northern	210001	-32.56^{0}S	151.17^{0}E	HUNTER	SINGLETON
	210006	-32.34^{0}S	150.10^{0}E	GOULBURN	COGGAN
	419005	-30.68^{0}S	150.78^{0}E	NAMOI	NORTH CUERINDI
Southern	410004	-35.07^{0}S	148.11^{0}E	MURRUMBIDGEE	GUNDAGAI
	410024	-35.17^{0}S	148.69^{0}E	GOODRADIGBEE	WEE JASPER (KASHMIR)
	410033	-36.16^{0}S	149.09^{0}E	MURRUMBIDGEE	MITTAGANG CROSSING
	410700	-35.32^{0}S	148.94^{0}E	COTTER	KIOSK
Central	409002	-36.01^{0}S	146.40^{0}E	MURRAY	COROWA
	410001	-35.10^{0}S	147.37^{0}E	MURRUMBIDGEE	WAGGA WAGGA
	412002	-33.83^{0}S	148.68^{0}E	LANCHAN	COWRA
Western	409005	-35.63^{0}S	144.12^{0}E	MURRAY	BARHAM
	422002	-29.95^{0}S	146.86^{0}E	BARWON	BREWARRINA

Streamflow stations are chosen with data records considered to be of appropriate length for the statistical analysis carried out in this study. It can be seen from Figure 12 that the locations of the streamflow stations provide good spatial coverage of NSW. However, most of the stations selected in this study are concentrated in the eastern part of NSW due to predominance of coastal rivers as well as the agricultural importance of this region.

The starting point for any statistical analysis should be the data collection; useful quality data are required for the development, calibration and validation of any model. In this study, historical streamflow data is collected from the Australian Bureau of Meteorology (BoM, 2000).

Observed monthly streamflow in the unit of cumec (cubic meter per second) is collected for 102 years, ranging from 1914 to 2015 for nine stations while 99, 101 and 88 years of data are collected for North Cuerindi, Wee Jasper and Mittangang Crossing stations respectively. These stations have less than 0.5% missing values, which are filled by the series mean of the streamflow data. Using this data, seasonal mean discharge data is derived for spring (September-October-November) season.

3.2.2 Climate Indices Data

Five climate drivers: ENSO based SST anomalies NINO3.4, EMI, IPO, PDO, and DMI(IOD) are selected for MLR analysis, considering the previous research works on rainfall and streamflow in this region as well as the concurrent and lagged correlation analysis in the preliminary stage of the current research.

ENSO phenomenon has two components- sea surface temperature and atmospheric pressure which are intensely correlated and can be represented by two types of indicators, the SLP indicator and the SST indicator (Duc et al. 2017). The most common of those is the Troup Southern Oscillation Index (SOI) which measures the difference between sea-level atmospheric pressures at Papeete (Tahiti) and Darwin (Troup 1965). Various SST anomalies are also available which have been derived using different areas of the equatorial Pacific Ocean (Kiem & Franks 2001). Generally, the SST anomalies are monitored in 3 geographic regions (Figure 13) of the equatorial Pacific and defined as NINO3 (5°S – 5°N, 150°– 90°W), NINO3.4 (5°S – 5°N, 170° – 120°W) and NINO4 (5°S – 5°N, 160° – 150°W) (Risbey et al. 2009). Hanley et al. (2003) compared the response of pressure based and SST based anomalies to the ENSO extreme event and found that Niño-3.4 and Niño-4 indices are equally sensitive to El Niño events whereas SOI is less sensitive to La Niña events than others.

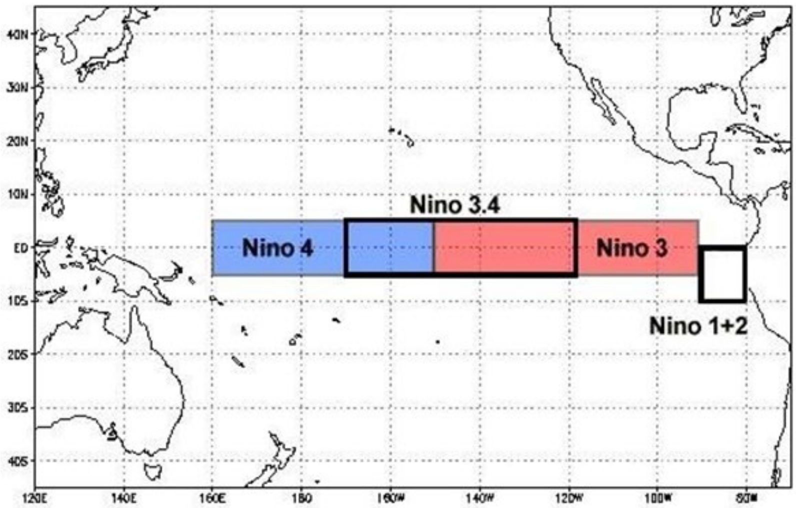


Figure 13. Map showing ENSO region (source: NOAA)

The El Nin˜o Modoki is an ocean-atmosphere coupled process, which results in unique tripolar sea level pressure pattern during the evolution, similar to the Southern

Oscillation phenomenon of El Nino (Ashok et al. 2007). Therefore, this phenomenon is named as El Nin˜o–Southern Oscillation (ENSO) Modoki and expressed by the following equation (Ashok et al. 2007)

$$EMI = SSTX - (0.5 * SSTY) - (0.5 * SSTZ) \quad (3\text{-}1)$$

Where,

X=165°E–140°W, 10°S–10°N, Y= 110°W–70°W, 15°S–5°N, Z=125°E–145°E, 10°S–20°N

The IOD represents the couples oceanic-atmospheric variability in the tropical Indian Ocean which is classified by SST anomalies of reverse sign in the east and west (Saji et al. 1999; Webster et al. 1999). The Dipole Mode Index (DMI) which is a measure of the IOD is defined as the difference in SST anomaly between the tropical western Indian Ocean (10ºS–10ºN, 50º–70ºE) the tropical south-eastern Indian Ocean (10ºS–equator, 90º–110ºE).

The IPO is described as the Pacific ENSO-like pattern of SST, which is found in the analysis of near-global inter-decadal SST (Folland et al. 1999). IPO has a cycle of 15-30 years and characterized with two phases namely, positive and negative phases (Henley et al. 2015; Salinger et al. 2001). While IPO is defined for the whole Pacific Basin, PDO is defined for the North Pacific, poleward of 20°N (Figure 14).

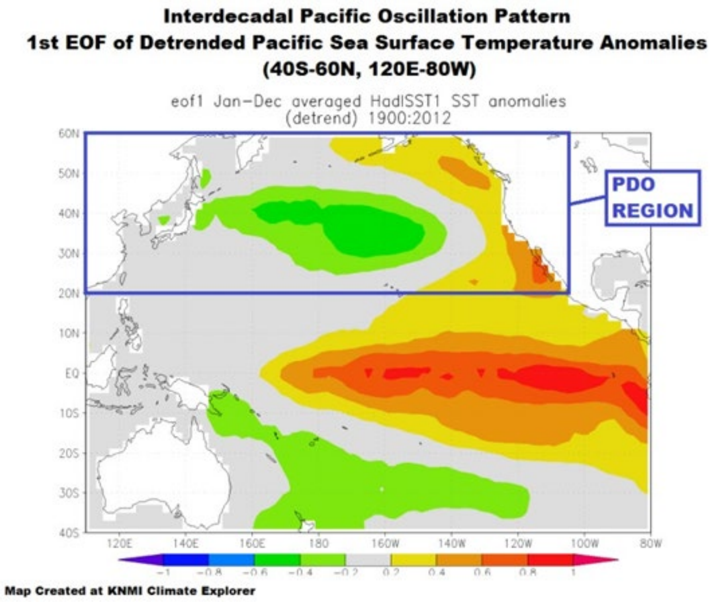


Figure 14. Map showing IPO and PDO region

(Source: Timmermann and Trenberth, 2014)

The oceanic and atmospheric climate indices data are obtained from the Climate Explorer website (http://climexp.knmi.nl) while the EMI data is collected from the website of JAMSTEC (http://www.jamstec.go.jp/frcgc/research/dl/iod/modoki) for the duration of 102 years (1914-2015). An overview of the used climatic variables is presented in Table 3.

Table 3. Overview of the Used Climate Indices and Data Source

Predictors	Predictor definition	Origin	Data period	Data Source
PDO	SSTA anomaly in North Pacific Ocean, (north of 20°N	Pacific Ocean	1914-2015	ERSST(http://climexp.knmi.nl/)
IPO	SST anomaly in North and South Pacific Ocean (includes South of 20°N Latitude)	Pacific Ocean	1914-2015	HadISST1 (http://climexp.knmi.nl/)
Nino3.4	Average SST anomaly over central Pacific ocean(5°S-5°N,120°-	Pacific Ocean	1914-2015	HadISST1(http://climexp.knmi.nl/)
IOD	West Pole Index(10°S-10°N,50°-70°E) -East Pole Index(10°S-0°,	Indian Ocean	1914-2015	HadISST (http://climexp.knmi.nl/)
EMI	Coupled ocean-atmosphere phenomenon in the	Pacific Ocean	1914-2015	HadISST (http://www.jamstec.go.jp/frcgc/research/d1/iod/modoki)

3.3 Data Preprocessing

Data preprocessing has got significant influences on model outputs, especially when the input variables have got different ranges of data. It is important to normalize the data to ensure that every input variable gets equal attention during the calibration of the model, which will result in significant improvement in the accuracy of the developed models. For Multiple Regression (MR) analysis, the basic statistical assumptions are very important. According to Osborne & Elaine (2002), MR modelling was established based on a number of assumptions; thus, if all these assumptions are not satisfied, the results may become highly erroneous. These assumptions include but not limited to the type of variables, no multicollinearity among the variables, additivity of the independent variables. Again, for GEP analysis, there are several assumptions that need to be maintained to get more accurate outcomes. One of the basic conditions of GEP modelling is that the input variables should be of equal length. Therefore, while setting the initial environment, this condition was kept in mind, and any missing value for any variable was replaced by the mean value of the corresponding data series.

Another problem of dealing with the long dataset is the presence of extreme or influential data points. These data points which are far outside the norm for a variable or population, are referred to as outliers (Jarrell, 1994; Rasmussen, 1988; Stevens, 1984). Outliers can have adverse effects on statistical analyses by increasing error variance and reducing the power of statistical tests. Outliers can be generated by different reasons, among which two main reasons are errors in the data and inherent variability of the data. There are arguments on whether to remove the outliers or not. In the study by Osborne and Amy (2004), the effect of outliers on correlation analysis was found to be very significant. It was revealed in their study that removal of the outliers had a very significant impact on the magnitude of the correlation and the cleaned correlations were more accurate. Although there are arguments among the researchers (Orr et al. 1991) for removal or modifying the outliers, this research (Osborne & Amy, 2004) suggested that removal of outliers enhanced the accuracy of the results in both correlation and t-tests conducted by the authors and in most cases the error was significantly reduced.

Considering the aforementioned research outcomes, in the current study, the outliers were determined through boxplot method (Figure 15) and then those data points were removed with a view to getting more accurate results from the developed models.

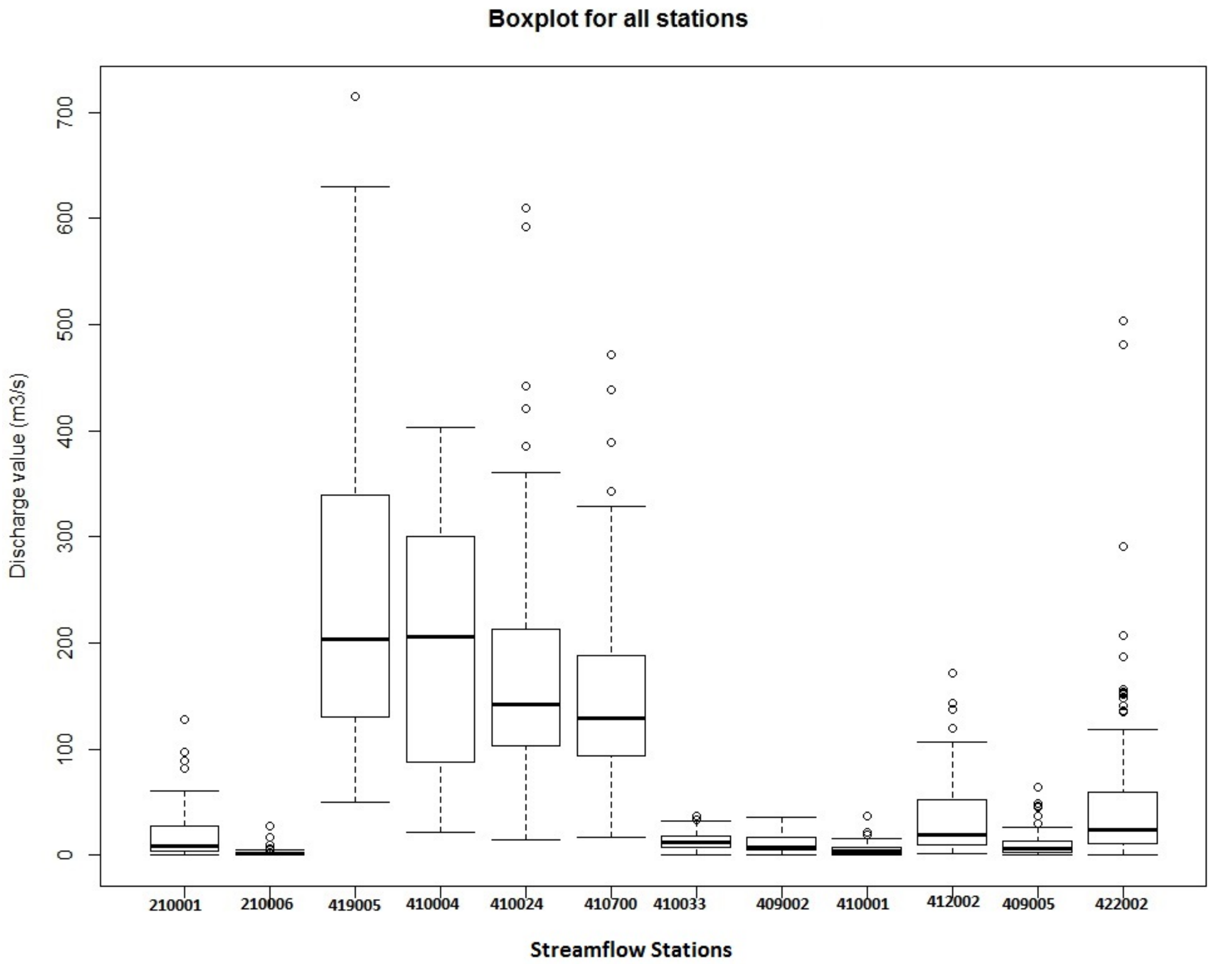


Figure 15. Boxplots to find out outliers for all 12 stations

A boxplot or a box-and-whisker diagram is a graphical representation of groups of numerical data through their quartiles (Figure 16) which extends from the 25^{th} percentile to the 75^{th} percentile including a line at 50^{th} percentile (median). This box is referred to Interquartile Range (IQR). The outliers are plotted as the individual points.

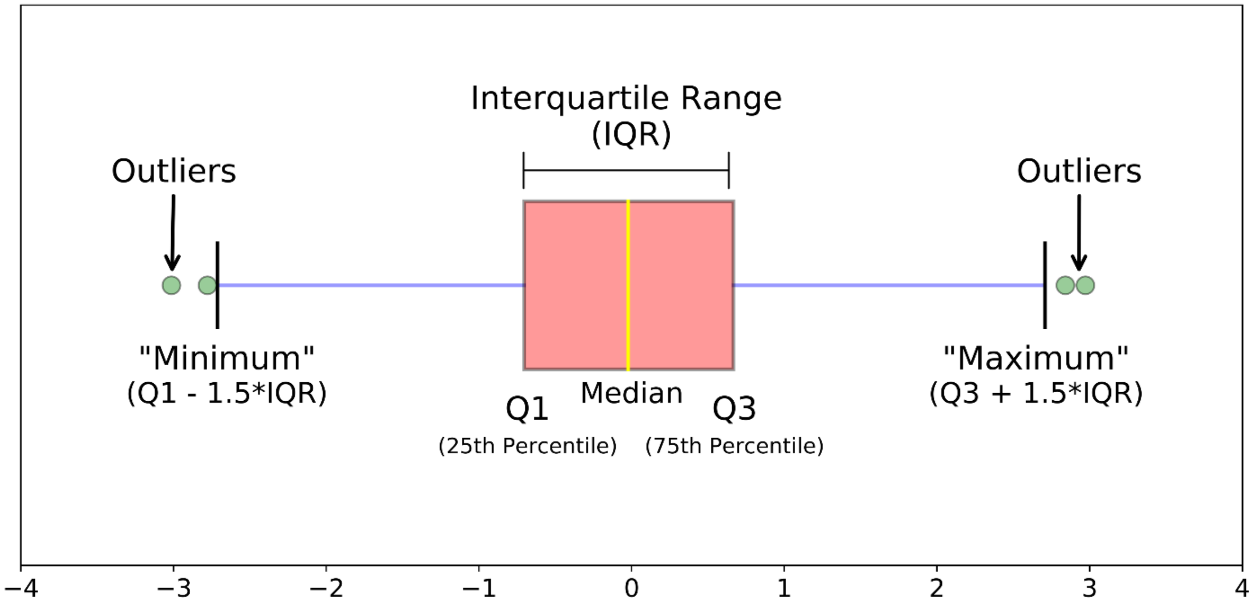


Figure 16. Different parts of Boxplot (Source: Galarnyk 2018)

3.4 Model Verification

3.4.1 Statistical Performance Analysis

Statistical performances of the models were evaluated using different functions which include Root Mean Square Error (RMSE), Root Relative Squared Error (RRSE), Relative Absolute Error (RAE), Mean Absolute Error (MAE), Nash-Sutcliffe Efficiency (NSE), Willmott index of agreement (d) and Pearson correlation (r). The closer the 'd' value to 1, the better the model fits the observations. The best models for each station was chosen considering its higher correlation value and lower errors which ensure the best fitness of the developed model.

Pearson Correlation Coefficient (r)

Correlation coefficients are used in statistics to measure how strong a relationship is between two variables. One of the most commonly used correlation coefficients is Pearson's correlation (also called Pearson's r). The equation which is used to calculate the coefficient is as follows:

$$\boldsymbol{r} = \frac{\sum_{i=1}^{n}(x_i-\bar{x}\ )(y_i-\bar{y}\ )}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x}\ )}\times\sqrt{\sum_{i=1}^{n}(y_i-\ \bar{y}\ )^2}} \qquad (3\text{-}2)$$

where, r is the correlation coefficient between variables $\bar{x}$ and $\bar{y}$; and are the average values of x and y, respectively and n is the number of data points. r ranges within the domain [-1, 1] where the values of 1 and -1 indicate positive and negative perfect linear correlation respectively, while r = 0 is an indication that there is no correlation between the two data series.

Mean Absolute Error (MAE)

The mean absolute error (MAE), E_i of an individual model i is evaluated by the equation:

$$E_i = \frac{1}{n}\sum_{j=1}^{n}\left|P_{(ij)} - T_j\right| \tag{3-3}$$

where P(ij) is the value predicted by the individual model i for record j (out of n records); and T_j is the target value for record j. For a perfect fit, $P_{(ij)} = T_j$ and $E_i = 0$. So, the E_i index ranges from 0 to infinity, with 0 corresponding to the ideal.

Root Mean Squared Error (RMSE)

The root mean squared error (RMSE) E_i of an individual model i is evaluated by the equation:

$$E_i = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\left(P_{(ij)} - T_j\right)^2} \tag{3-4}$$

where $P_{(ij)}$ is the value predicted by the individual model i for record j (out of n records); and T_j is the target value for record j. For a perfect fit, $P_{(ij)} = T_j$ and $E_i = 0$. So, the E_i index ranges from 0 to infinity, with 0 corresponding to the ideal.

Root Relative Squared Error(RRSE)

The RRSE fitness function of GeneXpro Tools is, as expected, based on the standard root relative squared error, which, on its turn, is based on the absolute error. The relative squared error is relative to what it would have been if a simple predictor had been used. More specifically, this simple predictor is just the average of the actual values. Thus, the relative squared error takes the total squared error and normalizes it by dividing by the total squared error of the simple predictor. By taking the square root of the relative squared error one reduces the error to the same dimensions as the quantity being predicted. Mathematically, the root relative squared error E_i of an individual program i is evaluated by the equation:

$$\boldsymbol{E}_i = \sqrt{\frac{\sum_{j=1}^{n}(\boldsymbol{P}_{ij}-\boldsymbol{T}_j)^2}{\sum_{j=1}^{n}(\boldsymbol{T}_j-\bar{\boldsymbol{T}})^2}} \qquad (3\text{-}5)$$

where P(ij) is the value predicted by the individual program i for fitness case j (out of n fitness cases or sample cases); T_j is the target value for fitness case j; and $\bar{T}$ is given by the formula:

$$\bar{T} = \frac{1}{n}\sum_{j=1}^{n} T_j \qquad (3\text{-}6)$$

For a perfect fit, the numerator is equal to 0 and $E_i = 0$. So, the RRSE index ranges from 0 to infinity, with 0 corresponding to the ideal. As it stands, E_i cannot be used directly as fitness since, for fitness proportionate selection, the value of fitness must increase with efficiency. Thus, for evaluating the fitness f_i of an individual program i, the following equation is used:

$$f_i = 1000.\frac{1}{(1+E_i)} \qquad (3\text{-}7)$$

which obviously ranges from 0 to 1000, with 1000 corresponding to the ideal (Ferreira 2006).

Relative Absolute Error (RAE)

The relative absolute error (RAE) is very similar to the relative squared error in the sense that it is also relative to a simple predictor, which is just the average of the actual values. In this case, though, the error is just the total absolute error instead of the total squared error. Thus, the relative absolute error takes the total absolute error and normalizes it by dividing by the total absolute error of the simple predictor. Mathematically, the relative absolute error Ei of an individual model i is evaluated by the equation:

$$E_i = \frac{\sum_{j=1}^{n}|\boldsymbol{P}_{ij}-\boldsymbol{T}_j|}{\sum_{j=1}^{n}|\boldsymbol{T}_j-\bar{\boldsymbol{T}}|} \qquad (3\text{-}8)$$

where $P_{(ij)}$ is the value predicted by the individual model i for record j (out of n records); T_j is the target value for record j; and $\bar{T}$ is given by the formula:

$$\bar{T} = \frac{1}{n}\sum_{j=1}^{n} T_j \qquad (3\text{-}9)$$

For a perfect fit, the numerator is equal to 0 and $E_i = 0$. So, the E_i index ranges from 0 to infinity, with 0 corresponding to the ideal.

Nash–Sutcliffe model efficiency coefficient (NSE)

The Nash-Sutcliffe efficiency (NSE) is a normalized statistic that determines the relative magnitude of the residual variance ("noise") compared to the measured data variance ("information") (Nash and Sutcliffe, 1970). The is used to assess the predictive power of hydrological models. It is defined as:

$$NSE = 1 - \frac{\sum_{t-1}^{T}(Q_m^t - Q_o^t)^2}{\sum_{t=1}^{T}(Q_o^t - \bar{Q}_o)^2} \qquad (3\text{-}10)$$

where $\bar{Q}_o$ is the mean of observed discharges, and Q_m is modelled discharge. Q_o^t is observed discharge at time t.

NSE indicates how well the plot of observed versus simulated data fits the 1:1 line.

Nash-Sutcliffe efficiencies range from -Inf to 1. Essentially, the closer to 1, the more accurate the model is.

- NSE = 1, corresponds to a perfect match of modelled to the observed data.
- NSE = 0, indicates that the model predictions are as accurate as the mean of the observed data,
- -Inf < NSE < 0, indicates that the observed mean is better predictor than the model.

3.4.2 Calibration and Validation of the Developed Models

MLR analysis served as the basis for the current study, which was followed by MNLR analysis and GEP analysis, respectively. Since there are no hard and fast rules to divide the entire data series into calibration and validation datasets, at the beginning for MLR and MNLR analysis, this study took 85 years of data for calibration dataset (1914-1998) and the rest of 17 years (1999-2015) of data was used to validate the models. Later on, for GEP analysis to improve the model performances, the whole data set was divided into two segments where the first 96 years (1914-2009) of data were used for calibrating the models and rest of 6 years (2010-2015) of data were used to assess the validity of the models. Considering the effect of the "millennium drought" period (which was 1994 to 2010 according to Bond et al. 2008) in Australia, longer data range was used for calibration period to prepare the models for any unusual phenomenon like droughts or flood. Finally, to compare the MLR models with GEP models, MLR analysis was redone considering the same calibration and validation datasets as GEP.

Chapter 4
Analysis of the Relationship between Streamflow and Climate Indices

4.1 Pearson Correlation Analysis

A detailed study of past research works revealed that different climate anomalies have impacts on the seasonal streamflow of NSW, which varies seasonally as well as spatially. Pearson correlation analysis is done for identifying the strength of the relation between the seasonal climate anomalies. To analyze the extent of the influences of different climatic variables on seasonal streamflow of NSW, Pearson correlation analysis is applied in two different phases. The first phase takes into account the concurrent relationship of climate indices and seasonal streamflow while the second phase is conducted in order to investigate the lagged relationship of the climate models and seasonal streamflow. Later on, the outcomes of the second phase are served as the basis for selecting the suitable lagged climate indices to use as inputs while developing MLR, MNLR and GEP forecasting models.

4.1.1 Concurrent Relationship

4.1.1.1 Seasonal Relationship between Climate Indices

To check the multicollinearity is a very important step of MLR modelling. Multicollinearity is observed when there is a significant correlation among the predictors. Therefore, it is essential to check if the predictors are highly correlated to avoid getting biased results from the developed models. If the multicollinearity is not verified at the beginning, the modelling will result in an abrupt change in parameter estimates in response to small changes in the input data. In this study, to avoid highly correlated predictors, concurrent correlation analysis was conducted for the seasonal indices (Table 4-7). Usually, climate indices which are originated from the same climatic phenomenon, are highly correlated and cannot be used in the same MLR model as the predictors. For instance, NINO3 and SOI are originated from the ENSO phenomenon; thus, while developing the MLR models, these two indices cannot be used in the same model as predictors.

Table 4. Concurrent correlations between Summer climate indices

Indices	NINO4	SOI	IPO	PDO	EMI	IOD	SAM	NINO3.4
NINO3	0.85**	-0.82**	0.25**	0.57**	0.48**	0.06	-0.21	-0.1
NINO4		-0.83**	0.28**	0.56**	0.81**	0.01	-0.2	0
SOI			-0.26**	-0.55**	-0.57**	0.11	0.30*	-0.06
IPO				0.39**	0.15	-0.17	0.05	-0.23
PDO					0.31**	-0.03	-0.08	0.03
EMI						-0.16	-0.29*	0.11
IOD							0.41**	0.06
SAM								-0.26*

* Correlation is significant at 5% level ** Correlation is significant at 1% level

Table 5. Concurrent correlations between Winter climate indices

Indices	NINO4	SOI	IPO	PDO	EMI	IOD	SAM	NINO3.4
NINO3	0.69**	-0.63**	0.12	0.39**	-0.1	0.47**	0.29*	-0.11
NINO4		-0.7**	0.32**	0.52**	0.54**	0.31**	0.18	-0.08
SOI			-0.22*	-0.42**	-0.33**	-0.45**	-0.19	0.13
IPO				0.48**	0.2*	-0.02	0.14	-0.26*
PDO					0.13	0.01	0.11	-0.04
EMI						-0.1	-0.12	0.03
IOD							0.27*	-0.13
SAM								0.22

* Correlation is significant at 5% level ** Correlation is significant at 1% level

Table 6. Concurrent correlations between Autumn climate indices

INDEX	Nino4	SOI	IPO	PDO	EMI	IOD	SAM	NINO3.4
NINO3	0.72**	-0.66**	0.33**	0.54**	.08	-0.08	0	-0.1
NINO4		-0.7**	0.35**	0.57**	0.69**	-0.07	-0.16	0.02
SOI			-0.37**	-0.53**	-0.32**	-0.06	0.18	0.1
IPO				0.56**	0.2*	-0.19	0.06	-0.19
PDO					0.19*	-0.21*	-0.01	0.01
EMI						-0.11	-0.24	0.1
IOD							0.15	-0.07
SAM								-0.13

* Correlation is significant at 5% level ** Correlation is significant at 1% level

Table 7. Concurrent correlations between Spring climate indices

INDEX	Nino4	SOI	IPO	PDO	EMI	IOD	SAM	NINO3.4
NINO3	0.86**	-0.74**	0.11	0.59**	0.41**	0.60**	-0.06	-0.12
NINO4		-0.81**	0.18	0.57**	0.74**	0.59**	-0.03	-0.1
SOI			-0.24**	-0.54**	-0.63**	-0.53**	-0.01	0.07
IPO				0.25**	0.09	0.01	0.08	-0.2
PDO					0.23*	0.25**	0.02	0.08
EMI						0.31**	0	-0.01
IOD							-0.13	-0.08
SAM								-0.06

* Correlation is significant at 5% level ** Correlation is significant at 1% level

4.1.1.2 Seasonal Relationship between Streamflow and Climate Indices

The linear relationships between spring, summer, autumn and winter streamflow and spring, summer, autumn and spring climate anomalies respectively have been explored by directing the Pearson correlation analysis for each station of the four study regions of NSW. The purpose of this analysis is to identify the season that shows significant correlations with the highest number of climate indices compared to all other seasons (Table 8-11). Furthermore, this analysis plays an important role to select the most influential climate indices on the particular seasonal streamflow.

Table 8. Pearson correlations between Summer streamflow and Summer climate indices

Region	Station	NINO3	NINO4	SOI	IPO	PDO	EMI	IOD	NINO3.4	SAM
NNSW	210001	-0.31**	-0.35**	0.31**	-0.27**	-0.26**			-0.30**	
	419005	-0.38**	-0.42**	0.38**	-0.20*	-0.25**	-0.28**		-0.39**	
	210006	-0.33**	-0.38**	0.28**	-0.28**	-0.22*	-0.27**		-0.34**	
SNSW	410024	-0.36**	-0.43**	0.33**		-0.20*	-0.36**		-0.39**	
	410700	-0.32**	-0.37**	0.34**		-0.22*	-0.27**		-0.34**	
	410033		-0.25*							
	410004						-0.31**			
CWNSW	409002						-0.20*	0.25**		
	410001						-0.32**	0.28**		
	412002							0.27**		
WNSW	409005	-0.32**	-0.31**	0.30**	-0.33**	-0.31**			-0.31**	
	422002	-0.39**	-0.51**	0.48**	-0.35**	-0.33**	-0.44**	0.21*	-0.45**	

* Correlation is significant at 5% level ** Correlation is significant at 1% level

Table 9. Pearson correlations between Autumn streamflow and Autumn climate indices

Region	Station	NINO3	NINO4	SOI	IPO	PDO	EMI	IOD	NINO3.4	SAM
NNSW	210001		-0.26**	0.23*	-0.3**	-0.2*			-0.22*	
	419005				-0.27**					
	210006		-0.19*	0.21*	-0.26**	-0.24*				
SNSW	410024		-0.29**	0.31**	-0.25**	-0.28**	-0.23*	-0.20*	-0.24**	
	410700		-0.31**	0.33**	-0.25**	-0.31**	-0.23*	-0.19*	-0.26**	
	410033		-0.26**	0.31**		-0.33**		-0.24*	-0.25*	
	410004		-0.3**	0.27**		-0.28**	-0.23*		-0.27**	
CWNSW	409002									
	410001		-0.3**	0.28**	-0.25**	-0.27**	-0.23*		-0.27**	
	412002			0.28**				-0.24*		
WNSW	409005		-0.24*	0.36**	-0.2*				-0.20*	
	422002	-0.23*	-0.31**	0.32**	-0.38**	-0.31**			-0.28**	

* Correlation is significant at 5% level ** Correlation is significant at 1% level

Table 10. Pearson correlations between Winter streamflow and Winter climate indices

Region	Station	NINO3	NINO4	SOI	IPO	PDO	EMI	IOD	NINO3.4	SAM
NNSW	210001	-0.19*	-0.27**	0.3**		-0.32**			-0.24*	
	419005				-0.27**			-0.21*	-0.24*	
	210006	-0.2*	-0.27**	0.34**		-0.24**	-0.21*	-0.18	-0.26**	
SNSW	410024	-0.22*	-0.3**	0.39**				-0.32**	-0.28**	-0.43**
	410700	-0.2*	-0.34**	0.41**				-0.32**	-0.25**	-0.46**
	410033		-0.26**	0.31**	-0.22*	-0.33**		-0.32**	-0.23*	-0.30*
	410004		-0.32**	0.35**	-0.19*			-0.25**	-0.23*	-0.34**
CWNSW	409002		-0.22*	0.34**				-0.28**		
	410001		-0.3**	0.34**				-0.26**	-0.21*	-0.33**
	412002		-0.23*	0.34**				-0.14	-0.26**	-0.35**
WNSW	409005		-0.27**	0.37**				-0.32**		
	422002		-0.26**	0.33**		-0.2*	-0.21*	-0.20*	-0.41**	

* Correlation is significant at 5% level ** Correlation is significant at 1% level

Table 11. Pearson correlations between Spring streamflow and Spring climate indices

Region	Station	NINO3	NINO4	SOI	IPO	PDO	EMI	IOD	NINO3.4	SAM
NNSW	**210001**	-0.35**	-0.42**	0.38**		-0.34**	-0.25**	-0.19*	-0.38**	
	419005	-0.43**	-0.54**	0.52**		-0.44**	-0.44**	-0.33**	-0.47**	
	210006	-0.27**	-0.36**	0.33**		-0.28**	-0.3**		-0.32**	
SNSW	**410024**	-0.39**	-0.49**	0.46**	-0.23*	-0.26**	-0.36**	-0.50**	-0.43**	
	410700	-0.33**	-0.43**	0.42**		-0.37**	-0.22*	-0.46**	-0.34**	
	410033	-0.28**	-0.37**			-0.26**		-0.46**	-0.29**	
	410004	-0.26**	-0.36**	0.32**	-0.25**	-0.26**	-0.19*	-0.34**	-0.28**	
CWNSW	**409002**	-0.28**	-0.36**	0.31**	-0.21*	-0.3**		-0.43**	-0.29**	
	410001	-0.27**	-0.37**	0.33**	-0.24*	-0.26**	-0.21*	-0.36**	-0.28**	
	412002	-0.19*	-0.32**	0.26**	-0.23*		-0.21*	-0.24*	-0.22*	
WNSW	**409005**	-0.43**	-0.54**	0.52**		-0.44**	-0.44**	-0.33**	-0.47**	
	422002	-0.36**	-0.47**	0.38**		-0.35**	-0.42**	-0.35**	-0.42**	

* Correlation is significant at 5% level   ** Correlation is significant at 1% level

Comparing the results of the concurrent analyses between seasonal streamflow and seasonal climate indices, it can be said that spring streamflow (Table 11) showed strong correlations for the greatest number of the indices compared to all other seasons. It is evident from Table 11 that all the climate indices (except IPO) have significant correlations with spring streamflow for all the regions. IPO shows significant correlations only for CWNSW and for two stations of SNSW regions which are geographically located close to each other. Therefore, it can be anticipated that in the central-west and southern parts of NSW, IPO has a strong influence on streamflow during spring. Again, among all climate indices, SAM showed the least number of significant correlations with the corresponding seasonal flow.

It was mentioned earlier, based on this concurrent correlation analyses between the seasonal streamflow and seasonal indices, the most significant season to forecast and the most influential climate indices will be selected. Following the outcome of this analysis, this study will consider only Spring season for further analyses and NINO3, NINO4, NINO3.4, EMI, IOD, PDO and IPO were found to be the most influential indices.

4.1.2 Single Lagged Relationships

For each selected station of four study regions of NSW, single lag correlation analysis is performed between spring streamflow at year 'n' and monthly (December$_{n-1}$ to August$_{n}$) values of the climate indices. The outcomes are presented in Table 12-15.

It is observed from the single lagged analysis that different region is influenced by different climatic variables. Lagged NINO3.4 and PDO have significant impacts on the spring streamflow of all four selected regions. CWNSW, including the two stations of Southern NSW (Gundagai and Wee Jasper Stations) which are nearly located to CWNSW stations, are also influenced by lagged IPO. Almost all the stations have signification correlations with lagged IOD indices whereas lagged EMI has impacts on a very limited number of stations.

NINO3.4 shows statistically significant correlations up to a lagged period of four months (April to August) with the correlations ranging from -0.19 to -0.46. These findings align with the study of Wang et al. (2009), who found a strong impact of NINO3.4 on spring rainfall in the same study region up to a lag of 2 months. Duc et al. (2017) explained that ENSO indices have a strong impact on rainfall during Austral spring. The lagged periods of IOD that have significant correlations are not consistent for all the stations while most of the stations have significant correlation up to 4 months' lag. It is observed that EMI has significant correlation only up to 3 months' lag for five stations with a maximum significant correlation of -0.41. The lagged relationship of PDO is quite different as for some stations it presents more significant correlations with longer lead times. The maximum lead times (up to 9 months) with significant correlations for spring streamflow are also obtained for this climate variable of the Pacific Ocean. This is similar to the assessment of Whiting et al. (2003) that discovered that PDO has a greater correlation with an annual rainfall of Sydney than that of SOI. Latif et al. (1997) also showed a strong relationship between PDO and Australian summer monsoon. Westra et al. (2008) evaluated correlation coefficients between seasonal inflows of a reservoir in Sydney and climate indices where spring inflow correlations for NINO3.4 and PDO were found to be -0.17 and -0.19 respectively.

Table 12. Pearson correlations (r) of single lagged climate indices and spring streamflow NNSW

		Lagged Months								
Stations	Indices	Dec_{n-1}	Jan_n	Feb_n	Mar_n	Apr_n	May_n	$June_n$	$July_n$	Aug_n
210001	NINO3	-0.07	-0.16	-0.18	-0.17	-0.21*	-0.30**	-0.36**	-0.35**	-0.27**
	NINO4	-0.16	-0.13	-0.17	-0.25**	-0.28**	-0.37**	-0.42**	-0.42**	-0.39**
	SOI	0.08	0.02	0.15	0.20*	0.20*	0.23*	0.29**	0.30**	0.29**
	PDO	-0.17	-0.26**	-0.33**	-0.29**	-0.30**	-0.35**	-0.30**	-0.37**	-0.34**
	EMI	-0.13	-0.01	-0.03	-0.14	-0.08	-0.1	-0.14	-0.07	-0.16
	IPO	0.11	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06
	NINO3.4	-0.08	-0.13	-0.15	-0.19	-0.22*	-0.35**	-0.43**	-0.39**	-0.31**
	IOD	-0.05	-0.05	-0.1	-0.08	0.03	-0.15	-0.22*	-0.12	-0.15
210006	NINO3	0.01	-0.02	-0.1	-0.06	-0.17	-0.23*	-0.25**	-0.27**	-0.19*
	NINO4	-0.05	0.01	-0.08	-0.09	-0.17	-0.22*	-0.27**	-0.35**	-0.35**
	SOI	0.02	-0.05	0.04	0.12	0.15	0.20*	0.38**	0.33**	0.29**
	PDO	-0.11	-0.19*	-0.22*	-0.21*	-0.24**	-0.30**	-0.23*	-0.32**	-0.31**
	EMI	-0.07	0.06	0	-0.06	-0.06	-0.09	-0.14	-0.18	-0.26**
	IPO	-0.01	-0.17	-0.17	-0.17	-0.17	-0.17	-0.17	-0.17	-0.17
	NINO3.4	0	0	-0.07	-0.06	-0.17	-0.28**	-0.31**	-0.35**	-0.26**
	IOD	0.11	0.06	-0.11	-0.11	-0.06	-0.06	-0.21*	-0.13	-0.12
419005	NINO3	-0.01	-0.07	-0.11	-0.11	-0.12	-0.25**	-0.36**	-0.35**	-0.29**
	NINO4	-0.14	-0.1	-0.15	-0.23*	-0.29**	-0.39**	-0.44**	-0.51**	-0.50**
	SOI	0.08	-0.15	0.06	0.11	0.19	0.33**	0.43**	0.51**	0.43**
	PDO	-0.21*	-0.27**	-0.26**	-0.22*	-0.29**	-0.33**	-0.31**	-0.41**	-0.36**
	EMI	-0.18	-0.07	-0.07	-0.19	-0.20*	-0.23*	-0.25**	-0.27**	-0.39**
	IPO	0.12	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06	-0.06
	NINO3.4	-0.05	-0.07	-0.1	-0.14	-0.18	-0.34**	-0.45**	-0.46**	-0.39**
	IOD	0.05	0.08	0.01	-0.05	0.13	-0.11	-0.11	-0.14	-0.14

* Correlation is significant at 5% level ** Correlation is significant at 1% level

Table 13. Pearson correlations (r) of single lagged climate indices and spring streamflow SNSW

		Lagged Months								
Stations	**Indices**	**Dec$_{n-}$**	**Jan$_n$**	**Feb$_n$**	**Mar$_n$**	**Apr$_n$**	**May$_n$**	**June$_n$**	**July$_n$**	**Aug$_n$**
410024	**NINO3**	0	-0.02	-0.11	-0.1	-0.1	-0.23*	-0.32**	-0.31**	-0.31**
	NINO4	-0.09	-0.07	-0.08	-0.15	-0.19	-0.29**	-0.38**	-0.47**	-0.49**
	SOI	0.02	-0.14	-0.11	0.11	0.17	0.36**	0.43**	0.48**	0.41**
	PDO	-0.21*	-0.23*	-0.18	-0.19*	-0.21*	-0.16	-0.18	-0.19*	-0.21*
	EMI	-0.05	-0.03	0.04	-0.06	-0.07	-0.08	-0.14	-0.26**	-0.32**
	IPO	-0.14	-0.23*	-0.23*	-0.23*	-0.23*	-0.23*	-0.23*	-0.23*	-0.23*
	NINO3.4	-0.02	-0.03	-0.06	-0.09	-0.11	-0.27**	-0.40**	-0.42**	-0.40**
	IOD	0.03	-0.06	0.01	-0.09	0.03	-0.14	-0.32**	-0.27**	-0.32**
410700	**NINO3**	-0.03	-0.04	-0.13	-0.14	-0.12	-0.18	-0.20*	-0.41**	-0.28**
	NINO4	-0.08	-0.06	-0.05	-0.11	-0.15	-0.22*	-0.30**	0.53**	-0.42**
	SOI	0.11	-0.11	-0.06	0.11	0.20*	0.37**	0.43**	-0.18	0.51**
	PDO	-0.29**	-0.31**	-0.30**	-0.32**	-0.26**	-0.19	-0.16	-0.17	-0.24*
	EMI	0.04	0.05	0.14	0.09	0.02	0.01	-0.07	-0.15	-0.19
	IPO	-0.06	-0.15	-0.15	-0.15	-0.15	-0.15	-0.15	-0.35**	-0.15
	NINO3.4	-0.03	-0.02	-0.06	-0.08	-0.1	-0.19*	-0.25**	-0.34**	-0.33**
	IOD	-0.13	-0.11	-0.08	-0.09	-0.03	-0.21*	-0.22*	-0.29**	-0.31**
410033	**NINO3**	-0.13	-0.1	-0.13	-0.14	-0.12	-0.18	-0.20*	-0.27**	-0.22*
	NINO4	-0.15	-0.15	-0.05	-0.11	-0.15	-0.22*	-0.30**	-0.41**	-0.33**
	SOI	0.15	0	-0.06	0.11	0.20*	0.37**	0.43**	0.53**	0.2
	PDO	-0.23*	-0.26**	-0.30**	-0.32**	-0.26**	-0.19	-0.16	-0.18	-0.24*
	EMI	0.07	0.03	0.14	0.09	0.02	0.01	-0.07	-0.17	-0.05
	IPO	-0.14	-0.15	-0.15	-0.15	-0.15	-0.15	-0.15	-0.15	-0.15
	NINO3.4	-0.16	-0.07	-0.06	-0.08	-0.1	-0.19*	-0.25**	-0.35**	-0.26**
	IOD	-0.33**	-0.2	-0.08	-0.09	-0.03	-0.21*	-0.22*	-0.34**	-0.36**
410004	**NINO3**	-0.22*	-0.20*	-0.29**	-0.25**	-0.23*	-0.23*	-0.32**	-0.26**	-0.24**
	NINO4	-0.25**	-0.22*	-0.22*	-0.26**	-0.26**	-0.27**	-0.38**	-0.41**	-0.40**
	SOI	0.18	0.05	0.04	0.17	0.19*	0.31**	0.43**	0.43**	0.32**
	PDO	-0.26**	-0.26**	-0.28**	-0.31**	-0.27**	-0.22*	-0.18	-0.22*	-0.21*
	EMI	-0.09	-0.09	-0.02	-0.1	-0.06	-0.03	-0.14	-0.20*	-0.19
	IPO	-0.13	-0.25**	-0.25**	-0.25**	-0.25**	-0.25**	-0.23*	-0.25**	-0.25**
	NINO3.4	-0.23*	-0.20*	-0.24*	-0.24**	-0.24**	-0.26**	-0.40**	-0.34**	-0.29**
	IOD	-0.07	-0.02	0	-0.06	0	-0.19*	-0.21*	-0.25**	-0.25**

* Correlation is significant at 5% level ** Correlation is significant at 1% level

Table 14. Pearson correlations (r) of single lagged climate indices and spring streamflow WNSW

		Lagged Months								
Stations	**Indices**	**Dec$_{n-}$**	**Jan$_n$**	**Feb$_n$**	**Mar$_n$**	**Apr$_n$**	**May$_n$**	**June$_n$**	**July$_n$**	**Aug$_n$**
409005	**NINO3**	-0.03	-0.01	-0.11	-0.1	-0.1	-0.19	-0.22*	-0.28**	-0.30**
	NINO4	-0.08	-0.04	-0.07	-0.13	-0.17	-0.23*	-0.27**	-0.41**	-0.38**
	SOI	0.08	-0.09	-0.08	0.13	0.22*	0.30**	0.44**	0.46**	0.36**
	PDO	-0.24**	-0.25**	-0.26**	-0.28**	-0.20*	-0.08	-0.06	-0.08	-0.18
	EMI	0.05	0.06	0.08	0.02	-0.03	-0.02	-0.02	-0.14	-0.13
	IPO	-0.05	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09	-0.09
	NINO3.4	-0.03	0.01	-0.06	-0.09	-0.11	-0.20*	-0.26**	-0.35**	-0.33**
	IOD	-0.15	-0.14	-0.07	-0.09	-0.11	-0.18	-0.30**	-0.44**	-0.46**
422002	**NINO3**	0.17	0.15	0.08	0.06	-0.05	-0.15	-0.29**	-0.32**	-0.28**
	NINO4	0.02	0.06	-0.02	-0.1	-0.16	-0.21*	-0.32**	-0.43**	-0.46**
	SOI	-0.05	-0.19	-0.15	-0.09	0.01	0.34**	0.44**	0.48**	0.37**
	PDO	-0.08	-0.16	-0.14	-0.13	-0.20*	-0.26**	-0.26**	-0.32**	-0.28**
	EMI	-0.12	0.01	-0.02	-0.15	-0.15	-0.17	-0.24*	-0.31**	-0.41**
	IPO	0.06	-0.07	-0.07	-0.07	-0.07	-0.07	-0.07	-0.07	-0.07
	NINO3.4	0.12	0.13	0.06	0.01	-0.09	-0.21*	-0.37**	-0.43**	-0.39**
	IOD	0.25**	0.13	0.03	-0.08	0.07	-0.05	-0.11	-0.19*	-0.23*

* Correlation is significant at 5% level ** Correlation is significant at 1% level

Table 15. Pearson correlations (r) of single lagged climate indices and spring streamflow CWNSW

		Lagged Months								
Stations	**Indices**	**Dec$_{n-}$**	**Jan$_n$**	**Feb$_n$**	**Mar$_n$**	**Apr$_n$**	**May$_n$**	**June$_n$**	**July$_n$**	**Aug$_n$**
409002	**NINO3**	-0.15	-0.12	-0.20*	-0.18	-0.15	-0.21*	-0.22*	-0.27**	-0.26**
	NINO4	-0.17	-0.13	-0.15	-0.18	-0.18	-0.21*	-0.23*	-0.38**	-0.35**
	SOI	0.12	-0.02	0.05	0.20*	0.18	0.23*	0.38**	0.44**	0.33**
	PDO	-0.24**	-0.24*	-0.27**	-0.31**	-0.25**	-0.13	-0.15	-0.14	-0.18
	EMI	-0.01	0	0.02	-0.04	-0.01	0.02	0.03	-0.13	-0.12
	IPO	-0.12	-0.21*	-0.21*	-0.21*	-0.21*	-0.21*	-0.21*	-0.21*	-0.21*
	NINO3.4	-0.16	-0.1	-0.16	-0.18	-0.16	-0.22*	-0.25**	-0.33**	-0.29**
	IOD	-0.08	-0.09	-0.02	-0.02	0.01	-0.1	-0.22*	-0.35**	-0.40**
410001	**NINO3**	-0.20*	-0.19	-0.29**	-0.25**	-0.21*	-0.21*	-0.23*	-0.26**	-0.23*
	NINO4	-0.24*	-0.20*	-0.21*	-0.25**	-0.25**	-0.27**	-0.33**	-0.42**	-0.40**
	SOI	0.17	0.04	0.02	0.16	0.20*	0.32**	0.38**	0.45**	0.35**
	PDO	-0.25**	-0.25**	-0.27**	-0.29**	-0.25**	-0.20*	-0.22*	-0.20*	-0.19*
	EMI	-0.1	-0.09	-0.01	-0.09	-0.07	-0.05	-0.1	-0.21*	-0.21*
	IPO	-0.12	-0.24*	-0.24*	-0.24*	-0.24*	-0.24*	-0.24*	-0.24*	-0.24*
	NINO3.4	-0.22*	-0.19	-0.23*	-0.23*	-0.22*	-0.25**	-0.30**	-0.35**	-0.29**
	IOD	-0.05	0	0.01	-0.05	0.02	-0.17	-0.21*	-0.25**	-0.25**
412002	**NINO3**	-0.20*	-0.21*	-0.25**	-0.19	-0.19	-0.19	-0.23*	-0.18	-0.29**
	NINO4	-0.24**	-0.21*	-0.20*	-0.25**	-0.27**	-0.28**	-0.33**	-0.36**	-0.50**
	SOI	0.15	0.04	0.09	0.17	0.21*	0.27**	0.29**	0.32**	0.43**
	PDO	-0.12	-0.16	-0.18	-0.20*	-0.21*	-0.22*	-0.21*	-0.20*	-0.36**
	EMI	-0.12	-0.11	-0.04	-0.14	-0.12	-0.09	-0.11	-0.22*	-0.39**
	IPO	-0.12	-0.23*	-0.23*	-0.23*	-0.23*	-0.23*	-0.23*	-0.23*	-0.06
	NINO3.4	-0.22*	-0.20*	-0.22*	-0.21*	-0.22*	-0.24**	-0.31**	-0.27**	-0.39**
	IOD	-0.04	0.02	-0.02	-0.17	-0.08	-0.20*	-0.17	-0.11	-0.14

* Correlation is significant at 5% level ** Correlation is significant at 1% level

Chapter 5
Multiple Linear Regression Analysis

5.1 Introduction

Multiple Linear Regression (MLR) modelling is one of the simplest and most commonly used techniques in forecasting hydrological and atmospheric variables. MLR technique was used by many studies for predicting flood and rainfall in different parts of the world (He et al. 2014; Nicholson 2014; Mekanik et al. 2013; Latt et al. 2014; Chavoshi et al. 2013). In the study of Rossel and Cadier (2009) MLR modelling was applied to predict monthly rainfall in Ecuador. They were succeeded to explain 60-82% of the monthly precipitation variance with their developed MLR models. MLR models were effectively used to forecast Indian summer monsoon rainfall using SST anomalies in the Indian Ocean as predictors (Sadhuram and Murthy, 2008). Ihara et al. (2007) found the combination of NINO3 and the zonal wind anomalies over the equatorial Indian Ocean are good predictors of rainfall while they were trying to explore the relationship between ENSO and Indian Ocean indices with Indian summer monsoon rainfall with the help of MLR method. In the current study, MLR modelling will be used as the benchmark analysis for the GEP analysis, which is the major focus of the study as a forecasting tool to predict long-term streamflow.

5.2 Methodology

There are several engineering applications for exploring relationships between two or more parameters. The regression analysis model is one of the popular statistical approaches and is highly recommended for this kind of analysis. The most commonly used form of linear regression is MLR analysis. MLR models establish a statistical relationship between two or more explanatory variables and a response variable and provide linear equation as output which represents the significant correlation among the variables. In every equation, the value of every independent variable (X) is associated with the value of the dependent (Y) variable. In many studies, climate forecasting has been undertaken using the MLR model, due to the fact that this model comprises many regressors to deal with the time series database.

Multiple regression models can be presented by the following equation:

$$Y = a + b_1X_1 + b_2X_2 + c \qquad (5\text{-}1)$$

where, Y refers to the dependent variable (i.e., spring streamflow for this study), X_1 and X_2 are two selected independent variables (e.g. NINO3.4, EMI, IOD, PDO or IPO for this study), b_1 and b_2 are the coefficients of two independent variables, a is constant and c is intercept or the error.

In the present study, to evaluate the goodness-of-fit of the models, F-test was used to verify the statistical significance of the overall fit. The next statistical criterion that needs to be satisfied while developing an MLR model is the evaluation of the statistical significance of the individual parameters of the model.

To check the existence of multicollinearity among the predictors is the key stage of MLR modelling. It occurs when predictors are highly correlated, a small change in the data or the model results in the remarkable change in parameter estimation. The Variance Inflation Factor (VIF) is used to ascertain the multicollinearity among the predictors. In order to verify multicollinearity among the predictors, tolerance (T) and VIF are used,

$$tolerance = 1 - R^2 \qquad (5\text{-}2)$$

$$VIF = \frac{1}{tolerance} \qquad (5\text{-}3)$$

Where, R^2 is the coefficient of multiple determinations:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \qquad (5\text{-}4)$$

Where SST is the total sum of squares, SSR is the regression sum of squares, and SSE is the error sum of squares. According to Quan et al. (2006), a tolerance of less than 0.20–0.10 or a VIF greater than 5–10 indicates a multicollinearity problem.

In order to ensure the independence of residuals error of the model Durbin-Watson (DW) test is performed, which assesses the serial correlation between errors. DW parameter has a range of 0 to 4; a value of less than 1 or greater than 3 is certainly a matter of concern (Field 2009).

The performance of the developed MLR models has been assessed by several statistical performance measures which are widely used for the evaluation of regression models. Statistical measures namely mean square error (RMSE), mean absolute error (MAE), Pearson correlation coefficient (r) and Willmott index of agreement (d) are exclusively chosen for this study. 'd' is defined as follows:

$$d = 1 - \frac{[\sum |\hat{y}_i - x_i|^2]}{[\sum (|\hat{y}_i - \bar{x}_1| + |x_i - \bar{x}_1|)^2]} \qquad (5\text{-}5)$$

Where, $\hat{y}_i$ refers to the predicted value corresponding to i^{th} observation and x_i refers to i^{th} value of observation. The closer the 'd' value to 1, the better the model fits the observations. The development of Multiple Linear Regression models and all the relevant statistical calculations are performed using the "R Studio 3.3.1" software.

5.3 Results and Discussion of MLR Analysis

Various models with different lagged months' indices are analyzed for all twelve stations in order to find out the best forecasting model for each of the four study regions. Multiple Regression Set for all the stations can be found in Table 16, 17, 18 and 19. For all the stations, the best models with lower errors while satisfying the statistical limits are selected. F-test is performed to evaluate the best model that fits the population of the sample data while the t-test is conducted to identify the significance of the individual parameters. The best model for each station with their regression coefficients, variance inflation factor (VIF), Durbin-Watson statistics (DW) and the Pearson correlation (r) are presented in Table 20.

Table 16. Multiple Regression Set for each station of NNSW

Region (Northern)	SINGLETON	COGGAN	NORTH CUERINDI
PDO- NINO3.4	Aug-Aug, Jul-Jul, Jun-Jun, May-May, Apr-Jun, Apr-May, Mar- May, Mar- Jun, Feb- May, Jan- May, Dec*-May,	Aug-Aug, Jul-Jul, Jun-Jun, May-May, Apr-Jun, Apr-May, Mar- May, Mar- Jun, Feb- May, Jan- May, Dec*-May,	Aug-Aug, Jul-Jul, Jun-Jun, May-May, Apr-Jun, Apr- May, Mar-May, Mar- Jun, Feb- May, Jan-May, Dec*-May,
IOD- NINO3.4	Jun-Jun	Jun-Jun	--------
EMI- NINO3.4	---------	--------	Aug- Aug, Jul- Jul, Jun-Jun, May-May, Apr- May

"*" denotes the month of the previous year.

Table 17. Multiple Regression Set for each station of SNSW

Region (Southern)	GUNDAGAI	WEE JASPER (KASHMIR)	MITTAGANG CROSSING	KIOSK
PDO-NINO3.4	Aug-Aug, Jul-Jul, Jun-Jun, May-May, Apr-Apr, May-May, Mar-Mar, Jan-Feb, Dec*-Feb	Aug-Aug, Jul-Jul, Apr-Jun, Apr- May, Mar-May, Feb- May, Jan-May, Dec*-May,	Aug-Aug, Aug-Jul, Aug-Jun, Apr- May, Mar-May, Feb- May, Jan-May, Dec*-May	Aug-Aug, Aug-Jul, Apr-Jun, Apr- May, Mar- May, Feb-May, Jan- May, Dec*-May
IOD-NINO3.4	Aug-Aug, Jul-Jul, Jun-Jun, May-May, May-Apr, May-Mar, May-Feb, May-Jan	Aug-Aug, Jul-Jul, Jun-Jun, Jun-May	Aug-Aug, Jul-Jul, Jun-Jun, May-May, Apr-May, Mar- May	Aug-Aug, Jul-Jul, Jun-Jun, May-May
IPO-NINO3.4	Aug-Aug, Jul-Jul, Jun-Jun, May-May, Apr-Apr, May-May, Mar-Mar	Aug-Aug, Jul-Jul, Jun-Jun, May-May, Apr-May, Mar- May, Feb-May, Jan- May	--------	--------
IOD- IPO	Aug-Aug, Jul-Jul, Jun-Jun, May-May, May-Apr, May-Mar, May-Feb, May-Jan	Aug-Aug, Jul-Jul, Jun-Jun, Jun-May, Jun-Apr, Jun-Mar, Jun-Feb, Jun-Jan	--------	--------
EMI-NINO3.4	-------	Aug-Aug, Jul-Jul, Jul-Jun, Jul-May	--------	--------

Table 18. Multiple Regression Set for each station of CWNSW

Region (Central)	COROWA	WAGGA WAGGA	COWRA
PDO- NINO3.4	Apr-Aug, Apr-Jul, Apr-Jun, Apr-May, Mar- May, Feb-May, Jan- May, Dec*-May	Aug-Aug, Jul-Jul, Jun-Jun, May-May, Apr-Apr, May-May, Mar-Mar, Jan-Feb, Dec*-Feb	Aug-Jul, , Jul-Jul, Jun-Jun, May-May, Apr-Apr, May-May, Mar-Mar, Mar-Feb
IOD- NINO3.4	Aug-Aug, Jul-Jul, Jun- Jun, Jun-May	Aug-Aug, Jul-Jul, Jun-Jun, Jun-May, Jun-Apr, Jun-Mar, Jun-Feb	---------
IPO- NINO3.4	Aug-Aug, Jul-Jul, Jun-Jun, May-May, Apr-May, Mar-May, Feb- May, Jan- May	Aug-Aug, Jul-Jul, Jun-Jun, May-May, Apr-Apr, May-May, Mar-Mar, Feb-Feb, Jan-Feb	---------
IOD- IPO	Aug-Aug, Jul-Jul, Jun-Jun, Jun-May, Jun-Apr, Jun-Mar, Jun-Feb, Jun-Jan	Aug-Aug, Jul-Jul, Jun-Jun, Jun-May, Jun-Apr, Jun-Mar, Jun-Feb, Jun-Jan	---------
EMI- NINO3.4	---------	Aug-Aug, Jul-Jul, Jul- Jun, Jul-May, Jul-Apr, Jul-Mar, Jul-Feb	Aug-Aug, Jul-Jul, Jul- Jun, Jul-May, Jul-Apr, Jul-Mar, Jul-Feb,
EMI- IPO	--------	Aug-Aug, Jul-Jul, Jul- Jun, Jul-May, Jul-Apr, Jul-Mar, Jul-Feb, Jul-Jan	---------

Table 19. Multiple Regression Set for each station of WNSW

Region (Western)	BARHAM	BREWARRINA
PDO- NINO3.4	Apr-Aug, Apr-Jul, Apr-Jun, Apr-May, Mar- May, Feb- May, Jan- May, Dec*-May	Aug-Aug, Jul-Jul, Jun-Jun, May-May, Apr-May
IOD- NINO3.4	Aug-Aug, Jul-Jul, Jun- Jun, Jun-May	Aug-Aug, Jul-Jul, Jul- Jun, Jul-May, Dec*-May
EMI- NINO3.4	---------	Aug-Aug, Jul-Jul, Jun-Jun, Jun-May

It is seen from Table 20 that VIF values for the selected models are close to 1, which refers that there is no multi-collinearity problem between the predictors. According to Field (2009), values less than 1 or greater than 3 for the DW test will indicate the presence of serial correlations between the model errors. So, it can be concluded from the results of Table 20 that the DW test of each selected model satisfies the statistical limits, which also establishes the goodness of fit of the models

Table 20. Equations of the best developed MLR models

Region	Station Name	Model	Constant	Coefficient	R	Durbin - Watson	VIF
NNSW	Singleton	PDO_{MARCH} $NINO3.4_{JUNE}$	18.81	-1.81 -13.64	0.41	2.13	1.12
	North Cuerindi	PDO_{JUL} $NINO3.4_{JUL}$	9.58	-1.67 -5.77	0.46	1.47	1.16
	Coggan	PDO_{JULY} $NINO3.4_{JULY}$	1.72	-0.15 -1.11	0.33	1.99	1.17
SNSW	Wee Jasper(Kashmir)	IOD_{JULY-} $NINO3.4_{JULY}$	14.70	-3.15 -4.80	0.42	1.76	1.17
	Kiosk	PDO_{AUG} $NINO3.4_{JULY}$	5.91	-0.54 -2.68	0.45	1.73	1.30
	Mittagang Crossing	PDO_{AUG} $NINO3.4_{JULY}$	13.16	-1.90 -2.54	0.35	1.16	1.30
	Gundagai	IPO_{JULY} $NINO3.4_{JULY}$	158.12	-18.76 -42.94	0.43	1.83	1.05
CNSW	Corowa	IPO_{JUNE} IOD_{JUNE}	262.05	-37.29 -85.21	0.30	1.71	1
	Wagga Wagga	IPO_{JULY} $NINO3.4_{JULY}$	176.47	-20.15 -54.26	0.43	1.91	1.05
	Cowra	PDO_{MAR} $NINO3.4_{FEB}$	39.32	-1.42 -11.11	0.27	1.62	1.44
WNSW	Barham	IOD_{JUNE} $NINO3.4_{JUNE}$	215.13	-67.47 -39.48	0.31	1.80	1.10
	Brewarrina	IOD_{JUL} $NINO3.4_{JUL}$	50.27	-2.74 -32.84	0.40	1.74	1.17

It is evident from the current analysis that the selected models are not only statistically significant but also have the potential to predict the spring streamflow of north-east New South Wales with the highest correlation 0.46 for the North Cuerindi station (Table 20). The model outcomes are really effective as best prediction skills are obtained from models having a lead time of four months, where the contribution of PDO is effective even 9 months before. 85 years' (from 1914 to 1998) of streamflow data was selected for the calibration of the models, while the remaining 17 years' (from 1998 to 2015) data were selected for validation in order to assess the future streamflow predictability of the developed MLR models.

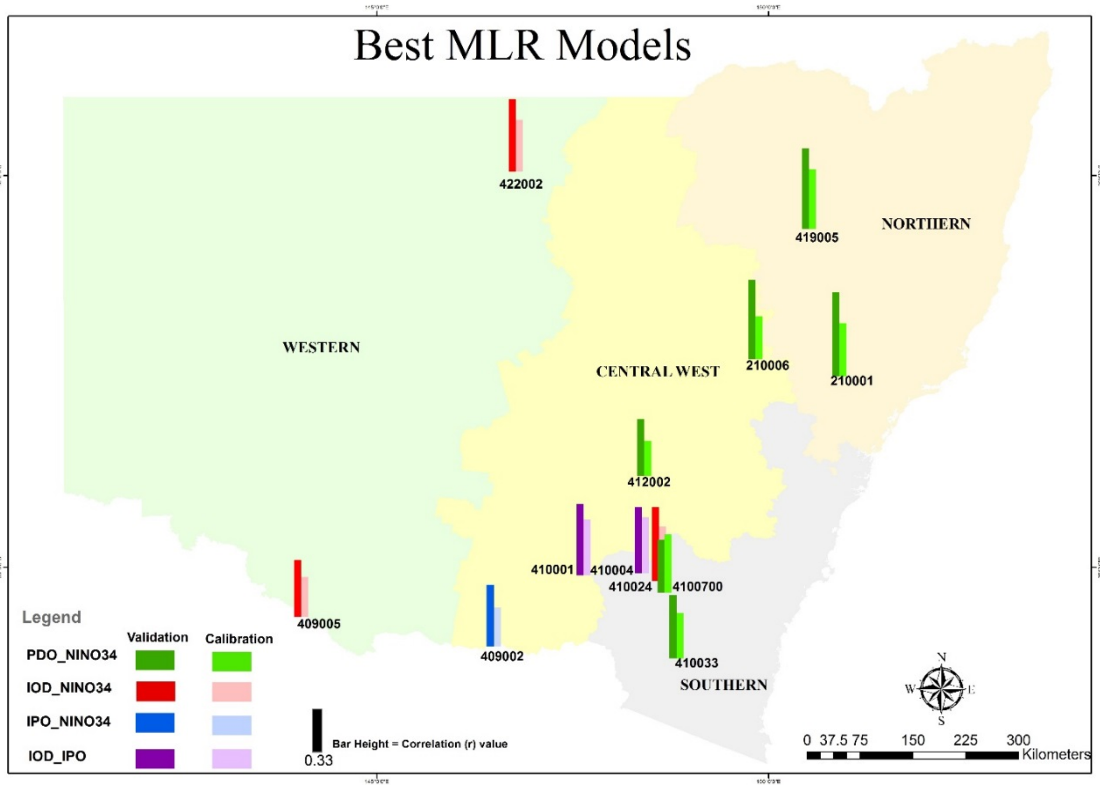


Figure 17. Comparative analysis of the influences of various MLR models on different study regions

The results of the MLR analysis depicts a clear view of the regional variation of influence of combined multiple indices throughout the study area (Figure 18). A good number of models combining PDO and NINO3.4 show statistically significant correlations with spring streamflow for the Northern NSW region, which implies the strong impact of these two indices in the northern part of the state. Thereby, for this region, the best model for forecasting spring streamflow is obtained with the combination of 3 Months lagged NINO3.4 and six months lagged PDO (at Singleton) with a good correlation of 0.41. In the Southern part of the state, significant correlations are obtained for PDO-NINO3.4 and IOD-NINO3.4 combinations, though two stations (at Wee Jasper and Gundagai) of this region are found to have significant correlations for IPO-NINO3.4 and IPO-IOD combinations, which are more similar to the findings of Wagga Wagga station at CNSW. One possible reason for this similarity can be the close geographical locations of these stations. IPO consisting models have significant contribution to forecast streamflow in the southern part of CWNSW. The central part of CWNSW has good correlations for only PDO-NINO3.4 combination. It is evident from the results that the IOD containing combined models show good performance with significant correlation for the southern part of the state. But in the coastal eastern part of the state IOD-NINO3.4 combinations are observed to be less effective, which align with the findings of Pepler et al. (2014), who stated that eastern seaboard rainfall is less influenced by tropical sea surface temperature variability such as the ENSO and IOD than inland because the effect of the IOD opposes ENSO during the cool season. In the western part of the country, IOD-NINO3.4 and PDO-NINO3.4 combined models can be used to forecast spring streamflow.

85 years' (from 1914 to 1998) of streamflow data is selected for the calibration of the models, while the remaining 17 years' (from 1999 to 2015) data were selected for validation in order to assess the future streamflow predictability of the developed MLR models. In order to determine the accuracy of the developed MLR models, validation test is performed. Table 21 shows the performance statistics of RMSE, MAE, index of agreement (d) and Pearson correlation (r) of the best-developed models for the calibration and validation periods. It is clearly evident from Table 21 that there is significant increment (except Kiosk station) of the correlation values from calibration to validation stage, as for example correlation value increased from 0.41 in calibration

stage to 0.65 in validation stage for Singleton station in NNSW. This station also provides the highest correlation in the validation period which is obtained for PDO_{JAN}-$NINO3.4_{MAY}$ combination (r=0.71) showing great improvement as the value was only 0.27 in the calibration period (not shown in Table 21).

Table 21. Performance test of the best MLR models for calibration and validation periods

Region	Station Name	Model	Calibration Period				Validation Period			
			R	RMSE	MAE	d	R	RMSE	MAE	d
NNSW	Singleton	PDO_{MARCH} $NINO3.4_{JUNE}$	0.41	19.23	13.81	0.51	0.65	12.09	10.98	0.70
	North Cuerindi	PDO_{JUL} $NINO3.4_{JUL}$	0.46	8.92	5.98	0.55	0.62	6.92	6.12	0.70
	Coggan	PDO_{JULY} $NINO3.4_{JULY}$	0.33	2.15	1.26	0.38	0.61	1.25	1.16	0.67
SNSW	Wee Jasper(Kashm ir)	IOD_{JULY}- $NINO3.4_{JULY}$	0.45	7.51	6.04	0.53	0.57	5.47	4.39	0.63
	Kiosk	PDO_{AUG} $NINO3.4_{JUL}$	0.45	3.98	3.04	0.57	0.41	4.24	3.74	0.52
	Mittagang Crossing	PDO_{AUG} $NINO3.4_{JUL}$	0.35	8.91	7.41	0.44	0.49	9.73	9.30	0.38
	Gundagai	IPO_{JUL} $NINO3.4_{JUL}$	0.43	71.40	58.00	0.55	0.51	72.37	66.03	0.40
CWNSW	Corowa	IPO_{JUN} IOD_{JUN}	0.30	139.33	114.87	0.40	0.48	131.01	126.17	0.30
	Wagga Wagga	IPO_{JUL} $NINO3.4_{JUL}$	0.43	85.84	69.26	0.55	0.55	80.53	71.44	0.44
	Cowra	PDO_{MAAR} $NINO3.4_{FEB}$	0.35	35.01	28.91	0.37	0.44	28.04	24.75	0.44
WNSW	Barham	IOD_{JUN} $NINO3.4_{JUN}$	0.31	106.05	90.67	0.43	0.44	95.51	82.49	0.37
	Brewarrina	IOD_{JUL} $NINO3.4_{JUL}$	0.40	44.47	34.59	0.49	0.56	38.55	36.04	0.57

The equations of the best developed models for each of the twelve stations are presented in Table 22.

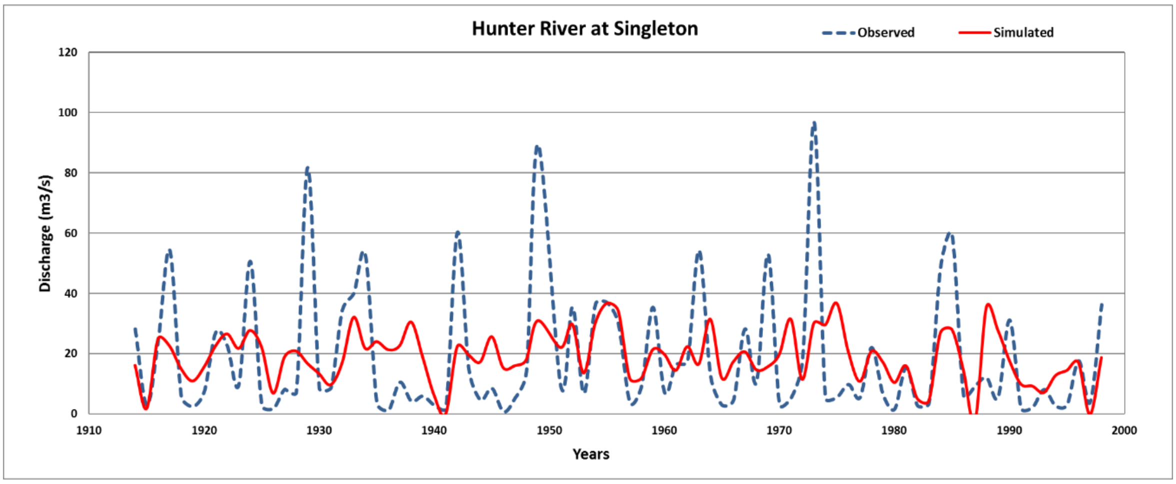
Table 22. Best Developed MLR Models for 12 Stations

Region	Station Name	Best Developed Model
NNSW	Singleton	$Q = 18.81 - 1.81PDO_{Mar} - 13.64NINO3.4_{Jun}$
	North Cuerindi	$Q = 9.58 - 1.67PDO_{Jul} - 5.77NINO3.4_{Jul}$
	Coggan	$Q = 1.72 - 0.15PDO_{Jul} - 1.11NINO3.4_{Jul}$
SNSW	Wee Jasper(Kashmir)	$Q = 14.70 - 3.15IOD_{Jul} - 4.80NINO3.4_{Jul}$
	Kiosk	$Q = 5.91 - 0.54PDO_{Aug} - 2.58NINO3.4_{Jul}$
	Mittagang Crossing	$Q = 13.16 - 1.90PDO_{Aug} - 2.54NINO3.4_{Jul}$
	Gundagai	$Q = 158.12 - 18.76IPO_{Jul} - 42.94NINO3.4_{Jul}$
CWNSW	Corowa	$Q = 262.05 - 37.29IPO_{Jun} - 85.21IOD_{Jun}$
	Wagga Wagga	$Q = 176.47 - 20.15IPO_{Jul} - 54.26NINO3.4_{Jul}$
	Cowra	$Q = 39.32 - 1.42PDO_{Mar} - 11.11NINO3.4_{Feb}$
WNSW	Barham	$Q = 215.13 - 67.47IOD_{Jun} - 39.48NINO3.4_{Jun}$
	Brewarrina	$Q = 50.27 - 2.74IOD_{Jul} - 32.84NINO3.4_{Jul}$

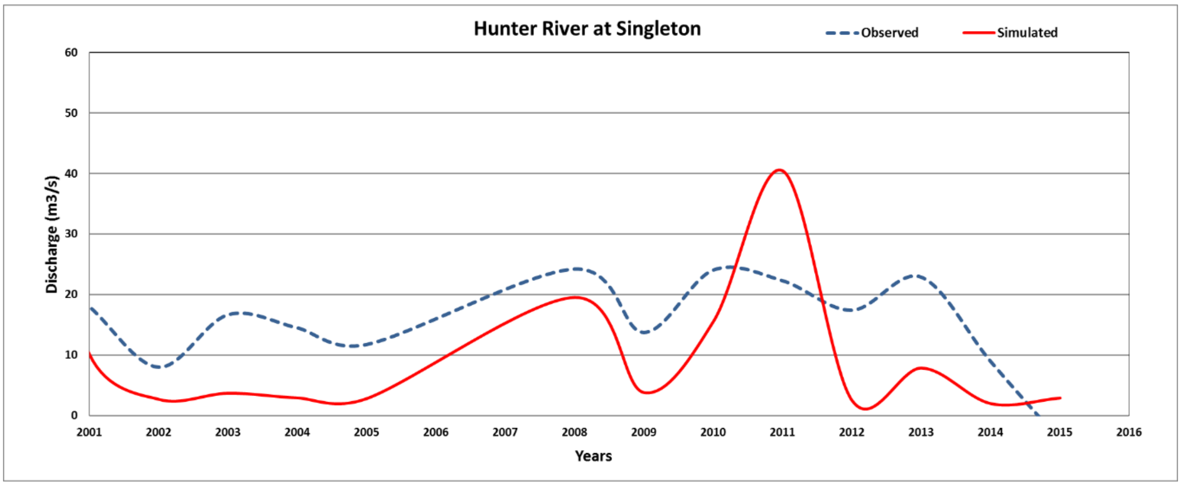
The best predictor models for each of the four study regions are given below:

$$Q_{Singleton} = 18.81 - 1.81PDO_{Mar} - 13.64NINO3.4_{Jun}$$
$$Q_{WeeJasper} = 14.70 - 3.15IOD_{Jul} - 4.80NINO3.4_{Jul}$$
$$Q_{WaggaWagga} = 176.47 - 20.15IPO_{Jul} - 54.26NINO3.4_{Jul}$$
$$Q_{Brewarrina} = 50.27 - 2.74IOD_{Jul} - 32.84NINO3.4_{Jul}$$

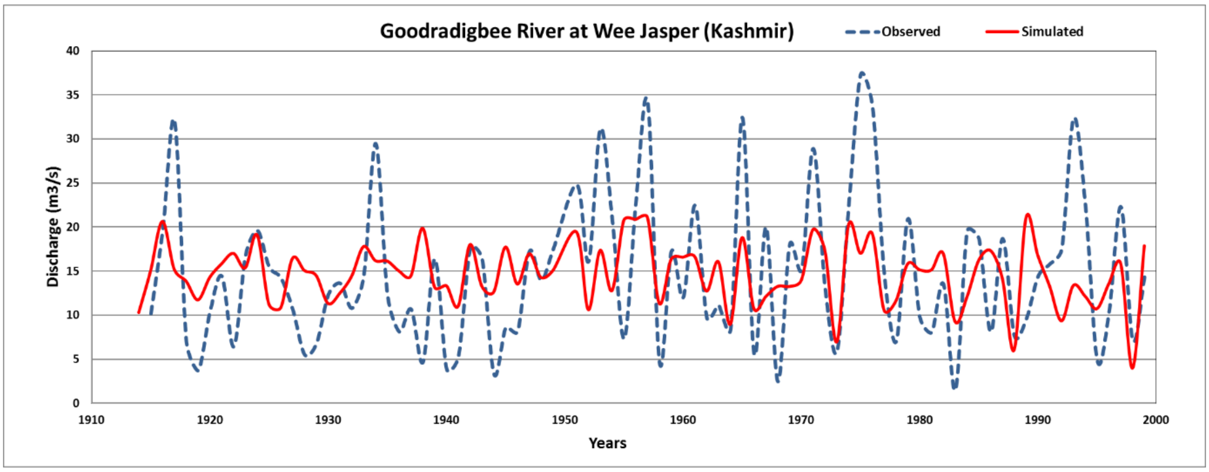
The ability of these four best MLR models to predict future streamflow has been explained through the time series plots of observed and simulated flow in Figure 18.



i(a)

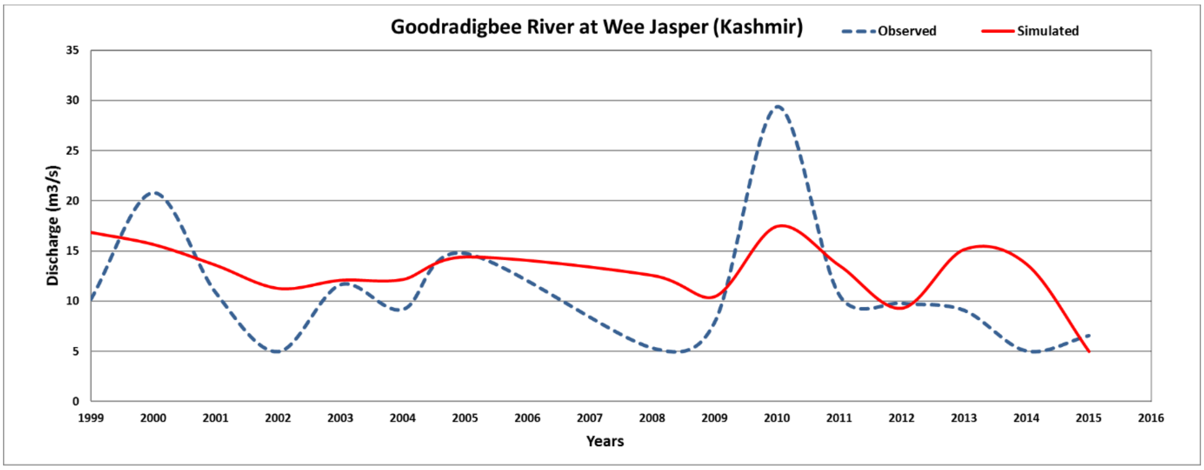


i(b)



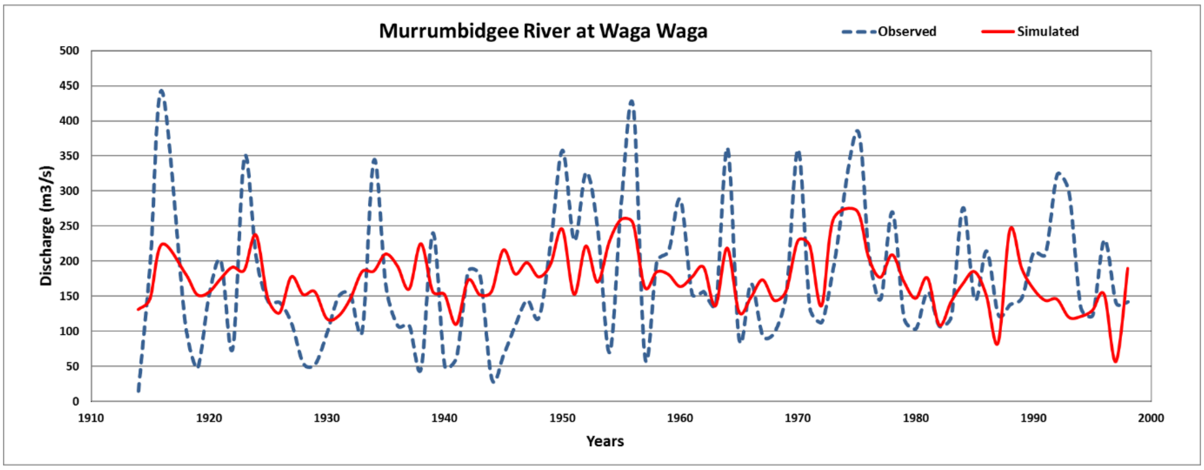
Goodradigbee River at Wee Jasper (Kashmir)
Observed
Simulated
Discharge (m3/s)
Years
0
5
10
15
20
25
30
35
40
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000

ii(a)



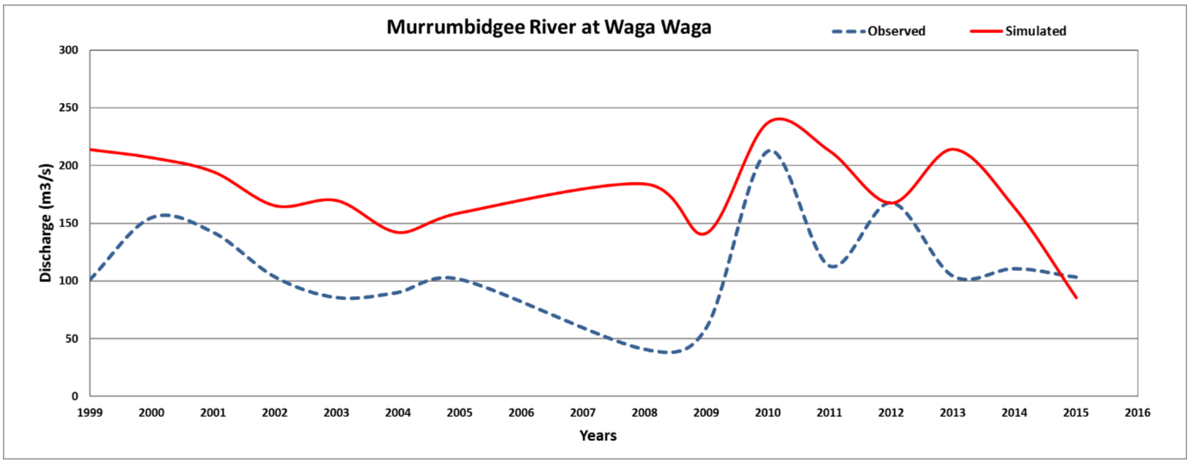
Goodradigbee River at Wee Jasper (Kashmir)
Observed
Simulated
Discharge (m3/s)
Years
0
5
10
15
20
25
30
35
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016

ii(b)

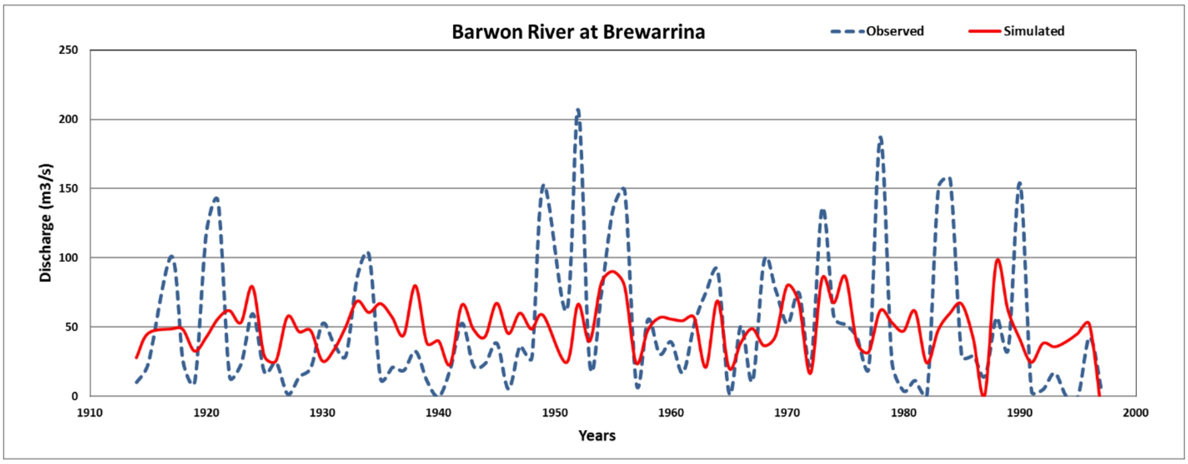


Murrumbidgee River at Waga Waga
Observed
Simulated
Discharge (m3/s)
Years
0
50
100
150
200
250
300
350
400
450
500
1910
1920
1930
1940
1950
1960
1970
1980
1990
2000

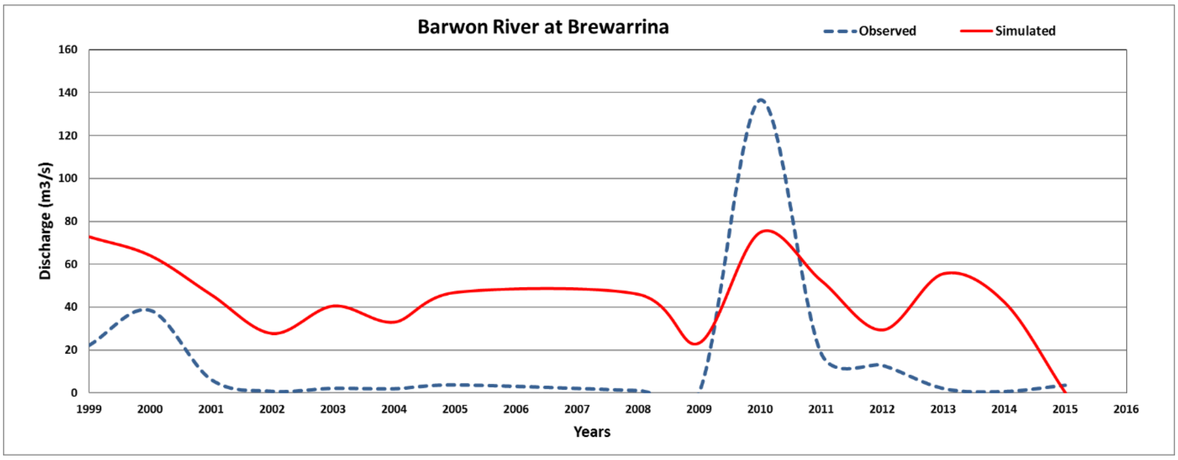
iii(a)



iii(b)



iv(a)



iv(b)

Figure 18. Comparison between the observed and simulated streamflow during the (a)calibration (1914-1998) and (b)validation (1999-2015) periods for (i) Singleton (NNSW), (ii) Wee Jasper (SNSW), (iii) Wagga Wagga (CWNSW), (iv) Brewarrina (WNSW) stations

In Figure 18 some over-estimation of the models can be observed during the validation stage which may be resulted from the “millennium drought” (Bond et al. 2008) periods that occurred from 1994 to 2010 over the continent. It was explored by (Kiem et al. 2009) that a combination of climate drivers in the Pacific Ocean (ENSO, PDO), IOD and SAM were responsible for the past three droughts in south-east Australia; the ‘Federation’ drought (1895-1902), the ‘World War II’ drought (1937-1945) and the ‘Big Dry’ (1994-2010). A model based on only two climate indices (NINO3.4 and PDO) is not likely to replicate an unusual phenomenon like “millennium drought”. This can also be the reason for not reflecting the unusually high rainfall in some years of NSW in the time series plots, where the difference between the simulated and observed flow is found to be high. Another reason is that some other climate indices might have been more influential at that time rather than the selected indices in this study. However, to get the best predictor model a few unusual events which seem to be outliers in a box-plot analysis, are removed from the calibration and validation periods.

The capability of the developed models for forecasting spring streamflow with higher accuracy has been ensured as the values of RMSE, MAE and d in the validation period show good agreement with the calibration period. The index of agreement (d) for both calibration and validation period is close to 0.5, which ensures the better forecasting ability of the models. Significant increment of the Pearson correlation values has explained that the combined models have greater skills for predicting streamflow than the single lagged indices. For instance, in case of Singleton station (Table 12), while the single NINO3.4 model with three months’ lead time has a correlation r= -0.43, the predictability is significantly enhanced by the contribution of six months’ lagged PDO to a correlation $r_{validation}$= 0.65 (Table 21).

5.4 Performance based comparative analysis of MLR models

A comparison of present study with the previous studies on forecasting streamflow is shown in Table 23. It is evident from the comparison that the combined MLR models of this study showed better agreement with streamflow compared to the single lagged climate index models of the current study and the previous research studies as well. However, the single lagged climate index models of current study outperformed the models developed by past research studies using single climate mode to forecast streamflow of Australia.

Table 23. Comparison with the previous studies based on the highest correlations between indices and spring streamflow for South-East Australia

INDICES	Kirono et al. (2010)	Chiew et al. (2003)	CURRENT STUDY	
			Single lagged correlation	MLR correlation
Nino3.4	-	-	-0.43[iv]	0.65[viii]
PDO	-	-	-0.41[v]	
Nino3	0.35[i]	-	0.36[vi]	
SOI	0.36[ii]	0.51[iii]	0.51[vii]	

i) 8 months lagged Nino3 ii) 12 months lagged SOI iii) Winter SOI

iv) 3 months lagged Nino3.4 v) 2 months lagged PDO vi) 3 months lagged Nino3.4

vii) 2 months lagged SOI viii) PDO_{MAR}& $Nino3.4_{JUN}$

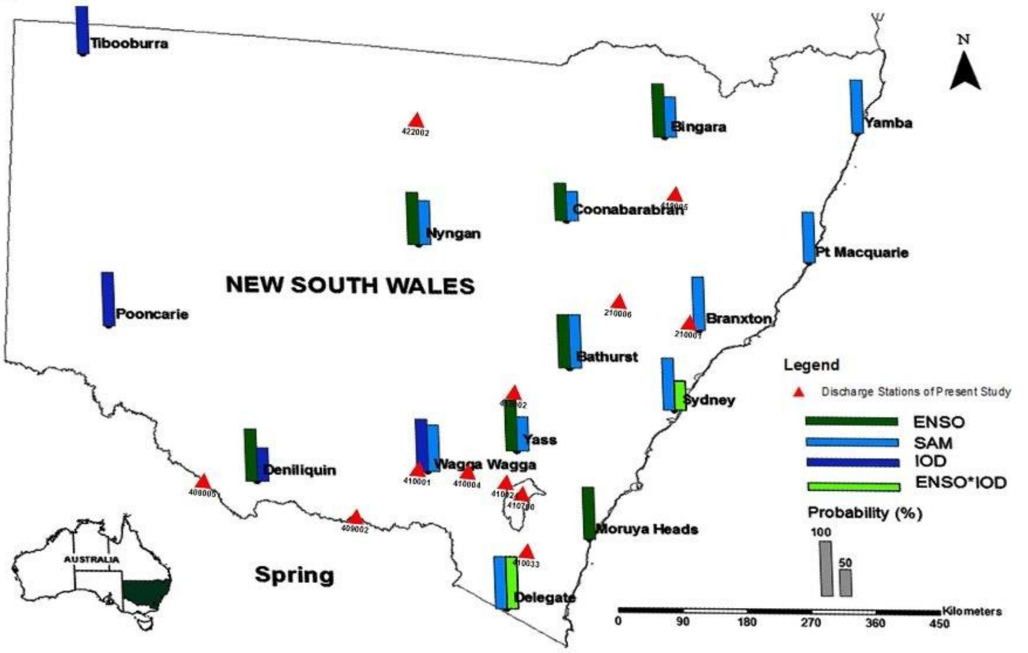


Figure 19. Influence of climate drivers on spring rainfall of NSW by Duc et al (2017) including the discharge stations of the current study

The variation of influences of different climate indices on different study regions of NSW is comparable to the recent study outcomes of Duc et al. (2017) who has done research on the association of climate indices with NSW rainfall using Bayesian Model Averaging (Figure 19). In his study, he has selected 15 sites across NSW, of which the results of the nearby stations which are within 161 kilometers of the current study's streamflow stations have been chosen for the current comparison. In the present study, PDO-NINO3.4 interactions influence spring streamflow almost across the whole state, which align with the findings of Duc et al. (2017) who explored that single IPO (PDO) cannot impact NSW rainfall significantly, but its association with ENSO is significantly influential on the rainfall of almost whole NSW.

In Wagga Wagga, evidence of strong IOD influence has been found in the study of Duc et al. (2017) which is consistent with current study as IOD combined models performed significantly near this area. Similar outcomes have been obtained for Mittagang Crossing station, where Duc et al. (2017) found ENSO_IOD combined impact on spring rainfall to be very strong (Posterior probability= 1) and the current study has significant correlation (r=0.42) with the same combination of indices. As a whole, Duc et al. (2017) found ENSO, SAM and IOD to be influential on spring rainfall of NSW whereas current study includes the influence of IPO and PDO with significant correlation values for many locations.

The only difference between Duc et al. (2017) and the present study is that SAM (Southern Annular Mode) has not been included in the current analysis, whereas Duc et al. (2017) found a strong influence of SAM on spring rainfall of almost all parts of NSW. At the preliminary stage of the present research, single concurrent correlation analysis was performed between SAM and seasonal streamflow of NSW, but no significant outcome was obtained for any season. Based on these results, SAM was not considered for further analysis in the present study. One reason for the poor results of SAM could be a shorter length of data availability (1957-present).

Duc et al. (2017) revealed that the greatest number of indices and their combined models have strong influences on the winter rainfall of NSW, whereas the current study found the greatest number of influential indices for spring streamflow. This satisfies the fact that winter is the main season for rainfall in Australia and streamflow is resulted from rainfall. So, the indices that impact winter rainfall have the potential for influencing spring streamflow.

5.5 Summary of MLR Analysis

In the current research, the MLR method was applied with a view to exploring the potential skills of combined multiple climate indicators to forecast the spring streamflow of NSW regions with a longer lead time than the usual practice. Before performing MLR analysis, first single correlation analyses were performed to identify potential climate predictors for the region. Through a single correlation analysis, several indices

(NINO3.4, IOD, EMI, PDO, and IPO) were found to have strong effects on spring streamflow of NSW with a lagged time of a maximum of three months. Some indices were found to give significant correlations with a lagged time of more than two months; however, in general, the correlation values decrease with the increase of lagged months. This study, through findings of five effective climate indices for the region, opened an opportunity to study with more than two indices which no one has ever done for this region. For the current study, to achieve better correlations (prediction capability) different combinations of two (out of the five significant) indices were tested in the MLR analysis. It was found that the same combination of indices did not turn out to be best for all the stations/regions, which is reasonable as the current study was dealing with a large region and the further the distance from a particular station the greater the likelihood of being influenced by other indexes. Also, the combined best models' lagged indices are not necessarily from the same month. In general, among the best combined dual indices, NINO3.4 is found to be significant for all the stations except one (Corowa), PDO is more significant towards the north-eastern and south-eastern coastal region, IPO is more significant towards the central-south, whereas IOD is more significant towards the west of NSW. The best correlation is obtained for Singleton station in NNSW for the PDO–NINO3.4 combined model with a correlation of 0.65 (in the validation period) for the prediction of spring streamflow with two months' lagged period. It is noteworthy that every time the combined model outperformed the models considering a single climate variable in terms of Pearson correlation (r), it was evidence of better predictive skills of the MLR models.

For this study, selections of the best models were based on the significant correlation values in both calibration and validation stages. However, while looking at the time series comparisons between the observed and simulated streamflow values, it is found that the developed models are unable to capture some unusual events like severe droughts or high floods. A simple MLR model consisting of only two climate indices is not expected to capture the complex relationships between streamflow and climate drivers very well and thus is not anticipated to provide a very good match with observed values. Moreover, in fact, rainfall and streamflow are also influenced by some other local and/or regional factors (i.e., temperature, humidity, wind speed, soil moisture, etc.), which are not possible to consider in such regression models. Thus, the next stage of the current study will include other non-linear techniques (i.e., MNLR and GEP) as some researchers (Mekanik et al. 2013; Abbot and Marohasy 2015; Rasel et al. 2016) successfully explained the non-linear relationship between rainfall and climate drivers using this technique, although they could not provide any output model which could be useful to stakeholders. Since the relationship between streamflow and remote climate drivers is likely to be non-linear (Piechota et al. 1998), a non-linear model is expected to give better results than a linear model. Also, an extension of the current research can be to explore the effectiveness of incorporating more than two indices in a MLR model.

Nevertheless, the developed MLR models have the potential to provide indications on the possibility of getting increased or decreased amounts of streamflow and expected magnitude in the future season. Currently, water users in Australia get the seasonal predictions of streamflow just at the beginning of the season, which does not give them enough time for prudent decision making. Moreover, those predictions are stochastic, i.e., the users do not get any estimation of expected magnitude. As a result, the farmers get inadequate information at the start of the cropping year. The pressure has been highly increased on water resource availability since the recent eastern Australian severe drought (2002-2007), (Murphy and Timbal 2008). According to Nicholls (2006) over the latter half of the 20th century, a huge decline in rainfall has been observed in eastern Australia which has been a matter of great concern to the scientist and researchers considering the large population and economic importance of this region. Therefore, Streamflow forecast several months ahead will be invaluable to the people associated with water management, allocation, irrigation, agricultural production etc. The developed MLR models in this study are expected to provide water users and planners with some insight which will enable them to take tactical cropping decisions three months in advance, and that will help them to avoid huge economic loss during severe droughts in the dry season and massive floods in wet seasons.

Chapter 6
Multiple Non-Linear Regression Analysis

6.1 Introduction

The primary source of water is mainly water available in the catchment in the form of groundwater and the rainfall obtained in the catchment. The limit of extraction of groundwater makes it an unreliable source making rainfall the best suitable source of water. So, predicting the amount of water usually depends upon predicting the rainfall associated in that region. Likewise, the issues of water resources in Australia is also closely related to the distribution of rainfall, and this governs the pattern of development, habitation and agriculture for the country (Risbey, 2011). As discussed in Chapter 3, water allocation is one of the biggest problems in Australia, resulted from mostly the unpredictable weather pattern of the country.

Rainfall is one of the most important factors in the hydrological cycle. It has a considerable effect on nature as well as human lives. The availability of the water on the earth surface depends on the spatial and temporal distribution of rainfall. Since water is required for the human being for their daily activities, it has become the most imperial thing for the human being. Along with daily activities, water is also required for agricultural activities. Since agricultural production is directly dependent on the availability of the water. The main source of the water for agriculture is underground water and rainfall. Since the underground water extraction is also dependable of the rainfall, it is essential to predict the rainfall for the planning of agriculture and flood mitigation. But it is hard to predict the exact amount of rainfall to the researchers. So, the prediction of the seasonal rainfall has become plausible to the researcher.

Australia is the smallest continent in the world and the sixth-largest country. In 2017, Agriculture contributed to 3% of the total Gross Domestic Production (GDP), producing 93% of the food consumed in Australia (National Farmers Federation, 2017). During 2002-2003, drought in Australia brought a reduction in Gross Domestic Production (GDP) reduction by 1.6% (Horridge et al., 2005). The main source of water for farming is streamflow which is directly associated with the amount of rainfall in the region. Farmers are more dependable on water from streamflow for the agricultural purpose. Australia has extensive topographic variations due to which there are high climatic variabilities making it hard to predict the streamflow in Australia. So, the farmers had to go through many problems in managing the water supply for agriculture. However, for the management of proactive risk management like drought management, the seasonal streamflow forecasts play a vital role. Government agencies, researchers, concerned parties have been trying hard to predict the amount of discharge in the streams using various methods in order to be prepared and plan for the effective utilisation of the water that the country might receive in any particular season. The prediction of rainfall, either monthly or seasonal, is essential for agricultural planning and flood mitigation strategies. However, accurate prediction of seasonal rainfall remains elusive to the scientists. Therefore, seasonal rainfall forecasting becomes plausible amongst the hydrologic researchers around the globe (Tennant and Hewitson 2002 and Frias et al. 2018).

Seasonal forecasting can be classified into two broad categories: the statistical approach and the dynamic approach. In the statistical approach, the statistical relationships between the predictors and the predictands are investigated (Jenicek et al. 2016). In the dynamic approach, seasonal meteorological estimates are used to build a hydrological model. However, there are methodological implications in using meteorological inputs in the current hydrological models (Crochemore et al. 2016). The climate model produces the outputs based on coarse grid scales, which has the potential to capture forecasting uncertainties and hence lead to bias. Furthermore, the data requirements of the dynamic models hinder the application of the modelling type. As a result, the statistical approach drew considerable attention to the practical users of the prediction models.

Streamflow is mainly dependable of catchment condition and climatic variables. Since the catchment condition is more complicated for the forecasting models, climatic variables are used for the prediction of the streamflow. With the advancement of technology, various weather prediction models are being developed for the prediction of rainfall. However, the skill for prediction of rainfall is more reliable for short-term prediction than seasonal rainfall prediction due to the complexity of rainfall phenomenon. The short-range forecasting like for a day or week can be predicted by atmospheric motion which is dependable of the initial condition but relatively intensive to the boundary. In case of the seasonal rainfall forecasting scenario is just reversed, atmospheric motion is observed by a lower boundary condition.

Long-term prediction of seasonal rainfall has the potential to help in the decision-making process for planning appropriate watershed management strategies (Crochemore et al. 2016). Moreover, the advanced prediction of rainfall can provide information to adopt the consequences of climate change (Winsemius et al. 2014). As a result, the urge for the application of seasonal rainfall forecasting is increasing day by day. Therefore, seasonal forecasting is routinely performed by different research institutes, to have a better understanding of climate change throughout the world. However, there exist limiting factors which act as the barriers for the wider application of the seasonal prediction models (Goddard 2010). For example, the seasonal predictions are affected by the predictors, predictands, region and season (Barnston et al. 2010). Nevertheless, the chaotic dynamics of the atmosphere may lead to the erroneous prediction of seasonal rainfall (Wang 2009). The uncertainties in the model parameterisation further hinder the prediction of seasonal rainfall. Australia's topography can be categorised by its arid and desert landscape. Various studies have been carried out to predict the rainfall distribution pattern in Australian. Rainfall, either intraseasonal or interseasonal, is linked to the climatic process, which is mostly originating in the tropical region. It is well established that large-scale atmospheric circulation patterns significantly affect the annual precipitation around the globe, including Australia. The atmospheric circulation configuration is dominated by the patterns of the sea surface temperature. Many researchers accept the capability of the El Nino Southern Oscillation (ENSO) in predicting time-series events. After analysing the role of ENSO on seasonal precipitation, Manzanas et al (2014). Manzanas et al (2014). found that September to October is the most skilful season to predict rainfall around eastern Australia. Hossain et al. (2015 and 2018) also identified the effects of ENSO and Indian Ocean Dipole (IOD) on west Australian rainfall. Therefore, the evaluation of the ENSO capability in time-series prediction is the fundamental requirement. Other climatic variables, such as sea surface temperature over the Atlantic and the Indian Ocean, have considerable impacts on the climate variability near the surrounding regions (Goddard et al. 2001). Recent studies also suggested that Indian Ocean Dipole (IOD) has considerable effects on the climate variability in the continental regions, including Australia Saji et al. (2003) and Ashok et al. (2003). Rasel et al. (2015) revealed the effects of SAM as a potential contributor of South Australian rainfall variability. The interaction of the ENSO phenomenon and IPO in the Pacific Ocean contribute to wetter or dryer seasons in Australia.

Till today, precipitation is the most challenging climatic phenomena, which can be predicted with least accuracy (Barnston 2010). On the other hand, most of the research studies on precipitation prediction have been conducted over a regional area of the world and in a particular season (Hossain et al. 2018; Mekanik et al. 2013; Kim et al. 2012; Lim et al. 2011). There exist only a few studies that concentrate on the precipitation analysis of the whole world (Branston et al. 2010; Wang et al. 2009; Manzanas 2014). Most of the studies conducted used number of various scores to evaluate predicted rainfall with the observed rainfall, such as correlations, ranked probability score and Brier skill score. However, there is still doubt regarding the accuracy of the predictions of the seasonal models, which has immense implications in the decision-making process (Rayner et al. 2005). Nevertheless, there exists overconfidence and lack of reliability in the prediction of seasonal rainfall using the currently available models (Langford et al. 2011). Therefore, it is necessary to assess the comprehensive performance of different models and their uncertainties in predicting seasonal rainfall.

A number of studies have been examined to identify appropriate modelling technique for the prediction of seasonal rainfall. However, only a single climate driver is not capable of replicating the accurate precipitation characteristics. Multi-predictors models have a higher prediction skill than single predictor models (Liu and Fan 2014). Nevertheless, there may exist dissimilar characteristics of seasonal rainfall patterns with the same rainfall totals (Tennant et al. 2002). On the other hand, there exist non-linear characteristics of the seasonal climate (Wang et al. 2009). Therefore, a closer look at the appropriate mechanism of seasonal rainfall formation becomes essentially important. Many researchers have tried to explore the underlying relationship between the large-scale climatic indices and rainfall.

Hossain et al. (2018) performed MNLR analysis to predict rainfall of Australian Capital Territory (ACT) using climate indices as predictors and compared the result with MLR analysis. The study found the cubic function to produce a maximum correlation between the dependent and independent variables while the correlation values ranged between 0.71 and 0.91. A similar approach will be applied in this section to develop seasonal streamflow forecast model using large scale climate indices as predictors. Later on, the results of this analysis will be compared with the results obtained from MLR analysis with a view to exploring the better predictor modelling technique.

6.2 Development of MNLR Models

There are several engineering applications for exploring relationships between two or more parameters. The regression analysis model is one of the popular statistical approaches and is highly recommended for this kind of analysis (Mekanik et al., 2013). In the previous stage (Chapter 5), the MLR technique was applied for forecasting seasonal streamflow prediction for the same region as the current study. In this chapter, to make the regressions more flexible, NMLR models were applied with a view to comparing the predictability of both these techniques.

Regression methods can explain the relationships between a response (dependent) variable and several regressor (independent) variables (Tabari et al., 2010). In MLR, the function is linear which can be explained by the following equation

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots . + \beta_n X_n \qquad (6\text{-}1)$$

Where Y is the dependent variable (e.g. streamflow for the current study) and $X_1, X_{2,} \ldots X_n$ are the independent variables (climate indices e.g. ENSO, PDO, IPO etc.). $\beta_1, \beta_2 \ldots\ldots \beta_n$ are the coefficients of the independent variables while α is the intercept or error and n is the number of observations.

Unlike traditional MLR methods, MNLR models are able to capture the arbitrary relationships between dependent and independent variables. The MNLR function is the non-linear combination of model parameters and depends on one or more independent variables (Bilgili 2010). The general form of a MNLR function can be represent by the following equation (Ivakhnenko, 1970):

$$Y = \alpha + \beta_1 X_i + \beta_2 X_j + \beta_3 X_i^2 + \beta_4 X_j^2 \ldots . + \beta_n X_i X_j \qquad (6\text{-}2)$$

Where α is the intercept and $\beta_1, \beta_2 \ldots\ldots \beta_n$ are the coefficients of the independent variables while n is the number of observations.

One of the major problems of statistical analysis was to establish the appropriate relationship between the dependent variable and a set of independent variables. In order to find out the suitable relationship of each independent variable (climate indices), a series of simple regression analysis between the streamflow and climate variables were performed (Haque et al. 2013). Based on the correlation values (Pearson correlation, r) of this analysis, the appropriate non-linear relationship for each variable is selected to develop the multiple non-linear equations for predicting streamflow. Different functions including the exponential, power, cubic, logarithmic, quadratic and linear functions are used to identify the best relation (Table 24).

Table 24. List of Non-Linear equations used in this study

Function	General Equation
Cubic	$y = ax^3 + bx^2 + cx + d$
Quadratic	$y = ax^2 + bx + c$
Logarithmic	$y = a \log bx$
Power	$y = ax^b$
Exponential	$y = ae^{bx}$

Where a, b, c and d are constants, and y and x are observed values.

The MNLR analysis in the present study was performed using Minitab software. Different combinations of input variables were used to calibrate and validate the MNLR models. At first, every MNLR model was calibrated using 85 years of data (1914-1998) which was followed by the validation of the models with the rest of 17 years (1999-2015) of data.

The performance of the developed MNLR models were assessed by two statistical performance measures, Pearson correlation value (r) and RMSE. A similar approach for validating the results was applied by Mekanik et al. (2013) while predicting rainfall using climate indices. The ideal value for the Pearson correlation is 1, which will refer to the best association between two variables, whereas a value of 0 will indicate there is no association. The lower value of the RMSE will indicate the better performance of the model.

6.3 Results of MNLR Analysis

6.3.1 Single Non-Linear Regression Analysis

Northern New South Wales (NNSW)

Singleton Station

Table 25. Single Non-Linear Regression analysis for Singleton station

Indices	Linear	Cubic	Power	Logarithmic	Quadratic	Exponential
PDO_{Aug}	0.294	0.301	*	*	0.300	0.300
PDO_{July}	0.303	0.305	*	*	0.303	0.290
PDO_{June}	0.247	0.252	*	*	0.252	0.236
PDO_{May}	0.257	0.268	*	*	0.267	0.245
PDO_{April}	0.299	0.246	*	*	0.230	0.232
PDO_{Mar}	0.230	0.243	*	*	0.238	0.238
PDO_{Feb}	0.207	0.364	*	*	0.327	0.318
PDO_{Jan}	0.203	0.318	*	*	0.300	0.245
PDO_{Dec}	0.131	0.176	*	*	0.155	0.141
$NINO3.4_{Aug}$	0.305	0.321	*	*	0.319	0.316
$NINO3.4_{July}$	0.359	0.386	*	*	0.359	0.347
$NINO3.4_{June}$	0.418	0.458	*	*	0.419	0.391
$NINO3.4_{May}$	0.278	0.279	*	*	0.279	0.268
$NINO3.4_{Feb}$	0.418	0.458	*	*	0.419	0.391
IOD_{Jun}	0.190	0.202	*	*	0.194	0.193

**Due to the Mathematical error, the logarithmic and power model cannot be generated.*

Coggan station

Table 26. Single Non-Linear Regression analysis for Coggan station

Indices	Linear	Cubic	Power	Logarithmic	Quadratic	Exponential
PDO_{Aug}	0.240	0.300	*	*	0.279	0.278
PDO_{July}	0.193	0.218	*	*	0.206	0.206
PDO_{June}	0.143	0.235	*	*	0.175	0.158
PDO_{May}	0.179	0.199	*	*	0.198	0.188
PDO_{April}	0.168	0.219	*	*	0.198	0.168
PDO_{Mar}	0.134	0.165	*	*	0.161	0.145
PDO_{Feb}	0.122	0.144	*	*	0.143	0.130
PDO_{Jan}	0.095	0.103	*	*	0.097	0.096
PDO_{Dec}	0.044	0.117	*	*	0.096	0.041
$NINO3.4_{Aug}$	0.289	0.364	*	*	0.357	0.345
$NINO3.4_{July}$	0.339	0.381	*	*	0.381	0.375
$NINO3.4_{June}$	0.306	0.235	*	*	0.319	0.315
$NINO3.4_{May}$	0.172	0.206	*	*	0.193	0.186
IOD_{Jun}	0.212	0.243	*	*	0.225	0.220

**Due to the Mathematical error, the logarithmic and power model cannot be generated.*

North Cuerindi

Table 27. Single Non-Linear Regression analysis for North Cuerindi station

Indices	Linear	Cubic	Power	Logarithmic	Quadratic	Exponential
PDO_{Aug}	0.311	0.467	*	*	0.390	0.395
PDO_{July}	0.340	0.424	*	*	0.394	0.394
PDO_{June}	0.239	0.267	*	*	0.241	0.243
PDO_{May}	0.200	0.237	*	*	0.200	0.201
PDO_{April}	0.173	0.292	*	*	0.222	0.191
PDO_{Mar}	0.081	0.229	*	*	0.229	0.095
PDO_{Feb}	0.136	0.397	*	*	0.379	0.196
PDO_{Jan}	0.146	0.327	*	*	0.265	0.173
PDO_{Dec}	0.070	0.089	*	*	0.089	0.073
$NINO3.4_{Aug}$	0.430	0.465	*	*	0.444	0.435
$NINO3.4_{July}$	0.482	0.520	*	*	0.488	0.472
$NINO3.4_{June}$	0.483	0.510	*	*	0.485	0.471
$NINO3.4_{May}$	0.264	0.334	*	*	0.299	0.299
EMI_{Aug}	0.337	0.379	*	*	0.351	0.357
EMI_{July}	0.241	0.258	*	*	0.253	0.252
EMI_{Jun}	0.267	0.279	*	*	0.268	0.267
EMI_{May}	0.209	0.271	*	*	0.255	0.233
EMI_{Apr}	0.189	0.261	*	*	0.216	0.215

**Due to the Mathematical error, the logarithmic and power model cannot be generated.*

Southern New South Wales (SNSW)

Gundagai Station

Table 28. Single Non-Linear Regression analysis for Gundagai station

Indices	Linear	Cubic	Power	Logarithmic	Quadratic	Exponential
PDO_{Aug}	0.242	0.351	*	*	0.304	0.263
PDO_{July}	0.238	0.283	*	*	0.251	0.246
PDO_{June}	0.229	0.326	*	*	0.259	0.239
PDO_{May}	0.168	0.329	*	*	0.255	0.179
PDO_{April}	0.21	0.442	*	*	0.257	0.221
PDO_{Mar}	0.246	0.451	*	*	0.326	0.268
PDO_{Feb}	0.193	0.264	*	*	0.212	0.199
PDO_{Jan}	0.174	0.254	*	*	0.186	0.178
PDO_{Dec}	0.177	0.222	*	*	0.177	0.176
$NINO3.4_{Aug}$	0.349	0.363	*	*	0.356	0.355
$NINO3.4_{Jul}$	0.375	0.399	*	*	0.393	0.387
$NINO3.4_{Jun}$	0.312	0.334	*	*	0.323	0.318
$NINO3.4_{May}$	0.239	0.322	*	*	0.321	0.259
$NINO3.4_{Apr}$	0.203	0.27	*	*	0.244	0.212
$NINO3.4_{Mar}$	0.199	0.225	*	*	0.203	0.201
$NINO3.4_{Feb}$	0.174	0.193	*	*	0.174	0.174
IOD_{Aug}	0.215	0.241	*	*	0.23	0.221
IOD_{July}	0.236	0.304	*	*	0.293	0.253
IOD_{June}	0.253	0.283	*	*	0.258	0.258
IOD_{May}	0.162	0.215	*	*	0.188	0.167
IOD_{April}	0.028	0.165	*	*	0.12	0.029
IOD_{Mar}	0.128	0.186	*	*	0.176	0.133
IPO	0.224	0.297	*	*	0.297	0.24
EMI_{Aug}	0.125	0.166	*	*	0.144	0.121
EMI_{July}	0.119	0.191	*	*	0.164	0.114

**Due to the Mathematical error, the logarithmic and power model cannot be generated.*

Wee Jasper Station

Table 29. Single Non-Linear Regression analysis for Wee Jasper station

Indices	Linear	Cubic	Power	Logarithmic	Quadratic	Exponential
PDO_{Aug}	0.211	0.257	*	*	0.228	0.219
PDO_{July}	0.177	0.201	*	*	0.186	0.181
PDO_{June}	0.181	0.255	*	*	0.189	0.185
PDO_{May}	0.155	0.263	*	*	0.226	0.164
PDO_{April}	0.206	0.415	*	*	0.267	0.219
PDO_{Mar}	0.207	0.406	*	*	0.329	0.23
PDO_{Feb}	0.191	0.264	*	*	0.222	0.199
PDO_{Jan}	0.238	0.306	*	*	0.266	0.248
PDO_{Dec}	0.234	0.281	*	*	0.235	0.236
$NINO3.4_{Aug}$	0.388	0.408	*	*	0.394	0.373
$NINO3.4_{Jul}$	0.411	0.417	*	*	0.411	0.406
$NINO3.4_{Jun}$	0.385	0.387	*	*	0.385	0.381
$NINO3.4_{May}$	0.273	0.400	*	*	0.377	0.302
$NINO3.4_{Apr}$	0.134	0.286	*	*	0.278	0.148
$NINO3.4_{Mar}$	0.148	0.268	*	*	0.268	0.162
$NINO3.4_{Feb}$	0.121	0.262	*	*	0.261	0.134
IOD_{Aug}	0.298	0.308	*	*	0.303	0.289
IOD_{July}	0.278	0.296	*	*	0.284	0.284
IOD_{June}	0.188	0.257	*	*	0.191	0.191
IOD_{May}	0.128	0.204	*	*	0.132	0.127
IOD_{April}	0.029	0.159	*	*	0.084	0.03
IOD_{Mar}	0.138	0.234	*	*	0.21	0.146
IPO	0.252	0.293	*	*	0.293	0.266
EMI_{Aug}	0.295	0.307	*	*	0.307	0.283
EMI_{July}	0.267	0.275	*	*	0.271	0.26

**Due to the Mathematical error, the logarithmic and power model cannot be generated.*

Mittagang Crossing Station

Table 30. Single Non-Linear Regression analysis for Mittagang Crossing station

Indices	Linear	Cubic	Power	Logarithmic	Quadratic	Exponential
PDO_{Aug}	0.279	0.365	*	*	0.345	0.317
PDO_{July}	0.219	0.269	*	*	0.253	0.237
PDO_{June}	0.143	0.248	*	*	0.186	0.152
PDO_{May}	0.162	0.260	*	*	0.173	0.167
PDO_{April}	0.265	0.380	*	*	0.323	0.292
PDO_{Mar}	0.379	0.462	*	*	0.44	0.424
PDO_{Feb}	0.309	0.360	*	*	0.346	0.334
PDO_{Jan}	0.266	0.339	*	*	0.338	0.293
PDO_{Dec}	0.255	0.308	*	*	0.307	0.275
$NINO3.4_{Aug}$	0.272	0.331	*	*	0.302	0.252
$NINO3.4_{Jul}$	0.288	0.306	*	*	0.301	0.273
$NINO3.4_{Jun}$	0.303	0.342	*	*	0.336	0.283
$NINO3.4_{May}$	0.242	0.252	*	*	0.252	0.25
$NINO3.4_{Apr}$	0.071	0.103	*	*	0.102	0.073
$NINO3.4_{Mar}$	0.156	0.172	*	*	0.172	0.161
$NINO3.4_{Feb}$	0.116	0.117	*	*	0.117	0.117
IOD_{Aug}	0.324	0.334	*	*	0.327	0.328
IOD_{July}	0.36	0.375	*	*	0.366	0.369
IOD_{June}	0.31	0.34	*	*	0.31	0.312
IOD_{May}	0.314	0.342	*	*	0.32	0.322
IOD_{April}	0.309	0.31	*	*	0.309	0.308
IOD_{Mar}	0.2	0.218	*	*	0.208	0.204
IPO	0.193	0.194	*	*	0.193	0.192
EMI_{Aug}	0.046	0.227	*	*	0.22	0.041
EMI_{July}	0.039	0.158	*	*	0.157	0.036

**Due to the Mathematical error, the logarithmic and power model cannot be generated.*

Kiosk Station

Table 31. Single Non-Linear Regression analysis for Kiosk station

Indices	Linear	Cubic	Power	Logarithmic	Quadratic	Exponential
PDO_{Aug}	0.288	0.357	*	*	0.332	0.318
PDO_{July}	0.187	0.213	*	*	0.212	0.197
PDO_{June}	0.151	0.218	*	*	0.188	0.16
PDO_{May}	0.131	0.179	*	*	0.178	0.139
PDO_{April}	0.187	0.27	*	*	0.227	0.199
PDO_{Mar}	0.271	0.358	*	*	0.338	0.302
PDO_{Feb}	0.254	0.285	*	*	0.273	0.267
PDO_{Jan}	0.257	0.272	*	*	0.269	0.265
PDO_{Dec}	0.187	0.195	*	*	0.193	0.182
$NINO3.4_{Aug}$	0.425	0.428	*	*	0.425	0.42
$NINO3.4_{Jul}$	0.428	0.43	*	*	0.428	0.418
$NINO3.4_{Jun}$	0.293	0.34	*	*	0.299	0.296
$NINO3.4_{May}$	0.218	0.274	*	*	0.251	0.23
$NINO3.4_{Apr}$	0.116	0.182	*	*	0.167	0.123
$NINO3.4_{Mar}$	0.087	0.144	*	*	0.129	0.092
$NINO3.4_{Feb}$	0.017	0.065	*	*	0.064	0.017
IOD_{Aug}	0.321	0.329	*	*	0.322	0.318
IOD_{July}	0.342	0.365	*	*	0.365	0.362
IOD_{June}	0.217	0.246	*	*	0.229	0.208
IOD_{May}	0.206	0.235	*	*	0.211	0.202
IOD_{April}	0.021	0.117	*	*	0.112	0.023
IOD_{Mar}	0.149	0.182	*	*	0.182	0.156
IPO	0.198	0.227	*	*	0.219	0.207
EMI_{Aug}	0.141	0.202	*	*	0.186	0.131
EMI_{July}	0.125	0.133	*	*	0.133	0.121

**Due to the Mathematical error, the logarithmic and power model cannot be generated.*

Central West New South Wales (CWNSW)

Corowa Station

Table 32. Single Non-Linear Regression analysis for Corowa station

Indices	Linear	Cubic	Power	Logarithmic	Quadratic	Exponential
PDO_{April}	0.232	0.456	*	*	0.305	0.251
PDO_{Mar}	0.295	0.419	*	*	0.321	-0.237
PDO_{Feb}	0.241	0.263	*	*	0.241	0.242
PDO_{Jan}	0.218	0.263	*	*	0.22	0.22
PDO_{Dec}	0.245	0.322	*	*	0.245	0.246
$NINO3.4_{Aug}$	0.317	0.318	*	*	0.317	0.312
$NINO3.4_{Jul}$	0.336	0.34	*	*	0.34	0.34
$NINO3.4_{Jun}$	0.23	0.248	*	*	0.248	0.238
$NINO3.4_{May}$	0.192	0.264	*	*	0.26	0.208
IOD_{Aug}	0.391	0.414	*	*	0.401	0.398
IOD_{Jul}	0.353	0.374	*	*	0.368	0.367
IOD_{Jun}	0.231	0.244	*	*	0.231	0.23
IPO	0.182	0.267	*	*	0.238	0.193

** Due to the Mathematical error, the logarithmic and power model cannot be generated.*

Wagga Wagga Station

Table 33. Single Non-Linear Regression analysis for Wagga Wagga station

Indices	Linear	Cubic	Power	Logarithmic	Quadratic	Exponential
PDO_{Aug}	0.18	0.297	*	*	0.219	0.191
PDO_{July}	0.058	0.153	*	*	0.108	0.061
PDO_{June}	0.052	0.185	*	*	0.069	0.053
PDO_{May}	0.057	0.15	*	*	0.146	0.06
PDO_{April}	0.183	0.314	*	*	0.221	0.191
PDO_{Mar}	0.279	0.331	*	*	0.312	0.293
PDO_{Feb}	0.255	0.266	*	*	0.258	0.258
PDO_{Jan}	0.242	0.242	*	*	0.242	0.241
PDO_{Dec}	0.25	0.281	*	*	0.26	0.244
$NINO3.4_{Aug}$	0.339	0.361	*	*	0.361	0.322
$NINO3.4_{July}$	0.35	0.361	*	*	0.359	0.338
$NINO3.4_{June}$	0.241	0.246	*	*	0.243	0.238
$NINO3.4_{May}$	0.199	0.259	*	*	0.253	0.211
$NINO3.4_{April}$	0.12	0.204	*	*	0.197	0.127
$NINO3.4_{Mar}$	0.112	0.231	*	*	0.193	0.12
$NINO3.4_{Feb}$	0.088	0.125	*	*	0.125	0.091
IOD_{Aug}	0.451	0.453	*	*	0.451	0.442
IOD_{July}	0.428	0.431	*	*	0.428	0.42
IOD_{June}	0.279	0.328	*	*	0.301	0.267
IOD_{May}	0.158	0.166	*	*	0.161	0.156
IOD_{April}	0.086	0.102	*	*	0.102	0.087
IOD_{Mar}	0.093	0.115	*	*	0.093	0.093
IPO	0.097	0.121	*	*	0.098	0.098
EMI_{Aug}	0.118	0.233	*	*	0.23	0.11
EMI_{July}	0.147	0.198	*	*	0.197	0.139

**Due to the Mathematical error, the logarithmic and power model cannot be generated.*

Cowra Station

Table 34. Single Non-Linear Regression analysis for Cowra station

Indices	Linear	Cubic	Power	Logarithmic	Quadratic	Exponential
PDO_{Aug}	0.22	0.244	*	*	0.24	0.234
PDO_{July}	0.222	0.243	*	*	0.226	0.228
PDO_{June}	0.214	0.259	*	*	0.23	0.225
PDO_{May}	0.209	0.383	*	*	0.323	0.251
PDO_{April}	0.192	0.379	*	*	0.254	0.216
PDO_{Mar}	0.194	0.427	*	*	0.335	0.247
PDO_{Feb}	0.176	0.343	*	*	0.284	0.212
PDO_{Jan}	0.158	0.303	*	*	0.208	0.174
PDO_{Dec}	0.133	0.239	*	*	0.153	0.141
$NINO3.4_{Aug}$	0.211	0.219	*	*	0.211	0.209
$NINO3.4_{July}$	0.287	0.305	*	*	0.305	0.301
$NINO3.4_{June}$	0.318	0.337	*	*	0.336	0.334
$NINO3.4_{May}$	0.251	0.362	*	*	0.348	0.303
$NINO3.4_{April}$	0.228	0.294	*	*	0.293	0.255
$NINO3.4_{Mar}$	0.227	0.271	*	*	0.27	0.248
$NINO3.4_{Feb}$	0.236	0.325	*	*	0.314	0.273
IOD_{Aug}	0.06	0.113	*	*	0.075	0.062
IOD_{July}	0.072	0.2	*	*	0.168	0.082
IOD_{June}	0.159	0.182	*	*	0.165	0.164
IOD_{May}	0.186	0.362	*	*	0.19	0.189
IOD_{April}	0.05	0.294	*	*	0.107	0.053
IOD_{Mar}	0.143	0.218	*	*	0.215	0.157
IPO	0.272	0.396	*	*	0.381	0.327
EMI_{Aug}	0.208	0.227	*	*	0.208	0.205
EMI_{July}	0.248	0.344	*	*	0.253	0.233

**Due to the Mathematical error, the logarithmic and power model cannot be generated.*

Western New South Wales (WNSW)

Barham Station

Table 35. Single Non-Linear Regression analysis for Barham station

Indices	Linear	Cubic	Power	Logarithmic	Quadratic	Exponential
PDO_{Aug}	0.228	0.351	*	*	0.296	0.25
PDO_{July}	0.212	0.259	*	*	0.229	0.22
PDO_{June}	0.202	0.298	*	*	0.23	0.21
PDO_{May}	0.139	0.303	*	*	0.221	0.148
PDO_{April}	0.181	0.4	*	*	0.223	0.19
PDO_{Mar}	0.219	0.41	*	*	0.304	0.239
PDO_{Feb}	0.175	0.255	*	*	0.197	0.182
PDO_{Jan}	0.161	0.249	*	*	0.174	0.165
PDO_{Dec}	0.156	0.197	*	*	0.158	0.155
$NINO3.4_{Aug}$	0.36	0.369	*	*	0.368	0.367
$NINO3.4_{July}$	0.39	0.41	*	*	0.407	0.402
$NINO3.4_{June}$	0.311	0.33	*	*	0.321	0.317
$NINO3.4_{May}$	0.229	0.317	*	*	0.317	0.251
$NINO3.4_{April}$	0.181	0.257	*	*	0.233	0.19
$NINO3.4_{Mar}$	0.187	0.219	*	*	0.198	0.191
$NINO3.4_{Feb}$	0.16	0.183	*	*	0.163	0.161
IOD_{Aug}	0.227	0.242	*	*	0.235	0.231
IOD_{July}	0.242	0.292	*	*	0.283	0.258
IOD_{June}	0.247	0.27	*	*	0.25	0.25
IOD_{May}	0.134	0.194	*	*	0.153	0.137
IOD_{April}	0.045	0.196	*	*	0.119	0.046
IOD_{Mar}	0.143	0.201	*	*	0.183	0.148
IPO	0.206	0.265	*	*	0.265	0.219
EMI_{Aug}	0.147	0.178	*	*	0.169	0.142
EMI_{July}	0.131	0.182	*	*	0.168	0.125

**Due to the Mathematical error, the logarithmic and power model cannot be generated.*

Brewarrina Station

Table 36. Single Non-Linear Regression analysis for Brewarrina station

Indices	Linear	Cubic	Power	Logarithmic	Quadratic	Exponential
PDO_{Aug}	0.311	0.377	*	*	0.365	0.359
PDO_{July}	0.321	0.412	*	*	0.407	0.378
PDO_{June}	0.276	0.320	*	*	0.305	0.299
PDO_{May}	0.275	0.304	*	*	0.297	0.291
PDO_{April}	0.277	0.418	*	*	0.335	0.335
$NINO3.4_{Aug}$	0.364	0.371	*	*	0.368	0.362
$NINO3.4_{July}$	0.408	0.419	*	*	0.411	0.401
$NINO3.4_{June}$	0.427	0.441	*	*	0.433	0.426
$NINO3.4_{May}$	0.305	0.352	*	*	0.346	0.333
IOD_{Aug}	0.229	0.237	*	*	0.229	0.225
IOD_{July}	0.197	0.216	*	*	0.215	0.209
IOD_{Dec}	0.107	0.142	*	*	0.107	0.352
EMI_{Aug}	0.256	0.275	*	*	0.256	0.255
EMI_{July}	0.241	0.266	*	*	0.254	0.256
EMI_{Jun}	0.229	0.273	*	*	0.229	0.230

**Due to the Mathematical error, the logarithmic and power model cannot be generated.*

Multiple Non-Linear Regression Analysis

Northern New South Wales (NNSW)

Singleton Station

Table 37. Multiple Non-Linear Regression Analysis for Singleton Station

Indices Combination	Correlations	
	Calibration	Validation
PDO_{Aug} -$NINO3.4_{Aug}$	0.372	0.646
PDO_{Jul}-$NINO3.4_{Jul}$	0.438	0.537
PDO_{Jun}-$NINO3.4_{Jun}$	0.471	0.568
PDO_{May}-$NINO3.4_{May}$	0.311	0.620
PDO_{Apr}-$NINO3.4_{Jun}$	0.466	0.581
PDO_{Mar}-$NINO3.4_{May}$	0.322	0.562
PDO_{Mar}-$NINO3.4_{Feb}$	0.473	0.588
PDO_{Feb}-$NINO3.4_{May}$	0.413	0.551
PDO_{Jan}-$NINO3.4_{May}$	0.404	0.584
PDO_{Jan}-$NINO3.4_{May}$	0.314	0.565
IOD_{Jun}-$NINO3.4_{Jun}$	0.478	0.437

Coggan Station

Table 38. Multiple Non-Linear Regression Analysis for Coggan Station

Indices Combination	Correlations	
	Calibration	Validation
PDO_{Aug} -$NINO3.4_{Aug}$	0.400	0.953
PDO_{Jul}-$NINO3.4_{Jul}$	0.391	0.821
PDO_{Jun}-$NINO3.4_{Jun}$	0.352	0.745
PDO_{May}-$NINO3.4_{May}$	0.253	0.325
PDO_{Apr}-$NINO3.4_{Jun}$	0.343	0.755
PDO_{Apr}-$NINO3.4_{May}$	0.255	0.733
PDO_{Mar}-$NINO3.4_{Jun}$	0.335	0.738
PDO_{Mar}-$NINO3.4_{May}$	0.246	0.796
PDO_{Feb}-$NINO3.4_{May}$	0.244	0.456
PDO_{Jan}-$NINO3.4_{May}$	0.223	0.476
PDO_{Feb}-$NINO3.4_{May}$	0.227	0.496
IOD_{June}-$NINO3.4_{June}$	0.380	0.372

North Cuerindi

Table 39. Multiple Non-Linear Regression Analysis for North Cuerindi Station

Indices Combination	Correlations	
	Calibration	Validation
PDO_{Aug} -$NINO3.4_{Aug}$	0.576	0.99
PDO_{Jul}-$NINO3.4_{Jul}$	0.582	0.771
PDO_{Jun}-$NINO3.4_{Jun}$	0.514	0.642
PDO_{May}-$NINO3.4_{May}$	0.359	0.400
PDO_{Apr}-$NINO3.4_{Jun}$	0.534	0.623
PDO_{Apr}-$NINO3.4_{May}$	0.382	0.948
PDO_{Mar}-$NINO3.4_{Jun}$	0.619	0.544
PDO_{Mar}-$NINO3.4_{May}$	0.410	0.383
PDO_{Feb}-$NINO3.4_{May}$	0.546	0.324
PDO_{Jan}-$NINO3.4_{May}$	0.479	0.324
PDO_{Feb}-$NINO3.4_{May}$	-0.016	0.289
EMI_{Aug}-$NINO3.4_{Aug}$	0.496	0.496
EMI_{July}-$NINO3.4_{July}$	0.547	0.580
EMI_{Jun}-$NINO3.4_{June}$	0.566	0.606
EMI_{May}-$NINO3.4_{May}$	0.435	0.435
EMI_{Apr}-$NINO3.4_{May}$	0.391	0.682

Central West New South Wales (CWNSW)

Corowa Station

Table 40. Multiple Non-Linear Regression Analysis for Corowa Station

Indices Combination	Correlations	
	Calibration	Validation
PDO_{April} -$NINO3.4_{Aug}$	0.504	0.078
PDO_{April}-$NINO3.4_{July}$	0.499	0.036
PDO_{April}-$NINO3.4_{June}$	0.464	-0.481
PDO_{April}-$NINO3.4_{May}$	0.484	-0.497
PDO_{Mar}-$NINO3.4_{May}$	0.494	0.117
PDO_{Feb}-$NINO3.4_{May}$	0.368	0.402
PDO_{Jan}-$NINO3.4_{May}$	0.363	0.307
PDO_{Dec}-$NINO3.4_{May}$	0.434	0.388
IOD_{Aug}-$NINO3.4_{Aug}$	0.439	0.164
IOD_{July} $NINO3.4_{July}$	0.443	0.101
$IOD_{June}NINO3.4_{June}$	0.319	0.634
$IOD_{June}NINO3.4_{May}$	0.342	0.737
IPO-$NINO3.4_{Aug}$	0.403	0.521
IPO-$NINO3.4_{July}$	0.409	0.561
IPO-$NINO3.4_{June}$	0.341	0.694
IPO-$NINO3.4_{May}$	0.345	0.686
IOD_{Aug} -IPO	0.525	0.639
IOD_{July}-IPO	0.482	0.631

Wagga Wagga Station

Table 41. Multiple Non-Linear Regression Analysis for Wagga Wagga Station

Indices Combination	Correlations	
	Calibration	Validation
PDO_{Aug} -$NINO3.4_{Aug}$	0.441	0.362
PDO_{July}-$NINO3.4_{July}$	0.423	0.481
PDO_{June}-$NINO3.4_{June}$	0.376	0.279
PDO_{May}-$NINO3.4_{May}$	0.373	-0.194
PDO_{April}-$NINO3.4_{Apr}$	0.469	-0.019
PDO_{Mar}-$NINO3.4_{Mar}$	0.43	0.35
PDO_{Feb}-$NINO3.4_{Feb}$	0.294	-0.134
IOD_{Aug} -$NINO3.4_{Aug}$	0.382	0.52
IOD_{July}-$NINO3.4_{July}$	0.451	0.541
IOD_{June}-$NINO3.4_{June}$	0.39	0.382
IOD_{June}-$NINO3.4_{May}$	0.399	-0.218
IOD_{June}-$NINO3.4_{Apr}$	0.354	-0.224
IOD_{June}-$NINO3.4_{Mar}$	0.36	-0.301
IOD_{June}-$NINO3.4_{Feb}$	0.344	-0.393
IPO -$NINO3.4_{Aug}$	0.44	0.345
IPO -$NINO3.4_{July}$	0.467	0.363
IPO -$NINO3.4_{June}$	0.401	0.175
IPO -$NINO3.4_{May}$	0.379	0.274
IPO -$NINO3.4_{Apr}$	0.346	-0.046
IPO -$NINO3.4_{Mar}$	0.319	-0.062
IPO -$NINO3.4_{Feb}$	0.298	-0.061
EMI_{Aug} -$NINO3.4_{Aug}$	0.376	-0.127
EMI_{July}-$NINO3.4_{July}$	0.418	0.506
EMI_{July}-$NINO3.4_{June}$	0.334	0.577
EMI_{July}-$NINO3.4_{May}$	0.325	0.586
EMI_{July}-$NINO3.4_{Ap}$r	0.288	-0.153
EMI_{July}-$NINO3.4_{Mar}$	0.253	-0.052
EMI_{July}-$NINO3.4_{Feb}$	0.231	-0.448
IOD_{Aug} -IPO	0.376	0.093

$IOD_{July}IPO$	0.394	-0.031
EMI_{Aug} -IPO	0.307	0.201
EMI_{July} -IPO	0.294	0.175

Cowra Station

Table 42. Multiple Non-Linear Regression Analysis for Cowra Station

Indices Combination	Correlations	
	Calibration	**Validation**
$PDOJ_{uly}$-NINO3.4_{July}	0.332	-0.909
PDO_{June}-NINO3.4_{June}	0.358	-0.782
PDO_{May}-NINO3.4_{May}	0.427	-0.023
PDO_{April}-NINO3.4_{Apr}	0.436	0.105
$PDO_{Ma}r$-NINO3.4_{Mar}	0.464	-0.316
PDO_{Mar}-NINO3.4_{Feb}	0.485	-0.372
EMI_{Aug} -NINO3.4_{Aug}	0.278	-0.859
EMI_{July}-NINO3.4_{July}	0.406	-0.739
EMI_{July}-NINO3.4_{June}	0.414	-0.732
EMI_{July}-NINO3.4_{May}	0.427	-0.411
EMI_{July}-NINO3.4_{Apr}	0.381	-0.321
EMI_{July}-NINO3.4_{Mar}	0.366	-0.386
EMI_{July}-NINO3.4_{Feb}	0.385	-0.443

Southern New South Wales (SNSW)

Table 43. Multiple Non-Linear Regression Analysis for Gundagai Station

Indices Combination	Correlations	
	Calibration	Validation
PDO_{Aug} -$NINO3.4_{Aug}$	0.434	0.192
PDO_{July}-$NINO3.4_{July}$	0.421	0.264
PDO_{June}-$NINO3.4_{June}$	0.394	-0.019
PDO_{May}-$NINO3.4_{May}$	0.388	-0.363
PDO_{April}-$NINO3.4_{Apr}$	0.508	0.167
PDO_{Mar}-$NINO3.4_{Mar}$	0.468	0.535
PDO_{Feb}-$NINO3.4_{Feb}$	0.3	0.029
PDO_{Jan}-$NINO3.4_{Feb}$	0.289	0.09
PDO_{Dec}-$NINO3.4_{Feb}$	0.259	0.324
IOD_{Aug}-$NINO_{Aug}$	0.379	0.316
IOD_{July}-$NINO_{July}$	0.45	0.362
IOD_{June}-$NINO_{June}$	0.402	0.168
IOD_{May}-$NINO_{May}$	0.366	-0.381
IOD_{May}-$NINO_{April}$	0.338	-0.193
IOD_{May}-$NINO_{Mar}$	0.314	-0.212
IOD_{May}-$NINO_{Feb}$	0.296	-0.282
IPO-$NINO_{Aug}$	0.458	0.144
IPO -$NINO_{July}$	0.477	0.2
IPO -$NINO_{June}$	0.426	0.154
IPO -$NINO_{May}$	-	-
IPO -$NINO_{April}$	0.376	0.01
IPO -$NINO_{Mar}$	0.346	0.017
IPO -$NINO_{Feb}$	0.329	-0.016
IOD_{Aug}- IPO	0.401	0.188
IOD_{July}- IPO	0.42	0.024
IOD_{June}- IPO	0.422	0.156
IOD_{May}- IPO	0.379	0.115

Table 44. Multiple Non-Linear Regression Analysis for Wee Jasper Station

Indices Combination	Correlations	
	Calibration	Validation
PDO_{Aug} -$NINO3.4_{Aug}$	0.434	0.192
PDO_{July}-$NINO3.4_{July}$	0.421	0.264
PDO_{April}-$NINO3.4_{June}$	0.394	-0.019
PDO_{April}-$NINO3.4_{May}$	0.388	-0.363
PDO_{March}-$NINO3.4_{May}$	0.508	0.167
PDO_{Jan}-$NINO3.4_{May}$	0.468	0.535
PDO_{Dec}-$NINO3.4_{May}$	0.3	0.029
IOD_{Aug}-$NINO_{Aug}$	0.379	0.316
IOD_{July}-$NINO_{July}$	0.45	0.362
IOD_{June}-$NINO_{June}$	0.402	0.168
IPO-$NINO_{Aug}$	0.458	0.144
IPO -$NINO_{July}$	0.477	0.2
IPO -$NINO_{June}$	0.426	0.154
IPO -$NINO_{May}$	-	-
IOD_{Aug}- IPO	0.401	0.188
IOD_{July}- IPO	0.42	0.024
IOD_{June}- IPO	0.422	0.156
IOD_{May}- IPO	0.379	0.115
EMI_{Aug} -$NINO3.4_{Aug}$	0.46	0.462
EMI_{July}-$NINO3.4_{July}$	0.448	0.667
EMI_{April}-$NINO3.4_{June}$	0.425	0.668
EMI_{April}-$NINO3.4_{May}$	0.43	0.285

Table 45. Multiple Non-Linear Regression Analysis for Mittagang Crossing Station

Indices Combination	Correlations	
	Calibration	Validation
PDO_{Aug} -$NINO3.4_{Aug}$	0.456	0.272
PDO_{Aug}-$NINO3.4_{July}$	0.434	0.506
PDO_{Aug}-$NINO3.4_{June}$	0.463	0.591
PDO_{April}-$NINO3.4_{May}$	0.407	0.772
PDO_{March}-$NINO3.4_{May}$	0.478	0.505
PDO_{Feb}-$NINO3.4_{May}$	0.406	0.461
PDO_{Jan}-$NINO3.4_{May}$	0.387	0.326
PDO_{Dec}-$NINO3.4_{May}$	0.383	0.294
IOD_{Aug}-$NINO_{Aug}$	0.408	-0.218
IOD_{July}-$NINO_{July}$	0.421	0.139
IOD_{June}-$NINO_{June}$	0.433	0.628
IOD_{May}-$NINO3.4_{May}$	0.395	0.505
IOD_{April}-$NINO3.4_{May}$	0.422	0.277
IOD_{March}-$NINO3.4_{May}$	0.353	0.633

Table 46. Multiple Non-Linear Regression Analysis for Kiosk Station

Indices Combination	Correlations	
	Calibration	Validation
PDO_{Aug} -$NINO3.4_{Aug}$	0.482	0.473
PDO_{Aug}-$NINO3.4_{July}$	0.477	0.425
PDO_{Aug}-$NINO3.4_{June}$	0.378	0.574
PDO_{April}-$NINO3.4_{May}$	0.359	0.231
PDO_{March}-$NINO3.4_{May}$	0.425	-0.009
PDO_{Feb}-$NINO3.4_{May}$	0.39	-0.049
PDO_{Jan}-$NINO3.4_{May}$	0.366	-0.045
PDO_{Dec}-$NINO3.4_{May}$	0.336	0.15
IOD_{Aug}-$NINO_{Aug}$	0.453	0.58
IOD_{July}-$NINO_{July}$	0.491	0.506
IOD_{June}-$NINO_{June}$	0.379	0.559
IOD_{May}-$NINO3.4_{May}$	0.332	0.033

Western New South Wales (WNSW)

Barham Station

Table 47. Multiple Non-Linear Regression Analysis for Barham Station

Indices Combination	Correlations	
	Calibration	Validation
PDO_{Dec}-$NINO3.4_{May}$	0.396	0.51
PDO_{Jan}-$NINO3.4_{May}$	0.339	0.221
PDO_{Feb}-$NINO3.4_{May}$	0.369	0.212
PDO_{Mar}-$NINO3.4_{May}$	0.405	0.252
PDO_{Apr}-$NINO3.4_{May}$	0.383	0.048
PDO_{May}-$NINO3.4_{June}$	0.354	0.846
PDO_{May}-$NINO3.4_{July}$	0.427	0.648
PDO_{May}-$NINO3.4_{Aug}$	0.437	0.586
IOD_{Aug}-$NINO3.4_{Aug}$	0.482	0.35
IOD_{July} $NINO3.4_{July}$	0.484	0.391
$IOD_{June}NINO3.4_{June}$	0.369	0.771
$IOD_{June}NINO3.4_{May}$	0.389	0.642

Table 48. Multiple Non-Linear Regression Analysis for Brewarrina Station

Indices Combination	Correlations	
	Calibration	Validation
PDO_{Aug}-$NINO3.4_{Aug}$	0.454	0.473
PDO_{Jul}-$NINO3.4_{Jul}$	0.523	0.534
PDO_{Jun}-$NINO3.4_{Jun}$	0.472	0.545
PDO_{May}-$NINO3.4_{May}$	0.407	0.016
PDO_{Apr}-$NINO3.4_{May}$	0.492	0.404
IOD_{Aug}-$NINO3.4_{Aug}$	0.381	0.701
IOD_{Jul}- $NINO3.4_{Jul}$	0.429	0.760
IOD_{Jul}-$NINO3.4_{Jun}$	0.462	0.704
IOD_{Jul}-$NINO3.4_{May}$	0.400	0.001
IOD_{Dec}-$NINO3.4_{May}$	0.361	0.179
EMI_{Aug} -$NINO3.4_{Aug}$	0.414	0.720
EMI_{July}-$NINO3.4_{July}$	0.465	0.767
EMI_{Jun}-$NINO3.4_{Jun}$	0.516	0.634
EMI_{Jun}-$NINO3.4_{May}$	0.446	0.145

6.3.2 Statistical Error Analysis of the Best Models

Northern New South Wales (NNSW)

Table 49. Statistical performance analysis of MNLR models for NNSW

Station	Model	Calibration Period				Validation Period			
		r	MAE	RMSE	d	r	MAE	RMSE	d
Singleton	PDO_{June} $NINO3.4_{June}$	0.47	12.67	17.96	0.58	0.56	12.92	13.71	0.83
Coggan	IOD_{June} $NINO3.4_{June}$	0.38	1.14	2.04	0.43	0.37	1.35	1.73	0.52
North Cuerindi	PDO_{April} $NINO3.4_{June}$	0.53	5.18	7.83	0.64	0.62	7.63	8.98	0.57

Southern New South Wales (SNSW)

Table 50. Statistical performance analysis of MNLR models for SNSW

Station	Model	Calibration Period					Validation Period			
		r	MAE	RMSE	VIF	d	r	MAE	RMSE	d
Gundagai	PDO_{Mar} $NINO3.4_{Mar}$	0.468	53.906	68.982	1.280	0.567	0.535	27.965	31.226	0.655
Wee Jasper	IPO $NINO3.4_{Jun}$	0.451	5.777	10.184	1.254	0.579	0.484	8.241	8.553	0.532
Mittagang Crossing	PDO_{Mar} $NINO3.4_{May}$	0.478	6.767	8.342	1.295	0.599	0.505	5.094	5.488	0.312
Kiosk	IOD_{Jun} $NINO3.4_{Jun}$	0.379	3.188	4.105	1.166	0.48	0.559	3.835	4.126	0.569

Central West New South Wales (CWNSW)

Table 51. Statistical performance analysis of MNLR models for CWNSW

Station	Model	Calibration Period						Validation Period			
		r	MAE	RMSE	NSE	VIF	d	r	MAE	RMSE	d
Corowa	**PDO_{Dec} $NINO3.4_{May}$**	0.434	109.752	131.3	0.19	1.2	0.55	0.388	57.6	26.17	0.453
Wagga Wagga	**PDO_{Mar} $NINO3.4_{Mar}$**	0.43	65.735	84.42	0.19	1.2	0.52	0.35	36.939	5.035	0.557
Cowra	**PDO_{Apr} $NINO3.4_{Apr}$**	0.436	23.583	31.05	0.24	1.3	0.6	0.105	18.142	17.66	0.177

Western New South Wales (WNSW)

Table 52. Statistical performance analysis of MNLR models for WNSW

Station	Model	Calibration Period						Validation Period			
		r	MAE	RMSE	NSE	VIF	d	r	MAE	RMSE	d
Barham	**PDO_{Dec} $NINO3.4_{May}$**	0.396	88.487	104.1	0.16	1.2	0.52	0.51	51.753	57.86	0.653
Brewarrina	**EMI_{Jun} $NINO3.4_{Jun}$**	0.51	30.73	41.00	0.63		0.63	0.61	35.02	40.99	0.64

6.3.3 Equations of the Best Developed Models

Table 53. Equations of the best-developed MNLR models for all twelve stations

Region	Station	Best Developed Model
NNSW	Singleton	$Q = 19.9093 + 0.405381 * PDO_{JUNE}^3 - 0.781302 * PDO_{JUNE}^2 - 2.70294 * PDO_{JUNE} + 12.3403 * NINO3.4_{JUNE}^3 - 1.24083 * NINO3.4_{JUNE}^2 - 24.9111 * NINO3.4_{JUNE}$
	Coggan	$Q = 1.47979 + 3.04563 * IOD_{JUNE}^3 + 0.875505 * IOD_{JUNE}^2 - 0.443079 * IOD_{JUNE} + 0.443079 * NINO3.4_{JUNE}^3 - 0.48093 * NINO3.4_{JUNE}^2 - 1.48439 * NINO3.4_{JUNE}$
	North Cuerindi	$Q = 8.11847 - 0.298752 * PDO_{APRIL}^3 + 0.813401 * PDO_{APRIL}^2 + 1.09289 * PDO_{AUG} + 4.37942 * NINO3.4_{JUNE}^3 - 0.118078 * NINO3.4_{JUNE}^2 - 11.1756 * NINO3.4_{JUNE}$
SNSW	Gundagai	$Q = -8.403 * PDO_{Mar}^3 + 9.54 * PDO_{Mar}^2 + 13.462 * PDO_{Mar} + 18.246 * NINO3.4_{Mar}^3 - 1.842 * NINO3.4_{Mar}^2 - 29.314 * NINO3.4_{Mar} + 136.08$
	Wee Jasper	$Q = -0.1546 * IPO^3 + 1.4408 * IPO^2 - 1.0806 * IPO + 1.4964 * NINO3.4_{Jun}^3 - 0.6662 * NINO3.4_{Jun}^2 - 6.3021 * NINO3.4_{Jun} + 13.5436$
	Mittagang Crossing	$Q = -0.4587 * PDO_{Mar}^3 + 1.1277 * PDO_{Mar}^2 - 1.337 * PDO_{Mar} + 2.1657 * NINO3.4_{May}^3 + 0.8696 * NINO3.4_{May}^2 - 3.5271 * NINO3.4_{May} + 10.7027$
	Kiosk	$Q = -4.838 * IOD_{Jun}^3 - 0.413 * IOD_{Jun}^2 - 0.521 * IOD_{Jun} + 2.05461 * NINO3.4_{Jun}^3 + 0.53274 * NINO3.4_{Jun}^2 - 3.70504 * NINO3.4_{Jun} + 5.65868$
CNSW	Corowa	$Q = -9.37485 * PDO_{Dec}^3 - 0.838266 * PDO_{Dec}^2 + 9.52596 * PDO_{Dec} - 0.824081 * NINO3.4_{May}^3 + 97.4374 * NINO3.4_{May}^2 - 44.5191 * NINO3.4_{Jun} + 217.189$

	Wagga Wagga	$Q = -9.02781 * PDO_{Mar}^{3} + 10.7172 * PDO_{Mar}^{2} + 14.8737 * PDO_{Mar} + 20.3952 * NINO3.4_{Mar}^{3} + 1.35147 * NINO3.4_{Mar}^{2} - 34.972 * NINO3.4_{Mar} + 150.128$
	Cowra	$Q = -3.8423 * PDO_{Apr}^{3} - 2.25516 * PDO_{Apr}^{2} + 9.06112 * PDO_{Apr} - 11.2937 * NINO3.4_{Apr}^{3} + 9.92076 * NINO3.4_{Apr}^{2} - 19.7842 * NINO3.4_{Apr} + 29.56$
WNSW	**Barham**	$Q = -4.71687 * PDO_{Dec}^{3} - 5.78292 * PDO_{Dec}^{2} - 3.54451 * PDO_{Dec} - 56.668 * NINO3.4_{May}^{2} - 65.8202 * NINO3.4_{May} + 192.641$
	Brewarrina	$Q = 44.6711 - 50.112 * EMI_{JUNE}^{3} - 6.60371 * EMI_{JUNE}^{2} + 17.5917 * EMI_{JUNE} + 14.0838 * NINO3.4_{JUNE}^{3} + 5.33871 * NINO3.4_{JUNE}^{2} - 51.1802 * NINO3.4_{JUNE}$

6.4 Detailed Discussion on the Result of MNLR Analysis

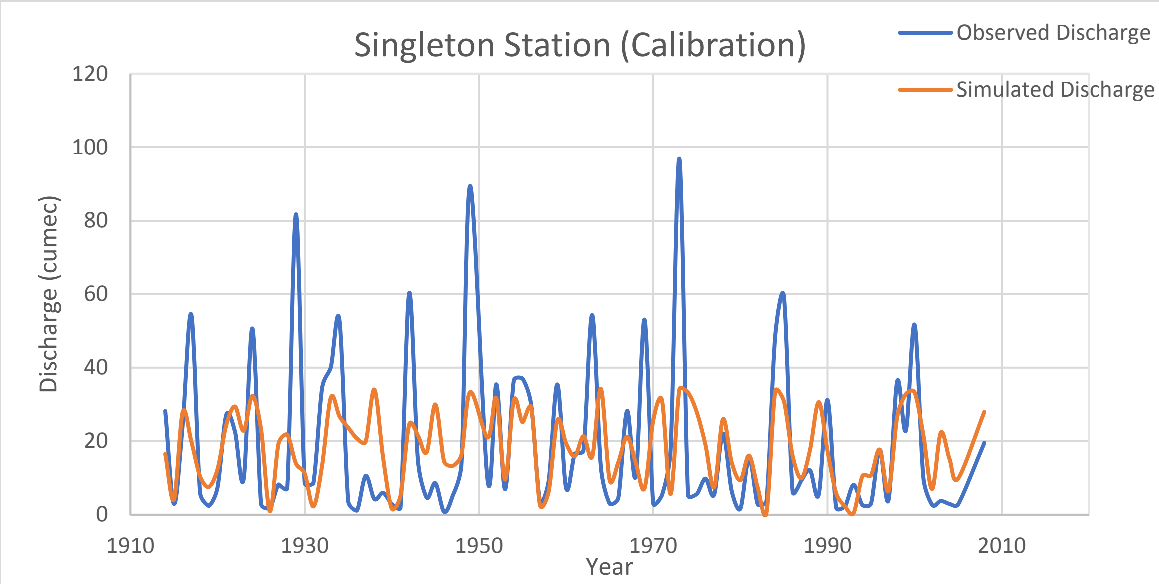
North New South Wales (NNSW)

Singleton Station

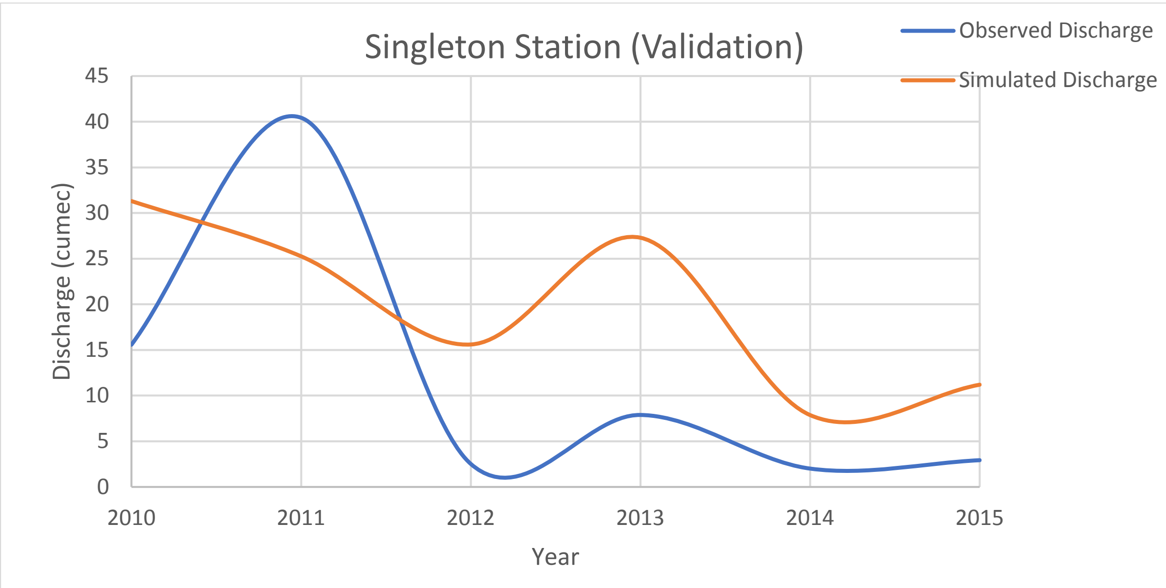
From the result shown in Table 25, it can be seen that the maximum value for the single correlation value was obtained for cubic function for all indices. NINO3.4$_{Feb}$ with a correlation value of 0.458 is the highest correlation value for single non-linear regression analysis. It was observed that for the single linear correlation analysis, the maximum value was 0.418.

From Table 37, the maximum value of correlation coefficient in the calibration stage was found to be 0.478 for IOD$_{June}$- NINO 3.4$_{June,}$ whereas in the validation stage, was 0.646 for PDO$_{Aug-}$ NINO 3.4$_{Aug}$. Since the difference between correlation values of calibration and validation period was high for both of these two combinations along with the low lag period months, these models were not considered as the best models. Hence the combination PDO$_{June}$-NINO3.4$_{June}$ with correlation coefficient 0.471 in calibration and 0.568 in validation period was selected as the best model, and further statistical performance analysis (Table 49) showed lower statistical errors for this model. This model was developed using cubic function for both indices, which was based on the outcomes of single non-linear regression analysis (Table 49). This model can predict future streamflow for three months in advance.

The time series plots (Figure 20) of observed and simulated flow for calibration and validation periods show that the model follows the trend of observed flow mostly, though they fail to predict extreme points, i.e. very high or low values.



(a)



(b)

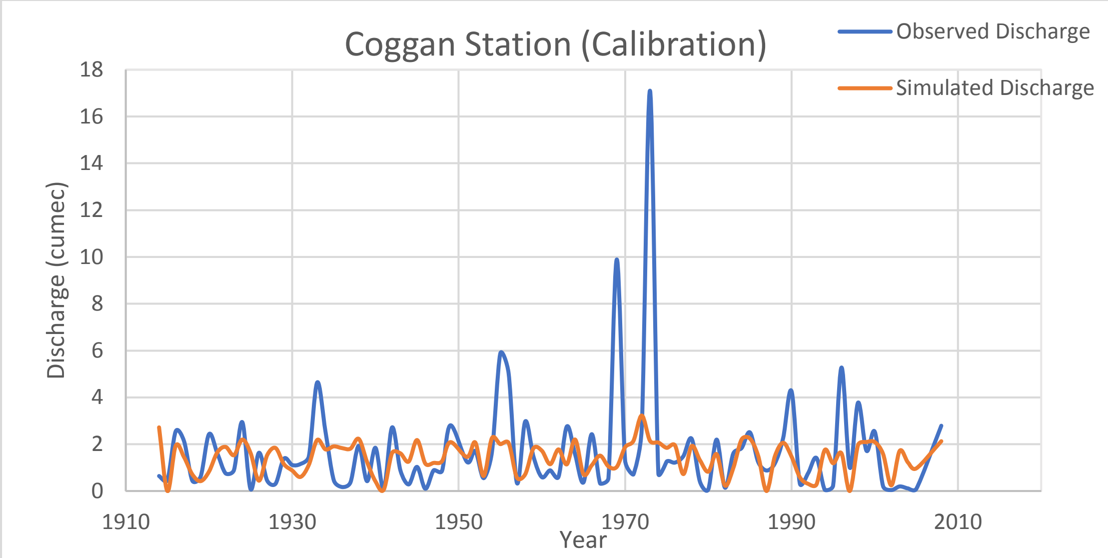
Figure 20. Comparison between the observed and simulated streamflow during the (a) calibration and (b) validation periods for Singleton Station

Coggan Station

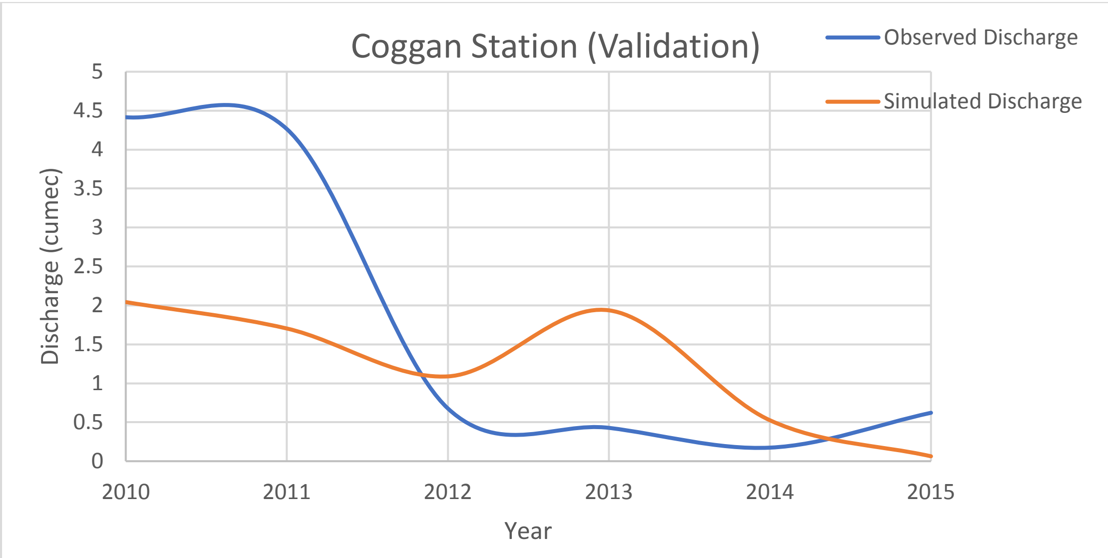
From the result shown in Table 26, it can be seen that the maximum value for the single non-linear regression analysis was obtained with quadratic function for NINO3.4$_{July}$ and the value was 0.381 while for all other indices maximum correlations were provided by cubic function. It was observed that for the single linear correlation analysis the maximum value was 0.339.

From Table 38, the maximum correlation coefficient was found for PDO$_{Aug}$- NINO 3.4$_{Aug}$ combination while the correlation values in calibration and validation periods were 0.40 and 0.953 respectively. But for this model, there is huge difference between the correlation values of calibration and validation periods and also the model could predict streamflow only one month in advance. Thus, this model was not considered as the best model. The combination of IODJune- NINO 3.4June was chosen as the best model for this station with correlation values 0.380 and 0.372 for calibration and validation periods respectively. This model was developed using the cubic function for both indices which was based on the outcomes of single non-linear regression analysis (Table 26). Further statistical performance analysis (Table 49) showed lower statistical errors for this model. This model can predict future streamflow for three months in advance.

The time series plots (Figure 21) of observed and simulated flow for calibration and validation periods show that the model follows the trend of observed flow mostly, though they fail to predict extreme points, i.e. very high or low values.



(a)



(b)

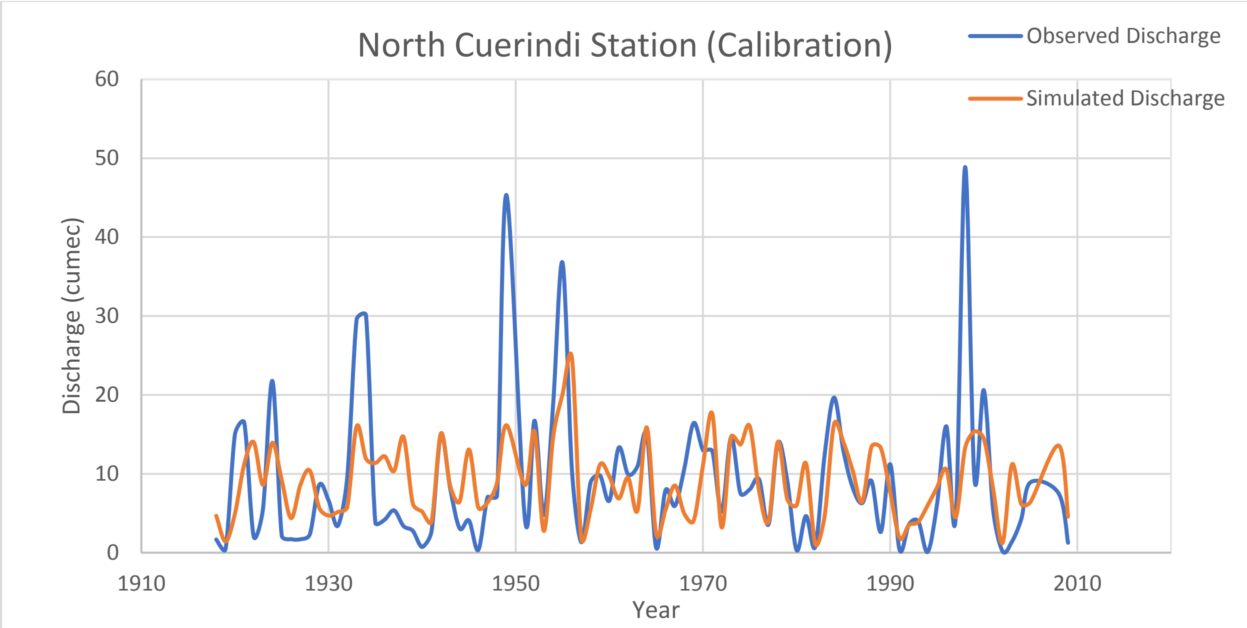
Figure 21. Comparison between the observed and simulated streamflow during the (a) calibration and (b) validation periods for Coggan Station

North Cuerindi Station

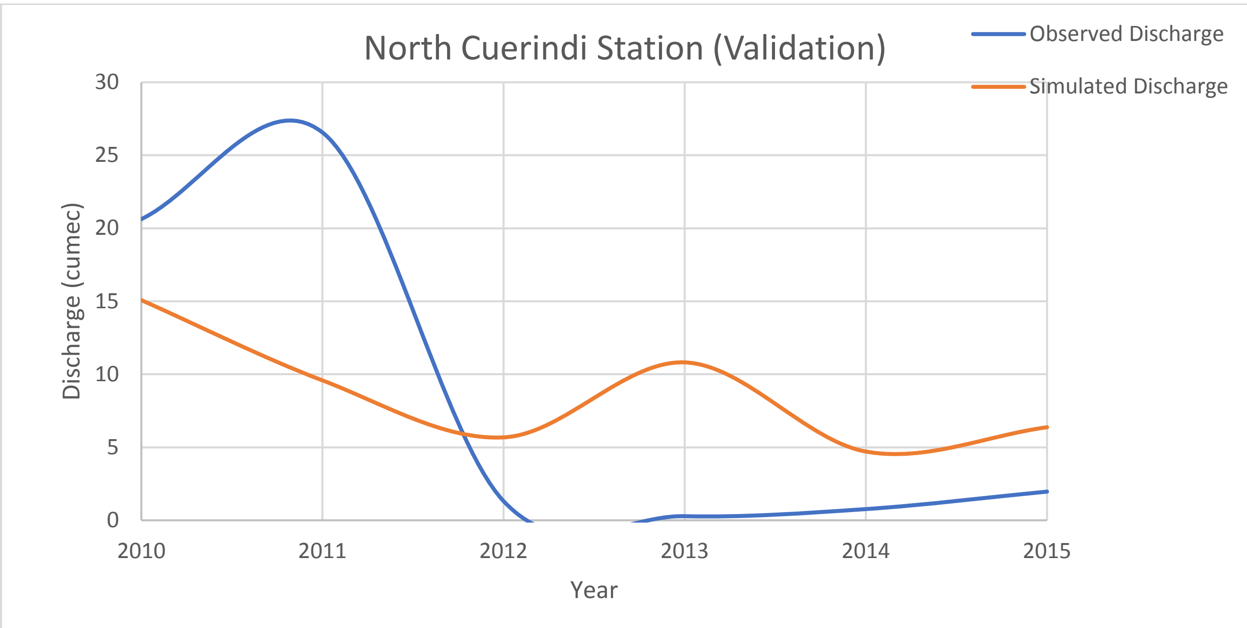
From single non-linear regression analysis (Table 27), it can be seen that the maximum value for the single correlation value was obtained for cubic function for all indices. $NINO3.4_{July}$ provided the highest correlation value for single non-linear regression analysis, which is 0.520. It was observed that for the single linear correlation analysis the maximum value was 0.483.

For MNLR analysis, the maximum value of correlation coefficient in the calibration stage was found to be 0.619 for PDO_{April}- NINO $3.4_{June,}$ whereas in the validation stage, correlation was 0.930 for PDO_{Aug-} NINO 3.4_{Aug} (Table 39). Since the difference between correlation values of calibration and validation period was high for both of these two combinations along with the low lag period months, these models were not considered as the best models. Hence the combination PDO_{April}-$NINO3.4_{June}$ with correlation coefficient 0.534 and 0.623 in calibration and validation periods respectively was selected as the best model and further statistical performance analysis (Table 49) showed lower statistical errors for this model. This model was developed using cubic function for both indices which was based on the outcomes of single non-linear regression analysis (Table 27). This model can predict future streamflow for three months in advance

The time series plots (Figure 22) of observed and simulated flow for calibration and validation periods show that the model follows the trend of observed flow mostly, though they fail to predict extreme points, i.e. very high or low values.



(a)



(b)

Figure 22. Comparison between the observed and simulated streamflow during the (a) calibration and (b) validation periods for North Cuerindi Station

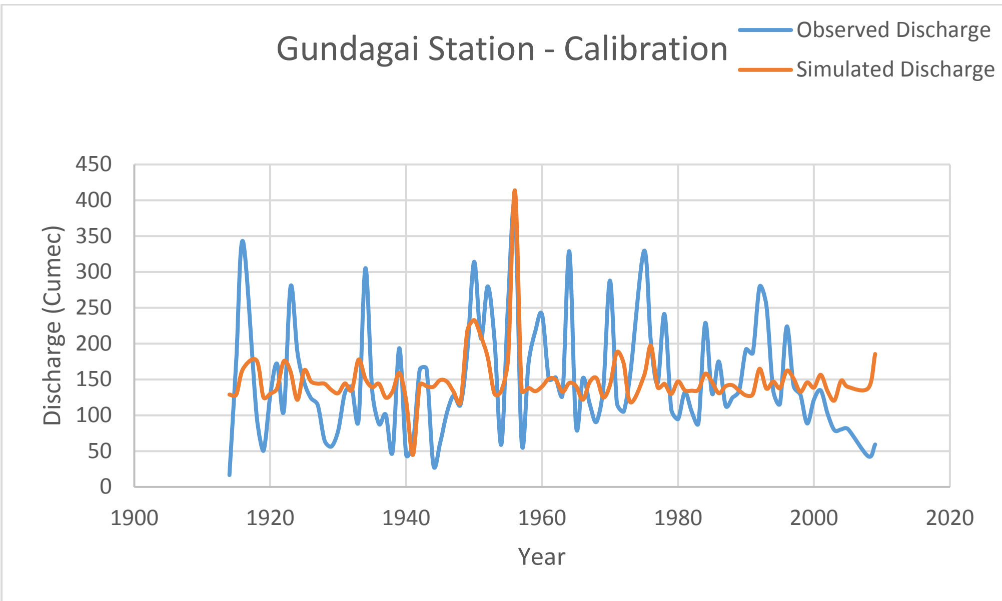
Southern New South Wales (SNSW)

Gundagai Station

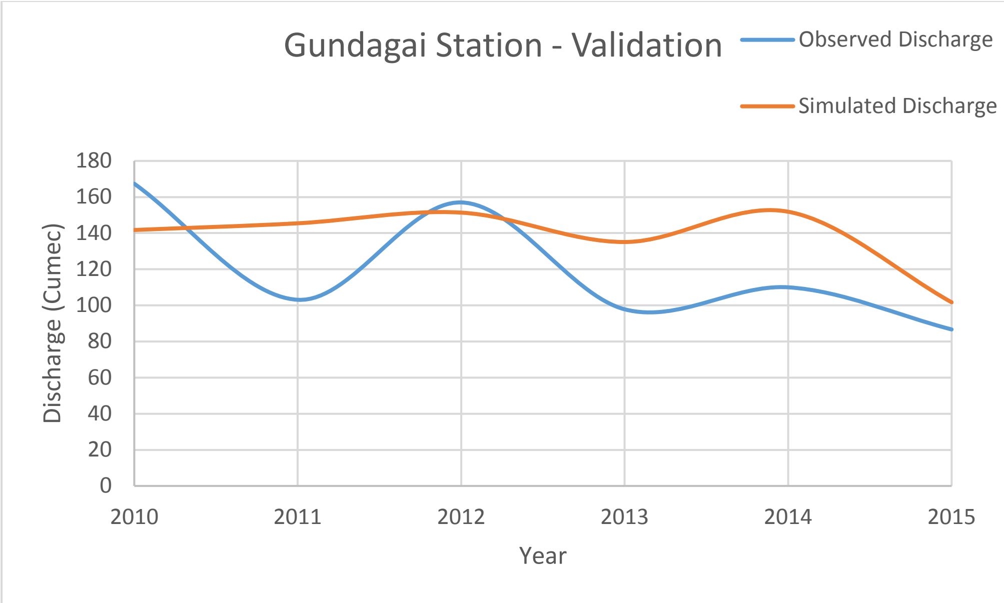
For single Non-Linear Regression analysis (Table 28), all indices showed the highest correlation for cubic function. The highest correlation value was obtained as 0.451 for PDO_{March}. In comparison, the linear regression could only yield the maximum correlation of 0.375.

Upon further proceeding, to MNLR the highest correlation value was obtained for PDO_{March}-$NINO3.4_{March}$ combination (Table 43) where correlation values for calibration and validation periods were 0.468 and 0.535 respectively which are very close to each other while the statistical performance was also satisfied by lower errors (Table 50). This model was developed using the cubic function for both indices, which was based on the outcomes of single non-linear regression analysis (Table 28). This model enables the prediction of streamflow five months in advance.

The time series plots (Figure 23) of observed and simulated flow for calibration and validation periods show that the model follows the trend of observed flow mostly, though they fail to predict extreme points, i.e., very high or low values.



(a)



(b)

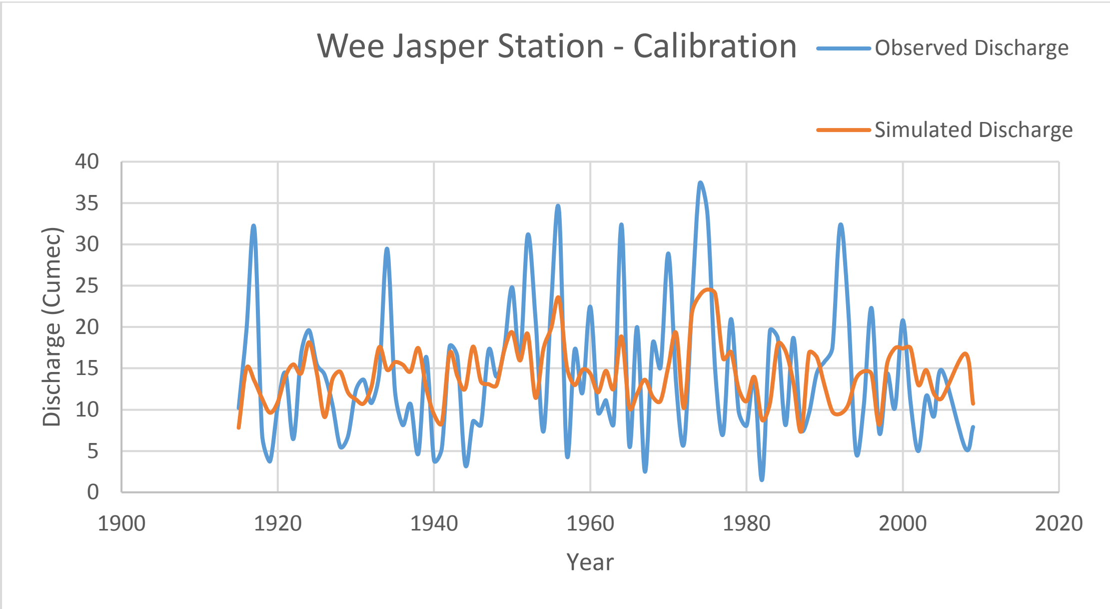
Figure 23. Comparison between the observed and simulated streamflow during the (a) calibration and (b) validation periods for Gundagai Station

Wee Jasper Station

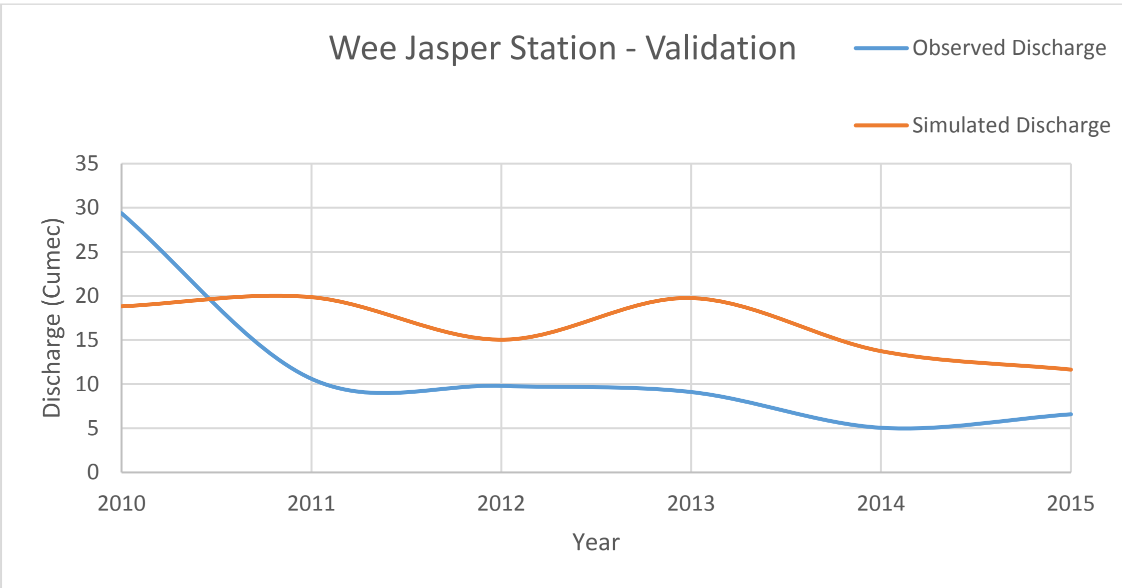
For single Non-Linear Regression analysis (Table 29), all indices showed the highest correlation for cubic function. The highest correlation value was obtained as 0.471 for $NINO3.4_{July}$. In comparison, the linear regression could only yield the maximum correlation of 0.411.

For MNLR analysis, the highest correlation value was obtained for IPO_{July}-$NINO3.4_{July}$ combination where the correlation values for the calibration period was 0.487 and for validation was 0.518 (Table 44). But since the lag period was only one month, the second-highest correlation was selected. The model was derived using IPO_{June}-$NINO3.4_{June}$ combination which resulted in correlation values 0.451 for calibration and 0.484 for validation which is very close to each other while the statistical performance was also satisfied by lower errors (Table 50). This model was developed using the cubic function for both indices, which was based on the outcomes of single non-linear regression analysis (Table 29). This model predicts streamflow a month in advance.

The time series plots (Figure 24) of observed and simulated flow for calibration and validation periods show that the model follows the trend of observed flow mostly, though they fail to predict extreme points, i.e., very high or low values.



(a)



(b)

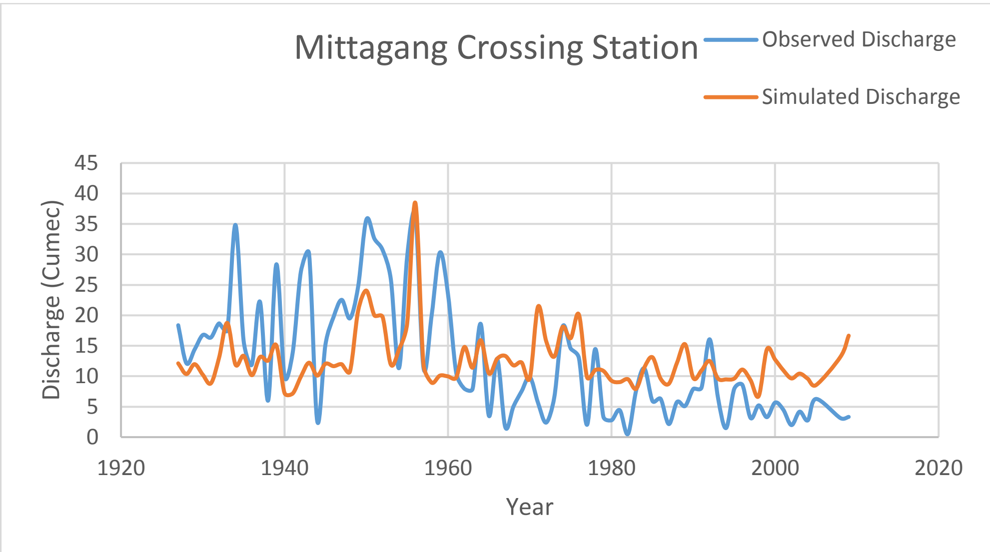
Figure 24. Comparison between the observed and simulated streamflow during the (a) calibration and (b) validation periods for Wee Jasper Station

Mittagang Crossing

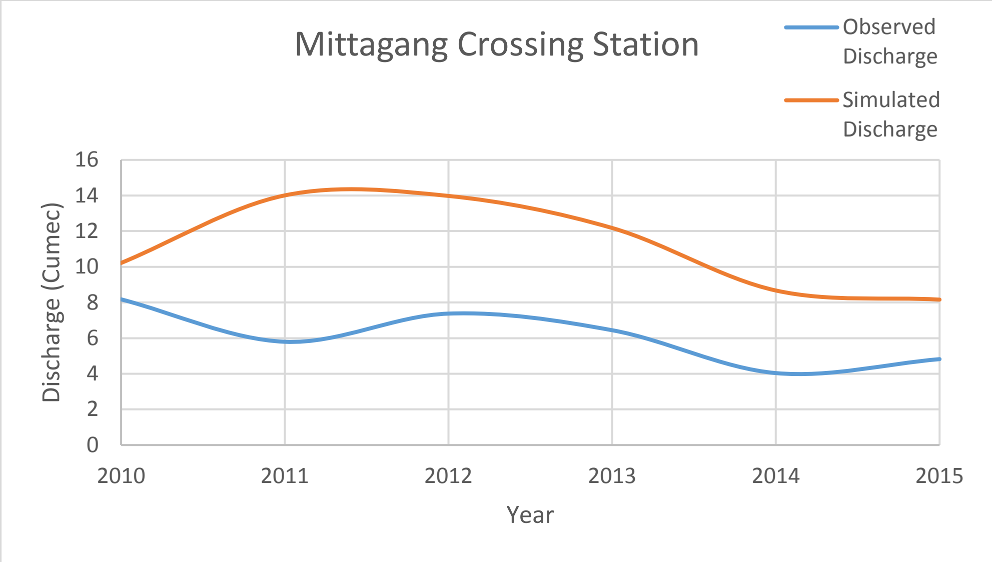
For this Station, upon performing the single non-linear regression, the highest correlations were obtained for cubic function for all indices while the maximum correlation was 0.462 for PDO_{March}. The single linear regression gave the highest correlation value of 0.379 for PDO_{March} (Table 30).

For Multiple Non-Linear Regression analysis, in Table 45, the highest correlation was obtained for PDO_{March}-$NINO3.4_{May}$ combination. The value of correlations for calibration was 0.478 and for validation was 0.505 along with lower statistical errors (Table 50). This model was developed using the cubic function for both indices, which was based on the outcomes of single non-linear regression analysis (Table 30). With this model, the prediction of streamflow could be made four months ahead.

The time series plots (Figure 25) of observed and simulated flow for calibration and validation periods show that the model follows the trend of observed flow mostly, though they fail to predict extreme points, i.e., very high or low values.



(a)



(b)

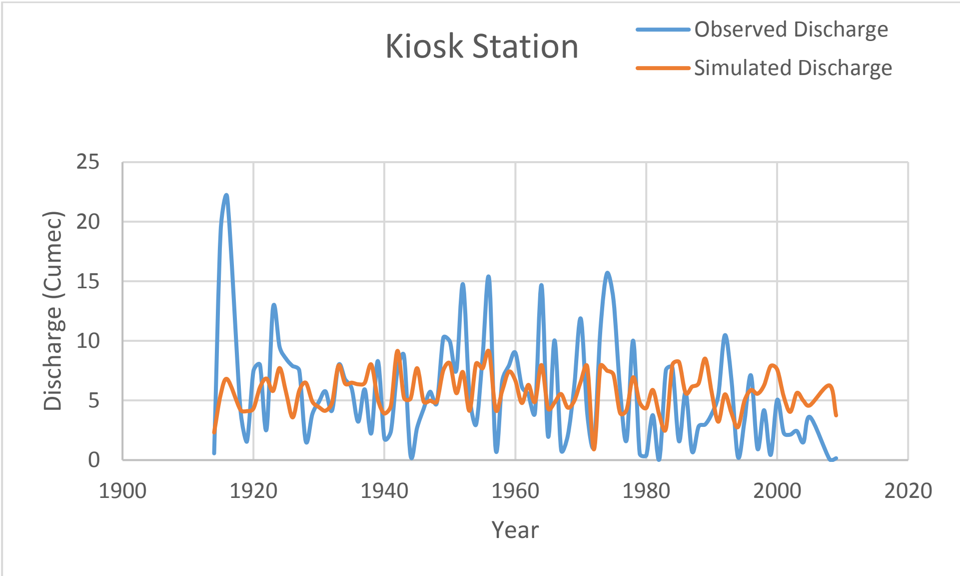
Figure 25. Comparison between the observed and simulated streamflow during the (a) calibration and (b) validation periods for Mittagang Crossing Station

Kiosk Station

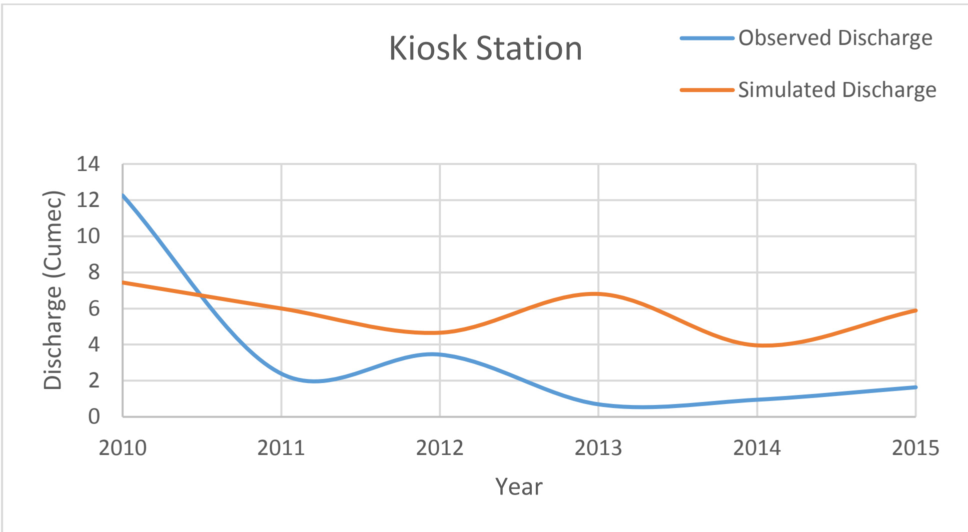
For Kiosk Station, upon performing the single non-linear regression, the highest correlations were obtained for cubic function for all indices while the maximum correlation was 0.430 for $NINO3.4_{July}$. The single linear regression gave the highest correlation value of 0.428 for $NINO3.4_{July}$ (Table 31).

Proceeding to MNLR analysis, as shown in Table 46, the best combination providing the highest correlation for calibration and validation periods was IOD_{July}_$NINO3.4_{July}$. The Pearson correlation values (r) were 0.491 and 0.506 for calibration and validation periods, respectively. Since the lag period was only one month for this combination, another combination with second highest correlation values (0.379 and 0.559 for calibration and validation period respectively) and lower statistical errors (Table 50) was selected which was the combination of IOD_{June} and $NINO3.4_{June}$. This model was developed using the cubic function for both indices, which was based on the outcomes of single non-linear regression analysis (Table 31). This model could predict streamflow for two months in advance.

The time series plots (Figure 26) of observed and simulated flow for calibration and validation periods show that the model follows the trend of observed flow mostly, though they fail to predict extreme points, i.e., very high or low values.



(a)



(b)

Figure 26. Comparison between the observed and simulated streamflow during the (a) calibration and (b) validation periods for Kiosk Station

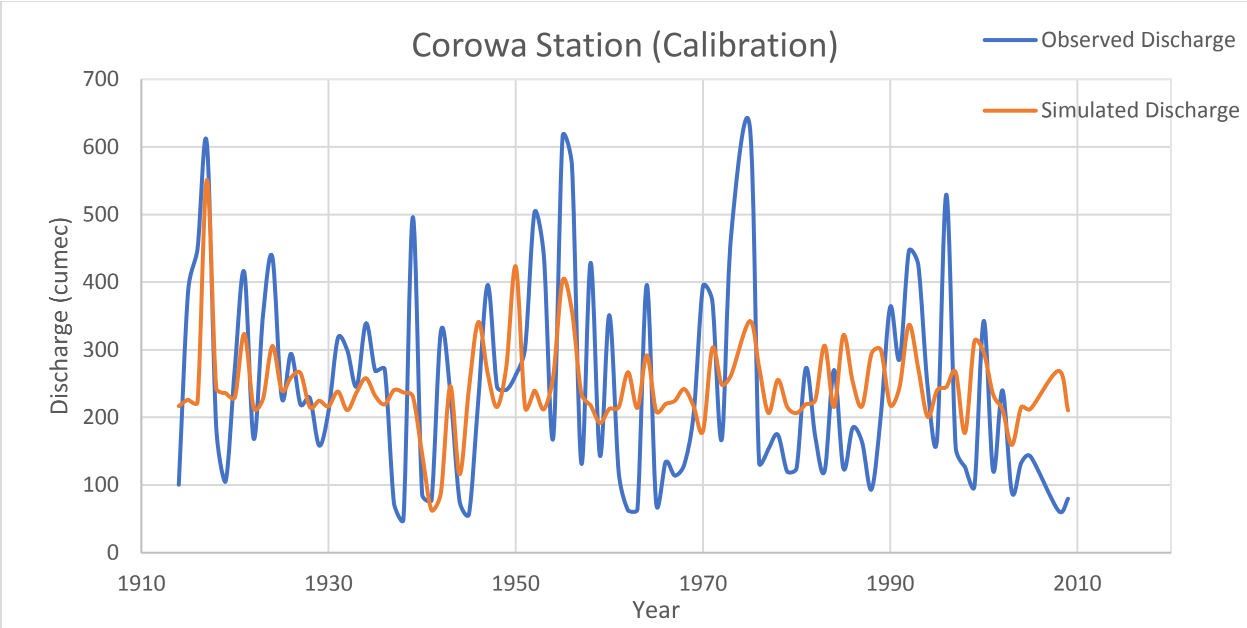
Central New South Wales (CNSW)

Corowa station

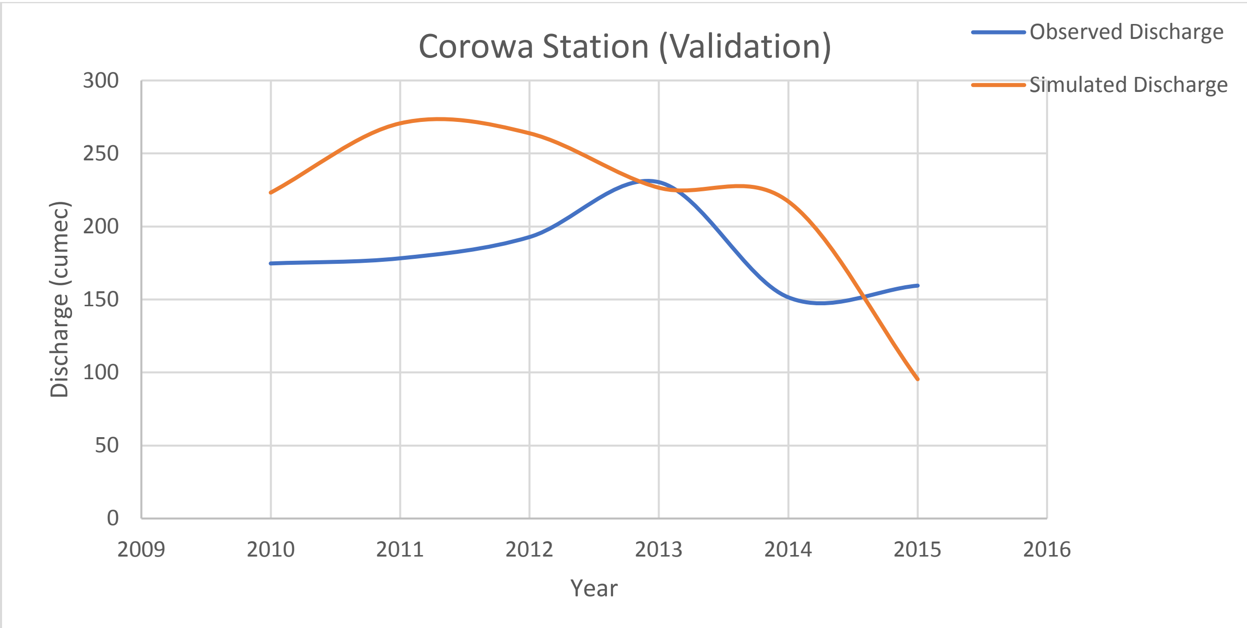
For single Non-Linear Regression analysis (Table 32), all indices showed the highest correlation for cubic function. The highest correlation value was obtained as 0.414 for IOD_{Aug}. It was observed, for single linear correlation analysis, the maximum value was 0.391.

Since cubic function produced the highest correlations for all indices, this function was chosen to form the combined non-linear models with the combination of two indices. The highest correlation for the combined model (Table 40) in the calibration stage was found for IOD_{June}- $NINO3.4_{May}$ model (0.525) and in the validation stage for IOD_{June}-NINO 3.4_{May} model (0.737). Since the difference of correlation values between calibration and validation periods are significant along with lower lagged months for both of these models, the best model was chosen as PDO_{Dec} – NINO 3.4_{May} with a better set of correlation values (0.434 in calibration period and 0.388 in validation period) and lower statistical errors (Table 51). The best model is able to predict streamflow for four months in advance.

The time series plots (Figure 27) of observed and simulated flow for calibration and validation periods show that the model follows the trend of observed flow mostly, though they fail to predict extreme points, i.e., very high or low values.



(a)



(b)

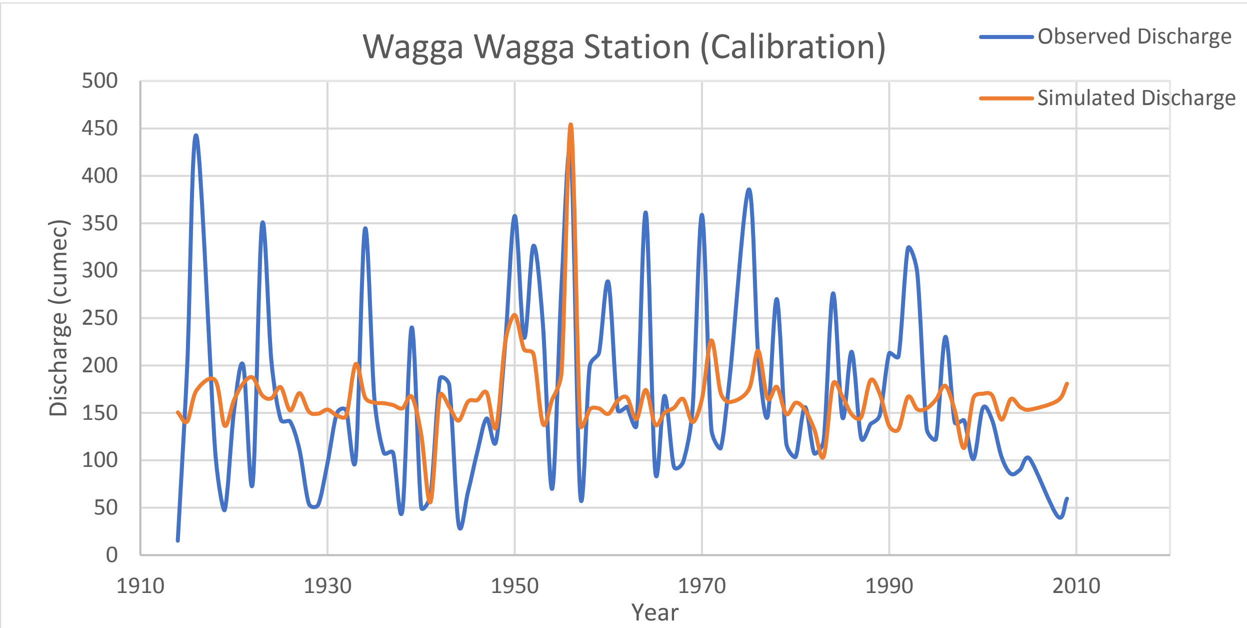
Figure 27. Comparison between the observed and simulated streamflow during the (a) calibration and (b) validation periods for Corowa Station

Wagga Wagga station

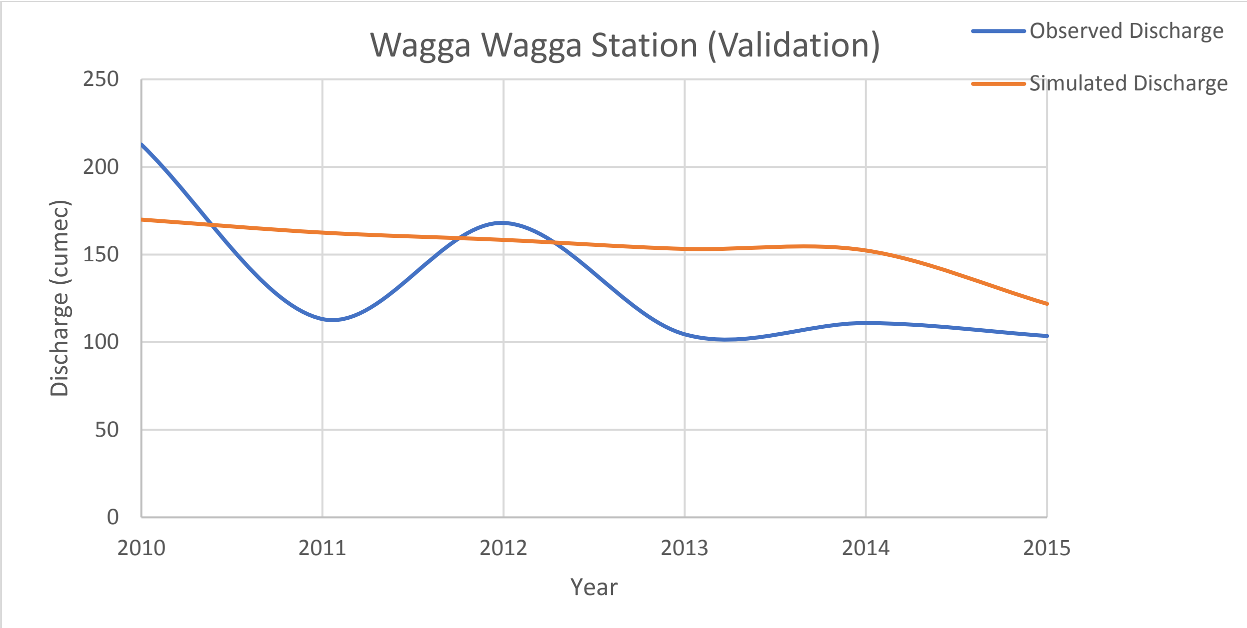
For single Non-Linear Regression analysis (Table 33), all indices showed the highest correlation for cubic function. The highest correlation value was obtained as 0.453 for IOD_{Aug}. It was observed, for single linear correlation analysis, the maximum value was 0.451.

Since cubic function produced the highest correlations for all indices, this function was chosen to form the combined non-linear models with the combination of two indices. The highest correlation for the combined model (Table 41) in the calibration stage was found for PDO_{April}-NINO3.4_{April} model (0.469) and in validation stage for IOD_{June}-NINO 3.4_{May} model (0.541). Since the difference of correlation values between calibration and validation periods are significant along with lower lagged months for both of these models, the best model was chosen as PDO_{Mar} – NINO 3.4_{Mar} with a better set of correlation values (0.430 in calibration period and 0.350 in validation period) and lower statistical errors (Table 51). The best model is able to predict streamflow for six months in advance.

The time series plots (Figure 28) of observed and simulated flow for calibration and validation periods show that the model follows the trend of observed flow mostly, though they fail to predict extreme points, i.e., very high or low values.



(a)



(b)

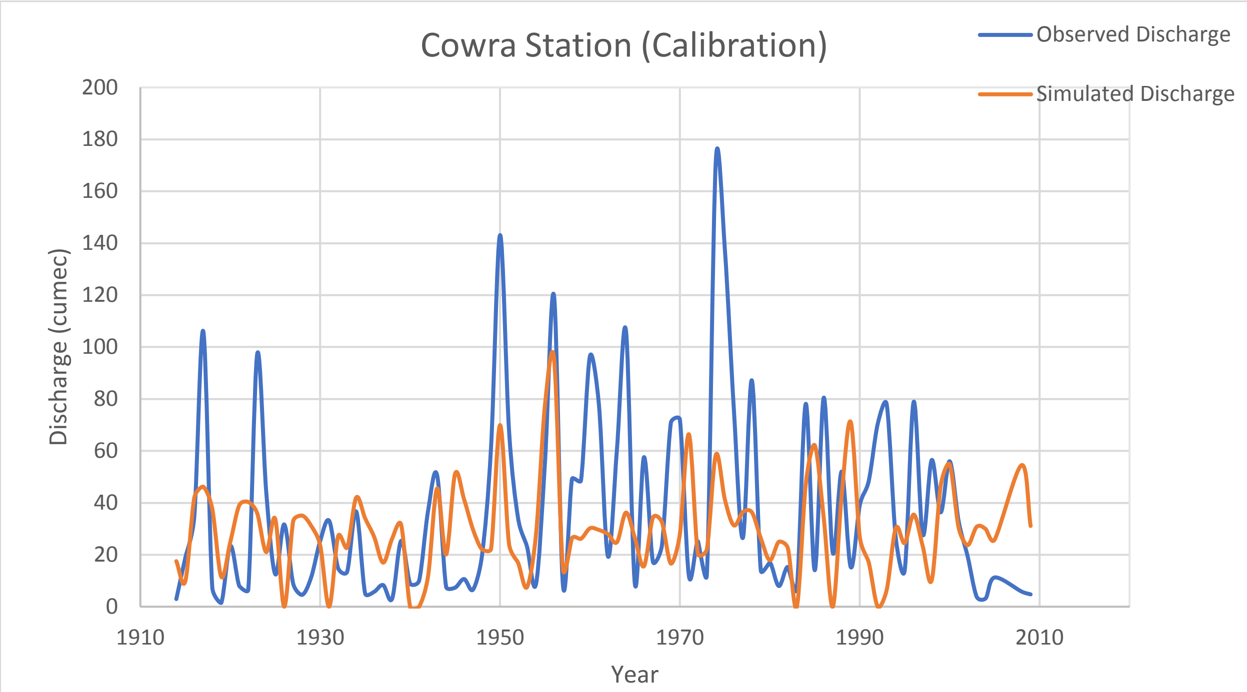
Figure 28. Comparison between the observed and simulated streamflow during the (a) calibration and (b) validation periods for WaggaWagga Station

Cowra Station

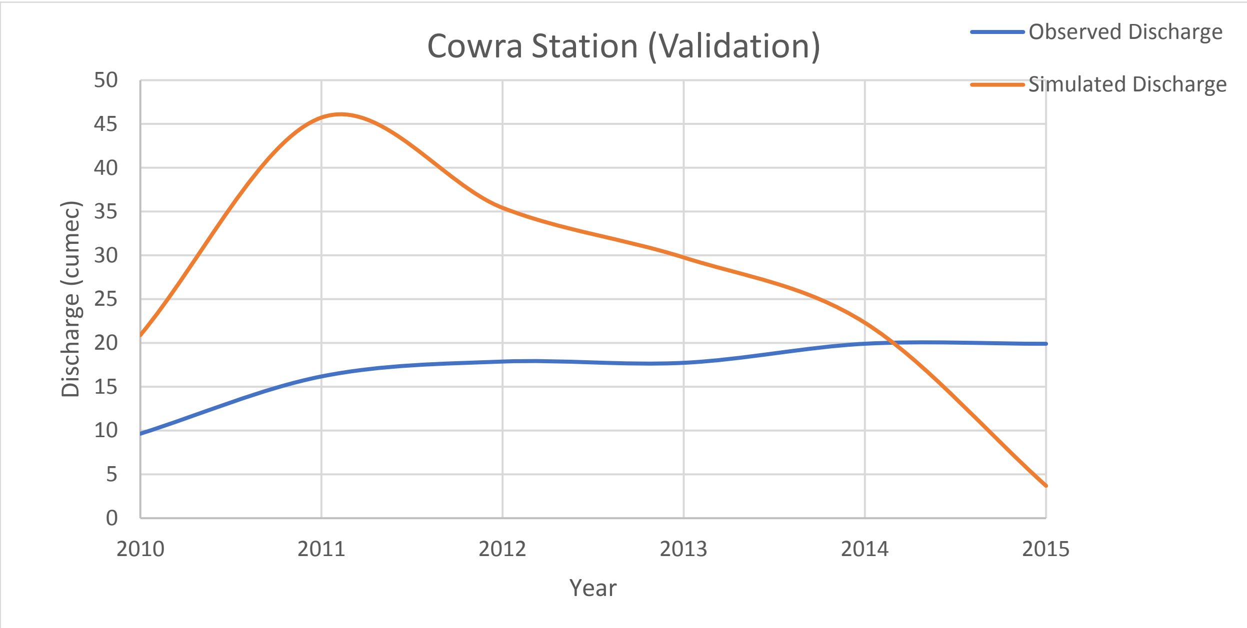
For single Non-Linear Regression analysis (Table 34), all indices showed the highest correlation for cubic function. The highest correlation value was obtained as 0.453 for PDO_{March}. It was observed, for single linear correlation analysis, the maximum value was only 0.248.

Since cubic function produced the highest correlations for all indices, this function was chosen to form the combined non-linear models with the combination of two indices. The highest correlation for the combined model (Table 42) in the calibration stage was found for PDO_{March}-$NINO3.4_{Feb}$ model (0.485) and in validation stage for PDO_{April}-$NINO\ 3.4_{April}$ model (0.105). Since the difference of correlation values between calibration and validation periods are significant along with lower lagged months for both of these models, the best model was chosen as PDO_{Mar} – $NINO\ 3.4_{Mar}$ with a better set of correlation values (0.436 in calibration period and 0.105 in validation period) and lower statistical errors (Table 51). The best model is able to predict streamflow for six months in advance.

The time series plots (Figure 29) of observed and simulated flow for calibration and validation periods show that the model follows the trend of observed flow mostly, though they fail to predict extreme points, i.e., very high or low values.



(a)



(b)

Figure 29. Comparison between the observed and simulated streamflow during the (a) calibration and (b) validation periods for Cowra Station

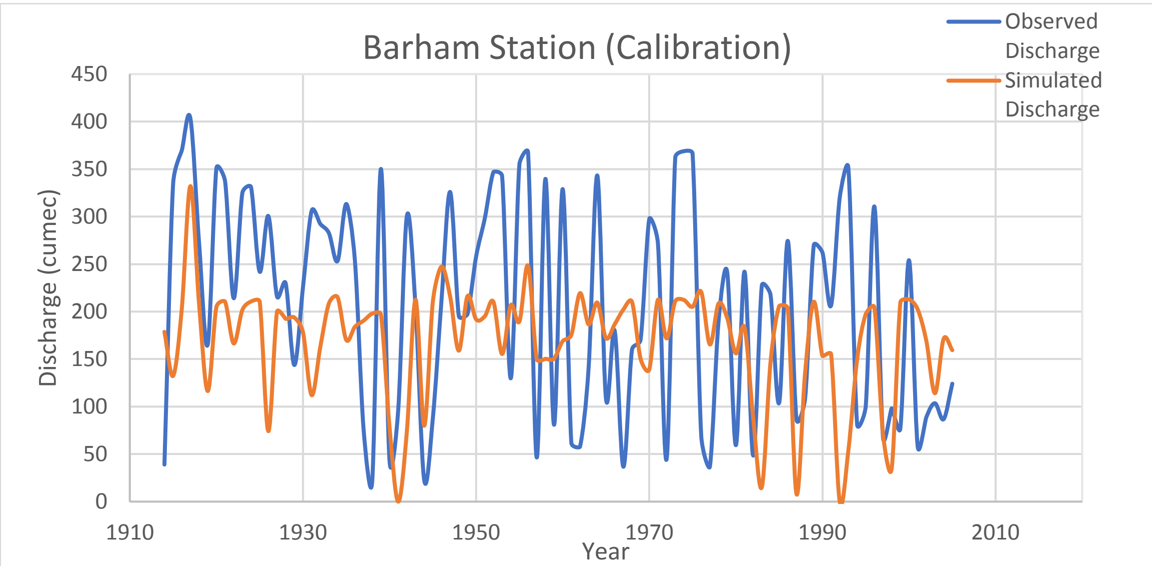
Western New South Wales (WNSW)

Barham station

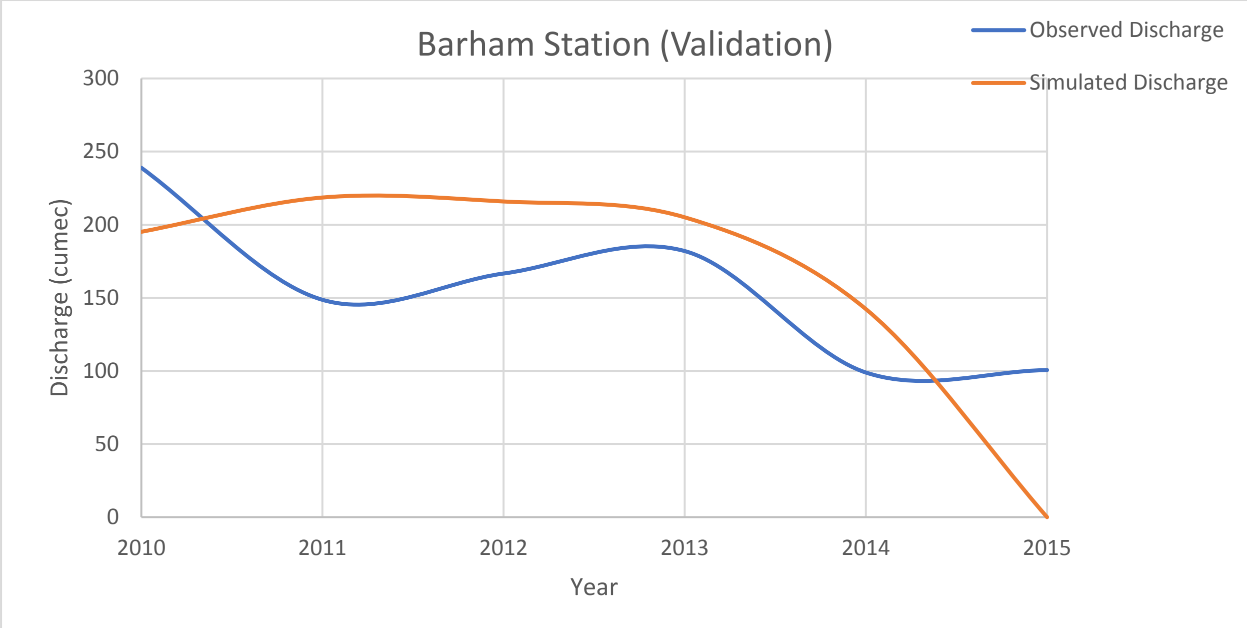
From the result shown in Table 35, it can be seen that the maximum correlation for single non-linear regression analysis was obtained for the quadratic function, which is 0.407 for NINO3.4$_{July}$. For all other indices, the highest correlation values were obtained for cubic function. It was observed that for the single linear correlation analysis, the maximum value was 0.39.

For MNLR analysis (Table 47), the highest correlation values in the calibration stage was found to be 0.484 for IOD$_{July}$- NINO 3.4$_{May}$ whereas in validation stage, was 0.846 for IOD$_{June}$- NINO 3.4$_{May}$. But since the lag period was only one month, the second-highest correlation was selected. The model was derived using IPO$_{June}$-NINO3.4$_{June}$ combination which resulted in correlation values 0.451 for calibration and 0.484 for validation which are very close to each other while the statistical performance was also satisfied by lower errors (Table 52). This model was developed using the cubic function for both indices, which was based on the outcomes of single non-linear regression analysis (Table 35). This model predicts streamflow three months in advance.

The time series plots (Figure 30) of observed and simulated flow for calibration and validation periods show that the model follows the trend of observed flow mostly, though they fail to predict extreme points, i.e., very high or low values.



(a)



(b)

Figure 30. Comparison between the observed and simulated streamflow during the (a) calibration and (b) validation periods for Barham Station

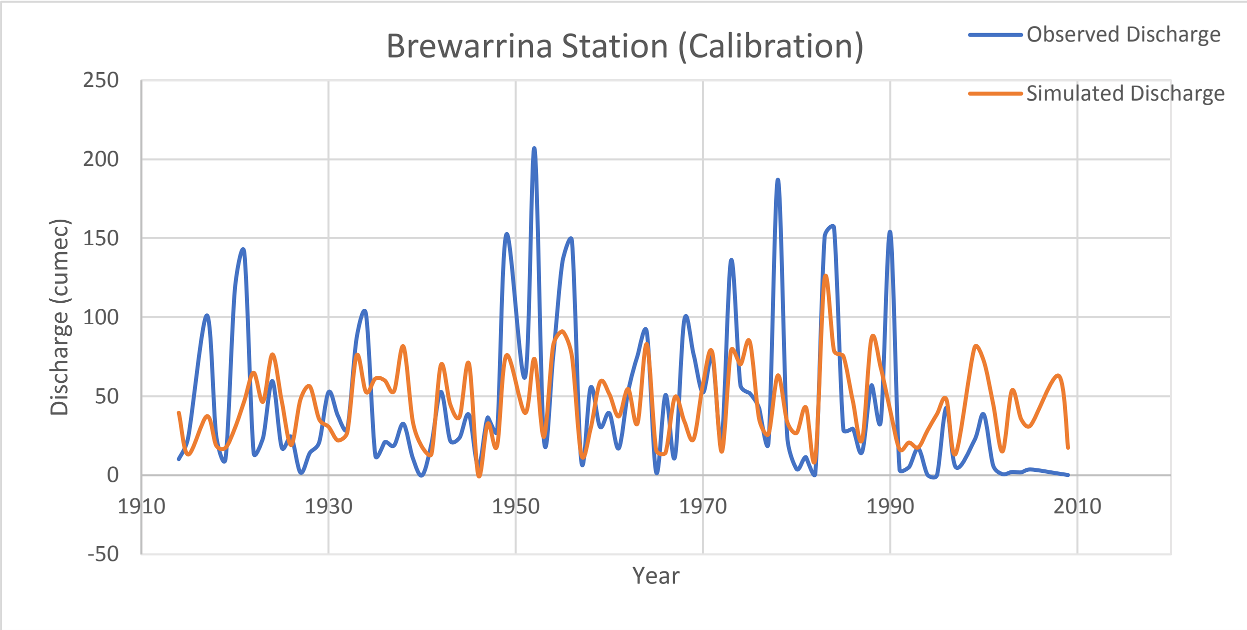
Brewarrina Station

For single non-linear regression analysis (Table 36) was the maximum correlation values were obtained for cubic function for all indices while the highest correlation was 0.441 for NINO3.4$_{June}$. It was observed that for single linear correlation analysis, the maximum value was 0.427 for NINO3.4$_{June}$.

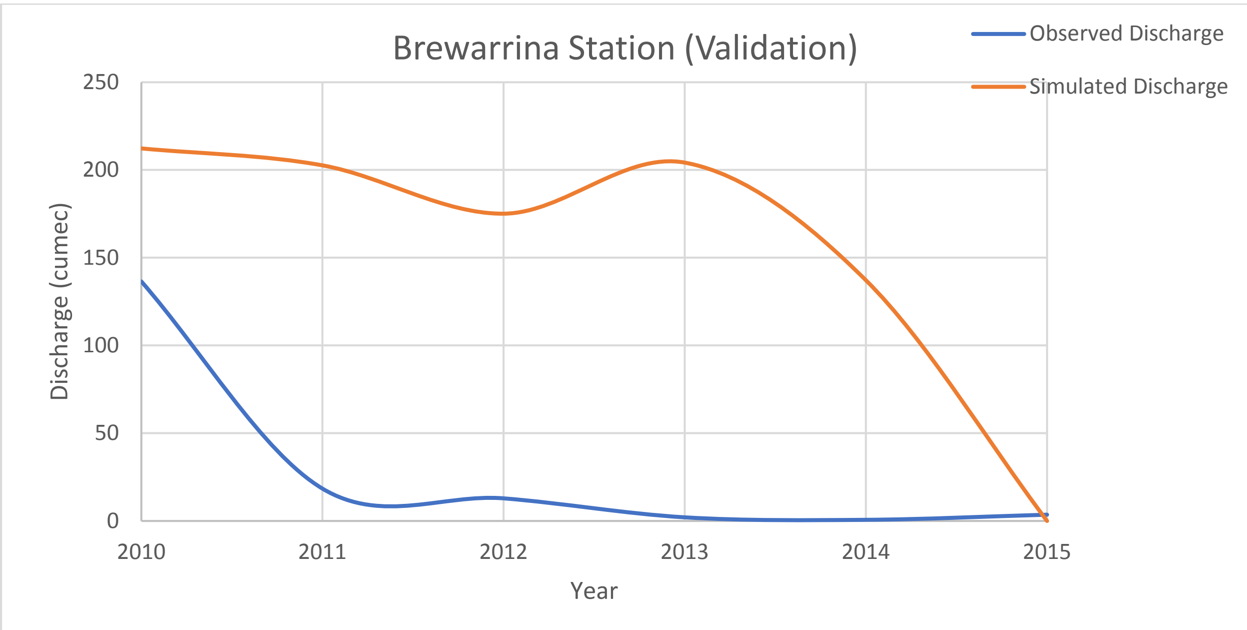
For MNLR analysis(Table 48) the highest correlation value was obtained for EMI$_{July}$-NINO3.4$_{July}$ combination where the correlation values for calibration and validation periods were 0.465 and

0.767 respectively which are close to each other while the statistical performance was also satisfied by lower errors (Table 52). This model was developed using the cubic function for both indices, which was based on the outcomes of single non-linear regression analysis (Table 36). This model predicts streamflow two months in advance.

The time series plots (Figure 31) of observed and simulated flow for calibration and validation periods show that the model follows the trend of observed flow mostly, though they fail to predict extreme points i.e., very high or low values.



(a)



(b)

Figure 31. Comparison between the observed and simulated streamflow during the (a) calibration and (b) validation periods for Brewarrina Station

6.5 Summary of MNLR Analysis

The MNLR analysis was carried out to explore the non-linear relationship between streamflow and climate indices. Five different non-linear functions along with the linear function were used to perform a single correlation analysis between spring streamflow and single lagged climate indices which was followed by Multiple Non-Linear Regression (MNLR) analysis. The function which provided the highest correlation among the used six functions (linear and five non-linear) were chosen to develop the MNLR equations. Finally, the best model for each station was selected based on its higher correlation values and better statistical performance.

From the single linear regression analysis, it is observed that all the indices for almost all stations (except Barham and Coggan stations which showed the highest correlation for quadratic function) showed highest correlations for the cubic function which implies that cubic function has comparatively more potential to explain the relationship between spring streamflow and lagged climate indices. For every station, the non-linear function had higher correlation values than the linear function, which referred to the underlying relationship between spring streamflow and lagged climate indices is more likely to be non-linear.

Based on the outcomes of single linear regression analysis, combined MNLR models were developed with the combination of two different climate indices. The non-linear function that showed the highest correlation in single non-linear regression analysis was used for the corresponding index in order to form the combined model. Out of the twelve stations, eight stations had the best models combined of PDO and NINO3.4 indices implying the stronger influence of these two indices on spring streamflow of NSW. Among the other four stations, Coggan and Kiosk had the best model consisting of IOD and NINO3.4 indices while Wee Jasper had the best model with IPO and NINO3.4 and Brewarrina had the best model with EMI and NINO3.4 indices. Hence, it was evident that PDO and ENSO indices have the strongest impact on spring streamflow of NSW.

Statistical performances of the developed models were analysed to ensure the reliability of the models. Different statistical measures, including MAE, RMSE, d and VIF, were used to check the reliability of the models. The models with higher correlation values and lower errors were selected as the best models. The Pearson correlation values in calibration and validation stages were quite similar, which implies the good performance of the models. The best-developed models were able to predict streamflow from three to six months in advance.

The ability of the best MNLR model from each station to predict future streamflow has been explained through the time series plots of observed and simulated flows in Figure 20-31. In the time series plot, some differences can be identified between the observed and simulated flow for a few years. The reason can be that a regression model based on only two climate indices (e.g. PDO and NINO3.4) are not expected to capture the unusual phenomenon like severe droughts (e.g. millennium drought from 1994-2010) and floods. Another reason is that some other climate indices might have been more influential at that time rather than the selected indices in this study.

Chapter 7
Gene Expression Programming

7.1 Introduction

High interannual variability of streamflow in Australia presents the challenge to the hydrologists and researchers to develop reliable streamflow forecast models that can help the water stakeholders in making low-risk decisions at the earlier stage of the crop period which may enhance the potential of economic benefit as well (White et al. 2004; Abawi et al. 2005).

Streamflow is largely dependent on the initial catchment and future climatic conditions (Robertson and Wang, 2009). As the remote climate drivers fluctuate at very low frequencies, they have better predictability of streamflow while comparing to initial catchment condition. Even though developing sophisticated models incorporating different hydrological and hydro-meteorological variables such as antecedent moisture content of soil, evapotranspiration, precipitation and temperature is possible, it will be economically preferable if a simple mathematical model using climate indices can simulate the streamflow with enough reliability. Many hydrologists established the existence of strong correlations between streamflow and large-scale climate drivers, though the nature of the relationship remained a difficult question to deal with. According to Piechota et al. (1998), the relationship between streamflow and remote climate drivers is more likely to be non-linear; thus a non-linear model is expected to give better solutions than a linear model.

There are five different forecasting methods in practice (Qi and Chang, 2011) which are time series analysis, regression analysis, artificial intelligence method (e.g. ANN, fuzzy logic etc.), the hybrid and Monte Carlo simulation methods. Currently, data-driven (DD) models have attained much popularity compared to the physical-based models due to their unique qualities such as minimum data requirement, the ability of fast application and time-saving. Some of the existing data-driven techniques include statistical methods like Simple Regression (SR), Multi-Linear Regression (MLR) and

Auto-Regressive Moving Average (ARMA) in addition to Artificial Intelligence (AI) methods like Artificial Neural Networks (ANNs), Adaptive Network-Base Fuzzy Inference System (ANFIS), Genetic Programming (GP) and Genetic Expression Programming (GEP) (Savic et al. 1999; Sajikumar and Thandaveswara 1999; Maier and Dandy 2000; Khu et al. 2001; Nourani et al. 2011; Traore and Guven 2012; Kişi et al. 2012; Kiafar et al. 2017; Sattar et al. 2016).

Though the difficulties in dealing with artificial intelligence models encourage the users to attempt comparatively simple statistical models (Adamowski 2012), the limitations of statistical models are evident when data become complex. Chiew et al. (2003) proposed non-linear regression-based models to attain higher correlation values between streamflow and climate indices as it enables to capture the underlying non-linear relationship between two variables. Again, one of the major advantages of artificial intelligence-based models like GEP and ANN over regression-based models is, they do not impose any fixed model structure on the data, rather they allow the data itself to identify the model structure by using artificial intelligence (Aziz et al. 2013).

Recently, artificial intelligence (AI) has got immense attention for its wide range of successful applications in the field of water resources engineering, agro-hydrology and agro-meteorology (Tayfur 2002, Coulibaly et al. 2001, Kişi 2006a, b; Kişi 2007b; Maier and Dandy 2000; Shiri and Kişi 2010, 2011; Supharatid 2003; Shiri et al. 2012). ANN and hybrid ANN techniques were employed Wang et al. (2006) for forecasting streamflow. Linear genetic programming and ANN models were applied to estimate daily pan evaporation by Guven and Kişi (2011). Short-term and long-term river flows were predicted by applying GEP and ANFIS in the study of Kişi and Shiri (2010). They also conducted a comparative analysis between GEP and ANFIS to predict groundwater table depth fluctuations (Shiri and Kişi 2011a). Again, they attempted to compare GEP, ANFIS and ANN models for predicting daily pan evaporation values (Shiri and Kişi 2011b). GP and ANN were exploited for predicting and modelling rainfall-runoff transformation by Dorado et al. (2003). GP and ANN were applied in the study of Rabunal et al. (2007) in order to determine the unit hydrograph of an urban basin.

Guven and Gunal (2008) used GEP to explore the maximum scour depth and location in downstream of grade control structures.

Throughout the years, researchers have been trying to develop the data-driven techniques from black-box to semi-explicit to transparent form. One of the main advantages of GEP models over some other data-driven models (for example ANN) is that the resultant model is not a complete “black-box”, rather the relationship between input (climate indices) and output (streamflow) variables can be explained with a mathematical expression (the combination of basic operators and functions). Being a transparent model GEP may help the users to understand the underlying hydrological process between the climate mode and streamflow without having much knowledge about the used software such as ANN, Genexpro tools etc.

GEP was found to give better performance than other data-driven methods such as ANN and ANFIS (Aytek and Alp, 2008, Azamathullah et al., 2011; Guven and Aytek, 2009; Kişi et al., 2012, Kişi et al., 2013; Shiri et al., 2012). Kişi et al. (2012) investigated the comparative performance of ANN, ANFIS, GEP and ARMA models to forecast lake levels in Turkey and concluded that GEP was the better performer among all other data-driven models. GEP along with ANFIS, Priestley-Taylor and Hargreaves-Samani models was employed to estimate daily evapotranspiration in Northern Spain where the results revealed the best performance of GEP model followed by ANFIS model (Shiri et al. 2012). GEP model was suggested as a feasible alternative to ANN, ANFIS and MLR time series when these models were applied to simulate the rainfall-runoff transformation process (Kişi et al. 2013). A wavelet-GEP model was applied by Kişi and Shiri (2011) to forecast precipitation.

GEP has been applied to various fields including artificial intelligence, artificial life, engineering, science, industrial, biological chemical processes, and financial markets to solve problems like symbolic regression, time series prediction, evolutionary neural networks and so on (Samadianfard, 2012).

GEP was deployed to solve a number of hydrological and hydraulic modelling problems, for instance, the stage-discharge relationship models were developed and compared with traditional methods by Guven and Aytek (2009) while he found the best outputs from GEP models. Many other investigations were carried out with GEP to establish functional relationships of sediment transport in sewer pipe systems (Ghani and Azamathulla, 2012), estimate the flow discharge in compound channels (Zahiri and Eghbali, 2012), measure evapotranspiration using daily climate variables (Aytek and Kişi, 2008) etc. Kişi and Shiri (2012) investigated the implementation of GEP models in many studies which include but not limited to modelling river suspended sediment load by using climate indices.GP was applied by Kişi et al. (2013) for rainfall-runoff modelling. Azamathulla and Ghani (2011) predicted longitudinal dispersion coefficients in streams using GP. GP and GEP were attempted by Zakaria et al. (2010), and Azamathulla et al. (2010) for their studies on sediment transport and bridge pier scour respectively. Several studies applied GP for rainfall-runoff modelling (Savic et al. 1999; Whigham and Crapper 1999; Babovic and Keijzer 2002). GP was also applied to develop sedimentary particle settling velocity equations (Babovic and Keijzer 2002). A study was carried out to predict velocity in compound channels using GP (Harris et al. 2003). Chezy resistance co-efficient in corrugated channels was determined by using GP (Giustolisi 2004).

Till to date, many researchers have studied the relationship between Australian rainfall, streamflow and climate indices. Dutta et al. (2006) indicated the necessity for exploring the skills of forecasting streamflow and rainfall with different lead time exploiting various climate indicators. According to him, streamflow forecast is more significant compared to rainfall forecast as it can be predicted with longer lead times. Thereby, streamflow forecast enables the water users to make the decision at the earlier stage of the year, which ultimately increase the potential of financial benefits. Again, it is evident from the study of Kirono et al. (2010) that statistically, significant lag relationships exist between atmospheric, oceanic variables (thermocline, SOI and NINO4) and winter, summer and spring runoff in the northern part of Moree of northern NSW, which is better than the relation with antecedent runoff.

Thus, considering the significant role of reliable streamflow predictor models in the field of agriculture as well as the economy of Australia, the current research study aims to explore the potential of GEP for developing reliable streamflow forecast models incorporating a combination of multiple large-scale climate drivers as predictor variables. NSW has been selected as the case study region considering the agricultural importance of this state in contributing most of Australia’s agricultural production. After a careful study of the previous research works, the authors confirm that no such study on long-term streamflow prediction of NSW region by applying GEP has been approached by any researcher till date.

The study intends to provide deterministic forecast as it can play more important roles in solving water management problems by enabling the water stakeholders to take more accurate decisions knowing the predicted amount of future streamflow, compared to the probabilistic approaches which have been attempted by many researchers till date (Piechota et al. 1998; Ruiz et al., 2007; Robertson & Wang 2009; Wang & Robertson 2011; Duc et al. 2017). Furthermore, the Bayesian joint probability (BJP) method used by Australian Bureau of Meteorology (BoM, 2000) to provide futuristic streamflow is again a probabilistic method. Communication of the concept of the probabilistic forecast remains a challenge, whereas end-user confidence is very important for the adaptation of a forecast model for decision making. Therefore, the interactions of multiple climate indices with seasonal streamflow have been explored in this study in order to obtain deterministic output models.

This particular study has focused on the prediction of spring streamflow only considering the outcomes of the past research studies in this field (McBride & Nicholls, 1983; Robertson & Wang, 2009) as well as the outcomes of the MLR and MNLR analyses which was conducted at the preliminary stage of this study. Thus, the predictability of the linear and non-linear techniques will be compared at the end of the study.

7.2 Details Methodology Involved in GEP analysis

7.2.1 GEP evolution

Evolutionary programming, which is a kind of machine learning-artificial intelligence, was developed to perform symbolic regression. There are three different variants of evolutionary programming- “Genetic Algorithms (GA)”, “genetic programming (GP)” and “gene expression programming (GEP)” (Ferreira 2006). Holland (1975) developed Genetic Algorithms (GAs) by coding those as symbolic strings of fixed length which work as ribonucleic acid (RNA) replicators in nature. Genetic programming (GP) was formulated by Cramer (1985) at first, and later on, it was promoted by Koza (1992, 1994) and Koza et al. (1999, 2003). GP was designed to use nonlinear parse trees of various size and shapes that act as protein replicators in nature. The disadvantage of this method was that many computations got wasted due to the non-functional algorithms generated from many of the mutations. The latest evolutionary programming method which is considered as the best method of this genre till date was developed by Ferreira (2001, 2006) who was inspired by Darwin’s theory of evolution. GEP programmed individuals as linear strings of fixed length which is afterwards represented by expression trees (simple diagram representation). One of the advantages of GEP over GA and GP is that genetic operators work at the chromosome level, which makes genetic diversity creation extremely simplified. Furthermore, the multigenic nature of GEP makes the evolution of complex programs composed of numerous sub-programs. GEP comprises the advantages of both GA and GP while overcoming some of their individual limitations which makes GEP 100 to 60000 times better performer than the old GP methods (Ferreira 2001, 2002, 2006). The most important advantages of GEP are (Ferreira 2001): (i) the chromosomes are small entities: linear, compact, relatively, small, easy to manipulate genetically (replicate, mutate, recombine, etc.); ii) the expression trees are exclusively the expression of their respective chromosomes; they are entities upon which selection acts, and according to fitness, they are selected to reproduce with modification.

7.2.2 GEP structure

GEP is a search technique which involves individuals of non-linear structures with different size and shape encoded in linear chromosomes of fixed lengths. The genotype, i.e., chromosomes (which are generally consisted of more than one gene of equal length) and the phenotype, i.e., expression trees or ETs (expression of the genetic information encoded in chromosomes) are the two important entities of GEP which are different both structurally and functionally. The Chromosomes of GEP are usually composed of multiple genes where each gene codes for a sub-ET and several sub-ETs connect with each other through linking function to form a more complex ET. The GEP genes consist of a head and a tail. The head contains both functions(nodes) and terminal symbols(leaves) while the tail contains only the terminal symbols. In each problem, the head length (h) is decided by the user, and tail length (t) is formulated by the equation:

$t= (n-1)\,h+1$, where n is the number of arguments of the functions.

In GEP during the reproduction process only the linear chromosomes are transmitted, as all the genetic modifications take place in the linear chromosomes. Furthermore, the unconstrained applications of important genetic operators such as mutation, transposition and recombination are allowed by the structural and functional organization of the linear chromosomes. In GEP chromosomes are selected based on their fitness values using the roulette wheel selection process where the fitter chromosomes have more chances to get passage to the next generation. Then these selected chromosomes are modified by using the genetic operators such as mutation, transposition, inversion, recombination etc. among which mutation is considered as the most efficient operator that can sometimes be used as the only genetic operation for modification purposes. This modification process continues with the new individuals until the required accuracy is achieved or a predefined number of generations is created (Ferreira 2001). The process of information decoding is termed as translation which involves a genetic code and a set of rules which are very simple in GEP. The genetic code in GEP determines one-to-one relationship between the symbols of the genes and the nodes in the ETs while the rules denote the spatial organization of nodes in the ETs and type of the links between sub-ETs.

The steps to predict streamflow are as follows:

- Initially random population is generated which is consisted of individual chromosomes of fixed length.
- Each chromosome in the initial population is expressed by the expression trees and evaluated using the predicted-observed data pairs of the training period as well as an appropriate fitness function (AE, RE and correlation coefficient (r) etc. In the current study RMSE was utilized as a fitness function to fit a curve to the target values following the steps of Shiri et al. (2012).
- The next step is to determine the set of terminals T and the set of functions F to create the chromosomes. The selection of appropriate functions which is not so obvious depends on the user's understanding and viewpoint. The functions used in this study which was selected based on a trial and error process have been explained in Table 54 However, a detailed study on the appropriate function selection process is beyond the scope of this study.
- The next major step is to select the chromosomal architecture which is composed of length of head (h), the number of genes per chromosome and genetic operators. The relevant values used in this study are stated in Table 54.
- The proper linking function (addition, multiplication, subtraction or division) needs to be chosen to connect the algebraic sub-trees which is suggested by Ferreira (2001a) to be "addition" or "multiplication", though it is evident from the recent studies that the "addition" linking function gives optimal outputs in case of linking parse trees (Shiri et al. 2012, Kişi and Shiri, 2011).
- Finally, the default values of the genetic operators need to be selected for the Genexpro program (Shiri and Kişi, 2011a). An overview of the used parameters is described in Table 54.

An outline of the construction of GEP models has been explained through Figure 32. The development of GEP models and all the relevant statistical calculations are performed using the"GeneXpro tools 5.0" software.

Table 54. List of Parameters Used in the Study

Parameter	Value
Function set	+, -, *, /, Pow (x, y), exp(x), ln(x), x^2, x^3, x^4, x^5
Genetic operator	Optimal evolution
Chromosomes	30
Head size	7-10
Number of genes	3-9
Linking Function	Addition, Multiplication
Fitness Function error type	R-square, R, RMSE, RRSE, RAE
Mutation rate	0.00138
Inversion rate	0.00546
One-point recombination rate	0.00277
Two-point recombination rate	0.00277
Gene recombination rate	0.00277
Gene transposition rate	0.00277
Numerical constants	±10

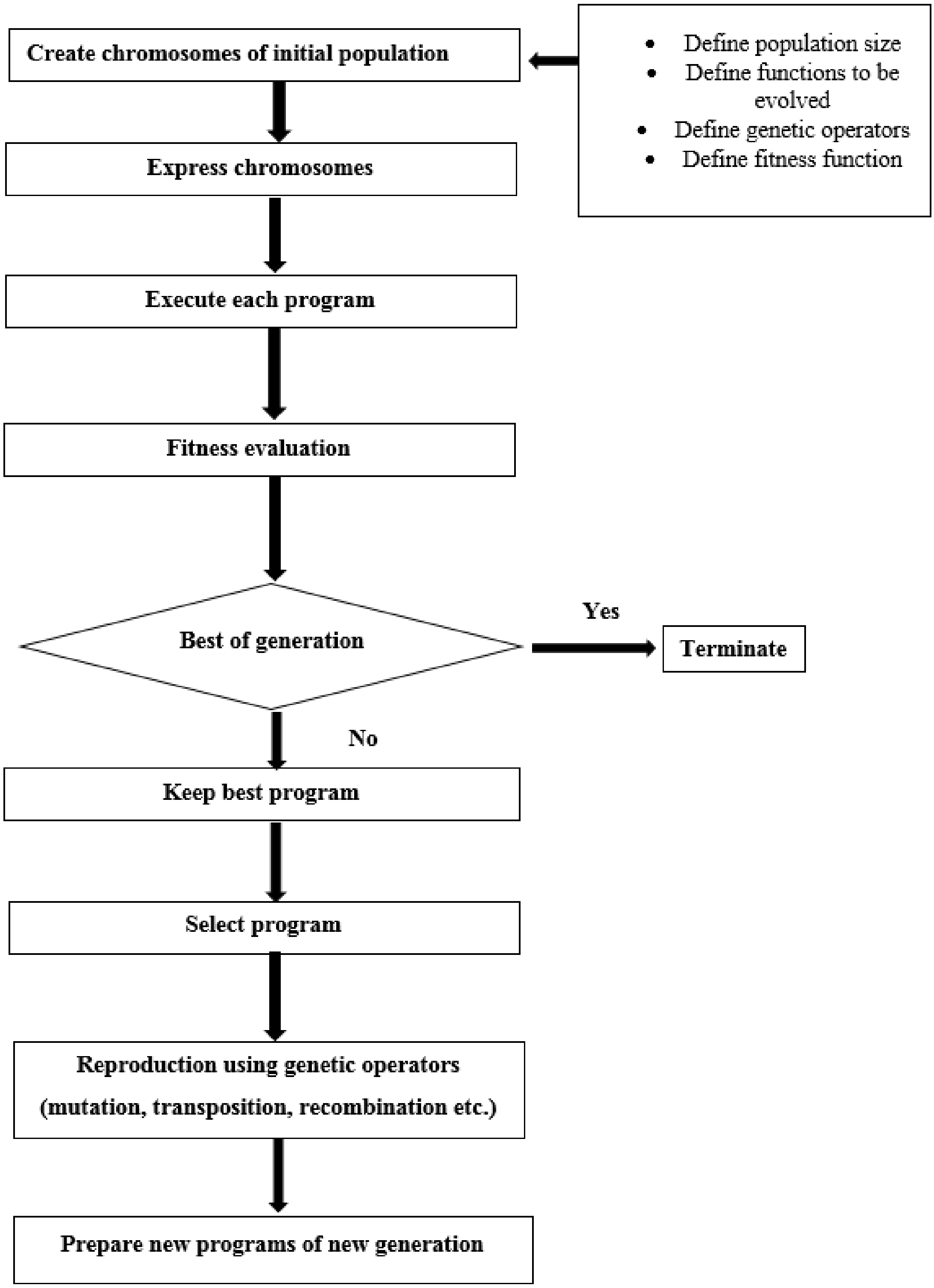


Figure 32. A flow diagram for construction of GEP Models

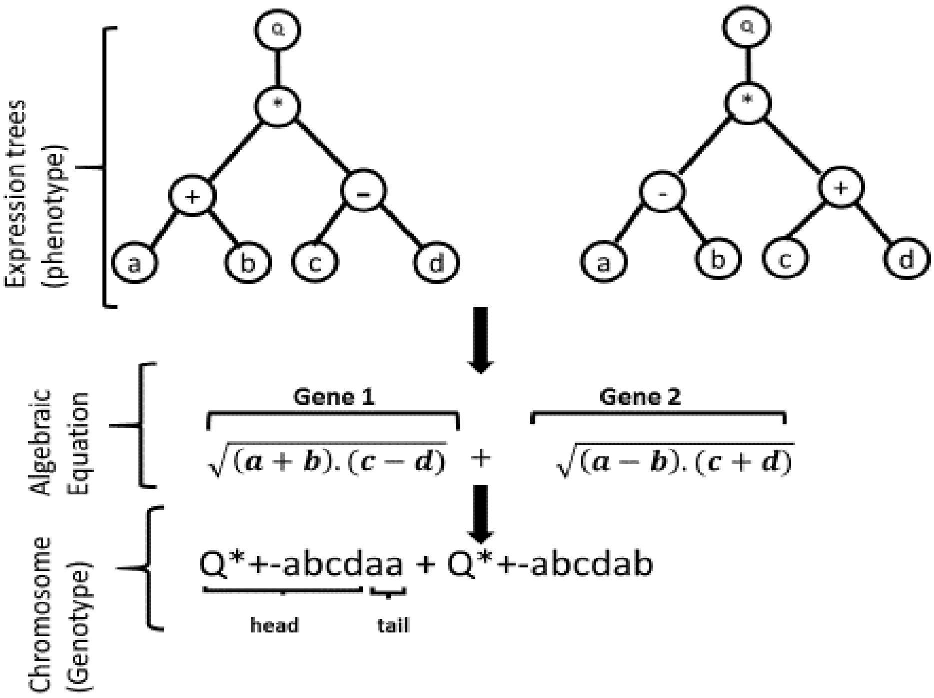


Figure 33. Example of Open Reading Format (ORF) to obtain the output equation

7.3 Results

7.3.1 Selection of Input variables

In the preliminary stage of this study, a detailed research was conducted to explore the relationship between climate indices and seasonal streamflow of NSW which revealed the comparatively more influential indices on streamflow of the study region and served the basis for identifying the different combination sets of multiple indices. Thus, for each of the 12 stations, a number of models were run with different input sets combining two different indices (this study was intended to use two indices in the combination set, though the future study will include more than two indices in the input set) with different lagged months in order to find out the best output models based on their higher Pearson correlation (r) values and lower errors.

7.3.2 Selection of best models

The best GEP model was selected through a trial and error process where initially a number of GEP models were developed with random parameters (head size, number of genes, linking function, genetic operators etc.) which went through continuous modification by changing the parameter values (for current study only head size, number of genes, linking function were modified) in order to obtain the best output model. Most of the stations showed better performances while the head size and number of genes were kept 10 and 9 respectively with some exceptions for some stations where better forecast models were obtained by keeping the head size and gene number 7 and 3 respectively. For all developed models the linking function was used as either addition or multiplication. Once the best models were found, the output equation was obtained from the Expression Trees (ETs) by using Open Reading Format (ORF) which is explained through an example in Figure 33. Equations for one best model from each of the four regions are presented in Table 55. It is noteworthy that the equations get more complicated with the increment of head size and number of genes. The author's endeavor was to keep the models simple for the easy understanding of the end users, thereby the head size and number of genes were kept up to 10 and 9, respectively.

Table 55. Equations of the Best-Developed Models for Four Regions

Region	Best developed models
NNSW	$Q_{\ Coggan} = \frac{d_0 * d_1^2}{1.55(0.88 + 2 * d_0)} + \frac{(4.07 + d_1) * d_0^2}{3.30 + d_1 + 13.029^{d_1}} + \frac{0.065}{d_0^5 - 0.86}$ ***Here, d_0=PDO_{Mar} and d_1=$NINO3.4_{Jun}$***
SNSW	$Q_{\ Gundagai} = \frac{6.44 * \ (d_1^3 - d_0 - 6.64) * (d_0^2 + d_0)}{4.87} + (d_1 - 8.9)^2 + d_1 - d_0 - d_0^3 - \frac{d_1 * (3.62 + \ d_1)}{d_0} + \left\{6.46 + \frac{d_0^3 - 34.57}{\exp(5.53 - d_0)}\right\}^{2} + \frac{0.35 * \ d_1 * \{(d_0 + d_1)\}^{2}}{\exp(d_0^4)} + \frac{d_1}{40.64 + d_1^4} + (d_1^2 - d_1^3)^{3} + \frac{6.11}{d_0^3 + d_0 + d_1 * (d_0 - 8.94)} - 11.22 d_1 * (d_0^2 - d_0 + d_1) - (d_0 * d_1 - 2.3) + \{\exp(5.45) + (8.76 + d_0)\}^{-6.16} + \frac{d_1^5 * (0.42 d_0^4) * \exp(-1.05 - d_1)}{exp(d_0) + 0.54 * d_0}$ ***Here, d_0=PDO_{Mar} and d_1=$NINO3.4_{Jun}$***
CWNSW	$Q_{\ Corowa} = \{(3.38 - d_0)^2 + d_1 + 10.85 + 6.50 d_1\}\frac{d_0^3}{d_0 - d_1} + (68.21 - 2d_0)d_0 + 29.7 d_1 + 44.70 d_0 d_1 + \frac{d_0}{d_0 + 2.38} - 1.85 d_1 {d_0}^4 + (d_0 - d_1)^4 + \frac{13.83 d_1 {d_0}^2}{8.47 * 10^{-2} - d_1} + {d_0}^2(d_0 - 11.34)) + \frac{1.19 d_0^3}{97.61 d_1^2\ (d_1 + 9.59)} + \ln\left(3d_0 - d_{1-}{d_0}^2\right)^{20} - 5.67 \ln\{4.49(d_0 + d_1 + 1.04)^{10}\} + 2d_0^3 - (d_0 - 0.62)^2 + 226.01$ ***Here, d_0=PDO_{Jun} and d_1=$NINO3.4_{Jun}$***
WWNSW	$Q_{\ Brewarrina} = d_0 - (d_0 - d_1)^3 + \exp(d_0^3)^{9.53} + \frac{d_0}{0.44 + \left(\frac{d_1 + 0.44}{d_0}\right)^{10} - \ d_0} + (2d_1 - d_0 - 6.04)^2 - 47.54 d_0^2 + \left(d_0^5 - \frac{0.25 d_1}{1.35 + d_1}\right)\left(\frac{d_1}{d_0} - 16.35\right) + \frac{d_0}{16 d_0^4 + 2.38 d_1 + 2.38 d_1 - 2.6} - 10.62 d_0 * (d_{1-}5.94) - (d_1 - 2.5)(4.53 - d_1) + \frac{d_1}{(d_0 - 0.38) + 2.70 * 10^8 * d_1^{20}} - \frac{1.11}{2d_1 + \ d_0 + 0.75 + \frac{d_1^5}{d_0 d_1}} + \ln d_1^4 + d_1 + \frac{2.43 d_{1-} d_0 - 0.41}{d_0}$ ***Here, d_0=EMI_{JuL} and d_1=$NINO3.4_{JuL}$***

7.3.3 Statistical Performances

Statistical performances of the models were evaluated using different functions which include RMSE, RRSE, RAE, MAE, NSE and Willmott index of agreement (d). The closer the ‘d’ value to 1, the better the model fits the observations. The best models for each station was chosen considering its higher correlation value and lower errors which ensure the best fitness of the developed model. The Pearson correlation (r) values and the statistical errors of the best models for all 12 stations have been documented in Table 56.

Table 56. Performance Test of the Best GEP Models for Calibration and Validation Periods

Region	Station Name	Model	Calibration							Validation						
			r	RRSE	RAE	RMSE	MAE	NSE	d	r	RRSE	RAE	RMSE	MAE	NSE	d
NNSW	Singleton	PDO_{Apr} $NINO3.4_{May}$	0.74	0.67	0.67	13.85	10.62	0.55	0.83	0.79	0.78	0.70	10.60	7.47	0.39	0.63
NNSW	Coggan	PDO_{Mar} $NINO3.4_{Jun}$	0.87	0.24	0.65	1.07	0.82	0.76	0.82	0.91	0.92	0.69	1.68	1.18	0.15	0.93
NNSW	North Cuerindi	PDO_{May} $NINO3.4_{May}$	0.76	0.64	0.72	5.82	4.86	0.58	0.86	0.82	0.90	0.80	9.68	7.99	0.19	0.39
SNSW	Gundagai	PDO_{Mar} $NINO3.4_{Mar}$	0.72	0.69	0.70	53.82	42.80	0.52	0.82	0.93	0.94	0.92	28.69	25.70	0.12	0.83
SNSW	Wee Jasper	PDO_{Jul} $NINO3.4_{Jul}$	0.71	0.70	0.70	5.69	4.47	0.50	0.50	0.87	0.50	0.56	4.03	3.30	0.75	0.92
SNSW	Kiosk	PDO_{Apr} $NINO3.4_{May}$	0.72	0.70	0.70	3.08	2.42	0.52	0.57	0.86	0.90	1.02	3.61	2.96	0.18	0.78
SNSW	Mittagang Crossing	IOD_{Mar} $NINO3.4_{May}$	0.74	0.67	0.62	6.40	4.88	0.45	0.83	0.78	7.58	5.94	10.74	7.27	-10.51	0.56

CWNSW	Corowa	PDO_{Jun} $NINO3.4_{Jun}$	0.70	0.72	0.69	104.02	82.02	0.49	0.79	0.83	2.09	2.17	53.67	43.91	-3.37	0.68
	Wagga Wagga	EMI_{JuL} IPO_{JuL}	0.74	0.64	0.62	59.38	45.20	0.60	0.86	0.72	0.91	0.94	37.47	34.38	0.17	0.75
	Cowra	PDO_{Mar} $NINO3.4_{Feb}$	0.89	0.46	0.44	16.50	12.35	0.78	0.94	0.57	4.88	5.81	17.00	15.36	-22.78	0.48
WNSW	Barham	PDO_{Apr} $NINO3.4_{May}$	0.67	0.75	0.69	84.22	69.24	0.42	0.77	0.84	0.99	1.00	47.87	39.95	0.02	0.51
	Brewarrina	EMI_{Jul} $NINO3.4_{Jul}$	0.84	0.54	0.51	25.92	18.72	0.71	0.91	0.97	0.67	0.80	32.33	28.82	0.56	0.77

7.3.4 Validating the models

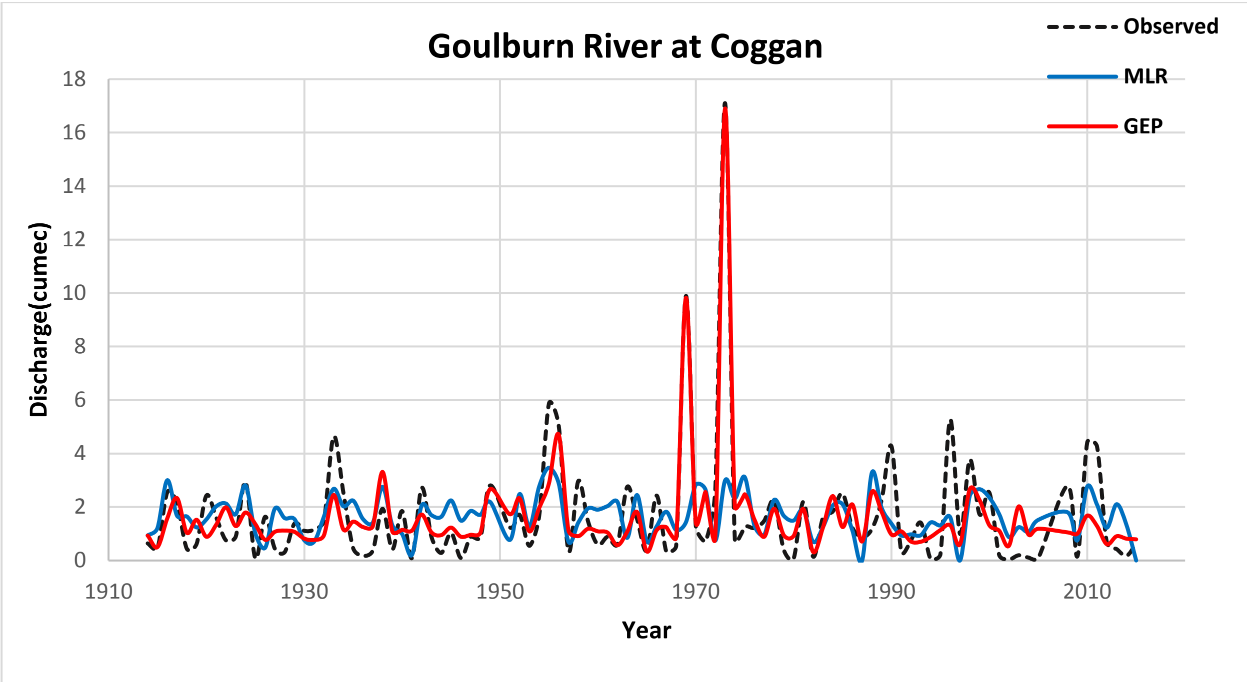
For the validation of the developed models, the whole data set was divided into two segments where the first 96 years of data were used for calibrating the models and rest of 6 years of data were used to assess the validity of the models. Considering the effect of the "millennium drought" period (which was 1994 to 2010 according to Bond et al. 2008) in Australia, longer data range was used for calibration period to prepare the models for any unusual phenomenon like droughts or flood. And the achievements because of such selection criteria have clearly been reflected on the results presented in Table 56 where it is evident that the Pearson correlation (r) values for calibration and validation stages are quite close and the errors are very low in both stages which ensure the acceptability of the developed models.

7.3.5 A brief comparison of GEP models with MLR models

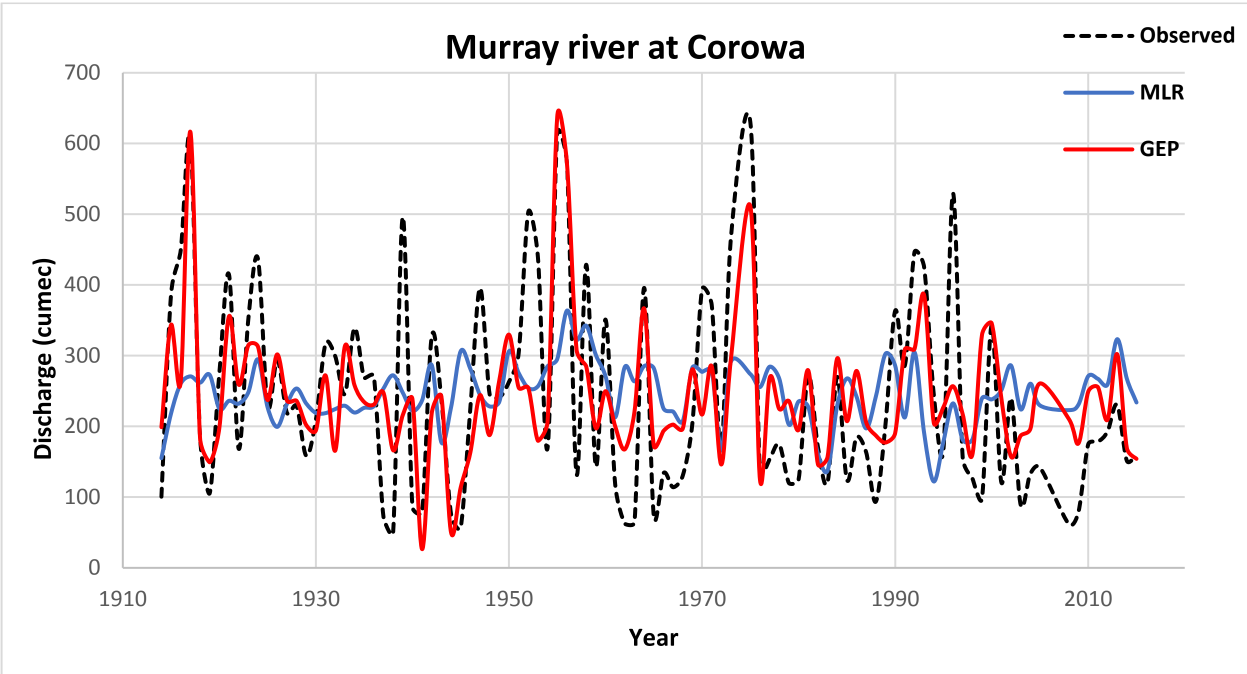
As mentioned earlier, a preliminary study was carried out to explore the linear relationships between climate indices and seasonal streamflow of the same study region where MLR technique was used to develop the linear regression models (In Table 57), the outputs of the current study have been compared with the outcomes of the MLR analyses with a view to exploring the better technique to forecast seasonal streamflow of NSW. For every single station, the non-linear GEP models has shown better performance than the linear MLR models in terms of both Pearson correlation (r) values and statistical errors. The time series plot in Figure 34 explains the much better performance of the developed GEP models compared to the MLR models. The GEP models were able to follow the trend of the actual observed data. Furthermore, while the MLR models failed to capture the high values, GEP models successfully captured almost all the high points. It is noteworthy that MLR models were developed for different calibration (1914-1998) and validation periods (1999-2015). Later on, the GEP models were developed with different data ranges in order to obtain better performances from the developed models.

Table 57. A brief comparison of MLR and GEP Models in Terms of Pearson correlation (r) Values

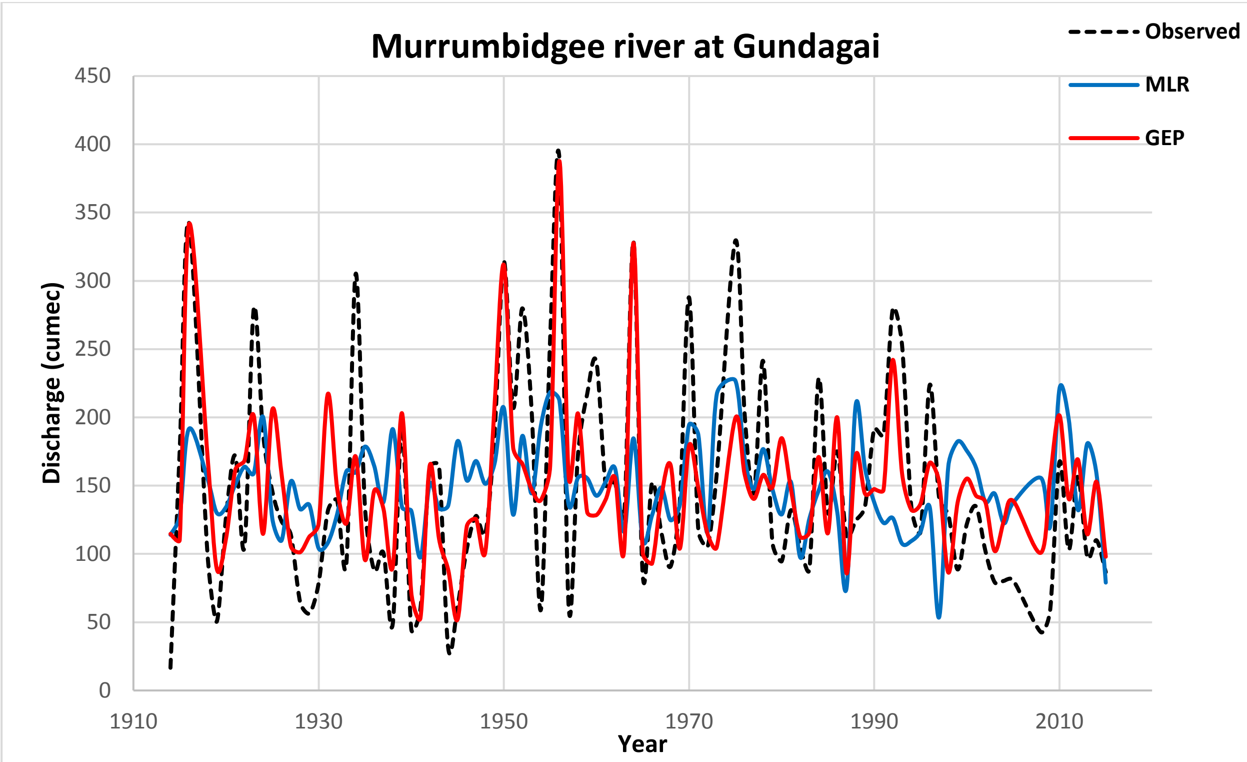
Region	Station Name	GEP			MLR		
		Best Models	Pearson Correlation (r)		Best Models	Pearson Correlation (r)	
			Calibration	Validation		Calibration	Validation
NNSW	Singleton	PDO_{Apr} $NINO3.4_{May}$	0.74	0.79	PDO_{Mar} $NINO3.4_{Jun}$	0.43	0.51
	Coggan	PDO_{Mar} $NINO3.4_{Jun}$	0.87	0.91	PDO_{Jul} $NINO3.4_{Jul}$	0.35	0.60
	North Cuerindi	PDO_{May} $NINO3.4_{May}$	0.76	0.82	PDO_{Jul} $NINO3.4_{Jul}$	0.51	0.56
SNSW	Gundagai	PDO_{Mar} $NINO3.4_{Mar}$	0.72	0.93	IPO_{Jul} $NINO3.4_{Jul}$	0.40	0.43
	Wee Jasper	PDO_{Jul} $NINO3.4_{Jul}$	0.71	0.87	IOD_{Jul} $NINO3.4_{Jul}$	0.44	0.58
	Kiosk	PDO_{Apr} $NINO3.4_{May}$	0.72	0.86	PDO_{Aug} $NINO3.4_{Jul}$	0.44	0.53
	Mittagang Crossing	IOD_{Mar} $NINO3.4_{May}$	0.74	0.78	PDO_{Aug} $NINO3.4_{Jul}$	0.33	0.64
CWNSW	Corowa	PDO_{Jun} $NINO3.4_{Jun}$	0.70	0.83	IPO_{Jun} IOD_{Jun}	0.29	0.87
	Wagga Wagga	IPO_{Jul} EMI_{JuL}	0.74	0.72	IPO_{Jul} $NINO3.4_{JuL}$	0.41	0.20
	Cowra	PDO_{Mar} $NINO3.4_{Feb}$	0.89	0.57	PDO_{Mar} $NINO3.4_{Feb}$	0.25	0.32
WNSW	Barham	PDO_{Apr} $NINO3.4_{May}$	0.67	0.84	IOD_{Jun} $NINO3.4_{Jun}$	0.33	0.78
	Brewarrina	EMI_{Jul} $NINO3.4_{Jul}$	0.84	0.97	IOD_{Jul} $NINO3.4_{Jul}$	0.41	0.59


Goulburn River at Coggan
Observed
MLR
GEP
Discharge(cumec)
18
16
14
12
10
8
6
4
2
0
1910
1930
1950
1970
1990
2010
Year

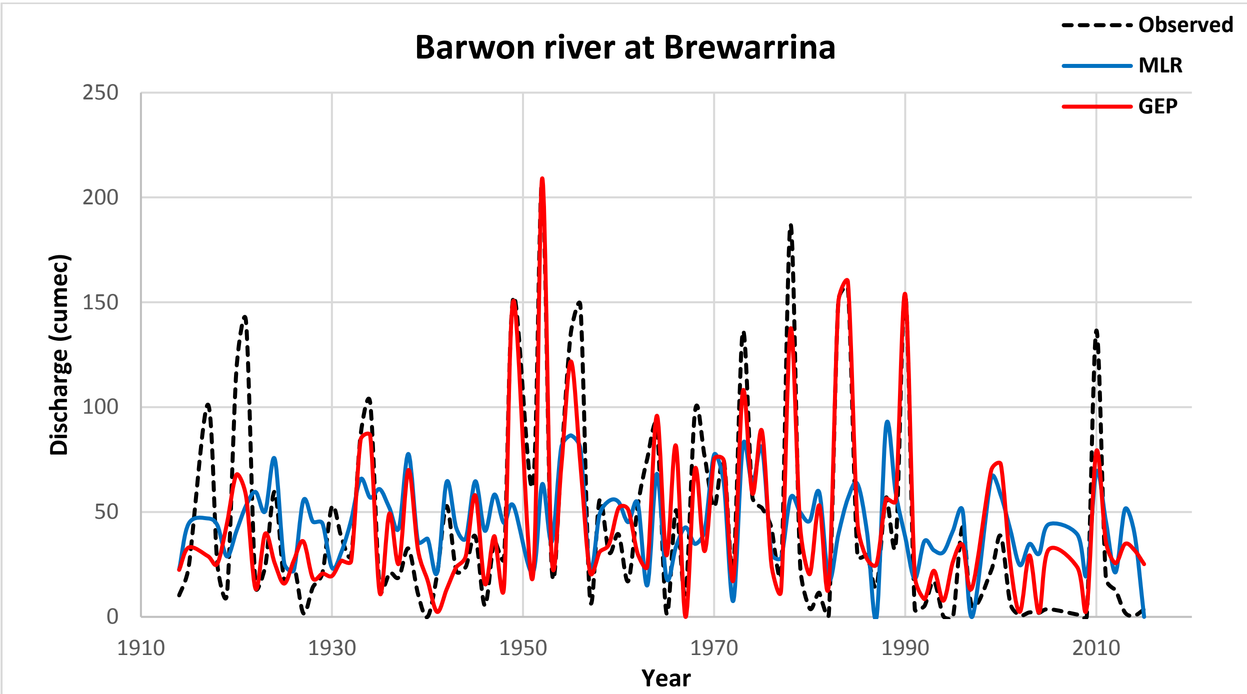
(i)


Murray river at Corowa
Observed
MLR
GEP
Discharge (cumec)
700
600
500
400
300
200
100
0
1910
1930
1950
1970
1990
2010
Year

(ii)

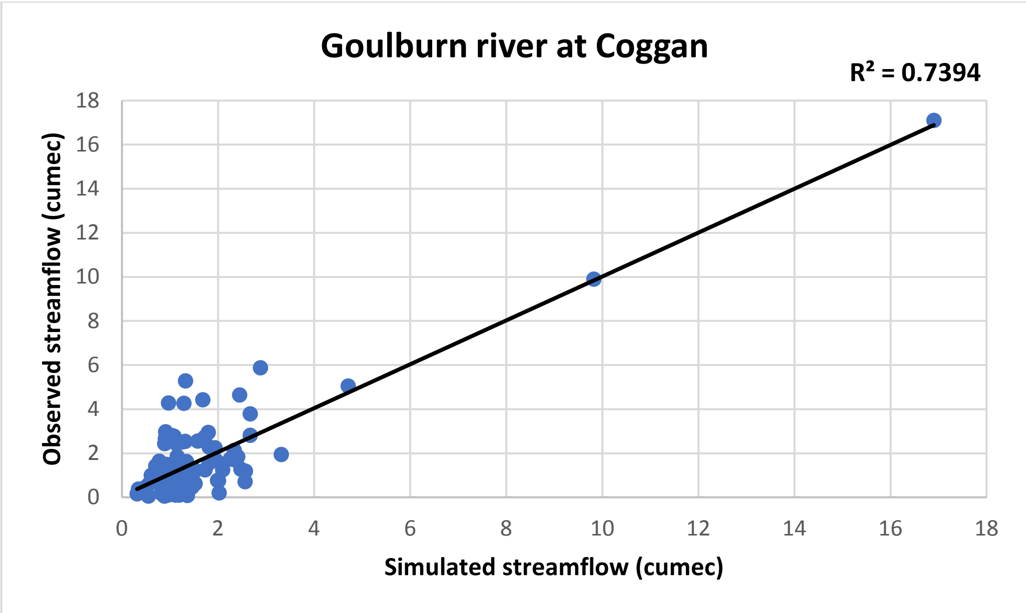

Murrumbidgee river at Gundagai
Observed
MLR
GEP
Discharge (cumec)
450
400
350
300
250
200
150
100
50
0
1910
1930
1950
1970
1990
2010
Year

(iii)

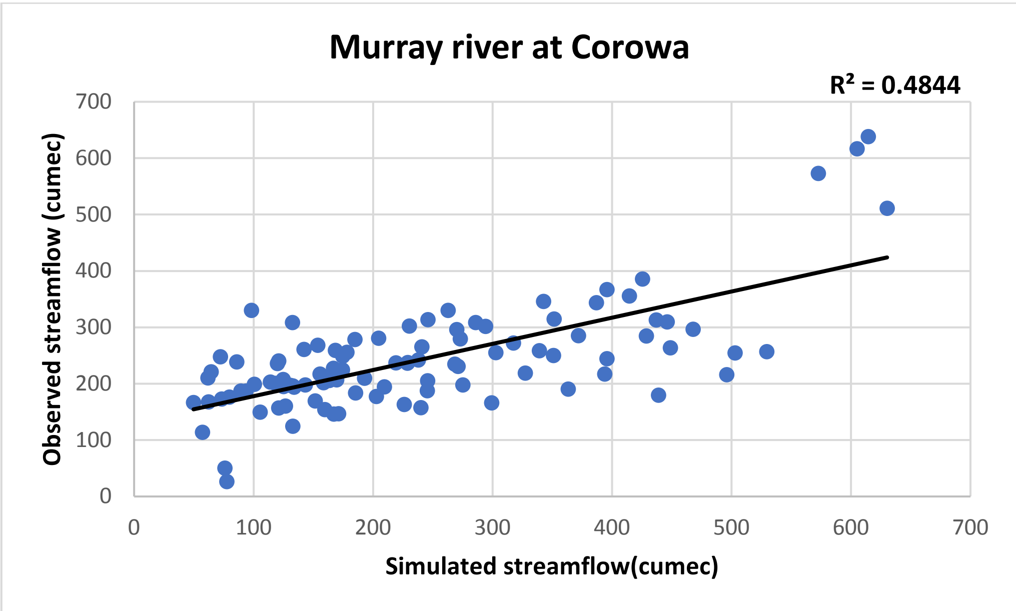

Barwon river at Brewarrina
Observed
MLR
GEP
Discharge (cumec)
250
200
150
100
50
0
1910
1930
1950
1970
1990
2010
Year

(iv)

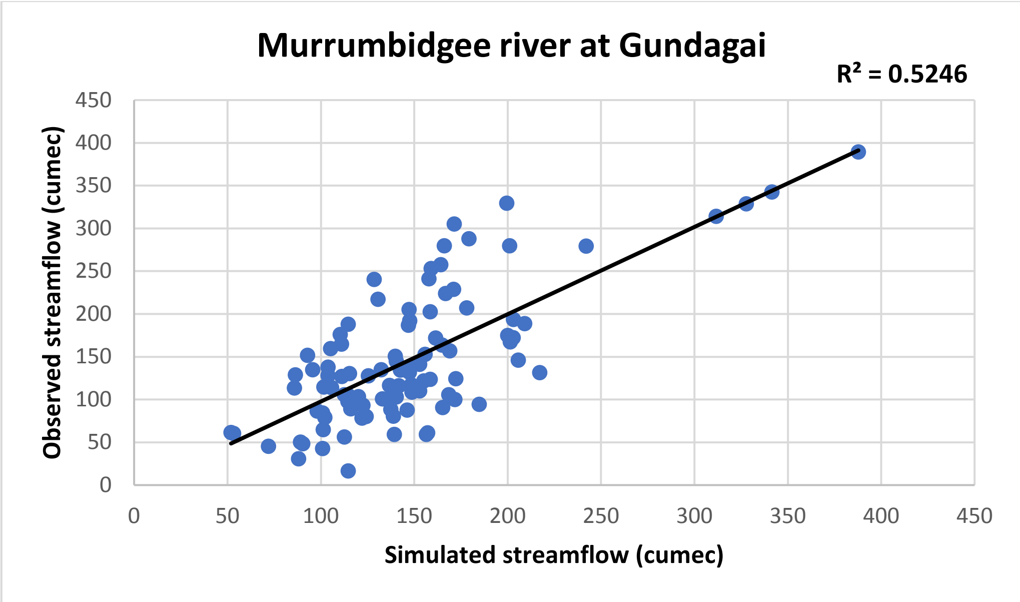
Figure 34. Comparison between observed streamflow and simulated streamflow by MLR and GEP (i) Coggan (NNSW), (ii) Corowa (CWNSW), (iii) Gundagai (SWNSW), (iv) Brewarrina (WNSW) stations



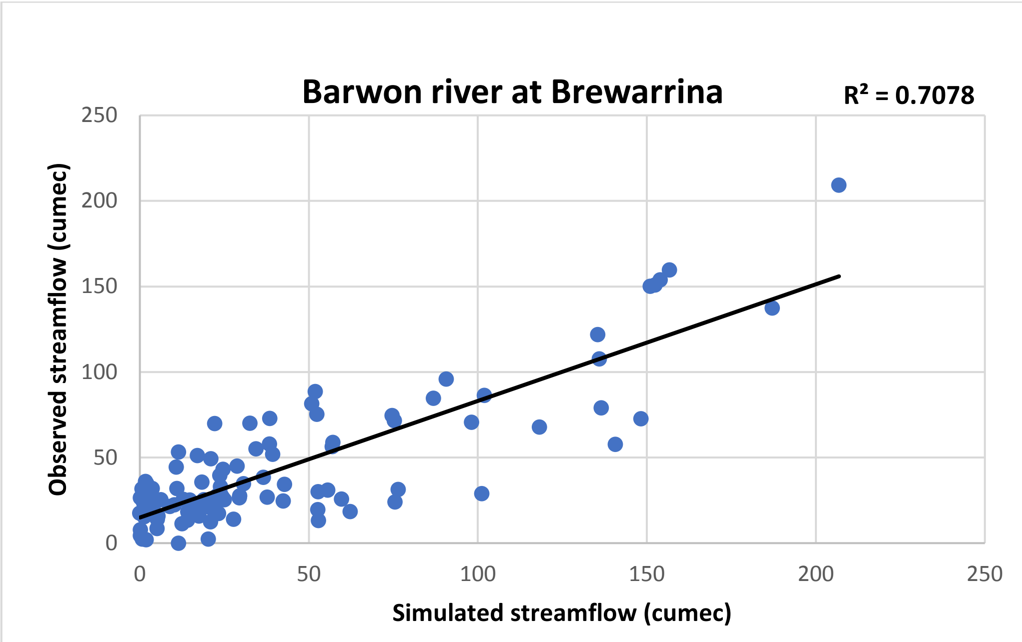
(i)



(ii)



(iii)



(iv)

Figure 35. Scatter plots of observed and simulated streamflow by GEP (i) Coggan (NNSW), (ii) Corowa (CWNSW), (iii) Gundagai (SWNSW), (iv) Brewarrina (WNSW) stations

7.4 Discussion

It is evident from the overall outcomes of the analyses that there is a distinct spatial variation of the most influential indices on seasonal streamflow across NSW. Different combination sets of indices were found to develop the best forecasting model for different regions. Though most of the stations showed the best performances with the input combination of PDO and NINO3.4, however, the best performances were achieved with the selected indices from different months. Significant performances of the developed models were also observed for many other different combination sets containing different indices verifying the influences of other indices on streamflow of NSW.

Analyzing the performances of the best-developed models presented in Table 56, it is observed that all the best models for NNSW were obtained with PDO_NINO3.4 combined models while the highest correlation was found to be 0.87 and 0.91 in calibration and validation stages respectively for the Goulburn river at Coggan station. For the other two stations of NNSW, the results are very satisfactory with significantly high correlation values (0.74~0.82) and lower errors. Again, SSNW was observed to be influenced by the same combination set comprising PDO and NINO3.4 indices with different lagged months except for Murrumbidgee river at Mittagong Crossing where the best-developed model was found with a different combination set comprising IOD and NINO3.4 which has comparatively lower correlation values, i.e.,0.74 and 0.67 in calibration and validation periods respectively. IN SNSW, the highest correlation was achieved for the Murrumbidgee river at Gundagai station, which is 0.72 and 0.93 in calibration and validation stages respectively. The other two stations of this region also showed significantly higher correlations which ranges between 0.71 and 0.87. PDO and NINO3.4 were found to be dominant on the streamflow of CWNSW as well, where two of the three stations provided best models with these indices showing significant correlations that range between 0.57 and 0.89. The other station in CWNSW, Murrumbidgee river at Wagga Wagga was found to be influenced by a different combination set consisting of EMI and IPO indices, though the correlation was relatively lower, i.e., 0.77 and 0.57 for calibration and validation periods respectively.

PDO and ENSO indices were predominant in WNSW where the Barwon River at Brewarrina station provided the highest correlation, i.e.,0.84 and 0.97 in calibration and validation periods respectively for EMI and NINO3.4 combination set. Barham river at Brewarrina station was influenced by PDO_NINO3.4 combined model with correlations 0.67 and 0.84 in calibration and validation stages respectively.

While considering the lagged months of different models, it is observed that the best models are able to predict the spring streamflow up to 5 months ahead. For instance, at Gundagai station in SNSW, the best-developed model consisted of 5 months lagged PDO and NINO3.4 indices which implies that this model can predict futuristic spring (September-November) streamflow at the end of March in the corresponding year. It is noteworthy that despite predicting the streamflow at such advanced stage of the year, the correlation values for this model was significantly high, i.e., 0.72 and 0.93 in calibration and validation periods respectively ensuring the very satisfactory statistical performance of the model.

Nevertheless, few of the best-developed models do not have the ability to predict spring streamflow with longer lead time (Table 56). For these stations, many other models were found which had longer lead time but relatively lower correlation values or higher errors, thus not chosen as the best model. Only one best model from each of the 12 stations, considering the higher correlation values and lower errors have been presented in Table 55. The scatter plots of these models are presented in Figure 35, which shows satisfactory performances of the models over 102 years as the values of the coefficient of determination (R^2) were significant. It has been noticed, in general, the correlation values decrease with the increase of lead time of selected indices.

The statistical performances of the best predictor models ensured the predictability of the developed models with high accuracy as all the values of RRSE, RAE, RMSE, MAE and d shows good agreement in both calibration and validation periods apart from a few exceptions. The index of agreement (d) for both the calibration and validation periods were close to 0.5, which ensured good forecasting ability of the models.

It has been observed during the evaluation process that if the number of generations increases the correlation values also increase. But after reaching an optimum level, the correlation value does not increase anymore. For a few stations, the number of generations became too large to find an optimum solution and apparently, the statistical performances of those models were not good. For instance, the RMSE values of these three stations are relatively higher than those of other stations, for instance, Murray river at Corowa stations has a very high RMSE value 104.02, thus, even though the Pearson correlation is significantly high, the overall performance of the model is not satisfactory. For Mittagong crossing and Cowra stations, the correlation values in validation stages are lower than those of the calibration stages, which is showing different behaviour from the rest of the stations. However, in most of the cases, the correlation values for calibration and validations periods are quite similar.

Based on the Pearson correlation values (r), the comparison of the results of GEP models with MLR models, presented in Table 57 shows that GEP models have better predictability. The highest correlations were 0.89 and 0.97 for GEP models, whereas, for MLR models, the highest correlations were 0.51 and 0.87 in calibration and validation periods, respectively (Table 57). For all the stations, GEP models outperformed MLR models in terms of Pearson correlation (r) values ensuring the better predictability of the GEP models. The greater skills of the combined models have been proved as the correlation values were higher for the combined models with multiple indices than that could be obtained from the single correlation analysis.

7.5 Summary of GEP Analysis

In this chapter, an endeavor has been made to explore the potential skills of GEP for developing streamflow predictor models with longer lead time than usual practice incorporating lagged climate indices. Twelve stations from NSW were chosen considering their agricultural importance and longer data records. Some preceding analyses of this study revealed the most influential indices in the study region, which included PDO(IPO), IOD, EMI and NINO3.4. These indices were exploited as the indicators for streamflow forecasting using GEP. A comparative analysis will be carried out in the following chapter between the outcomes of linear and non-linear techniques with a view to exploring the better predictor modelling technique for streamflow of NSW.

To develop the GEP models, different combinations of two (out of four significant) indices were selected as the input sets. It is observed that not the same combination set gave the best predictor model for all stations/regions which is quite reasonable as NSW is a large region to deal with and distance from a particular location may increase the likeliness of being influenced by other index. Again, the best combinations did not consist of the same lagged months indices. In general, most of the stations were evident to be influenced by PDO and NINO3.4 indices, though the lagged months of the indices were different. IOD was found to be dominant at Corowa station in SNSW with the combination of NINO3.4 whereas EMI strongly influenced Wagga Wagga (in CWNSW) and Brewarrina (in WNSW) stations. The highest correlation was obtained from Brewarrina station, which was 0.84 and 0.97 for calibration and validation stages respectively with one month lagged EMI and NINO3.4 indices. Though considering the longer lagged months, higher correlation values and lower statistical errors, the best performance was provided by Gundagai station in SNSW for PDO_NINO3.4 combined model, despite having significantly high correlation values (0.72 and 0.93 calibration and validation periods respectively), the model was able to predict spring streamflow five months in advance. It is remarkable that all the GEP models outperformed MLR models in terms of Pearson correlation (r) values which was almost twice of the MLR models. Thus, the better predictability of the non-linear GEP models over linear MLR models has been ensured. Nevertheless, for three (Mittagang Crossing, Corowa and Cowra) out of twelve stations, NSE values were negative along with other higher errors implying poor performances of the developed models. Further research will be carried out to explore the weakness of these models, and new models will be developed, incorporating new indices to obtain better models for these three stations. Future work will also include more than three variables (the current study was limited to only two variables) in one model to see the variation of influences on streamflow. The current practice of streamflow forecasting in Australia does not enable the water-stakeholders to take low-risk decisions as they do not get the streamflow forecast at the earlier stage of the crop period. Furthermore, those predictions are stochastic, i.e. the users do not get any in-depth information like the magnitude of the predicted amount of flood. The developed GEP models are able to provide the expected amount of futuristic spring streamflow up to 5 months in advance. This will surely help the water stakeholders to take tactical cropping decisions, thereby, will increase the potential of financial benefits. Further research work may be carried out to explore the influential indices on any particular region and their predictability to forecast seasonal streamflow.

Chapter 8
Model Comparisons

8.1 Introduction

The main objective of this study was to develop a streamflow forecast model using lagged climate indices. To accomplish this objective different modelling approaches had been attempted and compared at the end of this research, with a view to exploring the best modelling technique amongst the used ones in this study for predicting seasonal streamflow.

At the first stage, MLR technique was used as the basic modelling method. It is worth to mention that the whole input data set was divided into two parts in order to calibrate and validate the developed models. The division of the data set was a major challenge for this study since there are no hard and fast rules to divide such modelling data sets into calibration and validation periods. Therefore, in the beginning, random data range was chosen to develop the MLR models which had been changed until satisfactory results could be obtained from the developed models using the corresponding calibration and validation data sets. This trial and error method finally led to the selection of the calibration and validation data set that gave satisfactory results while applied for MLR models. This final data set had 85 years (1914-1998) of data in the calibration period while the rest of 17 years (1999-2015) of data was used for the validation period.

The same data set was used for the next stage of the study where the Multiple Non-Linear Regression (MNLR) method was exploited with a view to getting better streamflow forecast models compared to the MLR models which were developed at the first stage of this study.

In the final stage of this study, for further improvement of the streamflow forecast models, a much more advanced modelling technique, Gene Expression Programming (GEP) was applied. But when the same calibration and validation data set was used for GEP modelling, the models failed to produce satisfactory results. Thus, new data sets were tried to get the desired satisfactory outcomes from the GEP models. Finally, the combination of 96 years data in the calibration stage and the rest of 6 years data invalidation stage, gave the best results for the GEP models.

Since, one of the main aims of this research was to compare the different forecast modelling methods and identify the best one amongst the used methods, for a fair comparison, MLR and MNLR analysis was redone with the same data set that was applied for GEP modelling. Eventually, this data set also was able to produce satisfactory results for MLR and MNLR models. The reason for the best performance of this combination of data might be the inclusion of the unusual drought periods (2000-2009) in the calibration dataset which ultimately prepared the models for any unusual phenomenon, such as droughts, floods etc., and thus the model could perform equally good in the validation stage as well.

The results obtained at different stages of this study will be compared in this chapter. The first comparison will be between the results obtained from applying two different data sets for MLR modelling. This will be followed by the comparison of MNLR models using these two different data sets. Finally, GEP models developed using the second data set will be compared with the linear (MLR) and non-linear (MNLR) techniques.

At the end of this chapter, to verify the results derived from this research study, a comparative analysis will be carried out between the current study and the previous studies which were done on a similar topic.

8.2 Comparison of MLR Models for Different Datasets

As mentioned earlier, MLR analysis was done with two different sets of data i.e., the original data set was divided into calibration and validation groups following two different time ranges. At first, the whole data set was segmented into 85 years for calibration and 17 years for validation. Later on, the division was first 96 years for calibration and last 6 years for validation. For convenience, in this study, the first set and second set will be considered as “Dataset-1” and “Dataset-2” respectively.

The comparison between the MLR models developed with “Dataset-1” and “Dataset-2” is showed in Table 58. In general, most of the stations showed similar correlation values for both datasets. Nevertheless, there was significant increment in correlation values (in validation period) for “Dataset-2” in case of Maittagang Crossing, Corowa and Barham station whereas some stations also produced reduced correlation values with “Dataset-2”, for instance, Wagga Wagga station in validation period.

For both datasets, the greatest number of best models were found with PDO and NINO3.4 combined indices. However, strong influence of IPO and IOD combined models were also observed six out of twelve station had best models consisted of IPO or IOD along with NINO3.4.

Hence the strong influence of all these four indices on spring streamflow of NSW were clearly evident in this analysis.

Table 58. Comparison of MLR models develpoed with the different Datasets

Region	Station Name	Best Models	MLR models for Dataset-1		MLR models for Dataset-2	
			Pearson Correlation (r)		Pearson Correlation (r)	
			Calibration	Validation	Calibration	Validation
NNSW	Singleton	PDO_{Mar} $NINO3.4_{Jun}$	0.41	0.65	0.43	0.51
	Coggan	PDO_{Jul} $NINO3.4_{Jul}$	0.33	0.61	0.35	0.60
	North Cuerindi	PDO_{Jul} $NINO3.4_{Jul}$	0.46	0.62	0.51	0.56
SNSW	Gundagai	IPO_{Jul} $NINO3.4_{Jul}$	0.43	0.51	0.40	0.43
	Wee Jasper	IOD_{Jul} $NINO3.4_{Jul}$	0.45	0.57	0.44	0.58
	Kiosk	PDO_{Aug} $NINO3.4_{Jul}$	0.45	0.41	0.44	0.53
	Mittagang Crossing	PDO_{Aug} $NINO3.4_{Jul}$	0.35	0.49	0.33	0.64
CWNSW	Corowa	IPO_{Jun} IOD_{Jun}	0.30	0.48	0.29	0.87
	Wagga Wagga	IPO_{Jul} $NINO3.4_{JuL}$	0.43	0.55	0.41	0.20
	Cowra	PDO_{Mar} $NINO3.4_{Feb}$	0.35	0.44	0.25	0.32
WNSW	Barham	IOD_{Jun} $NINO3.4_{Jun}$	0.31	0.44	0.33	0.78
	Brewarrina	IOD_{Jul} $NINO3.4_{Jul}$	0.40	0.56	0.41	0.59

8.3 Comparison of MNLR Models for Different Datasets

NNSW

MNLR analysis was also done for two datasets to explore which dataset gives better results. Since most of the stations gave higher correlations with "Dataset-2", this dataset seemed promising to give a better result. Therefore, the trial with both datasets was applied only for the three stations of NNSW. For the rest of the stations, MNLR models were developed with "Dataset-2" only. Thus, the comparative analysis between developed MNLR models was carried out only for NNSW, which is shown in Table 59.

It is evident from Table 59 that all three stations are showing higher correlation values with "Dataset -2", except Singleton station which had a higher correlation with "Dataset-1" in the validation period.

It is noteworthy that, even though only the best models are presented in Table 59, all other combinations were also explored to identify which of the two datasets produce higher correlations. For most of the combinations, the higher correlations were obtained using "Dataset-2" apart from Singleton station where correlation value in the validation stage is higher with dataset-1 than dataset-2. For both datasets, the best MNLR models were developed with PDO and NINO3.4 combination, implying the strong influence of these two indices on spring streamflow of NNSW region.

Table 59. Comparison of MNLR models develpoed with the different Datasets for NNSW

Station Name	Model	**MNLR models for Dataset-1**		**MNLR models for Dataset-2**	
		Calibration	Validation	Calibration	Validation
Singleton	**PDO_{Jun}-$NINO3.4_{Jun}$**	0.46	0.71	0.47	0.57
Coggan	**PDO_{Jul}-$NINO3.4_{Jul}$**	0.38	0.76	0.40	0.82
North Cuerindi	**PDO_{Jul}-$NINO3.4_{Jul}$**	0.52	0.72	0.58	0.77

8.4 Comparison of MLR and MNLR models for Dataset-1

In the Table 60, a comparison between MLR and MNLR models developed using "Dataset-1" has been presented. As abovementioned, MNLR analysis using "Dataset-1" was developed for the NNSW region only. Hence, the comparison shown in the Table 60 is only for the three stations of NNSW region. It is clearly evident that for all three stations, MNLR models outperformed MLR models in terms of Pearson correlation (r) values. The highest correlation obtained from the MLR model was 0.65, whereas it was 0.76 for the MNLR model. It is noteworthy that both methods developed best models (based on correlation values) for the same combination of indices for each of the three stations of NNSW region. For both methods, the best models have been developed with PDO and NINO3.4 combination, which indicates the strong impact of these two indices on spring streamflow of NNSW region.

Table 60. Comparison of MLR and MNLR models developed with the Dataset-1

Station Name	Model	Calibration Period		Validation Period	
		MLR	MNLR	MLR	MNLR
		r	r	r	r
Singleton	PDO_{MARCH}_NINO3.4_{JUNE}	0.41	0.45	0.65	0.70
Coggan	PDO_{JUly}_NINO3.4_{JUly}	0.33	0.38	0.61	0.76
North Cuerindi	PDO_{JUly}_NINO3.4_{JUly}	0.46	0.52	0.62	0.72

8.5 Comparison of MLR and MNLR models for Dataset-2

"Dataset-2" was used for developing MLR and MNLR based forecast models for all 12 stations. The comparison of the performances of these models based on their correlation values has been presented in the Table 61. The result shows that in the calibration stage, the MNLR models outperformed MLR models almost all stations apart from Kiosk station where the correlation of MLR (0.44) model was slightly higher than MNLR model (0.38). In the validation stage, both methods showed similar performances based on their correlation values. Out of twelve stations, six stations showed higher correlations for MLR models while the other six stations showed higher correlations for MNLR models. However, for each of the twelve stations, correlation values were quite similar.

For both techniques, the greatest number of best models were found with PDO and NINO3.4 combined indices which indicates the strongest influence of these indices on spring streamflow of NSW. However, the strong influence of IPO and IOD along with EMI was also observed as some best models were also developed with either of these indices along with NINO3.4. Hence the strong influence of all these five indices on spring streamflow of NSW was clearly evident in this analysis.

Table 61. Comparison of MLR and MNLR models developed with the Dataset-2

Region	Station Name	MLR models for Dataset-2			MNLR models for Dataset-2		
		Best Models	Pearson Correlation (r)		Best Models	Pearson Correlation (r)	
			Calibration	Validation		Calibration	Validation
NNSW	Singleton	PDO_{Mar} $NINO3.4_{Jun}$	0.43	0.51	PDO_{June} $NINO3.4_{June}$	0.47	0.56
	Coggan	PDO_{Jul} $NINO3.4_{Jul}$	0.35	0.60	IOD_{June} $NINO3.4_{June}$	0.38	0.37
	North Cuerindi	PDO_{Jul} $NINO3.4_{Jul}$	0.51	0.56	PDO_{Apr} $NINO3.4_{June}$	0.53	0.62
SNSW	Gundagai	IPO_{Jul} $NINO3.4_{Jul}$	0.40	0.43	PDO_{Mar} $NINO3.4_{Mar}$	**0.47**	**0.54**
	Wee Jasper	IOD_{Jul} $NINO3.4_{Jul}$	0.44	0.58	IPO $NINO3.4_{Jun}$	**0.45**	**0.48**
	Kiosk	PDO_{Aug} $NINO3.4_{Jul}$	0.44	0.53	IOD_{Jun} $NINO3.4_{Jun}$	**0.38**	**0.56**
	Mittagang Crossing	PDO_{Aug} $NINO3.4_{Jul}$	0.33	0.64	PDO_{Mar} $NINO3.4_{May}$	**0.48**	**0.51**
CWNSW	Corowa	IPO_{Jun} IOD_{Jun}	0.29	0.87	PDO_{Dec}- $NINO3.4_{May}$	0.43	0.39
	Wagga Wagga	IPO_{Jul} $NINO3.4_{JuL}$	0.41	0.20	PDO_{Mar}- $NINO3.4_{Mar}$	0.43	0.35
	Cowra	PDO_{Mar} $NINO3.4_{Feb}$	0.25	0.32	PDO_{April}- $NINO3.4_{Apr}$	0.44	0.11
WNSW	Barham	IOD_{Jun} $NINO3.4_{Jun}$	0.33	0.78	PDO_{Dec}- $NINO3.4_{May}$	0.40	0.51
	Brewarrina	IOD_{Jul} $NINO3.4_{Jul}$	0.41	0.59	EMI_{June}_ $NINO3.4_{June}$	0.51	0.61

8.6 Comparison of MLR, MNLR and GEP models for Same Dataset

As mentioned earlier, a preliminary study was carried out to explore the linear relationships between climate indices and seasonal streamflow of NSW region where MLR technique was used to develop the linear regression models which were followed by the application of Multiple Non-Linear Regression (MNLR) technique to develop the forecast models with a view to comparing the potential of both linear and non-linear techniques. In the final stage of this study, an advanced technique, Gene Expression Programming (GEP) was exploited to develop the forecast models with a view to further improvement of the performances of output models.

Here, performances of all three methods, i.e., MLR, MNLR and GEP, have been compared and analyzed to explore the best streamflow forecasting technique among the used three techniques for NSW region. The Pearson correlation value (r) based comparative analysis has been presented in the Table 62. It is noteworthy that in this comparative analysis, “Dataset-2” was used to develop all the models for all three methods.

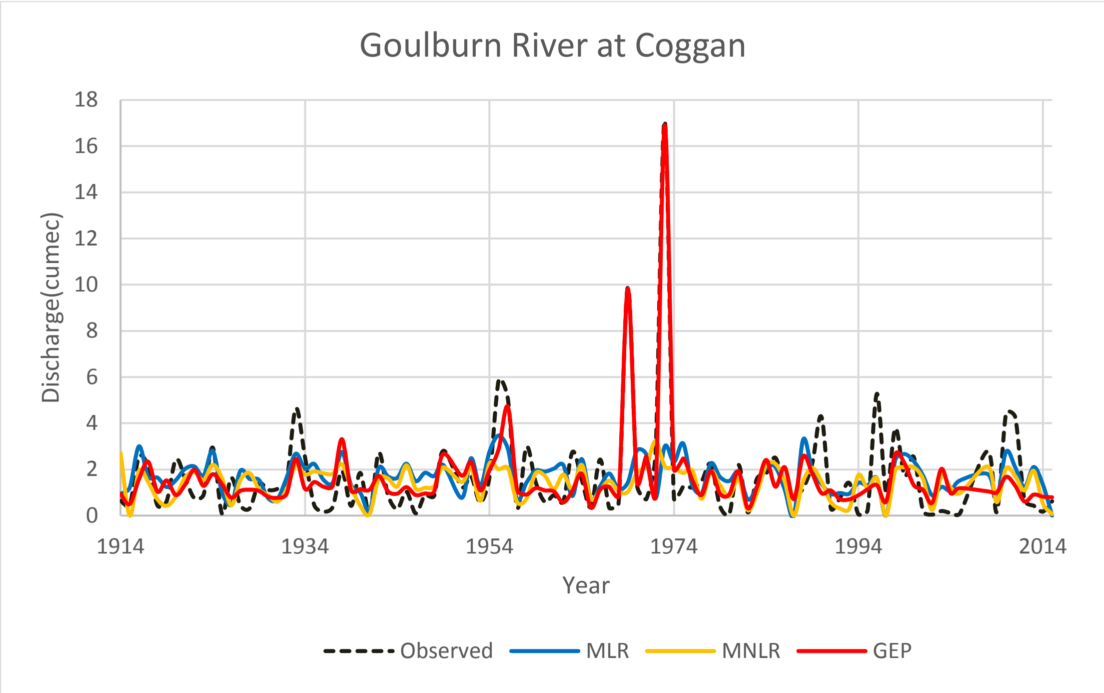
For every single station, the non-linear GEP models have shown better performance than the linear MLR and non-linear MNLR models in terms of both Pearson correlation (r) values and statistical errors. The time series plot in Figure 36 explains the much better performance of the developed GEP models compared to the MLR and MNLR models.

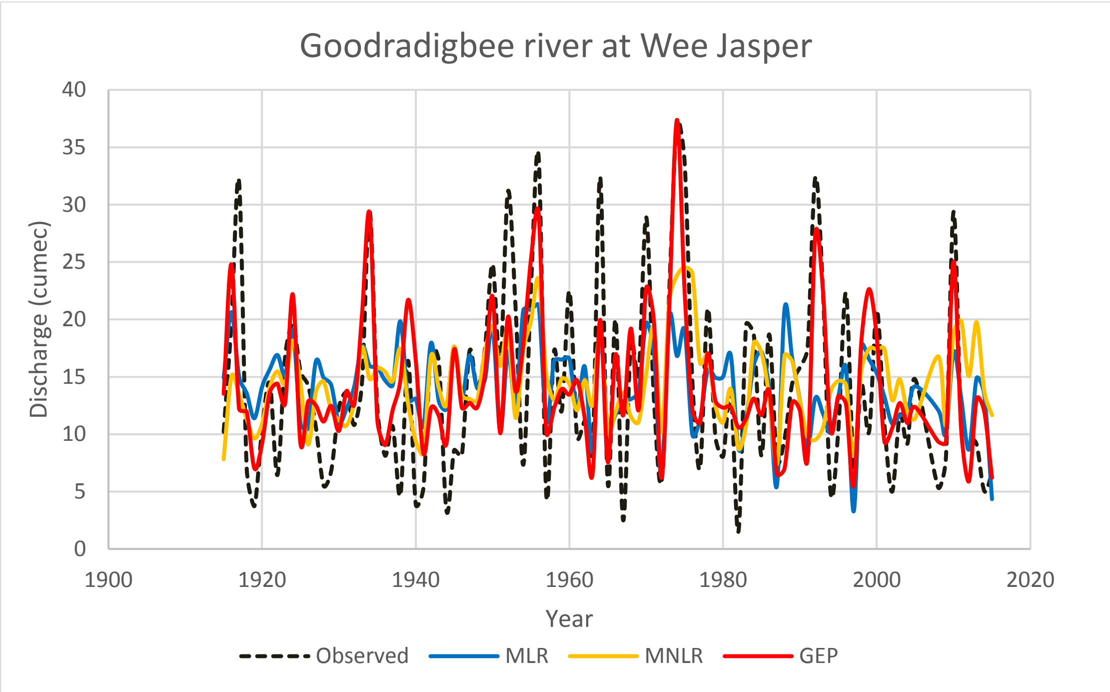
The GEP models were able to follow the trend of the actual observed data. Furthermore, while the MLR and MNLR models failed to capture the high values, GEP models successfully captured almost all the high points.

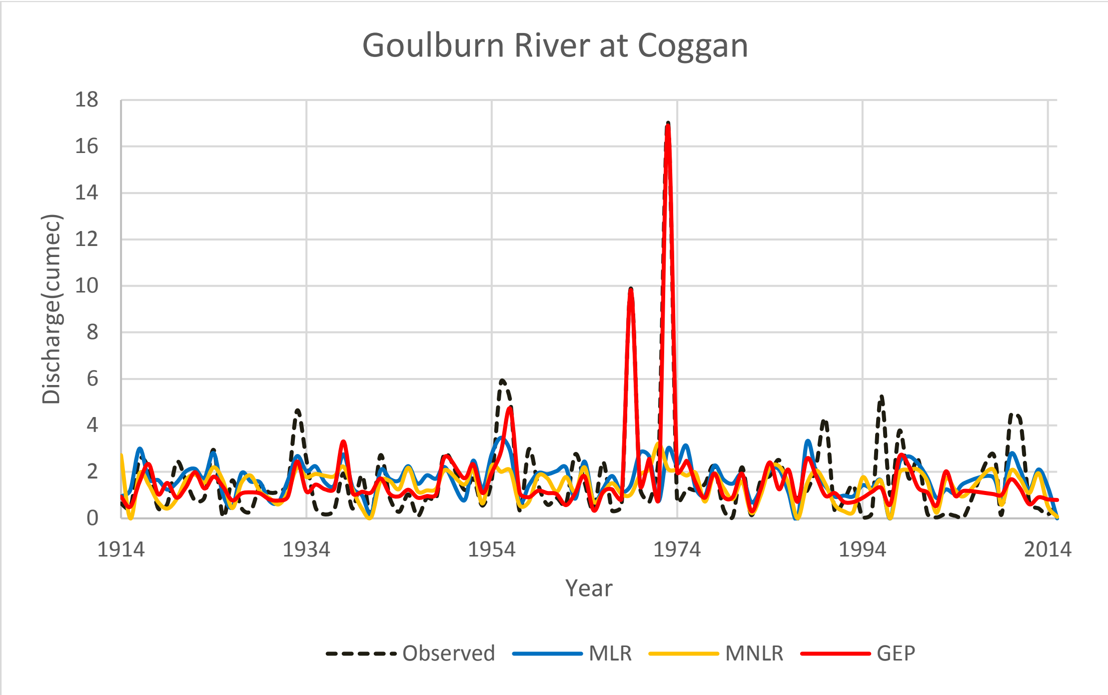
Based on the Pearson correlation values (r), the comparison of the results of GEP models with MLR and MNLR models, presented in Table 62 shows that GEP models have the best predictability. The highest correlations were 0.89 and 0.97 for GEP models, whereas for MLR models the highest correlations were 0.51 and 0.87 and for MNLR models the values were 0.53 and 0.62 in calibration and validation periods respectively. For all the stations, GEP models outperformed MLR and MNLR models in terms of Pearson correlation (r) values ensuring the better predictability of the GEP models.

Table 62. Performance comparison of the developed MLR, MNLR and GEP models

Region	Station Name	GEP			MLR			MNLR		
		Best Models	Pearson Correlation (r)		Best Models	Pearson Correlation (r)		Best Models	Pearson Correlation (r)	
			Calibration	Validation		Calibration	Validation		Calibration	Validation
NNSW	Singleton	PDO_{Apr} $NINO3.4_{May}$	0.74	0.79	PDO_{Mar} $NINO3.4_{Jun}$	0.43	0.51	PDO_{Jun} $NINO3.4_{Jun}$	0.47	0.56
	Coggan	PDO_{Mar} $NINO3.4_{Jun}$	0.87	0.91	PDO_{Jul} $NINO3.4_{Jul}$	0.35	0.60	IOD_{Jun} $NINO3.4_{Jun}$	0.38	0.37
	North Cuerindi	PDO_{May} $NINO3.4_{May}$	0.76	0.82	PDO_{Jul} $NINO3.4_{Jul}$	0.51	0.56	PDO_{Apr} $NINO3.4_{jun}$	0.53	0.62
SNSW	Gundagai	PDO_{Mar} $NINO3.4_{Mar}$	0.72	0.93	IPO_{Jul} $NINO3.4_{Jul}$	0.40	0.43	PDO_{Mar} $NINO3.4_{Mar}$	0.47	0.54
	Wee Jasper	PDO_{Jul} $NINO3.4_{Jul}$	0.71	0.87	IOD_{Jul} $NINO3.4_{Jul}$	0.44	0.58	IPO $NINO3.4_{Jun}$	0.45	0.48
	Kiosk	PDO_{Apr} $NINO3.4_{May}$	0.72	0.86	PDO_{Aug} $NINO3.4_{Jul}$	0.44	0.53	IOD_{Jun} $NINO3.4_{Jun}$	0.38	0.56
	Mittagang Crossing	IOD_{Mar} $NINO3.4_{May}$	0.74	0.78	PDO_{Aug} $NINO3.4_{Jul}$	0.33	0.64	PDO_{Mar} $NINO3.4_{May}$	0.48	0.51
CWNSW	Corowa	PDO_{Jun} $NINO3.4_{Jun}$	0.70	0.83	IPO_{Jun} IOD_{Jun}	0.29	0.87	PDO_{Dec} $NINO3.4_{May}$	0.43	0.39
	Wagga Wagga	IPO_{Jul} EMI_{JuL}	0.74	0.72	IPO_{Jul} $NINO3.4_{JuL}$	0.41	0.20	PDO_{Mar} $NINO3.4_{Mar}$	0.43	0.35
	Cowra	PDO_{Mar} $NINO3.4_{Feb}$	0.89	0.57	PDO_{Mar} $NINO3.4_{Feb}$	0.25	0.32	PDO_{Apr} $NINO3.4_{Apr}$	0.44	0.11
WNSW	Barham	PDO_{Apr} $NINO3.4_{May}$	0.67	0.84	IOD_{Jun} $NINO3.4_{Jun}$	0.33	0.78	PDO_{Dec} $NINO3.4_{May}$	0.40	0.51
	Brewarrina	EMI_{Jul} $NINO3.4_{Jul}$	0.84	0.97	IOD_{Jul} $NINO3.4_{Jul}$	0.41	0.59	EMI_{Jun} $NINO3.4_{Jun}$	0.51	0.61


Goulburn River at Coggan
Discharge(cumec)
18
16
14
12
10
8
6
4
2
0
1914
1934
1954
1974
1994
2014
Year
Observed
MLR
MNLR
GEP


Goodradigbee river at Wee Jasper
Discharge (cumec)
40
35
30
25
20
15
10
5
0
1900
1920
1940
1960
1980
2000
2020
Year
Observed
MLR
MNLR
GEP



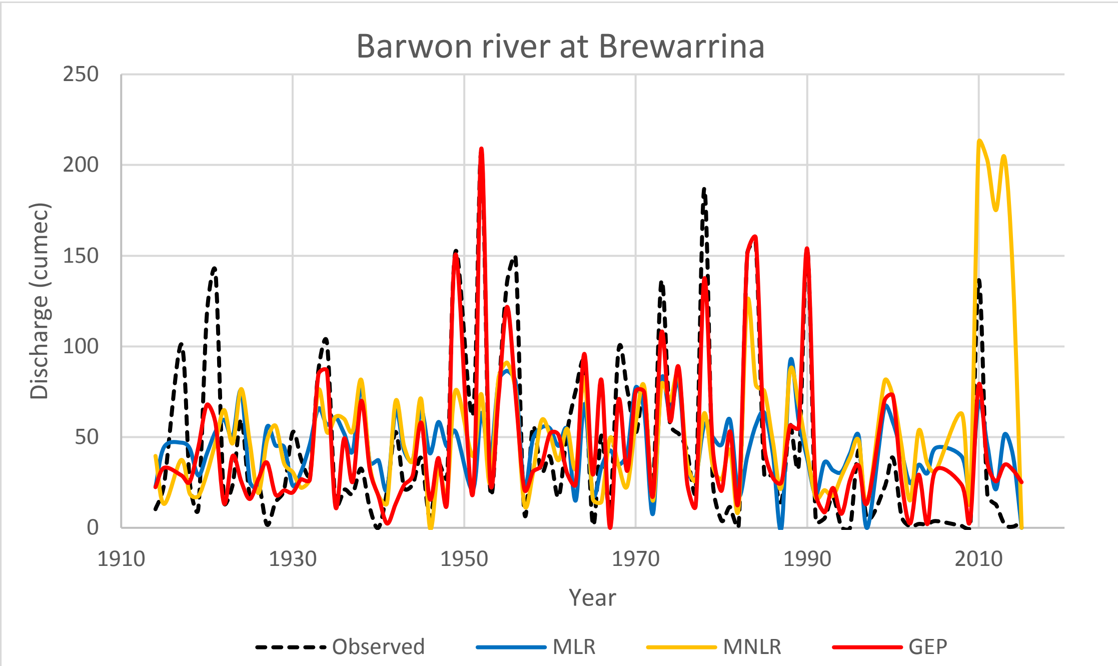


Figure 36. Comparison of the performances of developed MLR, MNLR and GEP models through time series plot

8.7 Comparison of Current Study with Previous Studies

In Table 63, the performances of developed MLR, MNLR and GEP models were compared with the works of many researchers who attempted similar analyses to forecast seasonal streamflow using climate indices in different regions of south-east Australia. The developed combined models in the current study outperformed the previously developed models in terms of Pearson correlation (r) values.

The comparison between the outcomes obtained from this study (MLR, MNLR and GEP models) and past research studies based upon highest correlation values presented in Table 63 depicts that the current study (all MLR, MNLR and GEP models) outperformed any of the past research works explaining the significant combined impact of multiple indices on spring streamflow of the study region. Pearson correlation (r) values are much higher for the multiple indices' models than that of the single-index model. For instance, at Singleton station during single correlation analyses, NINO3.4June and PDOMarch showed correlations −0.43 and −0.29, respectively (Esha and Imteaz, 2018), whereas, for MLR, MNLR and GEP analyses the correlation values for PDO_NINO3.4 combined models increased to 0.56, 0.62 and 0.93 respectively(Table 63). The variation of influences of different climate indices on different study regions of NSW is comparable with the recent study outcomes of Duc et al. (2017). They have reported association of climate indices with NSW rainfall using Bayesian model averaging. Among their studied sites, outcomes of the sites which are within 160 km of the current study's selected streamflow stations are similar to the findings of current study. They have reported that a single IPO cannot impact NSW rainfall significantly; however, its association with ENSO is significantly influential on the rainfall of almost the whole of NSW. The current study evidenced the strong influence of PDO–NINO3.4 on spring streamflow almost across the whole state (it is to be noted that IPO and PDO are similar as IPO acts on the whole Pacific basin and PDO is active in the North Pacific, poleward of 20°N). This finding is strongly supported by the findings of many past studies that suggested IPO or PDO phases modulate the frequency and magnitude of ENSO events (Power et al. 1999; Folland et al. 2002; Franks 2004; Verdon et al. 2004) which is influential on the streamflow volumes of many parts of the world (Kahya & Dracup 1993; Moss et al. 1994; Piechota & Dracup 1996; Piechota et al. 1998; Chiew et al. 1998; Dettinger & Diaz 2000; Kiem & Franks 2001; Wooldridge et al. 2001).

Table 63. Comparison of the present study with the previous studies based on the highest correlations between indices and spring streamflow for South-East Australia

INDICES	Kirono et al. (2010)	Chiew et al. (2003)	CURRENT STUDY			
			Single lagged correlation	MLR correlation	MNLR correlation	GEP correlation
Nino3.4	-	-	-0.43[iv]	0.56[viii]	0.62[ix]	0.93[ix]
PDO	-	-	-0.41[v]			
Nino3	0.35[i]	-	0.36[vi]			
SOI	0.36[ii]	0.51[iii]	0.51[vii]			

i) 8 months lagged Nino3 ii) 12 months lagged SOI iii) Winter SOI
iv) 3 months lagged Nino3.4 v) 2 months lagged PDO vi) 3 months lagged Nino3.4
vii) 2 months lagged SOI viii) PDO_{July} & $Nino3.4_{July}$ ix) PDO_{April} & $Nino3.4_{June}$
x) PDO_{March} & $Nino3.4_{March}$

8.8 Conclusion

In this chapter, a comparative analysis of the performances of developed MLR, MNLR and GEP models has been carried out. The throughout analysis revealed that GEP is the best performing model while predicting spring streamflow of NSW region. However, the non-linear technique (MNLR) was a better performer than the linear technique (MLR) while comparing the predictability of these two techniques. As mentioned earlier, two different datasets were used in this analysis between which “Dataset-2” produced better results while used as input dataset for the streamflow forecast models. Among the five used climate indices, PDO and NINO3.4 combined models provided the maximum number of best models implying the strongest impact of these two indices on spring streamflow on NSW region. All techniques (MLR, MNLR and GEP) applied in this study outperformed the results of any of the previous studies carried out for forecasting streamflow in this region.

Chapter 9
Conclusion

9.1 Summary of the research study

The main objective of this study was to develop a seasonal streamflow forecast model for NSW incorporating multiple large-scale climate indices as predictors with the application of different linear, non-linear and Artificial Intelligence (AI) based methods. NSW was chosen as the study area considering the agricultural importance of this region and its major contribution to Australia's economy. A reliable forecast model will enable the water stakeholders of this region to take low-risk decisions at the earlier stage of the crop period and thus will enhance agricultural production and mitigate the losses due to unusual extreme climatic phenomena like droughts and high floods.

Three different modelling techniques were applied to accomplish the aim of this study which included Linear MLR (Multiple Linear Regression), Non-linear MNLR (Multiple Non-Linear Regression) and Artificial Intelligence (AI)-based GEP (Gene Expression Programing) methods. Though MLR and MNLR are commonly used statistical techniques for forecasting different hydrological parameters such as rainfall, streamflow, etc., application of GEP is quite rare in this field. The novelty of this study is that this is the first time long-term seasonal streamflow prediction for Australian rivers is attempted using GEP. Among the modelling techniques, Artificial Intelligence (AI) based models are preferred over regression-based models, as they allow the data itself to identify the model structure rather than imposing any predefined structure on the data. Therefore, GEP was considered as the primary modelling technique in this study. GEP is chosen over ANN (Artificial Neural Networks) model, as ANN is a black-box model, whereas GEP is able to explain the developed forecast models with mathematical expressions.

Initial catchment conditions and climate indices are considered as the two main sources of predictability of Australian rainfall and streamflow. Since the incorporation of initial catchment condition is complex, this study exploited indices of large-scale climate anomalies as the predictor of seasonal streamflow as they showed significant concurrent and lagged correlations with seasonal streamflow of NSW. A number of climate indices were selected which included Interdecadal Pacific Oscillation (IPO)/ PDO (Pacific Decadal Oscillation), ENSO (El Nino Southern Oscillation), EMI (ENSO Modoki Index) originated from the Pacific Ocean, and IOD (Indian Ocean Dipole) originated from the Indian Ocean. This selection was based on the concurrent and single lagged concurrent analyses between streamflow and climate indices as well as on the outcomes of the previous research studies in the related field.

As a case study, twelve streamflow measuring stations were selected from four different regions of NSW in order to explore the spatial variation of influences of different climate indices on seasonal streamflow of this state. Monthly streamflow data for 102 years (1914-2016) were collected from the Bureau of Meteorology's (BoM) website. Monthly oceanic and atmospheric climate indices, i.e., IPO, PDO, ENSO (NINO3.4) and IOD data were obtained from Climate Explorer website (http://climexp.knmi.nl) while the EMI data was collected from the website of JAMSTEC (http://www.jamstec.go.jp/frcgc/research/dl/iod/modoki) for 102 years. This whole dataset was divided into two categories- calibration and validation periods. Since the aim was to explore the best predictor model, the range of calibration and validation periods was selected through a trial and error process which revealed two datasets (Dataset-1 and Dataset-2) to produce promising results for MLR and MNLR methods whereas GEP technique showed good results with Dataset-2 only. Therefore, the final comparison among the applied techniques was carried out with Dataset-2 only.

The research began with the investigation of the concurrent relationship between seasonal streamflow and seasonal climate drivers. It was revealed from the concurrent correlation analysis spring and summer seasons showed a statistically significant relation with climate indices. Based on the outcome of this analysis, since spring season seemed to have the strongest correlations with climate modes, it was chosen to carry on the analysis to develop spring streamflow forecast models using climate indices. Next single lagged correlation analysis was done between the single lagged climate index for nine antecedent months from December in the previous year until August of the current year and spring streamflow in order to gain an understanding of the extent of influence of climate indices on spring flow. This analysis helped to identify the comparatively more influential climate indices on spring streamflow with their corresponding lagged months.

Based on the outputs of single lagged correlation analysis, the combinations of multiple climate indices were selected, which served as the input data for developing the MLR models. To check the multicollinearity existed among the climate indices, concurrent correlation analysis between seasonal climate indices was carried out. While selecting the multiple indices for developing MLR models, indices originating from the same source were not used in the same model to avoid multicollinearity. Thus, multiple indices (two indices for this study) having statistically significant relationships with spring flow (obtained from single lagged correlation analysis) and originating from a different source to avoid multicollinearity, were selected for developing any MLR model. As mentioned earlier, the MLR analysis was done with both Datasets (Dataset-1 and Dataset-2). It paved a way to make a comparison between the MLR models developed with two different datasets as well. The MLR analysis depicts a clear view of regional variation in the influence of combined multiple models throughout the study area. Most numbers of models from all four regions showed statistically significant correlations for PDO and NINO3.4 combinations, implying the strong influence of these two indices on spring streamflow of NSW. Nevertheless, IPO and IOD combined models were also able to produce statistically significant outcomes for a few stations. After calibrating the models, the models were validated using the validation dataset. The correlation values in both calibration and validation periods were quite similar. To assess the reliability of the developed models, statistical performance test was conducted based on RMSE and MAE values. The best models were based on higher correlation values and lower statistical errors. It was evident that every time the combined models outperformed the models developed considering a single climate index in terms of Pearson correlation values. The correlation values obtained from the best models ranged between 0.51 and 0.65. The best models could provide a prediction of streamflow up to six months in advance, though the correlation values decrease with an increase in lagged months.

For this study, selections of the best models were based on the significant correlation values in both calibration and validation stages. However, while looking at the time series comparisons between the observed and simulated streamflow values, it is found that the developed models are unable to capture some unusual events like severe droughts or high floods. A simple multiple linear regression model consisting of only two climate indices is not expected to capture the complex relationships between streamflow and climate drivers very well and thus is not anticipated to provide a very good match with observed values. Moreover, in fact, rainfall and streamflow are also influenced by some other local and/or regional factors (i.e., temperature, humidity, wind speed, soil moisture, etc.), which are not possible to consider in such regression models.

Since the linear models could not produce very satisfactory results and it is assumed that the relationship between streamflow and climate indices is more inclined to non-linearity than linearity, a Multiple Non-Linear Regression Modelling (MNLR) approach was taken to capture the existing non-linear relationships between seasonal flow and climate modes. The same combination set of multiple indices were chosen for MNLR analysis. Again, for MNLR analysis, most of the stations showed statistically significant correlations with PDO and NINO3.4 combinations while influence IPO, IOD and EMI was also evident for few stations. Even though, in general, most stations have higher correlations for MNLR models than MLR models, the improvement was not very significant. Furthermore, the poor performance of MNLR models was observed in time series analysis while it failed to capture the extreme points, which is no better than MLR models.

The weak prediction skills of the developed MNLR models led to the application of a much more advanced technology which was Artificial Intelligence (AI) based Gene Expression Programming (GEP). The same combined sets of multiple indies as MLR and MNLR were used as input data for developing GEP models. The outcomes obtained from GEP models were very promising with their superior performances while compared to MLR or MNLR. Higher predictabilities of the developed models were ensured by higher correlation values which ranged between 0.57 and 0.97, which are mostly about twice the values achieved by MLR or MNLR. The reliability of the GEP models was even more enhanced while the simulated models were able to successfully capture the extreme points (i.e., very high and low) of the observed flow data and following the trend very well during the time series analysis of observed models versus simulated models. The developed GEP models are able to predict spring streamflow up to 5 months in advance with significantly high correlation values.

The comparative analysis between the outcomes obtained from this study (MLR, MNLR and GEP models) and past research studies based upon highest correlation values demonstrated that the current study (all MLR, MNLR and GEP models) outperformed any of the past research works confirming the significant combined impact of multiple indices on spring streamflow of the study region.

9.2 Conclusions

Large scale climate drivers have a strong relationship with seasonal streamflow of NSW while spring season appeared to have the strongest relationship with the lagged indices. The relationship is more inclined to non-linearity than linearity. Hence, the non-linear modelling technique is more reliable than linear modelling technique to develop streamflow forecast models incorporating lagged climate indices as predictors. In general, the Pacific Ocean climate drivers develop more accurate streamflow forecast models compared to Indian Ocean climate drivers. PDO and NINO3.4 have the strongest influence on spring streamflow across the whole of NSW. The Artificial Intelligence (AI) based Gene Expression Programming (GEP) has great potential to forecast seasonal streamflow of NSW using large-scale climate variables as the predictor. The accuracy and reliability of GEP are much higher than simple linear and non-linear regression modelling techniques (such as Multiple Linear Regression(MLR) and Multiple Non-Linear Regression (MNLR)) while predicting the seasonal flow of NSW. Along with its high prediction skill, GEP is able to generate output models as mathematical expressions which can be very useful to the water stakeholders even to someone without having much knowledge of the used software. This unique transparent quality of GEP makes it more suitable than other black-box modelling techniques such as ANN for developing streamflow forecast models. The research study reveals strong predictability of GEP models to forecast seasonal streamflow of NSW exploiting antecedent large-scale climate indices as predictors.

9.3 Future Recommendation

For further improvement of the predictability of GEP models, more than two climate indices can be incorporated into the model as the predictors of seasonal streamflow. This sort of study is mainly based on regional climate index/indices applicable for a region. However, a similar concept can be applied to other regions if any such index/indices are found to be effective for other regions. GEP modelling technique should be applied for other regions of Australia to explore the spatial variation of influence of climate modes on seasonal streamflow of this country. Again, this method can be used for other seasons as it will help to compare the seasonal variations of influence of different climate indices and thus will enable to identify the most influential indices for each season to predict streamflow. Hybrid models like Wavelet-GEP and Wavelet-ANN should be attempted to explore and compare their potential to forecast seasonal streamflow using lagged climate indices as the predictors. These areas of research will be taken into account for future research works as an extension of the current study.

References

Abawi, Y., Dutta, S., Zhang, X. and McClymont, D. (2005) ENSO-based streamflow forecasting and its application to water allocation and cropping decisions–an Australian experience, Regional hydrological impacts of climate change–Impact assessment and decision making, Brazil, pp. 346-354.

Abbot, J. & Marohasy, J. 2012. Application of artificial neural networks to rainfall forecasting in Queensland, Australia. *Advances in Atmospheric Sciences,* 29**,** 717-730

Abbot, J. and Marohasy, J., 2015. Using lagged and forecast climate indices with artificial intelligence to predict monthly rainfall in the Brisbane Catchment, Queensland, Australia. *International Journal of Sustainable Development and Planning*, *10*(1), pp.29-41.

Adamowski, J., Fung Chan, H., Prasher, S.O., Ozga-Zielinski, B. and Sliusarieva, A., 2012. Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resources Research*, *48*(1).

Allan, R.J., Lindesay, J.A. and Reason, C.J., 1995. Multidecadal variability in the climate system over the Indian Ocean region during the austral summer. *Journal of Climate*, *8*(7), pp.1853-1873.

Anderson, J., Van Den Dool, H., Barnston, A., Chen, W., Stern, W. & Ploshay, J. 1999. Present-day capabilities of numerical and statistical models for atmospheric extratropical seasonal simulation and prediction. *Bulletin of the American Meteorological Society,* 80**,** 1349-1361.

Arblaster, J.M.G.A., Meehl, G. and Moore, A., 2002. Interdecadal modulation of Australian rainfall. *Climate Dynamics*, *18*(6), pp.519-531.

Ashok, K., Guan, Z. and Yamagata, T. 2003. Influence of the Indian Ocean Dipole on the Australian winter rainfall, Geophysical Research Letters, 30 (15), CLM 6-1 - 6-4, doi: 10.1029/2003gl017926.

Ashok, K., Behera, S.K., Rao, S.A., Weng, H. and Yamagata, T., 2007. El Niño Modoki and its possible teleconnection. *Journal of Geophysical Research: Oceans*, *112*(C11).

Ashok, K. and Yamagata, T., 2009. The El Niño with a difference. Nature, 461(7263), pp.481-484.

Australian Government Bureau of Meteorology (2012). *Record-breaking La Niña events*. [online] Commonwealth of Australia 2020, Bureau of Meteorology. Available at: http://www.bom.gov.au/climate/enso/history/La-Nina-2010-12.pdf [Accessed 20 Jun. 2018].

Australian Government Bureau of Meteorology. (2020). *7-day Streamflow Forecasts: Water Information: Bureau of Meteorology*. [online] Available at: http://www.bom.gov.au/water/7daystreamflow/#panel=model_evaluation [Accessed 20 Jun. 2020].

Australian Government Bureau of Meteorology. (2020). *Seasonal Streamflow Forecasts: Water Information: Bureau of Meteorology*. [online] Available at: http://www.bom.gov.au/water/ssf/index.shtml [Accessed 20 Jun. 2020].

Aytek, A. and Alp, M. (2008) An application of artificial intelligence for rainfall-runoff modeling. Journal of Earth System Science, 117.2: 145-155.

Aytek, A. and Kisi, Ö. (2008) Genetic programming approach to suspended sediment modelling. Journal of Hydrology, 351 (3–4), 288–298.

Azamathulla, H. M. and Aminuddin, A. G. (2011) Genetic programming for predicting longitudinal dispersion coefficients in streams. Water Resources Management, 25.6: 1537-1544.

Azamathulla, H.M., Ghani, A.A., Zakaria, N.A. and Guven, A. (2010) Genetic programming to predict bridge pier scour. ASCE Journal of Hydraulic Engineering, 136(3):165–169.

Aziz, K., Rahman, A., Shamseldin, A.Y. and Shoaib, M., 2013. Co-active neuro fuzzy inference system for regional flood estimation in Australia. *J Hydrol Environ Res*, *1*(1), pp.11-20.

Babovic, V. and Keijzer, M. (2002) Rainfall-runoff modeling based on genetic programming. Nordic Hydrology, 33(5), 331–346.

Babovic, V. and Keijzer, M., 2002. Rainfall runoff modelling based on genetic programming. *Hydrology Research*, *33*(5), pp.331-346.

Bamston, A. G., ChelliahC, M. & Goldenberg, S. B. 1997. Documentation of a highly ENSO-related sst region in the equatorial pacific: Research note. *Atmosphere-ocean*, 35, 367-383.

Barnston, A.G., Li, S., Mason, S.J., DeWitt, D.G., Goddard, L. and Gong, X., 2010. Verification of the first 11 years of IRI's seasonal climate forecasts. *Journal of Applied Meteorology and Climatology*, *49*(3), pp.493-520.

Bilgili, M., 2010. Prediction of soil temperature using regression and artificial neural network models. *Meteorology and atmospheric physics*, *110*(1-2), pp.59-70.

Bond, N.R., Lake, P.S. and Arthington, A.H., 2008. The impacts of drought on freshwater ecosystems: an Australian perspective. *Hydrobiologia*, *600*(1), pp.3-16.

Bureau of Meteorology, <http://www.bom.gov.au/waterdata/> (Retrieved on 25.08.2016 and 11.12.2016)

Cai, W. and Rensch, P., 2012. The 2011 southeast Queensland extreme summer rainfall: A confirmation of a negative Pacific Decadal Oscillation phase?. *Geophysical Research Letters*, *39*(8).

Cai, W., Van Rensch, P., Cowan, T. & HendoN, H. H. 2012. An asymmetry in the IOD and ENSO teleconnection pathway and its impact on Australian climate. *Journal of Climate,* 25**,** 6318-6329.

Chavoshi, S., Sulaiman, W. N. A., Saghafian, B., Sulaiman, M. N. B. & Manaf, L. A. 2013. Flood prediction in southern strip of Caspian Sea watershed. *Water Resources*, 40, 593-605.

Chiew, F.H., 2006. Estimation of rainfall elasticity of streamflow in Australia. *Hydrological Sciences Journal*, *51*(4), pp.613-625.

Chiew, F.H., Piechota, T.C., Dracup, J.A. and McMahon, T.A., 1998. El Nino/Southern Oscillation and Australian rainfall, streamflow and drought: Links and potential for forecasting. *Journal of Hydrology*, *204*(1-4), pp.138-149.

Chiew, F.H.S. and Leahy, M.J., 2003. Inter-decadal Pacific oscillation modulation of the impact of El Niño/Southern Oscillation on Australian rainfall and streamflow.

Coulibaly, P., Anctil, F., Aravena, R. and Bobée, B., 2001. Artificial neural network modeling of water table depth fluctuations. *Water resources research*, *37*(4), pp.885-896.

CPTEC (Centro de Previsão de Tempo e Estudos Climáticos) 2006, El Niño e La Niña, http://wwwcptecinpebr/enos/artigosshtml.

Crochemore, L., Ramos, M.H. and Pappenberger, F., 2016. Bias correcting precipitation forecasts to improve the skill of seasonal streamflow forecasts.

Crueger, T., Zinke, J. and Pfeiffer, M., 2009. Patterns of Pacific decadal variability recorded by Indian Ocean corals. *International Journal of Earth Sciences*, *98*(1), pp.41-52.

D'Arrigo, R. and Ummenhofer, C.C., 2015. The climate of Myanmar: evidence for effects of the Pacific Decadal Oscillation. *International Journal of Climatology*, *35*(4), pp.634-640.

D'Arrigo, R., Cook, E.R., Wilson, R.J., Allan, R. and Mann, M.E., 2005. On the variability of ENSO over the past six centuries. *Geophysical Research Letters*, *32*(3).

Dettinger, M.D. and Diaz, H.F., 2000. Global characteristics of stream flow seasonality and variability. *Journal of Hydrometeorology*, *1*(4), pp.289-310.

Dorado, J., Rabunal, J. R., Pazos, A., Rivero, D., Santos, A. and Puertas, J. (2003) Prediction and modeling of the rainfall-runoff transformation of a typical urban basin using ANN and GP. Applied Artificial Intelligence, 17, 329–343.

Drecourt, J. P. (1999) Application of neural networks and genetic programming to rainfall-runoff modeling. D2K Technical Rep. No. 0699-1-1, Danish Hydraulic Institute, Hørsholm, Denmark.

Drosdowsky, W. & Chmabers, L. E. 2001. Near-global sea surface temperature anomalies as predictors of Australian seasonal rainfall. *Journal of Climate,* 14**,** 1677-1687.

Duc, H.N., Rivett, K., MacSween, K. and Le-Anh, L., 2017. Association of climate drivers with rainfall in New South Wales, Australia, using Bayesian Model Averaging. *Theoretical and Applied Climatology*, *127*(1-2), pp.169-185.

Dutta, S.C., Ritchie, J.W., Freebairn, D.M. and Abawi, G.Y., 2006. Rainfall and streamflow response to El Niño Southern Oscillation: a case study in a semiarid catchment, Australia. *Hydrological sciences journal*, *51*(6), pp.1006-1020.

Einfalt, T., Denolux, T. and Jacquet, G. (1990). A radar rainfall forecasting method designed for hydrological purposes. Journal of Hydrology, Vol. 114, pp. 229-244.

Erskine, W.D. and Warner, R.F., 1988. Geomorphic effects of alternating flood- and drought-dominated regimes on NSW coastal rivers. *Fluvial geomorphology of Australia*, pp.223-244.

Evan Hajania, Ataur Rahman, and Elias Ishak. 2017. Trends in extreme rainfall in the state of New South Wales, Australia. Hydrological Sciences JournaL.

F. Murphy, B. & Timbal, . B., 2007. A review of recent climate variability and climate change. International Journal OF Climatology*,* p. 859–879

Fan, Y., Allen, M.R., Anderson, D.L.T. and Balmaseda, M.A., 2000. How predictability depends on the nature of uncertainty in initial conditions in a coupled model of ENSO. *Journal of climate*, *13*(18), pp.3298-3313.

Fawcett, R., and R. C. Stone (2010), A comparison of two seasonal rainfall forecasting systems for Australia, Australian Meteorological and Oceanographic Journal, 60(2), 15-24.

Feng, J. and Li, J., 2011. Influence of El Niño Modoki on spring rainfall over south China. *Journal of Geophysical Research: Atmospheres*, *116*(D13).

Feng, M. & Meyers, G. 2003. Interannual variability in the tropical Indian Ocean: a two-year time-scale of Indian Ocean Dipole. *Deep Sea Research Part II: Topical Studies in Oceanography,* 50**,** 2263-2284.

Ferreira, C. (2001) Gene expression programming in problem solving. 6th Online World Conference on Soft Computing in Industrial Applications, September.

Ferreira, C. (2002) Gene expression programming in problem solving. In "Soft computing and industry", Eds: Roy R., Köppen M., Ovaska S., Furuhashi T., Hoffmann F., pp. 635-653. Springer, London.

Ferreira, C., 2006. *Gene expression programming: mathematical modeling by an artificial intelligence* (Vol. 21). Springer.

Field, A., 2009. Discovering Statistics Using SPSS, Thrid Edition.

Folland, C.K., Parker, D.E., Colman, A. and Washington, R., 1999. Beyond El Nino: decadal and interdecadal climate variability.

Folland, C.K., Renwick, J.A., Salinger, M.J. and Mullan, A.B., 2002. Relative influences of the interdecadal Pacific oscillation and ENSO on the South Pacific convergence zone. *Geophysical Research Letters*, *29*(13).

Franks, S. W. Multi-decadal climate variability, New South Wales, Australia. *Water Science and Technology* 49.7 (2004): 133-140.

Franks, S.W. and Kuczera, G., 2002. Flood frequency analysis: Evidence and implications of secular climate variability, New South Wales. *Water resources research*, *38*(5).

Franks, S.W., 2002a. Identification of a change in climate state using regional flood data. *Hydrology and Earth System Sciences*, *6*(1), pp.11-16.

Franks, S.W., 2002b. Assessing hydrological change: deterministic general circulation models or spurious solar correlation? *Hydrological Processes*, *16*(2), pp.559-564.

Frías, M.D., Iturbide, M., Manzanas, R., Bedia, J., Fernández, J., Herrera, S., Cofiño, A.S. and Gutiérrez, J.M., 2018. An R package to visualize and communicate uncertainty in seasonal climate prediction. *Environmental modelling & software*, *99*, pp.101-110.

Galarnyk, M., Understanding Boxplots, 2018. *URL: https://towardsdatascience.com/understanding-boxplots-5e2df7bcbd51 (visited on 09/12/2018)*.

Gallant, AJ, Hennessy, KJ & Risbey, J 2007, 'Trends in rainfall indices for six Australian regions: 1910-2005', Australian Meteorological Magazine, vol. 56, no. 4, pp.223-241.

Ghani, A.A. and Azamathulla, H.M. (2011) Gene expression programming for sediment transport in sewer pipe systems. Journal of Pipeline System Engineering Practice, 2(3),102–106.

Ghani, A.A. and Azamathulla, H.M., 2014. Development of GEP-based functional relationship for sediment transport in tropical rivers. *Neural Computing and Applications*, *24*(2), pp.271-276.

Giustolisi, O., 2004. Using genetic programming to determine Chezy resistance coefficient in corrugated channels. *Journal of Hydroinformatics*, *6*(3), pp.157-173.

Goddard, L., Aitchellouche, Y., Baethgen, W., Dettinger, M., Graham, R., Hayman, P., Kadi, M., Martínez, R. and Meinke, H., 2010. Providing seasonal-to-interannual climate information for risk management and decision-making. *Procedia Environmental Sciences*, *1*, pp.81-101.

Goddard, L., Mason, S. J., Zebiak, S. E., RopelewskI, C. F., Basher, R. & Cane, M. A. 2001. Current approaches to seasonal to interannual climate predictions. *International Journal of Climatology,* 21**,** 1111-1152.

Goddard, L., Mason, S.J., Zebiak, S.E., Ropelewski, C.F., Basher, R. and Cane, M.A., 2001. Current approaches to seasonal to interannual climate predictions. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *21*(9), pp.1111-1152.

Gudmundsson, L. and Seneviratne, S.I. (2015). *Runoff and Streamflow Data for Climate Extremes*. [online] ETH,Zurich: IACETH, Land-Climate Dynamics. Available at: https://www.wcrp-climate.org/images/grand_challenges/climate_extremes/Data_Workshop/WCRP_GCE_workshop_gudmundsson_2015_v3.pdf [Accessed 24 Sep. 2020].

Guven, A. and Aytek, A. (2009) A new approach for stage–discharge relationship: gene-expression programming. Journal of Hydrologic Engineering, 14 (8), 812–820.

Guven, A. and Gunal, M. (2008) A genetic programming approach for prediction of local scour downstream hydraulic structures. Journal of Irrigation and Drainage Engineering 134(2), DOI: 10.1061/(ASCE)0733-9437(2008)134:2(241).

Guven, A. and Kişi, Ö. (2011) Daily pan evaporation modeling using linear genetic programming technique. Irrigation Science, 29(2): 135-145

Guven, A. and Kişi, Ö., 2011. Daily pan evaporation modeling using linear genetic programming technique. *Irrigation science*, *29*(2), pp.135-145.

Halide, H. & Ridd, P. 2008. Complicated ENSO models do not significantly outperform very simple ENSO models. *International Journal of Climatology,* 28**,** 219-233.

Hanley, D.E., Bourassa, M.A., O'Brien, J.J., Smith, S.R. and Spade, E.R., 2003. A quantitative evaluation of ENSO indices. *Journal of Climate*, *16*(8), pp.1249-1258.

Haque, M. M., Rahman, A., Hagare, D., and Kibria, G. 2013. Principal Component Regression Analysis in Water Demand Forecasting: An Application to the Blue Mountains, NSW, Australia. Journal of Hydrology and Environment Research, Volume 1, No 1: 49-59.

Harris, E. L., Babovic, V. and Falconer, R. A. (2003) Velocity predictions in compound channels with vegetated floodplains using genetic programming. International Journal of River Management, 1(2): 117–123.

He, X., Guan, H., ZhanG, X. & Simmons, C. T. 2014. A wavelet-based multiple linear regression model for forecasting monthly rainfall. *International Journal of Climatology,* 34**,** 1898-1912.

Heller, G.Z., Stasinopoulos, D.M., Rigby, R.A. and Van Ogtrop, F., 2009. Randomly stopped sum models: a hydrological application. *International Workshop on Statistical Modelling*, pp.156-163

Henley, B.J., Gergis, J., Karoly, D.J., Power, S., Kennedy, J. and Folland, C.K., 2015. A tripole index for the Interdecadal Pacific Oscillation. *Climate dynamics*, *45*(11-12), pp.3077-3090.

Hoerling, M.P., Whitaker, J.S., Kumar, A. and Wang, W., 2001. The midlatitude warming during 1998–2000. *Geophysical research letters*, *28*(5), pp.755-758.

Horridge, M., Madden, J. & Wittwer, G., 2005. The Impact of the 2002-2003 drought on Australia. Journal of Plicy Modeling, 27(3), pp. 285-308.

Hossain, I., Rasel, H.M., Imteaz, M.A. and Mekanik, F., 2018. Long-term seasonal rainfall forecasting: efficiency of linear modelling technique. *Environmental Earth Sciences*, *77*(7), p.280.

Hossain, I.; Rasel, H.M.; Imteaz, M.A.; Moniruzzaman, M. Statistical correlations between rainfall and climate indices inWestern Australia. In Proceedings of the 21st International Congress on Modelling and Simulation, Gold Coast, Australia, 29 November–4 December 2015; pp. 1991–1997.

Huff, F. and Vogel, J. (1981). Real-time rainfall monitoring-prediction system and urban hydrology applications. Journal of Water Resources Planning and Management, Vol. WR2, pp. 419-435.

Ihara, C., KushniR, Y., Cane, M. A. & De La Peña, V. H. 2007. Indian summer monsoon rainfall and its link with ENSO and Indian Ocean climate indices. *International Journal of Climatology,* 27**,** 179-187.

Ivakhnenko, A.G., 1970. Heuristic self-organization in problems of engineering cybernetics. *Automatica*, *6*(2), pp.207-219.

Jain, A., and Kumar, A.M. (2007) Hybrid neural network models for hydrologic time series forecasting. Applied Soft Computing, 7, 585–592.

Jain, S. K. (2016). Long-term streamflow forecasting based on ensemble streamflow prediction technique: a case study in New Zealand. Water resources management, 30(7), 2295-2309.

Jenicek, M., Seibert, J., Zappa, M., Staudinger, M. and Jonas, T., 2016. Importance of maximum snow accumulation for summer low flows in humid catchments. *Hydrology and Earth System Sciences*, *20*(2), pp.859-874.

Kahya, E. and Dracup, J. A. (1993) US streamflow patterns in relation to the El Niño/Southern Oscillation. Water Resources Research, 29 (8), 2491–2503.

Kahya, E. and Dracup, J.A., 1993. US streamflow patterns in relation to the El Niño/Southern Oscillation. *Water Resources Research*, *29*(8), pp.2491-2503.

Karimi, S., Shiri, J., Kisi, O. and Shiri, A.A., 2016. Short-term and long-term streamflow prediction by using'wavelet–gene expression'programming approach. *ISH Journal of Hydraulic Engineering*, *22*(2), pp.148-162.

Karumuri, A.K. and Yamagata, T., 2009. Climate change: the El Niño with a difference. *Nature*, *461*, pp.481-484.

Khan,S.,Robinson, D.,m Beddek, R, Wang, B., Dassanayake, D. and Rana, T., 2005. *Hydro-climatic and economic evaluation of seasonal climate forecasts for risk based irrigation management*.

Khu, S.T., Liong, S.Y., Babovic, V., Madsen, H. and Muttil, N., 2001. Genetic programming and its application in real-time runoff forecasting 1. *JAWRA Journal of the American Water Resources Association*, *37*(2), pp.439-451.

Kiafar, H., Babazadeh, H., Marti, P., Kisi, O., Landeras, G., Karimi, S. and Shiri, J., 2017. Evaluating the generalizability of GEP models for estimating reference evapotranspiration in distant humid and arid locations. *Theoretical and Applied Climatology*, *130*(1-2), pp.377-389.

Kiem, A. S., and Verdon-Kidd, D. C. 2009. Climatic drivers of Victorian streamflow: Is ENSO the dominant influence?, Australian Journal of Water Resources, 13(1), 17-29.

Kiem, A.S. and Franks, S.W., 2001. On the identification of ENSO-induced rainfall and runoff variability: a comparison of methods and indices. *Hydrological sciences journal*, *46*(5), pp.715-727.

Kiem, A.S. and Franks, S.W., 2004. Multi-decadal variability of drought risk, eastern Australia. *Hydrological Processes*, *18*(11), pp.2039-2050.

Kiem, A.S. and Verdon-Kidd, D.C., 2009. Climatic drivers of Victorian streamflow: is ENSO the dominant influence?. *Australian Journal of Water Resources*, *13*(1), pp.17-29.

Kiem, A.S., Franks, S.W. and Kuczera, G., 2003. Multi-decadal variability of flood risk. *Geophysical Research Letters*, *30*(2).

Kiladis, G.N. and Diaz, H.F., 1989. Global climatic anomalies associated with extremes in the Southern Oscillation. *Journal of Climate*, *2*(9), pp.1069-1090.

Kim, H.M., Webster, P.J. and Curry, J.A., 2012. Seasonal prediction skill of ECMWF System 4 and NCEP CFSv2 retrospective forecast for the Northern Hemisphere Winter. *Climate Dynamics*, *39*(12), pp.2957-2973.

King, A.D., Alexander, L.V. and Donat, M.G., 2013. Asymmetry in the response of eastern Australia extreme rainfall to low-frequency Pacific variability. *Geophysical Research Letters*, *40*(10), pp.2271-2277.

Kirono, D.G., Chiew, F.H. and Kent, D.M., 2010. Identification of best predictors for forecasting seasonal rainfall and runoff in Australia. *Hydrological processes*, *24*(10), pp.1237-1247.

Kisi O. and Shiri J. (2010) A comparison of genetic programming and ANFIS, The International Symposium on Innovations in Intelligent Systems and Applications. PP 118–122, 21–24 June, Kayseri and Cappadocia, Turkey.

Kişi, Ö. (2007). “Evapotranspiration modeling from climate data using a neural computing technique.” Hydrol. Processes, 21(6), 1925–1934.

Kişi, Ö. 2006 Generalized regression neural networks for evapotranspiration modelling. Hydrological Sciences Journal, 51(6), pp.1092-1105.

Kisi, O. and Shiri, J. (2011) Precipitation forecasting using wavelet-genetic programming and wavelet-neuro-fuzzy conjunction models. Water Resources Management, 25(13): 3135–3152.

Kişi, Ö., 2006. Daily pan evaporation modelling using a neuro-fuzzy computing technique. *Journal of hydrology*, *329*(3-4), pp.636-646.

Kişi, Ö., 2007. Streamflow forecasting using different artificial neural network algorithms. *Journal of Hydrologic Engineering*, *12*(5), pp.532-539.

Kişi, O., Nia, A.M., Gosheh, M.G., Tajabadi, M.R.J. and Ahmadi, A. (2012) Intermittent streamflow forecasting by using several data driven techniques. Water Resources Management, 26(2), 457-474.

Kisi, O., Shiri, J. and Tombul, M. (2013) Modeling rainfall-runoff process using soft computing techniques. Computers and Geosciences. 51, 108–117.

Koza, J.R. (1992) Genetic programming II, automatic discovery of reusable subprograms. MIT Press, Cambridge, MA.

Koza, J.R. (1994) Genetic programming as a means for programming computers by natural selection. Statistics and Computing, 4(2), 87-112.

Koza, J.R., 1995, November. Survey of genetic algorithms and genetic programming. In *Wescon conference record* (pp. 589-594). WESTERN PERIODICALS COMPANY.

Koza, J.R., Andre, D., Keane, M.A. and Bennett III, F.H., 1999. Genetic programming III: Darwinian invention and problem solving (Vol. 3). Morgan Kaufmann Publishers.

Koza, J.R., Andre, D., Keane, M.A. and Bennett III, F.H., 1999. *Genetic programming III: Darwinian invention and problem solving* (Vol. 3). Morgan Kaufmann.

Koza, J.R., Bennett III, F.H., Andre, D. and Keane, M.A. (2003) Genetic programming problem solver with automatically defined stores loops and recursions. U.S. Patent 6,532,453.

Koza, J.R., Keane, M.A. and Streeter, M.J., 2003. Evolving inventions. *Scientific American*, *288*(2), pp.52-59.

Koza, JR & Koza, JR 1992, *Genetic programming: on the programming of computers by means of natural selection,* MIT press.

Langford, S. and Hendon, H.H., 2011. *Assessment of international seasonal rainfall forecasts for Australia and the*.

Langford, S.; Hendon, H.H. Assessment of international seasonal rainfall forecasts for Australia and the benefit of multi-model ensembles for improving reliability. In the Centre for Australian Weather and Climate Research Technical Report No. 039; The Centre for Australian Weather and Climate Research: Victoria, Australian, 2011.

Latif, M., Kleeman, R. and Eckert, C., 1997. Greenhouse warming, decadal variability or El Nino? An attempt to understand the anomalous 1990s. *Journal of Climate*, *10*(9), pp.2221-2239.

LATT, Z. Z. & WITTENBERG, H. 2014. Improving Flood Forecasting in a Developing Country: A Comparative Study of Stepwise Multiple Linear Regression and Artificial Neural Network. *Water Resources Management,* 28**,** 2109-2128.

Lim, E.P., Hendon, H.H., Anderson, D.L., Charles, A. and Alves, O., 2011. Dynamical, statistical–dynamical, and multimodel ensemble forecasts of Australian spring season rainfall. *Monthly weather review*, *139*(3), pp.958-975.

Liu, Y. and Fan, K., 2014. An application of hybrid downscaling model to forecast summer precipitation at stations in China. *Atmospheric research*, *143*, pp.17-30.

Lough, J. M. 2007. Tropical river flow and rainfall reconstructions from coral luminescence: Great Barrier Reef, Australia. *Paleoceanography,* 22.

Maier, H. R. & Dandy, G. C. 2000. Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modelling and software,* 15**,** 101-124.

Mantua, N.J. and Hare, S.R., 2002. The Pacific decadal oscillation. *Journal of oceanography*, *58*(1), pp.35-44.

Mantua, N.J. and Hare, S.R., 2002. The Pacific decadal oscillation. Journal of oceanography, 58(1), pp.35-44.

Mantua, N.J., Hare, S.R., Zhang, Y., Wallace, J.M. and Francis, R.C., 1997. A Pacific interdecadal climate oscillation with impacts on salmon production. *Bulletin of the american Meteorological Society*, *78*(6), pp.1069-1080.

Manzanas, R., Frías, M.D., Cofiño, A.S. and Gutiérrez, J.M., 2014. Validation of 40 year multimodel seasonal precipitation forecasts: The role of ENSO on the global skill. *Journal of Geophysical Research: Atmospheres*, *119*(4), pp.1708-1719.

McBride, J.L. and Nicholls, N., 1983. Seasonal relationships between Australian rainfall and the Southern Oscillation. *Monthly Weather Review*, *111*(10), pp.1998-2004.

McMahon, T.A., Finlayson, B.L., Haines, A.T. and Srikanthan, R., 1992. *Global runoff: continental comparisons of annual flows and peak discharges*. Catena Verlag.

Meehl, G.A., Hu, A., Santer, B.D. and Xie, S.P., 2016. Contribution of the Interdecadal Pacific Oscillation to twentieth-century global surface temperature trends. Nature Climate Change, 6(11), pp.1005-1008.

Mehrotra, R., Srikanthan, R. and Sharma, A., 2006. A comparison of three stochastic multi-site precipitation occurrence generators. *Journal of Hydrology*, *331*(1-2), pp.280-292.

Mekanik, F. and Imteaz, M.A., 2012, August. Forecasting Victorian spring rainfall using ENSO and IOD: A comparison of linear multiple regression and nonlinear ANN. In *2012 2nd International Conference on Uncertainty Reasoning and Knowledge Engineering* (pp. 86-89). IEEE.

Mekanik, F. and Imteaz, M.A., 2012. A multivariate artificial neural network approach for rainfall forecasting: case study of Victoria, Australia. In *Proceedings of the world congress on engineering and computer Science* (Vol. 1, pp. 557-561).

Mekanik, F., Imteaz, M.A., Gato-Trinidad, S. and Elmahdi, A., 2013. Multiple regression and Artificial Neural Network for long-term rainfall forecasting using large scale climate modes. *Journal of Hydrology*, *503*, pp.11-21.

Mekanik, F., Imteaz, M.A., Gato-Trinidad, S. and Elmahdi, A., 2013. Multiple regression and Artificial Neural Network for long-term rainfall forecasting using large scale climate modes. *Journal of Hydrology*, *503*, pp.11-21.

Meyers, G, Mcintosh, P, Pigot, L & Pook, M. 2007, 'The years of El Nino, La Nina, and interactions with the tropical Indian Ocean', Journal of Climate, vol. 20, no. 13, pp. 2872-2880.

MEYERS, G. 1996. Variation of Indonesian throughflow and the El Niño-Southern Oscillation. *Journal of Geophysical Research: Oceans (1978–2012),* 101**,** 12255-12263.

MEYERS, G., MCINTOSH, P., PIGOT, L. & POOK, M. 2007. The years of El Niño, La Niña and interactions with the tropical Indian Ocean. *Journal of Climate,* 20**,** 2872-2880.

Moss, M.E., Pearson, C.P. and McKerchar, A.I., 1994. The Southern Oscillation index as a predictor of the probability of low streamflows in New Zealand. *Water Resources Research*, *30*(10), pp.2717-2723.

Mühlenbein, H., 1997. Genetic algorithms.

Nazemosadat, M.J. and Cordery, I., 1997. The influence of geopotential heights on New South Wales rainfall. *Meteorology and Atmospheric Physics*, *63*(3-4), pp.179-193.

Nicholls, N., 2010. Local and remote causes of the southern Australian autumn-winter rainfall decline, 1958–2007. *Climate dynamics*, *34*(6), pp.835-845.

Nicholls, N., Butler, C.D. and Hanigan, I., 2006. Inter-annual rainfall variations and suicide in New South Wales, Australia, 1964–2001. International Journal of Biometeorology, 50(3), pp.139-143.

Nicholls, N., Glantz, M. H. and Katz, R. W 1991. Teleconnections Linking Worldwide Climate Anomalies: Scientific Basis and Societal Impact, *Cambridge University Press*.

Nourani, V., Kisi, Ö. and Komasi, M., 2011. Two hybrid artificial intelligence approaches for modeling rainfall–runoff process. Journal of Hydrology, 402(1-2), pp.41-59.

Osborne, J.W. and Overbay, A., 2004. The power of outliers (and why researchers should always check for them). *Practical Assessment, Research, and Evaluation*, *9*(1), p.6.

Osborne, J.W. and Overbay, A., 2004. The power of outliers (and why researchers should always check for them). *Practical Assessment, Research, and Evaluation*, *9*(1), p.6.

Osborne, J.W. and Waters, E., 2002. Four assumptions of multiple regression that researchers should always test. *Practical assessment, research, and evaluation*, *8*(1), p.2.

Owens, D.T., Clarkson, N.M. & Clewett, J.F., Persistence of Australian streamflow and its application to seasonal forecasts. *MODSIM 2003.International Congress on Modelling and Simulation. Modelling and Simulation Society of Australia and New Zealand, July 2003*, *01*(A02/09)

Owens, D.T., Clarkson, N.M. and Clewett, J.F., 2013, Persistence of Australian streamflow and its application to seasonal forecasts.

Palmer, J.G., Cook, E.R., Turney, C.S., Allen, K., Fenwick, P., Cook, B.I., O'Donnell, A., Lough, J., Grierson, P. and Baker, P., 2015. Drought variability in the eastern Australia and New Zealand summer drought atlas (ANZDA, CE 1500–2012) modulated by the Interdecadal Pacific Oscillation. Environmental Research Letters, 10(12), p.124002.

Pepler, A., Timbal, B., Rakich, C. and Coutts-Smith, A., 2014. Indian Ocean Dipole overrides ENSO's influence on cool season rainfall across the Eastern Seaboard of Australia. *Journal of Climate*, *27*(10), pp.3816-3826.

Pepler, A., Timbal, B., Rakich, C. and Coutts-Smith, A., 2014. Indian Ocean dipole overrides ENSO's influence on cool season rainfall across the eastern seaboard of Australia. Journal of Climate, 27(10), pp.3816-3826.

Piechota, T, C, Chiew, F, H, Dracup, J, A & McMahon, T, A 1998, 'Seasonal streamflow forecasting in eastern Australia and the El Niño–Southern Oscillation', Water Resources Research, 34(11), pp3035-3044.

Piechota, T.C. and Dracup, J.A., 1996. Drought and regional hydrologic variation in the United States: Associations with the El Niño-Southern Oscillation. *Water Resources Research*, *32*(5), pp.1359-1373.

Piechota, T.C., Chiew, F.H., Dracup, J.A. and McMahon, T.A., 1998. Seasonal streamflow forecasting in eastern Australia and the El Niño–Southern Oscillation. *Water Resources Research*, *34*(11), pp.3035-3044.

Power, S., Casey, T., Folland, C., Colman, A. and Mehta, V., 1999. Inter-decadal modulation of the impact of ENSO on Australia. *Climate Dynamics*, *15*(5), pp.319-324.

Qi, C. and Chang, N.B., 2011. System dynamics modeling for municipal water demand estimation in an urban region under uncertain economic impacts. *Journal of environmental management*, *92*(6), pp.1628-1641.

Quan, X., Hoerling, M., Whitaker, J., Bates, G. and Xu, T., 2006. Diagnosing sources of US seasonal forecast skill. *Journal of climate*, *19*(13), pp.3279-3293.

Rabunal, J.R., Puertas, J., Suarez, J. and Rivero, D., 2007. Determination of the unit hydrograph of a typical urban basin using genetic programming and artificial neural networks. *Hydrological Processes: An International Journal*, *21*(4), pp.476-485.

Rao, S. A., Behera, S. K., Masumoto, Y. & Yamagata, T. 2002. Interannual subsurface variability in the tropical Indian Ocean with a special emphasis on the Indian Ocean dipole. *Deep Sea Research Part II: Topical Studies in Oceanography,* 49**,** 1549-1572.

Rasel, H.M., Imteaz, M.A. and Mekanik, F., 2015, November. Evaluating the effects of lagged ENSO and SAM as potential predictors for long-term rainfall forecasting. In *Water Resources and Environment: Proceedings of the 2015 International Conference on Water Resources and Environment (Beijing, 25-28 July 2015)* (p. 125). CRC Press.

Rasel, H.M., Imteaz, M.A. and Mekanik, F., 2016. Investigating the influence of Remote Climate Drivers as the Predictors in Forecasting South Australian spring rainfall. *International Journal of Environmental Research*, *10*(1), pp.1-12.

Rasmussen, J.L., 1988. Evaluating outlier identification tests: Mahalanobis D squared and Comrey Dk. *Multivariate Behavioral Research*, *23*(2), pp.189-202.

Rayner, S., Lach, D. and Ingram, H., 2005. Weather forecasts are for wimps: why water resource managers do not use climate forecasts. *Climatic change*, *69*(2-3), pp.197-227.

Risbey, J. S., 2011. Dangerious Climate change and water resources in Australia. Regional Environmental Change, Volume I, pp. 197-203.

Risbey, J.S., Pook, M.J., McIntosh, P.C., Wheeler, M.C. and Hendon, H.H., 2009. On the remote drivers of rainfall variability in Australia. *Monthly Weather Review*, *137*(10), pp.3233-3253.

Robertson, D, E., and Wang, Q, J., 2009, 'A Bayesian joint probability approach to seasonal prediction of streamflows: Predictor selection and skill assessment', In H2009: 32nd Hydrology and Water Resources Symposium, Newcastle: Adapting to Change (p 1545) Engineers Australia.

Robertson, D.E. and Wang, Q.J., 2008. An investigation into the selection of predictors and skill assessment using the Bayesian joint probability (BJP) modelling approach to seasonal forecasting of streamflows. *Water for a Healthy Country flagship report, CSIRO Land and Water, Canberra*.

Robertson, D.E. and Wang, Q.J., 2009. A Bayesian joint probability approach to seasonal prediction of streamflows: Predictor selection and skill assessment. In *H2009:*

32nd Hydrology and Water Resources Symposium, Newcastle: Adapting to Change (p. 1545). Engineers Australia.

Ropelewski, C.F. and Halpert, M.S., 1987. Global and regional scale precipitation patterns associated with the El Niño/Southern Oscillation. *Monthly weather review*, *115*(8), pp.1606-1626.

Rossel, F. & Cadier, E. 2009. El Niño and prediction of anomalous monthly rainfalls in Ecuador. *Hydrological Processes,* 23**,** 3253-3260.

Ruiz, J.E., Cordery, I. and Sharma, A., 2007. Forecasting streamflows in Australia using the tropical Indo-Pacific thermocline as predictor. *Journal of hydrology*, *341*(3), pp.156-164.

Sadhuram, Y. & Murthy, T. R. 2008. Simple multiple regression model for long range forecasting of Indian summer monsoon rainfall. *Meteorology and Atmospheric Physics,* 99**,** 17-24.

Saffle, B. and Greene, D. (1978). The Role of Radar in the Flash Flood Watch Warning System- Johnstown Examined, 18th Conf. on Radar Meteorology, Atlanta, GA. American Meteorological Society, Boston, MA.

Saji, N. H., and Yamagata, T. 2003. Possible impacts of Indian Ocean dipole mode events on global climate. Climate Research 25: 151–169.

Saji, N.H. and Yamagata, T., 2003. Possible impacts of Indian Ocean dipole mode events on global climate. *Climate Research*, *25*(2), pp.151-169.

Saji, N.H., Goswami, B.N., Vinayachandran, P.N. and Yamagata, T., 1999. A dipole mode in the tropical Indian Ocean. *Nature*, *401*(6751), p.360.

Salinger, M.J., Renwick, J.A. and Mullan, A.B., 2001. Interdecadal Pacific oscillation and south Pacific climate. *International Journal of Climatology*, *21*(14), pp.1705-1721.

Samadianfard, S., 2012. Gene expression programming analysis of implicit Colebrook–White equation in turbulent flow friction factor calculation. *Journal of Petroleum Science and Engineering*, *92*, pp.48-55.

Sattar, A.M., 2016. Closure to "Gene Expression Models for the Prediction of Longitudinal Dispersion Coefficients in Transitional and Turbulent Pipe Flow" by Ahmed MA Sattar. *Journal of Pipeline Systems Engineering and Practice*, *7*(4), p.07016002.

Savic, D.A., Walters, G.A. and Davidson, J.W. (1999) A genetic programming approach to rainfall-runoff modelling. Water Resources Management. 13(3), 219–231.

Seo, D.J. and Smith, J.A. (1992). Radar-based short-term rainfall prediction. Journal of Hydrology, Vol. 131, pp. 341-367.

Sharma A. 2000. Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 1 – a strategy for system predictor identification. J. Hydrol. 239: 232–239.

Sharma, A., 2000. Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 3—A nonparametric probabilistic forecast model. *Journal of Hydrology*, *239*(1), pp.249-258.

Sharma, S., Srivastava, P., Fang, X. and Kalin, L., 2015. Long-range hydrologic forecasting in El Niño Southern Oscillation-affected coastal watersheds: Comparison of climate model and weather generator approach. *Journal of Hydrologic Engineering*, *20*(12), p.06015006.

Shiri, J. and Kişi, O. (2011a) Comparison of genetic programming with neuro-fuzzy systems for predicting short-term water table depth fluctuations. Computers & Geosciences, 37(10): 1692-1701.

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

Shiri, J. and Kisi, O., 2010. Short-term and long-term streamflow forecasting using a wavelet and neuro-fuzzy conjunction model. *Journal of Hydrology*, *394*(3-4), pp.486-493.

Shiri, J. and KişI, Ö., 2011. Comparison of genetic programming with neuro-fuzzy systems for predicting short-term water table depth fluctuations. *Computers & Geosciences*, *37*(10), pp.1692-1701.

Shiri, J., Kisi, O., Landeras, G., López, J. J., Nazemi, A. H. and Stuyt, L. C. P. M. (2012) Daily reference evapotranspiration modeling by using genetic programming approach in the Basque Country (Northern Spain). Journal of Hydrology, 414-415: 302-316.

Simpson, H.J., Cane, M.A., Lin, S.K., Zebiak, S.E. and Herczeg, A.L., 1993. Forecasting annual discharge of River Murray, Australia, from a geophysical model of ENSO. *Journal of Climate*, *6*(2), pp.386-390.

Steffen, W. 2015, Thirsty Country: Climate Change and Drought in Australia, Climate Council, Taschetto, AS & England, MH 2009, 'El Nino Modoki impacts on Australian rainfall', Journal of Climate, vol. 22, no. 11, pp. 3167-3174.

Stevens, J.P., 1984. Outliers and influential data points in regression analysis. *Psychological bulletin*, *95*(2), p.334.

Stone, R. and Auliciems, A., 1992. SOI phase relationships with rainfall in eastern Australia. *International Journal of Climatology*, *12*(6), pp.625-636.

Supharatid, S., 2003. Application of a neural network model in establishing a stage–discharge relationship for a tidal river. *Hydrological processes*, *17*(15), pp.3085-3099.

Suppiah, R., 2004. Trends in the southern oscillation phenomenon and Australian rainfall and changes in their relationship. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *24*(3), pp.269-290.

Tabari, H., Marofi, S. and Sabziparvar, A.A., 2010. Estimation of daily pan evaporation using artificial neural network and multivariate non-linear regression. *Irrigation Science*, *28*(5), pp.399-406.

Taschetto, A.S. and England, M.H., 2009. El Niño Modoki impacts on Australian rainfall. *Journal of Climate*, *22*(11), pp.3167-3174.

Tayfur, G., 2002. Artificial neural networks for sheet sediment transport. *Hydrological Sciences Journal*, *47*(6), pp.879-892.

Tennant, W.J. and Hewitson, B.C., 2002. Intra‐seasonal rainfall characteristics and their importance to the seasonal prediction problem. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *22*(9), pp.1033-1048.

Traore, S. and Guven, A., 2012. Regional-specific numerical models of evapotranspiration using gene-expression programming interface in Sahel. *Water resources management*, *26*(15), pp.4367-4380.

Troup, A. J. 1965. The 'southern oscillation', Quarterly Journal of the Royal Meteorological Society, 91(390), 490-506, doi:10.1002/qj.49709139009.

Troup, A. J. 1965. The ‘southern oscillation’, Quarterly Journal of the Royal Meteorological Society, 91(390), 490-506, doi:10.1002/qj.49709139009.

Vance, T.R., Roberts, J.L., Plummer, C.T., Kiem, A.S. and Van Ommen, T.D., 2015. Interdecadal Pacific variability and eastern Australian megadroughts over the last millennium. *Geophysical Research Letters*, *42*(1), pp.129-137.

Verdon, D.C., Wyatt, A.M., Kiem, A.S. and Franks, S.W., 2004. Multidecadal variability of rainfall and streamflow: Eastern Australia. *Water Resources Research*, *40*(10).

Verdon-Kidd, D. C. & Kiem, A. S. 2009a. Nature and causes of protracted droughts in southeast Australia: Comparison between the Federation, WWII, and Big Dry droughts. *Geophysical Research Letters,* 36.

Verdon-Kidd, D. C. & Kiem, A. S. 2009b. On the relationship between large-scale climate modes and regional synoptic patterns that drive Victorian rainfall. *Hydrology and Earth System Sciences,* 13**,** 467-479.

Wang, B., Lee, J.Y., Kang, I.S., Shukla, J., Park, C.K., Kumar, A., Schemm, J., Cocke, S., Kug, J.S., Luo, J.J. and Zhou, T., 2009. Advance and prospectus of seasonal prediction: assessment of the APCC/CliPAS 14-model ensemble retrospective seasonal prediction (1980–2004). *Climate Dynamics*, *33*(1), pp.93-117.

Wang, Q.J. and Robertson, D.E., 2011. Multisite probabilistic forecasting of seasonal flows for streams with zero value occurrences. *Water Resources Research*, *47*(2).

Wang, Q.J., Robertson, D.E. and Chiew, F.H.S., 2009. A Bayesian joint probability modeling approach for seasonal forecasting of streamflows at multiple sites. *Water Resources Research*, *45*(5).

Wang, W., Van Gelder, P.H., Vrijling, J.K. and Ma, J. (2006) Forecasting daily streamflow using hybrid ANN models. Journal of Hydrology, 324(1-4), 383-399.

Webster, P.J., Moore, A.M., Loschnigg, J.P. and Leben, R.R., 1999. Coupled ocean-atmosphere dynamics in the Indian Ocean during 1997-98. *Nature*, *401*(6751), p.356.

Westra, S. and Sharma, A., 2009. Probabilistic estimation of multivariate streamflow using independent component analysis and climate information. *Journal of Hydrometeorology*, *10*(6), pp.1479-1492.

Westra, S., Sharma, A., Brown, C. and Lall, U., 2008. Multivariate streamflow forecasting using independent component analysis. *Water Resources Research*, *44*(2).

Wheeler, MC, Hendon, HH, Cleland, S, Meinke, H & Donald, A 2009, 'Impacts of the Madden-Julian oscillation on Australian rainfall and circulation', *Journal of Climate*, vol. 22, no. 6, pp. 1482-1498.

Whigham, P.A. and Crapper, P.F., 1999. Time series modelling using genetic programming: An application to rainfall-runoff models. *Advances in genetic programming*, *3*(5), pp.89-104.

White, W. B., Gershunov, A., Annis, J. L., McKeon, G. and Syktus, J. I. (2004) Forecasting Australian drought using Southern Hemisphere modes of sea-surface temperature variability. Intedesimprnational Journal of Climatology, 24(15): 1911-1927.

Whiting, J.P., Lambert, M.F. and Metcalfe, A.V., 2003. Modelling persistence in annual Australia point rainfall. *Hydrology and Earth System Sciences Discussions*, *7*(2), pp.197-211.

Wijffels, S. & Meyers, G. 2004. An intersection of oceanic waveguides: Variability in the Indonesian Throughflow region. *Journal of Physical Oceanography,* 34**,** 1232-1253.

Wikipedia Contributors (2020). *National Farmers' Federation*. [online] Wikipedia. Available at: https://en.wikipedia.org/wiki/National_Farmers%27_Federation [Accessed 10 Sep. 2020].

Wilks, D. S. and Murphy, A. H. (1986) A decision-analytic study of the joint value of seasonal precipitation and temperature forecasts in a choice-of-crop problem, Atmosphere-Ocean, 24(4), 353-368.

Winsemius, H.C., Dutra, E., Engelbrecht, F.A., Archer Van Garderen, E., Wetterhall, F., Pappenberger, F. and Werner, M.G.F., 2014. The potential value of seasonal forecasts in a changing climate in southern Africa. *Hydrology and Earth System Sciences*, *18*(4), pp.1525-1538.

Wooldridge, S.A., Franks, S.W. and Kalma, J.D., 2001. Hydrological implications of the Southern Oscillation: variability of the rainfall-runoff relationship. *Hydrological Sciences Journal*, *46*(1), pp.73-88.

YAMAGATA, T., BEHERA, S. K., LUO, J. J., MASSON, S., JURY, M. R. & RAO, S. A. 2004. Coupled ocean-atmosphere variability in the tropical Indian Ocean. *Earth's Climate***,** 189-211.

Zahiri, A. and Eghbali, P. (2012) Gene expression programming for prediction of flow discharge in compound channels. Journal of Civil Engineering and Urbanism. 2(4), 164–169.

Zakaria, N.A., Azamathulla, H. M., Chang, C.K., Ghani, A. (2010) Gene-expression programming for total bed material load estimation—a case study. Science of Total Environment, 408(21):5078–5085.

Zealand, CM, Burn, DH, and Simonovic, SP, 1999, Short term streamflow forecasting using artificial neural networks, Journal of Hydrology, Vol. 214, pp. 32-48.

Zhang, L. and Delworth, T.L., 2015. Analysis of the characteristics and mechanisms of the Pacific decadal oscillation in a suite of coupled models from the Geophysical Fluid Dynamics Laboratory. Journal of climate, 28(19), pp.7678-7701.

Zhang, X.S., Amirthanathan, G.E., Bari, M.A., Laugesen, R.M., Shin, D., Kent, D.M., MacDonald, A.M., Turner, M.E. and Tuteja, N.K., 2016. How streamflow has changed across Australia since the 1950s: evidence from the network of hydrologic reference stations. *Hydrology and Earth System Sciences*, *20*(9), p.3947.

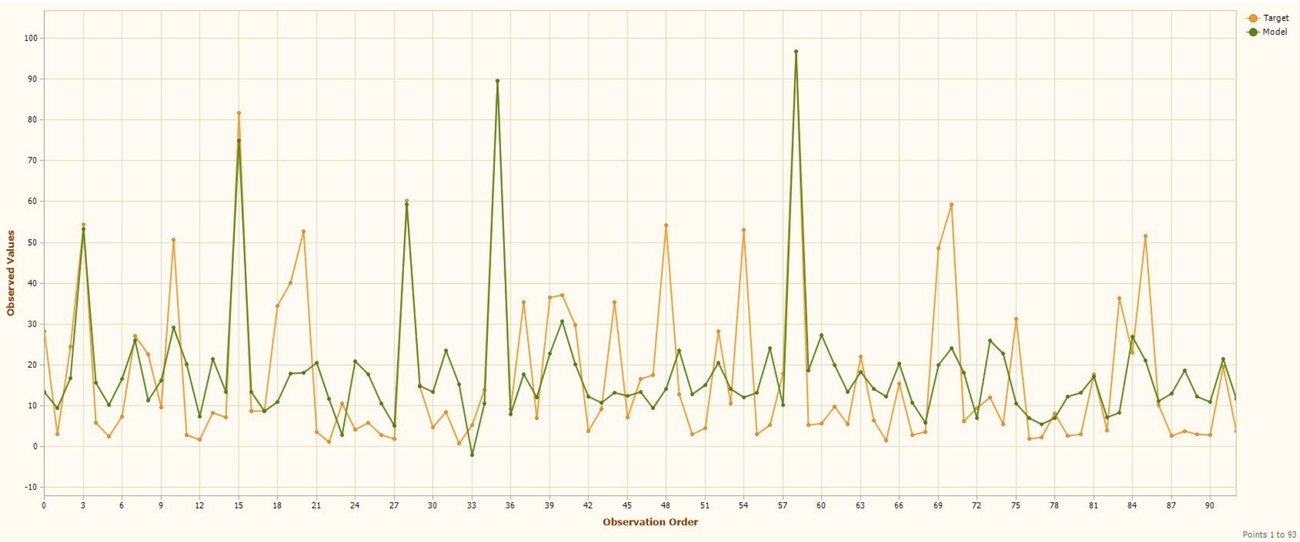
Zhang, Y., Wallace, J. M. & battisti, D. S. 1997. ENSO-like interdecadal variability: 1900-93. *Journal of Climate,* 10**,** 1004-1020.

Appendix A

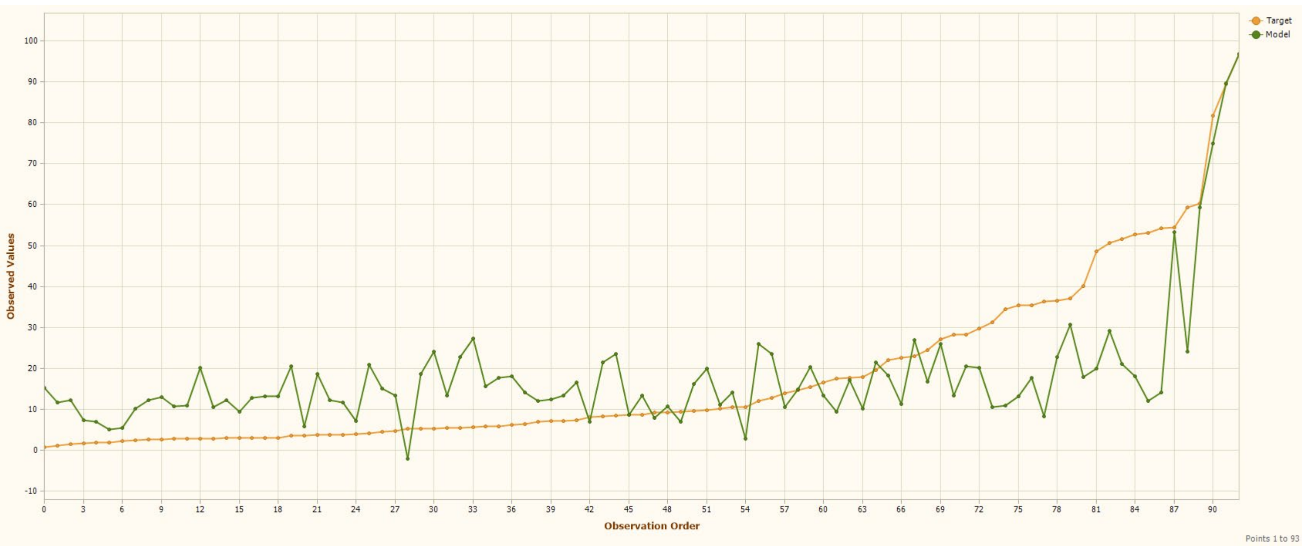
A1 Best GEP Model Output for Singleton Station (NNSW)

A1.1 Output results for Calibration period

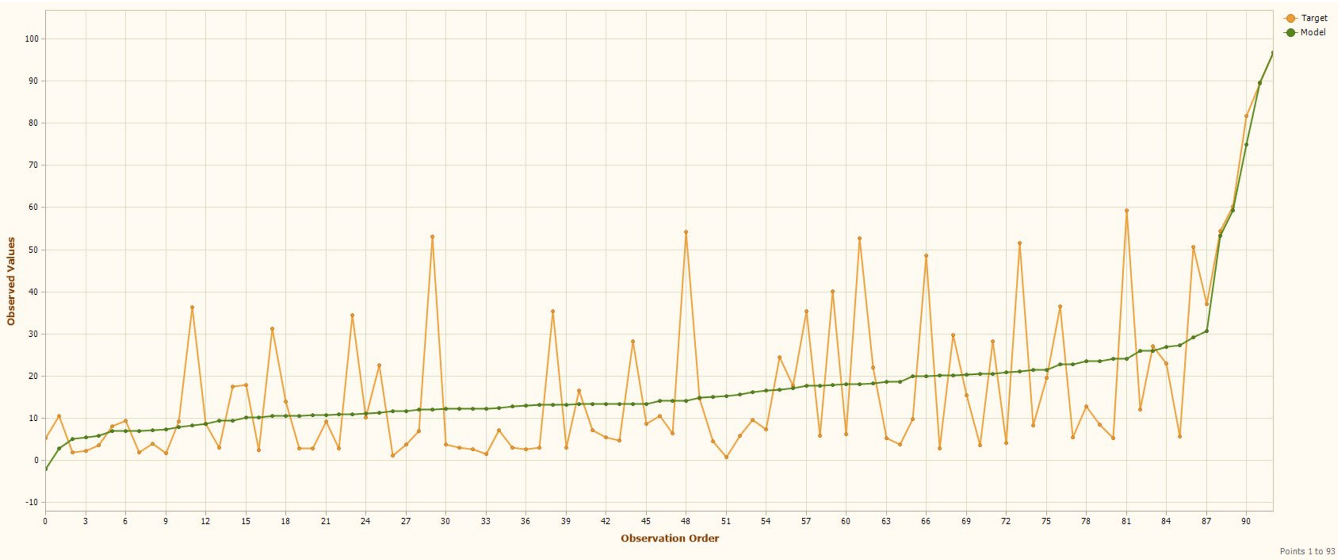
A1.1.1 Curve Fitting



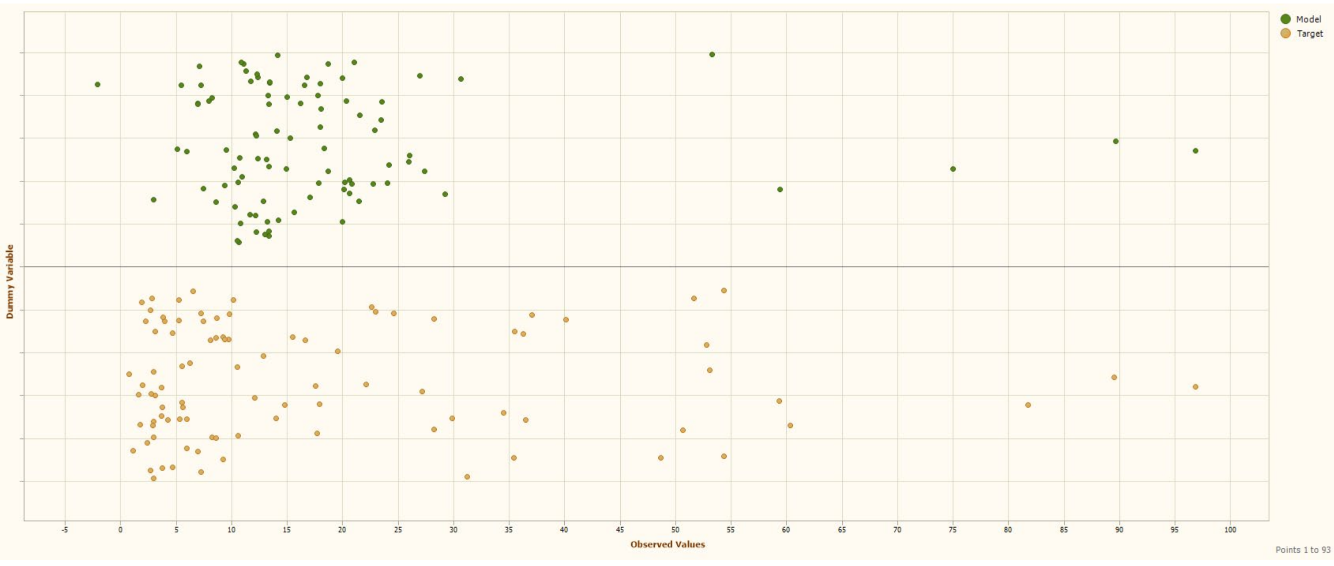
A1.1.2 Target sorted fitting



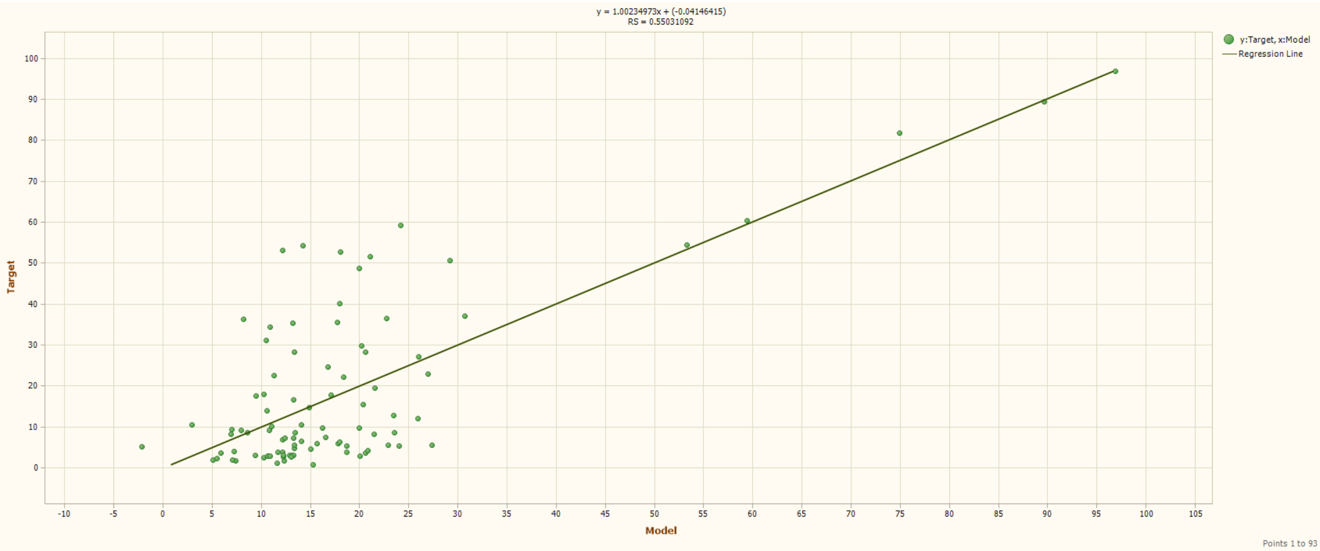
A1.1.3 Model sorted fitting



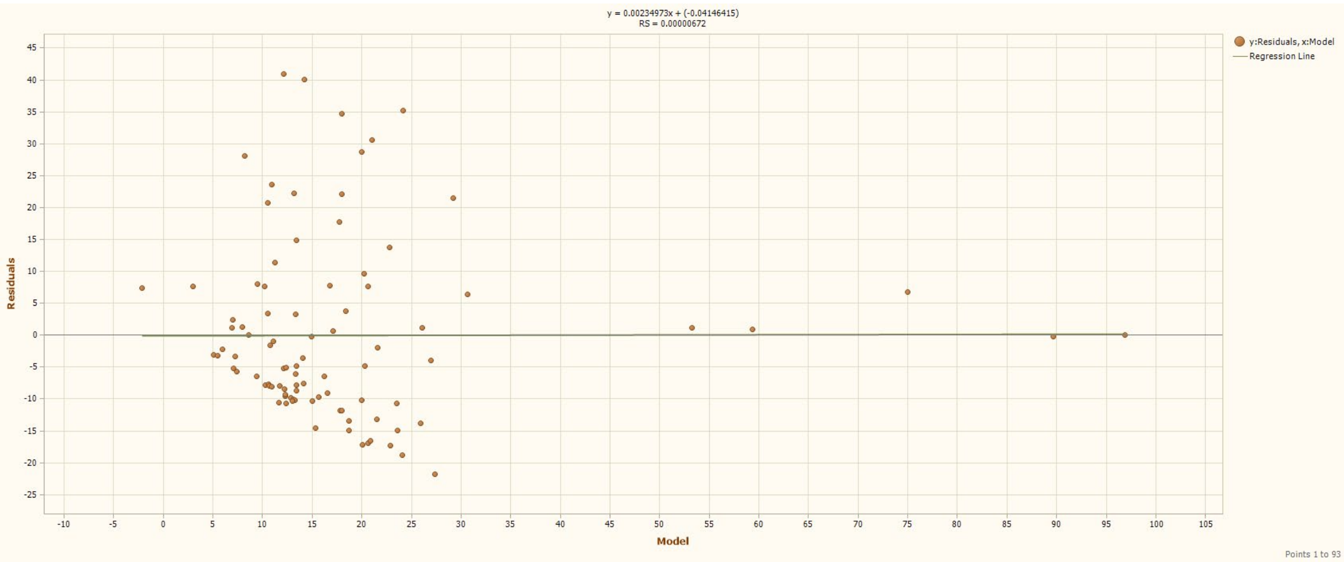
A1.1.4 Stacked distribution



A1.1.5 Scatter plot



A1.1.6 Residuals plot

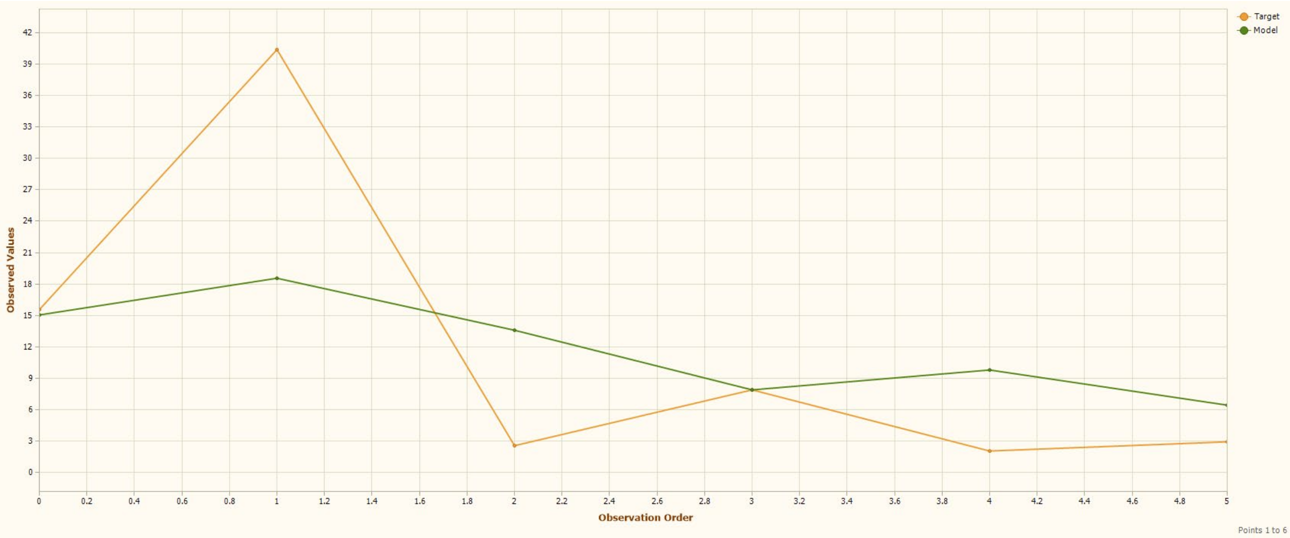


A1.1.7 Model performance

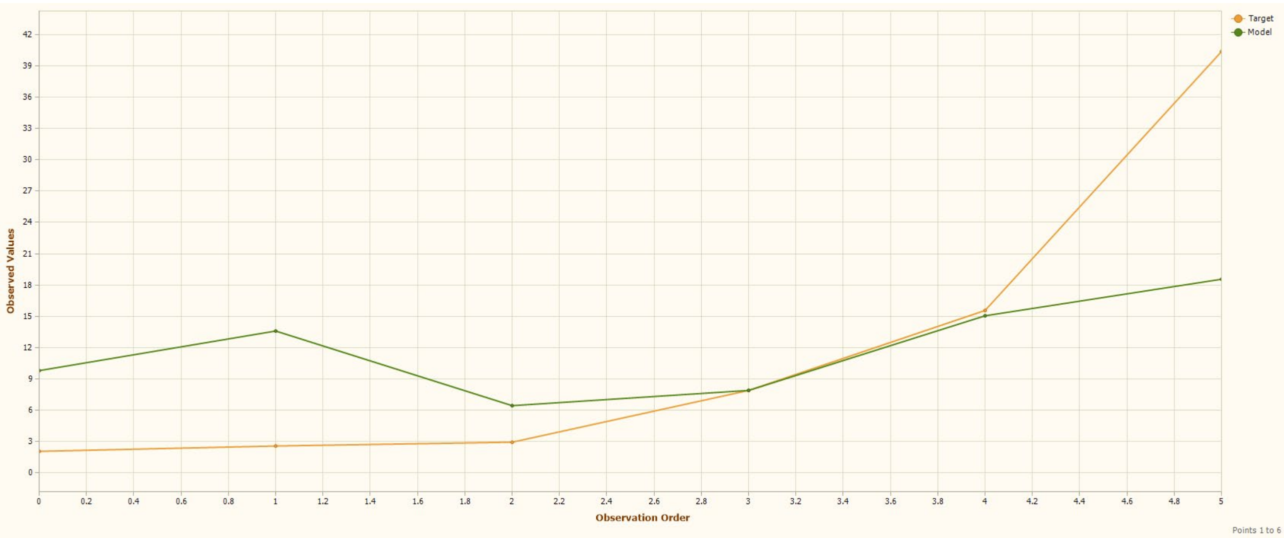
Fitness	67.3200044241749
MSE	191.945109895219
RMSE	13.8544256429207
MAE	10.6225830678003
RSE	0.449692111361631
RRSE	0.670590867341355
RAE	0.669938026087153
Correlation Coefficient	0.741829439115973
R-square	0.550310916739119
Calculation Errors	0

A1.2 Output results for Validation period

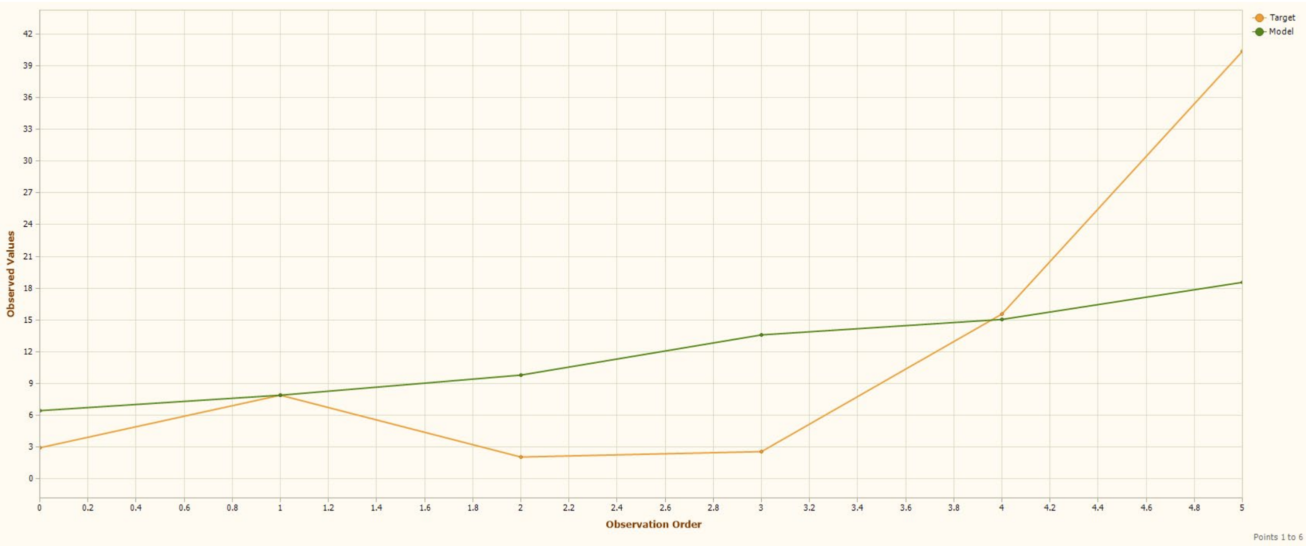
A1.2.1 Curve fitting



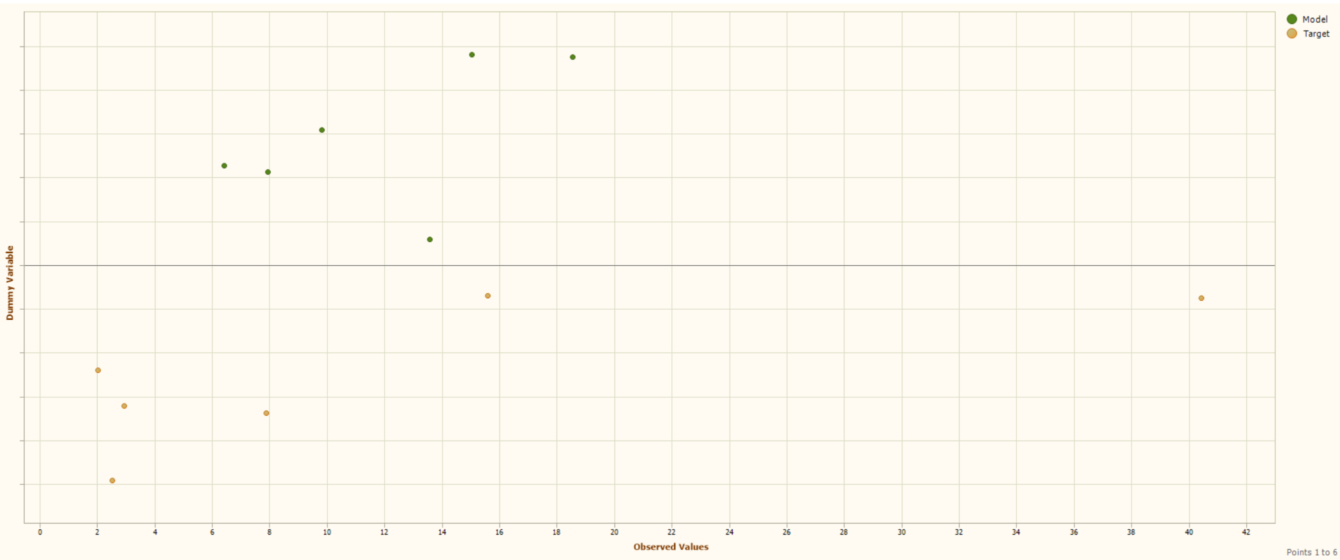
A1.2.2 Target sorted fitting



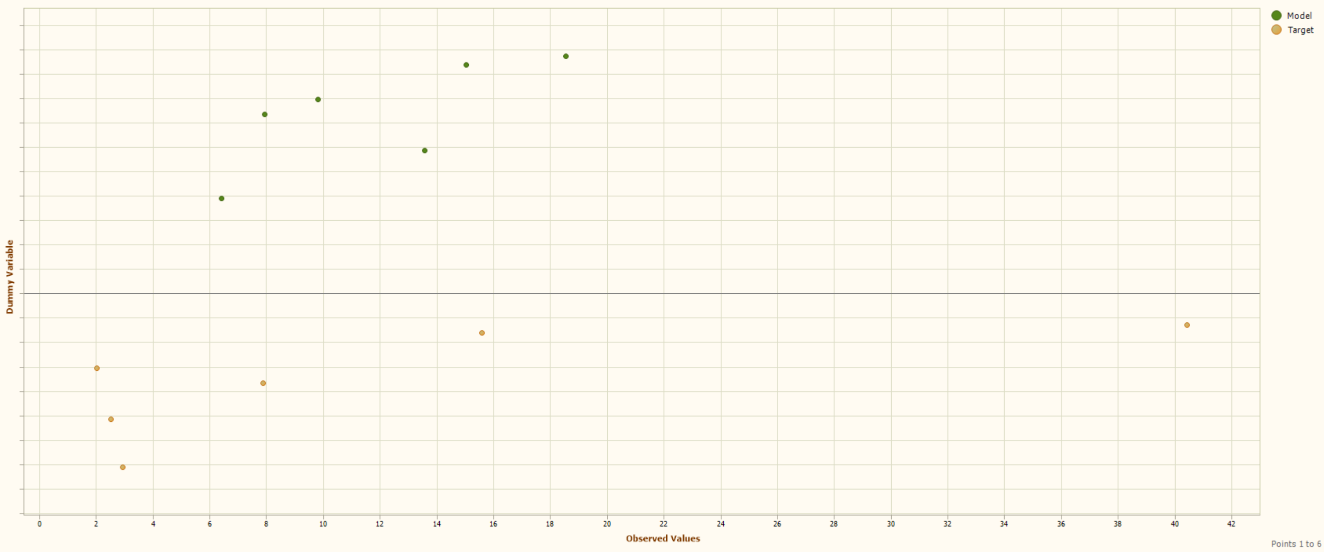
A1.2.3 Model sorted fitting



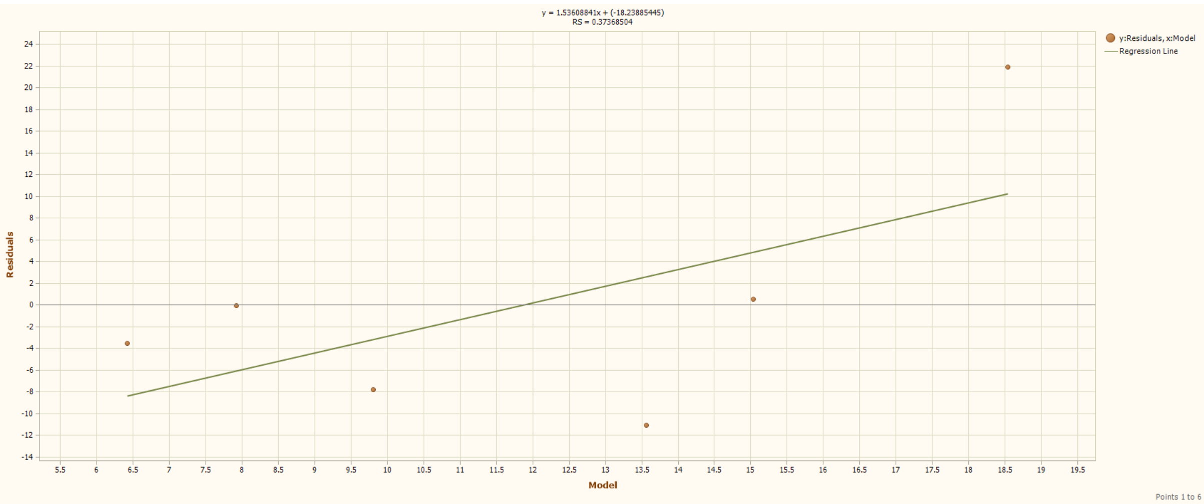
A1.2.4 Stacked distribution



A1.2.5 Scatter plot



A1.2.6 Residuals plot

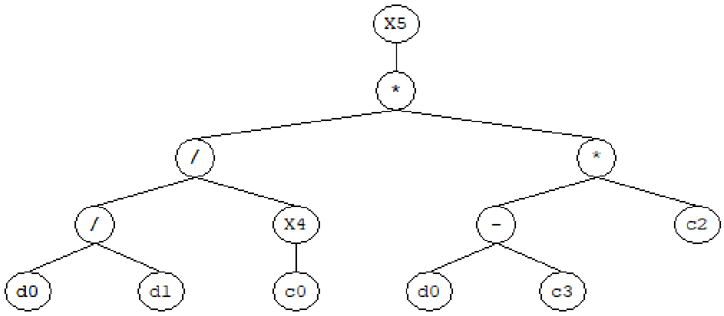


A1.2.7 Model performance

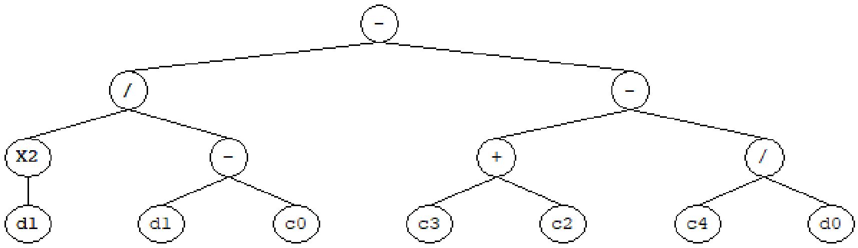
Fitness	86.1715174775266
MSE	112.460989012659
RMSE	10.6047625627667
MAE	7.46667904254414
RSE	0.607936373345048
RRSE	0.779702746785624
RAE	0.695177253859006
Correlation Coefficient	0.786918588245728
R-square	0.61924086452665
Calculation Errors	0

A1.3 Output model explained by Expression Tree

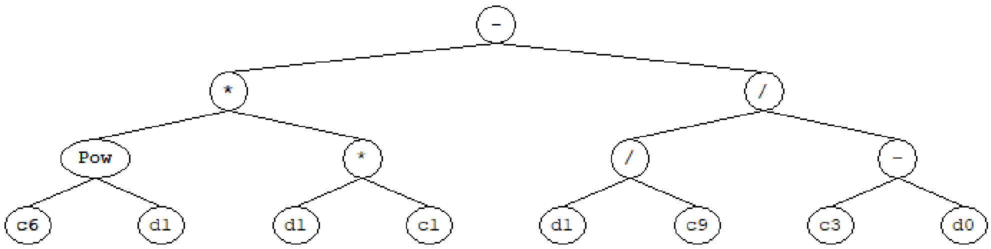
Sub-ET 1



Sub-ET 2



Sub-ET 3



A1.4 Output model explained by MATLAB

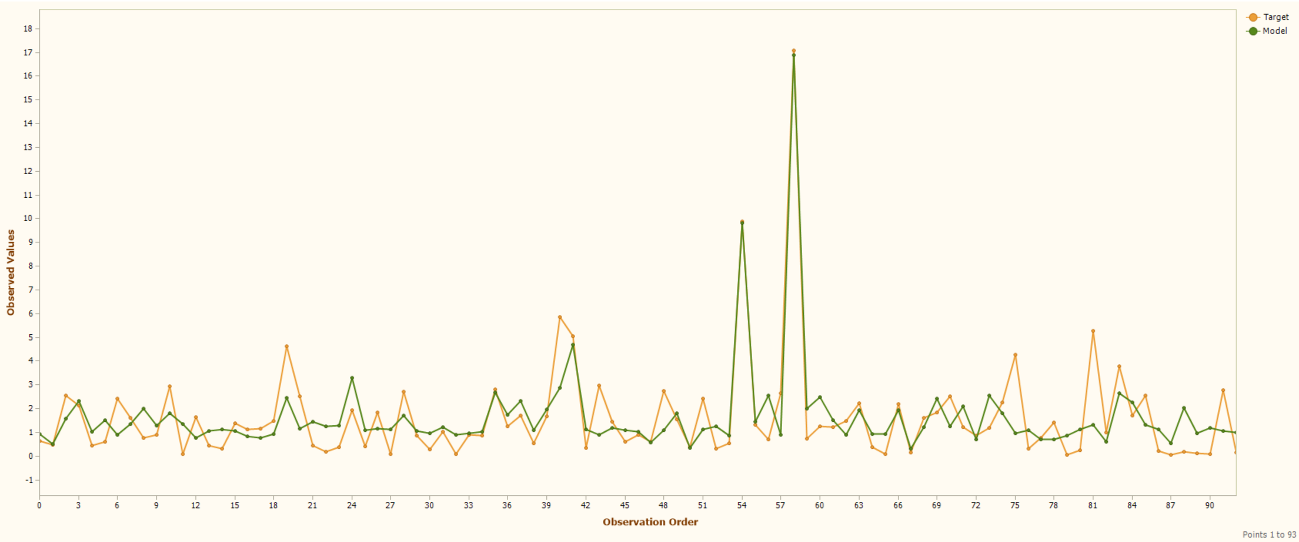
```
%-----------------------------------------------------------------------

% Regression model generated by GeneXproTools 5.0 on 23/09/2020 10:49:55 PM

% GEP File: C:\Users\resha\Google Drive\Final GeneXPro\FINAL\st01_5P_4N.gep

% Training Records:  93

% Validation Records:   6

% Fitness Function:  RMSE

% Training Fitness:  67.3200044241749

% Training R-square: 0.550310916739119

% Validation Fitness:   86.1715174775266

% Validation R-square:  0.61924086452665

%-----------------------------------------------------------------------


function result = gepModelQ1C(d)


G1C2 = -4.76412091433454;

G1C0 = -6.76543473616749;

G1C3 = -7.8735236671041;

G2C0 = -0.208719013227574;

G2C3 = -7.24332728049562;

G2C2 = -7.240425122837;

G2C4 = -0.212381382647828;

G3C6 = 0.784930507005318;

G3C1 = -10.9276076134536;

G3C9 = 0.975943839991455;

G3C3 = -0.991148106326487;


Five_PC = 1;

Four_NC = 2;
```

```
y = 0.0;

y = ((((d(Five_PC)/d(Four_NC))/(G1C0^4))*((d(Five_PC)-G1C3)*G1C2))^5);
y = y + (((d(Four_NC)^2)/(d(Four_NC)-G2C0))-((G2C3+G2C2)-(G2C4/d(Five_PC))));
y = y + ((realpow(G3C6,d(Four_NC))*(d(Four_NC)*G3C1))-((d(Four_NC)/G3C9)/(G3C3-d(Five_PC))));

result = y;
```

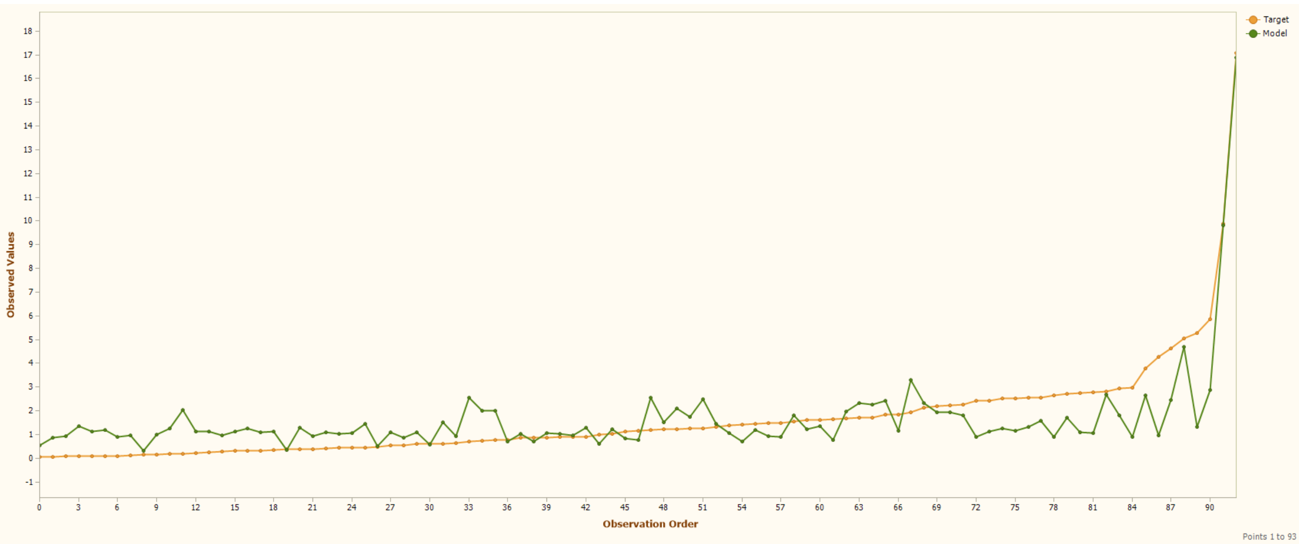
A2 Best GEP Model Output for Coggan Station (NNSW)

A2.1 Output results for Calibration period

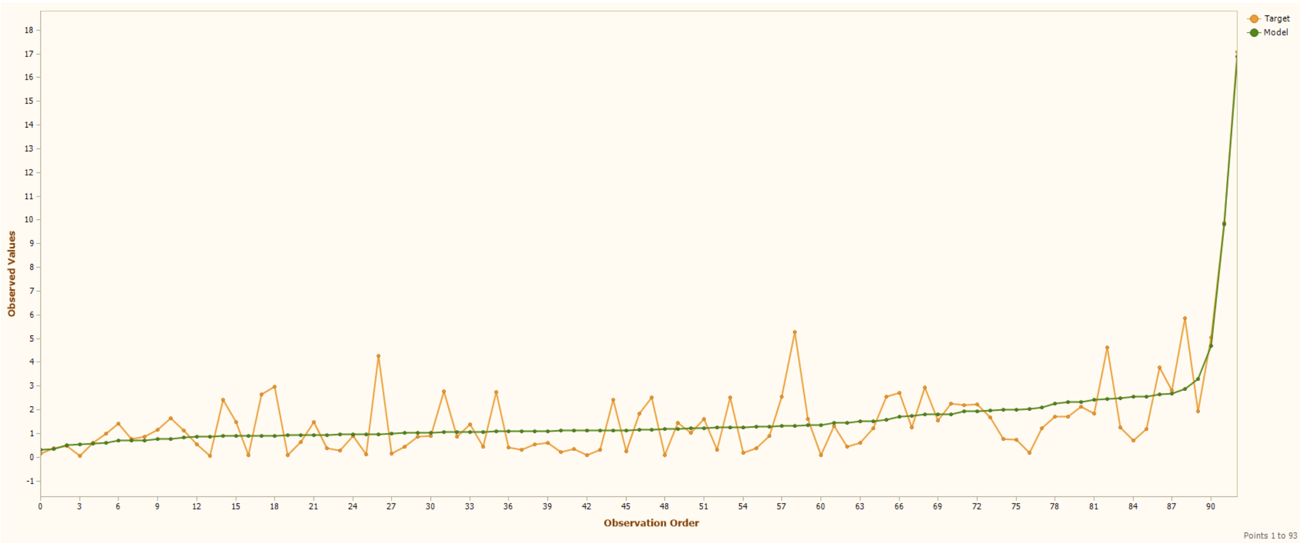
A2.1.1 Curve Fitting



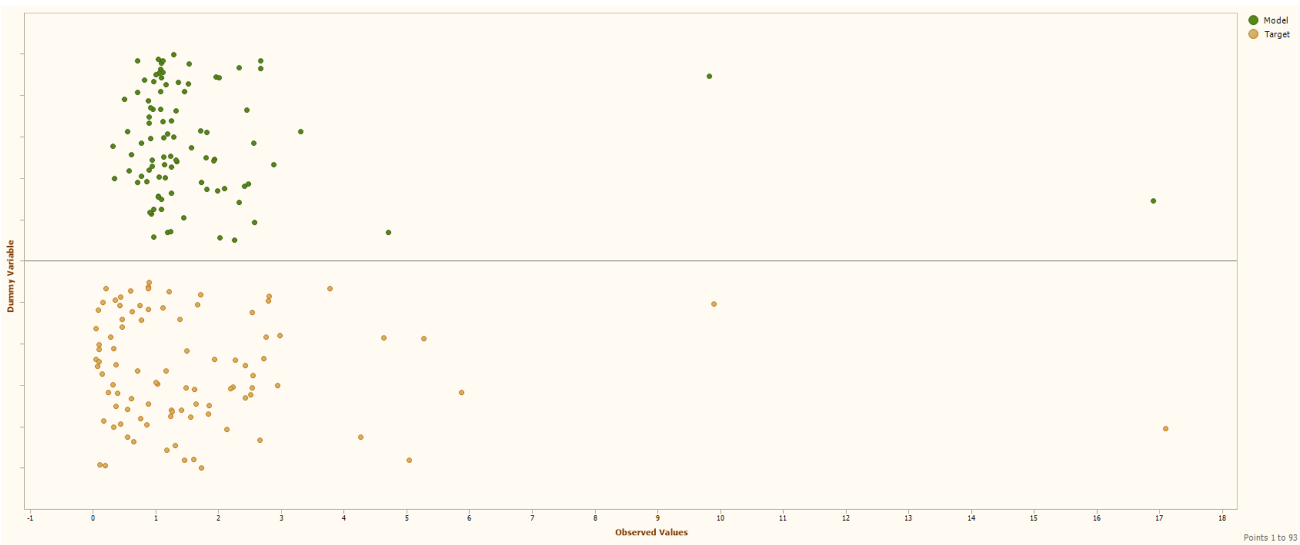
A2.1.2 Target sorted fitting



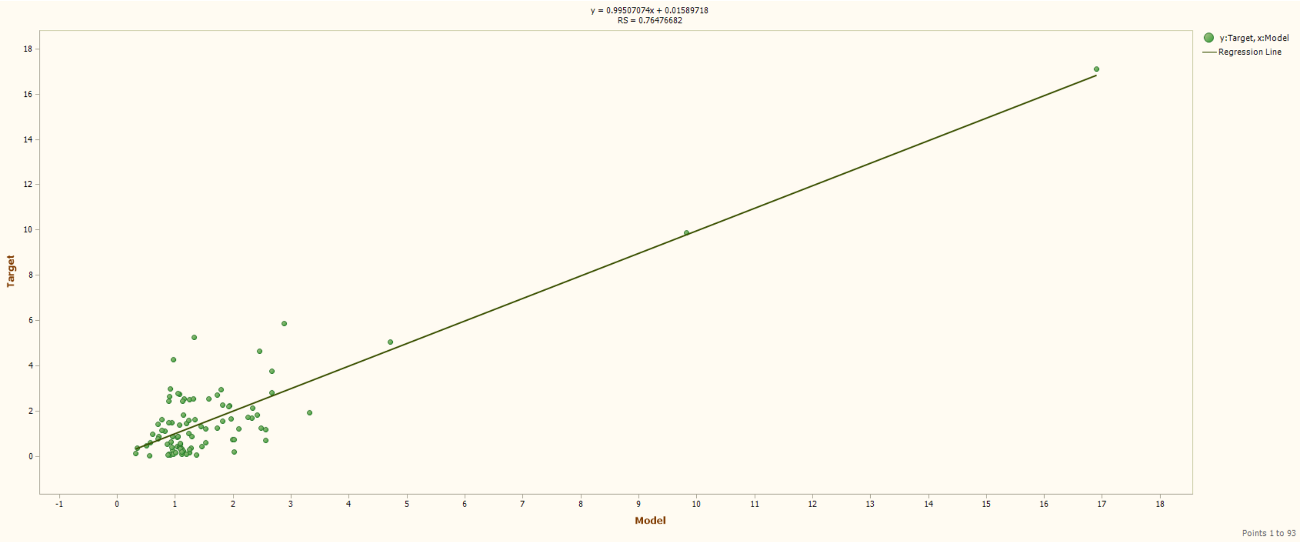
A2.1.3 Model sorted fitting



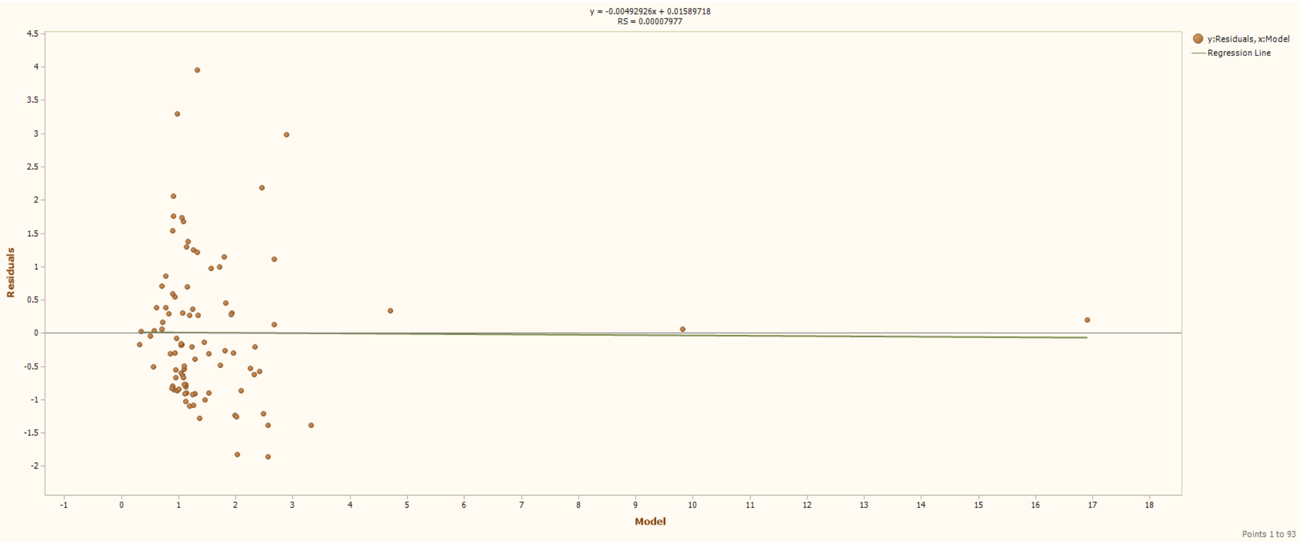
A2.1.4 Stacked distribution



A2.1.5 Scatter plot



A2.1.6 Residuals plot

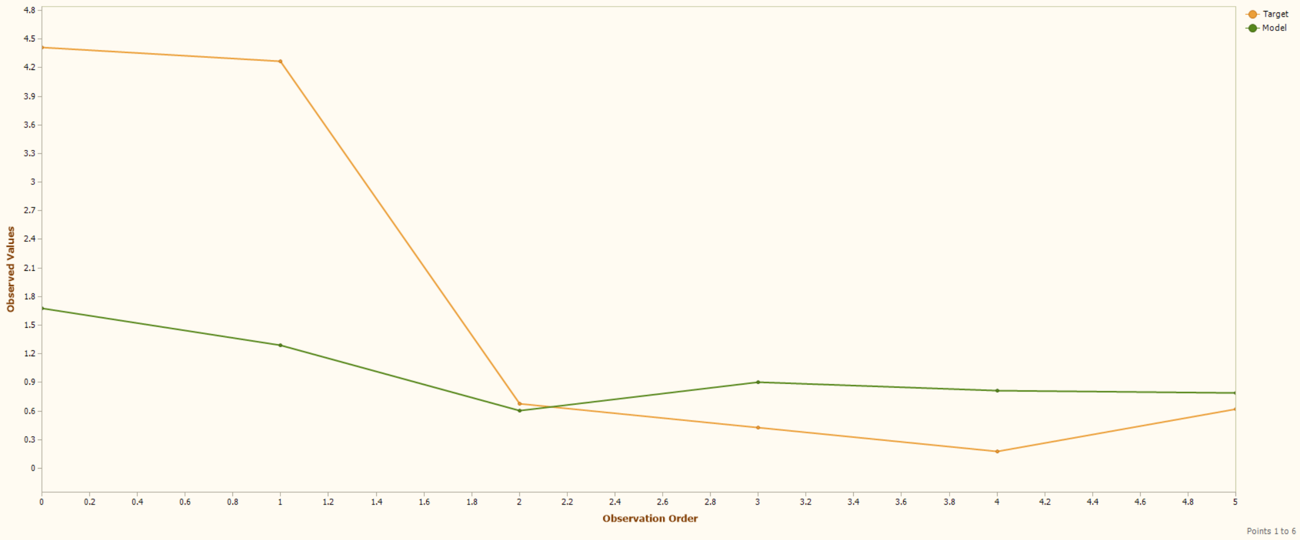


A2.1.7 Model performance

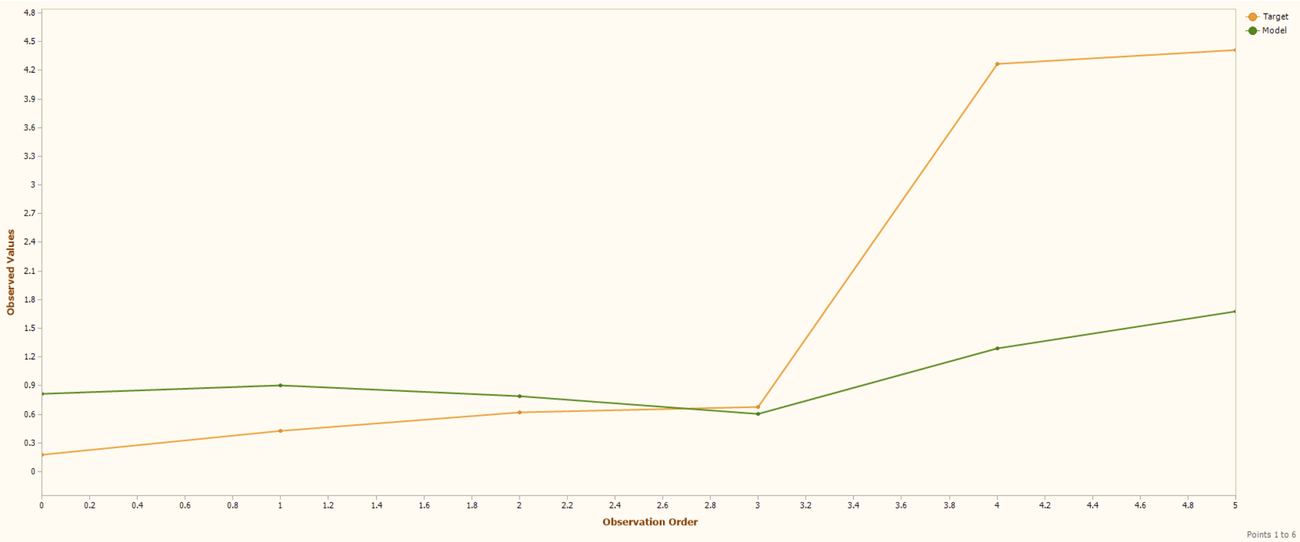
Fitness	482.635336522645
MSE	1.14909330259268
RMSE	1.07195769627009
MAE	0.816872739554996
RSE	0.235264338660963
RRSE	0.48504055362512
RAE	0.646793329952511
Correlation Coefficient	0.874509472062198
R-square	0.764766816726504
Calculation Errors	0

A2.2 Output results for Validation period

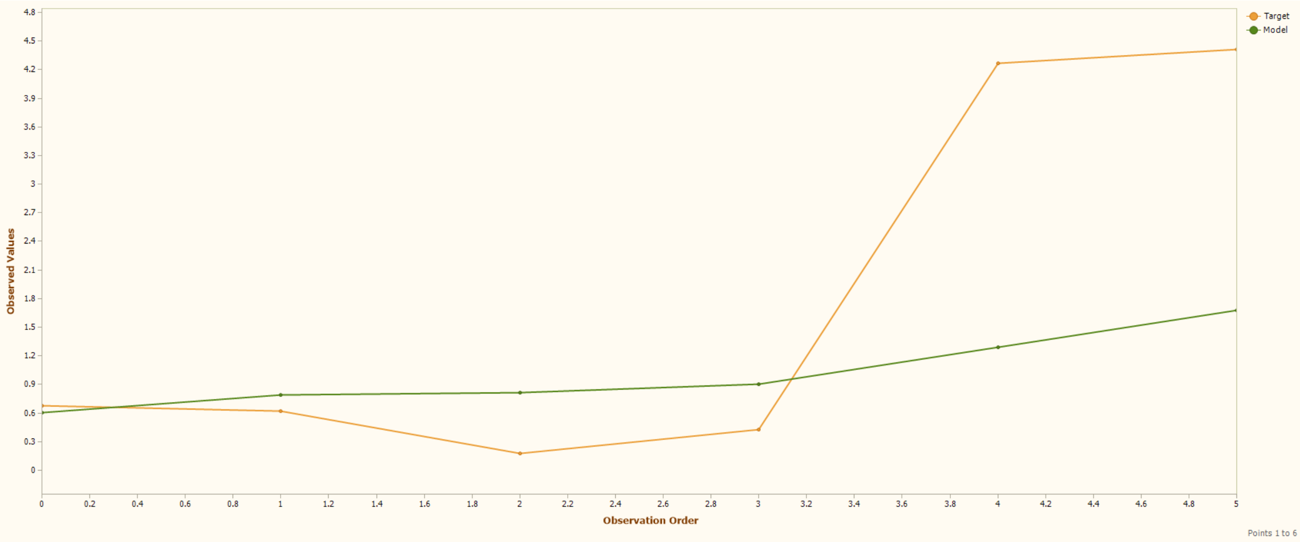
A2.2.1 Curve fitting



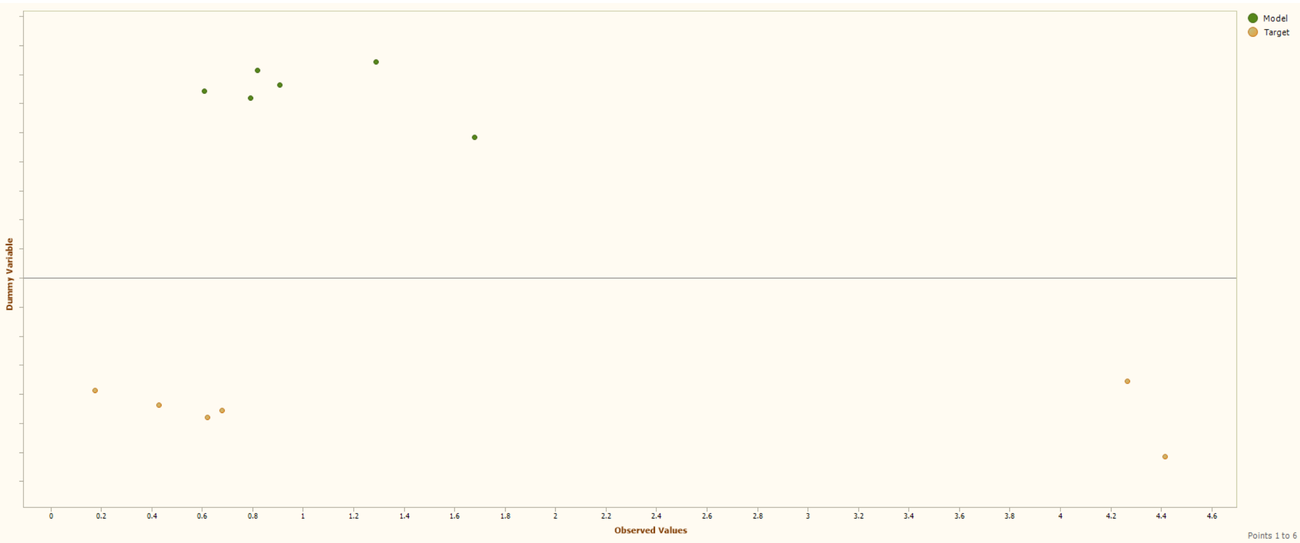
A2.2.2 Target sorted fitting



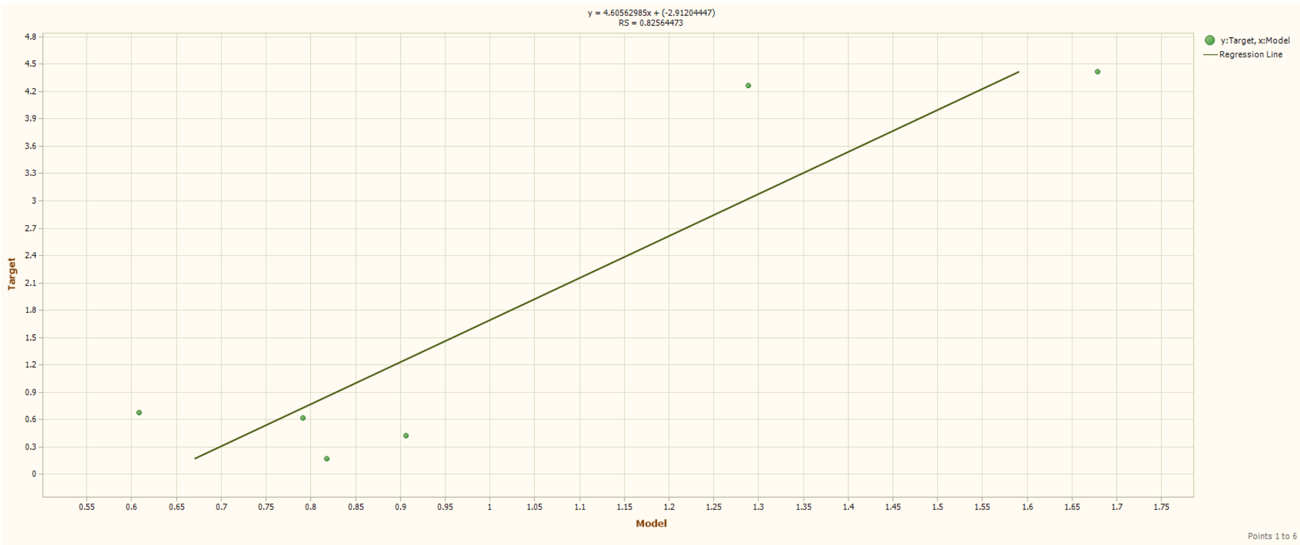
A2.2.3 Model sorted fitting



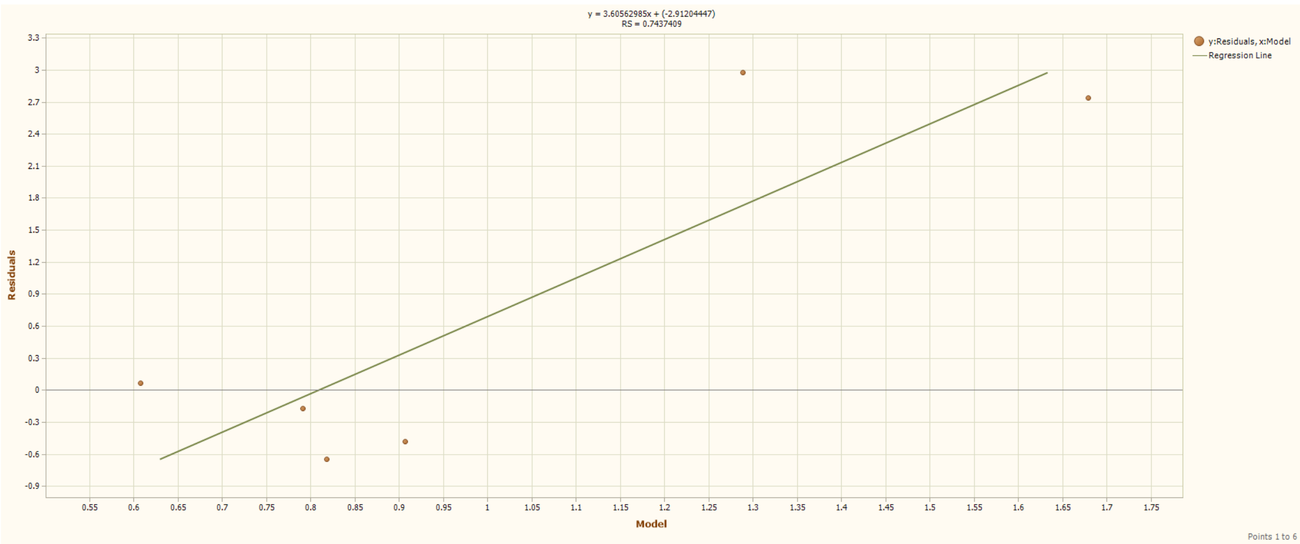
A2.2.4 Stacked distribution



A2.2.5 Scatter plot



A2.2.6 Residuals plot



A2.2.7 Model performance

Fitness 372.519939413869

MSE 2.8372709555991

RMSE 1.68442006506664

MAE 1.17910277178139

RSE 0.847677685775737

RRSE 0.920694132584615

RAE 0.686438051907575

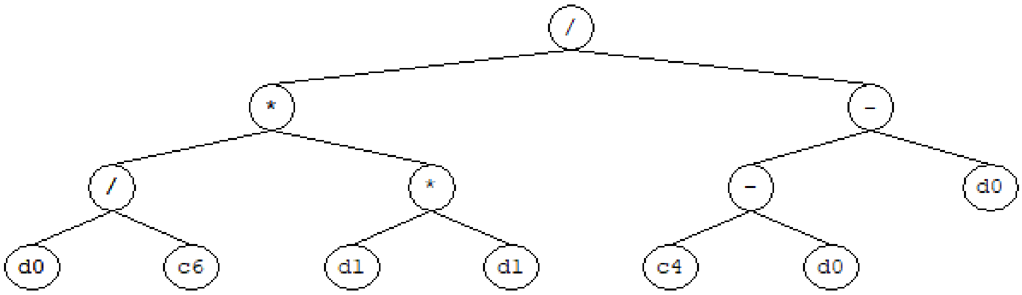
Correlation Coefficient 0.908649951813583

R-square 0.825644734930827

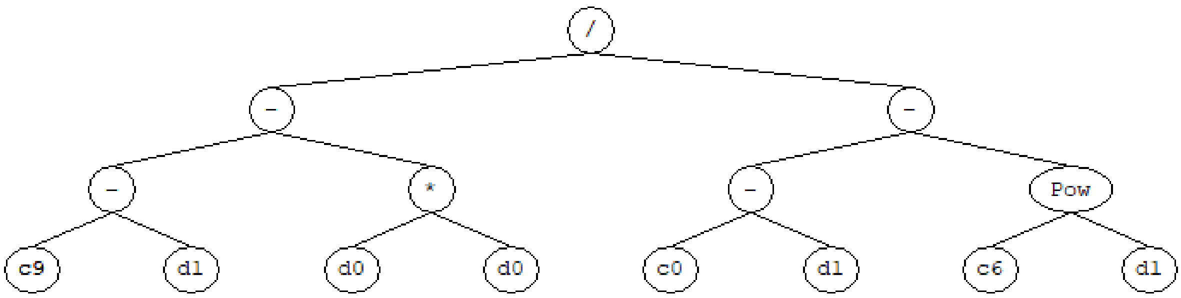
Calculation Errors 0

A2.3 Output model explained by Expression Tree

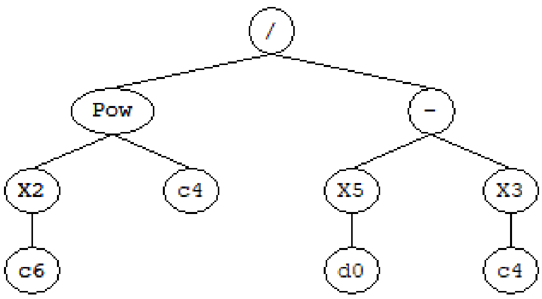
Sub-ET 1



Sub-ET 2



Sub-ET 3



A2.4 Output model explained by MATLAB

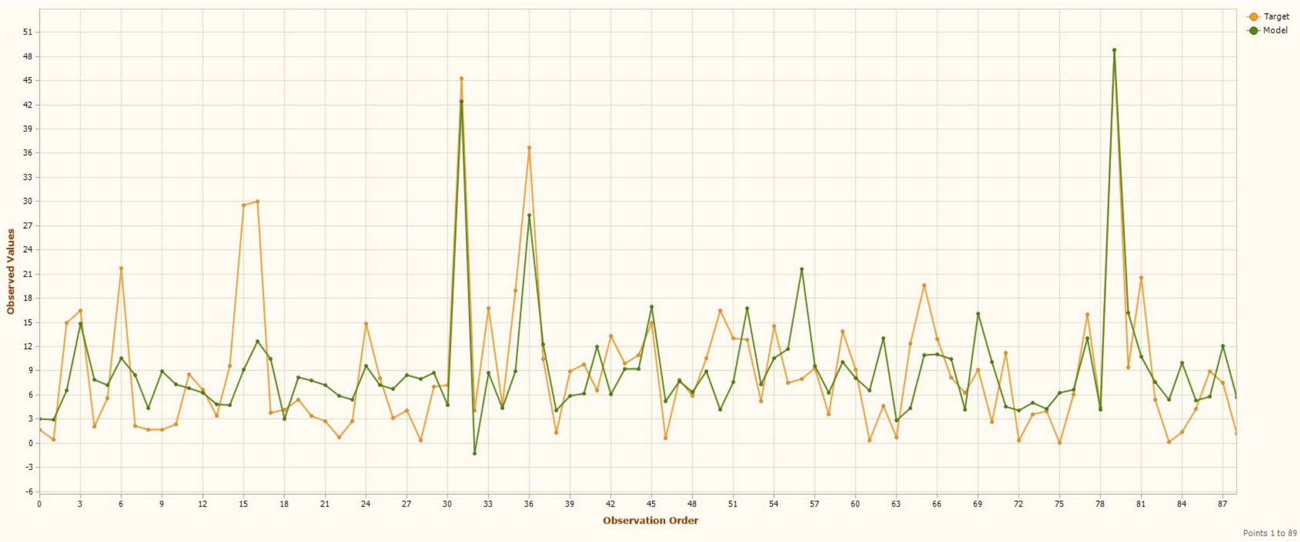
```
%-----------------------------------------------------------------------
% Regression model generated by GeneXproTools 5.0 on 24/09/2020 12:21:46 AM
% GEP File: C:\Users\resha\Google Drive\Final GeneXPro\FINAL\st2_6P_3N.gep
% Training Records:  93
% Validation Records:   6
% Fitness Function:  RMSE
% Training Fitness:  482.635336522645
% Training R-square: 0.764766816726504
% Validation Fitness:   372.519939413869
% Validation R-square:  0.825644734930827
%-----------------------------------------------------------------------

function result = gepModelQ2C(d)

G1C6 = -1.54510758084076;
G1C4 = -0.875769132040791;
G2C9 = -4.06526212069903;
G2C0 = -3.29559717102429;
G2C6 = 13.0293357653121;
G3C4 = -0.954386689813802;
G3C6 = -4.17687380978386;

Six_PC = 1;
Three_NC = 2;

y = 0.0;

y = (((d(Six_PC)/G1C6)*(d(Three_NC)*d(Three_NC)))/((G1C4-d(Six_PC))-d(Six_PC)));
y = y + (((G2C9-d(Three_NC))-(d(Six_PC)*d(Six_PC)))/((G2C0-d(Three_NC))-realpow(G2C6,d(Three_NC))));
```

```
y = y + (realpow((G3C6^2),G3C4)/((d(Six_PC)^5)-(G3C4^3)));

result = y;
```

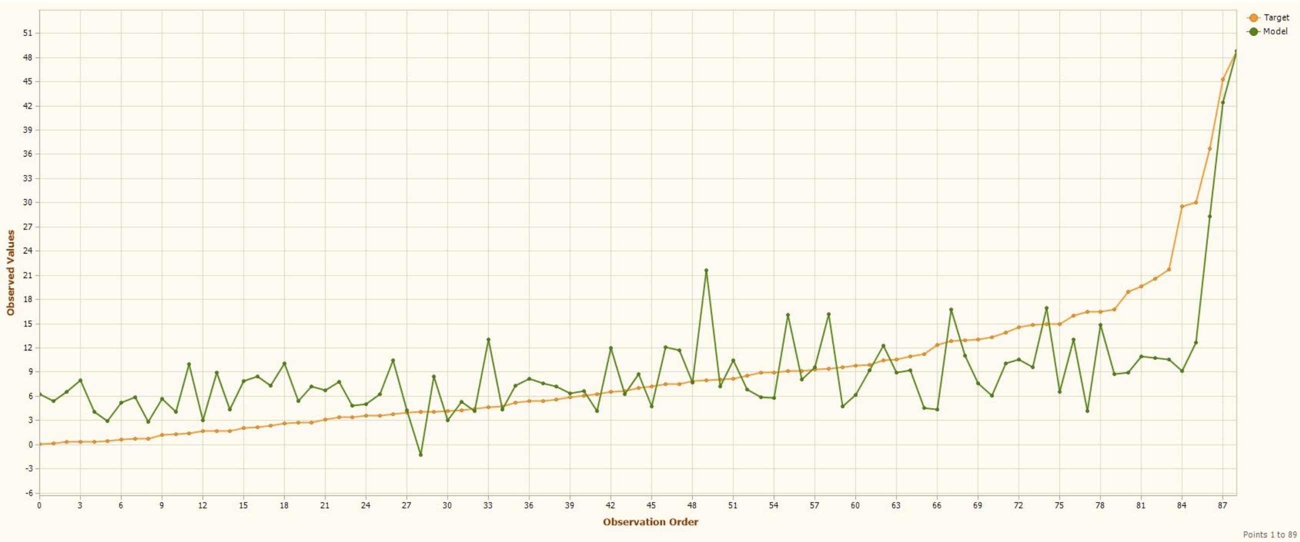
A3 Best GEP Model Output for North Cuerindi Station (NNSW)

A3.1 Output results for Calibration period

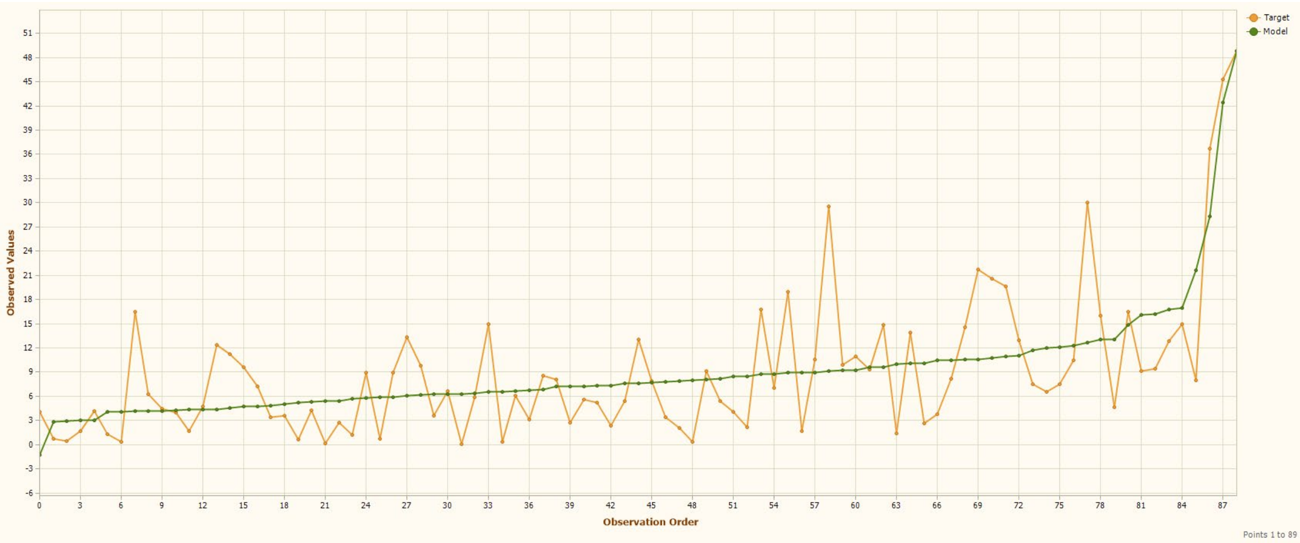
A3.1.1 Curve Fitting



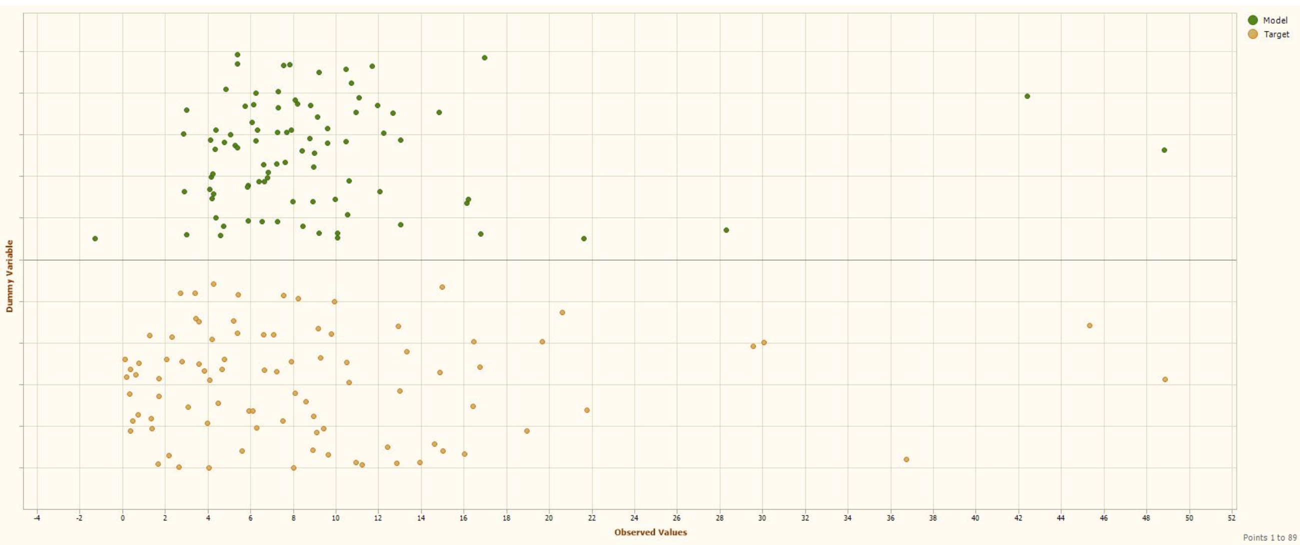
A3.1.2 Target sorted fitting



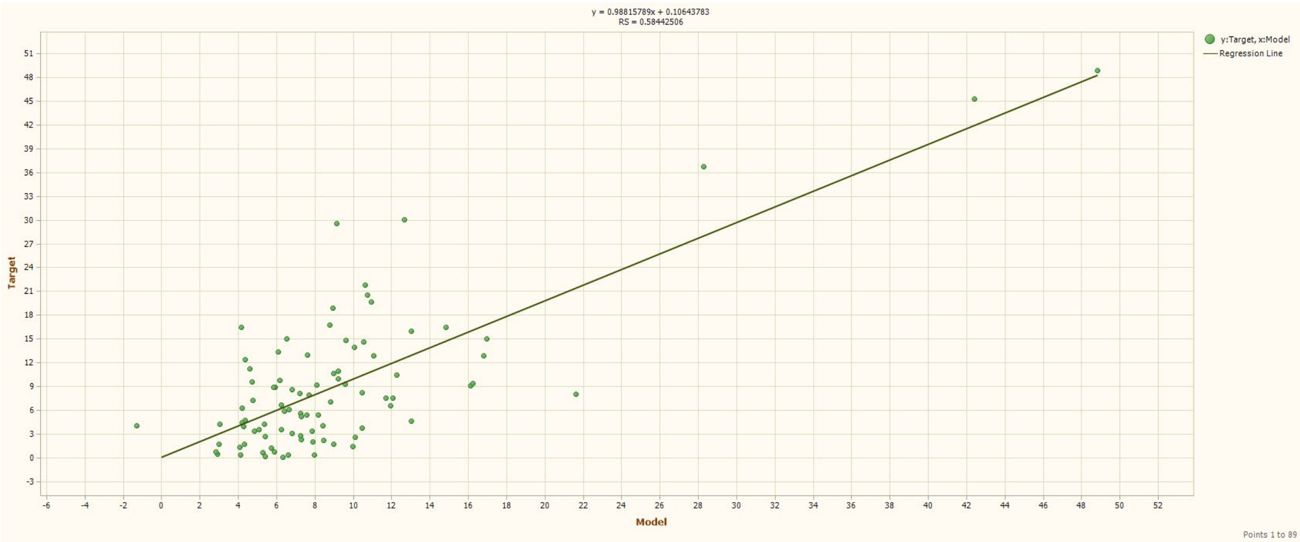
A3.1.3 Model sorted fitting



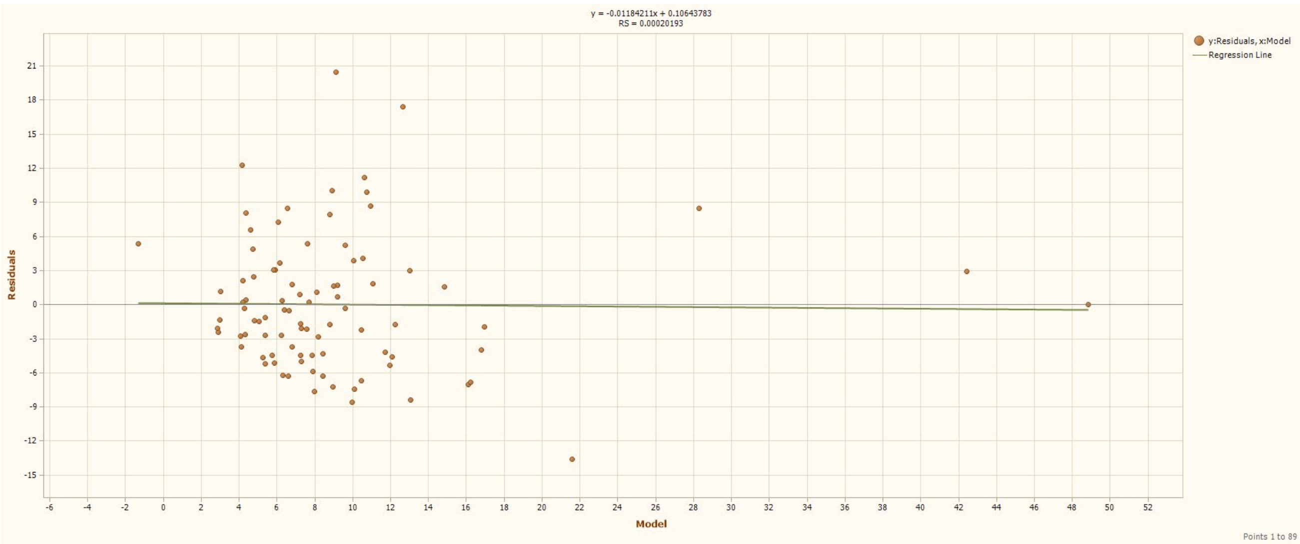
A3.1.4 Stacked distribution



A3.1.5 Scatter plot



A3.1.6 Residuals plot



A3.1.7 Model performance

Fitness 146.64905979453

MSE 33.8607658616522

RMSE 5.81900041773948

MAE 4.48574348227719

RSE 0.415658893841058

RRSE 0.644716134311108

RAE 0.724014609908479

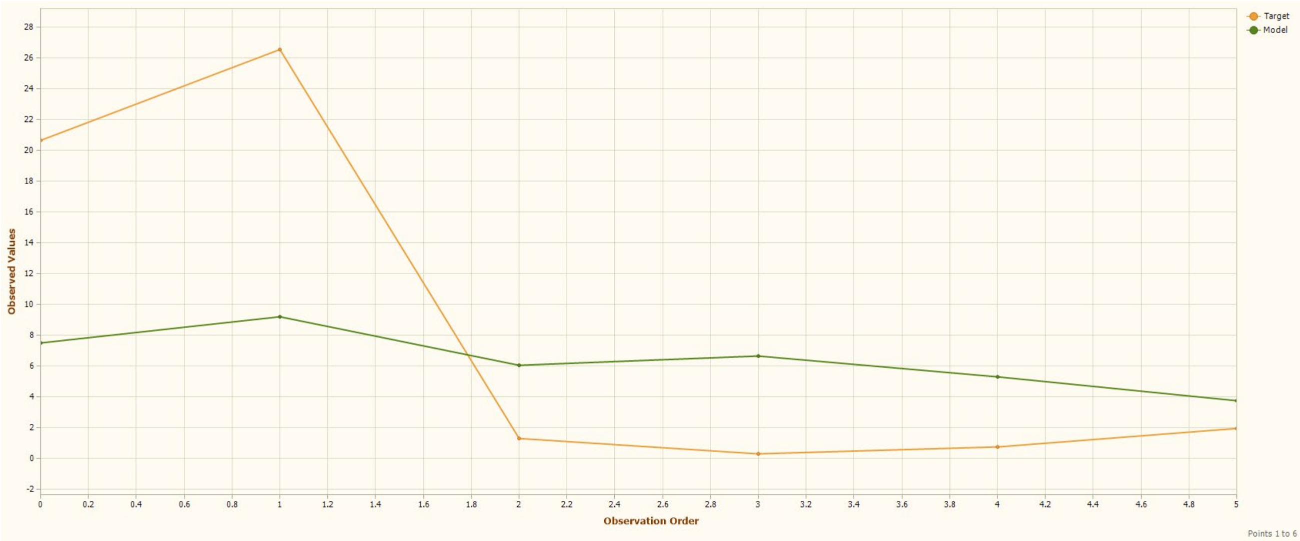
Correlation Coefficient 0.764476982722173

R-square 0.584425057111998

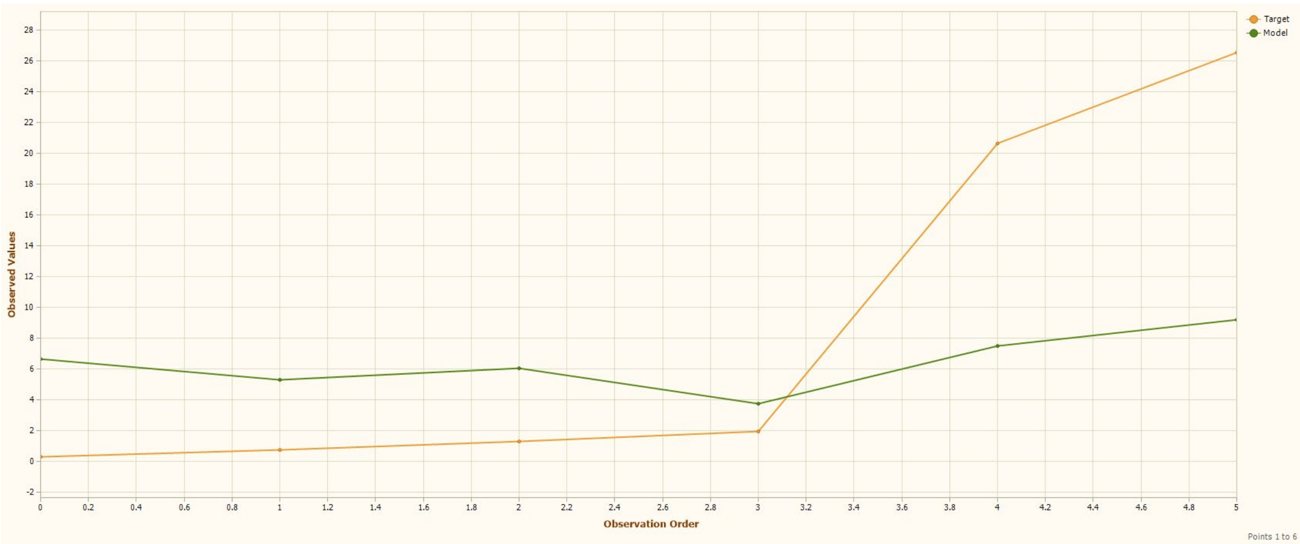
Calculation Errors 0

A3.2 Output results for Validation period

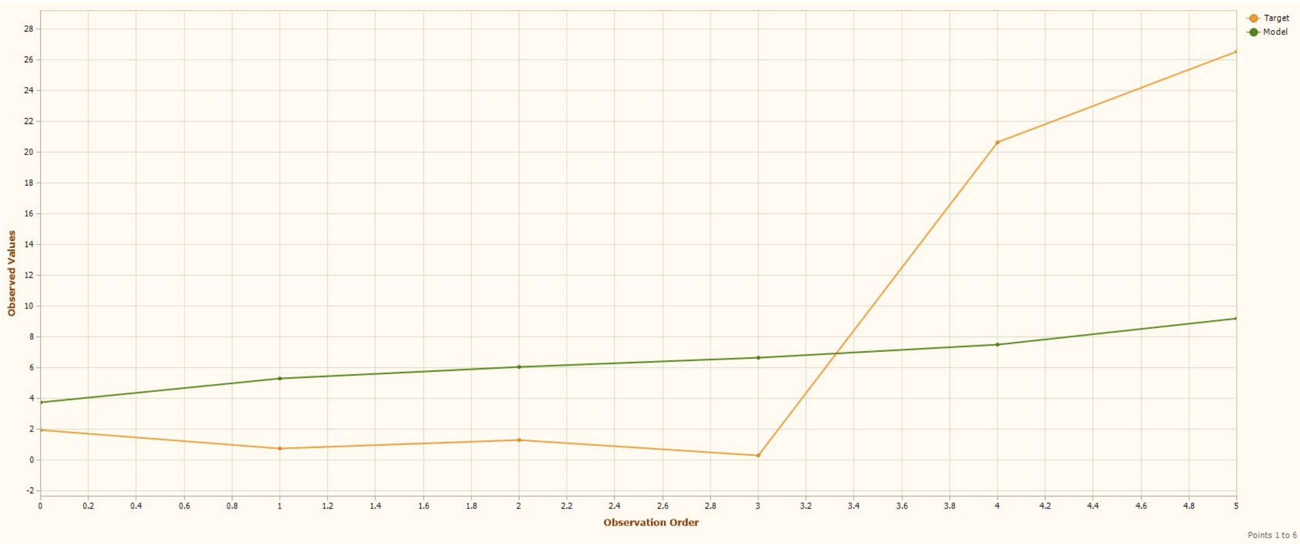
A3.2.1 Curve fitting



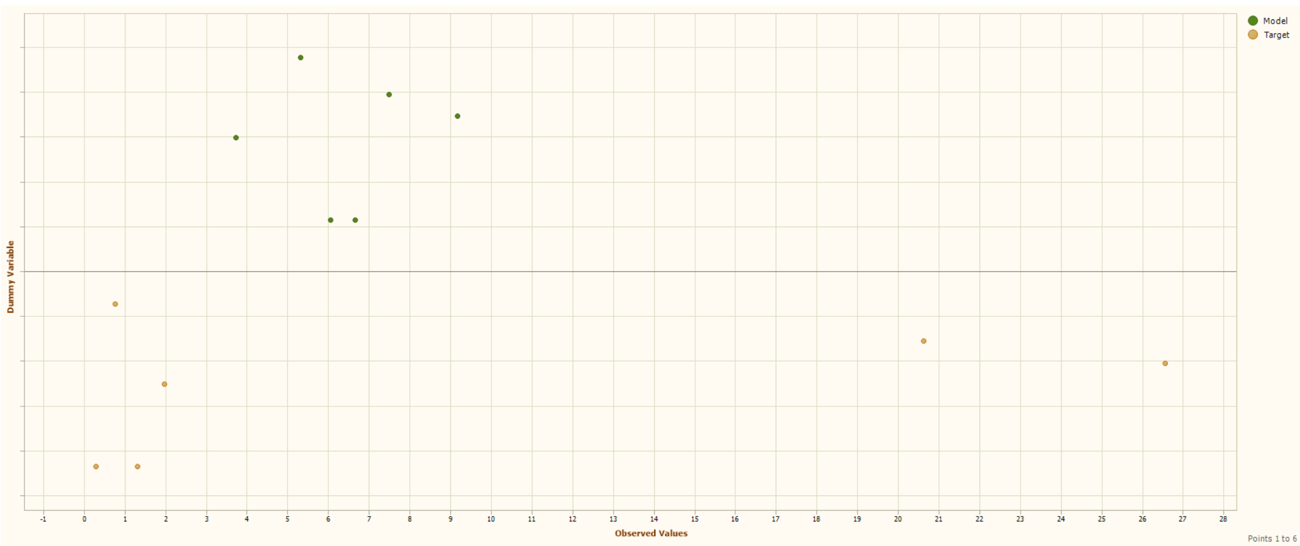
A3.2.2 Target sorted fitting



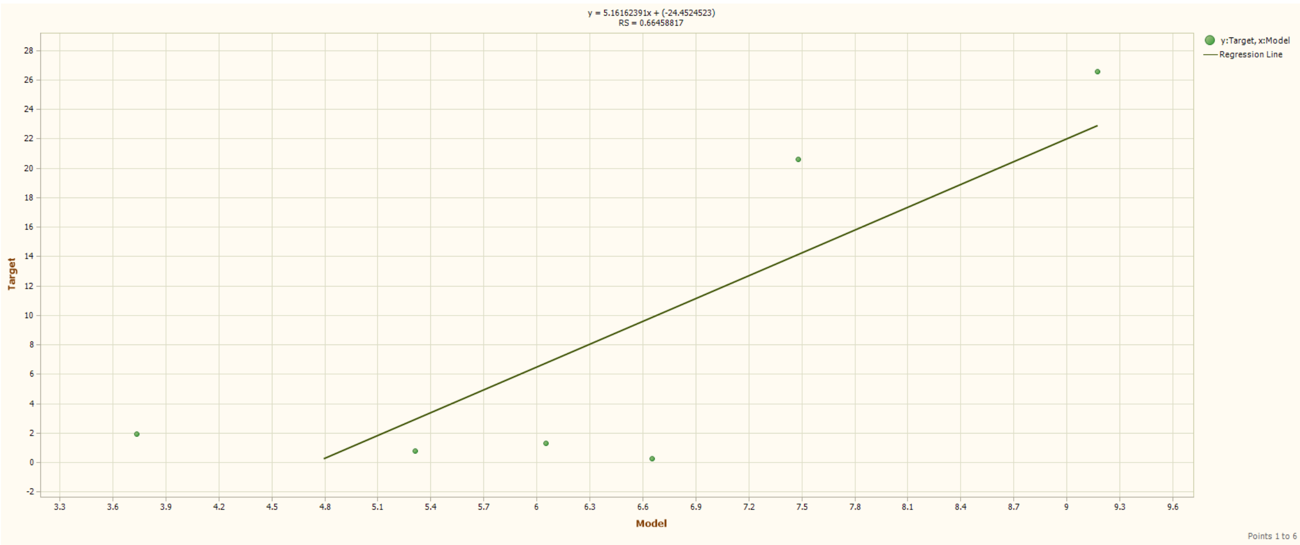
A3.2.3 Model sorted fitting



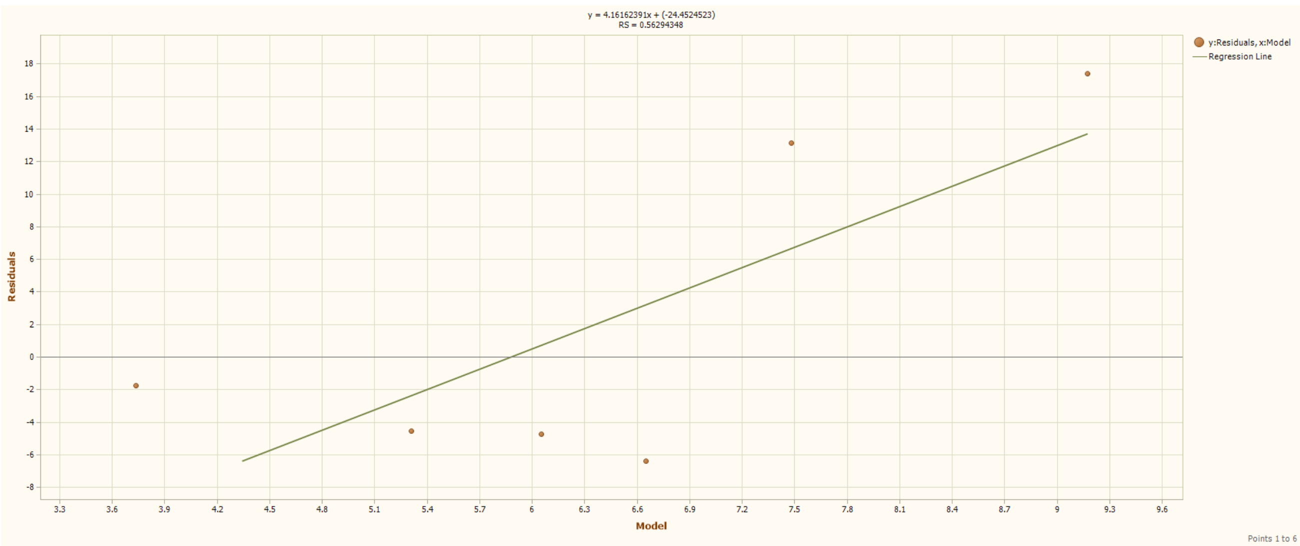
A3.2.4 Stacked distribution



A3.2.5 Scatter plot



A3.2.6 Residuals plot



A3.2.7 Model performance

Fitness 93.6447217390319

MSE 93.6764295785372

RMSE 9.67865845964911

MAE 7.99365719948064

RSE 0.808642855997251

RRSE 0.899245715028574

RAE 0.7988395002443

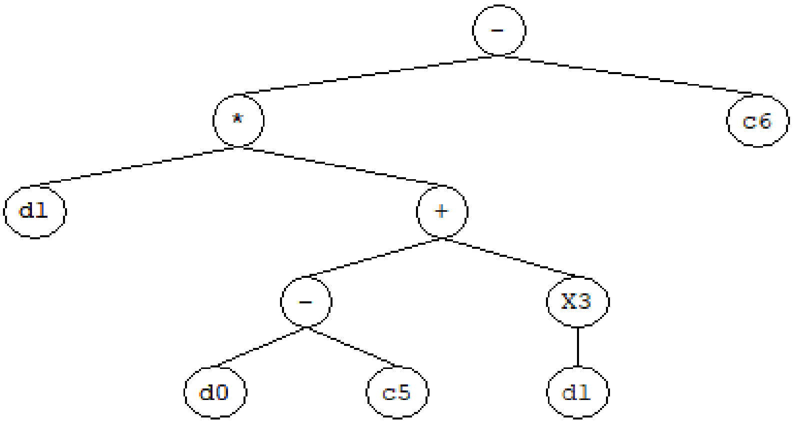
Correlation Coefficient 0.815222775750089

R-square 0.664588174101679

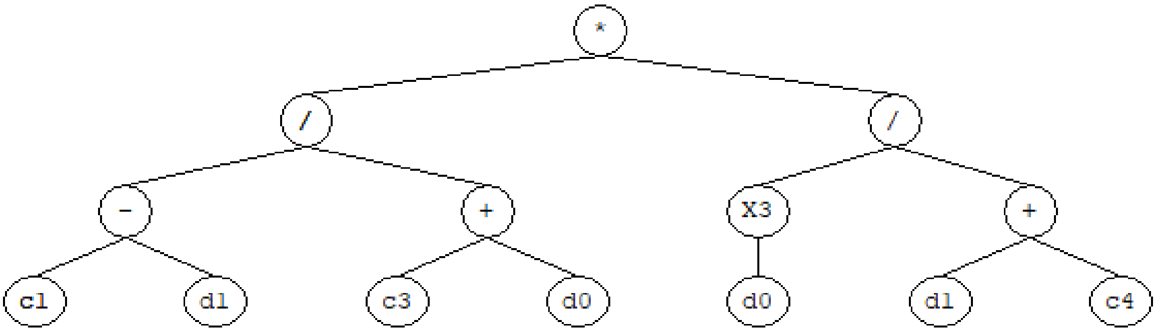
Calculation Errors 0

A3.3 Output model explained by Expression Tree

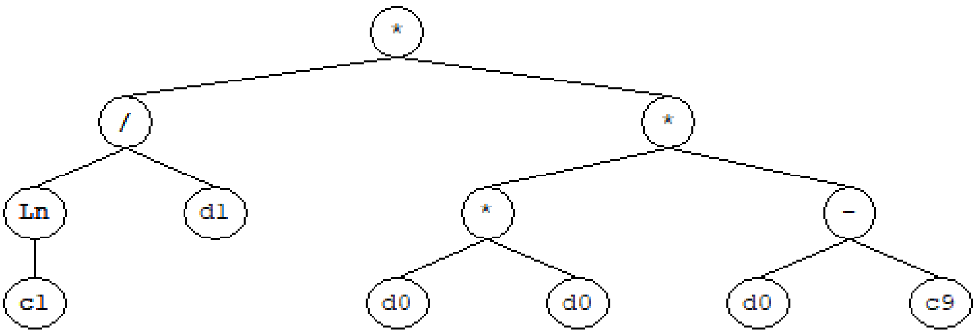
Sub-ET 1



Sub-ET 2



Sub-ET 3



A3.4 Output model explained by Matab

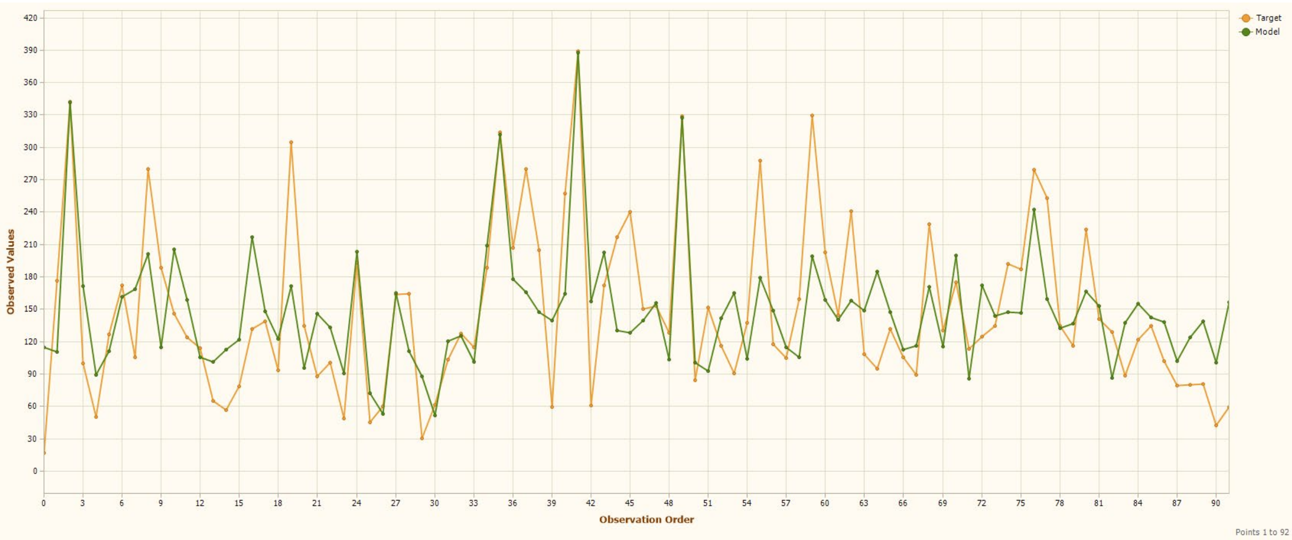
```
%------------------------------------------------------------------------

% Regression model generated by GeneXproTools 5.0 on 24/09/2020 12:48:20 AM

% GEP File: C:\Users\resha\Google Drive\Final GeneXPro\FINAL\st11_4P_4N.gep

% Training Records:  89

% Validation Records:   6

% Fitness Function:  RMSE

% Training Fitness:  146.64905979453

% Training R-square: 0.584425057111998

% Validation Fitness:   93.6447217390319

% Validation R-square:  0.664588174101679

%------------------------------------------------------------------------


function result = gepModelQ11C(d)


G1C6 = -6.46088332570378;

G1C5 = 4.74514669421048;

G2C1 = 0.907925656910916;

G2C3 = -0.442555087360802;

G2C4 = 1.59096311813514;

G3C1 = 0.934738896839175;

G3C9 = 0.903937118263785;


Four_PC = 1;

Four_NC = 2;


y = 0.0;


y = ((d(Four_NC)*((d(Four_PC)-G1C5)+(d(Four_NC)^3)))-G1C6);
```

```
y = y + (((G2C1-d(Four_NC))/(G2C3+d(Four_PC)))*((d(Four_PC)^3)/(d(Four_NC)+G2C4)));
y = y + ((reallog(G3C1)/d(Four_NC))*((d(Four_PC)*d(Four_PC))*(d(Four_PC)-G3C9)));

result = y;
```

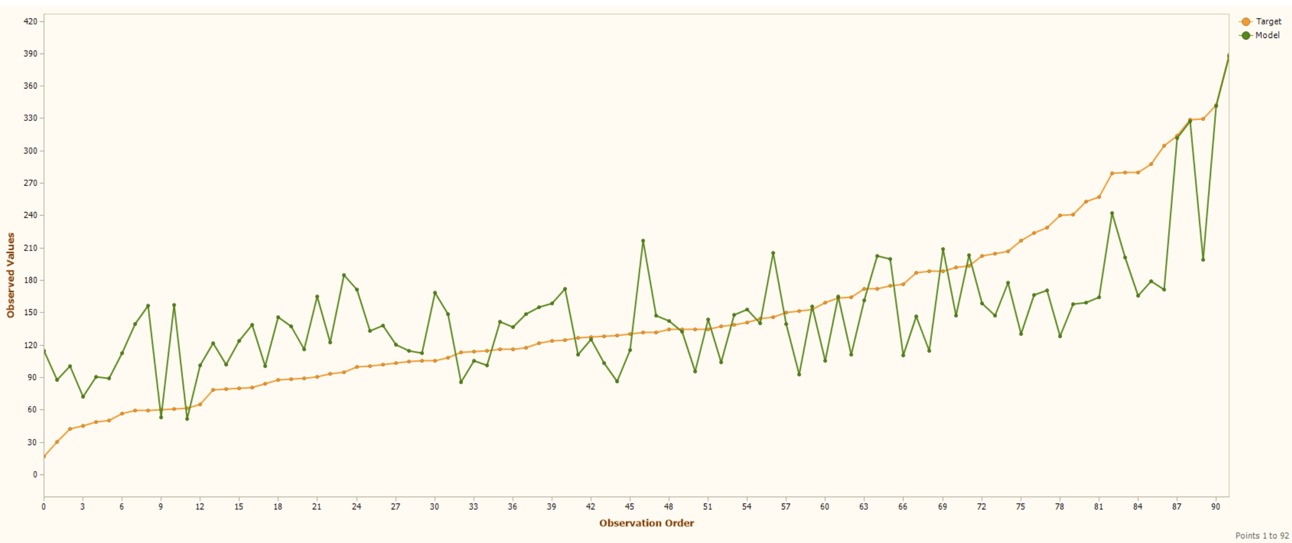
A4 Best GEP Model Output for Gundagai Station (SNSW)

A4.1 Output results for Calibration period

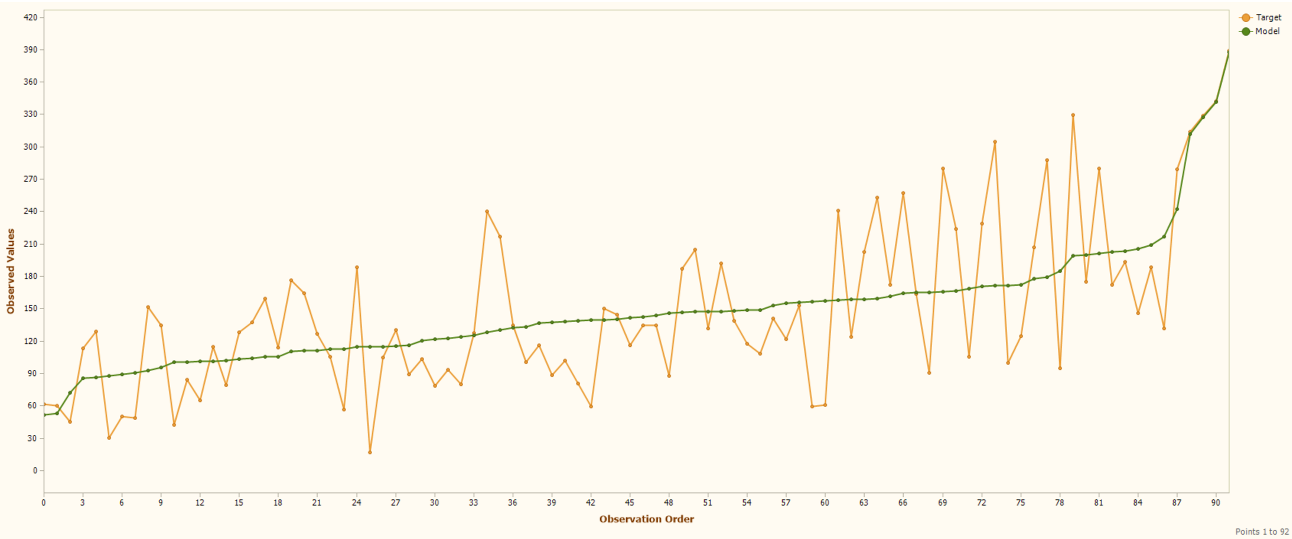
A4.1.1 Curve Fitting



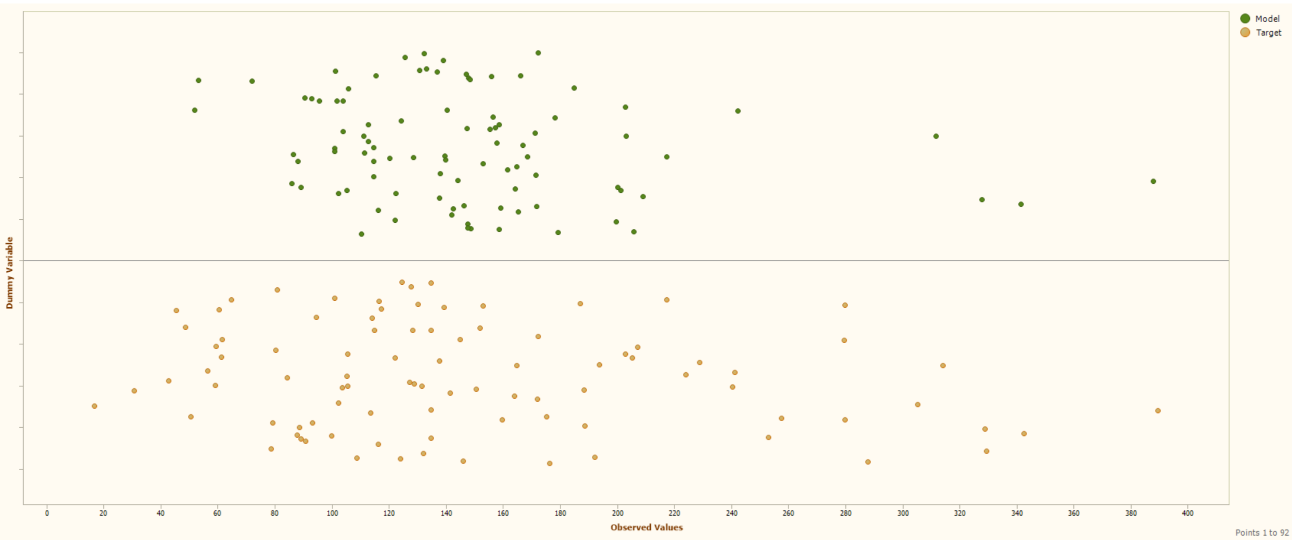
A4.1.2 Target sorted fitting



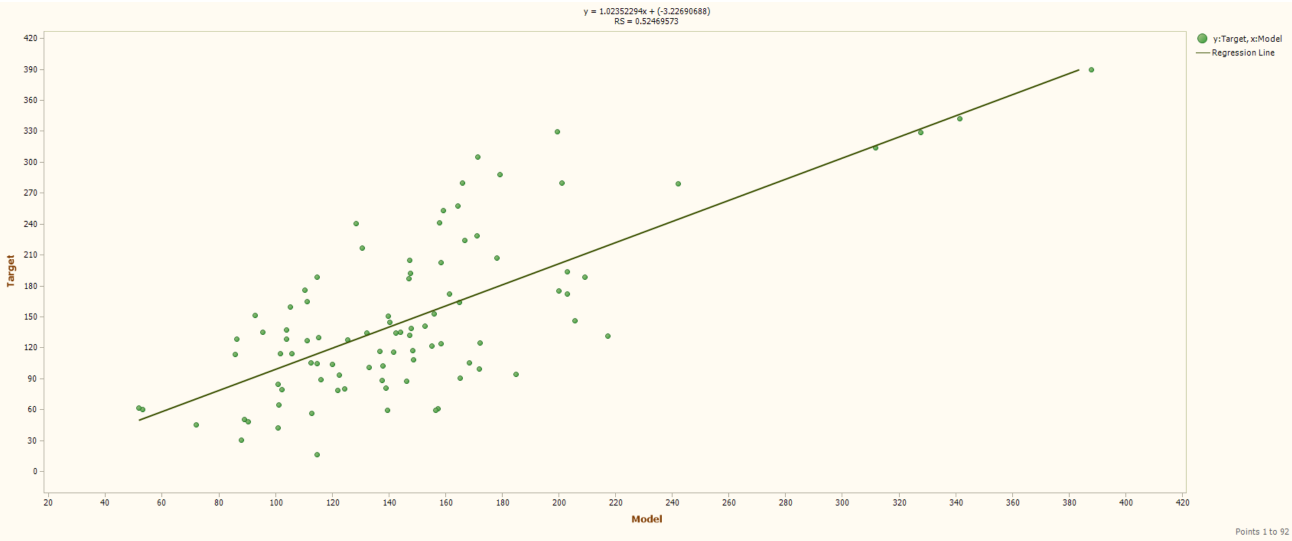
A4.1.3 Model sorted fitting



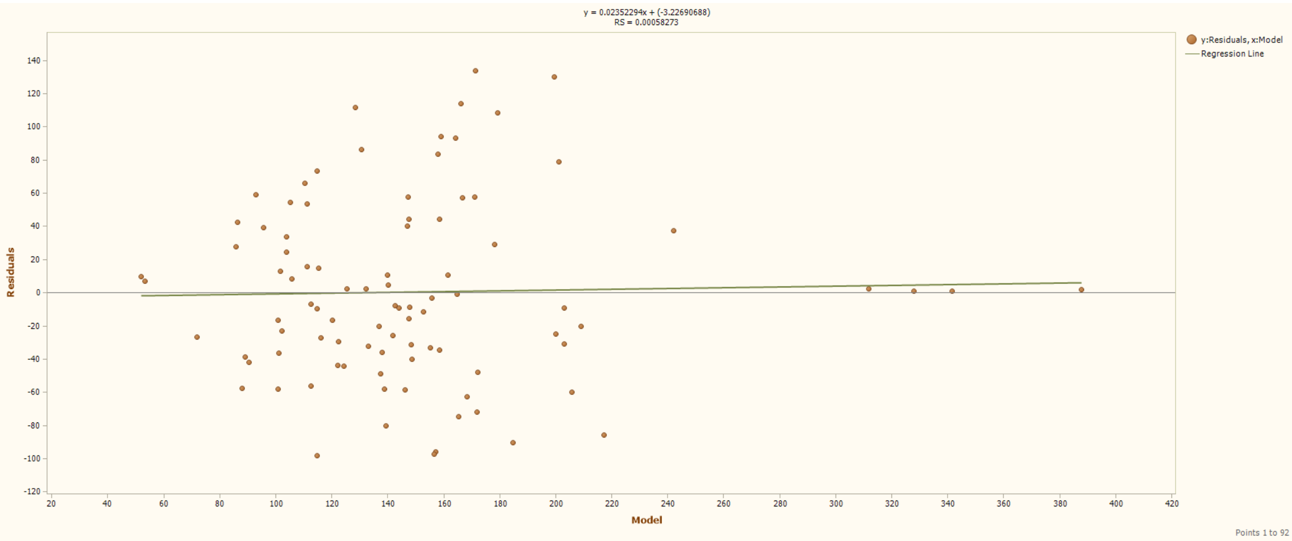
A4.1.4 Stacked distribution



A4.1.5 Scatter plot



A4.1.6 Residuals plot

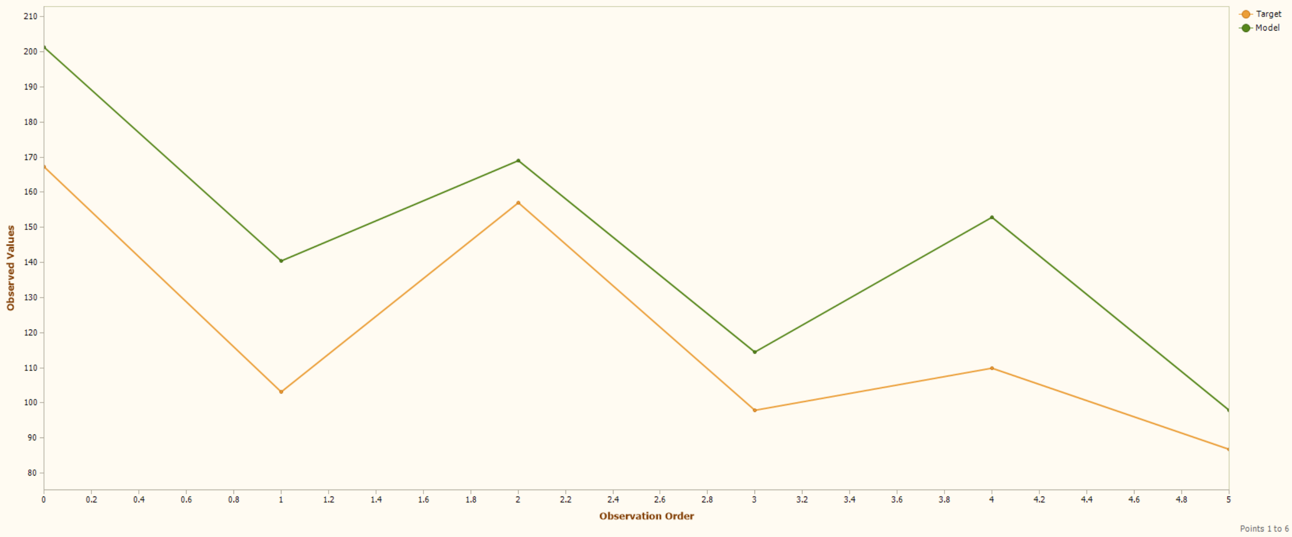


A4.1.7 Model performance

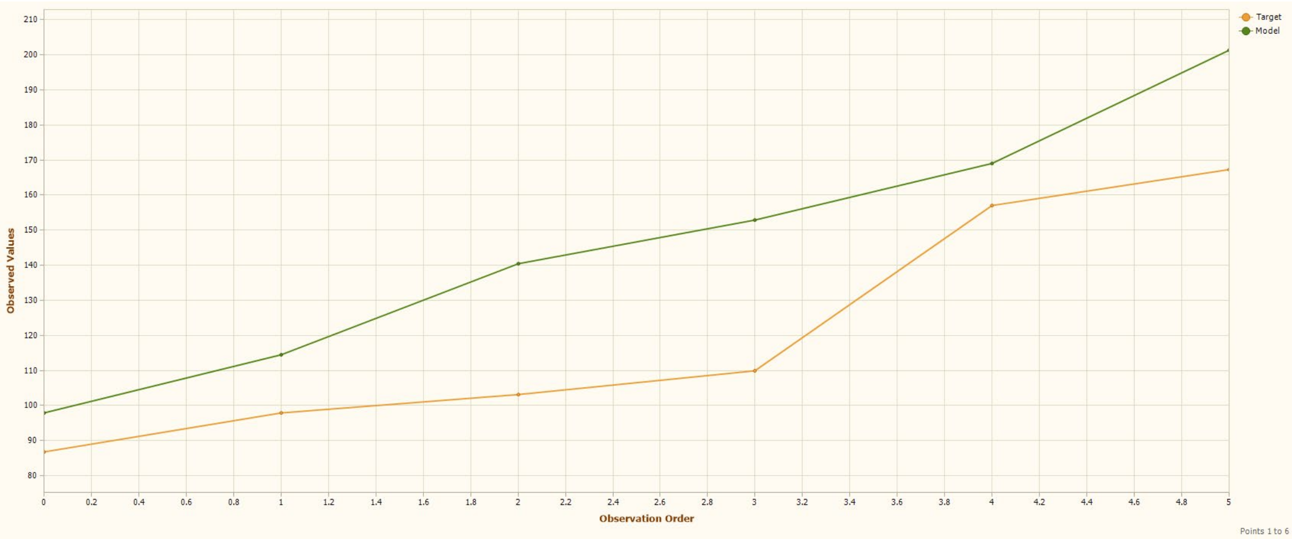
Fitness	18.2402230232923
MSE	2897.01124919828
RMSE	53.8238910633399
MAE	42.8029875773736
RSE	0.475592330687834
RRSE	0.689632025567138
RAE	0.701349310432205
Correlation Coefficient	0.724358842514304
R-square	0.524695732728662
Calculation Errors	0

A4.2 Output results for Validation period

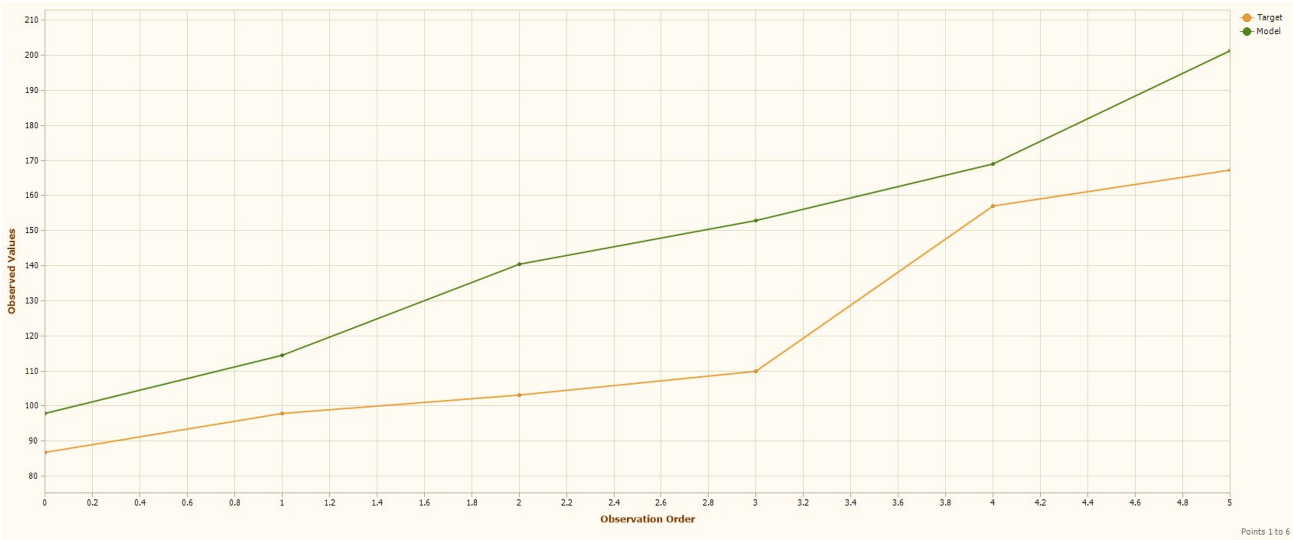
A4.2.1 Curve fitting



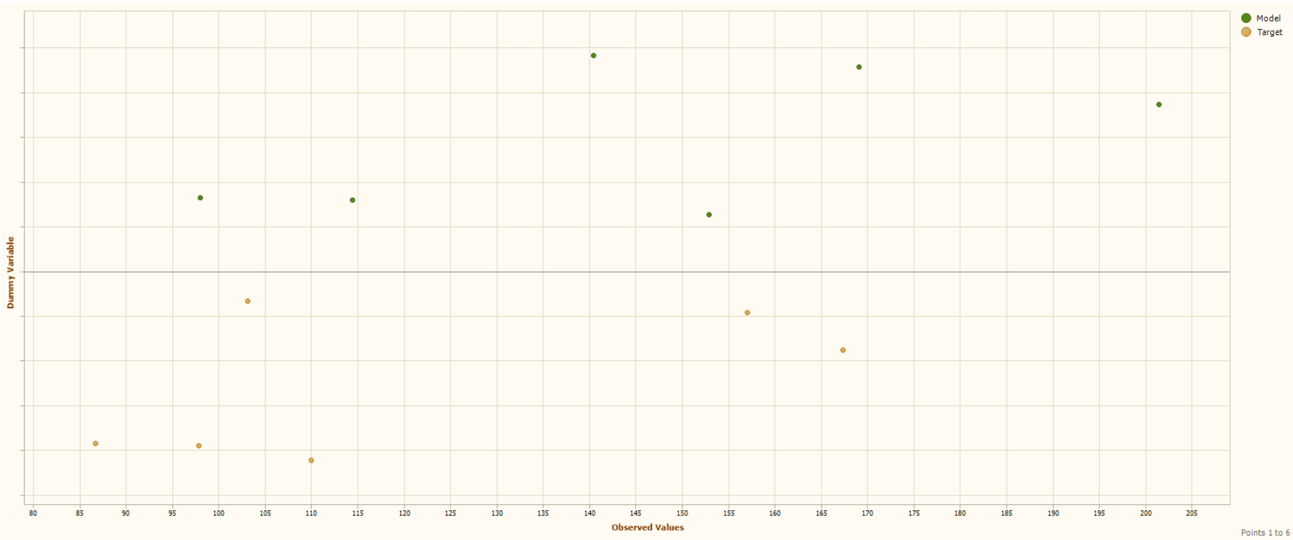
A4.2.2 Target sorted fitting



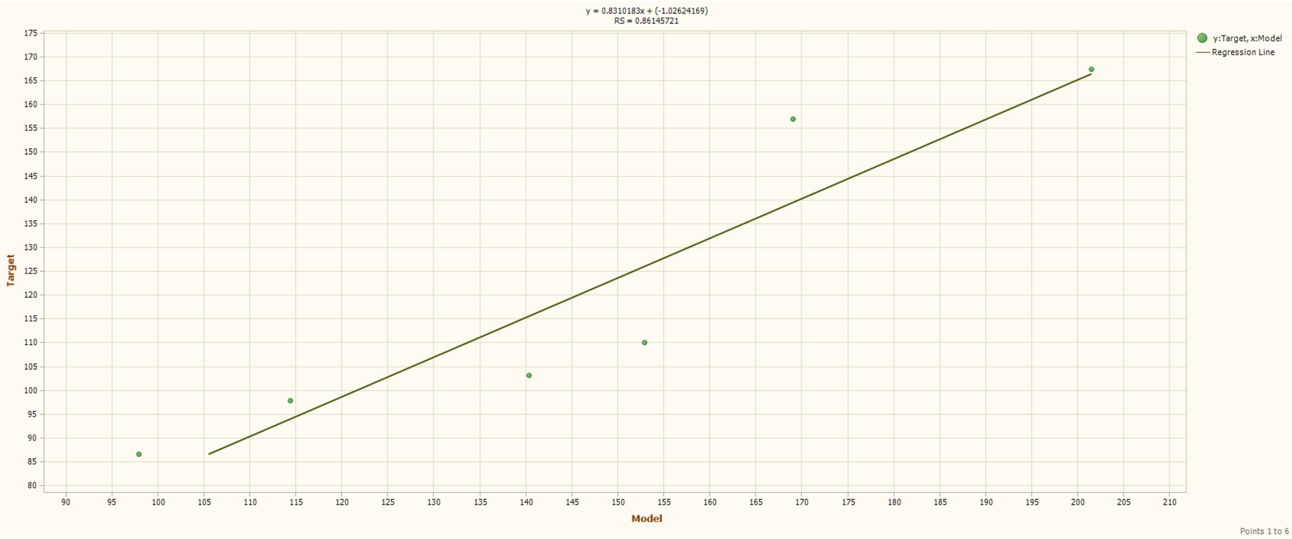
A4.2.3 Model sorted fitting



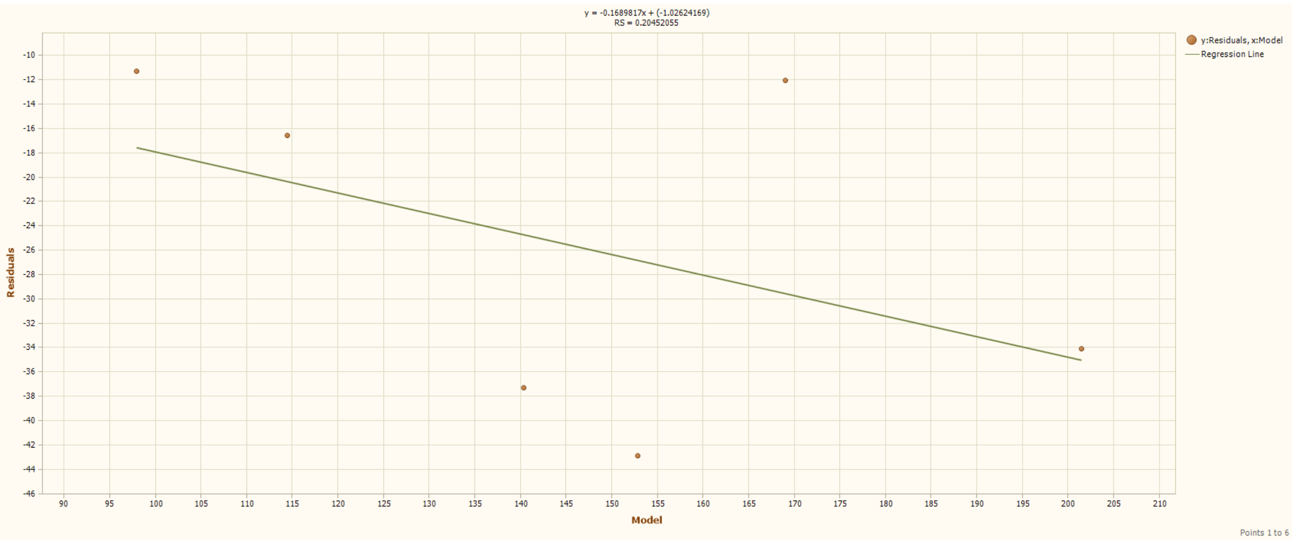
A4.2.4 Stacked distribution



A4.2.5 Scatter plot



A4.2.6 Residuals plot

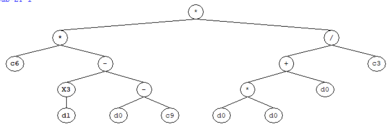


A4.2.7 Model performance

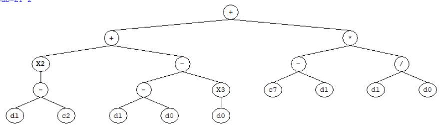
Fitness	33.6795201821979
MSE	823.209884513885
RMSE	28.691634399488
MAE	25.7019897351135
RSE	0.881653548098189
RRSE	0.938964082432437
RAE	0.920896421434122
Correlation Coefficient	0.928147193711262
R-square	0.861457213194092
Calculation Errors	0

A4.3 Output model explained by Expression Tree

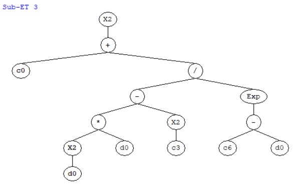
Sub-ET 1



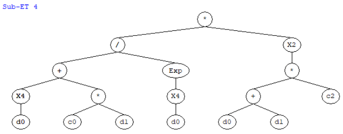
Sub-ET 2



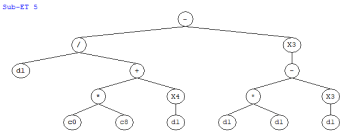
Sub-ET 3



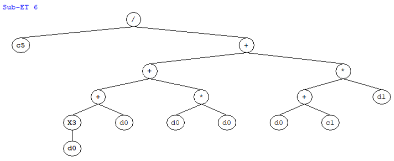
Sub-ET 4



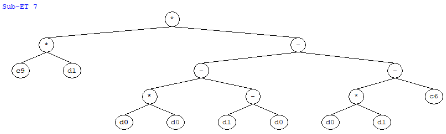
Sub-ET 5



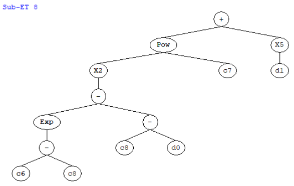
Sub-ET 6



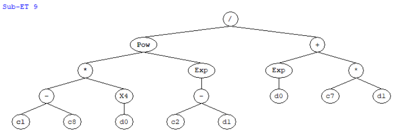
Sub-ET 7



Sub-ET 8



Sub-ET 9



A4.4 Output model explained by Matab

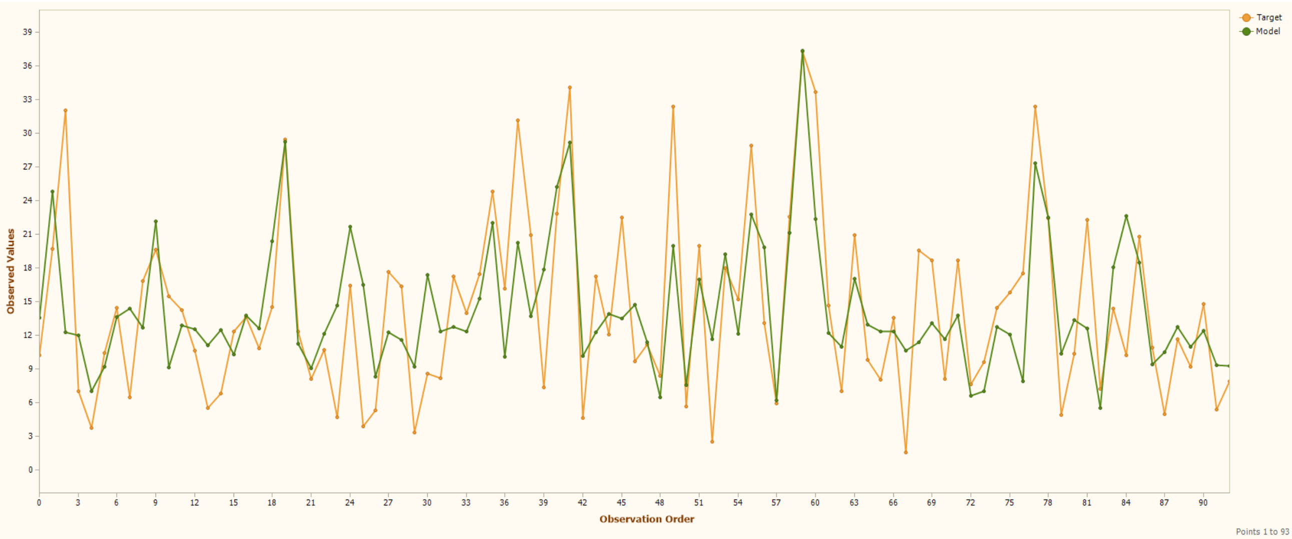
```
%------------------------------------------------------------------------
% Regression model generated by GeneXproTools 5.0 on 24/09/2020 1:30:11 AM
% GEP File: C:\Users\resha\Google Drive\Final GeneXPro\FINAL\st06_6P_6N.gep
% Training Records:  92
% Validation Records:   6
% Fitness Function:  RMSE
% Training Fitness:  18.2402230232923
% Training R-square: 0.524695732728662
% Validation Fitness:   33.6795201821979
% Validation R-square:  0.861457213194092
%------------------------------------------------------------------------

function result = gepModelQ6C(d)

G1C6 = -6.43852994018372;
G1C3 = -4.86755451027764;
G1C9 = 6.64409480835102;
G2C7 = -3.62437751945682;
G2C2 = 8.90140544873327;
G3C0 = 6.36768063447981;
G3C3 = -4.09708795588808;
G3C6 = 5.52622580339976;
G4C2 = 7.50121639519798;
G4C0 = 0.348274972022856;
G5C0 = -7.76722585678972e-04;
G5C8 = -5.23215556729666;
G6C5 = 6.11360511564837;
G6C1 = -8.94424402133345;
```

```
G7C9 = -11.2215366496599;

G7C6 = 2.30066567096449;

G8C7 = -3.0777589973697;

G8C8 = -2.19896653927834;

G8C6 = -3.30889278847621;

G9C7 = 0.542709547006177;

G9C1 = 6.53924680318613;

G9C8 = 6.11506263102109;

G9C2 = -1.04723253973531;


Six_PC = 1;

Six_NC = 2;


y = 0.0;


y = ((G1C6*((d(Six_NC)^3)-(d(Six_PC)-G1C9)))*(((d(Six_PC)*d(Six_PC))+d(Six_PC))/G1C3));

y = y + ((((d(Six_NC)-G2C2)^2)+((d(Six_NC)-d(Six_PC))-(d(Six_PC)^3)))+((G2C7-
d(Six_NC))*(d(Six_NC)/d(Six_PC))));

y = y + ((G3C0+((((d(Six_PC)^2)*d(Six_PC))-(G3C3^2))/exp((G3C6-d(Six_PC)))))^2);

y = y + ((((d(Six_PC)^4)+(G4C0*d(Six_NC)))/exp((d(Six_PC)^4)))*(((d(Six_PC)+d(Six_NC))*G4C2)^2));

y = y + ((d(Six_NC)/((G5C0*G5C8)+(d(Six_NC)^4)))-(((d(Six_NC)*d(Six_NC))-(d(Six_NC)^3))^3));

y = y + (G6C5/((((d(Six_PC)^3)+d(Six_PC))+(d(Six_PC)*d(Six_PC)))+((d(Six_PC)+G6C1)*d(Six_NC))));

y = y + ((G7C9*d(Six_NC))*(((d(Six_PC)*d(Six_PC))-(d(Six_NC)-d(Six_PC)))-((d(Six_PC)*d(Six_NC))-G7C6)));

y = y + (realpow(((exp((G8C6-G8C8))-(G8C8-d(Six_PC)))^2),G8C7)+(d(Six_NC)^5));

y = y + (realpow(((G9C1-G9C8)*(d(Six_PC)^4)),exp((G9C2-d(Six_NC))))/(exp(d(Six_PC))+(G9C7*d(Six_NC))));


result = y;
```

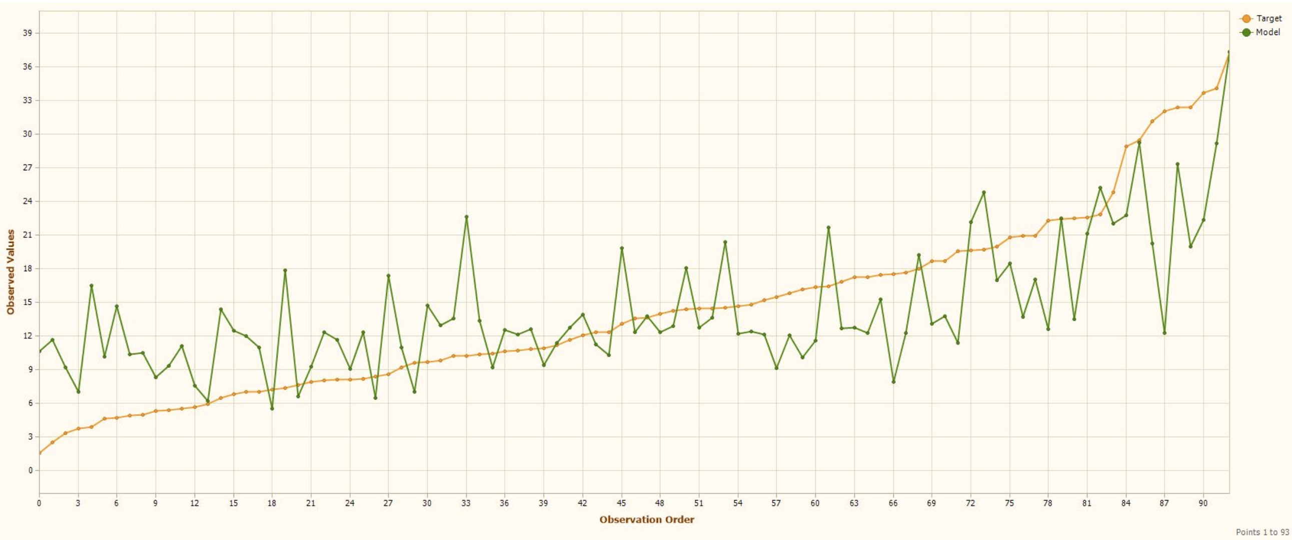
A5 Best GEP Model Output for Wee Jasper Station (SNSW)

A5.1 Output results for Calibration period

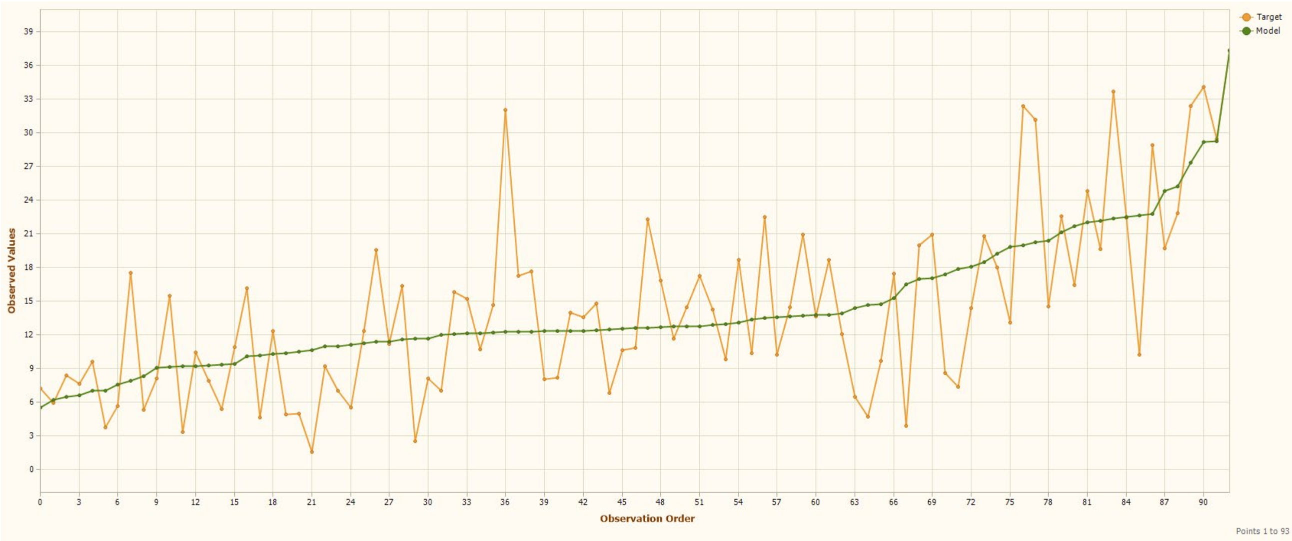
A5.1.1 Curve Fitting



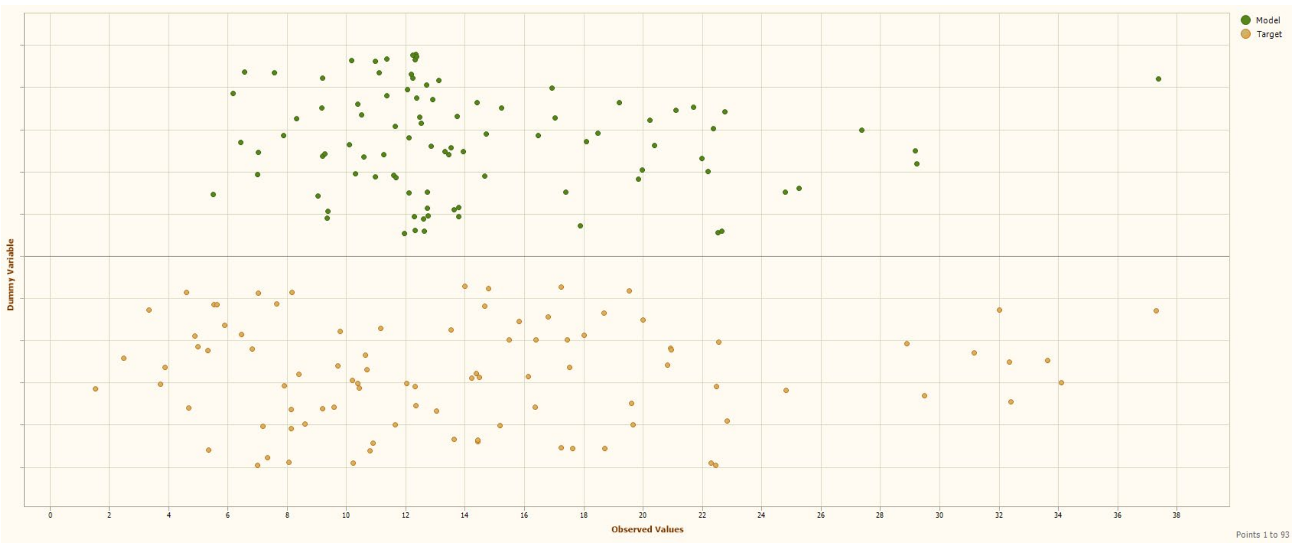
A5.1.2 Target sorted fitting



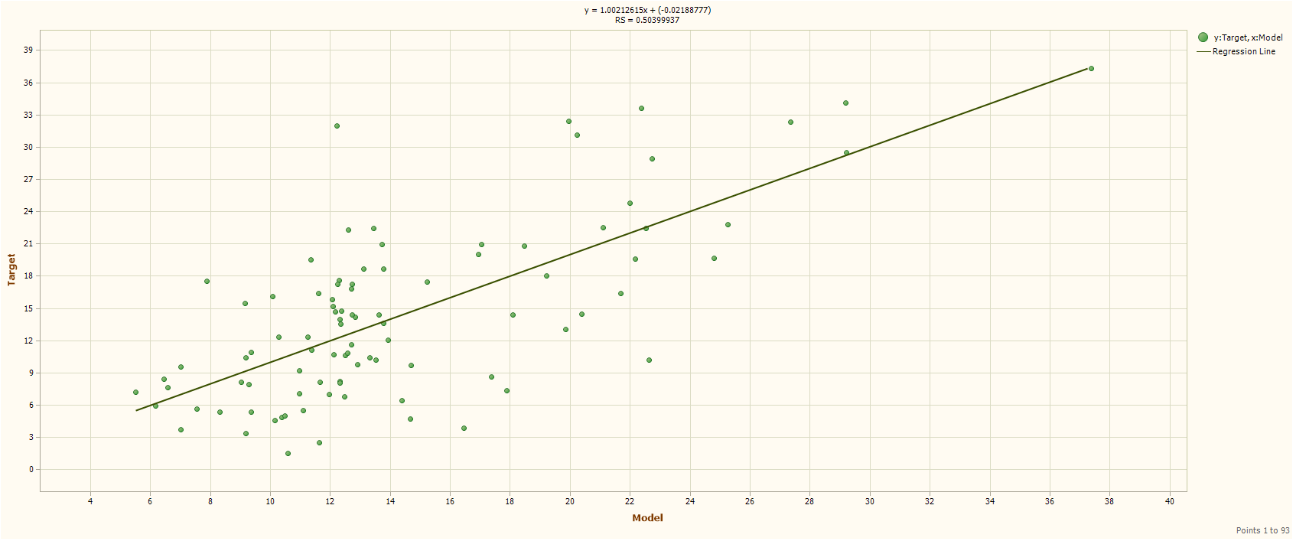
A5.1.3 Model sorted fitting



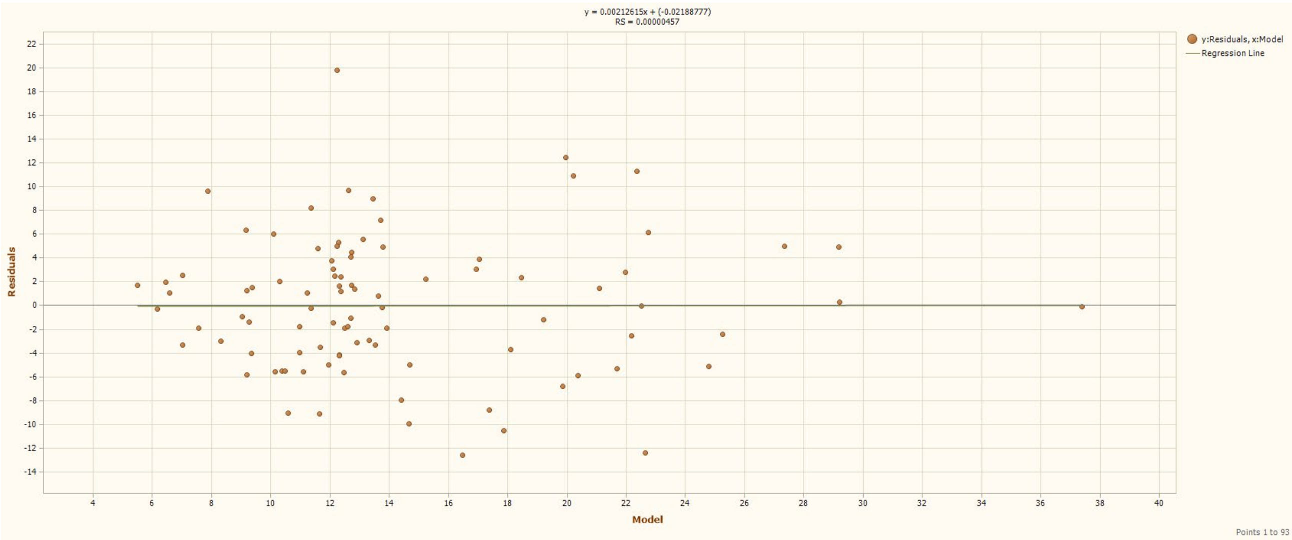
A5.1.4 Stacked distribution



A5.1.5 Scatter plot



A5.1.6 Residuals plot

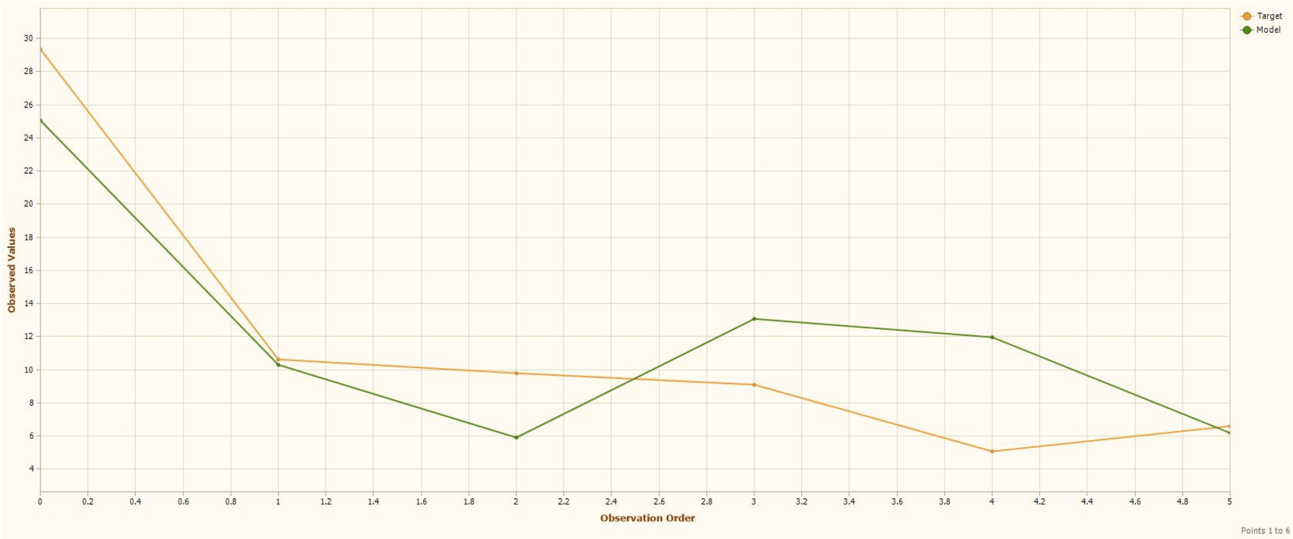


A5.1.7 Model performance

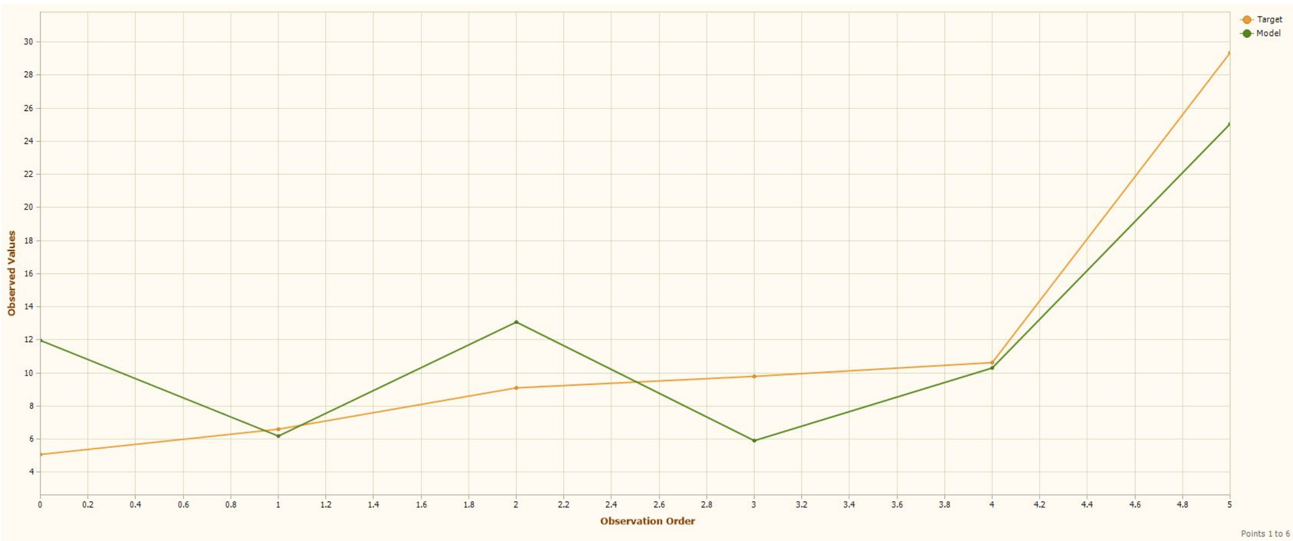
Fitness	149.559173836105
MSE	32.3341974865807
RMSE	5.68631668891038
MAE	4.47078781123361
RSE	0.496004011341018
RRSE	0.704275522321356
RAE	0.701064487688217
Correlation Coefficient	0.709929131358016
R-square	0.503999371550747
Calculation Errors	0

A5.2 Output results for Validation period

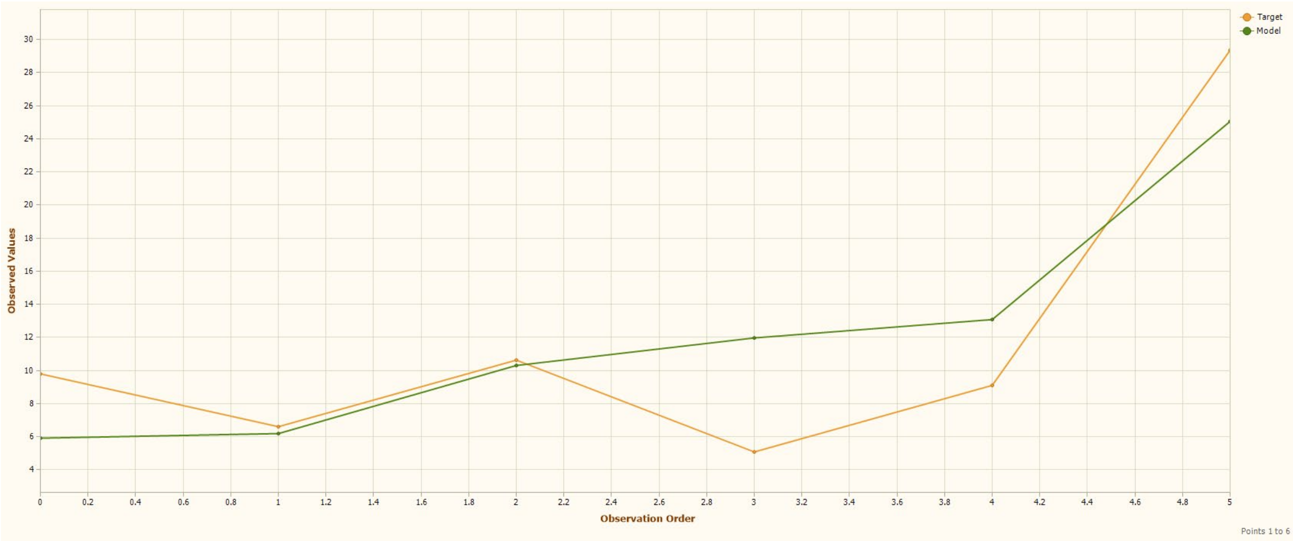
A5.2.1 Curve fitting



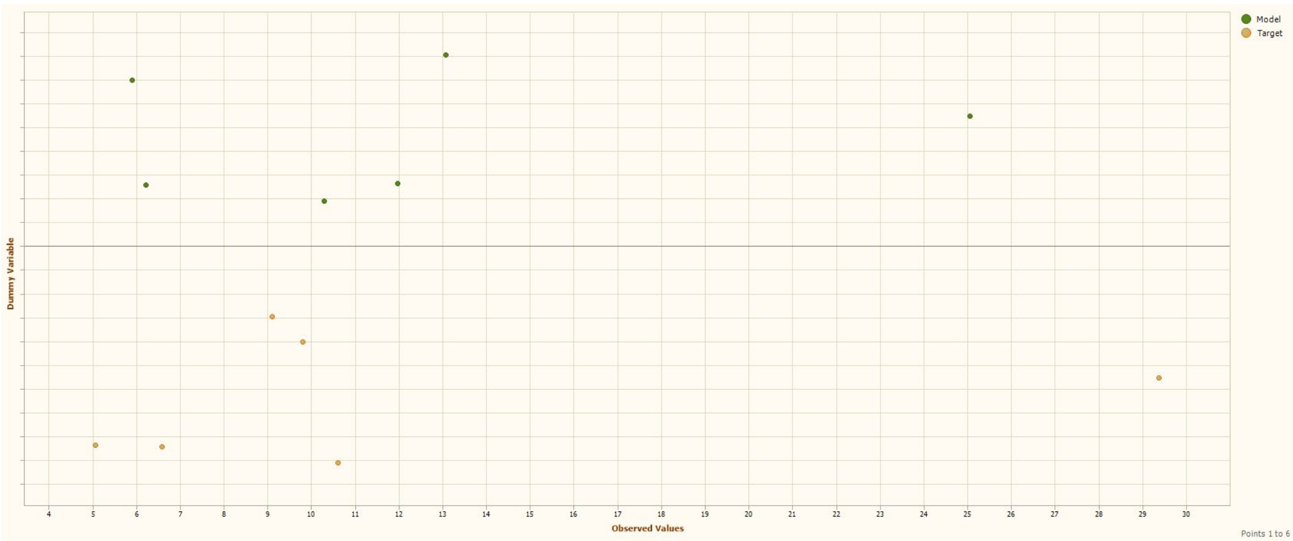
A5.2.2 Target sorted fitting



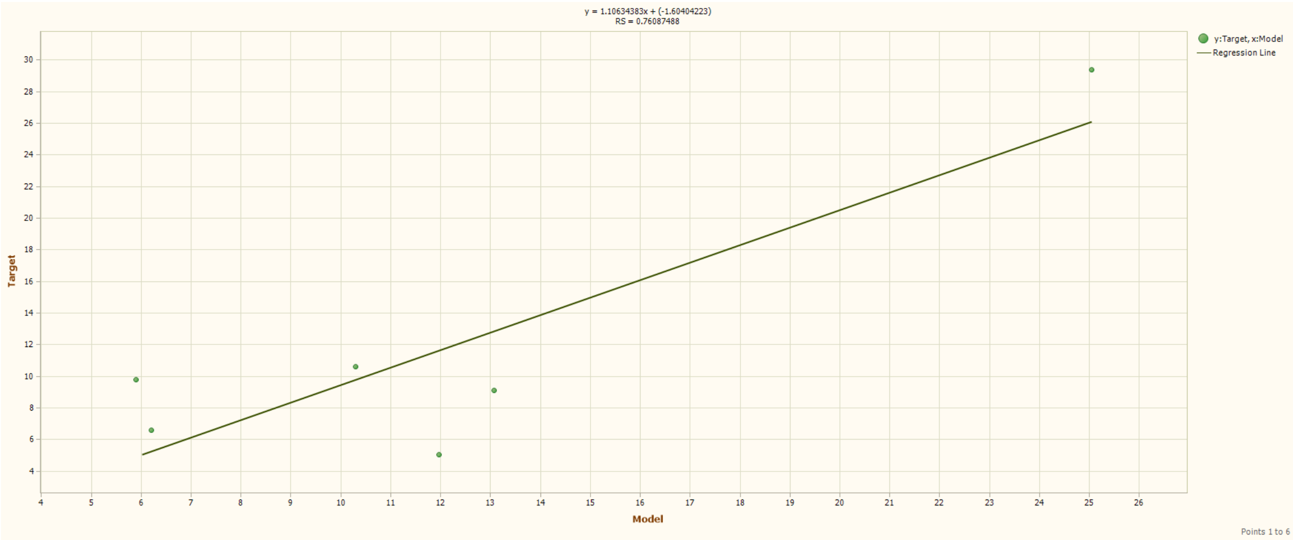
A5.2.3 Model sorted fitting



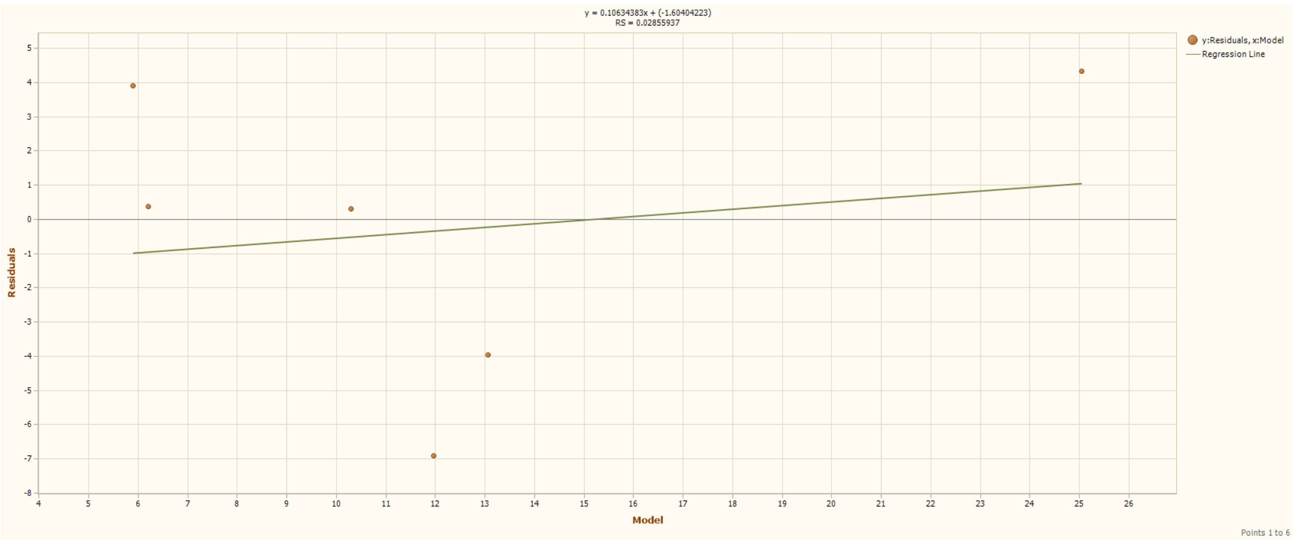
A5.2.4 Stacked distribution



A5.2.5 Scatter plot



A5.2.6 Residuals plot



A5.2.7 Model performance

Fitness 198.65228394696

MSE 16.2725225388221

RMSE 4.03392148396843

MAE 3.30059627687152

RSE 0.247710274566484

RRSE 0.497705007576259

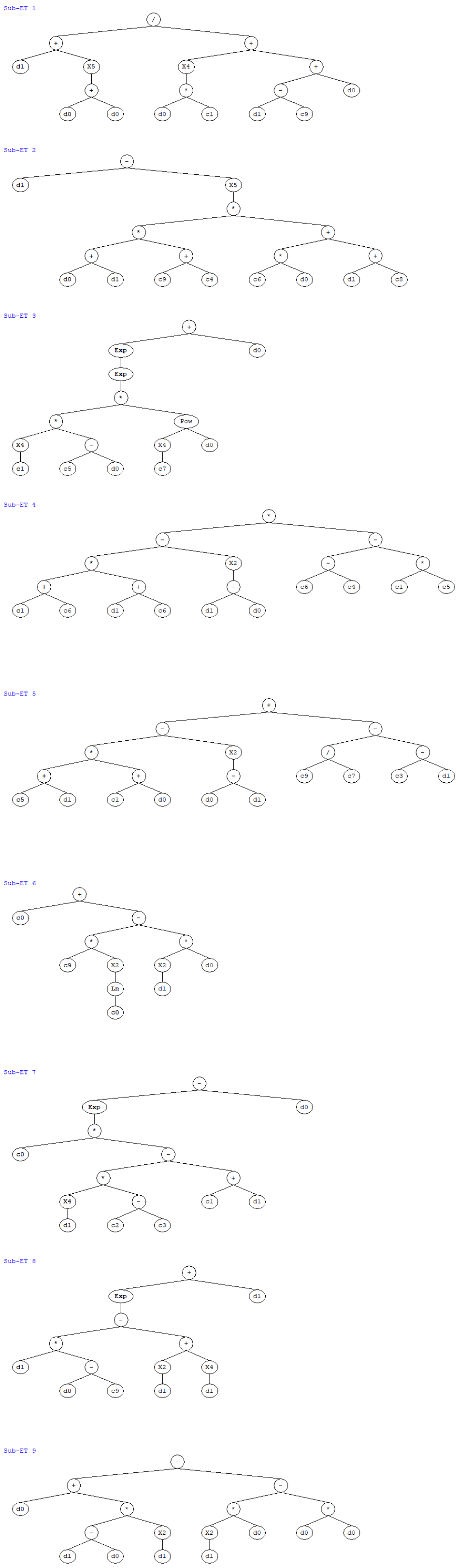
RAE 0.562047287951631

Correlation Coefficient 0.872281424377413

R-square 0.760874883313888

Calculation Errors 0

A5.3 Output model explained by Expression Tree



A5.4 Output model explained by MATLAB

```
%------------------------------------------------------------------------
% Regression model generated by GeneXproTools 5.0 on 24/09/2020 1:51:54 AM
% GEP File: C:\Users\resha\Google Drive\Final GeneXPro\SOUTH\station
7\h10g9\st07_2P_2N.gep
% Training Records:  93
% Validation Records:   6
% Fitness Function:  RMSE
% Training Fitness:  149.559173836105
% Training R-square: 0.503999371550747
% Validation Fitness:   198.65228394696
% Validation R-square:  0.760874883313888
%------------------------------------------------------------------------

function result = gepModel(d)

G1C1 = 2.02919395943907;
G1C9 = -0.395060156970892;
G2C9 = 1.04482345042268;
G2C4 = -1.03061006500443;
G2C6 = 8.10788201666699;
G2C8 = -5.81945921572754;
G3C1 = -69.2504444698088;
G3C5 = -1.5164121860644;
G3C7 = 389.098281285211;
G4C6 = -0.97334433035802;
G4C4 = -1.7999427288833;
G4C1 = -1.2880786371039;
G4C5 = -0.219702230850183;
G5C9 = 8.21667394968258;
G5C7 = 2.50386650251438;
G5C3 = 5.68711004901465;
G5C5 = -1.72012253274151;
G5C1 = -3.90518042417561;
```

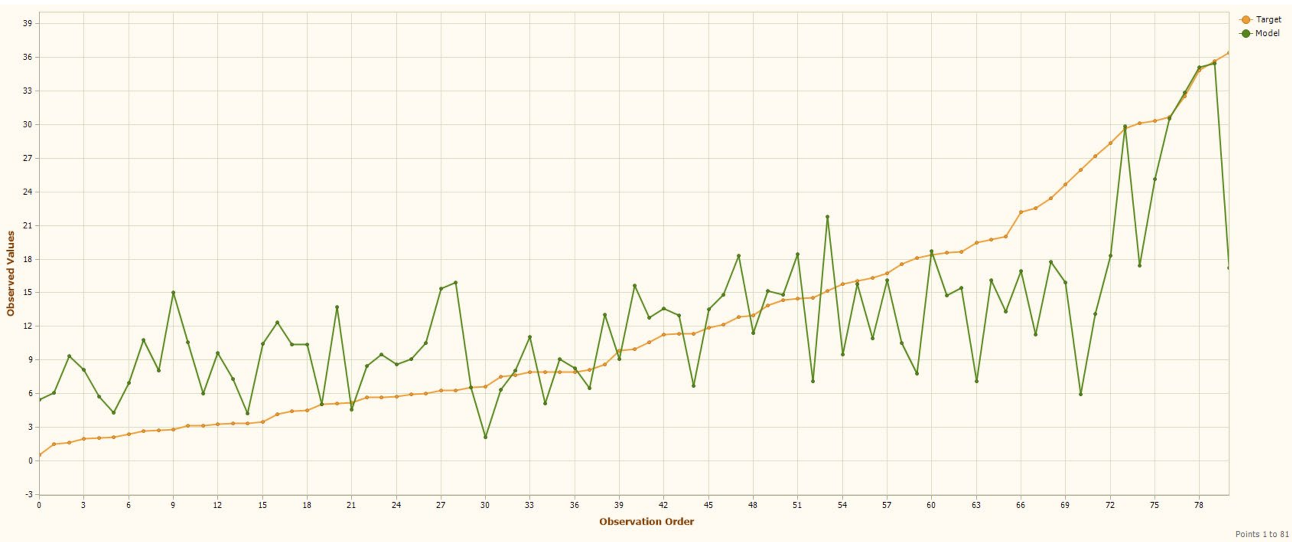
```
G6C0 = 8.46833477059208;
G6C9 = -0.729920407768304;
G7C0 = -84.2178973157296;
G7C1 = -0.265850915033099;
G7C2 = -3.98068900741016;
G7C3 = -7.98585653706473;
G8C9 = 2.28689086800411;

y = 0.0;

y = ((d(2)+((d(1)+d(1))^5))/(((d(1)*G1C1)^4)+((d(2)-G1C9)+d(1))));
y = y + (d(2)-((((d(1)+d(2))*(G2C9+G2C4))*((G2C6*d(1))+(d(2)+G2C8)))^5));
y = y + (exp(exp((((G3C1^4)*(G3C5-d(1)))*realpow((G3C7^4),d(1)))))+d(1));
y = y + ((((G4C1+G4C6)*(d(2)+G4C6))-((d(2)-d(1))^2))*((G4C6-G4C4)-(G4C1*G4C5)));
y = y + ((((G5C5+d(2))*(G5C1+d(1)))-((d(1)-d(2))^2))+((G5C9/G5C7)-(G5C3-d(2))));
y = y + (G6C0+((G6C9*(reallog(G6C0)^2))-((d(2)^2)*d(1))));
y = y + (exp((G7C0*(((d(2)^4)*(G7C2-G7C3))-(G7C1+d(2)))))-d(1));
y = y + (exp(((d(2)*(d(1)-G8C9))-((d(2)^2)+(d(2)^4))))+d(2));
y = y + ((d(1)+((d(2)-d(1))*(d(2)^2)))-(((d(2)^2)*d(1))-(d(1)*d(1))));

result = y;
```

A6 Best GEP Model Output for Mittagang Crossing (SNSW)

A6.1 Output results for Calibration period

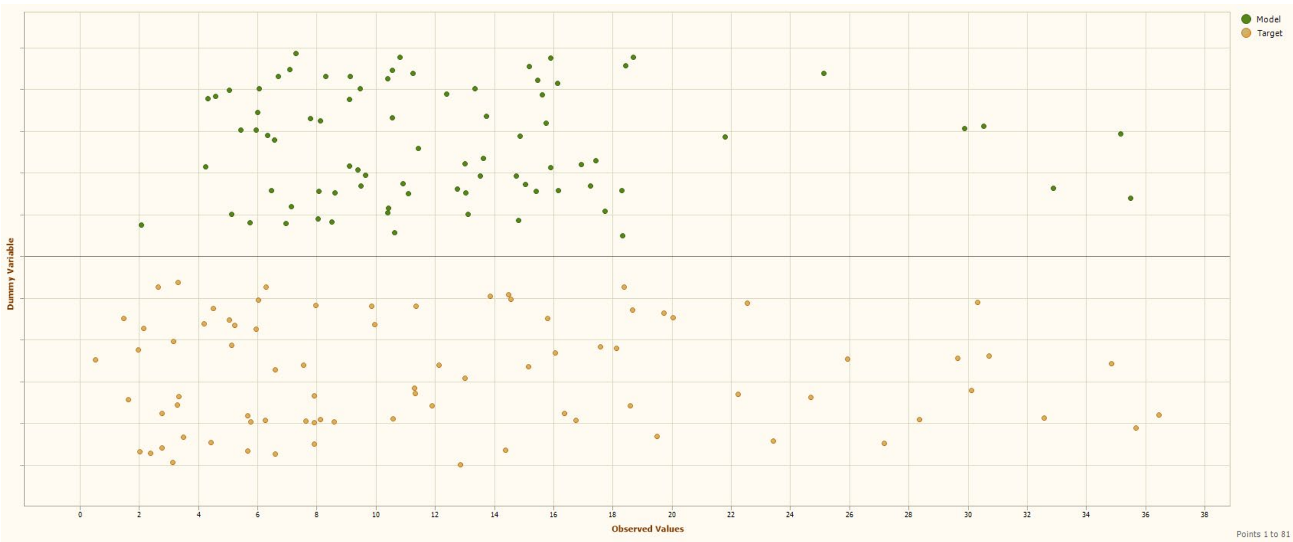
A6.1.1 Curve Fitting

A6.1.2 Target sorted fitting

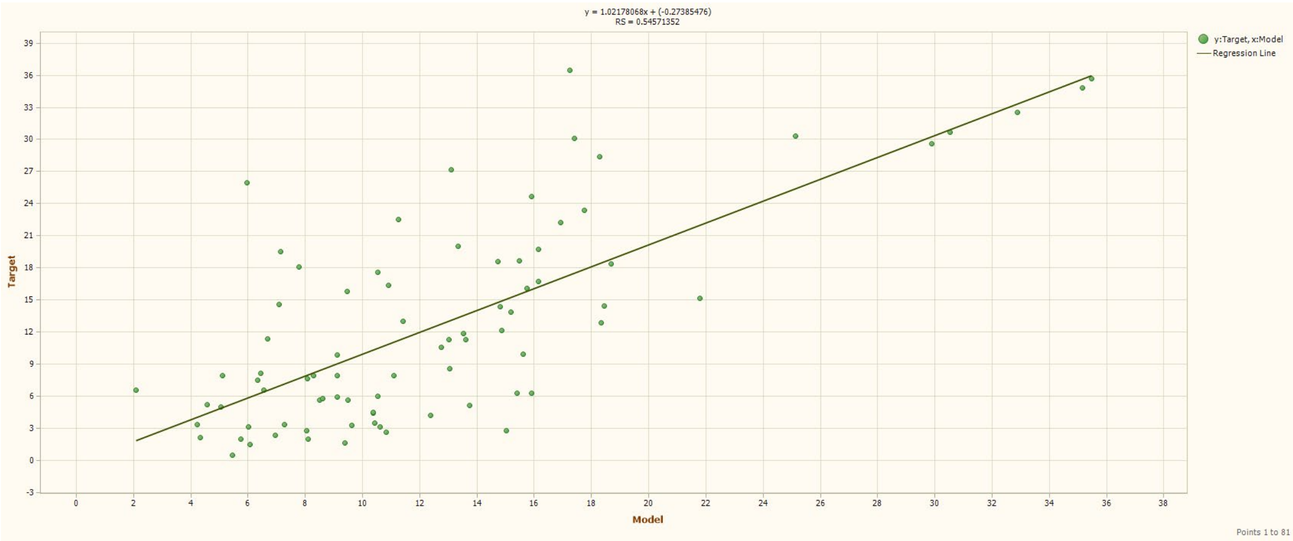


A6.1.3 Model sorted fitting

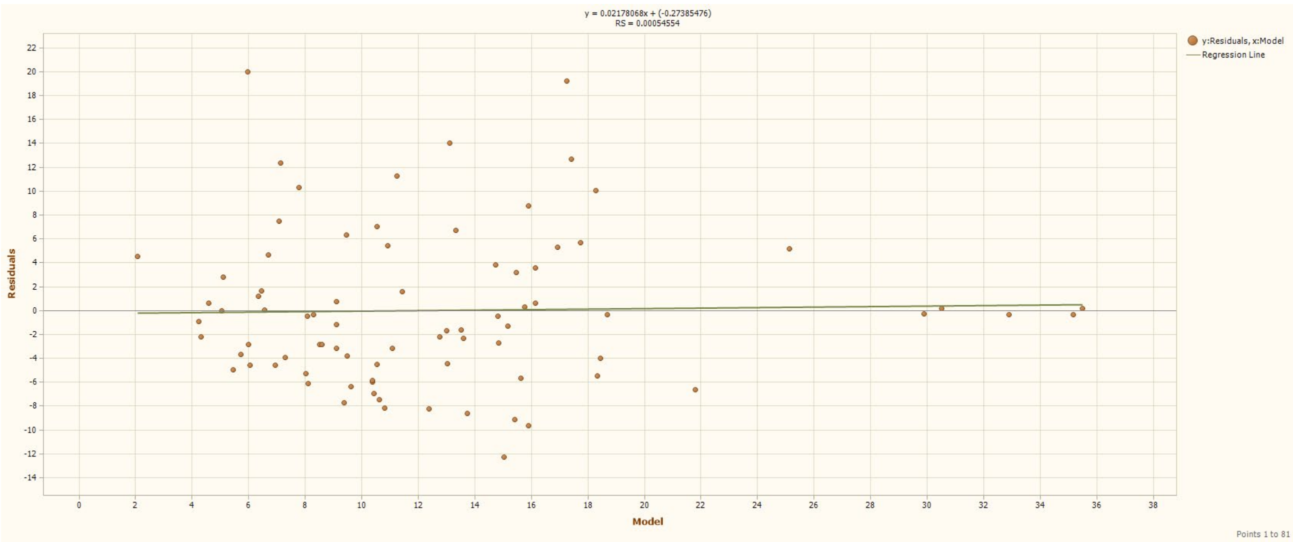
A6.1.4 Stacked distribution



A6.1.5 Scatter plot



A6.1.6 Residuals plot

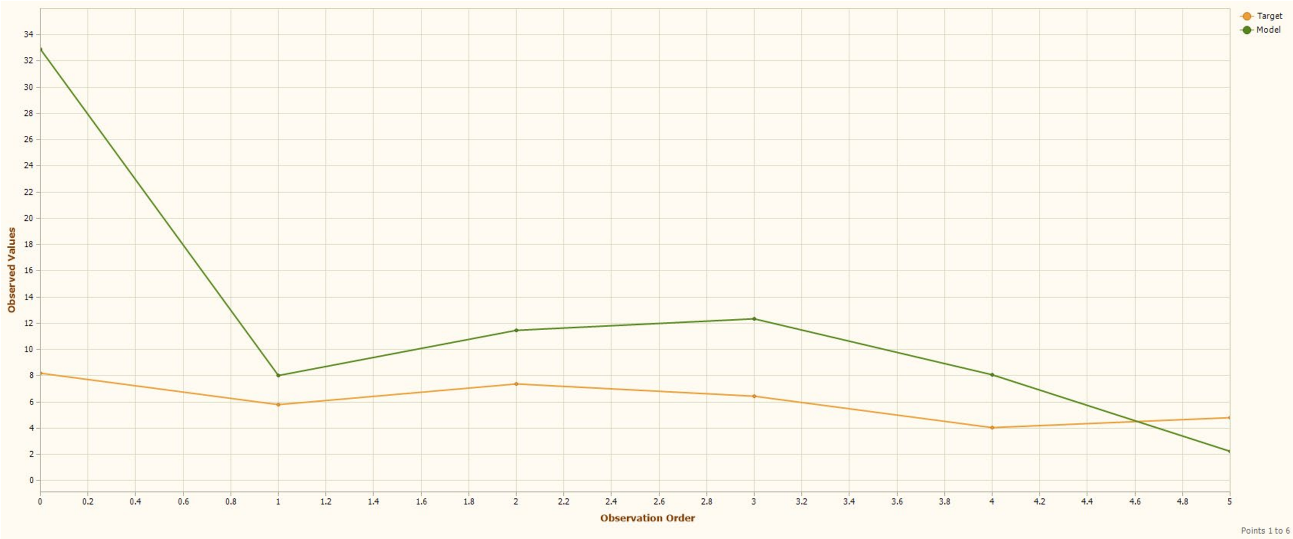


A6.1.7 Model performance

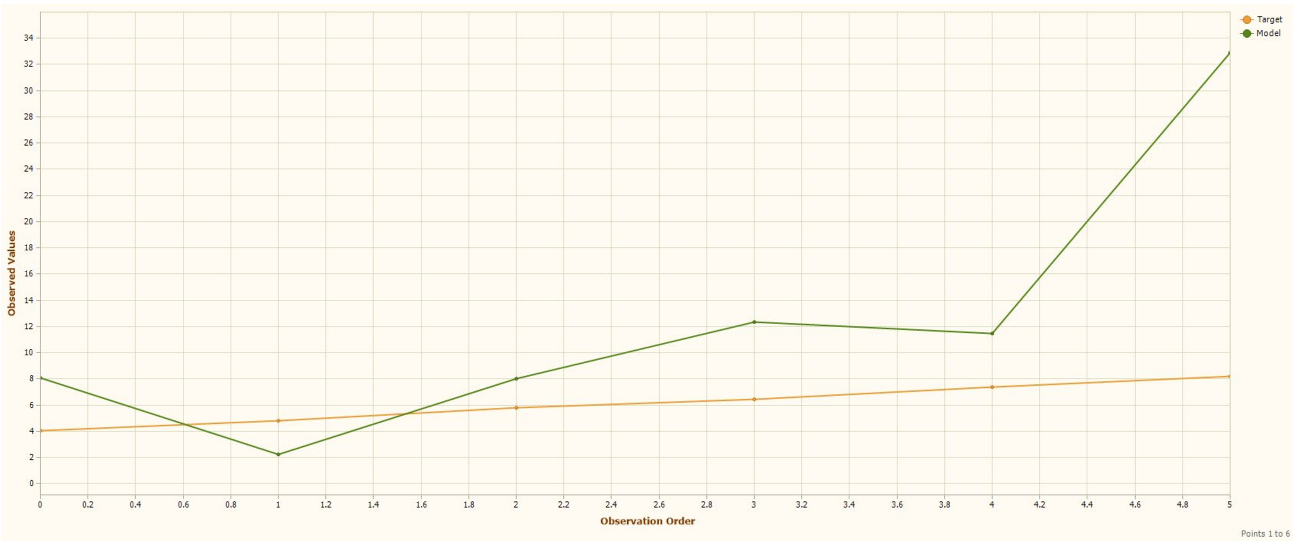
Fitness	135.074266362845
MSE	41.0026949729297
RMSE	6.40333467600513
MAE	4.88039565961857
RSE	0.454534464561648
RRSE	0.674191712023849
RAE	0.620088399188388
Correlation Coefficient	0.738724252724388
R-square	0.545713521563205
Calculation Errors	0

A6.2 Output results for Validation period

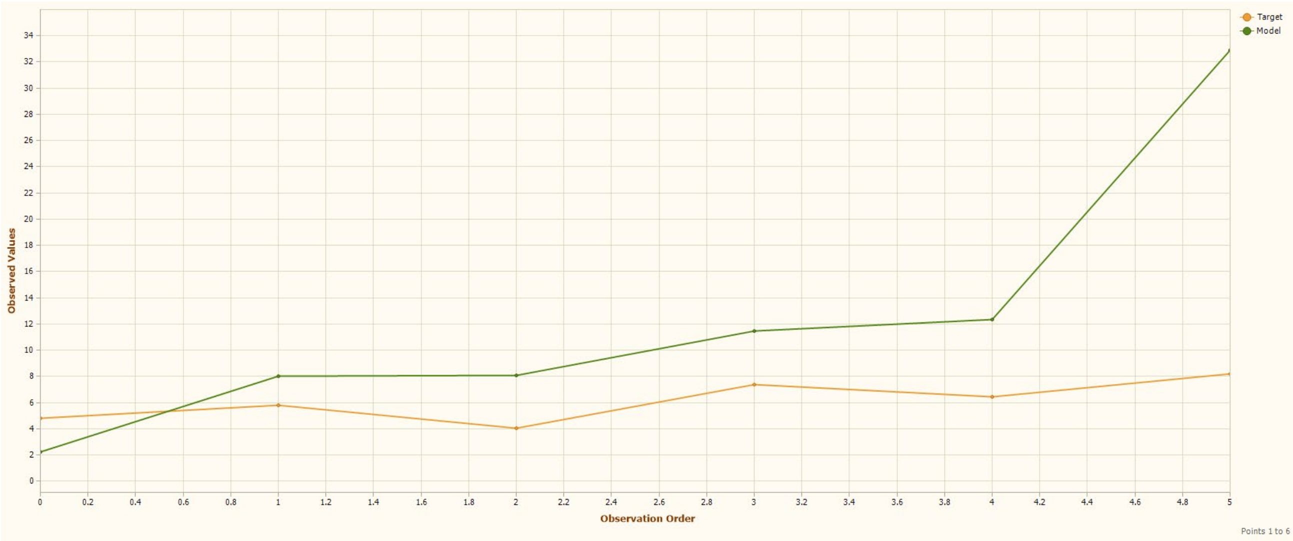
A6.2.1 Curve fitting



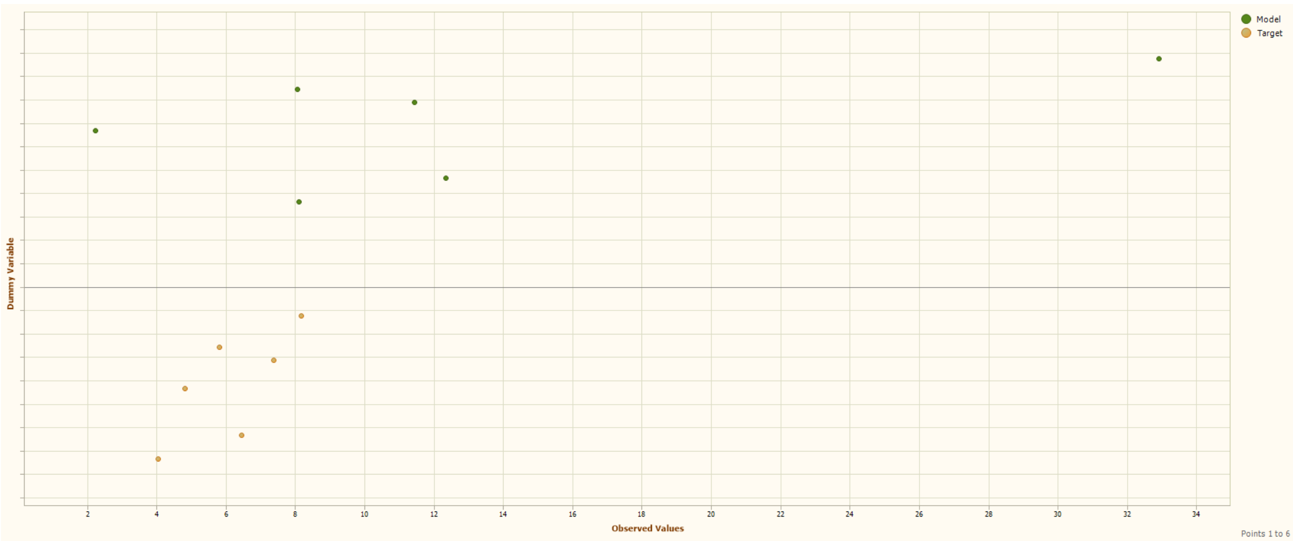
A6.2.2 Target sorted fitting



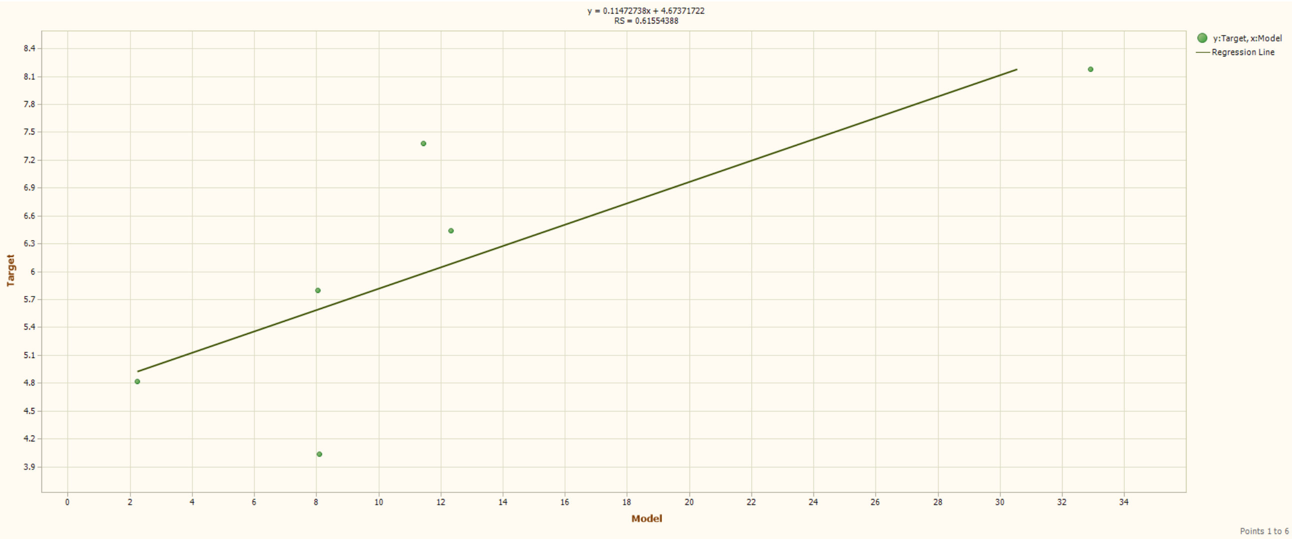
A6.2.3 Model sorted fitting



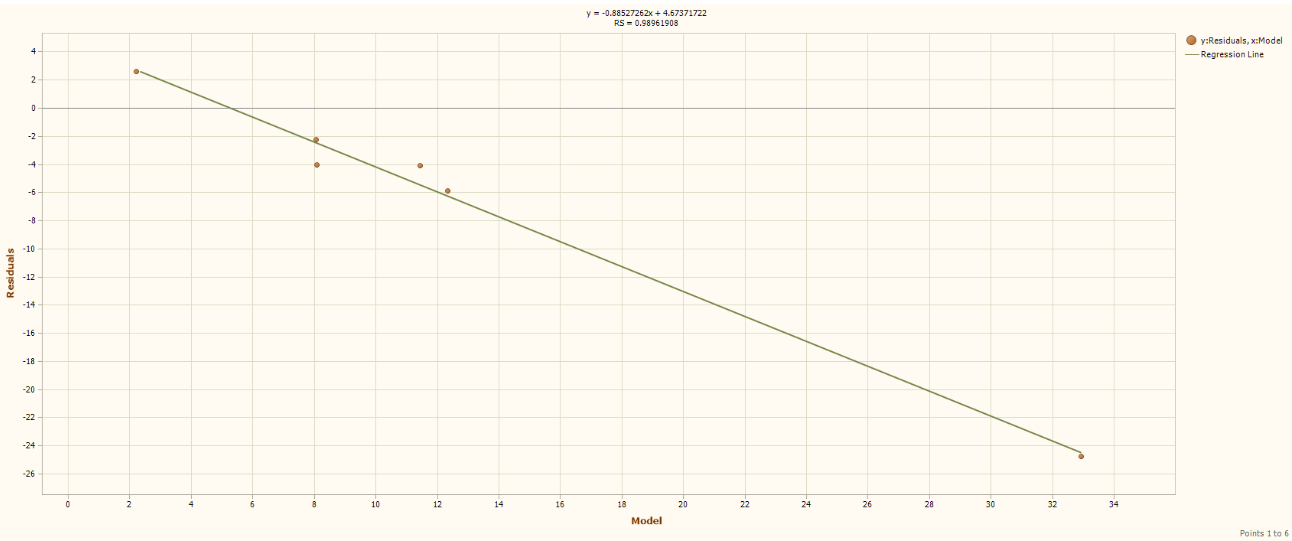
A6.2.4 Stacked distribution



A6.2.5 Scatter plot



A6.2.6 Residuals plot



A6.2.7 Model performance

Fitness 8.52070227534629E-03

MSE 115.264311997435

RMSE 10.7361218322742

MAE 7.266676347355

RSE 57.4543303087053

RRSE 7.57986347559804

RAE 5.93869318612283

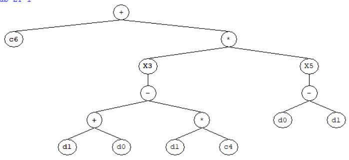
Correlation Coefficient 0.784566045134735

R-square 0.61554387917836

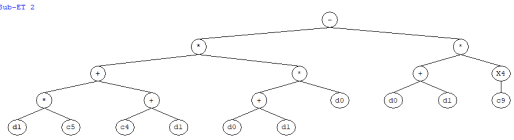
Calculation Errors 0

A6.3 Output model explained by Expression Tree

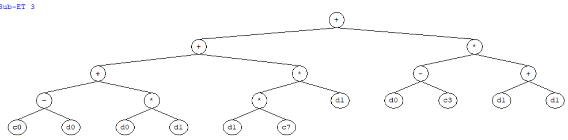
Sub-ET 1



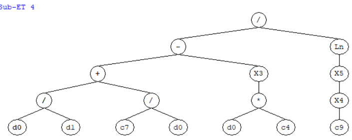
Sub-ET 2



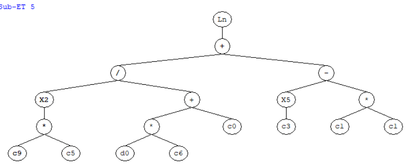
Sub-ET 3



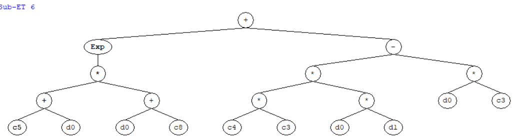
Sub-ET 4



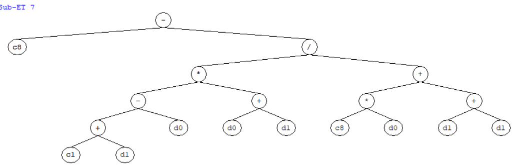
Sub-ET 5



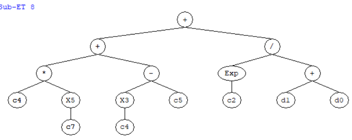
Sub-ET 6



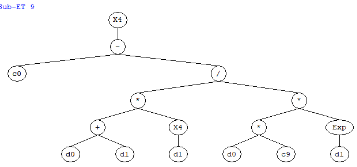
Sub-ET 7



Sub-ET 8



Sub-ET 9



A6.4 Output model explained by MATLAB

```
%------------------------------------------------------------------------

% Regression model generated by GeneXproTools 5.0 on 24/09/2020 4:01:48 AM

% GEP File: C:\Users\resha\Google Drive\Final GeneXPro\SOUTH\station 8\st08_6I_4N.gep

% Training Records:  81

% Validation Records:   6

% Fitness Function:  RMSE

% Training Fitness:  135.074266362845

% Training R-square: 0.545713521563205

% Validation Fitness:   8.52070227534629E-03

% Validation R-square:  0.61554387917836

%------------------------------------------------------------------------


function result = gepModelQ8C(d)


G1C6 = 0.352038906223837;

G1C4 = 0.274187613279256;

G2C9 = -1.6555446648697;

G2C5 = 28.6909954434334;

G2C4 = -9.89075983912494;

G3C3 = 6.31739576544455;

G3C0 = 4.28876759598874;

G3C7 = 13.050408090317;

G4C7 = 0.546463076497621;

G4C4 = -5.74836421579443;

G4C9 = 3.75538966743452;

G5C0 = -3.92330611422773;

G5C3 = 7.23974306492235;

G5C1 = 11.6994096468387;

G5C9 = 4.4911718312656;
```

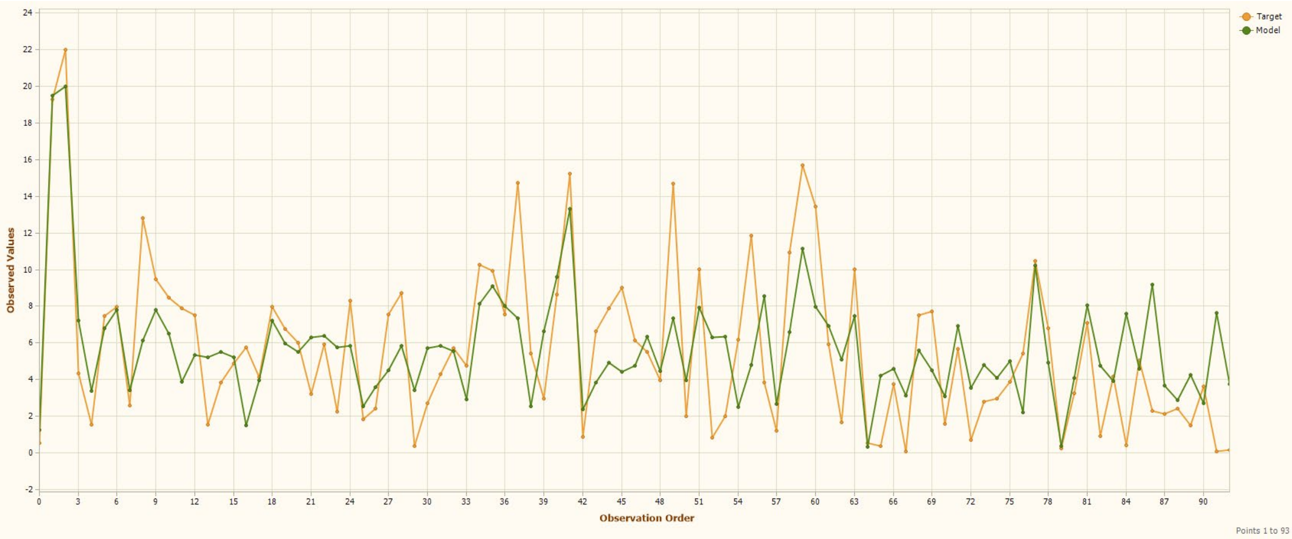
```
G5C5 = 11.8207427971022;

G5C6 = -7.7680436005704;

G6C3 = 1.31890315116497;

G6C5 = 0.905661284145816;

G6C8 = -11.570741760496;

G6C4 = 6.45639209732433;

G7C8 = 1.89602343821528;

G7C1 = 2.11735101232002;

G8C4 = -4.69954527420881;

G8C5 = 8.07280208385146;

G8C2 = -2.39515470752768;

G8C7 = 2.07495345927305;

G9C0 = -4.1300698873867;

G9C9 = -7.46202851756751;


Four_NC = 1;

Six_IC = 2;


y = 0.0;


y = (G1C6+((((d(Six_IC)+d(Four_NC))-(d(Six_IC)*G1C4))^3)*((d(Four_NC)-d(Six_IC))^5)));

y = y + ((((d(Six_IC)*G2C5)+(G2C4+d(Six_IC)))*((d(Four_NC)+d(Six_IC))*d(Four_NC)))-
((d(Four_NC)+d(Six_IC))*(G2C9^4)));

y = y + ((((G3C0-d(Four_NC))+(d(Four_NC)*d(Six_IC)))+((d(Six_IC)*G3C7)*d(Six_IC)))+((d(Four_NC)-
G3C3)*(d(Six_IC)+d(Six_IC))));

y = y + ((((d(Four_NC)/d(Six_IC))+(G4C7/d(Four_NC)))-((d(Four_NC)*G4C4)^3))/reallog(((G4C9^4)^5)));

y = y + reallog(((((G5C9*G5C5)^2)/((d(Four_NC)*G5C6)+G5C0))+((G5C3^5)-(G5C1*G5C1))));

y = y + (exp(((G6C5+d(Four_NC))*(d(Four_NC)+G6C8)))+(((G6C4*G6C3)*(d(Four_NC)*d(Six_IC)))-
(d(Four_NC)*G6C3)));

y = y + (G7C8-((((G7C1+d(Six_IC))-
d(Four_NC))*(d(Four_NC)+d(Six_IC)))/((G7C8*d(Four_NC))+(d(Six_IC)+d(Six_IC)))));

y = y + (((G8C4*(G8C7^5))+((G8C4^3)-G8C5))+(exp(G8C2)/(d(Six_IC)+d(Four_NC))));
```

```
y = y + ((G9C0-(((d(Four_NC)+d(Six_IC))*(d(Six_IC)^4))/((d(Four_NC)*G9C9)*exp(d(Six_IC)))))^4);

result = y;
```

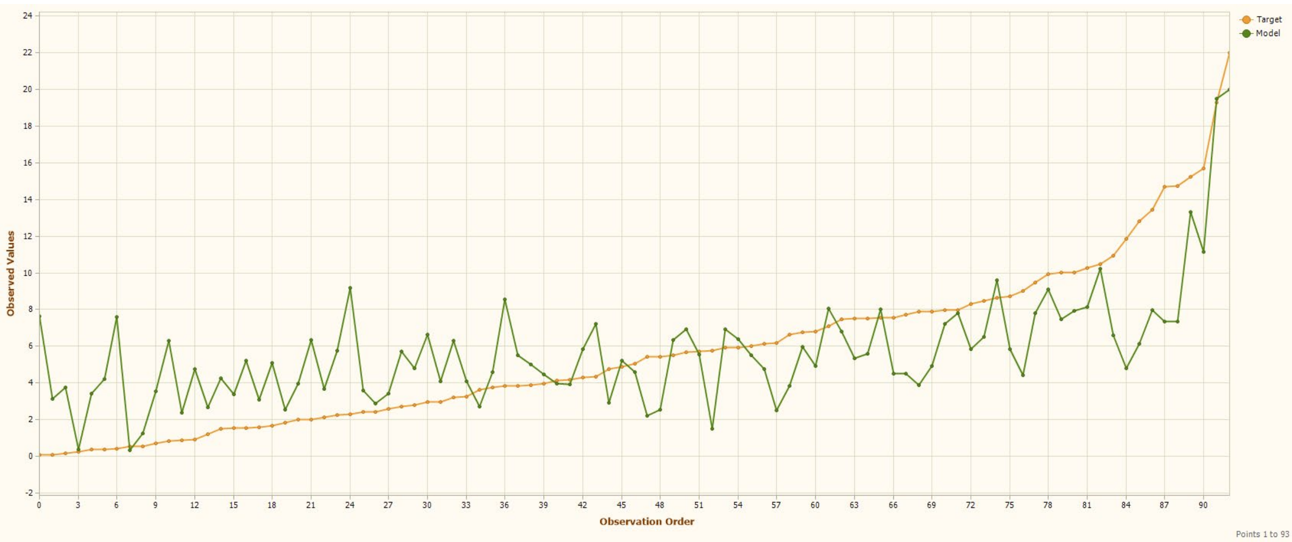
A7 Best GEP Model Output for Kiosk Station (SNSW)

A7.1 Output results for Calibration period

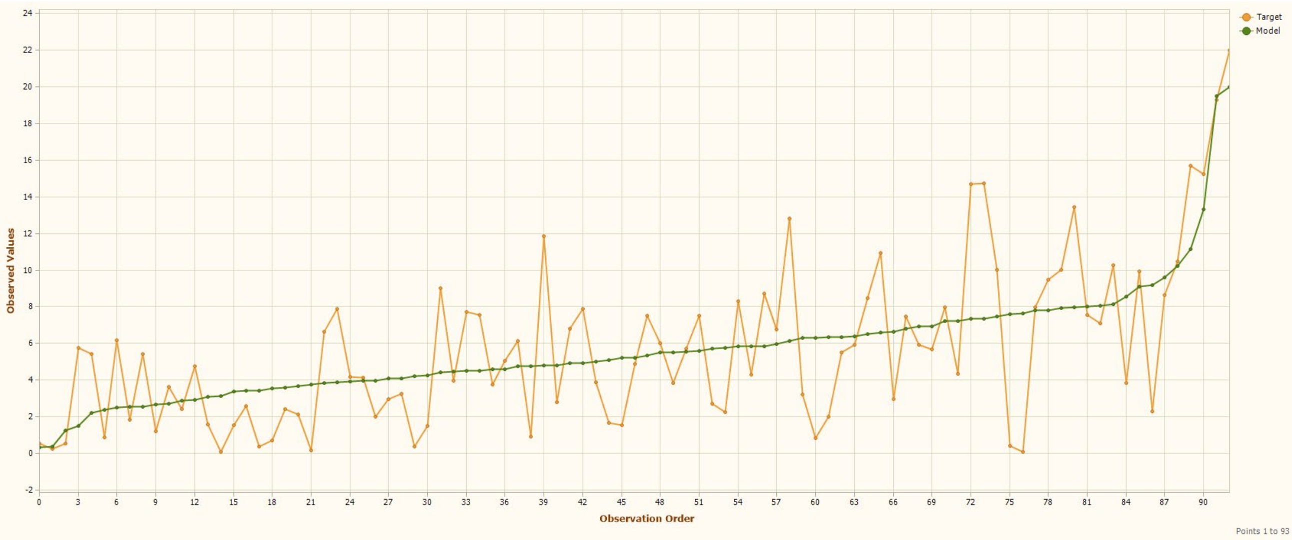
A7.1.1 Curve Fitting



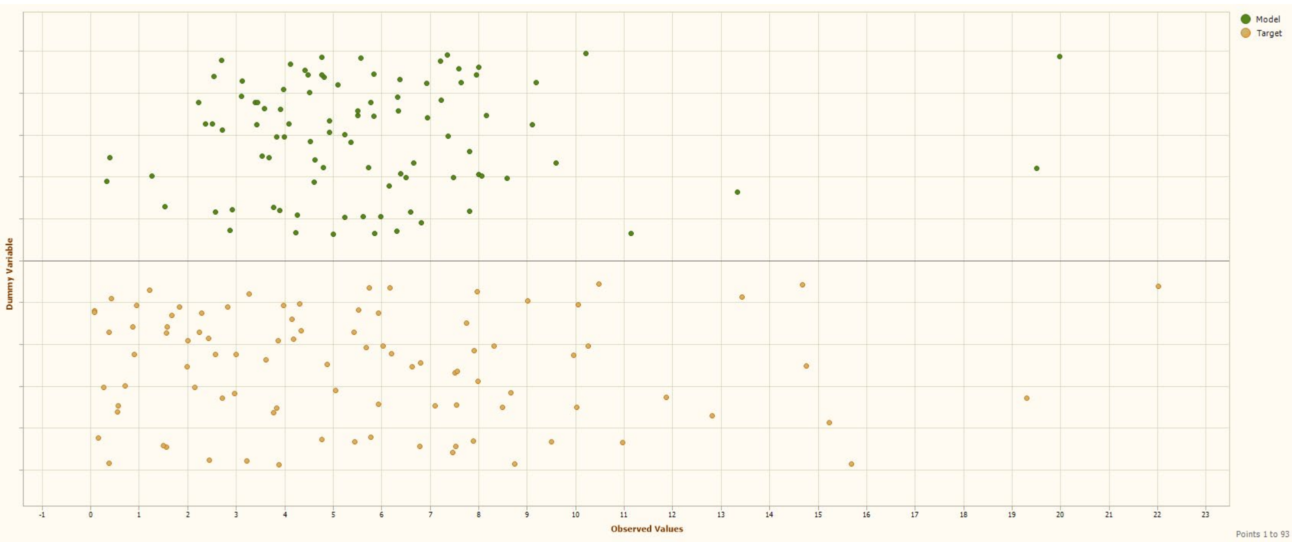
A7.1.2 Target sorted fitting



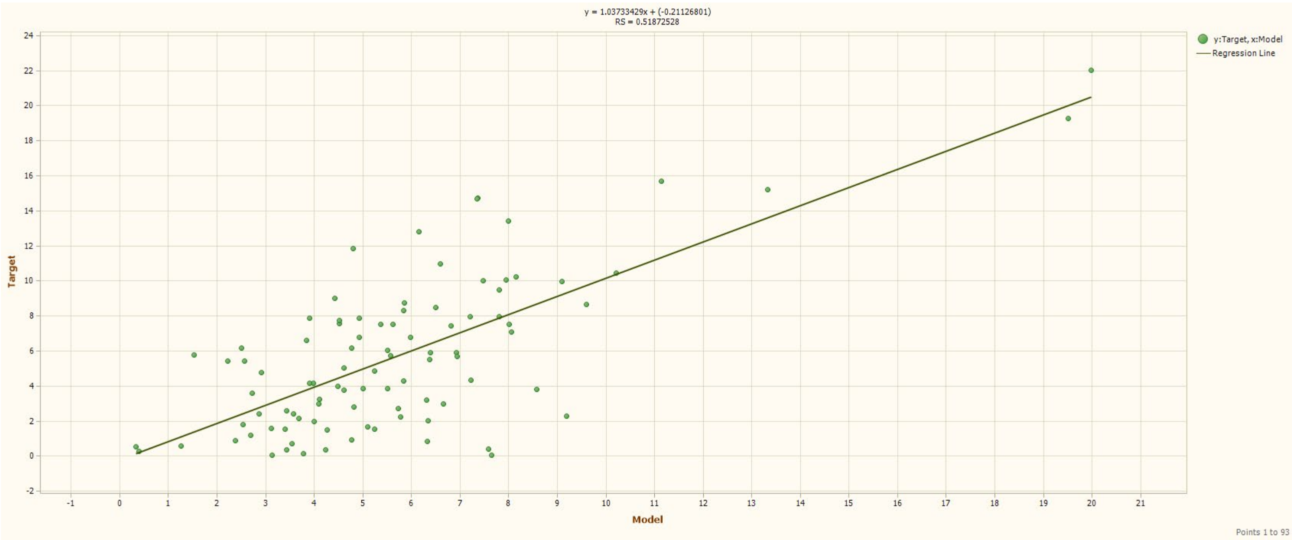
A7.1.3 Model sorted fitting



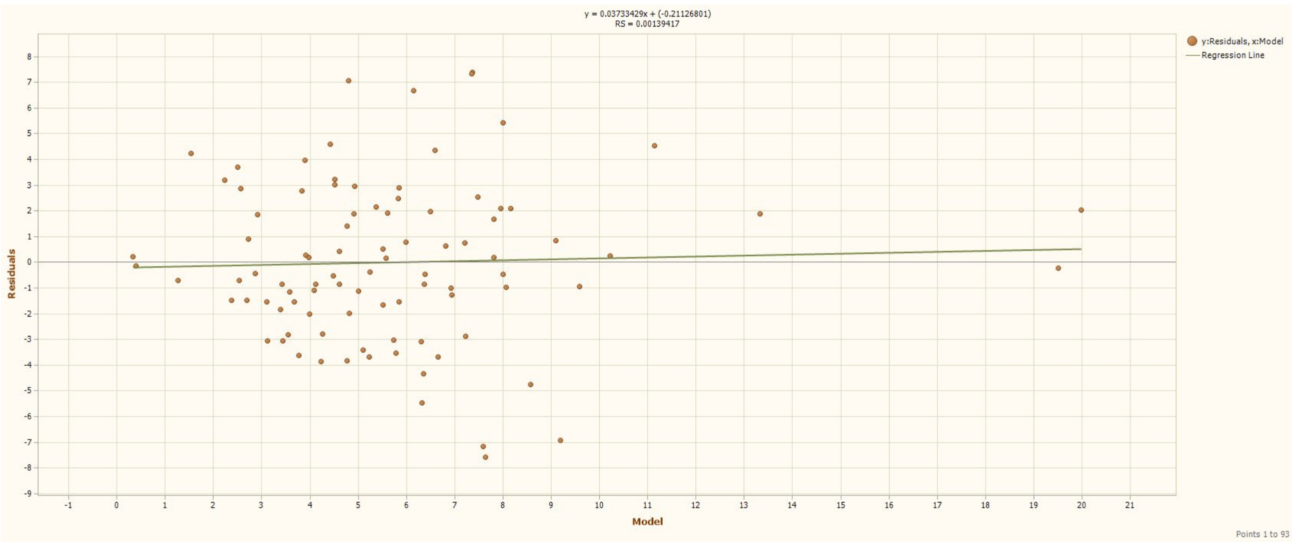
A7.1.4 Stacked distribution



A7.1.5 Scatter plot



A7.1.6 Residuals plot

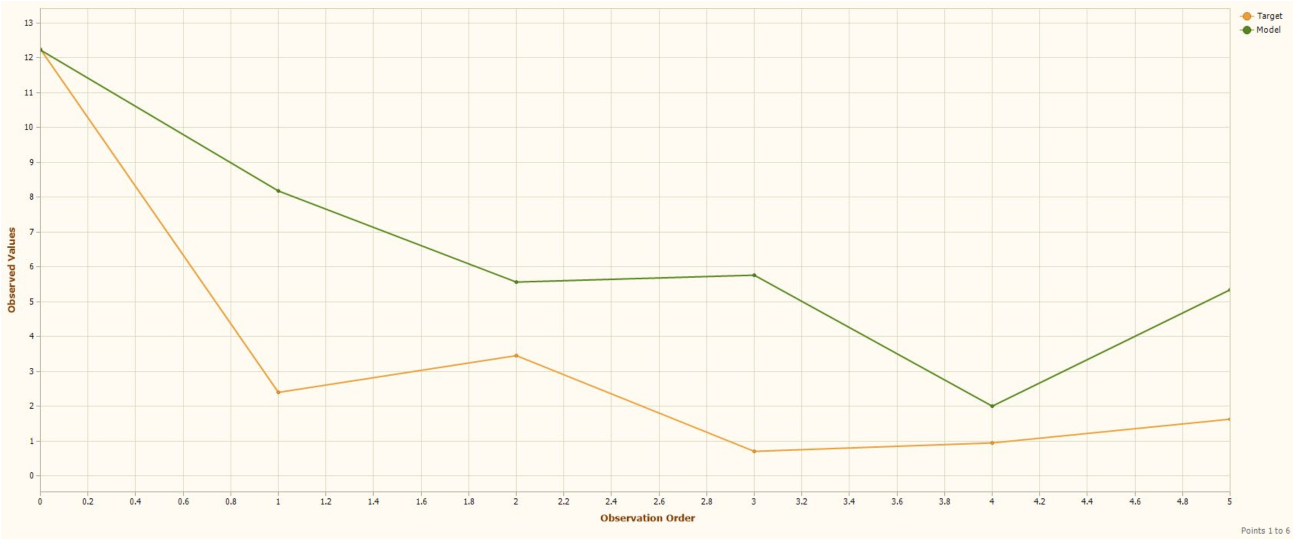


A7.1.7 Model performance

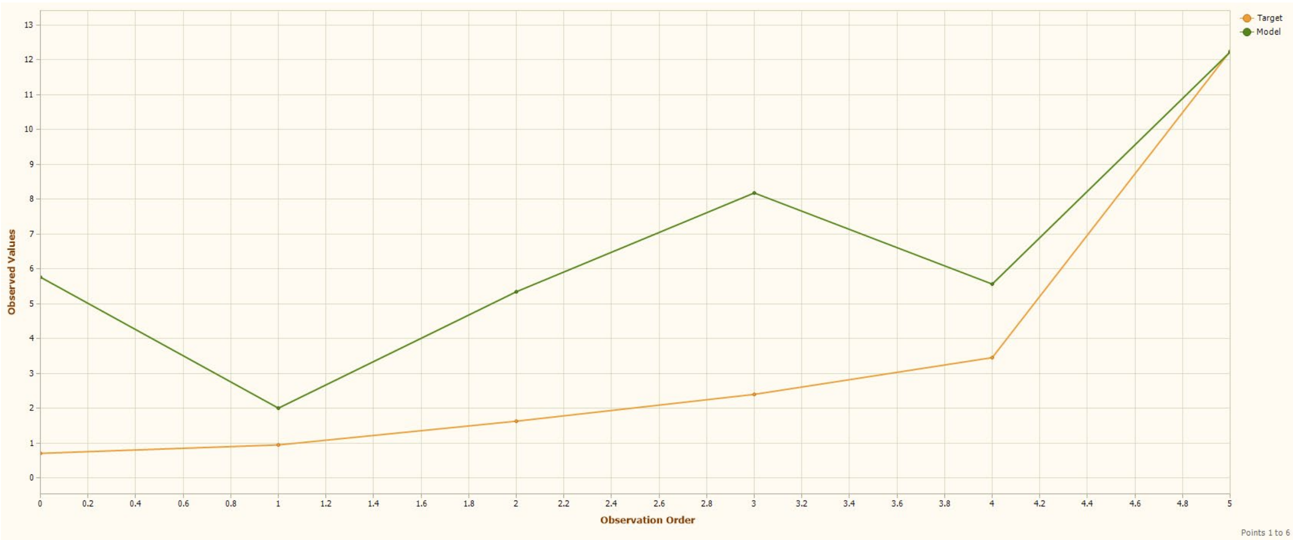
Fitness	245.121703630077
MSE	9.4839737972001
RMSE	3.07960611072262
MAE	2.4185239801213
RSE	0.481946765262886
RRSE	0.694223858177524
RAE	0.69913727832635
Correlation Coefficient	0.720225854722973
R-square	0.518725281811437
Calculation Errors	0

A7.2 Output results for Validation period

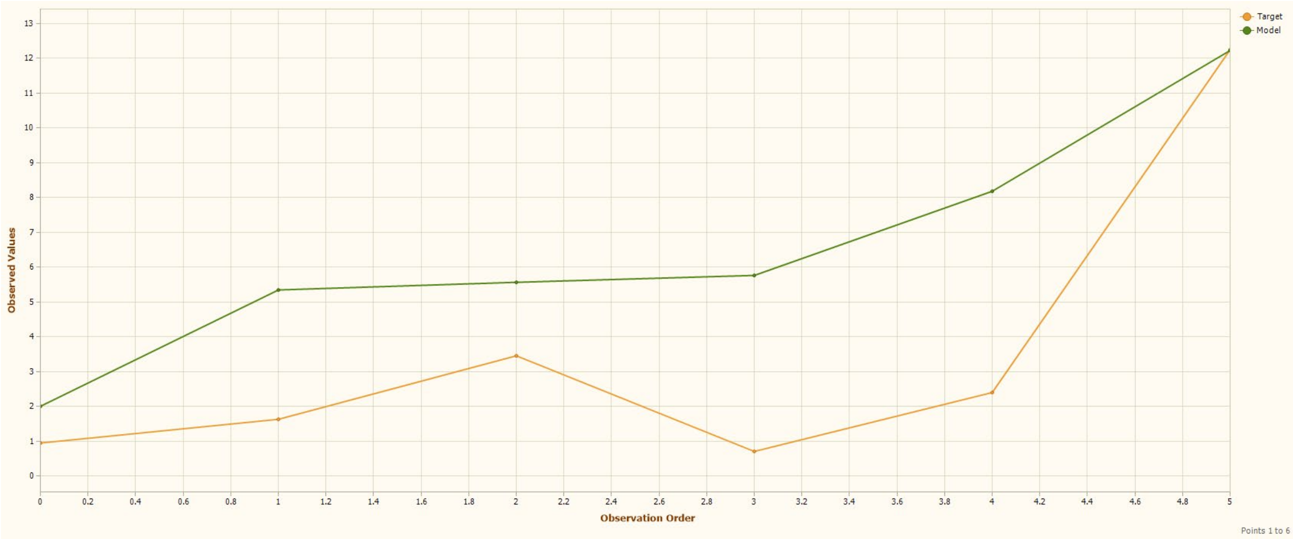
A7.2.1 Curve fitting



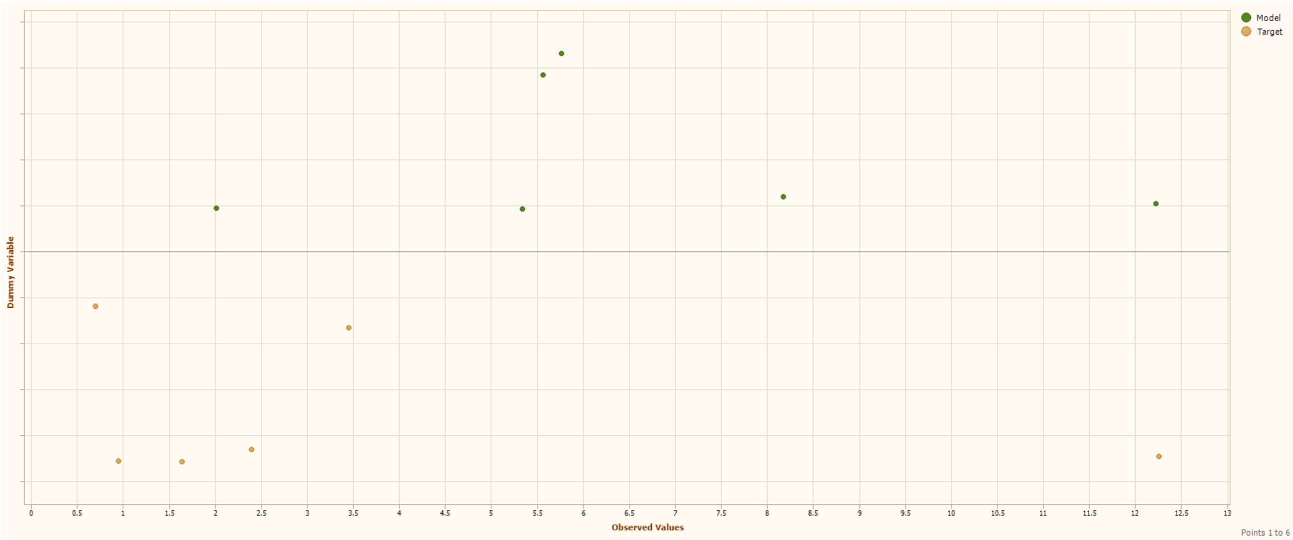
A7.2.2 Target sorted fitting



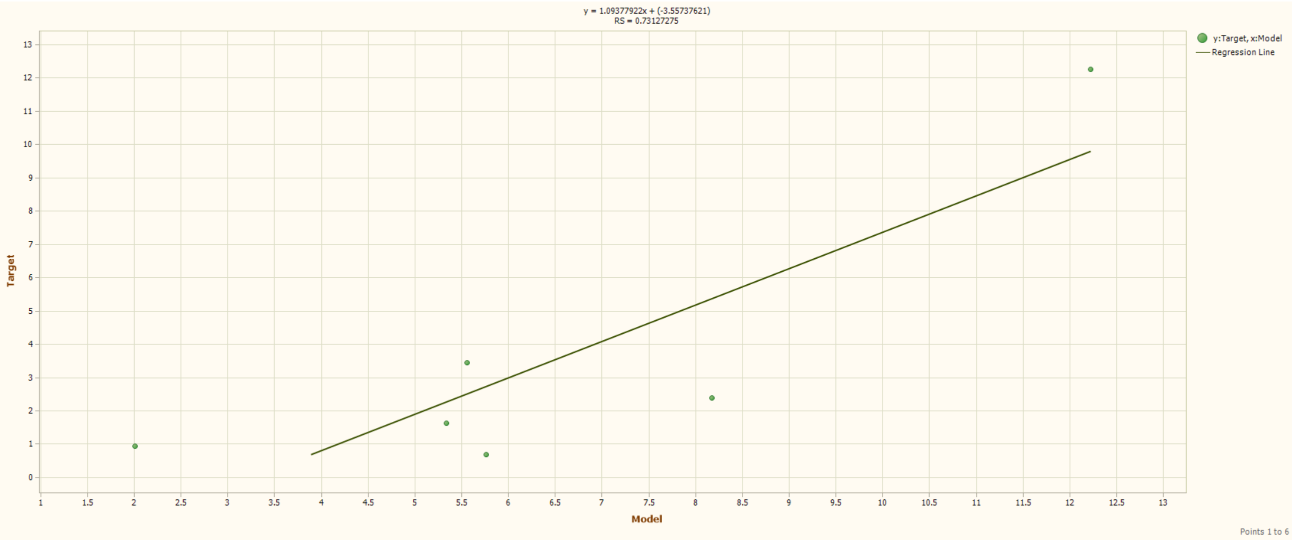
A7.2.3 Model sorted fitting



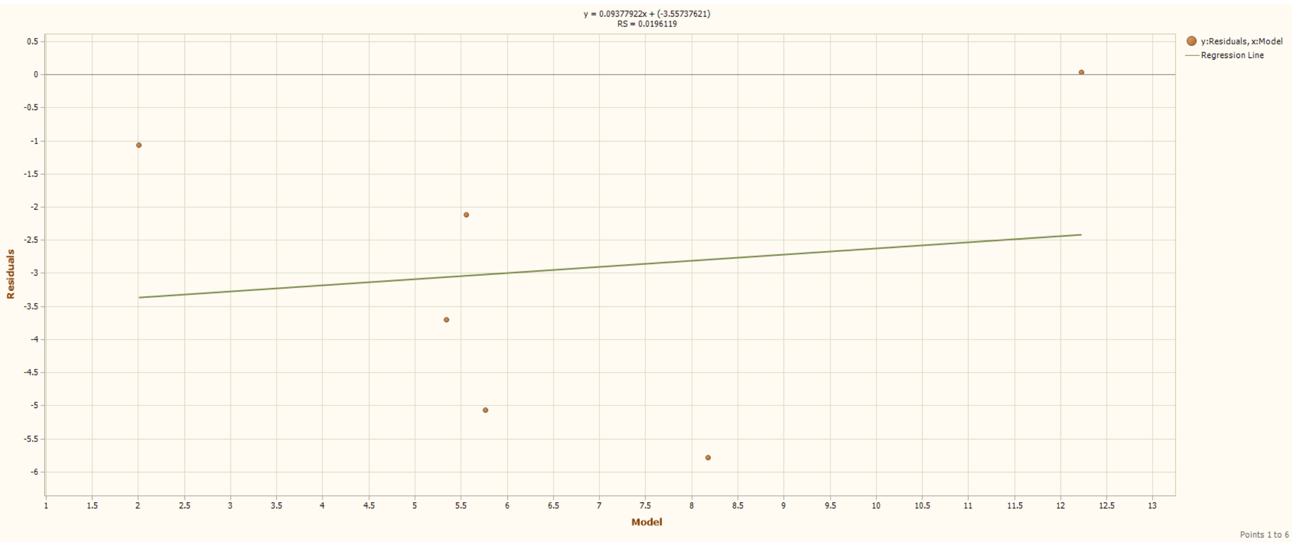
A7.2.4 Stacked distribution



A7.2.5 Scatter plot



A7.2.6 Residuals plot



A7.2.7 Data table

A7.2.8 Model performance

Fitness 216.750765540748

MSE 13.0580605500495

RMSE 3.61359385515991

MAE 2.95830392472734

RSE 0.818295208701081

RRSE 0.90459671052966

RAE 1.02078622876778

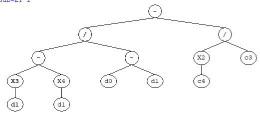
Correlation Coefficient 0.855144868746066

R-square 0.731272746542727

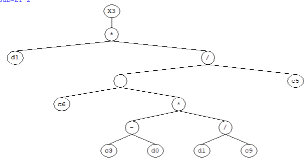
Calculation Errors 0

A7.3 Output model explained by Expression Tree

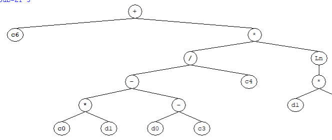
Sub-ET 1



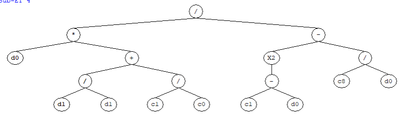
Sub-ET 2



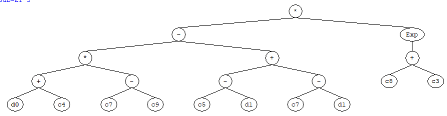
Sub-ET 3



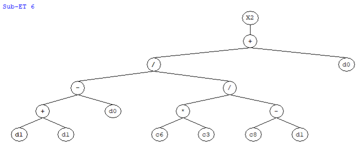
Sub-ET 4



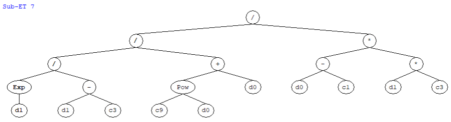
Sub-ET 5



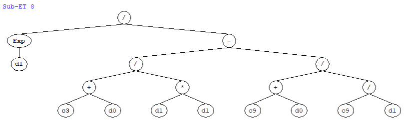
Sub-ET 6



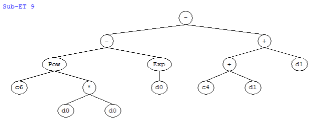
Sub-ET 7



Sub-ET 8



Sub-ET 9



A7.4 Output model explained by Matab

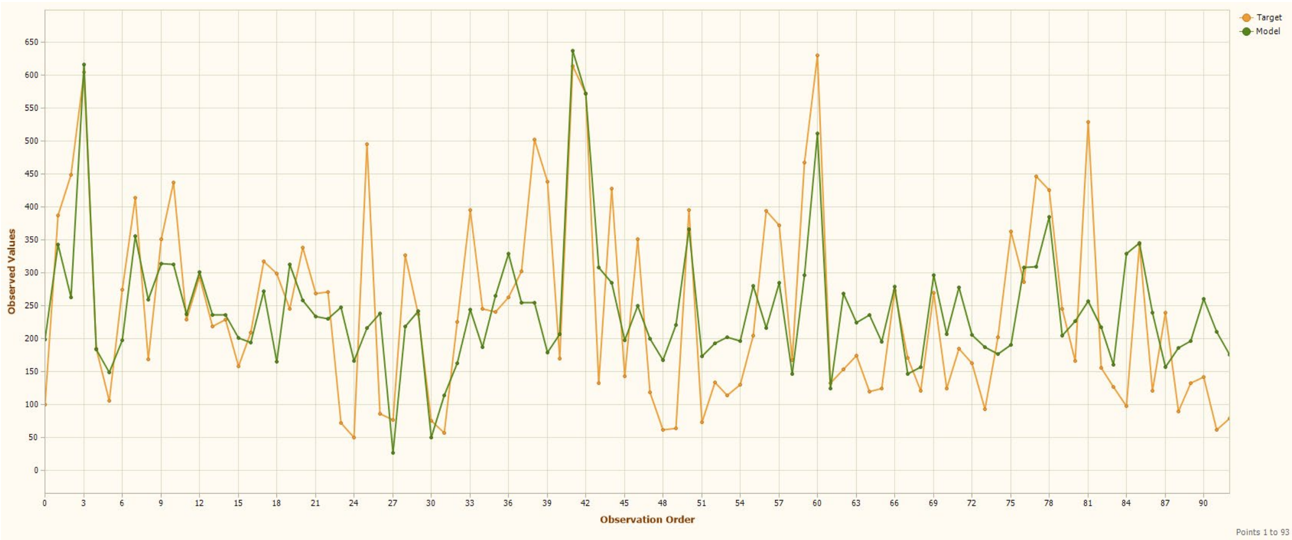
```
%------------------------------------------------------------------------
% Regression model generated by GeneXproTools 5.0 on 24/09/2020 2:21:35 AM
% GEP File: C:\Users\resha\Google Drive\Final GeneXPro\SOUTH\station 9\final\st09_5P_4N.gep
% Training Records:  93
% Validation Records:   6
% Fitness Function:  RMSE
% Training Fitness:  245.121703630077
% Training R-square: 0.518725281811437
% Validation Fitness:   216.750765540748
% Validation R-square:  0.731272746542727
%------------------------------------------------------------------------

function result = gepModel(d)

G1C3 = 1.20723919864344;
G1C4 = 0.791977515434463;
G2C5 = -6.26738181707205;
G2C6 = -6.87411275140917;
G2C3 = -8.66015096855453;
G2C9 = 5.53147984252449;
G3C6 = -9.74981514662645;
G3C4 = -5.43102850589641;
G3C0 = -5.18518381325724;
G3C3 = 0.771707784650686;
G4C8 = -5.71486425016022;
G4C1 = 3.08725400809015;
G4C0 = 6.65089877010407;
G5C8 = 0.614269386833822;
```

```
G5C3 = -2.62750350443245;

G5C4 = 8.19768777161619;

G5C7 = 7.47040886116749;

G5C9 = -5.00732147808073;

G5C5 = 7.86187322611164;

G6C6 = -3.89216430524613;

G6C3 = 0.669786237116415;

G6C8 = -0.86779060187744;

G7C1 = 1.04153176136133;

G7C3 = -7.4759170525861;

G7C9 = 4.58120204987008;

G8C3 = -1.28845608354311;

G8C9 = 2.40129516566367;

G9C6 = 1.21847203541812;

G9C4 = -3.14614135075788;


y = 0.0;


y = ((((d(2)^3)-(d(2)^4))/(d(1)-d(2)))-((G1C4^2)/G1C3));

y = y + ((d(2)*((G2C6-((G2C3-d(1))*(d(2)/G2C9)))/G2C5))^3);

y = y + (G3C6+((((G3C0*d(2))-(d(1)-G3C3))/G3C4)*reallog((d(2)*d(2)))));

y = y + ((d(1)*((d(2)/d(2))+(G4C1/G4C0)))/(((G4C1-d(1))^2)-(G4C8/d(1))));

y = y + ((((d(1)+G5C4)*(G5C7-G5C9))-((G5C5-d(2))+(G5C7-d(2))))*exp((G5C8+G5C3)));

y = y + (((((d(2)+d(2))-d(1))/((G6C6*G6C3)/(G6C8-d(2))))+d(1))^2);

y = y + (((exp(d(2))/(d(2)-G7C3))/(realpow(G7C9,d(1))+d(1)))/((d(1)-G7C1)*(d(2)*G7C3)));

y = y + (exp(d(2))/(((G8C3+d(1))/(d(2)*d(2)))-((G8C9+d(1))/(G8C9/d(2)))));

y = y + ((realpow(G9C6,(d(1)*d(1)))-exp(d(1)))-((G9C4+d(2))+d(2)));


result = y;
```

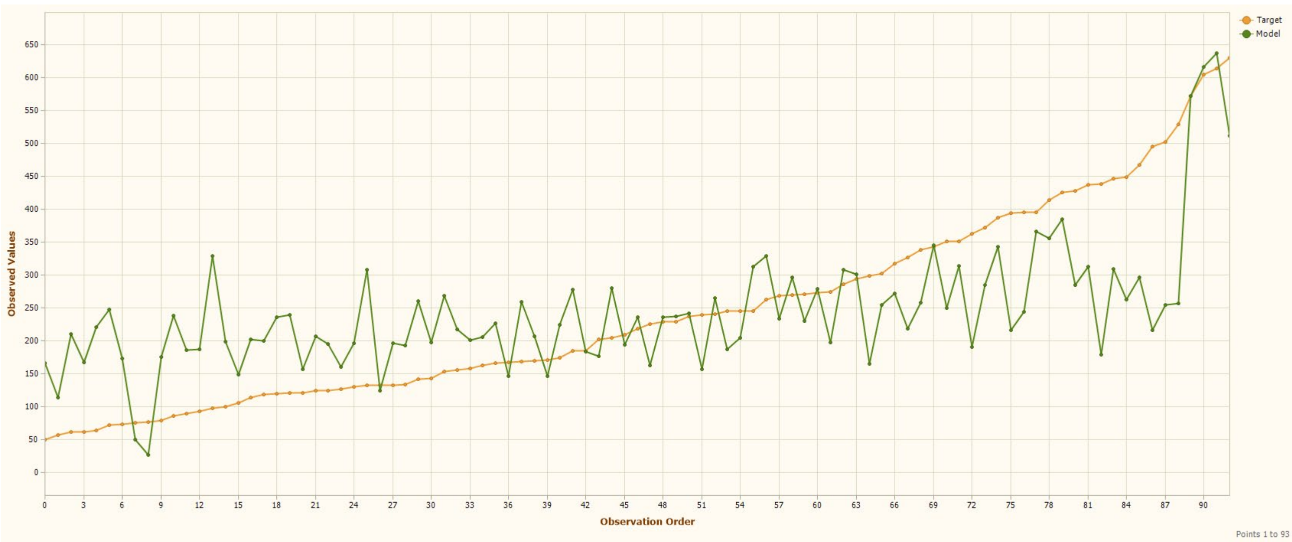
A8 Best GEP Model Output for Corowa (CWNSW)

A8.1 Output results for Calibration period

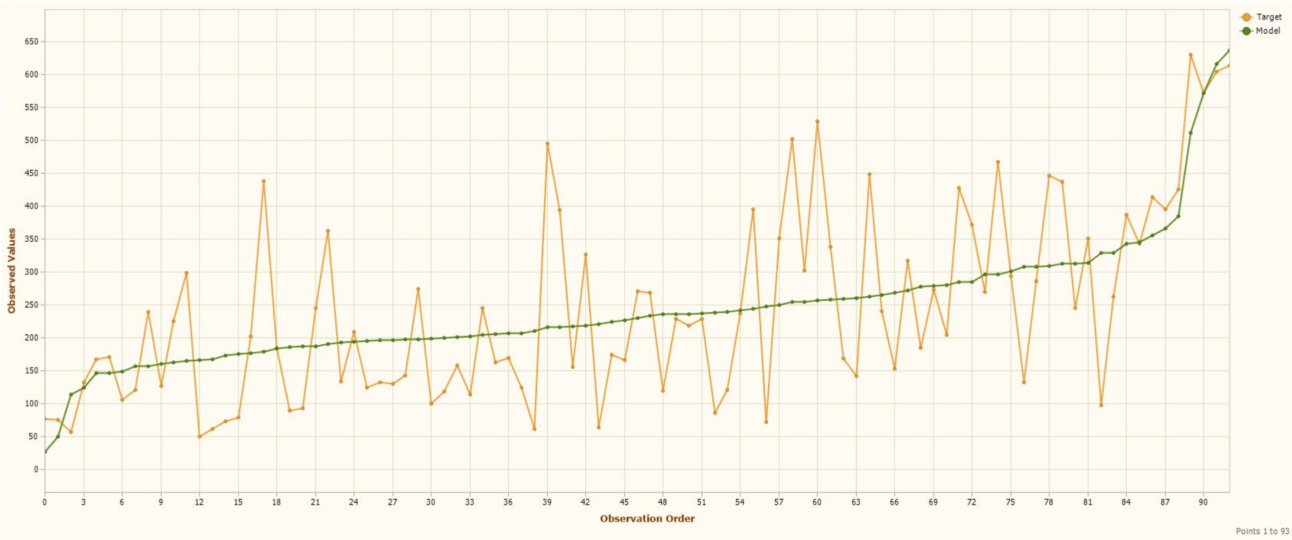
A8.1.1 Curve Fitting



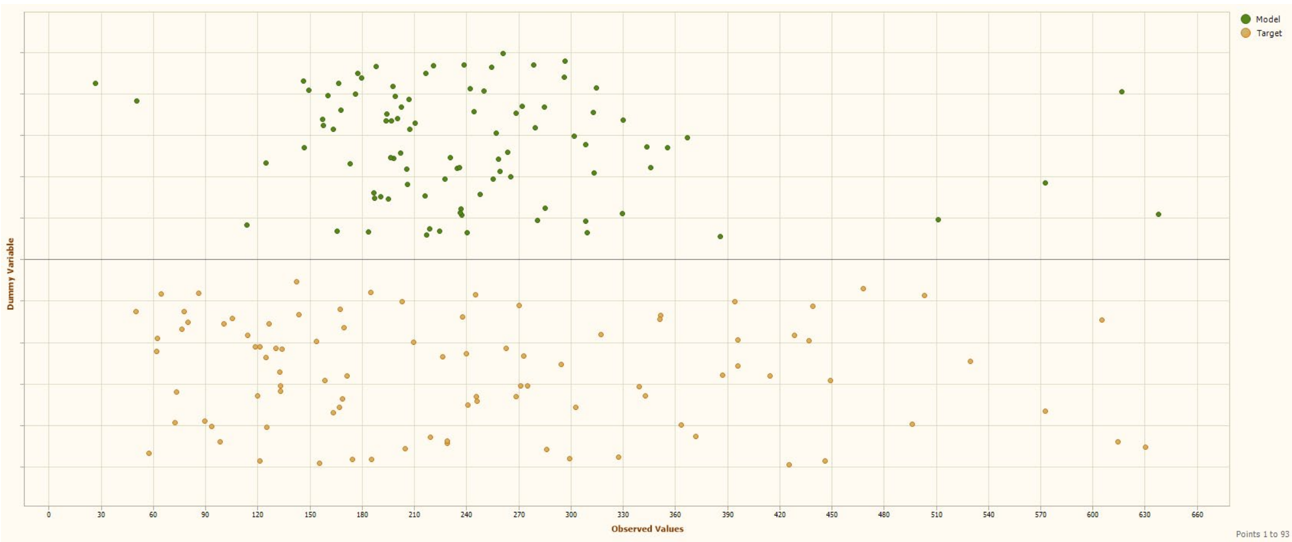
A8.1.2 Target sorted fitting



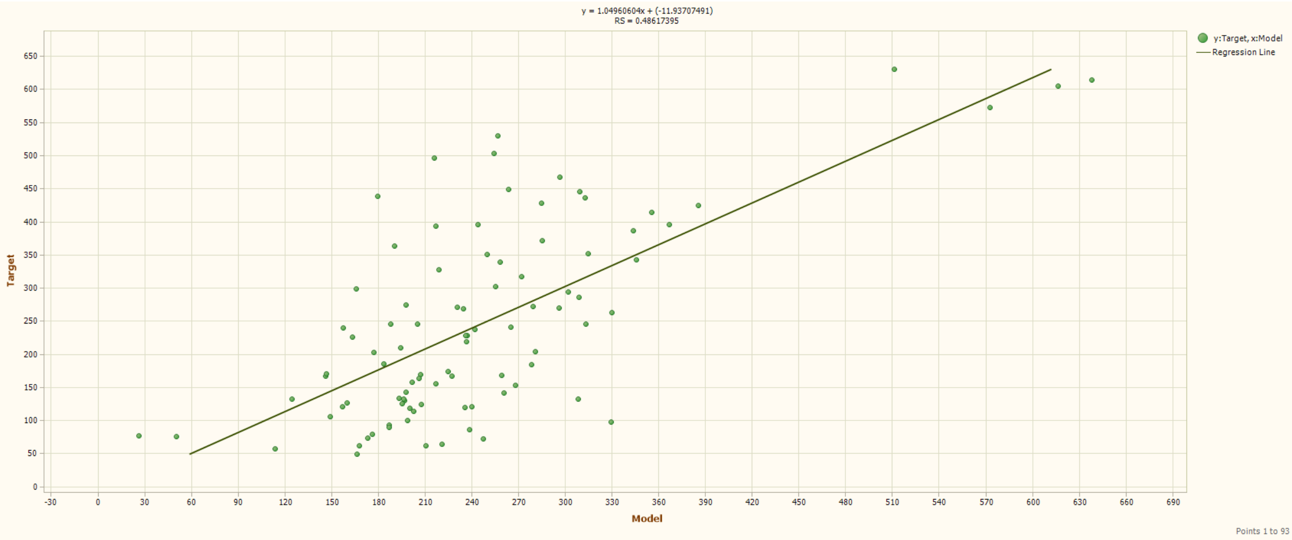
A8.1.3 Model sorted fitting



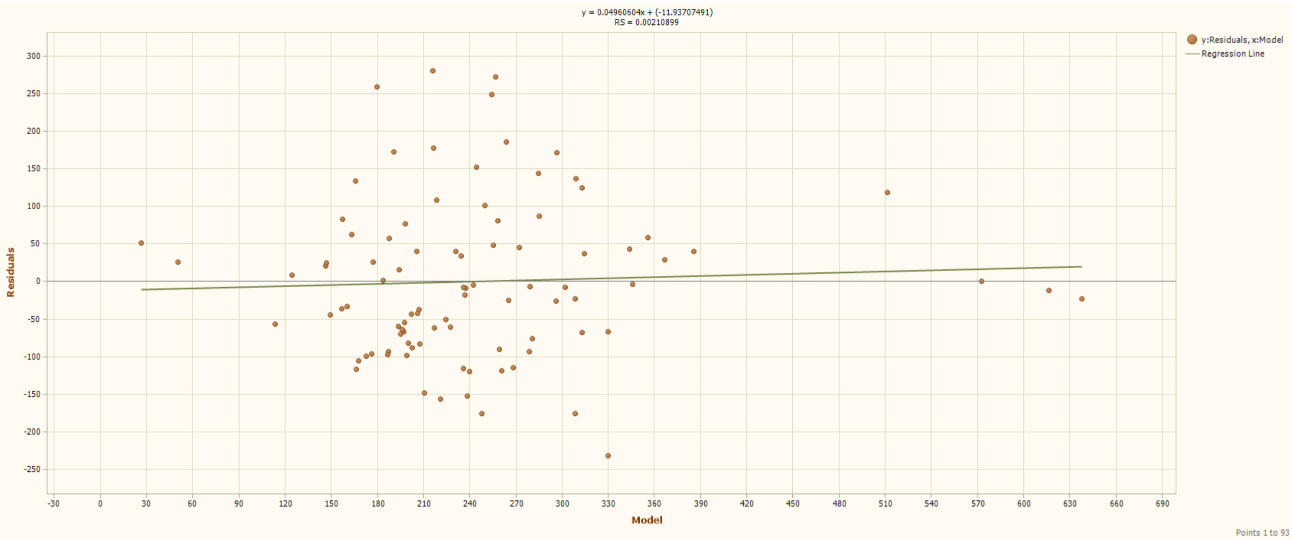
A8.1.4 Stacked distribution



A8.1.5 Scatter plot



A8.1.6 Residuals plot



A8.1.7 Model performance

Fitness 9.52180910046823

MSE 10820.5888211732

RMSE 104.022059300771

MAE 82.018262730314

RSE 0.514913761593019

RRSE 0.717574917059549

RAE 0.687624460698335

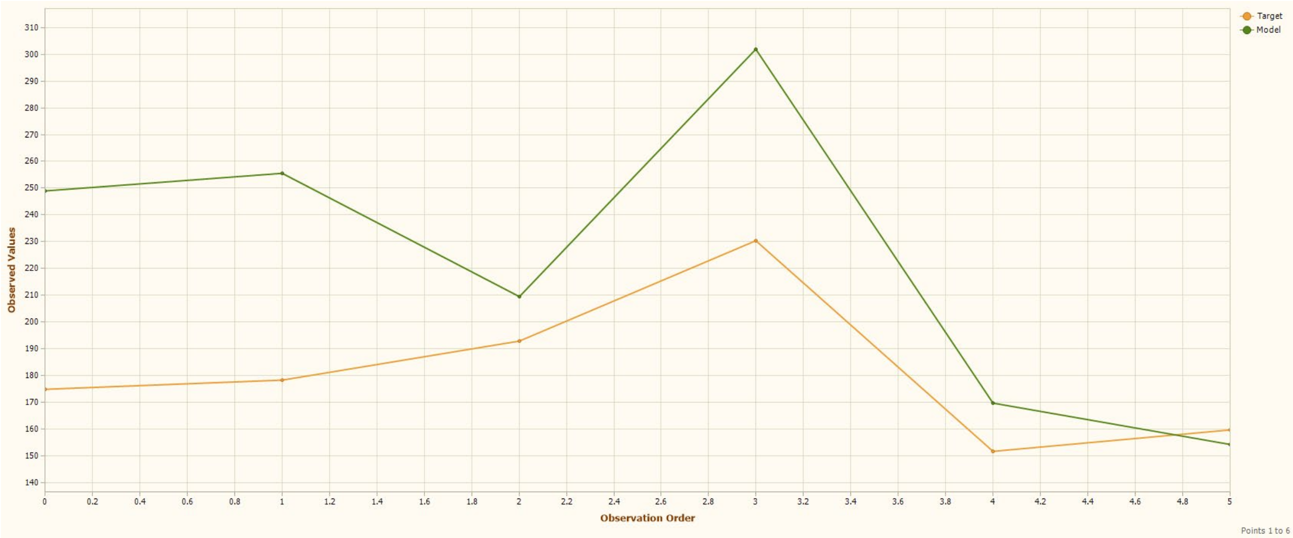
Correlation Coefficient 0.697261754763912

R-square 0.486173954656449

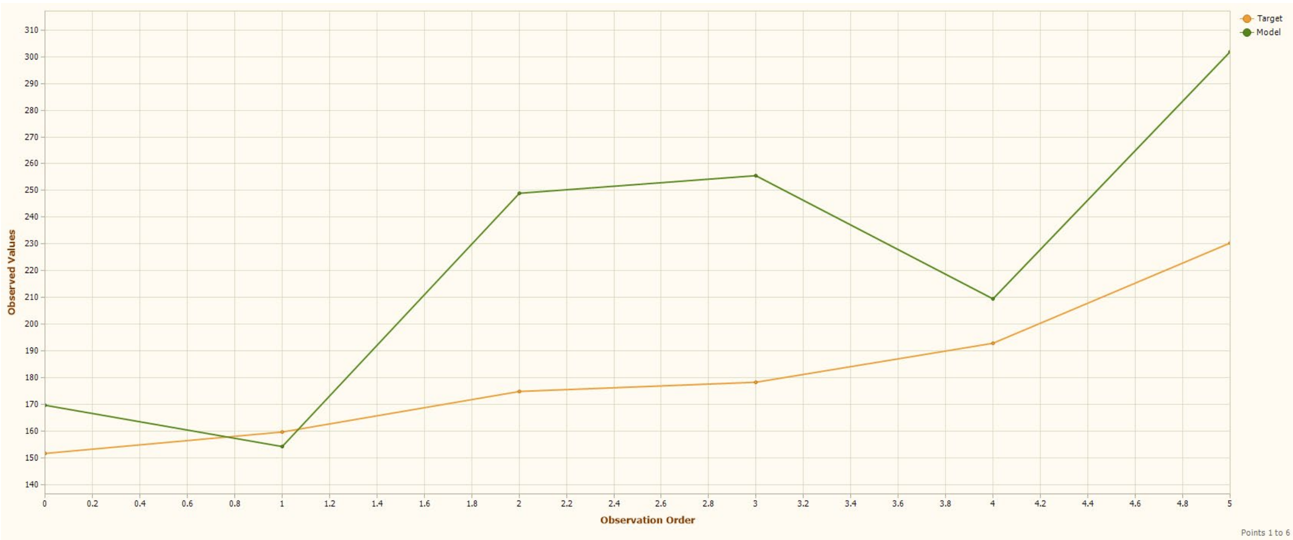
Calculation Errors 0

A8.2 Output results for Validation period

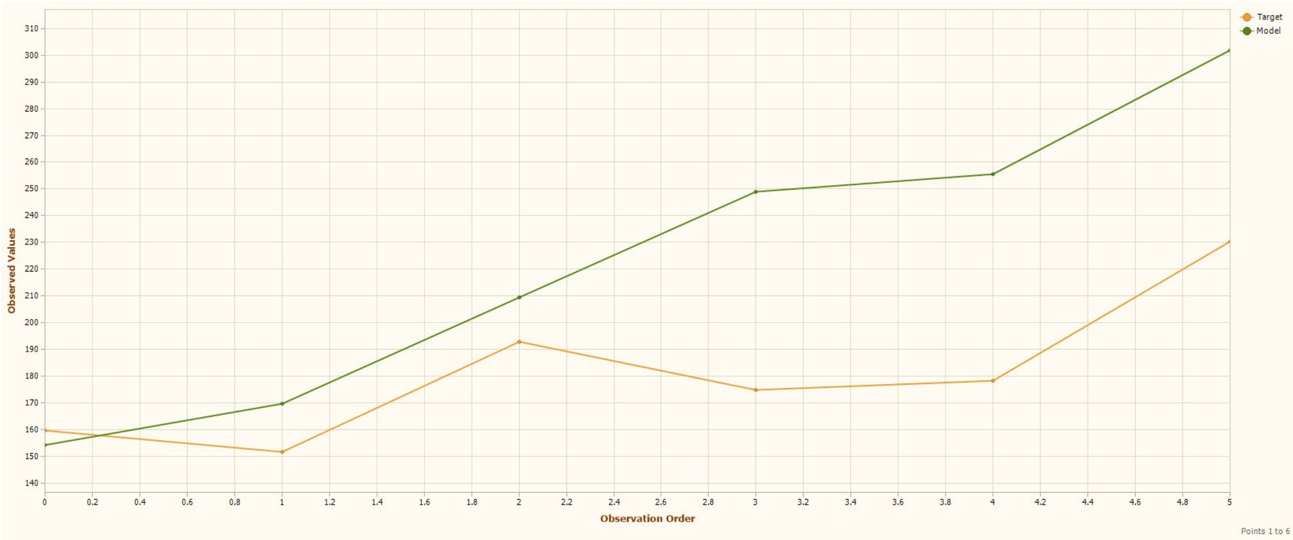
A8.2.1 Curve fitting



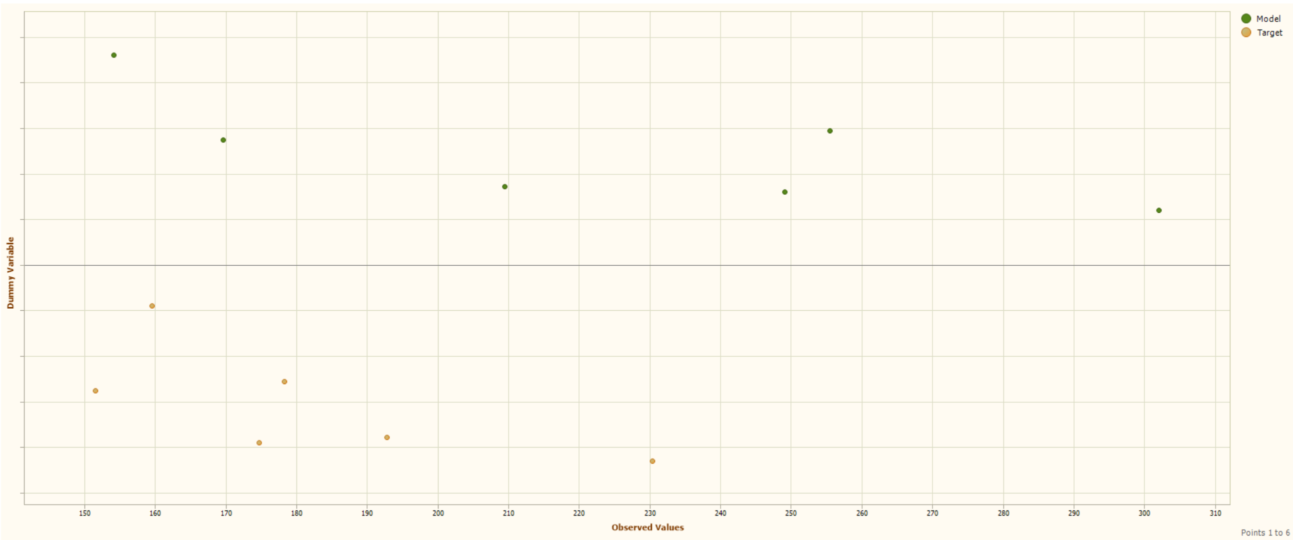
A8.2.2 Target sorted fitting



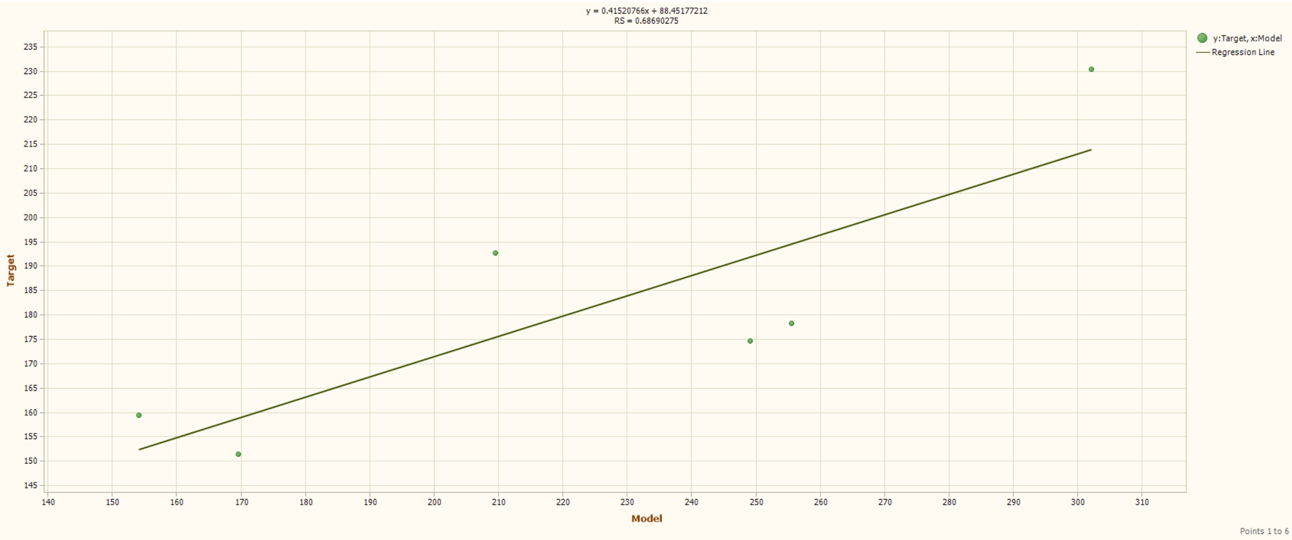
A8.2.3 Model sorted fitting



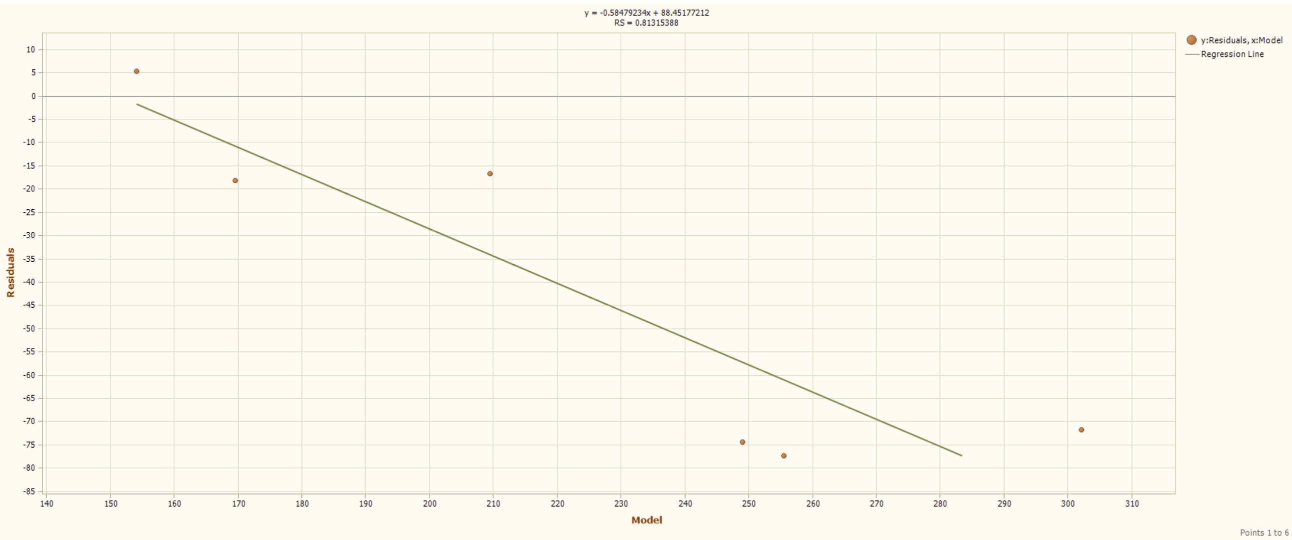
A8.2.4 Stacked distribution



A8.2.5 Scatter plot



A8.2.6 Residuals plot



A8.2.7 Model performance

Fitness 18.2920856569624

MSE 2880.30270065817

RMSE 53.6684516327625

MAE 43.9136605379604

RSE 4.36909613978975

RRSE 2.0902382973694

RAE 2.16698698823021

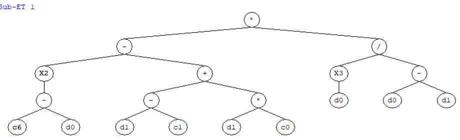
Correlation Coefficient 0.828795964705327

R-square 0.686902751111833

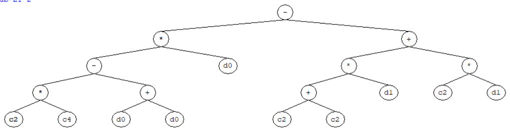
Calculation Errors 0

A8.3 Output model explained by Expression Tree

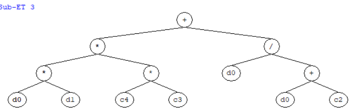
Sub-ET 1



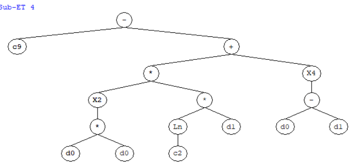
Sub-ET 2



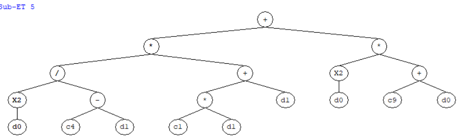
Sub-ET 3



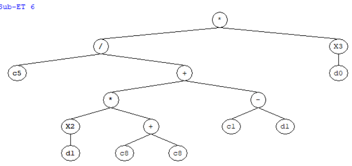
Sub-ET 4



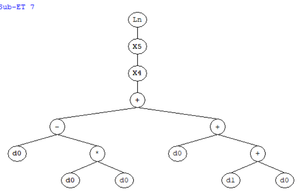
Sub-ET 5



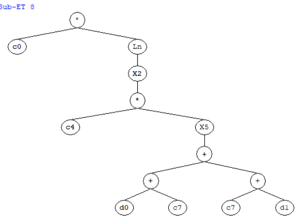
Sub-ET 6



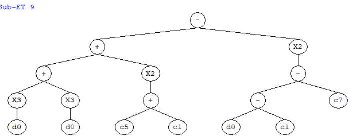
Sub-ET 7



Sub-ET 8



Sub-ET 9



A8.4 Output model explained by Matab

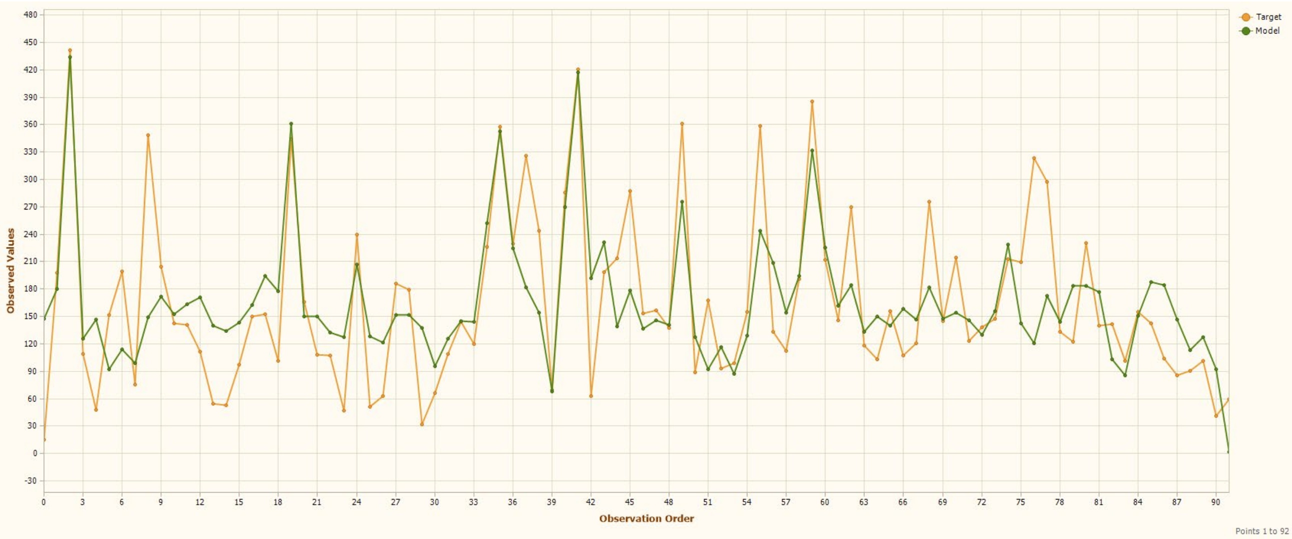
```
%------------------------------------------------------------------------
% Regression model generated by GeneXproTools 5.0 on 24/09/2020 2:30:44 AM
% GEP File: C:\Users\resha\Google Drive\Final GeneXPro\Central west\station 3\final\st03_3P_3N.gep
% Training Records:  93
% Validation Records:   6
% Fitness Function:  RMSE
% Training Fitness:  9.52180910046823
% Training R-square: 0.486173954656449
% Validation Fitness:   18.2920856569624
% Validation R-square:  0.686902751111833
%------------------------------------------------------------------------

function result = gepModel(d)

G1C6 = 3.37896986487187;
G1C1 = 10.847095819483;
G1C0 = 6.49540459303568;
G2C2 = -9.90341770840747;
G2C4 = -6.89046153447066;
G3C4 = 4.75857139805292;
G3C3 = 9.38901944029054;
G3C2 = 2.38013617389447;
G4C9 = 9.03866695150609;
G4C2 = 6.36856604969415;
G5C9 = -11.3406222017771;
G5C4 = 8.47246768082645e-02;
G5C1 = 12.8328771735375;
G6C5 = 1.19494078087582;
```

```
G6C1 = -9.59303262428663;

G6C8 = 9.8839888337962;

G8C0 = -5.68942911482263;

G8C4 = -2.11672314649688;

G8C7 = 0.518762779625843;

G9C7 = 5.81167897692779;

G9C5 = -9.54465429395428;

G9C1 = -5.19015516575132;


y = 0.0;


y = ((((G1C6-d(1))^2)-((d(2)-G1C1)+(d(2)*G1C0)))*((d(1)^3)/(d(1)-d(2))));

y = y + ((((G2C2*G2C4)-(d(1)+d(1)))*d(1))-(((G2C2+G2C2)*d(2))+(G2C2*d(2))));

y = y + (((d(1)*d(2))*(G3C4*G3C3))+(d(1)/(d(1)+G3C2)));

y = y + (G4C9-((((d(1)*d(1))^2)*(reallog(G4C2)*d(2)))+((d(1)-d(2))^4)));

y = y + ((((d(1)^2)/(G5C4-d(2)))*((G5C1*d(2))+d(2)))+((d(1)^2)*(G5C9+d(1))));

y = y + ((G6C5/(((d(2)^2)*(G6C8+G6C8))+(G6C1-d(2))))*(d(1)^3));

y = y + reallog(((((d(1)-(d(1)*d(1)))+(d(1)+(d(2)+d(1))))^4)^5));

y = y + (G8C0*reallog(((G8C4*(((d(1)+G8C7)+(G8C7+d(2)))^5))^2)));

y = y + ((((d(1)^3)+(d(1)^3))+((G9C5+G9C1)^2))-(((d(1)-G9C1)-G9C7)^2));


result = y;
```

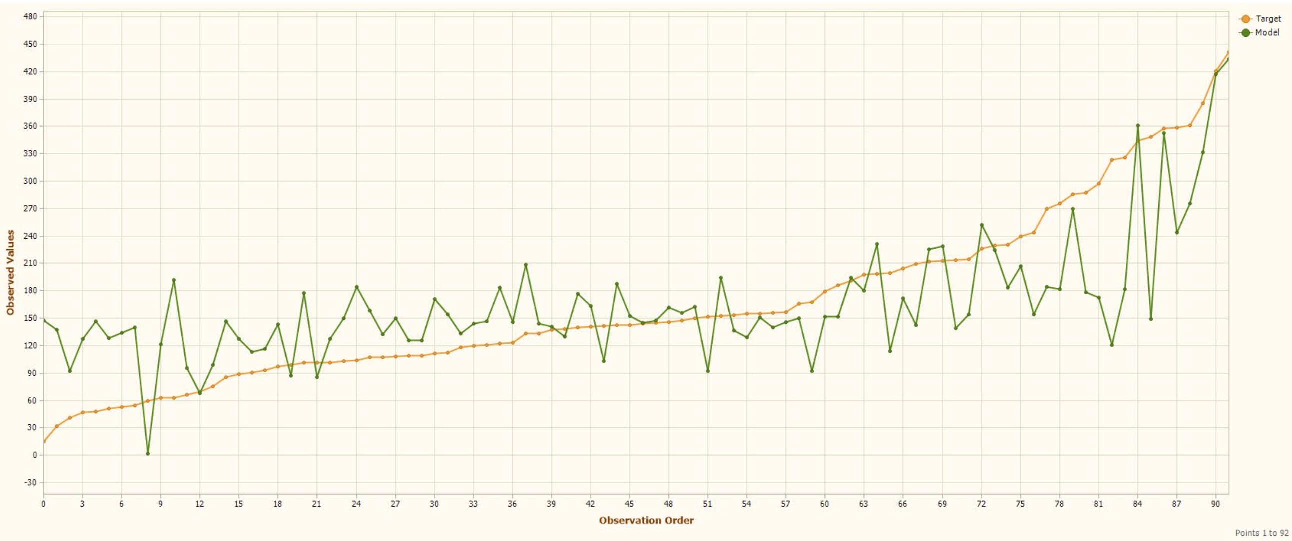
A9 Best GEP Model Output for Wagga Wagga (CWNSW)

A9.1 Output results for Calibration period

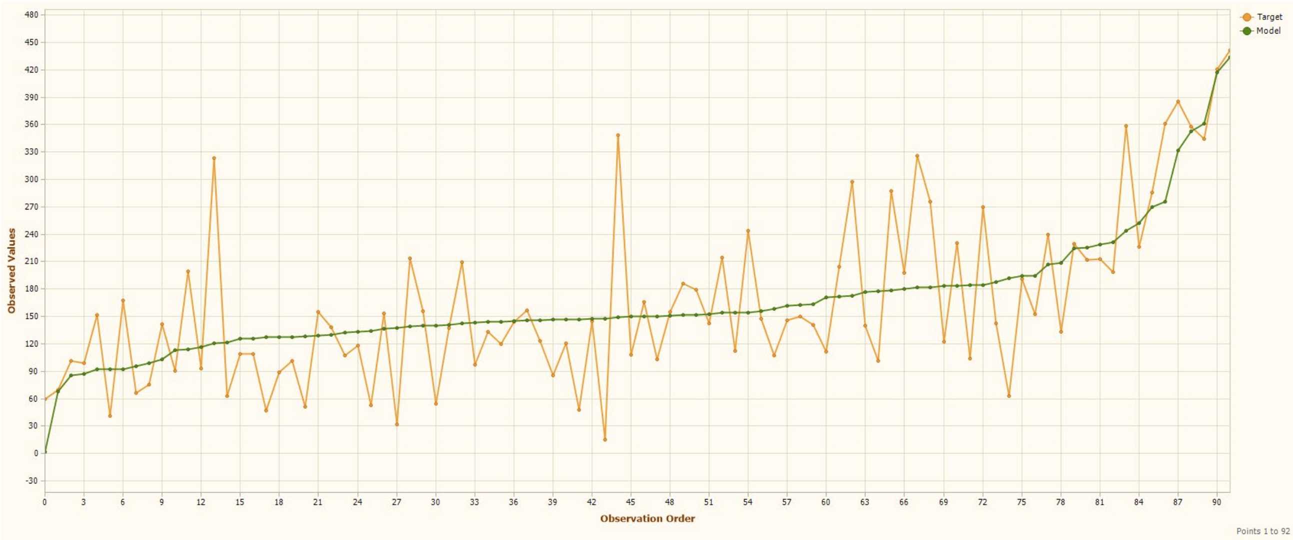
A9.1.1 Curve Fitting



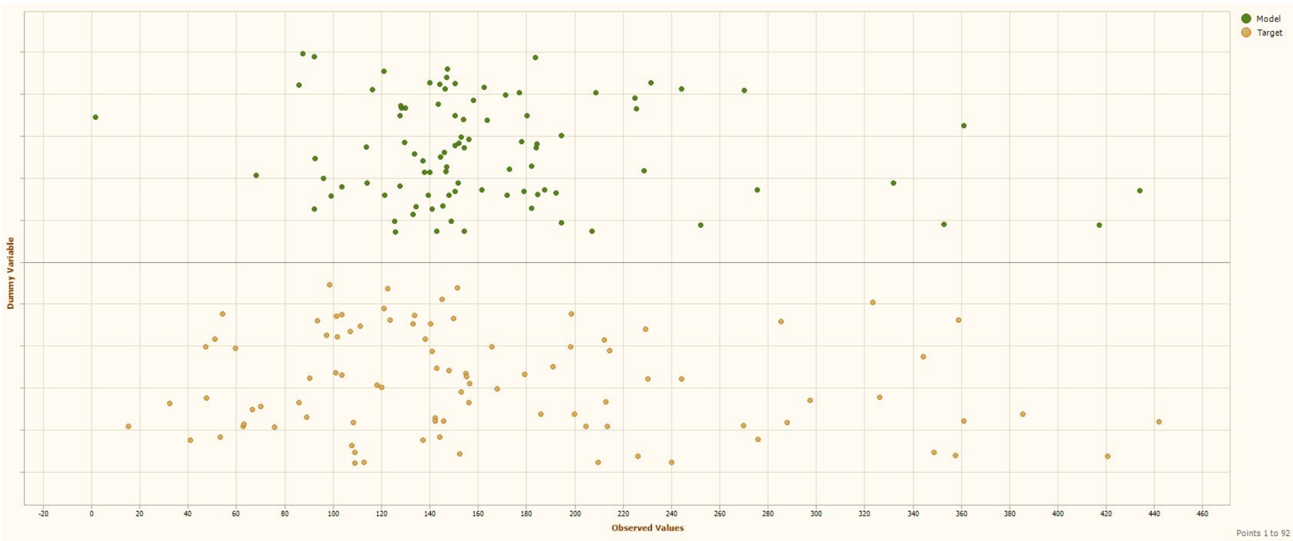
A9.1.2 Target sorted fitting



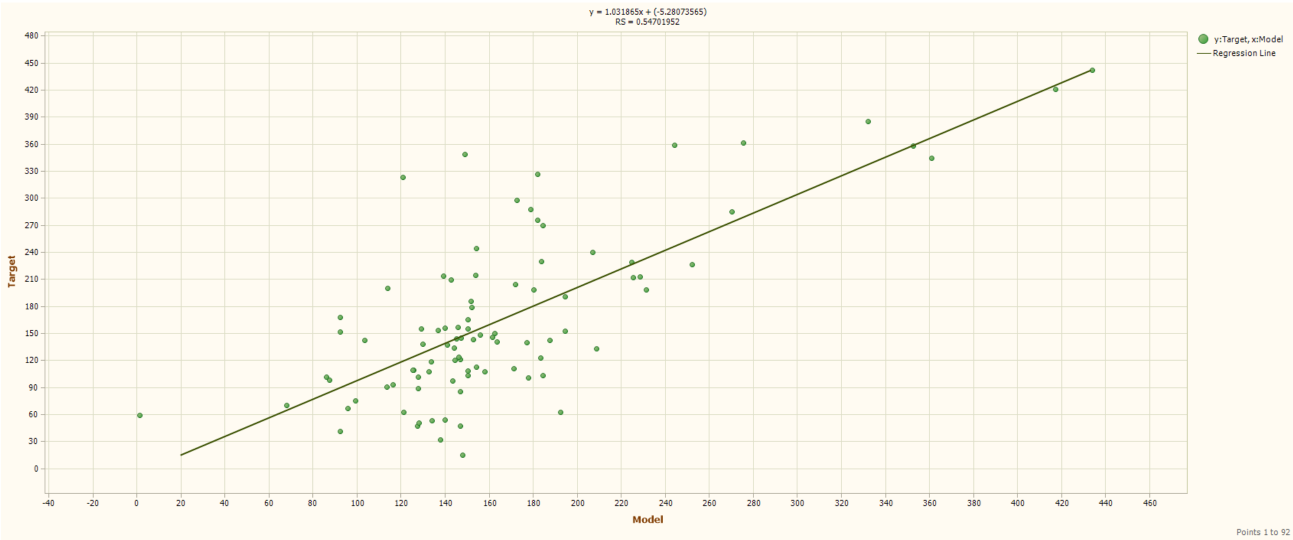
A9.1.3 Model sorted fitting



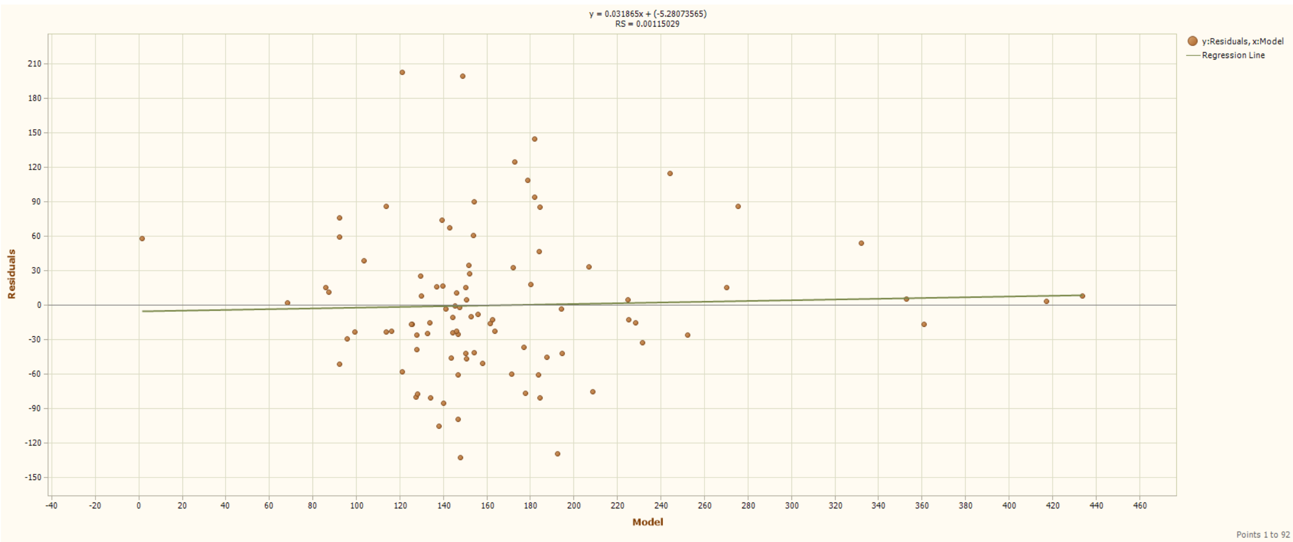
A9.1.4 Stacked distribution



A9.1.5 Scatter plot



A9.1.6 Residuals plot

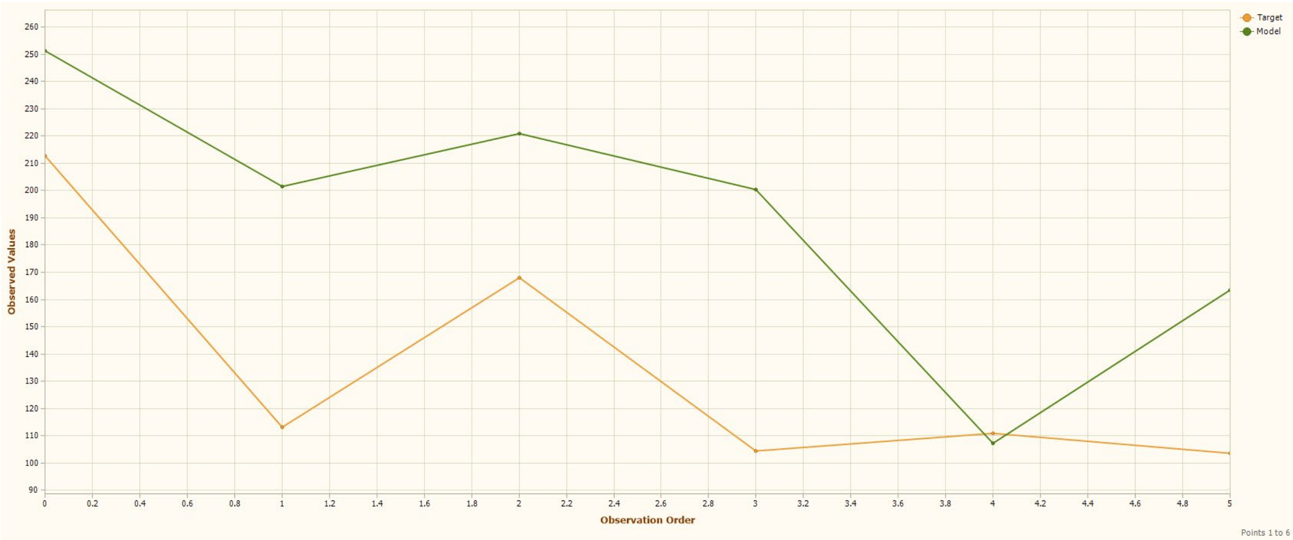


A9.1.7 Model performance

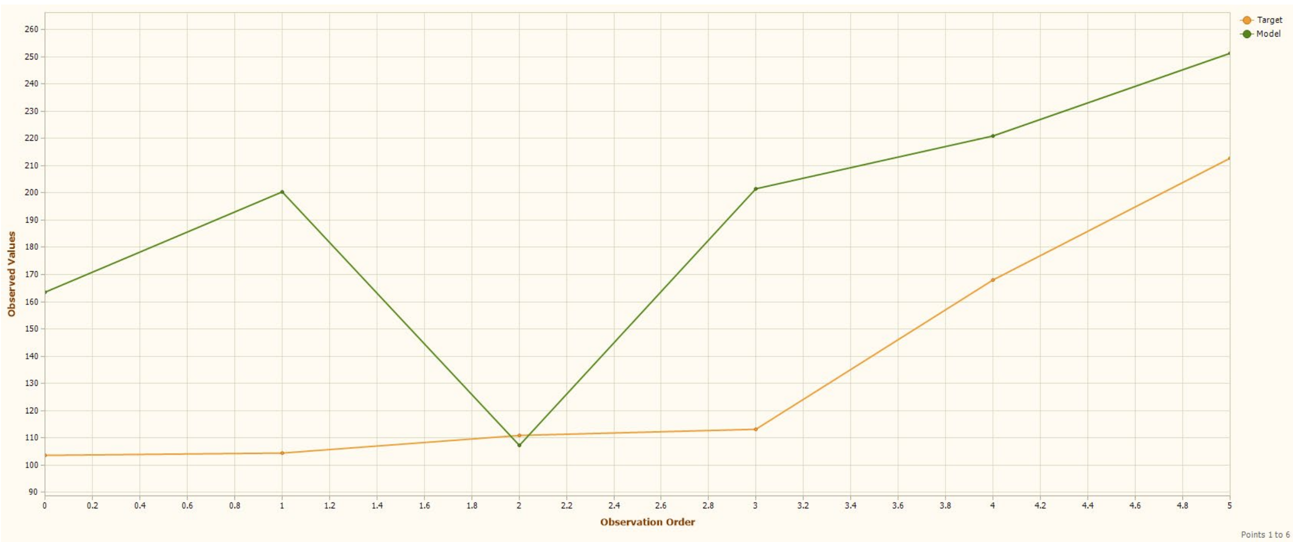
Fitness	15.6363904460653
MSE	3963.12789217514
RMSE	62.9533787192962
MAE	47.2664003345872
RSE	0.453502135819371
RRSE	0.673425672082206
RAE	0.644426774687492
Correlation Coefficient	0.739607680577755
R-square	0.547019521169606
Calculation Errors	0

A9.2 Output results for Validation period

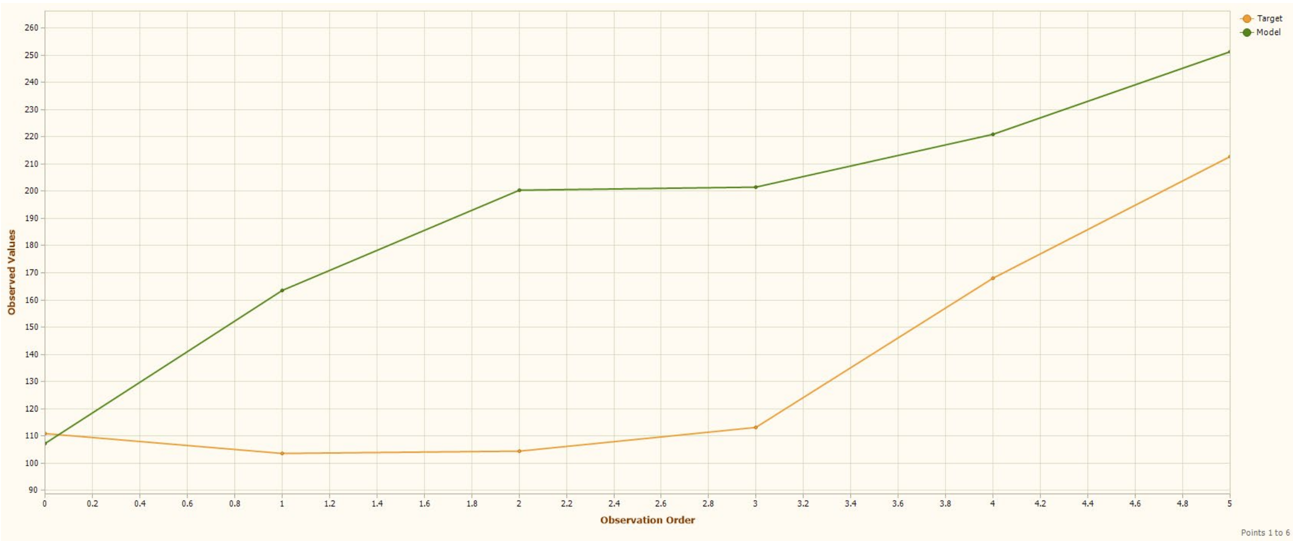
A9.2.1 Curve fitting



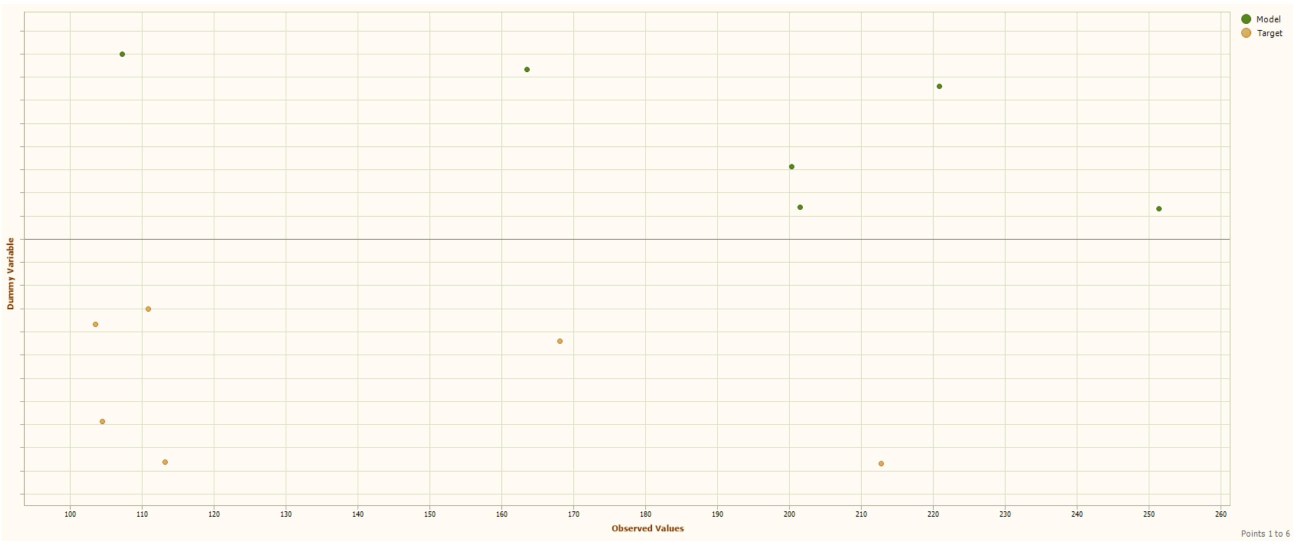
A9.2.2 Target sorted fitting



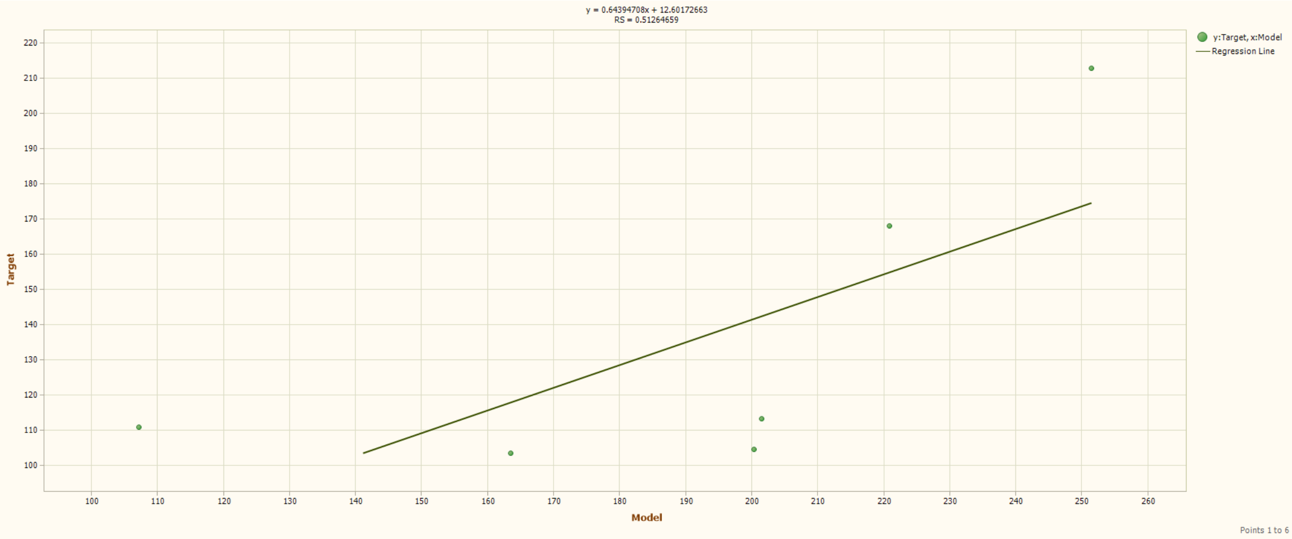
A9.2.3 Model sorted fitting



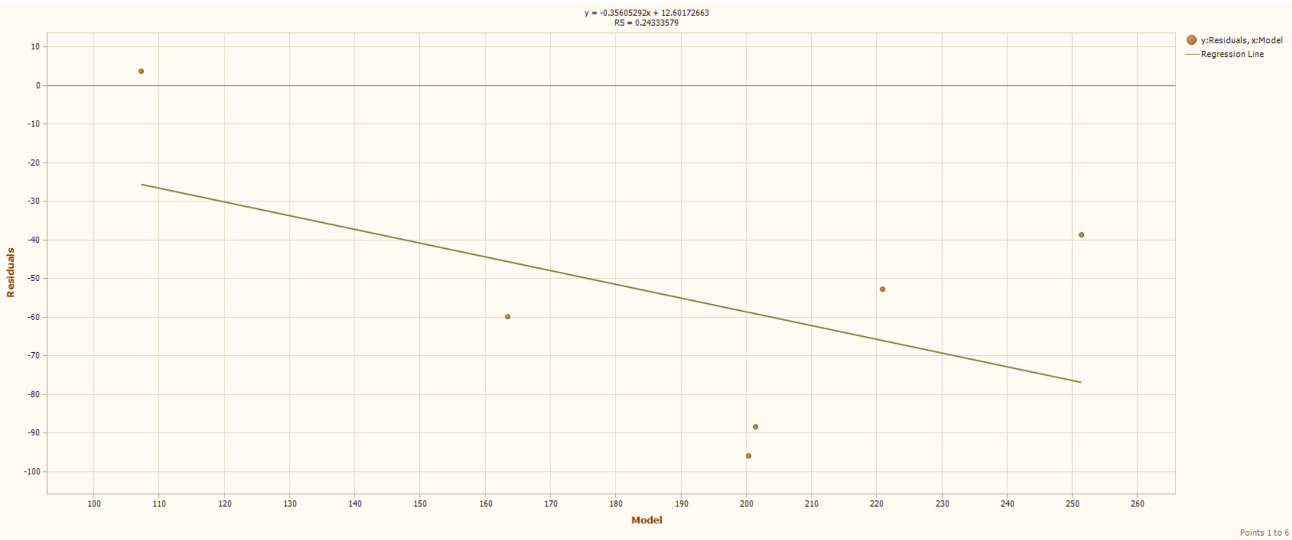
A9.2.4 Stacked distribution



A9.2.5 Scatter plot



A9.2.6 Residuals plot

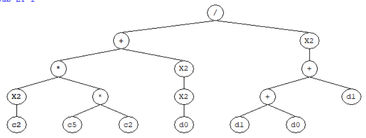


A9.2.7 Model performance

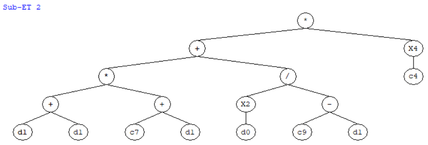
Fitness	15.2892709736612
MSE	4148.04810372354
RMSE	64.405342198637
MAE	56.5522967267102
RSE	2.45933621905974
RRSE	1.56822709422447
RAE	1.54422414841299
Correlation Coefficient	0.715993428236189
R-square	0.51264658927741
Calculation Errors	0

A9.3 Output model explained by Expression Tree

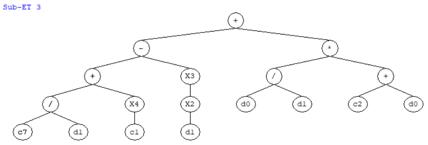
Sub-ET 1



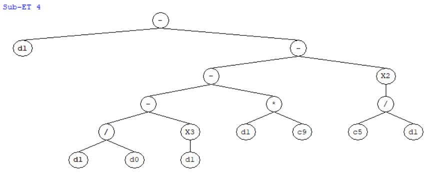
Sub-ET 2



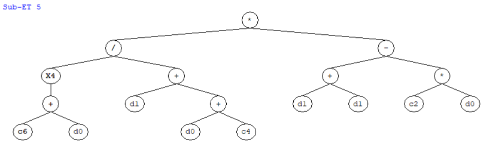
Sub-ET 3



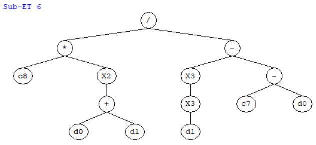
Sub-ET 4



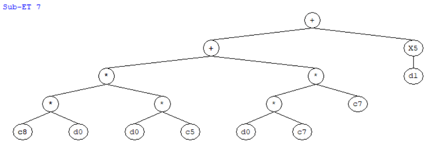
Sub-ET 5



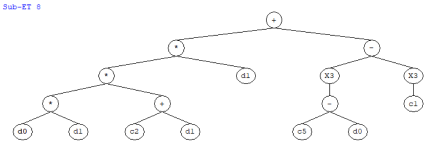
Sub-ET 6



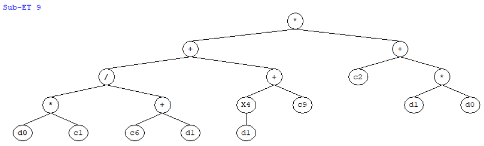
Sub-ET 7



Sub-ET 8



Sub-ET 9



A9.4 Output model explained by MATLAB

```
%-----------------------------------------------------------------------
% Regression model generated by GeneXproTools 5.0 on 24/09/2020 2:56:16 AM
% GEP File: C:\Users\resha\Google Drive\Final GeneXPro\Central west\station 5\st05_2E_IPO.gep
% Training Records:  92
% Validation Records:   6
% Fitness Function:  RMSE
% Training Fitness:  15.6363904460653
% Training R-square: 0.547019521169606
% Validation Fitness:   15.2892709736612
% Validation R-square:  0.51264658927741
%-----------------------------------------------------------------------

function result = gepModelQ5C(d_string)

G1C2 = -0.298172411175445;
G1C5 = 0.613361938676702;
G2C4 = -1.90283104948013;
G2C7 = -1.13646370064301;
G2C9 = -1.6009898925557;
G3C2 = 9.70066394064929;
G3C7 = 0.380586425608606;
G3C1 = 3.03491513537159;
G4C9 = 15.5187510405182;
G4C5 = -0.352933693749839;
G5C2 = 1.53348984827905;
G5C6 = 1.28141315990846;
G5C4 = -0.538233887894743;
```

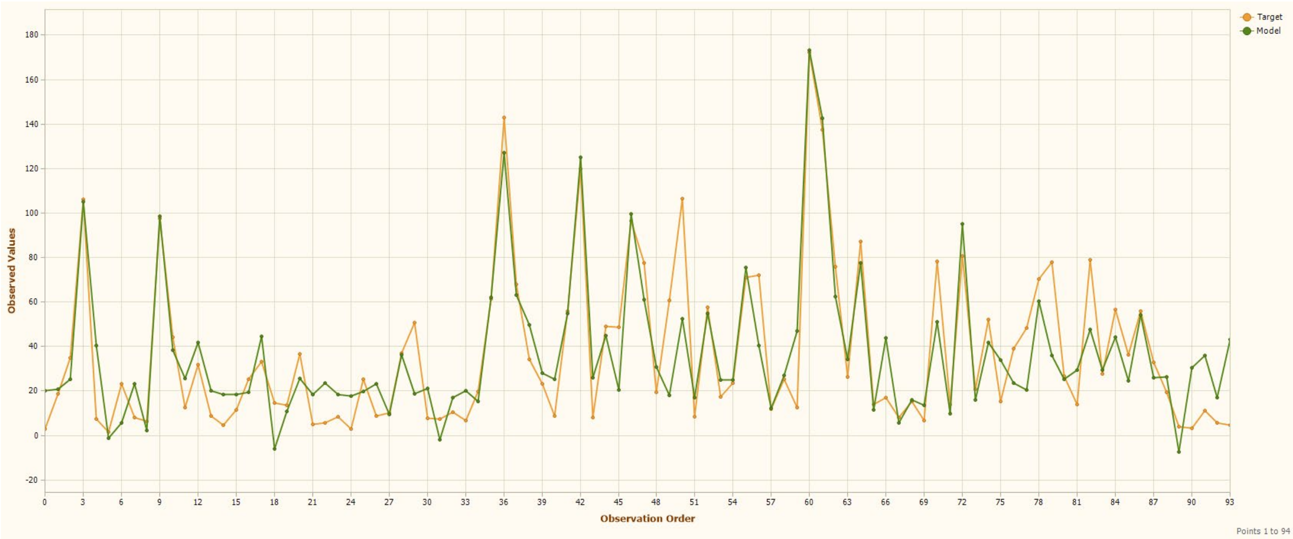
```
G6C8 = -1.66908470096463;

G6C7 = -7.91313823709688e-02;

G7C7 = 7.68558178745121;

G7C8 = -7.37819598637263;

G7C5 = 7.82870724703059;

G8C1 = -5.60473798058316;

G8C2 = -17.1379609355451;

G8C5 = -4.9572811128059;

G9C2 = 0.159740954202802;

G9C9 = -6.93649408443321;

G9C1 = 35.4234934384307;

G9C6 = 0.826136051515244;


Two_EC = 1;

All_IPOC = 2;


d = TransformCategoricalInputs(d_string);


y = 0.0;


y = ((((G1C2^2)*(G1C5*G1C2))+((d(Two_EC)^2)^2))/(((d(All_IPOC)+d(Two_EC))+d(All_IPOC))^2));

y = y + ((((d(All_IPOC)+d(All_IPOC))*(G2C7+d(All_IPOC)))+((d(Two_EC)^2)/(G2C9-d(All_IPOC))))*(G2C4^4));

y = y + ((((G3C7/d(All_IPOC))+(G3C1^4))-
((d(All_IPOC)^2)^3))+((d(Two_EC)/d(All_IPOC))*(G3C2+d(Two_EC))));

y = y + (d(All_IPOC)-((((d(All_IPOC)/d(Two_EC))-(d(All_IPOC)^3))-(d(All_IPOC)*G4C9))-
((G4C5/d(All_IPOC))^2)));

y = y + ((((G5C6+d(Two_EC))^4)/(d(All_IPOC)+(d(Two_EC)+G5C4)))*((d(All_IPOC)+d(All_IPOC))-
(G5C2*d(Two_EC))));

y = y + ((G6C8*((d(Two_EC)+d(All_IPOC))^2))/(((d(All_IPOC)^3)^3)-(G6C7-d(Two_EC))));

y = y + ((((G7C8*d(Two_EC))*(d(Two_EC)*G7C5))+((d(Two_EC)*G7C7)*G7C7))+(d(All_IPOC)^5));

y = y + ((((d(Two_EC)*d(All_IPOC))*(G8C2+d(All_IPOC)))*d(All_IPOC))+(((G8C5-d(Two_EC))^3)-(G8C1^3)));
```

```
y = y +
((((d(Two_EC)*G9C1)/(G9C6+d(All_IPOC)))+((d(All_IPOC)^4)+G9C9))*(G9C2+(d(All_IPOC)*d(Two_EC))));


result = y;


function outputData = TransformCategoricalInputs(inputData)

outputData(1) = str2double(inputData(1));

switch char(inputData(2))

   case ''

      outputData(2) = 0.110986439891304;

   otherwise

      outputData(2) = str2double(inputData(2));

end
```

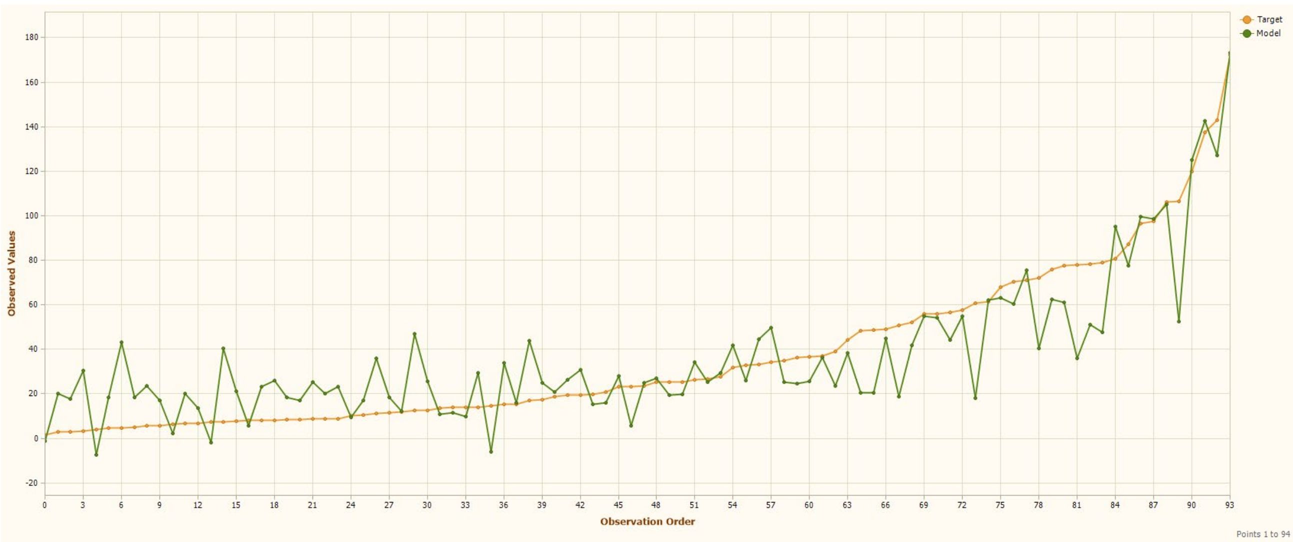
A10 Best GEP Model Output for Cowra Station (CWNSW)

A10.1 Output results for Calibration period

A10.1.1 Curve Fitting

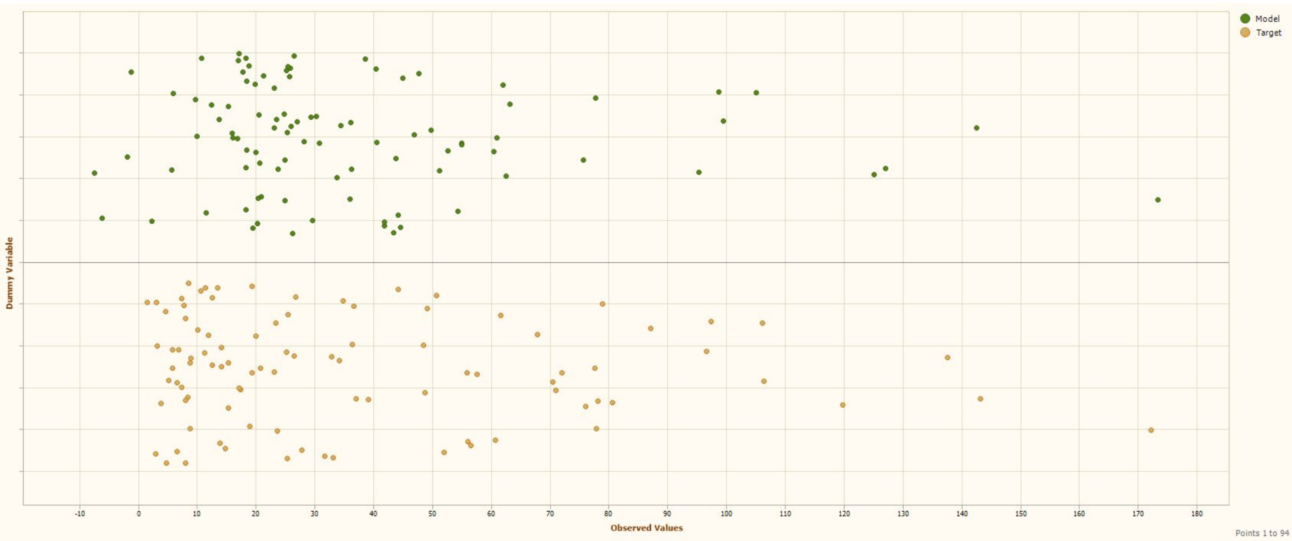


A10.1.2 Target sorted fitting

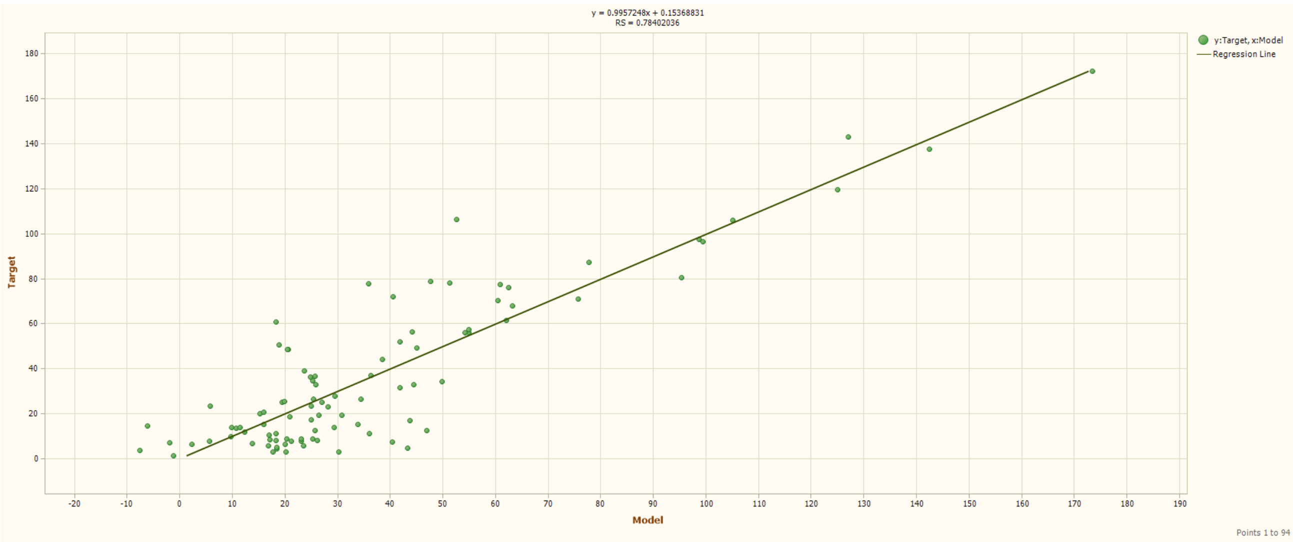


A10.1.3 Model sorted fitting

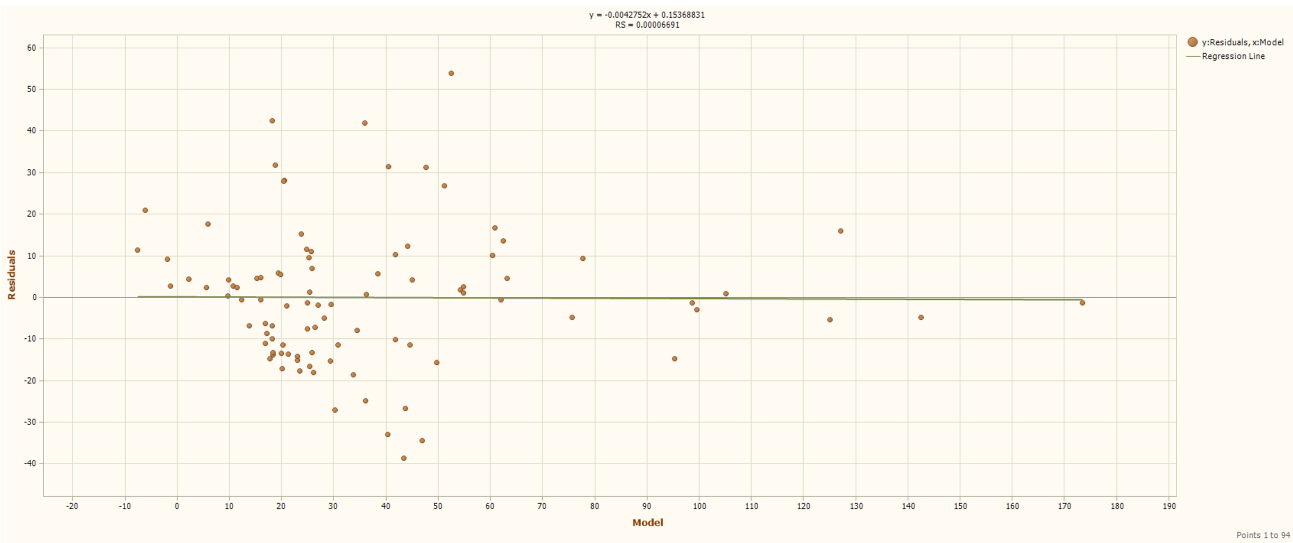
A10.1.4 Stacked distribution



A10.1.5 Scatter plot



A10.1.6 Residuals plot



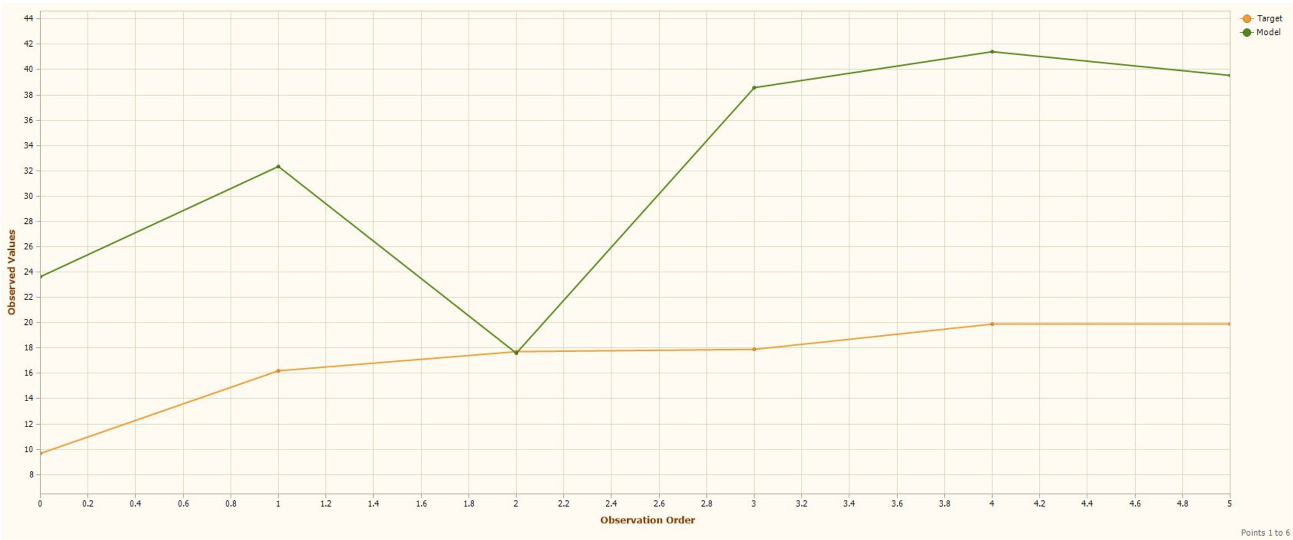
A10.1.7 Model performance

Fitness	57.1283554543532
MSE	272.396614625986
RMSE	16.5044422694615
MAE	12.3510036917747
RSE	0.215994105020991
RRSE	0.464751659513972
RAE	0.440590662415926
Correlation Coefficient	0.885449241961496
R-square	0.784020360090188
Calculation Errors	0

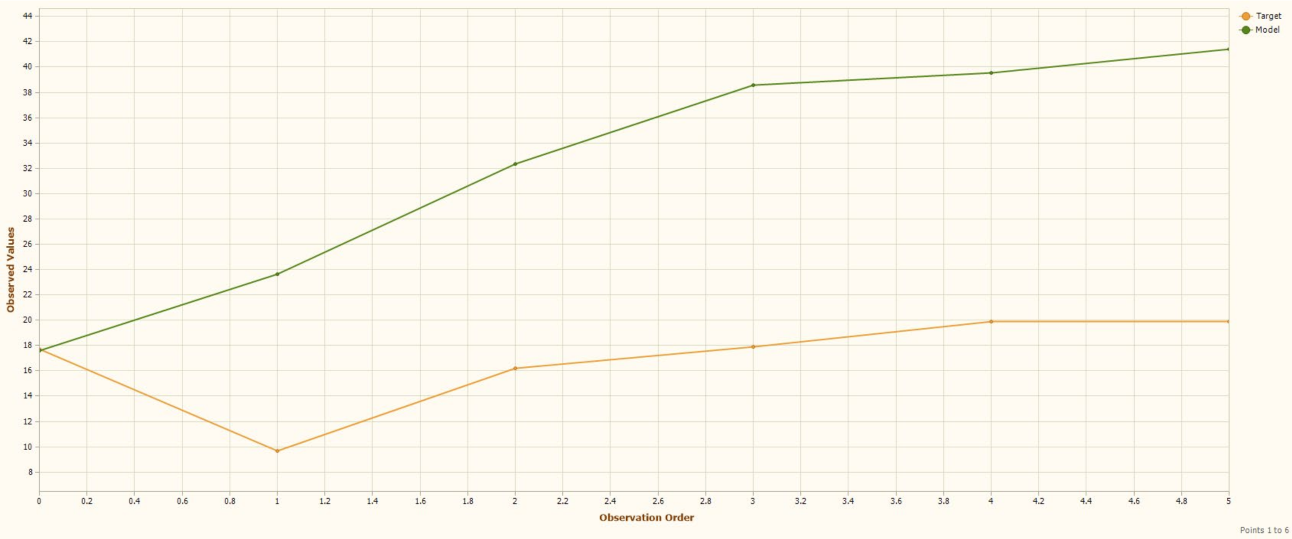
A10.2 Output results for Validation period

A10.2.1 Curve fitting

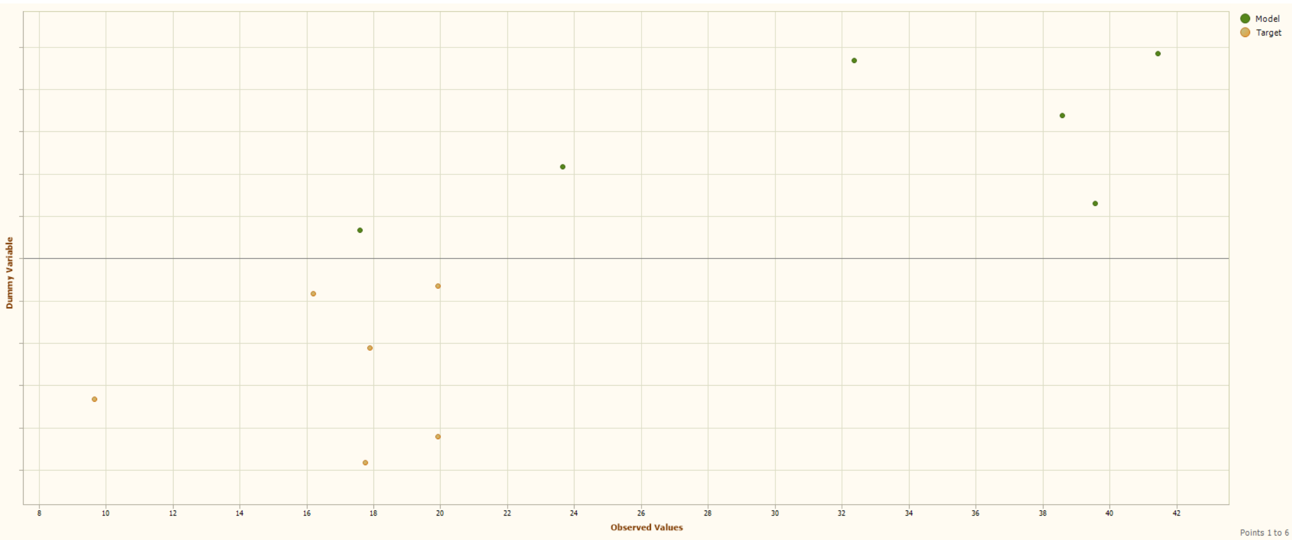
A10.2.2 Target sorted fitting



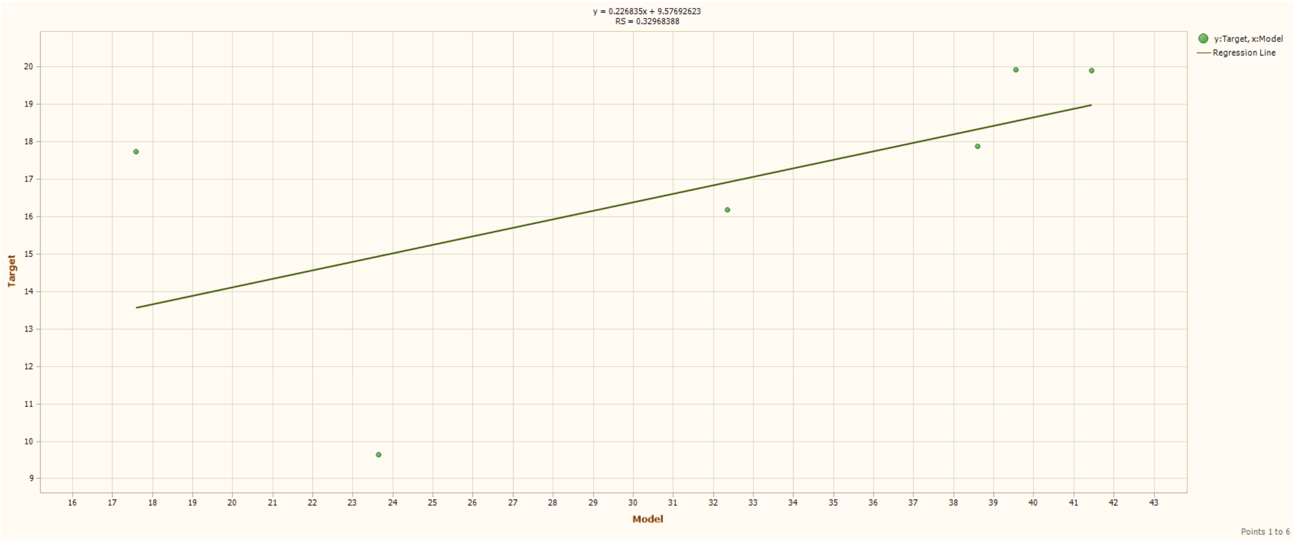
A10.2.3 Model sorted fitting



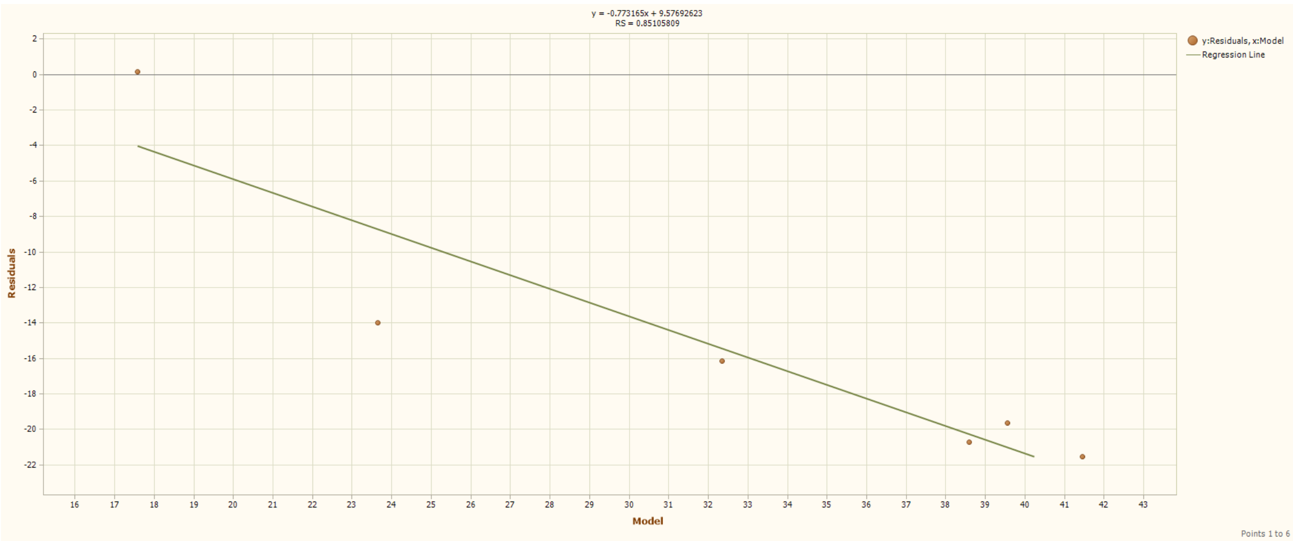
A10.2.4 Stacked distribution



A10.2.5 Scatter plot



A10.2.6 Residuals plot



A10.2.7 Model performance

Fitness 55.5367317011037

MSE 289.207471087483

RMSE 17.0061009960391

MAE 15.364756188316

RSE 23.796934649864

RRSE 4.87821018918456

RAE 5.8143792123248

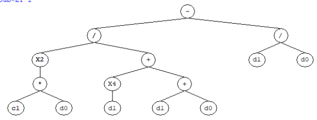
Correlation Coefficient 0.574181055221706

R-square 0.329683884175512

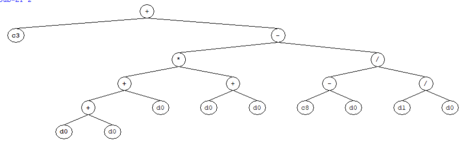
Calculation Errors 0

A10.3 Output model explained by Expression Tree

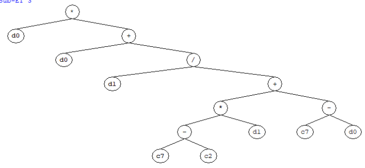
Sub-ET 1



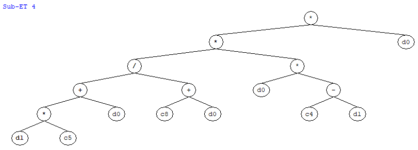
Sub-ET 2



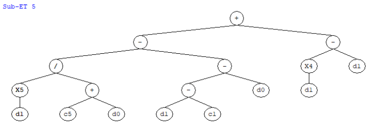
Sub-ET 3



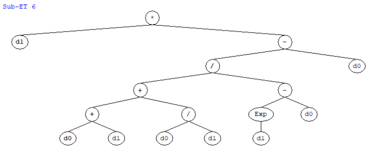
Sub-ET 4



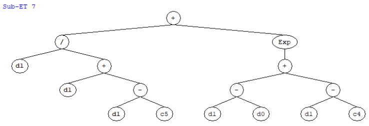
Sub-ET 5



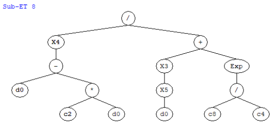
Sub-ET 6



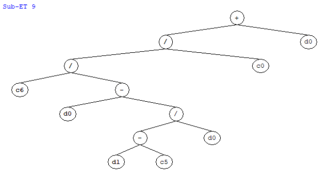
Sub-ET 7



Sub-ET 8



Sub-ET 9



A10.4 Output model explained by MATLAB

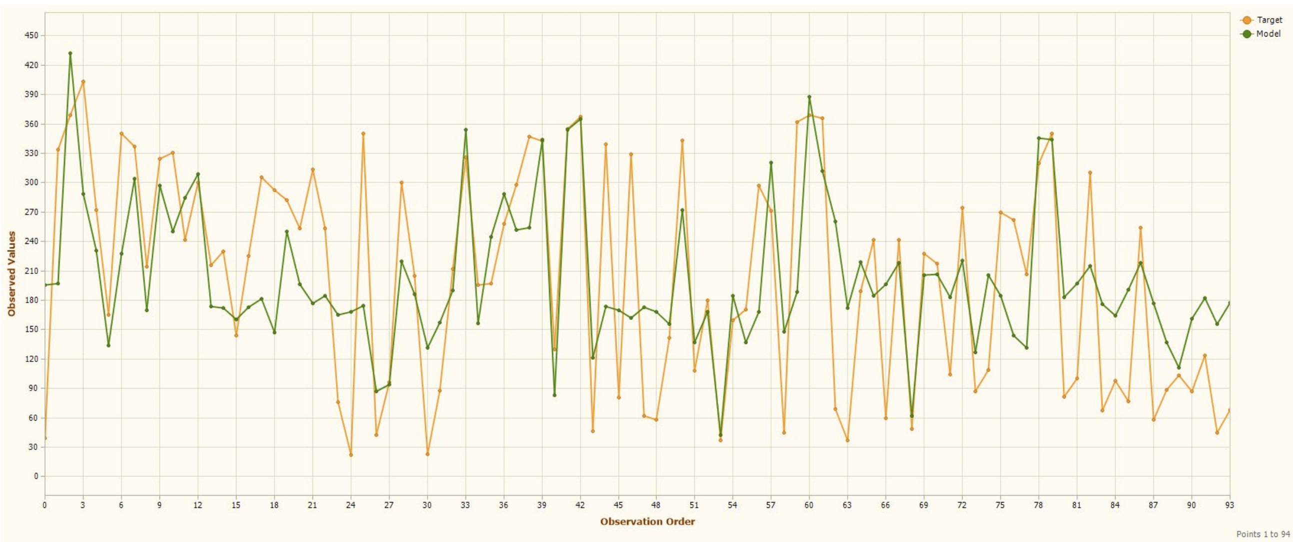
```
%------------------------------------------------------------------------
% Regression model generated by GeneXproTools 5.0 on 24/09/2020 3:11:28 AM
% GEP File: C:\Users\resha\Google Drive\Final GeneXPro\Central west\station 10\st10_6P_7N.gep
% Training Records:  94
% Validation Records:   6
% Fitness Function:  RMSE
% Training Fitness:  57.1283554543532
% Training R-square: 0.784020360090188
% Validation Fitness:   55.5367317011037
% Validation R-square:  0.329683884175512
%------------------------------------------------------------------------

function result = gepModelQ10C(d)

G1C1 = 2.15388204838582;
G2C3 = 8.98645030704631;
G2C8 = -0.495551920592748;
G3C7 = 9.94201483199561;
G3C2 = -4.82976470229194;
G4C8 = -8.26949088861504;
G4C4 = 3.71557145908994;
G4C5 = 16.8624098801291;
G5C5 = 1.31175263209875;
G5C1 = 7.00207871854284;
G7C5 = 2.51961229325873;
G7C4 = -0.757262858782164;
G8C2 = -1.18172183595872;
G8C8 = 7.83596301156651;
G8C4 = 7.74590289010285;
```

```
G9C0 = 4.1417502215211;

G9C6 = 5.82200421719728;

G9C5 = -4.35257579709894;


Six_PC = 1;

Seven_NC = 2;


y = 0.0;


y = ((((G1C1*d(Six_PC))^2)/((d(Seven_NC)^4)+(d(Seven_NC)+d(Six_PC))))-(d(Seven_NC)/d(Six_PC)));

y = y + (G2C3+((((d(Six_PC)+d(Six_PC))+d(Six_PC))*(d(Six_PC)+d(Six_PC)))-((G2C8-
d(Six_PC))/(d(Seven_NC)/d(Six_PC)))));

y = y + (d(Six_PC)*(d(Six_PC)+(d(Seven_NC)/(((G3C7-G3C2)*d(Seven_NC))+(G3C7-d(Six_PC))))));

y = y + (((((d(Seven_NC)*G4C5)+d(Six_PC))/(G4C8+d(Six_PC)))*(d(Six_PC)*(G4C4-d(Seven_NC))))*d(Six_PC));

y = y + ((((d(Seven_NC)^5)/(G5C5+d(Six_PC)))-((d(Seven_NC)-G5C1)-d(Six_PC)))+((d(Seven_NC)^4)-
d(Seven_NC)));

y = y + (d(Seven_NC)*((((d(Six_PC)+d(Seven_NC))+(d(Six_PC)/d(Seven_NC)))/(exp(d(Seven_NC))-d(Six_PC)))-
d(Six_PC)));

y = y + ((d(Seven_NC)/(d(Seven_NC)+(d(Seven_NC)-G7C5)))+exp(((d(Seven_NC)-d(Six_PC))+(d(Seven_NC)-
G7C4))));

y = y + (((d(Six_PC)-(G8C2*d(Six_PC)))^4)/(((d(Six_PC)^5)^3)+exp((G8C8/G8C4))));

y = y + (((G9C6/(d(Six_PC)-((d(Seven_NC)-G9C5)/d(Six_PC))))/G9C0)+d(Six_PC));


result = y;
```

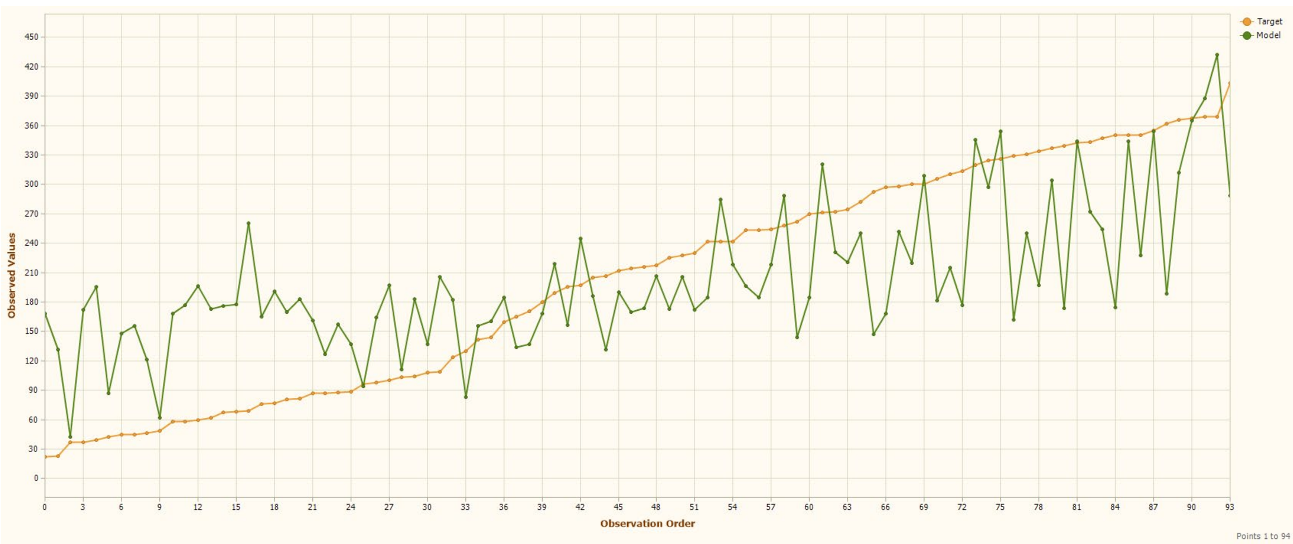
A11 Best GEP Model Output for Barham Station (WNSW)

A11.1 Output results for Calibration period

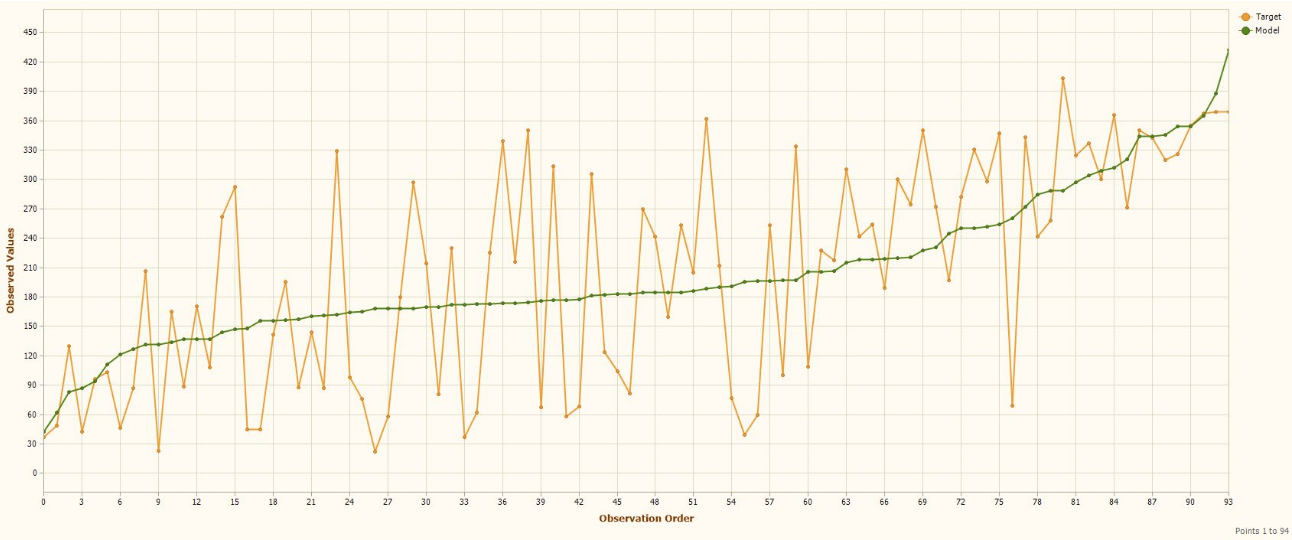
A11.1.1 Curve Fitting



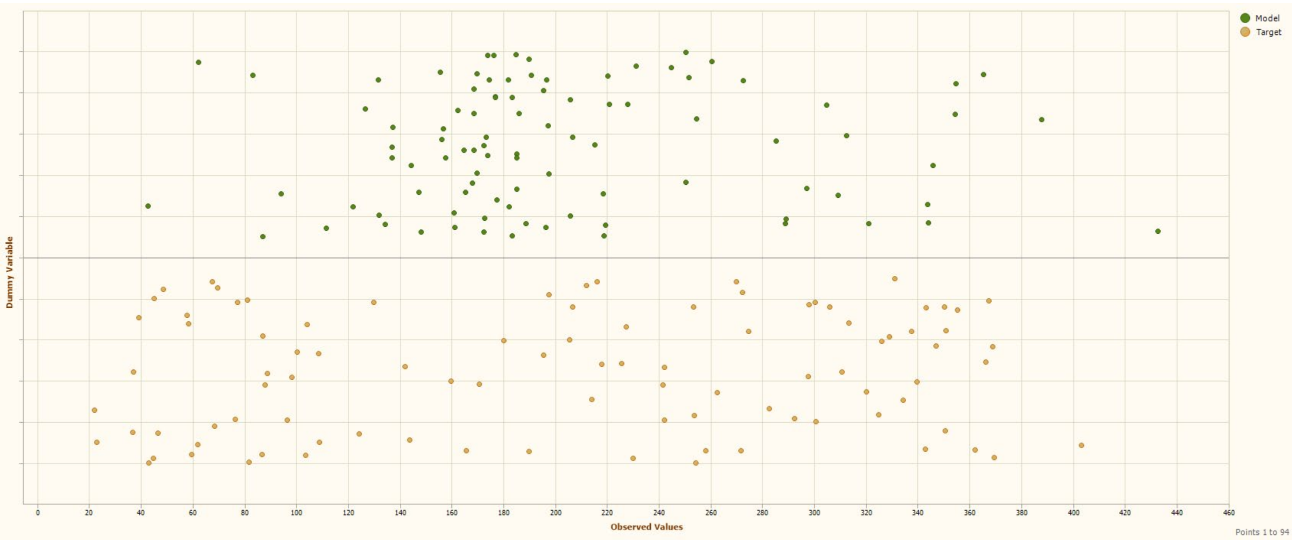
A11.1.2 Target sorted fitting



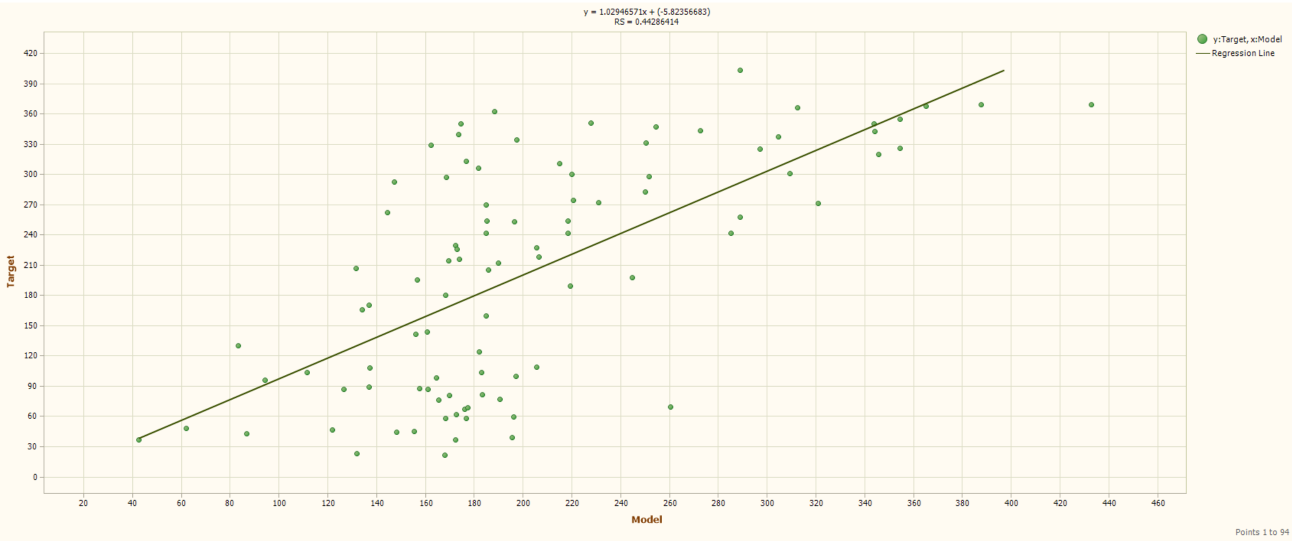
A11.1.3 Model sorted fitting



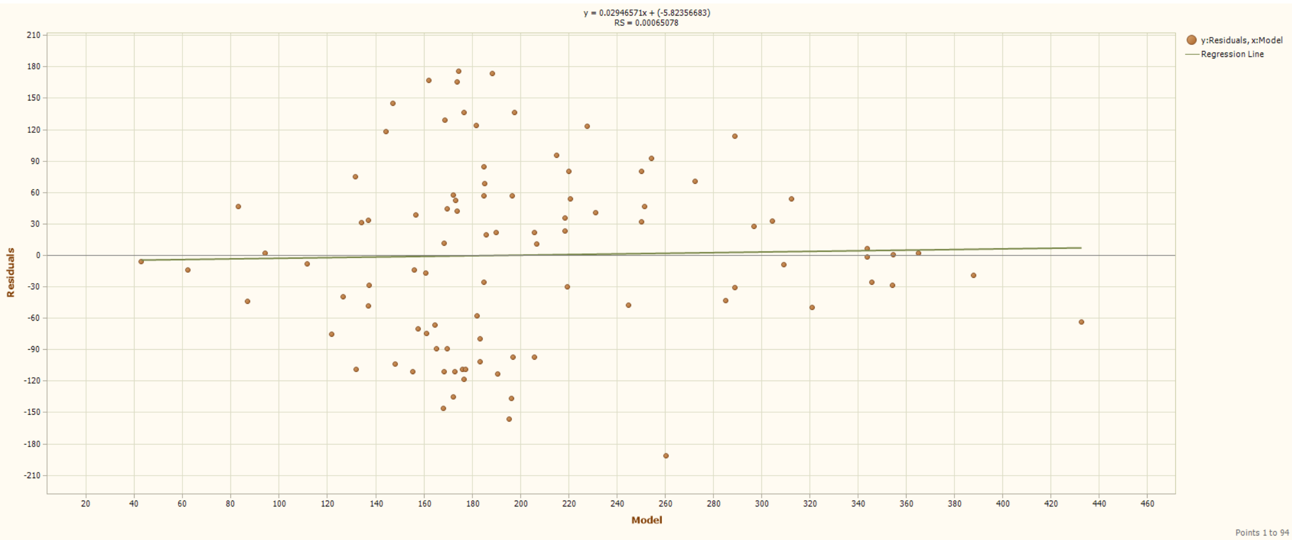
A11.1.4 Stacked distribution



A11.1.5 Scatter plot



A11.1.6 Residuals plot

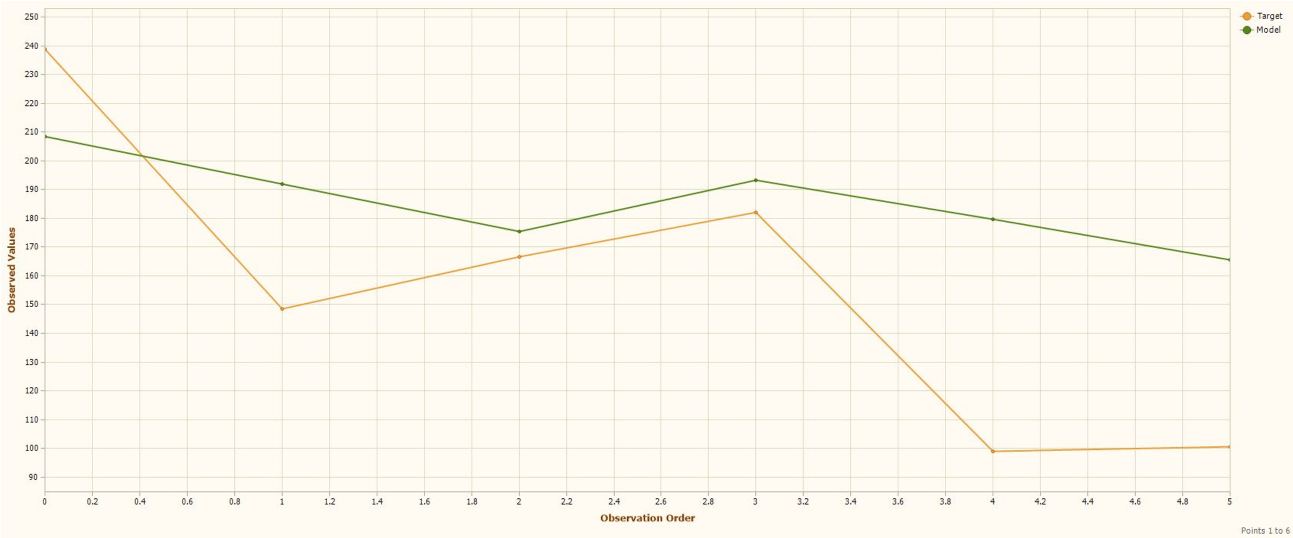


A11.1.7 Model performance

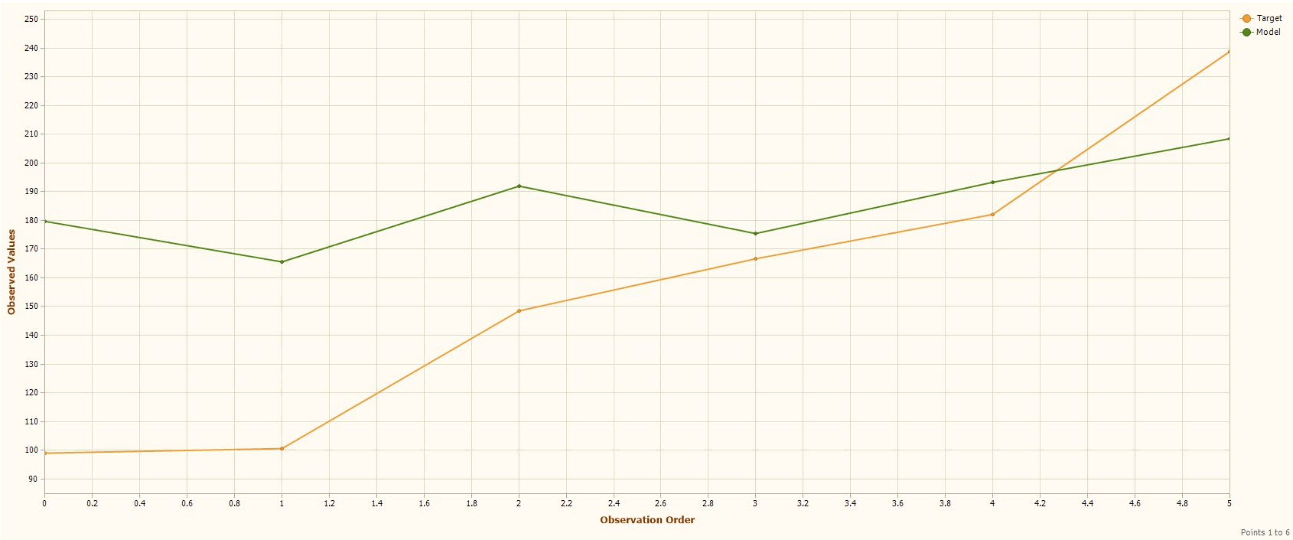
Fitness	11.734880586043
MSE	7092.34062660336
RMSE	84.2160354481458
MAE	69.2387897292566
RSE	0.557500503325499
RRSE	0.746659563205012
RAE	0.689892705049765
Correlation Coefficient	0.665480384478872
R-square	0.442864142126147
Calculation Errors	0

A11.2 Output results for Validation period

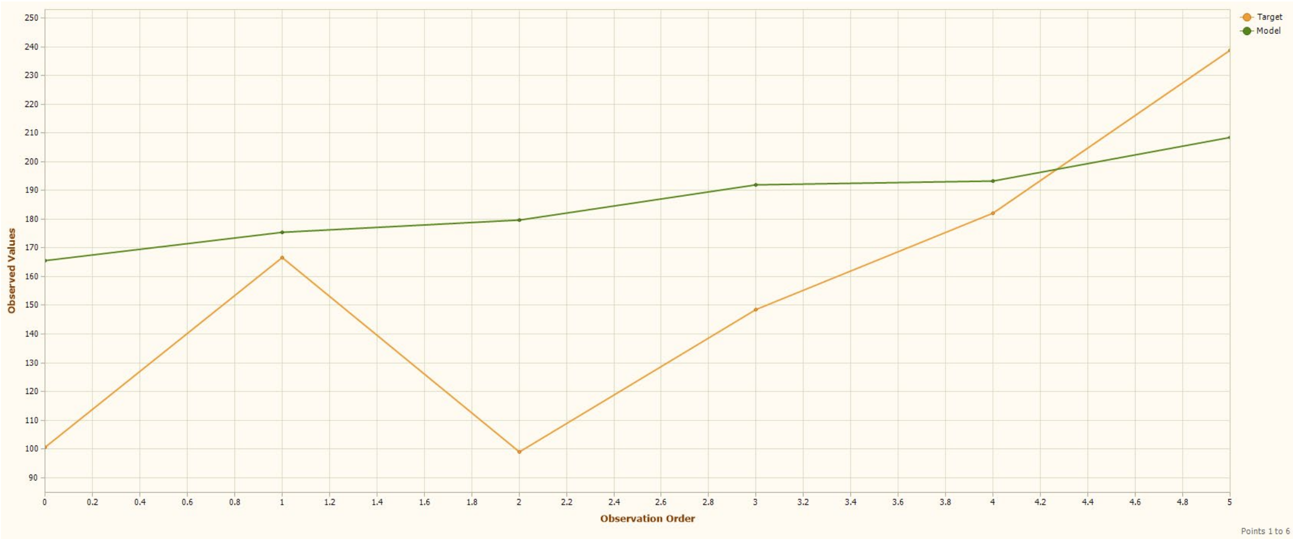
A11.2.1 Curve fitting



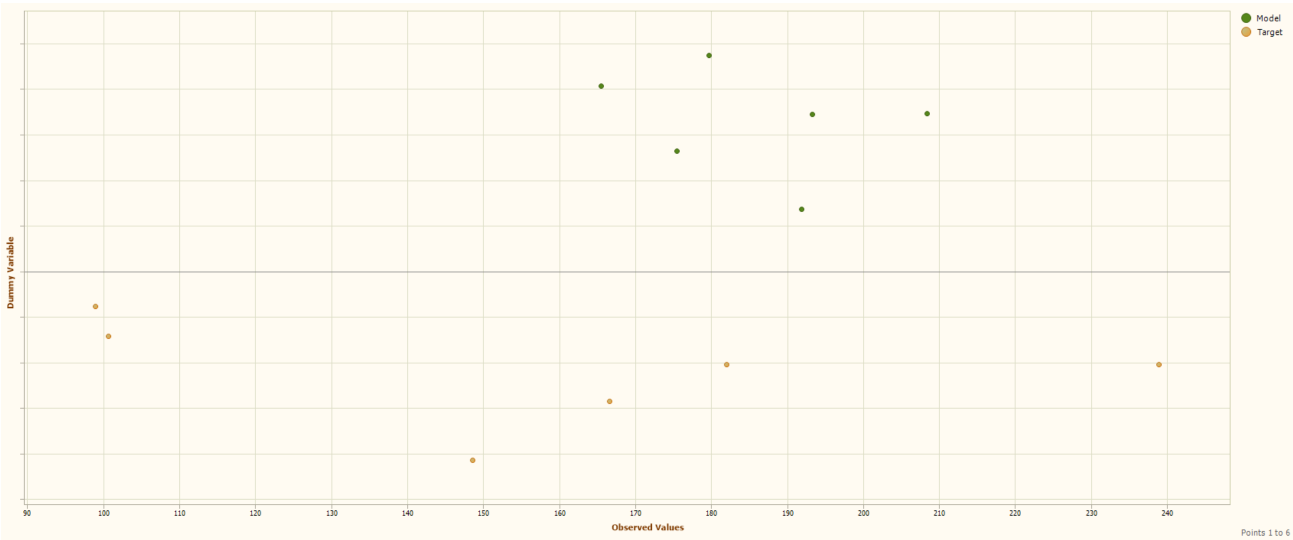
A11.2.2 Target sorted fitting



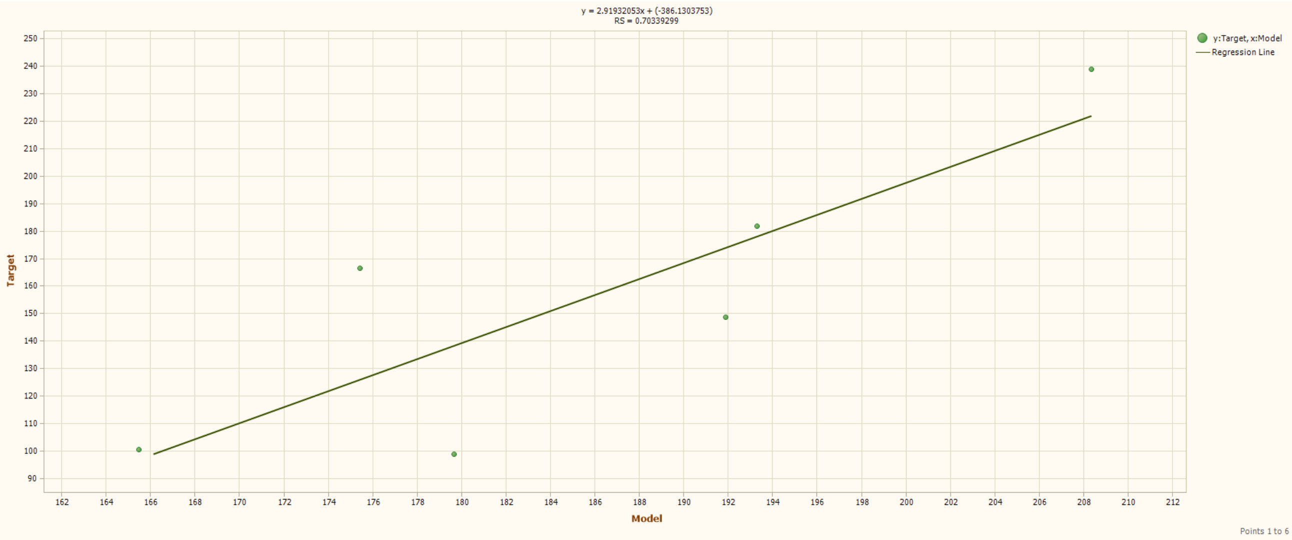
A11.2.3 Model sorted fitting



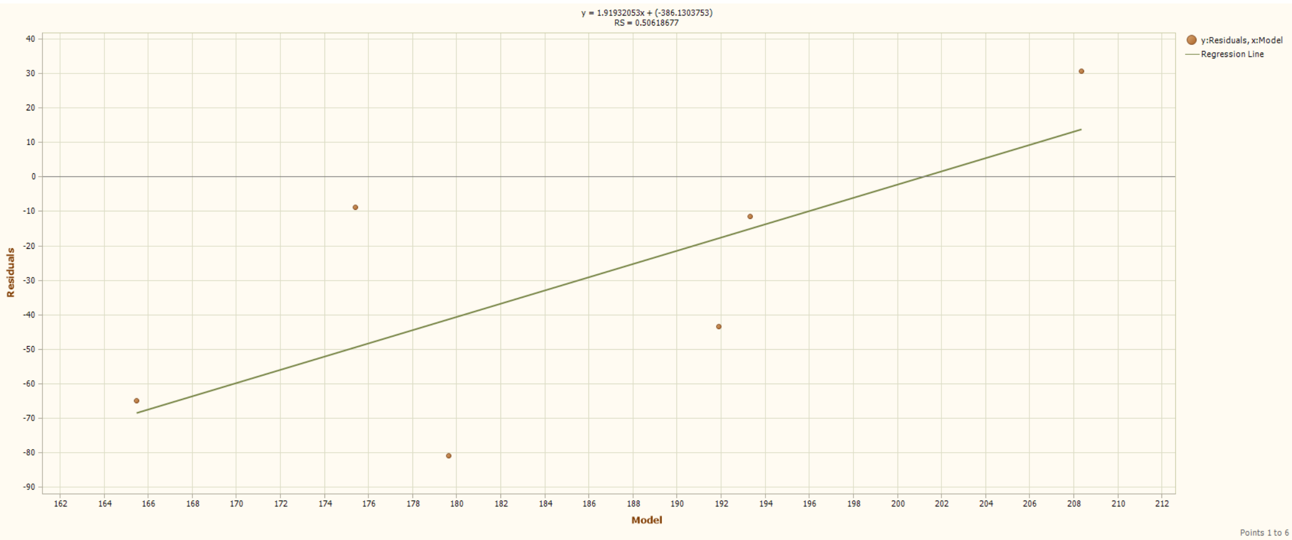
A11.2.4 Stacked distribution



A11.2.5 Scatter plot



A11.2.6 Residuals plot

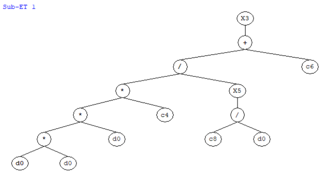


A11.2.7 Model performance

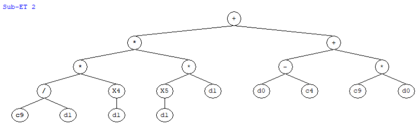
Fitness	20.4630195443411
MSE	2291.40699756531
RMSE	47.8686431556745
MAE	39.9455806317913
RSE	0.978701174197059
RRSE	0.989293270065585
RAE	1.00120730997167
Correlation Coefficient	0.838685272214799
R-square	0.703392985830012
Calculation Errors	0

A11.3 Output model explained by Expression Tree

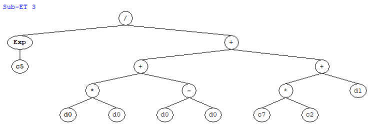
Sub-ET 1



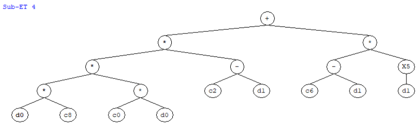
Sub-ET 2



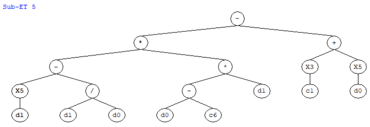
Sub-ET 3



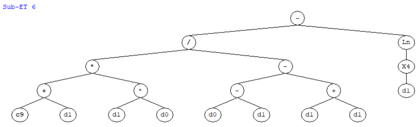
Sub-ET 4



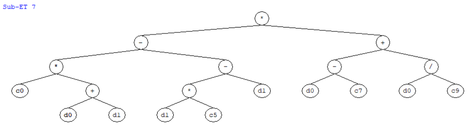
Sub-ET 5



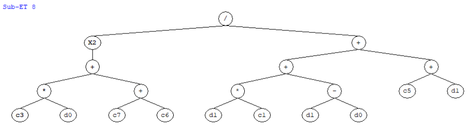
Sub-ET 6



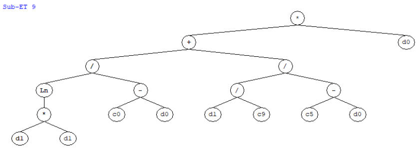
Sub-ET 7



Sub-ET 8



Sub-ET 9



A11.4 Output model explained by MATLAB

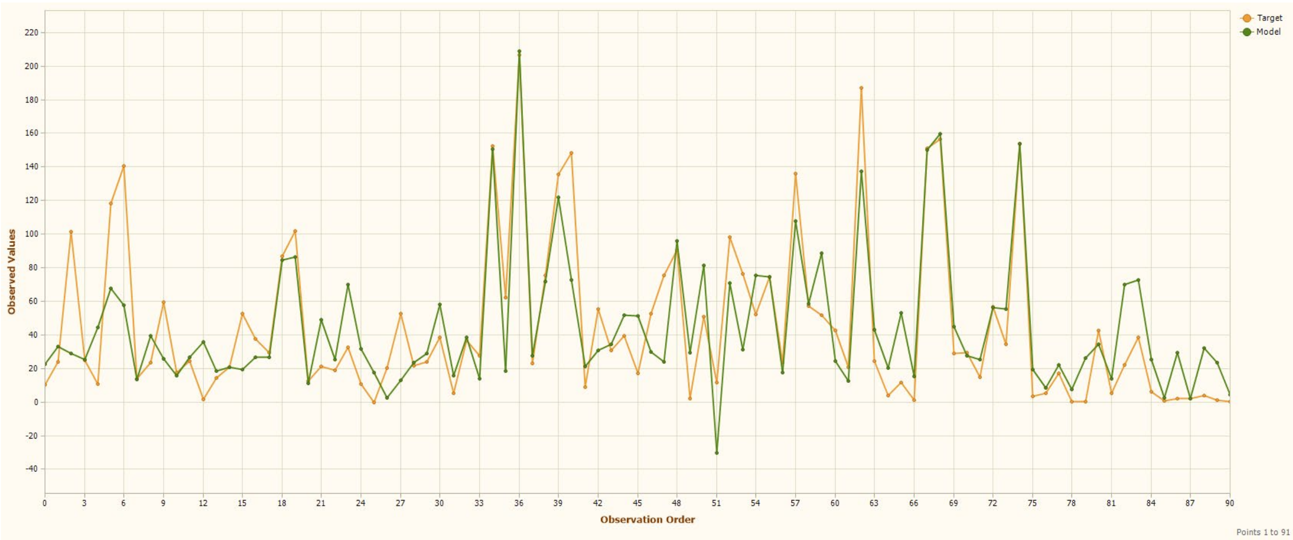
```
%------------------------------------------------------------------------

% Regression model generated by GeneXproTools 5.0 on 24/09/2020 3:20:09 AM

% GEP File: C:\Users\resha\Google Drive\Final GeneXPro\FINAL\st04_5P_4N.gep

% Training Records:  94

% Validation Records:   6

% Fitness Function:  RMSE

% Training Fitness:  11.734880586043

% Training R-square: 0.442864142126147

% Validation Fitness:   20.4630195443411

% Validation R-square:  0.703392985830012

%------------------------------------------------------------------------


function result = gepModelQ4C(d)


G1C6 = 7.01772617781985;

G1C4 = 16.2172564245718;

G1C8 = -8.19462899143966;

G2C4 = 10.842200475842;

G2C9 = 15.5303723879489;

G3C5 = 0.57007369242754;

G3C7 = -3.24746809839492;

G3C2 = 0.255708115503022;

G4C2 = 0.836263830088121;

G4C6 = 8.06398965177774;

G4C8 = 1.70072443834357;

G4C0 = 7.02838115874663;

G5C1 = 5.59413760495826;

G5C6 = 7.49776330437287;

G6C9 = -2.32088666808305;

G7C0 = 2.59376527435523;
```

```
G7C7 = 4.04541267117665;

G7C9 = 1.25068146290835;

G7C5 = -12.0351034426597;

G8C5 = 4.00789880865042;

G8C3 = 0.836112013123552;

G8C7 = -7.49038933103061;

G8C6 = 9.04324673183392;

G8C1 = -4.5302373635532;

G9C0 = -1.05650234325856;

G9C9 = 2.0863778778507;

G9C5 = -0.82002053193762;


Five_PC = 1;

Four_NC = 2;

y = 0.0;

y = ((((((d(Five_PC)*d(Five_PC))*d(Five_PC))*G1C4)/((G1C8/d(Five_PC))^5))+G1C6)^3);

y = y + ((((G2C9/d(Four_NC))*(d(Four_NC)^4))*((d(Four_NC)^5)*d(Four_NC)))+((d(Five_PC)-
G2C4)+(G2C9*d(Five_PC))));

y = y + (exp(G3C5)/(((d(Five_PC)*d(Five_PC))+(d(Five_PC)-d(Five_PC)))+((G3C7*G3C2)+d(Four_NC))));

y = y + ((((d(Five_PC)*G4C8)*(G4C0*d(Five_PC)))*(G4C2-d(Four_NC)))+((G4C6-
d(Four_NC))*(d(Four_NC)^5)));

y = y + ((((d(Four_NC)^5)-(d(Four_NC)/d(Five_PC)))*((d(Five_PC)-G5C6)*d(Four_NC)))-
((G5C1^3)+(d(Five_PC)^5)));

y = y + ((((G6C9+d(Four_NC))*(d(Four_NC)*d(Five_PC)))/((d(Five_PC)-d(Four_NC))-
(d(Four_NC)+d(Four_NC))))-reallog((d(Four_NC)^4)));

y = y + (((G7C0*(d(Five_PC)+d(Four_NC)))-((d(Four_NC)*G7C5)-d(Four_NC)))*((d(Five_PC)-
G7C7)+(d(Five_PC)/G7C9)));

y = y + ((((G8C3*d(Five_PC))+(G8C7+G8C6))^2)/(((d(Four_NC)*G8C1)+(d(Four_NC)-
d(Five_PC)))+(G8C5+d(Four_NC))));

y = y + (((reallog((d(Four_NC)*d(Four_NC)))/(G9C0-d(Five_PC)))+((d(Four_NC)/G9C9)/(G9C5-
d(Five_PC))))*d(Five_PC));


result = y;
```

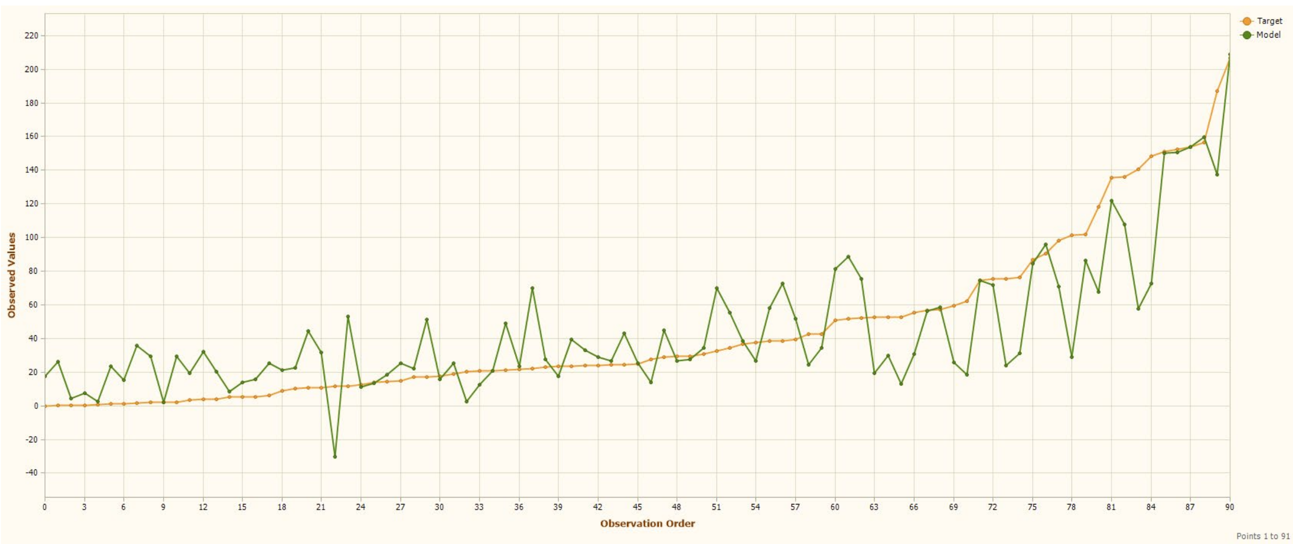
A12 Best GEP Model Output for Brewarrina Station (WNSW)

A12.1 Output results for Calibration period

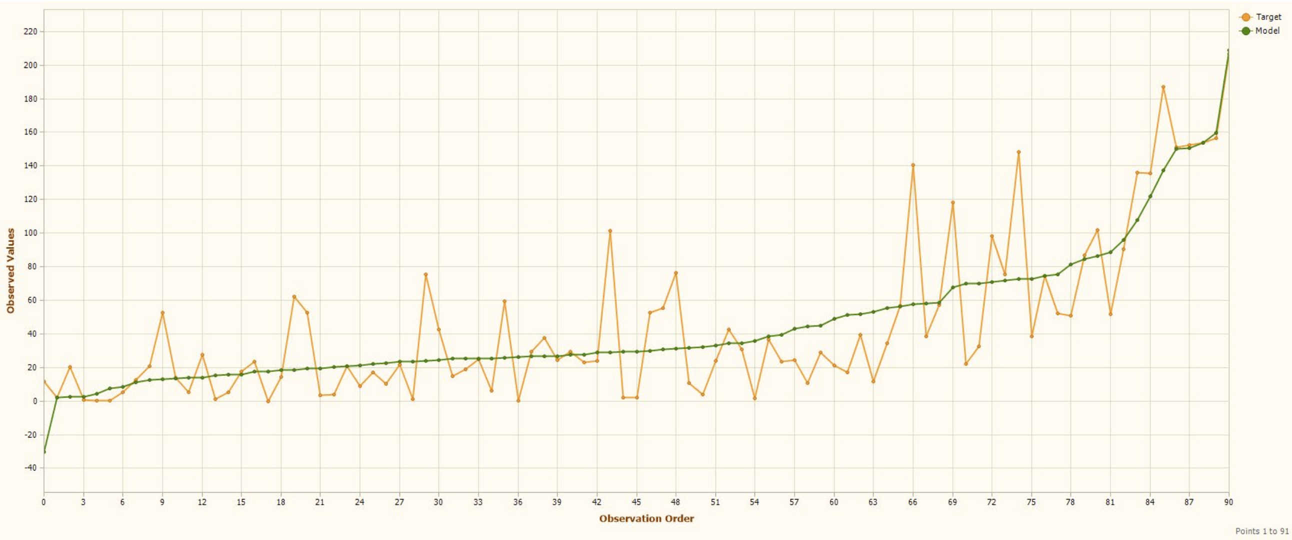
A12.1.1 Curve Fitting



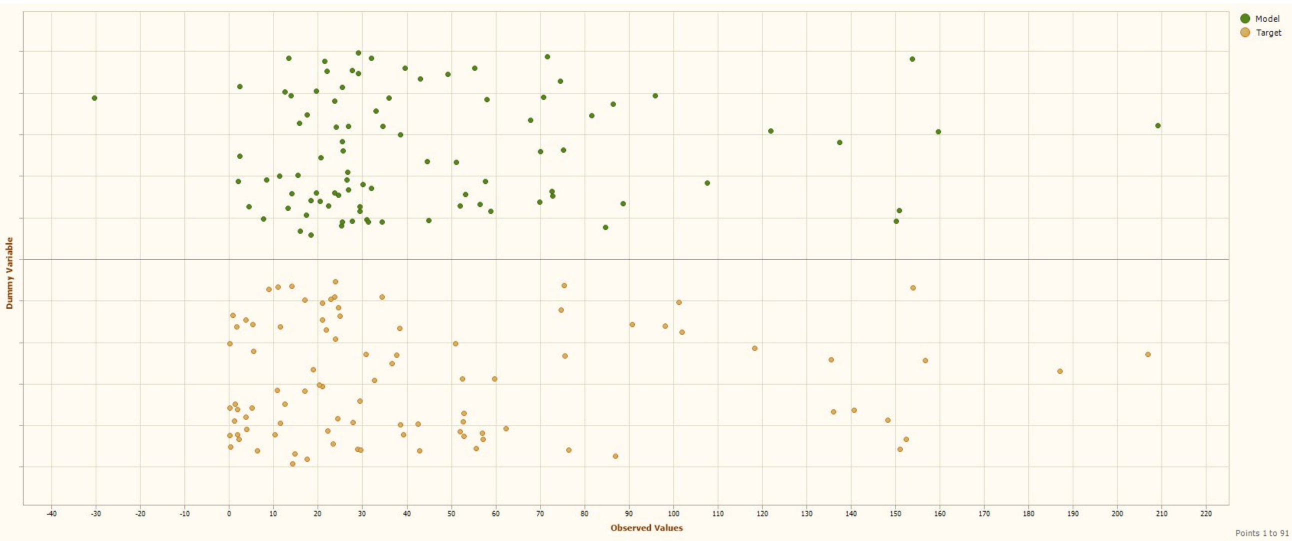
A12.1.2 Target sorted fitting



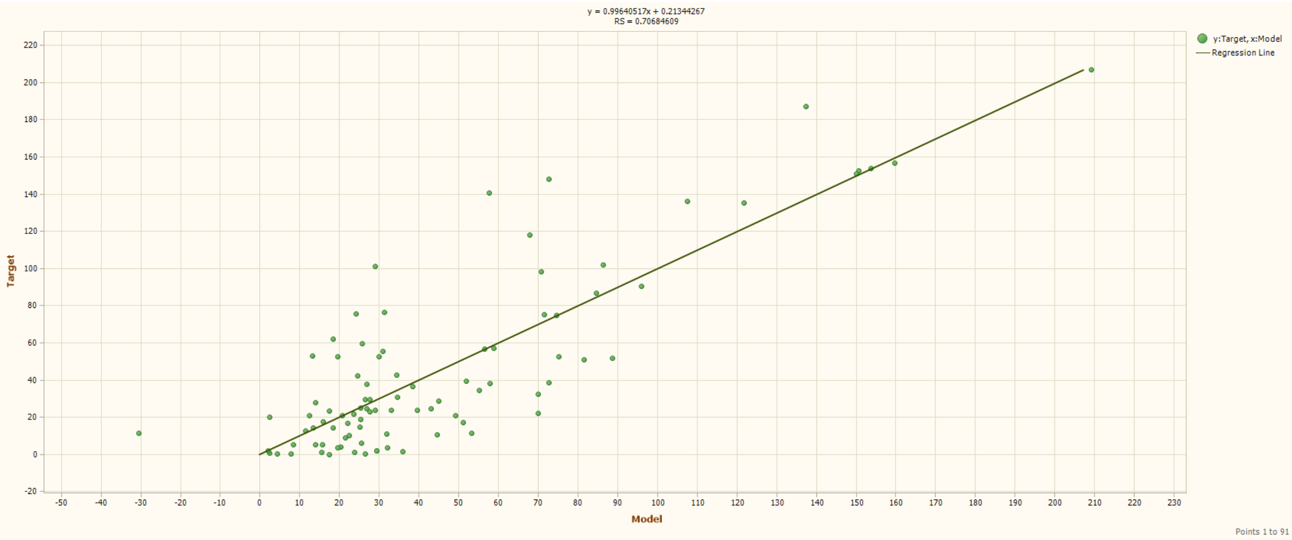
A12.1.3 Model sorted fitting



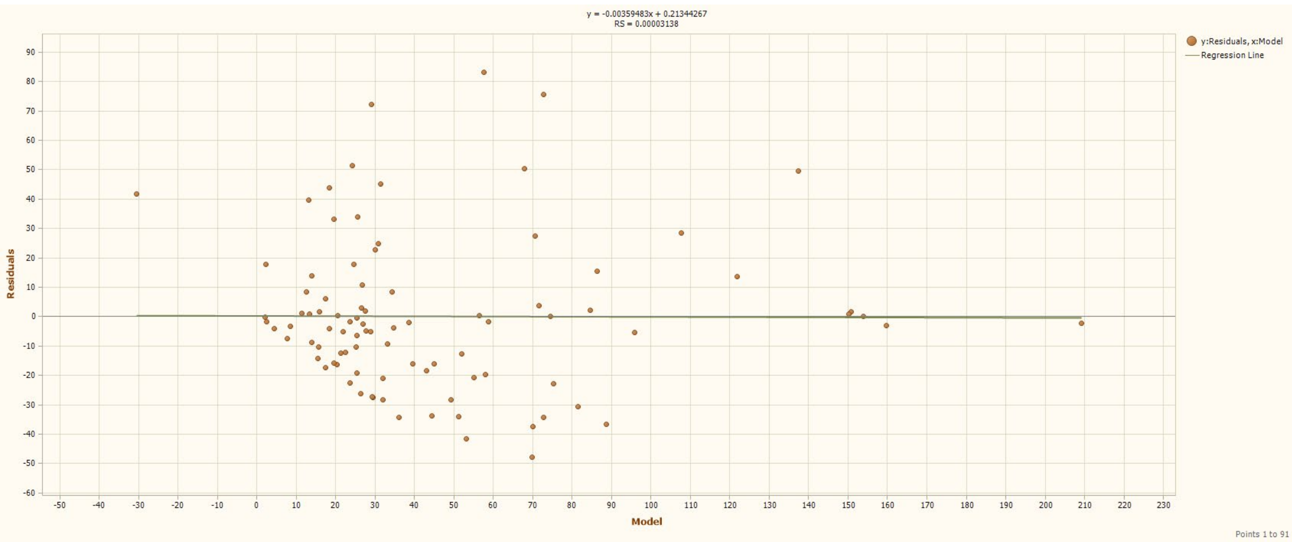
A12.1.4 Stacked distribution



A12.1.5 Scatter plot



A12.1.6 Residuals plot



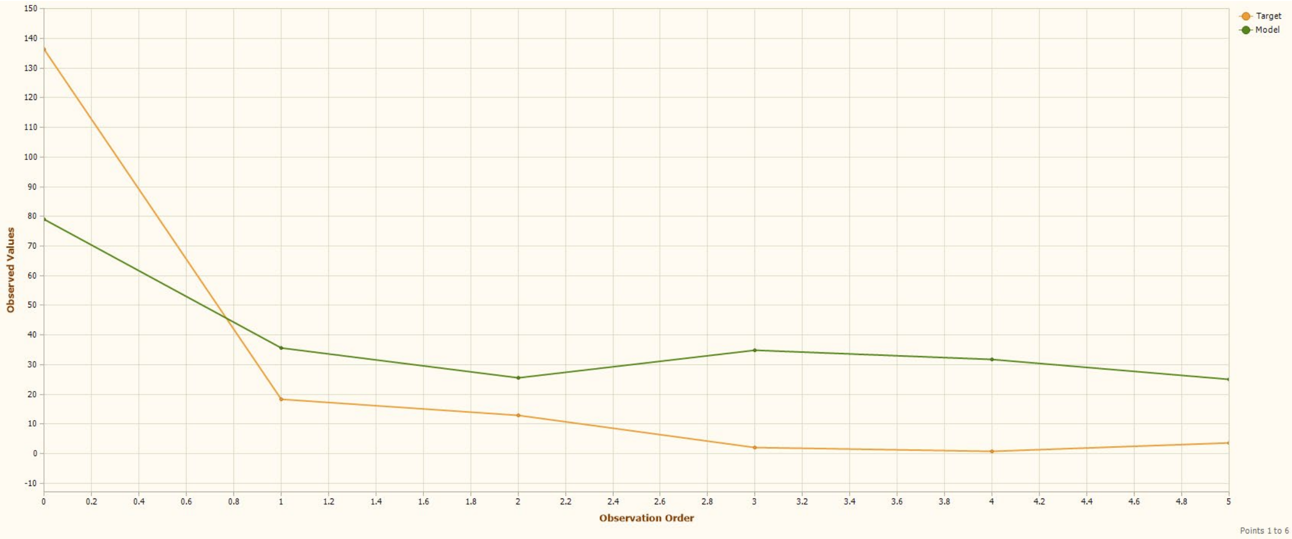
28 52.86348352 13.1319960036702 39.7314875163298

A12.1.7 Model performance

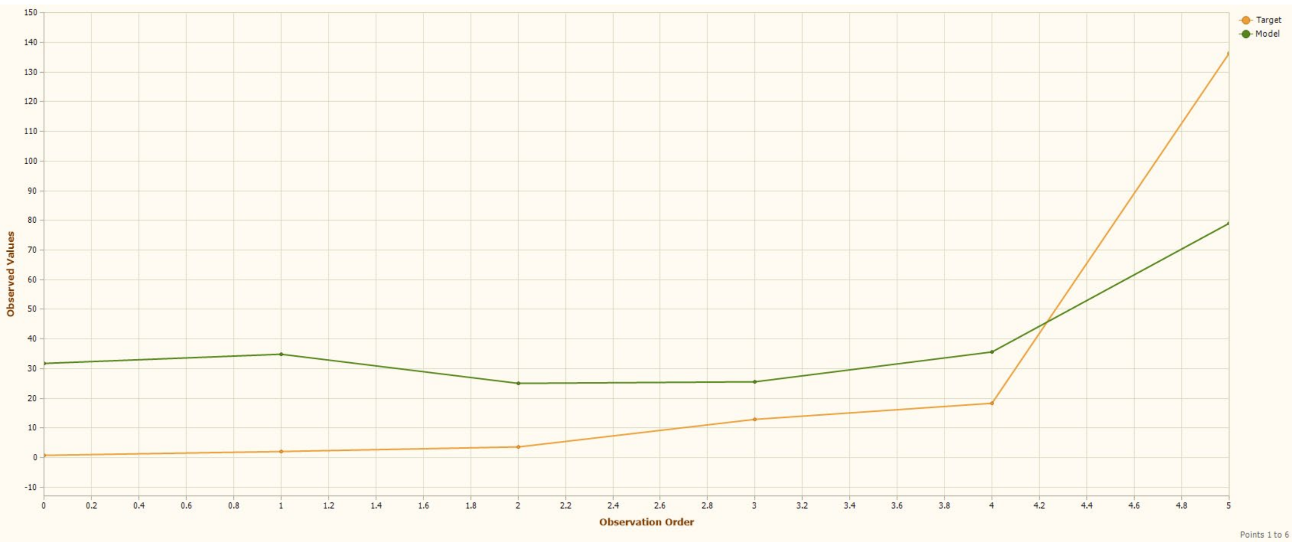
Fitness	37.1535071969118
MSE	671.605853981566
RMSE	25.9153594221953
MAE	18.7171276967771
RSE	0.293164243397578
RRSE	0.541446436314414
RAE	0.509720019066482
Correlation Coefficient	0.840741395477453
R-square	0.706846094069376
Calculation Errors	0

A12.2 Output results for Validation period

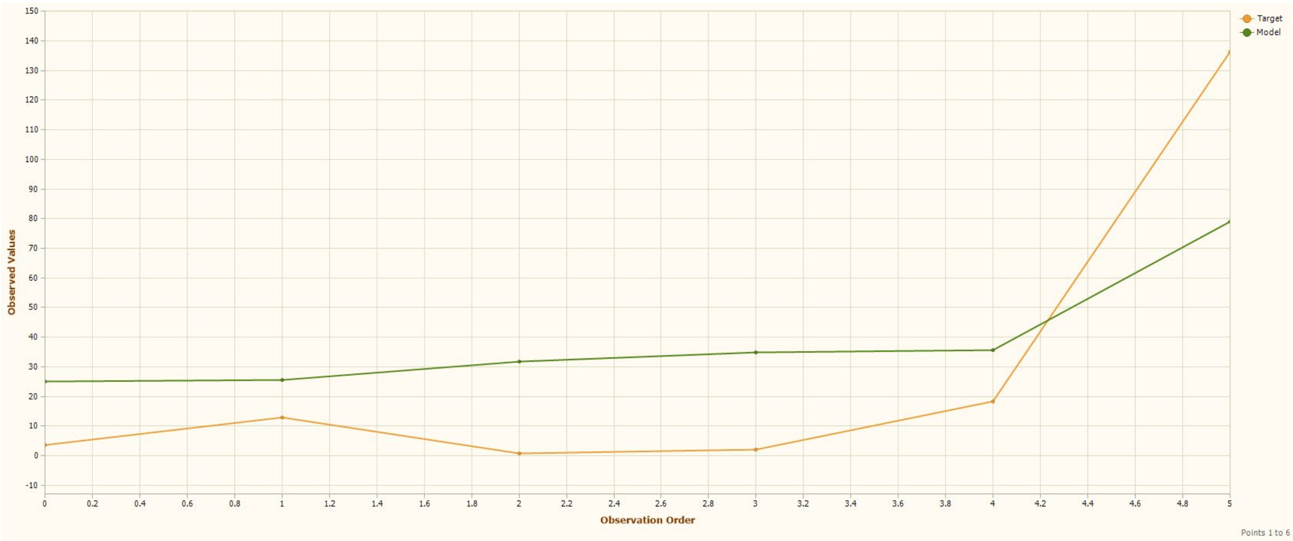
A12.2.1 Curve fitting



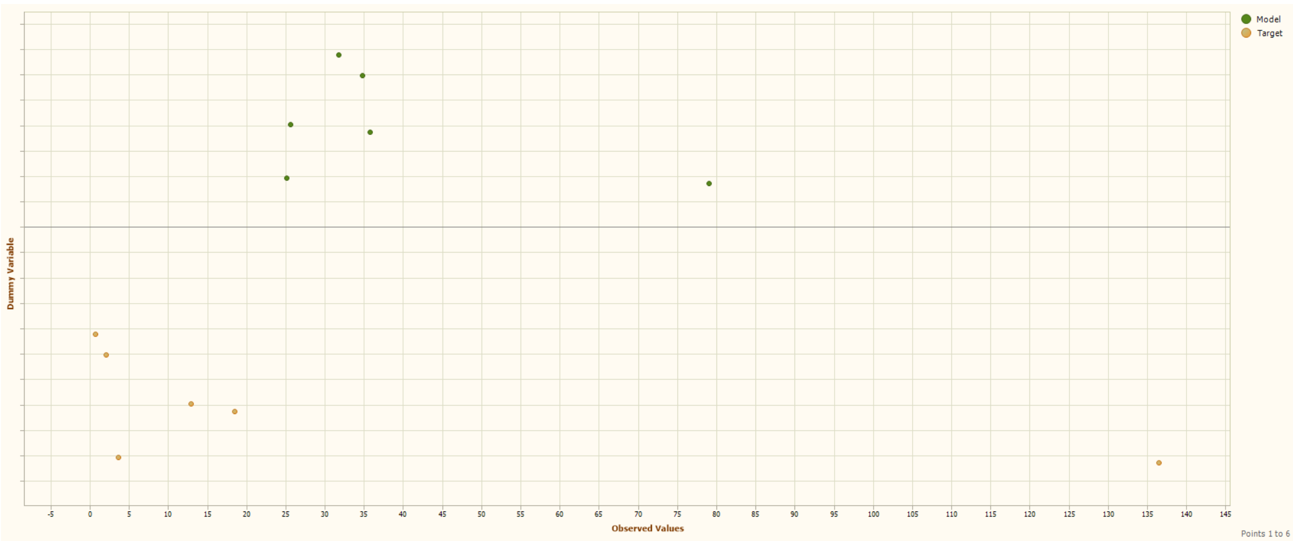
A12.2.2 Target sorted fitting



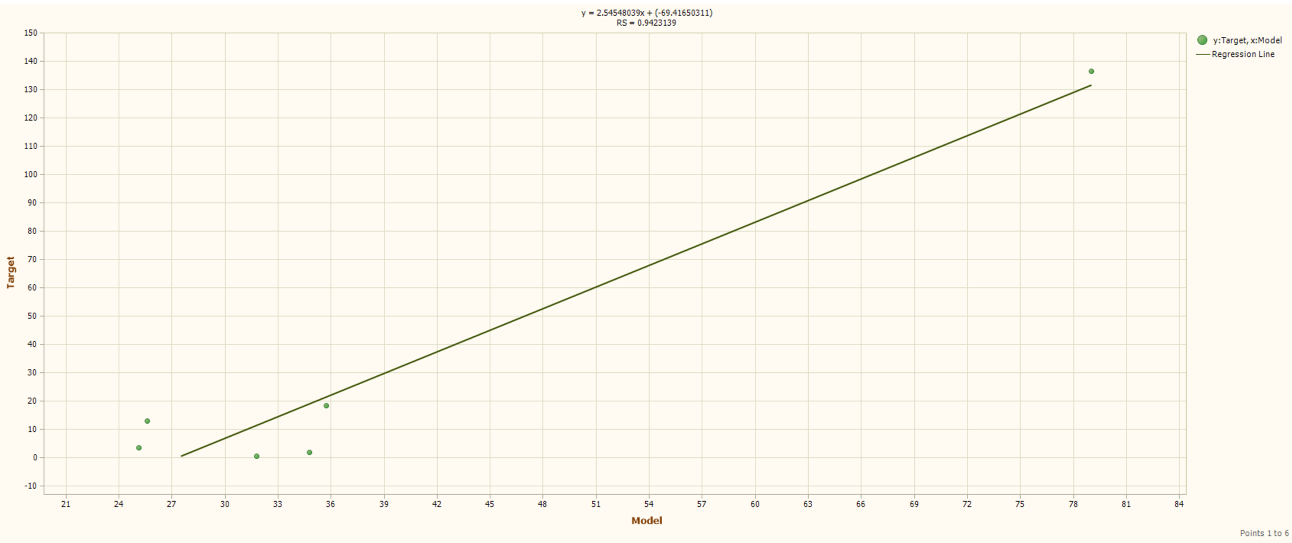
A12.2.3 Model sorted fitting



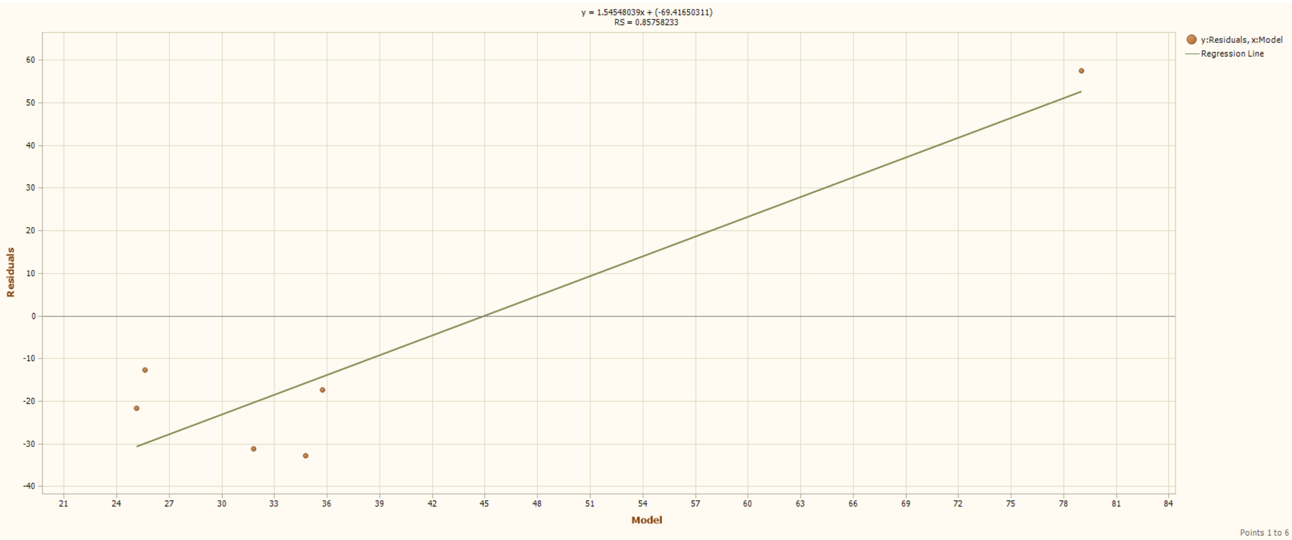
A12.2.4 Stacked distribution



A12.2.5 Scatter plot



A12.2.6 Residuals plot

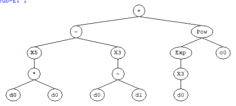


A12.2.7 Model performance

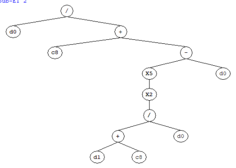
Fitness	30.0046321680741
MSE	1045.11169245501
RMSE	32.328187274498
MAE	28.8195955917715
RSE	0.444697906621199
RRSE	0.666856736204411
RAE	0.804458211614985
Correlation Coefficient	0.970728543084829
R-square	0.942313904359594
Calculation Errors	0

A12.3 Output model explained by Expression Tree

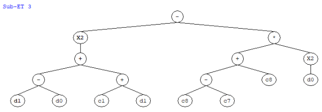
Sub-ET 1



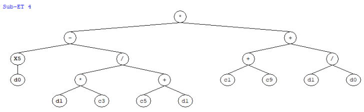
Sub-ET 2



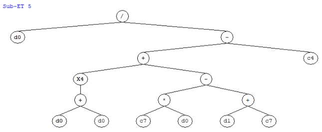
Sub-ET 3



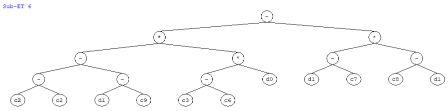
Sub-ET 4



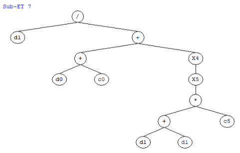
Sub-ET 5



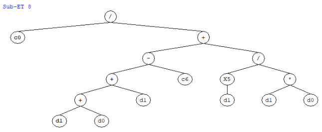
Sub-ET 6



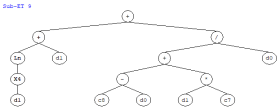
Sub-ET 7



Sub-ET 8



Sub-ET 9



A12.4 Output model explained by MATLAB

```
%------------------------------------------------------------------------

% Regression model generated by GeneXproTools 5.0 on 24/09/2020 3:28:21 AM

% GEP File: C:\Users\resha\Google Drive\Final GeneXPro\FINAL\st12_2E_2N.gep

% Training Records:  91

% Validation Records:   6

% Fitness Function:  RMSE

% Training Fitness:  37.1535071969118

% Training R-square: 0.706846094069376

% Validation Fitness:   30.0046321680741

% Validation R-square:  0.942313904359594

%------------------------------------------------------------------------


function result = gepModelQ12C(d)


G1C0 = 9.53449809062603;

G2C8 = 0.442547075687291;

G3C8 = 15.2988165846216;

G3C1 = -6.03820420271073;

G3C7 = -16.9383642438457;

G4C1 = -4.02629069231777;

G4C9 = -12.3236688828126;

G4C3 = -0.250886438635503;

G4C5 = 1.34861745461221;

G5C4 = 2.60186071352275;

G5C7 = -2.3820267342143;

G6C7 = 2.2537874172208;

G6C8 = 4.52776848963897;

G6C2 = -9.38459425641652;

G6C9 = -5.94369033317057;
```

```
G6C3 = 7.81753113860919;

G6C6 = 2.79611056495504;

G7C0 = -0.376852666967208;

G7C5 = -1.31964774620808;

G8C0 = -1.10925336768749;

G8C6 = -0.747868174799753;

G9C8 = -0.406758446555927;

G9C7 = 2.43077901369159;


Two_EC = 1;

Two_NC = 2;


y = 0.0;


y = ((((d(Two_EC)*d(Two_EC))^5)-((d(Two_EC)-d(Two_NC))^3))+realpow(exp((d(Two_EC)^3)),G1C0));

y = y + (d(Two_EC)/(G2C8+(((((d(Two_NC)+G2C8)/d(Two_EC))^2)^5)-d(Two_EC))));

y = y + ((((d(Two_NC)-d(Two_EC))+(G3C1+d(Two_NC)))^2)-(((G3C8-G3C7)+G3C8)*(d(Two_EC)^2)));

y = y + (((d(Two_EC)^5)-
((d(Two_NC)*G4C3)/(G4C5+d(Two_NC))))*((G4C1+G4C9)+(d(Two_NC)/d(Two_EC))));

y = y + (d(Two_EC)/((((d(Two_EC)+d(Two_EC))^4)+((G5C7*d(Two_EC))-(d(Two_NC)+G5C7)))-G5C4));

y = y + ((((G6C2-G6C2)-(d(Two_NC)-G6C9))*((G6C3-G6C6)*d(Two_EC)))-((d(Two_NC)-G6C7)*(G6C8-
d(Two_NC))));

y = y + (d(Two_NC)/((d(Two_EC)+G7C0)+((((d(Two_NC)+d(Two_NC))*G7C5)^5)^4)));

y = y + (G8C0/((((d(Two_NC)+d(Two_EC))+d(Two_NC))-G8C6)+((d(Two_NC)^5)/(d(Two_NC)*d(Two_EC)))));

y = y + ((reallog((d(Two_NC)^4))+d(Two_NC))+(((G9C8-d(Two_EC))+(d(Two_NC)*G9C7))/d(Two_EC)));


result = y;
```