PREDICTABILITY OF MONTHLY STREAMFLOW DISCHARGE USING REMOTE SENSING PRECIPITATION DATA BY DATA DRIVEN MODELS

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES
OF
MIDDLE EAST TECHNICAL UNIVERSITY

BY

MEHMET ALİ ÇOLAK

IN PARTIAL FULLFILLMENT OF THE REQUIREMENTS
FOR
THE DEGREE OF MASTER OF SCIENCE
IN
CIVIL ENGINEERING

AUGUST 2017

Approval of the thesis:

PREDICTABILITY OF MONTHLY STREAMFLOW DISCHARGE USING REMOTE SENSING PRECIPITATION DATA BY DATA DRIVEN MODELS

submitted by **MEHMET ALİ ÇOLAK** in partial fulfillment of the requirements for the degree of **Master of Science in Civil Engineering Department, Middle East Technical University** by,

Prof. Dr. Gülbin DURAL ÜNVER
Dean, Graduate School of **Natural and Applied Sciences** _____

Prof. Dr. İsmail Özgür YAMAN
Head of Department, **Dept. of Civil Engineering** _____

Asst. Prof. Dr. Mustafa Tuğrul YILMAZ
Supervisor, **Dept. of Civil Engineering, METU** _____

Examining Committee Members:

Prof. Dr. Melih YANMAZ
Dept. of Civil Engineering, METU _____

Asst. Prof. Dr. Mustafa Tuğrul YILMAZ
Dept. of Civil Engineering, METU _____

Assoc. Prof. Dr. Koray K. YILMAZ
Dept. of Geological Engineering, METU _____

Assoc. Prof. Dr. İsmail YÜCEL
Dept. of Civil Engineering, METU _____

Asst. Prof. Dr. Melih ÇALAMAK
Dept. of Civil Engineering, TED University _____

Date: 25.08.2017

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Name, Last Name: Mehmet Ali ÇOLAK

Signature :

ABSTRACT

PREDICTABILITY OF MONTHLY STREAMFLOW DISCHARGE USING REMOTE SENSING PRECIPITATION DATA BY DATA DRIVEN MODELS

ÇOLAK, Mehmet Ali

M.Sc., Department of Civil Engineering

Supervisor: Assist. Prof. Dr. Mustafa Tuğrul YILMAZ

August 2017, 131 pages

Predictability of stream flow has been the focus of many studies involving water resources management and hydroelectric energy production. Many hydrologic models have been developed to predict future and current streamflow at various time lags and locations. However, these physically-based models require reanalyzed future data sets (particularly precipitation forcing data) to predict future streamflow. Alternatively, data driven models can also provide predictions without the need of future projections by relying on the strong seasonality and autocorrelation that exist in the streamflow data. In this study, a data driven approach has been taken to predict monthly streamflow data sets utilizing precipitation data sets and using various linear and non-linear methods. Streamflow predictions of Coruh Basin have been performed using both the Tropical Rainfall Measurement Mission (TRMM) and the ground-based station precipitation (MGM) data sets between years 2000 – 2011. Predictions are validated using independent streamflow measurements acquired from General Directorate of State Hydraulic Works (DSI). A Simple Linear Regression Model (SLR), a Multiple Linear Regression Model (MLR), an Artificial Neural Network Model (MLP), and two Copula Models (Normal Copula and Frank Copula) are constructed and their predictions are cross-compared with the climatology- and persistence-based predictions. To further investigate the source of the predictive skills of these methods, separate predictions are made using the standardized anomaly

components of data sets [after climatology (long year monthly mean) components are removed and standardized by dividing by the standard deviation of the data] and complete data sets (normal/non-standardized data sets retaining both anomaly and climatology components). Results show the best predictions are obtained from the climatology-based predictions of the stations for the complete data sets while persistence-based predictions are also strong. Predictions using standardized anomaly data sets are improved when long-term climatology values added. These climatology added predictions show above 0.90 correlations, showing heavy majority of the predictive skill and the relation between the precipitation and the streamflow data sets are due to the strong seasonality impacting both variables.

Key Words: Monthly Streamflow Prediction, Rainfall-Runoff Modeling, Copulas, TRMM Data, Artificial Neural Networks (ANN), Climatology, Persistence

ÖZ

VERİ GÜDÜMLÜ YÖNTEMLER İLE UYDU YAĞIŞ VERİLERİ KULLANILARAK AYLIK AKARSU DEBİSİ TAHMİNİ

ÇOLAK, Mehmet Ali
Yüksek Lisans, İnşaat Mühendisliği Bölümü
Danışman:Yard. Doç. Dr. Mustafa Tuğrul YILMAZ
Ağustos 2017, 131 Sayfa

Akarsularda akım tahmini birçok bilimsel ve endüstriyel çalışmanın konusu olagelmiştir. Bu amaçla, farklı zaman ölçeklerini kullanan birçok hidrolojik model geliştirilmiştir. Ancak bu fiziksel modeller, özellikle yağış verilerini kullanan modeller, akarsu debisi tahmininde gelecekte olması düşünülen verilere ihtiyaç duyarlar. Bu modellere alternatif olarak, veri güdümlü modellerle de gelecek verilerine ihtiyaç duymadan tahminler yapılabilmektedir. Bu çalışmada, yağış verilerinden yararlanılarak, lineer ve lineer olmayan modeller kullanılarak, veri güdümlü yaklaşım ile aylık akarsu debisi tahmin edilmiştir. Debi tahminleri Çoruh Havzası için yapılmış, Tropik Yağış Ölçüm Misyonu (TRMM) ve Meteoroloji Genel Müdürülüğü (MGM) yer ölçüm istasyonlarının 2000-2011 yılları arası yağış ölçümlerinden faydalanılmıştır. Akarsu debisi tahminlerinin geçerliliği; Devlet Su İşleri (DSİ) tarafından sağlanan, yağış ölçümlerinden bağımsız olarak gerçekleştirilen akarsu debi ölçümlerinden yararlanılarak test edilmiştir. Bu çalışmada bir Basit Lineer Regresyon Modeli (SLR), bir Çoklu Lineer Regresyon (MLR) Modeli, bir Yapay Sinir Ağları Modeli (MLP) ile iki Kopula (NC ve FC) modeli oluşturulmuş ve debi tahminleri akarsu akımının klimatolojisi ve tutarlılığı ile karşılaştırılmıştır. Metodların akarsu debisi tahmin edebilirlik özelliklerini araştırmak için veri setlerinin standartlaştırılmış anomali bileşenleri [klimatoloji (uzun yıllar ortalaması) bileşenleri çıkarılıp ve standart sapmalarına bölünerek elde edilen veri setleri] ve bütün data setleri (klimatoloji ve anomali bileşenlerini içerisinde barındıran) ile birbirinden bağımsız tahminler yapılmıştır. Hesaplamalar sonucunda standartlaştırılmamış veri setleriyle en

iyi sonuçlar akarsu akımının klimatoloji tabanlı tahminlerinden elde edilmiş ve tutarlılık tabanlı akarsu akım tahminlerinin yaklaşık olarak aynı korelasyon değerlerine sahip olduğu gözlemlenmiştir. Uzun yıllar klimatoloji bileşenleri standartlaştırılmış veri setleriyle yapılan tahminlere eklendiğinde, tahminlerde ileri derecede gelişmeler olduğu gözlemlenmiştir. Klimatoloji bileşenlerinin eklenmiş olduğu tahminler, gerçek zamanlı gözlemlerle 0.90 değerinin üzerinde korelasyonlar göstermiştir. Bu durum yağış-akış ilişkisi ve debi tahminlerindeki birincil bileşenin güçlü mevsimsellik etkisi olduğunu göstermiş ve her iki tip veri seti (yağış ve akarsu debisi) içinde bu bileşenin ne kadar etkili olduğunu vurgulamıştır.

Anahtar Kelimeler: Aylık Akarsu Akım Tahmini, Yağış-Akım İlişkisi Modeli, Kopula, TRMM Verileri, Yapay Sinir Ağları Modeli (YSA), Klimatoloji, Tutarlılık

To My Mother…

Anneme…

ACKNOWLEDGMENTS

I would like to express sincere appreciation to my supervisor Assist. Prof. Dr. Mustafa Tuğrul YILMAZ for his guidance, support and critics, for pushing my limits, for trusting me through this research, and giving me the opportunity to work on this amazing project.

I would like to express sincere appreciation to Prof. Dr. Elçin KENTEL for her kind guidance and critics.

I am also thankful to my friends, of Civil Engineering Department, Mehdi Hesami Afshar and Burak Bulut for their kind assist and support.

TABLE OF CONTENTS

ABSTRACT……..……..V
ÖZ..…….....…..VII
ACKNOWLEDGMENTS……...……...……...……..... X
TABLE OF CONTENTS…….……...…...……XI
LIST OF TABLES……..……...……...…….....…..XIV
LIST OF FIGURES……..……..……...…....XVIII
ABBREVIATIONS……..…….…….... XXIII
CHAPTERS
1. INTRODUCTION.....1
1.1. Problem Definition and Motivation of the Study 1
1.2. Literature Review.....3
1.2.1. Review on Earlier Studies on Runoff Prediction Methods.....3
1.2.2. Studies Focusing on Rainfall – Runoff Predictions over Turkey.....8
1.3. Main Purpose of the Study.....9
2. METHODOLOGY.....11
2.1. Prediction Models.....11
2.1.1. Linear Regression Model.……..…....15
2.1.2. Multiple Linear Regression Model.……...…..17
2.1.3. ANN Model ...……....……20
2.1.4. Copula Models……...…….……24
2.2. Climatology and Persistence Benchmarks.……..……...26
2.2.1. Climatology…….....27
2.2.2. Persistence …….....…… 28
2.3. Complete and Standardized Anomaly Data sets Predictions.....…… 29
2.3.1. Complete Data sets Predictions.....30
2.3.2. Standardized Anomaly Data Sets Predictions.....31
2.4. Study Area.....…….....…….....…….....

2.4.1. General Layout of Basin……..……..……..…35
2.4.2. Basin Geography and Slope Properties ……..……..……37
2.4.3. Vegetation Cover and Climatology ….....……...……39
2.4.4. Regulated Regions of the Basin…. ...……...……40
2.5. Data Sets ……..…….....42
2.5.1. TRMM Precipitation Data…….....……44
2.5.2. MGM Precipitation Data…….....45
2.5.3. DSI Streamflow Data….....46
2.6. Runoff Predictions…….....……..……..…....47
2.6.1. Station E23A004……..…….....…..…....48
2.6.2. Station E23A005……..…….....…..…....49
2.6.3. Station E23A016……..…….....…..…....50
2.6.4. Station E23A020……..…….....…..…....51
2.6.5. Station E23A021……..…….....…..…....52
2.6.6. Station E23A023……..…….....……..…....53
2.7. Performance Statistics……..….....54
3. RESULTS and DISCUSSIONS…….....…..…57
3.1. Analysis of Precipitation Data…….....…….57
3.2. Analysis of Streamflow Data…….....…..…58
3.3. Streamflow Predictions…….....……60
3.3.1. Station E23A004…….....…….....61
3.3.1.1. Meteorology Data Predictions…….....……61
3.3.1.2. TRMM Single Pixel Data Predictions …….....…64
3.3.1.3. TRMM Catchment Average Data Predictions ……....67
3.3.2. Station E23A005 …….....…….....70
3.3.2.1. Meteorology Data Predictions…….....……70
3.3.2.2. TRMM Single Pixel Data Predictions …….....…73
3.3.2.3. TRMM Catchment Average Data Predictions ……....76
3.3.3. Station E23A016…….....…….....79
3.3.3.1. Meteorology Data Predictions…….....……79
3.3.3.2. TRMM Single Pixel Data Predictions …….....…82
3.3.3.3. TRMM Catchment Average Data Predictions ……....85

3.3.4. Station E23A020.....88
3.3.4.1. Meteorology Data Predictions.....88
3.3.4.2. TRMM Single Pixel Data Predictions.....91
3.3.4.3. TRMM Catchment Average Data Predictions 94
3.3.5. Station E23A021.....97
3.3.5.1. Meteorology Data Predictions.....97
3.3.5.2. TRMM Single Pixel Data Predictions 100
3.3.5.3. TRMM Catchment Average Data Predictions 103
3.3.6. Station E23A023.....106
3.3.6.1. Meteorology Data Predictions.....106
3.3.6.2. TRMM Single Pixel Data Predictions 109
3.3.6.3. TRMM Catchment Average Data Predictions 112
3.4. Discussion of Predictions.....115
4. SUMMARY and CONCLUSION.....121
REFERENCES 125

LIST OF TABLES

Table 2.1 Coruh Basin Upper and Middle Stream Dams..... ……..…….40

Table 3.1 Correlations Between MGM Stations and TRMM Data sets ……..…….57

Table 3.2 Correlations between the DSI Streamflow Observation Stations …….....……..…58

Table 3.3 DSI Streamflow Observation Stations Data Statistics and Autocorrelation Values …….....……..59

Table 3.4 Performance Statistics of Station E23A004 with Meteorology Data - Complete Data set Predictions ……....…….....……..62

Table 3.5 Performance Statistics of Station E23A004 with Meteorology Data – Anomaly Component Data set Predictions…...…….....…63

Table 3.6 Performance Statistics of Station E23A004 with TRMM Single Pixel Data - Complete Data set Predictions…….....……...65

Table 3.7 Performance Statistics of Station E23A004 with TRMM Single Pixel Data – Anomaly Component Data set Predictions……...……66

Table 3.8 Performance Statistics of Station E23A004 with TRMM Catchment Average Data - Complete Data set Predictions……....……..68

Table 3.9 Performance Statistics of Station E23A004 with Catchment Average Data – Anomaly Component Data set Predictions…….....……...69

Table 3.10 Performance Statistics of Station E23A005 with Meteorology Data - Complete Data set Predictions ……...……..……...71

Table 3.11 Performance Statistics of Station E23A005 with Meteorology Data – Anomaly Component Data set Predictions……..……...72

Table 3.12 Performance Statistics of Station E23A005 with TRMM Single Pixel Data - Complete Data set Predictions…….....…….74

Table 3.13 Performance Statistics of Station E23A005 with TRMM Single Pixel Data – Anomaly Component Data set Predictions……...……75

Table 3.14 Performance Statistics of Station E23A005 with TRMM Catchment Average Data - Complete Data set Predictions……...……....….77

Table 3.15 Performance Statistics of Station E23A005 with Catchment Average Data – Anomaly Component Data set Predictions……...……....78

Table 3.16 Performance Statistics of Station E23A016 with Meteorology Data - Complete Data set Predictions ……...……..……...80

Table 3.17 Performance Statistics of Station E23A016 with Meteorology Data – Anomaly Component Data set Predictions…….....……...81

Table 3.18 Performance Statistics of Station E23A016 with TRMM Single Pixel Data - Complete Data set Predictions……....…….....…….83

Table 3.19 Performance Statistics of Station E23A016 with TRMM Single Pixel Data – Anomaly Component Data set Predictions……...…....84

Table 3.20 Performance Statistics of Station E23A016 with TRMM Catchment Average Data - Complete Data set Predictions……..…....….86

Table 3.21 Performance Statistics of Station E23A016 with Catchment Average Data – Anomaly Component Data set Predictions……...……87

Table 3.22 Performance Statistics of Station E23A020 with Meteorology Data - Complete Data set Predictions ……..…….....89

Table 3.23 Performance Statistics of Station E23A020 with Meteorology Data – Anomaly Component Data set Predictions…….....…….....90

Table 3.24 Performance Statistics of Station E23A020 with TRMM Single Pixel Data - Complete Data set Predictions……...…….....…….92

Table 3.25 Performance Statistics of Station E23A020 with TRMM Single Pixel Data – Anomaly Component Data set Predictions……...…….....93

Table 3.26 Performance Statistics of Station E23A020 with TRMM Catchment Average Data - Complete Data set Predictions……...….....95

Table 3.27 Performance Statistics of Station E23A020 with Catchment Average Data – Anomaly Component Data set Predictions……...…….....96

Table 3.28 Performance Statistics of Station E23A021 with Meteorology Data - Complete Data set Predictions ……....……...98

Table 3.29 Performance Statistics of Station E23A021 with Meteorology Data – Anomaly Component Data set Predictions……...…….....……...…...99

Table 3.30 Performance Statistics of Station E23A021 with TRMM Single Pixel Data - Complete Data set Predictions……....…….....……...…….....101

Table 3.31 Performance Statistics of Station E23A021 with TRMM Single Pixel Data – Anomaly Component Data set Predictions……...…….102

Table 3.32 Performance Statistics of Station E23A021 with TRMM Catchment Average Data - Complete Data set Predictions……...……...104

Table 3.33 Performance Statistics of Station E23A021 with Catchment Average Data – Anomaly Component Data set Predictions……...……...…....105

Table 3.34 Performance Statistics of Station E23A023 with Meteorology Data - Complete Data set Predictions ……..…...……...…...107

Table 3.34 Performance Statistics of Station E23A023 with Meteorology Data – Anomaly Component Data set Predictions…….....……...…...108

Table 3.36 Performance Statistics of Station E23A023 with TRMM Single Pixel Data - Complete Data set Predictions……...…….....……...…...…...110

Table 3.37 Performance Statistics of Station E23A023 with TRMM Single Pixel Data – Anomaly Component Data set Predictions……...…...…...111

Table 3.38 Performance Statistics of Station E23A023 with TRMM Catchment Average Data - Complete Data set Predictions……...…….....113

Table 3.39 Performance Statistics of Station E23A023 with Catchment Average Data – Anomaly Component Data set Predictions……..…….....114

Table 3.40 Summary of the correlations of complete data set predictions with the streamflow validation timeseries...….....……...115

Table 3.41 Summary of the correlations of anomaly component data set predictions with the streamflow validation time series…...….....…...115

LIST OF FIGURES

Figure 2.1 Structure Of The Single Neuron Information Process In The Human Brain……...……...……..13

Figure 2.2 Sample Regression Line and Observed Data Scatter Plot...…….....15

Figure 2.3 General Layout of an ANN Model .....…….……..…20

Figure 2.4 Schematic Diagram of A Single Node in an ANN..…….....…21

Figure 2.5 Monthly Climatology Timeseries…….....…27

Figure 2.6 Input time series of StreamFlow and Precipitation to the models …...…30

Figure 2.7 Complete Data set Predictions …….....…31

Figure 2.8 Climatology and anomaly components of flow data.....….32

Figure 2.9 Streamflow Data Standardized Anomalies …….....…….32

Figure 2.10 Climatology and anomaly components of precipitation data ….....….33

Figure 2.11 Precipitation Data Standardized Anomalies ….....33

Figure 2.12 Climatology and Predicted Anomaly Components …….....…….34

Figure 2.13 Final StreamFlow Climatology Added Prediction …….....…35

Figure 2.14 Location of Coruh Basin …….....…36

Figure 2.15 Coruh Basin …….....…37

Figure 2.16 DEM Map of Coruh Basin …….....38

Figure 2.17 General Layout of Vegetation Cover of Coruh Basin ……....39

Figure 2.18 Sub-Basins and locations of TRMM measurement pixels and land observation stations ……..…43

Figure 2.19 TRMM pixels and Sub-basins …..…. .....…44

Figure 2.20 MGM Meteorological Station Locations over DEM Map …… ..…....45

Figure 2.21 DSI Streamflow Observation Station Locations …….....…46

Figure 2.22 Location map of Station E23A004 and related MGM Station and TRMM Pixels …….....…..…48

Figure 2.23 Location map of Station E23A005 and related MGM Station and TRMM Pixels …….....…..…49

Figure 2.24 Location map of Station E23A016 and related MGM Station and TRMM Pixels …….....…..…50

Figure 2.25 Location map of Station E23A020 and related MGM Station and TRMM Pixels …….....…..…51

Figure 2.26 Location map of Station E23A021 and related MGM Station and TRMM Pixels …….....…..…52

Figure 2.27 Location map of Station E23A023 and related MGM Station and TRMM Pixels …….....…..…53

Figure 3.1 Hypsometric Curve of the Coruh Basin …….....59

Figure 3.2 Station E23A004 DSI Streamflow Data v.s. Meteorology Precipitation Data Complete Data set Predictions …….....…..…..…61

Figure 3.3 Station E23A004 DSI Streamflow Data v.s. Meteorology Precipitation Data Anomaly Component Data set Predictions…….....62

Figure 3.4 Station E23A004 DSI Streamflow Data v.s. TRMM Single Pixel Precipitation Data Complete Data set Predictions.....64

Figure 3.5 Station E23A004 DSI Streamflow Data v.s. TRMM Single Pixel Precipitation Data Anomaly Component Data set Predictions.....65

Figure 3.6 Station E23A004 DSI Streamflow Data v.s. TRMM Catchment Average Precipitation Data Complete Data set Predictions.....67

Figure 3.7 Station E23A004 DSI Streamflow Data v.s. TRMM Catchment Average Precipitation Data Anomaly Component Data set Predictions.....68

Figure 3.8 Station E23A005 DSI Streamflow Data v.s. Meteorology Precipitation Data Complete Data set Predictions …….....……..…70

Figure 3.9 Station E23A005 DSI Streamflow Data v.s. Meteorology Precipitation Data Anomaly Component Data set Predictions…….....71

Figure 3.10 Station E23A005 DSI Streamflow Data v.s. TRMM Single Pixel Precipitation Data Complete Data set Predictions.....73

Figure 3.11 Station E23A005 DSI Streamflow Data v.s. TRMM Single Pixel Precipitation Data Anomaly Component Data set Predictions.....74

Figure 3.12 Station E23A005 DSI Streamflow Data v.s. TRMM Catchment Average Precipitation Data Complete Data set Predictions.....76

Figure 3.13 Station E23A005 DSI Streamflow Data v.s. TRMM Catchment Average Precipitation Data Anomaly Component Data set Predictions.....77

Figure 3.14 Station E23A016 DSI Streamflow Data v.s. Meteorology Precipitation Data Complete Data set Predictions …….....…79

Figure 3.15 Station E23A016 DSI Streamflow Data v.s. Meteorology Precipitation Data Anomaly Component Data set Predictions…….....80

Figure 3.16 Station E23A016 DSI Streamflow Data v.s. TRMM Single Pixel Precipitation Data Complete Data set Predictions.....82

Figure 3.17 Station E23A016 DSI Streamflow Data v.s. TRMM Single Pixel Precipitation Data Anomaly Component Data set Predictions.....83

Figure 3.18 Station E23A016 DSI Streamflow Data v.s. TRMM Catchment Average Precipitation Data Complete Data set Predictions.....85

Figure 3.19 Station E23A016 DSI Streamflow Data v.s. TRMM Catchment Average Precipitation Data Anomaly Component Data set Predictions.....86

Figure 3.20 Station E23A020 DSI Streamflow Data v.s. Meteorology Precipitation Data Complete Data set Predictions…….....…88

Figure 3.21 Station E23A020 DSI Streamflow Data v.s. Meteorology Precipitation Data Anomaly Component Data set Predictions....….....89

Figure 3.22 Station E23A020 DSI Streamflow Data v.s. TRMM Single Pixel Precipitation Data Complete Data set Predictions.....91

Figure 3.23 Station E23A020 DSI Streamflow Data v.s. TRMM Single Pixel Precipitation Data Anomaly Component Data set Predictions.....92

Figure 3.24 Station E23A020 DSI Streamflow Data v.s. TRMM Catchment Average Precipitation Data Complete Data set Predictions.....94

Figure 3.25 Station E23A020 DSI Streamflow Data v.s. TRMM Catchment Average Precipitation Data Anomaly Component Data set Predictions.....95

Figure 3.26 Station E23A021 DSI Streamflow Data v.s. Meteorology Precipitation Data Complete Data set Predictions …….....…97

Figure 3.27 Station E23A021 DSI Streamflow Data v.s. Meteorology Precipitation Data Anomaly Component Data set Predictions…….....98

Figure 3.28 Station E23A021 DSI Streamflow Data v.s. TRMM Single Pixel Precipitation Data Complete Data set Predictions.....100

Figure 3.29 Station E23A021 DSI Streamflow Data v.s. TRMM Single Pixel Precipitation Data Anomaly Component Data set Predictions.....101

Figure 3.30 Station E23A021 DSI Streamflow Data v.s. TRMM Catchment Average Precipitation Data Complete Data set Predictions.....103

Figure 3.31 Station E23A021 DSI Streamflow Data v.s. TRMM Catchment Average Precipitation Data Anomaly Component Data set Predictions.....104

Figure 3.32 Station E23A023 DSI Streamflow Data v.s. Meteorology Precipitation Data Complete Data set Predictions ……106

Figure 3.33 Station E23A023 DSI Streamflow Data v.s. Meteorology Precipitation Data Anomaly Component Data set Predictions.....107

Figure 3.34 Station E23A023 DSI Streamflow Data v.s. TRMM Single Pixel Precipitation Data Complete Data set Predictions.....109

Figure 3.35 Station E23A023 DSI Streamflow Data v.s. TRMM Single Pixel Precipitation Data Anomaly Component Data set Predictions.....110

Figure 3.36 Station E23A023 DSI Streamflow Data v.s. TRMM Catchment Average Precipitation Data Complete Data set Predictions.....112

Figure 3.37 Station E23A023 DSI Streamflow Data v.s. TRMM Catchment Average Precipitation Data Anomaly Component Data set Predictions.....113

Figure 3.38 Overall Average Model Correlations of Complete Data set Predictions 117

Figure 3.39 Overall average model correlations of Anomaly Component Data set Predictions..... 117

ABBREVIATIONS

ANN: …….…….....….Artificial Neural Network

DSI: ……....…….....….General Directorate of State Hydraulic Works

FC: ……...…Frank Copula

JAXA: …….……...…….….Japan Aerospace Exploration Agency

MGM: …….……..……..Turkish State Meteorological Service

MLP: …….……..Multi-Layer Perceptron ANN

MLR: …….….Multiple Linear Regression

NASA: …….……...…National Aeronautics and Space Administration

NC: …….……...……..…Normal Copula

SLR: …….....……....…….……...…...Simple Linear Regression

TRMM: ……...Tropical Rainfall Measurement Mission

CHAPTER 1

INTRODUCTION

1.1 Problem Definition and Motivation of the Study

Streamflow predictions in a specific catchment are desirable and have a major role in many industrial and scientific fields such as flood risk management for flood-sensitive areas, energy optimization and planning purposes for hydroelectric energy production, irrigation water management for agricultural fields, droughts, culture fishing, etc. Therefore, acquisition of reliable predictions of streamflow has a crucial importance, especially for human life and economy. Improved hydrological predictions are useful tools for management of water resources, disaster planning, and agriculture.

Hydropower is the most important renewable energy source in the world. Approximately 17% of world's electricity demand is supplied by the hydropower IEA, 2006). In Turkey, 32% of the electricity demand is supplied from the hydropower (EPDK, 2012). In this aspect efficient use of water resources is critical for the economy. Under many uncertain hydrologic and climatologic conditions, hydropower energy optimization and reservoir operation become very challenging. Having knowledge about the discharge of the river provides a vision for planning and maximization of the energy generation. Optimal hydropower energy generation in dams can be achieved with different approaches. Constructing a rule curve for optimal reservoir operation is one of them where rule curves describe how much storage must be maintained in a reservoir at different times in a year to ensure that discharge requirements can always be met (U.S. Army Corps of Engineers, 1985). Therefore,

having skillful streamflow data predictions are very important in long-term reservoir operations and management in the context of hydropower generation.

Floods are one of the most threatening disasters of all as it may cause loss of lives and property very quickly. Misuse of the land, destruction of the forest for a variety of reasons, wrong selection of the settlement and industrial area, infrastructure failure, rapid population growth, the corrupted balance between soil, water and plants are contributing factors to the damage of the floods, particularly in developing countries. Every year, over one million people are affected by floods and 40% of the damage of goods in the world is caused by the floods (World Disaster Report (2012), The International Federation of Red Cross and Red Crescent Societies). Therefore, skillful streamflow data predictions are very critical for accurate flood predictions and early warning systems.

However, the hydrological cycle is one of the most complex processes to understand. The composition of streamflow contains many different variables within itself e.g., precipitation, groundwater, soil moisture which may have to be estimated accurately over large regions. Most particularly precipitation rate is strongly non-Gaussian, discontinuous and at most weakly predictable at larger time scales (Yılmaz 2010) and runoff predictions are very difficult due to the existence of many uncertainties.

It has been around 40 years since remote sensing-based data sets have been used in hydrological data analysis studies. Today, many of the satellites can observe the earth surface for the purpose of retrieving information about rainfall, runoff, soil moisture content etc., while observation of the same variables using ground stations could be impractical. On the other hand, the added utility of remote sensing data in hydrological applications should be explicitly tested.

Stations are seemed to be the most accurate way of obtaining ground through data. However, it may not always be feasible to obtain good quality station-based data sets that represent large regions. On the other hand, remote sensing-based data sets are in many aspects very useful in estimation of ground conditions. It is known that predicted streamflow values obtained only rainfall data may not yield accurate predictions because streamflow composition has many other components like groundwater, soil

moisture, evaporation rate, snow melting etc. Models that are using only one component as an input cannot explain the complete variability. Despite, no precipitation falls during dry periods of the year, streamflow can be still observed, while the variability in streamflow may not be explained with only rainfall component. Potentially snow melting or groundwater feeds the streamflow on these periods. Addition of other data sets into the data driven approach may increase the model accuracy but the availability of additional data sets might be problematic to acquire over certain regions.

Actually, streamflow data itself in long-term comprises that unexplainable phenomenon in itself due to the strong seasonality. Climatology (long-term mean) component for a certain month or a season of streamflow can explain the phenomena because streamflow value in dry periods comprises snow melting or groundwater data. It is thought that, in a sort of way, the addition of climatology component of streamflow as an input to a rainfall-runoff prediction model will significantly increase the model accuracy.

1.2 Literature Review

1.2.1 Review on Earlier Studies on Runoff Prediction Methods

Streamflow prediction models based on precipitation data can be classified according to randomness, spatial variation, and time variability. These models can be grouped as deterministic, stochastic or statistic, neural network, conceptual or simplified physical models, and distributed physical models. Because the data sets might be problematic to acquire for constructing and calibrating conceptual and physical models (Kokkonen and Jakeman, 2001), data-driven hydrological methods have been begun to use in the field of hydrology (Besaw et al. 2009).

Conceptual models are simplified versions of hydrological processes (Chiew and McMahon, 1994). Physics-based rainfall-runoff models need huge amounts of data to train, validate and (VanderKwaak and Loague, 2001). Multiple linear regression

(MLR) and artificial neural networks (ANNs) are commonly used conceptual methods (Wang et al., 2008).

Various methods were developed in the past to simulate and predict the streamflow as the composition of historical rainfall and runoff recordings. Method accuracy is gaining importance, the core of method selection depends on the availability and quality of the historical data. In the following section, a literature review on the application of various regression analysis, copula methods, and ANN modeling for some hydrological researches are summarized.

Regression-based rainfall-runoff models (SLR, MLR) are simple data driven models which make a prediction from the historical data. Regression techniques provide a fast way of calculating the precipitation-based streamflow predictions.

Garen (1992) proposed several techniques to obtain accurate results from regression models which are using only known at forecast time, cross validation, principal components regression and systematic searching for optimal or near-optimal combinations of variables.

Post et al. (1999) constructed a daily lumped conceptual rainfall-runoff model which consisted of 16 small catchments. For many reasons like the variability of catchment topology, temperature etc., results were varying from catchment to catchment.

Rasouli et al. (2011) used a Global Circulation Forecasting model (NOAA), climate indices with local meteorologic - hydrologic observations to make a comparison study on streamflow prediction with three models; a neural network model (Bayesian NN), a support vector model, and a Gaussian model. They compared the performance of these models with a multiple linear regression model. As a result, they showed that non-linear models generally outperformed. On the other hand, they showed that the best streamflow predictions come from local observations.

ANN has been widely accepted as useful tools only during the 1990's (Callan, 2006). In this chapter, some of the applications of ANNs as a Rainfall-Runoff process calculation is summarized.

Minns et al. (1996) used a 3-layer feed-forward type ANN with a series of numerical experiments to test the closeness of fit that can be achieved. Models with one and two hidden layered ANN showed that an extra hidden layer improves the performance of model (Minns et al.,1996).

In the case study of Lineares-Rodriguez et al. (2015) a model for one day-ahead daily streamflow with ANN approach is presented. Different from early works, they used a new runoff index to test ANN input. This index obtained by combining two different variables which belong to the weather research and forecasting mesoscale model. ANN model simulation skills and model accuracy improved with this new index. With a persistence index (PI) of 0.81 and an R^2 value of 0.95 they proved that their model performance was satisfactory.

Birikundavyi et al. (2002) investigated the performance of neural networks as potential models capable of forecasting daily streamflow. By the mean square error and Nash coefficients-based comparison they showed that artificial neural networks results are superior to the ones obtained with a classic autoregressive model coupled with a Kalman filter.

Budu (2014) used artificial neural networks and multiple linear regression models with wavelets and moving average methods to predict daily reservoir inflows. In the study, real time observations are decomposed into subseries using different wavelet NN functions with discrete wavelet transform and then the proper subseries are taken as inputs to NN for forecasting reservoir inflow. Calibration is made by 7 years of data, and remaining data are used for the validation. The comparison made by the Nash–Sutcliffe efficiency coefficient, RMSE, and correlation coefficient with two NN models, one MLR model, and one moving average model. Results showed that wavelet model performance was the best and multiple linear regression approaches had efficient inflow hydrograph.

Garbrecht (2006) made Monthly Rainfall-Runoff Simulation and compared three alternative ANN design to investigate the effect of seasonal rainfall-runoff variations. The main difference of designs is the way in which the effects of seasonal climate and runoff variations were incorporated into the network. The important conclusion to the

investigation is that a separate network for each month reflects the seasonal rainfall runoff variations, ends up with better results (Garbrecht, 2006).

A copula is a multivariate distribution function. Recently, Copula functions have been a useful tool for various problems in hydrology and water resources studies and applications. Especially in the field of drought analysis copulas gives accurate results. Besides, copula functions have a significant role in the fields that flood frequency analysis, rainfall intensity-duration frequency analysis.

Lee et al. (2008) studied Copula usage in stochastic streamflow simulation. Different copula functions used in order to utilize non-parametric and parametric functions for fitting the distribution of the real time observed data and serial dependence structure. The advantages and disadvantages of different copula time series models were investigated by comparing the statistics of the generated data. The results showed that the benefits of using the copula functions are only marginally different from other method results.

Samaniego et al. (2010) find a procedure to measure dissimilarity that is estimated from coupled empirical copula functions of streamflow to calculate a metric. By using a multiscale parameter regionalization technique hydrologic model parameters were regionalized. Transfer function parameters were found via calibration of the model. The streamflow in an ungauged basin was found as an ensemble streamflow prediction. Reasonable results are obtained with the proposed technique in ungauged basins.

Sugimoto et al. (2016) used copulas as a tool to investigate the effect of climate change with existing long-term discharge records. They used copula asymmetry and copula distances to identify the nonsymmetrical property of discharge data and variability and interdependency. As a result of the study, they specified the climatic regime change periods of Neckar catchment.

Bezak et al. (2016) studied flash floods and rainfall. Copula-based IDF curves and empirical rainfall thresholds are used in the study. By using high-resolution rainfall data, they constructed an IDF curve on the Frank copula function for several rainfall

stations. After the analyses that showed rainfall characteristics triggering flash floods and landslides are different.

Persistence of streamflow indicates that sequences of river flow in the historical record that are similar to the recent past provide valuable information in the near future (Svensson, 2014). In the literature, there are several studies which are using persistence approach as a benchmark. Dralle et al. (2015), Glenn et al. (2004), Owens et al. (2003) made comprehensive discussion in their studies. Persistence approach is generally used for calculating the climatological indicators. Some meteorological studies of Namias J. (1952) and Van Den Dool (1982) are good examples on this topic.

Svensson (2014) studied on seasonal river flow forecast by using hydrological persistence and historical flow analogs. Results show that the forecasts based on the persistence of the previous month's flow generally outperform the analogs approach in her study.

Climatology of the streamflow indicates the mean of specific time periods, gives insights about the seasonal variations. In this study, monthly mean of historical streamflow data is used. Mean is the most commonly used statistical starting point of comparison.

One of the probable future work of this study considered as the effect of regulations on the streamflow predictions. Data used in this study is chosen in unregulated time periods. In this aspect study of Sayama et al. (2006) must be mentioned. In their study, investigation of alteration of the hydrological cycle caused by environmental change according to the actual conditions would not be qualified for make predictions in an ungauged basin. The motivation of the study was to investigate the effects of dam reservoirs to the flow regime changes. They end up such a conclusion that human factors make an alteration in hydrological cycle in a catchment and flood potential. This makes the real time observations not *directly* applicable for future predictions.

1.2.2 Studies Focusing on Rainfall-Runoff Predictions over Turkey

A study from Turkey was made by Bakis and Goncu (2015). They used regression analysis as a tool for filling the missing data of Zab River obtained from the flow gauging stations related to the correlations and drainage area-ratio method. As a result, they concluded, when compared with the observed data; both filling methods resulted in large deviations.

Kisi (2003) used an ANN model and an Auto-Regressive Model to forecast daily river flow, investigated the appropriate ANN architecture for hydrologic modeling for Filyos Stream with a 2-year period of streamflow data. Three different ANN models are used and the results are compared with an AR model. In the study, the comparison showed that ANNs provide the better results for daily streamflow forecasting.

Kentel (2009) used past precipitation and associated river flow data in an ANN algorithm for Guvenc Basin and made predictions of future river flows. Also in her study, she investigated impacts of various input patterns, a number of training cycles, and initial values assigned to the weights of the connections identified direct mapping from inputs to outputs without consideration of the complex relationships among the dependent and independent variables of the hydrological process. Also, Kentel (2009) used a fuzzy c-means algorithm to cluster the training and validation input vectors with regular and extreme events. Main reason behind this usage is that the user will have an idea about the risk of the ANN model to generate unreliable results.

Another ANN study on rainfall-runoff modeling is performed by Gumus et al. (2010). In this study, authors have assessed the ability of MLR, ANN model to predict the runoff over Euphrates River Basin using gauged based precipitation data. They obtained promising results with ANN model and concluded that ANN models can be easily used in rainfall-runoff modeling.

Kara and Yildiz (2014) performed a bivariate analysis of the relation between precipitation and runoff over Hirfanli Dam Basin. Two elliptical and three Archimedean copula functions have been tested to model the dependence structure between the hydrological variables. They used different graphical tools and numerical techniques for the appropriate model selection, parameter estimation and goodness of

fit. Their results indicated that Clayton and Student-t class copula functions are more appropriate to model the dependence structure between the precipitation and the runoff over their study area.

1.3 Main Purpose of the Study

The main purpose of this study is to investigate the predictability of runoff using precipitation data over Coruh Basin using Simple and Multiple Linear Regression, ANN, and Copula methods and their predictive skills are compared against the commonly used benchmarks, climatology- and persistence-based predictions. More specifically in this study;

- Intercomparison of the predictive skill of simple linear regression, multiple linear regression, ANN, and copula methods is done.
- The predictive skills of methods are compared against climatology and persistence benchmarks to better understand the added skill via these methods, quantifying the predictive skill of climatology components compared to the anomaly components and filling the gap between the inter-comparison of Copula, climatology- and persistence-based predictions of streamflow over Turkey.
- The predictions using above methods are made both the remote – sensing and station-based precipitation data to quantify the added utility of remote sensing-based precipitation in runoff predictions compared with the predictions using gauge-based precipitation data. Moreover, by this investigation, added utility of areal average precipitation data sets to single point station data in streamflow predictions are performed and compared.

CHAPTER 2

METHODOLOGY

2.1. Prediction Models

In this study, runoff predictions utilizing precipitation data are made. A Simple Linear Regression Model (SLR) and a Multiple Linear Regression Model (MLR) is constructed to simplify the relationship between the rainfall-runoff process and to get benefit from the ability to gain simple results of the regression models. An Artificial Neural Network Model (MLP) is constructed to get benefit from its ability to deal with the complexity and learning algorithm. Two Copula Models (Normal Copula and Frank Copula) are constructed to get benefit from its ability to separate marginal distribution from the dependent nature of the variables. All of the results that gained from these models will be compared with hydrologic Climatology and Persistence characteristics of the monthly state of streamflow. The following section provides brief information about the history of data driven prediction models which are used in this study

Regression analysis is the most common analysis method in hydrology. Analysis based on a relation made between a dependent variable and one independent variable. Analysis process shows the impact of changes in an independent variable on the dependent variable that we wish to predict. It is also a tool for explaining the relations of a variable is affected directly by another or both variables affected by common factors that are unknown (Web-1). Regression analysis can be used for the filling of the missing data in a series on a polynomial relationship between the observations or validation of the series (Fung, 2006).

The term “Regression” firstly used at the eighteen century by Francis Galton, who has many studies on genetics. While he was studying on sweet pea plant and trying to explain the phenomenon about the relationship between heights of the descendants of the plant, the study laid the foundations of modern linear regression analysis. After Galton, his lab researcher Karl Pearson developed the correlation concept. An Early form of his work was the least squared error approach and he stated that the main problem in regression analysis is to establish most suitable description between the variables (Stanton, 2001).

The method of least squares is a procedure to determine the best fit line to data. The method gets its name from the way the unknown statistics are computed. The method estimates parameters by minimizing the sum of the squared deviations between the variables. Results are obtained using basic calculus and specifically the property of a quadratic expression’s minimum value. The point must be considered in the least squared estimates is its high sensitivity to the outliers. This is a consequence of squaring the variations between the variables (Encyclopedia of Research Design, vol.1, Method of Least Squares, 2010).

Artificial Neural Networks (ANN) are widely used prediction models in hydrological studies. Methodology originally comes from the struggle that understanding the information processing of the human nervous system. Despite there are still too many unknowns about the information processing of the human brain, neural network approach to problem-solving tries to mirror that process of the neurons. A neuron collects the information from other neurons with a structure called dendrites. Then a neuron sends this information signal as an output through the structure called axon, every neuron connects to many other neurons through synapses. Synapses convert the signal from the axon into electrical signal form and these electrical signals send to the brain for the purpose of processing by connections of thousands of neurons. Every neuron can be thought as a single processing section. By changing the effect of a neuron over the other, learning occurs. (Stufflebeam,2008) Figure 2.1 shows the structure of the single neuron information process in the human brain.

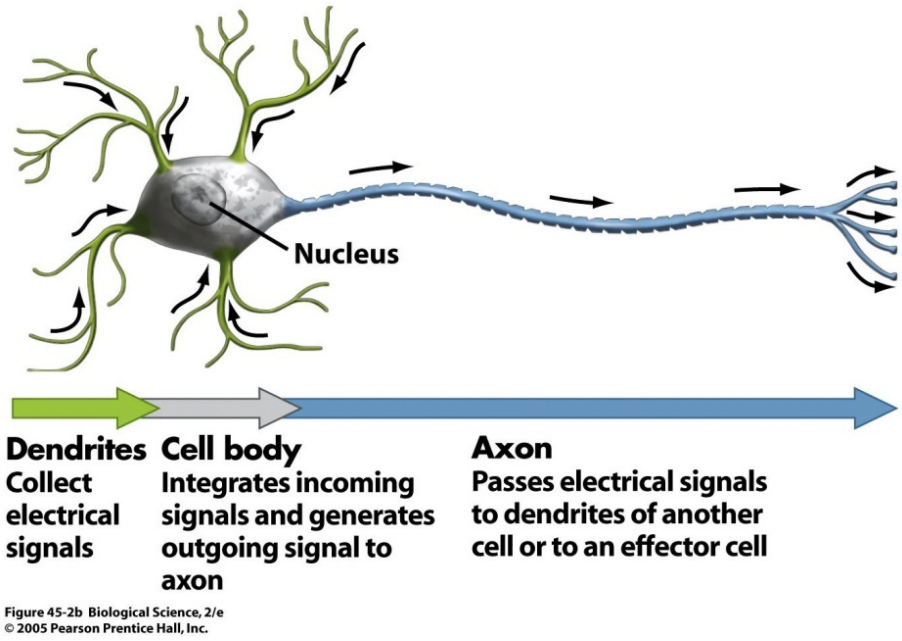


Figure 2.1 Structure of a single neuron information process in the human brain

Early understandings about neural networks belong to neurophysiologist Warren McCulloch (1943). After computers become advanced in the 1950s, it was possible to build some models which could imitate human neural network interactions and at 1980s ANN models started to use much scientific research. ANNs are the calculation tool of information through interactions among neurons (or nodes). Model runs inputs and outputs as variables of the nonlinear function. Mirroring the natural behavior of human neural network system ANN has the advantages of learning ability. Ability to response as a result of input data. Synapses in the biology known as the weights in the ANN. Weights are constants that use to calculate the output of the process. It is a practical tool to carry out complex hydrologic calculations (Reingold & Nightingale, 1999, History of Neural Networks).

ANN Models have a withstanding success in the use of rainfall-runoff modeling. A multilayer perceptron is the most commonly used ANN method which gives satisfactory results in modeling hydrologic process to predict the short run streamflow discharge.

Copula theory first appeared in Frechet's study (1951). On his study about the given distribution function of two random variables defined on the same probability space, the bivariate distribution function of marginal of this two random variables. In Sklar's 1959 study resulted in very important conclusions by constructing a new class of functions which he called as "Copula" for Frechet's work (Durante and Sempi, 2009). The copula is a latin word for "link".

Sklar and Schweizer published the book "Probabilistic Metric Spaces" in 1983. This book considered the main source of basic information on copulas. At the end of the 90's Copula Theory become popular (Durante and Sempi, 2009). Since the beginning of the 2000's Copulas have become a tool for hydrological studies especially in the field of drought analysis. Over a decade many important results still are to be discovered and derived (Salvadori, 2007).

The predictive skills of these methods have to be compared against a benchmark so that their true skill can be understood. In this study, Climatology and Persistence of monthly state of streamflow are used as benchmarks. To further investigate the source of the predictive skills of these methods, separate predictions are made using the standardized anomaly components of data sets (after -climatology/long year monthly mean- components are removed and standardized by dividing the standard deviation of the data) and complete data sets (normal/non-standardized data sets retaining both anomaly and climatology components). 80% of the data is used for training and 20% of the data is used for validation. Long data sets (42 years of observations) are divided into 34 years of training and 8 years of validation and short data sets (12 years of observations) are divided into 9 years of training and 3 years of validation

2.1.1. Linear Regression Model

Many problems in engineering and science involve exploring the relationships between two or more variables. While established relations are mostly used to predict the dependent variable using the independent one. Least squared based Linear Regression method (1806 Legendre, 1810 Gauss) is perhaps the oldest yet one of the most used prediction methods.

Regression analysis is a method to find a linear relationship between two variables. Figure 2.2 (also known as scatter plot) shows the linear relationship between the variables “streamflow” on the horizontal axis and “precipitation” on the vertical axis. The simplest relationship consists of a straight line.

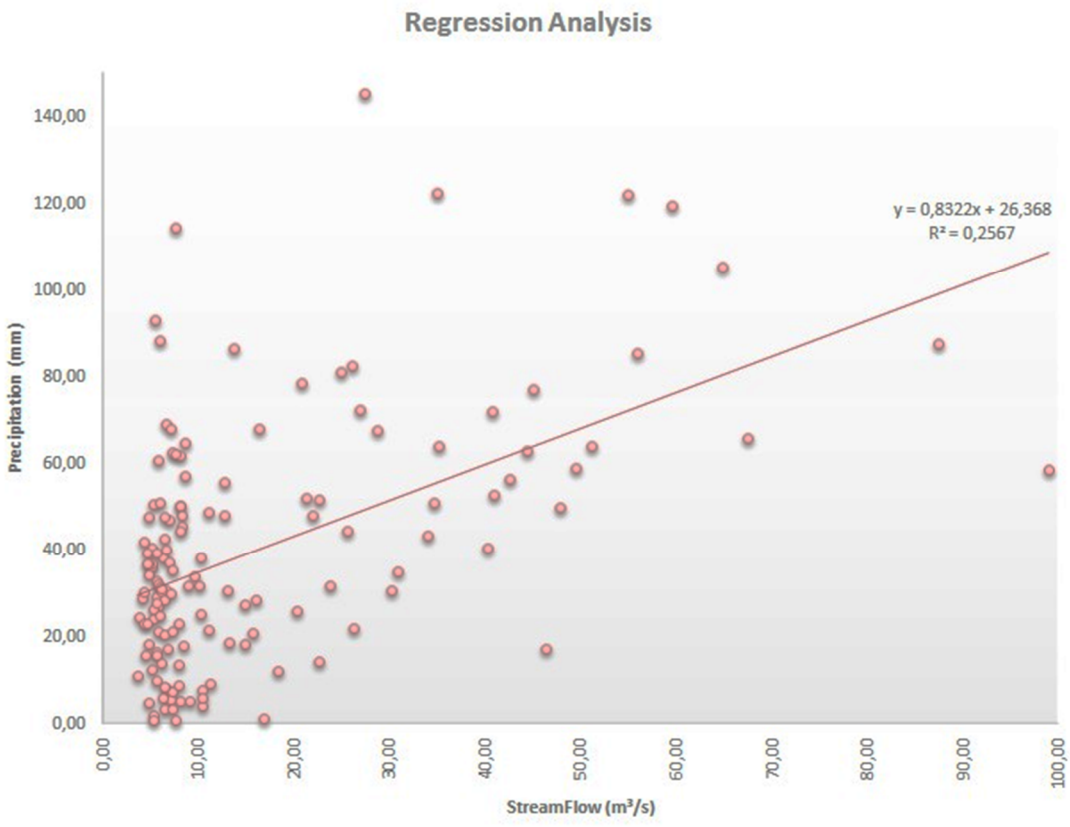


Figure 2.2 Sample regression line and observed data scatter plot

Regression calculations find the optimum values of the regression parameters (slope and intercept) that minimize the sum of the squared errors. A form of the regression formula is as follows (Levine et al. 2013).

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \quad (2.1)$$

Where; β_0 is intercepted for the population, β_1 is the slope for the population, ε_i is the random error in predictions *i,* Y_i is the dependent variable (sometimes referred to as the response variable) for observation *i* and X_i is the independent variable (sometimes referred to as the explanatory variable) for observation *i*

The solution for the calculation of the slope β_1,

$$\beta_1 = \frac{SSXY}{SSX} \quad (2.2)$$

where;

$$\text{SSXY} = \sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y}) \quad (2.3)$$

$$= \sum_{i=1}^{n} X_i Y_i - \frac{(\sum_{i=1}^{n} X_i)(\sum_{i=1}^{n} Y_i)}{n}$$

$$\text{SSX} = \sum_{i=1}^{n}(X_i - \dot{\bar{X}})^2 \quad (2.4)$$

$$= \sum_{i=1}^{n} X_i{}^2 - \frac{(\sum_{i=1}^{n} X_i)^2}{n}$$

Computation of the intercept,β_0,

$$\beta_0 = \bar{Y} - \beta_1 \bar{X} \tag{2.5}$$

where,

$$\bar{Y} = \frac{\sum_{i=1}^{n} Y_i}{n} \tag{2.6}$$

$$\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} \tag{2.7}$$

2.1.2. Multiple Linear Regression Model

As in the SLR model, again the purpose in MLR is to obtain least square estimates which minimize the total sum squared error term. When more than one predictor variable included in the calculations, effects of all variables can be observed from the predictions. Once the relations between the multiple variables and the dependent variable we can make more accurate predictions and use all the information. Due to lack of data, in this study, squared values of precipitation is used as a second predictor.

General MLR model (Shedden, 2016);

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon_i \tag{2.8}$$

where β_0 is the "intercept," β_1 …… β_k are "slopes" or "coefficients". These coefficients are fixed or constant but unknown values. ε is a random variable that is independent of Xs with "0" mean and variance of σ^2. Errors are not correlated with each other and not correlated with other predictors. Estimates can be calculated by using linear algebra;

By taking all observation into an n-dimensional vector, called “response vector” (Shedden, 2016);

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_n \end{pmatrix} \tag{2.9}$$

Then by putting all predictors into an n.p+1 matrix, design matrix is obtained;

$$X = \begin{bmatrix} \begin{pmatrix} 1 & X_{11} & X_{1p} \\ 1 & X_{21} & X_{2p} \\ . & . & . \\ 1 & X_n & X_{np} \end{pmatrix} \end{bmatrix} \tag{2.10}$$

It is important that first column of the design matrix be formed of 1s. After that packing the coefficients and intercepts into a p+1 dimensional matrix, we obtain a vector called slope vector.

$$\beta = \begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ . \\ \beta_p \end{pmatrix} \tag{2.11}$$

At last, we put the error terms into an n-dimensional another vector called error vector.

$$\varepsilon = \begin{pmatrix} \varepsilon \\ \varepsilon_1 \\ \varepsilon_2 \\ . \\ \varepsilon_n \end{pmatrix} \tag{2.12}$$

Model is become $Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k + \varepsilon_i$, can be written in algebraic form of ;

$$Y_i = X\beta + \varepsilon_i \tag{2.13}$$

Note that $X\beta$ is the matrix vector product. For estimating the β like in linear regression least squares approach will be used. By minimizing the error formulation;

$$\sum_i \left(Y_i - \alpha - \beta_1 X_{i,1} - \cdots \beta_p X_{i,p}\right) \tag{2.14}$$

Obtained formulation is,

$$\hat{\beta} = (X'X)^{-1}X'Y \tag{2.15}$$

Where $X'X$ and $(X'X)^{-1}$ are p+1 x p+1 symmetric vectors. $X'Y$ is a p+1 dimensional vector. Then,

$$\hat{Y} = X\hat{\beta} = |\mathrm{X}(X'X)^{-1}X'Y \tag{2.16}$$

$\hat{Y}$ is the best prediction obtained from the multiple linear regression calculations.

2.1.3. ANN Model

ANN Models mimic the neurons in the human brain. Connections generally called nodes or units. Information processing occurs in form of signals through these connections. Each unit has its own weight which shows the strength of the participation of the node into the calculation. Also, the weight of a node determines the effect of the variable on the process. In the process, a linear or nonlinear transformation function is applied to the nodes to determine its output signal. In this study, Multi-Layer Perceptron (MLP) is used. Figure 2.3 shows General Layout of an ANN Model (ASCE, 2000).

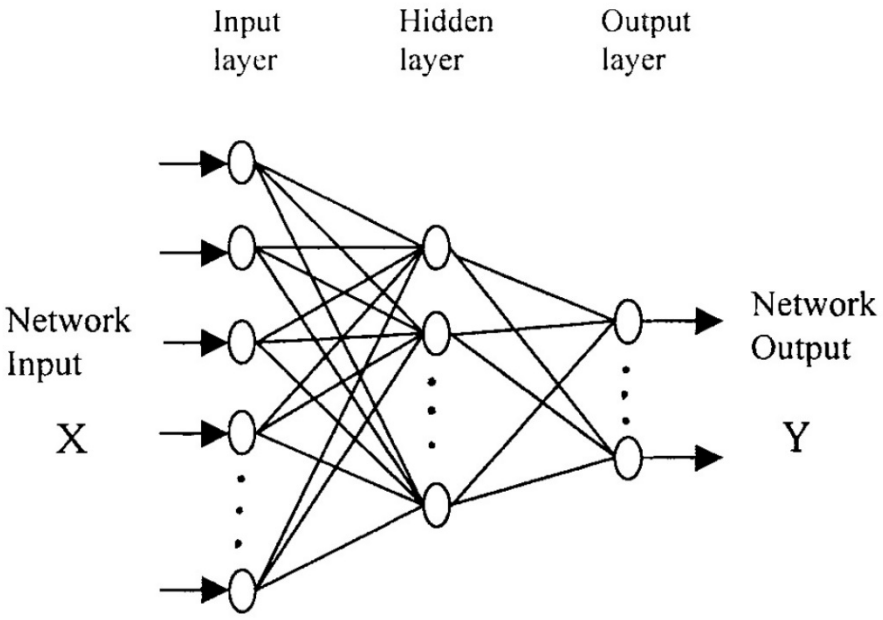


Figure 2.3 General layout of an ann model (ASCE, 2000)

Figure 2.4 shows the schematic diagram of a single node, b_j is the threshold value or generally called bias. W_{ij}s are weights and X_is are inputs for the node. f is the threshold

function and lastly y_j is the net output signal of the node. The below function defines the operation.

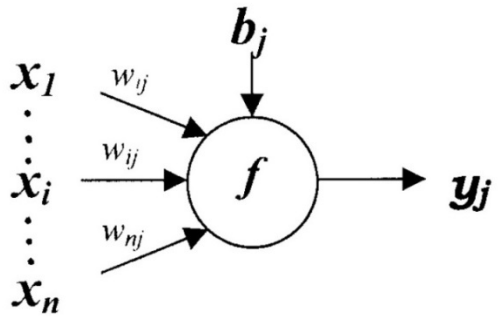


Figure 2.4 Schematic diagram of a single node in an ann (ASCE, 2000)

$$Y_j = f(X.W_j - b_j) \tag{2.17}$$

The function f in the figure is called an activation function or threshold function. Commonly used form of an activation function is sigmoid function and general form of a sigmoid function is as follows;

$$f_t = \frac{1}{1+e^{-t}} \tag{2.18}$$

MLP uses the back propagation algorithm for training, which is the most popular neural network type. In the algorithm, weights are iteratively changing to minimize the errors of the weights. The main strength of a network come to this capability of training. First, the input of the data set passes through the network then calculated output is compared with the measured or observed output. On this stage error calculation made and then this error term propagates backward through the network to each node. Weights are recalculated according to adjustment of the error term of each node.

Step by step calculations of a Feed Forward MLP algorithm as follows (ASCE, 2000), Input unit enters the process from the input layer and receives the signal, x_i, (i = 1, 2, …..n) and sends this signal to the next layer. Every hidden unit (H_j, j =1, 2,….. p) receives their weighted input values.

$$Hin_j = v_{oj} + \sum x_i v_{ij} \tag{2.19}$$

Where v_{ij} is the connection weight and v_{oj} is the bias value, perform a sigmoid function to calculate their output value;

$$H_j = f(Hin_j) \tag{2.20}$$

Then these values transfer to the following layer. Each output nodes (Y_k, where k = 1, 2, ...m) sums its weighted input signals,

$$Yin_j = w_{ok} + \sum h_i w_{jk} \tag{2.21}$$

and then again performs its threshold function to compute output signals.

$$Y_k = f(Yin_k) \tag{2.22}$$

At this stage Back-Propagation error calculated. Every output nodes that receive the desired value regarding the input value takes its error term.

$$\delta = (t_k - y_k) f'(Yin_k) \tag{2.23}$$

Weight correction and bias correction factors calculated,

Weight correction factor,

$$\lambda w_{jk} = \partial \delta_k H_j \tag{2.24}$$

Bias correction factor,

$$\lambda w_{0k} = \alpha \delta_k \tag{2.25}$$

Calculated δ_k error information term sends to nodes in the previous layer. Each hidden unit gets its delta inputs from next layer,

$$\delta in_j = \sum \delta_k w_{jk} \tag{2.26}$$

and multiplies by the derivative of threshold function for calculation of error term. After that again weight correction term to update v_{ij} and bias correction term to update v_{oj} are calculated.

$$\delta_j = \delta in_j f'(Hin_j) \tag{2.27}$$

$$\nabla v_{ij} = \partial \delta_j x_i \tag{2.28}$$

$$\nabla v_{oj} = \alpha \delta_j \tag{2.29}$$

Each output node restores by adding these corrections (ASCE, 2000).

New weights,

$$w_{jk}(new) = w_{jk}(old) + \lambda w_{jk} \tag{2.30}$$

New Bias,

$$v_{ij}(new) = v_{ij}(old) + \lambda v_{ij} \tag{2.31}$$

After new weights and biases are calculated error term between the desired output and recalculated one compare to each other, if error term is not in the acceptable range, correction of weights and biases are iterated respectively.

2.1.4. Copula Models

Copulas have wide usage areas, such as risk management, biology and recently used in hydrology and water resources management related fields. Especially in the field of drought analysis and flood risk analysis, there are many promising studies made by using copulas.

Copulas are simply the joint distribution of random variables, and also an extremely beneficial technique that allows forming a multivariate distribution function from univariate distribution functions (Bloomfield, 2013). Copula theory uses the marginal distributions to create a joint distribution. This is the main power of the Copula Theory.

The core of the Copula Theory comes from the Sklar's theorem (Bloomfield, 2013). Given two random variables X and Y, estimation of their univariate CDFs are trivial and the sample data sets are often sufficient, while the analytical solution for the joint CDFs is not immediately clear. At this point, copula functions link univariate CDFs and create their multivariate (e.g., bivariate or trivariate) CDFs (Afsar, 2016). For variables *X, Y*; $F_{XY}(X,Y)$ represents the joint cumulative distribution, $F_X(X)$ and $F_Y(Y)$ are the marginal distribution of *X* and Y, there exist a unique copula $C(F_X(X), F_Y(Y))$ (De Michelle, 2003).

$$F_{XY}(X,Y) = C(F_X(X), F_Y(Y)) \tag{2.32}$$

If there are N different random variables than general form of the formula can be shown (De Michelle, 2003),

$$F_{X_1X_2\ldots\ldots\ldots X_N}(X_1, X_2 \ldots\ldots X_N) = C(F_{X_1}(X_1), F_{X_2}(X_2) \ldots\ldots\ldots F_N(X_N)) \tag{2.33}$$

Copula Density :

Assume $F(X_1, X_2 \ldots\ldots X_N) = C[F_{X_1}(X_1), F_{X_2}(X_2) \ldots\ldots\ldots F_N(X_N)]$ and *F* and *C* are differentiable, joint probability density function of events (De Michelle, 2003),;

$$\frac{F(x_1,x_2,\ldots\ldots x_N)}{F_1(x_1)F_2(x_2)\ldots\ldots F_N(x_N)} = C\left[F_{X_1}(X_1), F_{X_2}(X_2) \ldots\ldots\ldots F_N(X_N)\right] \tag{2.34}$$

There are many classes of copulas available like Archimedean, Elliptic, Extreme Value etc. In this study, Normal Copula Function from the Elliptical copula family and Frank Copula Function from the Archimedean Copula Family are used.

Normal Copula (NC) Function:

$$C^N(x,y) = \frac{F(x,y)}{F_x(x)F_y(y)} = \int_{-\infty}^{\emptyset^{-1}(x)} \int_{-\infty}^{\emptyset^{-1}(y)} \frac{1}{2\pi(1-\rho^2)^{1/2}} e^{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}} dxdy \quad (2.35)$$

Where; ρ is the copula dependence parameter, x and y are two dependent univariate variables, $\emptyset$ is the CDF of standard univariate gaussian distribution, respectively.

Frank Copula (FC) Function:

$$C^F(x,y) = \frac{F(x,y)}{F_x(x)F_y(y)} = \frac{-1}{\alpha}\ln(1 + \frac{(e^{-\alpha X}-1)(e^{-\alpha Y}-1)}{e^{-\alpha}-1}) \quad (2.36)$$

Where; α is the copula dependence parameter, x and y are two dependent univariate variables, respectively.

2.2. Climatology and Persistence Benchmarks

In the absence of reliable predictors, a good estimate that can be made is the mean of the random variable. For example, for a Gaussian white noise process with "0" mean and "0" autocorrelation, given the availability of the entire time series of this variable, the best guess for the future predictions is the mean value (here "0"). Another example; given only monthly streamflow observations are available over a region between January 1901 and December 2000. A climatological prediction for January 2001 could be obtained as the average of 100 values of all January observations between 1901 and 2000.

In this way, a separate prediction can be made for each month, while these monthly predictions would be constant for any given year. Such a prediction would be valuable particularly over regions with strong seasonality as a result of snow melting,

groundwater feed or seasonally changing rainfall variability. There except for streamflow data, no other data available is present.

2.2.1. Climatology

Climatology type forecast is a simple way of making hydrologic predictions. It is making forecasts by using simple statistics. The idea behind the climatology method is forecast the same value for every event that occurs in the same time periods (prediction for month "A" is the average of all observations of "A"s), the mean.

(Web-2 can be checked on this topic).

Precipitation is random in time for a region, so as a result of randomness in hydrologic cycle streamflow has a tendency of vary over time. If there is no other prediction about streamflow, climatology concept gives the knowledge of basic variation of the streamflow over time. In lack of information, knowing the monthly streamflow variability provides efficient information. Figure 2.5 shows an example for monthly climatology time series.

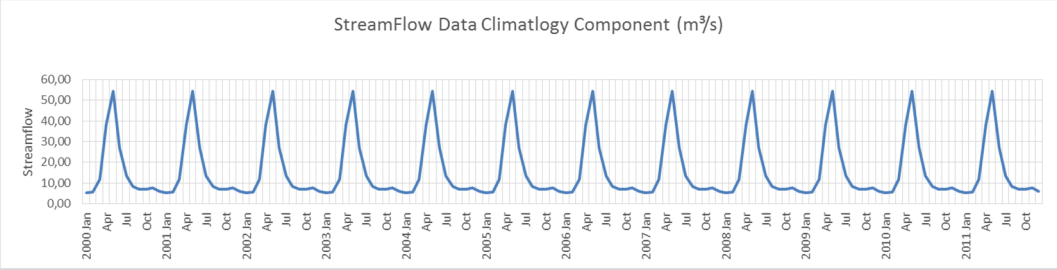


Figure 2.5 Example of a monthly climatology time series (E23A004)

General formula of Mean,

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n} \tag{2.37}$$

Where, x_i are the random values of same time periods, n is the number of observations.

2.2.2. Persistence

Among all of the forecasting methods, Persistence method is the easiest and most applicable one because the method predicts the tomorrow's hydrological event as same as today's event.

A disadvantage of persistence approach is it may not catch relationship between events as a rainfall-runoff model if the distribution of events has so much of variations in time. Persistence concept is very useful for short term predictions, but main power of persistence is to make long range forecasts when it used with the climatology of the event. In this study, observed standardized anomalies of the previous month is used to predict following month's streamflow predictions (following month's observed standardized anomaly taken as the following month's observed standardized anomaly) (Web-3 can be checked on this topic). The following section provides brief information about calculation of persistence,

First step, calculation of the Mean μ,

$$\mu_{i,t} = \frac{\sum_{i=1}^{n} x_{i,t}}{n} \qquad t = 1,2,3....12 \tag{2.38}$$

Second step, calculation of the Standard Deviation,

$$\sigma_{i,t} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(x_{i,t} - \mu_{i,t}\right)^2} \tag{2.39}$$

Third step, finding standardized anomalies of streamflow (Svensson,2014),

$$a_{i,t} = \frac{x_{i,t} - \mu_{i,t}}{\sigma_{i,t}} \tag{2.40}$$

Fourth step, calculation of persistence-based prediction,

$$p_{i,t+1} = a_{i,t} * \sigma_{t+1} + \mu_{t+1} \quad (2.41)$$

Where, $p_{i,t+1}$ is the following month's persistence-based prediction, $a_{i,t}$ is the standardized anomaly of the previous month, μ is the mean of the monthly observations, σ_{t+1} is the standard deviation of following month, n is the number of observations.

For example;

$$p_{i,february} = a_{i,january} * \sigma_{february} + \mu_{february} \quad (2.42)$$

In general, autocorrelation coefficient greater than 0.50 for a streamflow time series, it is expected that the persistence approach would provide better results than the climatology approach.

2.3. Complete and Standardized Anomaly Data Set Predictions

To further investigate the source of the predictive skills of these methods, separate predictions are made using the standardized anomaly components of data sets (after - climatology/long year monthly mean- components are removed and standardized by dividing the standard deviation of the data) and complete data sets (normal/non-standardized data sets retaining both anomaly and climatology components). 80% of the data is used for training and 20% of the data is used for validation. Long data sets (42 years of observations) are divided into 34 years of training and 8 years of validation and short data sets (12 years of observations) are divided into 9 years of training and 3 years of validation. Following section provides detailed information about complete

data set prediction methodology and standardized anomaly component data set predictions.

2.3.1. Complete Data Set Predictions

In this prediction methodology traditional way in Rainfall-Runoff models are followed. First input data sets (DSI Flow Station data sets v.s. MGM and TRMM Precipitation data sets) are introduced to models (SLR, MLR, ANN, Copulas). Figure 2.6 shows the input time series of streamflow and precipitation data sets to the models. In the presentation, only 12 years (2000-2011) of data is shown as a time series but in the calculations, 42 years (1970-2011) of data is used. Figure 2.6 shows an example input time series in traditional rainfall-runoff models.

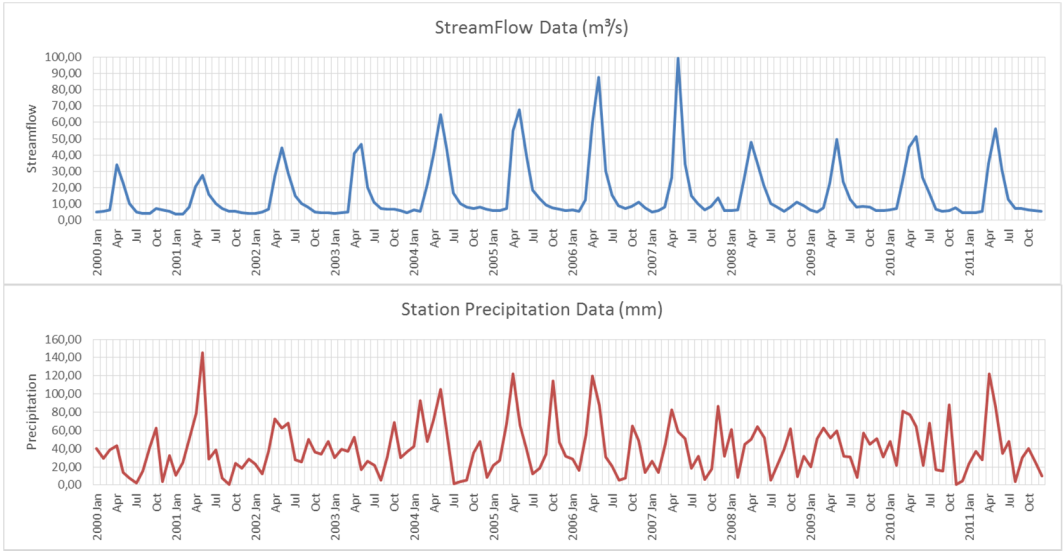


Figure 2.6 Example input time series of streamflow and precipitation to the models

80% of the data (34 years – 408 months) is used for training and 20% of the data (8 years – 96 months) of the data is used for validation.

Finally, predictions compared with climatology and hydrologic persistence of streamflow. Figure 2.7 shows an example of the complete data set predictions (results which are obtained from the data of this study as an example). In this figure, 3 years

of real observations, Climatology and Persistence validation time series and SLR and MLR validation time series are shown in the figure.

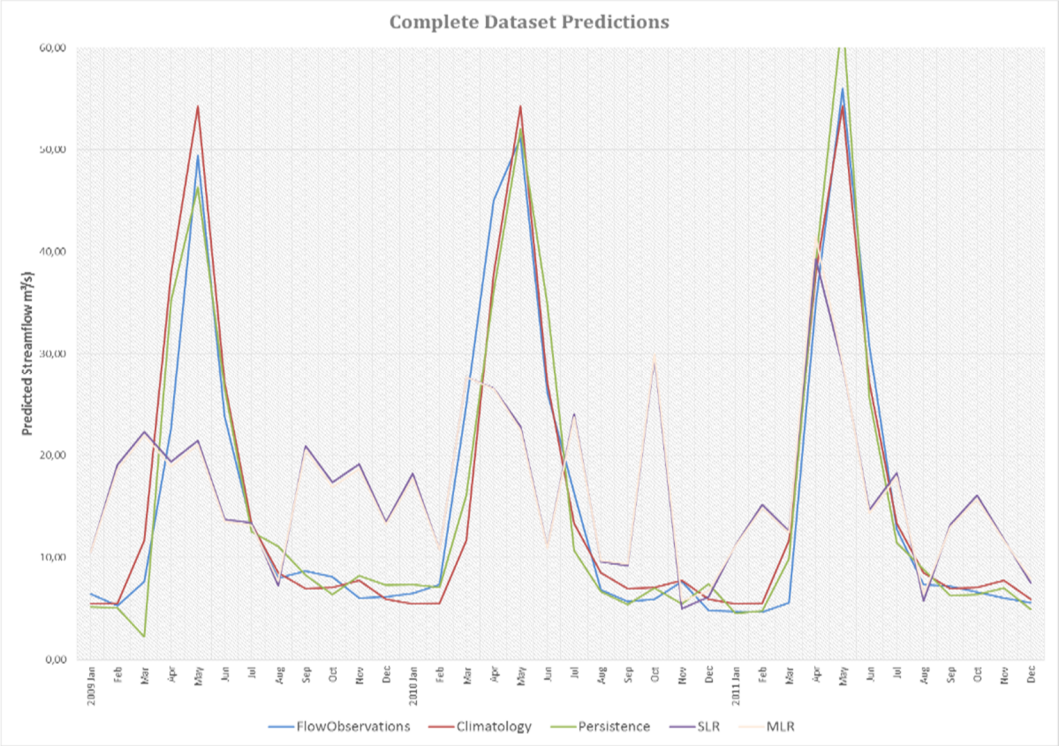


Figure 2.7 Example Complete Data set Predictions

2.3.2. Standardized Anomaly Data Set Predictions

In this methodology, predictions made with standardized anomaly components of data sets after –climatology (long year monthly mean), components are removed and standardized by dividing the standard deviation of the data. By this way, predictions are made with the deviations of the data. First monthly mean of 42 years of data (b) (12 years for short data sets) is calculated. Then, this monthly means of long years (climatology component) subtracted from the complete data set (a) (observations). Hereby, anomaly component of data has been calculated (c). Figure 2.8 shows an example of complete data and its climatology and anomaly components (i.e., complete dataset = climatology + anomaly).

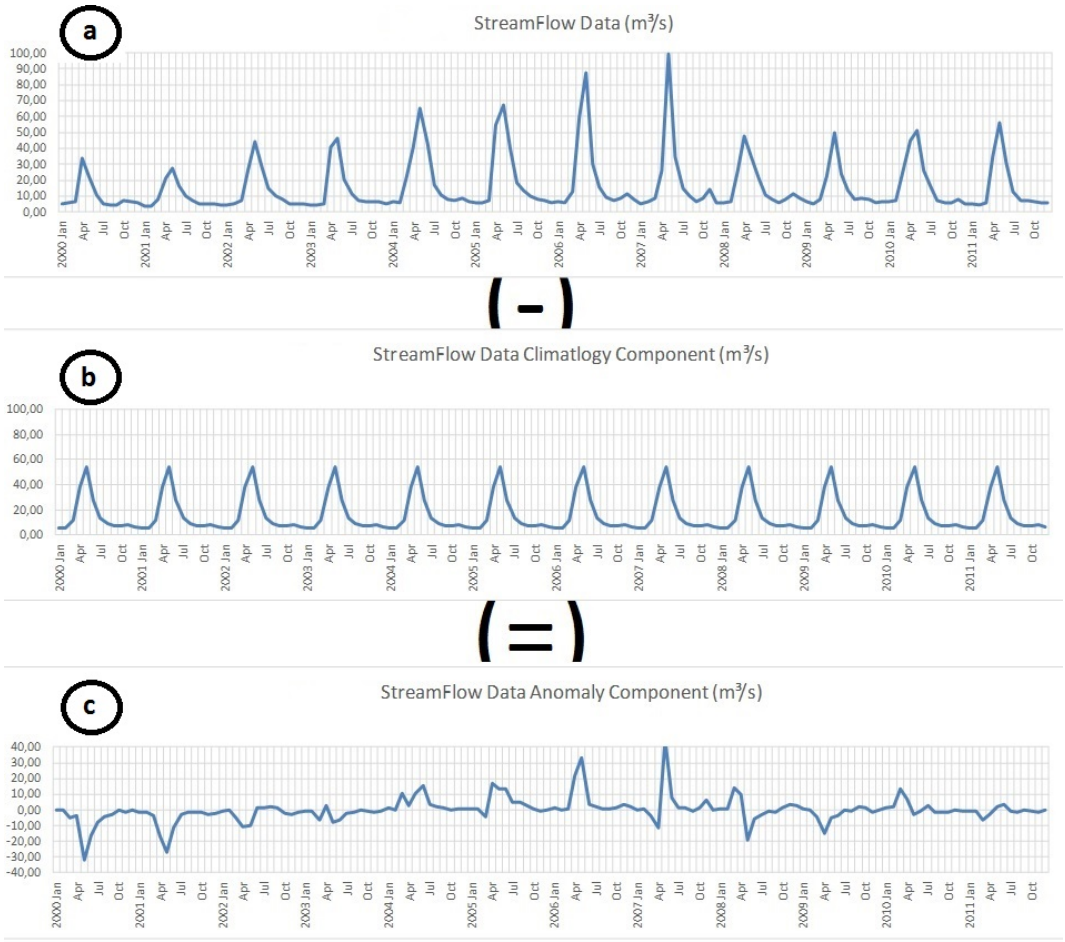


Figure 2.8 Climatology (b) and anomaly (c) components of complete (a) streamflow time series

Then anomaly component of the data is divided by the standard deviations of the data and standardized anomalies of flow data have been obtained. Figure 2.9 shows the standardized anomaly component of flow data.

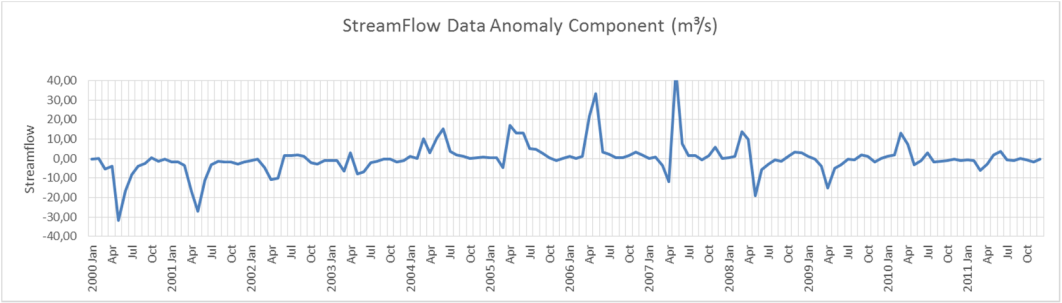


Figure 2.9 Streamflow data standardized anomalies

Same calculations made for the precipitation data as well. Figure 2.10 and Figure 2.11 shows calculations of climatology and anomaly components of precipitation data and precipitation anomalies of data (c = a - b).

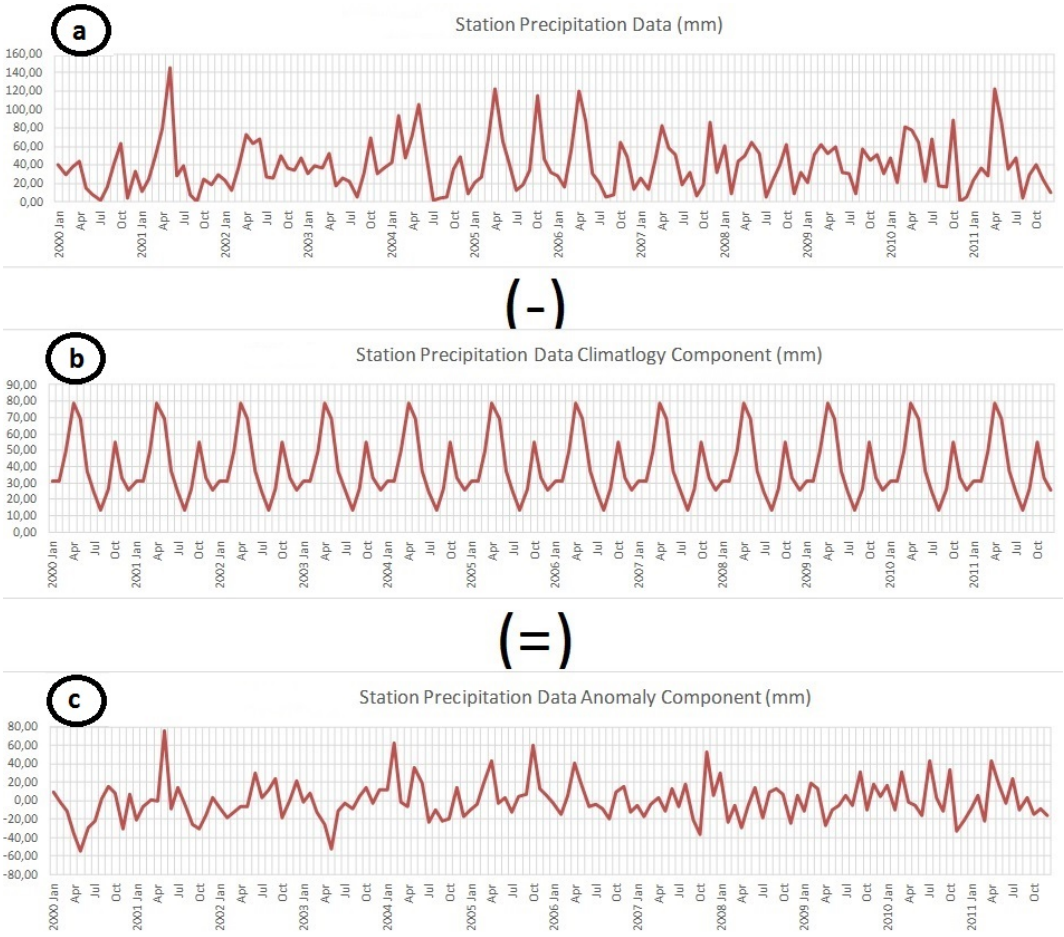


Figure 2.10 Climatology (b) and anomaly (c) components of complete (a) precipitation time series

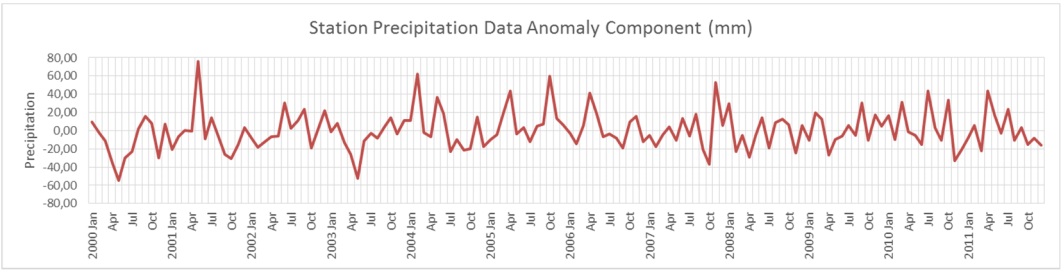


Figure 2.11 Precipitation data standardized anomalies

Once standardized anomalies of precipitation and flow data are obtained, in this methodology, these standardized anomalies introduced the models (SLR, MLR, ANN, Copulas) instead of the complete data sets. After predicted standardized anomalies of flow data obtained from models, next step is de-standardization. Predicted anomalies of flow data are de-standardized with multiplying with the standard deviation of the flow data. Figure 2.12 shows the acquiring steps of final predictions (only SLR output is used in the presentation).

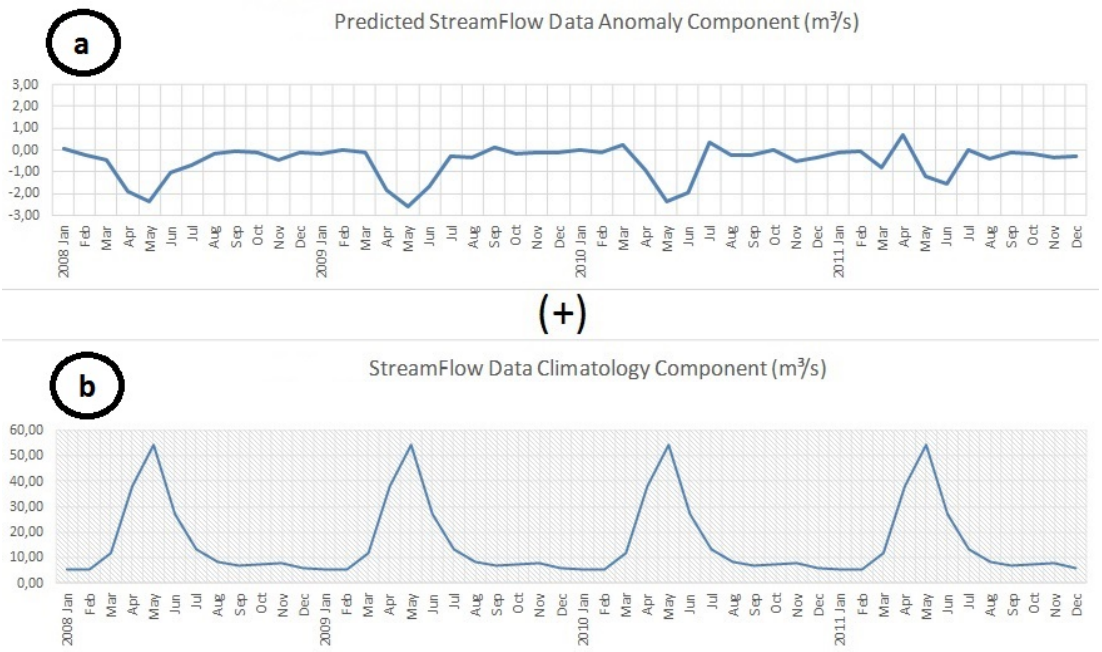


Figure 2.12 Anomaly (a) and climatology (b) components of predicted streamflow

Lastly, by adding the climatology component of flow data to predicted de-standardized anomalies of flow data (a+b), calculations are terminated. Figure 2.13 shows the SLR model final prediction with observations, climatology component, and persistence-based prediction together (results which are obtained from the data of this study as an example). Results show that predictions of SLR model which are using standardized data are improved when long-term climatology values added.

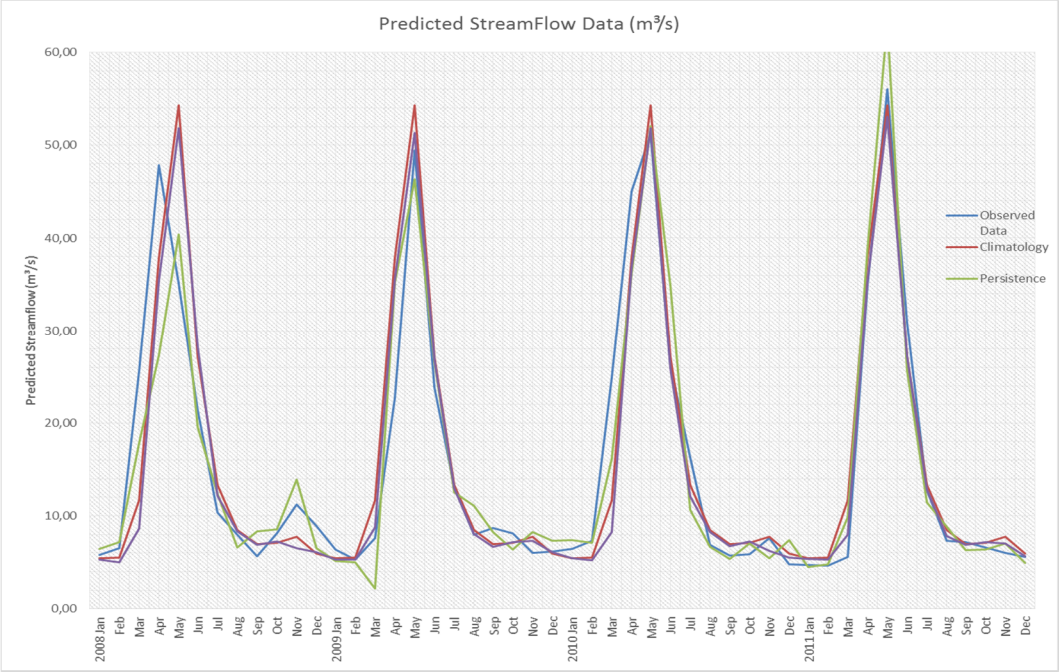


Figure 2.13 Example final streamflow climatology added prediction

2.4. Study Area

2.4.1. General Layout of Basin

Coruh River Basin located in northeastern region of Turkey. The basin has 19.872km^2 drainage area in Turkey and 2.090km^2 drainage area in Georgia (Web-4). Figure 2.14 shows the location of the Coruh Basin in Turkey. As a result of the steep elevation change in the basin, the Coruh River is the fastest flowing river in Turkey and the third fastest in the world.

The total length of the Coruh River (412km) and the steep topography change provides the basin one of the highest electricity generation potential in Turkey (27% of the total energy potential; Coruh River Basin Projects Report, 2009). In this study, Coruh Basin is selected as the study area because of its high hydropower generation potential (i.e., therefore prediction of streamflow is of interest for many users) while unregulated streamflow observations and long historical datasets are still available over the basin despite this potential.

Figure 2.14 Location of Coruh Basin

Coruh River is born from Civilikaya Hill, part of the Mescit Mountains and is located in the north of the Erzurum Plateau. Coruh River later flows through the Eastern Anatolian and Eastern Black Sea region that flows into the Black Sea near the city of Batumi, Georgia (Web-4). The total length of Coruh River is 431km, while 410km flows through Turkey 21km flows in Georgia. Tortum and Oltu streams are the main streams of the Coruh River from Turkey and Adzharis from Georgia. Coruh River has a mean annual runoff of 6.824hm^3. Observed mean discharge of the river is 200m^3/s, maximum discharge is 2.431m^3/s and minimum discharge is 38 m^3/s (Web-4). Coruh basin has the highest erosion potential among all of the basins of Turkey. This high erosion potential is driven by the speed of the river, which is one of the fastest flowing rivers in the world. This study covers the upper and middle stream of the Coruh River and one tributary river called Berhal Creek. Figure 2.15 shows the general layout of the Coruh Basin.

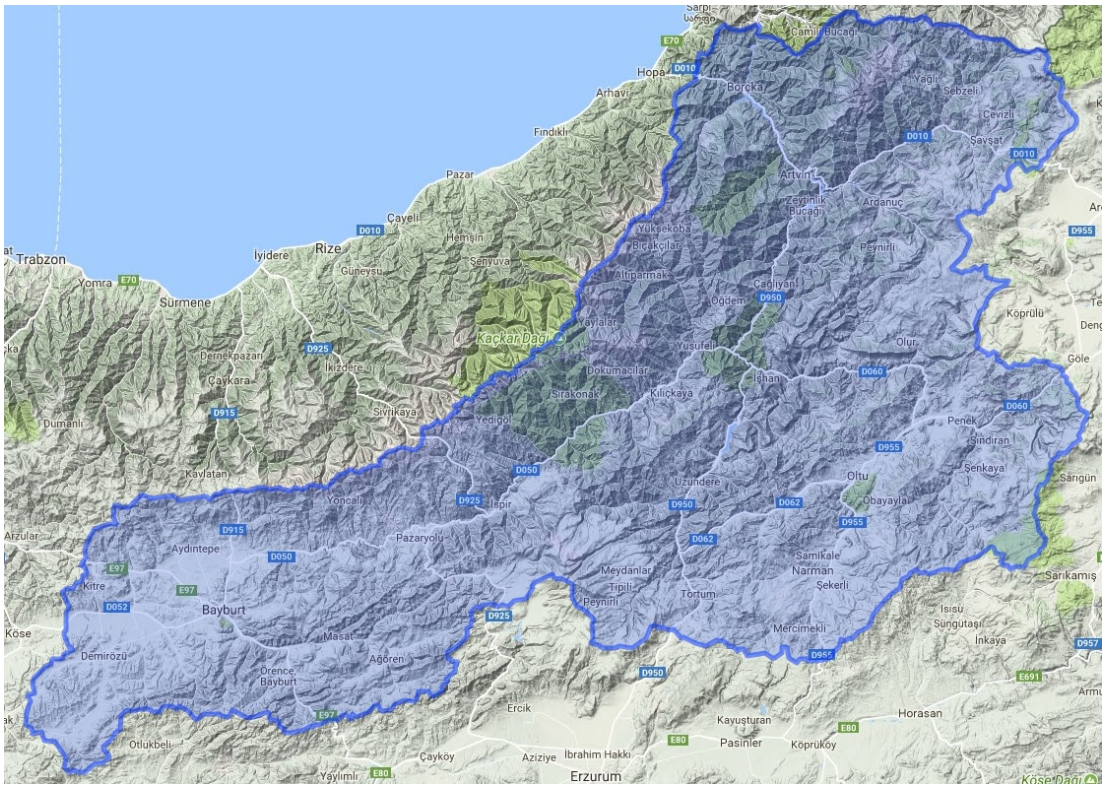


Figure 2.15 Coruh Basin

2.4.2. Basin Geography and Slope Properties

Coruh Basin is located in the orogenic belt. Toward north-northeast lies the black sea mountains and in the line of northeast-southwest lies the Mescit and Yalnizcam Mountains. From east to the west direction this mountain has a formation which reaches 4000m height due to east direction. Morphology of the mountains is generally formed of metamorphic rocks, mica schists, quartzite, granite, and schist. Coruh Basin mountains subside from the north of the Kackar Mountains. Highest mountains of the region as follows, Karadag (2300m), Arafek (2300m), Karayol (2750m), Ziyaret (2000m), Geberet (2413m), Horasan (2830m), Kurdevan (3050m), Ayvan (2000m), Arsiyan (3160m), Morete (2500m), and Karcal (3428m). This mountainside has deep valley formation where the Coruh River lies through this deep valley formation. (Figure 2.16)

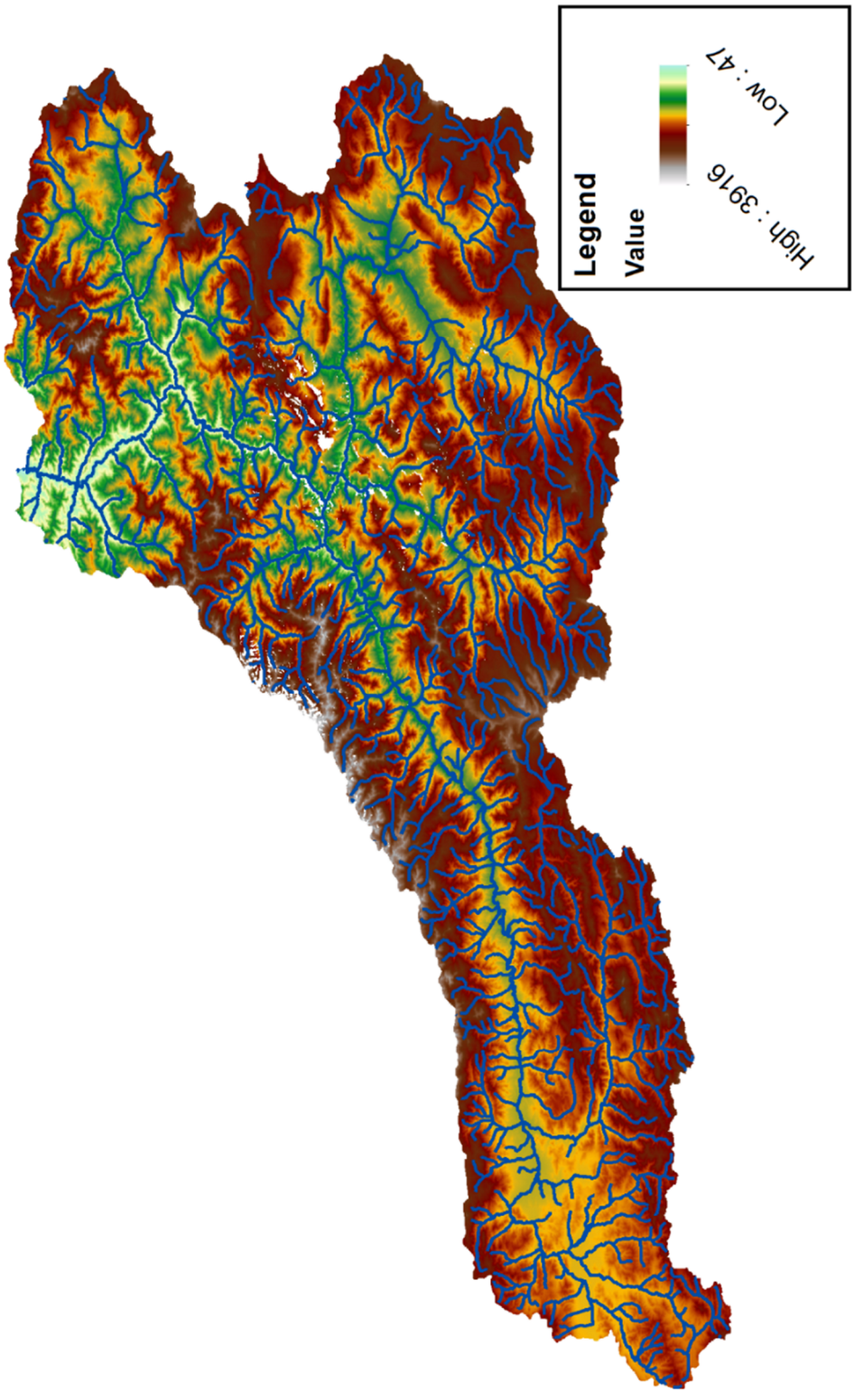


Figure 2.16 DEM Map of Coruh Basin

2.4.3. Vegetation Cover and Climatology

The Northern region of the basin has the characteristic of the black sea climate. Locations, that near East-Black Sea Region like Hopa and Kemalpasa have high annual precipitation. Average annual precipitation of this region is measured as 2754mm. Maximum streamflow values occur in May while maximum precipitation occur in April and October for any given year in Coruh Basin. Toward inland, the climate is in transition to continental climate characteristics. Some annual average meteorological station measurements are as follows; Borcka (1250mm), Artvin (689mm), Ardanuc (446mm), Yusufeli (295mm), Ispır (440mm), and Oltu (353mm) (Web-4). Figure 2.17 shows the general layout of the vegetation cover and heights of the basin.

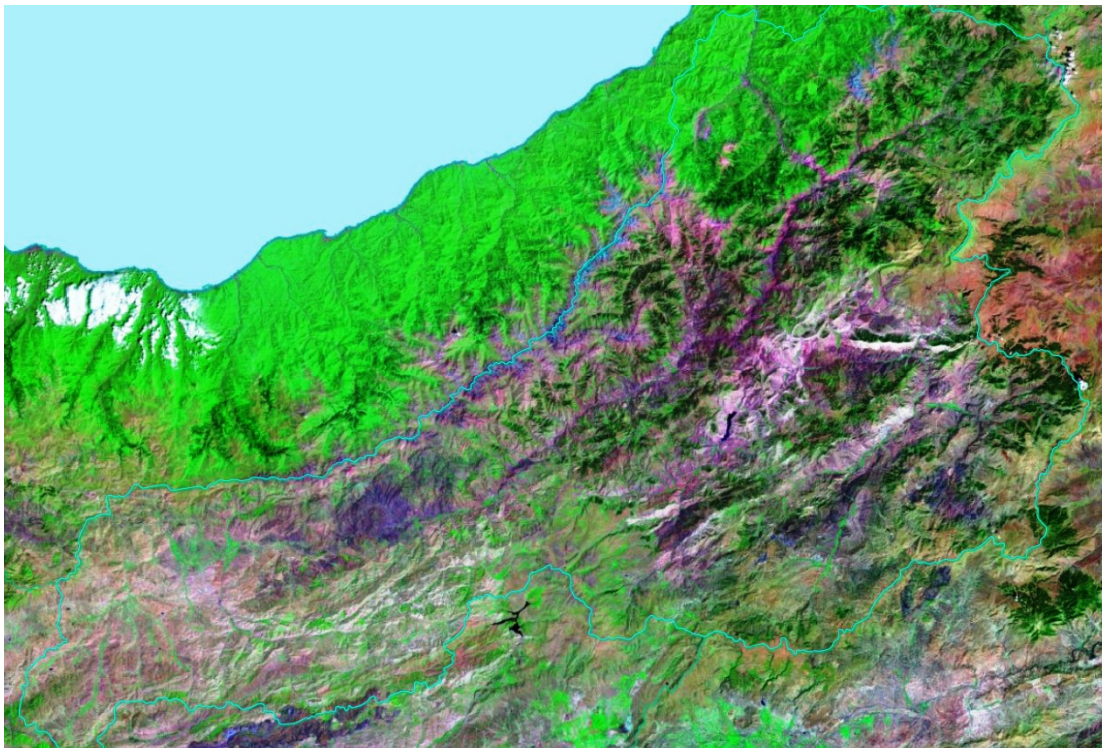


Figure 2.17 General layout of vegetation cover of Coruh Basin

Coruh Basin located in Colchic Region of Euro-Siberian Region Flora Zone. Mediterranean originated endemic taxons are also seen between the 200-400 m elevations (Web-4). Forest formation of basin spreads as, perhumid mild broad-leaved trees (Chestnut Forests, Beech Forest, and Mountain Alder Forests), moist and cold sandarac trees, dry forest and some bush formations.

2.4.4. Regulated Regions of the Basin

Coruh River, with the remarkable energy potential, has been attracted attention and many projects have been developed since 1962 by the Electrical Works Study Administrations (EİE). The energy potential of the river was calculated as %27 of the total energy potential of Turkey. Total head of the river that can be used to product energy is 1420m. There are 10 dams on the main stream (6 of them locates upper and middle sections of the stream) and 17 different dams and runoff-river type hydroelectric power plant projects on the tributary streams. These dams and power plants will regulate the river flow in the basin when the ongoing projects will be completed (Coruh River Basin Projects Report, 2009).

Table 2.1 gives brief information about the upper and middle stream dams. The following section provides information about the regulated regions and dams on the river.

Table 2.1 Coruh Basin upper and middle stream dams

Dams	Annual Mean Flow (m^3/s)	Drenaige Area (km^2)	Total Head (m)	Maximum Operation Level (m)	Installed Capacity (MW)	Annual Energy Production (GWh)
Laleli	879,16	4759,00	138,00	1480,00	102,00	244,55
Ispir	950,00	5100,00	195,00	1342,00	54,00	327,50
Gullubag	1288,00	5915,00	105,00	1147,00	96,00	284,50
Aksu	1492,00	6388,00	107,00	1042,00	130,00	344,40
Arkun	1743,56	6853,00	225,00	953,00	245,00	788,40
Yusufeli	3995,00	15250,00	196,30	712,00	558,00	1705,00

Upper Stream Projects of the Basin;

Laleli (Under Construction) Dam and HEPP:

Laleli Dam is the first dam on the river which is located in Bayburt province. The annual mean flow of the dam section is 879,16m^3 and area of the precipitation are 4759.00km^2. Total head of the project is 138.00m and annual planned capacity is 244.55GWh energy (Coruh River Basin Projects Report, 2009).

Ispir (2014) Dam and HEPP:

Ispir Dam is the second dam on the river which is located on the downstream of the Laleli Dam in Erzurum province. The annual mean flow of the dam section is 950.00m^3 and area of the precipitation are 5100.00km^2. Total head of the project is 195.00m and annual planned capacity is 327.50GWh energy (Coruh River Basin Projects Report, 2009).

Gullubag (2013) Dam and HEPP:

Gullubag Dam is the third dam on the river which is located in Erzurum province. The annual mean flow of the dam section is 1288.00m^3 and area of the precipitation are 5915.00km^2. Total head of the project is 105.00m and annual planned capacity is 284.50GWh energy (Coruh River Basin Projects Report, 2009).

Aksu (Under Construction) Dam and HEPP:

Aksu Dam is the fourth dam on the river which is located in Erzurum province. The annual mean flow of the dam section is 1492.00m^3 and area of the precipitation are 6388.00km^2. Total head of the project is 107.00m and annual planned capacity is 344.40GWh energy (Coruh River Basin Projects Report, 2009).

Arkun (2014) Dam and HEPP:

The fifth dam of the river, Arkun Dam, is located downstream of the Aksu Dam and located in Erzurum province. The annual mean flow of the dam section is 1743.56m^3 and area of the precipitation are 6853.00km^2. Total head of the project is 225.00m and annual planned capacity is 788.40GWh energy (Coruh River Basin Projects Report, 2009).

Middle Stream Projects of the Basin;

Yusufeli (Under construction) Dam and HEPP:

The first dam of the Middle Stream of the Basin, Yusufeli Dam, is located downstream of the Arkun Dam and located in Artvin province. The annual mean flow of the dam section is 3995.00m^3 and area of the precipitation are 15,250.00km^2. Total head of the project is 196.30m and annual planned capacity is 1705.00GWh energy (Coruh River Basin Projects Report, 2009).

2.5. Data Sets

In this study monthly runoff discharge predictions are made over 6 streamflow observation stations maintained by General Directorate of State Hydraulic Works (DSI) using precipitation data obtained from three ground-based stations that are maintained by Turkish State Meteorological Service (MGM) and remote sensing based precipitation data that are maintained by the Tropical Rainfall Measuring Mission (TRMM). MGM and DSI perform the measurement of meteorologic and hydraulic observations in Turkey. In this study, calculations made for monthly average of 42 years of MGM and DSI observations. Data period is chosen as 1970 -2011. TRMM is a collective mission between National Aeronautics and Space Administration (NASA) and the Japan Aerospace Exploration (JAXA), provided 17 years of precipitation data. In this study, calculations made for monthly average of 12 years. The time interval is chosen as 2000 -2011.

Coruh Basin is divided into 6 sub-basins with regarding the catchment areas of the observation stations. TRMM pixel grid and MGM precipitation data coupled with these flow observation stations. Figure 2.18 shows the Sub-Basins and locations of TRMM measurement pixels and ground-based observation stations. Information about the coupling the data sets are provided in the following section.

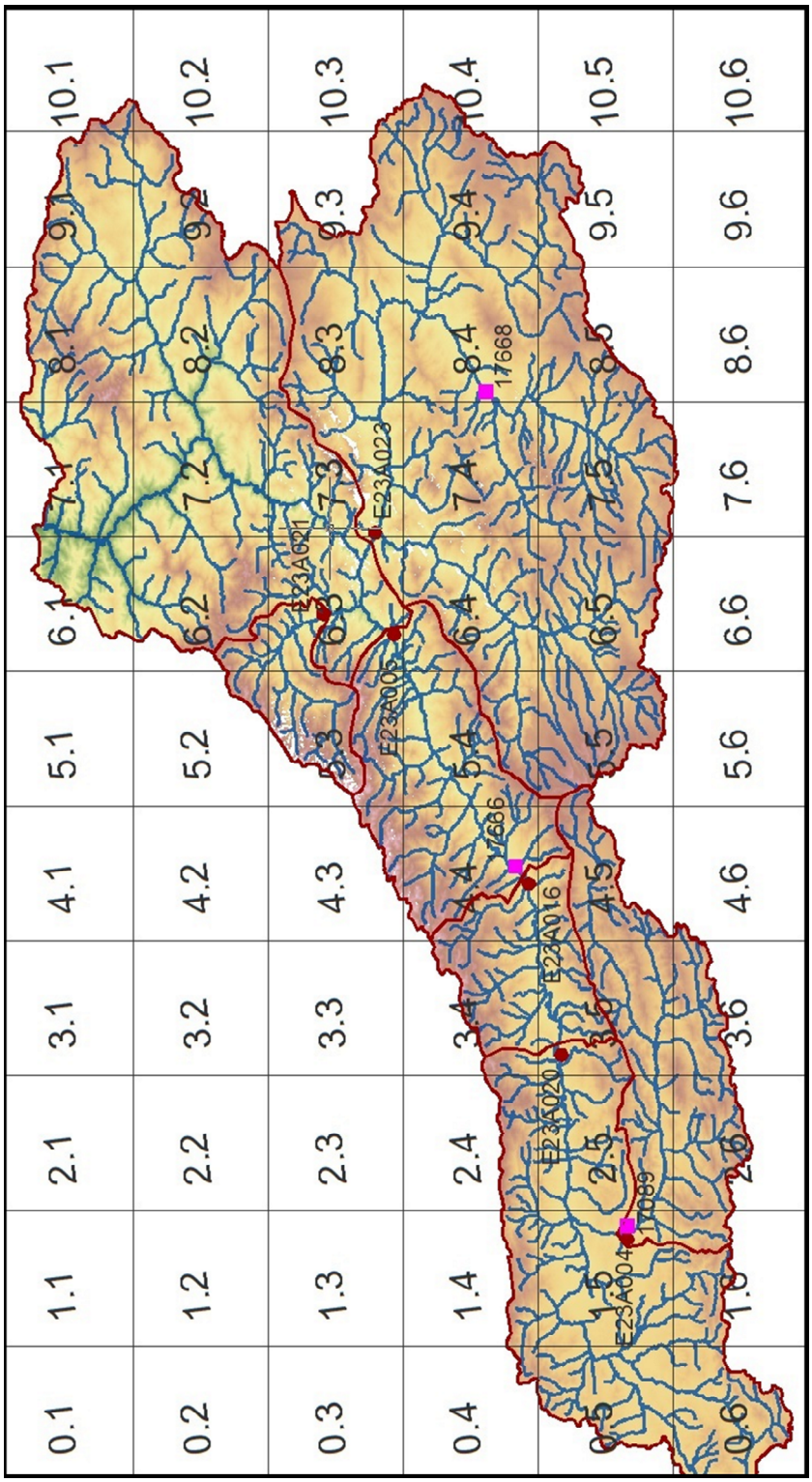


Figure 2.18 Sub-Basins and locations of TRMM measurement pixels and land observation stations

2.5.1. TRMM Precipitation Data

The Tropical Rainfall Measuring Mission (TRMM) is a collective mission between National Aeronautics and Space Administration (NASA) and Japan Aerospace Exploration (JAXA) provided 17 years of precipitation data. TRMM carried 5 instruments: 3-sensor rainfall and 2 related instruments. Names of sensors are as following; TMI – TRMM Microwave Imaging, VIRS – Visible Infrared Scanner, PR – Precipitation Radar, LIS – Lightning Imaging Sensor, CERES – Clouds and Earth's Radiant Energy System. For seventeen years TRMM satellite helped to understand the hydrologic cycle of the climate system of the tropical and sub-tropical regions of the earth. Launched in November 1997 and stopped receiving observations on April 15, 2015. TRMM delivered valuable scientific data about global precipitation and lightning (Web-5).

In this study, TRMM 3B42 V7 daily precipitation product is used by converting into monthly average precipitation value. 25kmx25km Pixels size is used. Figure 2.19 shows the TRMM pixels which cover the Coruh Basin (indices shown in Figure 2.19 is only for demonstration purposes).

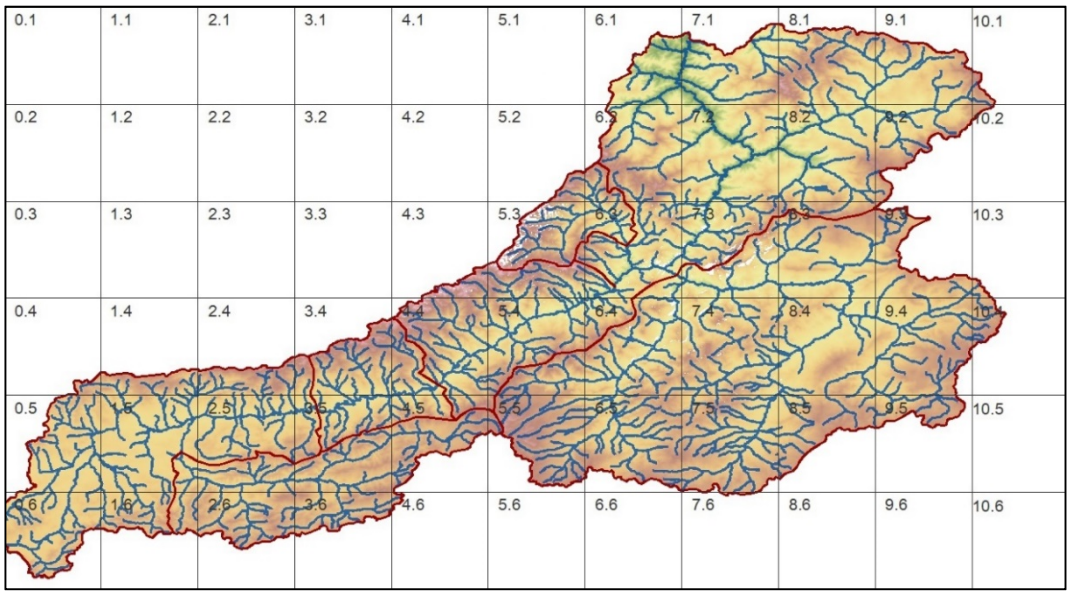
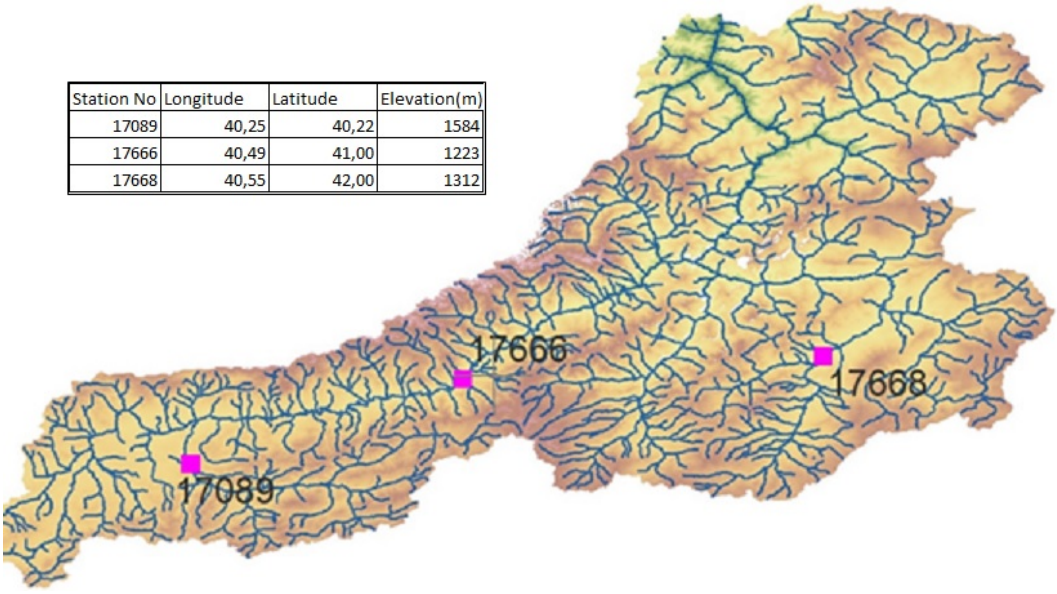


Figure 2.19 TRMM pixels and sub-basins

The TRMM data set became the standard satellite measurement for precipitation and improved our understanding of tropical cyclone structure and evolution, convective system properties, lightning-storm relationships, climate and weather modeling, and human impacts on rainfall. The data also added utility to calculations such as flood and drought risk mapping and weather forecasting.

2.5.2. MGM Gauge-based Precipitation Data

MGM performs the measurement of meteorologic observations in Turkey. MGM was founded in 1937 to make observations, provide forecast, climatological data and archive data. MGM provides meteorological data obtained from manual and automatic meteorologic observation stations all over the Turkey. In this study, daily precipitation data obtained from ground-based manual observation station data for the years 1970 to 2011 is used. Figure 2.20 shows the MGM observation station locations over digital elevation map.



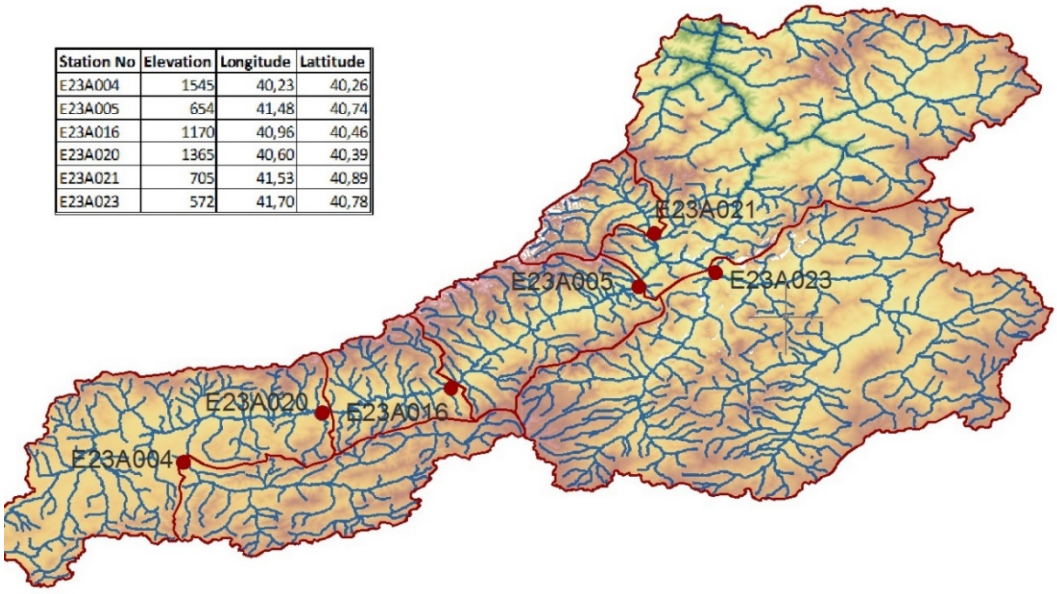
Station No	Longitude	Latitude	Elevation(m)
17089	40,25	40,22	1584
17666	40,49	41,00	1223
17668	40,55	42,00	1312

Figure 2.20 MGM meteorological station locations over DEM map

- 17089 – Bayburt Center Meteorologic Observation Station Data
- 17666 – Erzurum Oltu Meteorologic Observation Station Data
- 17668 – Erzurum Tortum Meteorologic Observation Station Data

2.5.3. DSI Gauge-based Streamflow Data

DSI) performs the measurement of hydrologic observations in Turkey. DSI was founded in 1954 and has a vision about protection, management, and development of water resources and responsible for providing hydrologic data sets to the public (www.DSI.gov.tr). In this study six DSI streamflow observation station data that covers the years 1970 to 2011 is used. Figure 2.21 shows DSI Streamflow Observation Station Locations.



Station No	Elevation	Longitude	Lattitude
E23A004	1545	40,23	40,26
E23A005	654	41,48	40,74
E23A016	1170	40,96	40,46
E23A020	1365	40,60	40,39
E23A021	705	41,53	40,89
E23A023	572	41,70	40,78

Figure 2.21 DSI streamflow observation station locations

- E23A004 – Bayburt Coruh River Flow Observation Station
- E23A005 – Peterek Coruh River Flow Observation Station
- E23A016 – Ispir Bridge Coruh River Flow Observation Station
- E23A020 – Laleli Coruh River Flow Observation Station
- E23A021 – Dutdere Parhal Creek Flow Observation Station
- E23A023 – Ishan Bridge Oltu Creek Flow Observation Station

2.6. Runoff Predictions

Monthly runoff predictions over each of six DSI stations are made using three different data sets. These data sets are respectively; DSI streamflow observation station data coupled with nearest MGM Station rainfall data, DSI streamflow observation station data coupled with TRMM Single Pixel (only pixel which covers the DSI streamflow observation station location) precipitation data and lastly DSI streamflow observation station data coupled with TRMM Catchment Average (all TRMM measurement pixels which cover the sub-basin of DSI streamflow observation station) rainfall.

Figure 18 shows the Sub-Basins and locations of TRMM measurement pixels and land observation stations. Following sections provide detailed information about the data coupling with respect to DSI observation stations.

2.6.1. Station E23A004

Three different rainfall-runoff data sets are coupled for the flow predictions of DSI streamflow observation station E23A004. Figure 2.22 shows the catchment area of the DSI streamflow observation station E23A004 and location, TRMM measurement pixels and MGM rainfall observation station locations. Below three scenarios are performed to investigate the relationship between precipitation and runoff in detail.

- DSI E23A004 Streamflow Data coupled with MGM 17089 station rainfall data
- DSI E23A004 Streamflow Data coupled with TRMM Pixel no 1.5 data (Pixel no 1.5 covers the station E23A004)
- DSI E23A004 Streamflow Data coupled with average of TRMM Pixels no 1.5, 1.6, 2.5, 2.6, 3.5, 3.6, 4.5, 4.6, 5.5 data (Pixels cover the catchment area of E23A004 station)

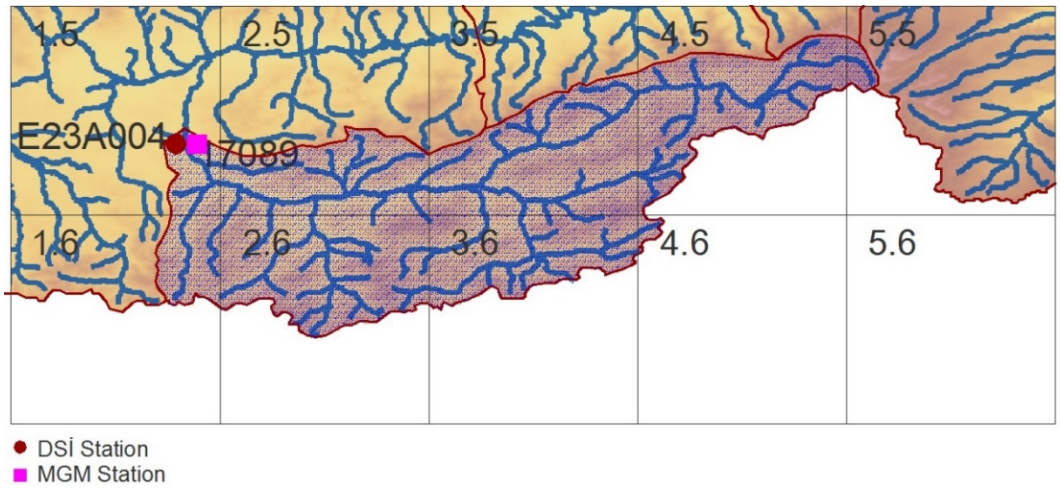


Figure 2.22 Location map of Station E23A004 and related MGM Station and TRMM Pixels

2.6.2. Station E23A005

Three different rainfall-runoff data sets are coupled for the flow predictions of DSI streamflow observation station E23A005. Figure 2.23 shows the catchment area of the DSI streamflow observation station E23A005 and location, TRMM measurement pixels and MGM rainfall observation station locations. Below three scenarios are performed to investigate the relationship between precipitation and runoff in detail.

- DSI E23A005 Streamflow Data coupled with MGM 17666 station rainfall data
- DSI E23A005 Streamflow Data coupled with TRMM Pixel no 6.3 data (Pixel no 6.3 covers the station E23A005)
- DSI E23A005 Streamflow Data coupled with average of TRMM Pixels no 0.4, 0.5, 0.6, 1.4, 1.5, 1.6, 2.4, 2.5, 2.6, 3.4, 3.5, 3.6, 4.3, 4.4, 4.5, 5.3, 5.4, 6.3, 6.4 data (Pixels cover the catchment area of E23A005 station)

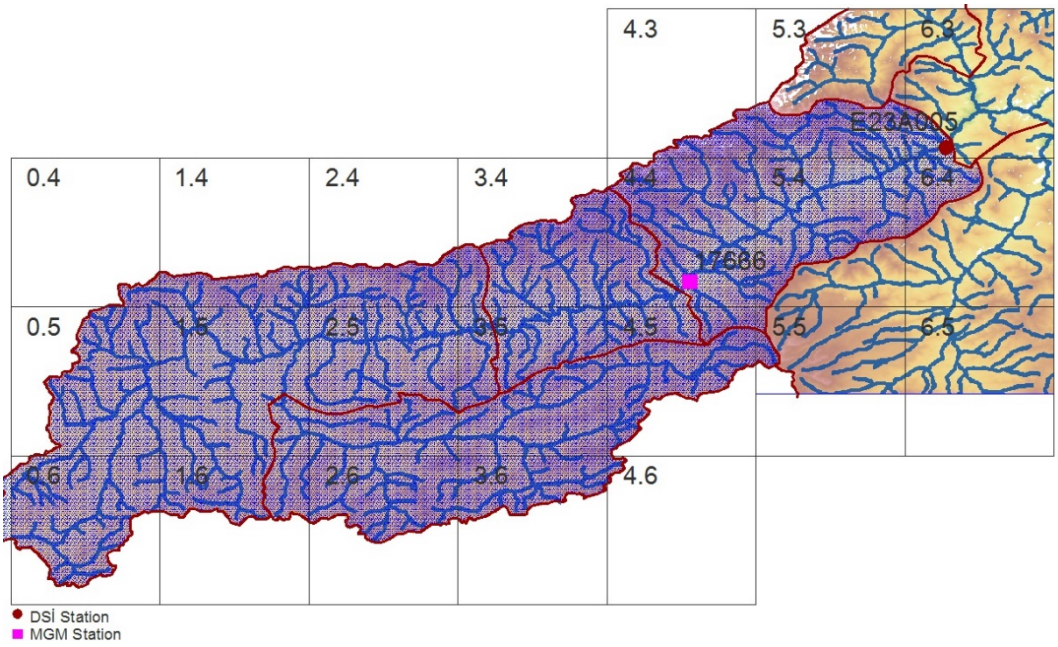


Figure 2.23 Location map of Station E23A005 and related MGM Station and TRMM Pixels

2.6.3. Station E23A016

Three different rainfall-runoff data sets are coupled for the flow predictions of DSI streamflow observation station E23A016. Figure 2.24 shows the catchment area of the DSI streamflow observation station E23A016 and location, TRMM measurement pixels and MGM rainfall observation station locations. Below three scenarios are performed to investigate the relationship between precipitation and runoff in detail.

- DSI E23A016 Streamflow Data coupled with MGM 17666 station rainfall data
- DSI E23A016 Streamflow Data coupled with TRMM Pixel no 4.4 data (Pixel no 4.4 covers the station E23A016)
- DSI E23A016 Streamflow Data coupled with average of TRMM Pixels no 0.4, 0.5, 0.6, 1.4, 1.5, 1.6, 2.4, 2.5, 2.6, 3.4, 3.5, 3.6, 4.3, 4.4, 4.5, data (Pixels cover the catchment area of E23A016 station)

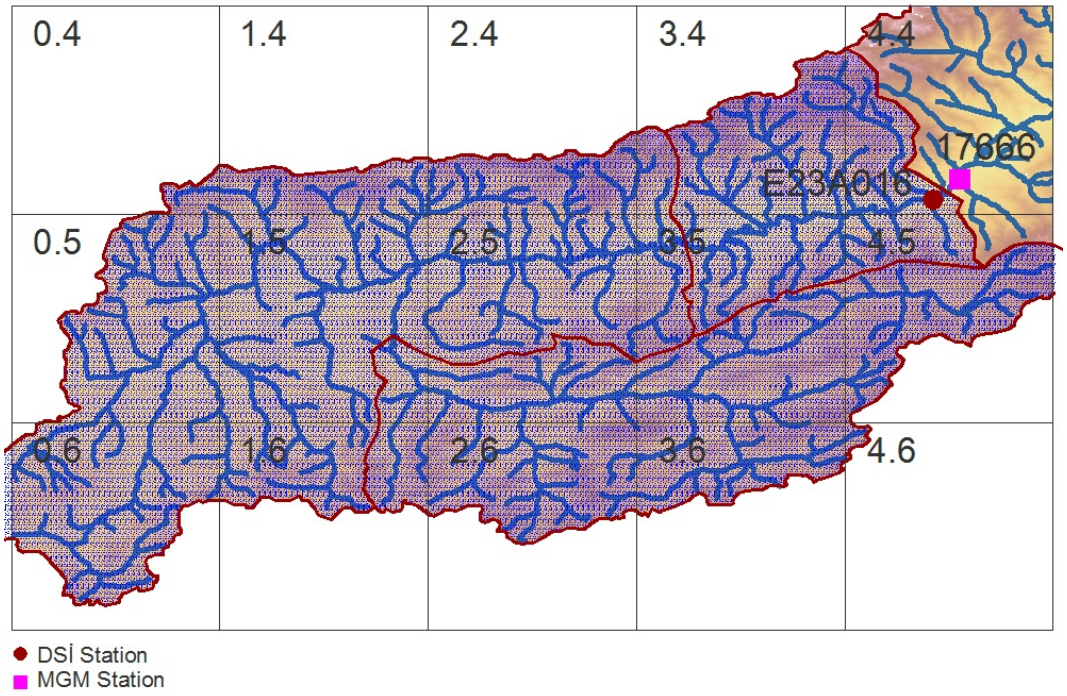


Figure 2.24 Location map of Station E23A016 and related MGM Station and TRMM Pixels

2.6.4. Station E23A020

Three different rainfall-runoff data sets are coupled for the flow predictions of DSI streamflow observation station E23A020. Figure 2.25 shows the catchment area of the DSI streamflow observation station E23A020 and location, TRMM measurement pixels and MGM rainfall observation station locations. Below three scenarios are performed to investigate the relationship between precipitation and runoff in detail.

- DSI E23A020 Streamflow Data coupled with MGM 17089 station rainfall data
- DSI E23A020 Streamflow Data coupled with TRMM Pixel no 3.5 data (Pixel no 3.5 covers the station E23A020)
- DSI E23A020 Streamflow Data coupled with average of TRMM Pixels no 0.4, 0.5, 0.6, 1.4, 1.5, 1.6, 2.4, 2.5, 2.6, 3.4, 3.5, 3.6, 4.5, data (Pixels cover the catchment area of E23A020 station)

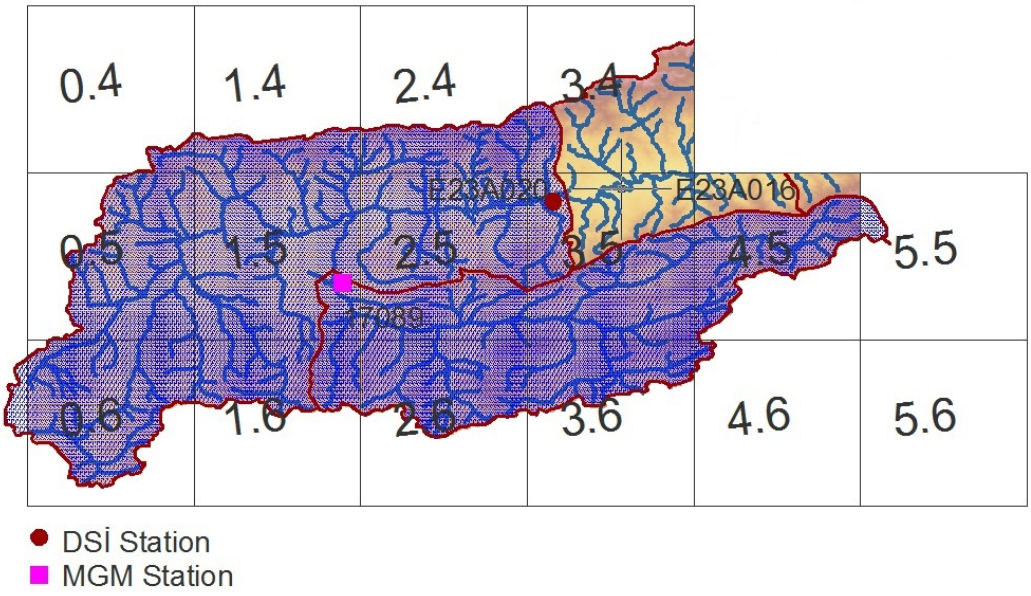


Figure 2.25 Location map of Station E23A020 and related MGM Station and TRMM Pixels

2.6.5. Station E23A021

Three different rainfall-runoff data sets are coupled for the flow predictions of DSI streamflow observation station E23A021. Figure 2.26 shows the catchment area of the DSI streamflow observation station E23A021 and location, TRMM measurement pixels and MGM rainfall observation station locations. Below three scenarios are performed to investigate the relationship between precipitation and runoff in detail.

- DSI E23A021 Streamflow Data coupled with MGM 17668 station rainfall data
- DSI E23A021 Streamflow Data coupled with TRMM Pixel no 6.3 data (Pixel no 6.3 covers the station E23A021)
- DSI E23A021 Streamflow Data coupled with average of TRMM Pixels no5.2, 5.3, 6.2, 6.3 data (Pixels cover the catchment area of E23A021 station)

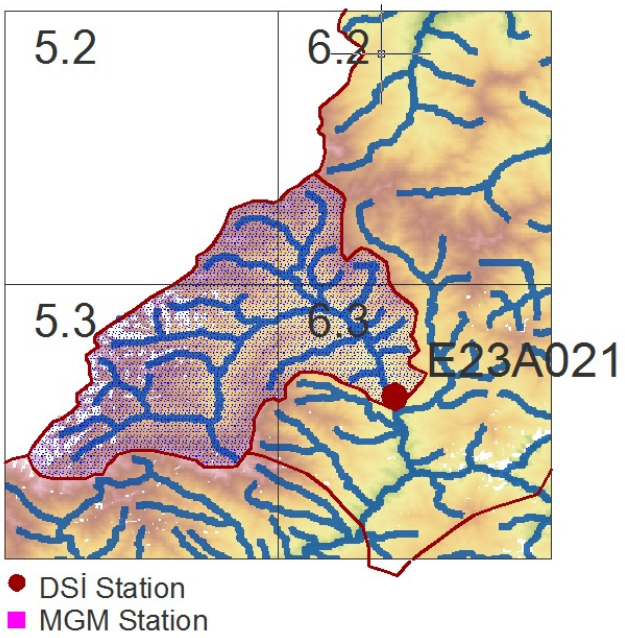


Figure 2.26 Location map of Station E23A021 and related MGM Station and TRMM Pixels

2.6.6. Station E23A023

Three different rainfall-runoff data sets are coupled for the flow predictions of DSI streamflow observation station E23A023. Figure 2.27 shows the catchment area of the DSI streamflow observation station E23A023 and location, TRMM measurement pixels and MGM rainfall observation station locations. Below three scenarios are performed to investigate the relationship between precipitation and runoff in detail.

- DSI E23A023 Streamflow Data coupled with MGM 17668 station rainfall data
- DSI E23A023 Streamflow Data coupled with TRMM Pixel no 7.3 data (Pixel no 7.3 covers the station E23A023)
- DSI E23A023 Streamflow Data coupled with average of TRMM Pixels no 5.4, 5.5, 6.4, 6.5, 7.3, 7.4, 7.5, 8.3, 8.4, 8.5, 9.3, 9.4, 9.5, 10.3, 10.4 data (Pixels cover the catchment area of E23A023 station)

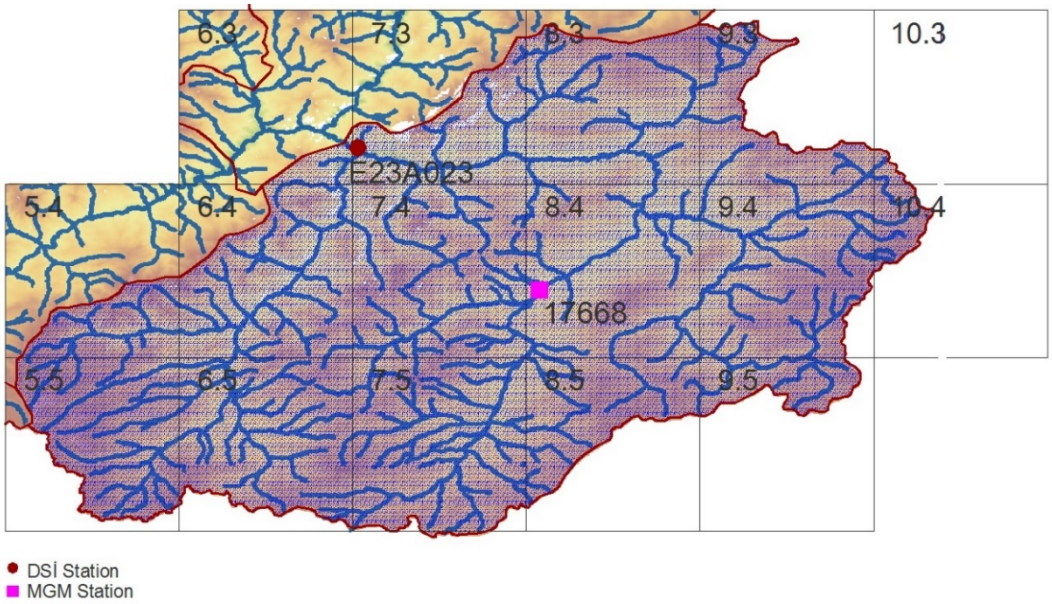


Figure 2.27 Location map of Station E23A023 and related MGM Station and TRMM Pixels

2.7. Performance Statistics

Generally, in statistical analysis, data split into two parts where some are reserved for the training while the remaining part is reserved for the validation. Statistician must be sure about the data set is the best set, which, lead the statistician to reliable results. This is really difficult and critical for forecasting. The point must be taken into consideration is that the training data set must include all kind of observation in it. The following section gives brief information on performance statistics to evaluate different model prediction performances.

Correlation Coefficient

The correlation coefficient is used to describe the relation between two or more variables. It is a degree of consistency between the variables. Correlation value between the observed and predicted streamflow data is calculated as follows;

$$r_{o,p} = \frac{n(\sum op) - (\sum o)(\sum p)}{\sqrt{[n(\sum o^2) - (\sum o)^2][n(\sum p^2) - (\sum p)^2]}} \tag{2.43}$$

Where, $r_{o,p}$ is the correlation coefficient between the observed data and predicted values of streamflow "*o*" is the observation, "*p*" is the predicted value and *n* is number of data.

Mean Error;

Mean Error is used to describe the average of all errors between the observed and predicted streamflow data. Where "*e*" is the error.

$$e_i = o_i - p_i \tag{2.44}$$

$$\mu_e = \frac{\sum_{i=1}^{n} e}{n} \tag{2.45}$$

Standard Deviation of Data and Standard Deviation of Error;

Standard Deviation of Data and Standard Error of Predicted Data is used to describe standardized deviations of data and error terms. Greater standard deviation in error than the standard deviation of data indicates the unuseful predictions.

$$\sigma_x = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)^2} \tag{2.46}$$

Root Mean Squared Error;

Root Mean Squared Error is used to describe the square root of the mean of the square of all errors.

$$RMSE = \sqrt{\frac{1}{n}\sum e^2} \tag{2.47}$$

CHAPTER 3

RESULTS and DISCUSSIONS

3.1. Analysis of Precipitation Data

MGM ground-based precipitation data sets and related TRMM observation data sets compared and correlation coefficients between full-time series, climatology components, and anomaly components are given in Table 3.1.

Table 3.1 Correlations between MGM station data and TRMM product

Stations	Correlations Between Precipitation Datasets								
	Climatology	CompleteTS	Anomaly Comp.	Climatology	CompleteTS	Anomaly Comp.	Climatology	CompleteTS	Anomaly Comp.
	MGM - TRMM Single Pixel (Climatology)	MGM - TRMM Single Pixel (Complete TimeSerie)	MGM - TRMM Single Pixel (Anomaly Component)	MGM - TRMM Catchment Average (Climatology)	MGM - TRMM Catchment Average (Complete TimeSerie)	MGM - TRMM Catchment Average (Anomaly)	TRMM Single Pixel - TRMM Catchment Average (Climatology)	TRMM Single Pixel - TRMM Catchment Average (Complete TimeSerie)	TRMM Single Pixel - TRMM Catchment Average (Anomaly Component)
E23A004	0,666	0,649	0,648	0,766	0,703	0,672	0,986	0,972	0,966
E23A005	0,640	0,656	0,668	0,683	0,656	0,659	0,947	0,903	0,891
E23A016	0,649	0,597	0,588	0,686	0,652	0,653	0,918	0,934	0,944
E23A020	0,712	0,664	0,654	0,700	0,654	0,632	0,991	0,978	0,975
E23A021	0,343	0,500	0,598	0,745	0,569	0,499	0,851	0,910	0,936
E23A023	0,350	0,545	0,661	0,802	0,756	0,735	0,790	0,891	0,945

In the calculations nearest MGM observation station data and TRMM pixel data which covers the related DSI streamflow stations are coupled. As the distance between the MGM station and TRMM Pixel (which covers the DSI Streamflow Station) increases, poor correlations are obtained. In that cases, TRMM Catchment Average data sets are provided relatively better correlations with the MGM ground-based precipitation observations.

On the other hand, correlation coefficient values between climatology components of data sets are expected to show higher values than the full-time series and anomaly component correlations. Anomaly component correlation coefficient values are expected to show lower values. In this aspect, MGM ground-based observations and TRMM satellite measurements show some inconsistency.

3.2. Analysis of Streamflow Data

Another correlation calculation is made for the DSI station observations between years 1970 and 2011. In the absence of data for a specific station or an ungauged region of the basin other station observations can provide insights about the region or can be used to complete missing data in the time series. Table 3.2 shows the correlations between the DSI Streamflow Observation stations.

Table 3.2 Correlations between the DSI streamflow observation station data

Correlations	E23A004	E23A005	E23A016	E23A020	E23A021	E23A023
E23A004	1,000	0,968	0,967	0,973	0,828	0,941
E23A005	0,968	1,000	0,979	0,970	0,883	0,935
E23A016	0,967	0,979	1,000	0,985	0,810	0,909
E23A020	0,973	0,970	0,985	1,000	0,788	0,909
E23A021	0,828	0,883	0,810	0,788	1,000	0,805
E23A023	0,941	0,935	0,909	0,909	0,805	1,000

It is expected that Persistence-Based predictions would provide better correlation results for a streamflow time series with a lag-1 autocorrelation coefficient greater than 0.50 than the Climatology-Based Predictions. Table 3.3 shows the lag-1 autocorrelations, mean values, standard deviations and % of the variability of the DSI Streamflow Observation stations data.

Table 3.3 DSI streamflow observation stations data statistics and autocorrelation values

Stations	Lag-1 Auto Correlation	Mean			Standard Deviation			% of Variability of Data		
		Full TimeSerie	Climatology	Anomaly	Full TimeSerie	Climatology	Anomaly	Full TimeSerie	Climatology	Anomaly
E23A004	0,60	15,512	15,959	-0,447	16,531	15,011	6,924	100	82,46	17,54
E23A005	0,67	69,194	69,194	0,000	73,146	66,465	30,541	100	82,57	17,43
E23A016	0,63	38,983	38,983	0,000	44,316	40,110	18,843	100	81,92	18,08
E23A020	0,62	29,387	29,387	0,000	33,671	30,374	14,530	100	81,38	18,62
E23A021	0,70	14,058	14,058	0,000	13,252	12,453	4,533	100	88,30	11,70
E23A023	0,55	33,506	32,708	0,798	33,847	28,301	18,566	100	69,91	30,09

Deviations of streamflow data show that climatology component of streamflow data contains more than 80% of variability while anomaly component contains approximately 20% of total variability.

The hypsometric curve of the basin (Figure 3.1) shows that the elevation of the basin changes between the 0m and 4000m and approximately 40% of the basin surface area is above the 2000m elevation. This condition causes great temperature changes between higher and lower places of the basin. Precipitation that falls to the basin is preserved in form of snow due to this temperature changes on the higher grounds and, with the increasing temperatures in April and May, melting snow participates to the streamflow forming peak values of streamflow.

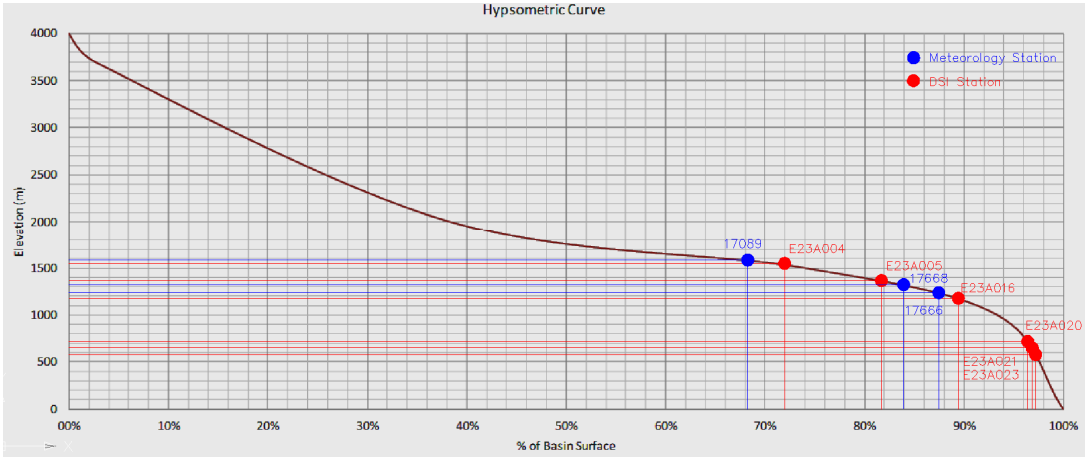


Figure 3.1 Hypsometric curve of the Coruh Basin

3.3. Streamflow Predictions

Streamflow predictions are made for each of 6 available streamflow observations. These predictions are made using 3 different precipitation data sets (MGM ground-based observations, TRMM Single Pixel and TRMM Catchment Average measurements). 6 different methods (SLR, MLR, ANN, Copula, Climatology, and Persistence) used with 2 different data set components (complete data sets and anomaly components only). Climatology and Persistence predictions form the benchmarks that the predictions are compared against.

3.3.1. Station E23A004

3.3.1.1. Predictions Using Meteorology Data

Figure 3.2 shows the validation time series of the predictions and performance statistics of validation time series are provided in Table 3.4. Overall correlations and RMSE values show a consistent pattern; the methods with higher correlation values have lower RMSE value. Among all the complete data set predictions, Persistence-Based predictions provided the best result for the streamflow data while Climatology-Based predictions of the streamflow have approximately 0.90 correlations with the validation data.

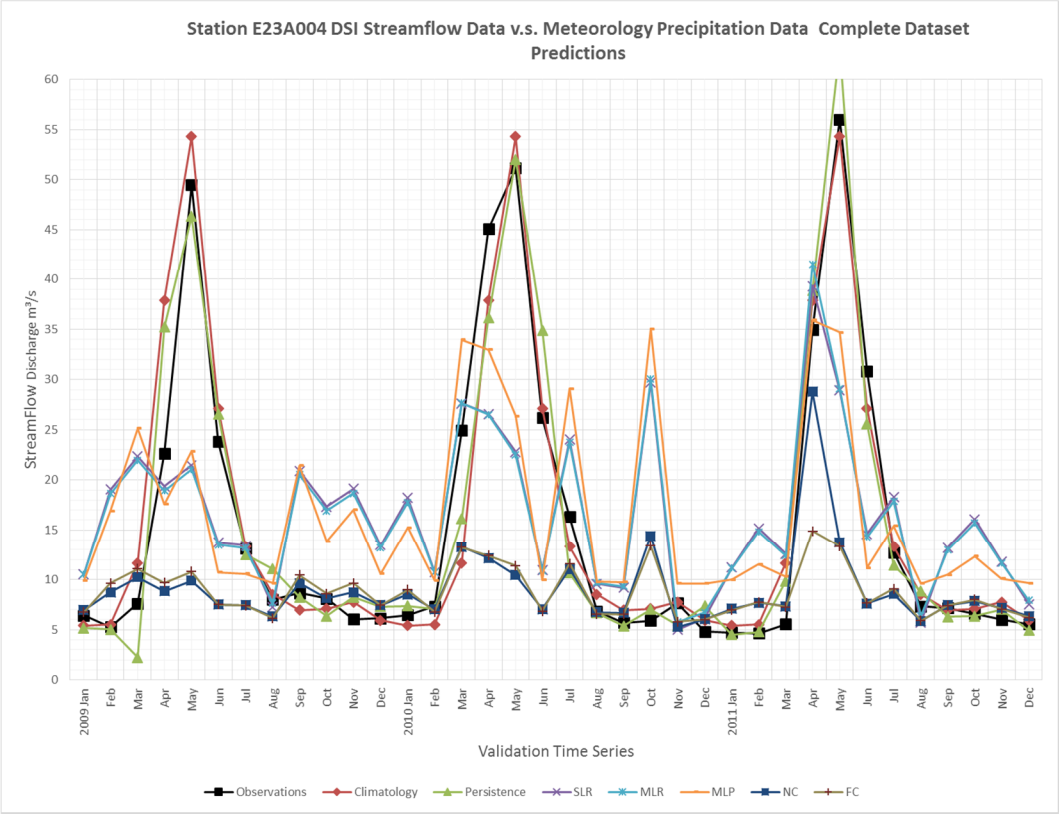


Figure 3.2 Station E23A004 DSI streamflow data v.s. meteorology precipitation data complete data set predictions

Table 3.4 Performance statistics of station E23A004 with meteorology data - complete data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,919	0,919	0,531	0,531	0,545	0,453	0,551
Standard Deviation of Observation	19,102	19,102	19,102	19,102	19,102	19,102	19,102
Standard Deviation of Errors	7,939	7,688	16,295	16,275	16,063	17,498	17,791
Mean Error	2,110	-0,094	1,436	1,416	1,244	8,790	9,227
RMSE	8,174	7,649	16,273	16,252	16,027	19,500	19,959
Mean of Observations	15,512	15,512	15,512	15,512	15,512	15,512	15,512

Similar results are obtained from the SLR, MLR and MLP models but showing under 0.60 correlation. Copula functions couldn't identify a skillful relationship for complete data sets as the other models. To further investigate the source of the predictive skills of these methods, additional predictions are made using the standardized anomaly components of data sets. Figure 3.3 shows the validation time series of the anomaly component predictions.

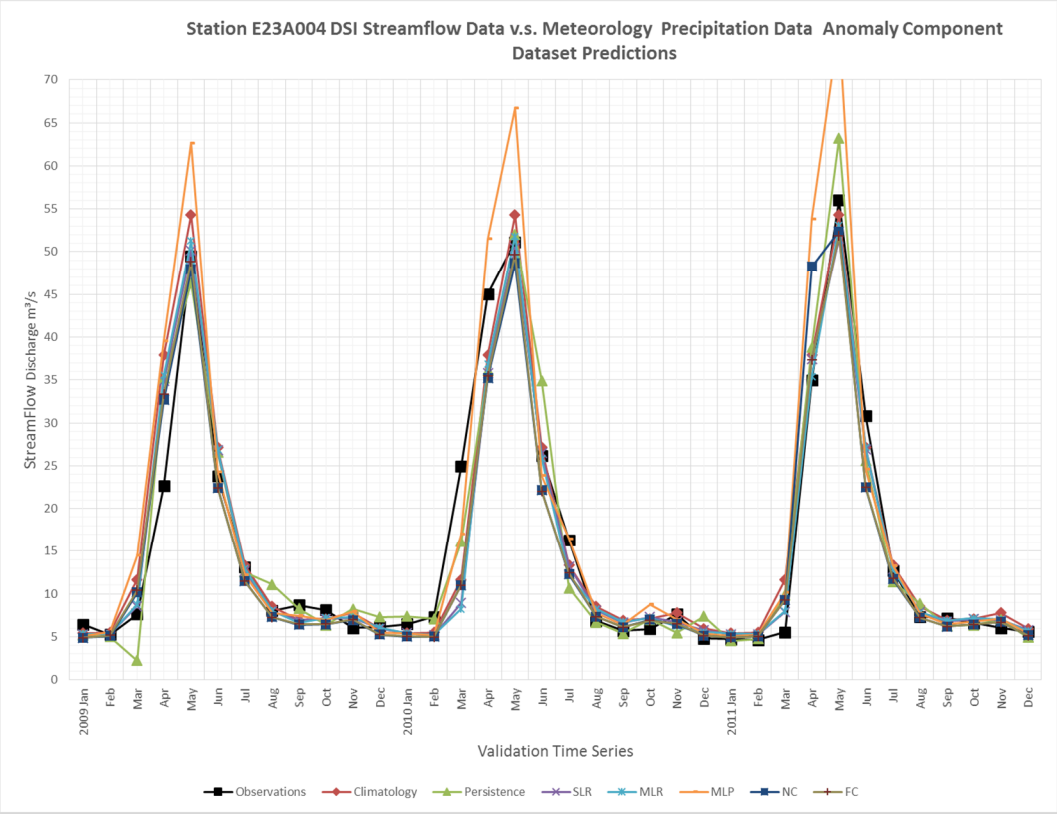


Figure 3.3 Station E23A004 DSI streamflow data v.s. meteorology precipitation data anomaly component data set predictions

Addition of climatology component to the standardized anomaly component predictions showed above 0.90 correlations, these improved predictions are shoving heavy majority of the predictive skill and the relation between the precipitation and the streamflow data sets are due to the strong seasonality.

Correlations of predictions which are obtained from the SLR, MLR, MLP and Copula models increased from 0.50 to above 0.90 with decreasing RMSE values and mean error. Performance statistics of validation time series are provided in Table 3.5.

Table 3.5 Performance statistics of station E23A004 with meteorology data – anomaly component data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,919	0,919	0,918	0,916	0,913	0,917	0,920
Standard Deviation of Observation	19,102	19,102	19,102	19,102	19,102	19,102	19,102
Standard Deviation of Errors	7,939	7,688	8,203	8,177	8,202	8,110	8,258
Mean Error	2,110	-0,094	3,023	2,916	0,009	3,356	3,658
RMSE	8,174	7,649	8,702	8,641	8,159	8,738	8,992
Mean of Observations	15,512	15,512	15,512	15,512	15,512	15,512	15,512

3.3.1.2. Predictions Using TRMM Single Pixel Data

Figure 3.4 shows the validation time series of the predictions and performance statistics of validation time series are provided in Table 3.6. Overall correlations and RMSE values show a consistent pattern; the methods with higher correlation values have lower RMSE value. Among all the complete data set predictions, Persistence-Based predictions provided the best result for the streamflow data while Climatology-Based predictions of the streamflow have over 0.90 correlations with the validation data.

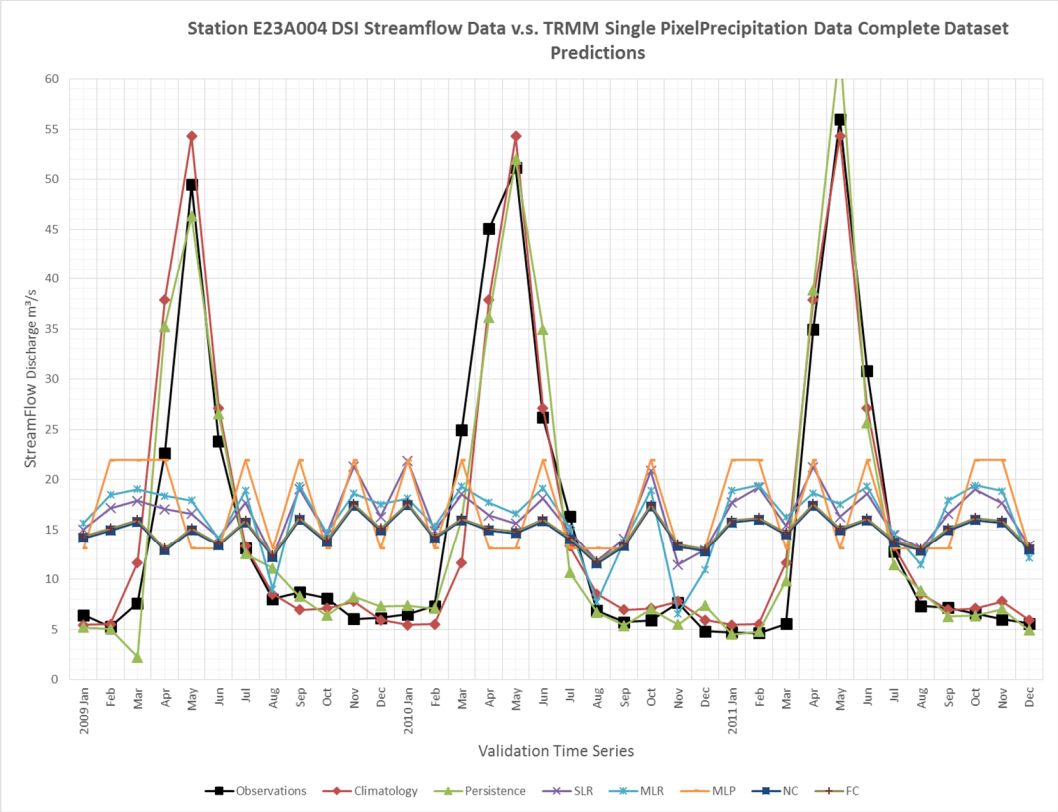


Figure 3.4 Station E23A004 DSI streamflow data v.s. TRMM single pixel precipitation data complete data set predictions

Table 3.6 Performance statistics of station E23A004 with TRMM single pixel data - complete data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,962	0,963	0,103	0,233	-0,123	0,111	0,111
Standard Deviation of Observation	14,926	14,926	14,926	14,926	14,926	14,926	14,926
Standard Deviation of Errors	4,171	4,215	14,894	14,517	16,101	14,834	14,834
Mean Error	-0,518	-0,178	-1,035	-0,822	-1,591	0,674	0,474
RMSE	4,145	4,160	14,722	14,337	15,955	14,642	14,634
Mean of Observations	15,512	15,512	15,512	15,512	15,512	15,512	15,512

Similar results are obtained from the SLR, MLR, MLP and Copula models but showing under 0.30 correlation and some artificial skills. These models could not identify a skillful relationship between the Rainfall and Runoff mainly due to lack of data. 9 years of monthly streamflow data and precipitation data could not train the models.

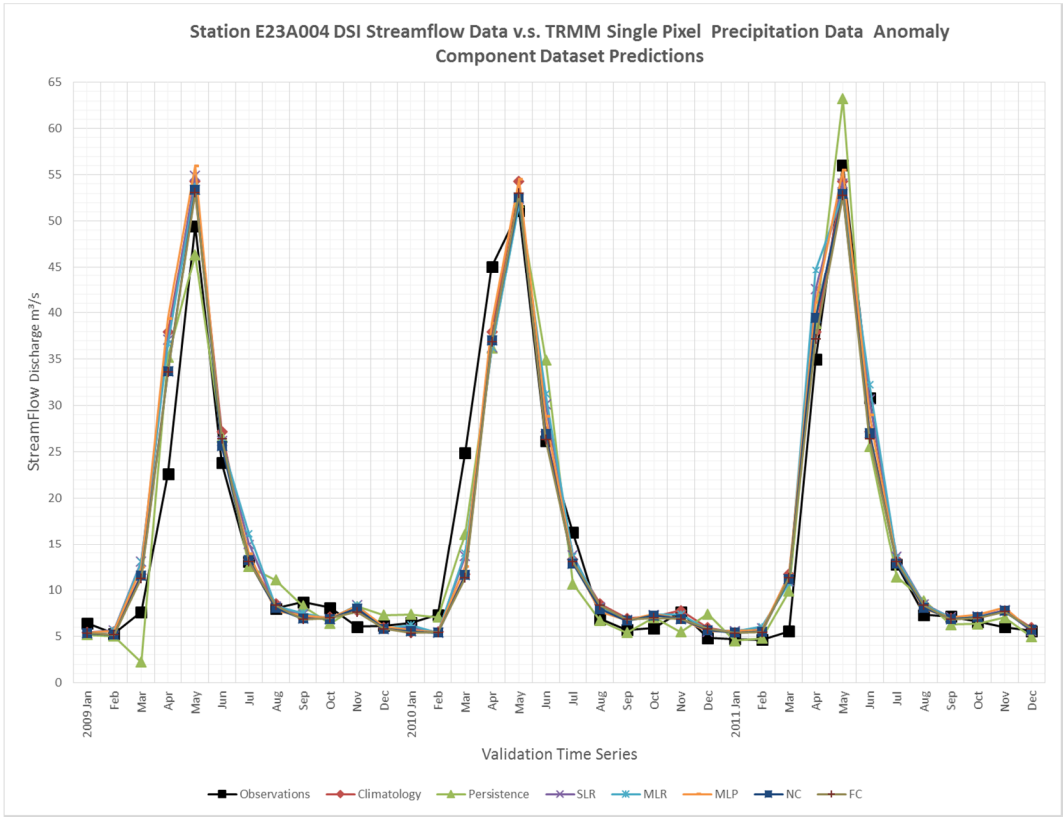


Figure 3.5 Station E23A004 DSI streamflow data v.s. TRMM single pixel precipitation data anomaly component data set predictions

Additional predictions are made using the standardized anomaly components of data sets. Figure 3.5 shows the validation time series of the anomaly component predictions.

Addition of Climatology component to the standardized anomaly component predictions showed above 0.90 correlations, these improved predictions are shoving heavy majority of the predictive skill and the relation between the precipitation and the streamflow data sets are due to the strong seasonality.

Correlations of predictions which are obtained from the SLR, MLR, MLP and Copula models increased from 0.20 to above 0.90 with decreasing RMSE values and mean error. Performance statistics of validation time series are provided in Table 3.7.

Table 3.7 Performance statistics of station E23A004 with TRMM single pixel data – anomaly component data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,962	0,963	0,963	0,962	0,962	0,967	0,967
Standard Deviation of Observation	14,926	14,926	14,926	14,926	14,926	14,926	14,926
Standard Deviation of Errors	4,171	4,215	4,137	4,200	4,310	3,823	3,834
Mean Error	-0,518	-0,178	-0,829	-0,923	-0,828	-0,110	-0,030
RMSE	4,145	4,160	4,163	4,243	4,330	3,771	3,781
Mean of Observations	15,512	15,512	15,512	15,512	15,512	15,512	15,512

3.3.1.3. Predictions Using TRMM Catchment Average Data

Figure 3.6 shows the validation time series of the predictions and performance statistics of validation time series are provided in Table 3.8. Overall correlations and RMSE values show a consistent pattern; the methods with higher correlation values have lower RMSE value. Among all the complete data set predictions, Persistence-Based predictions provided the best result for the streamflow data while Climatology-Based predictions of the streamflow have over 0.90 correlations with the validation data.

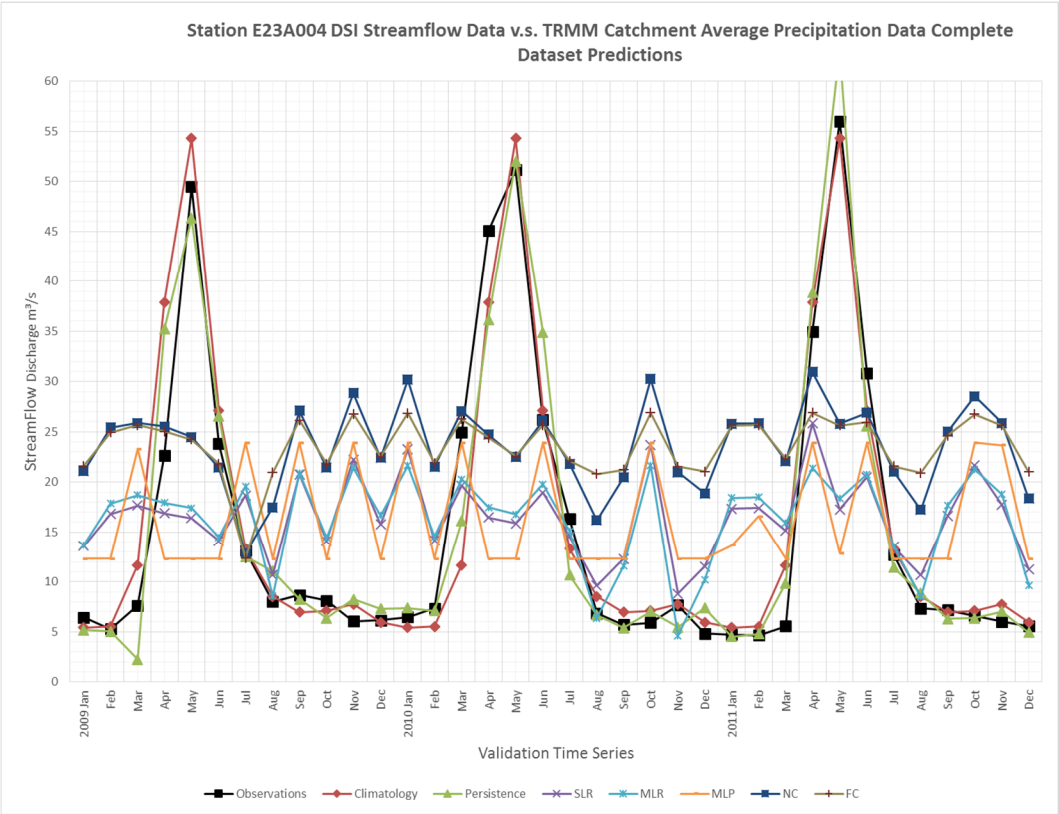


Figure 3.6 Station E23A004 DSI streamflow data v.s. TRMM catchment average precipitation data complete data set predictions

Table 3.8 Performance statistics of station E23A004 with TRMM catchment average data - complete data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,962	0,963	0,198	0,250	-0,043	0,196	0,166
Standard Deviation of Observation	14,926	14,926	14,926	14,926	14,926	14,926	14,926
Standard Deviation of Errors	4,171	4,215	14,674	14,475	16,107	14,693	14,724
Mean Error	-0,518	-0,178	-0,994	-0,793	-0,889	-8,119	-8,109
RMSE	4,145	4,160	14,503	14,294	15,907	16,608	16,629
Mean of Observations	15,512	15,512	15,512	15,512	15,512	15,512	15,512

Similar results are obtained from the SLR, MLR, MLP and Copula models but showing under 0.30 correlation. These models could not identify a skillful relationship between the Rainfall and Runoff mainly due to lack of data. 9 years of monthly streamflow data and precipitation data could not train the models.

Additional predictions are made using the standardized anomaly components of data sets. Figure 3.7 shows the validation time series of the anomaly component predictions.

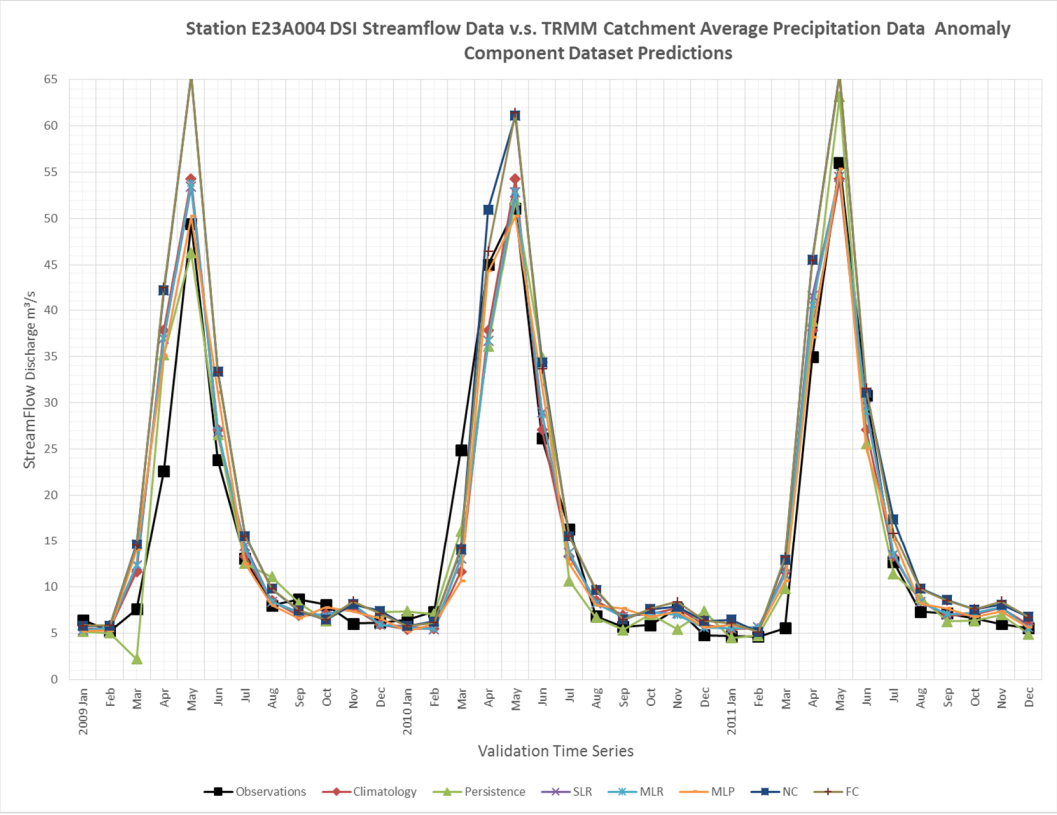


Figure 3.7 Station E23A004 DSI streamflow data v.s. TRMM catchment average precipitation data anomaly component data set predictions

Addition of Climatology component to the standardized anomaly component predictions showed above 0.90 correlations, these improved predictions are shoving heavy majority of the predictive skill and the relation between the precipitation and the streamflow data sets are due to the strong seasonality.

Correlations of predictions which are obtained from the SLR, MLR, MLP and Copula models increased from 0.20 to above 0.90 with decreasing RMSE values and mean error term. Performance statistics of validation time series are provided in Table 3.9.

Table 3.9 Performance statistics of station E23A004 with catchment average data – anomaly component data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,962	0,963	0,965	0,965	0,963	0,969	0,966
Standard Deviation of Observation	14,926	14,926	14,926	14,926	14,926	14,926	14,926
Standard Deviation of Errors	4,171	4,215	4,031	4,032	4,088	5,478	5,451
Mean Error	-0,518	-0,178	-0,705	-0,676	-0,575	-3,274	-3,127
RMSE	4,145	4,160	4,037	4,032	4,072	6,316	6,218
Mean of Observations	15,512	15,512	15,512	15,512	15,512	15,512	15,512

3.3.2. Station E23A005

3.3.2.1. Predictions Using Meteorology Data

Figure 3.8 shows the validation time series of the predictions and performance statistics of validation time series are provided in Table 3.10. Overall correlations and RMSE values show a consistent pattern; the methods with higher correlation values have lower RMSE value. Among all the complete data set predictions, Climatology-Based predictions provided the best result for the streamflow data while Persistence-Based predictions of the streamflow have approximately 0.90 correlations with the validation data.

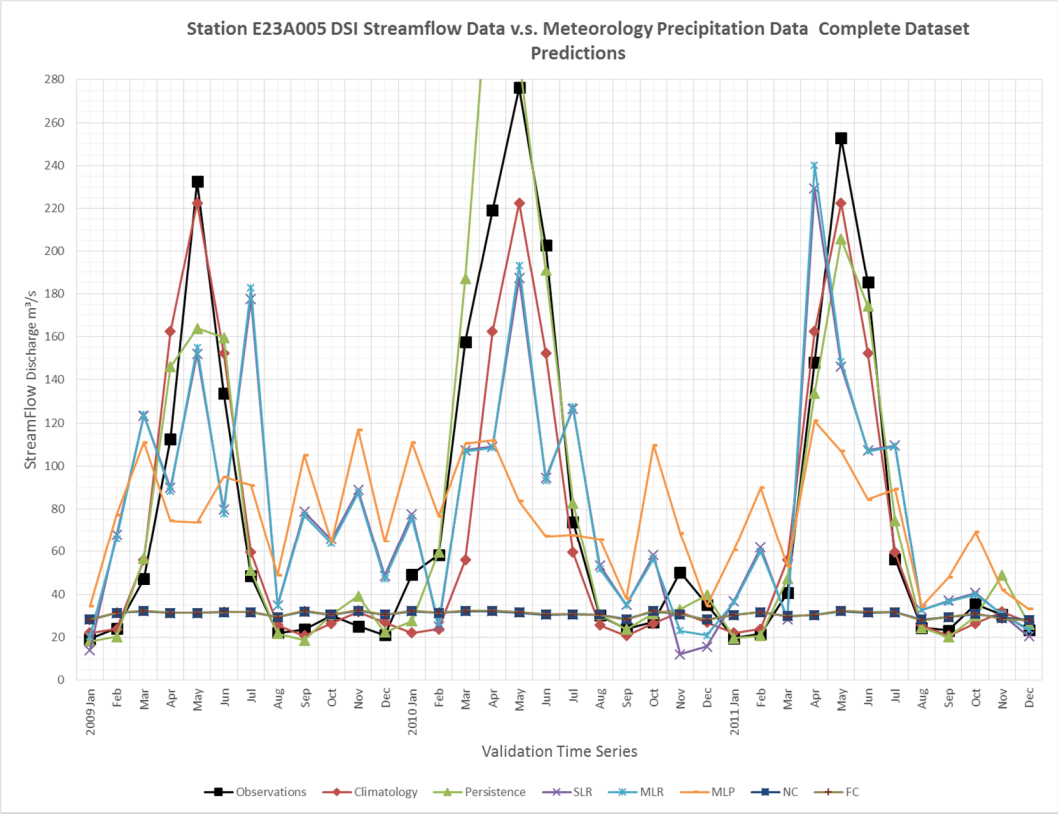


Figure 3.8 Station E23A005 DSI streamflow data v.s. meteorology precipitation data complete data set predictions

Table 3.10 Performance statistics of station E23A005 with meteorology data - complete data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,931	0,884	0,604	0,605	0,437	0,432	0,427
Standard Deviation of Observation	84,275	84,275	84,275	84,275	84,275	84,275	84,275
Standard Deviation of Errors	32,939	42,381	67,181	67,111	76,470	83,714	83,661
Mean Error	13,073	-2,448	5,049	4,519	9,650	51,692	51,675
RMSE	35,279	42,231	67,020	66,914	76,681	98,016	97,962
Mean of Observations	69,194	69,194	69,194	69,194	69,194	69,194	69,194

Similar results are obtained from the SLR and MLR models but showing 0.60 correlation. MLP model seems to identify the relationship but showing 0.40 correlation. Copula functions couldn't identify a skillful relationship for complete data sets.

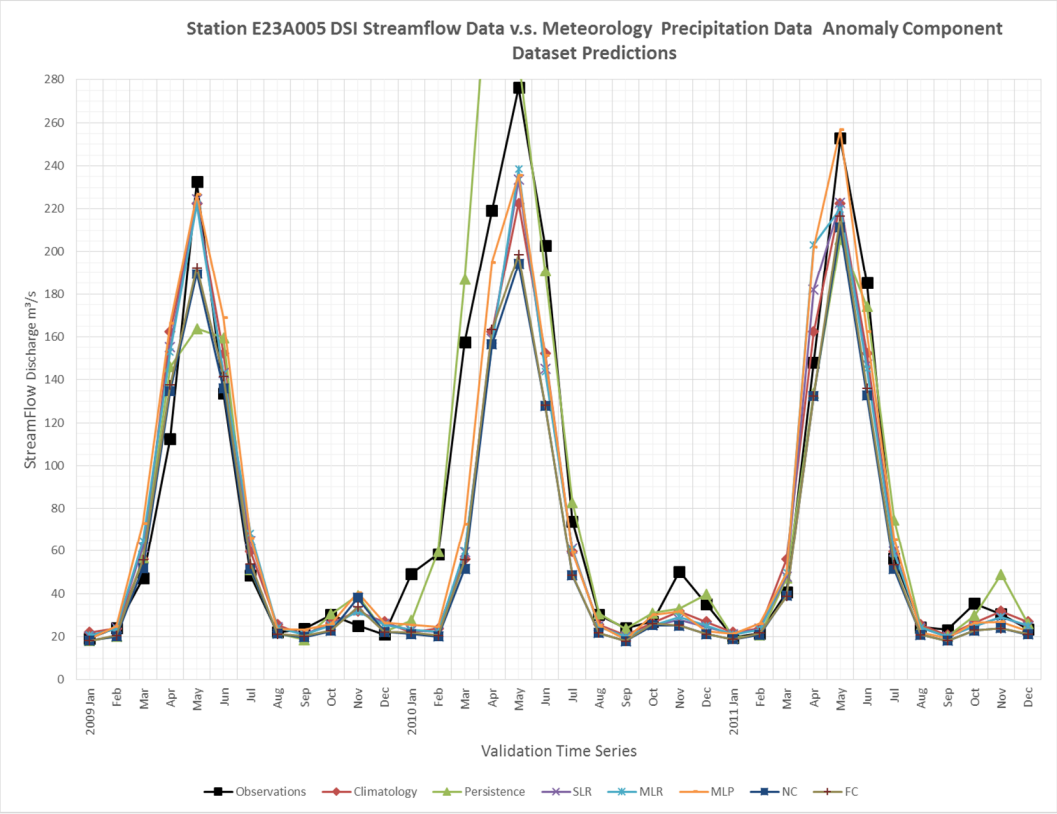


Figure 3.9 Station E23A005 DSI streamflow data v.s. meteorology precipitation data anomaly component data set predictions

To further investigate the source of the predictive skills of these methods, additional predictions are made using the standardized anomaly components of data sets. Figure 3.9 shows the validation time series of the anomaly component predictions.

Addition of Climatology component to the standardized anomaly component predictions showed above 0.90 correlations, these improved predictions are shoving heavy majority of the predictive skill and the relation between the precipitation and the streamflow data sets are due to the strong seasonality.

Correlations of predictions which are obtained from the SLR, MLR, MLP, and Copula models increased from 0.40 to above 0.90 with decreasing RMSE values and mean error term. Performance statistics of validation time series are provided in Table 3.11.

Table 3.11 Performance statistics of station E23A005 with meteorology data – anomaly component data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,931	0,884	0,928	0,924	0,938	0,939	0,941
Standard Deviation of Observation	84,275	84,275	84,275	84,275	84,275	84,275	84,275
Standard Deviation of Errors	32,939	42,381	33,382	34,062	29,447	33,392	33,040
Mean Error	13,073	-2,448	14,131	14,050	8,733	20,630	20,496
RMSE	35,279	42,231	36,089	36,682	30,567	39,103	38,735
Mean of Observations	69,194	69,194	69,194	69,194	69,194	69,194	69,194

3.3.2.2. Predictions Using TRMM Single Pixel Data

Figure 3.10 shows the validation time series of the predictions and performance statistics of validation time series are provided in Table 3.12. Overall correlations and RMSE values show a consistent pattern; the methods with higher correlation values have lower RMSE value. Among all the complete data set predictions, Climatology-Based predictions provided the best result for the streamflow data while Persistence-Based predictions of the streamflow have approximately 0.90 correlations with the validation data.

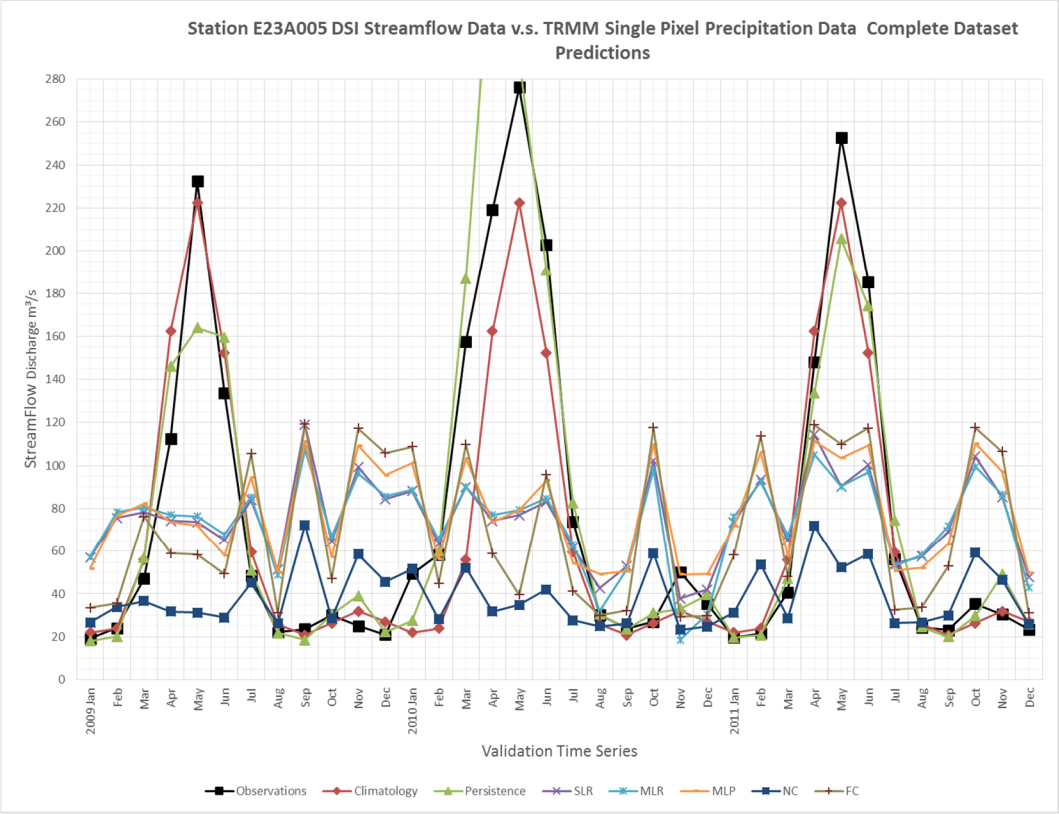


Figure 3.10 Station E23A005 DSI streamflow data v.s. TRMM single pixel precipitation data complete data set predictions

Table 3.12 Performance statistics of station E23A005 with TRMM single pixel data - complete data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,943	0,938	0,228	0,268	0,238	0,139	0,152
Standard Deviation of Observation	77,820	77,820	77,820	77,820	77,820	77,820	77,820
Standard Deviation of Errors	26,538	29,092	75,828	74,980	75,759	77,144	80,474
Mean Error	8,732	-3,210	2,980	4,332	0,546	38,938	8,002
RMSE	27,585	28,864	74,827	74,058	74,701	85,452	79,750
Mean of Observations	69,194	69,194	69,194	69,194	69,194	69,194	69,194

Similar results are obtained from the SLR, MLR, MLP and Copula models but showing under 0.30 correlation. These models could not identify a skillful relationship between the Rainfall and Runoff mainly due to lack of data. 9 years of monthly streamflow data and precipitation data could not be enough to train the models.

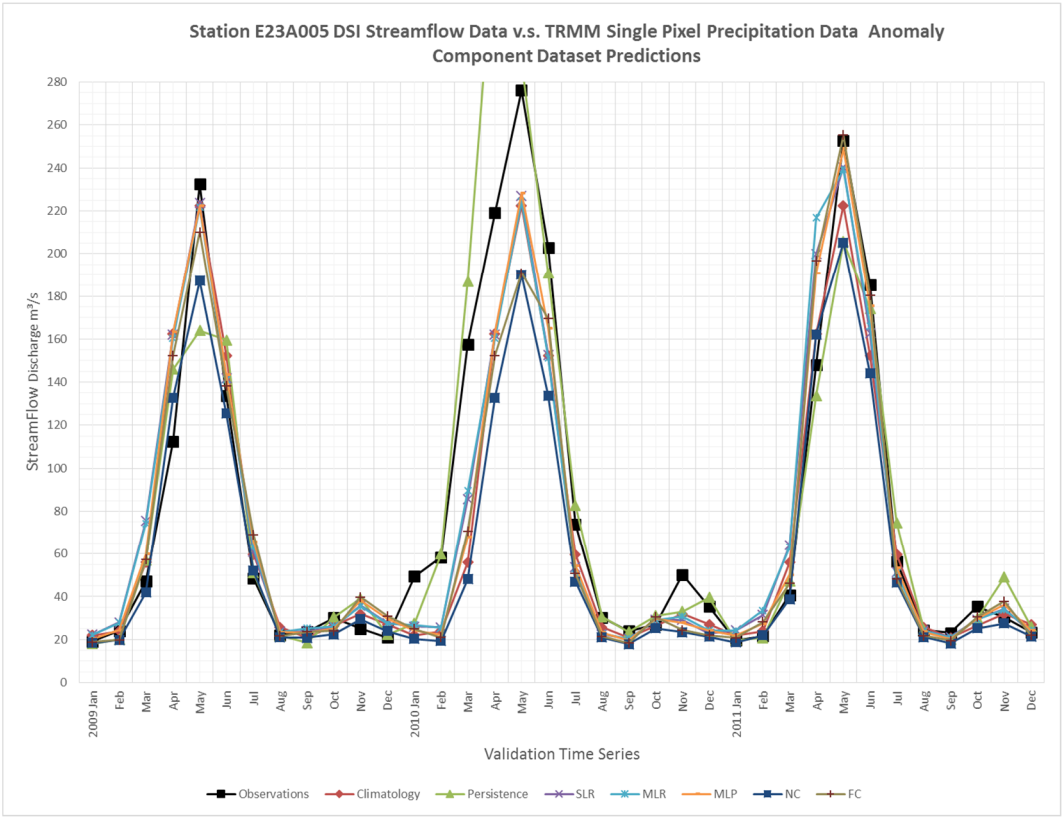


Figure 3.11 Station E23A005 DSI streamflow data v.s. TRMM single pixel precipitation data anomaly component data set predictions

Additional predictions are made using the standardized anomaly components of data sets. Figure 3.11 shows the validation time series of the anomaly component predictions.

Addition of Climatology component to the standardized anomaly component predictions showed above 0.90 correlations, these improved predictions are shoving heavy majority of the predictive skill and the relation between the precipitation and the streamflow data sets are due to the strong seasonality.

Correlations of predictions which are obtained from the SLR, MLR, MLP and Copula models increased from 0.20 to above 0.90 with decreasing RMSE values and mean error term. Performance statistics of validation time series are provided in Table 3.13.

Table 3.13 Performance statistics of station E23A005 with TRMM single pixel data – anomaly component data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,943	0,938	0,945	0,940	0,946	0,941	0,935
Standard Deviation of Observation	77,820	77,820	77,820	77,820	77,820	77,820	77,820
Standard Deviation of Errors	26,538	29,092	25,579	26,651	25,245	29,689	27,830
Mean Error	8,732	-3,210	5,286	4,785	5,942	18,326	8,206
RMSE	27,585	28,864	25,770	26,711	25,591	34,537	28,641
Mean of Observations	69,194	69,194	69,194	69,194	69,194	69,194	69,194

3.3.2.3. Predictions Using TRMM Catchment Average Data

Figure 3.12 shows the validation time series of the predictions and performance statistics of validation time series are provided in Table 3.14. Overall correlations and RMSE values show a consistent pattern; the methods with higher correlation values have lower RMSE value. Among all the complete data set predictions, Climatology-Based predictions provided the best result for the streamflow data while Persistence-Based predictions of the streamflow have approximately 0.90 correlations with the validation data.

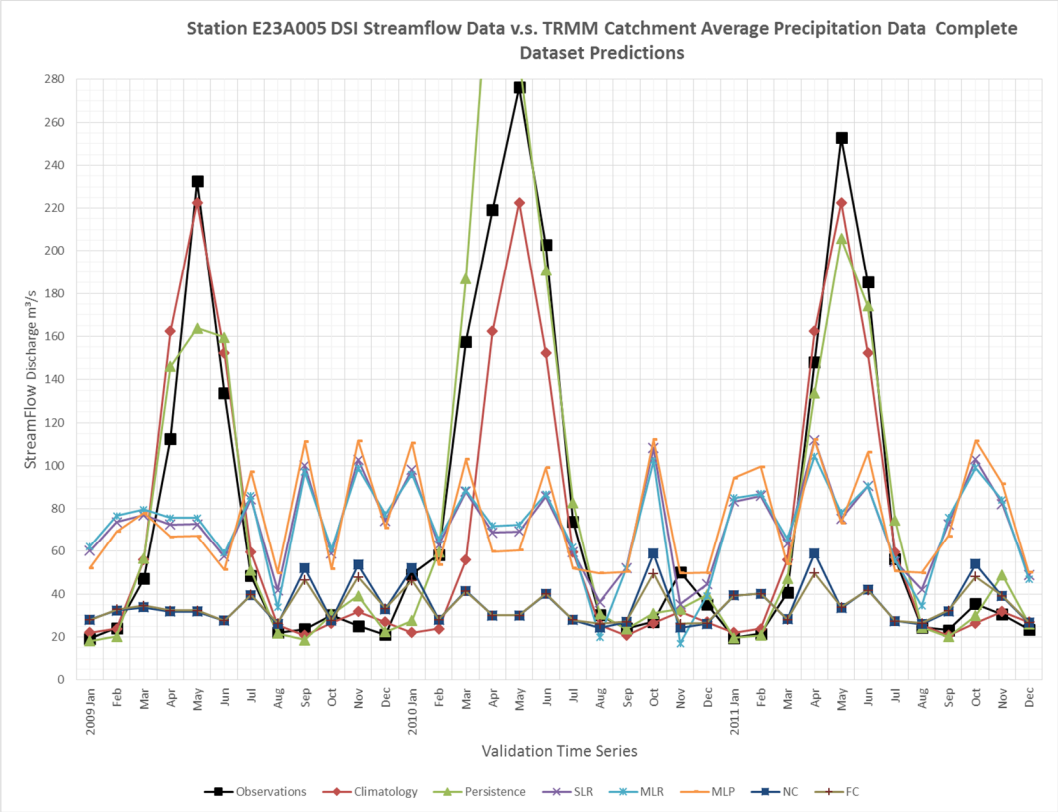


Figure 3.12 Station E23A005 DSI streamflow data v.s. TRMM catchment average precipitation data complete data set predictions

Table 3.14 Performance statistics of station E23A005 with TRMM catchment average data - complete data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,943	0,938	0,154	0,205	0,065	0,007	0,037
Standard Deviation of Observation	77,820	77,820	77,820	77,820	77,820	77,820	77,820
Standard Deviation of Errors	26,538	29,092	77,387	76,449	80,018	78,429	77,925
Mean Error	8,732	-3,210	5,687	6,706	3,233	42,552	43,444
RMSE	27,585	28,864	76,517	75,678	78,965	88,266	88,267
Mean of Observations	69,194	69,194	69,194	69,194	69,194	69,194	69,194

Similar results are obtained from the SLR, MLR, MLP and Copula models but showing under 0.20 correlation. These models could not identify a skillful relationship between the Rainfall and Runoff mainly due to lack of data. 9 years of monthly streamflow data and precipitation data could not train the models.

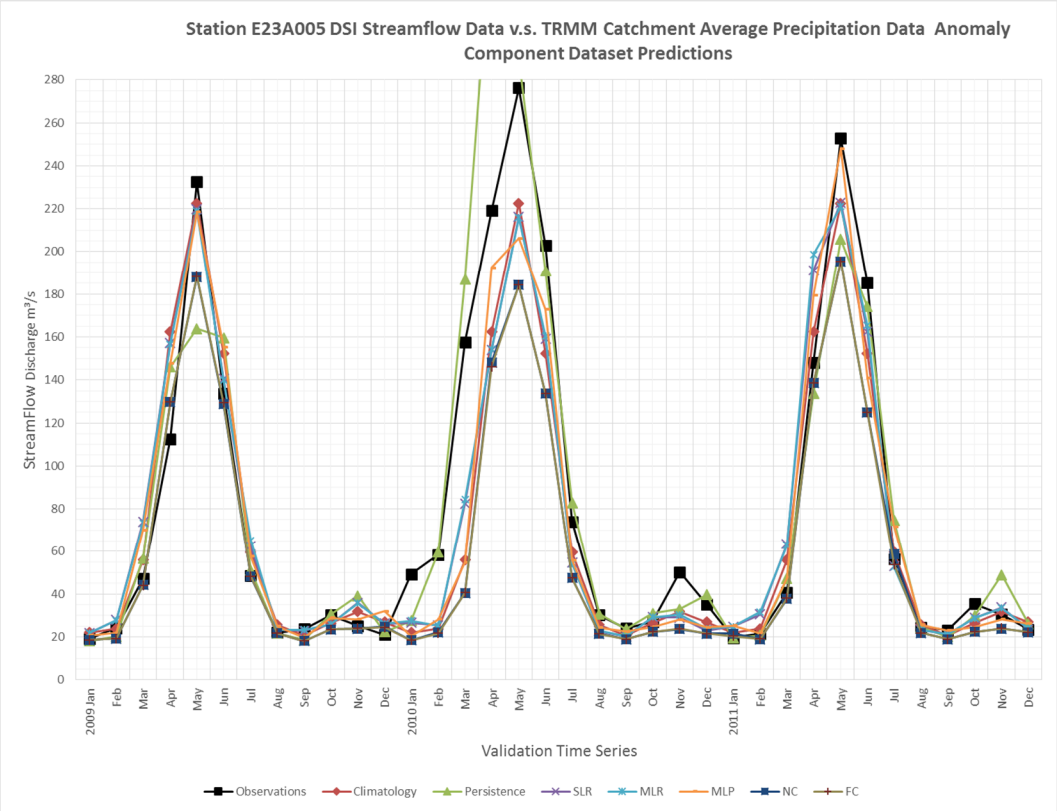


Figure 3.13 Station E23A005 DSI streamflow data v.s. TRMM catchment average precipitation data anomaly component data set predictions

Additional predictions are made using the standardized anomaly components of data sets. Figure 3.13 shows the validation time series of the anomaly component predictions.

Addition of Climatology component to the standardized anomaly component predictions showed above 0.90 correlations, these improved predictions are shoving heavy majority of the predictive skill and the relation between the precipitation and the streamflow data sets are due to the strong seasonality.

Correlations of predictions which are obtained from the SLR, MLR, MLP and Copula models increased from 0.20 to above 0.90 with decreasing RMSE values and mean error term. Performance statistics of validation time series are provided in Table 3.15.

Table 3.15 Performance statistics of station E23A005 with catchment average data – anomaly component data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,943	0,938	0,946	0,944	0,943	0,944	0,945
Standard Deviation of Observation	77,820	77,820	77,820	77,820	77,820	77,820	77,820
Standard Deviation of Errors	26,538	29,092	25,998	26,221	26,157	30,101	30,108
Mean Error	8,732	-3,210	6,814	6,436	7,308	19,708	19,896
RMSE	27,585	28,864	26,524	26,643	26,807	35,627	35,737
Mean of Observations	69,194	69,194	69,194	69,194	69,194	69,194	69,194

3.3.3. Station E23A016

3.3.3.1. Predictions Using Meteorology Data

Figure 3.14 shows the validation time series of the predictions and performance statistics of validation time series are provided in Table 3.16. Overall correlations and RMSE values show a consistent pattern; the methods with higher correlation values have lower RMSE value. Among all the complete data set predictions, Climatology-Based predictions provided the best result for the streamflow data while Persistence-Based predictions of the streamflow have approximately 0.90 correlations with the validation data.

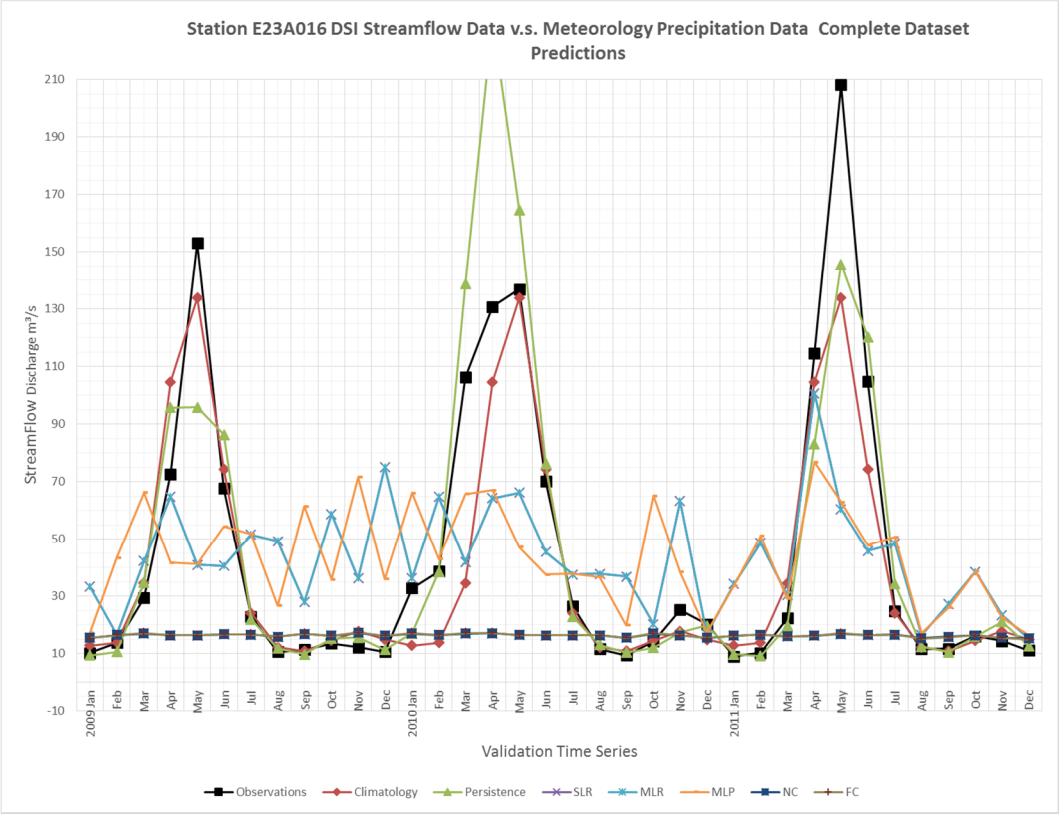


Figure 3.14 Station E23A016 DSI streamflow data v.s. meteorology precipitation data complete data set predictions

Table 3.16 Performance statistics of station E23A016 with meteorology data - complete data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,922	0,859	0,253	0,253	0,458	0,431	0,441
Standard Deviation of Observation	54,960	54,960	54,960	54,960	54,960	54,960	54,960
Standard Deviation of Errors	23,653	30,152	53,485	53,479	49,453	54,731	54,695
Mean Error	7,927	-0,825	4,709	4,712	5,485	30,686	30,675
RMSE	24,829	30,006	53,413	53,408	49,499	62,497	62,460
Mean of Observations	38,983	38,983	38,983	38,983	38,983	38,983	38,983

Similar results are obtained from the SLR and MLR models but showing 0.20 correlation. Copula functions couldn't identify a skillful relationship for complete data sets. MLP model is showed some artificial results. To further investigate the source of the predictive skills of these methods, additional predictions are made using the standardized anomaly components of data sets. Figure 3.15 shows the validation time series of the anomaly component predictions.

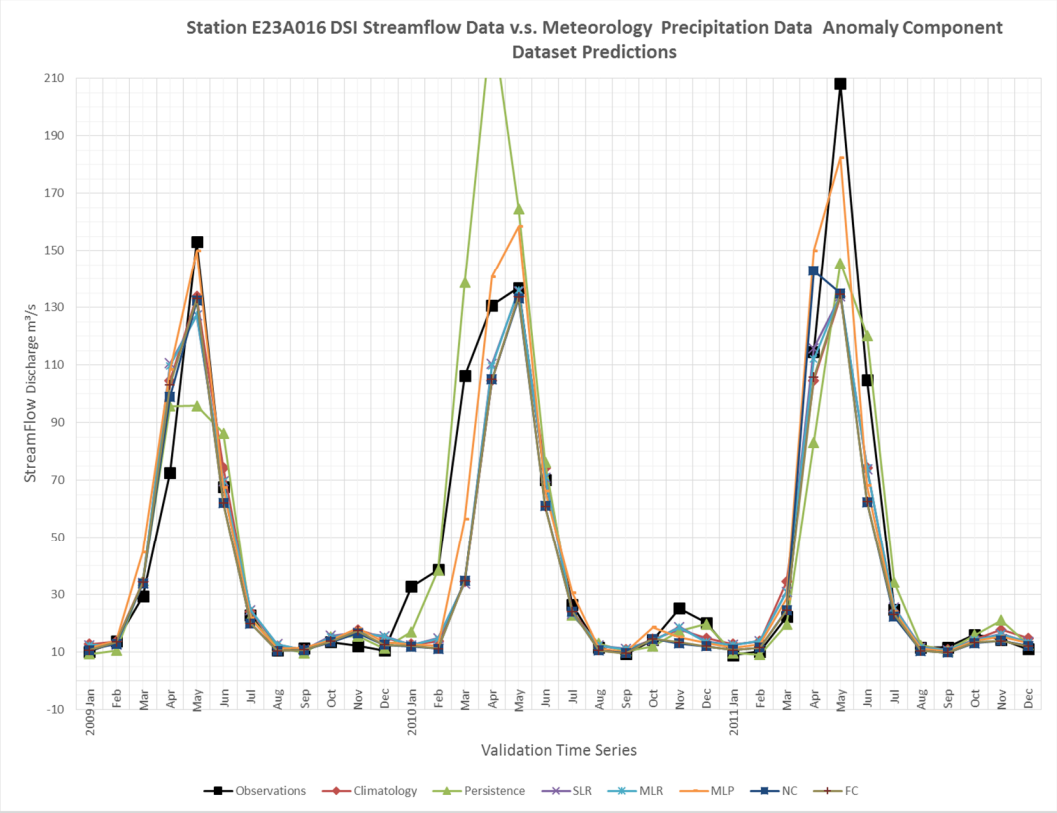


Figure 3.15 Station E23A016 DSI streamflow data v.s. meteorology precipitation data anomaly component data set predictions

Addition of Climatology component to the standardized anomaly component predictions showed above 0.90 correlations, these improved predictions are shoving heavy majority of the predictive skill and the relation between the precipitation and the streamflow data sets are due to the strong seasonality.

Correlations of predictions which are obtained from the SLR, MLR, MLP and Copula models increased from 0.20 to above 0.90 with decreasing RMSE values and mean error term. Performance statistics of validation time series are provided in Table 3.17.

Table 3.17 Performance statistics of station E23A016 with meteorology data – anomaly component data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,922	0,859	0,918	0,921	0,930	0,929	0,925
Standard Deviation of Observation	54,960	54,960	54,960	54,960	54,960	54,960	54,960
Standard Deviation of Errors	23,653	30,152	23,898	23,813	20,265	21,912	23,278
Mean Error	7,927	-0,825	8,204	8,334	2,713	9,029	9,991
RMSE	24,829	30,006	25,149	25,112	20,341	23,593	25,220
Mean of Observations	38,983	38,983	38,983	38,983	38,983	38,983	38,983

3.3.3.2. Predictions Using TRMM Single Pixel Data

Figure 3.16 shows the validation time series of the predictions and performance statistics of validation time series are provided in Table 3.18. Overall correlations and RMSE values show a consistent pattern; the methods with higher correlation values have lower RMSE value. Among all the complete data set predictions, Climatology-Based predictions provided the best result for the streamflow data while Persistence-Based predictions of the streamflow have approximately 0.90 correlations with the validation data.

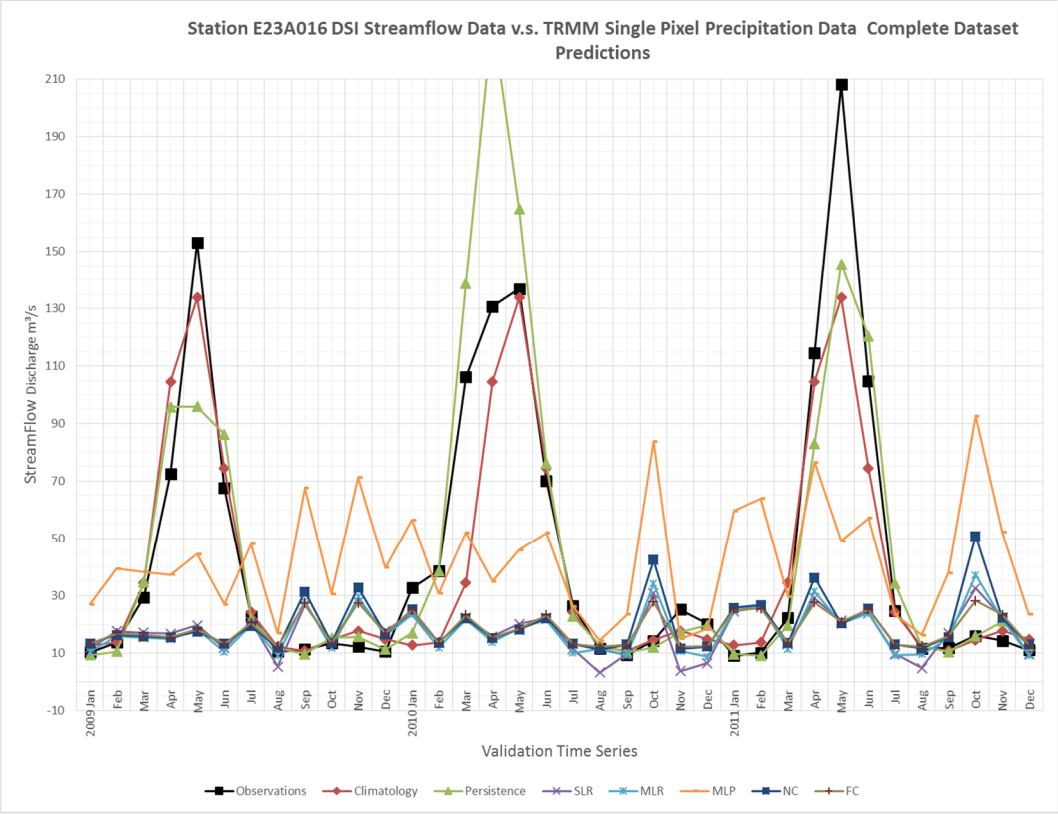


Figure 3.16 Station E23A016 DSI streamflow data v.s. TRMM single pixel precipitation data complete data set predictions

Table 3.18 Performance statistics of station E23A016 with TRMM single pixel data - complete data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,927	0,892	0,225	0,138	0,157	0,040	0,160
Standard Deviation of Observation	50,746	50,746	50,746	50,746	50,746	50,746	50,746
Standard Deviation of Errors	20,073	24,869	49,565	50,266	51,582	51,231	50,146
Mean Error	5,679	-1,803	27,350	27,035	2,337	24,915	26,246
RMSE	20,591	24,588	56,004	56,457	50,914	56,325	55,979
Mean of Observations	38,983	38,983	38,983	38,983	38,983	38,983	38,983

Similar results are obtained from the SLR, MLR, MLP and Copula models but showing under 0.20 correlation. These models could not identify a skillful relationship between the Rainfall and Runoff mainly due to lack of data. 9 years of monthly streamflow data and precipitation data could not train the models.

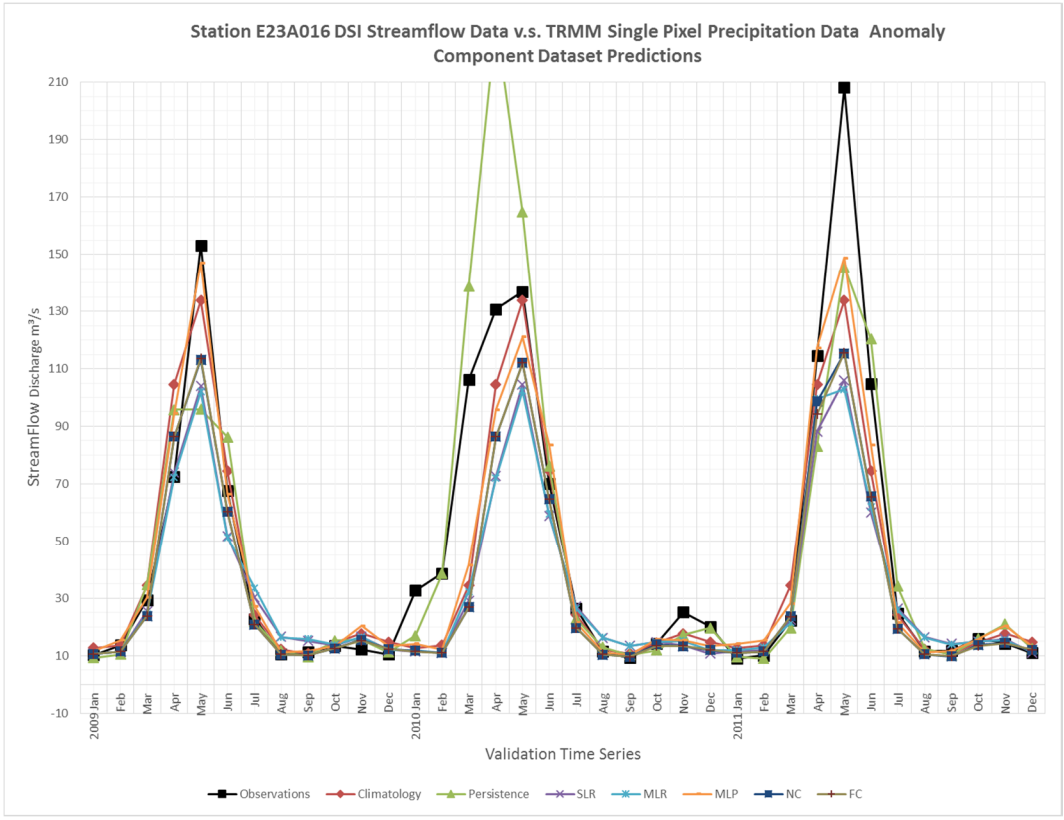


Figure 3.17 Station E23A016 DSI streamflow data v.s. TRMM single pixel precipitation data anomaly component data set predictions

Additional predictions are made using the standardized anomaly components of data sets. Figure 3.17 shows the validation time series of the anomaly component predictions.

Addition of Climatology component to the standardized anomaly component predictions showed above 0.90 correlations, these improved predictions are shoving heavy majority of the predictive skill and the relation between the precipitation and the streamflow data sets are due to the strong seasonality.

Correlations of predictions which are obtained from the SLR, MLR, MLP and Copula models increased from 0.20 to above 0.90 with decreasing RMSE values and mean error term. Performance statistics of validation time series are provided in Table 3.19.

Table 3.19 Performance statistics of station E23A016 with TRMM single pixel data – anomaly component data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,927	0,892	0,929	0,925	0,945	0,928	0,930
Standard Deviation of Observation	50,746	50,746	50,746	50,746	50,746	50,746	50,746
Standard Deviation of Errors	20,073	24,869	25,330	25,486	17,560	22,514	22,508
Mean Error	5,679	-1,803	12,300	11,996	5,236	11,749	11,882
RMSE	20,591	24,588	27,840	27,846	18,089	25,116	25,174
Mean of Observations	38,983	38,983	38,983	38,983	38,983	38,983	38,983

3.3.3.3. Predictions Using TRMM Catchment Average Data

Figure 3.18 shows the validation time series of the predictions and performance statistics of validation time series are provided in Table 3.20. Overall correlations and RMSE values show a consistent pattern; the methods with higher correlation values have lower RMSE value. Among all the complete data set predictions, Climatology-Based predictions provided the best result for the streamflow data while Persistence-Based predictions of the streamflow have approximately 0.90 correlations with the validation data.

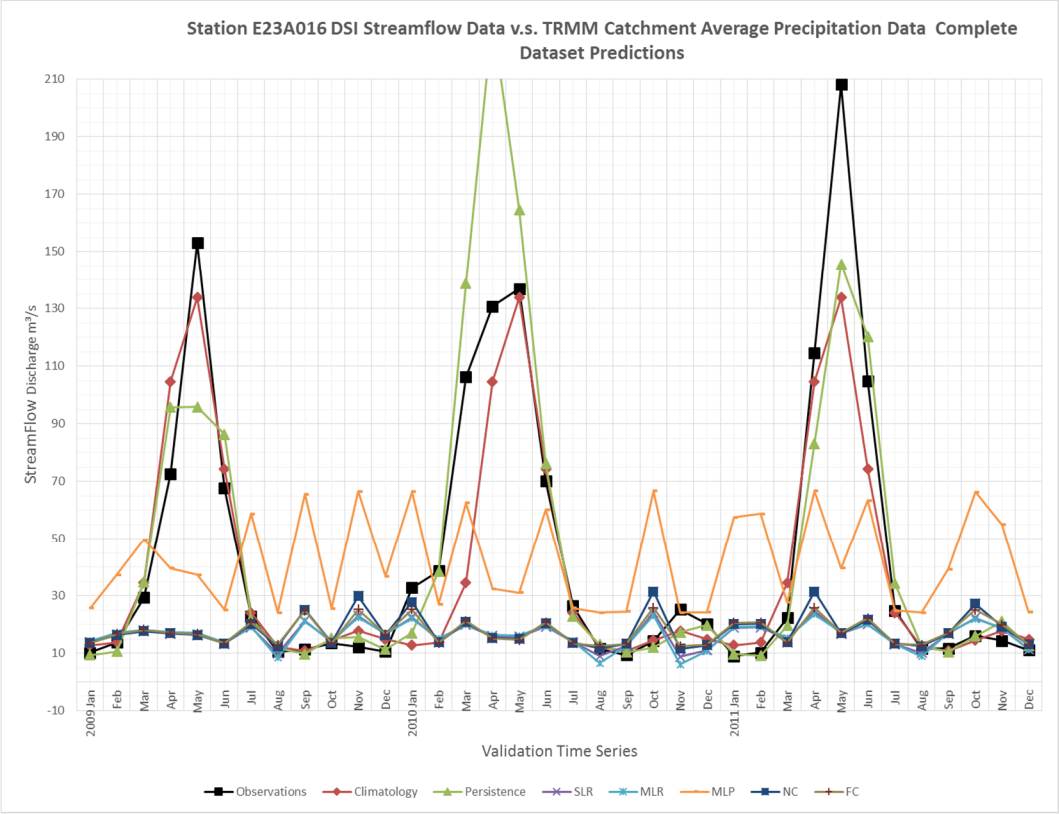


Figure 3.18 Station E23A016 DSI streamflow data v.s. TRMM catchment average precipitation data complete data set predictions

Table 3.20 Performance statistics of station E23A016 with TRMM catchment average data - complete data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,927	0,892	0,182	0,215	0,084	0,063	0,073
Standard Deviation of Observation	50,746	50,746	50,746	50,746	50,746	50,746	50,746
Standard Deviation of Errors	20,073	24,869	50,144	49,969	52,148	50,712	50,617
Mean Error	5,679	-1,803	28,245	28,382	2,766	26,771	27,146
RMSE	20,591	24,588	56,941	56,860	51,493	56,718	56,814
Mean of Observations	38,983	38,983	38,983	38,983	38,983	38,983	38,983

Similar results are obtained from the SLR, MLR, MLP and Copula models but showing under 0.20 correlation. These models could not identify a skillful relationship between the Rainfall and Runoff mainly due to lack of data. 9 years of monthly streamflow data and precipitation data could not train the models.

Additional predictions are made using the standardized anomaly components of data sets. Figure 3.19 shows the validation time series of the anomaly component predictions.

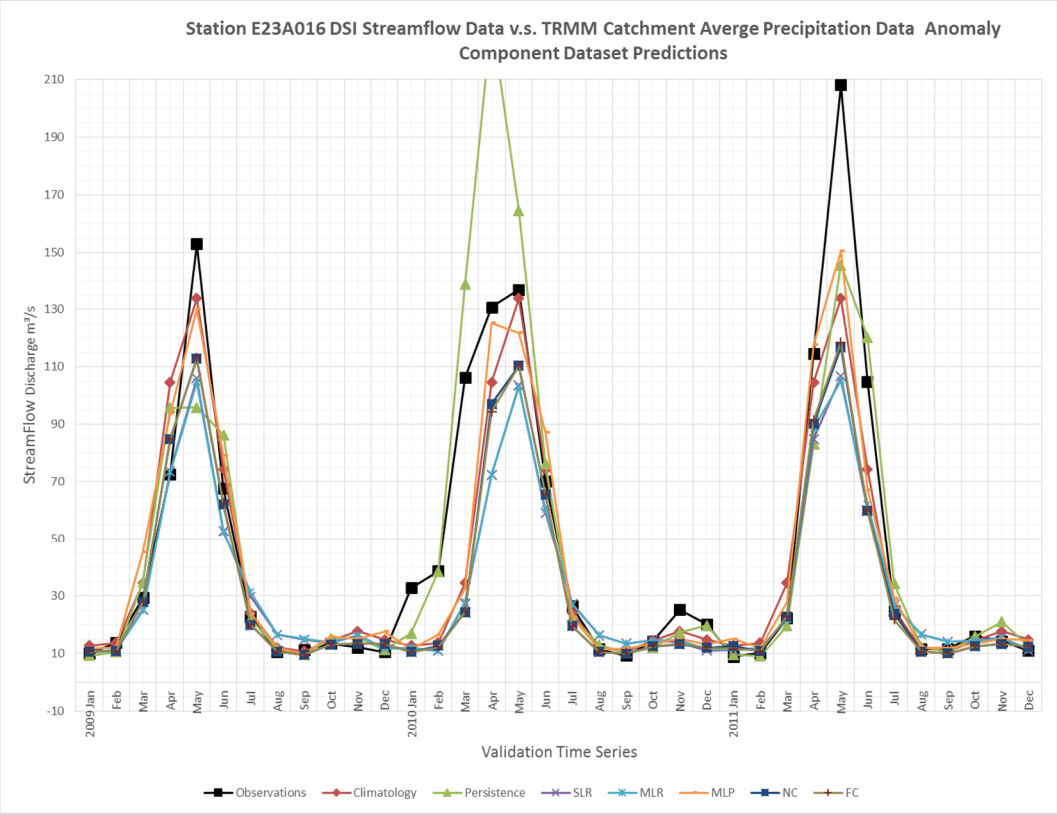


Figure 3.19 Station E23A016 DSI streamflow data v.s. TRMM catchment average precipitation data anomaly component data set predictions

Addition of Climatology component to the standardized anomaly component predictions showed above 0.90 correlations, these improved predictions are shoving heavy majority of the predictive skill and the relation between the precipitation and the streamflow data sets are due to the strong seasonality.

Correlations of predictions which are obtained from the SLR, MLR, MLP and Copula models increased from 0.20 to above 0.90 with decreasing RMSE values and mean error term. Performance statistics of validation time series are provided in Table 3.21.

Table 3.21 Performance statistics of station E23A016 with catchment average data – anomaly component data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,927	0,892	0,928	0,927	0,932	0,928	0,930
Standard Deviation of Observation	50,746	50,746	50,746	50,746	50,746	50,746	50,746
Standard Deviation of Errors	20,073	24,869	25,390	25,414	18,923	22,546	22,396
Mean Error	5,679	-1,803	12,434	12,271	4,767	11,787	11,840
RMSE	20,591	24,588	27,953	27,902	19,258	25,162	25,057
Mean of Observations	38,983	38,983	38,983	38,983	38,983	38,983	38,983

3.3.4. Station E23A020

3.3.4.1. Predictions Using Meteorology Data

Figure 3.20 shows the validation time series of the predictions and performance statistics of validation time series are provided in Table 3.22. Overall correlations and RMSE values show a consistent pattern; the methods with higher correlation values have lower RMSE value. Among all the complete data set predictions, Climatology-Based predictions provided the best result for the streamflow data while Persistence-Based predictions of the streamflow have approximately 0.90 correlations with the validation data.

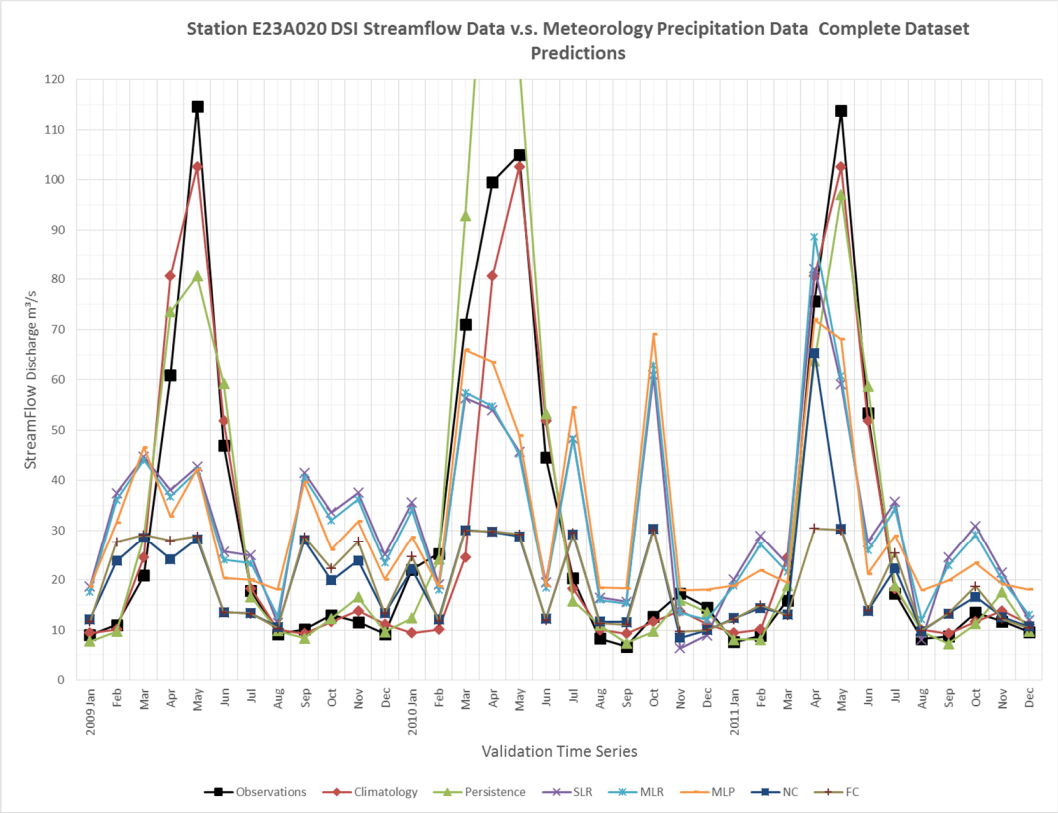


Figure 3.20 Station E23A020 DSI streamflow data v.s. meteorology precipitation data complete data set predictions

Table 3.22 Performance statistics of station E23A020 with meteorology data - complete data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,931	0,882	0,600	0,604	0,604	0,586	0,578
Standard Deviation of Observation	38,335	38,335	38,335	38,335	38,335	38,335	38,335
Standard Deviation of Errors	14,913	19,861	31,025	30,837	30,988	32,848	34,209
Mean Error	4,685	-1,217	2,069	1,994	2,114	13,878	14,694
RMSE	15,557	19,794	30,933	30,740	30,899	35,501	37,067
Mean of Observations	29,387	29,387	29,387	29,387	29,387	29,387	29,387

Similar results are obtained from the SLR, MLR and MLP models but showing 0.60 correlation. Copula functions couldn't identify a skillful relationship for complete data sets. To further investigate the source of the predictive skills of these methods, additional predictions are made using the standardized anomaly components of data sets. Figure 3.21 shows the validation time series of the anomaly component predictions.

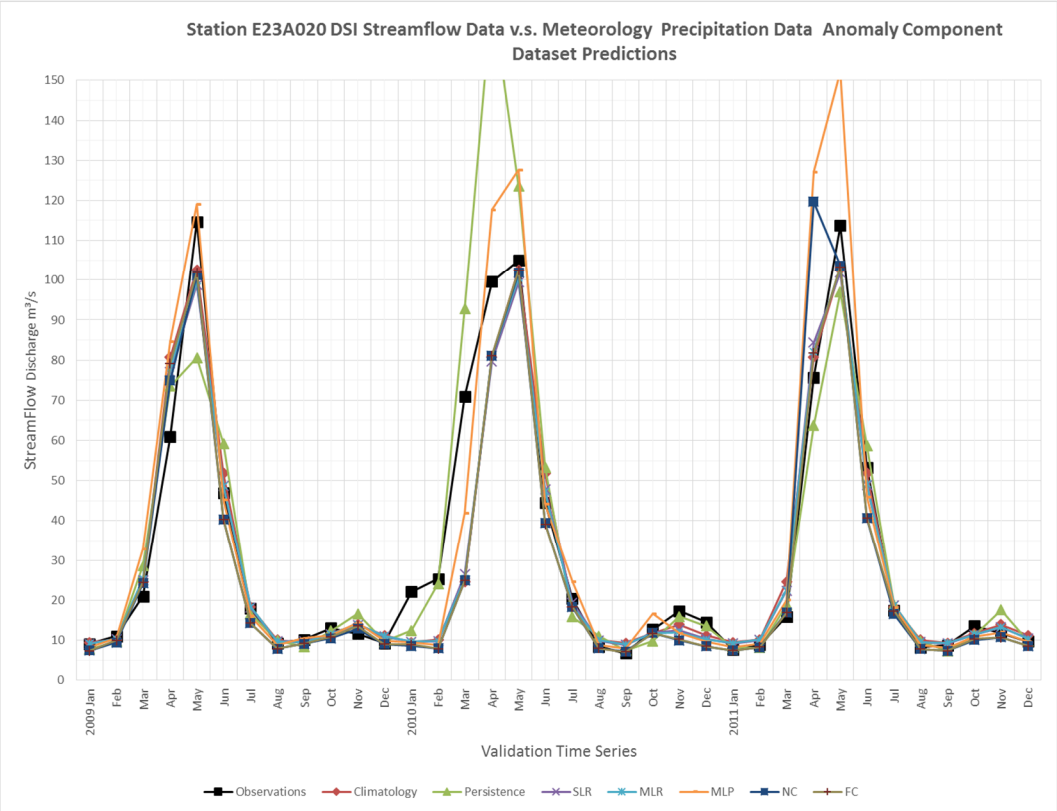


Figure 3.21 Station E23A020 DSI streamflow data v.s. meteorology precipitation data anomaly component data set predictions

Addition of Climatology component to the standardized anomaly component predictions showed above 0.90 correlations, these improved predictions are shoving heavy majority of the predictive skill and the relation between the precipitation and the streamflow data sets are due to the strong seasonality.

Correlations of predictions which are obtained from the SLR, MLR, MLP and Copula models increased from 0.60 to above 0.90 with decreasing RMSE values and mean error term. Performance statistics of validation time series are provided in Table 3.23.

Table 3.23 Performance statistics of station E23A020 with meteorology data – anomaly component data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,931	0,882	0,935	0,934	0,930	0,931	0,934
Standard Deviation of Observation	38,335	38,335	38,335	38,335	38,335	38,335	38,335
Standard Deviation of Errors	14,913	19,861	14,750	14,753	15,745	14,167	14,499
Mean Error	4,685	-1,217	5,259	5,153	-0,525	5,701	6,718
RMSE	15,557	19,794	15,587	15,555	15,671	15,203	15,911
Mean of Observations	29,387	29,387	29,387	29,387	29,387	29,387	29,387

3.3.4.2. Predictions Using TRMM Single Pixel Data

Figure 3.22 shows the validation time series of the predictions and performance statistics of validation time series are provided in Table 3.24. Overall correlations and RMSE values show a consistent pattern; the methods with higher correlation values have lower RMSE value. Among all the complete data set predictions, Climatology-Based predictions provided the best result for the streamflow data while Persistence-Based predictions of the streamflow have approximately 0.90 correlations with the validation data.

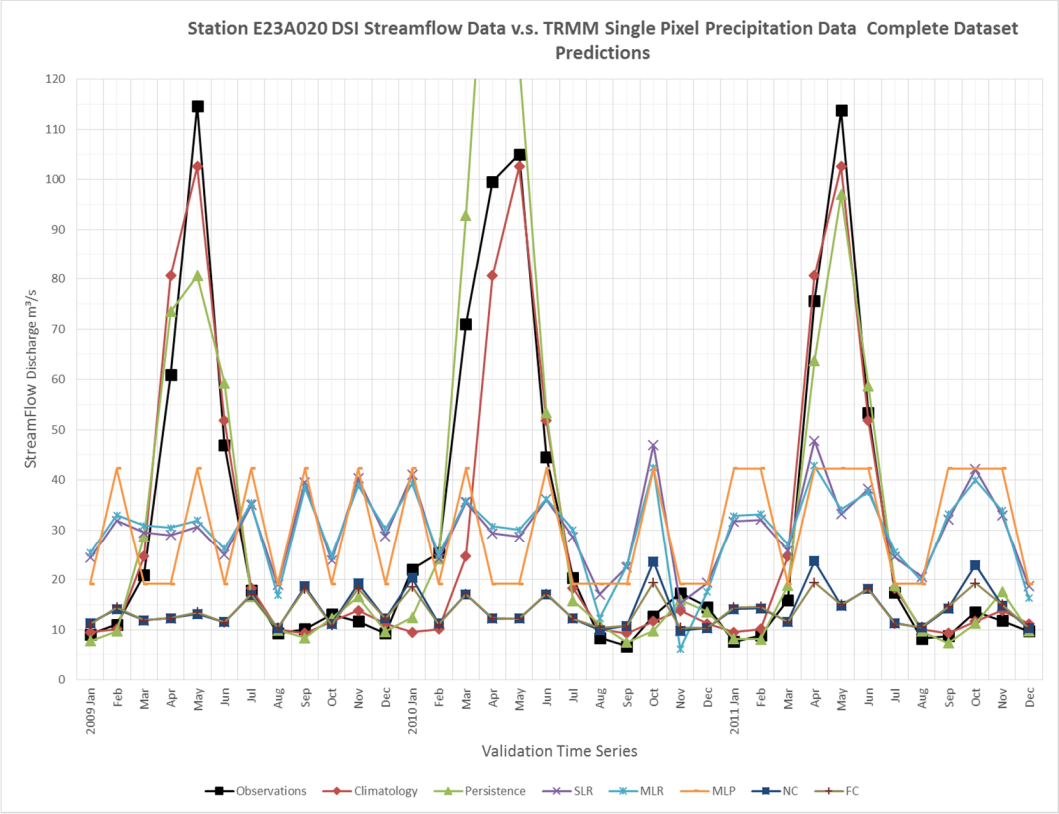


Figure 3.22 Station E23A020 DSI streamflow data v.s. TRMM single pixel precipitation data complete data set predictions

Table 3.24 Performance statistics of station E23A020 with TRMM single pixel data - complete data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,951	0,923	0,214	0,258	0,131	0,106	0,135
Standard Deviation of Observation	33,128	33,128	33,128	33,128	33,128	33,128	33,128
Standard Deviation of Errors	10,289	15,124	32,378	32,007	33,667	32,944	32,852
Mean Error	1,822	-2,379	1,265	1,652	1,236	17,144	17,486
RMSE	10,307	15,101	31,950	31,602	33,219	36,730	36,811
Mean of Observations	29,387	29,387	29,387	29,387	29,387	29,387	29,387

Similar results are obtained from the SLR, MLR, MLP and Copula models but showing under 0.30 correlation. These models could not identify a skillful relationship between the Rainfall and Runoff mainly due to lack of data. 9 years of monthly streamflow data and precipitation data could not train the models.

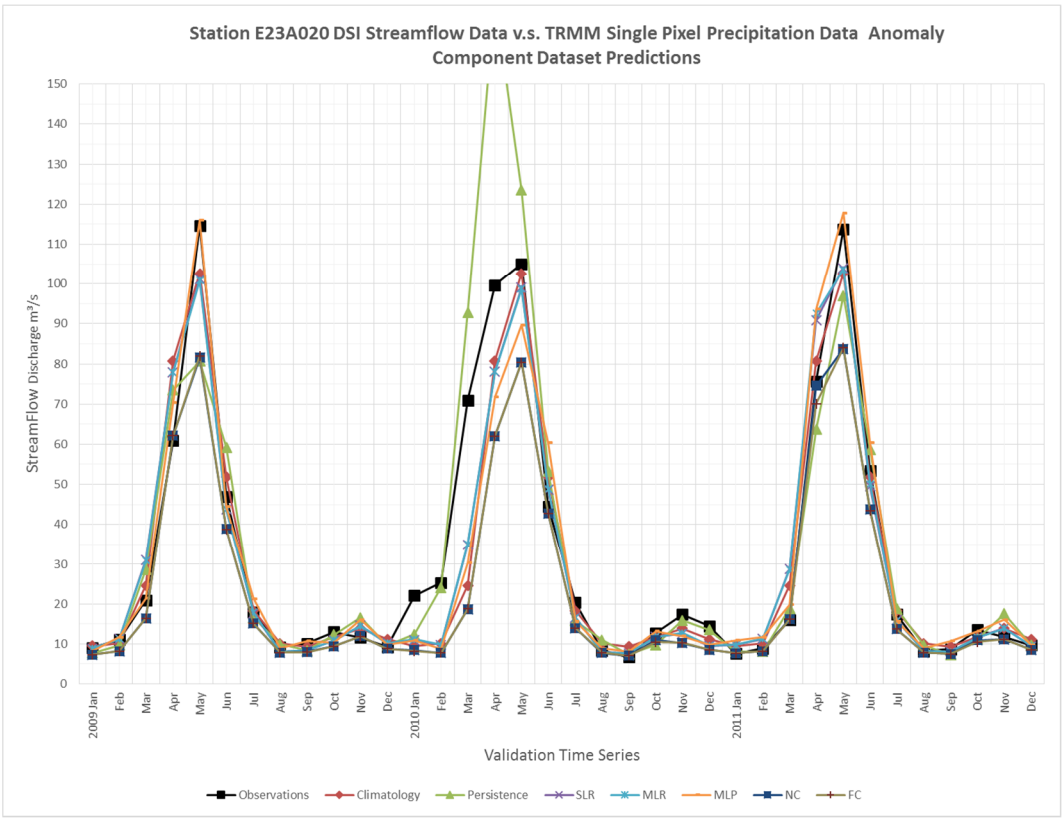


Figure 3.23 Station E23A020 DSI streamflow data v.s. TRMM single pixel precipitation data anomaly component data set predictions

Additional predictions are made using the standardized anomaly components of data sets. Figure 3.23 shows the validation time series of the anomaly component predictions.

Addition of Climatology component to the standardized anomaly component predictions showed above 0.90 correlations, these improved predictions are shoving heavy majority of the predictive skill and the relation between the precipitation and the streamflow data sets are due to the strong seasonality.

Correlations of predictions which are obtained from the SLR, MLR, MLP and Copula models increased from 0.30 to above 0.90 with decreasing RMSE values and mean error term. Performance statistics of validation time series are provided in Table 3.25.

Table 3.25 Performance statistics of station E23A020 with TRMM single pixel data – anomaly component data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,951	0,923	0,957	0,957	0,948	0,946	0,949
Standard Deviation of Observation	33,128	33,128	33,128	33,128	33,128	33,128	33,128
Standard Deviation of Errors	10,289	15,124	9,593	9,673	10,566	12,485	12,394
Mean Error	1,822	-2,379	2,061	2,015	1,383	7,756	7,895
RMSE	10,307	15,101	9,681	9,748	10,509	14,550	14,549
Mean of Observations	29,387	29,387	29,387	29,387	29,387	29,387	29,387

3.3.4.3. Predictions Using TRMM Catchment Average Data

Figure 3.24 shows the validation time series of the predictions and performance statistics of validation time series are provided in Table 3.26. Overall correlations and RMSE values show a consistent pattern; the methods with higher correlation values have lower RMSE value. Among all the complete data set predictions, Climatology-Based predictions provided the best result for the streamflow data while Persistence-Based predictions of the streamflow have approximately 0.90 correlations with the validation data.

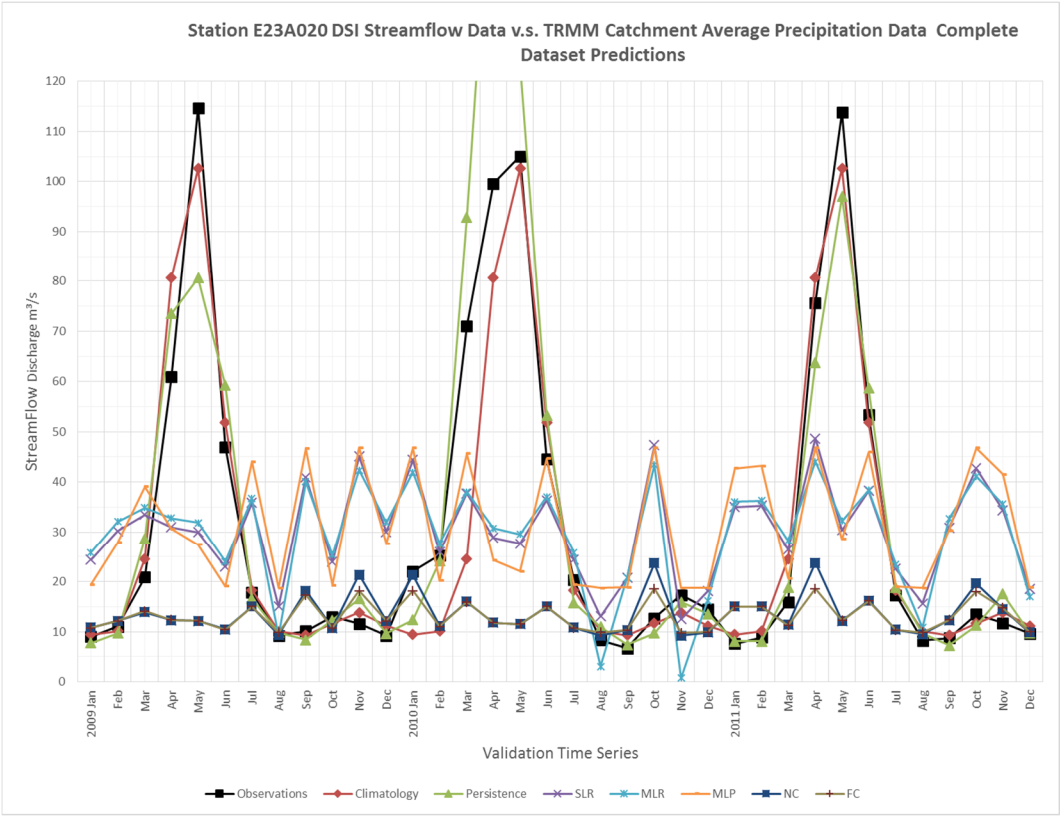


Figure 3.24 Station E23A020 DSI streamflow data v.s. TRMM catchment average precipitation data complete data set predictions

Table 3.26 Performance statistics of station E23A020 with TRMM catchment average data - complete data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,951	0,923	0,167	0,227	0,038	0,044	0,046
Standard Deviation of Observation	33,128	33,128	33,128	33,128	33,128	33,128	33,128
Standard Deviation of Errors	10,289	15,124	32,923	32,453	34,738	33,201	33,124
Mean Error	1,822	-2,379	1,314	1,890	0,319	17,719	18,187
RMSE	10,307	15,101	32,489	32,055	34,254	37,224	37,383
Mean of Observations	29,387	29,387	29,387	29,387	29,387	29,387	29,387

Similar results are obtained from the SLR, MLR, MLP and Copula models but showing under 0.20 correlation. These models could not identify a skillful relationship between the Rainfall and Runoff mainly due to lack of data. 9 years of monthly streamflow data and precipitation data could not train the models.

Additional predictions are made using the standardized anomaly components of data sets. Figure 3.25 shows the validation time series of the anomaly component predictions.

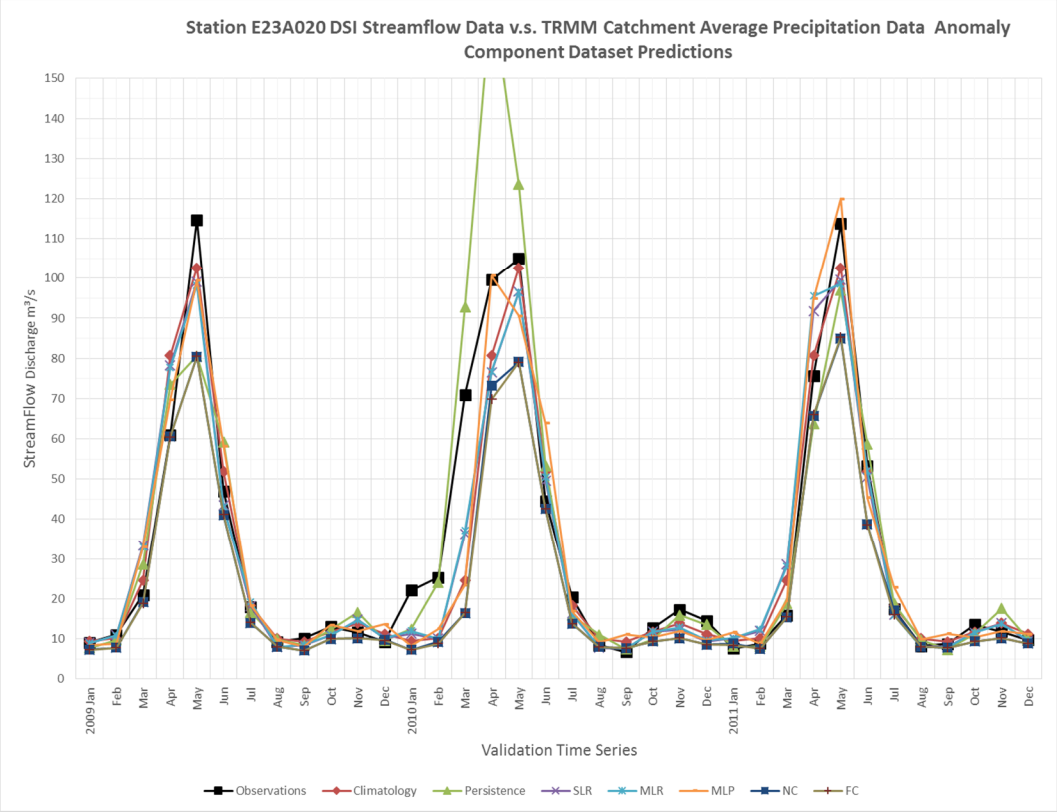


Figure 3.25 Station E23A020 DSI streamflow data v.s. TRMM catchment average precipitation data anomaly component data set predictions

Addition of Climatology component to the standardized anomaly component predictions showed above 0.90 correlations, these improved predictions are shoving heavy majority of the predictive skill and the relation between the precipitation and the streamflow data sets are due to the strong seasonality.

Correlations of predictions which are obtained from the SLR, MLR, MLP and Copula models increased from 0.20 to above 0.90 with decreasing RMSE values and mean error term. Performance statistics of validation time series are provided in Table 3.27.

Table 3.27 Performance statistics of station E23A020 with catchment average data – anomaly component data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,951	0,923	0,955	0,953	0,942	0,952	0,951
Standard Deviation of Observation	33,128	33,128	33,128	33,128	33,128	33,128	33,128
Standard Deviation of Errors	10,289	15,124	9,956	10,110	11,253	12,140	12,254
Mean Error	1,822	-2,379	2,146	1,976	0,774	7,886	8,012
RMSE	10,307	15,101	10,048	10,163	11,123	14,335	14,498
Mean of Observations	29,387	29,387	29,387	29,387	29,387	29,387	29,387

3.3.5. Station E23A021

3.3.5.1. Predictions Using Meteorology Data

Figure 3.26 shows the validation time series of the predictions and performance statistics of validation time series are provided in Table 3.28. Overall correlations and RMSE values show a consistent pattern; the methods with higher correlation values have lower RMSE value. Among all the complete data set predictions, Climatology-Based predictions provided the best result for the streamflow data while Persistence-Based predictions of the streamflow have approximately 0.90 correlations with the validation data.

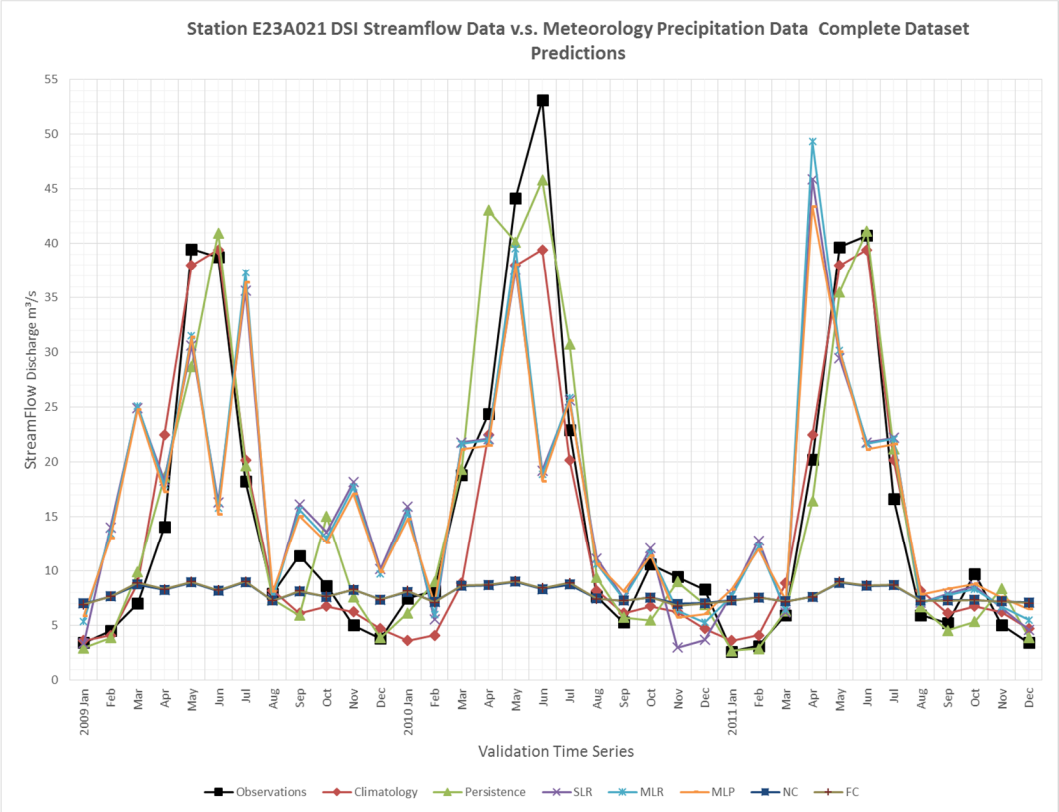


Figure 3.26 Station E23A021 DSI streamflow data v.s. meteorology precipitation data complete data set predictions

Table 3.28 Performance statistics of station E23A021 with meteorology data - complete data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,940	0,885	0,572	0,566	0,573	0,613	0,608
Standard Deviation of Observation	14,164	14,164	14,164	14,164	14,164	14,164	14,164
Standard Deviation of Errors	4,891	6,850	11,724	11,835	11,672	13,762	13,730
Mean Error	1,429	-0,189	-0,356	-0,526	-0,438	7,586	7,569
RMSE	5,071	6,817	11,668	11,785	11,619	15,651	15,616
Mean of Observations	14,058	14,058	14,058	14,058	14,058	14,058	14,058

Similar results are obtained from the SLR, MLR and MLP models but showing 0.60 correlation. Copula models couldn't identify a skillful relationship for complete data sets. To further investigate the source of the predictive skills of these methods, additional predictions are made using the standardized anomaly components of data sets. Figure 3.27 shows the validation time series of the anomaly component predictions.

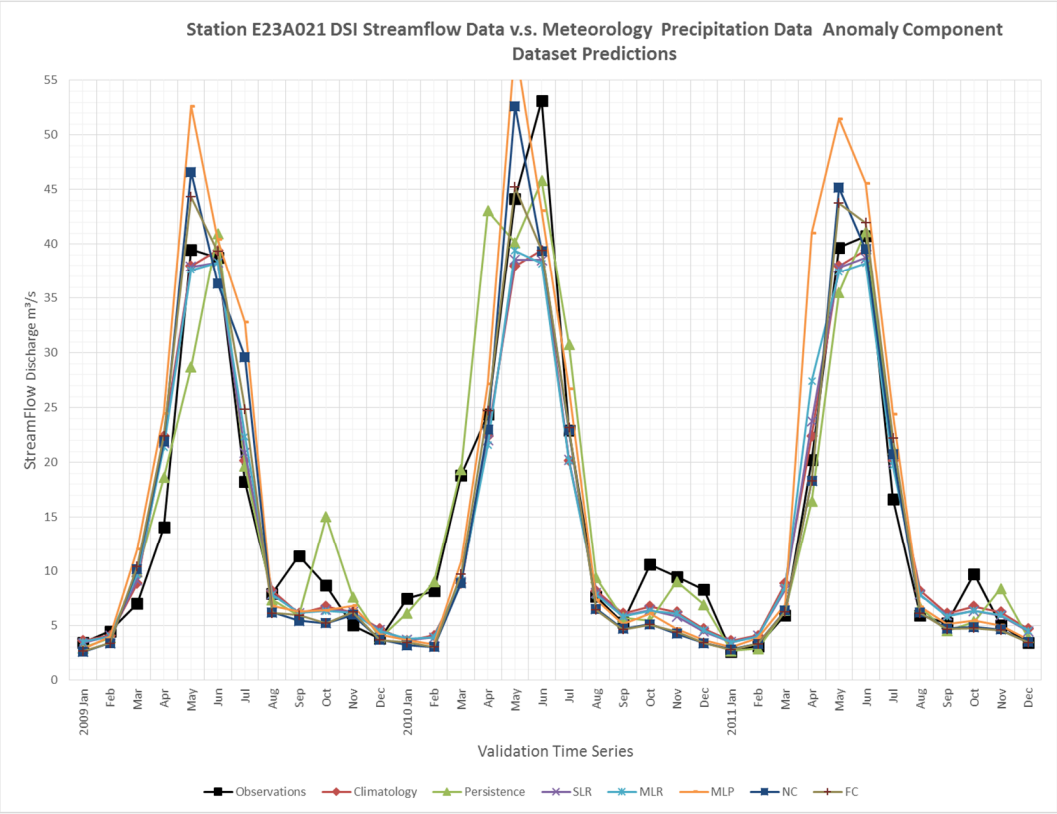


Figure 3.27 Station E23A021 DSI streamflow data v.s. meteorology precipitation data anomaly component data set predictions

Addition of Climatology component to the standardized anomaly component predictions showed above 0.90 correlations, these improved predictions are shoving heavy majority of the predictive skill and the relation between the precipitation and the streamflow data sets are due to the strong seasonality.

Correlations of predictions which are obtained from the SLR, MLR, MLP and Copula models increased from 0.50 to above 0.90 with decreasing RMSE values and mean error term. Performance statistics of validation time series are provided in Table 3.29.

Table 3.29 Performance statistics of station E23A021 with meteorology data – anomaly component data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,940	0,885	0,938	0,936	0,905	0,909	0,927
Standard Deviation of Observation	14,164	14,164	14,164	14,164	14,164	14,164	14,164
Standard Deviation of Errors	4,891	6,850	4,969	5,069	7,258	6,184	5,428
Mean Error	1,429	-0,189	1,635	1,605	-0,923	1,187	1,364
RMSE	5,071	6,817	5,207	5,292	7,278	6,265	5,570
Mean of Observations	14,058	14,058	14,058	14,058	14,058	14,058	14,058

3.3.5.2. Predictions Using TRMM Single Pixel Data

Figure 3.28 shows the validation time series of the predictions and performance statistics of validation time series are provided in Table 3.30. Overall correlations and RMSE values show a consistent pattern; the methods with higher correlation values have lower RMSE value. Among all the complete data set predictions, Climatology-Based predictions provided the best result for the streamflow data while Persistence-Based predictions of the streamflow have approximately 0.90 correlations with the validation data.

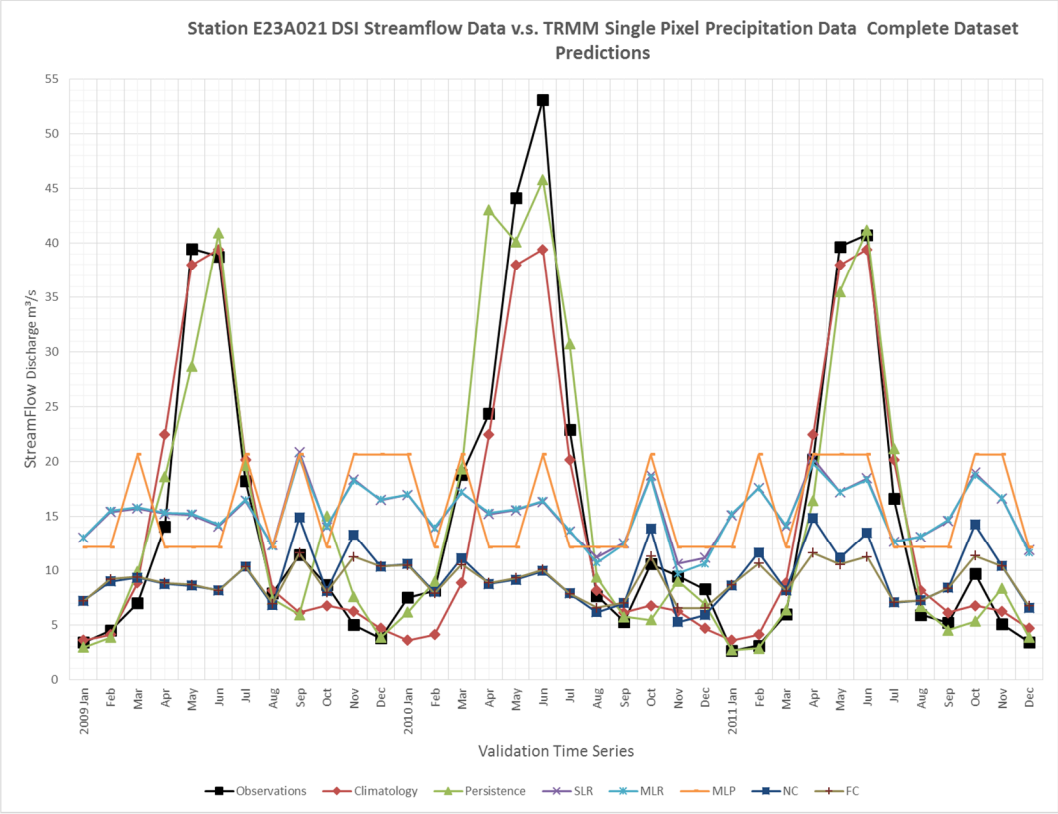


Figure 3.28 Station E23A021 DSI streamflow data v.s. TRMM single pixel precipitation data complete data set predictions

Table 3.30 Performance statistics of station E23A021 with TRMM single pixel data - complete data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,961	0,938	0,209	0,223	0,118	0,167	0,229
Standard Deviation of Observation	13,879	13,879	13,879	13,879	13,879	13,879	13,879
Standard Deviation of Errors	3,920	4,858	13,576	13,538	14,028	13,687	13,593
Mean Error	0,978	-0,255	-0,268	-0,205	-0,669	5,571	5,991
RMSE	3,987	4,797	13,389	13,350	13,848	14,601	14,681
Mean of Observations	14,058	14,058	14,058	14,058	14,058	14,058	14,058

Similar results are obtained from the SLR, MLR, MLP and Copula models but showing under 0.30 correlation. These models could not identify a skillful relationship between the Rainfall and Runoff mainly due to lack of data. 9 years of monthly streamflow data and precipitation data could not train the models.

Additional predictions are made using the standardized anomaly components of data sets. Figure 3.29 shows the validation time series of the anomaly component predictions.

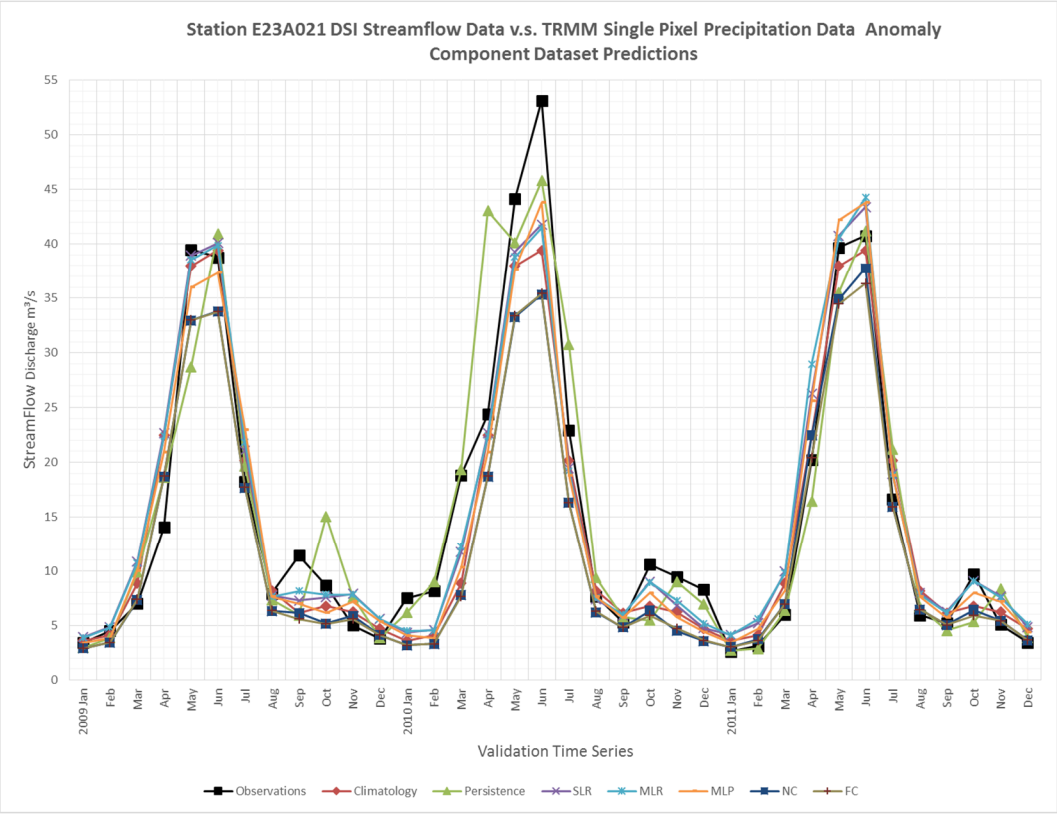


Figure 3.29 Station E23A021 DSI streamflow data v.s. TRMM single pixel precipitation data anomaly component data set predictions

Addition of Climatology component to the standardized anomaly component predictions showed above 0.90 correlations, these improved predictions are shoving heavy majority of the predictive skill and the relation between the precipitation and the streamflow data sets are due to the strong seasonality.

Correlations of predictions which are obtained from the SLR, MLR, MLP and Copula models increased from 0.20 to above 0.90 with decreasing RMSE values and mean error term. Performance statistics of validation time series are provided in Table 3.31.

Table 3.31 Performance statistics of station E23A021 with TRMM single pixel data – anomaly component data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,961	0,938	0,964	0,961	0,964	0,961	0,964
Standard Deviation of Observation	13,879	13,879	13,879	13,879	13,879	13,879	13,879
Standard Deviation of Errors	3,920	4,858	3,701	3,847	3,699	4,283	4,239
Mean Error	0,979	-0,255	0,009	-0,108	0,625	2,823	2,968
RMSE	3,987	4,796	3,650	3,795	3,700	5,080	5,126
Mean of Observations	14,058	14,058	14,058	14,058	14,058	14,058	14,058

3.3.5.3. Predictions Using TRMM Catchment Average Data

Figure 3.30 shows the validation time series of the predictions and performance statistics of validation time series are provided in Table 3.32. Overall correlations and RMSE values show a consistent pattern; the methods with higher correlation values have lower RMSE value. Among all the complete data set predictions, Climatology-Based predictions provided the best result for the streamflow data while Persistence-Based predictions of the streamflow have approximately 0.90 correlations with the validation data.

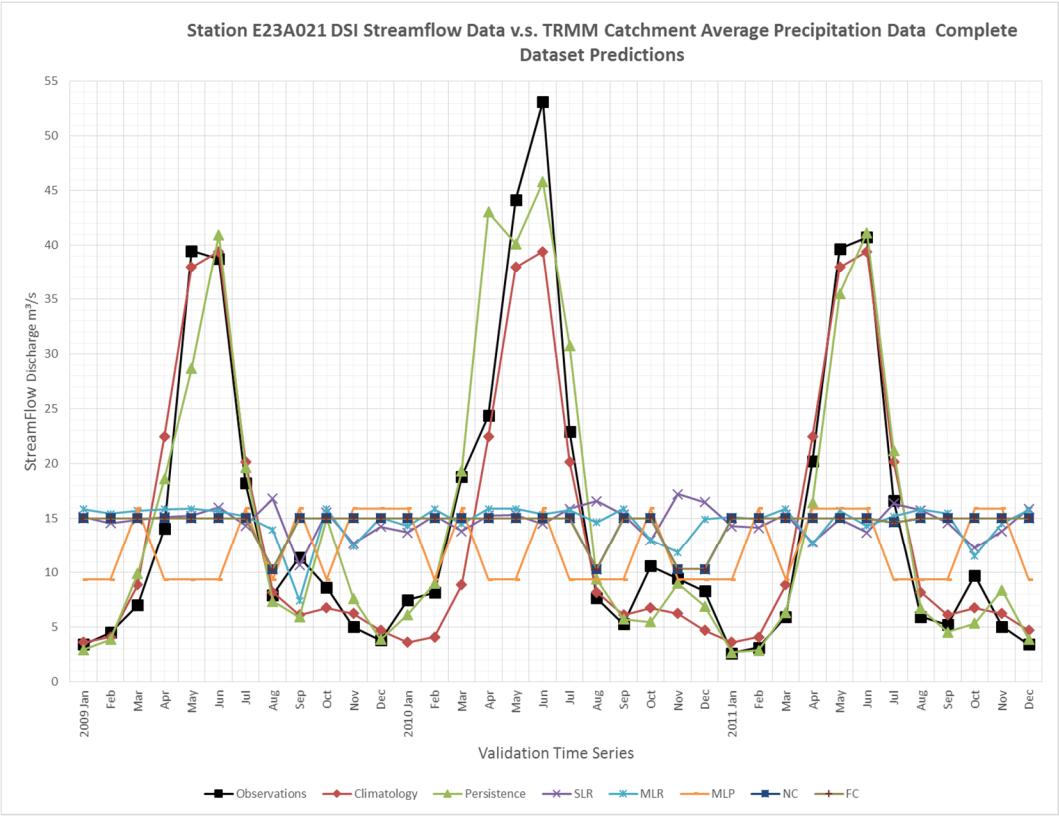


Figure 3.30 Station E23A021 DSI streamflow data v.s. TRMM catchment average precipitation data complete data set predictions

Table 3.32 Performance statistics of station E23A021 with TRMM catchment average data - complete data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,961	0,938	0,033	0,148	0,118	0,172	0,172
Standard Deviation of Observation	13,879	13,879	13,879	13,879	13,879	13,879	13,879
Standard Deviation of Errors	3,920	4,858	13,902	13,730	13,879	13,700	13,700
Mean Error	0,978	-0,255	0,255	0,325	2,943	0,532	0,532
RMSE	3,987	4,797	13,710	13,541	13,998	13,519	13,519
Mean of Observations	14,058	14,058	14,058	14,058	14,058	14,058	14,058

Similar results are obtained from the SLR, MLR, MLP and Copula models but showing under 0.20 correlation. These models could not identify a skillful relationship between the Rainfall and Runoff mainly due to lack of data. 9 years of monthly streamflow data and precipitation data could not train the models.

Additional predictions are made using the standardized anomaly components of data sets. Figure 3.31 shows the validation time series of the anomaly component predictions.

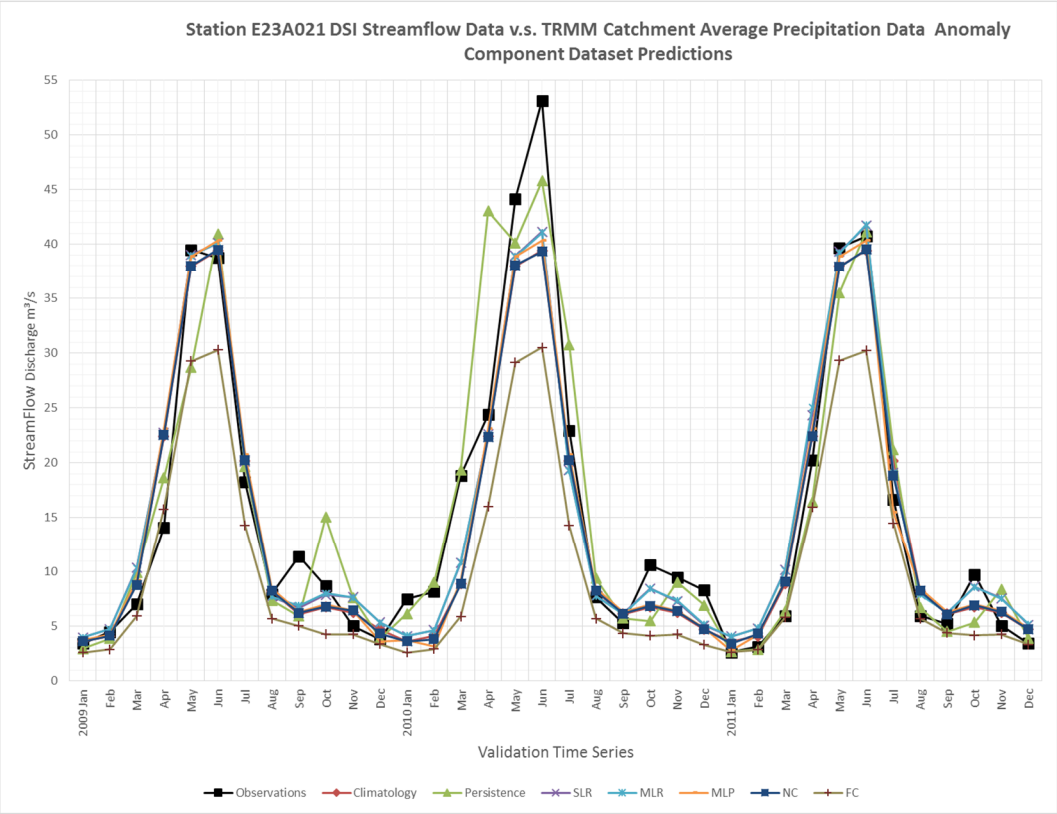


Figure 3.31 Station E23A021 DSI streamflow data v.s. TRMM catchment average precipitation data anomaly component data set predictions

Addition of Climatology component to the standardized anomaly component predictions showed above 0.90 correlations, these improved predictions are shoving heavy majority of the predictive skill and the relation between the precipitation and the streamflow data sets are due to the strong seasonality.

Correlations of predictions which are obtained from the SLR, MLR, MLP and Copula models increased from 0.20 to above 0.90 with decreasing RMSE values and mean error term. Performance statistics of validation time series are provided in Table 3.33.

Table 3.33 Performance statistics of station E23A021 with catchment average data – anomaly component data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,961	0,938	0,965	0,964	0,962	0,961	0,964
Standard Deviation of Observation	13,879	13,879	13,879	13,879	13,879	13,879	13,879
Standard Deviation of Errors	3,920	4,858	3,703	3,723	3,817	3,894	5,135
Mean Error	0,979	-0,255	0,267	0,246	0,828	1,005	4,763
RMSE	3,987	4,796	3,661	3,680	3,854	3,969	6,952
Mean of Observations	14,058	14,058	14,058	14,058	14,058	14,058	14,058

3.3.6. Station E23A023

3.3.6.1. Predictions Using Meteorology Data

Figure 3.32 shows the validation time series of the predictions and performance statistics of validation time series are provided in Table 3.34. Overall correlations and RMSE values show a consistent pattern; the methods with higher correlation values have lower RMSE value. Among all the complete data set predictions, Climatology-Based predictions provided the best result for the streamflow data while Persistence-Based predictions of the streamflow have approximately 0.90 correlations with the validation data.

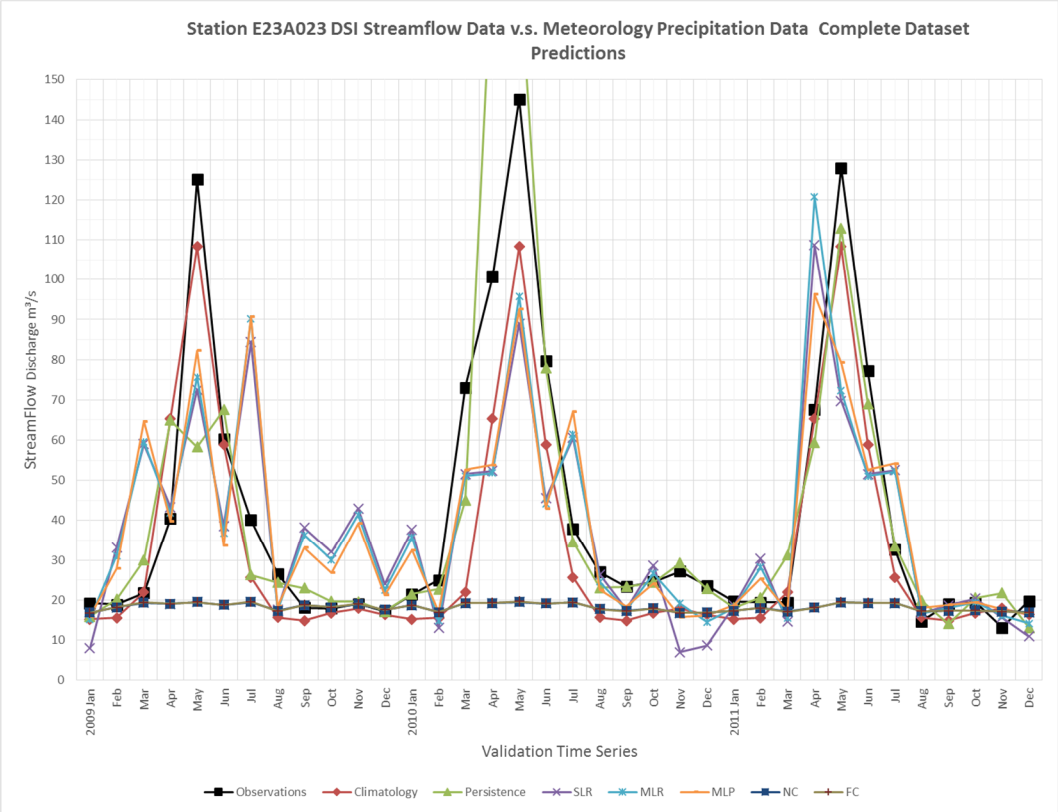


Figure 3.32 Station E23A023 DSI streamflow data v.s. meteorology precipitation data complete data set predictions

Table 3.34 Performance statistics of station E23A023 with meteorology data - complete data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,925	0,701	0,604	0,609	0,633	0,551	0,551
Standard Deviation of Observation	36,453	36,453	36,453	36,453	36,453	36,453	36,453
Standard Deviation of Errors	14,819	31,101	29,071	28,971	28,242	35,938	35,912
Mean Error	7,582	3,709	2,812	2,225	1,982	22,071	22,060
RMSE	16,577	31,160	29,056	28,905	28,164	42,015	41,986
Mean of Observations	33,506	33,506	33,506	33,506	33,506	33,506	33,506

Similar results are obtained from the SLR and MLR models but showing 0.60 correlation. Copula functions couldn't identify a skillful relationship for complete data sets. MLP model is also provided results that showing 0.60 correlation.To further investigate the source of the predictive skills of these methods, additional predictions are made using the standardized anomaly components of data sets. Figure 3.33 shows the validation time series of the anomaly component predictions.

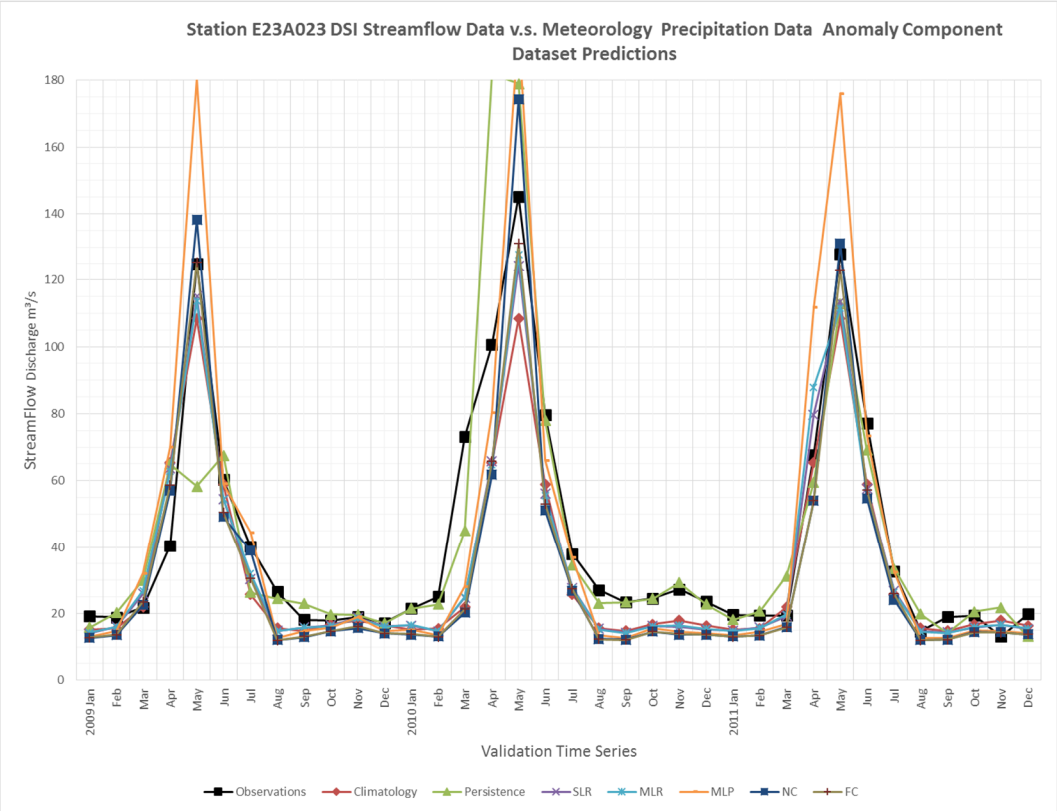


Figure 3.33 Station E23A023 DSI streamflow data v.s. meteorology precipitation data anomaly component data set predictions

Addition of Climatology component to the standardized anomaly component predictions showed above 0.90 correlations, these improved predictions are shoving heavy majority of the predictive skill and the relation between the precipitation and the streamflow data sets are due to the strong seasonality.

Correlations of predictions which are obtained from the SLR, MLR, MLP and Copula models increased from 0.50 to above 0.90 with decreasing RMSE values and mean error term. Performance statistics of validation time series are provided in Table 3.35.

Table 3.35 Performance statistics of station E23A023 with meteorology data – anomaly component data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,925	0,701	0,921	0,918	0,903	0,894	0,918
Standard Deviation of Observation	36,453	36,453	36,453	36,453	36,453	36,453	36,453
Standard Deviation of Errors	14,819	31,102	14,643	14,864	21,651	16,811	14,498
Mean Error	7,582	3,709	7,029	7,037	-0,674	7,132	8,495
RMSE	16,577	31,161	16,174	16,376	21,548	18,181	16,738
Mean of Observations	33,506	33,506	33,506	33,506	33,506	33,506	33,506

3.3.6.2. Predictions Using TRMM Single Pixel Data

Figure 3.34 shows the validation time series of the predictions and performance statistics of validation time series are provided in Table 3.36. Overall correlations and RMSE values show a consistent pattern; the methods with higher correlation values have lower RMSE value. Among all the complete data set predictions, Climatology-Based predictions provided the best result for the streamflow data while Persistence-Based predictions of the streamflow have approximately 0.90 correlations with the validation data.

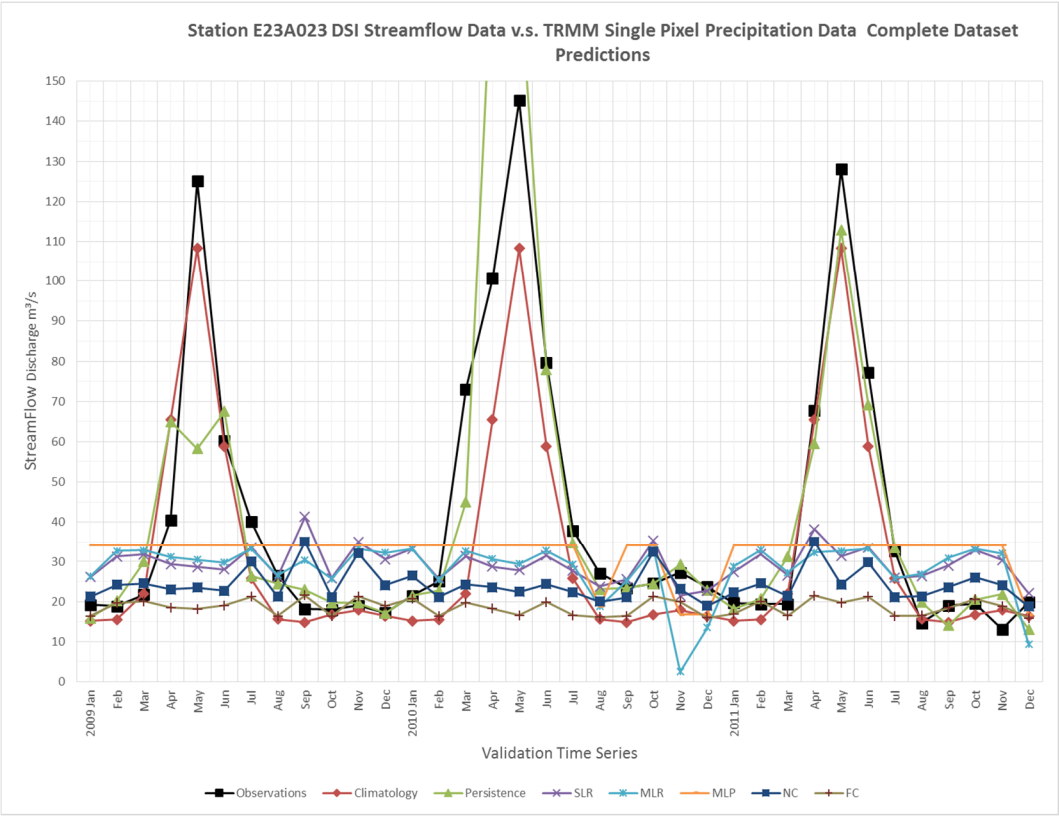


Figure 3.34 Station E23A023 DSI streamflow data v.s. TRMM single pixel precipitation data complete data set predictions

Table 3.36 Performance statistics of station E23A023 with TRMM single pixel data - complete data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,940	0,863	0,106	0,220	0,171	0,065	0,093
Standard Deviation of Observation	35,399	35,399	35,399	35,399	35,399	35,399	35,399
Standard Deviation of Errors	12,954	20,468	35,206	34,544	34,887	35,375	35,270
Mean Error	8,487	-0,661	11,796	12,953	8,921	16,900	22,642
RMSE	15,335	20,193	36,663	36,440	35,537	38,759	41,498
Mean of Observations	33,506	33,506	33,506	33,506	33,506	33,506	33,506

Similar results are obtained from the SLR, MLR, MLP and Copula models but showing under 0.20 correlation. These models could not identify a skillful relationship between the Rainfall and Runoff mainly due to lack of data and showed artificial skills in predictions. 9 years of monthly streamflow data and precipitation data could not train the models.

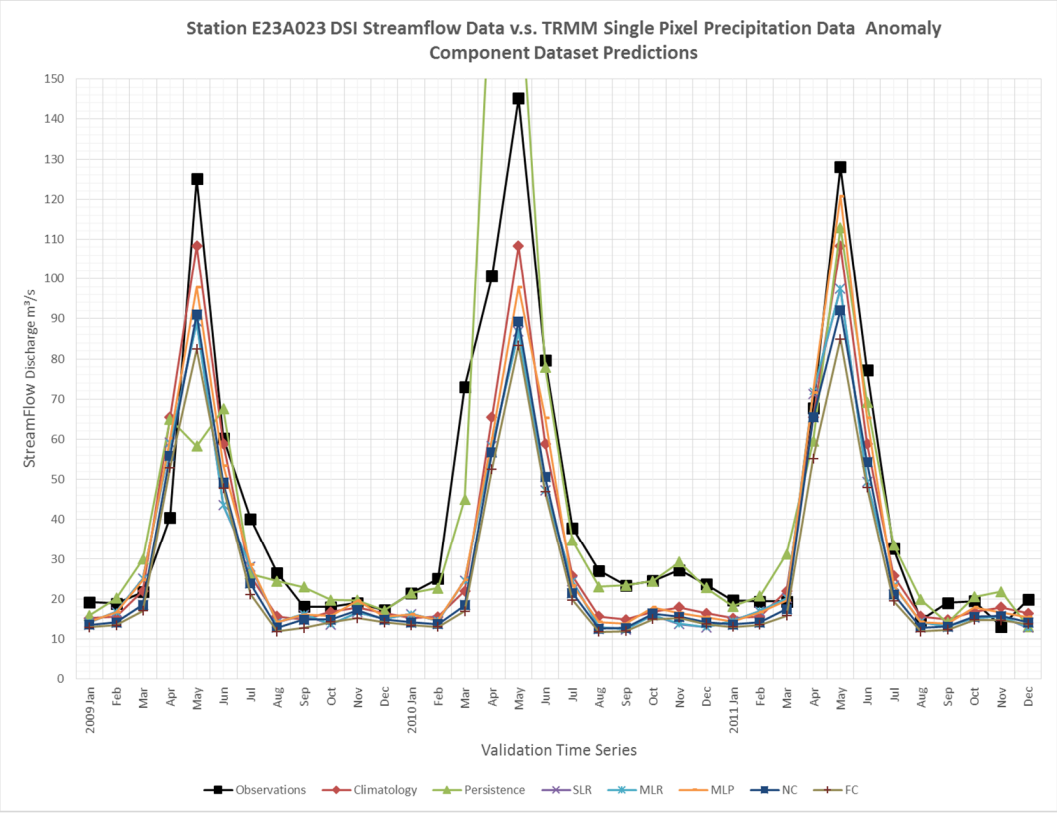


Figure 3.35 Station E23A023 DSI streamflow data v.s. TRMM single pixel precipitation data anomaly component data set predictions

Additional predictions are made using the standardized anomaly components of data sets. Figure 3.35 shows the validation time series of the anomaly component predictions.

Addition of Climatology component to the standardized anomaly component predictions showed above 0.90 correlations, these improved predictions are shoving heavy majority of the predictive skill and the relation between the precipitation and the streamflow data sets are due to the strong seasonality.

Correlations of predictions which are obtained from the SLR, MLR, MLP and Copula models increased from 0.20 to above 0.90 with decreasing RMSE values and mean error term. Performance statistics of validation time series are provided in Table 3.37.

Table 3.37 Performance statistics of station E23A023 with TRMM single pixel data – anomaly component data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,940	0,863	0,926	0,925	0,933	0,933	0,934
Standard Deviation of Observation	35,399	35,399	35,399	35,399	35,399	35,399	35,399
Standard Deviation of Errors	12,955	20,468	15,807	15,852	13,488	15,405	16,640
Mean Error	8,488	-0,660	11,874	11,857	8,781	12,714	14,898
RMSE	15,336	20,192	19,594	19,619	15,937	19,808	22,162
Mean of Observations	33,506	33,506	33,506	33,506	33,506	33,506	33,506

3.3.6.3. Predictions Using TRMM Catchment Average Data

Figure 3.36 shows the validation time series of the predictions and performance statistics of validation time series are provided in Table 3.38. Overall correlations and RMSE values show a consistent pattern; the methods with higher correlation values have lower RMSE value. Among all the complete data set predictions, Climatology-Based predictions provided the best result for the streamflow data while Persistence-Based predictions of the streamflow have approximately 0.90 correlations with the validation data.

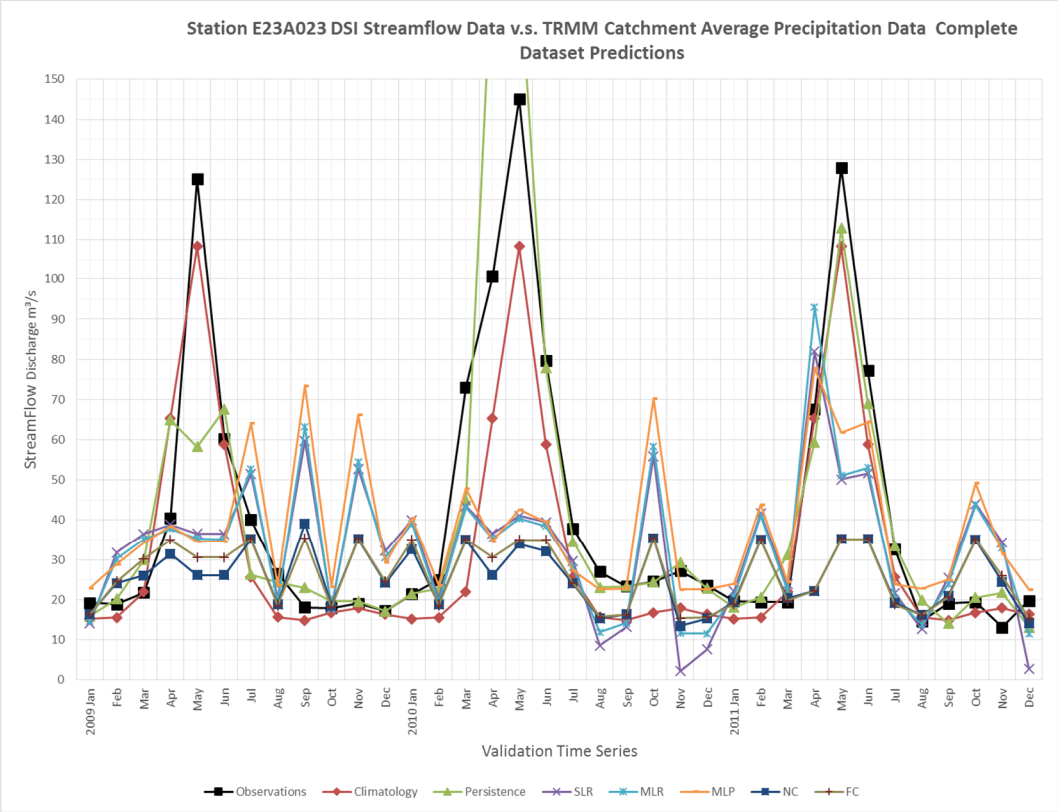


Figure 3.36 Station E23A023 DSI streamflow data v.s. TRMM catchment average precipitation data complete data set predictions

Table 3.38 Performance statistics of station E23A023 with TRMM catchment average data - complete data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,940	0,863	0,368	0,351	0,305	0,367	0,444
Standard Deviation of Observation	35,399	35,399	35,399	35,399	35,399	35,399	35,399
Standard Deviation of Errors	12,954	20,468	33,225	33,603	34,307	33,328	32,633
Mean Error	8,487	-0,661	8,420	7,705	3,363	15,699	14,954
RMSE	15,335	20,193	33,825	34,017	33,994	36,420	35,482
Mean of Observations	33,506	33,506	33,506	33,506	33,506	33,506	33,506

Similar results are obtained from the SLR, MLR, MLP and Copula models but showing under 0.40 correlation. These models could not identify a skillful relationship between the Rainfall and Runoff mainly due to lack of data. 9 years of monthly streamflow data and precipitation data could not train the models.

Additional predictions are made using the standardized anomaly components of data sets. Figure 3.37 shows the validation time series of the anomaly component predictions.

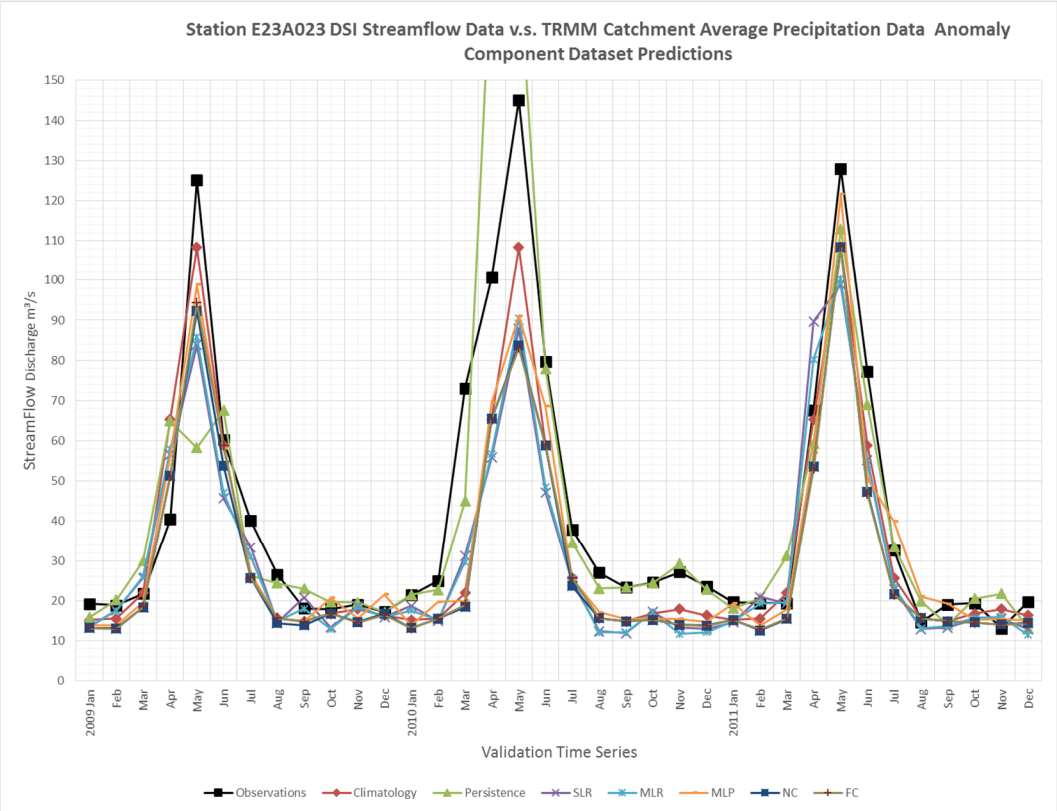


Figure 3.37 Station E23A023 DSI streamflow data v.s. TRMM catchment average precipitation data anomaly component data set predictions

Addition of Climatology component to the standardized anomaly component predictions showed above 0.90 correlations, these improved predictions are shoving heavy majority of the predictive skill and the relation between the precipitation and the streamflow data sets are due to the strong seasonality.

Correlations of predictions which are obtained from the SLR, MLR, MLP and Copula models increased from 0.30 to above 0.90 with decreasing RMSE values and mean error term. Performance statistics of validation time series are provided in Table 3.39.

Table 3.39 Performance statistics of station E23A023 with catchment average data – anomaly component data set predictions

Performance Statistics							
Model	**Climatology**	**Persistence**	**SLR**	**MLR**	**MLP**	**NC**	**FC**
Corelations	0,940	0,863	0,908	0,923	0,926	0,938	0,935
Standard Deviation of Observation	35,399	35,399	35,399	35,399	35,399	35,399	35,399
Standard Deviation of Errors	12,955	20,468	16,448	15,594	14,298	14,594	14,632
Mean Error	8,488	-0,660	10,658	10,907	8,739	12,176	11,907
RMSE	15,336	20,192	19,407	18,851	16,587	18,849	18,706
Mean of Observations	33,506	33,506	33,506	33,506	33,506	33,506	33,506

3.4. Discussion of Predictions

The summary of the correlations of complete data set predictions and anomaly component predictions with the streamflow validation time series are shown in Table 3.40 and Table 3.41.

Table 3.40 Summary of the correlations of complete data set predictions with the streamflow validation time series

	Overall Correlations of Complete Dataset Predictions with Validation Dataset							
	Station	Climatology	Persistence	SLR	MLR	MLP	NC	FC
Meteorology Data Predictions	E23A004	0,919	0,919	0,531	0,531	0,545	0,453	0,551
	E23A005	0,931	0,884	0,604	0,605	0,437	0,432	0,427
	E23A016	0,922	0,859	0,253	0,253	0,458	0,431	0,441
	E23A020	0,931	0,882	0,600	0,604	0,604	0,586	0,578
	E23A021	0,940	0,885	0,572	0,566	0,573	0,613	0,608
	E23A023	0,925	0,701	0,604	0,609	0,633	0,551	0,551
TRMM Single Pixel Data Predictions	E23A004	0,919	0,919	0,103	0,233	-0,123	0,111	0,111
	E23A005	0,931	0,884	0,228	0,268	0,238	0,139	0,152
	E23A016	0,922	0,859	0,225	0,138	0,157	0,040	0,160
	E23A020	0,931	0,882	0,214	0,258	0,131	0,106	0,135
	E23A021	0,940	0,885	0,209	0,223	0,118	0,167	0,229
	E23A023	0,925	0,701	0,106	0,220	0,171	0,065	0,093
TRMM Catchment Average Data Predictions	E23A004	0,919	0,919	0,198	0,250	-0,043	0,196	0,166
	E23A005	0,931	0,884	0,154	0,205	0,065	0,007	0,037
	E23A016	0,922	0,859	0,182	0,215	0,084	0,063	0,073
	E23A020	0,931	0,882	0,167	0,227	0,038	0,044	0,046
	E23A021	0,940	0,885	0,033	0,148	0,118	0,172	0,172
	E23A023	0,925	0,701	0,368	0,351	0,305	0,367	0,444

Table 3.41 Summary of the correlations of anomaly component data set predictions with the streamflow validation time series

	Overall Correlations of Anomaly Component Dataset Predictions with Validation Dataset							
	Station	Climatology	Persistence	SLR	MLR	MLP	NC	FC
Meteorology Data Predictions	E23A004	0,919	0,919	0,918	0,916	0,913	0,917	0,920
	E23A005	0,931	0,884	0,928	0,924	0,938	0,939	0,941
	E23A016	0,922	0,859	0,918	0,921	0,930	0,929	0,925
	E23A020	0,931	0,882	0,935	0,934	0,930	0,931	0,934
	E23A021	0,940	0,885	0,938	0,936	0,905	0,909	0,927
	E23A023	0,925	0,701	0,921	0,918	0,903	0,894	0,918
TRMM Single Pixel Data Predictions	E23A004	0,919	0,919	0,963	0,962	0,962	0,967	0,967
	E23A005	0,931	0,884	0,945	0,940	0,946	0,941	0,935
	E23A016	0,922	0,859	0,929	0,925	0,945	0,928	0,930
	E23A020	0,931	0,882	0,957	0,957	0,948	0,946	0,949
	E23A021	0,940	0,885	0,964	0,961	0,964	0,961	0,964
	E23A023	0,925	0,701	0,926	0,925	0,933	0,933	0,934
TRMM Catchment Average Data Predictions	E23A004	0,919	0,919	0,965	0,965	0,963	0,969	0,966
	E23A005	0,931	0,884	0,946	0,944	0,943	0,944	0,945
	E23A016	0,922	0,859	0,928	0,927	0,932	0,928	0,930
	E23A020	0,931	0,882	0,955	0,953	0,942	0,952	0,951
	E23A021	0,940	0,885	0,965	0,964	0,962	0,961	0,964
	E23A023	0,925	0,701	0,908	0,923	0,926	0,938	0,935

Overall results show that the best predictions are obtained from the climatology-based predictions of the stations for the complete data sets while persistence-based predictions have approximately 0.90 correlations with the observations. However, predictions which are using standardized anomaly component with different models are improved when long-term climatology values added as it is expected. These improved predictions indicate that the relation between the precipitation and the streamflow data sets are due to the strong seasonality. The results that the perfect knowledge of seasonality of the streamflow (i.e., climatology-based predictions) yield much better prediction that the prediction models (regression, ANN, and copula) may be primarily because the prediction models fail to accurately estimate the seasonality component of the streamflow while perfect knowledge of seasonality yield a much better prediction because the variability of the seasonality component of the streamflow data is much higher (~80%, Table 3.3) than the anomaly component (~20%).

Supporting these results, the anomaly predictions are made and then the perfect knowledge of the seasonality is added to yield a quantitative streamflow forecast, the skill (i.e., correlations) of the predictions increase to above 0.90. These results signify the choice of prediction model has secondary importance compared to accurate knowledge of seasonality of the streamflow data; perfect knowledge of the higher variability component of streamflow data (i.e., seasonality) primarily drives the skill of the predictions, while the inter annual deviations from these climatological components (i.e., anomalies) only marginally impact the overall accuracy of the complete time series predictions.

On the other hand, for drought related studies and the secondary energy generation (i.e., additional energy that can be generated in case additional hydropower potential is available in addition to the firm energy that is guaranteed to be supplied 95% of the time) the streamflow data anomalies (i.e., deviations from the long term means) are more important particularly for future energy generation planning studies (i.e., the future decisions about the quantity of water to be used for electricity generation compared to other usages like irrigation or the decision to generate more electricity

with excess water during high water seasons can be better done by using better anomaly predictions). This shows despite the seasonality component of the streamflow dominates the complete time series related statistics, anomaly component predictions have crucial effect for many applications. For such applications the skill and the choice of the prediction models (e.g., regression, ANN, copula) becomes very critical.

Correlations of real-time observations and predictions are improved from 0.20 to above 0.90 after long-term climatology values added. Added utility of climatology component show that the climatology component generates the main part of predictive skills. The main reason that streamflow climatology contains approximately 80% of total variability. These results showed that the main driver of the streamflow for the Coruh Basin is not the precipitation but snow-melting. Geological composition of the basin causes this condition. Despite, maximum streamflow values occur in May while maximum precipitation occurs in April and October for any given year in Coruh Basin. So, this condition shows that the main driver of streamflow of Coruh Basin is snow melting meaning that the temperature. Temperature data may provide better results than the precipitation data in streamflow predictions over Coruh Basin.

Among all the models that used for the calculations (SLR, MLR, MLP, NC, FC); MLR model provided the best results for the complete data set predictions while copula functions have the worst model performances. SLR and FC model performances are improved when the long-term climatology components added to the anomaly based predictions. Figure 3.38 and Figure 3.39 shows the averaged overall model correlation performances with different precipitation data sets and different components of the data sets.

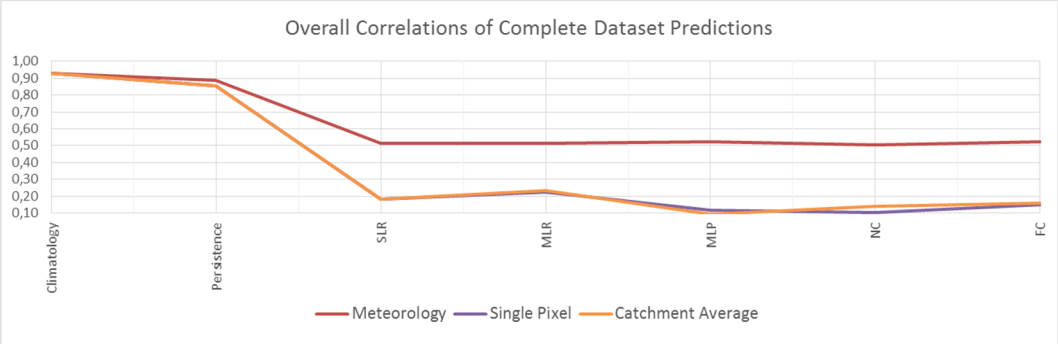


Figure 3.38 Overall average model correlations of complete data set predictions

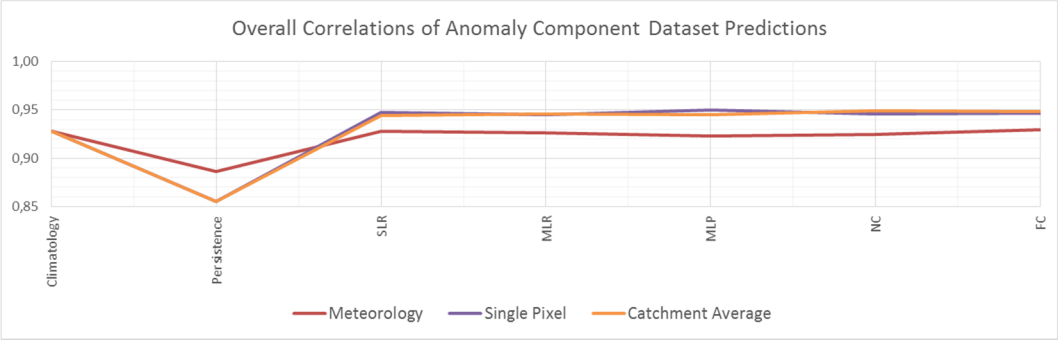


Figure 3.39 Overall average model correlations of anomaly component data set predictions

Considering the complete time series correlations between the TRMM single pixel estimates and station-based datasets are between 0.50 and 0.66 and correlations between the TRMM catchment average estimates and station-based datasets are between 0.57 and 0.76 (Table 3.1), the averaging operator applied in the vicinity of ground station increasing the correlations between the TRMM estimates and the station observations. On the other hand, this performance improvement is reduced with the increase of catchment area size; for larger catchments the TRMM catchment average precipitation estimates are less skillful than the single pixel estimates. These results imply TRMM individual pixel estimates may have random errors that can be reduced via spatial filtering techniques over relatively small catchments.

Model streamflow prediction performances in general vary; however, for TRMM single pixel vs TRMM catchment average precipitation-based prediction results only

marginally improve even though the catchment average TRMM precipitation estimates considerably better represent the station-based precipitation datasets than single pixel TRMM estimates (Table 3.40). This result shows TRMM precipitation errors are not low enough to be skillfully utilized in streamflow predictions over Coruh basin using the prediction methods investigated here. This signifies the local station-based precipitation observations are important in streamflow predictions as they have the ability to catch the local climatic variability.

Stations with smaller drainage area, like E23A021, has improved predictions utilizing the catchment average precipitation than the single pixel precipitation observations. For the predictions using TRMM catchment average precipitation data sets, the prediction skills and the errors are not very sensitive to the drainage area size; the prediction results are only marginally different. But results show that the importance of remote sensing data usage in rainfall-runoff modeling. Despite, local observations provided the best results, anomaly-based predictions which are obtained from the remote sensing data also provided strong results as with the local observations. Having short time series (Total 12 years – 9 years for training) cause some problems to train the models with the TRMM data.

Over certain basins, the monthly streamflow discharge rates are dominated by the seasonality of the snow melt and the ground water contribution: during the early spring months the streamflow discharge rates are heavily impacted by the snow melt contribution while the low discharge rate during summer months as a direct result of the reduced precipitation rates and ground water levels. Considering the total streamflow discharge rates time series signal can be decomposed into seasonality (i.e., low frequency and slowly varying component) and anomaly (i.e., high frequency and frequently varying component) components, the knowledge of the seasonality alone may give valuable predictive information related with the streamflow discharge rate itself. Given there is heavy snow melt and the precipitation seasonality over some basins in Turkey (i.e., the highest and the lowest flows occur during the spring and the summer months respectively, while the difference between the highest and the lowest values is great compared to the overall variability of the time series), the climatology

of the snow melt (primarily driven by temperature) and the precipitation heavily dominates the streamflow discharge rates. Accordingly, the climatology of the streamflow rates alone may serve a skillful streamflow discharge rate predictor for any year given the climatology of the historical data are available. This is why the climatology-based predictions proved to be very skillful compared to predictions made using other methods.

CHAPTER 4

SUMMARY and CONCLUSION

Streamflow data sets are predicted using precipitation data sets utilizing Simple Linear Regression (SLR), Multiple Linear Regression (MLR), Artificial Neural Network Model (MLP) and Copula (Normal Copula and Frank Copula) methods. The predictions are compared with benchmarks obtained using climatology- and persistence-based streamflow predictions, where climatology of the streamflow is heavily dominated by the seasonality of the snow melt (early spring) and the temperature over Coruh basin. To further investigate the source of the predictive skills of these methods, separate predictions are made using the standardized anomaly components of data sets (after climatology components are removed) and complete data sets (non-standardized data sets retaining both anomaly and climatology components). Separate monthly predictions are made utilizing station- (total 42 years observations) and remote sensing TRMM-based (total 12 years estimations) monthly precipitation data sets over six streamflow observation station locations over Coruh Basin in Turkey.

Results show that climatology benchmark-based streamflow predictions for the complete data sets have the highest correlations (>0.90) with the actual observations while persistence-based predictions are only marginally lower than these predictions. In general, MLP and Copula functions could not identify a skillful relationship between rainfall-runoff processes for both predictions using the complete data sets; however, the prediction skills are improved when anomaly predictions are made and the true seasonality (i.e., the component with highest variability) component is added to these anomaly predictions. These poor results using the prediction models are

primarily because of the poor streamflow seasonality component prediction of the predictions models. These results imply the significance of the estimation of true climatology in the streamflow estimates over Coruh river basin where the streamflow time series complete data set variability is heavily dominated by the climatology component rather than anomaly component.

Predictions using the ground station-based precipitation are more skillful than the predictions using the remote sensing-based precipitation data sets using the complete streamflow data sets. These results signify the importance of true precipitation data sets reflecting the local climatology of the study area. When the true seasonality is added to the anomaly-based streamflow predictions, the skills of the predictions become closer to the highest obtained climatology- and persistence-based streamflow predictions.

Climatology is the component with the highest skill (because both precipitation and streamflow data sets have very strong seasonality). So, in the absence of any precipitation or any other data set that can be used as predictor, the use of climatology or the persistence-based predictions may provide skillful-enough estimates.

To make better predictions, if the true long-term seasonality of the streamflow data sets are known, then anomaly based predictions using regression, ANN, and Copula models can be implemented and then the true seasonality information can be added to these anomaly predictions.

Because there are only a few studies investigating Rainfall-Runoff relation on Coruh Basin this study has valuable results for the applications over the basin. Results are valuable for dams that are planning to be built for hydropower energy generation purposes. Knowing the amount of discharges that come in different time intervals to the reservoir area, results of this study can be used for the reservoir operations and energy optimization for the individual dams that located in the upper and middle stream of the basin. Minimization of the amount of the spilled water or maximization of the secondary energy generation particularly during wet seasons are topics of future energy generation plans; hence such studies may benefit from investigations presented

in this study that utilizes streamflow and precipitation information to predict streamflow rates.

In this study only precipitation data sets are used as predictors; as a follow up, future studies could also include temperature data sets as another predictor where temperature also show very strong seasonality and primarily drive the snow melt over Coruh basin.

Averaging the TRMM Pixels in order to obtain TRMM catchment average data may cause some loss of information. Alternatively, another study with ANN and Copula models can be made by not averaging the TRMM pixels but taking every pixel data as a separate input. For example, if there are n pixels that cover the sub-basin of ground streamflow observation station, then there will be n sets of precipitation data sets that can be used as input to predictive models (Meher, 2014).

Regulations of the streamflow over Coruh River Upper and Middle streams began in year of 2012. Another future study can be made in the context of investigation of the regulation effects on streamflow predictions over Coruh Basin, where the climatology-based predictions might still prove skillful.

REFERENCES

Afshar, H. M., Sorman, A. U., & Yilmaz, M. T. (2016). Conditional Copula-Based Spatial–Temporal Drought Characteristics Analysis—A Case Study over Turkey. Water, 8(10), 426.

Arnell, N., Liu, C., Compagnucci, R., Cunha, L. da, Hanaki, k., Howe, C., Mailu, G., Shiklomanov, I., Stakhiv, E., *Publication Manual: Hydrology and Water Resources*.

Arpe, K. (1991). The hydrological cycle in the ECMWF short range forecasts. Dynamics of Atmospheres and Oceans, 16(1-2), 33-59.

Asati, S.R., Rathore, S. S., Comparative Study of Streamflow Prediction Models, International Journal of Life Sciences Biotechnology and Pharma Research, Gondia, 2012.

ASCE Task Committee, & ASCE Task Committee. (2000). Artificial neural networks in hydrology. II: Hydrologic applications. Journal of Hydrologic Engineering, 5(2), 124-137.

Bakiş, R., & Göncü, S. (2015). Completion of missing data in rivers flow measurement: case study of zab river basin. Anadolu university journal of science and technology–a Applied Sciences and Engineering, 16(1), 63-79.

Besaw, L. E., Rizzo, D. M., Bierman, P. R., & Hackett, W. R. (2010). Advances in ungauged streamflow prediction using artificial neural networks. Journal of Hydrology, 386(1), 27-37.

Bezak, N., Šraj, M., & Mikoš, M. (2016). Copula-based IDF curves and empirical rainfall thresholds for flash floods and rainfall-induced landslides. Journal of Hydrology, 541, 272-284.

Bloomfield, P., Statistics – 810 Course Notes, NC State University, 2013.

Brikundavyi, S., Labib, R., Trung, H. T., Rouselle, J., *Performance of Neural Networks in Daily Streamflow Forecasting*, Journal of Hydrologic Engineering, Montreal, 2002.

Budu, K. (2013). Comparison of wavelet-based ANN and regression models for reservoir inflow forecasting. Journal of Hydrologic Engineering, 19(7), 1385-1400.

Cabus, P. (2008). River flow prediction through rainfall–runoff modelling with a probability-distributed model (PDM) in Flanders, Belgium. Agricultural Water Management, 95(7), 859-868.

Callan R. *The Essence of Neural Networks*, Prentice Hall Europe, London 1999.

Chiew, F., & McMahon, T. (1994). Application of the daily rainfall-runoff model MODHYDROLOG to 28 Australian catchments. Journal of Hydrology, 153(1-4), 383-416.

Scaillet, O., Charpentier, A., & Fermanian, J. D. (2007). The estimation of copulas: Theory and practice.

ÇİMEN, M., & SAPLIOĞLU, K. (2007). Strem flow forecasting by fuzzy logic method. Süleyman Demirel University, Department of Civil Engineering, Isparta.

Dawson, C. W., & Wilby, R. L. (2001). Hydrological modelling using artificial neural networks.

De Michele, C., & Salvadori, G. (2003). A generalized Pareto intensity-duration model of storm rainfall exploiting 2-copulas. Journal of Geophysical Research: Atmospheres, 108(D2).

Devia, G. K., Ganasri, B. P., & Dwarakish, G. S. (2015). A review on hydrological models. Aquatic Procedia, 4, 1001-1007.

Dralle, D. N., Karst, N. J., & Thompson, S. E. (2016). Dry season streamflow persistence in seasonal climates. Water Resources Research, 52(1), 90-107.

Durante, F., & Sempi, C. (2010). Copula theory: an introduction. In Copula theory and its applications (pp. 3-31). Springer Berlin Heidelberg.

Eichert, B. (Ed.). (1982). Methods of hydrological computations for water projects: a contribution to the International Hydrological Programme: report (No. 38). United Nations Educational.

Encyclopedia of Research Design, vol.1, Method of Least Squares, (2010).

Feng, L., & Hong, W. (2008). On hydrologic calculation using artificial neural networks. Applied Mathematics Letters, 21(5), 453-458.

Fung, D.S., "*Methods for the Estimation of Missing Values in TimeSeries*", A thesis Submitted to the Faculty of Communications, Health and Science Edith Cowan University Perth, Western Australia (2006).

Garbrecht, J. D. (2006). Comparison of three alternative ANN designs for monthly rainfall-runoff simulation. Journal of hydrologic engineering, 11(5), 502-505.

Garen, D. C. (1992). Improved techniques in regression-based streamflow volume forecasting. Journal of Water Resources Planning and Management, 118(6), 654-670.

Gümüş, V., Kavşut, M. E., & Yenigün, K. (2011). Yağış-Akış Đlişkisinin Modellenmesinde YSA Kullanımının Değerlendirilmesi: Orta Fırat Havzası Uygulaması. New World Sciences Academy, 6(1).

Kellagher, R., Report – SC030219, *Rainfall – Runoff Management for Developments,* Flood and Coastal Erosion Risk Management Research and Development Programme, Oxon, 2013.

Kentel, E. (2009, December). Future River Flow Estimations of Gerede-Ulusu River, Ankara, Turkey. In AGU Fall Meeting Abstracts.

KİŞİ, Ö. (2005). Daily river flow forecasting using artificial neural networks and auto-regressive models. Turkish Journal of Engineering and Environmental Sciences, 29(1), 9-20.

Kokkonen, T. S., & Jakeman, A. J. (2001). A comparison of metric and conceptual approaches in rainfall-runoff modeling and its implications. Water Resources Research, 37(9), 2345-2352.

Lee, T., & Salas, J. D. (2008). Using copulas for stochastic streamflow generation. In World Environmental and Water Resources Congress 2008: Ahupua'A (pp. 1-10).

Levine, S., *Book: Using Statistics, Cha-13 Simple Linear Regression*, USA, 2014.

Linares-Rodriguez, A., Lara-Fanego, V., Pozo-Vazquez, D., & Tovar-Pescador, J. (2015). One-day-ahead streamflow forecasting using artificial neural networks and a meteorological mesoscale model. Journal of Hydrologic Engineering, 20(9), 05015001.

Meher, J. (2014). Rainfall and runoff estimation using hydrological models and Ann techniques (Doctoral dissertation).

Michele, C. de, *Modeling the Statistical Dependence in Hydrology Using "Copulas",* DIIAR, Politecnico di Milano, Milan, 2006.

Minns, A. W., & Hall, M. J. (1996). Artificial neural networks as rainfall-runoff models. Hydrological sciences journal, 41(3), 399-417.

Quesada-Molina, J. J., Rodriguez-Lallena, J. A., & Ubeda-Flores, M. (2003). What are copulas. Monografıas del Semin. Matem. Garcıa de Galdeano, 27, 499-506.

Del Moral, P., & Hadjiconstantinou, N. G. (2010). An introduction to probabilistic methods with applications. ESAIM: Mathematical Modelling and Numerical Analysis, 44(5), 805-829.

Mutlu, E., Chaubey, I., Hexmoor, H., & Bajwa, S. G. (2008). Comparison of artificial neural network models for hydrologic predictions at multiple gauging stations in an agricultural watershed. Hydrological processes, 22(26), 5097-5106.

Nadarajah, S., Afuecheta, E., & Chan, S. A Compendium of Copulas, 2009.

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

Ozturk, H. U. (2005). Discharge predictions using ann in sloping rectangular channels with free overfall (Doctoral dissertation, MSc Thesis, The Graduate School of Natural and Applied Sciences of Middle East Technical University (METU), Ankara).

Post, D. A., & Jakeman, A. J. (1999). Predicting the daily streamflow of ungauged catchments in SE Australia by regionalising the parameters of a lumped conceptual rainfall-runoff model. Ecological Modelling, 123(2), 91-104.

Radaideh, M. K. (1999). Statistical probability analysis for the estimation of hydrologic flooding.

Rasouli, K., Hsieh, W. W., & Cannon, A. J. (2012). Daily streamflow forecasting by machine learning methods with weather and climate inputs. Journal of Hydrology, 414, 284-293.

Abdul Rauf, U. (2014). A copula-based analysis of flood phenomena in Victoria, Australia.

Reingold E., Nightingale J., "History of Neural Networks" Department of Psychology, University of Toronto Mississauga, Canada 1999.

Salvadori, G., & De Michele, C. (2007). On the use of copulas in hydrology: theory and practice. Journal of Hydrologic Engineering, 12(4), 369-380.

Samaniego, L., Bárdossy, A., & Kumar, R. (2010). Streamflow prediction in ungauged catchments using copula-based dissimilarity measures. Water Resources Research, 46(2).

Sayama, T., Tachikawa, Y., Takara, K., & Ichikawa, Y. (2006). Distributed rainfall-runoff analysis in a flow regulated basin having multiple multi-purpose dams. IAHS PUBLICATION, 303, 371.

Sezer, U. (2009). Çoruh river development plan. In International Workshop on Transboundary Water Resources Management.

Shedden K., Statistics – 401 Course Notes, The University of Michigan, Fall 2016.

Smith, J., & Eli, R. N. (1995). Neural-network models of rainfall-runoff process. Journal of water resources planning and management, 121(6), 499-508.

Sodoudi, S., Noorian, A., Geb, M., & Reimer, E. (2010). Daily precipitation forecast of ECMWF verified over Iran. Theoretical and applied climatology, 99(1-2), 39-51.

Stanton, J. M. (2001). Galton, Pearson, and the peas: A brief history of linear regression for statistics instructors. Journal of Statistics Education, 9(3), 1-16.

Stufflebeam, R. (2008). Neurons, Synapses, Action Potentials, and Neurotransmission. Consortium on Cognitive Science Instruction (CCSI), Retrieved October 17, 2013.

Sugimoto, T., Bárdossy, A., Pegram, G. S. S., & Cullmann, J. (2015). Investigation of hydrological time series using copulas for detecting catchment characteristics and anthropogenic impacts. Hydrology & Earth System Sciences Discussions, 12(9).

Svensson, C., "*Seasonal river flow forecasts for the United Kingdom using persistence and historical analogs*", Hydrological Sciences Journal, 61:1, 19-35 (2016).

Tanu, M. M. (2012). Rainfall and Streamflow Variability in Ghana. State University of New York at Albany.

Van den Dool, H. M., & Nap, J. L. (1981). An explanation of persistence in monthly mean temperatures in the Netherlands. Tellus, 33(2), 123-131.

VanderKwaak, Joel E., and Keith Loague. "Hydrologic□response simulations for the R□5 catchment with a comprehensive physics□based model." Water resources research 37(4), 999-1013, 2001.

Voisin, N. (2010). A Medium Range Probabilistic Quantitative Hydrologic Forecast System for Global Application (Doctoral dissertation, University of Washington).

Wang, C., Chang, N. B., & Yeh, G. T. (2009). Copula-based flood frequency (COFF) analysis at the confluences of river systems. Hydrological processes, 23(10), 1471-1486.

Web-1 Regression Analysis (2017, August) Retrieved from:

https://en.wikipedia.org

Web-2 Climatologies and Standardized Anomalies (2017, August). Retrieved from: http://iridl.ldeo.columbia.edu

Web-3 Persistence Method: Today Equals Tomorrow (2017, August). Retrieved from: ww2010.atmos.uiuc.edu

Web-4 Information over Coruh Basin (2017, August). Retrieved from: www.ekopangea.com

Web-5 Information over TRMM Data (2017, August). Retrieved from: https://pmm.nasa.gov

Wilks, D. S. (2011). Statistical methods in the atmospheric sciences (Vol. 100). Academic press.

Williamson, R. A., Hertzfeld, H. R., & Cordes, J. (2002). The socio-economic value of improved weather and climate information. Space Policy Institute, Washington, DC. Available at http://www. gwu. edu/~ spi/assets/docs/Socio-EconomicBenefitsFinalREPORT2. pdf [Verified 6 June 2015].

Yaman, N. (2014). Modelling precipitation data of certain regions for turkey via hidden markov models (Doctoral Dissertation, METU).

Yilmaz, M. Tugrul, and Timothy DelSole. *Predictability of seasonal precipitation using joint probabilities*. Journal of Hydrometeorology 11(2), 533-541, 2010.