JULY 2023 M.Sc. in Civil Engineering MUHAMMED VEDAT GÜN

REPUBLIC OF TÜRKİYE
GAZİANTEP UNIVERSITY
GRADUATE SCHOOL OF NATURAL & APPLIED SCIENCES

INVESTIGATION OF IMPACT OF CLIMATE CHANGE ON SEDIMENT YIELD AND STREAMFLOW IN A SUBBASIN OF EUPHRATES RIVER USING STATISTICAL DOWNSCALING AND GENETIC PROGRAMMING

M.Sc. THESIS
IN
CIVIL ENGINEERING

BY
MUHAMMED VEDAT GÜN
JULY 2023

INVESTIGATION OF IMPACT OF CLIMATE CHANGE ON SEDIMENT YIELD AND STREAMFLOW IN A SUBBASIN OF EUPHRATES RIVER USING STATISTICAL DOWNSCALING AND GENETIC PROGRAMMING

M.Sc. Thesis
in
Civil Engineering
Gaziantep University

Supervisor

Prof. Dr. Aytaç GÜVEN

by
Muhammed Vedat GÜN
July 2023

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.

Muhammed Vedat GÜN

ABSTRACT

INVESTIGATION OF IMPACT OF CLIMATE CHANGE ON SEDIMENT YIELD AND STREAMFLOW IN A SUBBASIN OF EUPHRATES RIVER USING STATISTICAL DOWNSCALING AND GENETIC PROGRAMMING

GÜN, Muhammed Vedat
M.Sc. in Civil Engineering
Supervisor: Prof. Dr. Aytaç GÜVEN
July 2023
155 pages

Studying the hydrological impacts of climate change is a critical area of research, as it directly affects regional water resources, ecosystem health, and human societies. Global Circulation Models (GCMs), while indispensable for understanding broad climatic patterns, exhibit limitations in representing local-scale climate features due to their coarse spatial resolution. Statistical downscaling techniques are employed to evaluate climate change's hydrological consequences at finer scales effectively. These approaches enhance GCM outputs, generating high-resolution climate projections that facilitate a more nuanced understanding and anticipation of climate change's specific implications for hydrological systems, thereby guiding the development of adaptation and mitigation strategies. This study focuses on predicting monthly discharge and sediment yield for a sub-basin of the Euphrates River. The analysis leverages scenarios derived from the Canadian Second Generation Earth System Model (CanESM2), coupled with a statistical downscaling method and artificial intelligence genetic programming models. Watershed delineation was conducted using the Watershed Modeling System (WMS), followed by the application of Gene Expression Programming (GEP) and Linear Genetic Programming (LGP) for the statistical downscaling process. The results were compared with linear regression outcomes. The machine-coded LGP model yielded the best correlation (R), lowest Mean Squared Error (MSE) and Mean Absolute Error (MAE) values for both discharge and sediment yield during training and testing periods. Regarding future projections under RCP 2.6, 4.5, and 8.5 scenarios, while both models yielded optimal results under the RCP 8.5 scenario, the GEP model demonstrated superior performance when comparing predictions to the observed period of 2006 – 2020. Overall, the models indicate a mild decreasing trend for both discharge (Q) and sediment yield (Qs) from 2021 to 2100, providing valuable insights for future hydrological management and planning.

Keywords: Climate Change, GCM, Statistical Downscaling, AI, GEP, LGP, Discharge, Sediment, WMS

ÖZET

İKLİM DEĞİŞİKLİĞİNİN FIRAT NEHRİ ALT HAVZASINDAKİ SEDİMENT TAŞINIMI VE DEBİYE ETKİSİNİN İSTATİSTİKSEL ÖLÇEK KÜÇÜLTME VE GENETİK PROGRAMLAMA KULLANARAK ARAŞTIRILMASI

GÜN, Muhammed Vedat
Yüksek Lisans Tezi, İnşaat Mühendisliği
Danışman: Prof. Dr. Aytaç GÜVEN
Temmuz 2023
155 sayfa

İklim değişikliğinin hidrolojik etkilerini incelemek, bölgesel su kaynakları, ekosistem sağlığı ve insan toplumları üzerinde doğrudan etkisi olduğu için önemli bir araştırma alanıdır. Küresel Dolaşım Modelleri (GCM'ler), geniş iklimsel örüntüleri anlamak için vazgeçilmez olsa da, yerel ölçekteki iklim özelliklerini temsil etmede büyük uzamsal çözünürlüklerinden kaynaklanan sınırlamalar gösterirler. İklim değişikliğinin hidrolojik sonuçlarını daha küçük ölçeklerde etkili bir şekilde değerlendirmek için istatistiksel ölçek küçültme teknikleri kullanılır. Bu yaklaşımlar, GCM çıktılarını geliştirerek, hidrolojik sistemler için iklim değişikliğinin özel etkilerini daha hassas bir şekilde anlama ve öngörme imkanı sağlayan yüksek çözünürlüklü iklim projeksiyonları üretir ve böylece uyum ve hafifletme stratejilerinin geliştirilmesine rehberlik eder. Bu çalışma, Fırat Nehri'nin bir alt havzası için aylık akım ve sediment taşınımı verilerini tahmin etmeye odaklanmaktadır. Analiz, CanESM2 modeli ile elde edilen senaryolara dayanarak, istatistiksel ölçek küçültme yöntemi ve yapay zeka genetik programlama modelleri ile birleştirilmiştir. Havza tanımlaması, Watershed Modeling System (WMS) kullanılarak gerçekleştirilmiş ve istatistiksel ölçek küçültme için GEP ve LGP modelleri uygulanmıştır. Sonuçlar, daha sonra lineer regresyon sonuçlarıyla karşılaştırılmıştır. Makine kodlu LGP modeli, eğitim ve test süreçleri boyunca hem debi hem de sediment taşınımı için en iyi korelasyon (R), en düşük MSE ve MAE değerlerini vermiştir. RCP 2.6, 4.5 ve 8.5 senaryoları altındaki gelecek projeksiyonları açısından, her iki model de RCP 8.5 senaryosu altında en uygun sonuçları verirken, GEP modeli 2006 - 2020 arasındaki gözlenen dönem ile karşılaştırılan tahminlerde üstün performans göstermiştir. Genel olarak, modeller, 2021 ile 2100 arasındaki dönemde hem debi (Q) hem de sediment taşınımı (Qs) için hafif bir düşüş eğilimi göstermektedir ve bu da gelecekteki hidrolojik yönetim ve planlama için değerli içgörüler sunmaktadır.

Anahtar Kelimeler: İklim Değişikliği, GCM, İstatistiksel Ölçek Küçültme, Yapay Zeka, GEP, LGP, Akım, Sediment, WMS

''Dedicated to arzu & payiz''

ACKNOWLEDGEMENTS

I am sincerely grateful to Prof. Dr. Aytaç Güven, my supervisor, for his invaluable support and kindness throughout this thesis. His expertise shaped the quality of this work.

I would also like to express my appreciation to my colleague Research Assistant Abdulhadi Pala for his insightful discussions and constructive feedback.

Special thanks to my family, especially my dear sister and my lovely wife for their constant support.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

LIST OF SYMBOLS AND ABBREVIATIONS

GCM	Global Circulation Model
CanESM2	Canadian Second Generation Earth System Model
WMS	Watershed Modeling System
GEP	Gene Expression Programming
LGP	Linear Genetic Programming
LR	Linear Regression
SVR	Support Vector Regression
MLR	Multiple Linear Regression
Q	Discharge
Q_S	Sediment Yield
DEM	Digital Elevation Model
RCP	Representative Concentration Pathways
IPCC	Intergovernmental Panel on Climate Change
NASA	National Aeronautics and Space Administration
GHG	Greenhouse Gases
RCM	Regional Climate Model
R	Correlation Coefficient
MAE	Mean Absolute Error
MSE	Mean Square Error
RMSE	Root Mean Square Error
RRSE	Root Relative Squared Error
AIC	Akaike Information Criterion
SDM	Statistical Downscaling Method
TDD	Traditional Dynamical Downscaling
NN	Neural Network
ANN	Artificial Neural Network
ANFIS	Adaptive Neuro-Fuzzy Inference System
MLPNN	Multilayer Perceptron Neural Network
GRNN	General Regression Neural Network

NCEP	National Centers for Environmental Prediction
NCAR	National Center for Atmospheric Research
CMIP	Coupled Model Intercomparison Project
HaDCM2	Hadley Centre 2^{nd} Generation
HaDCM3	Hadley Centre 3^{rd} Generation
CGCM3	3rd form of the Coupled Global Climate Model
SWAT	Soil and Water Assessment Tool
SRES	Special Report on Emissions Scenarios
CCS	Carbon Capture and Storage
UNFCCC	United Nations Framework Convention on Climate Change
WATCH	Water and Global Change
FG	Fuzzy Genetic
SS	Stephens–Stewart
SRV	Stage Rating Curve
SSY	Suspended Sediment Yield
ET_0	Evapotranspiration
CPU	Central Processing Unit
AI	Artificial Intelligence

CHAPTER 1
INTRODUCTION

1.1 General

The history of scientific revelations regarding climate change can be traced back to the early 1800s, when the existence of ice ages and additional variations in paleoclimate were initially hypothesised, and the natural greenhouse effect was recognised for the first time. In 1824, distinguished French physicist Joseph Fourier explicated the Earth's innate "greenhouse effect." Subsequently, during the latter part of the 19th century, researchers posited that anthropogenic emissions of greenhouse gases held the potential to change the planet's energy equilibrium and climatic conditions. The greenhouse effect indicates to the process by which heat is retained around the Earth's surface through the action of greenhouse gases. These warmth-retaining substances can be analogised to a quilt encircling the Earth, maintaining its temperature higher than its absence. Greenhouse gases encompass methane, carbon dioxide, water vapour, and nitrous oxides. Global average temperatures have increased significantly due to the intensification of greenhouse gas emissions from human activities, such as deforestation, burning fossil fuels, and agriculture (NASA, 2021). According to the findings of the Intergovernmental Panel on Climate Change (IPCC) as outlined in their 2021 report, there has been an increase of approximately 1.2°C in the global mean temperature since the pre-industrial period (IPCC, 2021). This global warming phenomenon has far-reaching consequences on Earth's ecosystems and natural processes. Climate change has led to a multitude of concerning trends, including the alarming loss of approximately 28 trillion metric tons of ice between 1994 and 2017, which has contributed to a global sea level rise of 34.6 millimetres (Slater et al., 2021). The magnitude of these changes underscores the need for a comprehensive understanding of the various impacts of climate change.

Climate change negatively affects health through various pathways, and low-income countries are particularly vulnerable; adaptation requires public health strategies and improved surveillance, while mitigation through renewable energy use can reduce air pollution exposure and improve health (Haines et al., 2006).

A new socio-economic scenario, A2r, predicts that mitigating climate change could reduce the impacts on agricultural demand of water by 40%, potentially leading to annual cost reductions of around 10 billion US dollars by 2080 (Fischer et al., 2007).

The regional climate model RegCM4 predicts that extreme climate events in China will become more pronounced under the high emission RCP8.5 scenario, with droughts worsening in Southwest China and decreased extreme events under the RCP4.5 scenario, suggesting that reducing greenhouse gas emissions may mitigate climate change impacts (Ji & Kang, 2015).

As a reaction to the adversities presented by climate change, various mitigation strategies have been explored, including reducing greenhouse gas (GHG) emissions, enhancing carbon sinks, and developing renewable energy resources. Studies have shown the potential of carbon capture and storage (CCS) technologies to substantially mitigate carbon dioxide (CO2) emissions originating from power plants and industrial facilities (Leung et al., 2003). Additionally, research on afforestation and reforestation has demonstrated the effectiveness of these approaches in sequestering atmospheric carbon (Canadell & Raupach, 2008). Renewable energy sources, such as wind power and solar, have also been extensively studied as alternatives to fossil fuels, with the capacity to alleviate climate change through the reduction of GHG emissions (Jacobson & Delucchi, 2011).

As climate change continues to reshape the global environment, its implications for hydrology and river systems have emerged as vital components requiring attention.

A variety of studies have been dedicated to investigating the implications of climate change on hydrological systems across Europe. [e.g., (Andréasson et al., 2004; Caballero et al., 2007; Graham et al., 2007; Kleinn et al., 2005; Thodsen, 2007)]

(Hoang et al., 2016) assesses the hydrological impact of climate change on the Mekong River basin, revealing an intensifying hydrological cycle with increased annual and seasonal discharges, higher low flows, and more frequent high-flow events. The results emphasise the importance of adaptive water management and addressing hydrological extremes to mitigate climate change impacts and reduce associated risks.

Hydrological systems and their relationship with climate change are better understood using essential tools known as General Circulation Models (GCMs). GCMs represent a critical component in studying the Earth's climate system, as they provide valuable insights into the complex interactions between various components, including the atmosphere, oceans, and land surface. These models have evolved with continuous advancements in computational power, enhanced comprehension of the fundamental mechanisms, and refinements in model parameterisations. GCMs have gained a growing significance and serve as the primary foundation for conducting most impact studies (Ramirez-Villegas et al., 2013). Also, GCMs have been instrumental in simulating and projecting climate variability and change on a global scale, informing policy decisions pertaining to strategies aimed at both adapting to and mitigating climate change impacts.

The performance and reliability of GCMs have been the subject of extensive research and evaluation. Model intercomparison projects, such as the Coupled Model Intercomparison Project (CMIP), have facilitated the assessment of GCMs by comparing their simulations with historical observations and amongst each other (Taylor et al., 2012). These comparisons enable the identification of model strengths and weaknesses, informing further model development and improvement. In addition, the latest phase of the CMIP (CMIP6) has incorporated a range of new modelling experiments, addressing various aspects of climate change, including the role of aerosols, land-use changes, and the effects of different emission scenarios on future climate projections (Eyring et al., 2016).

Despite the advancements in GCMs, several challenges remain. One of the primary limitations is the coarse spatial resolution of GCMs, which often fails to capture small-scale processes and regional climate features adequately (Giorgi & Gutowski Jr, 2015). Regional climate models (RCMs) and downscaling techniques, such as statistical downscaling, have been developed to provide higher-resolution climate projections to address this issue.

Hydrological processes generally take place on a smaller spatial scale (Kundzewicz et al., 2007). The objective of downscaling is to reconcile the discrepancy between the scale of climate change projections and the requisite resolution for assessing their consequences. This process posits that macro-scale atmospheric conditions substantially affect micro-scale weather phenomena while disregarding potential feedback mechanisms from local to global scales. There are two prevailing strategies for downscaling. The first, dynamical downscaling, embeds a regional climate model (RCM) within the general circulation model (GCM) to represent atmospheric processes at an increased grid box resolution within a specified region of interest. The secondary approach, statistical downscaling, forms statistical connections between large-scale meteorological phenomena and observed weather patterns at the local-scale (Maraun et al., 2010). Reviewing the recent developments and applications of statistical downscaling models in Scandinavia, (Hanssen-Bauer et al., 2005) focused on linear techniques for predicting local temperature and precipitation using large-scale temperature fields and atmospheric circulation indices. The downscaled temperature and precipitation scenarios for the 21st century show increased warming and precipitation in most of Scandinavia, with more significant warming rates in winter and varying projections for summer precipitation across different regions.

In the extent of climate change impacts, the statistical downscaling technique has been applied to create future climate data projections and to perform comparisons between various models (Guven & Pala, 2022; Tofiq & Guven, 2014).

This study processed the statistical downscaling method using Gene Expression Programming (GEP) and machine-coded Linear Genetic Programming (LGP) models. The hydrological impacts of climate change at the regional level were examined, and future predictions of climatic data such as sediment yield and discharge were conducted. In this area, only for sediment yield (Dahl et al., 2018) and (Guven & Kişi, 2010) made future calculations using different methods. Only for discharge data, (Khoi et al., 2022) and (Tofiq & Guven, 2014) have done similar studies. For both sediment yield and discharge data (Wei et al., 2020) examined the effects of climate change between 1954 and 2018 without future predictions. (Kisi & Shiri, 2012) compared Gene Expression Programming (GEP), Adaptive Neuro-Fuzzy Inference System (ANFIS), and Artificial Neural Networks (ANNs) to estimate daily suspended sediment concentration in rivers. After determining the optimal combination of daily stream flow and rainfall data using ANNs, it was found that a model combining these inputs outperformed discharge-based models. Among all methods, GEP provided superior results, and within ANN models, the Levenberg–Marquardt algorithm was the most effective. This study has been rare in terms of establishing future projections for both sediment and discharge data and comparing different models in examining the climate change's effects on hydrology.

1.2 Scope and objectives of the study

This study investigates the impacts of climate change on a sub-basin of the Euphrates River using statistical downscaling to examine sediment yield and discharge across the years under various future GCM emission scenarios. GEP and LGP models are employed to predict monthly sediment yield and discharge data separately. To determine the best-performing models, model evaluation criteria are implemented. Finally, both models simulated future sediment yield and discharge projection and apposed with observed periods to designate the best-performing model and scenario.

The objectives of this research are as follows;

- Identification of large-scale GCM predictors that are most correlated for local sediment yield and discharge data.
- Evolving different statistical downscaling models, GEP and LGP and performing linear regression analyses for comparison
- Training and testing the models with the observed data in two different periods (calibration and validation) so that the genetic models give the best performance for future predictions
- Simulating future forecasts for both sediment yield and discharge data under different scenarios with models
- Comparing the models with the observed periods
- Examining how the sediment yield and discharge data give an impression of the future in monthly averages from past years to future years, in different periods, under different models and scenarios

1.3 Layout of thesis

This thesis contains six chapters. The content of the chapters is as follows:

Chapter One: General introduction and information about the study

Chapter Two: Literature review about the study

Chapter Three: Methodology of the study

Chapter Four: Case study

Chapter Five: Results and discussion

Chapter Six: Conclusion and recommendations.

CHAPTER 2
LITERATURE REVIEW

2.1 Climate Change

A long-term change in a region's average weather conditions is referred to as climate change. Over the past century, there has been an increase in the average global temperature by approximately 0.7°C (Trenberth et al., 2007), and by 2100, the Intergovernmental Panel on Climate Change (IPCC) estimates that it will have warmed by 1.4 to 5.8°C (Abdo et al., 2009). It is widely established that human-induced greenhouse gas emissions, rather than other natural sources, are the primary cause of current global warming (Solomon, 2007).

In relation to climate change, adaptation refers to the actions taken to minimise the adverse results of climate change and take advantage of any potential benefits. It involves a wide range of measures, including physical, technological, and policy-based approaches. A significant increase in weather-related risks to human settlements is expected due to climate change and global warming over the next century. Climate change mitigation has received a great deal of attention, but few actions have been taken to prepare for a future that many believe has already begun (Muller, 2007).

Türkiye, with its Mediterranean and snow-dominated climate, stands as one of the nations susceptible to the impacts of climate change (Yilmaz & Imteaz, 2014). The western and southern regions of Türkiye fall within the Mediterranean basin, an area acknowledged for its heightened vulnerability to global warming and forthcoming climate-related extremes (Sánchez et al., 2004).

Climate change forecasting refers to the systematic process of prognosticating forthcoming alterations in the Earth's climate, encompassing predictions related to alterations in sea level, temperature, atmospheric and oceanic circulation patterns, and precipitation. These forecasts are based on global climate models (GCMs) output and scenarios that assume different levels of greenhouse gas emissions and other factors that could influence the climate.

Scientists have attempted to comprehend the significance of climate change in the future scenario by generating various forecasting systems to investigate the concerns of global warming caused by climate change (Yerlikaya et al., 2020).

Forecasts of future climate change are used to inform policy decisions and to help governments, businesses, and individuals plan for and adjust to the climate change's effects.

Climate models are essential for studying the climate system's response to various factors, making seasonal to decadal-scale climate forecasts, and predicting future climate in the upcoming century and after.

2.2 Climate Change Impacts

Climate change impacts denote the repercussions of climate change on both the natural environment and human societies. These impacts can be physical, ecological, social, or economic. Some examples of climate change impacts include heightened occurrences of heatwaves and extreme weather events, escalating sea levels, shifting patterns of precipitation, as well as alterations in the amount and diversity of plants and animals. Climate change can also significantly impact human communities through the disruption of agriculture and food production, the spread of diseases, and the displacement of people from their homes.

The consequences of climate change are generally seen in two categories depending on whether they directly affect the economy or, more broadly, affect communities and people. The first one is called market impacts, such as energy consumption and agricultural production. The second one is non-market impacts (environment, health) (Jamet & Corfee-Morlot, 2009).

Governments and organisations around the world are taking steps to address climate change and its impacts. Many resources are being expended on studying the causes and effects of climate change (Houghton et al., 1990). It is possible to include initiatives to lower greenhouse gas emissions, such as transitioning to renewable energy sources and encouraging energy efficiency. It can also include measures to adapt to the impacts of climate change that are already being felt, such as by improving infrastructure to withstand extreme weather events better or by assisting communities affected by natural disasters. These steps are generally referred to as mitigation of climate change. Climate change mitigation allows society to move towards a more sustainable future (Kılkış et al., 2022).

Over the past two decades, Türkiye's climate policy has remained largely unchanged. With no mitigation commitments under UNFCCC, Türkiye's status as an Annex I party is peculiar (Turhan et al., 2016)

(Danandeh Mehr et al., 2020) analysed meteorological drought across Ankara using adjusted and downscaled outputs of three global climate models (GCMs).

As a result of global climatic change over the past several decades, (Yucel et al., 2015) examined whether snowmelt runoff for 15 streamflow stations in eastern Anatolia, Türkiye, has been consistent over the last several decades and future runoff changes in these basins.

Anticipated consequences of climate change on hydrological systems arise from modifications in precipitation, temperature, and reference evapotranspiration. These variables serve as fundamental inputs for the terrestrial component of the hydrological cycle (van Roosmalen et al., 2007).

Using dynamically downscaled data, (Fujihara et al., 2007) developed a method for inputting hydrologic simulations into the Seyhan River Basin in Türkiye. The study assessed the impact of climate change on the water resource systems of the basin using a scenario approach that considered the changes in water use. This approach takes into account the effects of climate change and alterations in water use; will enable the re-evaluation of upcoming irrigation projects and the construction of dams, as well as the reconsideration of current reservoir operation rules and the investigation of adaptation strategies for mitigating the effects of global warming.

2.3 Climate Change Impacts on Sediment yield

Climate change is causing significant alterations to various natural processes and ecosystems across the globe. One of the many impact areas is the alteration of sedimentation processes in water bodies. Several studies have investigated the impacts of climate change on sedimentation processes in river systems.

(Inman & Jenkins, 1999) examined the flow of water in the stream and sediment movement regarding the 20 most substantial streams that flow into the Pacific Ocean along the central and southern California coastline.

(Thodsen et al., 2008) modelled alterations in the transport of suspended sediment triggered by climate change, in five different scenarios, based on the projected changes in land use and cover for two Danish river catchments.

(Kisi & Shiri, 2012) used the Eel River near Dos Rios, California, USA as a case study to compare the effectiveness of different soft computing methods (ANNs, ANFIS and GEP) for predicting daily suspended sediment levels in rivers. The study began with employing an ANN model to select the most effective input combination from various combinations of daily stream flow and rainfall data. A quintuple-input ANN model combining precipitation-based and discharge-based inputs was then established, which performed better than all discharge-based models. When comparing the performance of ANN, ANFIS, and GEP models using this new merged input combination, it was found that the GEP model provided superior results. Among the three algorithms used for the ANN models - Levenberg–Marquardt, conjugate gradient, and gradient descent - the Levenberg–Marquardt algorithm was determined to be the most effective.

The objective of the investigation of (Bussi et al., 2014) was to utilise sediment proxy data from the bottom deposits of a reservoir to establish a sediment model. This sediment model aims to assess the climate change effects in a catchment area located in the Mediterranean region. The investigation presented a technique for creating a distributed sediment model by utilising proxy sediment information, such as sedimentation volume in reservoirs. If data on reservoir sedimentation is accessible, this methodology can be utilised to solve the problem of insufficient information on sediment erosion and transport for model calibration and validation. The broad applicability of this technique is due to the abundance of large reservoirs worldwide.

The objective of a study conducted by (Verma et al., 2015) was to examine the effects of climate change on hydrological processes in the Maumee River watershed located in the Lake Erie Basin. The Soil and Water Assessment Tool (SWAT) was used to model these processes, and the model was calibrated and validated using data from a 10-year period (1995-2005). Subsequently, the study employed temperature and precipitation data obtained from three distinct GCMs to project the future discharge, sediment, and nutrient loading in the watershed for two specific time periods: mid-century (2045-2055) and late-century (2089-2099). According to the findings of the study, the hydrological processes in the Maumee River watershed are vulnerable to anticipated climate fluctuations. Specifically, a future escalation of 2.9°C in mean annual temperatures, combined with a 3.2% reduction in annual precipitation, would cause an 8.5% decrease in water flow in the watershed during the mid-century period compared to the baseline period of 1995-2005. Downscaling projected data into a higher resolution to meet modelling requirements contributes to some extent of uncertainty in the assessment. Furthermore, the hydrological process modelling comes with uncertainties, such as the underestimation of daily pollutant peaks by the SWAT model used in this investigation. This uncertainty raises questions about the dependability of future projections.

(Praskievicz, 2016) employed downscaled climate modelling from RCMs to drive a basin-scale hydrologic model to examine the climate change's impacts on discharge and suspended-sediment transport for three snowmelt-dominated rivers in the interior Pacific Northwest. In this study, the author evaluated the potential effects on upcoming climate change on discharge and suspended-sediment load in three rivers using the SWAT basin-scale hydrologic model and downscaled climate forecasts. The analysis revealed that future climate change might result in a range of projected impacts, such as modifications in the annual pattern of river discharge, an increase in the intensity of the most substantial floods, and variations in suspended-sediment load influenced by differences in energy availability for transport and sediment availability during the winter and spring seasons. These alterations in hydrological mechanisms could have a big impact on the processes that control risks, water availability, water quality, river terrain, and habitation—all of which are crucial for managing rivers for ecological and socioeconomic purposes.

In the study of (Wagena et al., 2016), the impact of climate change on water availability and sediment transport was examined by analysing six downscaled and bias-corrected climate models from the CMIP5 database. This analysis was conducted for two future periods, specifically between 2041-2065 and 2075-2099. The assessment of the ensemble model mean, and discrepancy in climate change forecasting was accomplished by utilising watershed models to evaluate the impact of climate forcing. To mitigate the adverse effects of changes in discharge and sediment in both basins, it is crucial to consider the influence of climate change when designing irrigation systems and implementing flood control measures for optimal reservoir operation. This is because climate change is expected to cause an increase in sediment concentrations, especially during the early and late stages of the monsoon, and in some instances, a reduction during the highest point of the monsoon. Therefore, measures should be taken to reduce the adverse impacts of these changes.

(Azim et al., 2016) aimed to assess the impact of climate change on sediment yield in the Naran watershed of Pakistan. The study utilized observational data from 1961 to 2010 and future climate projections using HaDCM3 GCM predictors of SRES scenarios A2 and B2. Statistical downscaling techniques were applied to generate future precipitation and temperature time series for the periods 2011-2040 and 2041-2070. The downscaled data revealed mixed changes compared to observations. The study also simulated potential sediment yield for the future based on climate change, which indicated an expected increase in stream discharges during both the snowy and monsoon seasons.

The article by (Nkwasa et al., 2022) introduces an approach that utilises the SWAT+ model for forecasting and projecting sediment yield at a local level in areas where there is a lack of data using global datasets. The framework comprises four main components: (a) integrating topographical features from high/medium resolution digital elevation models, (b) incorporating crop phenology data, (c) implementing an areal limit to linearize sediment yield in big scale units, and (d) employing hydrological water balance adjustment. To examine the effectiveness of this framework, the researchers evaluated a model implementation with and without the structure under both the past and future climate estimations, using the Nile Basin as a test case.

2.4 Climate Change Impacts on Discharge

The effects of climate change on river discharge have long been a topic of interest in hydrological research. Early studies such as (Gleick, 1987) laid the groundwork by examining the potential impacts of global warming on water resources, highlighting the importance of understanding the spatial and temporal distribution of runoff and its implications for river discharge. In the study, a water balance model was developed and tested to evaluate the effects of climate change on the availability of water in

California's Sacramento Basin. The model, possessing the capability to accurately replicate the magnitude and temporal patterns of monthly and seasonal runoff as well as changes in soil moisture conditions, is utilized to assess the hydrological implications of diverse climate change scenarios. These findings hold significant ramifications for long-term water resource planning, agricultural water utilization, and industrial development. Building on this foundation, (Arnell, 1999) performed an evaluation of the climate change's effects on river flow regimes at a worldwide scale. This study investigates the effects of climate change on global hydrological systems and water resources using Hadley Centre climate simulations (HadCM2 and HadCM3) and a macro-scale hydrological model. The results indicate that mean annual runoff will increase in high latitudes, equatorial Africa, Asia, and Southeast Asia while falling in the mid-latitudes and most subtropical regions. The study also shows that climate change will affect the people living in countries with water stress, with varying results depending on the climate model used. Finally, the study covers concerns relating to the scope of analysis and formulation of water resource impact indices, emphasizing the wide variation in impact estimations.

As research progressed, (Milly et al., 2005) evaluated the long-term trends in global river discharge and identified the weight of anthropogenic climate change on these trends. The study demonstrates that an assemblage of 12 climate models successfully simulates regional sequence of twentieth-century streamflow changes. These models predict that by 2050, runoff will be substantially reduced in southern Africa, southern Europe, the Middle East, and mid-latitude western North America while greatly rising in eastern equatorial Africa, the La Plata basin, and high-latitude regions. Such variations in the supply of sustainable water could have a significant effect on local economies and ecosystems.

Subsequent research, such as the study by (Dai et al., 2009), investigated the relationship between climate change and river discharge, explicitly focusing on the impacts of changes in precipitation and evapotranspiration. A new dataset of historical monthly discharge for the world's 925 largest ocean-reaching rivers enables more reliable assessments of variation in continental freshwater streamflow. The study unveils that while precipitation is a dominant factor shaping discharge patterns, the direct anthropogenic influence on yearly river flow is relatively minimal in comparison to climatic drivers for most prominent rivers during the period of 1948-2004.

(Gudmundsson et al., 2012) matches nine large-scale hydrological models within the European Union Water and Global Change (WATCH) project framework to quantify uncertainties in capturing seasonal variability of geographically averaged yearly runoff percentiles in 426 small watersheds throughout Europe. The assemblage mean performs well in estimating geographically averaged time series of yearly low, mean, and high flows, but individual models exhibit a large spread in performance, implying the need for caution when using a single model for impact assessments.

(Hirabayashi et al., 2013) Used 11 climate models to project global flood risk for the end of the century, revealing increased flood frequency in regions like Southeast Asia, Peninsular India, eastern Africa, and the northern half of the Andes. It emphasizes the need for adaptation because, based on the level of warming, the world is becoming more vulnerable to flooding.

The study of (Lutz et al., 2014) employs a large-scale, high-resolution cryospheric-hydrological model to analyse the effects of climate change on upcoming water availability and characterize the upstream hydrological systems of large Asian rivers. The research concludes that despite significant differences between basins and tributaries, runoff is projected to increase until at least 2050 due to increased precipitation and accelerated melt. These conclusions have implications for climate change policies, emphasising the need to focus on handling seasonal changes in the availability of water.

(Cervi et al., 2018) examined the response of low-yield streams in the Italian Northern Apennines to climate change, as their flows closely resembles meteoric water recharge trends, making them sensitive to water shortages. A hydrological rainfall-runoff model was validated and used to generate baseline and future periods using weather data from five RCM-GCM groups. The results reveal no modifications in the average annual discharges, although seasonal discharges are predicted to rise in winter and spring while falling in summer and autumn. Climate change has a considerable impact on low-flow indices and droughts, with higher rates and length of droughts expected.

2.5 Downscaling

Downscaling is a process of converting information from a coarse spatial resolution to a fine spatial resolution. This can be done using statistical methods or physically-based models to generate high-resolution datasets from existing low-resolution datasets. Downscaling is widely used to develop high-resolution projections of future climate change and its impacts.

There are several downscaling methods, but it is unclear whether one yields the best accurate estimates of the time series of daily rainfall (Dibike & Coulibaly, 2005).

(Salathé Jr et al., 2007) addresses four distinct but interconnected ecological systems of the Pacific Northwest and the socioeconomic and/or political systems that support them. Because hydrologic procedures are crucial to climate impacts across all sectors, downscaling climate scenarios for hydrologic simulations serves as the foundation for statistical assessments. Numerous strategies have been developed, which are based on experimental methods to data from climate simulations. These methods rely on the correlation among the observed values and simulations conducted under comparable climatic conditions.

Employing five downscaling models and past precipitation data, (Chiew et al., 2010) evaluated the properties of the precipitation downscaled from three GCMs. The researchers evaluate the performance of the SIMHYD daily rainfall-runoff model by comparing the modelled daily runoff generated using the downscaled daily precipitation with the observed past runoff aspects. Additionally, they compare the variations in projected precipitation and runoff characteristics, as simulated by the model, for the period 2046-2065 to the baseline period of 1961-2000.

Two categories are frequently used to categorize downscaling techniques: statistical and dynamical.

2.5.1 Dynamical Downscaling

Dynamical downscaling is a method for generating high-resolution climate simulations using an RCM that is "nested" within a larger-scale GCM. Dynamical downscaling is considered a valuable method not only within the climate research field but also for related fields such as environmental and hydrological studies (Hong & Kanamitsu, 2014).

A primary focus was placed on identifying areas requiring further examination, such as the selection and sensitivity of the model domain and resolution, strategies for implementing lateral boundary conditions, and the internal variability of RCMs (Jacob & Podzun, 1997).

The preliminary and peripheral boundary conditions from the RCM are continually integrated with the output from the GCM; this is called Traditional Dynamical Downscaling (TDD) (Z. Xu et al., 2019). TDD has been extensively applied in numerous research over the past years, including those involving hydrology (Feng et al., 2011; Gao et al., 2011; Giorgi, Hostetler, et al., 1994) air quality and atmospheric chemistry (Tan et al., 2015; Varghese et al., 2011) and meteorology, (Giorgi, Shields Brodeur, et al., 1994; Leung et al., 2003; Y. Xu et al., 2006).

Late research suggested that correcting GCM bias could greatly enhance dynamical downscaling simulations (Bruyère et al., 2014; Z. Xu & Yang, 2012).

(Giorgi & Gutowski Jr, 2015) evaluated the difficulties and potential future developments associated with regional climate models (RCM) or dynamical downscaling endeavours.

Due to the enormous computing and workforce required, the ensemble downscaling simulations necessitate multi-institutional collaboration, which hinders the application of TDD in climate change and impact studies (Z. Xu et al., 2019).

2.5.2 Statistical Downscaling

In order to establish quantifiable correlations between predictors (large-scale atmospheric variables) and predictands (local surface variables), statistical downscaling methods were created (Fistikoglu & Okkan, 2011).

(Wilby & Wigley, 1997) conducts a comprehensive review of the current generation of downscaling techniques, specifically examining four key categories: regression-based methods, circulation-based approaches, stochastic weather generators, and limited-area climate models. The study subsequently presents the findings of an international experiment that sought to evaluate a variety of rainfall models utilised for downscaling comparatively. The results demonstrate that while circulation-based downscaling methods are effective in modelling the characteristics of daily precipitation as observed and generated by models, they are only capable of capturing a portion of the daily rainfall variability changes related to climate changes as derived from models. Finally, the study delves into the various ongoing challenges that must be addressed in the advancement of climate downscaling.

The use of statistical downscaling methods (SDMs) in coupled atmosphere-ocean GCMs is common for improving future climate projections (Vrac et al., 2007). However, statistical downscaling models (SDMs) are inadequate in their ability to accurately simulate the small-scale variations in climate dynamics that are not resolved by GCMs.

Therefore, (Seguí et al., 2010) suggests an alternative validation process that involves a two-step generalised approach to compare the accuracy of any SDM against RCM simulations that are derived from the same GCM fields. The proposed validation process involves a comparison of past station-based measurements with models derived from a statistical model. The first step is to compare the measurements with the models derived from a statistical model that has been calibrated and driven by reanalysis fields. After that, the comparison of the same measurements, but this time driven by historical GCM fields. After evaluating the performance of the SDM under these two scenarios, the next step is to assess its ability to generate projections that are suitable with RCM simulations used as pseudo-observations under various future emissions scenarios.

(Tatli et al., 2004) examined a statistical downscaling technique for monthly total precipitation which is essential to local-scale climate variability analysis. The study implements an alternative to traditional statistical pre-processing techniques such as principal component analysis, principal factor analysis, and independent component analysis. Lastly, the study concludes by highlighting some potential areas for improvement and suggesting avenues for future research.

(Sennikovs & Bethers, 2009) aimed to offer useful meteorological input data for hydrological models to forecast future changes in river discharge.

(Tofiq & Guven, 2014) aimed to establish a novel method for forecasting the highest monthly discharge rates using linear genetic programming (LGP) through statistical downscaling. Various General Circulation Model scenarios were assessed to analyse the effect of global warming and climate change on flood discharge estimation. The study successfully downscaled the highest monthly discharges solely based on surface weather variables (NCEP and CGCM3) without relying on rainfall-runoff models, yielding satisfactory outcomes.

As an alternative to statistical downscaling for impact studies application of "Change factors" is one of the easy and popular method. (Arnell, 2003; Arnell & Reynard, 1996; Diaz-Nieto & Wilby, 2004; Eckhardt & Ulbrich, 2003; Pilling & Jones, 1999; Prudhomme et al., 2002). Initially, a benchmark climatology is settled for the specific location or area of attention. This benchmark may vary based on the intended application and could take the form of an indicative long-term average, such as the one from the period 1961-1990 or a substantive meteorological record, such as daily maximum temperatures. Subsequently, the study calculates the changes in the matching temperature variable for the GCM grid box nearest to the intended location or region (Wilby et al., 2004).

(Guven & Pala, 2022) used statistical downscaling to investigate the hydrological impacts of climate change at the local level. After selecting the most effective predictors, they used GMDH, GEP, SVM and LR techniques. They created future projections of average monthly precipitation. In order to assess the effectiveness of the models created, coefficient of determination (R^2), RMSE, MAE and AIC model evaluations are used. Finally, they concluded that the SVM model performs the best with the highest R^2 and lowest RMSE and MAE, but GEP has the lowest AIC.

(Guven et al., 2022) employs statistical downscaling techniques to assess the hydrological impacts of climate change on local hydropower capacity in Turkey's Göksu River basin. By analysing observed variables alongside statistically downscaled emission scenarios from the Canadian Second Generation Earth Systems Model (CanESM2) and employing Gene Expression Programming (GEP), the research evaluates the projected monthly precipitation and GCM projections under three Representative Concentration Pathways (RCP2.6, RCP4.5, and RCP8.5). The GEP model outperforms the Linear Regression (LR) model, and the findings shows an overall drop in monthly mean precipitation among the studied scenarios, with RCP8.5 providing the most suitable projections for the case study.

2.6 Genetic Programming

Genetic programming presents a method of traversing the domain of potential computer programs to identify a highly suitable individual program capable of addressing a diverse range of issues from various disciplines (Koza, 1992).

In the academic literature, tree, linear, and graphical representation formats are the most widely employed in Genetic Programming.

2.6.1 Linear Genetic Programming

Linear Genetic Programming (LGP) is a variant of Genetic Programming (GP) that uses a linear representation of computer programs rather than the tree-based representation used in traditional GP. (Nordin, 1994) determined that genetic operators in imperative programming act on a linear genome. It is essential to note that the term "linear GP" pertains to the linear structure of genomes, and as such, the LGP method does not necessarily generate linear models for non-linear systems. In fact, it is frequently utilised for evolving highly non-linear models.

Based on the nature of the instructions, two types of LGP systems are distinguished: (i) Machine–coded LGP and (ii) Interpreted LGP. This classification is determined by the ability of the central processing unit of a computer to execute the instructions directly. The machine–coded LGP makes use of commands that the CPU of a computer can carry out immediately, without using an interpreter (compiler), resulting in a significantly reduced runtime compared to the interpreted LGP. The latter type involves the execution of each instruction through a higher-level virtual machine written in a proficient programming language (Mehr, 2018).

Study of (Guven, 2009) employs the use of linear genetic programming (LGP) and two variations of Neural Networks (NNs) for forecasting the daily flow rates at a station on the Schuylkill River in the United States. The purpose of this study is to introduce LGP as a potential method for predicting river discharge time-series data and to assess the performance of the proposed LGP models against well-established Multilayer Perceptron Neural Network (MLPNN) and General Regression Neural Network (GRNN) techniques. To achieve this goal, various LGP and NN models were developed and evaluated using widely recognised performance measures. The conclusions indicate that LGP and NN methods exhibit a satisfactory level of agreement with observed data in predicting the daily discharge time series.

(Guven & Kisi, 2013) conducted an investigation to evaluate the accuracy of several methods, including LGP, FG, ANFIS, ANN, and SS, in developing models for pan evaporation. They used monthly climatic data from two stations in Turkey. The results of the above models' comparison showed that LGP performed the best in accuracy. In the next section of the research, LGP again gave the best results for monthly pan evaporation with high precision.

(Guven & Kişi, 2010) have conducted a study to assess the precision of the LGP method in determining suspended sediment levels. The authors used daily discharge and suspended sediment data from two stations in the USA to model the simulations. Compared to GEP and ANN models, LGP performed best, giving less RMSE and a better determination coefficient. They concluded that the utilisation of LGP has proven effective in the representation of daily suspended sediment levels.

The research of (Kisi & Guven, 2010) examines the precision of LGP in modelling daily reference evapotranspiration (ET0). The precision of LGP is contrasted against that of support vector regression (SVR), artificial neural network (ANN), and several empirical models, including the California irrigation management system Penman, Hargreaves, Ritchie, and Turc methods. The comparison results demonstrate that LGP is a better choice than SVR and ANN techniques.

(Huang et al., 2021) conducted a thorough examination of the design aspects of linear genetic programming regarding its application to dynamic job shop scheduling. Emphasis is placed on evaluating aspects such as the number of registers and the step size variation. The results show that linear genetic programming exhibits a strong performance compared to standard genetic programming and is capable of generating efficient dispatch regulations in a compact form.

Strengths of LGP

LGP can be applied to problems with multiple outputs by utilising multiple registers to produce the solution. The use of low-level language in LGP allows for direct execution on the computer processor, eliminating the need for an interpreter and resulting in faster program evolution.

Weaknesses of LGP

The number of registers used in LGP influences its performance, typically equal to the number of problem attributes. If the problem has limited attributes, additional registers may be needed to achieve a complex expression. The number of supplementary registers must match the desired expression complexity to avoid poor results (Oltean et al., 2009).

2.6.2 Gene Expression Programming

Gene Expression Programming (GEP) is a computational technique that employs a genetic algorithm to simulate the process of biological evolution in order to generate a computer program that can model a particular system or phenomenon. Through this method, a wide range of models can be generated, such as decision trees, neural networks, and polynomial constructs.

An investigation led by (Guven & Aytek, 2009) examined two locations along the Schuylkill River in the United States, where GEP was evaluated in relation to more conventional techniques, specifically, Multiple Linear Regression (MLR) and Stage Rating Curve (SRV) methods. Compared to the other two, GEP showed to be more accurate.

(Kisi & Guven, 2010a) proposes the utilisation of LGP for the purpose of estimating suspended sediment concentration. This study conducts a comparison of the LGP method with alternative techniques such as NN, ANFIS and rating curve models. When the results are compared, it has seen that LGP performs the best among them.

(Guven & Talu, 2010) made an investigation to explore the application of GEP in developing a model that establishes the association between discharge and SSY (suspended sediment yield). The GEP models put forth in this study are clear and uncomplicated, making them suitable even for untrained eyes. The GEP formulae proposed in this study provide a pragmatic approach for predicting sediment and achieving precise outcomes, thus promoting the incorporation of GEP in other areas of water engineering research.

(Traore & Guven, 2012) examines the capability of GEP in constructing models for predicting reference crop evapotranspiration (ET_0) utilising decadal climatic data from

Burkina Faso, a country in the Sahel region, and GEP has been found to be an effective tool for accurately computing evapotranspiration in the region.

(Fernando et al., 2012) presents an exploratory technique for aggregating outputs from multiple rainfall-runoff models using GEP to conduct symbolic regression. The technique of combining multiple models using GEP in this paper employs the simultaneous simulation of river flows from four conventional rainfall-runoff models to develop a set of consolidated estimates of river flows for four separate catchment areas. The GEP combination approach is capable of synthesising the outputs from individual models that may have lower accuracy, resulting in a more accurate forecast of river flows.

The study of (Seckin & Guven, 2012) examined the prediction of peak flood discharges for 543 ungauged locations throughout Türkiye by utilising linear genetic programming (LGP), GEP, and logistic regression (LR) models. The findings showed that the GEP and LGP models were superior to the LR model in reference to the correlation coefficient between the predicted and observed peak flow.

(Mehr, 2018) introduces a new hybrid method, comprising the integration of genetic algorithm and GEP, for estimating the monthly streamflow of an intermittent stream. In this study, both auto-correlation and partial auto-correlation functions were utilised on the streamflow records to identify the optimal predictors, specifically the number of lags, for the model. Additionally, an evolutionary search method of GEP was employed to further aid in this identification process. In order to assess the efficacy of the GEP-GA hybrid model, it was juxtaposed to various benchmark models, including traditional genetic programming, gene expression programming, linear regression and GEP-linear regression models, which were developed within the scope of this study. The study results indicated that the GEP-GA hybrid model had superior performance compared to all the benchmark models, which motivated its potential usage in practical applications.

CHAPTER 3
METHODOLOGY

3.1 General

The chapter outlines the methodology used in this study. Initially, the process of downscaling is described, and statistical downscaling is chosen as the technique due to its simplicity and low computational requirements. Next, large-scale weather factors, or predictors, are defined and obtained from CanESM2 data with three future emission scenarios, RCP 2.6, RCP 4.5 and RCP 8.5. These predictors are available for download from CGCM box grids. Then, LGP and GEP are explored as strategies for the calibration and validation processes. Lastly, evaluation criteria are specified for comparing the predicted and observed monthly discharge and precipitation levels, and the best-performing model is selected.

3.2 Downscaling

Downscaling is the process of taking data with a coarse spatial resolution and producing a dataset with a finer spatial resolution while preserving the essential features of the original dataset. Downscaling is commonly used in meteorology, climate modelling, and remote sensing to convert data into a format that is easier to work with or to make predictions about future conditions.

There are basically two types of downscaling. Dynamical and statistical downscaling. Dynamical downscaling is a technique used in climate science to obtain high-resolution climate information at regional scales. The technique involves using a high-resolution atmospheric model to simulate the behaviour of the atmosphere at smaller spatial scales and then using these simulations to provide regional climate information that can be used for various applications.

Statistical downscaling is a specific type that uses statistical methods to estimate the relationships between large-scale climate variables and smaller-scale variables that are of interest, such as precipitation, temperature, or wind speed. Statistical downscaling is often used to create high-resolution climate projections from low-resolution global climate models.

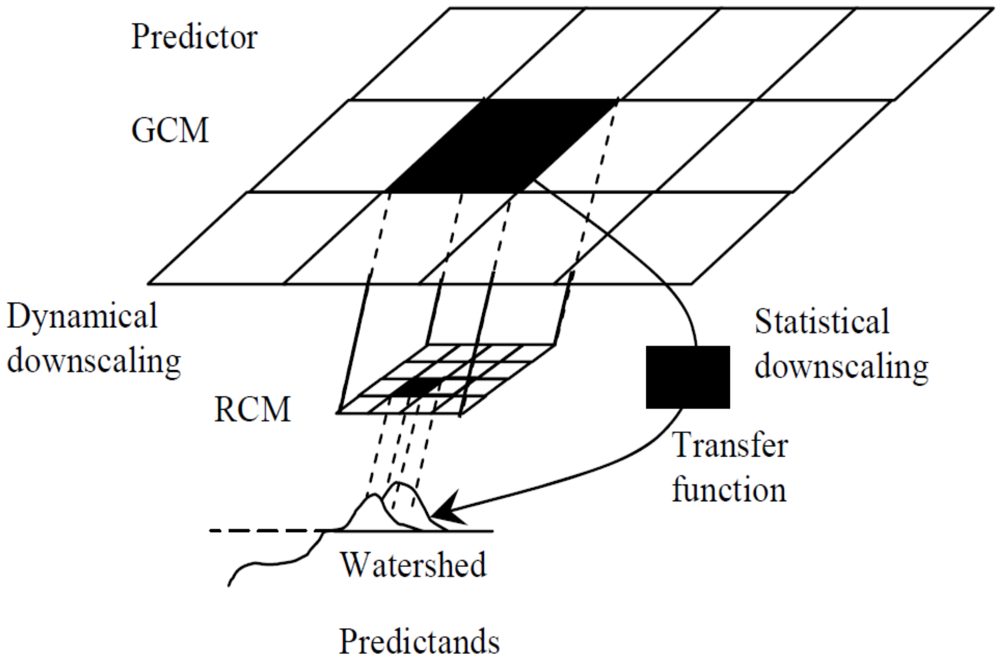

Figure 3.1 A schematic representation of downscaling process (Wetterhall, 2005).

Statistical downscaling methods are generally simpler to implement and require less computational resources than dynamical downscaling methods. This makes them more accessible to researchers who may not have access to high-performance computing resources. Table 3.1 shows the advantages and disadvantages of both statistical and dynamical downscaling.

Table 3.1 Summary of the relative advantages of statistical and dynamical downscaling methods (Fowler et al., 2007), adapted from (Wilby & Wigley, 1997).

	Statistical downscaling	Dynamical downscaling
Advantages	• Comparatively cheap and computationally efficient • Can provide point-scale climatic variables from GCM-scale output • Can be used to derive variables not available from RCMs • Easily transferable to other regions • Based on standard and accepted statistical procedures • Able to directly incorporate observations into method	• Produces responses based on physically consistent processes • Produces finer resolution information from GCM-scale output that can resolve atmospheric processes on a smaller scale
Disadvantages	• Require long and reliable observed historical data series for calibration • Dependent upon choice of predictors • Non-stationarity in the predictor-predictand relationship • Climate system feedbacks not included • Dependent on GCM boundary forcing; affected by biases in underlying GCM • Domain size, climatic region and season affects downscaling skill	• Computationally intensive • Limited number of scenario ensembles available • Strongly dependent on GCM boundary forcing

For the above reasons, statistical downscaling is used in this study. In statistical downscaling, a quantitative relationship is established between the input data, known as predictors, and the output data, referred to as predictands, which typically include climate variables such as discharge and sediment yield. The predictors are used as the basis for statistical downscaling techniques, allowing for the estimation of regional-scale climate variables based on the large-scale climate variables represented in the predictors. Figure 3.2 represents the flow diagram of the downscaling process in this thesis.

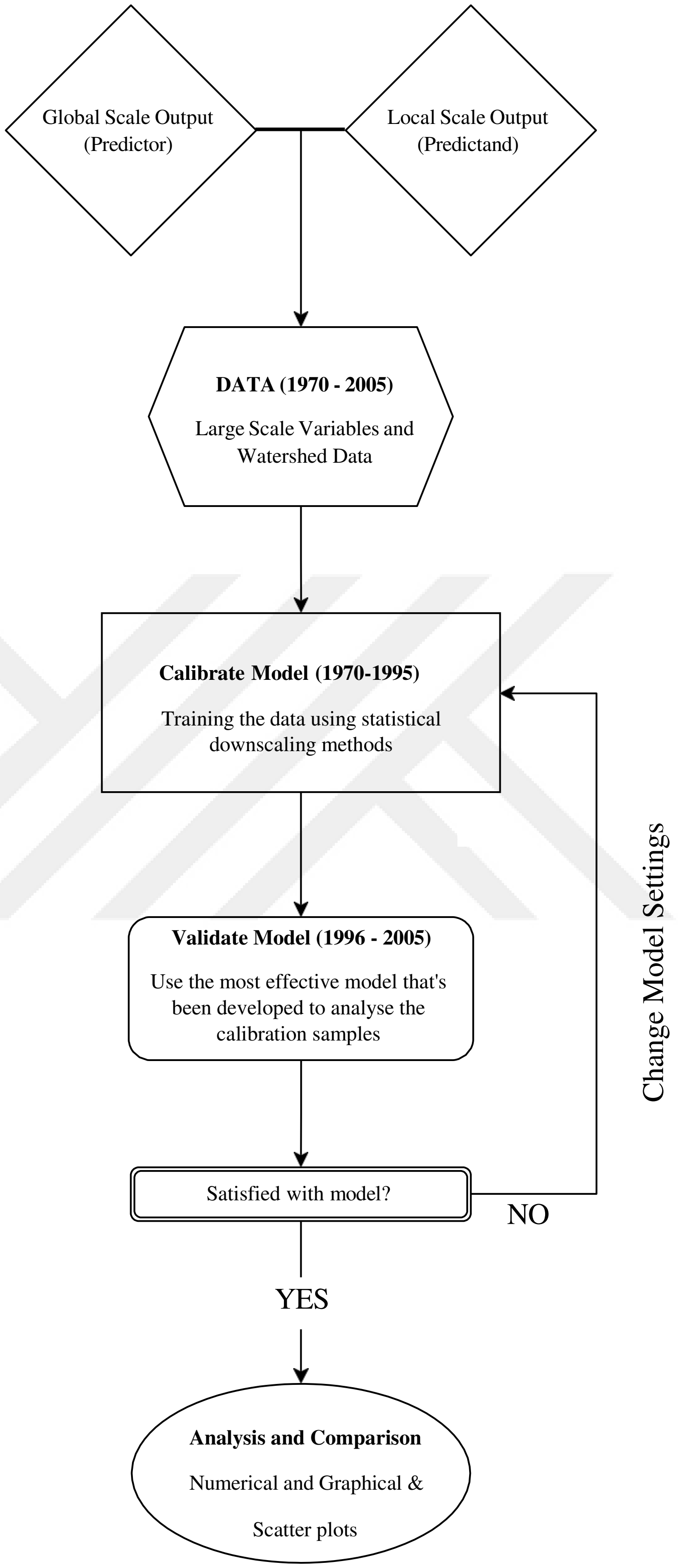

Figure 3.2 Flow diagram of downscaling process for this study

3.3 Gene Expression Programming (GEP)

Gene Expression Programming (GEP) is a computational approach introduced by Cândida Ferreira in 1999. It is a type of genetic programming that automatically creates computer programs or mathematical models from a given set of input and output data.

GEP is inspired by the concept of biological evolution, and it uses genetic operators such as mutation, crossover, and selection to evolve a population of computer programs. Each program is represented as a string of symbols composed of functions, constants, and variables. The string of symbols is decoded into a tree-like structure, representing the computer program.

GEP has been used effectively across a number of different issue categories, including classification, prediction, optimisation, and data analysis. GEP has the advantage of being able to handle non-linear, non-parametric, and non-stationary problems. It also has the ability to discover complex relationships between input and output data without requiring prior knowledge or assumptions about the problem.

The encoding of symbols in genes is a critical factor in facilitating GEP's efficient mutation of valid expressions. This encoding method is commonly referred to as the Karva Language, as proposed by Ferreira in 1996. By utilising the Karva Language, GEP is capable of rapidly generating a diverse range of functional expressions that can be effectively mutated and evaluated, leading to enhanced performance in complex optimisation problems. In GEP, a gene comprises a specified quantity of symbols encoded in the Karva language. The gene consists of two distinct segments, namely the head and the tail. The head component is responsible for encoding the functions utilized in the expression, whereas the tail serves as a supplementary reservoir of terminal symbols. These additional terminal symbols are employed when the head segment does not provide an adequate number of terminals for the function arguments. As such, while the head can contain functions, variables, and constants, the tail is limited to variables and constants (i.e. terminals). This structural arrangement provides a means to optimise GEP's expression capabilities by efficiently utilising both functional and terminal components within a given gene. In the context of GEP, a chromosome comprises one or multiple genes. The quantity of genes within a chromosome is a configurable parameter for the analysis. When multiple genes are present in a chromosome, they are combined using a linking function to form the final function. This linking function may be static or evolving, depending on the specific requirements of the problem. GEP can generate increasingly complex functions that exhibit greater accuracy in problem-solving applications by utilising a dynamic linking function. To illustrate all, Figure 3.3 illustrates an instance of a chromosome alongside the corresponding encoded tree structure.

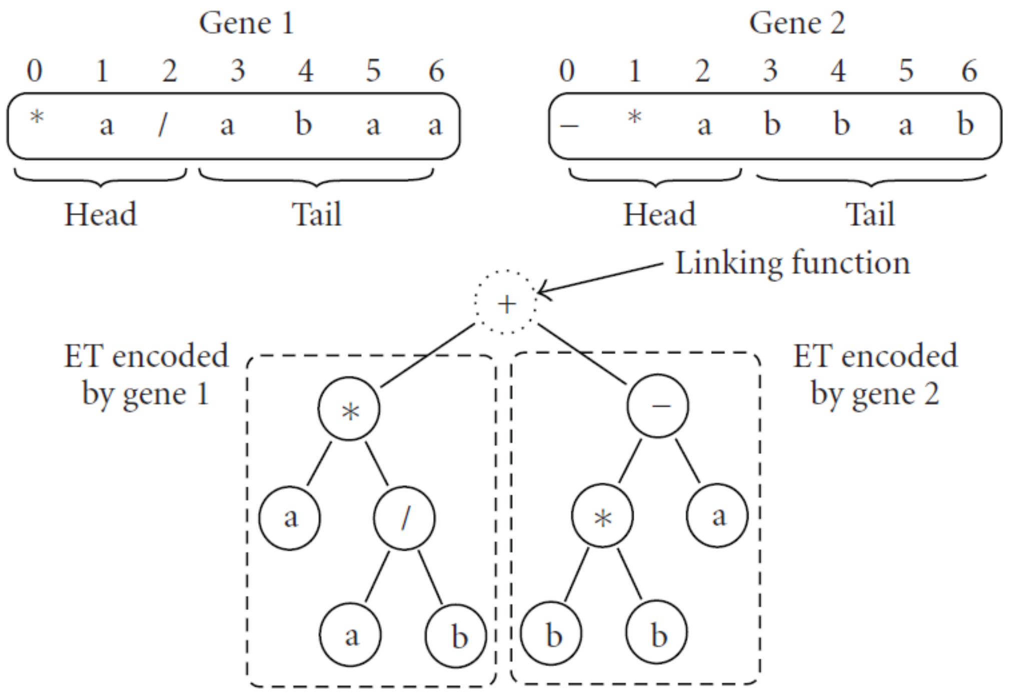

Figure 3.3 Chromosome with two genes: head size 3, tail size 4 (Browne & dos Santos, 2010).

GEP involves four essential steps for obtaining solutions to problems. This study emphasises that the first critical step is the identification of a set of functions to be used. To this end, a set of fundamental functions, including addition, subtraction, multiplication, division, power, and natural logarithm, were selected from the available functions in the GEP program. The second step is determining the chromosome structure, which involves specifying the number of genes and their size. The optimal approach in this study was to use four genes per chromosome for each GEP model. The third step involves selecting the linking function established in this study is addition. The final important step is to evaluate fitness using a specific measure. In this particular investigation, the MSE and Explained Variance (R^2) calculated from the training set were utilized as the fitness function. These four essential steps constitute the foundation of GEP, and their careful implementation is vital for obtaining effective solutions in problem-solving applications.

3.4 Linear Genetic Programming (LGP)

The concept of Genetic Programming (GP) originated as an evolutionary methodology designed to generate programs by employing expressions derived from the functional programming language LISP (Angeline, 1994). A new variant of GP called Linear Genetic Programming (LGP) has been introduced by Brameier and Banzhaf, which uses imperative programming language sequences of instructions to evolve genetic programs. LGP represents genetic programs as linear sequences of C instructions, and it has the benefit of removing noneffective codes (introns) before executing during fitness evaluation. Removing introns does not alter the behaviour of individuals in the population or during evolution, but it significantly speeds up the execution process. Linear Genetic Programming (LGP) is sometimes referred to as machine–coded LGP because it utilises C or C++ directly as programming languages, unlike tree-based Gene Expression Programming (GEP) (Guven & Kişi, 2010). Using machine–coded LGP as a modelling tool has several advantages. Firstly, it offers a high speed, allowing many runs to be conducted within realistic timeframes. Secondly, the method is designed to prevent overfitting and generate robust solutions. Finally, the models generated by LGP exhibit efficient execution times when invoked by integrated software. However, choosing the appropriate number of registers is essential when using LGP, as an inappropriate selection can lead to catastrophic issues in the evolved program (Kisi & Guven, 2010a).

LGP uses a variable-length sequence of basic C language instructions to represent each program. These instructions manipulate predefined sets of registers (r[i]) or constants (c) within the program, as described by (Brameier, 2004) and (Oltean & Grosan, 2003).

An example of the C code generated by the LGP program can be as follows:

$$r[0] -= v[0];$$

$$r[0] -= r[1];$$

$$r[1] += v[1];$$

$$r[2] = -v[3];$$

$$r[0] /= -0.23;$$

$$r[0] *= r[2];$$

In the programs generated by LGP, v[i] represents the input and output variables, while r[1] and r[2] are temporary computation variables. The value that remains in r[0] after program execution is the output of these temporary computation variables programs.

3.5 Linear Regression Analysis

The notion of linear regression was originally conceptualised by Sir Francis Galton in 1894. Linear regression involves the use of statistical techniques to determine and quantify the association among variables within a given dataset.

There are several important reasons for utilising a linear regression model. Providing a descriptive analysis that assesses the strength of the relationship between the dependent and predictor variables, enabling adjustment for the effects of confounding variables, helping to estimate the risk factors that are most significant in affecting the dependent variable, analysing how changes in the independent variable would affect the dependent variable, and quantifying new cases for prediction purposes (Kumari & Yadav, 2018).

An equation that exhibits linearity in its coefficients is known as a linear regression equation.

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_i X_i \tag{3.1}$$

In a non-linear regression equation, the coefficients can also be expressed as powers.

$$Y = \alpha X\beta_1{}^{\beta_1} X\beta_2{}^{\beta_2} \dots X\beta_i{}^{\beta_i} \quad (3.2)$$

The hydrological model is based on the assumption that there is a linear correlation between two variables. These models are categorised as simple linear regression models and generally follow one of the following general formats:

$$Y = \alpha + \beta X \quad (3.3)$$

Where Y is the dependent variable, X is the independent variable (predictor variable) and α, β are regression coefficients.

3.6 Model Evaluation Criteria

In this thesis, two different downscaling methods were used to forecast monthly average sediment yield and discharge data of the basin for both calibration (1970 – 1995) and validation (1996-2005) periods. In order to comprehend the performance of each model and ascertain which one performs the best, it is necessary to employ evaluation criteria for the proposed time frames.

The criteria used to evaluate models can be divided into two categories: graphical and quantitative evaluation. The ultimate goal is to select the model with the best performance based on measurements of both categories and then use this model to predict sediment and discharge under different emission scenarios in future periods.

3.6.1 Graphical Evaluation

The graphical evaluation comprises two components, specifically the scatter plot and the graph depicting the monthly average graphs of the basin.

3.6.1.1 Scatterplot

A scatter plot is a graph that displays the relationship between two continuous variables. It consists of a horizontal axis (x-axis) representing predicted variables and a vertical axis (y-axis) representing the observed variables. By examining the distribution and pattern of the points on the graph, we can infer the strength and direction of the relationship between the two variables.

3.6.1.2 Monthly Data Graphs

During the calibration and validation phases, graphs displaying monthly data were constructed using observed and predicted datasets for each month. In addition, to investigate the impact of climate change on the basin, a monthly data line for sediment and discharge was established as a reference point for future predictions.

3.6.1.3 Annual Average Data Graphs

The future predicted data obtained as output from the models are displayed for comparison with the observed data using annual average graphs. These data pertain to those models and scenarios that have been determined to perform well based on scatterplots.

3.6.2 Quantitative Evaluation

Some methods are used for this thesis to measure the performance of the statistical downscaling models. The selection of evaluation measurements depends on the particular problem being studied. In regression analysis, for instance, metrics such as Correlation Coefficient (R), Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE) are commonly employed to quantify the level of forecast errors and the effectiveness of the model.

3.6.2.1 Correlation Coefficient (R)

The correlation coefficient is a statistical construct that assists in determining a relationship between anticipated and actual outcomes derived from a statistical investigation. The computed correlation coefficient value elucidates the precision between the forecasted and actual values. The value of the correlation coefficient invariably resides between -1 and +1. A positive correlation coefficient indicates a congruous and identical relationship between the two variables. Conversely, a negative coefficient signifies a divergence between the two variables.

$$R = \frac{\sum_{i=1}^{M}(O_i - \bar{O}_i)\,(P_i - \bar{P}_i)}{\sqrt{\sum_{i=1}^{M}(O_i - \bar{O}_i)^2 \sum_{i=1}^{M}(P_i - \bar{P}_i)^2}} \tag{3.4}$$

Where O_i is the mean of the observed value, and P_i is the mean of the predicted value.

3.6.2.2 Mean Absolute Error (MAE)

Mean Absolute Error (MAE) serves as a prevalent assessment metric employed in regression models. It pertains to the arithmetic mean of the absolute magnitudes of individual prediction errors computed across the entire test set. Prediction errors are determined as the disparities between the predicted values and the true values for each instance. The MAE encapsulates the average magnitude of deviations between the model's predictions and the actual observations, thereby providing valuable insights into the accuracy and precision of the regression model's estimations.

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|O_i - P_i| \tag{3.5}$$

Where P_i is predicted data, O_i is observed data, N is the total number of data, and i is the subscript.

3.6.2.3 Root Mean Squared Error (RMSE)

The Root Mean Square Error (RMSE) is a statistical measure representing the standard deviation of the prediction errors, also known as residuals. Residuals capture the degree to which data points deviate from the regression line, while RMSE reflects the extent to which these residuals are dispersed. Put differently, RMSE provides information on the degree of data clustering around the line of best fit. RMSE is widely used in climatology, forecasting, and regression analysis to evaluate experimental findings.

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(O_i - P_i)^2} \tag{3.6}$$

Where P_i is predicted data, O_i is observed data, N is the total number of data, and i is the subscript.

CHAPTER 4
CASE STUDY

4.1 Study Area

The study area is a sub-basin of the Euphrates River located in the eastern part of Türkiye. This sub-basin, covering an area of 10247 square kilometres, is situated between the provinces of Erzincan and Erzurum in the east-west direction and between the provinces of Tunceli and Bayburt in the north-south direction. Figure 4.1 shows the exact location of the area of study.

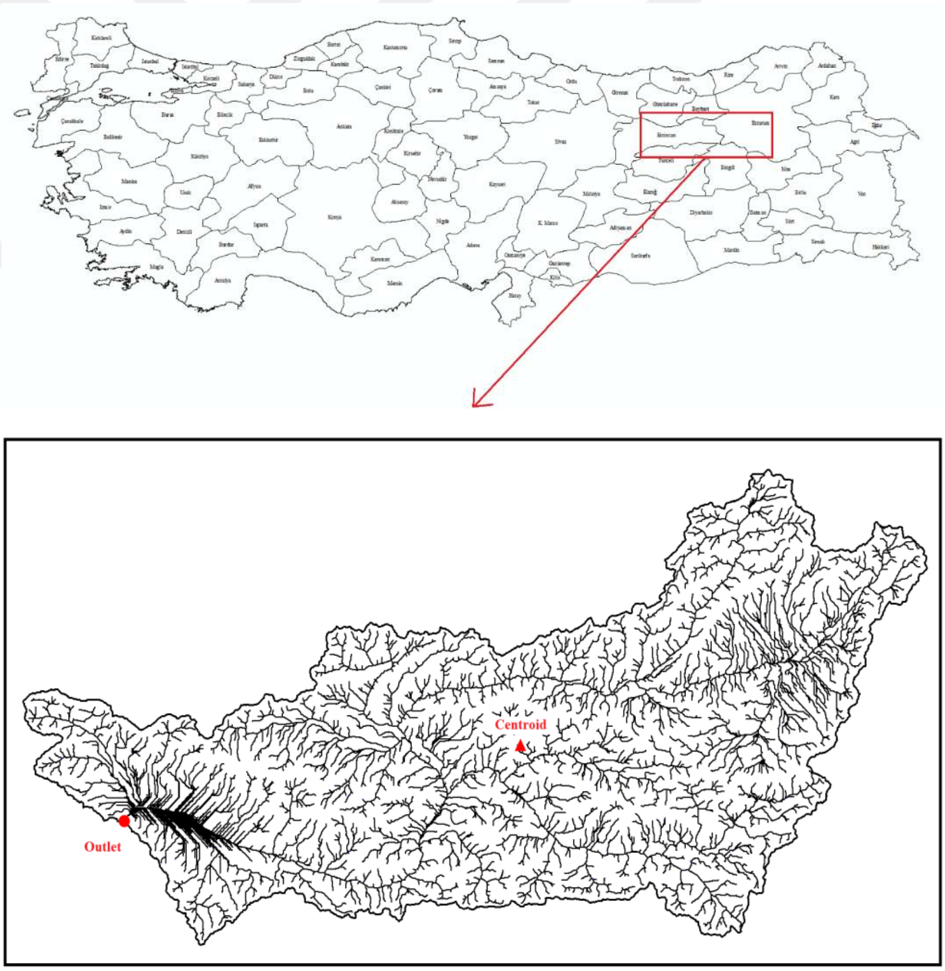

Figure 4.1 The location of the basin on the map of Türkiye (Watershed image from WMS)

The sediment yield & discharge station named Fırat N. Kemahboğazı, located on the Euphrates River, serves as the outlet point of the basin.

Information about station

Station No.	Station Name	Altitude (m)	Latitude (N)	Longitude (E)
E21A019	Fırat N. Kemahboğazı	1123	39º 40' 57"	39º 23' 35"

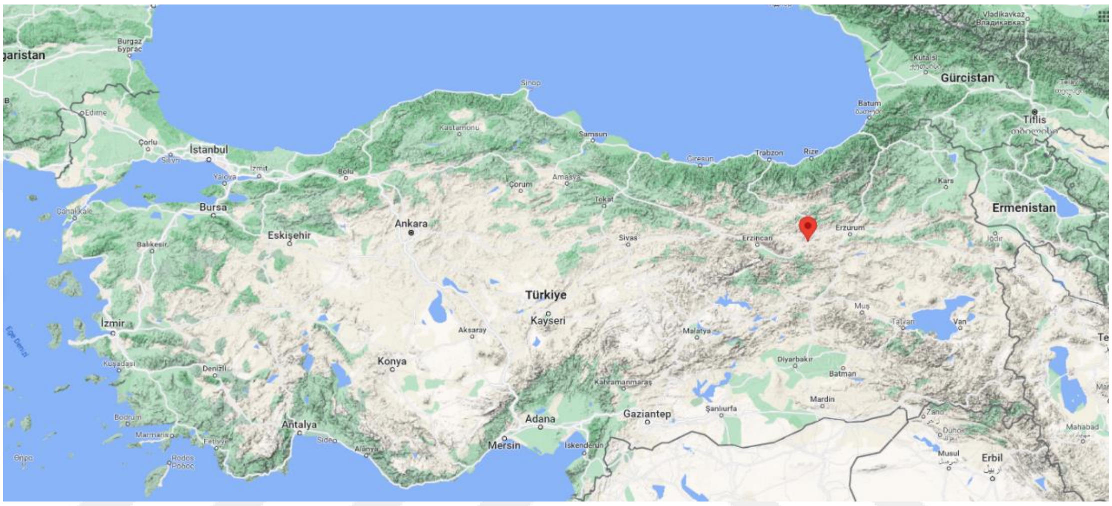

Figure 4.2 Location of the centroid of the basin

By choosing this station as the outlet point of the basin, watershed delineation is performed using Watershed Modeling System (WMS) software. WMS is an extensive hydrologic analysis software created by the Environmental Modeling Research Laboratory of Brigham Young University in partnership with the U.S. Army Corps of Engineers Waterways Experiment Station. It is currently being developed by Aquaveo. WMS provides advanced functionality for carrying out automatic watershed delineation and calculating critical watershed characteristics such as area, slope, and runoff distances. It functions as a visual interface for a range of hydraulic and hydrologic models, and its coordinate system management permits the visualisation and overlaying of data in real-world coordinates. Additionally, the software includes various display tools for examining terrain surfaces and the ability to export images for use in reports and presentations. Version 11.1 includes HEC-1, HEC-HMS, HEC-RAS, TR20, TR55, XP-SWMM and fourteen other models.

In the software, after choosing the projection and related area, WMS uses digital elevation model (DEM) to specify the boundary of the watershed. A Digital Elevation Model (DEM) portrays the Earth's topographical surface, comprising only the ground and excluding other objects such as trees or buildings.

WMS uses TOPAZ (Topographic Parameterization Program) to compute the flow direction and flow accumulation. The flow direction indicates the flow path from each cell in the DEM to the adjacent cell with the minimum elevation. The flow accumulation is calculated from the flow direction cell, which represents the count of cells that contribute to the drainage of each cell in the DEM.

WMS serves a couple of wizards in the software, allowing you to progress step by step in a user-friendly interface. In this study, for watershed delineation, the hydrological wizard is used. After choosing the projection and selecting the related area, the wizard takes you to the next step to choose the data you will use. DEM data is sufficient enough for delineation. After downloading data, wizard takes you to a step where you run TOPAZ and thus, creating flow accumulation cells and directions. In the next step, the outlet point is to be chosen. With the help of the real-world map that WMS lets you add, finding the corresponding outlet station is easy. Clicking delineate watershed button is the final step on the next page.

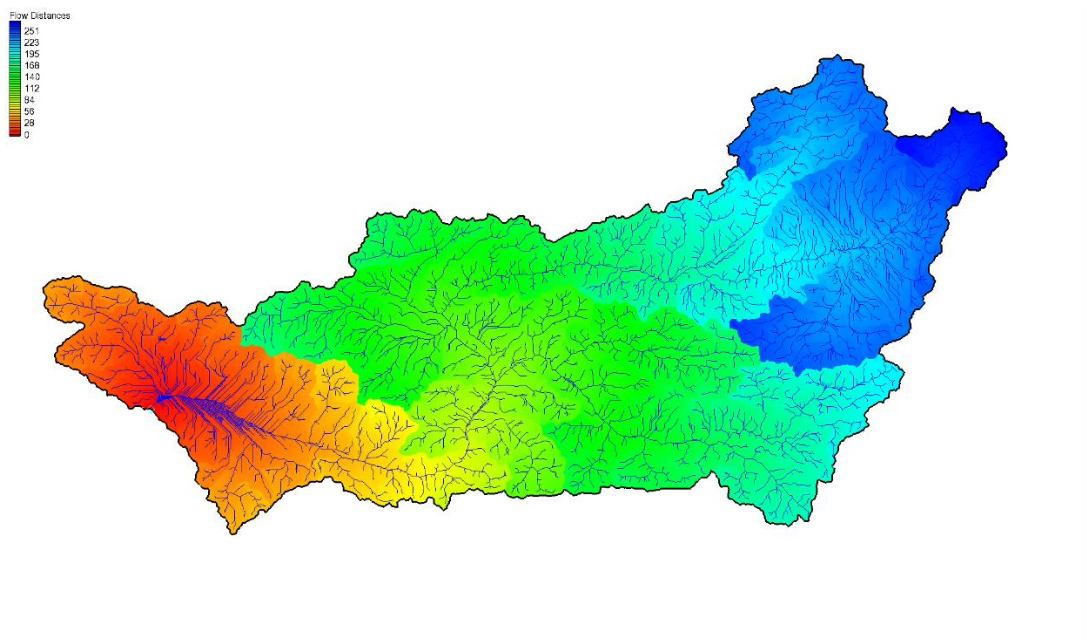

Figure 4.3 Flow distance contours and stream networks of the watershed

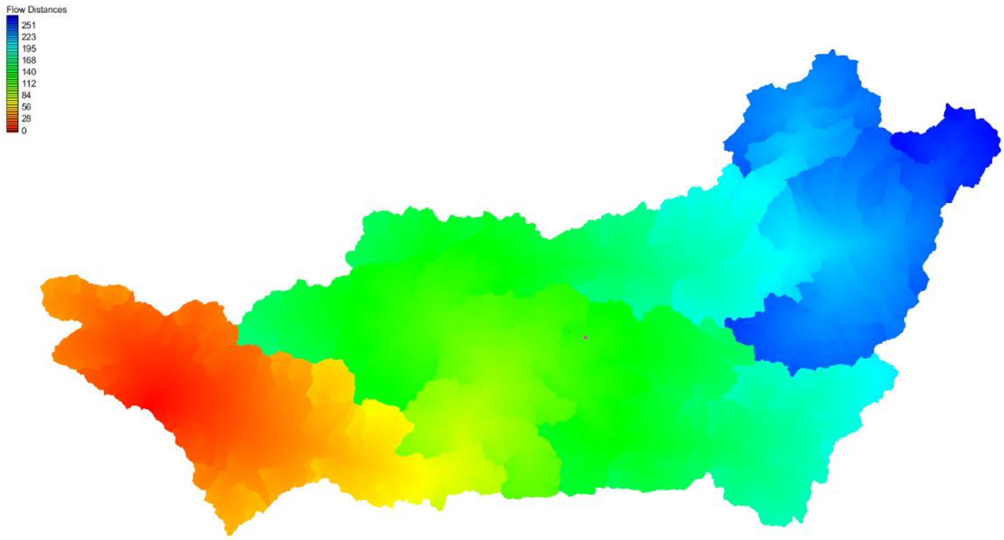

Figure 4.4 Flow distance contours of the watershed without stream networks

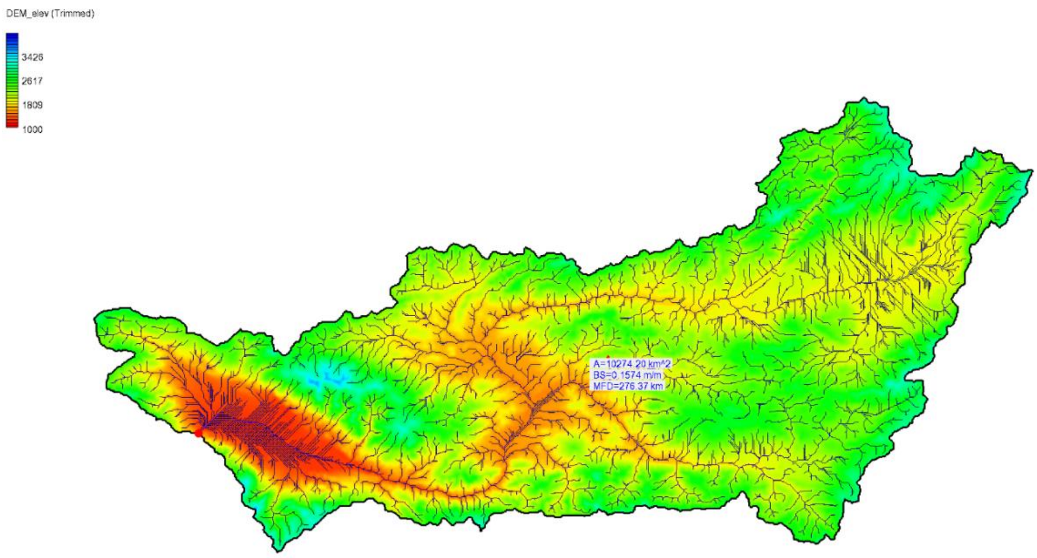

Figure 4.5 Digital Elevation Model (DEM) contours of the watershed

Table 4.1 Geomorphological characteristics of the basin

Basin Area (km2)	10274.20
Basin Slope (m/m)	0.1574
Average Overland Flow (km)	0.92
North Aspect	0.49
South Aspect	0.51
Basin Length (km)	193.91
Perimeter (km)	838.91
Shape Factor (mi^2/mi^2)	3.66
Sinuosity Factor (MSL/L)	1.42
Mean Basin Elevation (m)	2007.35
Max. Flow Distance (km)	276.37
Max. Flow Slope (m/m)	0.0070
Max. Stream Length (km)	274.54
Max. Stream Slope (m/m)	0.0057
Distance from Centroid to Stream (km)	0.99
Centroid Stream Distance (km)	127.03
Centroid Stream Slope (m/m)	0.0061

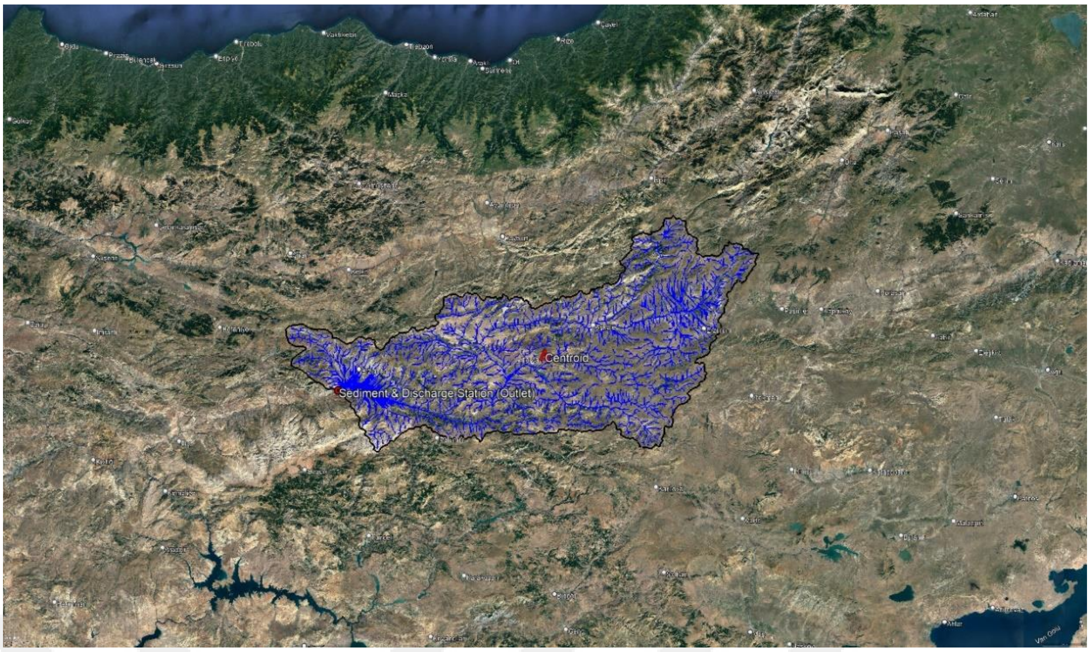

Figure 4.6 Basin located in Google Earth Map from WMS

4.2 Acquisition of Local Data

The data is obtained from the General Directorate of State Hydraulic Works (DSI). It is measured monthly between 1970 – 2020. The station is located in the southwest of Erzincan province, approximately 20 km from the city centre. The total annual charts of the station's discharge and sediment yield are shown below in Figure 4.7 and Figure 4.8. Observed monthly data of the station are listed in Appendix A.

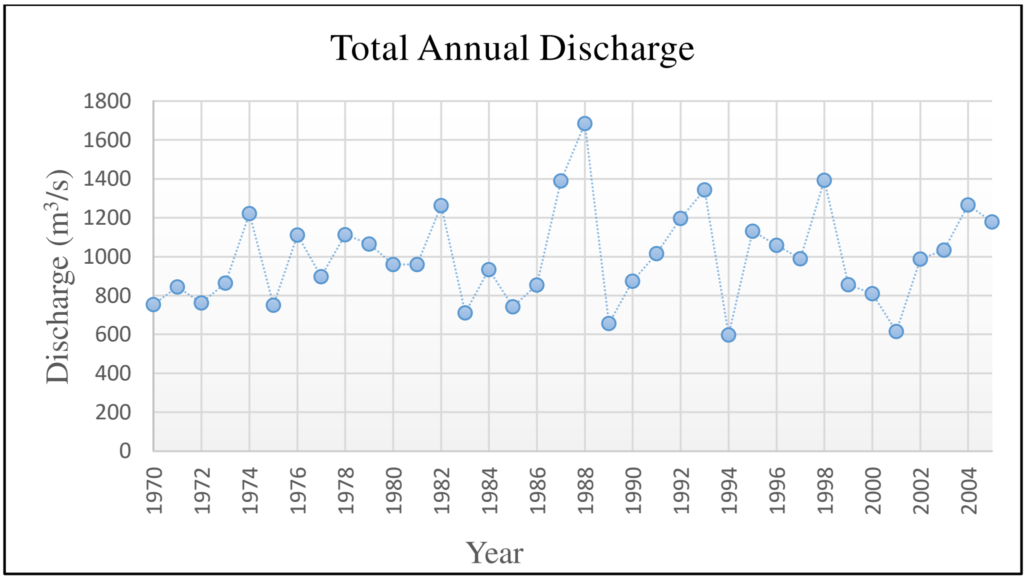

Figure 4.7 Total Annual Discharge of Fırat N. Kemahboğazı Station (1970 – 2005)

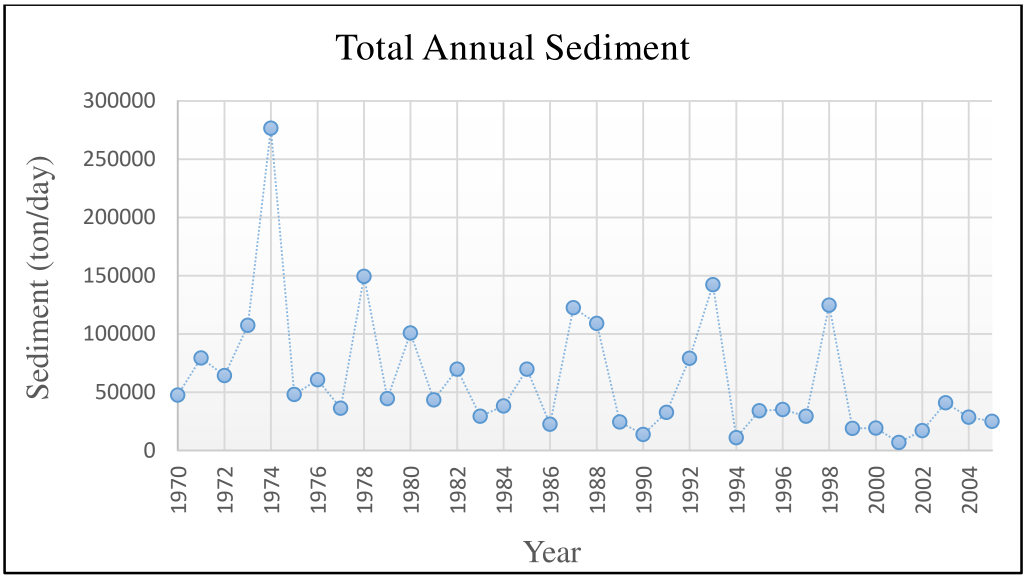

Figure 4.8 Total Annual Sediment of Fırat N. Kemahboğazı Station (1970 – 2005)

4.3 Monthly Sediment and Discharge Data (Predictand)

The monthly data taken from this station will be used as predictand for both sediment and discharge.

The monthly sediment and discharge data are divided in two. Calibration period (1970-1995) to develop statistical downscaling models, and the validation period (1996-2005) to examine the model's effectiveness and compare the outcomes of downscaling.

Figure 4.9 and Figure 4.10 illustrates the mean annual discharge and mean annual sediment yield for the station.

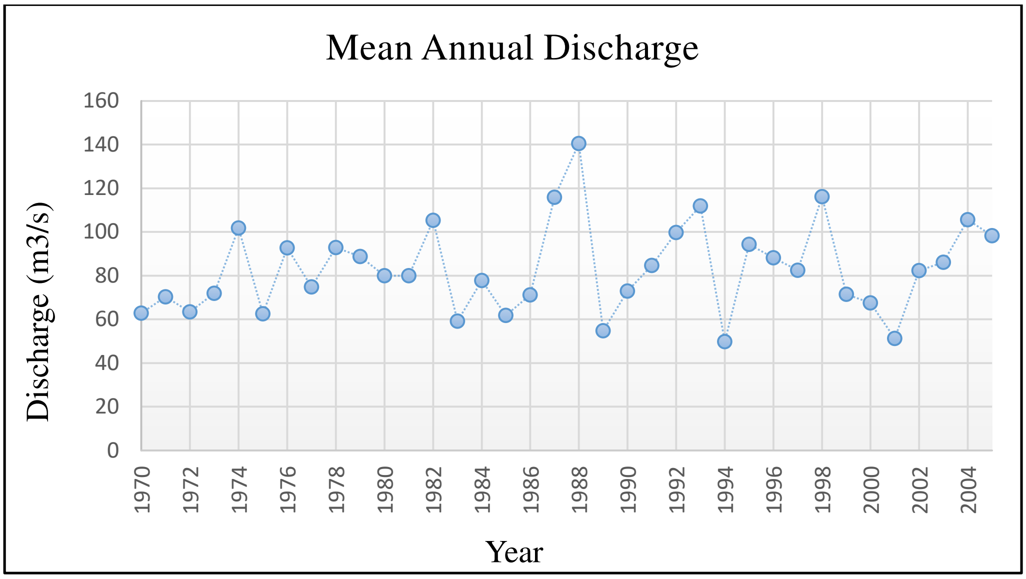

Figure 4.9 Mean Annual Discharge of Fırat N. Kemahboğazı Station (1970 – 2005)

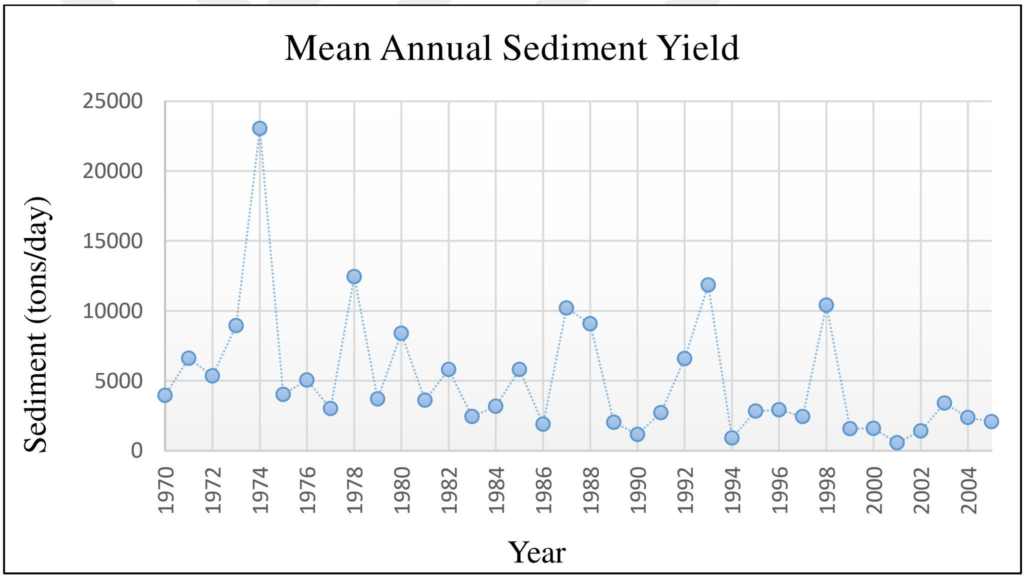

Figure 4.10 Mean Annual Sediment Yield of Fırat N. Kemahboğazı Station (1970 – 2005)

4.4 Large-Scale Variables (Predictors)

In order to develop spatial downscaling techniques, it is necessary to utilise predictor variables. These variables provide daily information about the large-scale state of the atmosphere. On the other hand, the predictand describes conditions at a local scale such as temperature, precipitation, discharge or sediment observed at a specific station.

Information about large-scale predictor variables has been prepared for a broad North American region. Near-observed data has been obtained from the NCEP Reanalysis dataset (Kalnay et al., 1996) for use during the calibration and validation procedures of the downscaling model. GCM data for baseline and climate scenario periods are derived from various GCM experiments.

The Canadian Centre for Climate Modelling and Analysis (CCCma) of Environment and Climate Change Canada created the fourth generation coupled global climate model known as CanESM2 (Second Generation Canadian Earth System Model). The Fifth Assessment Report (AR5) of the IPCC is Canada’s contribution.

In a 128x64 grid based on the T42 Gaussian grid, standardised daily data are extracted for each grid cell and stored in a single-column text file for the entire global area. This grid has virtually consistent horizontal resolution along the latitude of approximately 2.8125° and 2.8125° of longitude.

In files with the names BOX_iiiX_jjY, where iii stands for the longitudinal index and jj for the latitudinal index, predictions are kept for each grid cell. This organization of the predictors makes it simple to utilize them as input for statistical downscaling models.

In this study, the CanESM2 model is employed, and the data set is downloaded as a zip file from Canadian Climate Data and Scenarios (https://climate-scenarios.canada.ca/) by choosing the relevant grid cell. This can be done by manually putting the centroid of the watershed or simply selecting the cell in the rectangular map on the website, as shown in Figure 4.11.

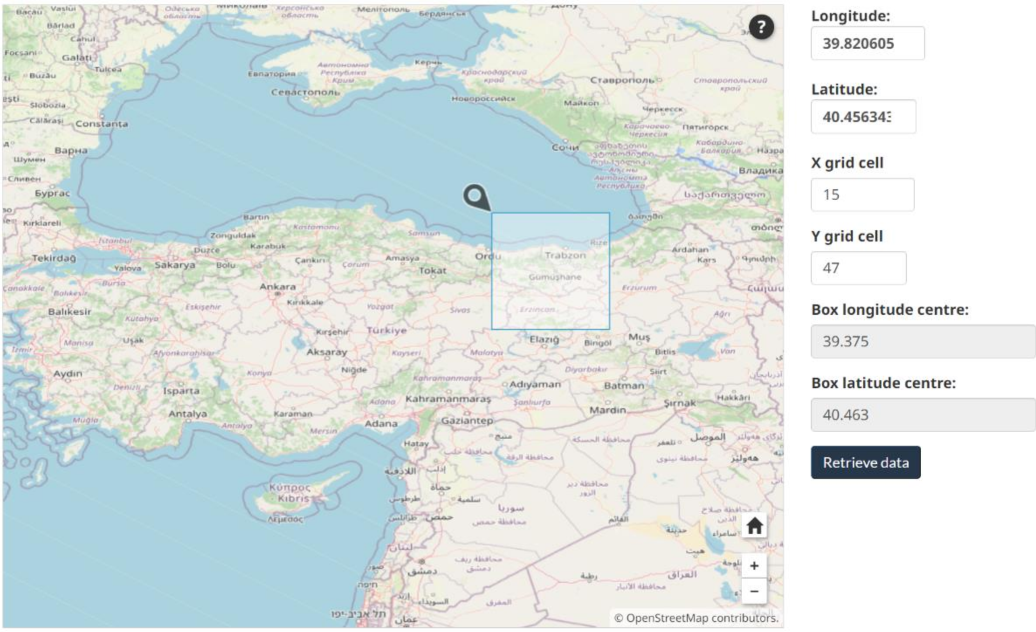

Figure 4.11 The grid box, which is the centroid of the basin, is located in the Environmental Canadian Global Climate Model

By clicking the retrieve data button, a zip file named “BOX_017X_46Y” was downloaded regarding the specific grid cell chosen. The file contains five folders and two text files, as shown in Figure 4.12.

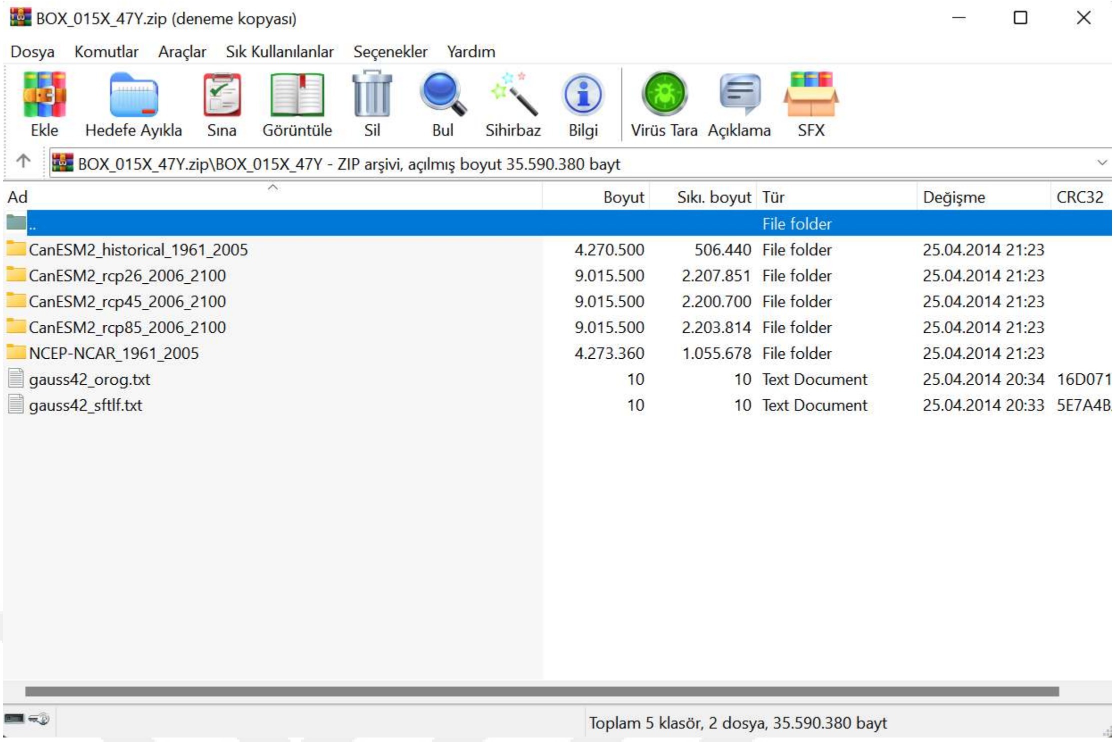

Figure 4.12 File includes CanESM2 large-scale factors

Table 4.2 contains a list of 26 large-scale factors, which have been included as input variable sets or predictors in the CanESM2 folders. After putting all 26 variables in one excel file, the given daily data was converted to monthly mean data. Then, the variables used as "predictors" in the first folder were separated into calibration (1970-1995) and validation (1996-2005) sets. The following three folders consist of future projections outputs, including the RCP2.6, RCP4.5, and RCP8.5 scenarios.

Table 4.2 Description of CanESM2 variables (Predictors)

No	CanESM2 Predictors	Description	Abbreviation
1	**ceshmslpgl**	Mean sea level pressure	V1
2	**ceshp1_fgl**	1000 hPa Wind speed	V2
3	**ceshp1_ugl**	1000 hPa Zonal wind component	V3
4	**ceshp1_vgl**	1000 hPa Meridional wind component	V4
5	**ceshp1_zgl**	1000 hPa Relative vorticity of true wind	V5
6	**ceshp1thgl**	1000 hPa Wind direction	V6
7	**ceshp1zhgl**	1000 hPa Divergence of true wind	V7
8	**ceshp5_fgl**	500 hPa Geopotential	V8
9	**ceshp5_ugl**	500 hPa Wind speed	V9
10	**ceshp5_vgl**	500 hPa Zonal wind component	V10
11	**ceshp5_zgl**	500 hPa Meridional wind component	V11
12	**ceshp5thgl**	500 hPa Relative vorticity of true wind	V12
13	**ceshp5zhgl**	500 hPa Wind direction	V13
14	**ceshp8_fgl**	500 hPa Divergence of true wind	V14
15	**ceshp8_ugl**	850 hPa Geopotential	V15
16	**ceshp8_vgl**	850 hPa Wind speed	V16
17	**ceshp8_zgl**	850 hPa Zonal wind component	V17
18	**ceshp8thgl**	850 hPa Meridional wind component	V18
19	**ceshp8zhgl**	850 hPa Relative vorticity of true wind	V19
20	**ceshp500gl**	850 hPa Wind direction	V20
21	**ceshp850gl**	850 hPa Divergence of true wind	V21
22	**ceshprcpgl**	Total precipitation	V22
23	**ceshs500gl**	500 hPa Specific humidity	V23
24	**ceshs850gl**	850 hPa Specific humidity	V24
25	**ceshshumgl**	1000 hPa Specific humidity	V25
26	**ceshtempgl**	Air temperature at 2 m	V26

CHAPTER 5
RESULTS AND DISCUSSION

5.1 General

In this study, as mentioned in the third chapter, the statistical downscaling method was used. The predictor required for this method is the large-scale weather factors (26 input sets) obtained from the Second Generation Earth System Model (CanESM2), as mentioned in the previous chapter. Local sediment yield and discharge data obtained from the General Directorate of State Hydraulic Works is used for predictand. The data sets are based on the monthly time scale, and the data were partitioned into two distinct components, namely calibration and validation. The calibration period, spanning the years between 1970 and 1995, is utilised to train the model. The validation period includes the period between 1996 and 2005 and is essential for evaluating the model's performance. After selecting the most effective predictors, Gene Expression Programming (GEP) and Linear Genetic Programming (LGP) models were used to predict monthly discharge and sediment yield on the basin. In the next stage, the prediction results were obtained and compared with each other to demonstrate the best model performance. Finally, for comparison, Linear Regression is performed.

In the following sections of this chapter, a detailed analysis of the models described and discussed in the previous sections, along with their comparisons, are presented. Firstly, the most effective set of predictors is introduced, followed by graphical and statistical results of calibration and validation periods, as well as the future predictions of discharge and sediment yield for three different RCP scenarios.

5.2 The Selection of the Most Effective Predictors

In order to enhance the association between the predictor variables and predictands, some normalisation techniques were applied to the data sets in addition to the standard ones. Specifically, each data set for both predictors and predictands was transformed by applying the natural logarithm function (Ln) and standardisation using the z-value. The variables were modified by implementing these normalisation methods to facilitate more accurate and reliable predictions. In order to identify the most impactful predictors and assess the linear relationship between input and output variables, a Pearson rank correlation coefficient analysis was utilised. The correlation analysis revealed that the large-scale weather factors were significantly correlated with local data at a confidence level of 99%. Those factors with the highest correlation coefficients were selected to identify the most effective predictors for the downscaling models. The correlation coefficients (R) between the predictors and predictands (both discharge and sediment yield) were calculated and listed in the following Tables 5.1, 5.2 and 5.3.

Table 5.1 Correlation coefficient (R) between non-normalized local variables and 26 GCM predictors

No.	GCM Predictor	R		No.	GCM Predictor	R	
		Q	Qs			Q	Qs
1	V1	-0,269	-0,150	14	V14	-0,357	-0,209
2	V2	-0,214	-0,117	15	V15	-0,142	-0,034
3	V3	-0,106	-0,004	16	V16	0,029	0,062
4	V4	0,025	0,061	17	V17	0,184	0,059
5	V5	0,375	0,212	18	V18	-0,095	0,010
6	V6	-0,096	0,007	19	V19	-0,166	-0,149
7	V7	-0,091	-0,051	20	V20	-0,071	-0,150
8	V8	-0,230	-0,086	21	V21	-0,351	-0,278
9	V9	-0,240	-0,086	22	V22	0,420	0,319
10	V10	0,217	0,151	23	V23	0,172	0,027
11	V11	0,177	0,076	24	V24	0,131	0,008
12	V12	-0,311	-0,173	25	V25	0,135	0,009
13	V13	0,057	0,010	26	V26	0,121	0,005

Table 5.2 Correlation coefficient (R) between Ln Q & Ln Qs (natural logarithm variables) and 26 GCM predictors

No.	GCM Predictor	R		No.	GCM Predictor	R	
		Q	Qs			Q	Qs
1	V1	-0,251	-0,288	14	V14	-0,356	-0,386
2	V2	-0,205	-0,221	15	V15	-0,112	-0,125
3	V3	-0,079	-0,086	16	V16	0,086	0,060
4	V4	0,081	0,056	17	V17	0,150	0,177
5	V5	0,348	0,363	18	V18	-0,058	-0,062
6	V6	-0,062	-0,065	19	V19	-0,179	-0,174
7	V7	-0,051	-0,076	20	V20	-0,126	-0,125
8	V8	-0,219	-0,219	21	V21	-0,374	-0,423
9	V9	-0,243	-0,220	22	V22	0,456	0,435
10	V10	0,200	0,235	23	V23	0,168	0,169
11	V11	0,187	0,179	24	V24	0,091	0,112
12	V12	-0,311	-0,294	25	V25	0,095	0,114
13	V13	0,054	0,069	26	V26	0,077	0,091

Table 5.3 Correlation coefficient (R) between standardised Q and Qs variables (z value) and 26 GCM predictors

No.	GCM Predictor	R		No.	GCM Predictor	R	
		Q	Qs			Q	Qs
1	V1	-0,269	-0,150	14	V14	-0,357	-0,209
2	V2	-0,214	-0,117	15	V15	-0,142	-0,034
3	V3	-0,106	-0,004	16	V16	0,029	0,062
4	V4	0,025	0,061	17	V17	0,184	0,059
5	V5	0,375	0,212	18	V18	-0,095	0,010
6	V6	-0,096	0,007	19	V19	-0,166	-0,149
7	V7	-0,091	-0,051	20	V20	0-,071	-0,150
8	V8	-0,230	-0,086	21	V21	-0,351	-0,278
9	V9	-0,240	-0,086	22	V22	0,420	0,319
10	V10	0,217	0,151	23	V23	0,172	0,027
11	V11	0,177	0,076	24	V24	0,131	0,008
12	V12	-0,311	-0,173	25	V25	0,135	0,009
13	V13	0,057	0,010	26	V26	0,121	0,005

The tables above show that correlations between predictors and predictands have been examined in three scenarios. As can be observed, the highest correlation was obtained when the natural logarithms of both data sets were taken. Therefore, in this thesis, data sets normalised by taking the natural logarithm have been used. The variables that provided the highest correlation were selected and shown in Table 5.4.

Table 5.4 Final most effective GCM predictors

No.	GCM Predictor	R		No.	GCM Predictor	R	
		Q	Qs			Q	Qs
1	V1	-0,251	-0,288	6	V10	0,200	0,235
2	V2	-0,205	-0,221	7	V12	-0,311	-0,294
3	V5	0,348	0,363	8	V14	-0,356	-0,386
4	V8	-0,219	-0,219	9	V21	-0,374	-0,423
5	V9	-0,243	-0,220	10	V22	0,456	0,435

The correlation coefficient's magnitude is an indicator of the degree of association between the variables, while its sign indicates the direction of the relationship, whether direct or inverse. Despite several large-scale variables showing a weaker correlation with local discharge and sediment yield data, they were still selected. This is because the set of GCM variables can describe the conditional process of the local-scale variable, which relies on several intermediate processes.

5.3 Statistical results of the models – calibration

To calibrate and train the downscaling models, discharge and sediment yield data were used along with the large-scale data of the calibration period (1970-1995).

5.3.1 Statistical results of the models for discharge (Q)

The GEP model generated an equation for the calibration process, and it uses the same equation to generate the validation process and forecast future projections. The generated equation for the first predictand, discharge, is shown below.

$$Ln(Q) = (V5 * V22) + (V22) + (2.4369617) + \left(\frac{1}{V12}\right)^2 + \left(\sqrt{(V2 + 2.1358698)} - V14 + \frac{1}{V12}\right) \quad (5.1)$$

As observed, the model utilised the V2, V5, V12, V14 and V22 predictors to generate the equation. These predictors have been selected by GEP automatically among the previously selected ten most effective predictors. Figure 5.1 shows the relative importance of predictors GEP model used for discharge (Q).

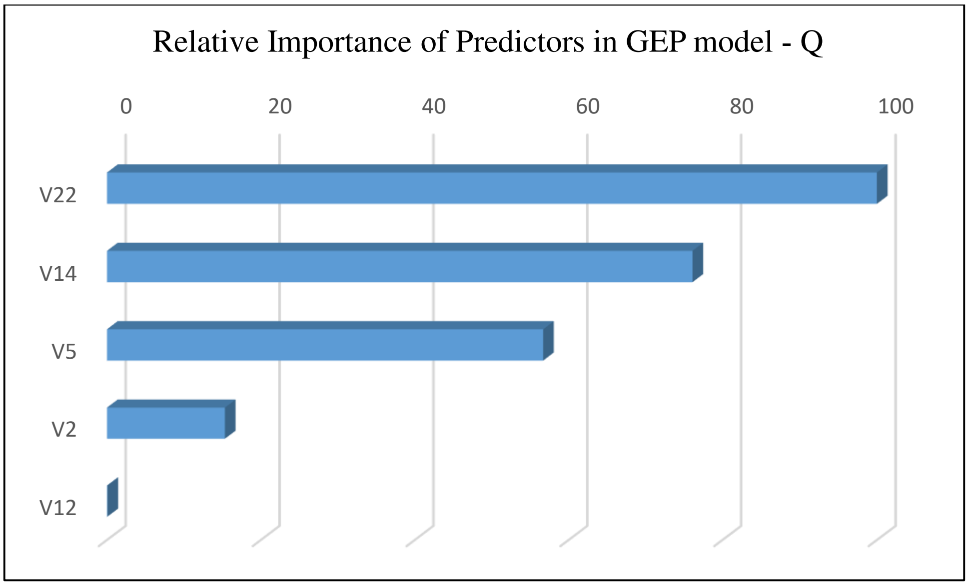

Figure 5.1 Relative importance of predictors in GEP model for discharge (Q)

As can be inferred from Figure 5.1, the most effective predictor is Total precipitation (V22), as 500 hPa Divergence of true wind (V14), 1000 hPa Relative vorticity of true wind (V5), 1000 hPa Wind speed (V2) and 500 hPa Relative vorticity of true wind (V12) have relatively less importance in GEP model for Q.

The LGP model generated a C code from the calibration data to predict the validation and future periods. Due to the extensive nature of the C code produced by the machine-coded LGP model software, it has been included in Appendix B. The LGP model used all ten selected most efficient predictors in the analyses. Figure 5.3 shows the relative importance of these predictors with respect to each other in the model.

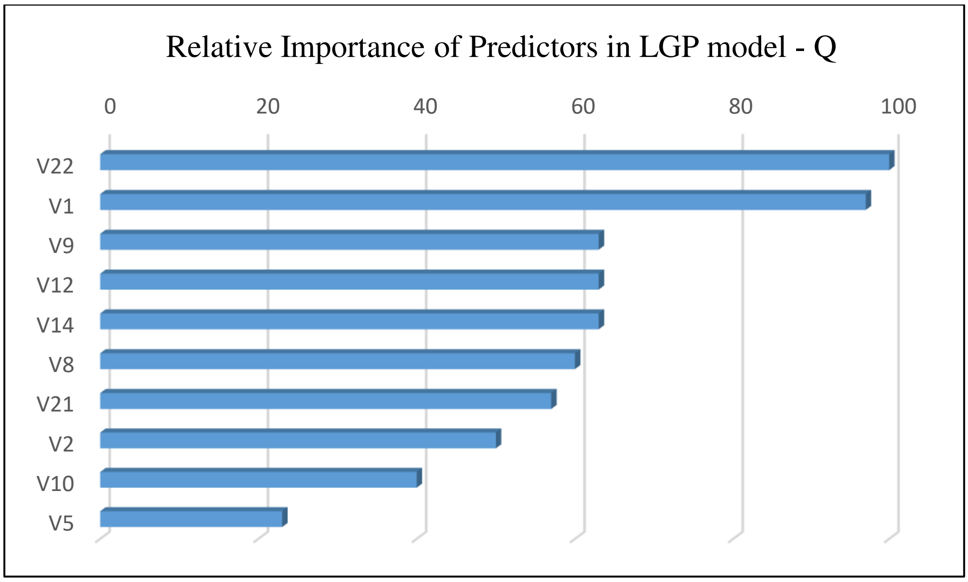

Figure 5.3 Relative importance of predictors in LGP model for discharge (Q)

For discharge data given to the model, as can also be noticed from Figure 5.3, while total precipitation (V22) and mean sea level pressure (V1) are the most effective predictors, the other selected predictors also have a considerable level of importance.

In order to observe the main relationship between the observed and predicted variables, a scatter plot was prepared. The correlation factor (R) between GEP, LGP and linear regression models for the calibration period is presented for the normalised variable (Ln). In Figures 5.4-5.6, the observed values are represented along the vertical axis, while the predicted values are represented along the horizontal axis. The linear equations of the relationship between observed and predicted values and correlation factor (R) for all models are also shown in Figures 5.4-5.6.

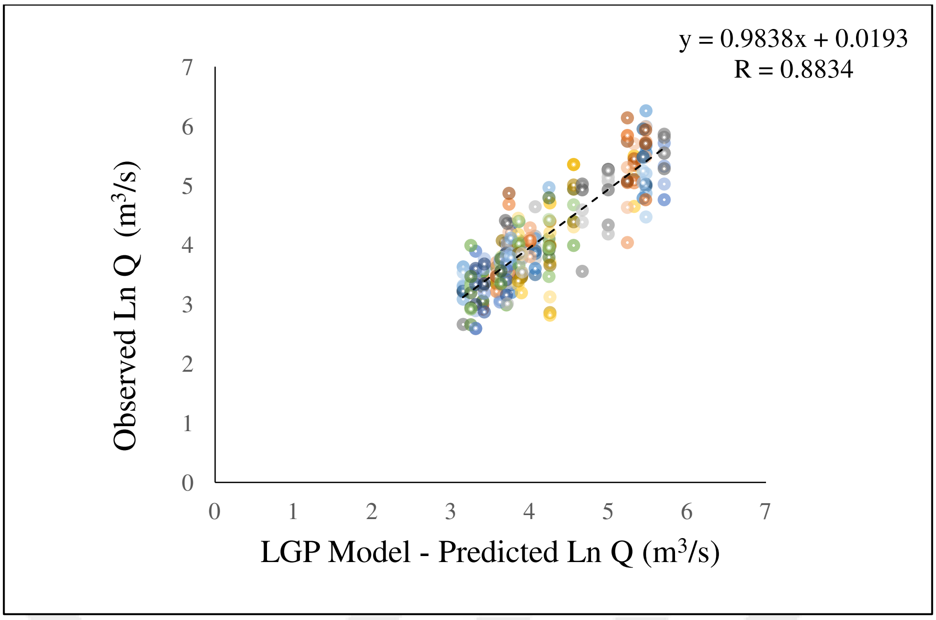

Figure 5.4 Scatter plot of the observed and predicted Ln Q values for the calibration period by LGP model

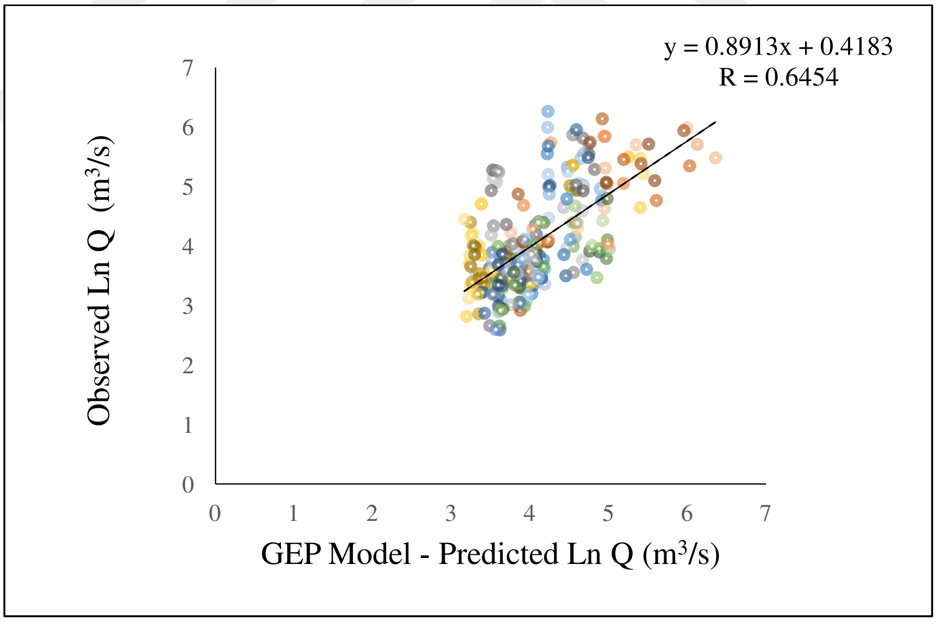

Figure 5.5 Scatter plot of the observed and predicted Ln Q values for calibration period by GEP model

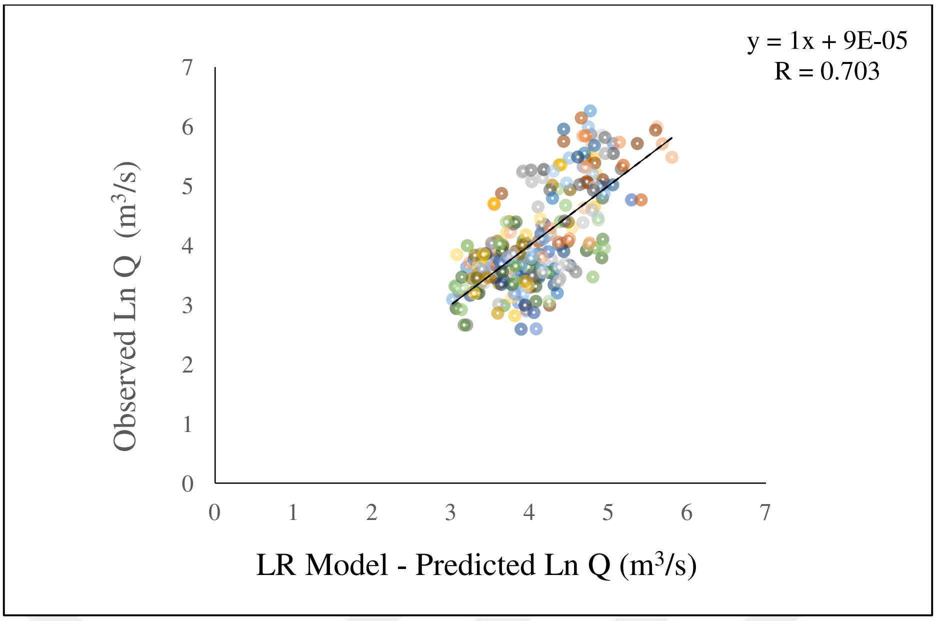

Figure 5.6 Scatter plot of the observed and predicted Ln Q values for calibration period by Linear Regression model

As seen in the figures, for the LGP model, the scatterplot shows a moderately strong linear relationship between the predicted and observed values, with a correlation factor of 0.8834. Moving from left to right along the horizontal axis, there appears to be an upward trend in the values, suggesting that higher predicted values tend to correspond to higher observed values. This trend is reinforced by the proximity of the data points to the fitted line, indicating that the model effectively captures the underlying patterns and relationships within the data.

Turning to the GEP model, the scatterplot reveals a weaker linear relationship between predicted and observed values, with a correlation coefficient of 0.6454. While there is still a discernible trend in the data, it is less pronounced than in the LGP model. The scatter of data points around the fit line is also somewhat spread out, indicating a higher degree of variability in the relationship between predicted and observed values. Nonetheless, the overall pattern suggests that the model still captures some meaningful signal in the data, albeit with a lower degree of accuracy than LGP and LR models.

Finally, the LR model displays a scatterplot with a correlation coefficient of 0.703, indicating a moderate linear relationship between the predicted and observed values. It is observed that it has a similar structure to the LGP and GEP and has a performance in between the two, suggesting that the model is capturing a reasonable degree of the underlying relationship between predicted and observed values.

Table 5.5 compares the statistical results of LGP, GEP, and LR models, along with the observed values for the calibration period (1970 – 1995).

Comparing the mean values, the LR model had the weakest performance compared to the other models in this area by making the farthest estimate (66.69 m^3/s) from the observed value (82.13 m^3/s), while the LGP model provided the closest value (79.91 m^3/s). Overall, all models underestimated the mean value compared to the observed data. Similarly, the models significantly underestimated the standard deviation compared to the observed data. The exact opposite situation can be seen in the comparison of minimum values. The GEP model provided the farthest estimate (23.70 m^3/s) from the observed minimum value (13.36 m^3/s), while the LR model gave the closest estimate (20.46 m^3/s). Contrary to their predictions for mean values, the models have predicted minimum values higher than the observed values. The GEP model performed well by making a close estimate of 9.82% higher than the observed value in predicting the maximum value, while the LGP model showed weak performance by making a prediction far below the observed value by 42.75%.

LGP model has lesser MAE (25.41 m^3/s) and RMSE (47.49 m^3/s) and a higher correlation factor (R) with the observed data than the other models. Following the LGP model Linear Regression model has a 38.03 m^3/s MAE and 68.76 m^3/s RMSE, GEP model has 40.71 m^3/s MAE and 78. 48 m^3/s RMSE.

Table 5.5 Comparison of statistical results of observed and predicted Q (m^3/s) for the calibration period (1970 – 1995)

Data	**Mean**	**Min**	**Max**	**Std. Deviation**	**MAE**	**RMSE**	**R**
Observed	82.13	13.36	523.88	88.62	-	-	-
GEP Model	70.92	23.70	575.34	64.97	40.71	75.48	0.645
LGP Model	79.91	23.34	299.91	74.16	25.41	47.49	0.883
Linear Regression	66.69	20.46	330.59	45.52	38.03	68.76	0.703

Figures 5.7, 5.8 and 5.9 provide a visual representation of the predicted mean monthly discharge (Q) of LGP, GEP and LR models and observed data during the calibration period.

The graphs show that while the general trends of the models are similar, they exhibit differences in proximity to the observed data. As can be easily inferred from the figures, all models overestimated Q in January, March, August and October. On the other hand, in April, May, and June, all models underestimated Q. In February, LGP and GEP models overestimated Q, while the LR model made a more accurate prediction just below the observed value. In September, LGP underestimated Q by 5.14%, while other models made higher predictions by 28.4% and 2.61% for GEP and LR, respectively. A similar scenario occurred in November, with the LGP model underestimating Q while other models overestimated the discharge (Q).

In February, LGP and GEP stayed over the observed value, LR model made a lower prediction closer to the observed value. In July, while LGP and LR models made predictions higher than the observed value, the GEP model made a prediction lower than the observed value. Finally, in December, LGP and LR models made very close predictions with deviations of 3.16% and 0.25%, respectively, just below the observed value. However, the GEP model made a prediction higher than the observed value with a deviation of 11.92%. Overall, the superiority of the LGP model to the other models is evident in the Figures 5.7-5.10 and Table 5.5, and it can make highly accurate predictions in monthly averages.

In Figure 5.10, all models can be observed and compared visually.

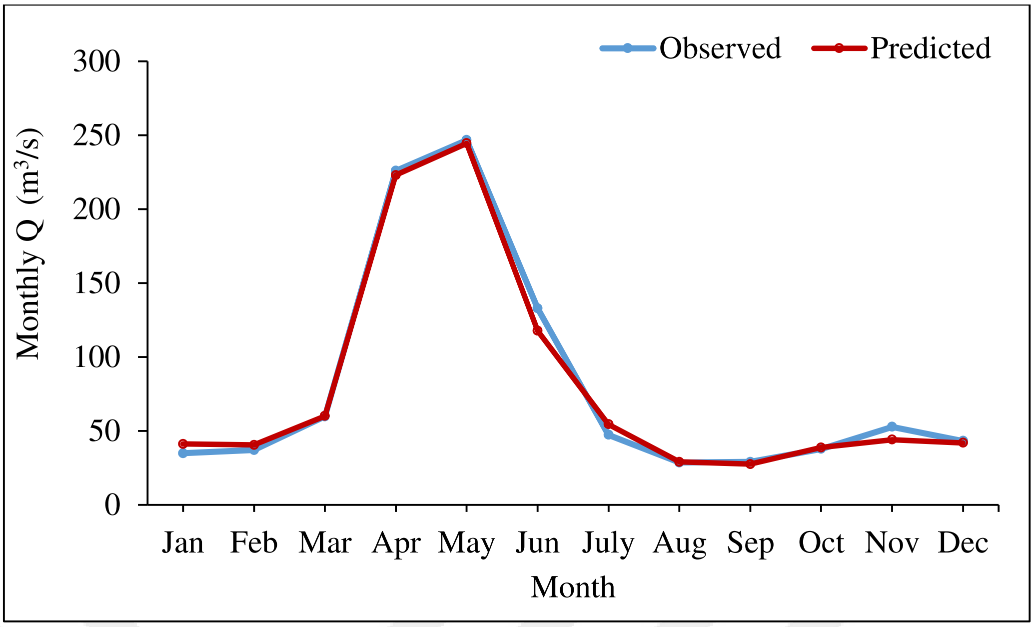

Figure 5.7 Observed and predicted monthly Q values for the calibration period by the LGP model

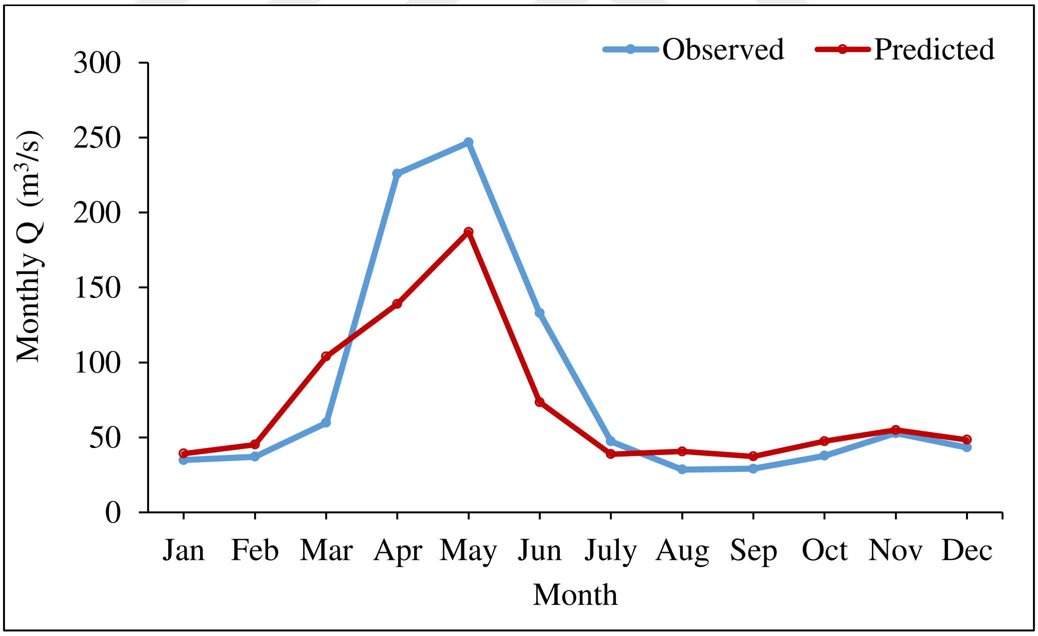

Figure 5.8 Observed and predicted monthly Q values for the calibration period by GEP model

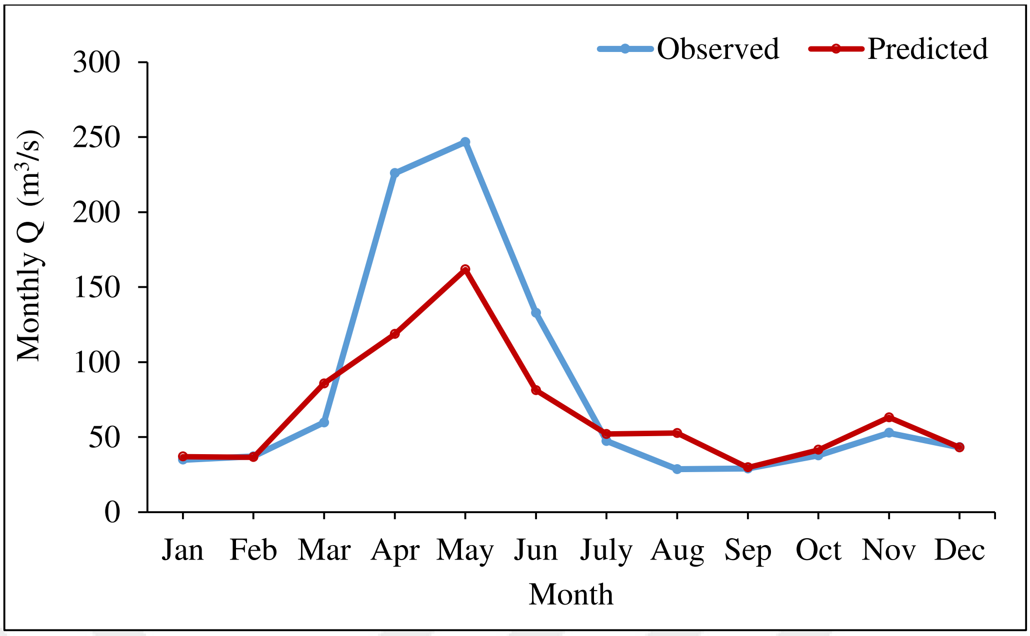

Figure 5.9 Observed and predicted monthly Q values for the calibration period by Linear Regression model

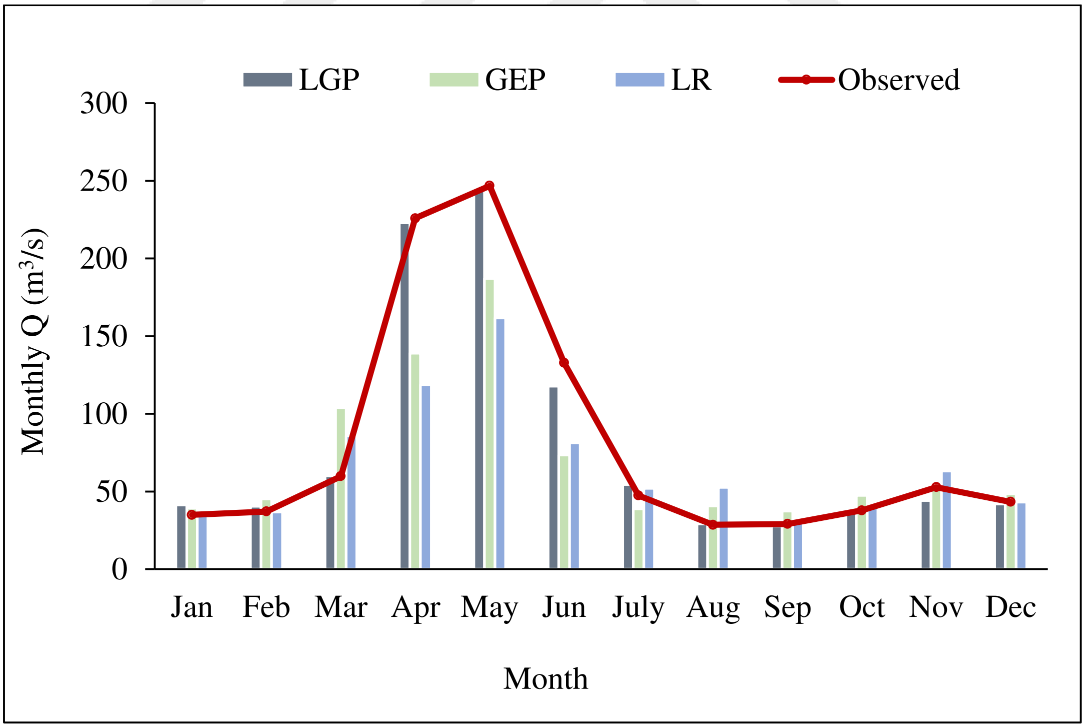

Figure 5.10 Observed and predicted monthly Q values for the calibration period by all models

5.3.2 Statistical results of the models for sediment yield (Qs)

The equation generated by the GEP model for the second predictand, sediment yield (Qs), is as follows:

$$Ln(Qs) = \left((2 * V22) + \left(V5 + \left(3 * (V10 * V22)\right)\right) + 6.2984642\right) - V9 \quad (5.2)$$

With sediment yield (Qs) data being the predictand, the GEP model used only four of the ten selected predictors. These are Total precipitation (V22), 500 hPa Wind speed (V9), 500 hPa Zonal wind component (V10) and 1000 hPa Relative vorticity of true wind (V5). Figure 5.11 shows the relative importance of predictors GEP model used for sediment yield (Qs).

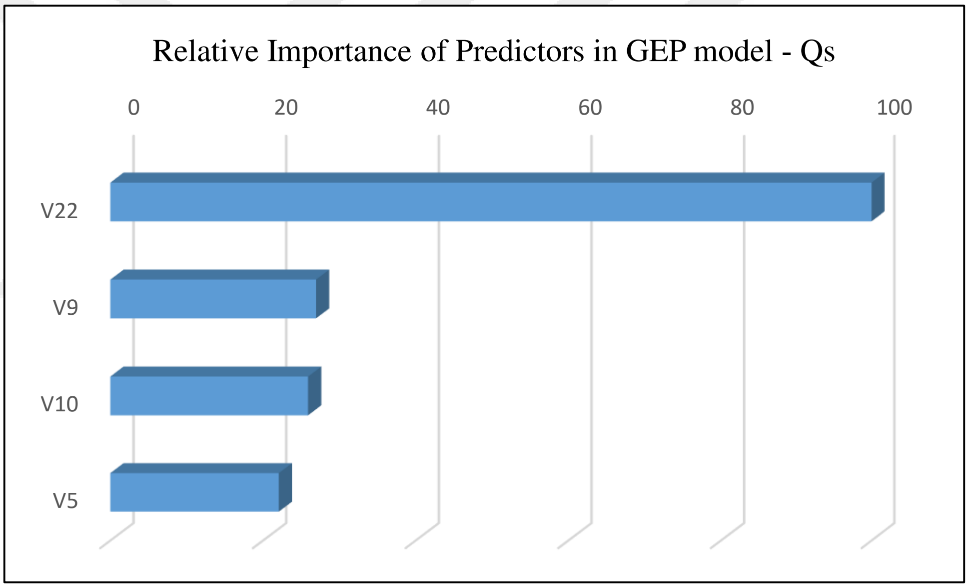

Figure 5.11 Relative importance of predictors in GEP model for sediment yield (Qs)

LGP model generated a C code from the given calibration data to predict the validation and future periods. C code produced by the LGP model is shown in Appendix C.

Figure 5.12 shows the relative importance of these predictors with respect to each other in the model.

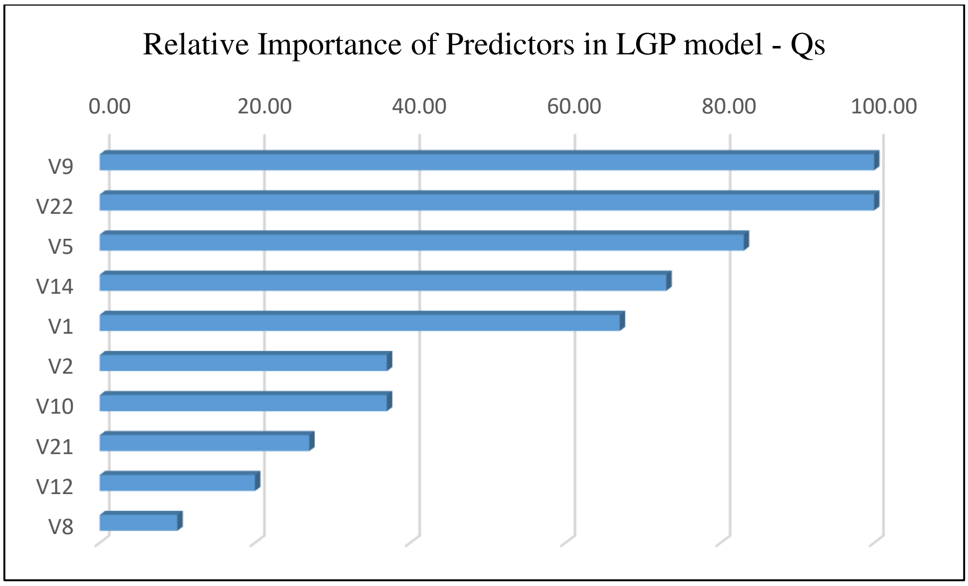

Figure 5.12 Relative importance of predictors in LGP model for sediment yield (Qs)

As observed in the results obtained by the model with sediment yield data, total precipitation (V22) is still of great importance, but 500 hPa wind speed (V9) also has the same significance level. In addition to these, it can be inferred from the figure that 1000 hPa relative vorticity of true wind (V5), 500 hPa divergence of true wind (V14), and mean sea level pressure (V1) also have noticeable importance.

Figures 5.13-5.15 illustrates the scatterplot of the observed and predicted variables of Qs. Correlation factor R is presented for the Ln values. The linear equations of the relationship between observed and predicted values and correlation factor (R) for all models are also shown in Figures 5.4-5.6.

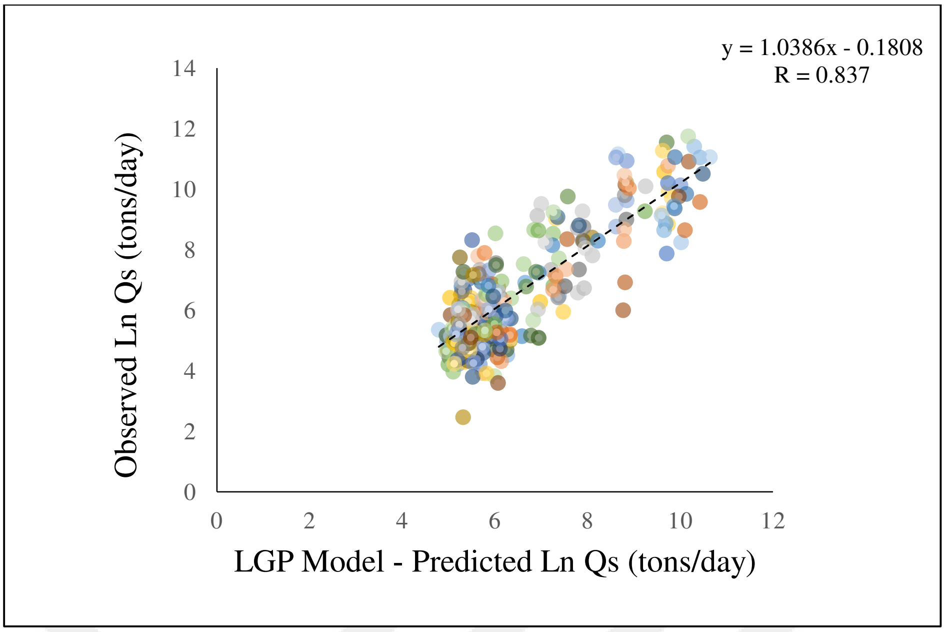

Figure 5.13 Scatter plot of the observed and predicted Ln Qs values for calibration period by LGP model

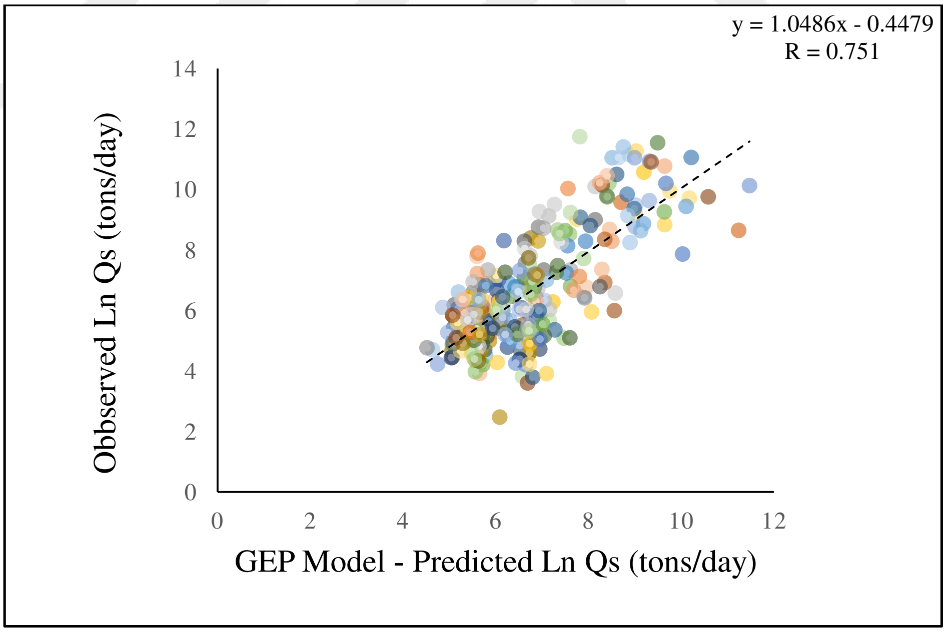

Figure 5.14 Scatter plot of the observed and predicted Ln Qs values for calibration period by GEP model

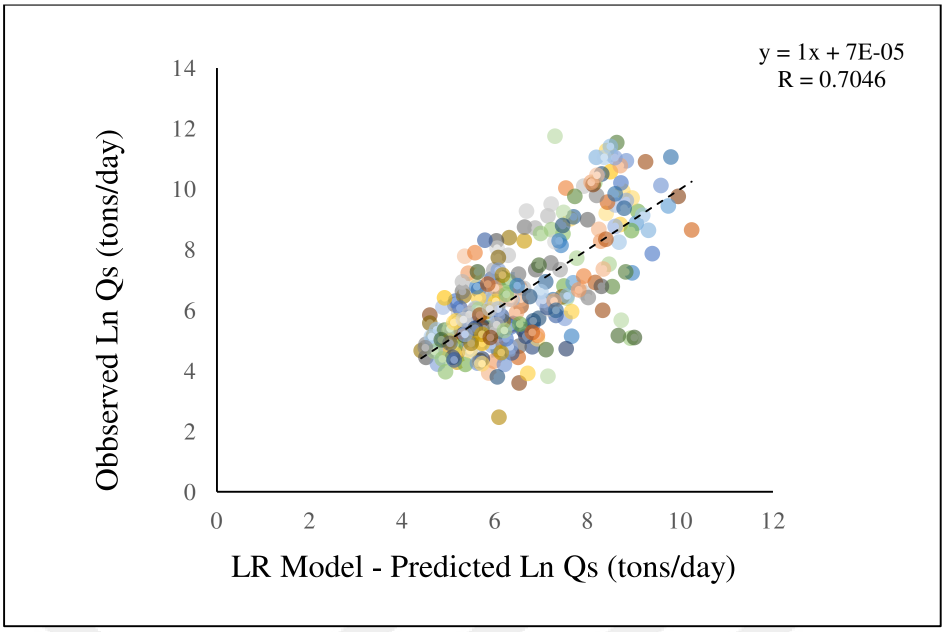

Figure 5.15 Scatter plot of the observed and predicted Ln Qs values for calibration period by Linear Regression model

As seen in Figures 5.13-5.15, similar to the discharge data, the LGP model performs better than the others. However, in this case, the GEP model performed better than LR and demonstrated a correlation of 0.751. The fact that the data in all models continue to increase from left to right and are close to the fit line is also an indication that the models performed well during the calibration period (1970 – 1995).

Table 5.6 compares the statistical values of observed and predicted data for the models. All models have underestimated the mean values of the data. The LGP model has the closest prediction to the observed data, while the LR model has the farthest value. The situation is the opposite regarding minimum values, and all models have made mostly unsuccessful predictions by overestimating the values. Similar patterns have been observed for maximum and standard deviation, and all models have underestimated the observed data. However, the GEP model has predicted the closest values for both, while the LR model has been the farthest.

Looking at the RMSE and MAE values, the LGP model has again provided the best performance with the lowest MAE and RMSE values, followed by the LR and GEP models, respectively. It is noteworthy that although LR lags behind GEP in correlation, it has shown slightly better performance than GEP in terms of MAE and RMSE.

Table 5.6 Comparison of statistical results of observed and predicted Qs (tons/day) for the calibration period (1970 – 1995)

Data	Mean	Min	Max	Std. Deviation	MAE	RMSE	R
Observed	5512.41	11.75	126037.70	15601.00	-	-	-
GEP	2703.17	91.73	96240.27	8227.52	4884.32	15484.23	0.751
LGP	2929.68	120.01	41758.04	6661.71	3806.41	12264.77	0.837
Linear Regression	1816.18	82.30	28184.55	3254.59	4798.76	15105.13	0.705

For Figures 5.16-5.18, it can be observed from the graphs that the models mostly underestimate Qs. This is true for all models in April, May, June, and December, as well as for LGP and GEP in February, March, July, and November. Generally, LR makes higher predictions during these months but provides closer results than LGP and GEP. In contrast to the general behaviour, all models overestimated Qs in August and September. In January, the GEP model underestimated the observed by 10.23%, LGP overestimated the observed by 13.4%, and LR slightly overestimated discharge just by 0.69%. It has been seen that the LR model gave results so close to the values observed in January, overestimating by 0.69% and in February, overestimating by 0.01%. In October, GEP overestimated Qs, while LGP and LR made low and relatively closer predictions.

As can be understood from Table 5.6 and Figures 5.16-5.18 in general, the LGP model outperforms the GEP and LR models once again. This can also be visually observed in Figure 5.19, where all the models are compared.

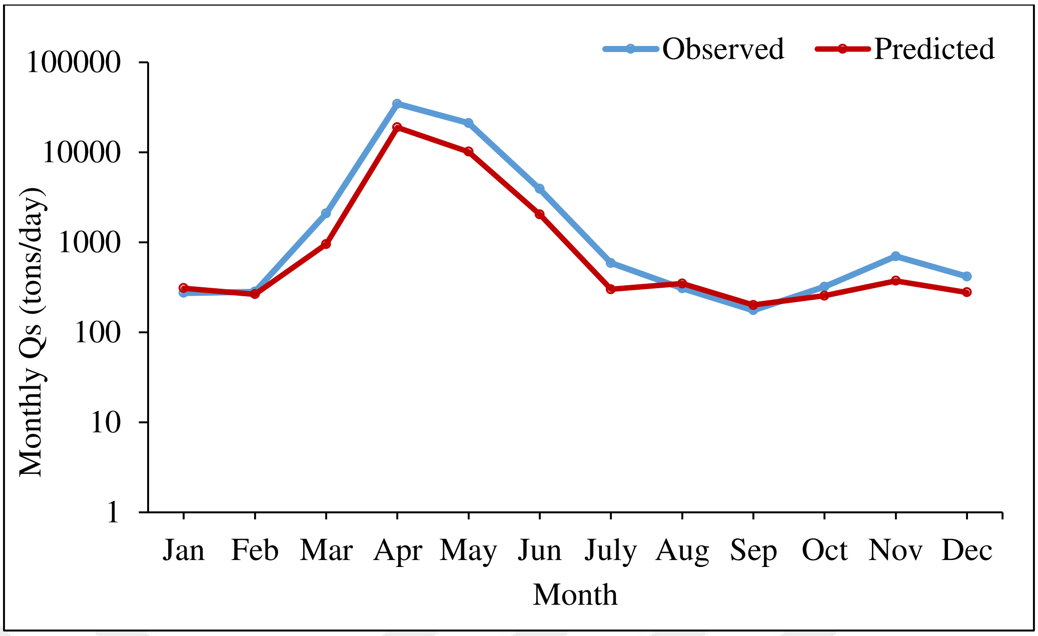

Figure 5.16 Observed and predicted monthly Qs values for the calibration period by LGP model in logarithmic scale

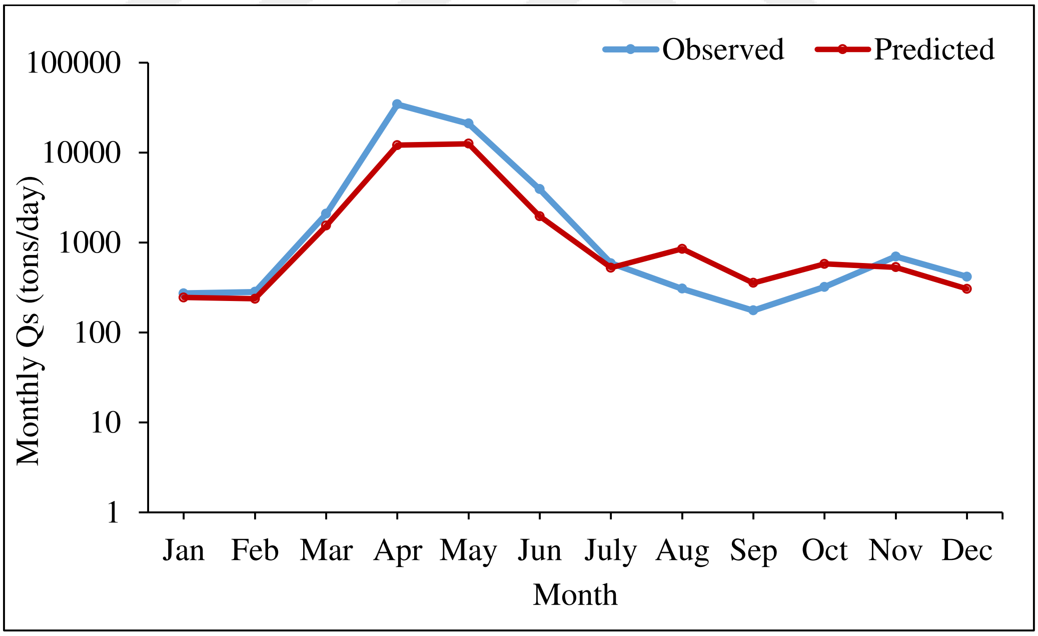

Figure 5.17 Observed and predicted monthly Qs values for the calibration period by GEP model in logarithmic scale

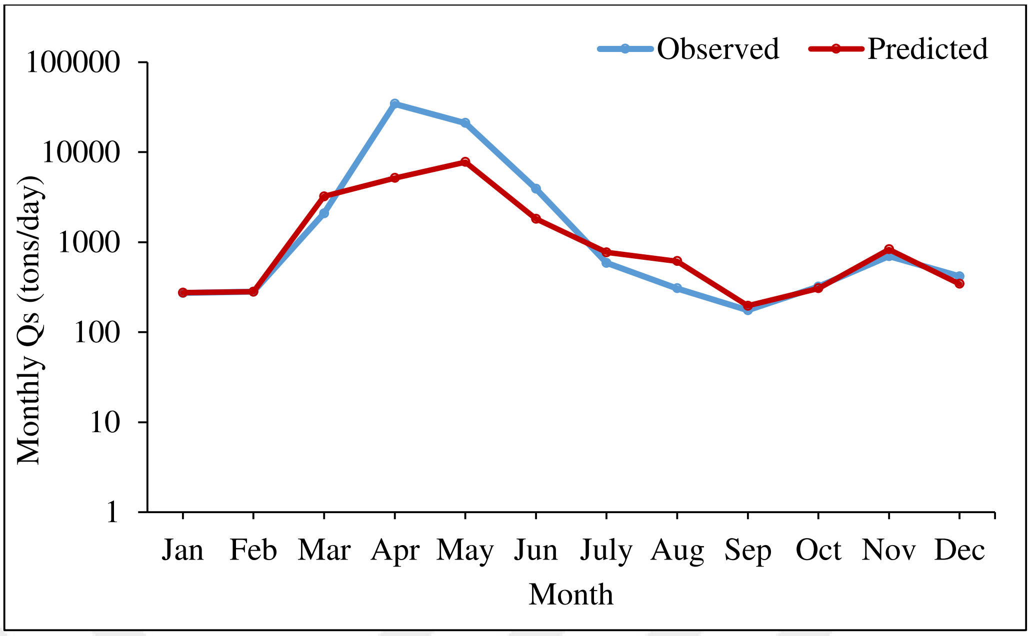

Figure 5.18 Observed and predicted monthly Qs values for the calibration period by Linear Regression model in logarithmic scale

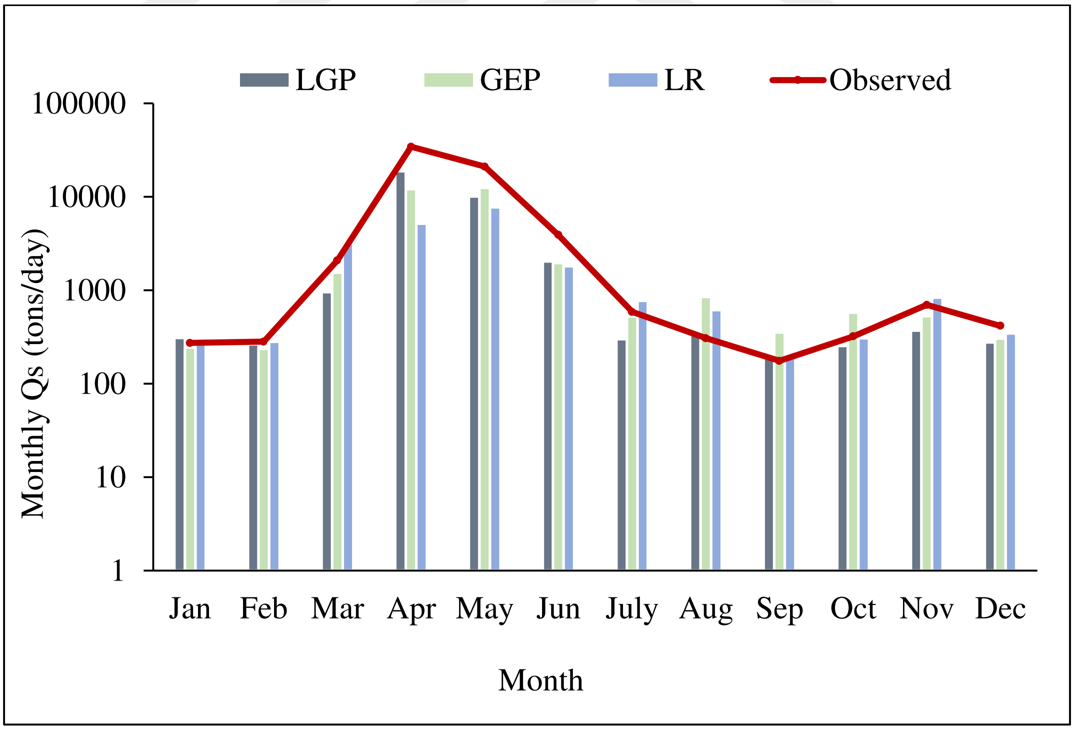

Figure 5.19 Observed and predicted monthly Qs values for the calibration period by all models in logarithmic scale

5.4 Statistical results of the models – validation

The validation process involves applying the formulation derived from the calibration period to a separate dataset to assess the model's ability to generalise beyond the calibration period. This involves testing the model's ability to accurately predict future outcomes by comparing its predictions against observed data not included in the calibration process. By doing so, we can evaluate the model's robustness and reliability in predicting future outcomes.

5.4.1 Statistical results of the models for discharge (Q)

The best-performing model during the calibration period was the LGP model. The validation process was created using the C code, shown in Appendix B, generated by the LGP model during the calibration period, and the outputs were shown in Figure 5.20. Similarly, the validation process for the GEP model was created using equation 5.1 generated by the GEP model, and its outputs were shown in Figure 5.21 on a scatter plot. Also, the linear regression model is included in the scatterplots for comparison. The observed (vertical axis) and predicted (horizontal axis) results of the models compared to each other with the correlation factor (R) shown on the distribution graphs are presented in Figures 5.20-5.22. As can be seen from the clustering of values around the fit line in Figures 5.20-5.22, it can be said that the models performed well. The correlation between the observed value and the predictions of the LGP model is 0.8414. This value is 0.684 for GEP and 0.745 for the Linear Regression model.

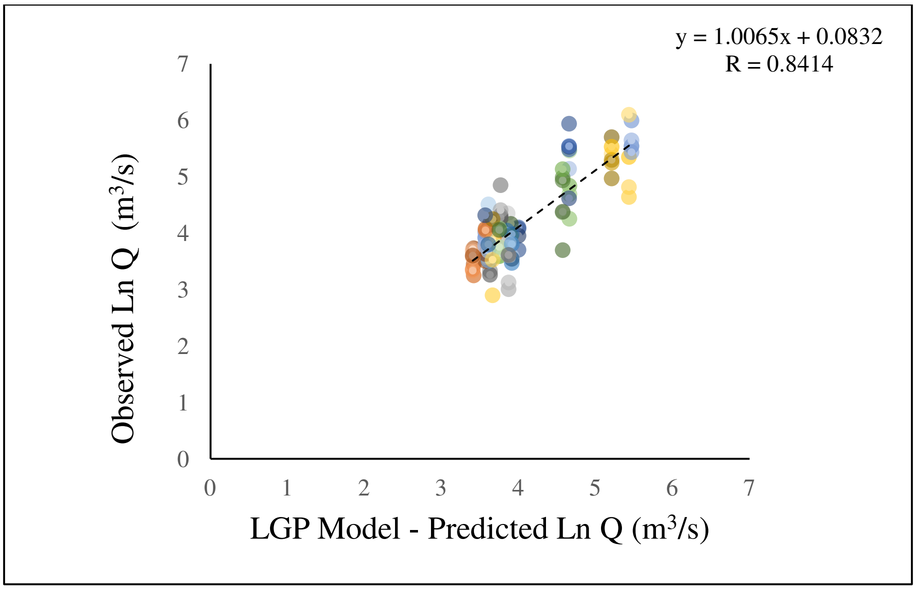

Figure 5.20 Scatter plot of the observed and predicted Ln Q values for the validation period by LGP model

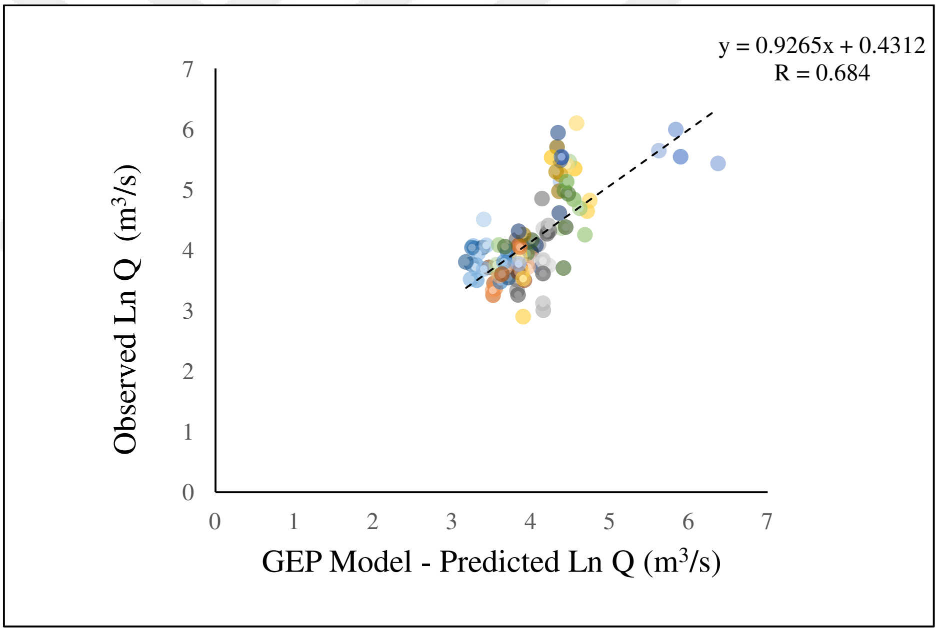

Figure 5.21 Scatter plot of the observed and predicted Ln Q values for the validation period by GEP model

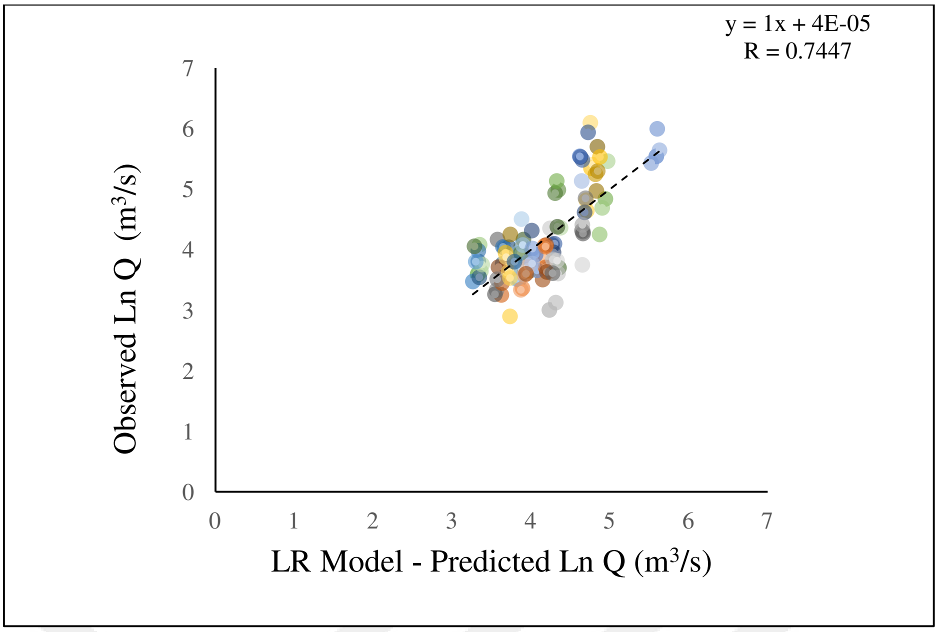

Figure 5.22 Scatter plot of the observed and predicted Ln Q values for the validation period by Linear Regression model

Table 5.7 compares the statistical results of observed and predicted values during the validation period (1996-2005). This comparison shows that the LR model provides the best result in predicting the mean value. Although the GEP model performs poorly in predicting the mean value, it performs best in predicting the minimum value. Furthermore, all models underestimated the mean values and overestimated the minimum values. The GEP model, again, exhibits the best performance for predicting the maximum value and standard deviation.

Table 5.7 also demonstrates that the LGP model performs best with the lowest MAE and RMSE. LR and GEP models follow LGP in terms of performance, respectively, in line with the correlation factor (R) values.

Table 5.7 Comparison of statistical results of observed and predicted Q (m^3/s) for the validation period (1996 – 2005)

Data	Mean	Min	Max	Std. Deviation	MAE	RMSE	R
Observed	84.92	18.10	442.89	82.90	-	-	-
GEP	66.32	24.01	585.88	67.76	36.69	74.21	0.684
LGP	69.75	30.14	237.34	56.72	27.14	51.79	0.841
Linear Regression	73.08	26.16	279.15	47.38	32.48	58.14	0.745

As observed in Figures 5.23-5.25, the models generally underestimated the discharge values compared to the observed data. It can be seen that the LGP model underestimated the observed in all months except January and September. All models underestimated Q in April, May, June, July, and December. In September, all models overestimated Q, but LGP has the closest prediction. The LGP model underestimated Q in February, August and October, whereas the GEP and LR models overestimated Q values. The LGP and GEP models underestimated the discharge (Q) values in March and November, while the LR model overestimated them. LGP model showed the best performance in April, October and December, with underestimation percentages of 7.61%, 7.51%, and 6.63%, respectively. The worst performance of the LGP model is in March and May, with underestimations by 35.32% and 38.14%, and in September, with overestimation by 21.92%. The GEP model provided very close estimates just below the observed values in March and November by 6.344%, 2.04%, and just over to the observed in August by 2.6%, but produced inaccurate estimates in April and July below the observed value by 59.394%, 46.158%, and over to the observed value in September by 57.186%. The LR model showed the weakest performance among the three models in general, with the closest estimate in October just above the observed by 2.84% and the furthest estimate with an overestimation percentage of 61.804% in September. Overall, the LGP model showed the best performance in predicting the observed values, while the LR model performed relatively inadequately in predicting the observed values.

Figure 5.26 demonstrates observed and predicted discharge values for all models, where each model can be compared with others in one graph.

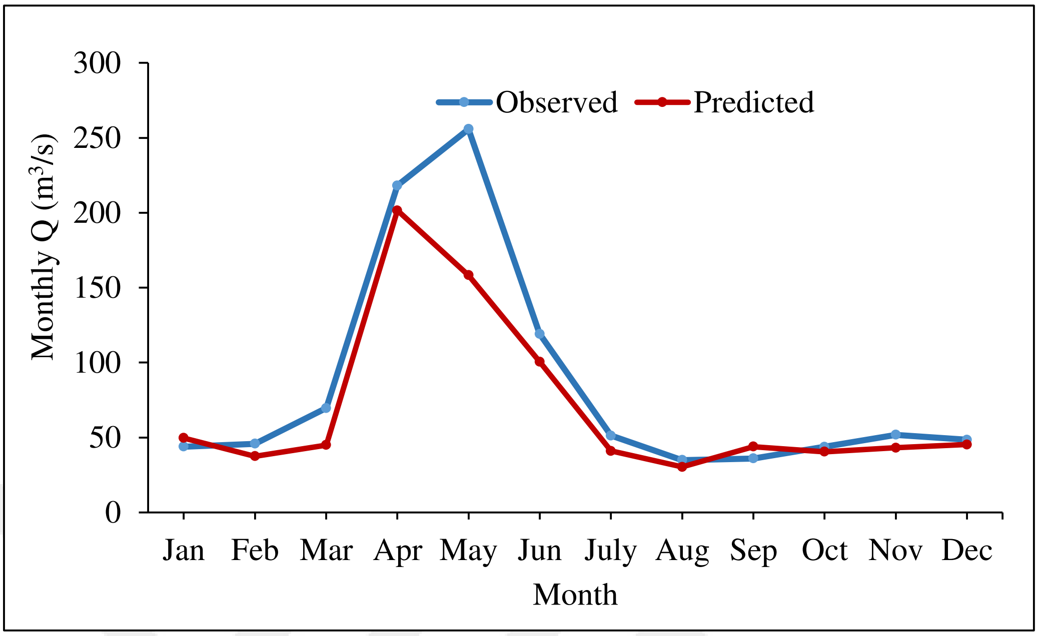

Figure 5.23 Observed and predicted monthly Q for the validation period by LGP model

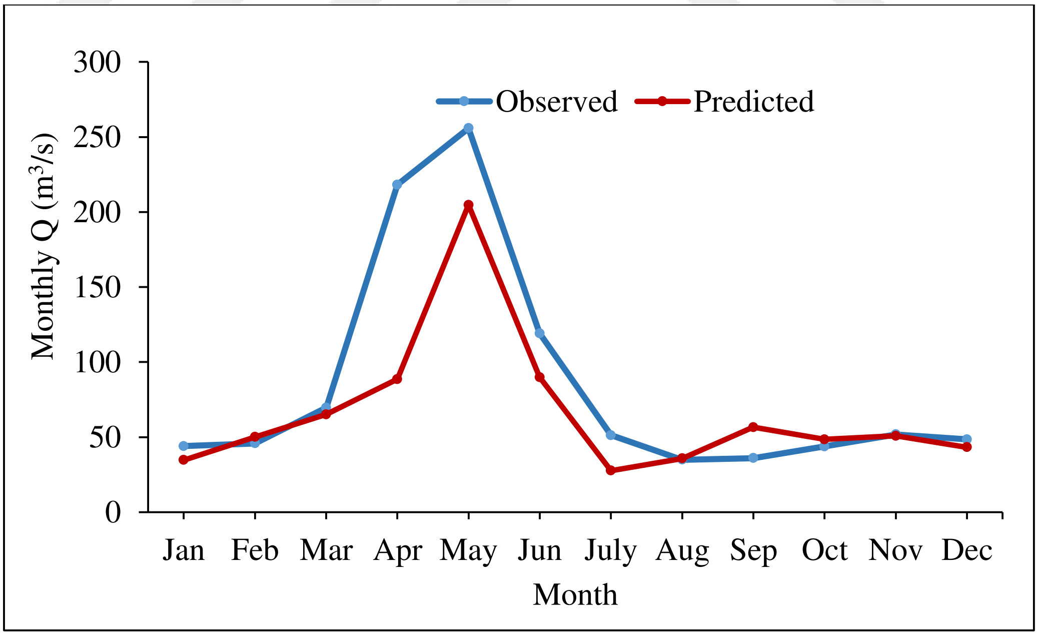

Figure 5.24 Observed and predicted monthly Q for the validation period by GEP model

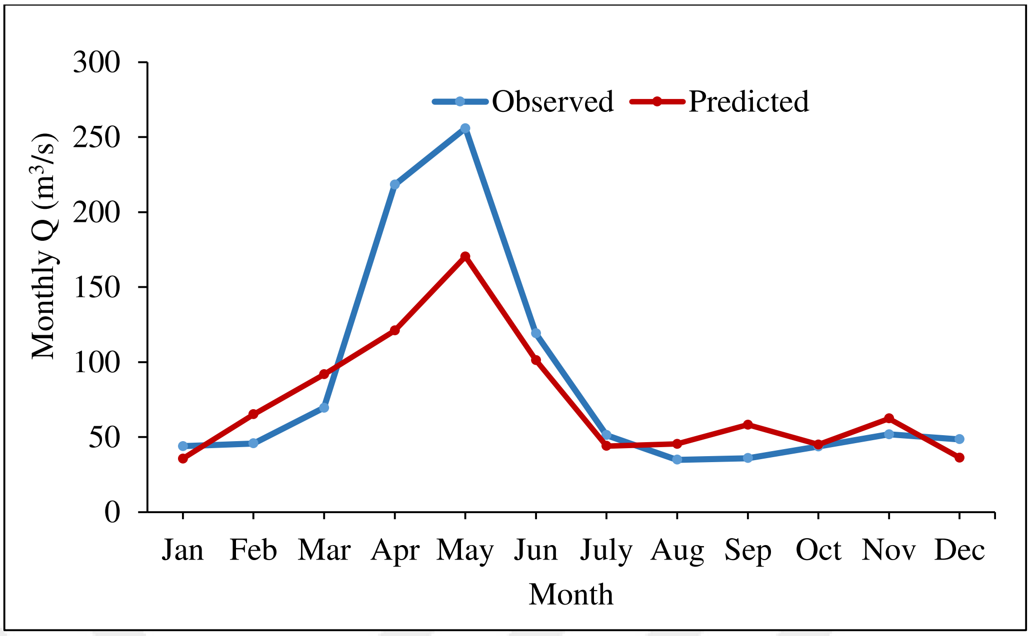

Figure 5.25 Observed and predicted monthly Q for the validation period by Linear Regression model

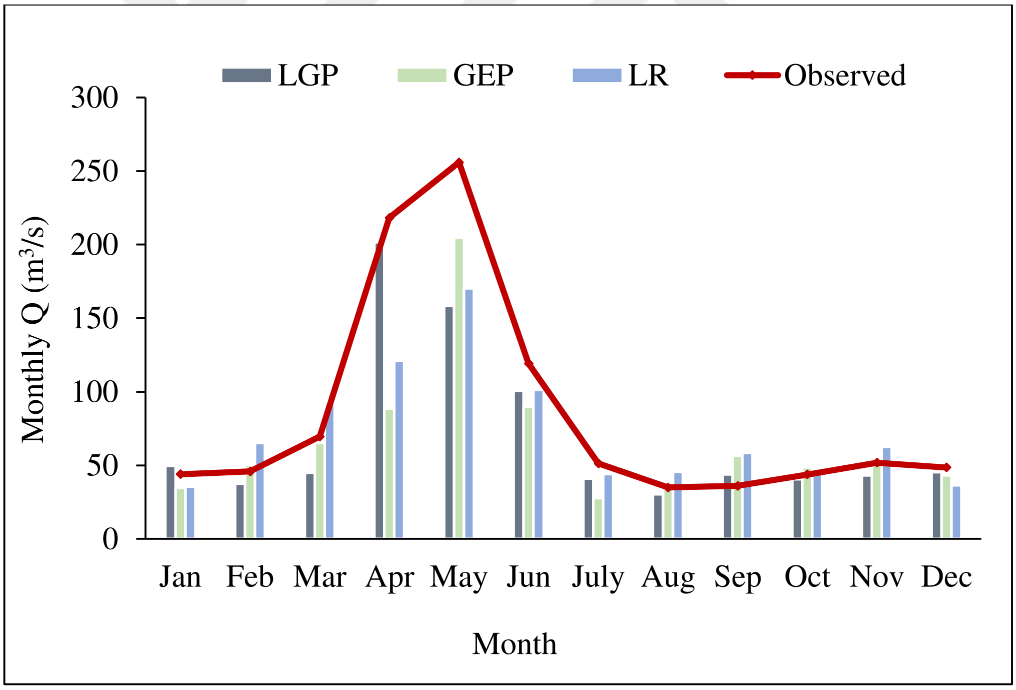

Figure 5.26 Observed and predicted monthly Q for the validation period by all models

5.4.2 Statistical results of the models for sediment yield (Qs)

The best-performing model during the calibration period for sediment yield was, again, the LGP model. Using equation 5.2, the validation process of the GEP model and using the C code in Appendix C, the validation process of the LGP model was created. The outputs of LGP and GEP and linear regression models for the validation period were shown in correlation factor (R). A scatter plot was created using observed (vertical axis) and predicted Qs (horizontal axis), as illustrated in Figures 5.27-5.29. Linear equations of each model were indicated on the graphs.

It can be seen from Figures 5.27-5.29; all models showed a decent performance. With a correlation factor of R=0.7875, the LGP model performed the best. Following LGP, LR and GEP models performed similarly with correlation factors R=0.705 and R=0.6733, respectively.

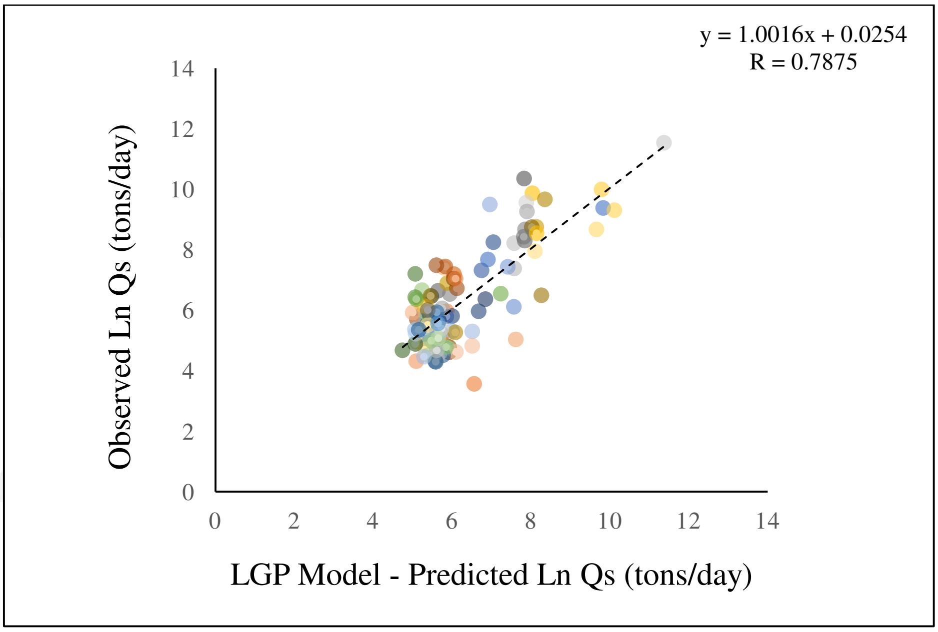

Figure 5.27 Scatterplot of the observed and predicted Ln Qs values for the validation period by LGP model

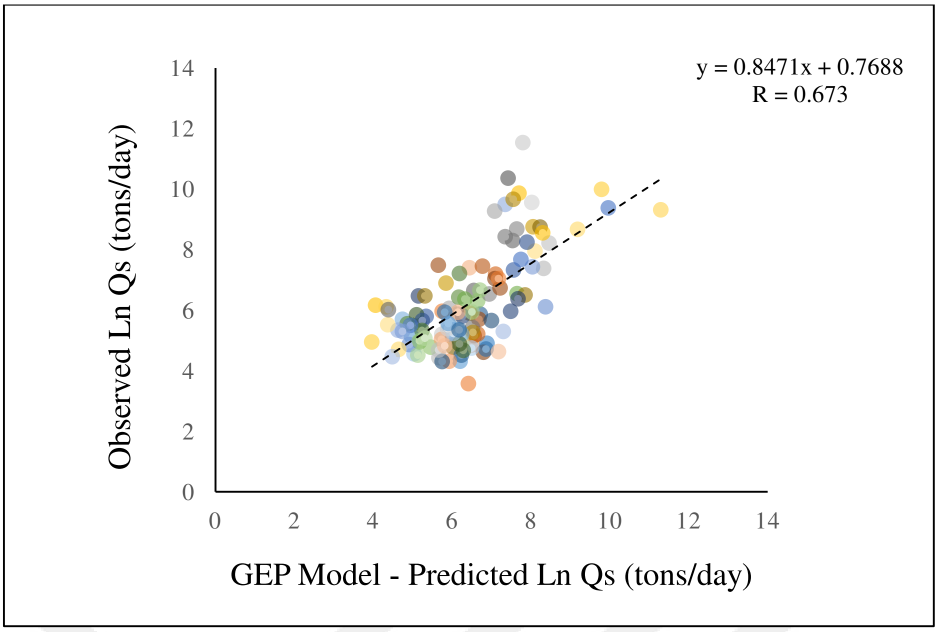

Figure 5.28 Scatterplot of the observed and predicted Ln Qs values for the validation period by GEP model

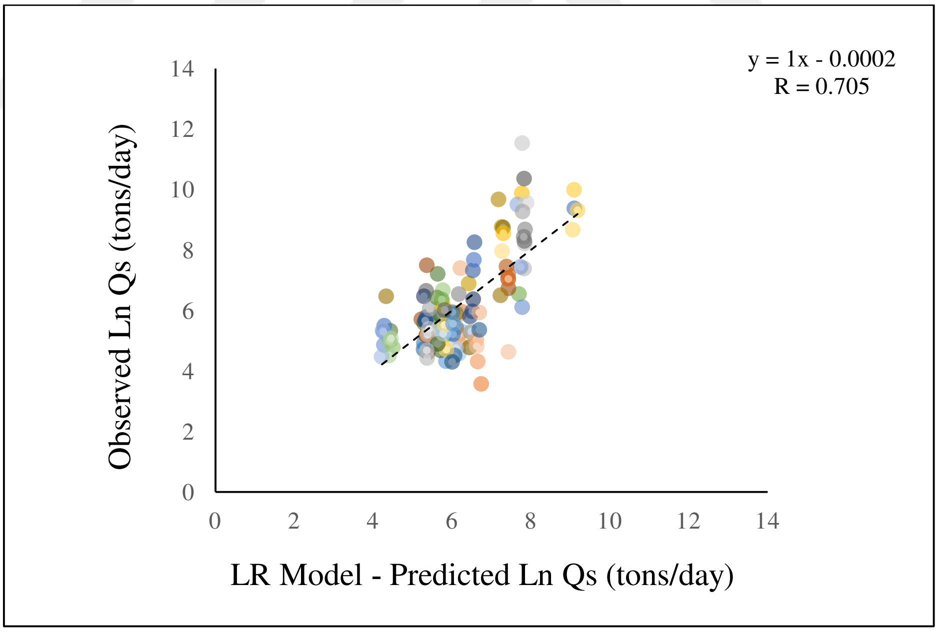

Figure 5.29 Scatterplot of the observed and predicted Ln Qs values for the validation period by Linear Regression model

Table 5.8 provides a comparison of the statistical results of observed and predicted sediment yield values during the validation period spanning from 1996 to 2005. Similar to previous scenarios, all models underestimated the observed mean value.

While the minimum observed value was 35.40 tons/day, the LGP model predicted 115.28 tons/day, the GEP model predicted 52.97 tons/day, and the LR model predicted 68.07 tons/day, all of which were above the minimum value. In predicting the maximum and standard deviation values, all models also made predictions below the observed value. The LGP model provided the closest predictions except for the minimum value. GEP model was superior to the LR model in estimating all statistical values.

In contrast to other statistics, the GEP model exhibited the weakest performance with MAE=2880.9 and RMSE=11871.73 values. The LGP model showed the best performance with MAE=1593.34 and RMSE=4325.66 values. The LR model performed between these two models, which is consistent with the correlation factor (R) values.

Table 5.8 Comparison of statistical results of observed and predicted Qs (tons/day) for the validation period (1996 – 2005)

Data	**Mean**	**Min**	**Max**	**Std. Deviation**	**MAE**	**RMSE**	**R**
Observed	2890.33	35.40	102010.30	10348.66	-	-	-
GEP	2014.00	52.97	80179.85	7760.17	2880.90	11871.73	0.673
LGP	2084.14	115.28	87534.24	8693.8	1593.34	4325.66	0.788
Linear Regression	1023.18	68.07	9724.34	1689.62	2282.65	10028.62	0.705

Figures 5.30-5.32 present comparisons of observed and predicted sediment yield values for each model during the validation period covering 1996-2006. Figure 5.33 allows for a comparison of all models in a single graph. It can be seen that all models have made overestimations during February, August, and September, while underestimations have been made during April, June, and December. LGP has shown the best performance among the three models in February, April, May, August, September, November, and December. The GEP model has made more accurate predictions than LGP and LR models in January, March, June, July, and October.

In January, which is the closest to the model's predictions, LGP has made an overestimation by 28.29%, while GEP and LR models have made underestimations by 11.47% and 12.08%, respectively. In September, the farthest from the models' predictions, the LGP model made a high overestimation by 135.38%, the LR model made an even higher overestimation by 155.10%, and the GEP model made the highest overestimation by 158.64%.

In general, as shown in Figures 5.30-5.33 and Table 5.8, LGP has been the most successful model among the three models for predicting sediment yield in both calibration and validation periods.

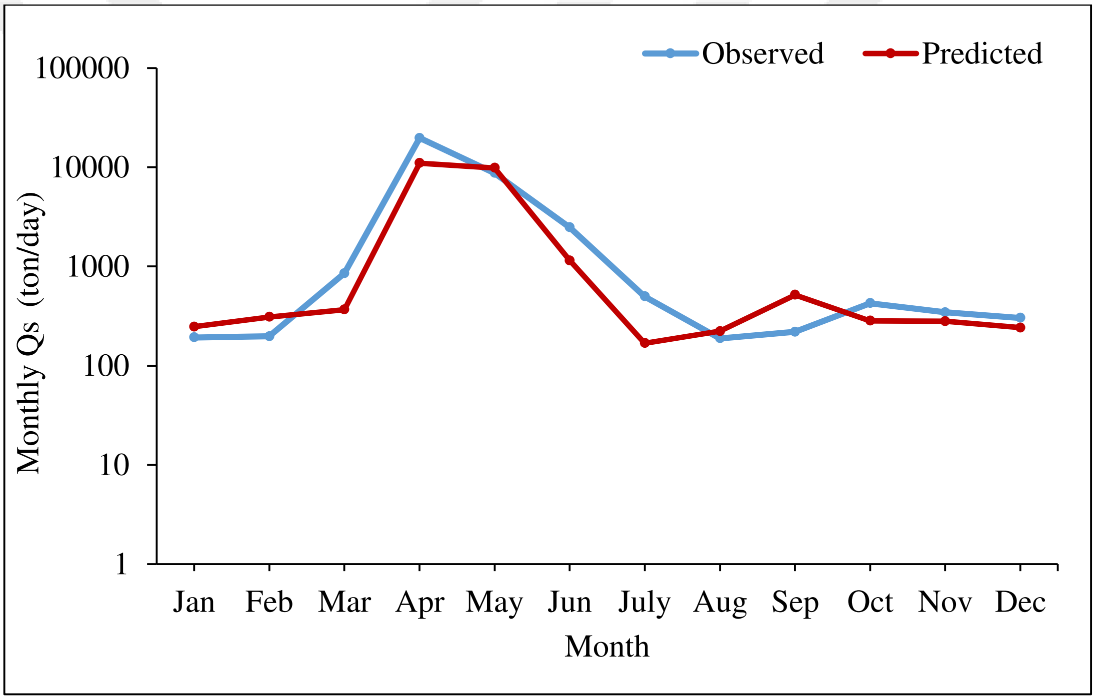

Figure 5.30 Observed and predicted monthly Qs for the validation period by LGP model in logarithmic scale

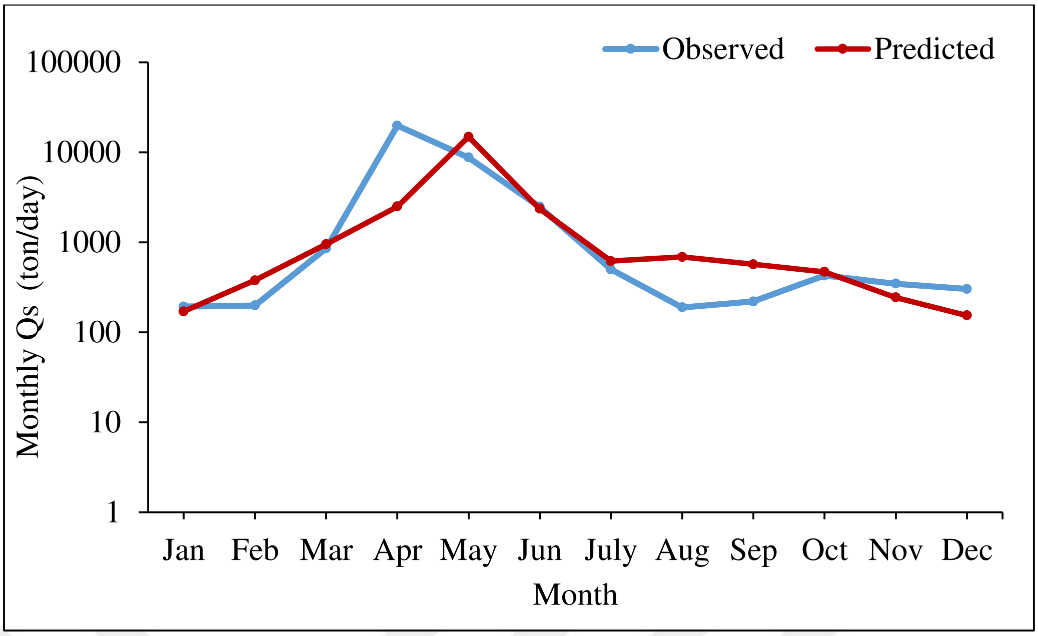

Figure 5.31 Observed and predicted monthly Qs for the validation period by GEP model in logarithmic scale

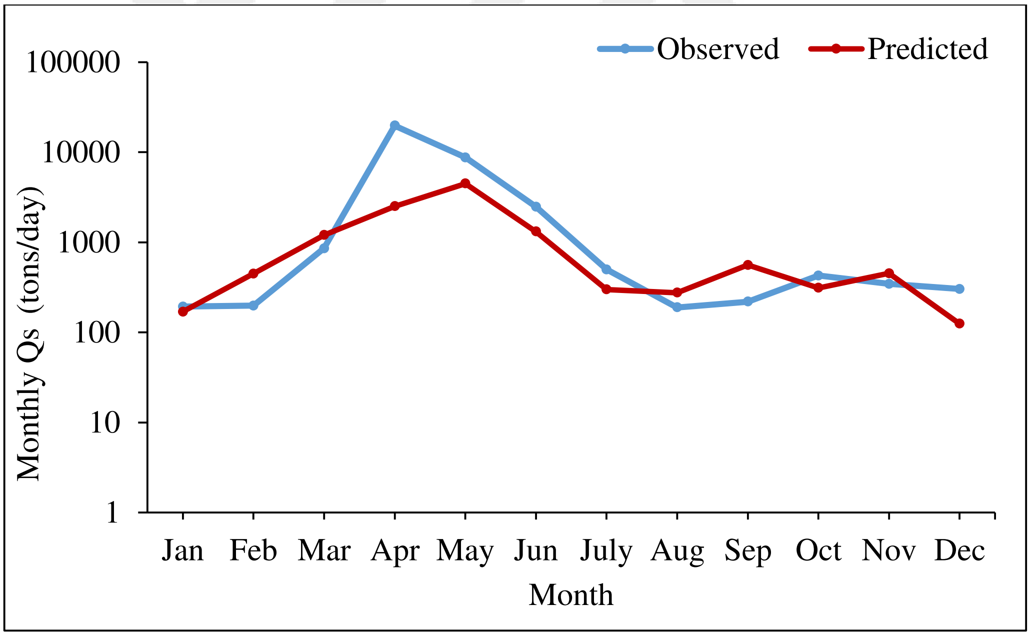

Figure 5.32 Observed and predicted monthly Qs for the validation period by Linear Regression model in logarithmic scale

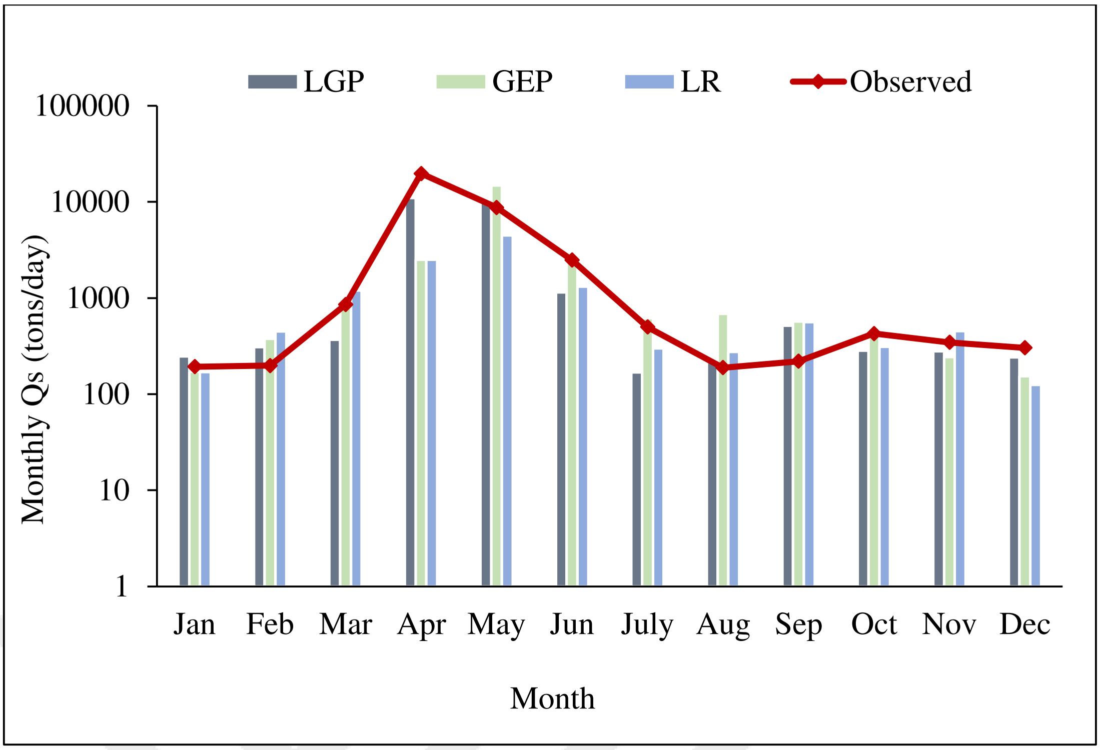

Figure 5.33 Observed and predicted monthly Qs for the validation period by all models in logarithmic scale

5.5 Future projection of discharge under different emission scenarios

Different output results were obtained from GEP and LGP models for three different scenarios, RCP2.6, RCP4.5, and RCP8.5. The outputs of the models for all three scenarios (horizontal axis) were compared to observed values (vertical axis) for the period 2006-2020, and scatterplots were created (Figures 5.34-5.39). It was determined that the R values were 0.408 for RCP2.6, 0.398 for RCP4.5, and 0.4585 for RCP8.5 in the GEP model, and 0.278 for RCP2.6, 0.347 for RCP4.5, and 0.351 for RCP8.5 in the LGP model. Based on these values, it is seen that the GEP model provides the best correlation for discharge (Q) in the RCP8.5 scenario. The GEP model has been superior in comparison to the LGP model by providing higher correlations in all scenarios. It is also observed that RCP8.5 performs better than other scenarios in both models. Among the data generated by the model, some values are exceedingly high, to the extent that they cannot be deemed as extreme, and have been considered outliers and removed from the dataset due to their disruptive impact on comparisons, as they are thousands or even tens of thousands of times higher than the average. These values constitute 1.4% of the total data. An outlier is an observation that deviates so much from other observations as to arouse suspicion that it was generated by a different mechanism (Hawkins, 1980).

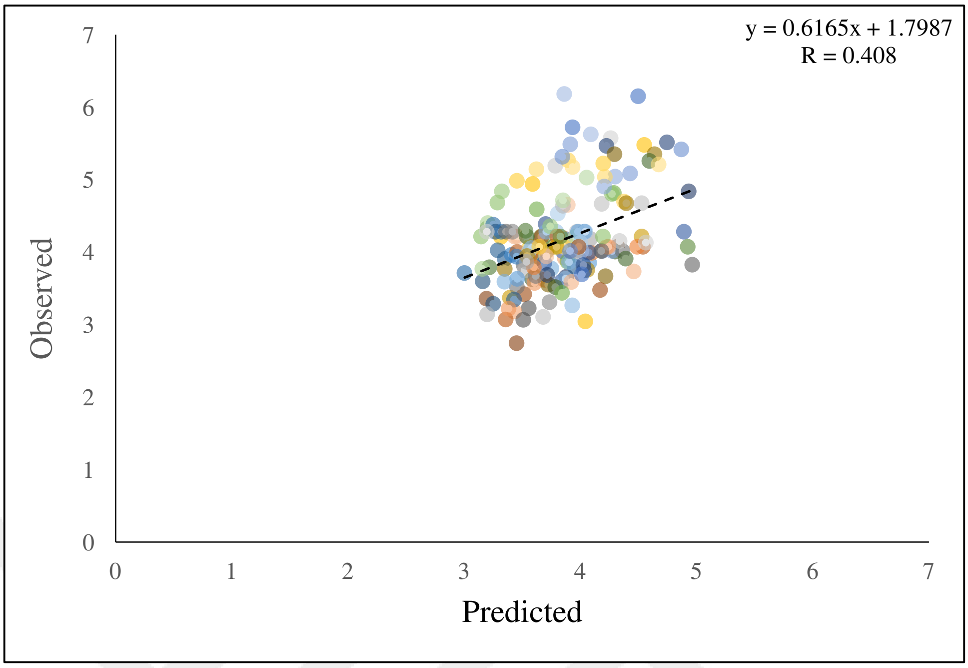

Figure 5.34 Scatterplot of the observed and projected discharge under CanESM2 RCP2.6 scenario by GEP model between 2006 – 2020 period

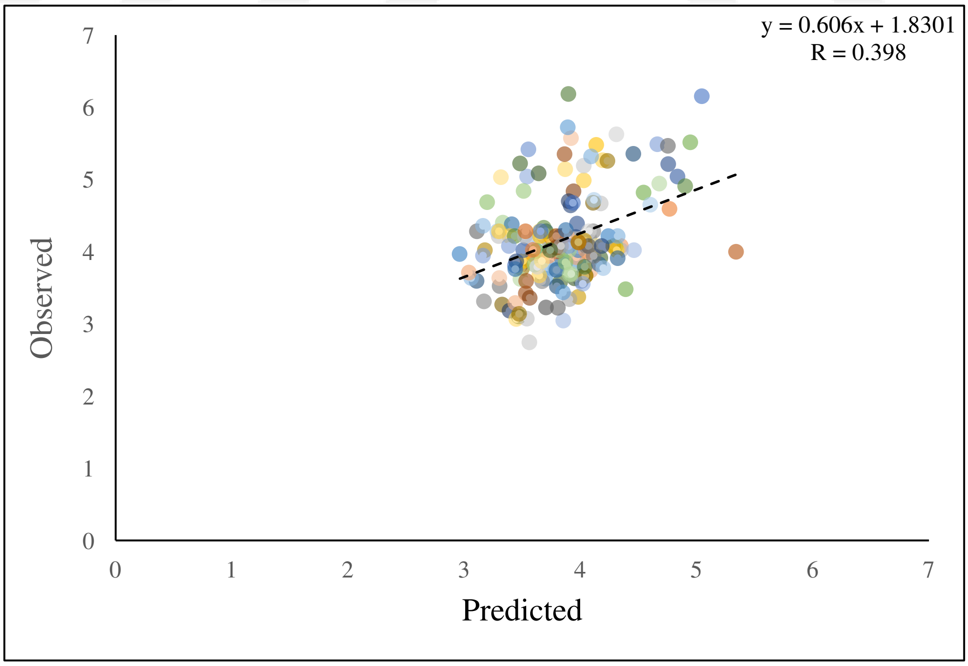

Figure 5.35 Scatterplot of the observed and projected discharge under CanESM2 RCP4.5 scenario by GEP model between 2006 – 2020 period

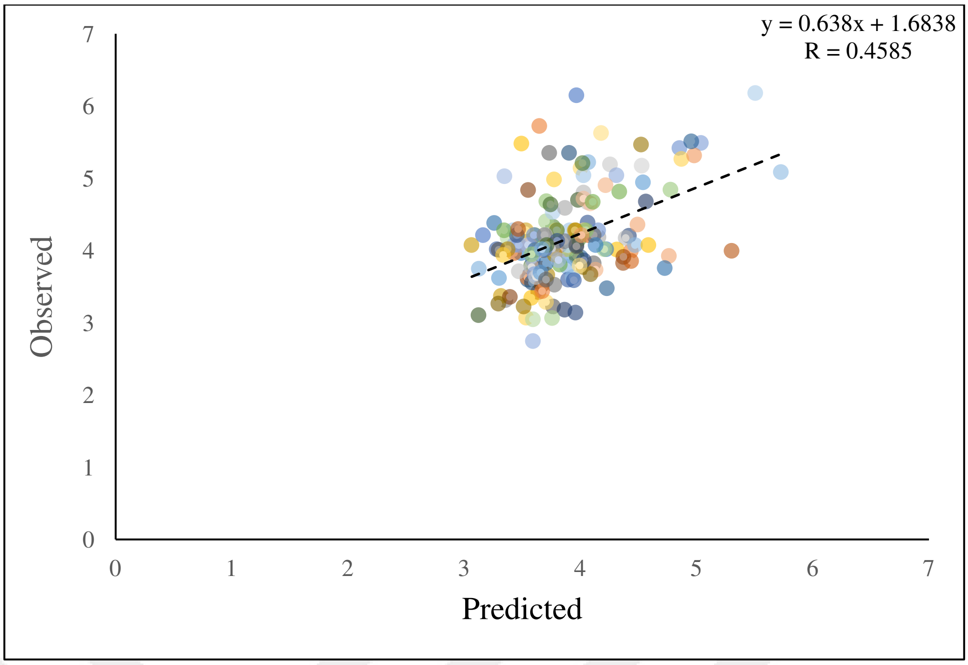

Figure 5.36 Scatterplot of the observed and projected discharge under CanESM2 RCP8.5 scenario by GEP model between 2006 – 2020 period

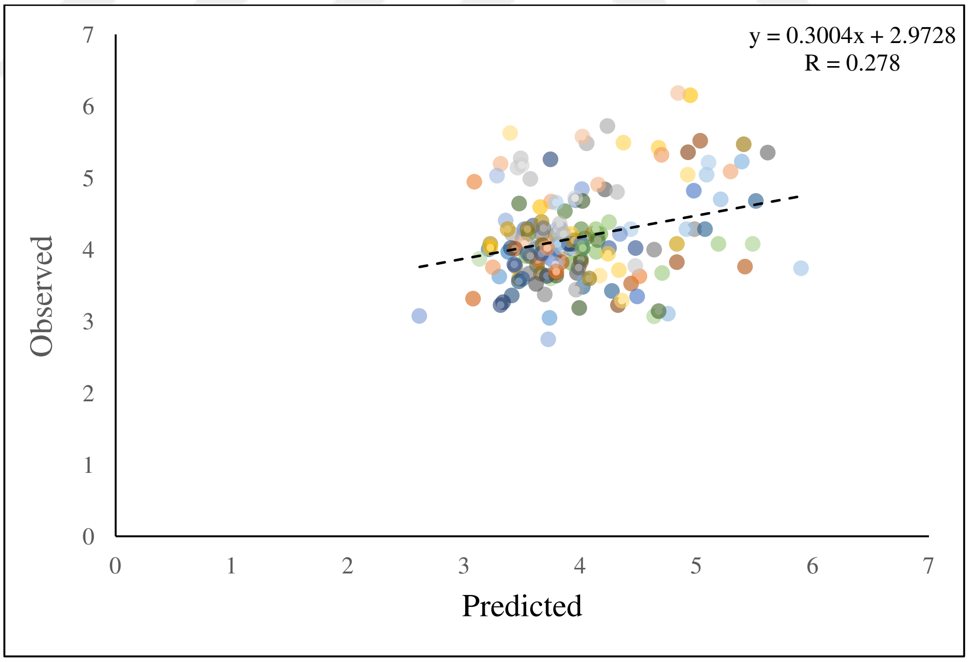

Figure 5.37 Scatterplot of the observed and projected discharge under CanESM2 RCP2.6 scenario by LGP model between 2006 – 2020 period

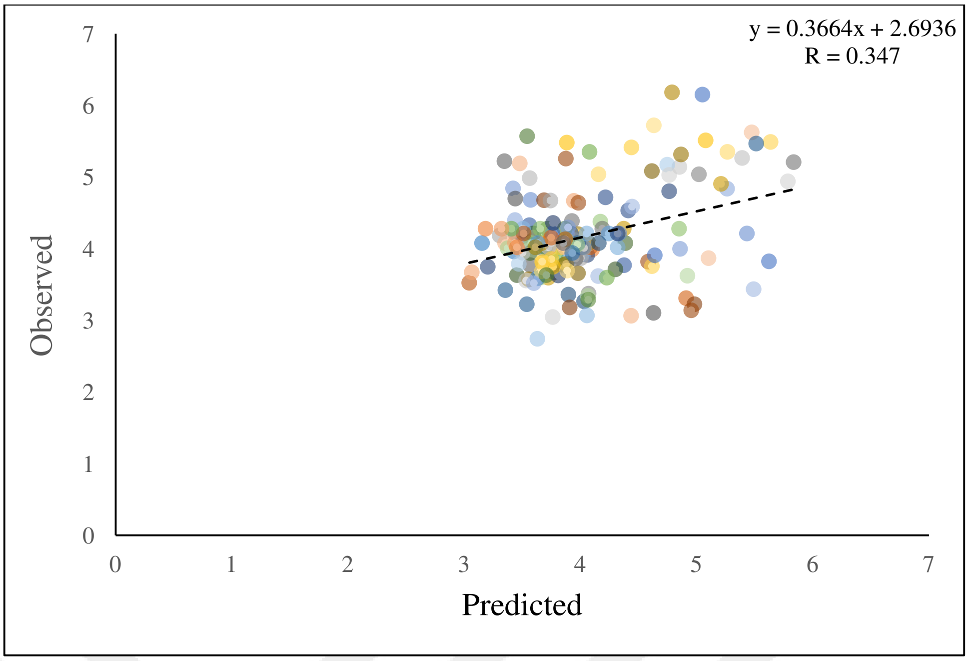

Figure 5.38 Scatterplot of the observed and projected discharge under CanESM2 RCP4.5 scenario by LGP model between 2006 – 2020 period

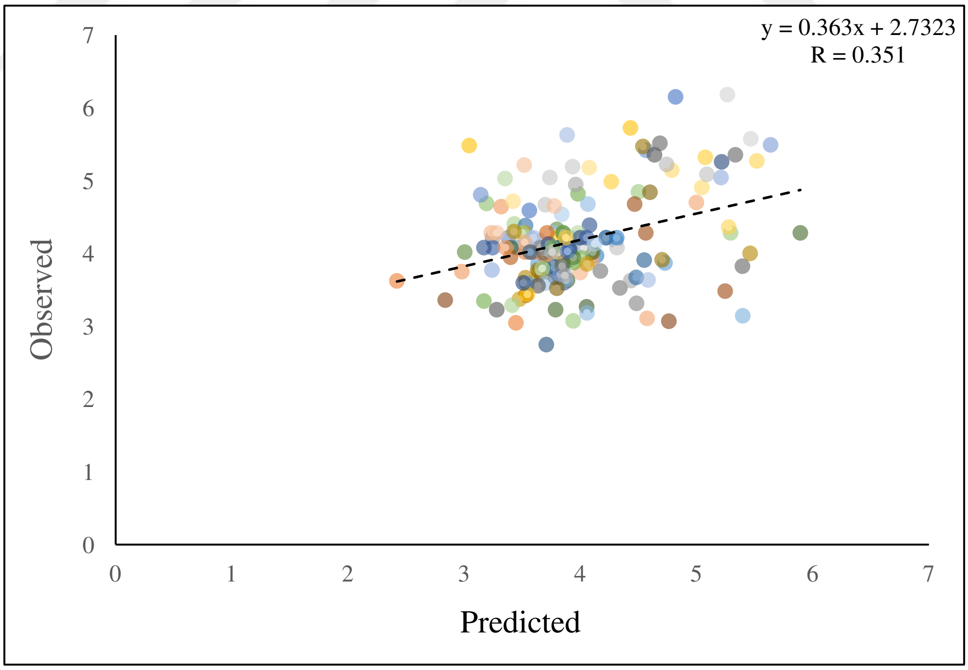

Figure 5.39 Scatterplot of the observed and projected discharge under CanESM2 RCP8.5 scenario by LGP model between 2006 – 2020 period

Figure 5.40 shows a comparison graph for the GEP RCP8.5 scenario, which has the best correlation between the projected 2006-2020 and future observed 2006-2020 periods. The comparison graphs between predicted and observed discharge values for the 2006-2020 period are listed in Appendix D for all models and scenarios.

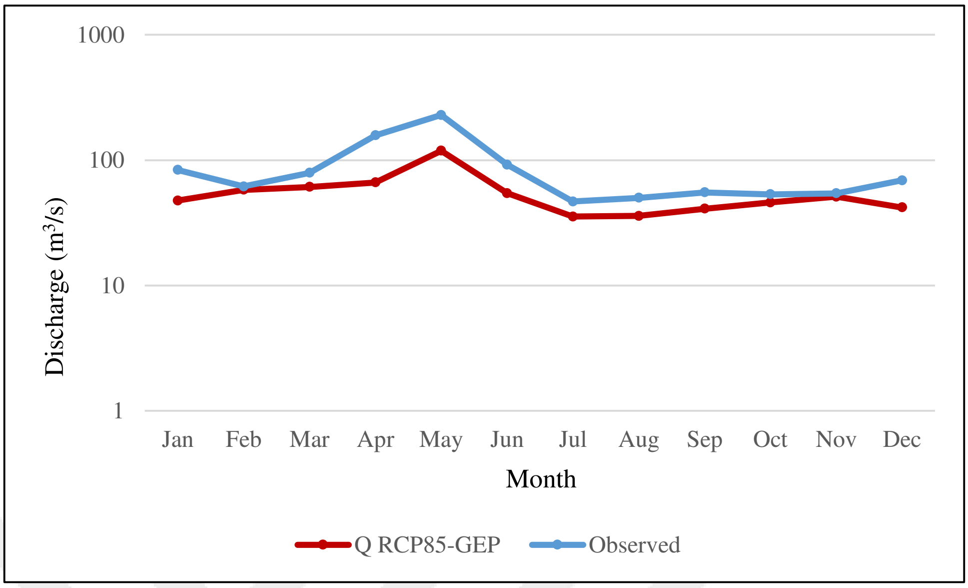

Figure 5.40 Observed and predicted monthly Q for the 2006-2020 period by GEP model under RCP8.5 scenario, in logarithmic scale

For all scenarios, downscaled results were divided into four time periods: the 2020s (2021-2040), 2040s (2041-2060), 2060s (2061-2080), 2080s (2081-2100), and were compared to two observed data periods which are observed calibration and validation periods combined (1970-2005) and future observed period (2006-2020). The average values for each month of the results of the RCP8.5 scenario of the GEP model, which showed the best performance for discharge, have been separately calculated for these four different periods and compared with observed data in Figure 5.41. In general, discharge values tend to be higher during April and May across all periods and observed data. The difference between the observed and other periods is also the highest in these months. The increment and decrement rates between critical months provide further insight into the variations in discharge values across the periods, offering a comprehensive understanding of the water flow dynamics. For all projected periods (2006-2020, 2020s, 2040s, 2060s and 2080s), the increment rates are critical between the months April-May (79.17%, 90.56%, 31.78%, 11.88%, 28.39%) and for the observed periods (1970-2005, 2006-2020) the increment rates are critical between March-April by 248.1% and 72.63%, respectively. Decrement rates between months are critical between May and June for all projected periods and observed periods by

54.11%, 56.47%, 45.74%, 29.40%, 59.84% and 56.84%, 49.66%. Figure 5.41 also shows that in all months, in every period of the GEP model projected, discharge values are lower than the future observed (2006-2020) values. When compared with the observed data between 1970-2005, it has been observed that predicted discharge values are lower in the GEP model for the months of April, May, June, July, November, and December. Conversely, in January, February, August, September, and October, the opposite situation is observed, with the 1970-2005 observed period staying lower than the values of the GEP model for all periods. Comparing the projected 2006-2020 period with the 2006-2020 observed data, we see that the predicted values are always lower than the observed values, particularly in April, May, and June. When comparing the periods of the GEP model under the RCP8.5 scenario, the 2020s tend to have higher discharge values, followed by the 2060s. The 2040s and 2080s generally show lower discharge values. The most significant average difference between the observed period of 2006-2020 and GEP models is in May, with percentage differences of 63.35%, 28.75%, 87.55%, 98.38% and 44.5%. Comparing with the observed period of 1970-2005, percentage differences are 109.18%, 57.34%, 107.46%, 105.64%, and 53.44% for April. The projection of average monthly discharge values for all models under every scenario of CanESM2 is listed in Appendix E.

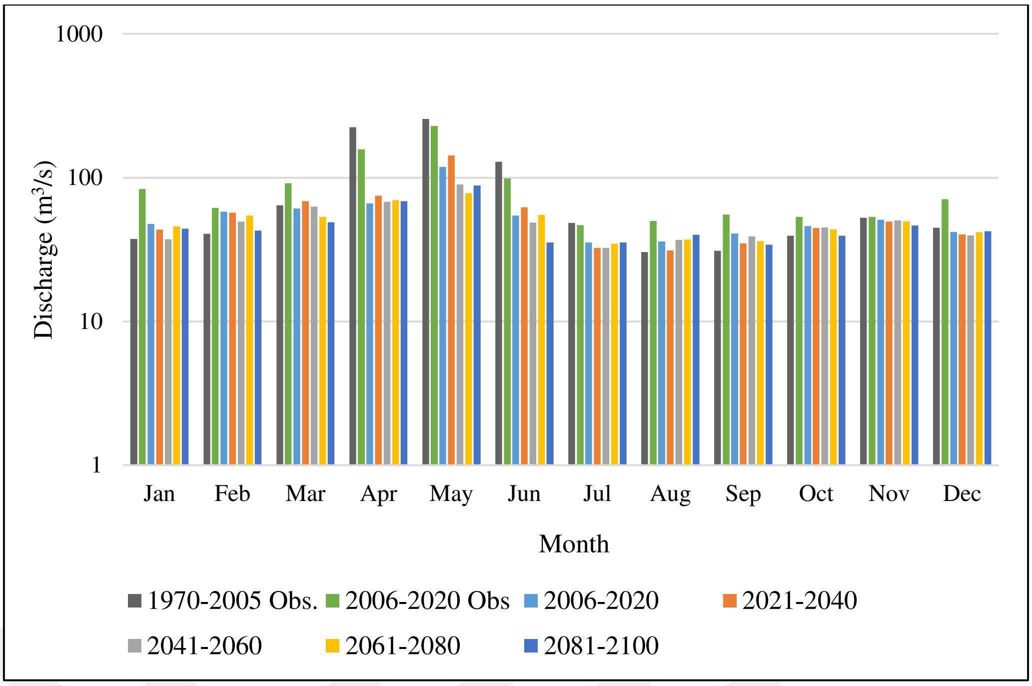

Figure 5.41 The projection of average monthly discharge Q values under CanESM2 RCP8.5 scenario for different periods by the GEP model in logarithmic scale

Figure 5.42 displays annual average discharge data between 1970 and 2100. The continuous data in the 1970-2100 period pertains to the GEP model RCP8.5 scenario, while the observed data is separately shown for the 1970-2020 period. Although no specific pattern is displayed, the observed data values are generally higher. The lowest annual average during the observed period (1970-2020) was 49.741 m^3/s, which occurred in 1994. The highest annual average during this period was 140.440 m^3/s in 1988. The lowest predicted annual average for the GEP model is 37.228 m^3/s, which is expected to occur in 2097. The highest predicted annual average in this period was 101.193 m^3/s in 1999.

Figure 5.43 displays the annual average discharge data values for May, which has been identified as the critical month due to having the highest values in all observed and simulated periods. Similar to Figure 5.42, the continuous data representing the 1970-2100 period pertains to the GEP model RCP8.5 scenario and is exhibited in the same graph as the observed data for 1970-2020. Generally, between 1970 and 2020, the observed data appears to exceed the predicted data. For the observed period, the highest value for May is 483.234 m^3/s, which belongs to the year 2019. The smallest value during this period is 56.918 m^3/s, recorded in 1989. In the data pertaining to the

GEP RCP8.5 scenario, the highest value is 585.881 m^3/s, observed in 1998, while the lowest value is 22.281 m^3/s, observed in 2047.

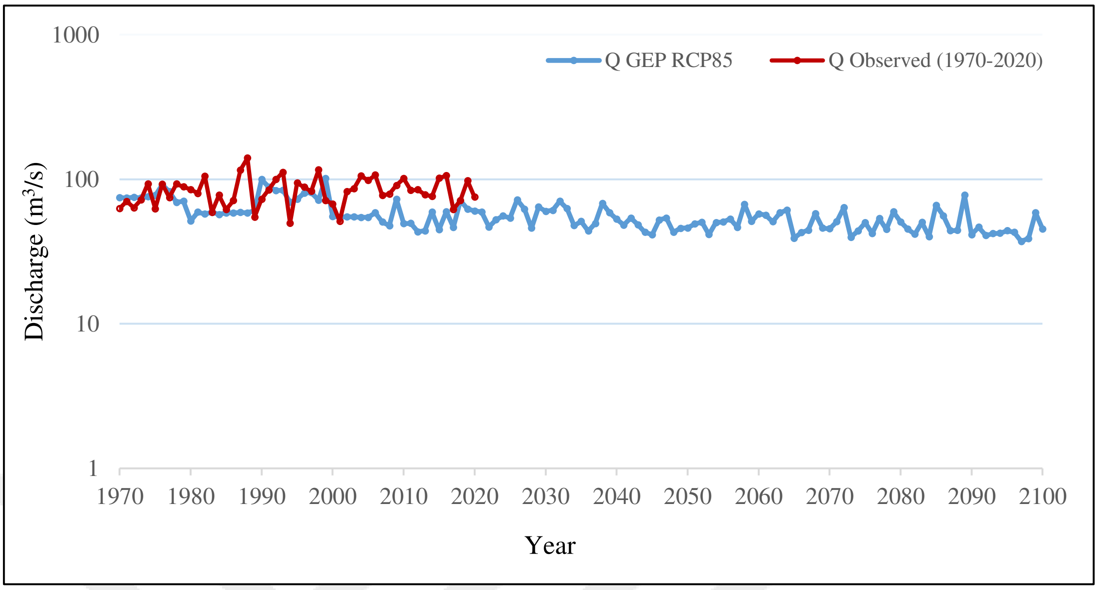

Figure 5.42 The projection of annual average discharge (Q) values under CanESM2 RCP8.5 scenario by GEP model and observed period of 1970-2020 in logarithmic scale

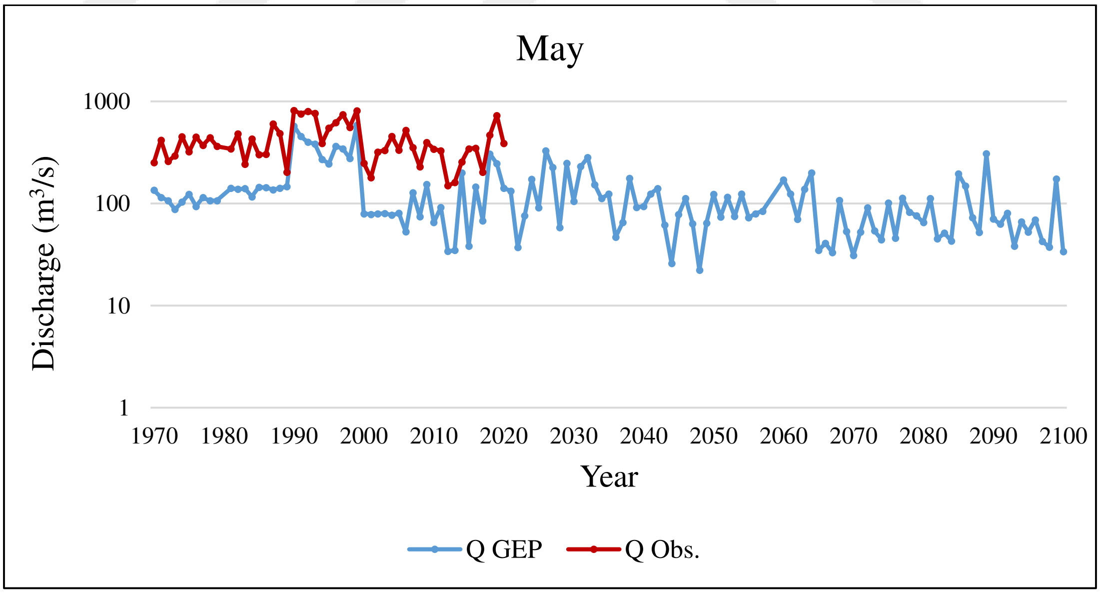

Figure 5.43 The projection of annual average discharge (Q) values under CanESM2 RCP8.5 scenario by GEP model and observed period of 1970-2020 for May in logarithmic scale

5.6 Future projection of sediment yield under different emission scenarios

Different output results were obtained from GEP and LGP models for three different scenarios, RCP2.6, RCP4.5, and RCP8.5. The outputs of the models for all three scenarios (horizontal axis) were compared to observed values (vertical axis) for the period 2006-2020, and scatterplots were created (Figures 5.44-5.49). It was determined that the R values were 0.303 for RCP2.6, 0.3795 for RCP4.5, and 0.427 for RCP8.5 in the GEP model, and 0.2915 for RCP2.6, 0.3905 for RCP4.5, and 0.377 for RCP8.5 in the LGP model. Based on these values, it is seen that the GEP model provides the best correlation for sediment yield in the RCP8.5 scenario than all scenarios in both models. For sediment yield, 2.46% of data for the GEP RCP8.5 scenario is considered outliers and removed from the dataset.

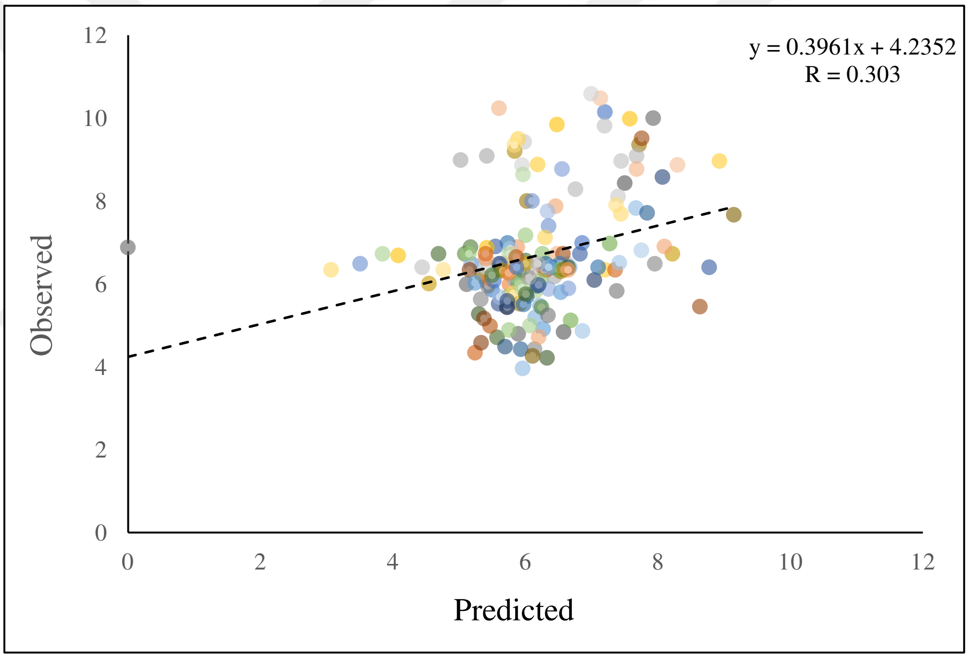

Figure 5.44 Scatterplot of the observed and projected sediment yield under CanESM2 RCP2.6 scenario by GEP model between 2006 – 2020

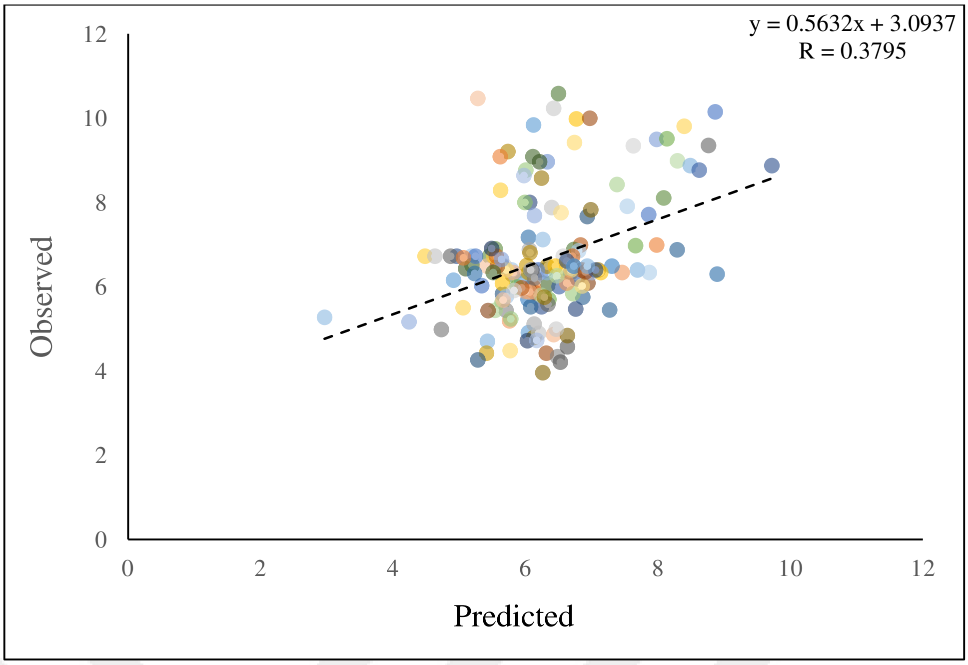

Figure 5.45 Scatterplot of the observed and projected sediment yield under CanESM2 RCP4.5 scenario by GEP model between 2006 – 2020

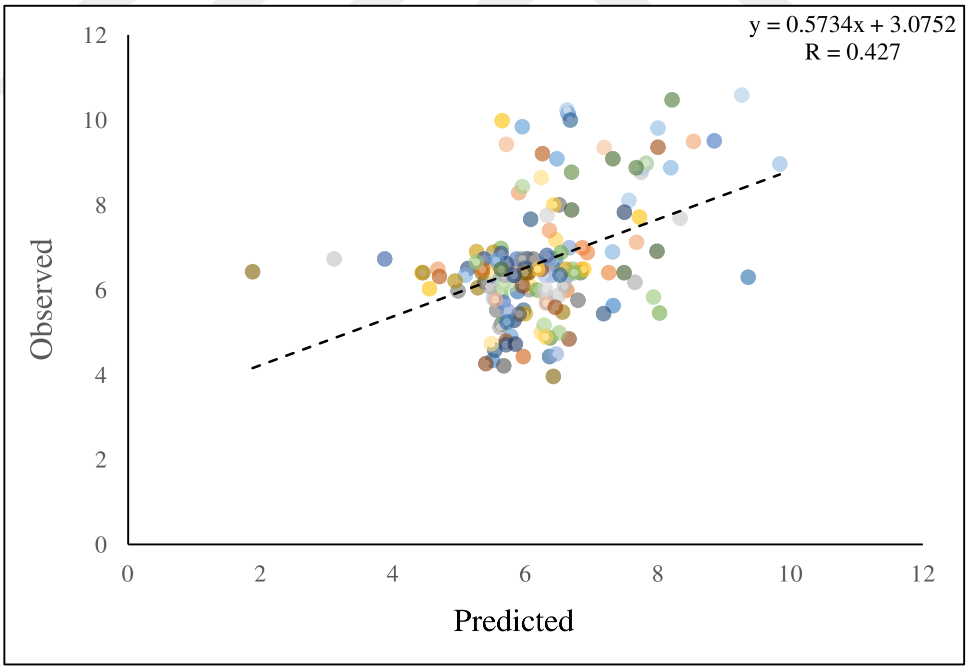

Figure 5.46 Scatterplot of the observed and projected sediment yield under CanESM2 RCP8.5 scenario by GEP model between 2006 – 2020

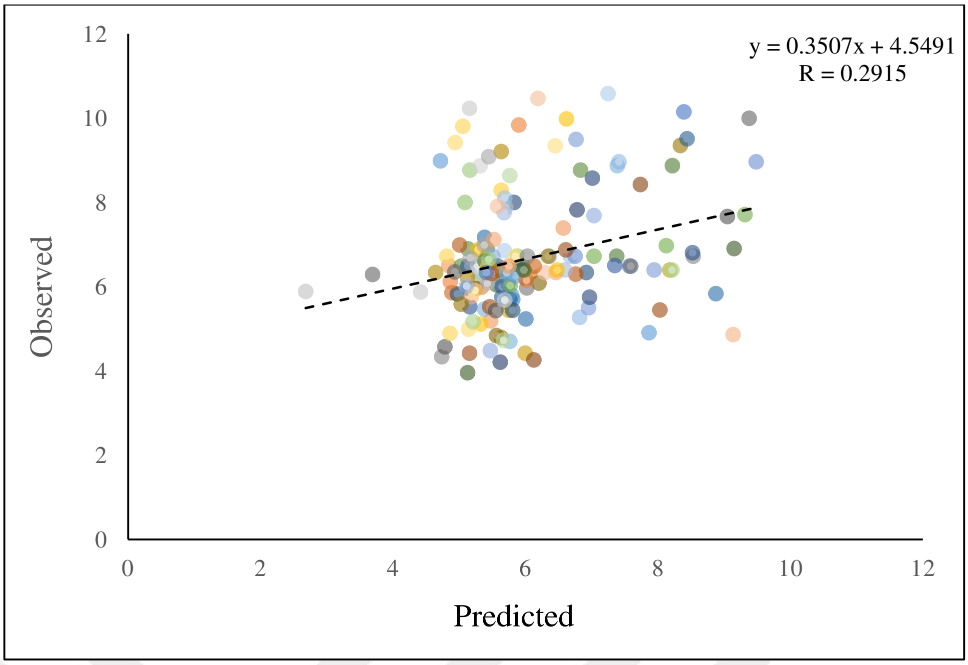

Figure 5.47 Scatterplot of the observed and projected sediment yield under CanESM2 RCP2.6 scenario by LGP model between 2006 – 2020 period

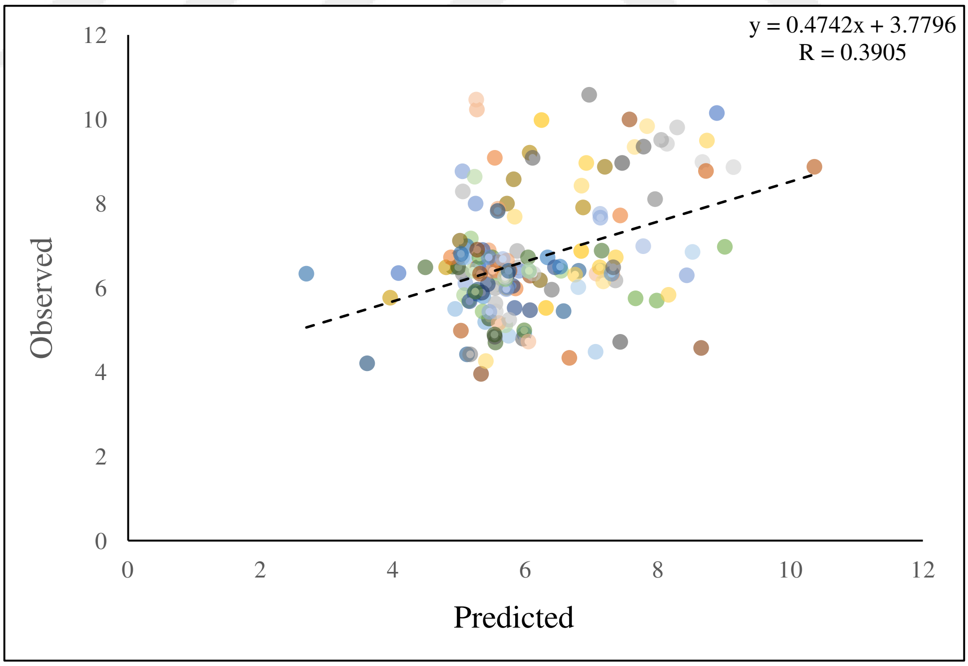

Figure 5.48 Scatterplot of the observed and projected sediment yield under CanESM2 RCP4.5 scenario by LGP model between 2006 – 2020

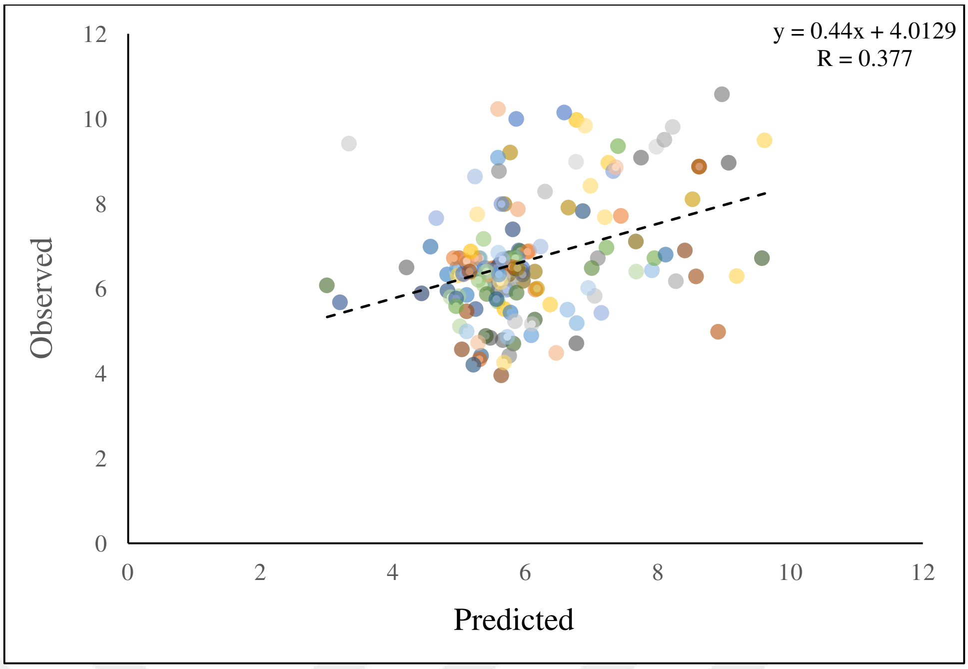

Figure 5.49 Scatterplot of the observed and projected sediment yield under CanESM2 RCP8.5 scenario by LGP model between 2006 – 2020

Figure 5.50 shows a comparison graph for the GEP RCP8.5 scenario, which has the best correlation between the projected 2006-2020 and future observed 2006-2020 periods. The comparison graphs between predicted and observed sediment yield values for the 2006-2020 period are listed in Appendix F for all models and scenarios.

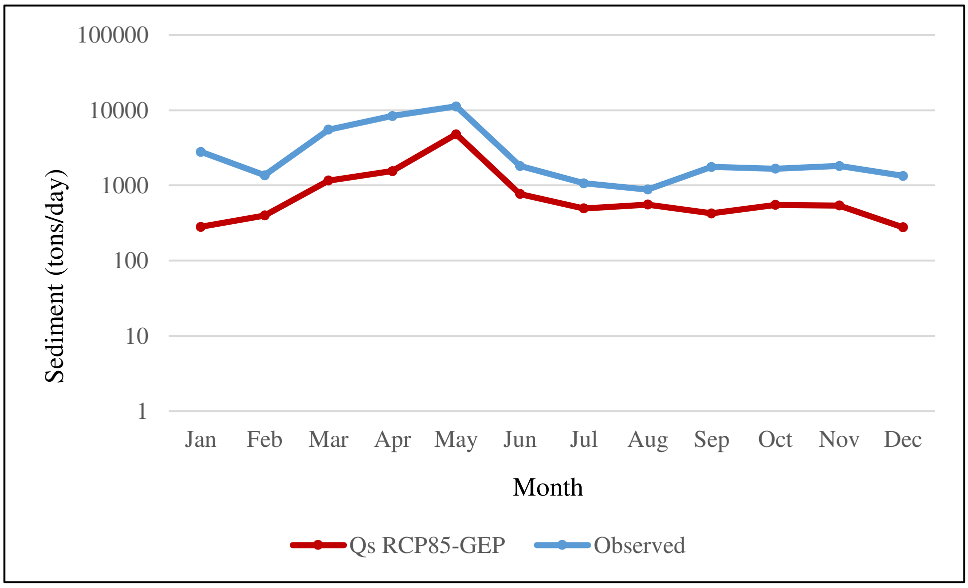

Figure 5.50 Observed and predicted monthly Qs for the 2006-2020 period by GEP model under RCP8.5 scenario, in logarithmic scale

Downscaled results were divided into four time periods: the 2020s (2021-2040), 2040s (2041-2060), 2060s (2061-2080), 2080s (2081-2100), and were compared to observed data for the periods 1970-2005 and 2006-2020. The average values of each month for the results of the GEP model under the RCP8.5 scenario, showing the best performance for sediment yield, have been calculated for four different periods and compared with observed data (2006-2020) in Figure 5.51.

It is observed that the months with the highest data values for all periods are March, April, May, and June. Sediment yield values, which increase with a mild slope from January to March, exhibit a significant rise during March-April and April-May. For the observed periods of 1970-2005 and 2006-2020, and the projected periods of 2006-2020, the 2020s, 2040s, 2060s, and 2080s, the increase percentages are as follows: for March-April, sequentially, 33.48%, 65.9%, 43.43%, 89.59%, 189.73%, 51.13%, and 1658.16%; for April-May, sequentially, 208.93%, 176.72%, 114.99%, 57.99%, 126.83%, 34.77% increase, and 41.9% decrease. A significant decline is also observed from May to June (83.94%, 61.24%, 70%, 40.73%, 70.09%, 81.36%, 80.11%). Values that decrease from June to July experience no major fluctuations in the subsequent months. The results of the sediment yield predictions, divided into 20-year periods and compared with the observed values between 2006-2020, are clearly below the observed values for each month, as shown in Figure 5.51. The difference is particularly dramatic in March and April, with the percentage difference between observed (2006-2020) and predicted values being 163.35%, 85.97%, 171.66%, and 164.73%, 79.46% for January and 130.62%, 72.58%, 141.24%, 151.7%, 82.64% for March for the projected periods, 2006-2020, 2020s, 2040s, 2060s, 2080s, respectively. When comparing the observed period of 1970-2005 with the simulated values, the differences between the average monthly values of observed values and the average monthly values of simulated data are observed to be very high in April and May. These values have been calculated as percentage differences for April, sequentially, at 180.52%, 91.12%, 182.73%, 181.83%, 90.42%, and for May at 114.43%, 61.95%, 142.76%, 154.18%, 67.17%. The projection of average monthly sediment yield values for all models under every scenario of CanESM2 is listed in Appendix G.

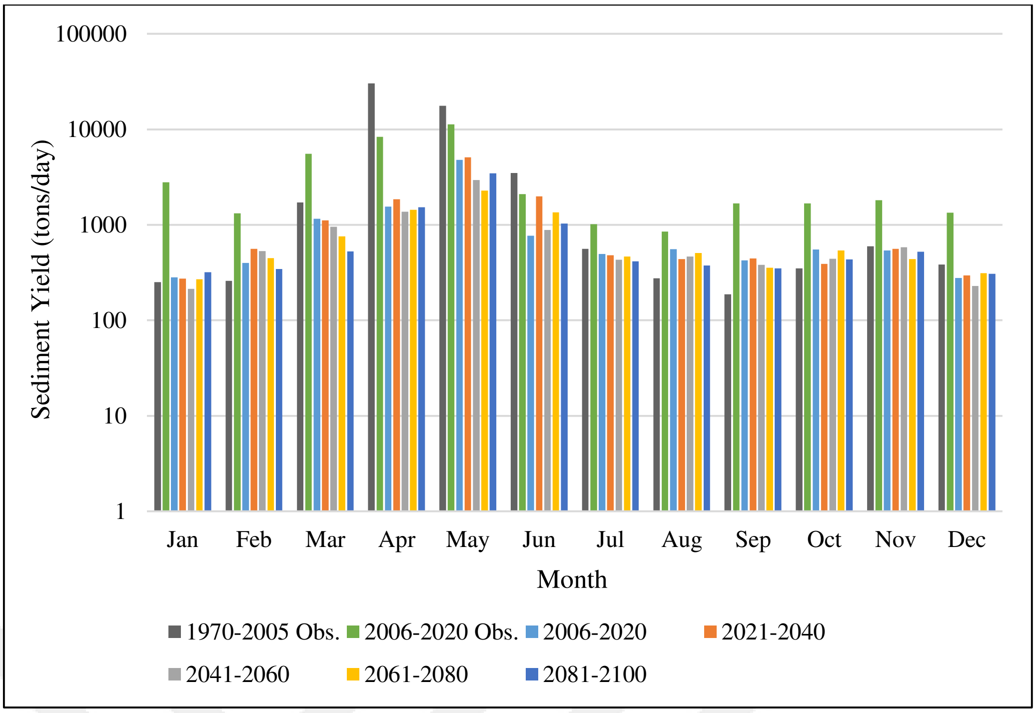

Figure 5.51 The projection of average monthly sediment yield values under CanESM2 RCP8.5 scenario for different periods by the GEP model in logarithmic scale

Figure 5.52 displays the annual average sediment yield (Qs) data of GEP RCP8.5 between 1970-2100 and observed data between 1970-2020. It is understood that the observed data is well above the data projected by the GEP model under the RCP8.5 scenario in most years of the 1970-2020 period. The highest annual average of the period that includes the observed data is 17386.910 tons/day, which belongs to 1974 and differs significantly from other years. The lowest annual average value of 569,717 tons/day in the same period belongs to 2001. In the annual averages of the data between 1970-2100, which is the period estimated by the RCP8.5 scenario of the GEP model, the lowest value belongs to the year 2097 with 293,755 tons/day, while the highest annual average belongs to the year 1976 with 9918.332 tons/day. It has also been observed that a slightly decreasing trend dominates the entire graph.

Figure 5.53 shows the annual average values of sediment yield for May, which has been considered a critical month due to having the highest values in all of the simulated periods of 2006-2020, 2020s, 2040s, 2060s and 2080s. Generally, it is observed that between 1970 and 2020, with the exceptions of the years 1990 and 1999, the observed data is significantly higher than the predicted data. In May, the observed sediment yield reached its highest value in 1974, with 69442.1 tons/day, and its lowest in 1989, with 399.1 tons/day. The highest value obtained by the GEP model under the RCP8.5 scenario between 1970 and 2100 was 80179.853 tons/day in 1999, while the lowest sediment yield value is 20.822 tons/day, which is expected to be recorded in 2049.

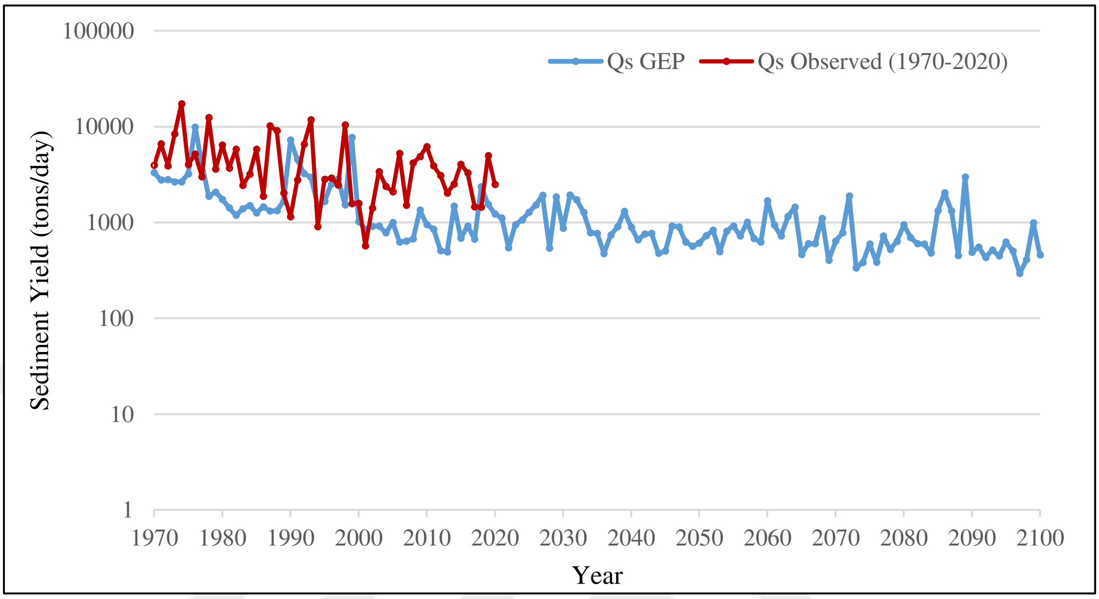

Figure 5.52 The projection of annual average sediment yield values under CanESM2 RCP8.5 scenario by GEP model and observed period of 1970-2020 in logarithmic scale

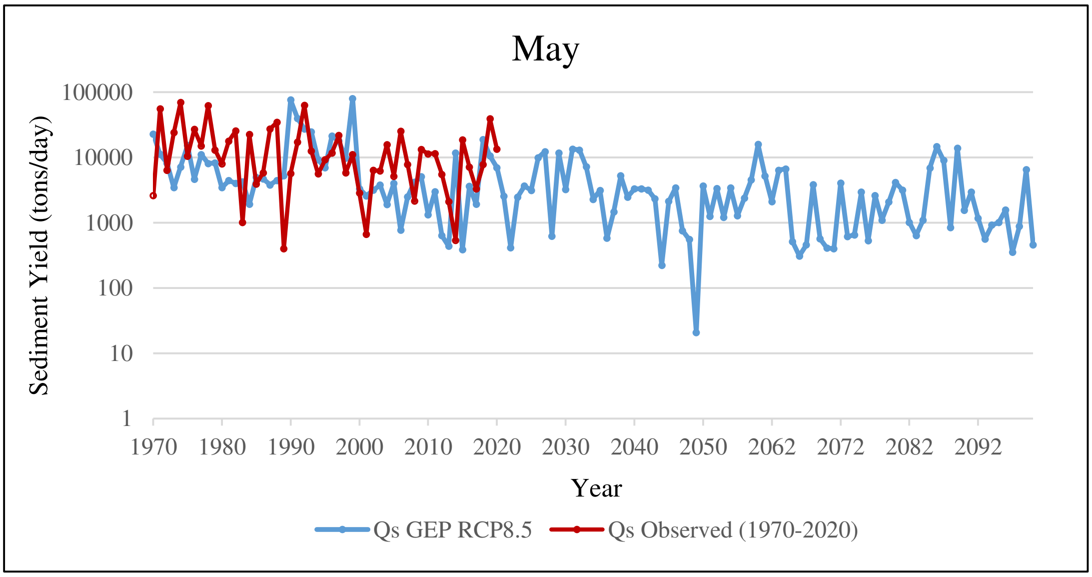

Figure 5.53 The projection of annual average sediment yield values under CanESM2 RCP8.5 scenario by GEP model and observed period of 1970-2020 for May, in logarithmic scale

CHAPTER 6
CONCLUSION AND RECOMMENDATIONS

6.1 Conclusion

Climate change presents a critical challenge in hydrology, as it significantly impacts the behaviour and characteristics of water resources, necessitating adaptive strategies to safeguard ecosystems, human communities, and water infrastructure. This thesis examines climate change and its effects on a watershed located on the Euphrates River in the northeastern region of Türkiye through hydrological measurements. The investigation was carried out using different climate models and statistical methods. Firstly, the watershed was delineated using WMS software, and the impact area was determined based on the outlet of the watershed. The statistical downscaling method and GCM were used. Downscaling techniques can translate the GCM outputs from their coarse spatial resolution into a finer resolution that is more appropriate for forecasting climate change at the local scale. The CanESM2 model was selected to generate future projections at the local scale.

Inputs consisting of 26 variables belonging to the CanESM2 model were downloaded based on the location of the centroid of the watershed. To observe the relationship between the downloaded variables and the regional discharge (Q) and sediment yield (Qs) data obtained from the general directorate of state hydraulic works and to determine the most effective variables, a correlation analysis was conducted using the "Pearson rank correlation coefficient" method. After the analysis, the ten most effective variables were selected from the 26 variables, and normalisation methods were used to obtain better correlations. Among these methods, the natural logarithm method showed the best performance.

The GEP and LGP artificial intelligence models were used for downscaling and monthly predictions of discharge and sediment yield data in the basin. The prediction results were compared, and linear regression analysis was used for comparison. Both discharge (Q) and sediment yield (Qs) data were analysed under three different scenarios, RCP2.6, RCP4.5, and RCP8.5, using GEP and LGP models with CanESM2 for future predictions.

The LGP model provides correlation (R) values of 0.883 for discharge (Q) and 0.837 for sediment yield (Qs) during the calibration period (1970-1995), and 0.841 for Q and 0.788 for Qs during the validation period (1996-2005), while the GEP model provides correlation values of 0.645 for Q and 0.751 for Qs during the calibration period, and 0.684 for Q and 0.673 for Qs during the validation period. This comparison shows that the LGP model exhibits significantly better performance than the GEP model for both discharge (Q) and sediment yield(Qs) during the aforementioned periods.

Outputs from the LGP and GEP models were obtained for the RCP2.6, RCP4.5, and RCP8.5 scenarios between 2006 and 2100 for future projections. The outputs between 2006 and 2020 were compared with the observed data for the same period, and the results were presented with scatterplots. In these comparisons, the projections of the GEP model were closer to the observed values than those of the LGP model for all three scenarios. The RCP8.5 scenario mostly provided the best results for the LGP and GEP models.

The obtained projection results were divided into 20-year periods, namely; 2020s (2021-2040), 2040s (2041-2060), 2060s (2061-2080), 2080s (2081-2100) and monthly averages discharge and sediment values were graphically compared with each other and with the observed monthly average values from 2006 to 2020. It was clearly observed that the monthly average values of these four periods (2020s, 2040s, 2060s, 2080s) were consistently lower than the monthly average values of the observed period (2006-2020) for both discharge (Q) and sediment (Qs). The largest difference for discharge (Q) was observed in April and May. Furthermore, for discharge (Q), it was observed that there was a slightly increasing trend from January to April in all periods (the 2020s, 2040s, 2060s, and 2080s), a more pronounced increase from April to May, followed by a slight decrease from May to July, and then no significant change afterwards. Additionally, for discharge (Q), the GEP model indicated the lowest values in July and the highest in May, according to the RCP8.5 scenario. Similarly, for sediment yield (Qs), a slight increase was observed between January and May, followed by a slight decrease from May to July. Again, no significant trend was observed after July. The lowest values for sediment yield were observed in January and the highest in May.

In order to view all observed and predicted years together, the values observed between 1970 and 2020 and the values predicted by the GEP model under the RCP8.5 scenario between 2021 and 2100 have been used. These are annual average values covering all results from 1970 to 2100. For discharge (Q), there was a rise from 1970 to 1988, with the maximum value observed in 1988 at 140.440 m^3/s. The trend then decreased, with the lowest value observed in 1994 at 49.741 m^3/s. While the general trend decreased from 2021 to 2100, there was an exception in 2089, where the average value reached a maximum of 77.737 m^3/s. The year with the lowest average value for the predicted period under the GEP model RCP8.5 scenario was 2097 (37.228 m^3/s). When the same comparison is made for sediment yield (Qs), the minimum value was observed in 2001 at 561.717 tons/day, following a decreasing trend after the peak value observed in 1974 at a whopping 17386.910 tons/day until 2020. The annual averages for 2021 to 2100 predicted by the GEP model RCP8.5 scenario showed the lowest value in again, 2097 at 293.77 tons/day, while the highest value of 2975 tons/day was observed in 2089, which was the same year as the maximum value for Q. In the annual average values of the data projected by the GEP model RCP8.5 scenario, the general trend of sediment yield decreases slightly between 2021 – 2100.

6.2 Recommendations

1. The impact of climate change on hydropower potential can be assessed using observed and projected data, which can facilitate the engineer in the construction of the hydraulic structures.
2. Monthly data is used in this study. Daily data can be suggested to use if available.
3. Data such as precipitation, temperature, humidity, wind speed and direction, solar radiation can be included as predictands in the study, or a combination of these data sets can be used by employing the models used in the study or recommended above.
4. In this research, only CanESM2 predictors were used to predict the discharge and sediment yield data. It can be suggested to consider other GCM models like Coupled Model Intercomparison Project Phase 6 (CMIP6) predictors,

which include CanESM5 (high ECS), MPI-ESM1.2-HR (medium ECS), and NorESM2-MM (low ECS) or NCEP variables.

REFERENCES

Abdo, K. S., Fiseha, B. M., Rientjes, T. H. M., Gieske, A. S. M., & Haile, A. T. (2009). Assessment of climate change impacts on the hydrology of Gilgel Abay catchment in Lake Tana basin, Ethiopia. *Hydrological Processes: An International Journal*, *23*(26), 3661–3669.

Andréasson, J., Bergström, S., Carlsson, B., Graham, L. P., & Lindström, G. (2004). Hydrological change–climate change impact simulations for Sweden. *AMBIO: A Journal of the Human Environment*, *33*(4), 228–234.

Angeline, P. J. (1994). *Genetic programming: On the programming of computers by means of natural selection: John R. Koza, A Bradford Book, MIT Press, Cambridge MA, 1992, ISBN 0-262-11170-5, xiv+ 819pp., US $55.00*. Elsevier.

Arnell, N. W. (1999). Climate change and global water resources. *Global Environmental Change*, *9*, S31–S49.

Arnell, N. W. (2003). *Effect of climate change on river flows and groundwater recharge: UKCIP02 scenarios*.

Arnell, N. W., & Reynard, N. S. (1996). The effects of climate change due to global warming on river flows in Great Britain. *Journal of Hydrology*, *183*(3–4), 397–424.

Azim, F., Shakir, A. S., & Kanwal, A. (2016). Impact of climate change on sediment yield for Naran watershed, Pakistan. *International Journal of Sediment Research*, *31*(3), 212–219.

Browne, N. P., & dos Santos, M. V. (2010). Adaptive representations for improving evolvability, parameter control, and parallelization of gene expression programming. *Applied Computational Intelligence and Soft Computing*, *2010*, 1–19.

Bruyère, C. L., Done, J. M., Holland, G. J., & Fredrick, S. (2014). Bias corrections of global models for regional climate simulations of high-impact weather. *Climate Dynamics*, *43*, 1847–1856.

Bussi, G., Francés, F., Horel, E., López-Tarazón, J. A., & Batalla, R. J. (2014). Modelling the impact of climate change on sediment yield in a highly erodible Mediterranean catchment. *Journal of Soils and Sediments*, *14*, 1921–1937.

Caballero, Y., Voirin-Morel, S., Habets, F., Noilhan, J., LeMoigne, P., Lehenaff, A., & Boone, A. (2007). Hydrological sensitivity of the Adour-Garonne river basin to climate change. *Water Resources Research*, *43*(7).

Canadell, J. G., & Raupach, M. R. (2008). Managing forests for climate change mitigation. *Science*, *320*(5882), 1456–1457.

Cervi, F., Petronici, F., Castellarin, A., Marcaccio, M., Bertolini, A., & Borgatti, L. (2018). Climate-change potential effects on the hydrological regime of freshwater springs in the Italian Northern Apennines. *Science of the Total Environment*, *622*, 337–348.

Chiew, F. H. S., Kirono, D. G. C., Kent, D. M., Frost, A. J., Charles, S. P., Timbal, B., Nguyen, K. C., & Fu, G. (2010). Comparison of runoff modelled using rainfall from different downscaling methods for historical and future climates. *Journal of Hydrology*, *387*(1–2), 10–23.

Dahl, T. A., Kendall, A. D., & Hyndman, D. W. (2018). Impacts of projected climate change on sediment yield and dredging costs. *Hydrological Processes*, *32*(9), 1223–1234.

Dai, A., Qian, T., Trenberth, K. E., & Milliman, J. D. (2009). Changes in continental freshwater discharge from 1948 to 2004. *Journal of Climate*, *22*(10), 2773–2792.

Danandeh Mehr, A., Sorman, A. U., Kahya, E., & Hesami Afshar, M. (2020). Climate change impacts on meteorological drought using SPI and SPEI: Case study of Ankara, Turkey. *Hydrological Sciences Journal*, *65*(2), 254–268.

Diaz-Nieto, J., & Wilby, R. L. (2004). The impact of climate change on low flows in the River Thames UK: A comparison of statistical downscaling and change factors. *Climatic Change*.

Dibike, Y. B., & Coulibaly, P. (2005). Hydrologic impact of climate change in the Saguenay watershed: Comparison of downscaling methods and hydrologic models. *Journal of Hydrology*, *307*(1–4), 145–163.

Eckhardt, K., & Ulbrich, U. (2003). Potential impacts of climate change on groundwater recharge and streamflow in a central European low mountain range. *Journal of Hydrology*, *284*(1–4), 244–252.

Eyring, V., Bony, S., Meehl, G. A., Senior, C. A., Stevens, B., Stouffer, R. J., & Taylor, K. E. (2016). Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development*, *9*(5), 1937–1958.

Feng, L., Zhou, T., Wu, B., Li, T., & Luo, J.-J. (2011). Projection of future precipitation change over China with a high-resolution global atmospheric model. *Advances in Atmospheric Sciences*, *28*, 464–476.

Fernando, A. K., Shamseldin, A. Y., & Abrahart, R. J. (2012). Use of gene expression programming for multimodel combination of rainfall-runoff models. *Journal of Hydrologic Engineering*, *17*(9), 975–985.

Fischer, G., Tubiello, F. N., Van Velthuizen, H., & Wiberg, D. A. (2007). Climate change impacts on irrigation water requirements: Effects of mitigation, 1990–2080. *Technological Forecasting and Social Change*, *74*(7), 1083–1107.

Fistikoglu, O., & Okkan, U. (2011). Statistical downscaling of monthly precipitation using NCEP/NCAR reanalysis data for Tahtali River Basin in Turkey. *Journal of Hydrologic Engineering*, *16*(2), 157–164.

Fowler, H. J., Blenkinsop, S., & Tebaldi, C. (2007). Linking climate change modelling to impacts studies: Recent advances in downscaling techniques for hydrological modelling. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *27*(12), 1547–1578.

Fujihara, Y., Tanaka, K., Watanabe, T., & Kojiri, T. (2007). Assessing the impact of climate change on the water resources of the Seyhan River Basin, Turkey. *The Final Report of ICCAP*, 89–94.

Gao, Y., Vano, J. A., Zhu, C., & Lettenmaier, D. P. (2011). Evaluating climate change over the Colorado River basin using regional climate models. *Journal of Geophysical Research: Atmospheres*, *116*(D13).

Giorgi, F., & Gutowski Jr, W. J. (2015). Regional dynamical downscaling and the CORDEX initiative. *Annual Review of Environment and Resources*, *40*, 467–490.

Giorgi, F., Hostetler, S. W., & Brodeur, C. S. (1994). Analysis of the surface hydrology in a regional climate model. *Quarterly Journal of the Royal Meteorological Society*, *120*(515), 161–183.

Giorgi, F., Shields Brodeur, C., & Bates, G. T. (1994). Regional climate change scenarios over the United States produced with a nested regional climate model. *Journal of Climate*, *7*(3), 375–399.

Gleick, P. H. (1987). The development and testing of a water balance model for climate impact assessment: Modeling the Sacramento basin. *Water Resources Research*, *23*(6), 1049–1061.

Graham, L. P., Hagemann, S., Jaun, S., & Beniston, M. (2007). On interpreting hydrological change from regional climate models. *Climatic Change*, *81*(Suppl 1), 97–122.

Gudmundsson, L., Tallaksen, L. M., Stahl, K., Clark, D. B., Dumont, E., Hagemann, S., Bertrand, N., Gerten, D., Heinke, J., & Hanasaki, N. (2012). Comparing large-scale hydrological model simulations to observed runoff percentiles in Europe. *Journal of Hydrometeorology*, *13*(2), 604–620.

Guven, A. (2009). Linear genetic programming for time-series modelling of daily flow rate. *Journal of Earth System Science*, *118*, 137–146.

Guven, A., & Aytek, A. (2009). New approach for stage–discharge relationship: Gene-expression programming. *Journal of Hydrologic Engineering*, *14*(8), 812–820.

Guven, A., & Kişi, Ö. (2010). Estimation of suspended sediment yield in natural rivers using machine-coded linear genetic programming. *Water Resources Management*, *25*, 691–704.

Guven, A., & Kisi, O. (2013). Monthly pan evaporation modeling using linear genetic programming. *Journal of Hydrology*, *503*, 178–185.

Guven, A., & Pala, A. (2022). Comparison of different statistical downscaling models and future projection of areal mean precipitation of a river basin under climate change effect. *Water Supply*, *22*(3), 2424–2439.

Guven, A., Pala, A., & Sheikhvaisi, M. (2022). Investigation of impact of climate change on small catchments using different climate models and statistical approaches. *Water Supply*, *22*(3), 3540–3552.

Guven, A., & Talu, N. E. (2010). Gene expression programing for estimating suspended sediment yield in Middle Euphrates Basin, Turkey. *CLEAN–Soil, Air, Water*, *38*(12), 1159–1168.

Haines, A., Kovats, R. S., Campbell-Lendrum, D., & Corvalán, C. (2006). Climate change and human health: Impacts, vulnerability and public health. *Public Health*, *120*(7), 585–596.

Hanssen-Bauer, I., Achberger, C., Benestad, R. E., Chen, D., & Førland, E. J. (2005). Statistical downscaling of climate scenarios over Scandinavia. *Climate Research*, *29*(3), 255–268.

Hawkins, D. M. (1980). *Identification of outliers* (Vol. 11). Springer.

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., Kim, H., & Kanae, S. (2013). Global flood risk under climate change. *Nature Climate Change*, *3*(9), 816–821.

Hoang, L. P., Lauri, H., Kummu, M., Koponen, J., Van Vliet, M. T., Supit, I., Leemans, R., Kabat, P., & Ludwig, F. (2016). Mekong River flow and hydrological extremes under climate change. *Hydrology and Earth System Sciences*, *20*(7), 3027–3041.

Hong, S.-Y., & Kanamitsu, M. (2014). Dynamical downscaling: Fundamental issues from an NWP point of view and recommendations. *Asia-Pacific Journal of Atmospheric Sciences*, *50*, 83–104.

Houghton, J. T., Jenkins, G. J., & Ephraums, J. J. (1990). Climate change: The IPCC scientific assessment. *American Scientist;(United States)*, *80*(6).

Huang, Z., Mei, Y., & Zhang, M. (2021). Investigation of Linear Genetic Programming for Dynamic Job Shop Scheduling. *2021 IEEE Symposium Series on Computational Intelligence (SSCI)*, 1–8.

Inman, D. L., & Jenkins, S. A. (1999). Climate change and the episodicity of sediment flux of small California rivers. *The Journal of Geology*, *107*(3), 251–270.

Jacob, D., & Podzun, R. (1997). Sensitivity studies with the regional climate model REMO. *Meteorology and Atmospheric Physics*, *63*, 119–129.

Jacobson, M. Z., & Delucchi, M. A. (2011). Providing all global energy with wind, water, and solar power, Part I: Technologies, energy resources, quantities and areas of infrastructure, and materials. *Energy Policy*, *39*(3), 1154–1169.

Jamet, S., & Corfee-Morlot, J. (2009). *Assessing the impacts of climate change: A literature review*.

Ji, Z., & Kang, S. (2015). Evaluation of extreme climate events using a regional climate model for China. *International Journal of Climatology*, *35*(6), 888–902.

Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., & Woollen, J. (1996). The NCEP/NCAR 40-year reanalysis project. *Bulletin of the American Meteorological Society*, *77*(3), 437–472.

Khoi, D. N., Sam, T. T., Chi, N. T. T., Linh, D. Q., & Nhi, P. T. T. (2022). Impact of future climate change on river discharge and groundwater recharge: A case study of Ho Chi Minh City, Vietnam. *Journal of Water and Climate Change*, *13*(3), 1313–1325.

Kisi, O., & Guven, A. (2010a). A machine code-based genetic programming for suspended sediment concentration estimation. *Advances in Engineering Software*, *41*(7–8), 939–945.

Kisi, O., & Guven, A. (2010b). Evapotranspiration modeling using linear genetic programming technique. *Journal of Irrigation and Drainage Engineering*, *136*(10), 715–723.

Kisi, O., & Shiri, J. (2012). River suspended sediment estimation by climatic variables implication: Comparative study among soft computing techniques. *Computers & Geosciences*, *43*, 73–82.

Kılkış, Ş., Krajačić, G., Duić, N., & Rosen, M. A. (2022). Effective mitigation of climate change with sustainable development of energy, water and environment systems. In *Energy Conversion and Management* (Vol. 269, p. 116146). Elsevier.

Kleinn, J., Frei, C., Gurtz, J., Lüthi, D., Vidale, P. L., & Schär, C. (2005). Hydrologic simulations in the Rhine basin driven by a regional climate model. *Journal of Geophysical Research: Atmospheres*, *110*(D4).

Koza, J. R. (1992). *Genetic Programming: On the Programming of Computers by Means of Natural Selection. Cambridge, Massachusetts, London, MA*. MIT Press.

Kumari, K., & Yadav, S. (2018). Linear regression analysis study. *Journal of the Practice of Cardiovascular Sciences*, *4*(1), 33.

Kundzewicz, Z. W., Mata, L. J., Arnell, N. W., Doll, P., Kabat, P., Jimenez, B., Miller, K., Oki, T., Zekai, S., & Shiklomanov, I. (2007). *Freshwater resources and their management*.

Leung, L. R., Qian, Y., & Bian, X. (2003). Hydroclimate of the western United States based on observations and regional climate simulation of 1981–2000. Part I: Seasonal statistics. *Journal of Climate*, *16*(12), 1892–1911.

Lutz, A. F., Immerzeel, W. W., Shrestha, A. B., & Bierkens, M. F. P. (2014). Consistent increase in High Asia's runoff due to increasing glacier melt and precipitation. *Nature Climate Change*, *4*(7), 587–592.

Maraun, D., Wetterhall, F., Ireson, A. M., Chandler, R. E., Kendon, E. J., Widmann, M., Brienen, S., Rust, H. W., Sauter, T., & Themeßl, M. (2010). Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics*, *48*(3).

Markus Brameier. (2004). *Brameier M (2004) On linear genetic programming. Ph.D. thesis.* University of Dortmund.

Mehr, A. D. (2018). An improved gene expression programming model for streamflow forecasting in intermittent streams. *Journal of Hydrology*, *563*, 669–678.

Milly, P. C., Dunne, K. A., & Vecchia, A. V. (2005). Global pattern of trends in streamflow and water availability in a changing climate. *Nature*, *438*(7066), 347–350.

Muller, M. (2007). Adapting to climate change: Water management for urban resilience. *Environment and Urbanization*, *19*(1), 99–113.

Nkwasa, A., Chawanda, C. J., & Van Griensven, A. (2022). Regionalization of the SWAT+ model for projecting climate change impacts on sediment yield: An application in the Nile basin. *Journal of Hydrology: Regional Studies*, *42*, 101152.

Nordin, P. (1994). 14 A Compiling Genetic Programming System that Directly. *Advances in Genetic Programming*, *1*, 311.

Oltean, M., & Grosan, C. (2003). A comparison of several linear genetic programming techniques. *Complex Systems*, *14*(4), 285–314.

Oltean, M., GROŞAN, C., DIOŞAN, L., & MIHĂILĂ, C. (2009). Genetic programming with linear representation: A survey. *International Journal on Artificial Intelligence Tools*, *18*(02), 197–238.

Pilling, C., & Jones, J. A. A. (1999). High resolution climate change scenarios: Implications for British runoff. *Hydrological Processes*, *13*(17), 2877–2895.

Praskievicz, S. (2016). Impacts of projected climate changes on streamflow and sediment transport for three snowmelt-dominated rivers in the interior Pacific Northwest. *River Research and Applications*, *32*(1), 4–17.

Prudhomme, C., Reynard, N., & Crooks, S. (2002). Downscaling of global climate models for flood frequency analysis: Where are we now? *Hydrological Processes*, *16*(6), 1137–1150.

Ramirez-Villegas, J., Challinor, A. J., Thornton, P. K., & Jarvis, A. (2013). Implications of regional improvement in global climate models for agricultural impact research. *Environmental Research Letters*, *8*(2), 024018.

Salathé Jr, E. P., Mote, P. W., & Wiley, M. W. (2007). Review of scenario selection and downscaling methods for the assessment of climate change impacts on hydrology in the United States Pacific Northwest. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *27*(12), 1611–1621.

Sánchez, E., Gallardo, C., Gaertner, M. A., Arribas, A., & Castro, M. (2004). Future climate extreme events in the Mediterranean simulated by a regional climate model: A first approach. *Global and Planetary Change*, *44*(1–4), 163–180.

Seckin, N., & Guven, A. (2012). Estimation of peak flood discharges at ungauged sites across Turkey. *Water Resources Management*, *26*, 2569–2581.

Seguí, P. Q., Ribes, A., Martin, E., Habets, F., & Boé, J. (2010). Comparison of three downscaling methods in simulating the impact of climate change on the hydrology of Mediterranean basins. *Journal of Hydrology*, *383*(1–2), 111–124.

Sennikovs, J., & Bethers, U. (2009). Statistical downscaling method of regional climate model results for hydrological modelling. *Proceedings of the 18th World IMacS/MODSIM Congress*, 3962–3968.

Slater, T., Lawrence, I. R., Otosaka, I. N., Shepherd, A., Gourmelen, N., Jakob, L., Tepes, P., Gilbert, L., & Nienow, P. (2021). Earth's ice imbalance. *The Cryosphere*, *15*(1), 233–246.

Solomon, S. (2007). IPCC (2007): Climate change the physical science basis. *Agu Fall Meeting Abstracts*, *2007*, U43D-01.

Tan, J., Zhang, Y., Ma, W., Yu, Q., Wang, J., & Chen, L. (2015). Impact of spatial resolution on air quality simulation: A case study in a highly industrialized area in Shanghai, China. *Atmospheric Pollution Research*, *6*(2), 322–333.

Tatli, H., Nüzhet Dalfes, H., & Sibel Menteş, Ş. (2004). A statistical downscaling method for monthly total precipitation over Turkey. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *24*(2), 161–180.

Taylor, K. E., Stouffer, R. J., & Meehl, G. A. (2012). An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society*, *93*(4), 485–498.

Thodsen, H. (2007). The influence of climate change on stream flow in Danish rivers. *Journal of Hydrology*, *333*(2–4), 226–238.

Thodsen, H., Hasholt, B., & Kjærsgaard, J. H. (2008). The influence of climate change on suspended sediment transport in Danish rivers. *Hydrological Processes: An International Journal*, *22*(6), 764–774.

Tofiq, F. A., & Guven, A. (2014). Prediction of design flood discharge by statistical downscaling and General Circulation Models. *Journal of Hydrology*, *517*, 1145–1153.

Traore, S., & Guven, A. (2012). Regional-specific numerical models of evapotranspiration using gene-expression programming interface in Sahel. *Water Resources Management*, *26*, 4367–4380.

Trenberth, K. E., Jones, P. D., Ambenje, P., Bojariu, R., Easterling, D., Klein Tank, A., Parker, D., Rahimzadeh, F., Renwick, J. A., & Rusticucci, M. (2007). *Observations. Surface and atmospheric climate change. Chapter 3*.

Turhan, E., Cerit Mazlum, S., Şahin, Ü., Şorman, A. H., & Cem Gündoğan, A. (2016). Beyond special circumstances: Climate change policy in Turkey 1992–2015. *Wiley Interdisciplinary Reviews: Climate Change*, *7*(3), 448–460.

van Roosmalen, L., Christensen, B. S., & Sonnenborg, T. O. (2007). Regional differences in climate change impacts on groundwater and stream discharge in Denmark. *Vadose Zone Journal*, *6*(3), 554–571.

Varghese, S., Langmann, B., Ceburnis, D., & O'Dowd, C. D. (2011). Effect of horizontal resolution on meteorology and air-quality prediction with a regional scale model. *Atmospheric Research*, *101*(3), 574–594.

Verma, S., Bhattarai, R., Bosch, N. S., Cooke, R. C., Kalita, P. K., & Markus, M. (2015). Climate change impacts on flow, sediment and nutrient export in a Great Lakes watershed using SWAT. *CLEAN–Soil, Air, Water*, *43*(11), 1464–1474.

Vrac, M., Stein, M. L., Hayhoe, K., & Liang, X.-Z. (2007). A general method for validating statistical downscaling methods under future climate change. *Geophysical Research Letters*, *34*(18).

Wagena, M. B., Sommerlot, A., Abiy, A. Z., Collick, A. S., Langan, S., Fuka, D. R., & Easton, Z. M. (2016). Climate change in the Blue Nile Basin Ethiopia: Implications for water resources and sediment transport. *Climatic Change*, *139*, 229–243.

Wei, X., Cai, S., Ni, P., & Zhan, W. (2020). Impacts of climate change and human activities on the water discharge and sediment load of the Pearl River, southern China. *Scientific Reports*, *10*(1), 16743.

Wetterhall, F. (2005). *Statistical downscaling of precipitation from large-scale atmospheric circulation: Comparison of methods and climate regions* [PhD Thesis].

Wilby, R. L., Charles, S. P., Zorita, E., Timbal, B., Whetton, P., & Mearns, L. O. (2004). Guidelines for use of climate scenarios developed from statistical downscaling methods. *Supporting Material of the Intergovernmental Panel on Climate Change, Available from the DDC of IPCC TGCIA*, *27*.

Wilby, R. L., & Wigley, T. M. (1997). Downscaling general circulation model output: A review of methods and limitations. *Progress in Physical Geography*, *21*(4), 530–548.

Xu, Y., Zhang, Y., Lin, E., Lin, W., Dong, W., Jones, R., Hassell, D., & Wilson, S. (2006). Analyses on the climate change responses over China under SRES B2 scenario using PRECIS. *Chinese Science Bulletin*, *51*, 2260–2267.

Xu, Z., Han, Y., & Yang, Z. (2019). Dynamical downscaling of regional climate: A review of methods and limitations. *Science China Earth Sciences*, *62*, 365–375.

Xu, Z., & Yang, Z.-L. (2012). An improved dynamical downscaling method with GCM bias corrections and its validation with 30 years of climate simulations. *Journal of Climate*, *25*(18), 6271–6286.

Yerlikaya, B. A., Ömezli, S., & Aydoğan, N. (2020). Climate change forecasting and modeling for the year of 2050. *Environment, Climate, Plant and Vegetation Growth*, 109–122.

Yilmaz, A. G., & Imteaz, M. A. (2014). Climate change and water resources in Turkey: A review. *International Journal of Water*, *8*(3), 299–313.

Yucel, I., Güventürk, A., & Sen, O. L. (2015). Climate change impacts on snowmelt runoff for mountainous transboundary basins in eastern Turkey. *International Journal of Climatology*, *35*(2), 215–228.

APPENDICES

APPENDIX A

Table A.1 Tables of monthly discharge and sediment yield (predictands) of Fırat N. Kemahboğazı Station

Y	M	Q (m^3/s)	Qs (tons/day)
1970	1	42.700	575.000
1970	2	45.200	286.300
1970	3	63.500	1527.300
1970	4	244.450	38833.650
1970	5	117.500	2597.650
1970	6	54.400	891.500
1970	7	31.900	1022.000
1970	8	27.000	246.400
1970	9	25.500	483.000
1970	10	31.200	171.200
1970	11	34.800	120.000
1970	12	36.000	644.800
1971	1	26.000	90.900
1971	2	36.200	964.500
1971	3	104.300	9077.800
1971	4	104.500	6829.500
1971	5	303.300	55535.800
1971	6	81.300	5015.400
1971	7	24.500	170.100
1971	8	32.400	849.200
1971	9	24.700	199.200
1971	10	28.400	105.800
1971	11	34.700	254.200
1971	12	43.800	198.300
1972	1	21.700	299.800
1972	2	30.500	50.300
1972	3	53.100	13387.300
1972	4	183.800	20745.700
1972	5	152.400	6368.900
1972	6	140.700	21159.000
1972	7	28.200	596.050
1972	8	20.000	305.900
1972	9	25.100	341.100

Table A.1 Continued…

Y	M	Q (m^3/s)	Qs (tons/day)
1972	10	37.574	125.700
1972	11	38.938	622.000
1972	12	29.563	185.500
1973	1	28.600	109.300
1973	2	48.483	4523.100
1973	3	43.290	644.000
1973	4	238.098	47602.100
1973	5	205.796	24146.200
1973	6	107.188	8228.800
1973	7	34.345	307.300
1973	8	18.600	882.000
1973	9	14.321	149.700
1973	10	28.797	361.300
1973	11	56.500	1904.600
1973	12	39.500	18499.000
1974	1	24.200	146.900
1974	2	97.900	260.600
1974	3	192.400	97364.000
1974	4	300.347	102368.400
1974	5	349.000	69442.100
1974	6	74.000	860.300
1974	7	28.636	255.300
1974	8	21.300	156.900
1974	9	37.900	5249.700
1974	10	25.000	106.600
1974	11	31.400	45.300
1974	12	38.870	226.700
1975	1	34.500	181.700
1975	2	19.862	135.100
1975	3	33.259	528.800
1975	4	218.000	26802.300
1975	5	197.700	10504.800
1975	6	81.000	8734.500
1975	7	31.803	458.000
1975	8	21.276	122.600
1975	9	22.000	106.200
1975	10	32.000	134.800
1975	11	29.889	108.900
1975	12	29.237	287.900
1976	1	20.910	74.900
1976	2	31.230	276.300

Table A.1 Continued…

Y	M	Q (m^3/s)	Qs (tons/day)
1976	3	36.905	382.300
1976	4	210.087	24791.000
1976	5	353.977	26948.900
1976	6	211.571	3425.400
1976	7	66.565	3733.700
1976	8	34.804	84.700
1976	9	27.600	130.700
1976	10	33.000	72.000
1976	11	40.189	337.700
1976	12	44.800	390.400
1977	1	34.900	110.300
1977	2	42.500	580.700
1977	3	49.400	715.700
1977	4	164.550	16445.800
1977	5	256.650	15075.400
1977	6	139.700	2234.000
1977	7	36.500	282.200
1977	8	27.500	36.500
1977	9	27.000	84.800
1977	10	40.700	261.200
1977	11	37.400	112.400
1977	12	41.000	168.900
1978	1	41.100	735.600
1978	2	41.800	402.500
1978	3	47.300	415.300
1978	4	233.700	78071.850
1978	5	334.900	62244.100
1978	6	150.300	5723.350
1978	7	77.900	566.700
1978	8	36.000	196.100
1978	9	34.400	312.200
1978	10	33.200	99.200
1978	11	37.977	438.138
1978	12	45.100	225.000
1979	1	43.300	281.500
1979	2	47.950	422.300
1979	3	60.833	3829.167
1979	4	156.200	9879.000
1979	5	256.250	13021.700
1979	6	212.250	10209.350
1979	7	54.967	993.600

Table A.1 Continued…

Y	M	Q (m^3/s)	Qs (tons/day)
1979	8	34.800	175.700
1979	9	25.150	149.650
1979	10	32.208	159.756
1979	11	82.400	4822.500
1979	12	58.750	545.300
1980	1	31.791	260.590
1980	2	43.500	225.200
1980	3	143.533	29990.700
1980	4	310.267	54253.733
1980	5	24.200	8016.700
1980	6	160.000	4412.800
1980	7	54.700	783.400
1980	8	34.400	1179.200
1980	9	27.600	143.100
1980	10	32.208	159.756
1980	11	42.300	625.500
1980	12	54.900	600.000
1981	1	41.800	305.300
1981	2	37.000	125.700
1981	3	60.600	1386.100
1981	4	182.233	14298.900
1981	5	202.100	17827.500
1981	6	171.800	3975.700
1981	7	67.400	1128.500
1981	8	30.200	220.400
1981	9	26.479	209.965
1981	10	46.500	2637.400
1981	11	49.500	842.000
1981	12	44.200	346.800
1982	1	32.625	264.316
1982	2	30.900	193.300
1982	3	50.300	1831.300
1982	4	401.845	36205.300
1982	5	344.875	25518.550
1982	6	187.100	3972.900
1982	7	40.782	471.800
1982	8	35.500	232.700
1982	9	28.900	175.400
1982	10	35.000	267.500
1982	11	33.100	248.000
1982	12	42.143	347.236

Table A.1 Continued…

Y	M	Q (m^3/s)	Qs (tons/day)
1983	1	30.100	96.100
1983	2	23.600	67.900
1983	3	44.500	160.000
1983	4	87.550	18750.550
1983	5	102.700	1006.200
1983	6	170.050	6264.750
1983	7	22.800	11.750
1983	8	18.100	98.000
1983	9	26.665	207.085
1983	10	33.522	168.734
1983	11	108.741	2400.200
1983	12	41.815	122.000
1984	1	40.203	252.900
1984	2	43.019	442.900
1984	3	52.366	288.200
1984	4	151.344	7074.300
1984	5	313.778	22634.300
1984	6	138.669	6079.125
1984	7	38.775	500.900
1984	8	36.843	229.700
1984	9	24.423	66.800
1984	10	37.171	212.300
1984	11	35.811	196.400
1984	12	20.373	198.100
1985	1	38.847	646.400
1985	2	37.267	316.686
1985	3	32.129	174.600
1985	4	258.587	62384.600
1985	5	156.477	3937.900
1985	6	76.698	833.800
1985	7	17.480	72.100
1985	8	13.357	66.900
1985	9	14.294	52.700
1985	10	32.508	875.300
1985	11	34.803	210.800
1985	12	29.085	190.250
1986	1	38.853	145.500
1986	2	34.616	285.300
1986	3	54.951	877.300
1986	4	147.767	3795.000
1986	5	160.124	5835.150

Table A.1 Continued…

Y	M	Q (m^3/s)	Qs (tons/day)
1986	6	195.945	10583.000
1986	7	53.584	379.800
1986	8	20.033	120.000
1986	9	18.999	103.300
1986	10	40.017	44.200
1986	11	52.536	197.000
1986	12	37.103	216.600
1987	1	35.627	75.700
1987	2	47.328	84.200
1987	3	52.145	162.200
1987	4	293.223	89477.100
1987	5	465.070	27397.500
1987	6	194.131	2454.700
1987	7	81.712	537.100
1987	8	28.167	125.500
1987	9	32.095	80.500
1987	10	38.588	274.200
1987	11	68.118	951.500
1987	12	54.127	745.200
1988	1	36.197	380.800
1988	2	39.873	262.800
1988	3	83.765	1424.300
1988	4	523.877	62870.900
1988	5	345.013	34512.700
1988	6	190.861	3109.800
1988	7	110.853	1279.400
1988	8	49.243	558.400
1988	9	54.381	218.900
1988	10	41.954	145.700
1988	11	130.663	2673.400
1988	12	78.601	1521.500
1989	1	46.454	606.300
1989	2	47.567	429.900
1989	3	121.226	5589.600
1989	4	130.774	11625.700
1989	5	56.918	399.100
1989	6	65.752	1533.100
1989	7	16.780	276.600
1989	8	13.427	169.700
1989	9	18.552	150.700

Table A.1 Continued…

Y	M	Q (m^3/s)	Qs (tons/day)
1989	10	43.497	1094.100
1989	11	38.518	1364.000
1989	12	56.885	1030.000
1990	1	34.520	281.990
1990	2	41.995	284.600
1990	3	59.208	1043.200
1990	4	156.236	3969.100
1990	5	240.481	5669.500
1990	6	80.261	608.900
1990	7	27.933	99.100
1990	8	36.667	78.000
1990	9	38.394	199.700
1990	10	44.400	425.700
1990	11	58.346	543.800
1990	12	56.447	574.000
1991	1	31.619	268.500
1991	2	42.470	319.100
1991	3	49.764	412.600
1991	4	237.719	6622.600
1991	5	301.104	17221.500
1991	6	98.742	883.600
1991	7	65.347	2289.900
1991	8	24.332	2248.800
1991	9	29.806	80.100
1991	10	42.266	1421.400
1991	11	58.797	573.100
1991	12	33.021	370.700
1992	1	31.898	350.000
1992	2	30.508	204.500
1992	3	81.186	1497.800
1992	4	190.503	5082.600
1992	5	399.912	63144.800
1992	6	147.504	4194.600
1992	7	85.756	1325.600
1992	8	43.224	1285.000
1992	9	44.372	329.900
1992	10	39.633	182.800
1992	11	57.753	740.000
1992	12	43.955	577.300
1993	1	47.130	294.500

Table A.1 Continued…

Y	M	Q (m^3/s)	Qs (tons/day)
1993	2	36.156	105.300
1993	3	38.823	415.700
1993	4	385.574	126037.700
1993	5	379.743	12536.600
1993	6	152.106	1240.400
1993	7	54.993	205.200
1993	8	39.750	134.500
1993	9	43.228	155.800
1993	10	61.439	254.500
1993	11	61.017	398.500
1993	12	43.316	341.900
1994	1	24.437	134.600
1994	2	40.114	134.900
1994	3	56.151	222.600
1994	4	120.703	1773.900
1994	5	117.407	5622.400
1994	6	35.018	787.000
1994	7	29.879	249.600
1994	8	17.674	49.900
1994	9	28.443	70.100
1994	10	47.313	680.400
1994	11	45.031	905.400
1994	12	34.721	193.800
1995	1	46.849	116.500
1995	2	45.433	181.800
1995	3	81.085	4056.000
1995	4	241.336	17170.500
1995	5	302.473	9233.800
1995	6	138.389	1557.100
1995	7	47.260	415.800
1995	8	28.451	68.500
1995	9	35.272	180.700
1995	10	44.221	205.700
1995	11	72.941	629.500
1995	12	47.496	162.600
1996	1	42.528	169.600
1996	2	58.185	387.000
1996	3	59.963	353.400
1996	4	209.364	19220.400
1996	5	254.299	11778.500

Table A.1 Continued…

Y	M	Q (m^3/s)	Qs (tons/day)
1996	6	126.155	693.500
1996	7	43.108	351.800
1996	8	41.711	298.600
1996	9	64.433	767.200
1996	10	46.302	362.200
1996	11	58.777	330.200
1996	12	53.330	263.900
1997	1	58.436	296.400
1997	2	45.450	2099.500
1997	3	41.958	127.100
1997	4	103.482	1606.800
1997	5	399.879	21788.100
1997	6	70.204	448.700
1997	7	55.955	548.400
1997	8	25.786	135.300
1997	9	27.872	182.300
1997	10	55.193	688.900
1997	11	60.144	981.200
1997	12	45.773	327.700
1998	1	41.142	255.700
1998	2	37.609	95.400
1998	3	77.425	1634.900
1998	4	442.885	102010.300
1998	5	281.055	5799.500
1998	6	235.232	13266.700
1998	7	56.710	379.700
1998	8	40.844	286.100
1998	9	26.028	100.200
1998	10	38.541	190.400
1998	11	52.063	118.500
1998	12	64.188	641.800
1999	1	51.736	344.200
1999	2	47.795	200.800
1999	3	45.416	179.000
1999	4	123.182	3677.700
1999	5	227.964	11020.400
1999	6	108.863	1692.700
1999	7	56.884	781.300
1999	8	31.331	111.300
1999	9	33.469	174.300

Table A.1 Continued…

Y	M	Q (m^3/s)	Qs (tons/day)
1999	10	37.863	227.100
1999	11	40.564	193.600
1999	12	51.606	286.500
2000	1	35.884	204.900
2000	2	41.223	115.300
2000	3	42.270	102.000
2000	4	221.893	14186.600
2000	5	169.583	2845.400
2000	6	78.294	199.600
2000	7	33.192	141.900
2000	8	28.800	74.900
2000	9	20.117	35.400
2000	10	51.883	426.700
2000	11	50.109	475.500
2000	12	36.244	242.800
2001	1	44.980	140.900
2001	2	33.185	91.000
2001	3	74.674	835.300
2001	4	143.811	4000.000
2001	5	100.641	662.800
2001	6	40.347	389.000
2001	7	33.604	107.100
2001	8	28.027	182.500
2001	9	22.671	74.400
2001	10	18.097	83.700
2001	11	38.680	141.200
2001	12	36.681	128.700
2002	1	32.181	118.600
2002	2	40.615	182.500
2002	3	73.577	1321.200
2002	4	188.955	5840.400
2002	5	240.680	6341.800
2002	6	145.598	1510.600
2002	7	58.909	616.100
2002	8	39.437	211.800
2002	9	46.502	374.000
2002	10	38.017	180.400
2002	11	44.932	109.600
2002	12	38.073	207.000
2003	1	34.491	91.100

Table A.1 Continued…

Y	M	Q (m^3/s)	Qs (tons/day)
2003	2	37.695	73.500
2003	3	71.332	1141.200
2003	4	297.895	31373.500
2003	5	251.973	6241.500
2003	6	79.595	582.600
2003	7	39.756	132.500
2003	8	35.741	211.700
2003	9	36.720	152.500
2003	10	34.052	107.300
2003	11	55.341	451.100
2003	12	59.113	199.200
2004	1	53.434	146.400
2004	2	58.876	376.200
2004	3	127.419	1712.900
2004	4	198.893	4573.700
2004	5	378.586	15742.700
2004	6	138.222	3821.600
2004	7	90.373	1344.400
2004	8	40.958	186.100
2004	9	45.462	123.700
2004	10	48.123	218.700
2004	11	43.307	247.700
2004	12	42.704	86.800
2005	1	44.784	159.700
2005	2	57.164	259.600
2005	3	81.863	1139.500
2005	4	251.610	10557.600
2005	5	254.220	5151.700
2005	6	168.991	2150.200
2005	7	44.705	579.700
2005	8	36.600	1907.000
2005	9	36.982	210.600
2005	10	69.992	1782.800
2005	11	74.529	407.200
2005	12	57.683	644.500

APPENDIX B

The C Code generated from machine-coded linear genetic programming (LGP) model to predict discharge (Q)

```
#define TRUNC(x)(((x)>=0) ? floor(x) : ceil(x))

#define C_FPREM (_finite(f[0]/f[1]) ? f[0]-(TRUNC(f[0]/f[1])*f[1]) : f[0]/f[1])

#define C_F2XM1 (((fabs(f[0])<=1) && (!_isnan(f[0]))) ? (pow(2,f[0])-1) :
((!_finite(f[0]) && !_isnan(f[0]) && (f[0]<0)) ? -1 : f[0]))


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


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


  L0:    f[0]+=0.1756083965301514f;
  L1:    f[0]*=v[5];
  L2:    cflag=(f[0] < f[3]);
  L3:    if (!cflag) f[0] = f[2];
  L4:    f[0]+=0.1058487892150879f;
```

```
f[0]*=0.1058487892150879f;
f[2]+=f[0];
f[0]/=v[9];
f[0]+=v[9];
f[2]+=f[0];
f[0]/=f[0];
f[0]-=v[3];
f[0]-=v[0];
f[0]/=1.505081653594971f;
f[0]-=0.7361507415771484f;
f[2]*=f[0];
f[0]=sqrt(f[0]);
f[0]-=v[1];
tmp=f[3]; f[3]=f[0]; f[0]=tmp;
f[0]-=-0.6496260166168213f;
f[0]=sqrt(f[0]);
f[3]*=f[0];
f[2]/=f[0];
f[0]*=1.501374244689941f;
f[0]/=0.03275442123413086f;
f[0]*=v[7];
f[0]=cos(f[0]);
```

```
f[0]=sin(f[0]);
f[0]-=f[2];
cflag=(f[0] < f[3]);
tmp=f[1]; f[1]=f[0]; f[0]=tmp;
if (cflag) f[0] = f[2];
f[3]-=f[0];
f[0]+=f[0];
cflag=(f[0] < f[2]);
f[0]=fabs(f[0]);
f[0]+=v[3];
f[0]-=v[9];
f[0]*=0.2174909114837647f;
f[0]+=f[0];
if (!cflag) f[0] = f[2];
f[3]-=f[0];
f[2]+=f[0];
f[0]*=f[0];
f[0]=cos(f[0]);
f[0]=fabs(f[0]);
f[0]-=0.1387641429901123f;
f[2]/=f[0];
f[0]+=v[0];
```

```
  L49:  f[0]=sqrt(f[0]);

  L50:  f[0]-=-1.641227006912231f;

  L51:  f[0]+=f[3];

  L52:  cflag=(f[0] < f[0]);

  L53:  if (cflag) f[0] = f[2];

  L54:  f[0]-=-0.09100413322448731f;

  L55:  f[0]*=pow(2,TRUNC(f[1]));

  L56:  f[0]-=v[3];

  L57:  f[0]-=-1.642645597457886f;

  L58:  f[0]-=-1.642645597457886f;

  L59:  f[0]=sqrt(f[0]);

  L60:  f[0]+=f[0];

  L61:  f[0]+=v[9];

  L62:


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


  return f[0];

}
```

APPENDIX C

The C Code generated from machine-coded linear genetic programming (LGP) model to predict sediment yield (Qs)

```
#define TRUNC(x)(((x)>=0) ? floor(x) : ceil(x))

#define C_FPREM (_finite(f[0]/f[1]) ? f[0]-(TRUNC(f[0]/f[1])*f[1]) : f[0]/f[1])

#define C_F2XM1 (((fabs(f[0])<=1) && (!_isnan(f[0]))) ? (pow(2,f[0])-1) :
((!_finite(f[0]) && !_isnan(f[0]) && (f[0]<0)) ? -1 : f[0]))


float DiscipulusCFunction(float v[])

{

  long double f[8];

  long double tmp = 0;

  int cflag = 0;


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


  L0:    f[0]+=v[7];

  L1:    f[0]+=0.6851940155029297f;

  L2:    f[1]-=f[0];

  L3:    f[0]+=v[0];

  L4:    f[0]+=f[1];
```

```
f[0]+=0.6851940155029297f;
f[1]-=f[0];
f[0]*=f[0];
f[0]+=f[0];
f[3]-=f[0];
f[1]/=f[0];
f[3]*=f[0];
f[0]*=f[3];
f[0]+=v[2];
f[3]/=f[0];
f[3]-=f[0];
f[0]+=v[4];
f[3]+=f[0];
f[0]-=f[0];
f[0]+=v[0];
f[1]+=f[0];
f[1]*=f[0];
f[0]+=f[0];
f[3]-=f[0];
f[0]+=v[4];
f[0]+=0.7051223516464233f;
f[1]/=f[0];
f[0]*=f[0];
f[3]/=f[0];
```

```
L29:  f[0]+=f[0];
L30:  f[3]/=f[0];
L31:  f[1]/=f[0];
L32:  f[0]=sqrt(f[0]);
L33:  f[0]+=v[4];
L34:  f[0]+=0.768571138381958f;
L35:  f[1]/=f[0];
L36:  f[0]/=f[0];
L37:  f[0]+=f[3];
L38:  f[0]+=v[7];
L39:  f[1]+=f[0];
L40:  f[0]-=f[0];
L41:  f[0]+=0.7051223516464233f;
L42:  f[0]+=v[4];
L43:  f[0]+=0.7051223516464233f;
L44:  f[0]*=f[0];
L45:  f[0]+=0.6851940155029297f;
L46:  f[1]/=f[0];
L47:  f[0]*=f[0];
L48:  f[0]+=f[0];
L49:  f[0]-=f[1];
L50:  f[0]+=0.7051223516464233f;
L51:  f[1]/=f[0];
L52:  f[0]-=f[0];
```

```
L53:  f[0]+=0.6851940155029297f;
L54:  f[1]-=f[0];
L55:  f[0]/=f[0];
L56:  f[0]+=v[4];
L57:  f[0]+=0.7051223516464233f;
L58:  f[1]/=f[0];
L59:  f[0]-=f[0];
L60:  f[0]+=v[2];
L61:  f[0]*=f[0];
L62:  f[0]+=v[1];
L63:  f[0]=fabs(f[0]);
L64:  f[0]=sqrt(f[0]);
L65:  f[0]=sqrt(f[0]);
L66:  f[0]+=v[0];
L67:  f[0]+=v[2];
L68:  f[0]=fabs(f[0]);
L69:  f[0]+=v[2];
L70:  f[0]+=f[1];
L71:  f[0]*=f[1];
L72:  f[0]+=v[9];
L73:  f[0]+=v[2];
L74:  f[0]=fabs(f[0]);
L75:  f[0]-=f[1];
L76:  f[0]*=f[0];
```

```
  L77:  f[0]+=f[0];

  L78:  f[0]=sqrt(f[0]);

  L79:  f[0]=sqrt(f[0]);

  L80:  f[0]+=0.7051223516464233f;

  L81:  f[0]-=f[1];

  L82:  f[0]*=f[0];

  L83:  f[0]+=f[0];

  L84:  f[0]+=f[1];

  L85:  f[0]=sqrt(f[0]);

  L86:  f[0]=sqrt(f[0]);

  L87:  f[0]+=0.7051223516464233f;

  L88:  f[0]*=f[0];

  L89:  f[0]+=v[9];

  L90:  f[0]+=v[9];

  L91:  f[0]+=v[2];

  L92:


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


  return f[0];
}
```

APPENDIX D

Tables and Figures for the comparison of observed and predicted monthly discharge values in the 2006-2020 period for all models and scenarios

Table D.1 Monthly discharge values for observed and predicted by the GEP model under RCP2.6 scenario in 2006-2020 period

Month	Q GEP RCP2.6	Q Obs.
Jan	44.708	83.501
Feb	62.303	61.749
Mar	70.099	91.420
Apr	68.654	157.821
May	81.012	229.123
Jun	57.749	98.818
Jul	34.086	46.822
Aug	33.760	50.050
Sep	34.572	55.263
Oct	41.612	53.352
Nov	51.171	53.447
Dec	39.849	70.666

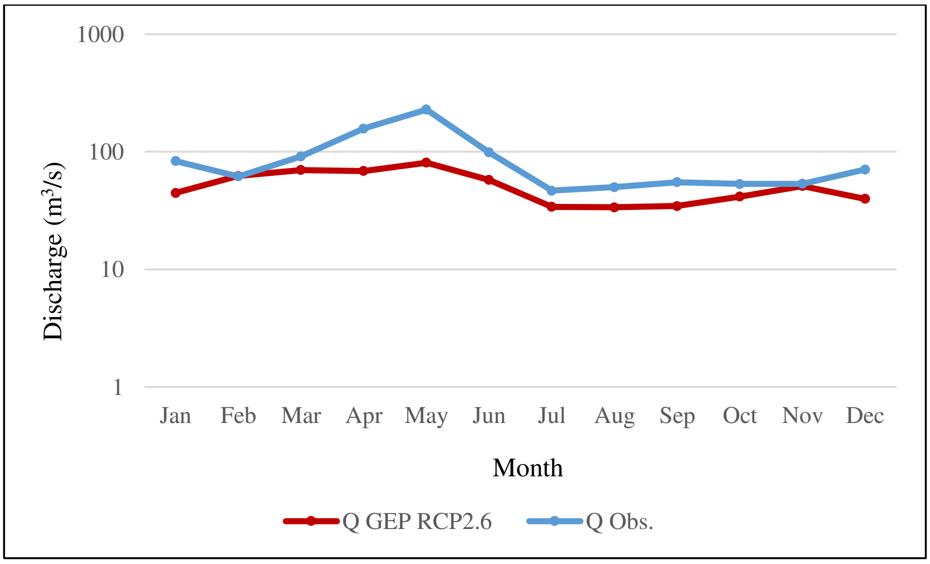

Figure D.1 Observed and predicted monthly Q values for the 2006-2020 period by GEP model under RCP2.6 scenario, in logarithmic scale

Table D.2 Monthly discharge values for observed and predicted by LGP model under RCP2.6 scenario in 2006-2020 period

Month	Q LGP RCP2.6	Q Obs.
Jan	41.037	86.105
Feb	88.835	61.749
Mar	47.042	91.420
Apr	125.813	157.821
May	117.107	229.123
Jun	59.426	98.818
Jul	63.030	46.822
Aug	44.095	50.050
Sep	88.457	55.263
Oct	38.077	53.352
Nov	43.114	53.447
Dec	42.809	66.898

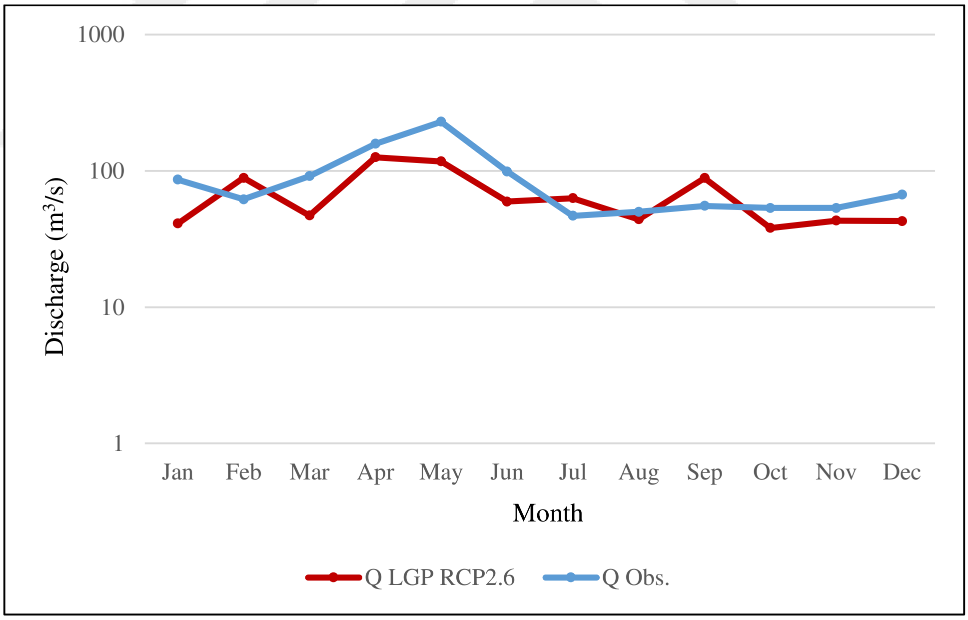

Figure D.2 Observed and predicted monthly Q values for the 2006-2020 period by LGP model under RCP2.6 scenario, in logarithmic scale

Table D.3 Monthly discharge values for observed and predicted by GEP model under RCP4.5 scenario in 2006-2020 period

Month	Q GEP RCP4.5	Q Obs.
Jan	39.783	80.283
Feb	59.746	61.749
Mar	58.681	91.420
Apr	65.846	156.469
May	87.649	229.123
Jun	54.490	97.181
Jul	31.664	46.822
Aug	37.402	50.050
Sep	34.389	55.263
Oct	49.822	53.352
Nov	50.525	52.852
Dec	48.308	70.666

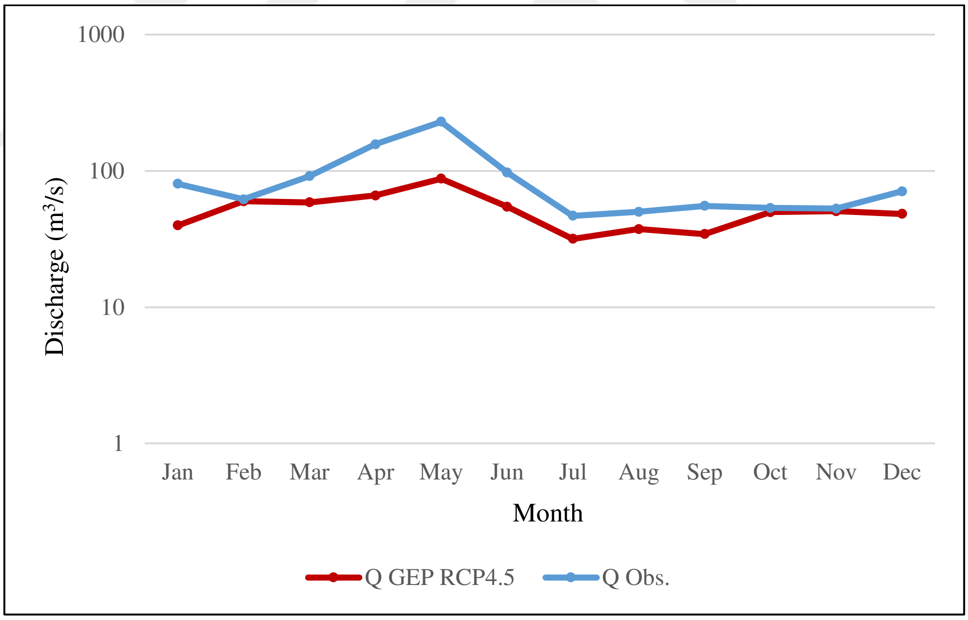

Figure D.3 Observed and predicted monthly Q values for the 2006-2020 period by GEP model under RCP4.5 scenario, in logarithmic scale

Table D.4 Monthly discharge values for observed and predicted by LGP model under RCP4.5 scenario in 2006-2020 period

Month	Q LGP RCP4.5	Q Obs.
Jan	44.645	78.536
Feb	46.171	60.702
Mar	56.104	91.420
Apr	136.945	157.821
May	154.814	229.123
Jun	60.596	97.059
Jul	53.573	46.822
Aug	43.443	50.050
Sep	85.037	55.263
Oct	45.839	54.380
Nov	47.309	52.852
Dec	62.500	70.666

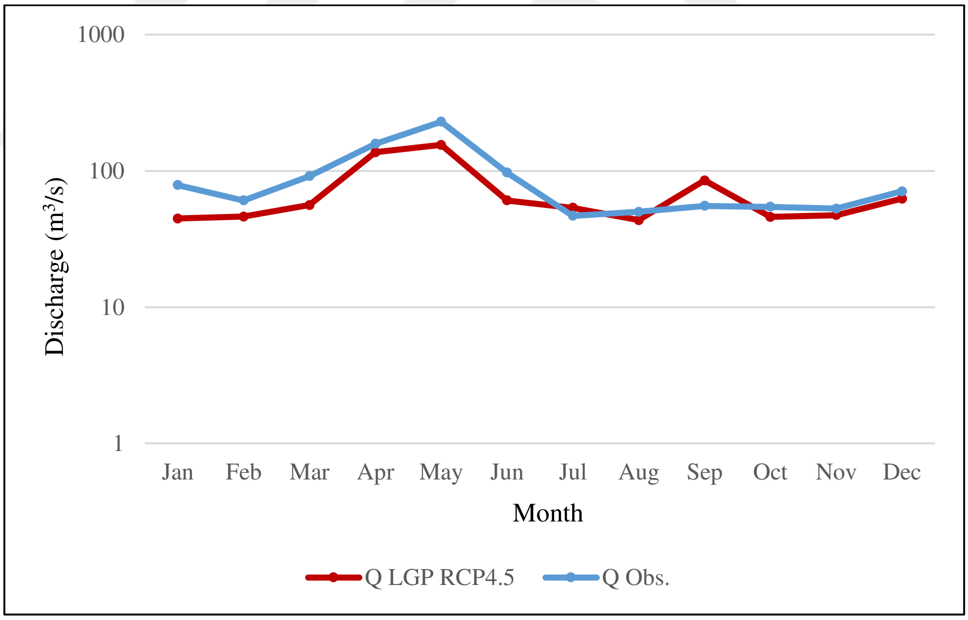

Figure D.4 Observed and predicted monthly Q values for the 2006-2020 period by LGP model under RCP4.5 scenario, in logarithmic scale

Table D.5 Monthly discharge values for observed and predicted by GEP model under RCP8.5 scenario in 2006-2020 period

Month	Q GEP RCP8.5	Q Obs.
Jan	47.613	83.501
Feb	58.030	61.749
Mar	61.077	79.162
Apr	66.358	157.821
May	118.895	229.123
Jun	54.558	92.168
Jul	35.429	46.822
Aug	35.932	50.050
Sep	40.986	55.263
Oct	45.927	53.352
Nov	51.050	54.422
Dec	41.925	68.916

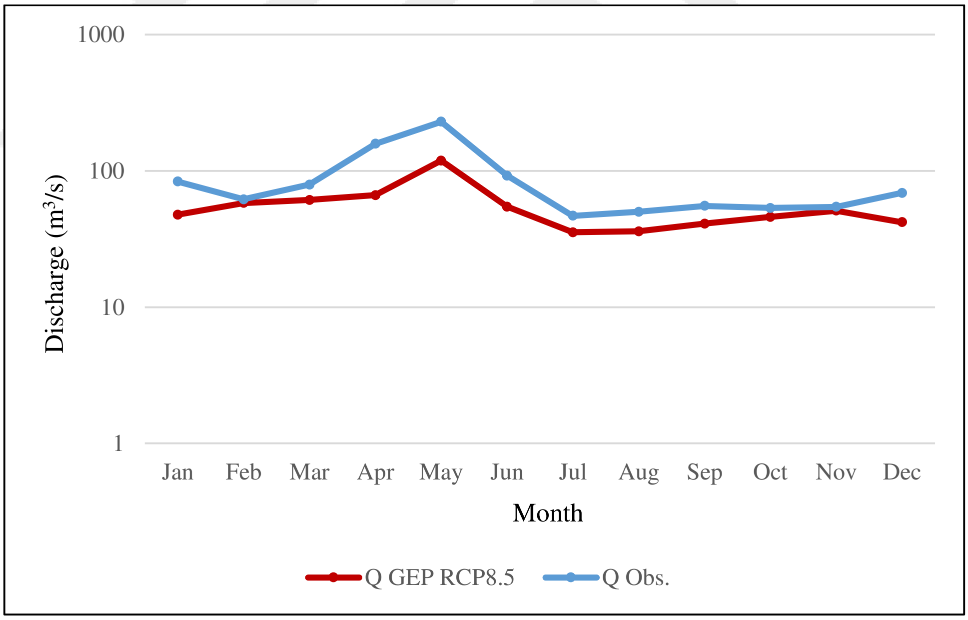

Figure D.5 Observed and predicted monthly Q values for the 2006-2020 period by GEP model under RCP8.5 scenario, in logarithmic scale

Table D.6 Monthly discharge values for observed and predicted by LGP model under RCP8.5 scenario in 2006-2020 period

Month	Q LGP RCP8.5	Observed
Jan	76.659	83.501
Feb	55.560	61.749
Mar	56.612	94.677
Apr	93.403	157.821
May	137.334	229.123
Jun	61.778	98.818
Jul	51.519	46.822
Aug	44.047	50.050
Sep	89.539	55.263
Oct	39.904	53.352
Nov	48.286	54.422
Dec	45.962	70.666

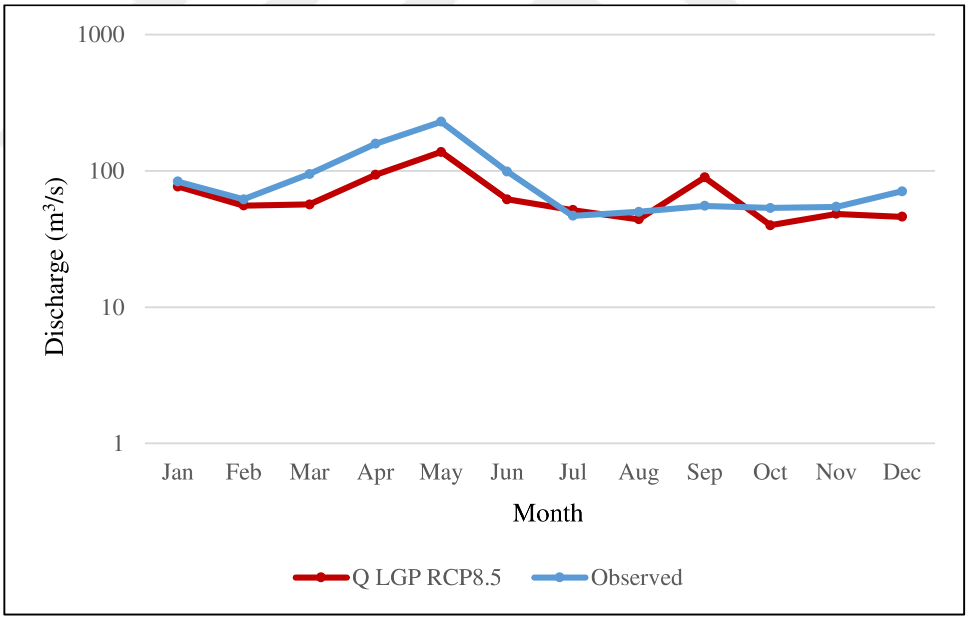

Figure D.6 Observed and predicted monthly Q values for the 2006-2020 period by LGP model under RCP8.5 scenario, in logarithmic scale

APPENDIX E

Tables and Figures of averaged monthly Q results under different CanESM2 scenarios for 20 - years of time periods by GEP and LGP models

Table E.1 Monthly Q results under CanESM2 RCP2.6 scenario for four future periods by GEP model

Month	2006-2020	2021-2040	2041-2060	2061-2080	2081-2100	2006-2020 Obs.	1970-2005 Obs.
Jan	44.708	42.910	33.503	37.381	34.494	83.501	34.879
Feb	62.303	48.709	43.744	43.140	42.631	61.749	38.603
Mar	70.099	54.454	61.643	50.979	53.896	91.420	62.145
Apr	68.654	64.028	68.977	62.864	54.970	157.821	225.895
May	81.012	75.371	86.053	85.679	96.605	229.123	256.626
Jun	57.749	50.276	50.820	48.132	46.405	98.818	132.901
Jul	34.086	27.033	27.745	25.097	28.096	46.822	47.360
Aug	33.760	32.762	31.487	28.121	27.974	50.050	28.579
Sep	34.572	34.872	35.780	29.870	30.303	55.263	29.061
Oct	41.612	45.599	37.314	43.887	41.135	53.352	37.781
Nov	51.171	51.807	44.241	41.678	43.429	53.447	52.776
Dec	39.849	43.120	37.702	34.126	36.797	70.666	43.262

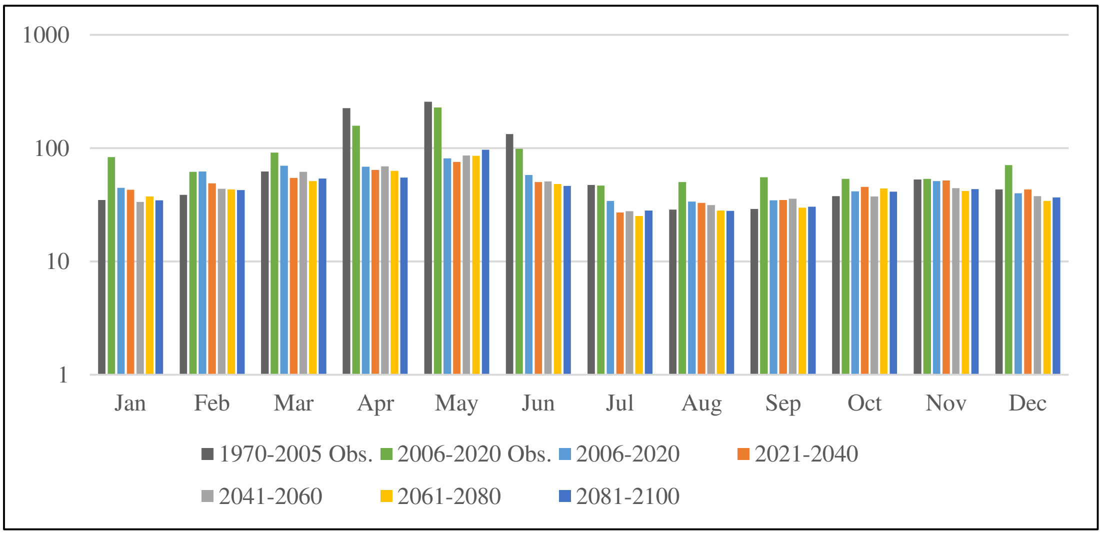

Figure E.1 The projection of average monthly discharge (Q) values under CanESM2 RCP2.6 scenario for different time periods by the GEP model in logarithmic scale

Table E.2 Monthly Q results under CanESM2 RCP4.5 scenario for four future periods by GEP model

Month	2006-2020	2021-2040	2041-2060	2061-2080	2081-2100	2006-2020 Obs	1970-2005 Obs.
Jan	39.783	40.776	42.546	40.711	44.558	83.501	34.879
Feb	59.746	55.089	56.558	53.622	64.308	61.749	38.603
Mar	58.681	69.113	60.225	61.047	62.654	91.420	62.145
Apr	65.846	77.093	66.009	81.608	81.630	157.821	225.895
May	87.649	71.967	97.179	80.464	93.209	229.123	256.626
Jun	54.490	67.147	41.318	53.484	58.226	98.818	132.901
Jul	31.664	29.789	33.660	33.332	33.103	46.822	47.360
Aug	37.402	35.731	35.097	35.915	35.852	50.050	28.579
Sep	34.389	37.107	34.320	35.129	36.328	55.263	29.061
Oct	49.822	56.489	48.129	45.248	44.750	53.352	37.781
Nov	50.525	53.376	59.217	51.692	61.666	53.447	52.776
Dec	48.308	42.044	44.263	39.767	41.737	70.666	43.262

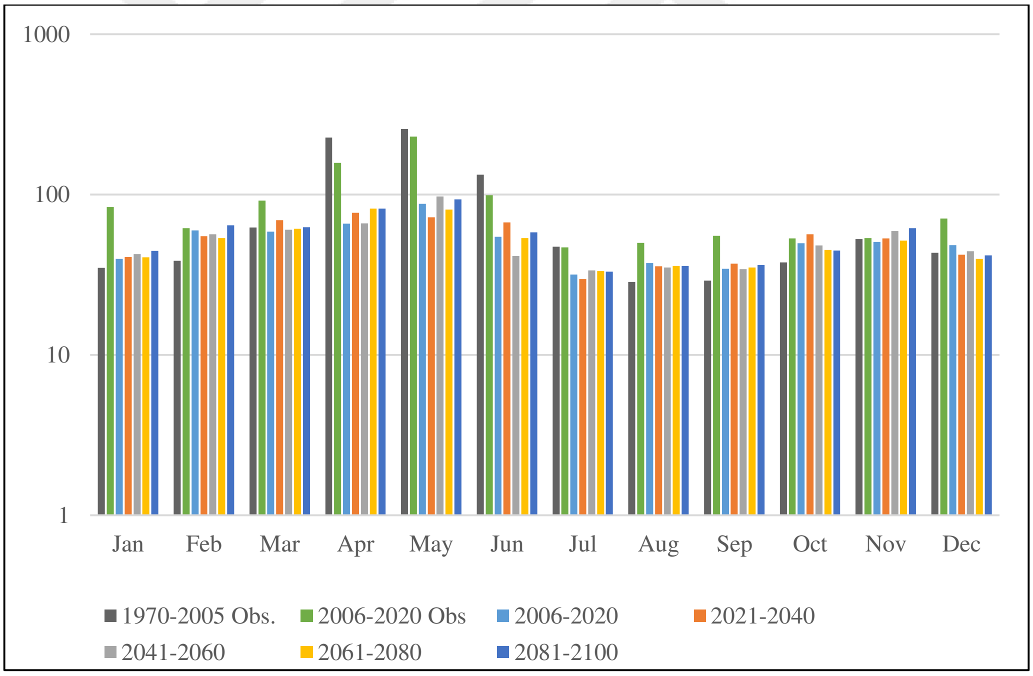

Figure E.2 The projection of average monthly discharge (Q) values under CanESM2 RCP4.5 scenario for different time periods by the GEP model in logarithmic scale

Table E.3 Monthly Q results under CanESM2 RCP8.5 scenario for four future periods by GEP model

Month	2006-2020	2021-2040	2041-2060	2061-2080	2081-2100	2006-2020 Obs	1970-2005 Obs.
Jan	47.613	43.462	37.279	45.743	44.169	83.501	34.879
Feb	58.030	56.996	49.536	54.443	42.748	61.749	38.603
Mar	61.077	68.508	63.026	53.467	48.856	91.420	62.145
Apr	66.358	74.900	67.990	69.744	68.542	157.821	225.895
May	118.895	142.728	89.598	78.031	88.003	229.123	256.626
Jun	54.558	62.134	48.619	55.093	35.346	98.818	132.901
Jul	35.429	32.386	32.439	34.775	35.356	46.822	47.360
Aug	35.932	31.218	36.854	37.162	40.105	50.050	28.579
Sep	40.986	34.941	39.022	36.078	34.236	55.263	29.061
Oct	45.927	44.643	45.169	43.800	39.398	53.352	37.781
Nov	51.050	49.448	50.536	49.756	46.583	53.447	52.776
Dec	41.925	40.238	39.654	41.799	42.279	70.666	43.262

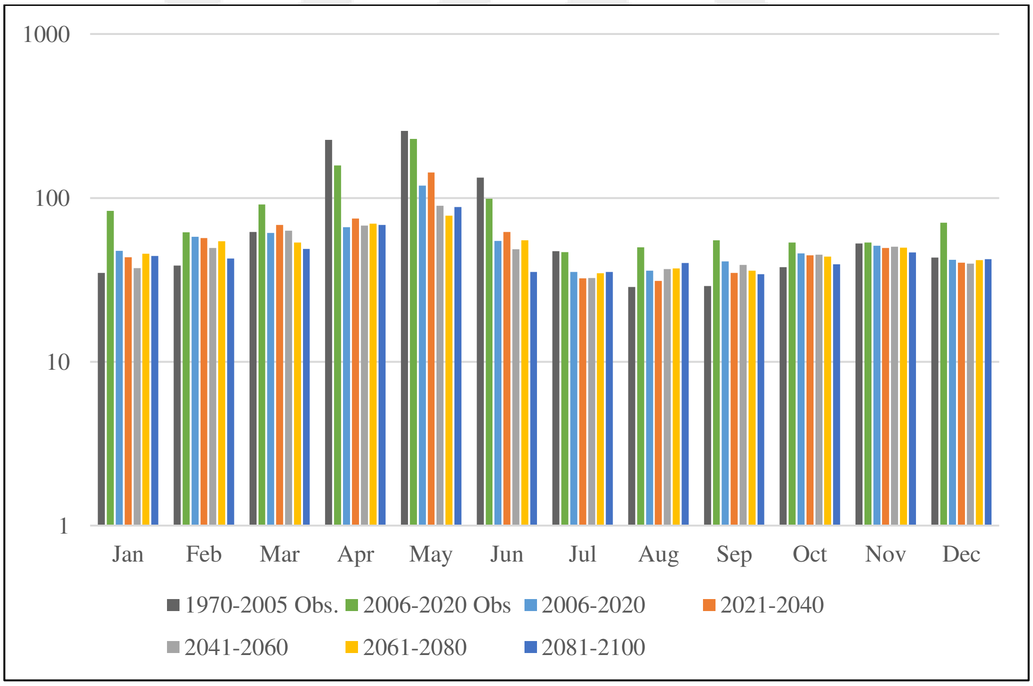

Figure E.3 The projection of average monthly discharge (Q) values under CanESM2 RCP8.5 scenario for different time periods by the GEP model in logarithmic scale

Table E.4 Monthly Q results under CanESM2 RCP2.6 scenario for four future periods by LGP model

Month	2006-2020	2021-2040	2041-2060	2061-2080	2081-2100	2006-2020 Obs	1970-2005 Obs.
Jan	41.037	54.528	45.006	39.953	63.487	83.501	34.879
Feb	88.835	59.224	48.511	63.718	83.280	61.749	38.603
Mar	47.042	57.469	59.746	72.294	52.098	91.420	62.145
Apr	125.813	109.716	126.010	127.836	70.901	157.821	225.895
May	117.107	164.842	136.664	141.026	117.979	229.123	256.626
Jun	59.426	60.506	62.176	80.152	74.411	98.818	132.901
Jul	63.030	55.317	54.068	54.353	52.919	46.822	47.360
Aug	44.095	45.167	43.173	35.575	39.514	50.050	28.579
Sep	88.457	85.157	98.327	80.072	83.358	55.263	29.061
Oct	38.077	53.177	48.928	67.084	50.318	53.352	37.781
Nov	43.114	51.821	46.900	45.365	44.324	53.447	52.776
Dec	42.809	45.193	49.129	42.160	42.462	70.666	43.262

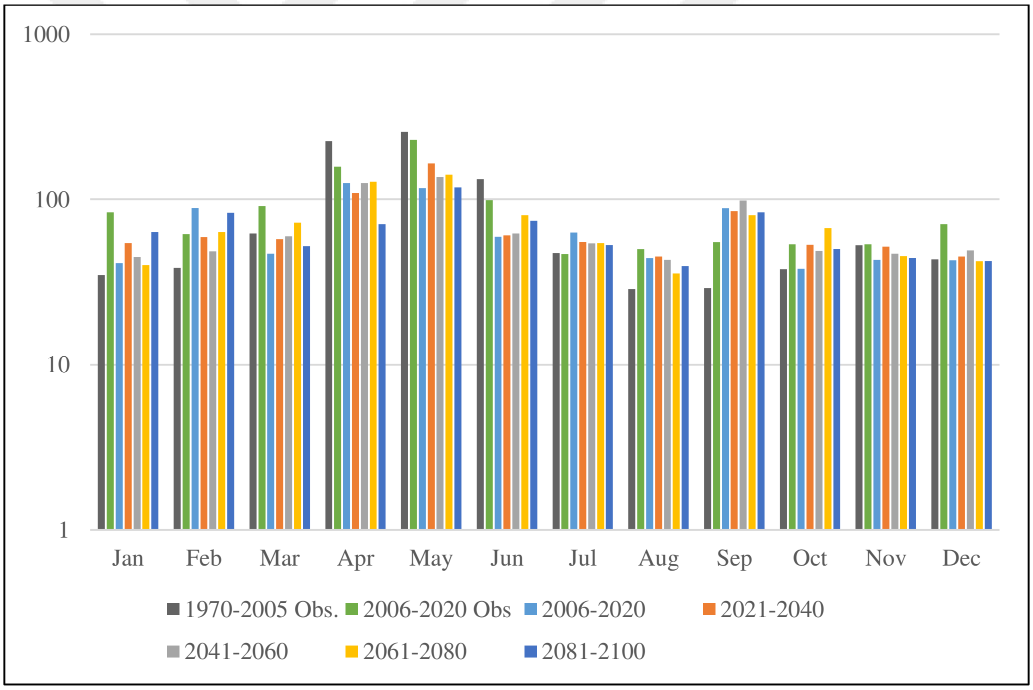

Figure E.4 The projection of average monthly discharge (Q) values under CanESM2 RCP2.6 scenario for different time periods by the LGP model in logarithmic scale

Table E.5 Monthly Q results under CanESM2 RCP4.5 scenario for four future periods by LGP model

Month	2006-2020	2021-2040	2041-2060	2061-2080	2081-2100	2006-2020 Obs	1970-2005 Obs.
Jan	44.645	44.306	40.298	44.150	42.802	83.501	34.879
Feb	46.171	46.887	48.260	43.795	48.019	61.749	38.603
Mar	56.104	66.469	59.678	53.344	50.031	91.420	62.145
Apr	136.945	119.146	137.755	120.130	90.840	157.821	225.895
May	154.814	146.155	158.868	114.138	124.927	229.123	256.626
Jun	60.596	65.034	50.669	52.838	65.167	98.818	132.901
Jul	53.573	55.702	61.006	61.511	58.983	46.822	47.360
Aug	43.443	40.780	46.035	43.895	48.367	50.050	28.579
Sep	85.037	86.111	101.968	79.805	105.728	55.263	29.061
Oct	45.839	47.058	42.582	38.305	37.563	53.352	37.781
Nov	47.309	44.814	50.986	47.277	52.025	53.447	52.776
Dec	62.500	44.217	45.550	41.951	42.453	70.666	43.262

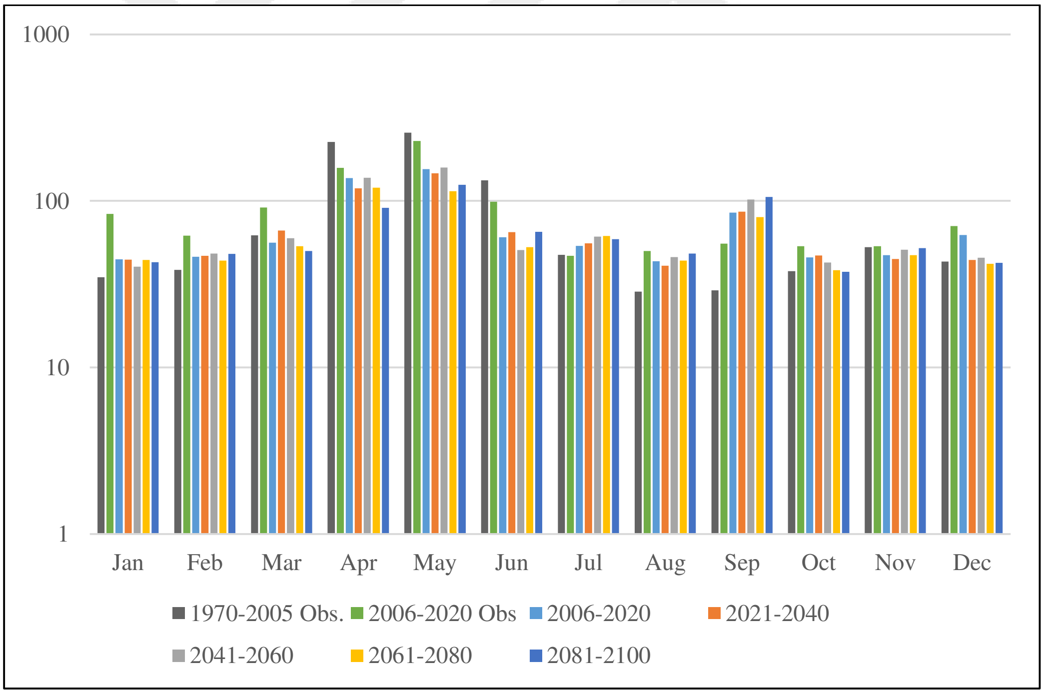

Figure E.5 The projection of average monthly discharge (Q) values under CanESM2 RCP4.5 scenario for different time periods by the LGP model in logarithmic scale

Table E.6 Monthly Q results under CanESM2 RCP8.5 scenario for four future periods by LGP model

Month	2006-2020	2021-2040	2041-2060	2061-2080	2081-2100	2006-2020 Obs	1970-2005 Obs.
Jan	76.659	61.480	85.434	78.112	106.920	83.501	34.879
Feb	55.560	55.693	45.769	56.584	42.180	61.749	38.603
Mar	56.612	62.346	43.463	65.469	40.911	91.420	62.145
Apr	93.403	141.751	116.462	118.282	99.546	157.821	225.895
May	137.334	159.408	159.829	129.160	119.487	229.123	256.626
Jun	61.778	81.449	58.618	42.377	50.205	98.818	132.901
Jul	51.519	54.920	58.781	68.245	61.241	46.822	47.360
Aug	44.047	41.146	40.008	44.301	41.899	50.050	28.579
Sep	89.539	99.836	97.467	85.233	83.425	55.263	29.061
Oct	39.904	38.265	38.195	54.117	47.679	53.352	37.781
Nov	48.286	47.062	47.662	46.462	45.518	53.447	52.776
Dec	45.962	43.869	43.814	45.006	44.965	70.666	43.262

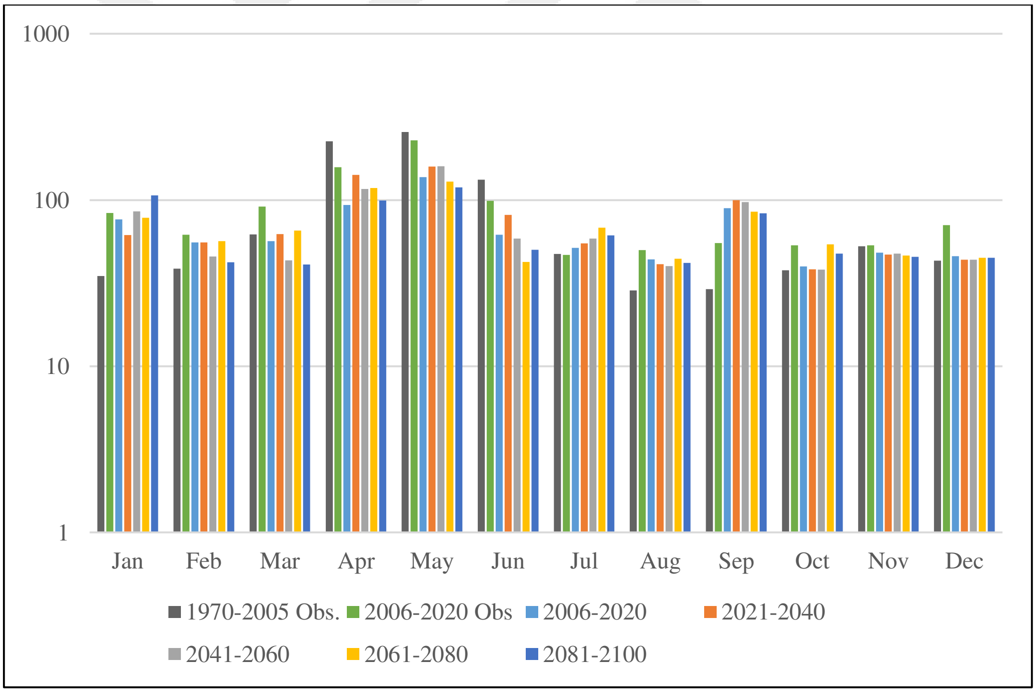

Figure E.6 The projection of average monthly discharge (Q) values under CanESM2 RCP8.5 scenario for different time periods by the LGP model in logarithmic scale

APPENDIX F

Tables and Figures for the comparison of observed and predicted monthly sediment yield values in 2006-2020 period for all models and scenarios

Table F.1 Monthly sediment yield values for observed and predicted by GEP model under RCP2.6 scenario in 2006-2020 period

Month	Qs GEP RCP2.6	Qs Obs.
Jan	309.011	2791.534
Feb	542.308	1315.825
Mar	1240.337	5536.644
Apr	1884.082	8367.325
May	2349.713	11276.306
Jun	1254.372	2101.516
Jul	488.728	1067.311
Aug	451.101	847.379
Sep	351.944	1681.025
Oct	330.739	1739.331
Nov	374.461	1817.919
Dec	298.822	1336.654

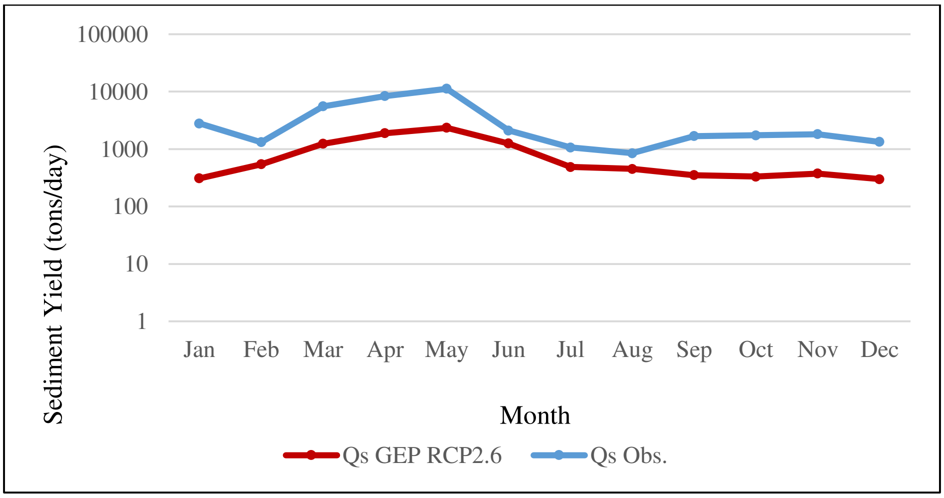

Figure F.1 Observed and predicted monthly Qs values for the 2006 – 2020 period by GEP model under RCP2.6 scenario, in logarithmic scale

Table F.2 Monthly sediment yield values for observed and predicted by LGP model under RCP2.6 scenario in 2006-2020 period

Month	Qs LGP RCP2.6	Qs Obs.
Jan	525.638	2791.534
Feb	1326.449	1315.825
Mar	962.111	5536.644
Apr	3697.208	8798.754
May	3255.292	11276.306
Jun	805.668	2101.516
Jul	221.937	1016.984
Aug	188.104	847.379
Sep	571.289	1919.909
Oct	270.867	1671.630
Nov	354.754	1817.919
Dec	514.751	1336.654

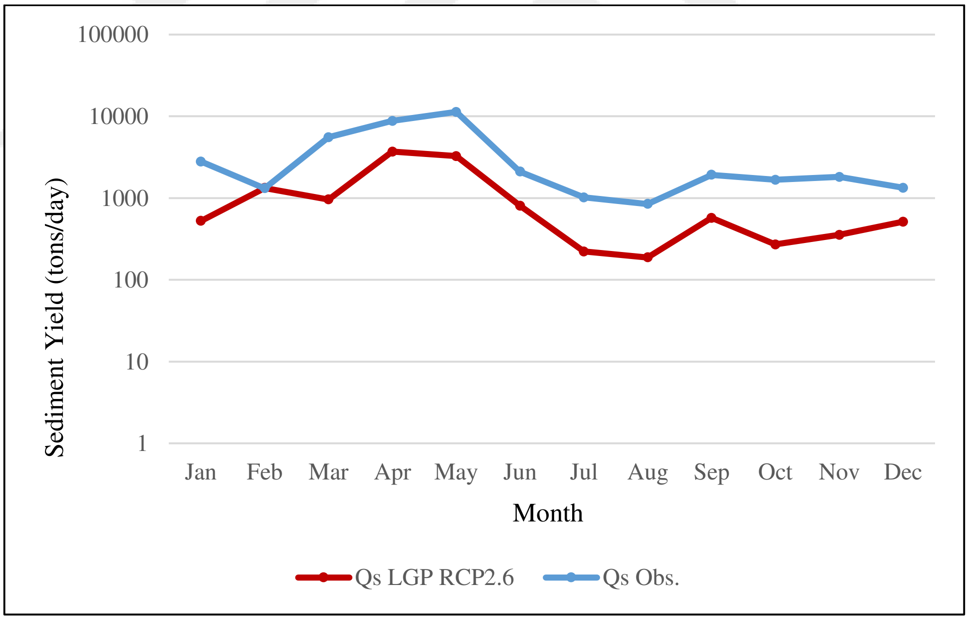

Figure F.2 Observed and predicted monthly Qs values for the 2006-2020 period by LGP model under RCP2.6 scenario, in logarithmic scale

Table F.3 Monthly sediment yield values for observed and predicted by GEP model under RCP4.5 scenario in 2006-2020 period

Month	Qs GEP RCP4.5	Qs Obs.
Jan	282.8241455	2596.68
Feb	554.0753204	1315.83
Mar	712.1210304	5536.64
Apr	2923.645239	8462.38
May	2788.522512	11276.3
Jun	1200.696957	2135.87
Jul	551.3602355	1039.4
Aug	575.5073009	847.379
Sep	374.8819128	1681.03
Oct	575.4100124	1763.61
Nov	679.086836	1817.92
Dec	364.2649141	1336.65

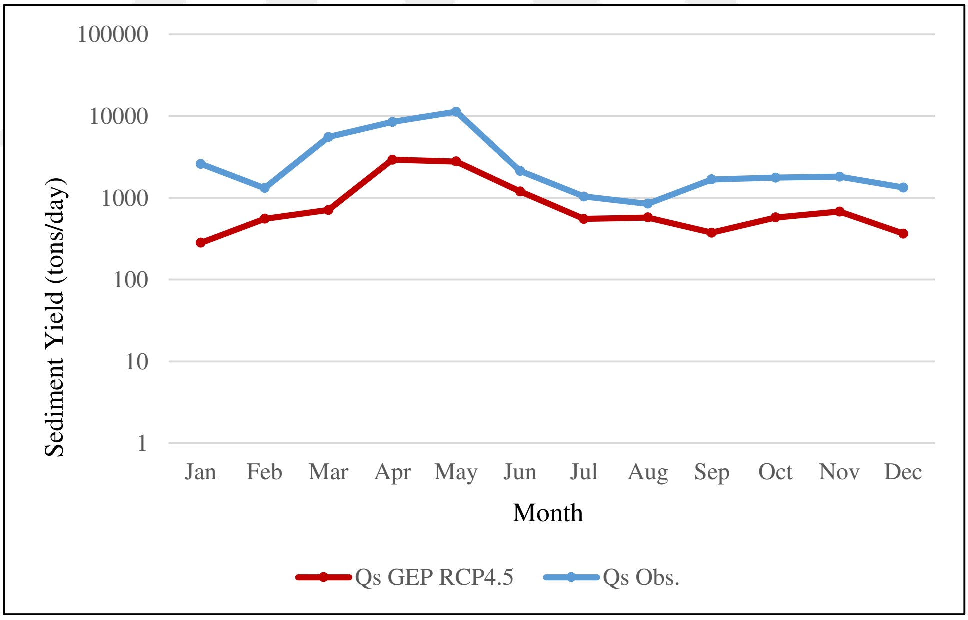

Figure F.3 Observed and predicted monthly Qs values for the 2006-2020 period by GEP model under RCP4.5 scenario, in logarithmic scale

Table F.4 Monthly sediment yield values for observed and predicted by LGP model under RCP4.5 scenario in 2006-2020 period

Month	Qs LGP RCP4.5	Qs Obs.
Jan	438.319	2596.678
Feb	820.934	1315.825
Mar	547.829	5536.644
Apr	4984.677	8799.101
May	2563.643	11276.306
Jun	1145.711	2135.868
Jul	158.355	1095.328
Aug	217.244	888.958
Sep	1051.682	1480.143
Oct	366.398	1671.630
Nov	419.603	1789.150
Dec	549.409	1336.654

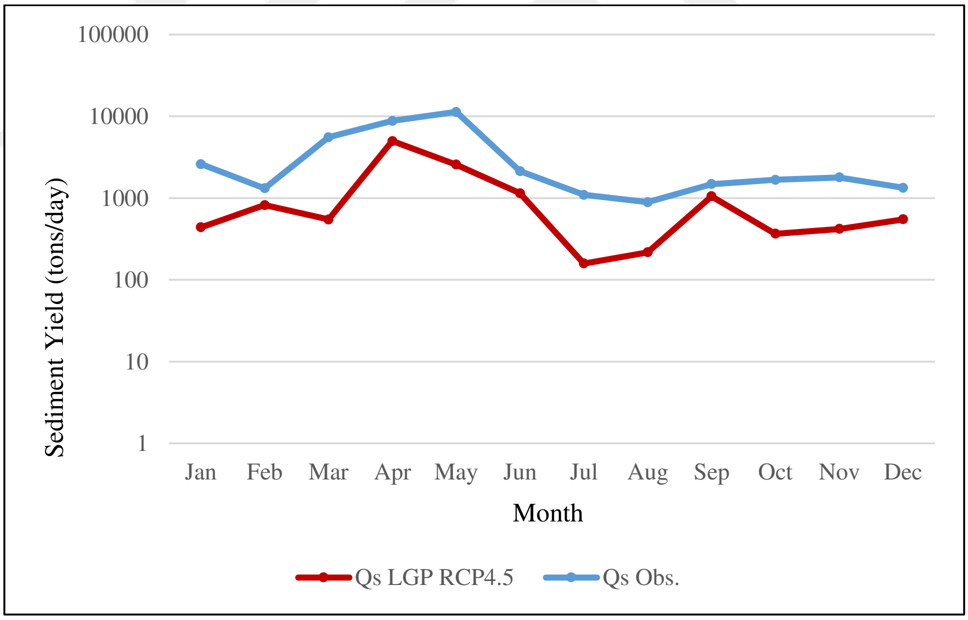

Figure F.4 Observed and predicted monthly Qs values for the 2006-2020 period by LGP model under RCP4.5 scenario, in logarithmic scale

Table F.5 Monthly sediment yield values for observed and predicted by GEP model under RCP8.5 scenario in 2006-2020 period

Month	Qs GEP RCP8.5	Qs Obs.
Jan	281.611	2791.534
Feb	398.080	1365.499
Mar	1161.944	5536.644
Apr	1550.973	8367.325
May	4791.360	11276.306
Jun	769.327	1810.566
Jul	493.525	1067.311
Aug	556.666	881.741
Sep	425.810	1762.806
Oct	549.950	1671.630
Nov	540.659	1817.919
Dec	277.397	1336.654

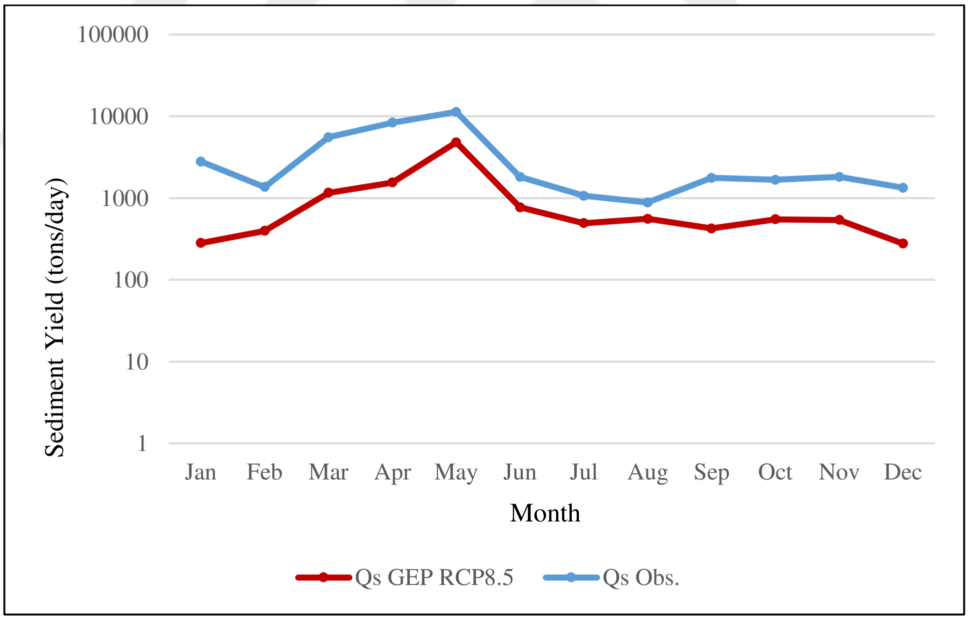

Figure F.5 Observed and predicted monthly Qs values for the 2006-2020 period by GEP model under RCP8.5 scenario, in logarithmic scale

Table F.6 Monthly sediment yield values for observed and predicted by LGP model under RCP8.5 scenario in 2006-2020 period

Month	Qs LGP RCP8.5	Qs Obs.
Jan	1491.541	2791.534
Feb	612.435	1388.627
Mar	847.742	3665.059
Apr	1650.139	8367.325
May	4306.875	11276.306
Jun	590.994	1874.639
Jul	162.582	1067.311
Aug	187.090	847.379
Sep	1314.477	1768.670
Oct	280.601	1623.119
Nov	342.441	1931.409
Dec	326.272	1336.654

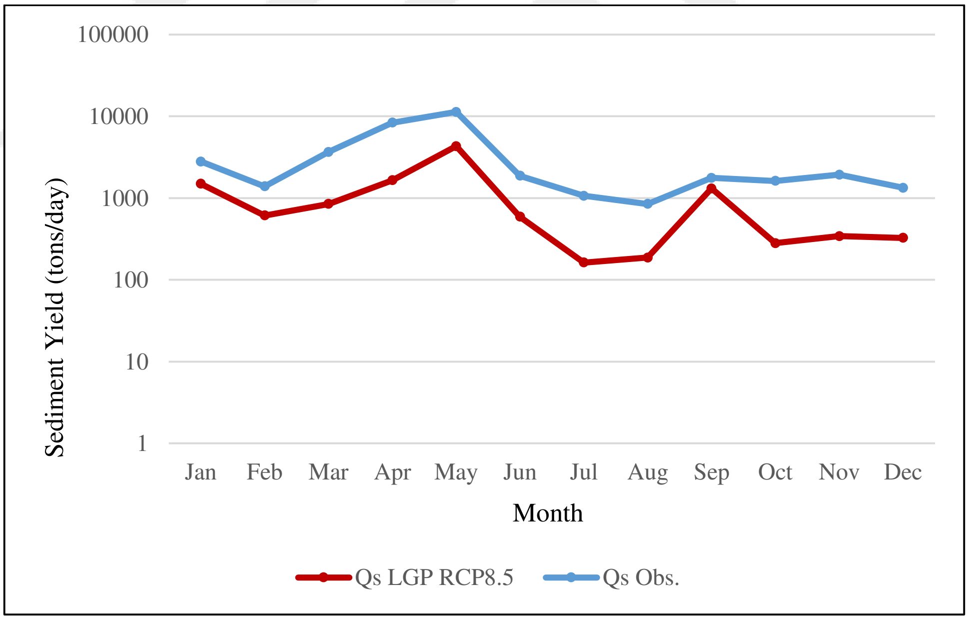

Figure F.6 Observed and predicted monthly Qs values for the 2006-2020 period by LGP model under RCP8.5 scenario, in logarithmic scale

APPENDIX G

Tables and Figures of averaged monthly Qs results under different CanESM2 scenarios for 20 - years of time periods by GEP and LGP models

Table G.1 Monthly Qs results under CanESM2 RCP2.6 scenario for four future periods by GEP model

Month	2006-2020	2021-2040	2041-2060	2061-2080	2081-2100	2006-2020 Obs.	1970-2005 Obs.
Jan	309.011	264.712	264.197	290.835	282.060	2791.534	250.372
Feb	542.308	422.077	572.563	395.136	394.692	1315.825	259.529
Mar	1240.337	925.817	1247.948	1180.614	1183.881	5536.644	1723.402
Apr	1884.082	1944.318	2788.294	2064.894	1651.203	8367.325	30300.244
May	2349.713	2328.572	3211.198	4075.671	4277.005	11276.306	17604.782
Jun	1254.372	1672.251	1837.519	1709.997	1598.965	2101.516	3501.974
Jul	488.728	619.373	568.482	620.930	768.520	1016.984	561.734
Aug	451.101	552.555	584.551	457.554	614.960	847.379	275.609
Sep	351.944	374.420	343.158	334.971	354.817	1681.025	187.883
Oct	330.739	534.215	471.465	596.989	539.545	1671.630	350.801
Nov	374.461	527.991	440.141	370.047	383.241	1817.919	597.090
Dec	298.822	302.733	317.519	274.485	266.188	1336.654	383.991

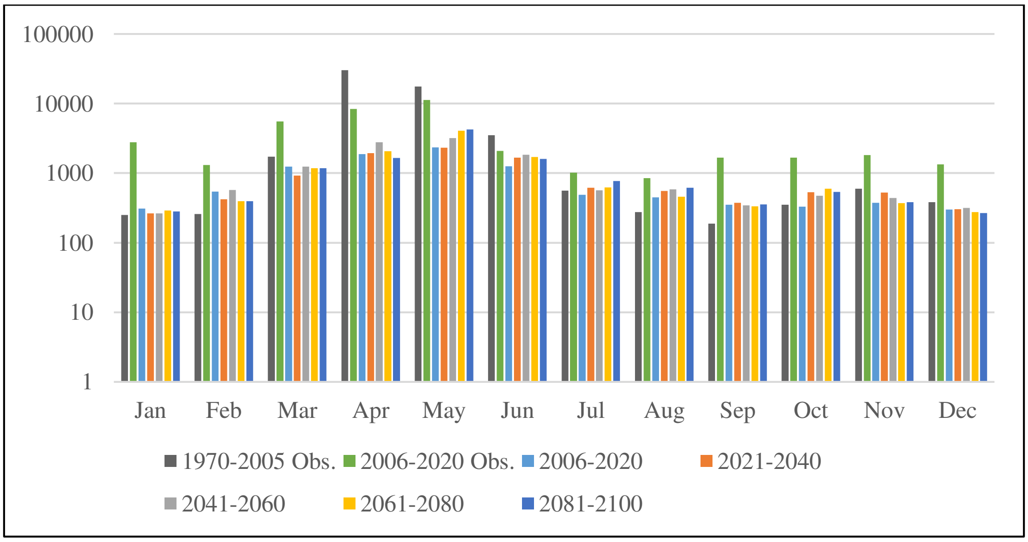

Figure G.1 The projection of average monthly sediment yield (Qs) values under CanESM2 RCP2.6 scenario for different time periods by the GEP model in logarithmic scale

Table G.2 Monthly Qs results under CanESM2 RCP4.5 scenario for four future periods by GEP model

Month	2006-2020	2021-2040	2041-2060	2061-2080	2081-2100	2006-2020 Obs.	1970-2005 Obs.
Jan	282.824	300.920	234.763	206.477	282.366	2791.534	250.372
Feb	554.075	525.932	563.901	450.387	411.717	1315.825	259.529
Mar	712.121	1433.491	1005.030	1169.996	805.132	5536.644	1723.402
Apr	2923.645	1465.072	1573.819	3322.567	1972.223	8367.325	30300.244
May	2788.523	2075.106	3351.386	2976.440	3110.656	11276.306	17604.782
Jun	1200.697	1563.891	694.121	1164.635	1240.066	2101.516	3501.974
Jul	551.360	501.763	417.413	464.756	497.436	1016.984	561.734
Aug	575.507	523.382	438.110	454.880	429.723	847.379	275.609
Sep	374.882	346.652	336.747	372.916	454.230	1681.025	187.883
Oct	575.410	1003.808	480.865	484.689	510.564	1671.630	350.801
Nov	679.087	612.450	364.470	493.708	693.416	1817.919	597.090
Dec	364.265	246.252	258.687	266.508	256.159	1336.654	383.991

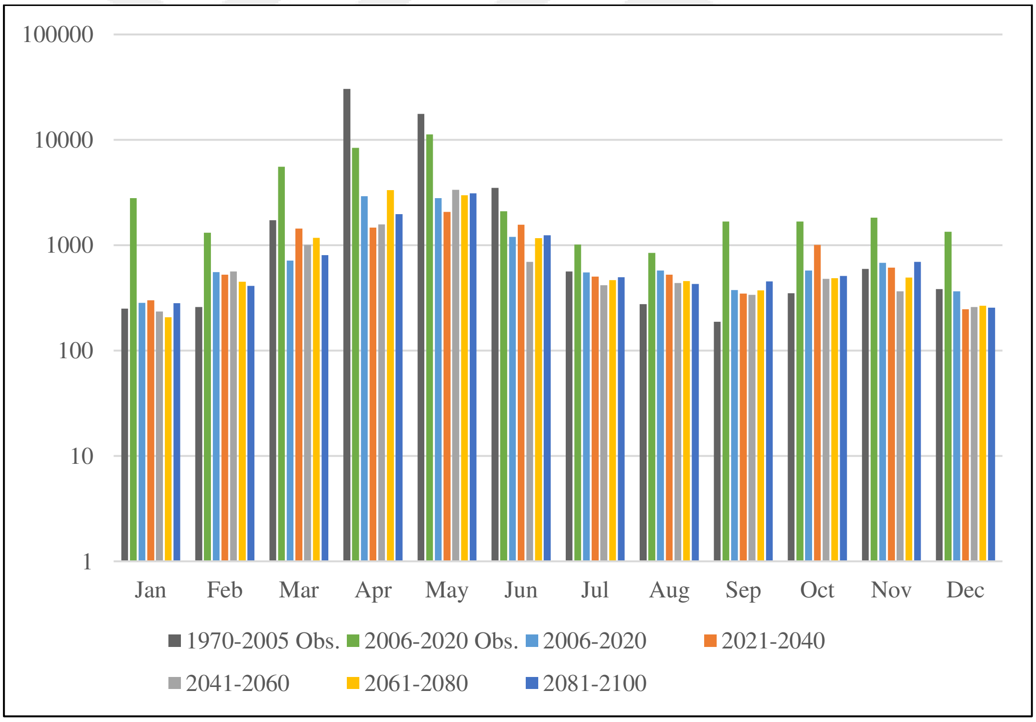

Figure G.2 The projection of average monthly sediment yield (Qs) values under CanESM2 RCP4.5 scenario for different time periods by the GEP model in logarithmic scale

Table G.3 Monthly Qs results under CanESM2 RCP8.5 scenario for four future periods by GEP model

Month	2006-2020	2021-2040	2041-2060	2061-2080	2081-2100	2006-2020 Obs.	1970-2005 Obs.
Jan	281.611	273.833	212.863	269.944	319.561	2791.534	250.372
Feb	398.080	561.904	532.567	446.964	344.999	1315.825	259.529
Mar	1161.944	1114.094	953.327	760.413	526.125	5536.644	1723.402
Apr	1550.973	1848.291	1367.345	1441.688	1524.349	8367.325	30300.244
May	4791.360	5114.560	2939.706	2277.699	3457.668	11276.306	17604.782
Jun	769.327	1982.152	881.945	1349.909	1034.154	2101.516	3501.974
Jul	493.525	479.654	429.633	465.973	416.381	1016.984	561.734
Aug	556.666	439.262	464.238	505.680	376.026	847.379	275.609
Sep	425.810	444.576	380.950	355.397	350.344	1681.025	187.883
Oct	549.950	391.287	441.671	540.881	433.412	1671.630	350.801
Nov	540.659	558.824	582.063	439.255	524.102	1817.919	597.090
Dec	277.397	295.396	228.124	311.626	307.937	1336.654	383.991

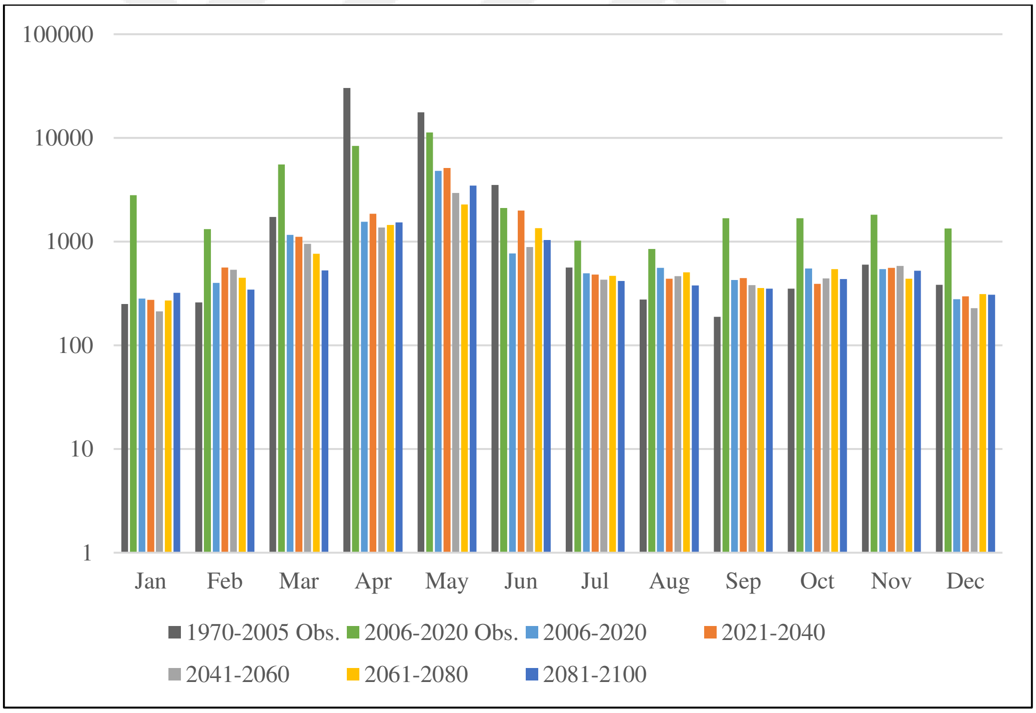

Figure G.3 The projection of average monthly sediment yield (Qs) values under CanESM2 RCP8.5 scenario for different time periods by the GEP model in logarithmic scale

Table G.4 Monthly Qs results under CanESM2 RCP2.6 scenario for four future periods by LGP model

Month	2006-2020	2021-2040	2041-2060	2061-2080	2081-2100	2006-2020 Obs.	1970-2005 Obs.
Jan	525.638	517.506	466.249	520.508	564.729	2791.534	250.372
Feb	1326.449	627.117	745.624	992.136	1575.259	1315.825	259.529
Mar	962.111	1089.063	1855.771	818.371	1411.481	5536.644	1723.402
Apr	3697.208	3619.869	6626.952	5679.380	3990.280	8367.325	30300.244
May	3255.292	7772.107	3543.107	7231.103	2570.923	11276.306	17604.782
Jun	805.668	926.441	699.578	1802.892	1153.136	2101.516	3501.974
Jul	221.937	172.758	186.882	174.719	142.084	1016.984	561.734
Aug	188.104	155.719	203.952	168.704	180.147	847.379	275.609
Sep	571.289	881.799	730.924	655.866	916.506	1681.025	187.883
Oct	270.867	325.361	430.717	289.345	312.962	1671.630	350.801
Nov	354.754	348.725	363.430	354.045	289.596	1817.919	597.090
Dec	514.751	568.244	350.258	323.746	313.745	1336.654	383.991

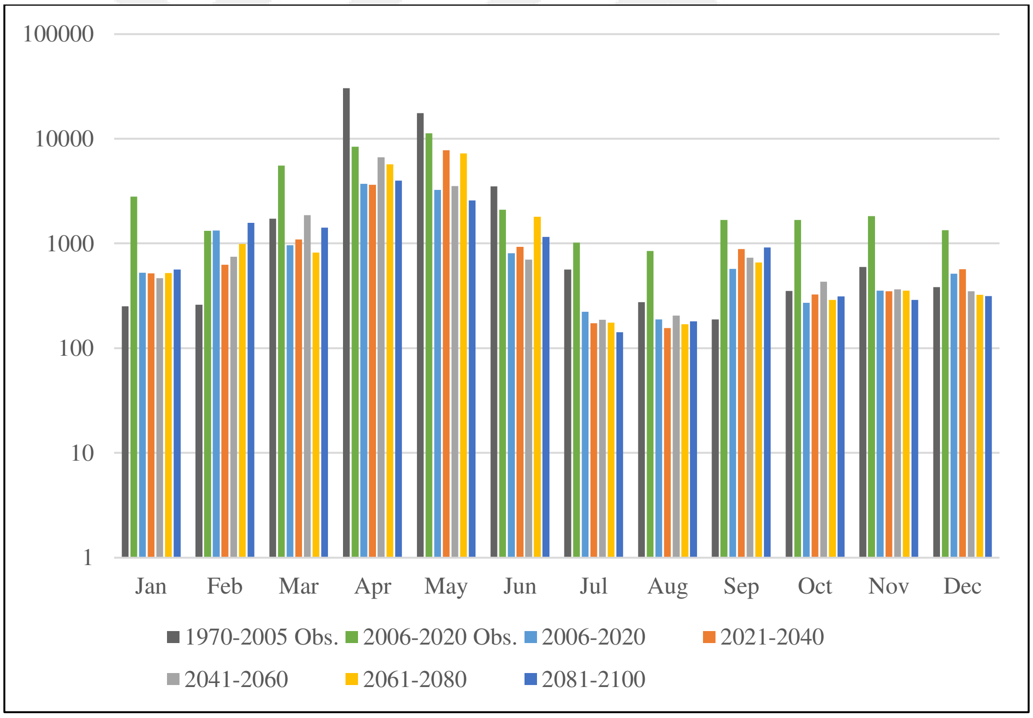

Figure G.4 The projection of average monthly sediment yield (Qs) values under CanESM2 RCP2.6 scenario for different time periods by the LGP model in logarithmic scale

Table G.5 Monthly Qs results under CanESM2 RCP4.5 scenario for four future periods by LGP model

Month	2006-2020	2021-2040	2041-2060	2061-2080	2081-2100	2006-2020 Obs.	1970-2005 Obs.
Jan	438.319	323.710	268.314	267.264	332.860	2791.534	250.372
Feb	820.934	522.486	516.670	434.353	424.786	1315.825	259.529
Mar	547.829	844.976	704.754	475.792	537.944	5536.644	1723.402
Apr	4984.677	2639.544	3824.487	4374.978	2170.789	8367.325	30300.244
May	2563.643	3240.762	4127.908	3516.484	2554.881	11276.306	17604.782
Jun	1145.711	1101.849	526.781	605.405	903.263	2101.516	3501.974
Jul	158.355	150.590	174.868	182.586	153.933	1016.984	561.734
Aug	217.244	231.699	221.982	187.929	198.524	847.379	275.609
Sep	1051.682	712.784	2767.272	989.102	1121.845	1681.025	187.883
Oct	366.398	443.150	279.696	275.350	273.722	1671.630	350.801
Nov	419.603	505.756	391.522	453.920	500.361	1817.919	597.090
Dec	549.409	291.947	286.831	310.683	375.175	1336.654	383.991

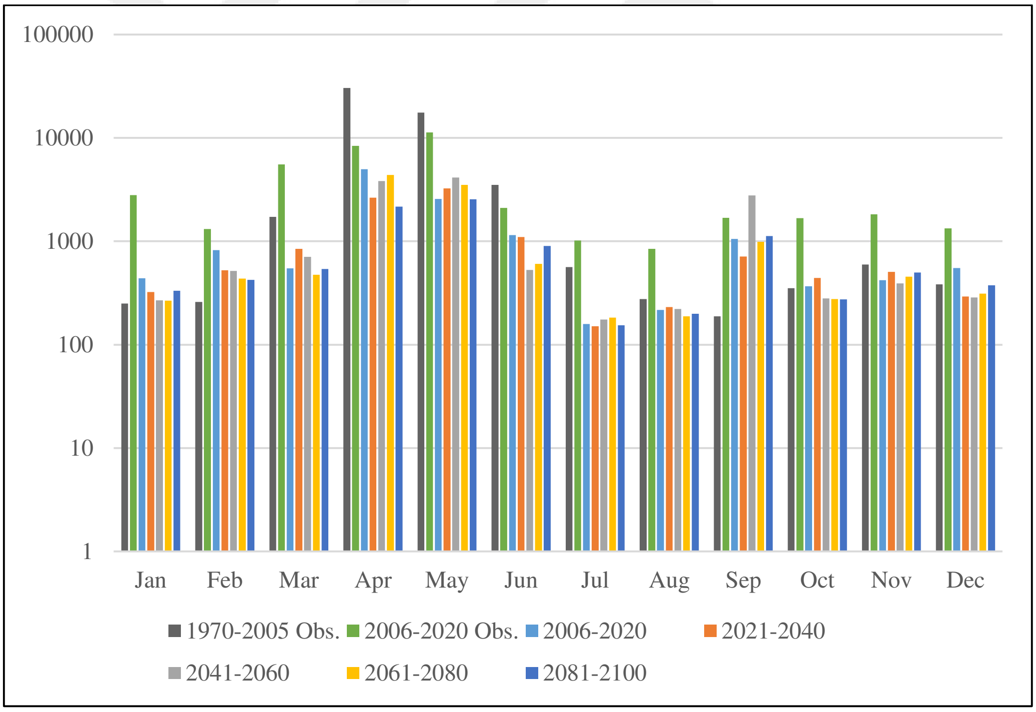

Figure G.5 The projection of average monthly sediment yield (Qs) values under CanESM2 RCP4.5 scenario for different time periods by the LGP model in logarithmic scale

Table G.6 Monthly Qs results under CanESM2 RCP8.5 scenario for four future periods by LGP model

Month	2006-2020	2021-2040	2041-2060	2061-2080	2081-2100	2006-2020 Obs.	1970-2005 Obs.
Jan	1491.541	534.198	854.709	712.674	1302.540	2791.534	250.372
Feb	612.435	508.211	468.645	386.966	268.961	1315.825	259.529
Mar	847.742	1249.050	642.071	799.308	1014.097	5536.644	1723.402
Apr	1650.139	8838.704	5248.427	6756.753	3135.332	8367.325	30300.244
May	4306.875	5816.705	4086.355	3345.891	2373.215	11276.306	17604.782
Jun	590.994	1352.140	503.095	1141.062	614.883	2101.516	3501.974
Jul	162.582	190.415	182.045	170.857	185.691	1016.984	561.734
Aug	187.090	146.807	202.014	191.564	217.624	847.379	275.609
Sep	1314.477	965.987	1219.658	787.685	586.857	1681.025	187.883
Oct	280.601	257.945	269.211	296.460	242.159	1671.630	350.801
Nov	342.441	347.595	306.656	314.436	424.485	1817.919	597.090
Dec	326.272	391.307	291.503	388.244	404.274	1336.654	383.991

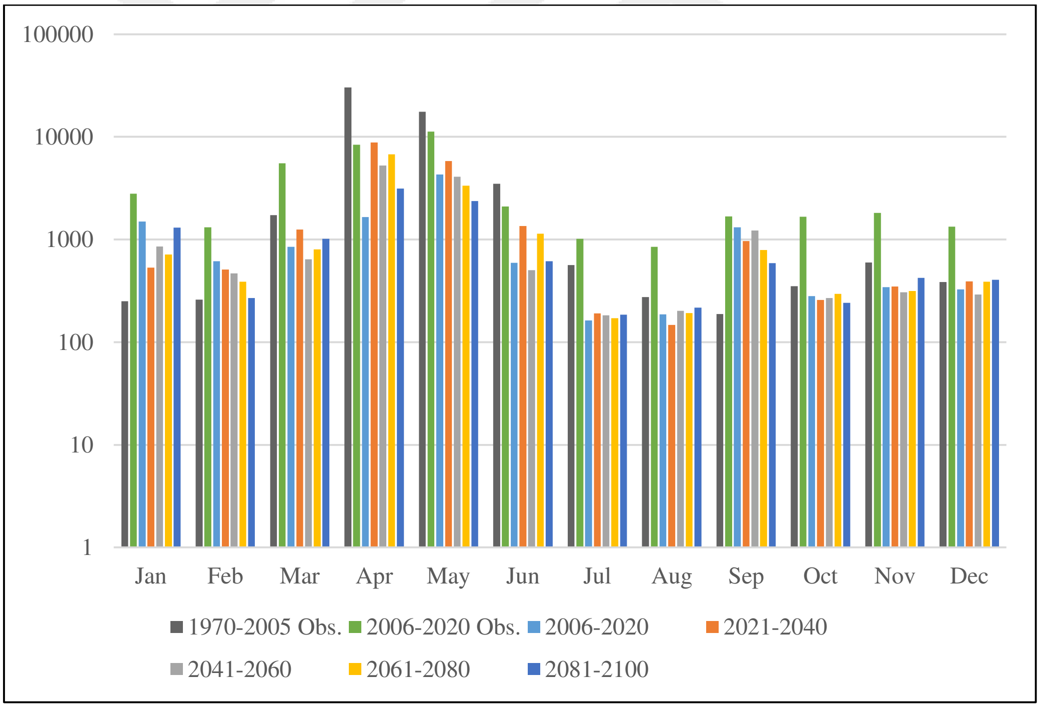

Figure G.6 The projection of average monthly sediment yield (Qs) values under CanESM2 RCP8.5 scenario for different time periods by the LGP model in logarithmic scale

CURRICULUM VITAE

PERSONAL INFORMATION

Name, Surname: Muhammed Vedat GÜN

EDUCATION

- Gaziantep University, Master of Science in Civil Engineering, Gaziantep / Türkiye (2020 – ongoing)
- Gaziantep University, Bachelor's Degree in Civil Engineering, Gaziantep / Türkiye (2013 – 2018)

EXPERIENCE

- Research Assistant, Batman University, Batman / Türkiye (2021 – ongoing)
- Civil Engineer, Limak Group of Companies, Gebze & Sakarya / Türkiye (2018 – 2020)