RECONSTRUCTION OF RIVER FLOWS BY ERA5 REANALYSIS DATASET AND LONG SHORT-TERM MEMORY MODELS

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

BY

İSMAİL ÜNLÜTÜRK

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

SEPTEMBER 2024

Approval of the thesis:

RECONSTRUCTION OF RIVER FLOWS BY ERA5 REANALYSIS DATASET AND LONG SHORT-TERM MEMORY MODELS

submitted by **İSMAİL ÜNLÜTÜRK** in partial fulfillment of the requirements for the degree of **Master of Science** i**n Civil Engineering, Middle East Technical University** by,

Prof. Dr. Naci Emre Altun
Dean, **Graduate School of Natural and Applied Sciences** _____

Prof. Dr. Erdem Canbay
Head of the Department, **Civil Engineering** _____

Assoc. Prof. Dr. Ali Ercan
Supervisor, **Civil Engineering, METU** _____

Examining Committee Members:

Prof. Dr. A. Burcu Altan Sakarya
Civil Engineering, METU _____

Assoc. Prof. Dr. Ali Ercan
Civil Engineering, METU _____

Prof. Dr. M. Tuğrul Yılmaz
Civil Engineering, METU _____

Prof. Dr. İsmail Yücel
Civil Engineering, METU _____

Asst. Prof. Dr. Kutay Yılmaz
Civil Engineering, Yaşar University _____

Date: 03.09.2024

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

Name Last name : İsmail Ünlütürk

Signature :

ABSTRACT

RECONSTRUCTION OF RIVER FLOWS BY ERA5 REANALYSIS DATASET AND LONG SHORT-TERM MEMORY MODELS

Ünlütürk, İsmail
Master of Science, Civil Engineering
Supervisor: Assoc. Prof. Dr. Ali Ercan

September 2024, 89 pages

Although long-term and continuous river flow data is necessary for water resources management, flood and drought analysis, design and management of hydraulic infrastructures, climate change research, and predicting river flows, there are limited river flow data around the world. Climate Reanalysis datasets provide the most complete picture currently possible of past weather and climate. Recently, deep learning has shown great success in various fields including hydrology and climate research.

Within the above framework, the scope of this study is to develop a methodology to reconstruct historical daily and monthly river flows by a Long Short-Term Memory (LSTM) model, utilizing the ERA5 Reanalysis climate dataset as inputs. As a demonstration, the proposed methodology is used at selected stations in the Seyhan and Euphrates Watersheds in Türkiye.

During the training period (2000-2011), correlation coefficient (R) is within 0.96-0.99 for monthly flows and 0.92-0.98 for daily flows, Nash-Sutcliffe Model Efficiency (NSE) value is within 0.87-0.99 for monthly flows and 0.80-0.96 for daily

flows, and Kling-Gupta Efficiency (KGE) value is within 0.81-0.95 for monthly flows and 0.83-0.92 for daily flows. On the other hand, during the test period (2012-2015), R ranges within 0.92-0.97 for monthly flows and 0.87-0.91 for daily flows, NSE value ranges within 0.84-0.92 for monthly flows, and 0.73-0.82 for daily flows, and KGE value ranges within 0.82-0.90 for monthly flows, and 0.72-0.86 for daily flows.

To conclude, the proposed river flow reconstruction methodology can successfully reconstruct river flows at selected locations if sufficiently long training and test datasets are available.

Keywords: Reconstruction, River flow, LSTM, ERA5 Climate Reanalysis, Deep Learning

ÖZ

ERA5 REANALİZ VERİ SETİ VE UZUN KISA SÜRELİ BELLEK MODELLERİ İLE NEHİR AKIMLARININ YENİDEN OLUŞTURULMASI

Ünlütürk, İsmail
Yüksek Lisans, İnşaat Mühendisliği
Tez Yöneticisi: Doç. Dr. Ali Ercan

Eylül 2024, 89 sayfa

Uzun vadeli ve devamlı nehir akım verileri, su kaynakları yönetimi, taşkın ve kuraklık analizi, hidrolik altyapıların tasarımı ve yönetimi, iklim değişikliği araştırmaları ve nehir akışlarının tahmin edilmesi için gerekli olmasına rağmen, dünya çapında sınırlı nehir akım verisi bulunmaktadır. İklim reanalizi veri setleri, geçmişteki hava durumu ve iklimin şu anda mümkün olan en eksiksiz resmini sunmaktadır. Son zamanlarda derin öğrenme, hidroloji ve iklim araştırmaları da dahil olmak üzere çeşitli alanlarında büyük başarı göstermiştir.

Bu çerçevede, bu çalışmanın kapsamı, ERA5 Reanaliz iklim veri setini girdi olarak kullanan bir Uzun Kısa Süreli Bellek (LSTM) modeli ile tarihsel günlük ve aylık nehir akımlarını yeniden oluşturmak için bir metodoloji geliştirmektir. Örnek olarak, önerilen metodoloji Türkiyedeki Seyhan ve Fırat Havzalarında seçilen istasyonlarda kullanılmıştır.

LSTM modeli kalibrasyon döneminde (2000-2011), Korelasyon Katsayısı (R) aylık akışlar için 0,96-0,99 ve günlük akışlar için 0,92-0,98 aralığında; Nash-Sutcliffe Model Verimliliği (NSE) değeri aylık akışlar için 0,87-0,99 ve günlük akışlar için

0,80-0,96 aralığında; Kling-Gupta Verimliliği (KGE) değeri ise aylık akışlar için 0,81-0,95 ve günlük akışlar için 0,83-0,92 aralığındadır. Öte yandan, test döneminde (2012-2015), R aylık akışlar için 0,92-0,97 ve günlük akışlar için 0,87-0,91 aralığında; NSE değeri aylık akışlar için 0,84-0,92 ve günlük akışlar için 0,73-0,82 aralığında; KGE değeri ise aylık akışlar için 0,82-0,90 ve günlük akışlar için 0,72-0,86 aralığında değişmektedir.

Sonuç olarak, önerilen nehir akımı yeniden yapılandırma metodolojisi, yeterince uzun kalibrasyon ve test veri setleri mevcutsa, seçilen konumlardaki nehir akımları başarılı bir şekilde yeniden oluşturulabilinir.

Anahtar Kelimeler: Yeniden Yapılandırma, Nehir Akımı, LSTM, ERA5 İklim Reanalizi, Derin Öğrenme



To My Family

ACKNOWLEDGMENTS

First of all, I would like to express my most profound gratitude to my supervisor Assoc. Prof. Dr. Ali Ercan for guiding and sharing his experiences with me to conduct this study. Also, I would like to thank Assoc. Prof. Dr. Ali Ercan for his friendly and sincere approach to all my concerns and his mentorship, which has made an invaluable contribution to my academic and professional development.

I would like to thank my thesis committee members Prof. Dr. Ayşe Burcu Altan Sakarya, Prof. Dr. Mustafa Tuğrul Yılmaz, Prof. Dr. İsmail Yücel and Asst. Prof. Dr. Kutay Yılmaz.

I would also like to thank all the research assistants and staff at the METU Hydraulics Laboratory, especially Research Assistant Hasan Hüseyin Yıldırım, for their continuous support and assistance as well as for their warm and welcoming attitude.

In addition, I would like to express my sincere thanks to all my close friends who made this challenging process more enjoyable.

Finally, I would like to extend my deepest thanks to my mother, Rahime Ünlütürk, and my father, Hüdayi Ünlütürk, who have supported me throughout my life and have always been by my side.

TABLE OF CONTENTS

ABSTRACT v

ÖZ vii

ACKNOWLEDGMENTS x

TABLE OF CONTENTS xi

LIST OF TABLES xiv

LIST OF FIGURES xv

LIST OF ABBREVIATIONS xix

CHAPTERS

1 INTRODUCTION 1

1.1 Introduction 1

1.2 Scope of the Study 4

1.3 Contents of the Chapters 5

2 LITERATURE REVIEW 7

2.1 Reconstruction of River Flow Datasets 7

2.2 River Flow Modeling 10

2.2.1 Conceptual Models 10

2.2.2 Physical Models 11

2.2.3 Data-Driven Models 11

2.3 Comparison of Different Hydrological Models with LSTM Model 14

3 STUDY AREA AND DATASETS 17

3.1 Study Area: Seyhan and Euphrates River Watersheds 17

3.2 Datasets 20

3.2.1 Observed River Flow Data 20
3.2.2 ERA5 Dataset 22
4 METHODOLOGY 25
4.1 Deep Learning Models 25
4.1.1 LSTM Model 26
4.2 Normalization 29
4.3 Evaluation Metrics 30
Root Mean Square Error (RMSE) 30
Correlation Coefficient (R) 31
Nash-Sutcliffe Model Efficiency Coefficient (NSE) 31
Kling-Gupta Efficiency (KGE) 32
4.4 Plots and Diagrams 32
Time Series Plot 33
Scatter Plot 33
Taylor Diagram 33
Flow Frequency Curve 34
5 RESULTS AND DISCUSSION 37
5.1 Flow Reconstruction in the daily timescale 37
5.1.1 Daily flows at D18A19 at Adana Feke in Seyhan Watershed 40
5.1.2 Daily Flows at D18A27 at Niğde Elekgölü in Seyhan Watershed 57
5.1.3 Daily Flows at D21A167 at Erzurum Hinis in Euphrates Watershed 65
6 CONCLUDING REMARKS AND FUTURE DIRECTIONS 73
REFERENCES 77
APPENDICES

A. Input Variable Description in ERA5.....89

LIST OF TABLES

TABLES

Table 3.1 Information on flow stations that were used in this study. 21
Table 5.1 Input variable cases for river flow simulation at station D18A19 in Seyhan Watershed. 38
Table 5.2 Input variable cases for river flow simulation at station D18A27 in Seyhan Watershed. 39
Table 5.3 Input variable cases for river flow simulation at station D21A167 in the Euphrates Watershed. 39
Table 5.4 Performance statistics of the input variable cases for flow station D18A19 in Seyhan Watershed. 42
Table 5.5 Performance statistics of the input variable cases for flow station D18A27 in Seyhan Watershed. 58
Table 5.6 Performance statistics of the input variable cases for flow station D21A167 in Euphrates Watershed. 66
Table A.1 Modified version of input variable descriptions in ERA5 (Hersbach et al., 2023). 89

LIST OF FIGURES

FIGURES

Figure 3.1 Hydrometeorological map and border of Seyhan Watershed (SYGM, 2020b). 18
Figure 3.2 Main River tributaries and border of Euphrates River Basin (SYGM, 2020a). 19
Figure 3.3 River network of Seyhan Watershed and location and selected grids for stations D18A19, and D18A27. 23
Figure 3.4 River network of Euphrates-Tigris Watershed and location and selected grids for station D21A167. 24
Figure 4.1 Venn diagram of artificial intelligence, machine learning and deep learning (Sarker, 2021). 26
Figure 4.2 LSTM Memory Cell structure (Fan et al., 2020). 27
Figure 4.3 Sample Taylor Diagram. 34
Figure 5.1 Taylor diagram of River flow simulations during a) the training period of 2000-2011, and b) the test period of 2012-2015. 43
Figure 5.2 River flow simulation timeseries and performance statistics at station D18A19 during **a)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in daily timescale; **b)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in monthly timescale. 44
Figure 5.3 River flow simulation scatter plots at station D18A19 during **a)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in daily timescale; **b)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in monthly timescale. 45
Figure 5.4 Flow frequency curves at station D18A19 during **a)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in daily timescale; **b)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in monthly timescale. 46

Figure 5.5 Flow frequency curves of observed data at station D18A19 **a)** in daily timescale (left figure); **b)** in monthly timescale (right figure). 47
Figure 5.6 River flow simulation timeseries and performance statistics at station D18A19 during second test period of 1991-1994. 48
Figure 5.7 River flow simulation scatter plots of station D18A19 during second test period of 1991-1994. 48
Figure 5.8 Flow frequency curve at station D18A19 during test period of 1991-1994. 49
Figure 5.9 Flow frequency curves of observed data at station D18A19 during training and the second test period. 49
Figure 5.10 Reconstructed flow data at station D18A19 since 1940 **a)** in daily timescale; **b)** in monthly timescale. 50
Figure 5.11 River flow simulation timeseries and performance statistics at station D18A19 during **a)** the training period of 1991-2011 (left figure); **b)** the test period of 2012-2015 (right figure) in daily timescale. 52
Figure 5.12 River flow simulation timeseries and performance statistics at station D18A19 during **a)** the training period of 1991-2011 (left figure); **b)** the test period of 2012-2015 (right figure) in daily timescale. 52
Figure 5.13 Flow frequency curves at station D18A19 during **a)** the training period of 2000-2011 (left figure); **b)** the test period of 2012-2015 (right figure) in daily timescale. 53
Figure 5.14 Flow frequency curves of observed data during the training and test periods at station D18A19 with longer training size. 53
Figure 5.15 River flow simulation timeseries and performance statistics at station D18A19 during **a)** the training period of 2004-2015 (left figure); **b)** the test period of 2000-2003 (right figure) in daily timescale. 55
Figure 5.16 River flow simulation timeseries and performance statistics at station D18A19 during **a)** the training period of 2004-2015 (left figure); **b)** the test period of 2000-2003 (right figure) in daily timescale. 55

Figure 5.17 Flow frequency curves at station D18A19 during **a)** the training period of 2004-2015 (left figure); **b)** the test period of 2000-2003 (right figure) in daily timescale. 56
Figure 5.18 Flow frequency curves of observed data during the training and test periods at station D18A19. 56
Figure 5.19 Taylor diagrams at station D18A27 during **a)** the training period of 2000-2011, and **b)** the test period of 2012-2015. 59
Figure 5.20 River flow simulation timeseries and performance statistics at station D18A27 during **a)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in daily timescale; **b)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in monthly timescale. 60
Figure 5.21 River flow simulation scatter plots at station D18A27 during **a)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in daily timescale; **b)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in monthly timescale. 61
Figure 5.22 Flow frequency curves at station D18A27 during **a)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in daily timescale; **b)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in monthly timescale. 62
Figure 5.23 Flow frequency curves of observed data at station D18A27 **a)** in daily timescale (left figure); **b)** in monthly timescale (right figure). 63
Figure 5.24 Reconstructed flow data at station D18A27 since 1940 **a)** in daily timescale; **b)** in monthly timescale. 64
Figure 5.25 Taylor diagrams at station D21A167 during **a)** the training period of 2000-2011, and **b)** the test period of 2012-2015. 67
Figure 5.26 River flow simulation timeseries and performance statistics at station D21A167 during **a)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in daily timescale; **b)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in monthly timescale. 68

Figure 5.27 River flow simulation scatter plots at station D21A167 during **a)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in daily timescale; **b)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in monthly timescale. 69
Figure 5.28 Flow frequency curves at station D21A167 during **a)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in daily timescale; **b)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in monthly timescale. 70
Figure 5.29 Flow frequency curves of observed data at station D21A167 **a)** in daily timescale (left figure); **b)** in monthly timescale (right figure). 71
Figure 5.30 Reconstructed flow data at station D21A167 since 1940 **a)** in daily timescale; **b)** in monthly timescale. 72

LIST OF ABBREVIATIONS

ABBREVIATIONS

ADAM	Adaptive Moment Estimation
ANFIS	Adaptive Neuro Fuzzy Inference System
ANN	Artificial Neural Network
CNN	Convolutional Neural Network
DSİ	State Hydraulic Works
ECMWF	European Centre for Medium-Range Weather Forecasts
GCMs	General Circulation Models
GRU	Gated Recurrent Unit
HBV	Hydrologiska Byråns Vattenbalansavdelning
HEC-HMS	Hydrologic Engineering Center - Hydrologic Modeling System
IFS	Integrated Forecasting System
KGE	Kling-Gupta Efficiency
LSTM	Long Short-Term Memory
MLP	Multilayer Perceptron
NSE	Nash-Sutcliffe Model Efficiency
NWM	National Water Model
PSO	Particle Swarm Optimization
R	Correlation Coefficient
RE	Relative Error
RMSE	Root Mean Square Error
RMSprop	Root Mean Square Propagation
RNN	Recurrent Neural Network
SAC-SMA	Sacramento Soil Moisture Accounting Model
SGDM	Stochastic Gradient Descent Method
SVM	Support Vector Machine
SVR	Support Vector Regression

SWAT	Soil and Water Assessment Tool
VIC	Variable Infiltration Capacity
XGBoost	Extreme Gradient Boosting
20CR	20th Century Reanalysis

CHAPTER 1

INTRODUCTION

1.1 Introduction

Long-term and continuous river flow data is crucial for water resources management, flood/drought analysis, design and management of hydraulic infrastructures, and predicting river flows (Machiwal and Jha, 2006; Wood et al., 2011; Lin et al. 2019; O'Connor et al., 2021; Miao et al., 2022). Long-term river flow datasets are also beneficial for evaluating and managing natural water resources, greenhouse gas contribution from rivers to the global carbon budget and understanding the impact of climate change on the terrestrial water cycle (Catalán et al., 2016; Miao et al., 2022). According to Ravindranath et al. (2019), regulations for dams and reservoirs are based on streamflow records spanning a few decades; however, variables influencing the variability of streamflow might change over longer times, so using long-term climate records and extending the streamflow data is crucial to understanding how water supply variability occurs over longer durations.

However, the availability of river flow stations is limited by practical, economic, and political reasons (Fekete & Vörösmarty, 2007; Lin et al., 2019). The recent decline in active river flow stations has further complicated our understanding of hydrologic processes in ungauged or scarcely gauged basins (Shiklomanov et al., 2002; Sivapalan, 2003; Hannah et al., 2011; Lin et al., 2019). Achieving long-term continuous flow is a crucial problem in the World including Türkiye. Only a few river flow gauges in Türkiye have continuous and long-duration data. Most gauges either have short-duration data or are missing for long durations. Therefore, river flows need to be reconstructed.

Reconstructing streamflow is challenging because of the numerous non-linear hydrological processes and the wide range of temporal and spatial scales, from local watersheds to global patterns and from months to millennia, so understanding and correctly representing natural hydrological regimes is exceedingly difficult (Miao et al., 2022). Traditional methods, such as the water balance and regression approaches, often have difficulty in representing accurately natural flows in regions with significant human influence, limited observational data, and rapidly changing environmental conditions (Hernández-Henríquez et al., 2010; Miao et al., 2022). Hydrological models that integrate physical and climatic characteristics are effective in reconstructing flows in regulated river systems (Piniewski, 2016; Lin et al., 2019; Miao et al., 2022). While hydrologic modeling can help predict flows in ungauged basins, challenges exist due to hydrologic scaling, equifinality, the lack of high-quality input datasets, difficulties in model parameterization, epistemic uncertainty (uncertainty mainly due to model structure, parameter values, input data, and modeling assumptions), and limitations in computational resources (Samaniego et al., 2010; Bierkens et al., 2015; Zheng et al., 2018; Beck et al., 2019; Shen, 2018; Lin et al., 2019).

There are several methods available for extending observation data with the simulation of river flows. Scientists and engineers have been trying to model river flows in the most precise way for years. Robust models should be created to reconstruct unavailable and missing river flow data. The rainfall-runoff models are one of the basic models to investigate river flows. Discharge of the river is tried to predict by using just precipitation as an input in the simple rainfall runoff models. Moreover, different variables can be used to improve models' accuracy such as temperature, wind speed etc. (Fan et al., 2020; Yokoo et al., 2022; Won et al., 2023).

Until now, many models have been created to reconstruct river flows using different techniques primarily by physical models, conceptual models, and data-driven models (Lin et al., 2019; Ghiggi et al., 2019; O'Connor et al., 2021; Miao et al., 2022; Wang et al., 2023). These models have advantages and disadvantages relative to each other as discussed in the literature review chapter, section 2.1. In recent years, the use of

data-driven models has been increasing. Long Short-Term Memory (LSTM) model, a kind of deep learning model, has been one of the most popular techniques to model river flows in recent years. LSTM models have been observed to give better results in most cases, whether compared with physical or other machine learning models (Hu et al., 2018; Kratzert et al., 2018).

Unlike physically based hydrological models, the LSTM is not predicated on the idea of water/mass conservation or the governing equations so LSTM must acquire these physical laws and concepts solely through data-driven calibration (Kratzert et al, 2018). On the other hand, because the LSTM model is a kind of black box model, the results may vary depending on many reasons such as input variable selection, data size, and river characteristics, so studies do not give the same results and other approaches may be better (Xu et al., 2020; Mao et al., 2021).

River flow models need reliable datasets to get accurate results. However, it is not always possible to access observed data. Even when observation data is available, it may not be at sufficient length and continuity. Therefore, global historical reanalysis datasets, which combine model and observational data for obtaining long-term climate data, have been frequently used recently. ERA-Interim, ERA5, and ERA5-Land, by the European Centre for Medium-Range Weather Forecasts (ECMWF), are popular reanalysis datasets.

Within this framework, the purpose of this study is to develop a methodology to reconstruct historical daily and monthly river flows by an LSTM model (Hochreiter & Schmidhuber, 1997), a recurrent neural network-type deep learning model, utilizing ERA-5 Reanalysis climate dataset (Hersbach et al., 2020) as inputs. As a demonstration, the proposed methodology is used at selected stations in the Seyhan and Euphrates Watersheds in Türkiye. Root Mean Square Error (RMSE), Correlation Coefficient (R), Nash-Sutcliffe model efficiency coefficient (NSE), and Kling-Gupta efficiency (KGE) are some of the statistical evaluation metrics utilized in this study. In addition to these statistical metrics, time series plots, scatter plots, Taylor plots, and frequency diagrams are used to investigate model performance against observations.

1.2 Scope of the Study

The scope of this study is to develop a methodology for reconstructing historical daily and monthly river flows in data-scarce regions combining an LSTM model with climate datasets from Reanalysis Datasets. LSTM (Hochreiter & Schmidhuber, 1997) is a recurrent neural network-type deep learning model. The ERA-5 Reanalysis climate dataset (Hersbach et al., 2020) is employed as the forcing data. The river flow dataset in preceding timesteps was not used as input in this study, because the purpose of this study, which distinguishes it from other studies, is to reconstruct historical river flows that do not exist. Flow data in preceding timesteps do not exist in the past, so it cannot be used as an input within the proposed reconstruction approach. As a case study, the proposed methodology is applied to selected stations in the Seyhan and Euphrates watersheds in Türkiye. The ERA-5 reanalysis datasets are used to attain climate data over the study regions and are used as input to the LSTM model. ERA-5 reanalysis is a global atmospheric reanalysis dataset produced by ECMWF. Reanalysis combines observations from various sources with a numerical weather prediction model to generate a consistent, gridded dataset of past weather conditions. ERA-5 specifically provides data on a wide range of atmospheric variables including temperature, humidity, wind, and pressure with high spatial and temporal resolution. The ERA-5 approximately provides 31 km grid-resolution climate data in hourly intervals. It is widely used in climate research, weather forecasting, and various environmental applications due to its comprehensive coverage and quality. Daily observed discharge data published by Turkish State Hydraulic Works (DSİ) were used as the target data. Different input variables were used in different combinations to investigate the impacts of the climate variables on streamflow modeling. Moreover, various sets of hyperparameters of LSTM were employed to understand the effects of hyperparameters on the model robustness and achieve the best results.

There are many statistical model evaluation metrics used for evaluating a model's accuracy and for comparing models with each other. Consequently, RMSE, R, NSE,

and KGE were used as statistical model evaluation metrics to compare results based on different statistical terms. Furthermore, time series plots and scatter plots of observed and predicted results were plotted to better understand how well the model predicts streamflow. Additionally, by examining Taylor diagrams of the best results from each different case of input variables, the importance of input data is demonstrated and compared. Lastly, flow frequency curves were plotted to examine the flow regimes at selected locations.

1.3 Contents of the Chapters

The thesis consists of 6 chapters:

The first chapter offers an introduction to the subject and purpose of the study.

The second chapter includes a literature review on the reconstruction of river flows; river flow modeling approaches including conceptual, physical, and data-driven models; a comparison of the LSTM model with different hydrological models, and river flow modeling studies by LSTM in Türkiye.

The third chapter contains information about the Seyhan and Euphrates River Basins and datasets used in this study including observed river flow data and variables used in the ERA5 Dataset.

Chapter 4 briefly summarizes the proposed methodology, mainly LSTM and its features, as well as the statistical evaluation metrics (e.g., RMSE, R, NSE, KGE) and visualization plots (e.g., scatter plot, Taylor diagram, and flow frequency curve) used to compare observation and simulations.

In Chapter 5, the results and discussions on the performance of various LSTM model inputs, as well as the training and test period performance of the best model parameters, are presented. Additionally, the river flows have been reconstructed from 1940 to the present.

Finally, concluding remarks and future directions and recommendations are provided in Chapter 6.

CHAPTER 2

LITERATURE REVIEW

2.1 Reconstruction of River Flow Datasets

Long-term continuous river flow data is essential for the management of water resources, flood and drought analysis, design and management of hydraulic infrastructures, predicting river flows, and understanding the impact of climate change on the water cycle (Machiwal and Jha, 2006; Lin et al. 2019; O'Connor et al., 2021; Miao et al., 2022). Coverage of river flow stations is limited due to practical, economic, and political constraints (Fekete & Vörösmarty, 2007; Lin et al., 2019). The recent decrease in active river flow stations has worsened the understanding of the hydrologic processes in ungauged or scarcely gauged basins (Shiklomanov et al., 2002; Sivapalan, 2003; Hannah et al., 2011; Lin et al., 2019). Therefore, river flow dataset reconstruction is an important research area.

Reconstruction of natural flows is complicated because of the involvement of non-linear hydrological processes and a wide range of space and time scales (Miao et al., 2022). Traditional reconstruction methods, including the water balance and regression methods, often struggle to define natural flows in areas with significant anthropogenic effects, limited observations, and fast-evolving environmental conditions (Hernández-Henríquez et al., 2010; Miao et al., 2022). Hydrological models have effectively reconstructed the flow in regulated river systems by integrating physical and climatic characteristics, (Piniewski, 2016; Lin et al.,2019; Miao et al., 2022). Although hydrologic modeling can help by predicting flows in ungauged basins; there are still challenges because of hydrologic scaling, equifinality, the lack of high-quality model input datasets, challenges in model parameterization, and the limitations in computational resources (Samaniego et al.,

2010; Zheng et al., 2018; Beck et al., 2019; Bierkens et al., 2015; Shen, 2018; Lin et al., 2019).

There are attempts to reconstruct a global dataset for naturalized river flows (Lin et al., 2019) and an observation-based global gridded runoff dataset (Ghiggi et al., 2019). Lin et al. (2019) used the Variable Infiltration Capacity land surface model and routing application for parallel computation of the river routing model to estimate naturalized river flows at 2.94 million reaches (median reach length 6.8 km) in the World. The precipitation forcing is from a 0.1° global product that merged gauge-, reanalysis-, and satellite-based data (Beck et al., 2019). Other input variables including min and max 2-m air temperatures, and 10-m wind speed were attained from the Climate Forecasting System Reanalysis. They also utilized machine learning-derived, global runoff characteristics maps for grid-by-grid model calibration and bias correction. Finally, they evaluated 35-year daily and monthly model simulations over 14,000 gauges globally and found that 35% (64%) have a percentage bias within ±20% (±50%), and 29% (62%) have a monthly KGE ≥0.6 (0.2). On the other hand, Ghiggi et al. (2019) reconstructed a global gridded monthly runoff data during 1902-2014, called GRUN. River flow observations were used to train a machine learning model that predicts monthly runoff rates using precipitation and temperature from 20th Century Reanalysis (20CR). Ghiggi et al. (2019) concluded that GRUN outperforms a group of state-of-the-art global hydrological models, especially for the reproduction of the seasonality, dynamics, and anomalies of runoff. They also concluded that precipitation and temperature are sufficient as predictors to reproduce significant aspects of terrestrial water dynamics at the monthly timescale.

O'Connor et al. (2021) reconstructed 250 years (1766–2016) monthly flow dataset using both a conceptual hydrological model (GR2M) (Mouelhi et al., 2006) and an empirical based Artificial Neural Network (ANN) (Fritsch et al., 2019) for 51 Irish catchments. Also, a mixed ensemble was derived from both GR2M and ANN model. Gridded historical meteorological data and bias corrected to observations (Walsh, 2012) in each catchment were used with derived potential evaporation to force

GR2M and ANN models, which were trained and verified using river flow observations. O'Connor et al. (2021) aimed that the reconstructed data set is appropriate for analyzing hydrological responses to arterial drainage, tracking hydrological variability and change, or testing the robustness of water plans and interpreting hydrological droughts.

With the motivation of sustainable management of water resources, Miao et al. (2022) utilized Variable Infiltration Capacity (VIC) land -surface model (Liang et al., 1994) and the Lohmann routing model (Lohmann et al., 1996) with flow direction correction to reconstruct monthly river flows during the time period 1961-2018 for 330 catchments of China. The VIC model was performed on a 0.25^{o}x0.25^{o} grid with inputs of daily precipitation, max and min temperatures, and wind speed from about 2400 stations. Monthly river flow data during 1961-1979 from 330 flow stations was used to train and test the models. Miao et al. (2022) claimed that previous reconstructions from sparse and poor-quality gauge observations have resulted in large biases in simulation of the interannual and seasonal variability of natural flows. They obtained (NSE) > 0.7 for 83% of the total catchments and (KGE) > 0.7 for 56% of the total catchments.

Wang et al. (2023) utilized the meteorological, runoff, and atmospheric circulation data to reconstruct monthly runoff in Hotan River Basin in China based on machine learning methods. They used 1958-1977 duration as the training period and 1978-1998 as the validation period to reconstruct runoff from 1999 to 2019. They showed gradient boosting, random forest, AdaBoost, and bagging methods can successfully simulate runoff with NSE of 0.84, 0.82, 0.78, and 0.78.

2.2 River Flow Modeling

There are mainly three distinct methods to model river flow: Conceptual models, physical models and data-driven models.

2.2.1 Conceptual Models

Conceptual models define hydrological processes by using simple mathematical equations and these models are relatively straightforward models because simulations are relied on simple equations instead of solving governing partial differential equations (Aghakouchak & Habib, 2010). Hydrologiska Byråns Vattenbalansavdelning (HBV) model (Bergström, 1976), GR4J model (Perrin et al., 2013), and Identification of unit Hydrographs and Component flows from Rainfall, Evaporation and Streamflow (IHACRES) model (Jakeman et al., 1990; Dye & Croke, 2003) are some of the well-known conceptual models.

Das et al. (2008) studied different HBV structures to model daily discharge during the 1961-1970 calibration period and 1971-1990 validation period and NSE value is ranging from 0.61 to 0.87 in the calibration period and 0.55 to 0.87 in validation period. Additionally, Cantoni et al. (2022) used LISFLOOD, a physical and conceptual hybrid model, and GR4J conceptual model to investigate flood situations in six different catchments over a nine-year calibration and a five-year validation period, the observed rainfall data and ERA5 rainfall data used separately to compare the effect on the model accuracy and GR4J model gave better mean KGE value (0.65 with observed rainfall, 0.59 with ERA5) in calibration period, on other hand, LISFLOOD achieved better mean KGE value (0.32 with observed rainfall, 0.26 with ERA5) in validation period.

2.2.2 Physical Models

Hydrological models are essential for various tasks, such as forecasting floods and droughts, assessing water resources, understanding the effect of human activities on water resources, and enhancing knowledge of the hydrological cycle (Beck et al., 2017). According to Bourdin et al. (2012), hydrological models are commonly applied in diverse areas, including land use and land cover changes, climate change impacts, water quality, flood hazard planning, operational forecasting, water quantity and resource planning, prediction in ungauged basins, and scientific inquiry.

Hydrological components are simulated and analyzed using physical models based on physical principles and rules. Commonly used physical hydrological models include the Soil and Water Assessment Tool (SWAT), Hydrologic Engineering Center - Hydrologic Modeling System (HEC-HMS), and MIKE-SHE are commonly used physical hydrological models (Abdulkareem et al., 2018; Keller et al., 2023).

Khoi (2016) compares HEC-HMS and SWAT models on a daily and monthly basis for calibration (1982-1990) and validation (1991-2009) periods in the Srepok River Catchment. The models used topography, soil, land use data, and meteorological data such as precipitation, minimum temperature, and maximum temperature as input for simulating streamflow. The SWAT model produced the best results in both the calibration and validation periods. For daily simulation during the calibration period, the R^2 was 0.85 and the NSE was 0.69, while for monthly simulation, the R^2 was 0.92 and the NSE was 0.75. During the validation period, the daily simulation had an R^2 of 0.81 and an NSE of 0.77, and the monthly simulation had an R^2 of 0.90 and an NSE of 0.85.

2.2.3 Data-Driven Models

Data-driven models, often referred to as black box models, identify the relationship between input and output data using statistical techniques and artificial intelligence (AI) without considering physical equations or principles. Machine learning, deep

learning and regression models are examples of data-driven techniques commonly used in hydrological modeling. Linear regression, ANN, support vector machine (SVM), multilayer perceptron (MLP), decision tree, random forest, convolutional neural network (CNN), recurrent neural network (RNN), LSTM are classified as data-driven methods (Choi et al., 2019; Van et al., 2020). These models can solve both linear and nonlinear complex relationship between input variables and output variables. Data-driven models can develop robust solutions with less data because they do not require physical conditions or parameters. According to Jehanzaip et al. (2022), while data-driven models require only a small number of parameters and limited data, they do not establish any relationship with the physical characteristics of the catchment.

Deep learning applications have shown great success in recent years due to the availability of large training datasets, advances in parallel computing, better regularization techniques (e.g., batch normalization, dropout, data augmentation, early stopping), robust optimization algorithms (e.g., AdaGrad, RMSProp, Adam, AdaBoost), and software platforms that easily integrate into GPU frameworks (Sengupta et al., 2020).

In Mao et al. (2021), the ANN and LSTM models were compared at different time scales for runoff modeling, using statistical evaluation criteria such as NSE, R^2, and RMSE. The ANN model performed better on a monthly time scale, while the LSTM model outperformed ANN on a daily time scale. This suggests that the LSTM model is more suitable for complex situations, whereas the ANN model can be used in simpler circumstances.

Hao and Bai (2023) examined daily streamflow simulations using various data driven models, including support vector regression (SVR), extreme gradient boosting (XGBoost), and LSTM models, under different input scenarios (only antecedent discharge; average of rainfall stations; each rainfall station; average of rainfall stations and antecedent discharge; each rainfall station and antecedent discharge). The data was split into 14 years of training, 5 years of validation and 4

years of testing. The results indicated that the LSTM model was superior to the other models in all simulations except for those where only antecedent flow was used as input.

Alsavaf and Teksoy (2022) generated missing data by using RNN model with past flow records and some statistical terms of the records. They also utilized the neighbor records to improve training skills. They observed that low and mid flows can be predicted with admissible accuracy.

Haznedar et al. (2023) investigated a streamflow forecasting by using a hybrid model which is a combination of the particle swarm algorithm (PSO), that is an optimization algorithm, and LSTM model, and compare with the LSTM and Adaptive neuro fuzzy inference system (ANFIS) models in daily basis at three different stations in Seyhan Watershed. 24 years of data, which is divided into %80 training and %20 in test, were used and the findings indicate that LSTM-PSO gives the best results in all stations.

In another study (Apaydin et al., 2022), preceding river flows were used as input to forecast the subsequent flow. Apaydin et al. (2022) modeled daily streamflow by using different RNN methods, which are LSTM, BiLSTM, GRU, and simple RNN, in Nalli River during 1997-2015 with considering 70% training and %30 test period using past seven flow data as input to predict the next day flow record. Their results indicated that the LSTM model is superior to the other models with a correlation coefficient value of 0.96 and an NSE value of 0.92 in test simulations.

The river flow dataset in preceding timesteps is utilized as input in various studies such as for runoff simulations, river flow forecasting, and flood studies (Kisi, 2007; Kisi and Cimen, 2011; Hao and Bai, 2023; Hu et al., 2018; Le et al., 2019; Li et al., 2023; Song et al., 2019). On the other hand, it was not used as input in this study, because the purpose of this study, which distinguishes it from many other studies, is to reconstruct historical river flows that do not exist. Flow data in preceding timesteps do not exist in the past, so it cannot be used as an input to the proposed reconstruction method.

2.3 Comparison of Different Hydrological Models with LSTM Model

Many studies compare physical models and data-driven models. For example, in the Yeongsan River basin, the SWAT model was compared with an LSTM model, and the LSTM model outperformed the SWAT model, achieving an R^2 of 0.98 and an NSE of 0.94. This indicates that the LSTM model can be more effective than the hydrological SWAT model, which requires a wide range of topographical and meteorological data (Kim & Kim, 2021).

However, this comparison does not always yield similar results. Under different conditions, for example, varying inputs or data length, the outcomes may differ. In the pioneering work of Kratzert et al. (2018), the LSTM model was compared with the Sacramento Soil Moisture Accounting Model (SAC-SMA) coupled with Snow-17. They evaluated the models using NSE values and three decomposition metrics: correlation coefficient, variance bias, and total volume bias. The study found that the LSTM model could estimate rainfall using meteorological measurements with accuracy close to that of the well-known SAC-SMA + Snow-17 model. When NSE values were considered, the SAC-SMA + Snow-17 model performed better in the southeastern catchments, while the LSTM model showed slightly better performance in the northern, more snow-influenced catchments. This indicates that model performance can vary across different regions.

Fan et al. (2020) proposed an LSTM model to simulate runoff in the Poyang Lake Basin and compared its performance with ANN and SWAT models. Their findings showed that the ANN can achieve performance comparable to the SWAT in most situations while the LSTM model can perform much better. Their results also demonstrated that even though LSTM with precipitation data as the only input can reach the NSE ranged from 0.60 to 0.92 during the test period, the performance can be improved by feeding the model with more meteorological variables (where NSE ranged from 0.74 to 0.94 for the test period).

In another investigation, Han and Morrison (2021) studied to predict hourly runoff by using ANN, SVR and LSTM in Russian River basin in California and compare with the physical model, The National Water Model (NWM), the inputs of the model are precipitation, land surface temperature, soil moisture and baseflow, then R, NSE, PBIAS and RMSE evaluation metrics are checked in order to compare the models and the findings indicated that SVR and LSTM models gave better results than ANN models like R value is more than 0.9 in both model, and SVR, LSTM and NWM were comparable with each other, so in some cases data-driven methods can be an alternative to complex physical models.

In addition, some studies combine multiple models to investigate river flow more accurately. For example, Li et al., 2022 constructed a hybrid model which combines CNN which is beneficial for recognizing a spatial pattern in precipitation, and LSTM which utilizes a time series relationship between input and output to simulate runoff. Then they compared their simulation performance with that of the SWAT model by considering NSE, R^2, and relative error (RE). The results indicate that the hybrid model is as good as SWAT model due to the ability of the hybrid model to capture nonlinear relationships.

CHAPTER 3

STUDY AREA AND DATASETS

This study was implemented on Seyhan Watershed and Euphrates Watershed by using ERA5 dataset and observed flow data.

3.1 Study Area: Seyhan and Euphrates River Watersheds

Seyhan basin is located between 34º 45' and 37º 00' East longitude and between 36º 30' and 39º 15' North latitude. The Seyhan River Basin, which spans 20.731 km^2, has an average annual precipitation depth of 624 mm and an annual flow of 211,07 m^3/sec (SYGM, n.d.). The Seyhan River and its tributaries provide the surface water for the Seyhan River basin and the Göksu and Zamantı Rivers are the primary tributaries of the Seyhan River (SYGM, 2020b). Between 2071-2100, the effects of climate change are predicted to diminish the basin's overall water potential by 30% (SYGM, n.d.). The geographical regions of Central Anatolia and the Mediterranean contain the Seyhan River rainfall area. The Central Anatolia Region has a continental climate, while the Mediterranean Region has a Mediterranean climate. Summers are hot and dry in both regions. In continental climates, winters are cold and usually snowy, but in Mediterranean climates winters are pleasant and rainy. In Figure 3.1, hydrometeorological map of Seyhan Watershed is shown. Some hydrological studies have been carried out on the basin by using HEC-HMS and classical methods so far (SYGM, 2020b).

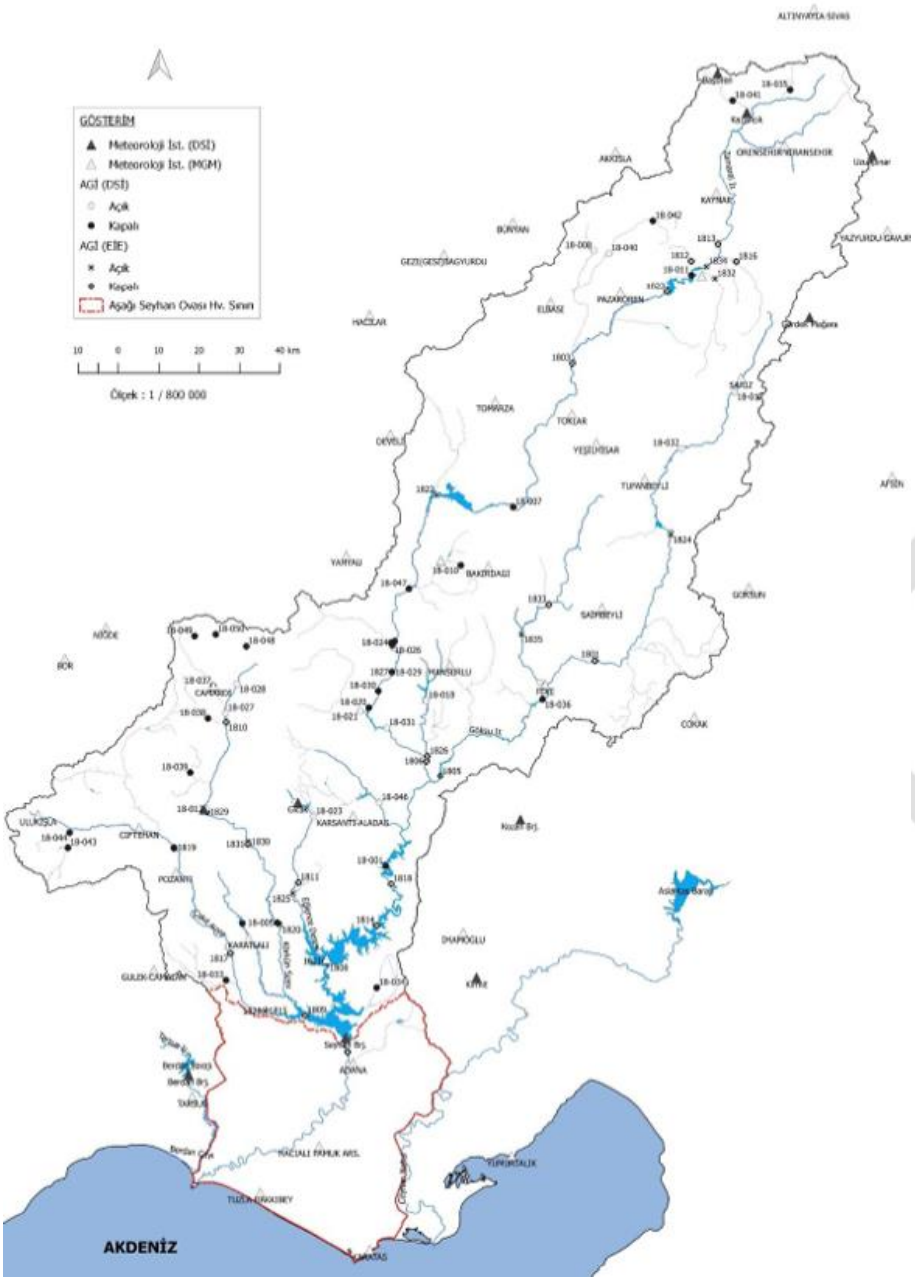


Figure 3.1 Hydrometeorological map and border of Seyhan Watershed (SYGM, 2020b).

Euphrates river basin is located on 36º 46' 36" and 43º 48' 10" East longitude and between 40º 20' 26" and 36º 39' 17" North latitude. The Euphrates watershed covers 122.010 km^2 which is 16% of Turkey. Since the basin covers a very large area, it is possible to see different types of climates. In the Southeastern Anatolia region, summers are hot and dry, winters are cold. In the Eastern Anatolia region, summers are cool, and winters are quite cold. In Figure 3.2 illustrates the main river tributaries and border of the Euphrates Watershed (SYGM, 2020a).

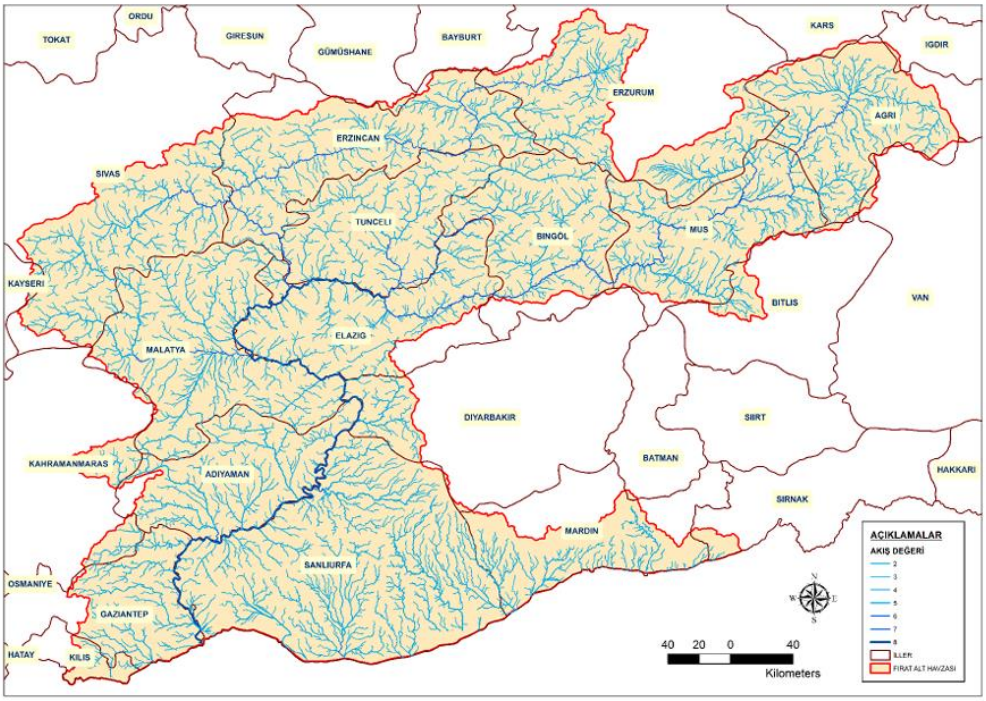


Figure 3.2 Main River tributaries and border of Euphrates River Basin (SYGM, 2020a).

3.2 Datasets

3.2.1 Observed River Flow Data

River flow data is crucial for managing water resources adequately. Different flow levels can be observed depending on the size of the river tributaries and climate conditions. Flow levels can be categorized as low, moderate, high and flood.

In Türkiye, the observed river flow dataset published by DSİ was used as the output target variable in the simulations. This is a point data measured by stream gauges, with river flow observations recorded in cubic meters per second (m^3/s). The data is available from 1959 to 2015. but not all years are covered for each station. The range of available data varies from station to station.

In this study, three different river flow stations were used for modeling, and then past unavailable river flow data from 1940 was reconstructed based on the best model results. Two of the stations - D18A19, and D18A27 - are from the Seyhan Watershed while one station, D21A167, is from the Euphrates Watershed.

The basin name where the station is located, the name of the station, station no, location information (latitude and longitude), precipitation area and approximate height of the stations are tabulated in Table 3.1.

The river network of the Seyhan and Euphrates Watersheds and the locations of the stations D18A19, D18A27, and D21A167 are shown in Figures 3.3 and 3.4, respectively.

Table 3.1 Information on flow stations that were used in this study.

Station No	Name of the Stations	Basin's Name	Latitude	Longitude	Drainage area (km^2)	Elevation (m)
D18A019	Asmaca Ç. Feke	Seyhan Watershed	37°49'21"	35°54'8"	619	550
D18A027	Ecemiş Ç. Elekgölü	Seyhan Watershed	37°46'21"	35°0'56"	1833	1550
D21A167	Kisre Ç. Necmettin Ağa Komu	Euphrates Watershed	39°22'39"	41°41'50"	250	1650

3.2.2 ERA5 Dataset

The ERA5 dataset (Hersbach et al., 2020) was used to obtain the input variables to the LSTM model in this study. This dataset, produced by ECMWF, combines both observed and model-generated climate data. ERA5 is available at hourly and monthly intervals, covering the entire globe from 1940 to the present, with a spatial resolution of 0.25° x 0.25° (approximately 31 km). In this study, daily climatic input variables were used. As a result, the hourly data were aggregated to a daily time resolution.

Tarek et al. (2020) compared precipitation and temperature data from ERA5 with ERA-Interim, the predecessor of ERA5, and observational data to evaluate its potential use as input for hydrological models. The study found that ERA5 reduces biases compared to ERA-Interim and produces results that are as reliable as observational data across most of North America. Additionally, the precipitation data from the ERA5 and ERA5-Land datasets showed similar patterns with no significant differences. Although other reanalysis and satellite-based datasets sometimes outperformed ERA5 in specific circumstances, each dataset has its own advantages, such as longer time intervals, varied resolutions, or global coverage (Hassler and Lauer, 2021; Xu et al., 2022). Considering these factors, the ERA5 dataset appears to be more suitable for this study.

A review of the literature reveals that many variables have been used as inputs for modeling river flows. In this study, the input variables include precipitation, 1-day lagged precipitation, temperature, potential evaporation, surface runoff, surface net solar radiation, and the u and v components of wind. Descriptions of these input variables can be shown in Appendix A.

As mentioned before, the ERA5 dataset consists of grided data. However, it is not practical to use all grids covering a basin as input data for a single station. Therefore, the maps of the river basins were divided into 0.25° x 0.25° grids to accurately

determine the number of input grids in the upstream influencing the flow stations. The location of each station was also examined, and the data from the grids surrounding each station were used as inputs for each variable. Special consideration was given to where the river tributaries feed into the station. For example, for the D18A27 station, three grids were sufficient as input. Figures 3.3 and 3.4 illustrate which grids were used for the flow stations in the Seyhan and Euphrates watersheds.

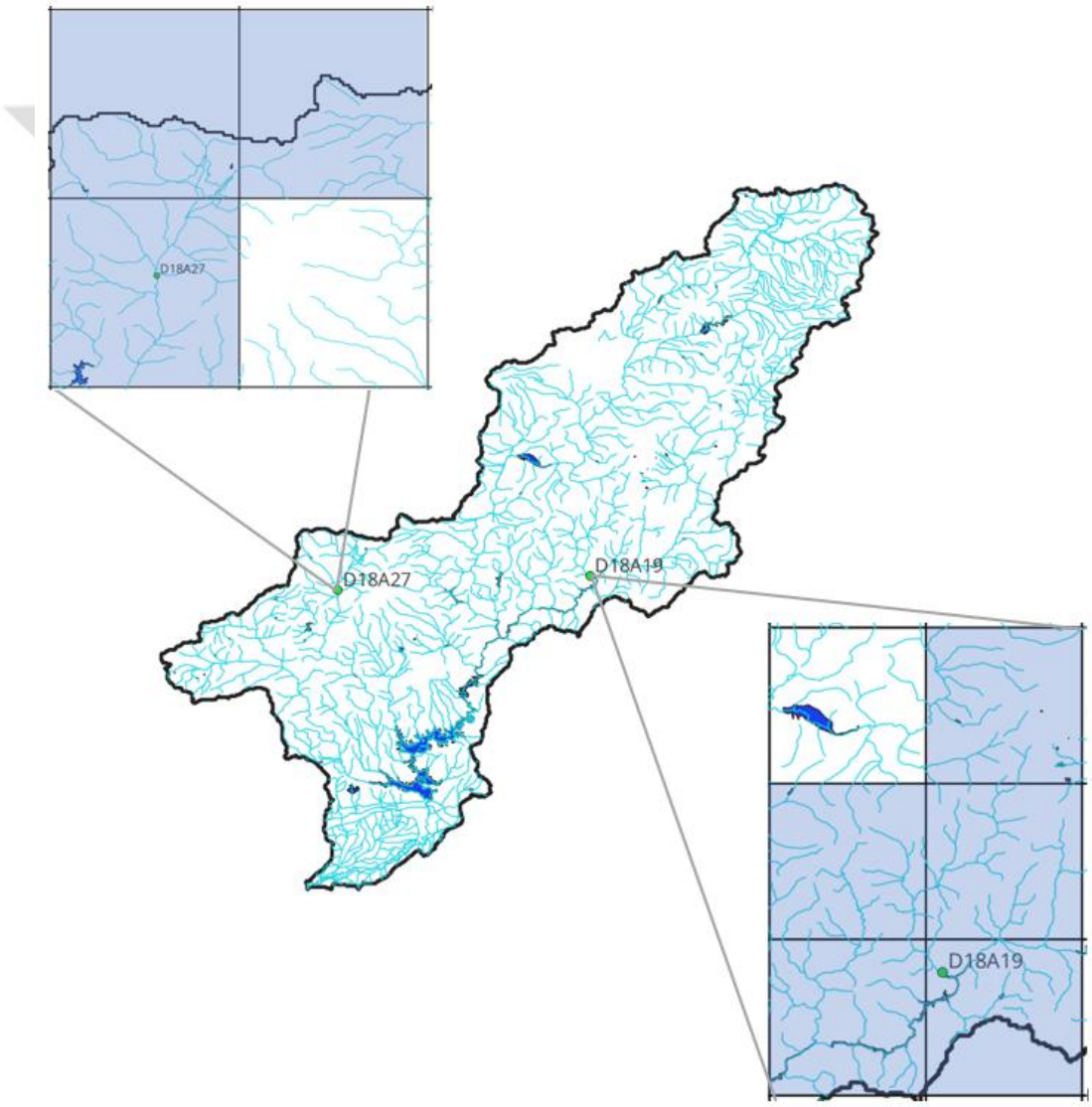


Figure 3.3 River network of Seyhan Watershed and location and selected grids for stations D18A19, and D18A27.

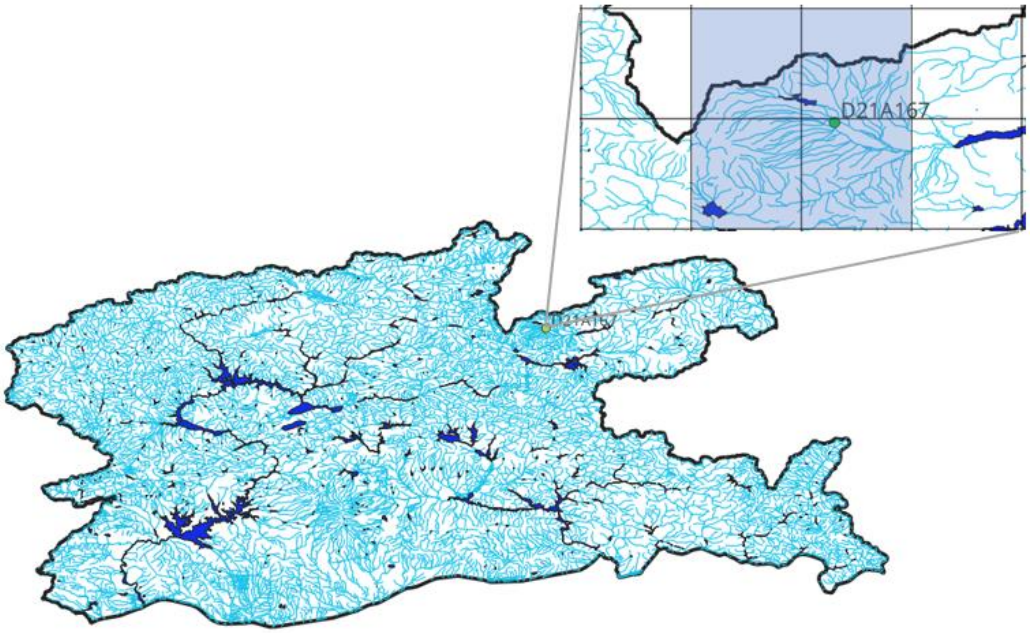


Figure 3.4 River network of Euphrates-Tigris Watershed and location and selected grids for station D21A167.

CHAPTER 4

METHODOLOGY

In this study, an LSTM model, a state-of-the-art deep learning model, was used as primary method for modeling river flows. In addition, this investigation was carried out using the Deep Learning Toolbox of MATLAB R2023a.

4.1 Deep Learning Models

As shown in Figure 4.1, Deep Learning is a subset of machine learning, which in turn is a subset of artificial intelligence. Deep learning, a state-of-the-art development in artificial neural network research, has had a significant impact across various fields, enhancing the ability of neural networks to recognize relevant features and understand complex data distributions (Shen, 2018). It employs neural networks to extract meaningful information from the data (Beale et al., 2018).

Deep learning is generally superior to traditional machine learning due to its ability to create new features from the training data by combining existing ones. This capability allows deep learning to solve complex problems more effectively. Additionally, deep learning's full-cycle learning process, from raw input data to outcomes, enhances the precision of the results (Ahmed et al., 2023). There are various types of deep learning models, such as CNN, gated recurrent unit (GRU), RNN, and LSTM, and the choice of model depends on its intended application. Deep learning techniques are versatile and can be applied in many fields, including classification, regression, time series prediction, and speech recognition (Affonso et al., 2017; Wu et al., 2019; Hua et al., 2019; Zhang et al., 2018; Graves et al., 2013). Deep learning models do not have strict rules regarding which inputs should be used, nor does using more inputs necessarily improve the model's performance. To find

the optimal solution, it is important to experiment with different variables, data lengths, and parameter sets.

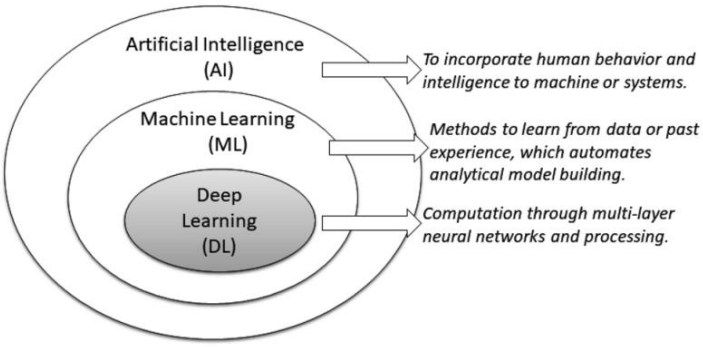


Figure 4.1 Venn diagram of artificial intelligence, machine learning and deep learning (Sarker, 2021).

4.1.1 LSTM Model

The LSTM model was first developed in 1997 by Hochreiter & Schmidhuber (1997). LSTM is a special type of recurrent neural network RNN models. In an RNN, the output of one step is used as the input for the next step. The main difference between LSTM and RNN models is that LSTM can address long-term dependency problems due to its unique structure. Although the structure of an LSTM is similar to that of an RNN, LSTM models have four neural network layers, unlike the single basic neural network layer found in RNNs, such as the tanh layer, and include a cell state (Olah, 2015).

The tanh and sigmoid layers are activation functions with different roles (Fan et al., 2020). The tanh layer is used to decide whether new information should be added to the cell state and to generate the final output. The range of the tanh output value is -1 to 1. The sigmoid layer controls the output of the gates and determines how much information should pass through them. The range of the sigmoid output value is 0 to 1. Figure 4.2 illustrates an LSTM model, which consists of four interacting layers.

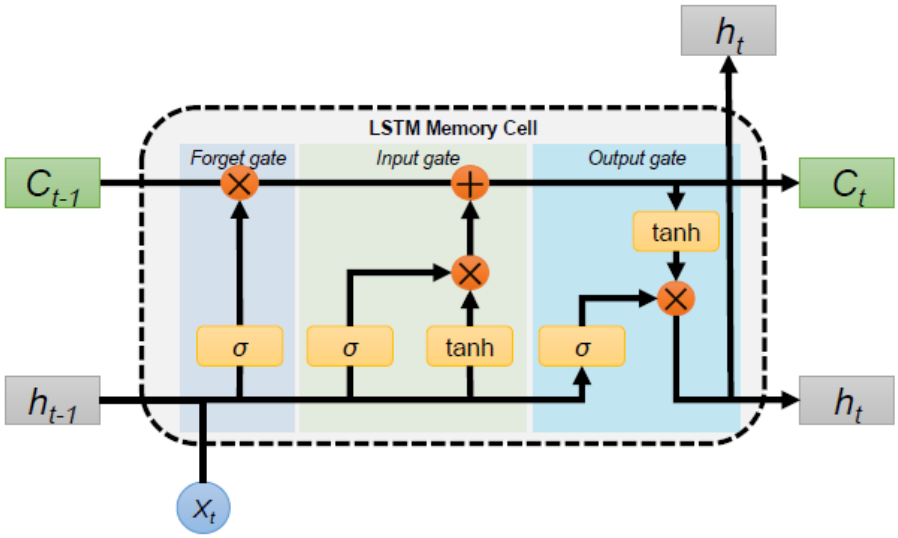


Figure 4.2 LSTM Memory Cell structure (Fan et al., 2020).

Another important feature that distinguishes LSTM from RNN is the presence of a cell state, which enables the model to retain and recall past data. To manage the cell state, three gates are used: the forget gate, the input gate, and the output gate (Zhang et al., 2023). The forget gate is the first to decide which data should be discarded from the cell state. The input gate, the second gate, determines which new data should be added to the cell state. Finally, the output gate generates the output (Olah, 2015). Additionally, the hidden state represents the output at that time step. The LSTM equations are shown below (Olah, 2015):

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (4.1)$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_c) \quad (4.2)$$

$$\widetilde{C_t} = tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (4.3)$$

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C_t} \quad (4.4)$$

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (4.5)$$

$$h_t = o_t * \tanh(C_t) \quad (4.6)$$

where f_t, i_t, and o_t are the forget gate, input gate and output gate at time t, respectively. x_t is the input, C_t is the cell state, $\widetilde{C_t}$ is a vector which has a range between -1 and 1 and h_t is the hidden state. tanh and σ are the activation functions. W_f, W_i, W_c, W_o are the weight matrices and b_f, b_c, b_o are bias vectors.

Hyperparameters are parameters that must be set before training simulations and have a significant impact on model accuracy. In this study, several hyperparameters for the LSTM model were considered, including hidden units, the number of epochs, the initial learning rate, and dropout values, while other hyperparameters were set to their default values. Hidden units refer to computational units within LSTM layers. Increasing the number of hidden units can help the model learn complex patterns more effectively, but it also increases computation time. The number of epochs denotes how many times the entire dataset is passed through the LSTM network, during which parameters are updated. The initial learning rate specifies the step size used in optimizing the parameters, influencing the rate of convergence to the minimum loss. Dropout helps prevent overfitting by randomly deactivating some neurons during training.

In addition, optimizers, which are techniques used to adjust parameters in order to minimize loss, were carefully considered. Stochastic Gradient Descent Method (SGDM), Root Mean Square Propagation (RMSprop), and Adaptive Moment Estimation (ADAM) were examined during the simulations. Among these, the ADAM optimizer, being the most popular optimization method, was used most frequently.

4.2 Normalization

In data driven models, different normalization techniques are used to improve model performance, increase the convergence rate, and reduce the overfitting problems. There are many normalization techniques such as min max normalization, z-score normalization, sigmoid normalization, median normalization, tanh estimators and decimal scaling normalization (Bhanja & Das, 2018). Z-score normalization and min-max normalization techniques are mostly used in runoff models using deep learning techniques (Fan et al., 2020; Kratzert et al., 2018; Liu et al., 2021; Yin et al., 2022; Hu et al., 2020; Gao et al., 2020).

In this study, z-score normalization was used to pre-process the data because z-score normalization has resistance to outliers by using the mean and standard deviation of datasets. Outliers have a huge impact on the scaling range in the min-max normalization and outliers may diminish the usefulness of min-max normalization, so more robust z-score normalization was used to avoid disturbing normalized data considering the presence of high flows. The formula of z-score normalization follows:

$$x_{normalized} = \frac{x - \mu}{\sigma} \tag{4.7}$$

Where $x_{normalized}$ is normalized data, x is actual data, μ is the mean value of the data and σ is the standard deviation of the data. In addition, after the simulation is completed, the results are denormalized to obtain the actual values by using the inverse of the formula.

4.3 Evaluation Metrics

In data-driven models, various statistical evaluation metrics are used to assess model performance and determine how well the predictions align with the observed data. Different metrics apply distinct statistical approaches to evaluate models. In this study, RMSE, R, NSE, and KGE are employed. The use of multiple metrics allows for a more comprehensive evaluation, as results may not always be consistent across metrics. Generally, these metrics are not directly comparable. For example, a high NSE value might be achieved with large coefficients of variation and small errors, but the KGE can be severely affected by large biases so If KGE is considered as the only metric, the model can be interpreted as unsatisfactory even it performs well. (Knoben et al., 2019).

Root Mean Square Error (RMSE)

RMSE is one of the most popular statistical evaluation metrics. It calculates the average value of the difference the between predicted value $\hat{y}_i$ and observed or actual value y_i. RMSE represents the square root of the mean error, which ensures that the metric has the same units as the observed values and reflects the typical or "standard" error in the case of normally distributed errors (Hodson, 2022). RMSE is particularly sensitive to large errors because it squares the errors before averaging. RMSE value never can be a negative value. Consequently, RMSE values are always non-negative, ranging from 0 to infinity, where 0 indicates a perfect fit with the observed values. It is defined as follows:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{4.8}$$

Where n is the number of observed values, y_i is the actual value and $\hat{y}_i$ is the predicted value for the i-th value.

Correlation Coefficient (R)

The R-value was developed by Karl Pearson in 1896, measures the statistical linear relationship between two variables. In this study, the correlation coefficient between observed and predicted values was calculated. The R-value ranges from -1 to 1. A value close to 1 indicates a strong linear relationship and suggests the robustness of the model. It is formulated as follows:

$$R = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}} \tag{4.9}$$

where x_i and y_i are the i-th value of the observed and predicted value, $\bar{x}$ and $\bar{y}$ are the means of the observed and predicted data. Typically, R values greater than 0.7 are considered acceptable (Santhi et al., 2001; Moriasi et al., 2007).

Nash-Sutcliffe Model Efficiency Coefficient (NSE)

The NSE value is commonly used in hydrological models, and climate studies to assess model accuracy. It is defined as:

$$NSE = 1 - \frac{\sum_{i=1}^{n}(Q_{sim,i} - Q_{obs,i})^2}{\sum_{i=1}^{n}(Q_{obs,i} - \bar{Q}_{obs})^2} \tag{4.10}$$

Where n is the number of observed values, $Q_{sim,i}$ and $Q_{obs,i}$ are the simulated and observed values at i-th step, $\bar{Q}_{obs}$ is the mean of the observations.

An NSE value of 1 indicates perfect agreement between predictions and observations. Conversely, a negative NSE value suggests that the model has poor accuracy.

According to Knoben et al. (2019), an NSE value greater than 0.5 is considered satisfactory for flow simulations. Conversely, Newman et al. (2015) set a higher

benchmark, considering an NSE value above 0.55 as acceptable for hydrological models. Moriasi et al. (2007) categorized NSE performance as follows: very good if greater than 0.75, good if between 0.65 and 0.75, satisfactory if between 0.5 and 0.65, and poor if less than 0.5, specifically for monthly time steps.

Kling-Gupta Efficiency (KGE)

The KGE was developed by Harald Kling and Hoshin Vijai Gupta for evaluating hydrological models (Gupta et al., 2009). It was designed to enhance NSE value by incorporating various model statistics, such as correlation coefficient, bias ratio and variability ratio to assess model performance more accurately. The formula for KGE is:

$$KGE = 1 - \sqrt{(r-1)^2 + (\alpha-1)^2 + (\beta-1)^2} \quad (4.11)$$

Where r is the correlation coefficient, α is the variability which is the standard deviation of simulation divided by standard deviation of observation and β is the bias term which is mean of simulation divided by mean of observation. If KGE value is equal to 1, it means the perfect model. Knoben et al. (2019) consider that if KGE > -0.41, the model has reasonable performance.

4.4 Plots and Diagrams

In this study, time series plots, scatter plots and Taylor diagrams were utilized to visualize, understand, and compare the simulations and examine the accuracy of the results of the model.

Time Series Plot

Time series plots display data over time, with the x-axis representing the time sequence and the y-axis showing the observed or simulated variables. These plots help in understanding the patterns over time and comparing variables.

Scatter Plot

Scatter plots visualize the relationships between the variables. In the scatter plots, the predicted and observed flow data are used to assess the quality of the simulations. If the points align closely with the reference line (1:1 line), it indicates a perfect prediction. Scatter plots also make it easier to identify extreme forecasts, as well as any underpredictions or overpredictions.

Taylor Diagram

The Taylor diagram was developed by Karl E. Taylor to demonstrate the similarity between the observed data and the simulations each other. Figure 4.3 shows the sample Taylor diagram where four 4 different cases are compared with the observed data. The dotted black line shows the correlation coefficient, green counters indicate the centered root mean square error, the dotted blue arches represent standard deviation of each simulation data, and the solid black arch indicates the standard deviation of the observed data. Each shape's location on the diagram illustrates how closely the simulated data pattern fits with the observations.

The Taylor diagram provides an idea about how the simulation results resemble the observed data by considering statistical properties like correlation coefficient and standard deviation. Hereby, multiple statistical data can be analyzed in a single visualization.

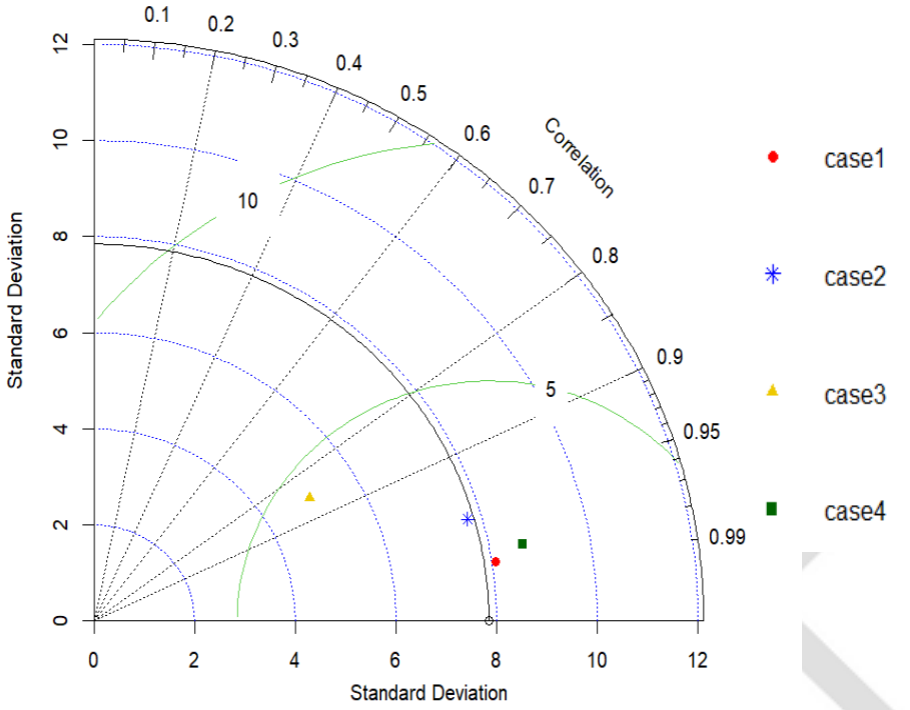


Figure 4.3 Sample Taylor Diagram.

Flow Frequency Curve

Flow frequency curves were plotted to visualize the distribution of the flow data. These curves help provide information on the variability of flow conditions by showing how often different flow rates occur. Flow frequency curves are commonly used in various fields such as flood and drought analysis, water management studies, comparing actual and simulated data in hydrological models, and also dam design and planning (Büchele et al., 2006; Peyras et al., 2012). X-axis represents the flow (m^3/s), and the y-axis indicates the probability of exceedance. In addition, observed and simulated frequency curves were plotted for both training and test simulations to examine the difference of simulated and observed flow data. Probability of exceedance formula is as follows:

$$P = \frac{m}{n+1} \tag{4.12}$$

P indicates the probability, m represents the rank of the flow value, 1 corresponds to the largest value, and n is the total number of data points.

CHAPTER 5

RESULTS AND DISCUSSION

In this study, historical daily and monthly river flows were reconstructed using the ERA5 reanalysis climate dataset and the LSTM model at two flow observation stations in the Seyhan Watershed and one flow observation station in the Euphrates watershed.

Using daily inputs, the LSTM model was tested with different inputs to identify the variables that significantly affect the flow since different variables may impact the flow in varying proportions depending on the locations. Additionally, various hyperparameter sets were considered to determine the optimum model performance since parameters can substantially affect the results. Several hyperparameters affect the model accuracy, computational efficiency, and memory requirements. In this regard, epoch number, dropout value, initial learning rate, and hidden unit number were varied to find their optimal values, while other hyperparameters were kept as default values.

5.1 Flow Reconstruction in the daily timescale

Precipitation, temperature, potential evaporation, surface net solar radiation, east (u) and north (v) components of wind, surface runoff data from the ERA5 dataset, time, and 1-day lag of precipitation were evaluated with different combinations at different stations. In Tables 5.1, 5.2, and 5.3 various combinations of input variables for each flow station are shown. The input combinations were dynamically varied based on the test results in each case to achieve optimal model results by considering four different statistical evaluation metrics. Two distinct river stations from Seyhan Watershed and one river station from Euphrates Basin were simulated to model and reconstruct the river flows since 1940 by using the LSTM model. These stations were

chosen based on the availability of data, and their locations with respect to dams. River stations without dam effects were chosen for analysis. In addition, input grid numbers were decided by considering the tributaries of the river.

The LSTM model gives location-independent results, meaning it is not a valid model for a single region. It is possible to reach satisfactory results in other basins as well as in the Seyhan Basin by using the same input data and hyperparameters. To illustrate this, simulations were also conducted for a station in a different basin (Euphrates Basin) with different topography and climate conditions.

Table 5.1 Input variable cases for river flow simulation at station D18A19 in Seyhan Watershed.

Cases	Input Variables
Case1	Precipitation
Case2	Time + Precipitation
Case3	Time + 1Day Lag of Precipitation + Precipitation
Case4	Time + Temperature + Precipitation
Case5	Temperature + Precipitation
Case6	Time + Temperature + Potential Evaporation + Precipitation
Case7	Time + Temperature + Potential Evaporation + U comp. of wind + V comp. of Wind + Precipitation
Case8	Time + Temperature + Potential Evaporation + Surface Runoff + Precipitation
Case9	Time + Temperature + Potential Evaporation + Surface Net Solar Radiation + Surface Runoff + Precipitation
Case10	Time + Temperature + Potential Evaporation + Surface Net Solar Radiation + U comp. of wind + V comp of Wind + Surface Runoff + Precipitation
Case11	Surface Runoff

Table 5.2 Input variable cases for river flow simulation at station D18A27 in Seyhan Watershed.

Cases	Input Variables
Case1	Precipitation
Case2	Time + Precipitation
Case3	Time + 1Day Lag of Precipitation + Precipitation
Case4	Time + Temperature + Precipitation
Case5	Time + Surface Runoff + Precipitation
Case6	Time + U comp. of wind + V comp. of Wind + Precipitation
Case7	Time + Temperature + Surface Runoff + Precipitation
Case8	Time + Temperature + Potential Evaporation + Surface Runoff + Precipitation
Case9	Time + Temperature + Potential Evaporation + Surface Net Solar Radiation + U comp. of wind + V comp of Wind + Surface Runoff + Precipitation
Case10	Surface Runoff

Table 5.3 Input variable cases for river flow simulation at station D21A167 in the Euphrates Watershed.

Cases	Input Variables
Case1	Precipitation
Case2	Time + Precipitation
Case3	Time + 1 Day Lag of Precipitation + Precipitation
Case4	Time + Temperature + 1 Day Lag of Precipitation + Precipitation
Case5	Time + Temperature + Surface Runoff + 1Day Lag of Precipitation + Precipitation
Case6	Time + Temperature + Potential Evaporation + 1 Day Lag of Precipitation + Precipitation
Case7	Time + Temperature + Potential Evaporation + Precipitation
Case8	Time + Temperature + Potential Evaporation + Surface Net Solar Radiation + 1 Day Lag of Precipitation + Precipitation
Case 9	Time + Temperature + Potential Evaporation + U comp. of wind + V comp. of Wind + 1 Day Lag of Precipitation + Precipitation
Case 10	Surface Runoff

5.1.1 Daily flows at D18A19 at Adana Feke in Seyhan Watershed

A 12-year training period was used for the simulations. Data from 5 computational grids were selected from the ERA5 dataset based on the location of the flow station (see Figure 3.3), considering which river tributaries affect the station. Table 5.4 demonstrates the performance statistics of the input variable cases with Case 9 selected as the optimal scenario. It used a combination of time, temperature, potential evaporation, surface net solar radiation, surface runoff, and precipitation as input. Also, the corresponding hyperparameters are as follows: the hidden unit is 128, the epoch number is 100, the initial learning rate is 0.002, the dropout is 0.2, and the optimizer is selected as ADAM. These parameters were selected from Case 9 because the RMSE, R, and NSE values were superior to other cases.

It is evident from the Taylor diagrams in Figure 5.1, that predictions are very close to the observed data during the training phase and tested with a high correlation. Moreover, most simulations with different combinations of inputs reached similar correlation results in each other for both training and test simulations except for the first 3 cases.

The time series graphs of the predicted and observed river flows during training and test periods are demonstrated in Figure 5.2. During the training period, R is 0.97, NSE value is 0.94, and KGE value is 0.95 for daily flows; R is 0.99, NSE value is 0.99, and KGE value is 0.95 for monthly flows. On other hand, during the test period, R is 0.90, NSE value is 0.79, and KGE value is 0.72 for daily flows; R is 0.96, NSE value is 0.90, and KGE value is 0.82 for monthly flows. As shown in these figures, the training simulations have highly accurate results. In contrast, the accuracy of the test simulations is slightly lower than training simulations, but the performance statistics remain very good.

Figure 5.3 represents the scatter plot of observed and predicted runoff during training and test simulations. When these figures and time series graphs are examined, the

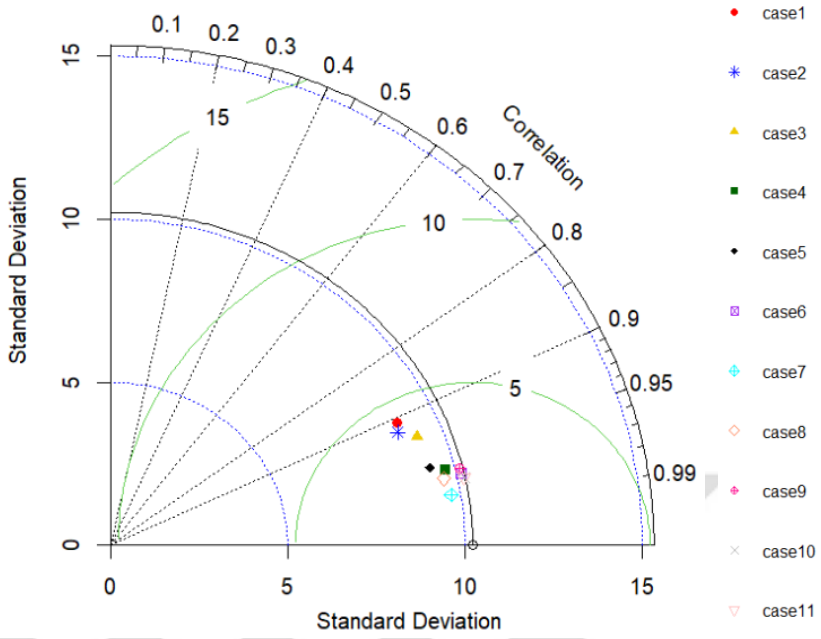
model underpredicts river flows, especially in the test simulations, this situation can be noticed obviously.

Figure 5.4 displays the flow frequency curves of the training and test periods of the observed and predicted flows, while Figure 5.5 demonstrates the probability of exceedance of observed data during both training and test periods. The close alignment of these probabilities further supports the success of the simulations.

Table 5.4 Performance statistics of the input variable cases for flow station D18A19 in Seyhan Watershed.

	TRAINING RESULTS				TEST RESULTS			
	RMSE	R	NSE	KGE	RMSE	R	NSE	KGE
Case1	4.3652	0.9065	0.8175	0.8314	7.5084	0.7646	0.5843	0.6557
Case2	4.0647	0.9195	0.8418	0.8393	7.6427	0.7557	0.5693	0.6362
Case3	3.6901	0.9330	0.8696	0.8842	7.6135	0.7585	0.5726	0.6345
Case4	2.4540	0.9710	0.9423	0.9428	5.9441	0.8625	0.7395	0.7642
Case5	2.6698	0.9668	0.9317	0.9066	6.2100	0.8551	0.7156	0.6943
Case6	2.2491	0.9757	0.9516	0.9711	5.6119	0.8792	0.7678	0.7778
Case7	1.6625	0.9873	0.9735	0.9511	5.9798	0.8635	0.7363	0.7302
Case8	2.2154	0.9770	0.9530	0.9352	6.1691	0.8843	0.7194	0.6156
Case9	2.4268	0.9725	0.9436	0.9472	**5.3899**	**0.9007**	**0.7858**	**0.7226**
Case10	2.1140	0.9789	0.9572	0.9652	5.9540	0.8747	0.7386	0.6861
Case11	3.4454	0.9437	0.8863	0.9414	6.1317	0.8837	0.7228	0.6235

a)



b)

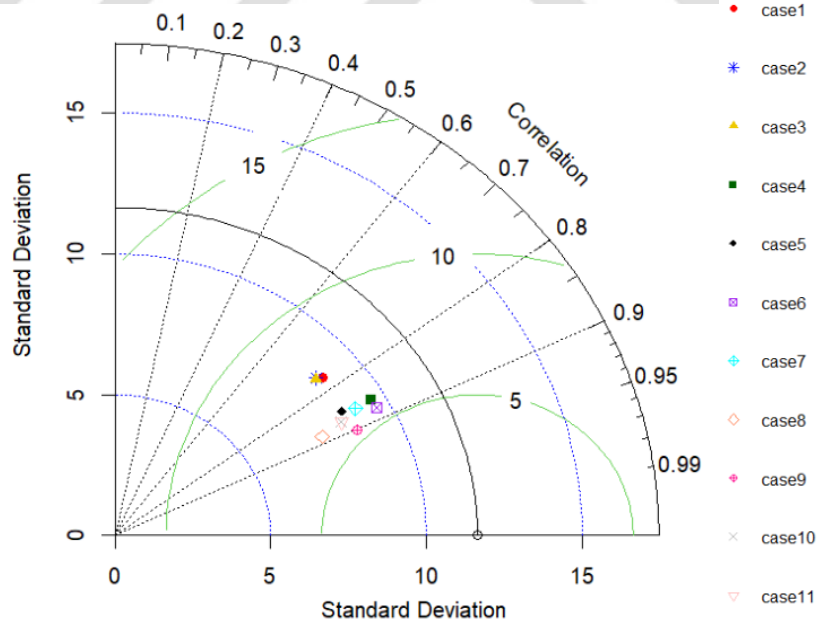
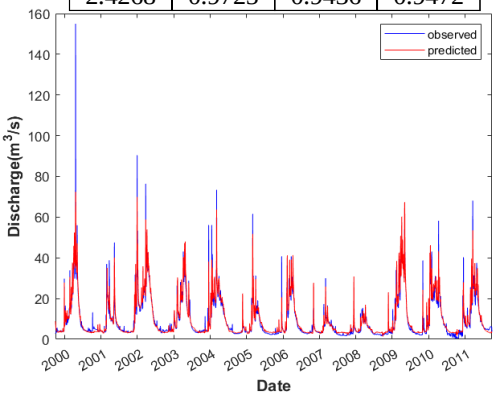


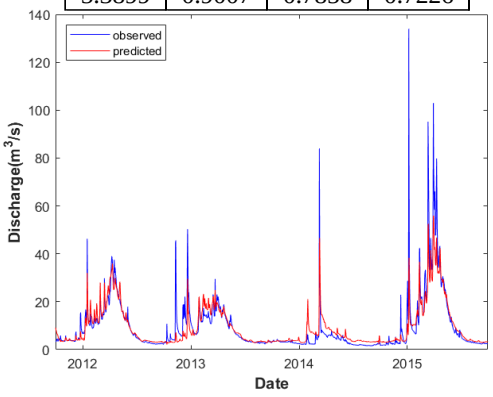
Figure 5.1 Taylor diagram of River flow simulations during a) the training period of 2000-2011, and b) the test period of 2012-2015.

a)

RMSE	R	NSE	KGE
2.4268	0.9725	0.9436	0.9472

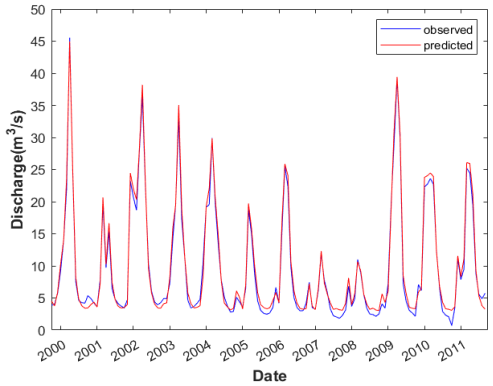


RMSE	R	NSE	KGE
5.3899	0.9007	0.7858	0.7226



b)

RMSE	R	NSE	KGE
0.9735	0.9957	0.9879	0.9475



RMSE	R	NSE	KGE
3.0204	0.9592	0.9014	0.8182

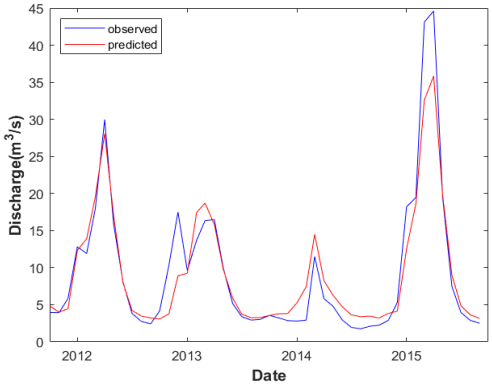


Figure 5.2 River flow simulation timeseries and performance statistics at station D18A19 during **a)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in daily timescale; **b)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in monthly timescale.

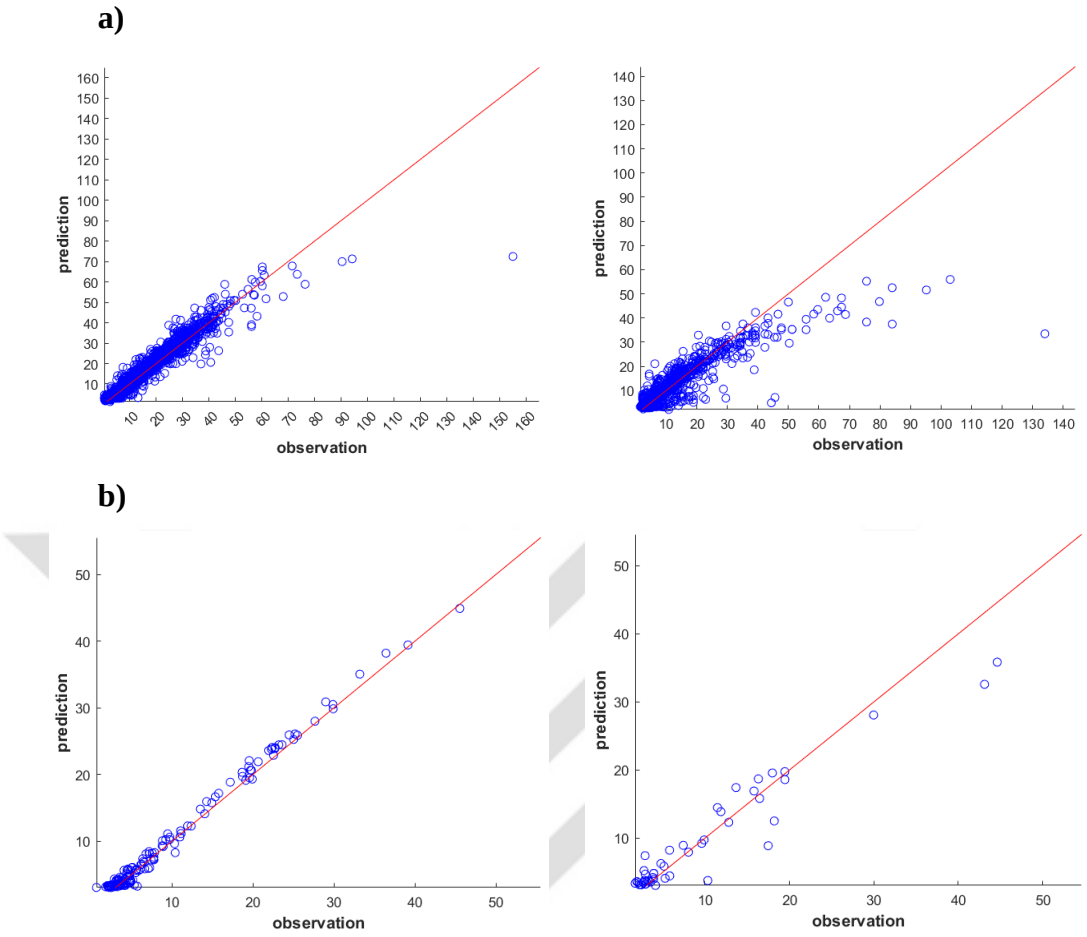


Figure 5.3 River flow simulation scatter plots at station D18A19 during **a)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in daily timescale; **b)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in monthly timescale.

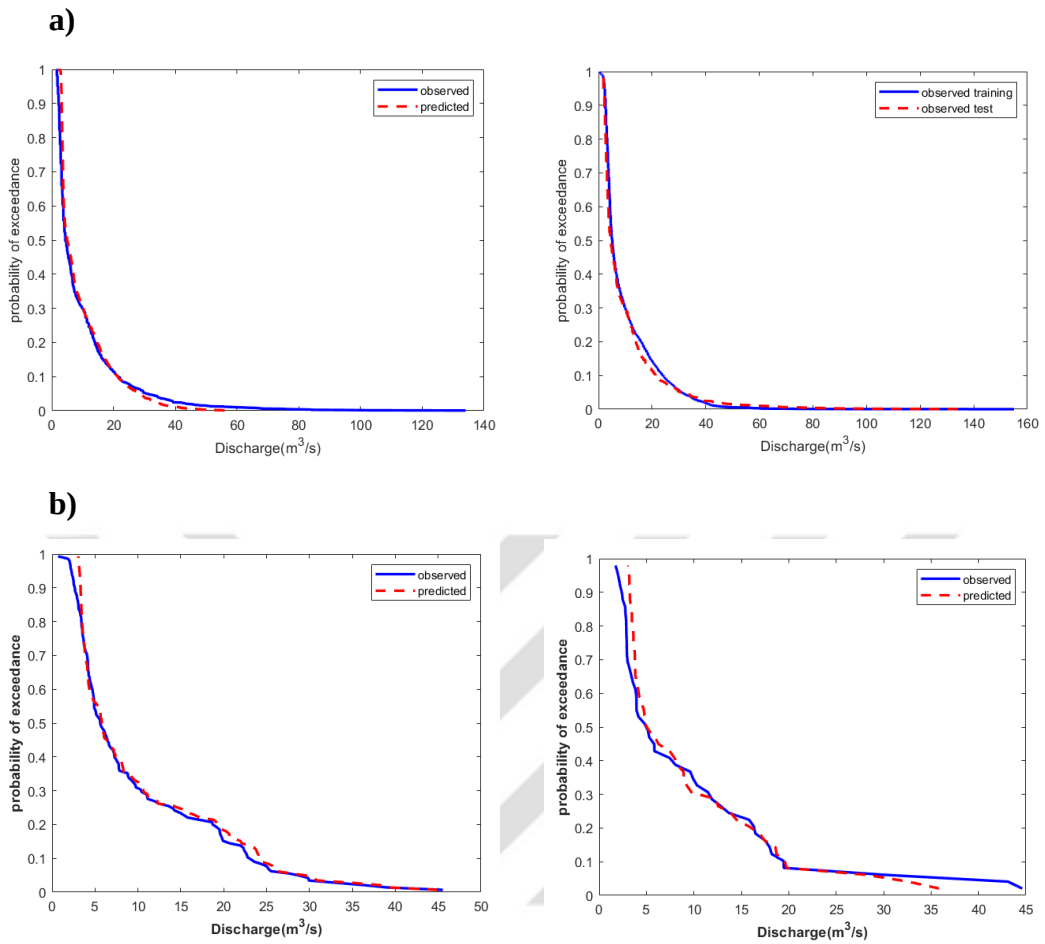


Figure 5.4 Flow frequency curves at station D18A19 during **a)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in daily timescale; **b)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in monthly timescale.

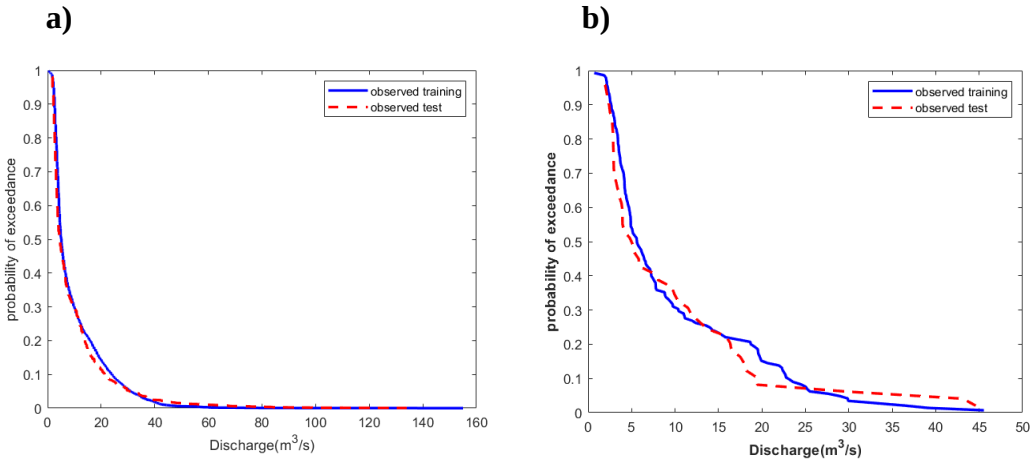


Figure 5.5 Flow frequency curves of observed data at station D18A19 **a)** in daily timescale (left figure); **b)** in monthly timescale (right figure).

Moreover, at this flow observation station, sufficient flow data is available to estimate other years using the set of variables and hyperparameters that give the best test results. The water years from 1991 to 1994 were used as a secondary test period, and the results indicate that the model is able to predict other years satisfactorily. RMSE is 4.95, R value is around 0.88, NSE is more than 0.74 and KGE exceed 0.85. Figures 5.6 and 5.7 demonstrate the time series and scatter plots of the second test simulation. During the second test period (1991-1994), R is 0.88 (compared to 0.90), NSE value is 0.74 (compared to 0.79), and KGE value is 0.86 (compared to 0.72) for daily flows.

Furthermore, Figures 5.8 and 5.9 indicate the flow frequency curves for both the predicted and observed data from the second test simulations, as well as comparison the observed data from the second test and the training period respectively.

According to the best simulation results of the station, historical data, which is from 1940 to end of the 1999 water year, were reconstructed. Figure 5.10 demonstrates the reconstructed river flow data of D18A19.

RMSE	R	NSE	KGE
4.9538	0.8793	0.7439	0.8556

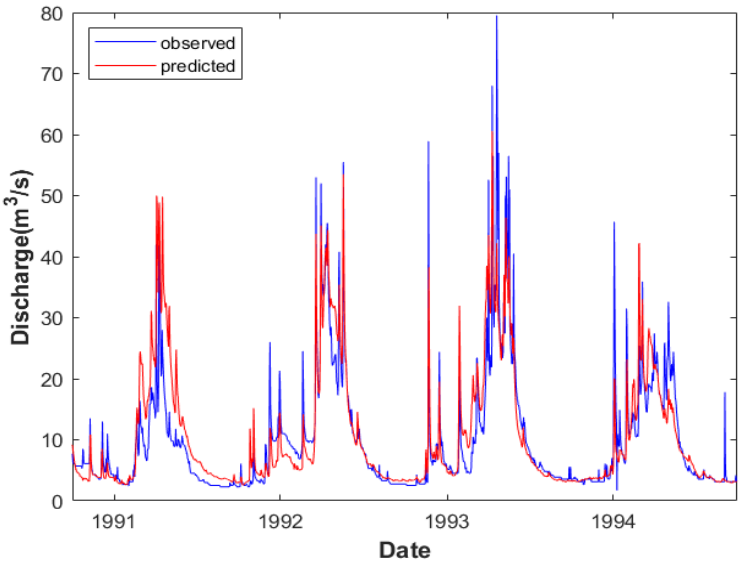


Figure 5.6 River flow simulation timeseries and performance statistics at station D18A19 during second test period of 1991-1994.

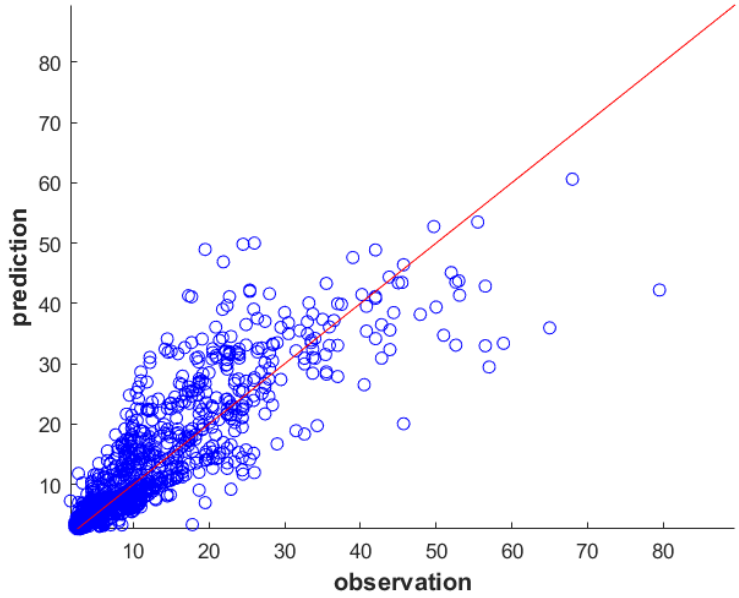


Figure 5.7 River flow simulation scatter plots of station D18A19 during second test period of 1991-1994.

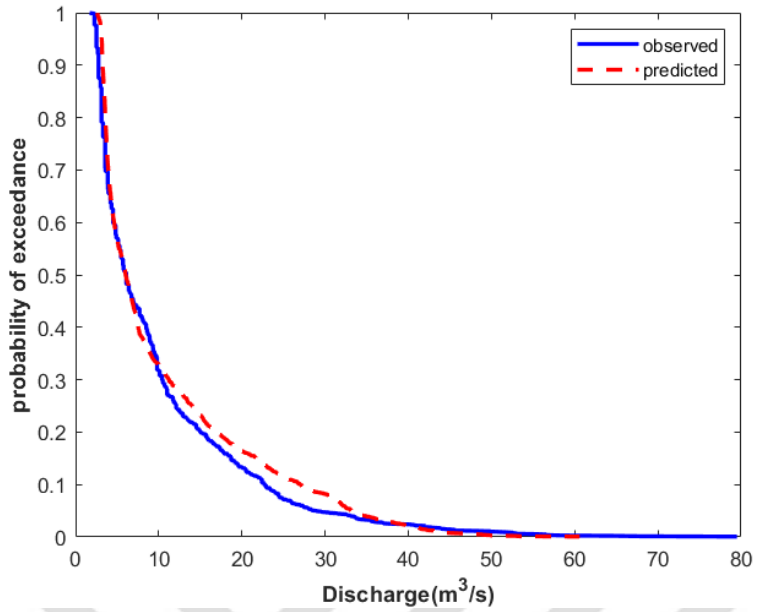


Figure 5.8 Flow frequency curve at station D18A19 during test period of 1991-1994.

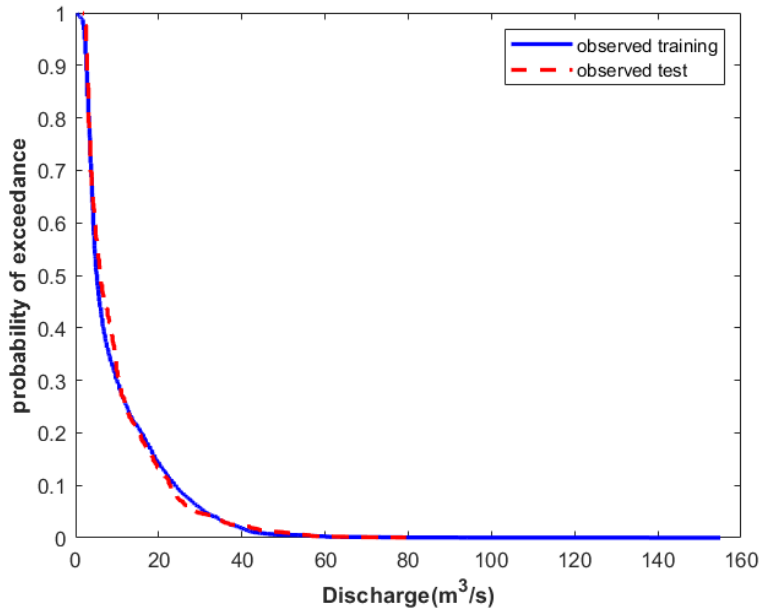


Figure 5.9 Flow frequency curves of observed data at station D18A19 during training and the second test period.

a) b)

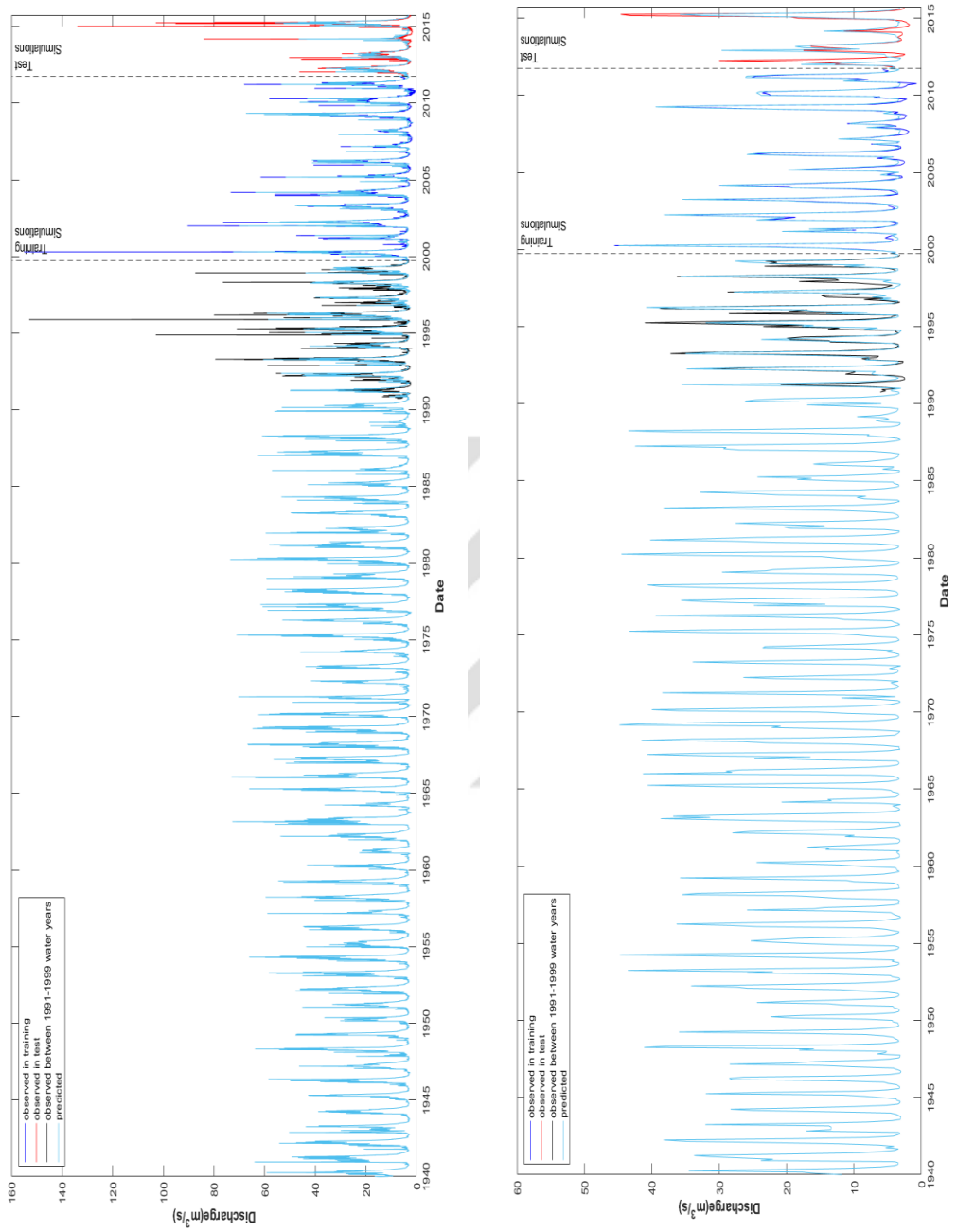


Figure 5.10 Reconstructed flow data at station D18A19 since 1940 **a)** in daily timescale; **b)** in monthly timescale.

5.1.1.1 D18A19 with longer training size

D18A19 station was modeled with longer training data in order to examine the effect of training data size on the model robustness. Therefore, instead of 12 years of training data size, 21 years of training data was (1991-2011 water years) used to simulate the river flow. The input variables that gave the best results in shorter training simulations were used in these simulations. On the other hand, different hyperparameters were tried because different parameter combinations may give better values when using longer data. However, just the epoch number slightly affected the results, indicating that additional training iterations can lead to better simulation results due to more data available. Therefore, the best results were obtained with the same parameters except that the epoch number was increased to 200 instead of 100. The best results are RMSE is 1.95, R value is 0.98, NSE is 0.97 and KGE is 0.97 in training simulations. The test results are RMSE is 5.2, R value is more than 0.90 NSE is around 0.80 and KGE is 0.76. When comparing the results with those obtained using 12 years of training data, there is no significant improvement. The simulations with 12 years of training data yield almost identical results. Hence, the 12 years of training data size is enough to get a robust model to simulate the river flow. Figures 5.11 and 5.12 show the time series and scatter plots. During the longer training period (1991-2011), R is 0.98 (compared to 0.97), NSE value is 0.97 (compared to 0.94), and KGE value is 0.97 (compared to 0.95) for daily flows. On other hand, during the test period, R is 0.90, NSE value is 0.79, and KGE value is 0.72 for daily flows.

Figure 5.13 represents the flow frequency curves of observed and simulated data for both training and test simulations and Figure 5.14 shows the flow frequency curves of observed training and test data.

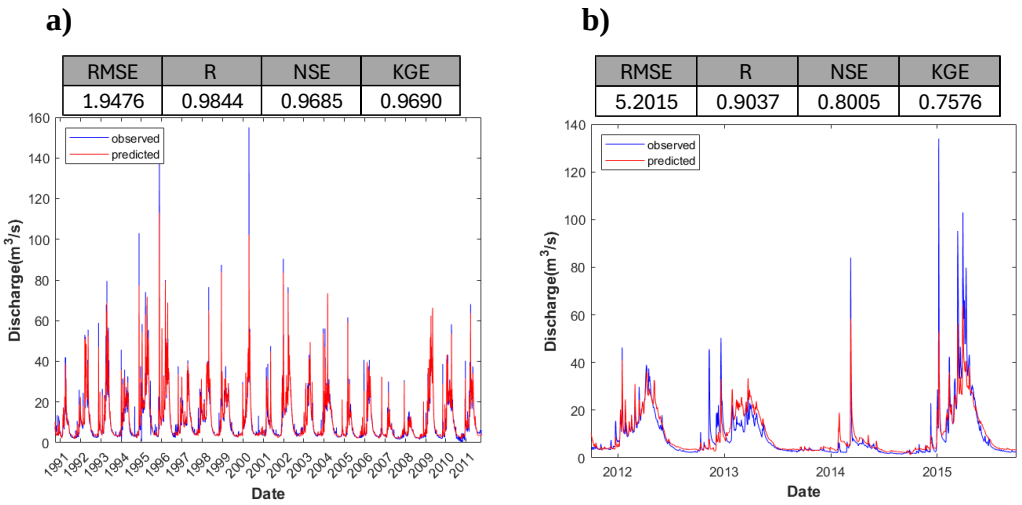


Figure 5.11 River flow simulation timeseries and performance statistics at station D18A19 during **a)** the training period of 1991-2011 (left figure); **b)** the test period of 2012-2015 (right figure) in daily timescale.

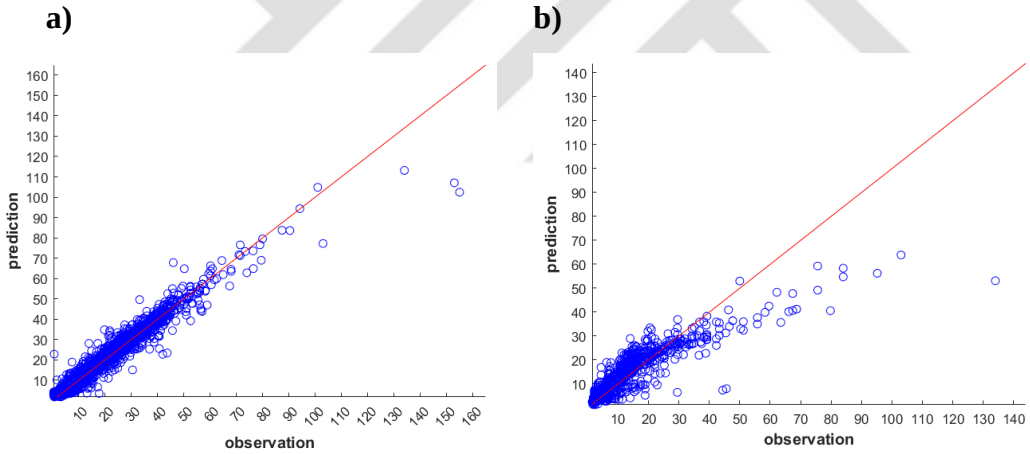


Figure 5.12 River flow simulation timeseries and performance statistics at station D18A19 during **a)** the training period of 1991-2011 (left figure); **b)** the test period of 2012-2015 (right figure) in daily timescale.

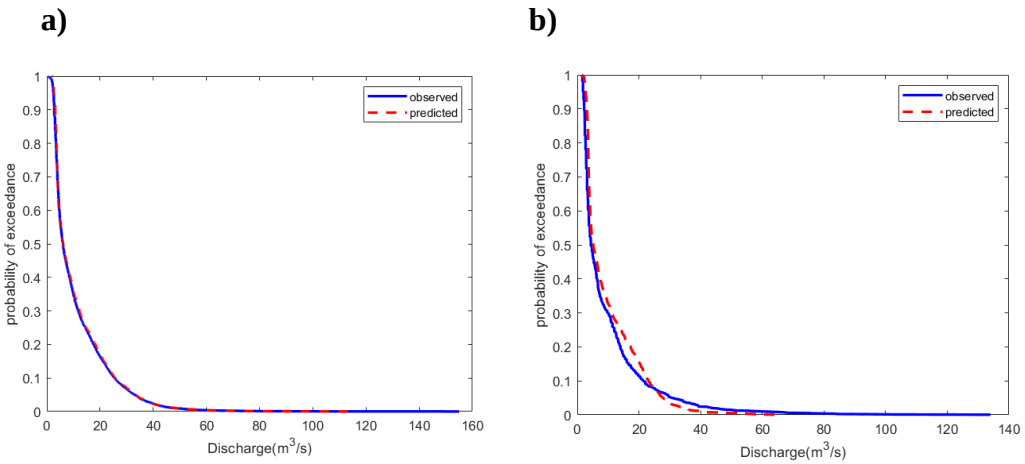


Figure 5.13 Flow frequency curves at station D18A19 during **a)** the training period of 2000-2011 (left figure); **b)** the test period of 2012-2015 (right figure) in daily timescale.

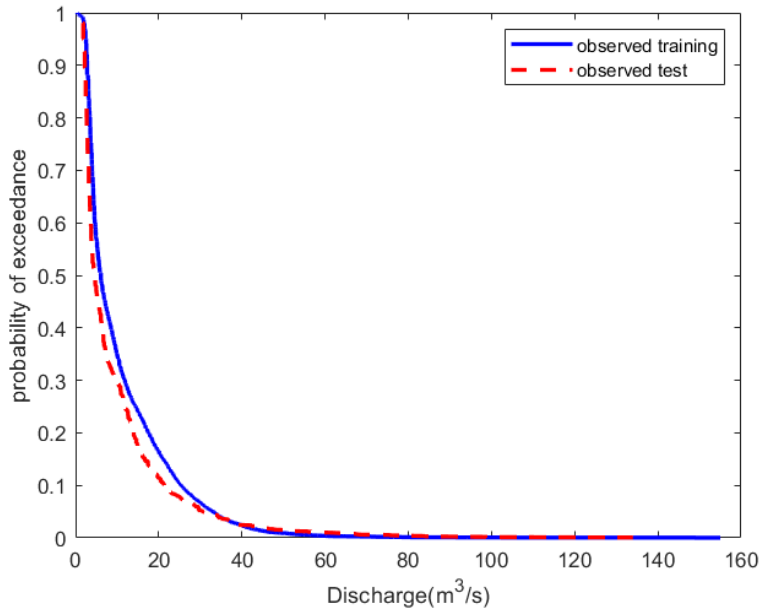


Figure 5.14 Flow frequency curves of observed data during the training and test periods at station D18A19 with longer training size.

5.1.1.2 Flow Station D18A19 with different training years

To examine the effects of simulating different years during the training period on the model accuracy, the years 2004-2015 were used for the training phase and the years 2000-2003 were used for testing the model (instead of the original simulation above for 2000-2011 for training, and 2012-2015 for test periods). The same input case and hyperparameter combination that performed best in the initial simulation was used. An improvement was observed in the test results. RMSE value is close to 2.64, R is 0.97, NSE is 0.93 and KGE is 0.93 in training simulations. Also, in test simulations, RMSE is 4.64 (compared to 5.39), R is 0.93 (compared to 0.90), NSE is 0.84(compared to 0.79), and KGE is 0.87 (compared to 0.72).

This is an indication of the versatility of data driven methods. Different results are obtained when different ranges of years are simulated during the training period and different years are tested, but as can be seen in these simulations, when a robust model is established, no extreme differences in results are observed or satisfactory results are obtained in each model.

However, this slight improvement seems to depend on the higher frequency of high flows in the simulated flow data during the new training period (2004-2015 instead of 2000-2011). The increased frequency of high flows in the training period has led to better predictions during the test period. Figures 5.15 and 5.16 show the time series and scatter plots of the simulations. Figure 5.17 shows the flow frequency curve of both training and test simulations. Besides, the difference between the probability of exceedance curves of observed training and test data is the evidence that the test simulation results are slightly worse than training simulation results and it shows in Figure 5.18.

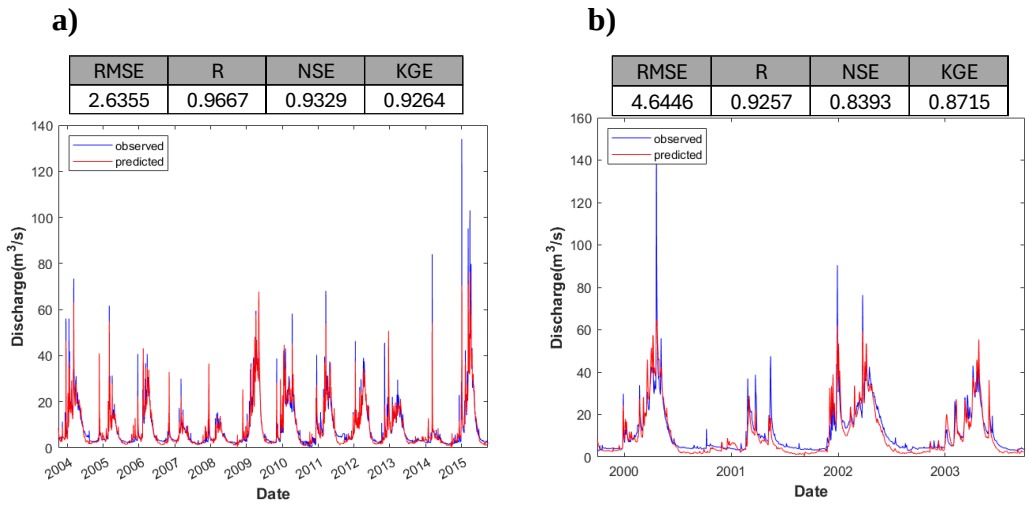


Figure 5.15 River flow simulation timeseries and performance statistics at station D18A19 during **a)** the training period of 2004-2015 (left figure); **b)** the test period of 2000-2003 (right figure) in daily timescale.

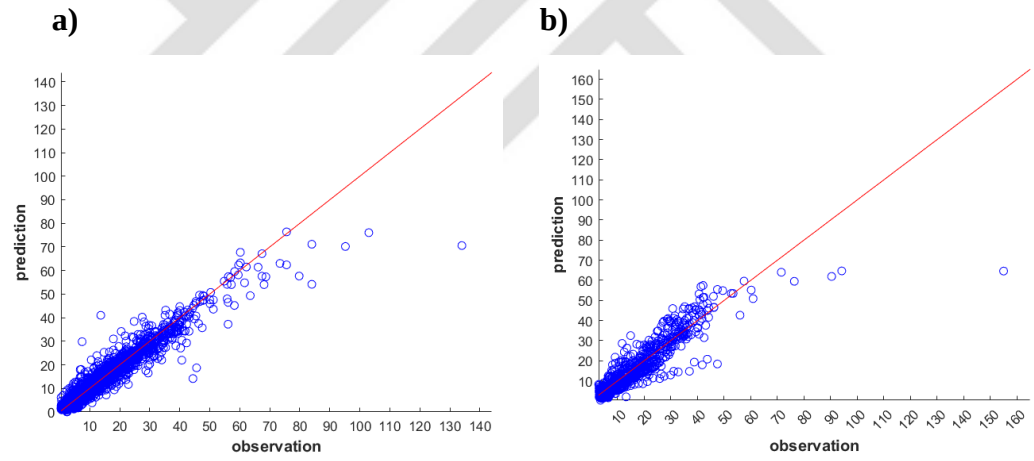


Figure 5.16 River flow simulation timeseries and performance statistics at station D18A19 during **a)** the training period of 2004-2015 (left figure); **b)** the test period of 2000-2003 (right figure) in daily timescale.

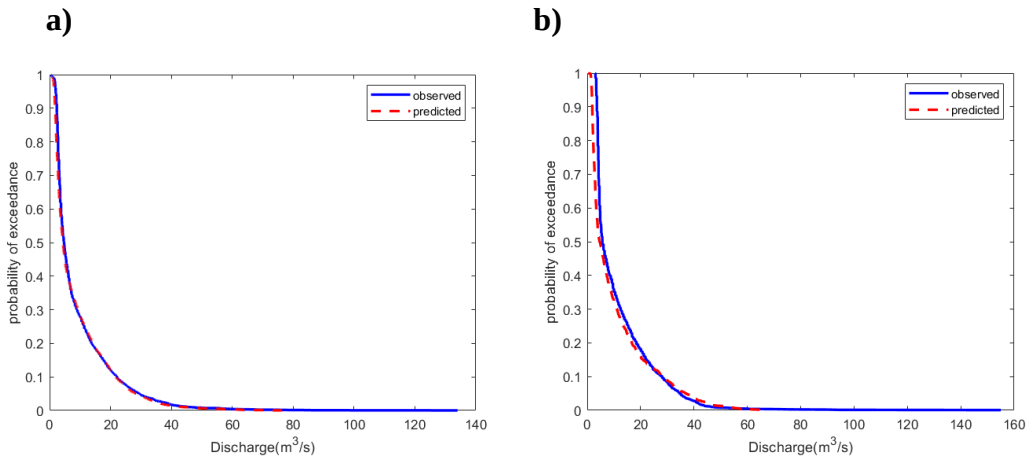


Figure 5.17 Flow frequency curves at station D18A19 during **a)** the training period of 2004-2015 (left figure); **b)** the test period of 2000-2003 (right figure) in daily timescale.

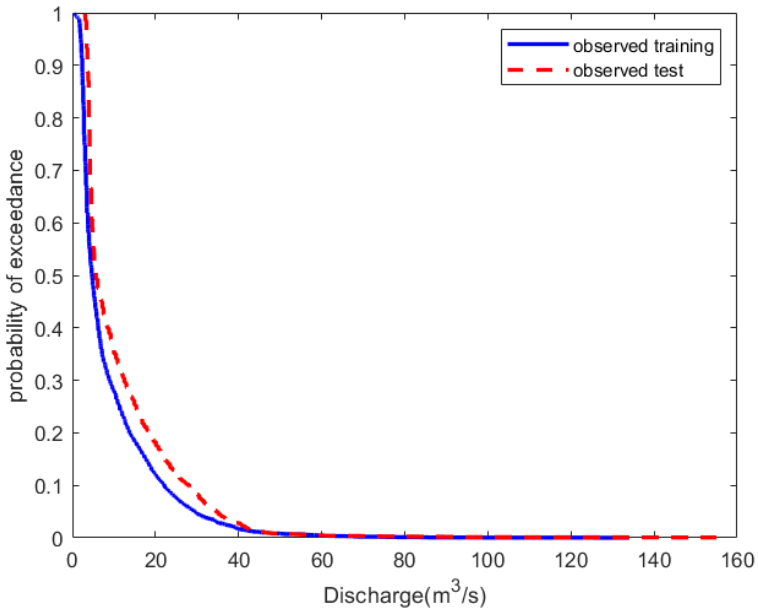


Figure 5.18 Flow frequency curves of observed data during the training and test periods at station D18A19.

5.1.2 Daily Flows at D18A27 at Niğde Elekgölü in Seyhan Watershed

Station D18A27 was evaluated with 10 different input variables for various simulation cases. Table 5.5 shows the best results of the different cases. The best test results were observed at Case 2 by using time and precipitation as input with the hyperparameters which are 128 hidden units, 150 epochs 0.002 initial learning rate, 0.1 dropout and ADAM optimizer. Also, 3 grids of the ERA5 dataset were chosen for the input data based on the station's location (see Figure 3.3).

Once the Taylor diagrams in Figure 5.19 are examined, the model, with different input variables, performs well in predicting the observed flow data. On the other hand, during test simulations, the other cases approximate case 1 which has poorer performance, but the results are still good enough for all different input combinations. Moreover, in training simulations, almost all case results are close to the standard deviation of the observed value, whereas in test simulations, the results deviate significantly from the standard deviation of the observed value. This situation explains the difference in the accuracy between training and test results.

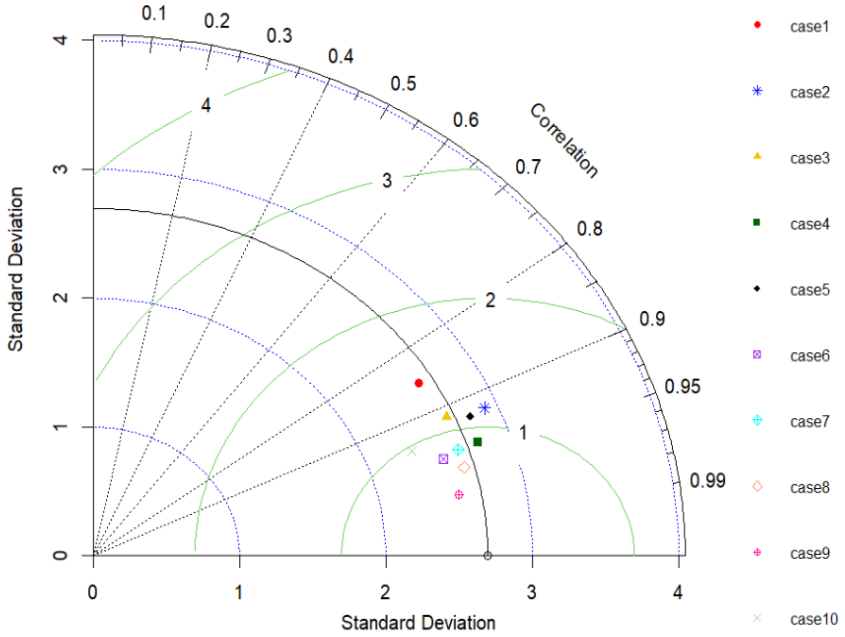
The below Figures 5.20 and 5.21 show the time series and scatter plots of the best test simulation of D18A27 station. During the training period, R is 0.92, NSE value is 0.80, and KGE value is 0.83 for daily flows; R is 0.96, NSE value is 0.87, and KGE value is 0.81 for monthly flows. On other hand, during the test period, R is 0.87, NSE value is 0.73, and KGE value is 0.86 for daily flows; R is 0.92, NSE value is 0.84, and KGE value is 0.90 for monthly flows.

Moreover, Figure 5.22 represents the flow frequency curve of the simulations. Besides, Figure 5.23 represents the flow frequency curves of observed training and test data, and it clearly shows that frequency of higher flows in the test data is greater than in the training data. Reconstructed flow data since 1940 can be seen in Figure 5.24.

Table 5.5 Performance statistics of the input variable cases for flow station D18A27 in Seyhan Watershed.

	TRAINING RESULTS				TEST RESULTS			
	RMSE	R	NSE	KGE	RMSE	R	NSE	KGE
Case1	1.4500	0.8566	0.7107	0.8189	1.9668	0.7825	0.6001	0.7524
Case2	1.2048	0.9187	0.8003	0.8272	**1.6026**	**0.8664**	**0.7345**	**0.8560**
Case3	1.1294	0.9133	0.8245	0.8882	1.6459	0.8569	0.7199	0.7877
Case4	0.8974	0.9482	0.8892	0.9184	1.7841	0.8495	0.6709	0.6778
Case5	1.1462	0.9216	0.8193	0.8477	1.6624	0.8597	0.7143	0.7368
Case6	0.8167	0.9539	0.9082	0.9101	2.1472	0.8211	0.5233	0.4811
Case7	0.8823	0.9494	0.8929	0.8964	1.8779	0.8417	0.6354	0.5980
Case8	0.7088	0.9650	0.9309	0.9544	2.0950	0.8084	0.5462	0.5504
Case9	0.5236	0.9823	0.9623	0.9325	1.8746	0.8662	0.6367	0.5627
Case10	0.9624	0.9371	0.8726	0.8485	1.8324	0.8278	0.6528	0.6971

a)



b)

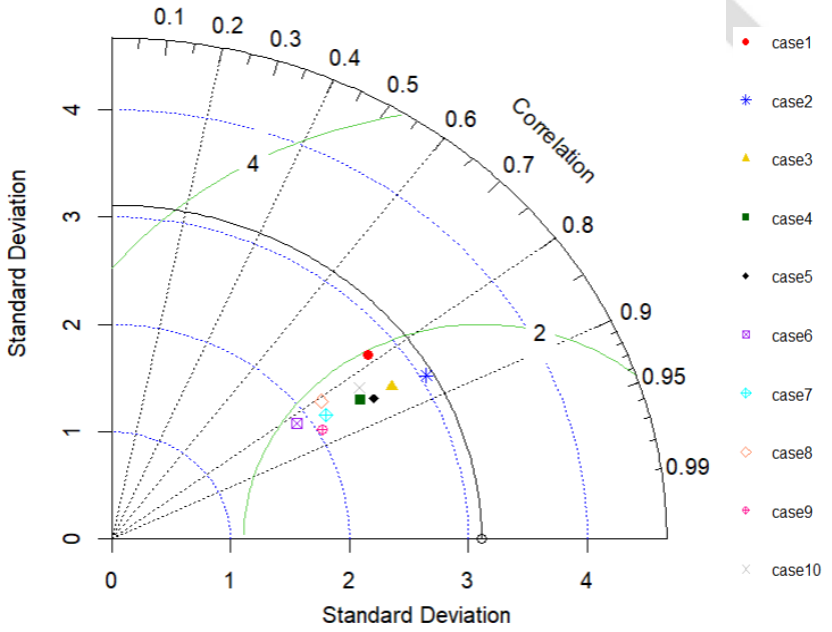
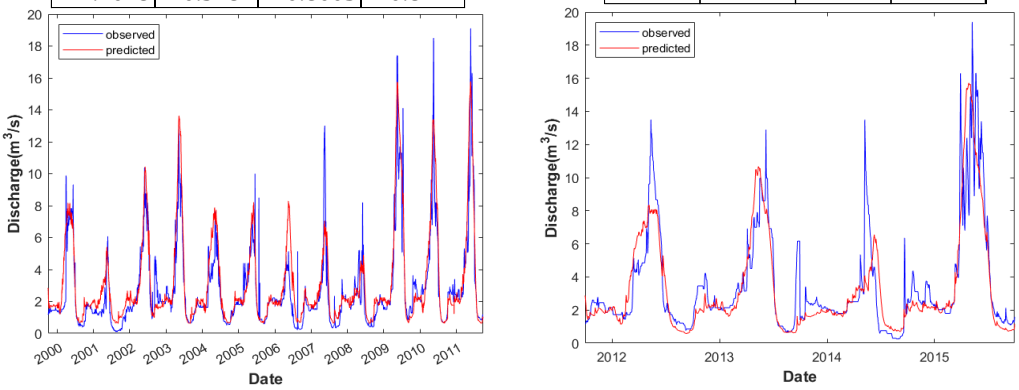


Figure 5.19 Taylor diagrams at station D18A27 during **a)** the training period of 2000-2011, and **b)** the test period of 2012-2015.

a)

RMSE	R	NSE	KGE
1.2048	0.9187	0.8003	0.8272

RMSE	R	NSE	KGE
1.6026	0.8664	0.7345	0.8560



b)

RMSE	R	NSE	KGE
0.9075	0.9578	0.8685	0.8147

RMSE	R	NSE	KGE
1.1558	0.9228	0.8358	0.9027

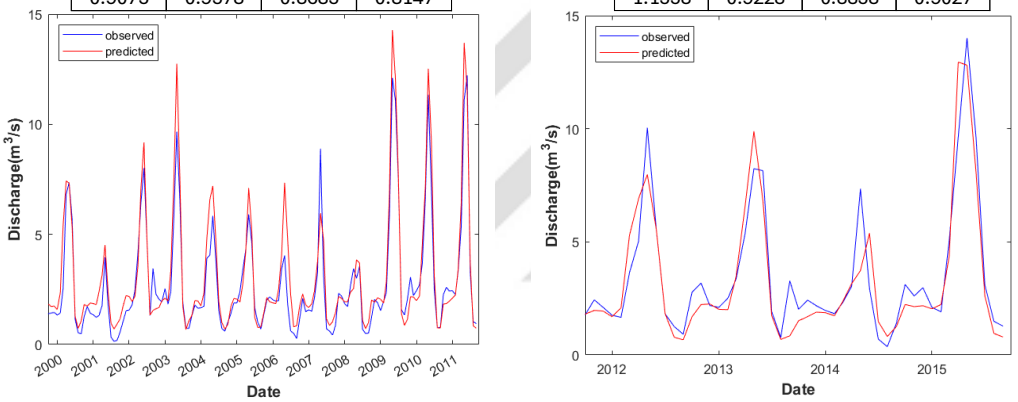


Figure 5.20 River flow simulation timeseries and performance statistics at station D18A27 during **a)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in daily timescale; **b)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in monthly timescale.

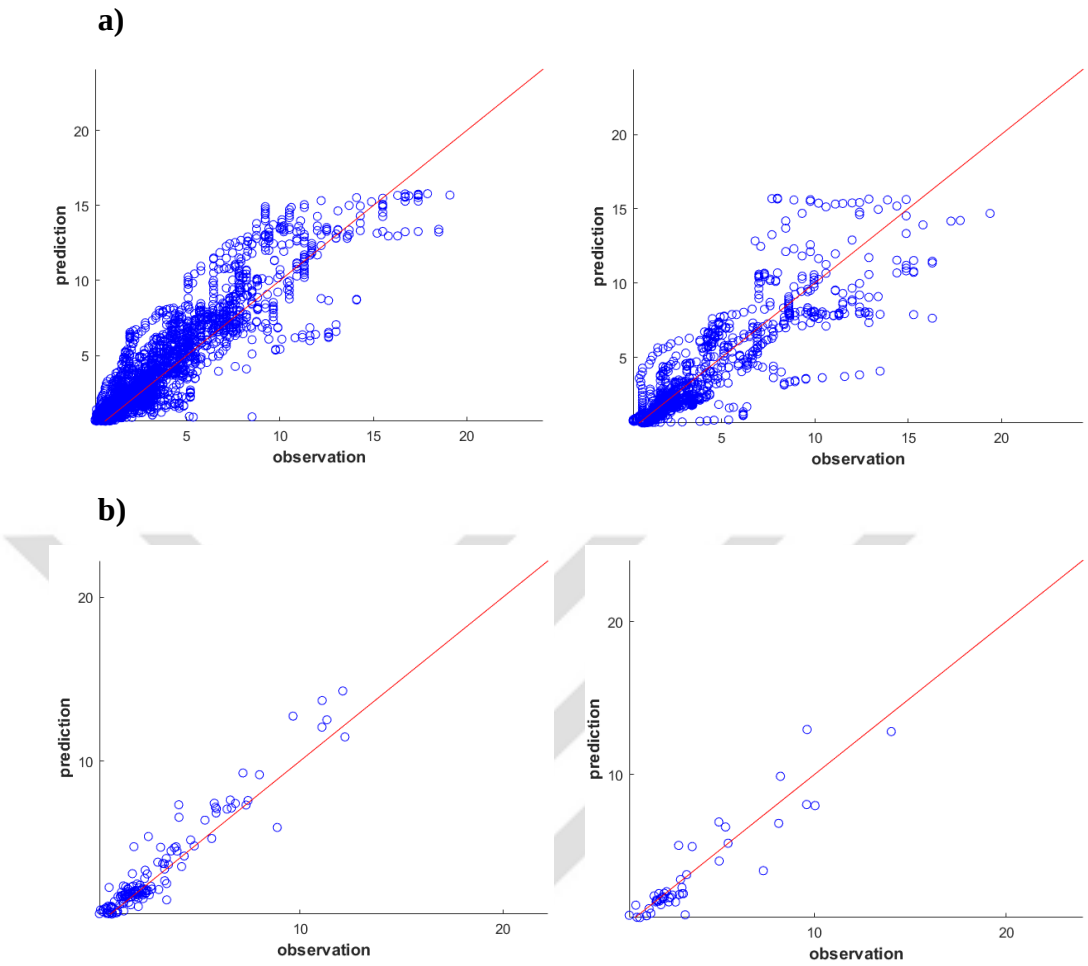


Figure 5.21 River flow simulation scatter plots at station D18A27 during **a)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in daily timescale; **b)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in monthly timescale.

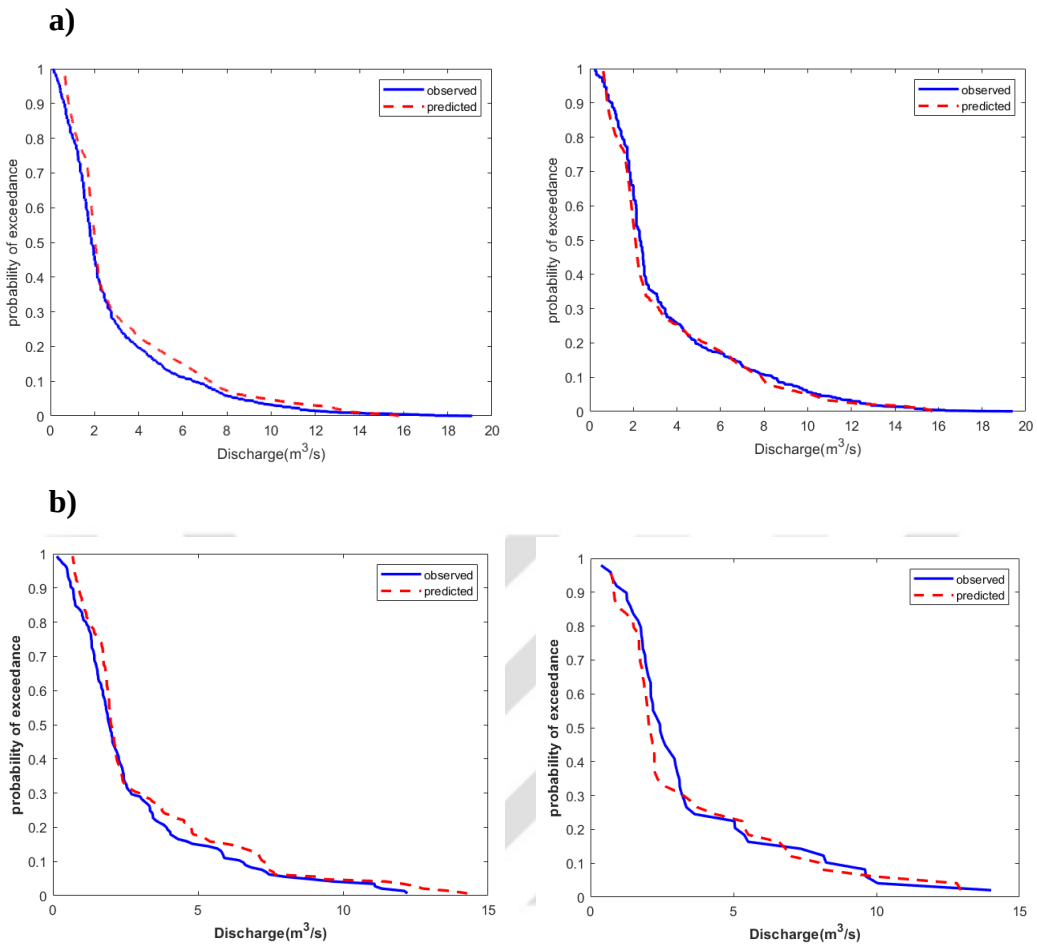


Figure 5.22 Flow frequency curves at station D18A27 during **a)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in daily timescale; **b)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in monthly timescale.

a) **b)**

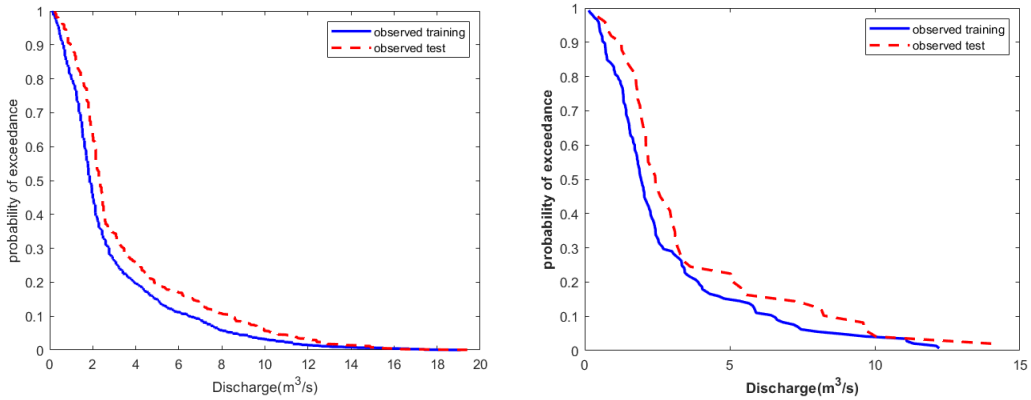


Figure 5.23 Flow frequency curves of observed data at station D18A27 **a)** in daily timescale (left figure); **b)** in monthly timescale (right figure).

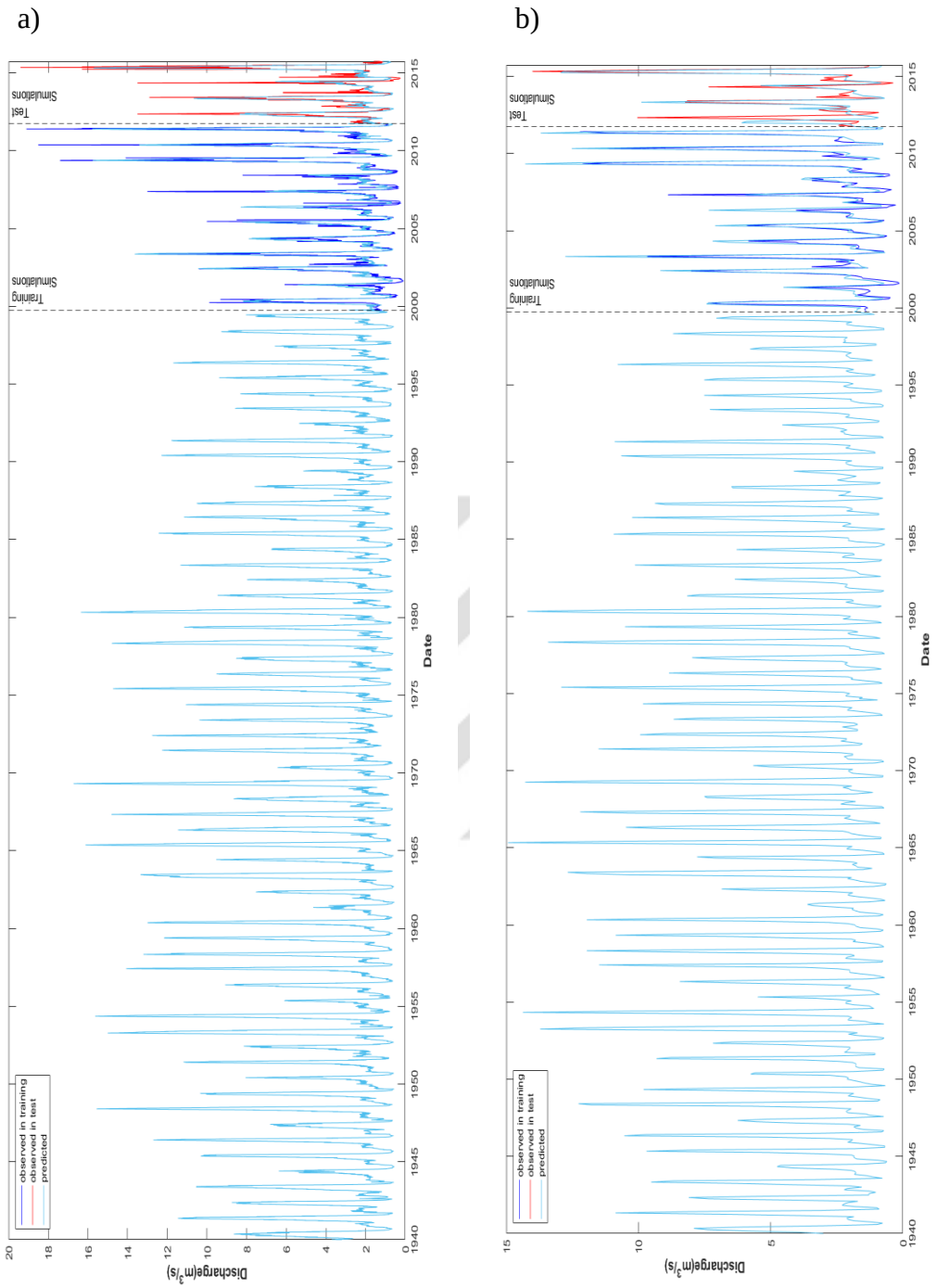


Figure 5.24 Reconstructed flow data at station D18A27 since 1940 **a)** in daily timescale; **b)** in monthly timescale.

5.1.3 Daily Flows at D21A167 at Erzurum Hinis in Euphrates Watershed

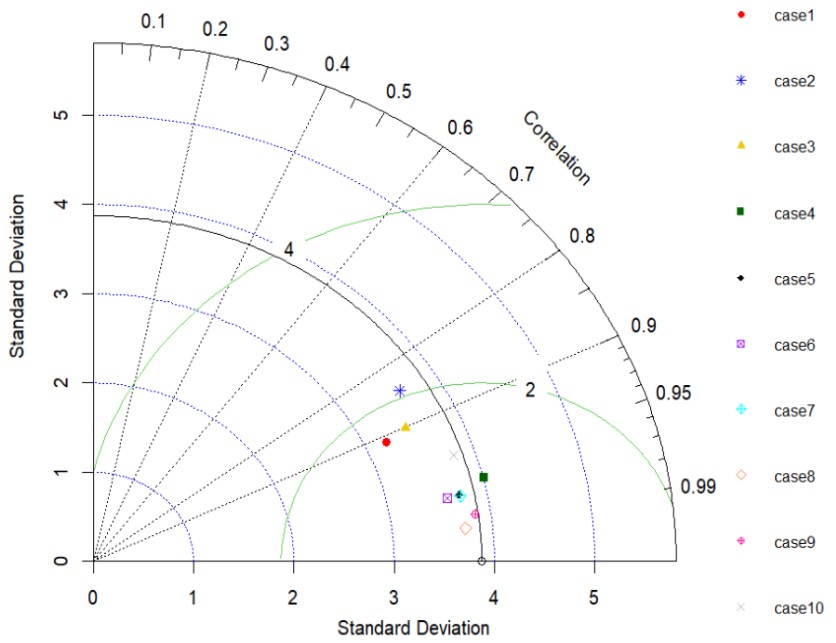
In order to find the optimum results, 10 different input combinations have been tried. Table 5.6 indicates the best results of each input variation. Moreover, the best results are obtained with case 6, which is the combination of time, temperature, potential evaporation, 1-day lag of precipitation and precipitation. In addition, hidden units were chosen as 128, epochs number is 200, and initial learning, dropout, and optimizer were selected as 0.002, 0.2, and ADAM, respectively. The highest results were achieved at this station, indicating that different regions produce distinct outcomes. This variation occurs because the impact of input variables may differ depending on the station's location.

When examining the Taylor diagrams in Figure 5.25, the models with most of the input variations show very good correlations and predictions close to the observed values, except for the first three cases during training simulations. The correlation coefficient values are close to 1, which represents a perfect condition. In the test phase, the other simulations are similar to the first cases, but they still exhibit high correlation values. In Figures 5.26 and 5.27, time series and scatter plots are demonstrated for the training and test phases of the simulation. During the training period, R is 0.98, NSE value is 0.96, and KGE value is 0.92 for daily flows; R is 0.99, NSE value is 0.99, and KGE value is 0.94 for monthly flows. On other hand, during the test period, R is 0.91, NSE value is 0.82, and KGE value is 0.81 for daily flows; R is 0.97, NSE value is 0.92, and KGE value is 0.95 for monthly flows. Although the statistical metrics are better than the other stations, the model struggled to predict high flows and generally underestimated river flows in the simulation for this station as can be seen from the time series graphs and scatter plots. In Figure 5.28, flow frequency curves are seen for both training and test data points. Moreover, Figure 5.29 shows the probability of exceedance curves of observed data. They have an almost perfect match. Therefore, the LSTM model simulates the river flows with high accuracy. Reconstructed flow data since 1940 can be seen in Figure 5.30.

Table 5.6 Performance statistics of the input variable cases for flow station D21A167 in Euphrates Watershed.

	TRAINING RESULTS				TEST RESULTS			
	RMSE	R	NSE	KGE	RMSE	R	NSE	KGE
Case1	1.6422	0.9095	0.8202	0.8048	2.1265	0.8042	0.6420	0.6762
Case2	2.0910	0.8474	0.7086	0.8220	2.0015	0.8448	0.6829	0.8324
Case3	1.6758	0.9019	0.8128	0.8515	1.8228	0.8640	0.7370	0.8299
Case4	0.9490	0.9719	0.9400	0.9464	1.5289	0.9046	0.8150	0.8904
Case5	0.8654	0.9759	0.9501	0.9269	2.1790	0.8700	0.6241	0.5020
Case6	0.7944	0.9804	0.9579	0.9209	**1.4931**	**0.9143**	**0.8235**	**0.8147**
Case7	0.7646	0.9805	0.9610	0.9582	1.8811	0.8571	0.7199	0.7391
Case8	0.4115	0.9949	0.9887	0.9594	1.6966	0.8824	0.7721	0.8313
Case9	0.5353	0.9904	0.9809	0.9872	1.7160	0.8770	0.7669	0.8396
Case10	1.2303	0.9487	0.8991	0.9414	2.3213	0.8140	0.5734	0.5074

a)



b)

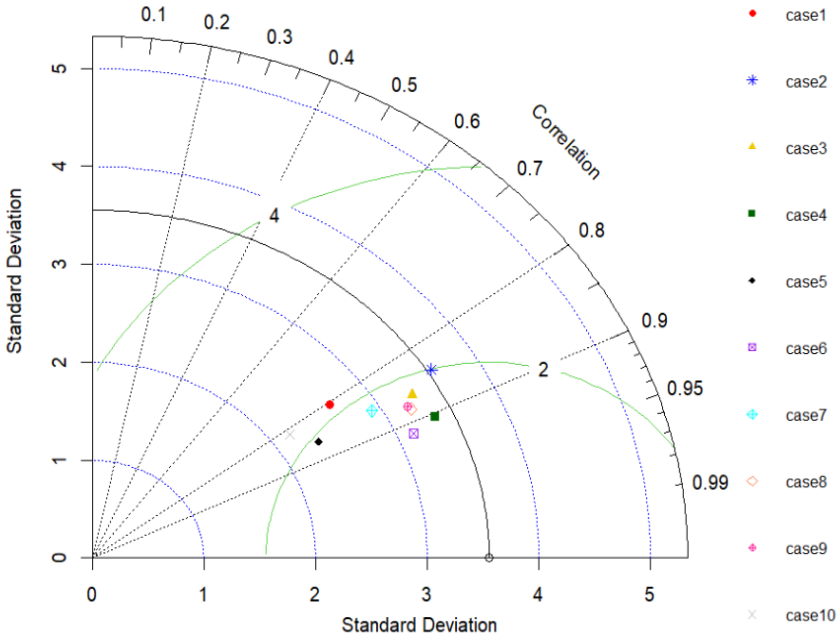
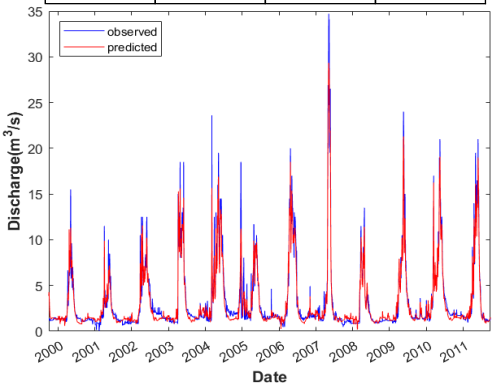


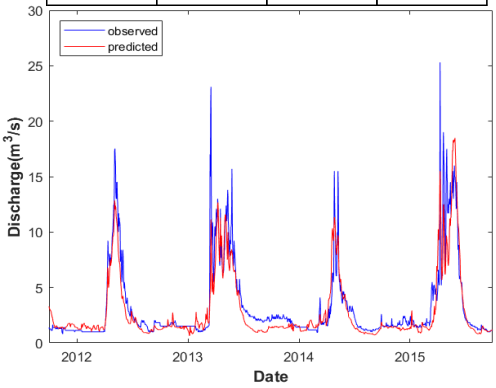
Figure 5.25 Taylor diagrams at station D21A167 during **a)** the training period of 2000-2011, and **b)** the test period of 2012-2015.

a)

RMSE	R	NSE	KGE
0.7944	0.9804	0.9579	0.9209

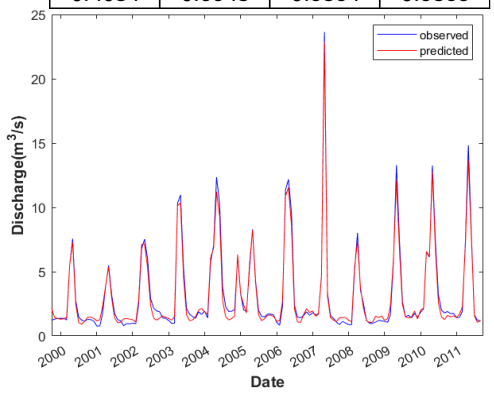


RMSE	R	NSE	KGE
1.4931	0.9143	0.8235	0.8147



b)

RMSE	R	NSE	KGE
0.4084	0.9948	0.9864	0.9365



RMSE	R	NSE	KGE
0.8906	0.9680	0.9185	0.8473

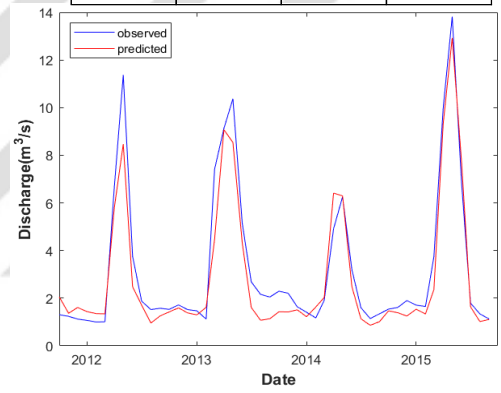
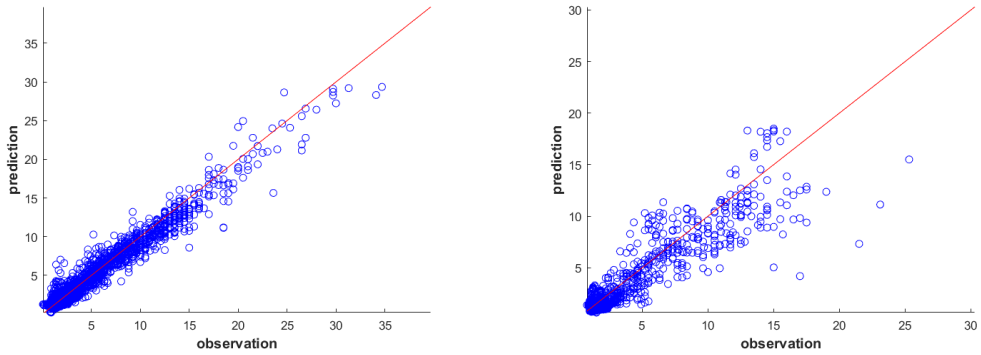


Figure 5.26 River flow simulation timeseries and performance statistics at station D21A167 during **a)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in daily timescale; **b)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in monthly timescale.

a)



b)

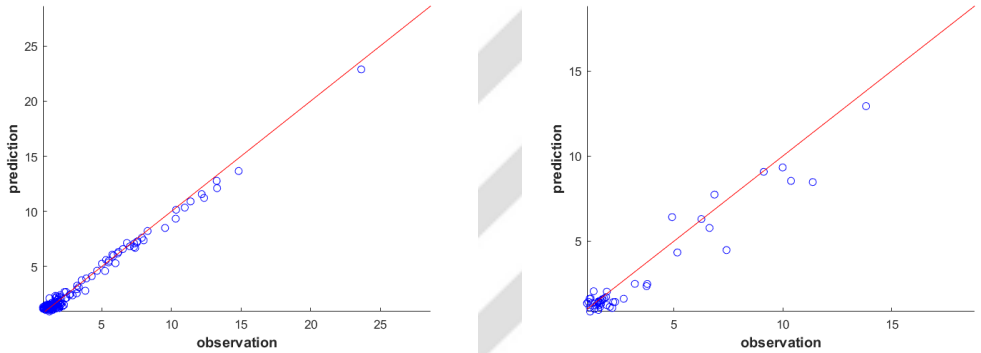


Figure 5.27 River flow simulation scatter plots at station D21A167 during **a)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in daily timescale; **b)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in monthly timescale.

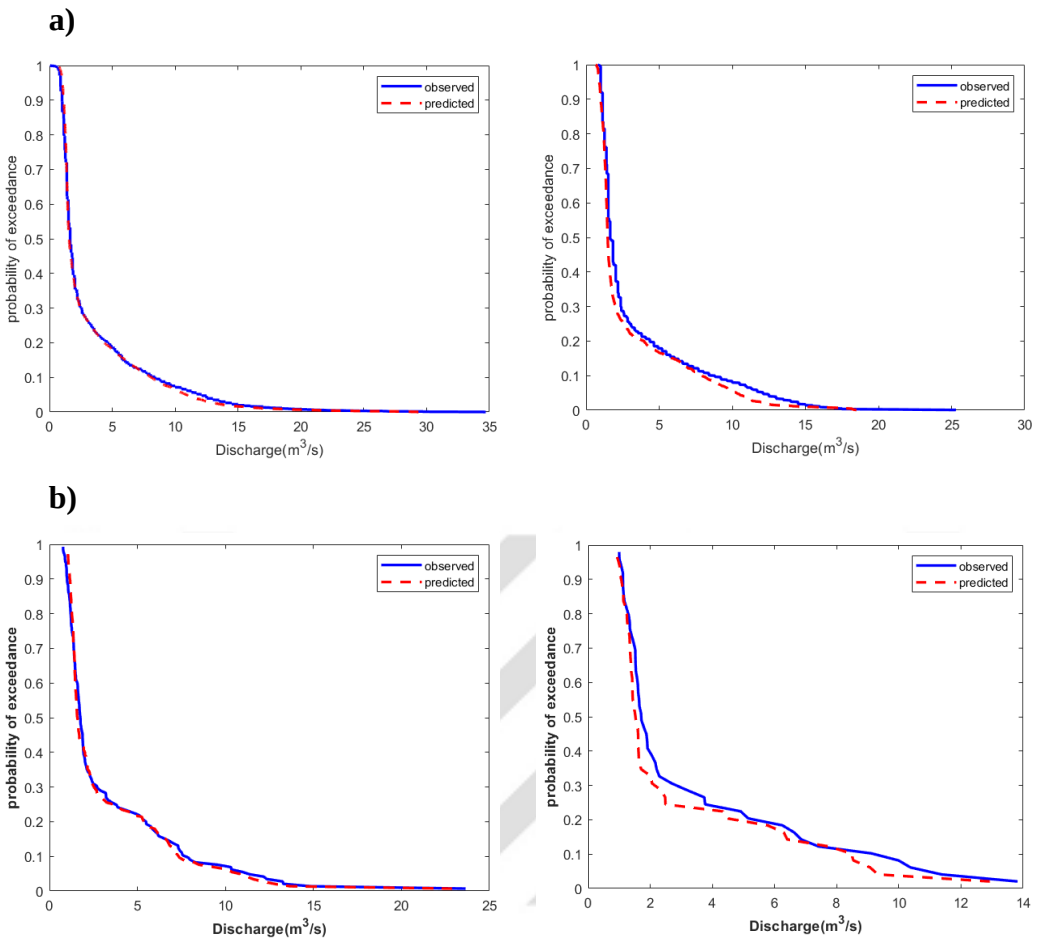


Figure 5.28 Flow frequency curves at station D21A167 during **a)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in daily timescale; **b)** the training period of 2000-2011 (left figure), and the test period of 2012-2015 (right figure) in monthly timescale.

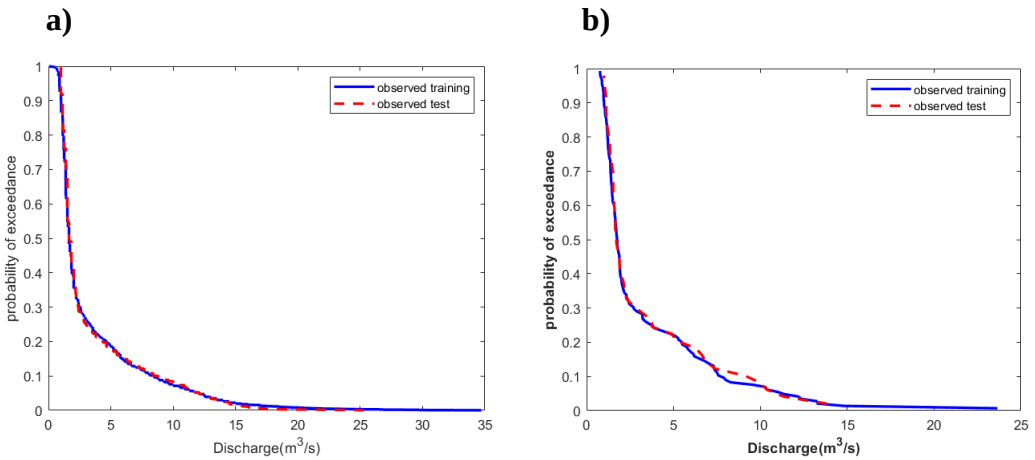


Figure 5.29 Flow frequency curves of observed data at station D21A167 **a)** in daily timescale (left figure); **b)** in monthly timescale (right figure).

.

a) b)

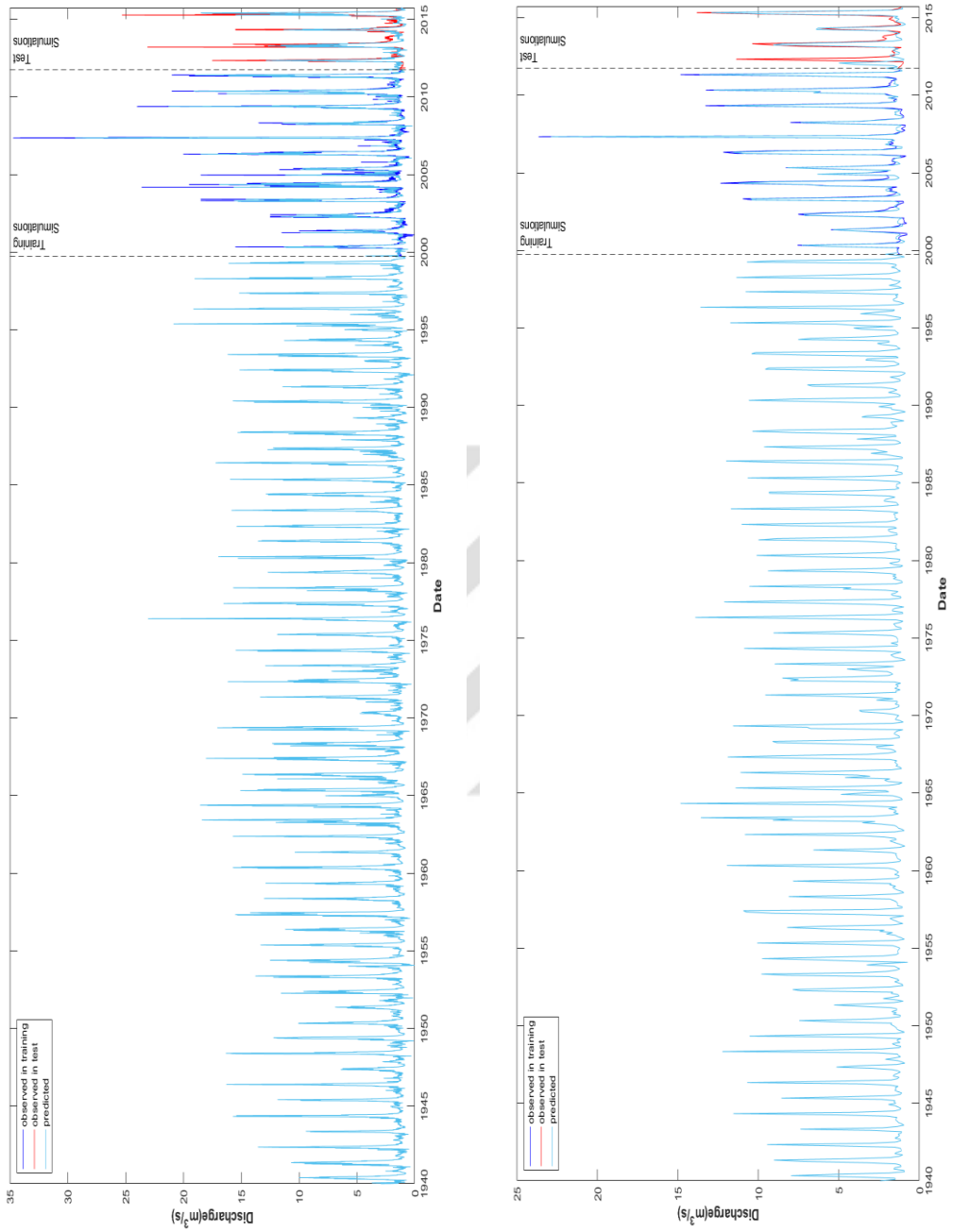


Figure 5.30 Reconstructed flow data at station D21A167 since 1940 **a)** in daily timescale; **b)** in monthly timescale.

CHAPTER 6

CONCLUDING REMARKS AND FUTURE DIRECTIONS

The purpose of this study was to develop a methodology to reconstruct historical daily and monthly river flows by an LSTM model, a recurrent neural network-type deep learning model, utilizing the ERA-5 Reanalysis climate dataset (Hersbach et al., 2020) as inputs. ERA-5 Reanalysis provides approximately 31 km grid resolution climate data in hourly time intervals and daily datasets were obtained from the hourly dataset. Reanalysis datasets, including ERA-5, provide the most complete picture currently possible of past weather and climate. They combine observations with state-of-the-art weather model simulations and are globally complete and consistent in time. Flow modeling component was chosen as LSTM, as deep learning has recently shown great success in various fields of science including hydrology and climate research due to the availability of large data sets, advances in parallel computing, better regularization and optimization techniques, and new software platforms that allow easy integration into a GPU framework (Sengupta et al., 2020).

As a demonstration, river flows were modeled with the LSTM method by the proposed methodology at two flow stations in the Seyhan Watershed and one station in Euphrates Watershed in Türkiye. Then the historical daily and monthly river flows were reconstructed since 1940, since when ERA5 dataset is available. Daily observed discharge data collected by Turkish State Hydraulic Works were used as the target data. Different input variables were used in different combinations to investigate the impacts of the climate variables on streamflow modeling. Moreover, various sets of hyperparameters of LSTM were employed to understand the effects of hyperparameters on the model robustness and to achieve the best results. RMSE, R, NSE, and KGE were used as statistical model evaluation metrics between observations and simulations. By examining Taylor diagrams, the importance of input data was demonstrated. Furthermore, flow frequency curves were examined to

investigate river flows during the training and test periods at selected stations. The following conclusions can be drawn from this study:

- At the three flow stations analyzed, during the training period, R is within 0.96-0.99 for monthly flows and 0.92-0.98 for daily flows, NSE value is within 0.87-0.99 for monthly flows and 0.80-0.96 for daily flows, and KGE value is within 0.81-0.95 for monthly flows and 0.83-0.92 for daily flows. On the other hand, during the test period, R ranges within 0.92-0.97 for monthly flows and 0.87-0.91 for daily flows, NSE value ranges within 0.84-0.92 for monthly flows and 0.73-0.82 for daily flows, and KGE value ranges within 0.82-0.90 for monthly flows and 0.72-0.86 for daily flows.
- The effect of the training period length on the test results was examined by increasing the training period from 12 years to 21 years. No significant improvement in the model's performance was observed. This demonstrates that 12 years on a daily basis is a sufficient length for modeling river flows using the LSTM and historical climate reanalysis data for the particular station.
- To examine the effects of simulating different years during the training period on the model accuracy with the same length, the years 2004-2015 were used for the training phase, and the years 2000-2003 were used for testing the model (instead of the original simulation for 2000-2011 as training, and 2012-2015 as test periods). The same input variable case and hyperparameter combination were used. An improvement was observed in the test results: RMSE became 4.64 (compared to 5.39), R became 0.93 (compared to 0.90), NSE became 0.84(compared to 0.79), and KGE became 0.87 (compared to 0.72). This slight improvement seems to depend on the higher frequency of high flows in the simulated flow data during the new training period (2004-2015 instead of 2000-2011). The increased frequency of high flows in the training period has led to better predictions during the test period.
- It was observed that the LSTM model has difficulty in predicting high flows. This is because the model encounters limited examples of such extreme

events during training, making it challenging to re-estimate rare data accurately. On other hand, different inputs affect the model at various locations due to variations in topographic and climatic conditions. Therefore, the choice of input variable should be investigated carefully. It is clear from the simulations that the most important input variable after precipitation is temperature. The effect of other variables may vary for different stations.

- With the LSTM model, river flows can be modeled and reconstructed by solely the climate reanalysis dataset satisfactorily at arbitrary stations. The proposed approach does not need either the physical governing equations that are required for the physical or conceptual models, or the physical parameters such as those needed for topography, land use land cover, or soil information. Therefore, the proposed approach herein can reconstruct river periods for long durations, for example from 1940 to 2000 after training and testing the LSTM model. Some other studies only focus on obtaining missing river flows for shorter durations, in this regard, they can also use the earlier river flow data as input to the data driven model.
- To obtain better simulation results in the future, river flows can be also modeled by using observed historical climate datasets with sufficient time and space resolution. Models at the locations with abundant climate observations can be compared with proposed approach for consistency.
- This study utilized 31km resolutions ERA5 Climate Reanalysis, but in the future, the proposed methodology can utilize dynamical downscaling of climate reanalysis datasets, which may capture the high flows better.
- The climate reanalysis dataset is available on hourly basis, so if hourly flow data was accessible in the future, hourly models can also be built, and hourly flow data can be reconstructed. Hourly flows are particularly needed for better modeling of floods.
- Artificial intelligence is evolving rapidly, and new methods are constantly being developed. On the other hand, especially noticeable in this study, the LSTM model had difficulty in capturing high flows, so in order to improve

the simulation performance, new artificial intelligence methods can be employed to increase the accuracy of the simulations, or hybrid models that combine physical models with data-driven approaches can be developed.

REFERENCES

Abdulkareem, J. H., Pradhan, B., Sulaiman, W. N. A., & Jamil, N. R. (2018). Review of studies on hydrological modelling in Malaysia. *Modeling Earth Systems and Environment*, *4*, 1577-1605.

Affonso, C., Rossi, A. L. D., Vieira, F. H. A., & de Leon Ferreira, A. C. P. (2017). Deep learning for biological image classification. *Expert systems with applications*, *85*, 114-122.

Aghakouchak, A., & Habib, E. (2010). Application of a conceptual hydrologic model in teaching hydrologic processes. *International Journal of Engineering Education*, *26*(4 (S1)), 963-973.

Ahmed, S. F., Alam, M. S. B., Hassan, M., Rozbu, M. R., Ishtiak, T., Rafa, N., ... & Gandomi, A. H. (2023). Deep learning modelling techniques: current progress, applications, advantages, and challenges. *Artificial Intelligence Review*, *56*(11), 13521-13617.

Alsavaf, Y., & Teksoy, A. (2022). Applicability of recurrent neural networks to retrieve missing runoff records: challenges and opportunities in Turkey. *Environmental Monitoring and Assessment*, *194*(1), 28.

Apaydin, H., Feizi, H., Akcakoca, F., & Sattari, M. T. (2022, May). Daily Streamflow Modelling in the Nalli River Using Recurrent Neural Networks. In *International Conference "New Technologies, Development and Applications"* (pp. 813-822). Cham: Springer International Publishing.

Beale, M. H., Hagan, M. T., & Demuth, H. B. (2018). Deep learning toolbox. *User's Guide, The MathWorks, Inc., Natick, MA*.

Beck, H. E., Pan, M., Roy, T., Weedon, G. P., Pappenberger, F., Van Dijk, A. I., ... & Wood, E. F. (2019). Daily evaluation of 26 precipitation datasets using Stage-IV gauge-radar data for the CONUS. *Hydrology and Earth System Sciences*, *23*(1), 207-224.

Beck, H. E., Van Dijk, A. I., De Roo, A., Dutra, E., Fink, G., Orth, R., & Schellekens, J. (2017). Global evaluation of runoff from 10 state-of-the-art hydrological models. *Hydrology and Earth System Sciences*, *21*(6), 2881-2903.

Bergström, S. (1976). *Development and application of a conceptual runoff model for Scandinavian catchments*.

Bhanja, S., & Das, A. (2018). Impact of data normalization on deep neural network for time series forecasting. *arXiv preprint arXiv:1812.05519*.

Bierkens, M. F., Bell, V. A., Burek, P., Chaney, N., Condon, L. E., David, C. H., ... & Sudicky, E. (2015). Hyper-resolution global hydrological modelling: what is next?. *Hydrological processes*, *29*(2).

Bourdin, D. R., Fleming, S. W., & Stull, R. B. (2012). Streamflow modelling: a primer on applications, approaches and challenges. *Atmosphere-Ocean*, *50*(4), 507-536.

Büchele, B., Kreibich, H., Kron, A., Thieken, A., Ihringer, J., Oberle, P., ... & Nestmann, F. (2006). Flood-risk mapping: contributions towards an enhanced assessment of extreme events and associated risks. *Natural Hazards and Earth System Sciences*, *6*(4), 485-503.

Cantoni, E., Tramblay, Y., Grimaldi, S., Salamon, P., Dakhlaoui, H., Dezetter, A., & Thiemig, V. (2022). Hydrological performance of the ERA5 reanalysis for flood modeling in Tunisia with the LISFLOOD and GR4J models. *Journal of Hydrology: Regional Studies*, *42*, 101169.

Catalán, N., Marcé, R., Kothawala, D. N., & Tranvik, L. J. (2016). Organic carbon decomposition rates controlled by water retention time across inland waters. *Nature Geoscience*, *9*(7), 501-504.

Choi, C., Kim, J., Han, H., Han, D., & Kim, H. S. (2019). Development of water level prediction models using machine learning in wetlands: A case study of Upo wetland in South Korea. *Water*, *12*(1), 93.

Das, T., Bárdossy, A., Zehe, E., & He, Y. (2008). Comparison of conceptual model performance using different representations of spatial variability. *Journal of Hydrology*, *356*(1-2), 106-118.

Dye, P. J., & Croke, B. F. (2003). Evaluation of streamflow predictions by the IHACRES rainfall-runoff model in two South African catchments. *Environmental Modelling & Software*, *18*(8-9), 705-712.

Fan, H., Jiang, M., Xu, L., Zhu, H., Cheng, J., & Jiang, J. (2020). Comparison of long short term memory networks and the hydrological model in runoff simulation. *Water*, *12*(1), 175.

Fekete, B. M., & Vörösmarty, C. J. (2007). The current status of global river discharge monitoring and potential new technologies complementing traditional discharge measurements. *IAHS publ*, *309*, 129-136.

Fritsch, S., Guenther, F. and Wright, M.N. (2019). Neuralnet: Training of neural networks, version 1.44.2., R package. Retrieved from CRAN.R-project.org/package=neuralnet

Gao, S., Huang, Y., Zhang, S., Han, J., Wang, G., Zhang, M., & Lin, Q. (2020). Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. *Journal of Hydrology*, *589*, 125188.

Ghiggi, G., Humphrey, V., Seneviratne, S. I., & Gudmundsson, L. (2019). GRUN: an observation-based global gridded runoff dataset from 1902 to 2014. *Earth System Science Data*, *11*(4), 1655-1674.

Graves, A., Mohamed, A. R., & Hinton, G. (2013, May). Speech recognition with deep recurrent neural networks. In *2013 IEEE international conference on acoustics, speech and signal processing* (pp. 6645-6649). Ieee.

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of hydrology*, *377*(1-2), 80-91.

Han, H., & Morrison, R. R. (2022). Data-driven approaches for runoff prediction using distributed data. *Stochastic Environmental Research and Risk Assessment*, *36*(8), 2153-2171.

Hannah, D. M., Demuth, S., Lanen, H. V., Looser, U., Prudhomme, C., Rees, G., ... & Tallaksen, L. M. (2011). Large-scale river flow archives: importance, current status and future needs.

Hao, R., & Bai, Z. (2023). Comparative study for daily streamflow simulation with different machine learning methods. Water, 15(6), 1179.

Hassler, B., & Lauer, A. (2021). Comparison of reanalysis and observational precipitation datasets including ERA5 and WFDE5. *Atmosphere*, *12*(11), 1462.

Haznedar, B., Kilinc, H. C., Ozkan, F., & Yurtsever, A. (2023). Streamflow forecasting using a hybrid LSTM-PSO approach: the case of Seyhan Basin. *Natural Hazards*, *117*(1), 681-701.

Hernández-Henríquez, M. A., Mlynowski, T. J., & Déry, S. J. (2010). Reconstructing the natural streamflow of a regulated river: A case study of La Grande Rivière, Québec, Canada. *Canadian Water Resources Journal*, *35*(3), 301-316.

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., ... & Thépaut, J. N. (2020). The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*, *146*(730), 1999-2049.

Hersbach, H., Bell, B., Berrisford, P., Biavati, G., Horányi, A., Muñoz Sabater, J., Nicolas, J., Peubey, C., Radu, R., Rozum, I., Schepers, D., Simmons, A., Soci, C., Dee, D., Thépaut, J-N. (2023): ERA5 hourly data on single levels from 1940 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS), DOI: 10.24381/cds.adbb2d47 (Accessed on 03-19-2024).

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.

Hodson, T. O. (2022). Root mean square error (RMSE) or mean absolute error (MAE): When to use them or not. *Geoscientific Model Development Discussions*, *2022*, 1-10.

Hu, C., Wu, Q., Li, H., Jian, S., Li, N., & Lou, Z. (2018). Deep learning with a long short-term memory networks approach for rainfall-runoff simulation. *Water*, *10*(11), 1543.

Hu, Y., Yan, L., Hang, T., & Feng, J. (2020). Stream-flow forecasting of small rivers based on LSTM. *arXiv preprint arXiv:2001.05681*.

Hua, Y., Zhao, Z., Li, R., Chen, X., Liu, Z., & Zhang, H. (2019). Deep learning with long short-term memory for time series prediction. *IEEE Communications Magazine*, *57*(6), 114-119.

Jehanzaib, M., Ajmal, M., Achite, M., & Kim, T. W. (2022). Comprehensive review: Advancements in rainfall-runoff modelling for flood mitigation. *Climate*, *10*(10), 147.

Jakeman, A. J., Littlewood, I. G., & Whitehead, P. G. (1990). Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. *Journal of hydrology*, *117*(1-4), 275-300.

Keller, A. A., Garner, K., Rao, N., Knipping, E., & Thomas, J. (2023). Hydrological models for climate-based assessments at the watershed scale: A critical review of existing hydrologic and water quality models. *Science of the Total Environment*, *867*, 161209.

Khoi, D. N. (2016). Comparison of the HEC-HMS and SWAT hydrological models in simulating the stream flow. Journal of Science and Technology, 53(5A), 189-195.

Kim, C., & Kim, C. S. (2021). Comparison of the performance of a hydrologic model and a deep learning technique for rainfall-runoff analysis. *Tropical Cyclone Research and Review*, *10*(4), 215-222.

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

Kisi, O., & Cimen, M. (2011). A wavelet-support vector machine conjunction model for monthly streamflow forecasting. Journal of Hydrology, 399(1-2), 132-140.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall–runoff modelling using long short-term memory (LSTM) networks. *Hydrology and Earth System Sciences*, *22*(11), 6005-6022.

Knoben, W. J., Freer, J. E., & Woods, R. A. (2019). Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences*, *23*(10), 4323-4331.

Le, X. H., Ho, H. V., Lee, G., & Jung, S. (2019). Application of long short-term memory (LSTM) neural network for flood forecasting. *Water*, *11*(7), 1387.

Li, J., Yuan, X., & Ji, P. (2023). Long-lead daily streamflow forecasting using Long Short-Term Memory model with different predictors. *Journal of Hydrology: Regional Studies*, *48*, 101471.

Li, X., Xu, W., Ren, M., Jiang, Y., & Fu, G. (2022). Hybrid CNN-LSTM models for river flow prediction. *Water Supply*, *22*(5), 4902-4919.

Liang X, Lettenmaier DP, Wood EF, et al. (1994). A simple hydrologically based model of land surface water and energy fluxes for general circulation models. J Geophys Res Atmos 99, 14415–28.

Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., ... & Wood, E. F. (2019). Global reconstruction of naturalized river flows at 2.94 million reaches. *Water resources research*, *55*(8), 6499-6516.

Liu, Y., Zhang, T., Kang, A., Li, J., & Lei, X. (2021). Research on runoff simulations using deep-learning methods. *Sustainability*, *13*(3), 1336.

Lohmann, D. A. G., Nolte-Holube, R. A. L. P. H., & Raschke, E. (1996). A large-scale horizontal routing model to be coupled to land surface parametrization schemes. *Tellus A*, *48*(5), 708-721.

Machiwal, D., & Jha, M. K. (2006). Time series analysis of hydrologic data for water resources planning and management: a review. *Journal of Hydrology and Hydromechanics*, *54*(3), 237-257.

Mao, G., Wang, M., Liu, J., Wang, Z., Wang, K., Meng, Y., ... & Li, Y. (2021). Comprehensive comparison of artificial neural networks and long short-term memory networks for rainfall-runoff simulation. *Physics and Chemistry of the Earth, Parts a/b/c*, *123*, 103026.

Miao, C., Gou, J., Fu, B., Tang, Q., Duan, Q., Chen, Z., ... & Wu, J. (2022). High-quality reconstruction of China's natural streamflow. *Science Bulletin*, *67*(5), 547-556.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., & Veith, T. L. (2007). Model evaluation guidelines for systematic quantification of accuracy in watershed simulations. *Transactions of the ASABE*, *50*(3), 885-900.

Mouelhi, S., Michel, C., Perrin, C. and Andréassian, V. (2006). Stepwise development of a two-parameter monthly water balance model. Journal of Hydrology, 318(1–4), 200–214.

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., ... & Duan, Q. (2015). Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrology and Earth System Sciences*, *19*(1), 209-223.

O'Connor, P., Murphy, C., Matthews, T., & Wilby, R. L. (2021). Reconstructed monthly river flows for Irish catchments 1766–2016. *Geoscience Data Journal*, *8*(1), 34-54.

Olah, C. (2015). Understanding lstm networks.

Perrin, C., Michel, C., & Andréassian, V. (2003). Improvement of a parsimonious model for streamflow simulation. *Journal of hydrology*, *279*(1-4), 275-289.

Peyras, L., Carvajal, C., Felix, H., Bacconnet, C., Royet, P., Becue, J. P., & Boissier, D. (2012). Probability-based assessment of dam safety using combined risk analysis and reliability methods–application to hazards studies. *European Journal of Environmental and Civil Engineering*, *16*(7), 795-817.

Piniewski, M. (2016). Natural streamflow simulation for two largest river basins in Poland: a baseline for identification of flow alterations. *Proceedings of the International Association of Hydrological Sciences*, *373*, 101-107.

Ravindranath, A., Devineni, N., Lall, U., Cook, E. R., Pederson, G., Martin, J., & Woodhouse, C. (2019). Streamflow reconstruction in the upper Missouri River basin using a novel Bayesian network model. *Water Resources Research*, *55*(9), 7694-7716.

Samaniego, L., Kumar, R., & Attinger, S. (2010). Multiscale parameter regionalization of a grid-based hydrologic model at the mesoscale. *Water Resources Research*, *46*(5).

Santhi, C, J. G. Arnold, J. R. Williams, W. A. Dugas, R. Srinivasan, and L. M. Hauck. (2001). Validation of the SWAT model on a large river basin with point and nonpoint sources. J. American Water Resources Assoc. 37(5): 1169-1188.

Sarker, I. H. (2021). Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions. *SN Computer Science*, *2*(6), 420.

Sengupta, S., Basak, S., Saikia, P., Paul, S., Tsalavoutis, V., Atiah, F., ... & Peters, A. (2020). A review of deep learning with special emphasis on architectures, applications and recent trends. *Knowledge-Based Systems*, *194*, 105596.

Shen, C. (2018). A transdisciplinary review of deep learning research and its relevance for water resources scientists. Water Resources Research, 54(11), 8558-8593.

Shiklomanov, A. I., Lammers, R. B., & Vörösmarty, C. J. (2002). Widespread decline in hydrological monitoring threatens pan-Arctic research. *Eos, Transactions American Geophysical Union*, *83*(2), 13-17.

Sit, M., Demiray, B. Z., Xiang, Z., Ewing, G. J., Sermet, Y., & Demir, I. (2020). A comprehensive review of deep learning applications in hydrology and water resources. *Water Science and Technology*, *82*(12), 2635-2670.

Sivapalan, M. (2003). Prediction in ungauged basins: a grand challenge for theoretical hydrology.

Song, T., Ding, W., Wu, J., Liu, H., Zhou, H., & Chu, J. (2019). Flash flood forecasting based on long short-term memory networks. *Water*, *12*(1), 109.

SYGM. (2020a). *Fırat-Dicle Havzası Taşkın Yönetim Planı Yönetici Özeti.*

https://www.tarimorman.gov.tr/SYGM/Belgeler/Ta%C5%9Fk%C4%B1n%20Y%C3%B6netim%20Planlar%C4%B1/F%C4%B1rat%20Alt%20Havzas%C4%B1%20Ta%C5%9Fk%C4%B1n%20Y%C3%B6netim%20Plan%C4%B1%20Y%C3%B6netici%20%C3%96zeti.pdf

SYGM. (2020b). *Seyhan Havzası Taşkın Yönetim Planı Yönetici Özeti.*

https://www.tarimorman.gov.tr/SYGM/Belgeler/Ta%C5%9Fk%C4%B1n%20Y%C3%B6netim%20Planlar%C4%B1%2026.12.2022/Seyhan_Yonetici_Ozeti_v3.pdf

SYGM. (n.d.). *Seyhan River Basin*.

https://www.tarimorman.gov.tr/SYGM/Belgeler/havza%20tan%C4%B1t%C4%B1m%2023.03.2023/ingilizce/SEYHAN%20RIVER%20BASIN.pdf. Accessed August 19, 2024.

Tarek, M., Brissette, F. P., & Arsenault, R. (2020). Evaluation of the ERA5 reanalysis as a potential reference dataset for hydrological modelling over North America. *Hydrology and Earth System Sciences*, *24*(5), 2527-2544.

Van, S. P., Le, H. M., Thanh, D. V., Dang, T. D., Loc, H. H., & Anh, D. T. (2020). Deep learning convolutional neural network in rainfall–runoff modelling. *Journal of Hydroinformatics*, *22*(3), 541-561.

Wang, S., Sun, M., Wang, G., Yao, X., Wang, M., Li, J., ... & Yang, Y. (2023). Simulation and Reconstruction of Runoff in the High-Cold Mountains Area Based on Multiple Machine Learning Models. *Water*, *15*(18), 3222.

Won, J., Seo, J., Lee, J., Choi, J., Park, Y., Lee, O., & Kim, S. (2023). Streamflow Predictions in Ungauged Basins Using Recurrent Neural Network and Decision Tree-Based Algorithm: Application to the Southern Region of the Korean Peninsula. *Water*, *15*(13), 2485.

Wood, E. F., Roundy, J. K., Troy, T. J., Van Beek, L. P. H., Bierkens, M. F., Blyth, E., ... & Whitehead, P. (2011). Hyperresolution global land surface modeling: Meeting a grand challenge for monitoring Earth's terrestrial water. *Water Resources Research*, *47*(5).

Wu, H., Liu, Q., & Liu, X. (2019). A review on deep learning approaches to image classification and object segmentation. *Computers, Materials & Continua*, *60*(2).

Xu, J., Ma, Z., Yan, S., & Peng, J. (2022). Do ERA5 and ERA5-land precipitation estimates outperform satellite-based precipitation products? A comprehensive comparison between state-of-the-art model-based and satellite-based precipitation products over mainland China. *Journal of Hydrology*, *605*, 127353.

Xu, W., Jiang, Y., Zhang, X., Li, Y., Zhang, R., & Fu, G. (2020). Using long short-term memory networks for river flow prediction. *Hydrology Research, 51*(6), 1358-1376.

Yin, H., Wang, F., Zhang, X., Zhang, Y., Chen, J., Xia, R., & Jin, J. (2022). Rainfall-runoff modeling using long short-term memory based step-sequence framework. *Journal of Hydrology*, *610*, 127901.

Yokoo, K., Ishida, K., Ercan, A., Tu, T., Nagasato, T., Kiyama, M., & Amagasaki, M. (2022). Capabilities of deep learning models on learning physical relationships: Case of rainfall-runoff modeling with LSTM. *Science of The Total Environment*, *802*, 149876.

Zhang, A., Lipton, Z. C., Li, M., & Smola, A. J. (2023). *Dive into deep learning*. Cambridge University Press.

Zhang, D., Lin, J., Peng, Q., Wang, D., Yang, T., Sorooshian, S., ... & Zhuang, J. (2018). Modeling and simulating of reservoir operation using the artificial neural network, support vector regression, deep learning algorithm. *Journal of Hydrology*, *565*, 720-736.

Zheng, F., Maier, H. R., Wu, W., Dandy, G. C., Gupta, H. V., & Zhang, T. (2018). On lack of robustness in hydrological model development due to absence of guidelines for selecting calibration and evaluation data: Demonstration for data-driven models. *Water Resources Research*, *54*(2), 1013-1030.

APPENDICES

A. Input Variable Description in ERA5

Table A.1 Modified version of input variable descriptions in ERA5 (Hersbach et al., 2023).

Input variables	Units	Description
10m u-component of wind	m s-1	This parameter is the eastward component of the 10m wind. When comparing this parameter to observations, it should be considered that the ECMWF Integrated Forecasting System (IFS) only provides average representation of the local terrain, vegetation, and buildings; wind measurements differ in small space and time scales.
10m v-component of wind	m s-1	This parameter is the northward component of the 10m wind. When comparing this parameter to observations, it should be considered that the ECMWF Integrated Forecasting System (IFS) only provides average representation of the local terrain, vegetation, and buildings; wind measurements differ in small space and time scales.
2m temperature	K	The air temperature at a height of 2m above the surface land, sea, or inland waters is represented by this parameter.
Potential evaporation	m	It generally describes the quantity of evaporation in the current atmospheric circumstances. Potential evaporation in the ECMWF Integrated Forecasting System (IFS) relies on surface energy balance calculations.
Surface net solar radiation	J m-2	This parameter is the total quantity of solar radiation, often referred to as shortwave radiation, that reaches a horizontal plane at the surface of the Earth minus the amount that is reflected by the planet's surface.
Surface runoff	m	Water from rainfall, melting snow, or deep in the soil, remains stored in the soil. On the other hand, the water drains away, either over the surface (surface runoff), or under the ground (sub-surface runoff).
Total precipitation	m	This parameter is the total amount of liquid and frozen water, including rain and snow, that falls to the Earth's surface.