ASSESSING THE MID-TERM WEATHER FORECASTS IN HYDROLOGICAL MODELLING

Master's Thesis

Abdishakur Dahir ABDULLAHI

Eskişehir 2025

ASSESSING THE MID-TERM WEATHER FORECASTS IN HYDROLOGICAL MODELLING

Abdishakur Dahir ABDULLAHI

MASTER'S THESIS

Department of Civil Engineering

Programme in Hydraulic Engineering

Supervisor: Assoc. Prof. Dr. Ali Arda ŞORMAN

Eskişehir

Eskişehir Technical University

Institute of Graduate Programs

July 2025

FINAL APPROVAL FOR THESIS

This thesis titled ASSESSING THE MID-TERM WEATHER FORECASTS IN HYDROLOGICAL MODELLING has been prepared and submitted by Abdishakur Dahir ABDULLAHI in partial fulfillment of the requirements in "Eskişehir Technical University Directive on Graduate Education and Examination" for the Degree of Master's in Civil Engineering Department has been examined and approved on 09/07/2025.

Committee Members		**Title, Name and Surname**	**Signature**
Member	:	Assoc. Prof. Dr. Ali Arda ŞORMAN	
Member	:	Prof. Dr. Nermin ŞARLAK	
Member	:	Assoc. Prof. Dr. Gökçen UYSAL	

Prof. Dr. Harun BÖCÜK
Director of the Institute of Graduate Programs

09/07/2025

SUPERVISOR APPROVAL

Master's student Abdishakur Dahir ABDULLAHI, whom I supervise, has completed this thesis titled ASSESSING THE MID-TERM WEATHER FORECASTS IN HYDROLOGICAL MODELLING. According to my inspections, the work is scientifically and ethically appropriate for the student to the thesis defense exam.

Supervisor

Assoc. Prof. Dr. Ali Arda ŞORMAN

ABSTRACT

ASSESSING THE MID-TERM WEATHER FORECASTS IN HYDROLOGICAL MODELLING

Abdishakur Dahir ABDULLAHI

Department of Civil Engineering

Programme in Hydraulic Engineering

Eskişehir Technical University, Institute of Graduate Programs, July 2025

Supervisor: Assoc. Prof. Dr. Ali Arda ŞORMAN

Accurate hydrological forecasting is critical for water resource management, flood prediction, hydropower generation, and risk assessment and mitigation but is fundamentally challenged by uncertainties arising from meteorological forcing, model structure, and initial conditions. This study provides an evaluation of medium range weather forecasting in hydrological modelling from European Centre for Medium-Range Weather Forecasts (ECMWF) to reduce these uncertainties for two mountainous basins. Using twenty hydrological models within the HydrOlOgical Prediction LAboratory (HOOPLA) framework, this research systematically compares four forecasting configurations: Open-Loop (OL) and Data Assimilated (DA) for both deterministic and ensemble forecasts. The Ensemble Kalman Filter (EnKF) was used for data assimilation, and multi-model (MM) combinations were generated using a Simple Averaging Method (SAM). Performance, assessed by the Kling-Gupta Efficiency (KGE), revealed a clear hierarchy. Ensemble forecasts consistently outperformed deterministic ones, and DA significantly enhanced forecast skill over OL simulations by correcting initial model states. The study demonstrates that an integrated approach combining all techniques yields the most reliable results. The DA multi-model ensemble proved superior, effectively mitigating multiple uncertainty sources and maintaining high accuracy across the 10-day forecast horizon. These findings offer a robust, evidence-based framework for improving operational hydrological forecasting systems.

Keywords: Medium range, Uncertainty, Ensemble forecasting, Data assimilation, Multi-model.

ÖZET

HİDROLOJİK MODELLEMEDE ORTA VADELİ HAVA TAHMİNLERİNİN DEĞERLENDİRİLMESİ

Abdishakur Dahir ABDULLAHI

İnşaat Mühendisliği Anabilim Dalı

Hidrolik Mühendisliği Bilim Dalı

Eskişehir Teknik Üniversitesi, Lisansüstü Eğitim Enstitüsü, Temmuz 2025

Danışman: Doç. Dr. Ali Arda ŞORMAN

Doğru hidrolojik tahminleme, su kaynaklarının yönetimi, taşkın erken uyarı sistemleri ve hidroelektrik enerji üretimi açısından kritik bir öneme sahiptir. Ancak, meteorolojik girdileri, model yapısı ve başlangıç koşullarının belirsizlikleri nedeniyle bu tahminlerin güvenilirliği çoğu zaman sınırlı kalmaktadır. Bu çalışmada, dağlık iki havzada, belirsizlikleri azaltmayı hedefleyen bir yaklaşım sunulmuştur. Avrupa Orta Vadeli Hava Tahminleri Merkezi (ECMWF) tarafından sağlanan orta vadeli meteorolojik tahminlerin, hidrolojik modelleme süreçlerindeki etkisi değerlendirilmiştir. HydrOlOgic Prediction LAboratory (HOOPLA) arayüzü kullanılarak yirmi farklı hidrolojik model ile dört farklı tahmin yapılandırması sistematik biçimde karşılaştırılmıştır: Open-Loop (OL) ve Veri Asimilasyonu (DA) yaklaşımları altında oluşturulan deterministik ve ensemble tahminler. Veri asimilasyonu için Ensemble Kalman Filtresi (EnKF) kullanılırken, çoklu model (multi-model) kombinasyonlar Basit Ortalama Yöntemi (Simple Average Method, SAM) ile oluşturulmuştur. Kling-Gupta Efficiency (KGE) ile yapılan performans değerlendirmesi, yöntemler arasında belirgin bir başarı hiyerarşisi ortaya koymuştur. Ensemble tahminler, deterministik tahminlere kıyasla tutarlı bir şekilde daha yüksek performans sergilemiş; veri asimilasyonu ise başlangıç koşullarını iyileştirerek OL tahminlerine göre önemli ölçüde daha başarılı sonuçlar vermiştir. Çalışma, veri asimilasyonu ile çok modelli ensemble yaklaşımının bir arada kullanıldığı bütünleşik bir yöntemin, 10 günlük tahmin ufku boyunca en yüksek doğruluğu sağladığını ve çoklu belirsizlik kaynaklarını etkili biçimde azaltabildiğini göstermektedir. Bu bulgular, operasyonel hidrolojik tahmin sistemlerinin doğruluğunu ve güvenilirliğini artırmak için sağlam bir çerçeve sunmaktadır.

Anahtar Sözcükler: Orta vadeli tahminleri, Ensemble tahminleri, Belirsizlik, Veri asimilasyonu, Çoklu model.

ACKNOWLEDGEMENTS

Completing this thesis has been a long journey of growth, reflection, and perseverance. Along the way, I have been fortunate to receive the support, wisdom, and encouragement of many remarkable individuals.

I would like to express my sincere gratitude to my supervisor, Assoc. Prof. Dr. Ali Arda ŞORMAN, for his invaluable guidance, insightful feedback, and unwavering support throughout this research.

My thanks also go to the academic staff and colleagues at Eskişehir Technical University, whose knowledge, encouragement, and advice have contributed significantly to my academic development.

I extend my deepest gratitude to Turkiye Bursları for the invaluable opportunity to pursue my Master studies in Turkiye. Their comprehensive scholarship has been pivotal in enabling my research and studies in Turkiye, and I am truly thankful for this significant investment in my academic future.

Last but not least, I extend my heartfelt thanks to my family—my mother Maryan Yusuf, my wife Khadra Ahmed, my son Mohamed Abdishakur, and all my brothers and sisters—for their patience, love, and belief in me, especially during the most challenging times. This achievement belongs to you as much as it does to me.

Abdishakur Dahir ABDULLAHI

09/07/2025

STATEMENT OF COMPLIANCE WITH ETHICAL PRINCIPLES AND RULES

I hereby truthfully declare that this thesis is an original work prepared by me; that I have behaved in accordance with the scientific ethical principles and rules throughout the stages of preparation, data collection, analysis and presentation of my work; that I have cited the sources of all the data and information that could be obtained within the scope of this study, and included these sources in the references section; and that this study has been scanned for plagiarism with “scientific plagiarism detection program” used by Eskişehir Technical University, and that “it does not have any plagiarism” whatsoever. I also declare that, if a case contrary to my declaration is detected in my work at any time, I hereby express my consent to all the ethical and legal consequences that are involved.

Abdishakur Dahir ABDULLAHI

CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

GLOSSARY OF SYMBOLS AND ABBREVIATIONS

ANN	:	Artificial Neural Networks
BMA	:	Bayesian Model Averaging
Ctg	:	A Thermal Inertia Factor for the Snowpack
DA	:	Data Assimilation
DDS	:	Dynamically Dimensioned Search
DSİ	:	State Hydraulic Works
ECMWF	:	European Centre for Medium-Range Weather Forecasts
ENIAC	:	Electronic Numerical Integrator and Computer
EnKf	:	Ensemble Kalman Filter
EPS	:	Ensemble Prediction Systems
GR4H	:	Génie Rural À 4 Paramètres Horaire
HBV	:	Hydrologiska Byråns Vattenbalansavdelning
HOOPLA	:	Hydrological Prediction Laboratory
IFS	:	Integrated Forecasting System
IMS	:	Moderate Resolution Imaging Spectroradiometer
Kf	:	A Degree-Day Melt Factor
KGE	:	Kling-Gupta Efficiency
MARRMoT	:	Modular Assessment of Rainfall-Runoff Models Toolbox
MEPS	:	Meteorological Ensemble Prediction System
MGM	:	Turkish Meteorological Services
MM	:	Multi-Model
MMCMs	:	Multi-Model Combination Methods
MMSE	:	Modified Multi-Model SuperEnsemble
MSE	:	Multi-Model SuperEnsemble
NCEP	:	National Centers for Environmental Prediction
NNM	:	Neutral Network Methods
NSE	:	Nash-Sutcliffe Efficiency
NSEsqrt	:	Square Root of Nash–Sutcliffe Efficiency
NWP	:	Numerical Weather Prediction
OL	:	Open-Loop
PET	:	Potential Evapotranspiration
PF	:	Particle Filter

r^2	:	Coefficient Of Determination
RMSE	:	Root Mean Square Error
SAM	:	Simple Average Method
SAR	:	Snow Accounting Routine
SCE	:	Shuffled Complex Evolution
SWE	:	Snow Water Equivalent
TIGGE	:	International Grand Global Ensemble
WAM	:	Weighted Average Method
WMO	:	World Meteorological Organization
WRF	:	Weather Research and Forecasting

1. INTRODUCTION

1.1. Background

Hydrological forecasting is a vital tool for effective management of water resources, flood prediction, hydropower generation, and risk assessment and mitigation (Dion et al., 2021). Accurate and reliable forecasts are essential for informed decision making in these areas, impacting everything from daily operations to long-term planning. However, hydrological modelling and forecasting are inherently uncertain. This uncertainty stems from various sources including uncertainties in input data, meteorological forcing, model parameters and structure, initial conditions and forecasting (A. Gupta and Govindaraju, 2023; Liu and Gupta, 2007; Montanari, 2007; Panchanathan et al., 2024; Thiboult et al., 2016).

These uncertainties pose significant challenges, potentially leading to flawed decision making in water allocation, risk assessment, emergency planning, and climate change adaptation strategies. Inaccurate forecasts can result in infrastructure deficiencies, ecosystem disruptions, and substantial economic losses across various sectors (Panchanathan et al., 2024). Therefore, understanding, quantifying, and mitigating uncertainty is paramount in hydrological modeling to ensure the reliability and practical value of forecasts.

Traditionally, discharge forecasts have often relied on deterministic methods, which provide a single, most likely prediction without accounting for inherent uncertainties. However, probabilistic forecasts offer a range of possible outcomes with associated probabilities, explicitly acknowledging and quantifying uncertainty. This capability has led to a shift in focus over the last decade toward probabilistic approaches as they are better equipped to capture the full spectrum of potential outcomes (Krzysztofowicz, 2001; Velázquez et al., 2011).

Several strategies have been employed to address these uncertainties and improve forecast accuracy. Ensemble Prediction Systems (EPS) play a crucial role in reducing uncertainties associated with meteorological input data. EPS generates multiple forecasts using slightly perturbed initial conditions in Numerical Weather Prediction (NWP) models, thus providing a probabilistic representation of potential future weather scenarios. This is particularly valuable in hydrology, as precipitation and temperature are critical drivers of discharge (Anctil and Ramos, 2019; Cloke and Pappenberger, 2009; Leutbecher and Palmer, 2008; Shu et al., 2023). Another approach to minimize

uncertainties arising from the model structure is the use of multi-model (MM) hydrological ensembles. Instead of relying on a single model, this approach combines predictions from multiple hydrological models, each with its own structure and parametrization, to provide a more robust and reliable forecast (Thiboult et al., 2016; Velázquez et al., 2011; Wang et al., 2021). Furthermore, data assimilation (DA) techniques are employed to address uncertainties in the initial conditions of the input data. By incorporating observational data, such as streamflow measurements, precipitation, and temperature, into the model, these techniques refine the initial state of the model and reduce the uncertainty associated with the starting conditions.

Despite these advancements, the application of these uncertainty reduction techniques, particularly the combination of EPS, MM ensembles, and DA, has been limited in the Turkish basins. This study aims to address this challenge by developing an integrated framework that combines probabilistic forecasting, MM, and DA techniques to improve the reliability and skill of discharge forecasts in two mountainous basins. By improving the accuracy and reliability of forecasts in these basins, this study seeks to contribute to more effective water resource management and hydropower generation in Turkiye. Specifically, this research will investigate the following research questions: (1) How does the use of MM ensemble approach, compare to a single-model approach, impact the accuracy and reliability of discharge forecasts? (2) What is the added value of incorporating DA techniques to improve the representation of the initial conditions and reduce forecast uncertainty? (3) How does the combined use of EPS, MM ensembles, and DA contribute to a more comprehensive understanding of the forecast uncertainty in the study basins?

1.2. Problem Statement

Accurate weather forecasting is crucial for effective hydrological modeling, particularly in regions or basins prone to extreme weather events. The reliability and skill of these forecasts are significantly affected by uncertainties arising from various sources, including model structure, initial conditions, and the forecasts themselves. Addressing these uncertainties requires the integration of EPS input data, MM approach for model structures, and DA techniques for initial conditions. The use of MM, DA, and EPS have shown good performance in recent studies, but are not commonly studied in Turkish Basins.

Hydrological forecasting in Turkish basins face significant challenges owing to their topography, climate variability, and limited application of advanced uncertainty reduction techniques. Although EPS, MM ensembles, and DA have shown promise in improving forecast accuracy elsewhere, their combined application has not been thoroughly investigated in the Turkish context. Specifically, studies that systematically evaluate the performance of multiple hydrological models driven by EPS data and further refined through DA in these basins are lacking. This study addresses this critical need by developing and evaluating an integrated framework that leverages the strengths of EPS, MM ensembles, and DA to improve the accuracy and reliability of discharge forecasts.

1.3. Objectives

The primary objective of this study was to evaluate the efficacy of mid-term weather forecasts in hydrological modelling using EPS, multi-hydrological models, and DA techniques. The aim of integrating these methods is to enhance the accuracy and reliability of hydrological predictions and to reduce uncertainty.

The specific aims of this study were as follows:

1. To evaluate the performance of medium-range European Center for Medium-Range Weather Forecasts (ECMWF) EPS forecasts in 20 lumped hydrological models in the two mountainous basins.
2. To assess the impact of ensemble forecast on the forecast performance comparing with deterministic forecasts of the basins.
3. To implement and evaluate the impact of the Ensemble Kalman Filter (EnKF) DA technique on the accuracy of discharge forecasts by comparing open-loop and DA forecasts.
4. To quantify the reduction in uncertainty achieved using a MM ensemble compared with a single-model approach for discharge forecasts in the study basins.

1.4. Thesis Organization

This thesis is structured into five chapters to present the assessing of mid-term weather forecasting in hydrological modelling on reducing uncertainty. **Chapter 1** introduces the study, providing the essential background and context for the research

problem. **Chapter 2** presents a literature review on uncertainty in hydrological forecasting. This chapter examines studies and practical applications of key uncertainty reduction techniques, including EPS, DA, and MM approaches, establishing s foundation for this work. **Chapter 3** details the methodology and materials employed. It describes Çukurkışla and Kayabaşı study areas, the meteorological and hydrological data used, and ECMWF mid-term deterministic and ensemble data. The chapter explains the use of the HOOPLA toolbox for calibrating twenty hydrological models using the Shuffled Complex Evolution algorithm, and it also shows open-loop and DA forecast and MM approach. **Chapter 4** presents the calibration, validation and forecast results. In addition, it clearly details the open-loop and data assimilated forecast in deterministic, ensemble, MM. **Chapter 5** concludes this study's findings. It also addresses the limitations of the research and provides recommendations for future work in this field.

2. LITERATURE REVIEW

2.1. Sources of Uncertainty in Hydrological Modelling

Hydrological modelling and forecasting uncertainty refer to the lack of precision or confidence in the prediction of hydrological processes such as discharge, snow, and soil moisture. Uncertainty rises from multiple sources, such as the model's structure, parameter estimation, input data quality, initial condition, and the intrinsic variability of natural systems (Montanari, 2007; Panchanathan et al., 2024). Uncertainties in hydrological modelling and forecasting can be summarized as measurement, input, structural, parameter, and forecast uncertainties (A. Gupta and Govindaraju, 2023; Panchanathan et al., 2024). Measurement uncertainty relates to errors in precipitation, temperature, evapotranspiration, or streamflow from the measurement process (McMillan et al., 2018). Input uncertainty relates to errors in data manipulation, such as interpolation and scaling. Structural uncertainty relates to errors that emerge from the formation and processing of hydrological models. The uncertainty rates differ for each hydrological model. Parameter uncertainties arise from the inadequacy of the model's parameterization to fully represent the hydrological process, often leading to equifinality (Panchanathan et al., 2024).

Accurate and timely flood forecasting is crucial for mitigating flood risks and facilitating effective disaster management (Panchanathan et al., 2024; Verkade and Werner, 2011). Traditional flood forecasting systems often rely on a single NWP model and hydrological model. However, both NWPs and hydrological models are subject to various sources of uncertainty (Ajami et al., 2006; Cloke and Pappenberger, 2009). Such uncertainties can affect the accuracy and reliability of flood forecasts. To address these limitations, researchers have explored ensemble forecasting approaches, MM approaches, and DA.

Thiboult et al. (2016) investigated three sources of uncertainty in hydrological forecasting: hydrological model error, initial conditions, and meteorological forcing uncertainties. Their study evaluated the effectiveness of using tools such as the EnKF, MM approaches, and the Meteorological Ensemble Prediction System (MEPS) to improve forecast accuracy and reliability. They observed that it enhances hydrological forecast reliability and accuracy by addressing various aspects of uncertainty.

Expanding on the theme of enhancing forecast reliability, Llauca et al. (2023) explored how DA techniques to evaluate the performance and usage of discharge less

than daily time steps using EnKF and Particle Filter (PF) algorithms. They used the GR4H model with 100 ensemble members. They observed that the use of DA provides more accurate and reliable streamflow predictions.

2.2. Ensemble Prediction Systems (EPS)

Ensemble Prediction Systems (EPS) have emerged as a powerful tool for quantifying uncertainty in weather forecasts and improving flood predictions (Cloke and Pappenberger, 2009; Wu et al., 2020). The EPS generates multiple forecasts by perturbing the initial conditions of an NWP model, providing a range of possible future weather scenarios.

NWP employs mathematical methods to forecast weather by numerically solving the fundamental equations governing atmospheric dynamics. At its core, NWP is built upon physical laws, such as the conservation of mass, momentum, and energy, to simulate atmospheric processes. Since its first operational implementation in the mid-20th century, NWP has become a cornerstone of modern meteorology (Bauer et al., 2015; Pu and Kalnay, 2018). The conceptual origins of NWP can be traced back to Vilhelm Bjerknes in 1904, who postulated that weather forecasting could be grounded in physics, provided that the initial conditions and physical laws were known. However, these ideas could only be practically implemented after the advent of digital computers. In 1950, the first successful NWP forecast was produced using the Electronic Numerical Integrator and Computer **(**ENIAC) by Jule Charney and his colleagues. Since then, NWP models have progressed from simplistic two-dimensional barotropic models to complex three-dimensional models that can simulate a wide range of atmospheric processes (Lynch, 2008). The accuracy of NWP has improved notably over the past decades for different hydrological variables, such as precipitation (Bélair et al., 2009), temperature (Mathiesen and Kleissl, 2011), and soil moisture (Dillon et al., 2016).

Velázquez et al. (2011) demonstrated that combining multiple hydrological model structures with meteorological ensembles yields more accurate and reliable predictions. Their findings indicate this approach outperforms predictions from a single hydrological model with meteorological ensemble predictions or multiple hydrological models with a deterministic meteorological forecast. These findings are supported by Brochero et al. (2011) who demonstrated that ensemble predictions derived from a combination of various hydrological model structures and meteorological ensembles outperform those

generated by a single model or deterministic forecasts. The integration of diverse models allows for a wide representation of hydrological processes.

In addition, the use of probabilistic weather forecasts is another critical aspect of the EPS. Matus et al. (2020) noted that EPS can extend lead times and better quantify predictability, making them particularly appealing for flood forecasting. This is further supported by Gelfan et al. (2015), who discussed the shift from deterministic methodologies to ensemble-based approaches, aimed to improve the reliability and accuracy of forecasts.

ECMWF EPS is a critical tool in meteorological forecasting, particularly in hydrology. One of the primary strengths of the ECMWF EPS is its ability to provide probabilistic forecasts that account for uncertainties in weather predictions. Palmer (2019) emphasized that the ECMWF EPS has evolved over more than 25 years, significantly enhancing its probabilistic forecasting capabilities from single deterministic forecasting through improved model physics and parameterizations. This evolution has been crucial for hydrological applications, where understanding the range of possible outcomes is essential for effective water resource management and flood forecasting. (Roulin and Vannitsem (2015) investigated the impact of errors in forcing, initial conditions, and model structure on hydrological forecasts and demonstrated that integrating ECMWF precipitation forecasts into hydrological models significantly improved forecast reliability. Moreover, the ECMWF has made progress in integrating high-resolution data into forecasting systems. The introduction of the Integrated Forecasting System (IFS) has allowed for improved vertical and horizontal resolutions, which are crucial for accurately capturing the dynamics of weather systems. The operational ensemble forecasts now utilize a horizontal grid spacing of approximately 18 km and 137 vertical levels with 50 perturbed members of 15 days range lead time, which is a significant improvement over the previous configurations (Lang et al., 2021).

2.3. Multi-model Approach in Hydrological Modelling

Multi-model (MM) approach involves the integration of more than one model, either in parallel or sequentially, to improve the performance of the streamflow. Although there are many hydrological models, no model outperforms the others (Shamseldin et al., 1997) and each has its own structural and parameter uncertainty (Panchanathan et al., 2024). By combining the outputs of hydrological models with different structural, it is

possible to increase their strengths and mitigate their individual weaknesses, which is leading to an improved overall forecast accuracy and reduced uncertainty (Thiboult and Anctil, 2015).

Different models employ varied structures, including lumped and distributed representations, and diverging parameterizations for critical hydrological processes like evapotranspiration and infiltration. This structural diversity allows MM approaches to encompass the complexity and uncertainty inherent in hydrological systems, thereby mitigating overconfidence that may arise from reliance on outcomes from a single model. According to Yu et al. (2018), MM ensembles help mitigate the impact of outliers or underperforming models by combining diverse approaches, which leads to more stable and reliable hydrological forecasts.

MM combination methods (MMCMs) are utilized to improve the accuracy of hydrological outputs in simulation or forecasting (Todorović et al., 2024). Different methods such as SAM, WAM, MM SuperEnsemble (MSE), Modified MM SuperEnsemble (MMSE), Bayesian Model Averaging (BMA), Artificial Neural Network (ANN), Best Model selection, and Cascading are utilized for a multiple hydrological models' combination (Ajami et al., 2006; Andraos, 2024; Chevuturi et al., 2023; Todorović et al., 2024).

Shamseldin et al. (1997) first investigated MM approach on hydrological modelling using Simple Average Method (SAM), Weighted Average Method (WAM), and Neutral Network Methods (NNM) of MM combinations over 11 catchments and they have seen that MM outperformed any of the single-model simulations. Perrin et al. (2001) conducted a comparative assessment of 19 lumped hydrological models across 429 catchments in various climates, including France, the U.S., Australia, Brazil, and the Ivory Coast. Their study examined the relationship between model complexity, defined by the number of optimized parameters, and predictive performance. They found that the complex models performed well in calibration but lacked stability in verification, indicating over parameterization. They also found that combining different model structures could lead to better overall performance compared to using individual models.

Ajami et al. (2006) observed that MM simulations obtained from uncalibrated single-model simulations generally outperformed even the best-calibrated individual model simulations, and more sophisticated MM combination techniques incorporating

bias correction worked better than simple MM averages or combinations without bias correction.

Besides that, Velázquez et al. (2011) investigated if the MM approach reduces the uncertainty in the hydrological ensemble forecasting. They integrated 16 lumped hydrological models with deterministic and probabilistic forecasting meteorological input data. They observed that the combined ensemble prediction from different hydrological models and meteorological ensembles outperformed single hydrological models.

Thiboult and Anctil (2015) investigated on comparing the performance of a MM approach, containing 20 lumped hydrological models, on a deterministic and probabilistic forecasting in Quebec. They showed that the MM ensembles with probabilistic forecasting outperforms the single models with deterministic and probabilistic forecasting.

Dion et al. (2021) investigated the use of a MM framework for hydrological ensemble forecasting to improve streamflow predictions. Their study incorporated eight lumped hydrological models, with initial states updated through EnKF DA, and forecasts driven by ECMWF NWP over five snowmelt-dominated catchments in Quebec, Canada. The study highlighted that a MM strategy provides a better representation of uncertainty than a single model approach and enhances forecast reliability by reducing biases and addressing under-dispersion issues.

Nikhil Teja et al. (2023) examined improving the flood forecasts using multiple numerical weather Predictions ensemble; ECMWF and National Centers for Environmental Prediction with 51 and 31 ensemble members, and multiple hydrological models; GR4J, HBV, SIMHYD, and HEC-HMS. They observed that addressing uncertainties is more crucial than the input data uncertainty and using multiple hydrological models produces more reliable forecasts than using a single model. Supporting this, Chevuturi et al. (2023) found that weighted blending of MM hydrological simulations, especially when combined with bias-correction, improved performance over individual models for evapotranspiration, soil moisture, and streamflow variables.

Todorović et al. (2024) investigated the effect of 10 different MM combination methods using 29 lumped hydrological models from Modular Assessment of Rainfall-Runoff Models Toolbox (MARRMoT) for hydrological modeling. They found that MM combination methods generally improved model efficiency, especially for runoff dynamics and high flows, with the Granger-Ramanathan (Diks and Vrugt, 2010)

performed best. However, MMCMs struggled to reproduce hydrological signature distributions accurately, often underestimating peak flows and overestimating low flows, highlighting a persistent challenge in hydrological modeling. Similarly, (Wan et al., 2021) found that Granger-Ramanathan method outperformed the other methods and the MM combination is less efficient in terms of low-flow simulations. In contrast, Andraos (2024) found that MM approaches, especially those using artificial neural networks, can substantially improve low flow forecasting accuracy. These fusion techniques outperform individual models by leveraging the strengths of each, which is especially beneficial during hydrologically challenging periods like droughts.

Thébault et al. (2024) employed a MM approach within a variable spatial framework to simulate streamflow, using 13 hydrological model structures, three objective functions, and two spatial frameworks (lumped and semi-distributed) across 121 catchments in France. Their findings show that the mixed MM approach outperforms single models and provides higher performance scores. MM approach reduced the spread in performance, indicating lower uncertainty compared to individual model.

2.4. Data Assimilation in Hydrological Modelling

Data assimilation (DA) involves improving the accuracy of hydrological models by combining model predictions with real-time observations such as precipitation, temperature, and streamflow, with different initial states. It provides a framework for integrating various sources of data, such as streamflow measurements, satellite data, and meteorological observations, into the model to reduce uncertainties and improve forecast accuracy. DA in MM approach helps to reduce uncertainties by reducing the errors in the initial states (DeChant and Moradkhani, 2014; Panchanathan et al., 2024). In the context of streamflow forecasting, DA seeks to enhance predictions by integrating errors between forecasting models and discharge measurements through the updating of model states (Llauca et al., 2023). Proper use of DA may help in handling uncertainties from model inputs, initialization and propagation of states, model structures, and even model parameters (Sun et al., 2016). DA reduces uncertainty by incorporating real-time observations, updates model states for better initial conditions, and corrects for input, structural, and parameter -related errors. Together, these improvements lead to more reliable streamflow predictions, particularly in short-term forecasting (Avellaneda et al., 2020).

The application of streamflow DA typically involves three main steps: designing the DA experiment by selecting variables to be perturbed and assimilated, such as rainfall and real-time streamflow measurements; and quantifying model errors, as uncertainties in rainfall, model states, and discharge can significantly impact results. Finally, the chosen DA algorithm is applied to an Open Loop hydrological model, with methods like the EnKF and PF commonly used for probabilistic hydrologic predictions and operational flood forecasting (Avellaneda et al., 2020; Llauca et al., 2023). Many different DA methods are used in hydrological models such as Kalman Filter, EnKF, PF, Four-Dimensional Variational Assimilation (DeChant and Moradkhani, 2014; Llauca et al., 2023; Pu and Kalnay, 2018; Sun et al., 2016).

Thiboult et al. (2016) investigated three probabilistic tools for reducing uncertainty in hydrometeorological forecasting: an MM approach, EnKF, and meteorological ensemble forcing. Each tool addressed a specific aspect of uncertainty: the MM approach quantifies and reduces hydrological model error, the EnKF deciphers initial condition uncertainty, and the meteorological ensemble accounts for forcing uncertainty. They also showed that EnKF performed better for the first lead days to reduce uncertainty.

Dion et al. (2021) found that DA using EnKF improved the ensemble streamflow predictions by reducing bias, enhancing spread, and increasing forecast accuracy compared to open-loop simulations, particularly in the short term (up to 2-3 days). However, it was not fully effective in correcting systematic biases for all seasons and under-dispersion remained an issue, especially during the spring freshet period.

Llauca et al. (2023) evaluated the application of streamflow DA techniques, specifically the EnKF and PF, in a lumped hydrological model for flood forecasting in the data-sparse Vilcanota River Basin in Peru. It's worth noting that both EnKF and PF improved streamflow predictions compared to the Open Loop (OL) simulations, with EnKF showing a slight edge in performance.

Supporting that, Sabzipour et al. (2023) examined the application of the EnKF DA technique to improve short-term streamflow forecasting on the Lac-Saint-Jean catchment in Canada. They conducted a sensitivity analysis to evaluate the effects of EnKF hyperparameters (temperature, precipitation, and inflow uncertainties) and the updating of three state variables (vadose zone, saturated zone, and snowpack) on short-term forecast skill. They found that forecast performance is sensitive to individual hyperparameters, particularly temperature uncertainty, which varies between seasons.

Additionally, the choice of state variables to update also impacts forecast skill depending on the season. The vadose zone state variable was found to be the most important and sensitive.

2.5. Hydrological Forecasting in Turkiye

There are studies that focus on hydrological forecasting on short-term, mid-term and seasonal forecasting. But the studies focusing on quantifying and reducing the uncertainties of hydrological forecasting in Turkiye are limited. Şorman et al. (2009) investigated how incorporating satellite-derived snow-covered area (SCA) data from Moderate Resolution Imaging Spectroradiometer (MODIS) into a hydrological model calibration improves snowmelt runoff simulation in a headwater basin of the Euphrates River in eastern Turkey. They applied the HBV model and integrated NWP data from the Mesoscale Model 5 (MM5) with a 1-day lead time to forecast short-term runoff, which provided promising results for operational flood mitigation and reservoir management.

Yucel et al. (2015) investigated the use of the Weather Research and Forecasting (WRF) hydrometeorological modeling system to simulate major flood events in the western Black Sea Region of Turkiye, using precipitation inputs from the WRF model with and without DA, as well as satellite rainfall estimates. The WRF-Hydro model was calibrated using a stepwise approach in one sub-basin and the calibrated parameters were then transferred to neighboring ungauged basins to assess transferability. The results showed that the WRF-Hydro system, when properly configured and calibrated, was able to reasonably simulate major flood events, with assimilated precipitation input and model calibration providing the best performance in terms of error reduction and improved hydrograph characteristics. Their study had limited calibration data with one sub-basin only using two events and two stream gauge stations due to limited streamflow data availability across events.

Ertas et al. (2016) employed the HBV hydrological model to simulate daily discharge in the Karasu Basin, a sub-region of the Upper Euphrates Basin in Turkiye. Their results indicated good model performance, with Nash-Sutcliffe efficiency values of 0.85 for calibration (2001-2008) and 0.71 for validation (2009-2014). However, their analysis also revealed a significant increase in forecast uncertainty over time, as seen in the root mean square error (RMSE) values, which ranged from 12 m^3/s on the first forecast

day to 43 m^3/s by the ninth day. While HBV performed well for short-term forecasts, its accuracy declined at longer lead times due to rising meteorological uncertainty.

Uysal et al. (2021) studied short-term (1-2 day) hydrological forecast system in snow-dominated mountainous basins, specifically the headwaters of the Euphrates and Seyhan Basins in Turkiye. They employed the HBV model for simulating the precipitation-runoff relationship, utilizing NWP data from MM5 and WRF models. The study showed that WRF-based forecasts generally performed better than MM5-based forecasts, with the 2015 water year predictions showing particularly high performance in both study basins, suggesting potential for improved water resource management and economic benefits.

Doğan et al. (2023) investigated how incorporating snow cover data from satellite products, MODIS and Interactive Multi-sensor Snow and Ice Mapping System (IMS), into hydrological models reduces parameter uncertainty and improves runoff forecasting, especially in snow-dominated basins like the Upper Coruh River Basin in Turkiye. They applied HBV model, calibrated through multi-criteria approaches using runoff and snow cover data, significantly enhancing the reliability of the predictions. Their study also integrated NWP data from the WRF model with a 2-days lead time to forecast short-term runoff, providing critical insights for reservoir management during snowmelt periods.

3. MATERIALS AND METHODS

3.1. Study Area

This study investigated two distinct basins, Upper Seyhan and Upper Aras, selected to represent regional diversity. The specific study sites correspond to the outlets monitored by Çukurkışla (Upper Seyhan) and Kayabaşı (Upper Aras) discharge stations and are henceforth referred to by these station names. Detailed descriptions of these basins are provided in the following sections. Notably, these basins have also been studied in recent such as studies by Eylen (2024) and Traore (2024).

3.1.1. Upper Seyhan (Çukurkışla)

The Seyhan River Basin locates in southern Turkiye, in the provinces of Adana, Nigde, and Kayseri. The Seyhan Basin, one of the twenty five basins in Turkiye, is located between 36º 30'' and 39º 15'' 0 N and 34º 45'' – 37º 00''E. The basin encompasses the catchment areas of the Seyhan River and its tributaries, the Göksu and Zamantı Rivers. Covering an area of 22,042 km², which constitutes 2.82% of Turkiye's total surface area, the Seyhan Basin is bordered by Ceyhan to the east, Konya and Berdan to the west, the Develi Basin and Kulmac Mountains to the north, and the Mediterranean Sea to the south. The location and topographic elevation of the basin are illustrated in Figure 3.1.

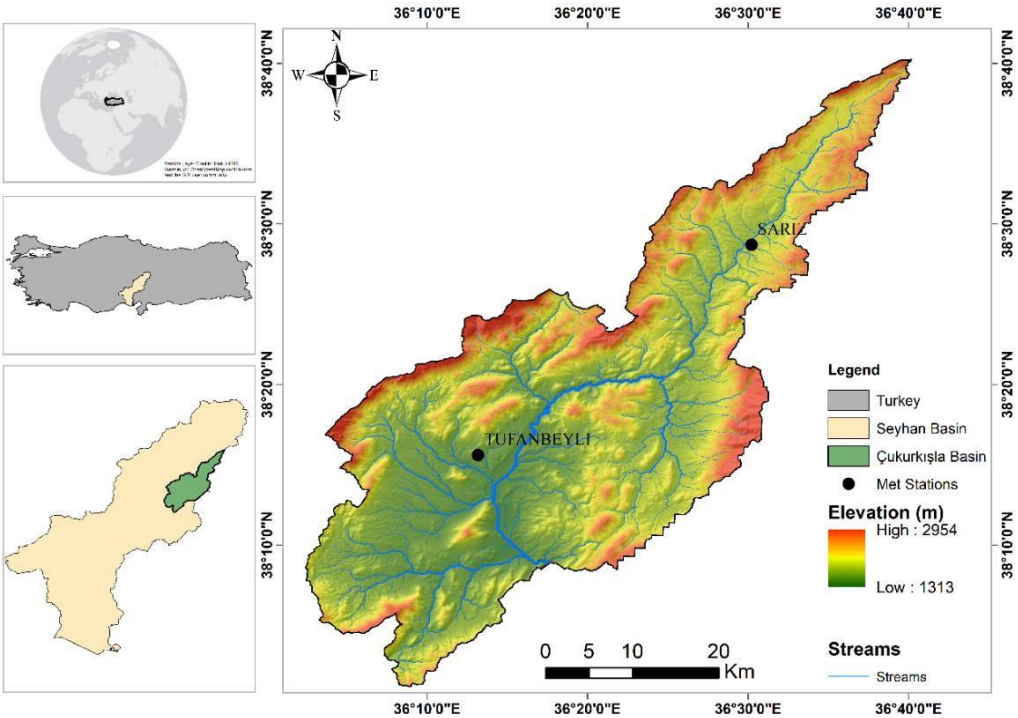

Figure 3.1. *Çukurkışla study area map*

Çukurkışla sub-basin is situated in the upper region of the Göksu River, with drainage area of 1522.4 km^2, which constitutes approximately 6.90% of the total basin area. The catchment's outlet station (E18A024) is operated by State Hydraulic Works (DSİ). The elevation within Çukurkışla Basin ranges from 1313 to 2954 meters above sea level.

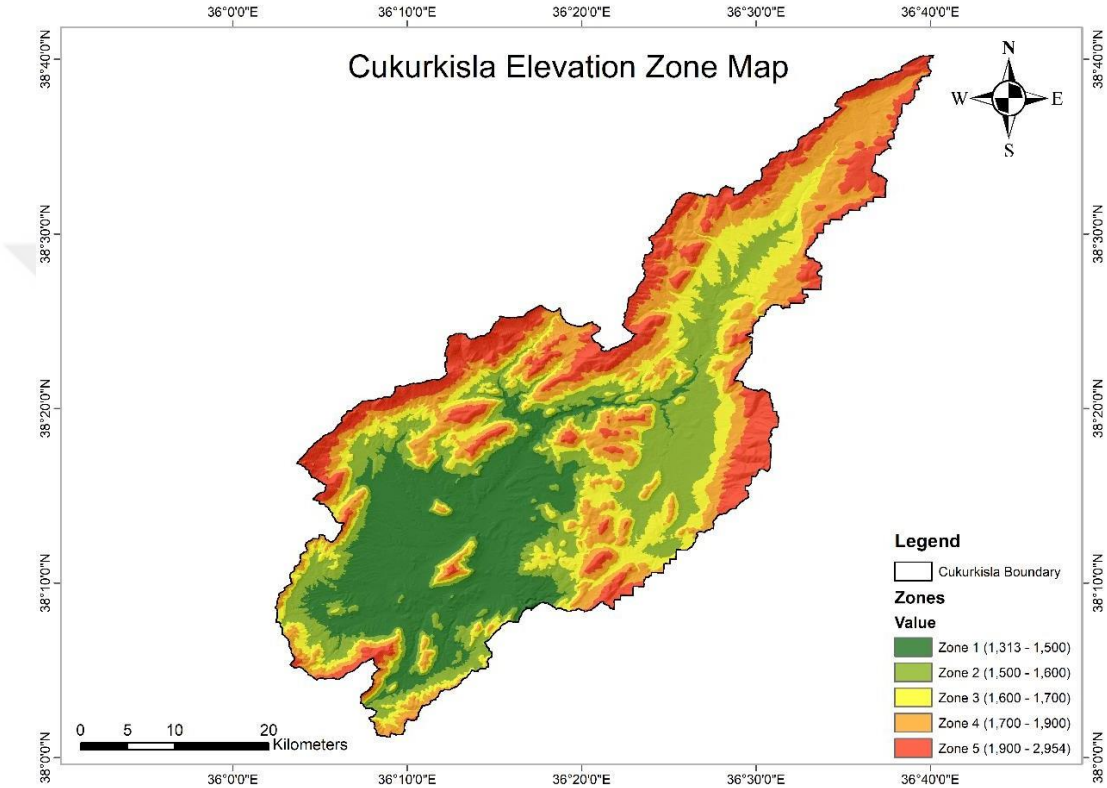

Figure 3.2. *Çukurkışla zone map*

The basin elevation was divided into five zones with equal areas, as shown in Figure 3.2. This shows that more than half of the basin elevation is between 1313 and 1650 m. In addition, 90% of the basin elevation is below 2000 m. Less than 20% of the basin elevation was in zone 5. Figure 3.3 shows the hypsometric curve distribution of the basin elevation, illustrating the relationship between relative elevation and cumulative area. The curve is steep in upper elevation, with less than 10% is above 2000 meters elevation. The elevation decreases gradually at higher area percentages, showing uniform mid-lower terrain distribution. Most of the basin elevations are flat. The median elevation of the basin is 1650 m, which is higher than around 300 meters from the lowest elevation. The basin relief, the difference between the maximum and minimum elevations, is 1641 meters.

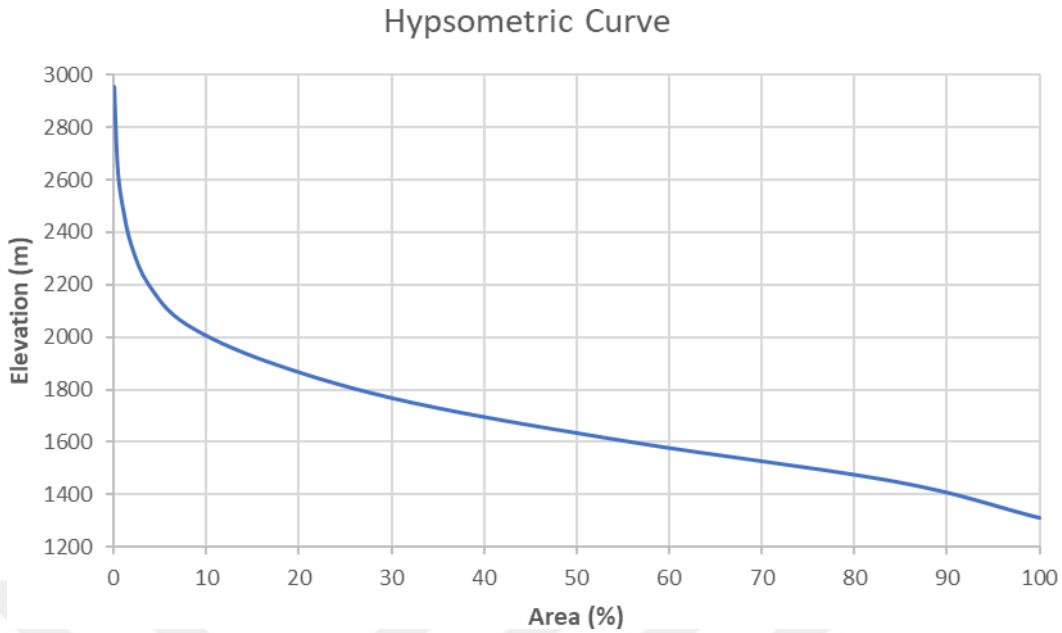

Figure 3.3. *Hypsometric curve of Çukurkışla Basin*

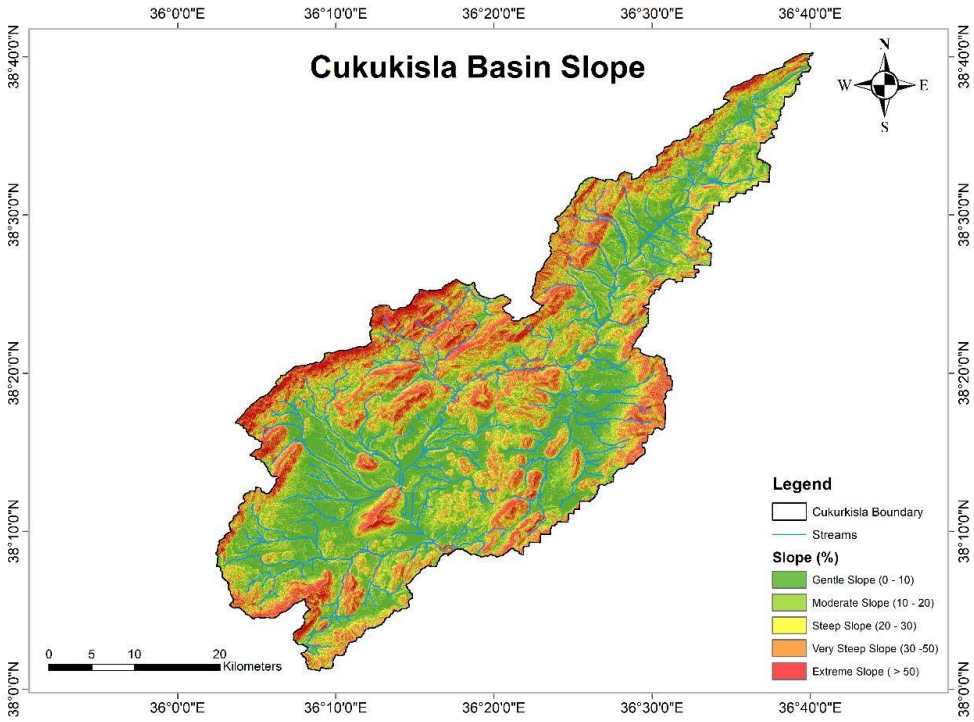

Figure 3.4. *Slope map of Çukurkışla Basin*

The slope of the basin was categorized as gentle, moderate, steep, very steep, or extreme as shown in Figure 3.4. Approximately 50% of the basin slope is less than 20% slope, in gentle and moderate slope. The mean slope of the basin is 22%.

3.1.2. Upper Aras (Kayabaşı)

The Aras Basin, one of Turkiye's 25 hydrological basins, is in northeastern Turkiye. From its source in the Bingöl Mountains, the Aras River flows through northeastern Turkiye, delineates part of the Turkish Armenian border, and continues into Iran and Azerbaijan to meet the Kura River before reaching the Caspian Sea. This basin encompasses portions of the provinces of Ağrı, Ardahan, Erzurum, Iğdır, and Kars, with a 27,548 km^2 drainage area inside Turkiye.

Kayabaşı is a sub-basin within the Aras Basin, situated in the upper part and spanning the provinces of Erzurum and Kars. The sub-basin is located between 39° 52' - 39° 18' north latitude and 41° 10' - 41° 55' east longitude. The State Hydraulic Works (DSİ) operates Kayabaşı Streamflow Station (D24A096) to monitor hydrological parameters. Figure 3.5 shows the location and the topographical elevation of the basin.

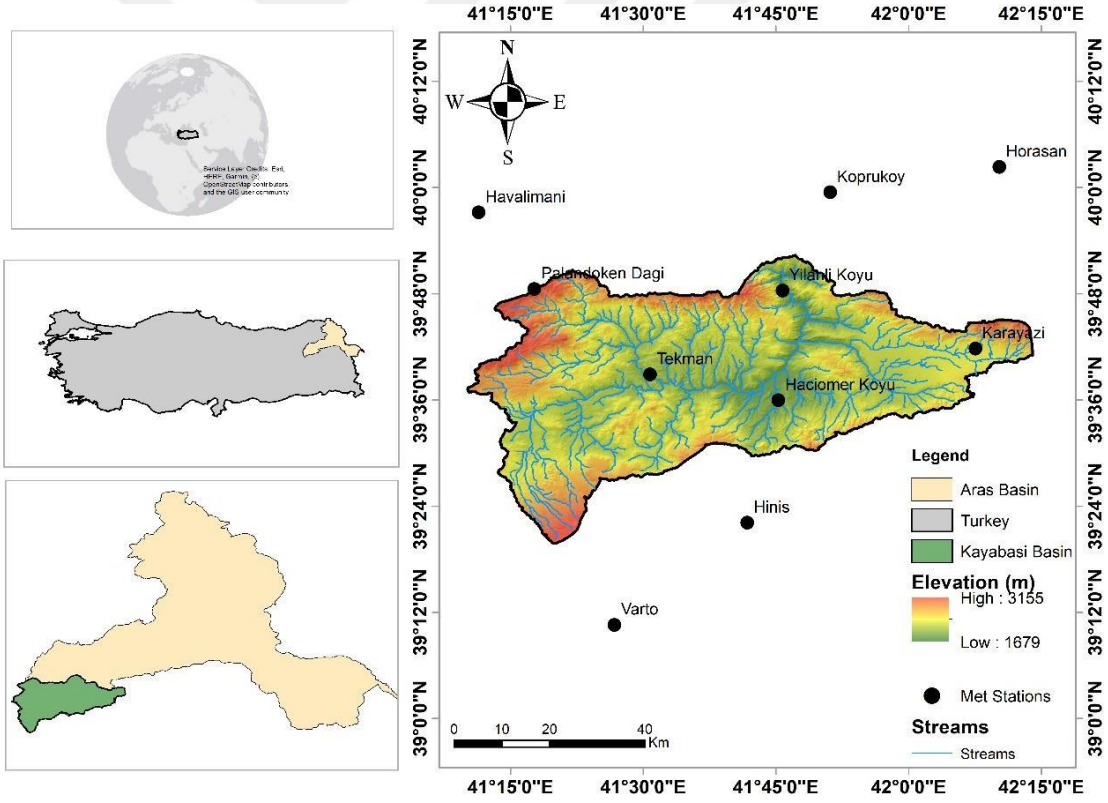

Figure 3.5. *Kayabaşı Basin Map*

Kayabaşı sub-basin covers a drainage area of approximately 2727 km², with elevations ranging between 1679 and 3155 m and an average elevation of 2220 m. Figure 3.6 shows Kayabaşı Basin elevation categorized into five zones to effectively illustrate the elevation diversity within the area. The spatial distribution of elevation was visualized

using a conventional hypsometric. Lower elevations (1679-2000 m) are concentrated in the central valley. Progressively higher intermediate zones, (2000-2150 m) and (2150-2250 m), are primarily located on the basin slopes. The highest elevation bands, (2250-2400 m) and red (>2400 m), correspond predominantly to the northern and southern watershed divides. This shows that 50% of the basin elevation is greater than 2220 m above sea level.

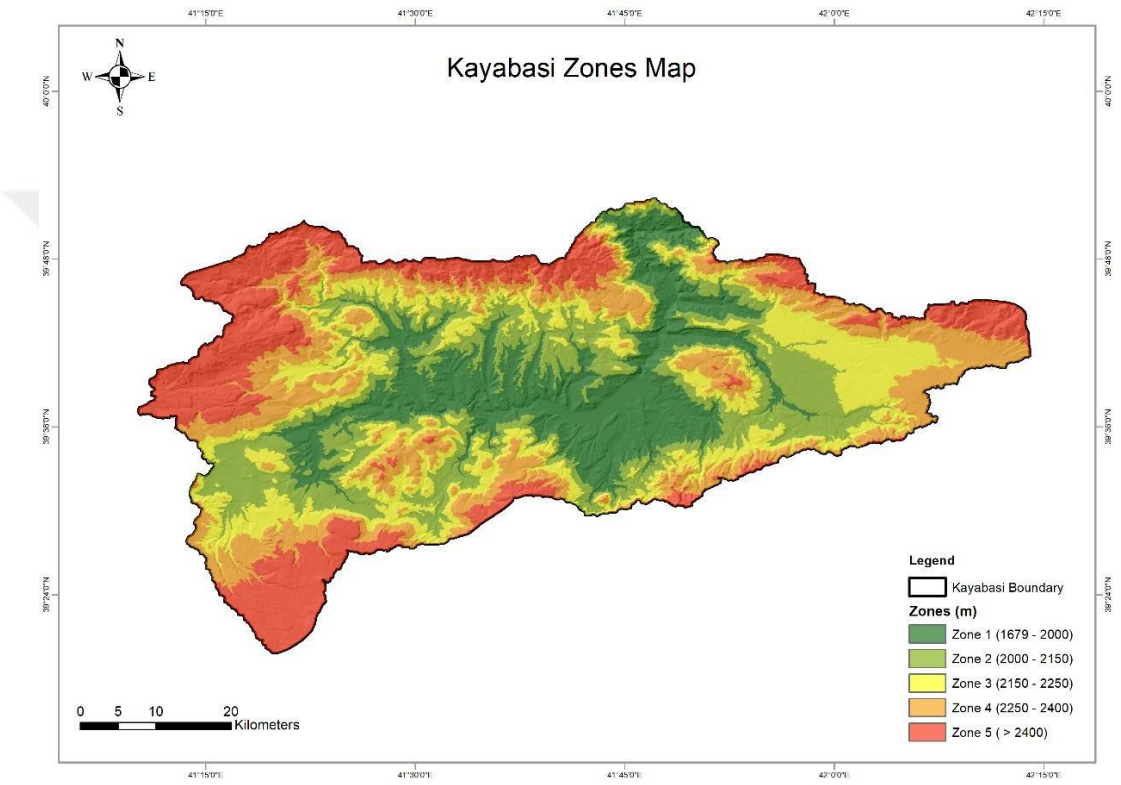

Figure 3.6. *Kayabaşı elevation zone map.*

Figure 3.7 shows the hypsometric curve distribution of the basin elevation, illustrating the relationship between relative elevation and cumulative area. The curve is steep in upper elevation, with less than 10% is above 2500 meters elevation. The elevation decreases gradually at higher area percentages, showing uniform mid-lower terrain distribution. The elevation of this basin is not low as Çukurkışla Basin. The median elevation of the basin is 2200 m, which is higher than around 521 meters from the lowest elevation. Basin relief is 1476 meters, which is lower than that of Çukurkışla Basin.

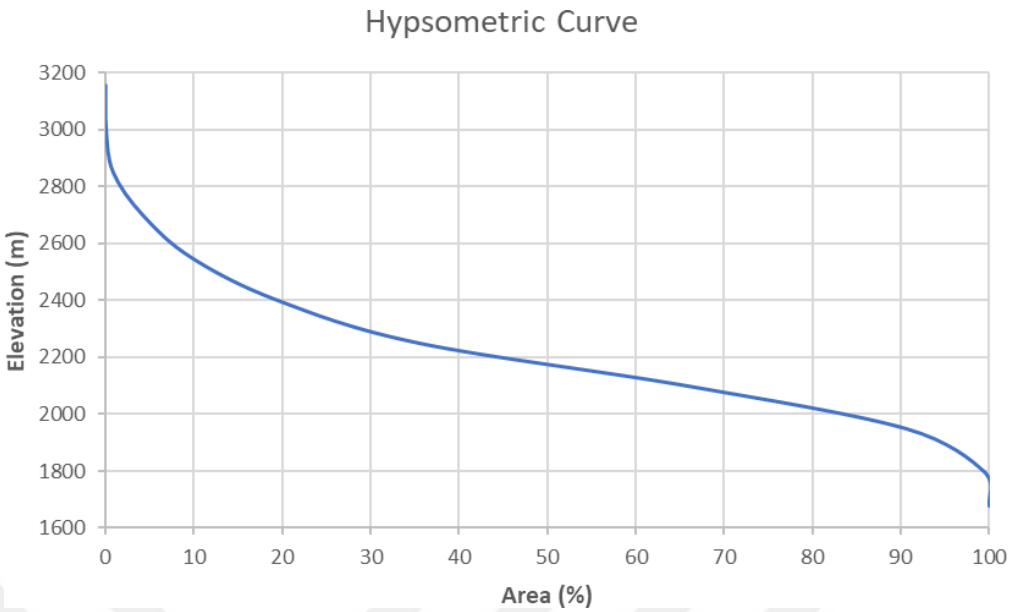

Figure 3.7. *Hypsometric curve for Kayabası Basin*

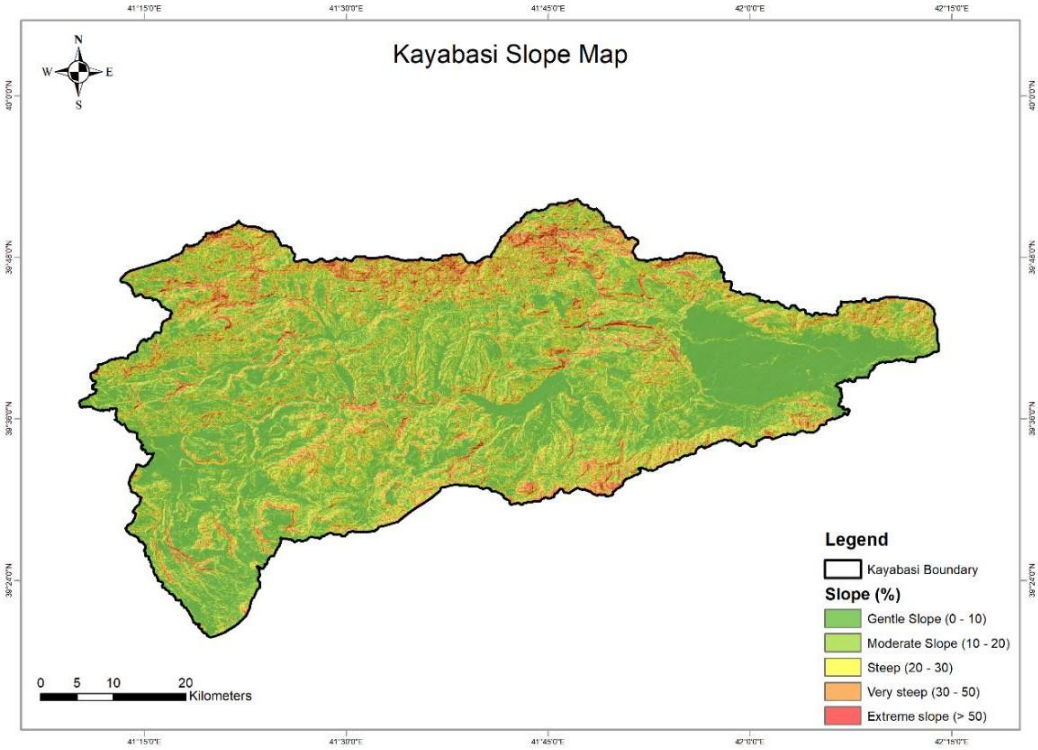

Figure 3.8. *Slope map of Kayabaşı Basin*

The average slope percentage is 19% which is a moderate slope. The basin slope was categorized as gentle slope (0-10%), moderate slope (10-20%), steep slope (20-30%), very steep slope (30-50%), and extremely steep (>50%). Figure 3.8 shows the areas of each category. Most of the basin's drainage area is under a gentle slope of 1000 km^2.

3.2. Flow Chart

The framework used in this study is illustrated in Figure 3.9. After preparation of the input data, the calibration and validation is conducted using the HOOPLA Toolbox. Following this setup, two parallel forecast simulations are executed: Open Loop (OL) and Data Assimilation (DA). The OL, which is a direct forecast without any real-time correction, produces both deterministic and ensemble runs, while the DA scheme employs an Ensemble Kalman Filter (EnKF) with 50 initial states to generate its own set of deterministic and larger ensemble simulations. Finally, the outputs from the various models are combined into a MM result using a simple averaging method.

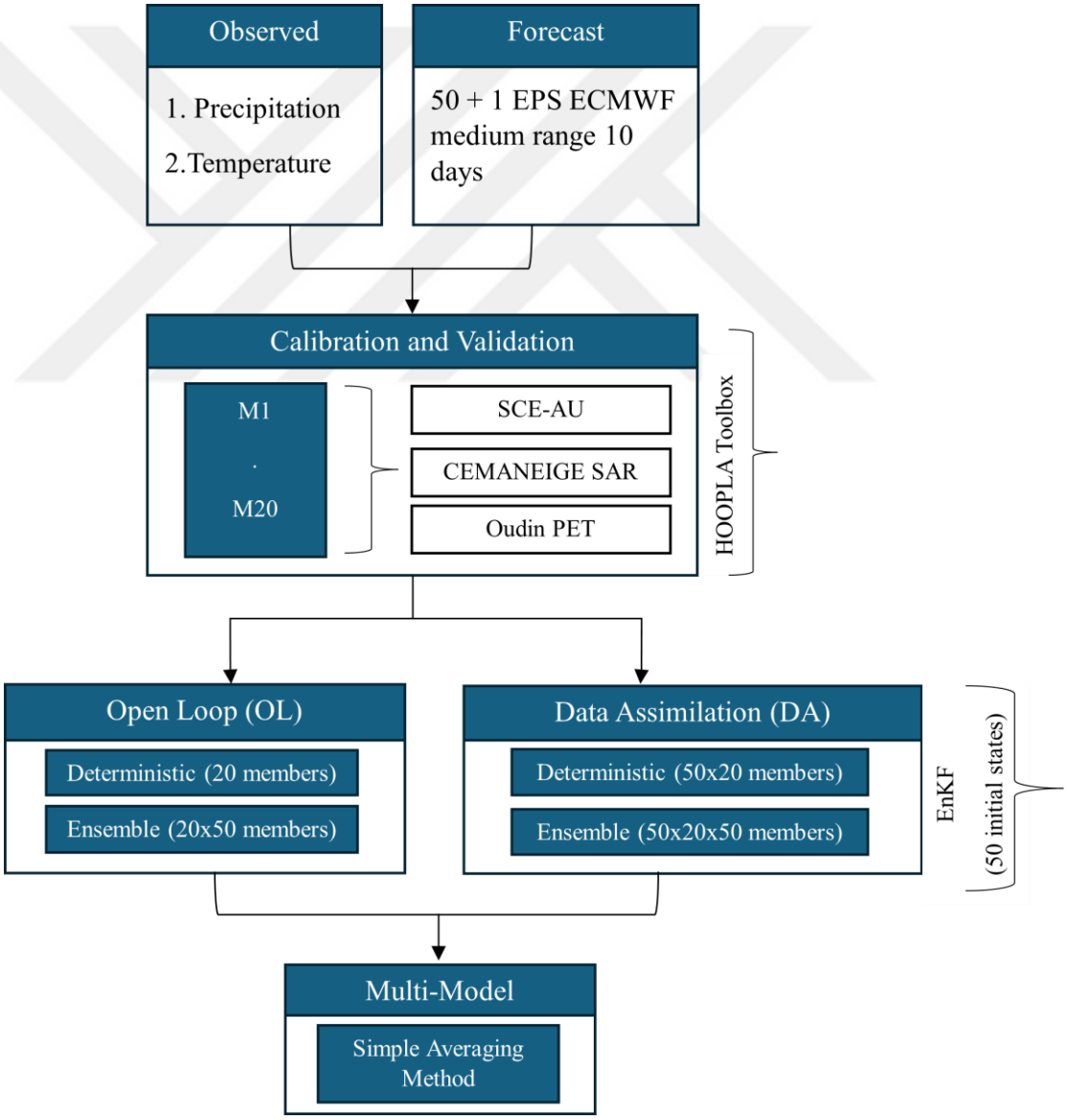

Figure 3.9. *Methodology overview*

3.3. Goodness of Fit

3.3.1. Coefficient of determination

The coefficient of determination, r^2, is defined as the squared value of the coefficient of correlation (Krause et al., 2005) and represents the amount of variance in observed data that is explained by the model (Althoff et al., 2021). It is proposed as a measure of how close the observed-predicted regression line approaches the ideal fit (E. Coffey et al., 2004). It calculates as:

$$R^2 = [\frac{\sum_{i=1}^{n} (O_i - O) + (P_i - P)}{\sqrt{\sum_{i=1}^{n} (O_i - O)^2} \sqrt{\sum_{i=1}^{n} (P_i - P)^2}}]^2 \ , \quad 0 \leq R^2 \leq 1 \qquad (3.1)$$

Where O_i = observed discharge, P_i = simulated discharge, O = mean of observed discharge, P= mean of simulated discharge.

The coefficient of determination, r^2, can take values ranging from 0 to 1, and it indicates how well the predicted values account for the dispersion in the observed data. A value of 0 indicates no correlation, while a value of 1 indicates perfect correlation, i.e., the dispersion of the predicted values is equivalent to that of the observed values. r^2 has some drawbacks, such as oversensitivity to extreme values and insensitivity to systematic under- or overprediction.

3.3.2. Nash-Sutcliffe efficiency

The Nash-Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970) is a widely used goodness-of-fit criterion in hydrological modeling. It is an improvement over r^2 as it is more sensitive to systematic under- or over-prediction. It is defined as one minus the mean squared error (MSE) normalized by the variance of the observed values (σ2). It calculates as:

$$NSE = 1 - \frac{\sum_{i=1}^{n} |O_i - P_i|^2}{\sum_{i=1}^{n} |O_i - B_i|^2}, \qquad -\infty < NSE \leq 1 \qquad (3.2)$$

where B_i is the benchmark series at the time-step i. In its original form B_i = O.

The NSE ranges from -∞ to 1, with 1 indicating a perfect fit. It penalizes the model when the slope and interception of the fitted regression line depart from 1 and 0, respectively. However, the NSE has some limitations, such as underestimating peak flows and underestimating observed flow variability, which can result in inflated NSE value (Althoff et al., 2021). Therefore, it is recommended to use a proper benchmark series instead of threshold values to assess model performance. The Kling-Gupta efficiency

index is suggested as a more reliable criterion for hydrological modeling (H. V Gupta et al., 2009).

3.3.3. Square root of Nash–Sutcliffe efficiency

The square root variant of the Nash–Sutcliffe Efficiency, denoted as NSEsqrt, is a modified form of the traditional NSE, developed to address its sensitivity to high discharge values and improve the assessment of low-flow simulations (Pushpalatha et al., 2012). The NSEsqrt applies a square root transformation to observed and simulated values, reducing extreme value influence and improving low-flow period evaluation, while traditional NSE emphasizes peak flows due to squared errors. The NSEsqrt is calculated using the following equation:

$$NSEsqrt = 1 - \frac{\sum_{i=1}^{n} (\sqrt{Q_{sim,t}} - \sqrt{Q_{obs,t}})^2}{\sum_{i=1}^{n} (\sqrt{Q_{obs,t}} - \sqrt{Q_{obs}})^2} \quad (3.3)$$

Where $Q_{sim,\ i}$ is the simulated discharge at time step *i*, $Q_{obs,i}$ is the observed discharge at time step *i*, Q_{obs} is the mean of the square root of observed discharge, *n* is the total number of observations. The transformation preserves the NSE structure while ensuring that both low and medium flow values are better represented in the efficiency computation.

3.3.4. Kling-Gupta efficiency

The Kling-Gupta efficiency (KGE), proposed by (H. V Gupta et al., 2009), is another goodness-of-fit criterion used to evaluate the performance of hydrological models. It was proposed to overcome the limitations of other criteria, such as the Nash-Sutcliffe efficiency index (NSE), which can lead to underestimation of flow variability. The KGE is formulated by calculating the Euclidean distance of three components: correlation, bias, and measure of variability.

$KGE = 1 - ED, \ -\infty < KGE \leq 1$

$$ED = \sqrt{[S_r.(r-1)^2 + [S_a.(a-1)^2 + [S_\beta.(\beta-1)^2} \quad (3.4)$$

where "ED is the Euclidian distance, r is the linear correlation coefficient between O_i and P_i, α is the variability ratio or ratio between the standard deviation of simulated values and standard deviation of observed values (σ_P/σ_O), β is the ratio between mean of the simulated and observed values (μ_P/μ_O or P/O). s_r, s_α, and s_β are scaling factors that can

re-scale the criteria space before computing the ED, that is, emphasizing different components. Conventionally, the scaling factors have equal weight, i.e., $S_r = S_\alpha = S_\beta = 1$" (Althoff et al., 2021). The index ranges from $-\infty$ to 1, with 1 indicating a perfect fit.

3.4. Data

3.4.1. Hydro-meteorological data

The Turkish Meteorological Services (MGM) and State Hydraulics Works (DSI) collect and provide precipitation, temperature, humidity, wind, evapotranspiration, and discharge data for all Turkish Basins. We obtained precipitation and temperature data from the MGM and stream data from the DSI.

3.4.1.1. Çukurkışla

Daily precipitation and temperature data from four stations – two inside and two outside the basin – were obtained from the MGM. Pınarbaşı and Göksun stations occur outside of the basin while Sarız and Tufanbeyli stations are inside the basin. To confirm the effect of each station on the basin, Thiessen Polygons were used for the distribution. Sarız and Tufanbeyli stations cover all the areas of the basin, 40% and 60% respectively. Streamflow data from the Göksu station (E18A024) were obtained from the DSI. Table 3.1 lists the meteorological stations within and around Çukurkışla Basin and their altitudes.

Table 3.1. *Meteorological stations for Çukurkışla Basin*

Station No	City	Station Name	Latitude	Longitude	Altitude (m)
18053	Adana	Tufanbeyli	38.26	36.2195	1400
17866	Kahramanmaraş	Göksun	38.024	36.4823	1344
17802	Kayseri	Pınarbaşı	38.7251	36.3904	1542
17840	Kayseri	Sarız	38.4781	36.5035	1599

The span of the obtained data for these stations was from 2000 to 2019 water years. The average annual total precipitation of the basin is 500.8 mm, whereas the discharge is 165.40 mm. The average annual temperature is 9.44 °C. The years 2002, 2009, and 2019 had the highest annual precipitation and discharges. The lowest precipitation was recorded in 2017 (324 mm), whereas the lowest discharge was recorded in 2014 (60.35 mm). According to the temperature records, the lowest average annual temperature was recorded in 2012 at 8.14 °C, with an average precipitation of 514 mm and high discharge

of 204.72 mm. The highest average annual temperature was 11.38 °C in 2010. Figure 3.10 shows the annual total precipitation, discharge, evaporation, and average temperature of Çukurkışla basin.

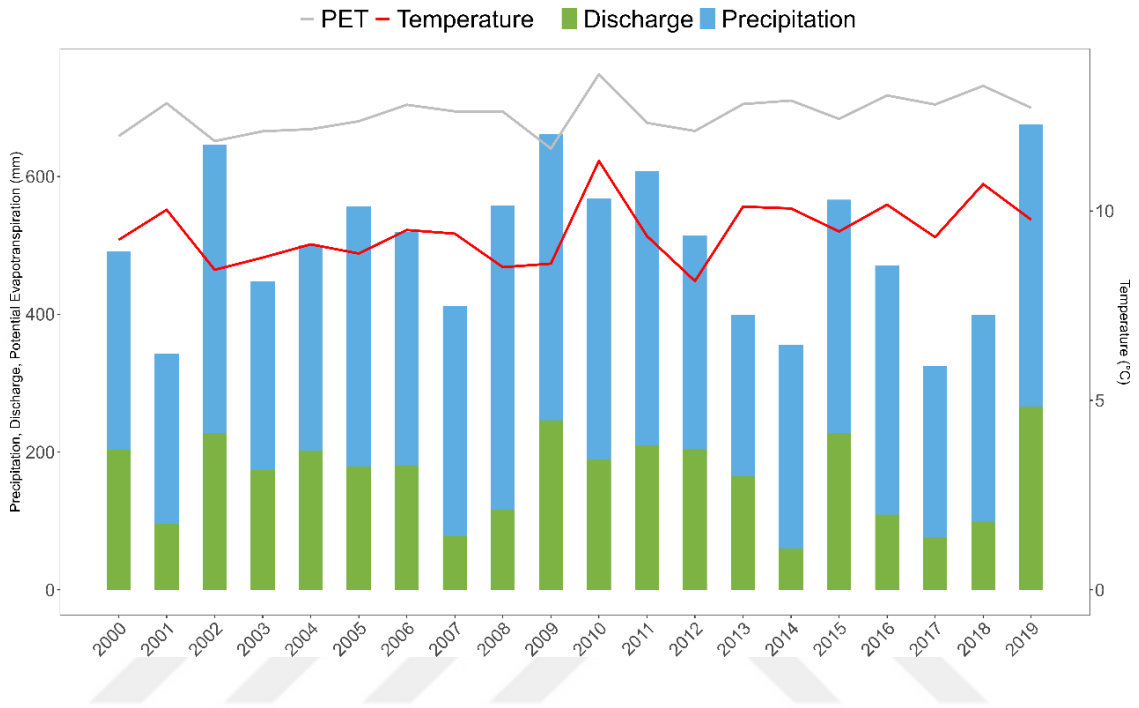

Figure 3.10. *Hydro-meteorological data for Çukurkışla Basin during the 2000-2019 water years.*

Figure 3.11 shows the monthly average temperature, precipitation, PET, and discharge from 2000 to 2019 water years. The lowest temperatures were observed in January, below freezing degrees. The temperature goes up in mid of February, which the snow melting starts. according to the discharge, more than 65% of the average basin discharge was observed between February and May. March and April had the highest average total monthly discharge, exceeding 30 mm/day.

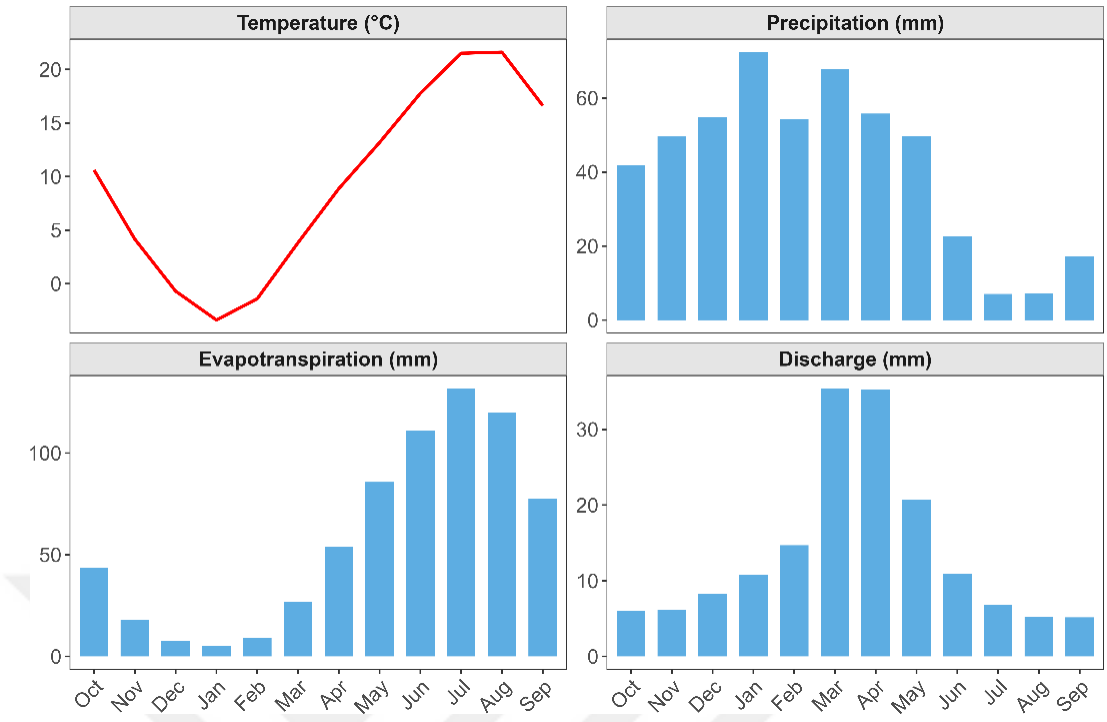

Figure 3.11. *Monthly temperature, precipitation, PET, and discharge average of Çukurkışla Basin*

Based on the data in Figure 3.12, the annual precipitation and temperature values for each year were categorized into four groups: Dry/Cold, Wet/Cold, Dry/Hot, and Wet/Hot years. These categories were determined by comparing the yearly values of precipitation and temperature with the respective medians for each variable. The data revealed interesting patterns, showing how these two climatic variables influenced each other over time. For instance, in 2010, the combination of high precipitation and high temperature placed it in the wet/hot category. However, 2017 was noted as the driest year, yet its temperature remained close to the median, indicating that while precipitation was notably low, the temperature did not exhibit extreme variations. Furthermore, 2019 stands out as the wettest year recorded in the dataset, yet its temperature surpasses the median for the Wet/Hot category. This suggests that although precipitation was abundant, the climate was still relatively warm compared to other years within the same category. In contrast, many of the years fall into the Wet/Cold and Dry/Hot categories, which appear to dominate the dataset. This could imply a recurring trend in the regional climate, where either wet condition with lower temperatures or dry conditions with higher temperatures are more common. As shown in Figure 3.12, years such as 2006, 2009, 2011, and 2015 consistently appear in the wet/cold category, whereas years such as 2001, 2007, 2013, and 2014 frequently fall into the dry/hot category.

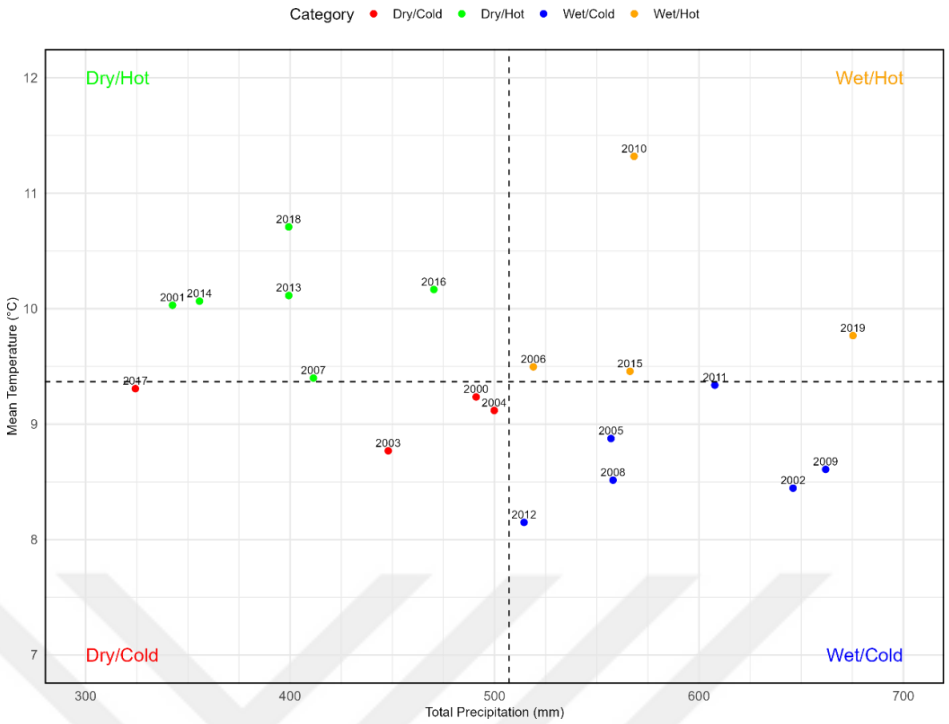

Figure 3.12. *Categorization of precipitation and temperature of Çukurkışla Basin.*

Furthermore, the potential evapotranspiration (PET) was calculated using the method described by Oudin et al. (2005). The average annual potential evapotranspiration was 690.4 mm. The highest PET was observed at 748.39 mm in 2010, while the lowest PET was observed at 640.5 mm.

3.4.1.2. Kayabaşı

Daily precipitation and temperature data from 15 meteorological stations (five inside and eight outside the basin) were obtained from the MGM, covering the 2008 to 2019 water years. Table 3.2 lists the meteorological stations of Kayabaşı Basin. Basin-wide precipitation and temperature were calculated for all stations using the Thiessen Polygon method. Additionally, streamflow data from Kayabaşı streamflow station (D24A096) were acquired from the DSI, covering the water years from 2008 to 2019. The average annual precipitation for the basin is 519.20 mm, whereas the average annual temperature is 6.14 ºC. The average annual discharge for the basin is 241.70 mm.

Table 3.2. *Meteorological Stations for Kayabaşı Basin.*

Station No.	City	Station Name	Latitude	Longitude	Altitude (m)
17096	Erzurum	Havalimanı	39.9529	41.1897	1758
17672	Erzurum	Palandöken Dağı	39.8088	41.2947	2973
17687	Erzurum	Ilıca havzası	39.8877	41.0766	2094
17690	Erzurum	Horasan	40.0383	42.1705	1540
17740	Erzurum	Hınıs	39.3688	41.6957	1715
17778	Bingöl	Varto	39.1763	41.4455	1650
18177	Ağrı	Karlıova	39.2936	41.0106	1828
18203	Erzurum	Çat	39.6058	40.975	1907
18204	Erzurum	Karayazı	39.6964	42.1256	2246
18366	Erzurum	Köprüköy	39.9908	41.8522	1685
18370	Erzurum	Tekman	39.6478	41.5125	1980
19072	Erzurum	Hacıömer Köyü	39.5992	41.7542	1832
19254	Erzurum	Yılanlı Köyü	39.8057	41.7623	1919

As shown in Figure 3.13, 2010 experienced the highest precipitation, whereas 2014 recorded the lowest. Precipitation and discharge exhibited a strong positive correlation, with peaks in 2010 (692 mm and 393 mm, respectively) and troughs in 2014 (327 mm and 124 mm, respectively). This suggests a strong relationship between precipitation and discharge, as years with higher precipitation generally corresponded to increased runoff. The highest temperature was recorded in 2010 at 9.68 °C.

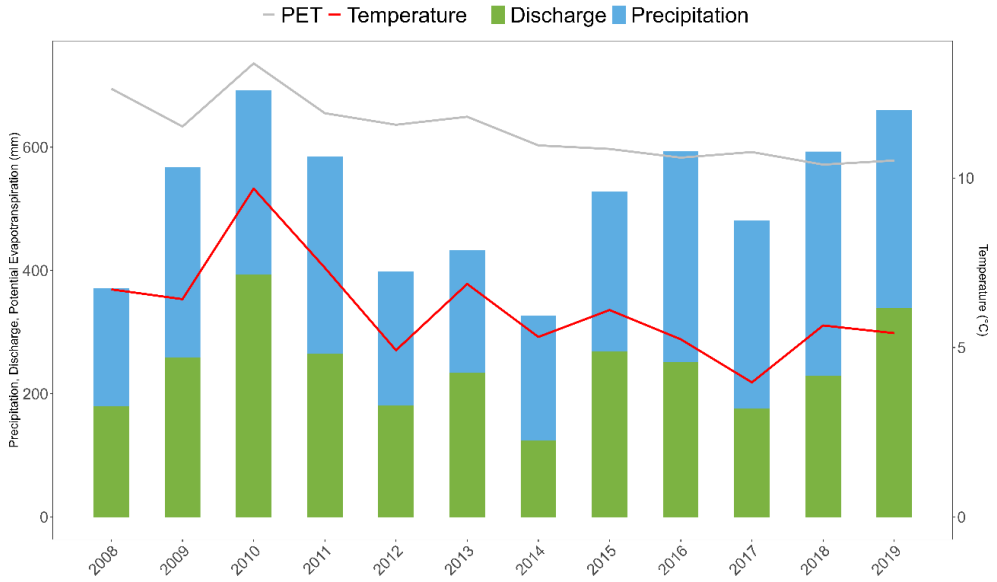

Figure 3.13. *Hydro-meteorological data for Kayabaşı Basin during the 2008-2019 water years.*

Figure 3.14 shows the average monthly temperature and monthly percentages of precipitation, potential evapotranspiration, and discharge from 2008 to 2019 water years. The monthly average temperature shows lower temperatures in the first months of the water year. The average lowest temperature was observed in December, January, and February, at less than 0°C. The temperature increase starts mid of March, which affects the snow melting. According to discharge, more than 60% of the basin discharge is observed between March to June. April and May had the highest average total monthly discharge, exceeding 60 mm/day. This shows snow melting effect on this basin which is higher than Çukurkışla Basin. As shown in Figure 3.15, the precipitation and temperature in each year were categorized. It shows that 2016,2018 and 2019 are wet and cold. Some other years, 2008, 2013 and 2018, were observed to be hot and dry. It is clear the diversity of precipitation and temperature throughout each year.

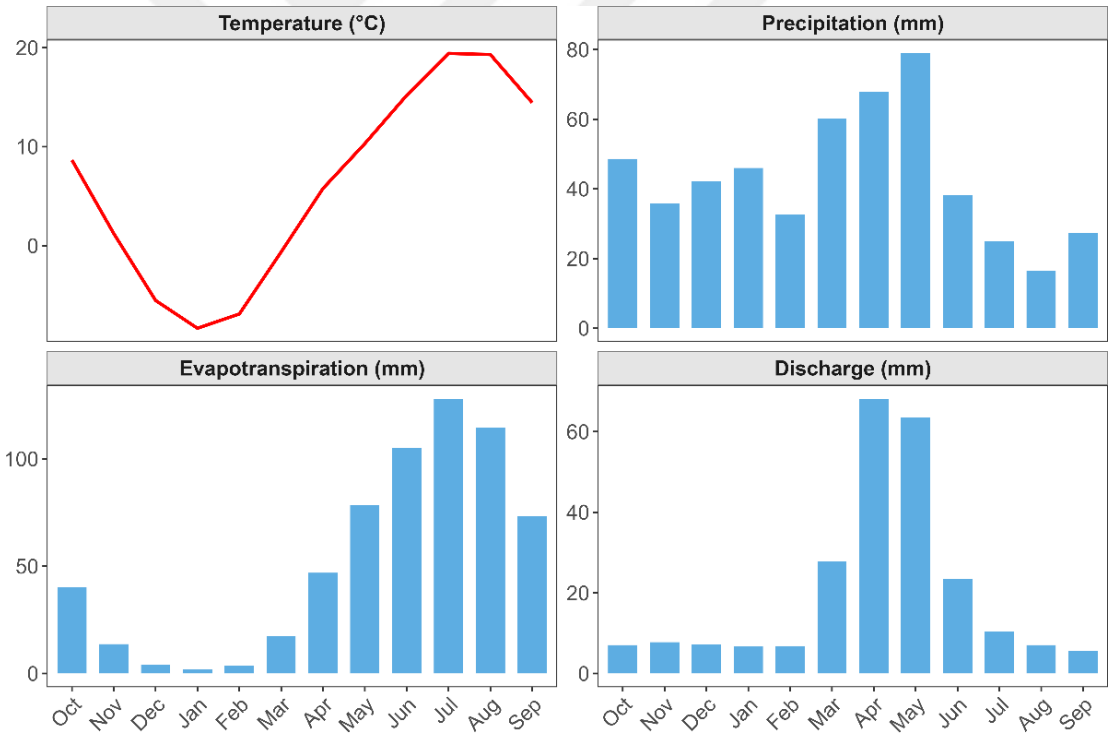

Figure 3.14. *Monthly temperature, precipitation, PET, and discharge average of Kayabaşı Basin*

Figure 3.15. *Categorization of precipitation and temperature of Kayabaşı Basin*

3.4.2. Forecast data

Medium-ensemble weather forecasts from ECMWF serve as essential inputs for hydrological forecasting. The ECMWF produces high-resolution numerical weather predictions that are widely used for probabilistic hydrological applications, allowing for improved flood forecasting, water resource management, and climate impact assessments.

These forecast data were obtained from The International Grand Global Ensemble (TIGGE) archive (www.apps.ecmwf.int), a comprehensive resource that provides ensemble forecast data from leading global weather prediction centers. TIGGE is a key initiative under the World Weather Research Programme (WWRP), aimed at fostering advancements in ensemble forecasting through open data-sharing and MM comparison. By leveraging ensemble forecasts, uncertainties in meteorological predictions can be quantified, improving decision-making for extreme weather events and water-related applications.

For this study, we utilized the ECMWF Ensemble Forecast (ENS) product, which consists of 51 ensemble members: one control forecast and 50 perturbed members. These ensemble members are designed to account for initial condition and model uncertainties, providing a probabilistic range of possible weather scenarios. The forecast data were retrieved in GRIdded Binary version 2 (GRIB2) format, a standard format for storing and transmitting gridded meteorological data. Additionally, ECMWF offers its weather datasets in Network Common Data Form (NETCDF) format, a widely used format for climate and weather data storage due to its self-descriptive structure and efficient compression capabilities.

Precipitation and temperature data (mean, maximum, and minimum) for the period of January to June in 2018 and 2019 for both study basins were downloaded. This period was selected as the most annual discharges for these basins, which were observed between these months as shown in Figures 3.11 and 3.14. The dataset includes six-hourly time steps with a forecast lead time of 10 days, which enables us to analyze medium range hydrological forecasts.

To process and manage the GRIB2 files, we utilized ecCodes, an open-source software package developed by ECMWF for handling GRIB, BUFR, and GTS messages. EcCodes is capable of decoding, encoding, and manipulating weather data in multiple formats, including:

- WMO FM-92 GRIB (editions 1 and 2) – the standard format for numerical weather prediction models.
- WMO FM-94 BUFR (editions 3 and 4) – used for observational data such as satellite and surface measurements.
- WMO GTS abbreviated headers (decoding only) – supporting data exchange within the Global Telecommunication System (GTS).

More details about ecCodes and its functionalities can be found at confluence.ecmwf.int/display/ECC/ecCodes+Home. After decoding the GRIB2 files, the data were systematically organized, aggregated, and converted into time-series format using the R programming language. R provides robust tools for handling large spatiotemporal datasets and performing statistical analyses, making it well-suited for hydrological applications. These processes involved:

1. Extracting relevant meteorological variables (precipitation and temperature) for Çukurkışla and Kayabaşı Basins through the study period.

2. Converting six-hourly data into daily aggregated values, ensuring consistency in temporal resolution for further hydrological modelling.
3. Quality controlling for the forecast data regarding the observed precipitation and temperature data.

3.4.2.1. Çukurkışla

The temperature and precipitation forecasts for Çukurkışla Basin were downloaded at a spatial resolution of 0.25º × 0.25º grids, as shown in Figure 3.16. To ensure data reliability, the forecasted meteorological variables were subjected to a quality control process, where they were compared against observed historical data from ground-based meteorological stations. This step is crucial for identifying any systematic biases, inconsistencies, or outliers in the dataset. Common quality control measures included: Screening for missing or erroneous values, assessing temporal consistency, evaluating forecast bias.

After quality control, the gridded forecast data were aggregated into basin-wide forecast time-series to provide spatially representative meteorological inputs for hydrological modeling. To evaluate the accuracy of precipitation forecasts for Çukurkışla in 2018, the monthly accumulated precipitation was analyzed by comparing observed and forecasted values. This comparison helps assess the reliability of the forecast data and identify any systematic biases across different lead times. The forecasted precipitation data were generated with a 10-day lead time, allowing for an assessment of forecast accuracy over this period. January exhibited the highest accumulated precipitation for both observed and forecasted datasets across all lead times. While the forecasted precipitation generally followed the observed trend across all months, there were noticeable biases, particularly at longer lead times, where precipitation was slightly overestimated. Figure 3.17 shows the monthly accumulated precipitation for observed and forecasted data in 2018, highlighting variations and forecast performance throughout the year.

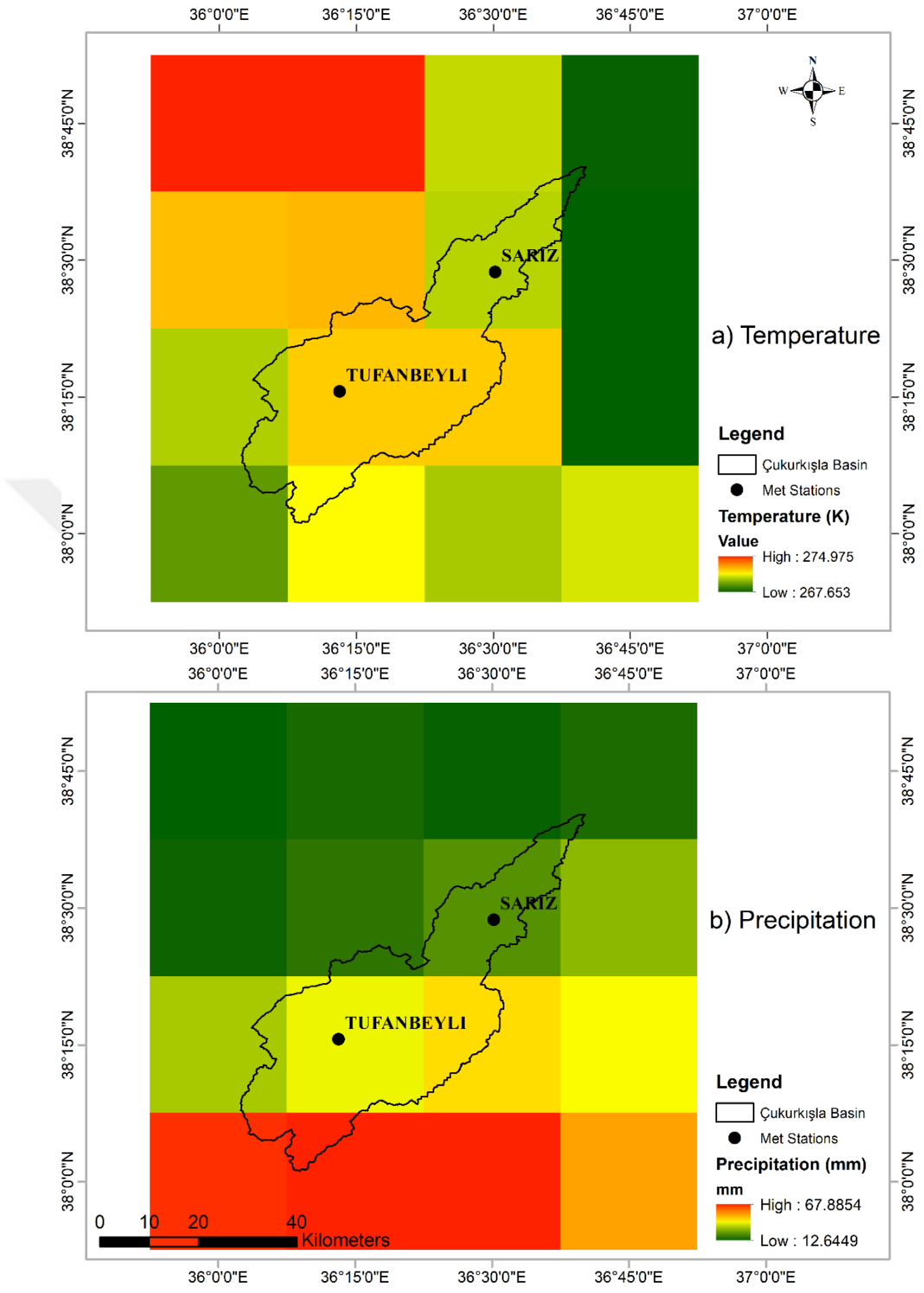

Figure 3.16. *The 0.25º × 0.25º Grids for Çukurkışla Basin in 15/03/2019 a) Temperature in Kelvin b) Precipitation in mm*

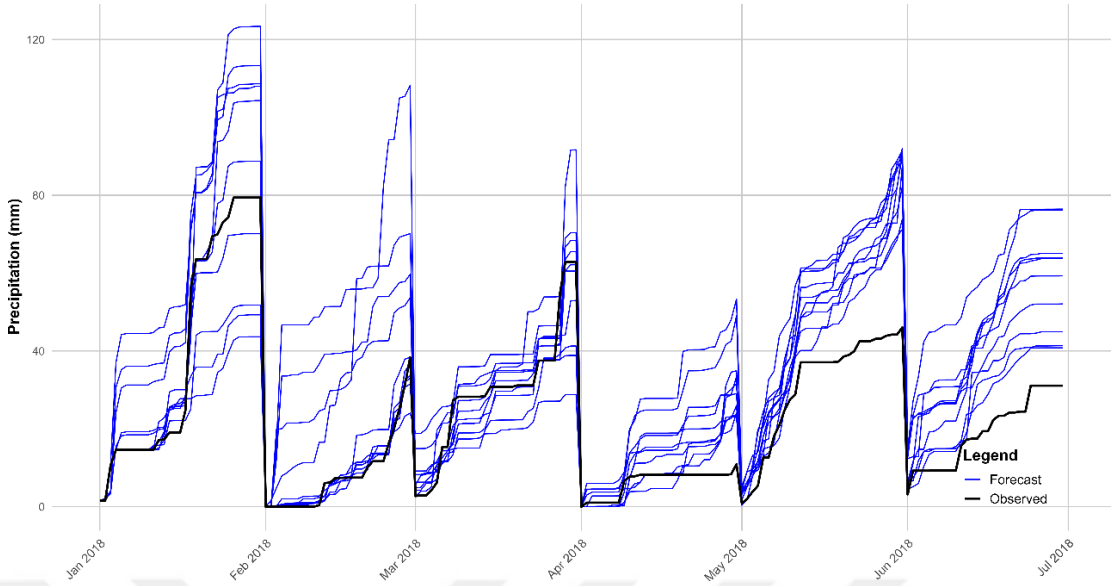

Figure 3.17. *Monthly observed and forecast precipitation accumulation for Çukurkışla Basin in 2018.*

To evaluate the relation between the observed and forecasted precipitation, correlation is calculated across all lead times. As lead time increases, the correlation between observed and forecasted precipitation declines, starting at 0.64 on the first lead day and dropping to 0.079 by the last lead day, as illustrated in Figure 3.18. This trend indicates that forecast accuracy diminishes over longer lead times, highlighting the challenge of predicting precipitation with extended forecasts.

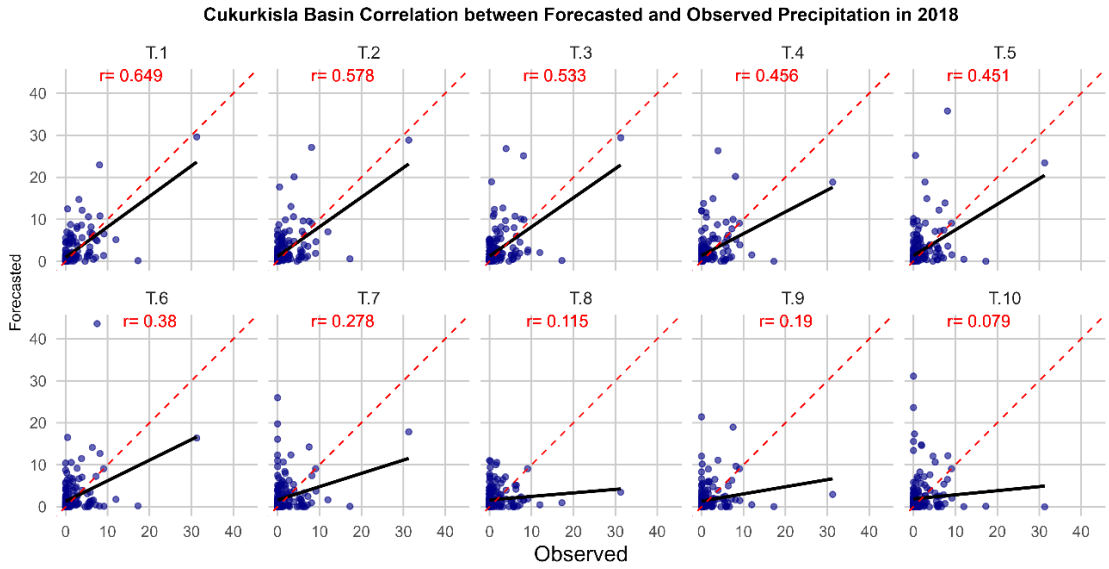

Figure 3.18. *The correlation between observed and forecast precipitation of Çukurkışla Basin through different lead-times in early half of 2018*

Figure 3.19 compares precipitation forecasts with observed precipitation (black line) for Çukurkışla Basin in early 2018. It displays an ensemble prediction system (EPS), showing the 50-member range (blue shaded area) and ensemble mean (blue line), alongside a deterministic forecast (red line). The comparison is presented for forecast lead times of Day 1 (top), Day 5 (middle), and Day 10 (bottom).

Ensemble and deterministic precipitation forecasts are compared with observed basin precipitation data. The ensemble spread for the first lead days is narrow, particularly during periods of little or no precipitation. At first lead time, both the ensemble mean, and deterministic forecast generally capture the observed precipitation well in timing major precipitation events.

At day five, forecast skill decreases. While forecasts attempt to capture major wet periods seen in observations, alignment in timing and magnitude is less precise. The ensemble mean and deterministic forecasts often appear smoother and fail to capture observed peaks. The ensemble spread is significantly wider than at lead day one, representing increased forecast uncertainty at this longer lead time.

At lead day ten, forecasts (ensemble and deterministic) show a weak relationship with observed precipitation. Both struggle to predict timing and magnitude, often showing low-amplitude variations or missing events. The ensemble spread is wide, ranging from zero to significant precipitation, indicating low predictability and high uncertainty for precipitation ten days ahead.

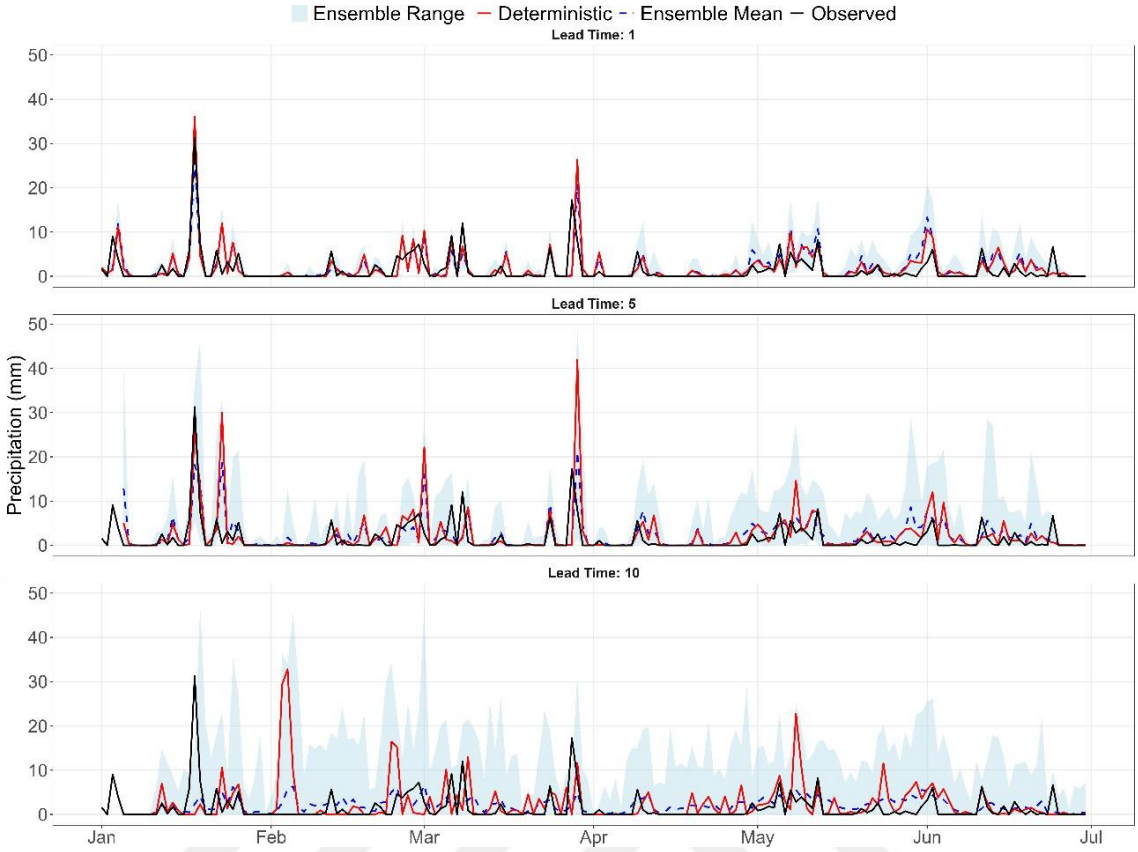

Figure 3.19. *Ensemble and deterministic precipitation comparison plots in Çukurkışla in 2018 with the observed in lead times 1, 5, and 10*

Figure 3.20 compares temperature forecasts with observed temperatures (black line) for Çukurkışla basin in early 2018. Like the previous precipitation, it shows EPS performance through the 50-member range (blue shaded area) and the ensemble mean (blue line), alongside the deterministic forecast (red line). The comparison covers three forecast lead times: Day 1 (top), Day 5 (middle), and Day 10 (bottom).

At day one lead time, temperature forecasts show good accuracy, with ensemble mean and deterministic forecast aligning with observed temperatures, capturing trends and daily fluctuations, but there is underestimation bias. The narrow ensemble spread indicates strong agreement among members and high forecast confidence. The observed temperature consistently falls within this range, reinforcing the reliability of early forecasts.

By day five, the forecast skill decreases. The ensemble mean and deterministic forecast follow temperature trends, smoothing early forecasts fluctuations. Minor deviations from observed temperatures become noticeable. The ensemble spread widens, reflecting increased uncertainty in longer-range forecasting, especially during dynamic

temperature changes. Observed temperatures generally remain within the ensemble spread, capturing probable outcomes despite a slight decline in accuracy.

At a 10-day lead time, forecast skill declines further, though forecasts still capture the seasonal warming trend. The ensemble mean and deterministic forecasts smooth observed variability, often lagging or missing short-term fluctuations. The ensemble spread widens, reflecting increased uncertainty in long-range predictions. While accuracy diminishes, observed temperature consistently falls within the ensemble range, highlighting its value in providing realistic outcomes.

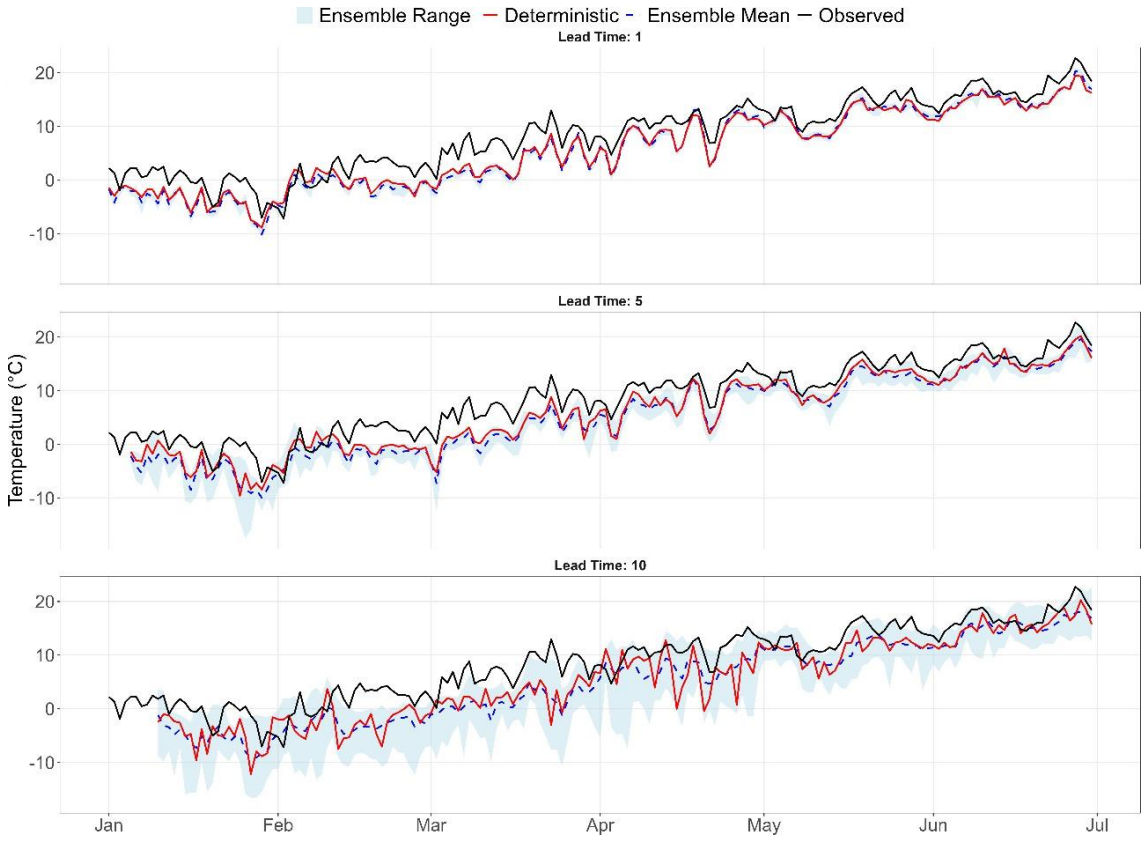

Figure 3.20. *Ensemble and deterministic temperature comparison plots in Çukurkışla in 2018 with the observed in lead times 1, 5, and 10*

Similarly, to evaluate the accuracy of precipitation forecasts for Çukurkışla in 2019, January received the highest precipitation for both observed and forecast. The overall forecast precipitation trend for each month through all the lead days agreed with the trend of the observed precipitation although some months forecast precipitation is overestimated or underestimated for different lead days. Figure 3.21 highlighted the monthly accumulated precipitation for both observed and forecasted data in 2019.

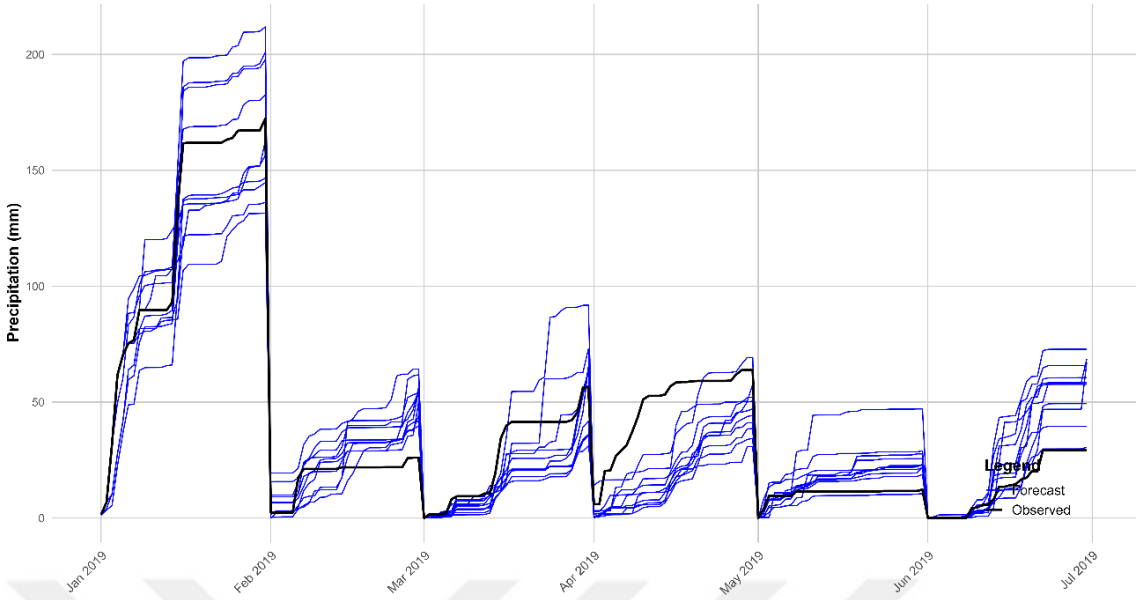

Figure 3.21. *Monthly observed and forecast precipitation accumulation for Çukurkışla Basin in 2019*

Similar to 2018, the correlation between observed and forecasted precipitation in 2019 declines as lead time increases. In 2019, the correlation starts at 0.739 on the first lead day and decreases to 0.376 on the last lead day, as illustrated in Figure 3.22. Compared to 2018, the correlation in 2019 is higher across all lead times.

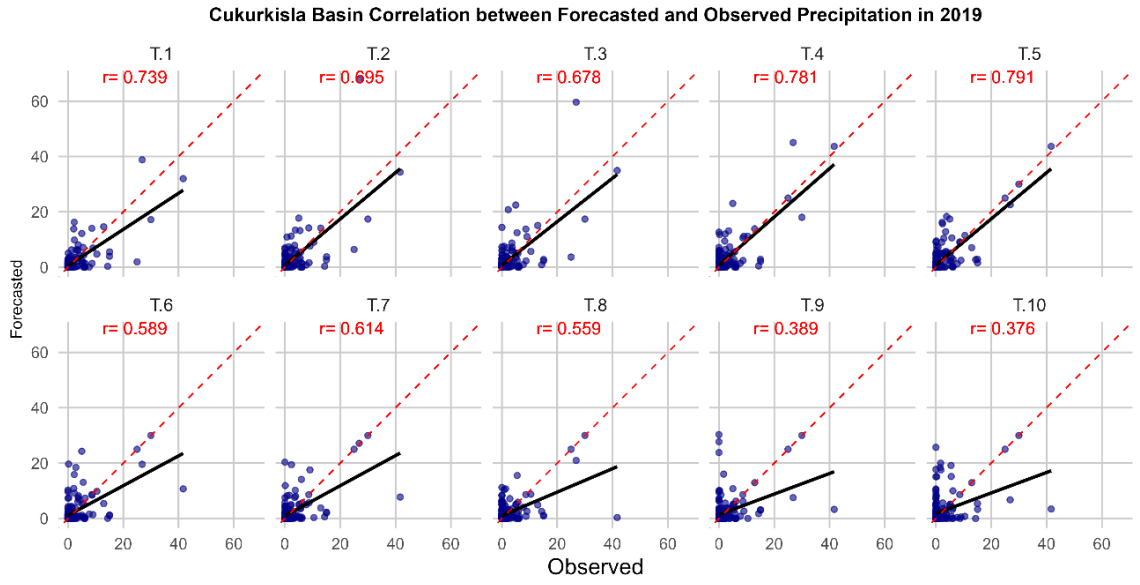

Figure 3.22. *The correlation between observed and forecast precipitation of Çukurkışla Basin through different lead-times in early half of 2019*

Figure 3.23 compares precipitation forecasts with observed precipitation (black line) for Çukurkışla Basin in early 2019. It displays ensemble, showing the 50-member

range (blue shaded area) and ensemble mean (blue line), alongside a deterministic forecast (red line). The comparison is presented for forecast lead times of Day 1 (top), Day 5 (middle), and Day 10 (bottom).

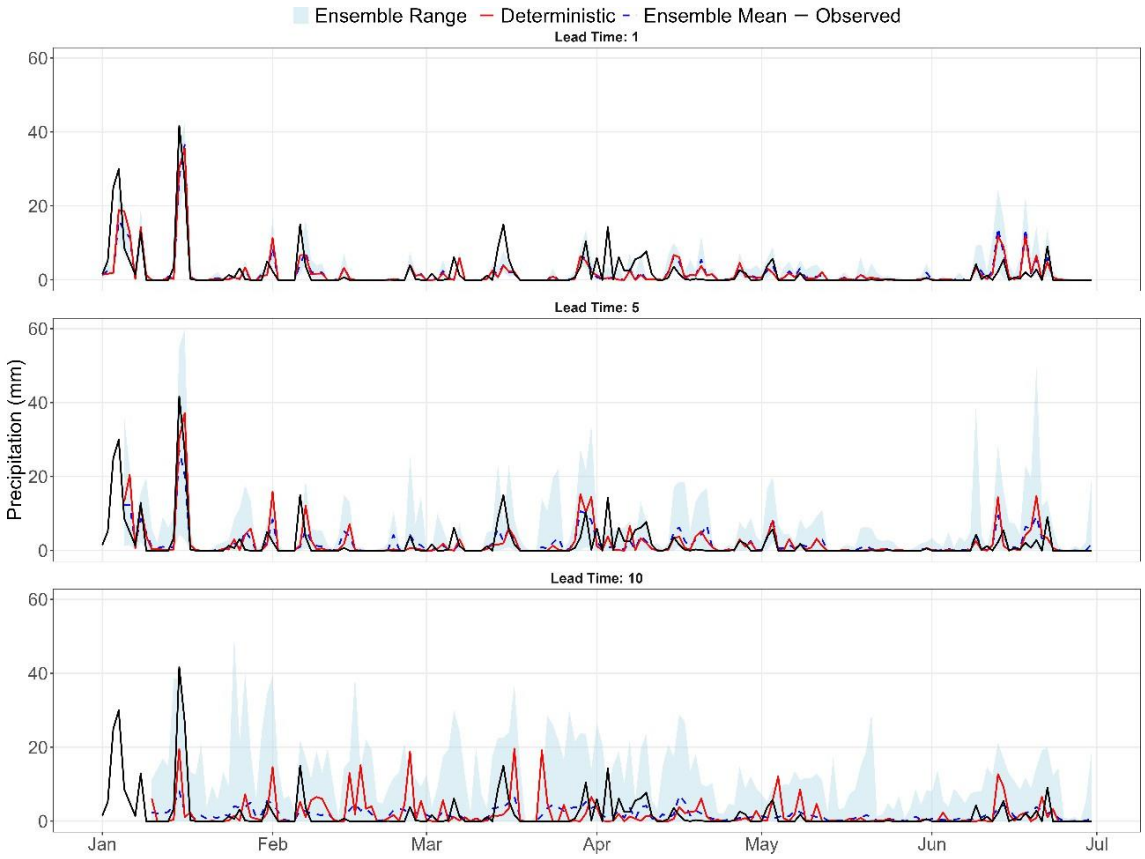

Figure 3.23. *Ensemble and deterministic precipitation comparison plots in Çukurkışla in 2019 with the observed in lead times 1, 5, and 10*

Ensemble and deterministic precipitation forecasts are compared with observed basin precipitation data. The ensemble spread for the first lead day is narrow, particularly during periods of little or no precipitation, indicating higher confidence. At this first lead time, both the ensemble mean, and deterministic forecast capture the timing of observed major precipitation events well, though discrepancies in peak magnitude can occur.

At day five, forecast skill decreases noticeably. While forecasts attempt to identify major wet periods seen in observations, the alignment in timing and magnitude is less precise. The ensemble mean and deterministic forecasts often appear smoother and can fail to capture the full intensity of observed peaks. The ensemble spread is significantly wider than at lead day one, visually representing the increased forecast uncertainty at this longer lead time.

At day ten, forecasts (ensemble mean and deterministic) show a weak relationship with the observed precipitation. Both struggle significantly to predict the specific timing and magnitude of rainfall events, often showing low-amplitude variations or missing observed events entirely. The ensemble spread is very wide, frequently ranging from zero to potentially significant precipitation amounts, indicating low predictability and high uncertainty for specific rainfall details ten days ahead.

Figure 3.24 compares temperature forecasts with observed temperature (black line) for Çukurkışla Basin in early 2019. It displays an ensemble prediction system (EPS), showing the 50-member range (blue shaded area) and ensemble mean (blue line), alongside a deterministic forecast (red line). The comparison is presented for forecast lead times of Day 1 (top), Day 5 (middle), and Day 10 (bottom).

Ensemble and deterministic temperature forecasts are compared with observed basin temperature data. The ensemble spread for the first lead day is very narrow, indicating high confidence among ensemble members. At this first lead time, both the ensemble mean, and deterministic forecast track the observed temperature extremely closely, capturing daily fluctuations and trends with high accuracy, but the temperature is underestimated in some months.

At day five, forecast skill remains high, though slightly reduced compared to day one. While forecasts (ensemble mean and deterministic) continue to capture the overall observed temperature trend effectively, they begin to smooth out some finer daily variations. The ensemble spread is wider than at lead day one, representing a moderate increase in forecast uncertainty, but the observed temperature generally remains well within this range.

At lead day ten, forecasts (ensemble mean and deterministic) show a further decrease in skill regarding specific daily values but still successfully capture the main temperature trends over the period. Deviations from observed short-term fluctuations are more apparent. The ensemble spread is wider, reflecting significant uncertainty ten days ahead, yet it consistently encompasses the observed temperature, demonstrating its utility in defining the range of outcomes.

Figure 3.24. *Ensemble and deterministic temperature comparison plots in Çukurkışla in 2019 with the observed in lead times 1, 5, and 10*

Comparing the precipitation (Figures 3.19 and 3.23) and temperature (Figures 3.20 and 3.24) forecasts for Çukurkışla across 2018 and 2019 reveals consistent patterns in forecast behavior and variable predictability. Temperature forecasts consistently demonstrated higher skill and lower uncertainty compared to precipitation forecasts at all lead times (Day 1, 5, and 10) in both years. While both variables showed the expected decrease in forecast accuracy and increase in ensemble spread with longer lead times, the degradation was much more pronounced for precipitation.

3.4.2.2. Kayabaşı

The temperature and precipitation forecasts for Kayabaşı Basin were downloaded at a spatial resolution of 0.25º × 0.25º grids, as shown in Figure 3.25. To ensure data reliability, the forecasted meteorological variables were subjected to a quality control process, where they were compared against observed historical data from ground-based meteorological stations.

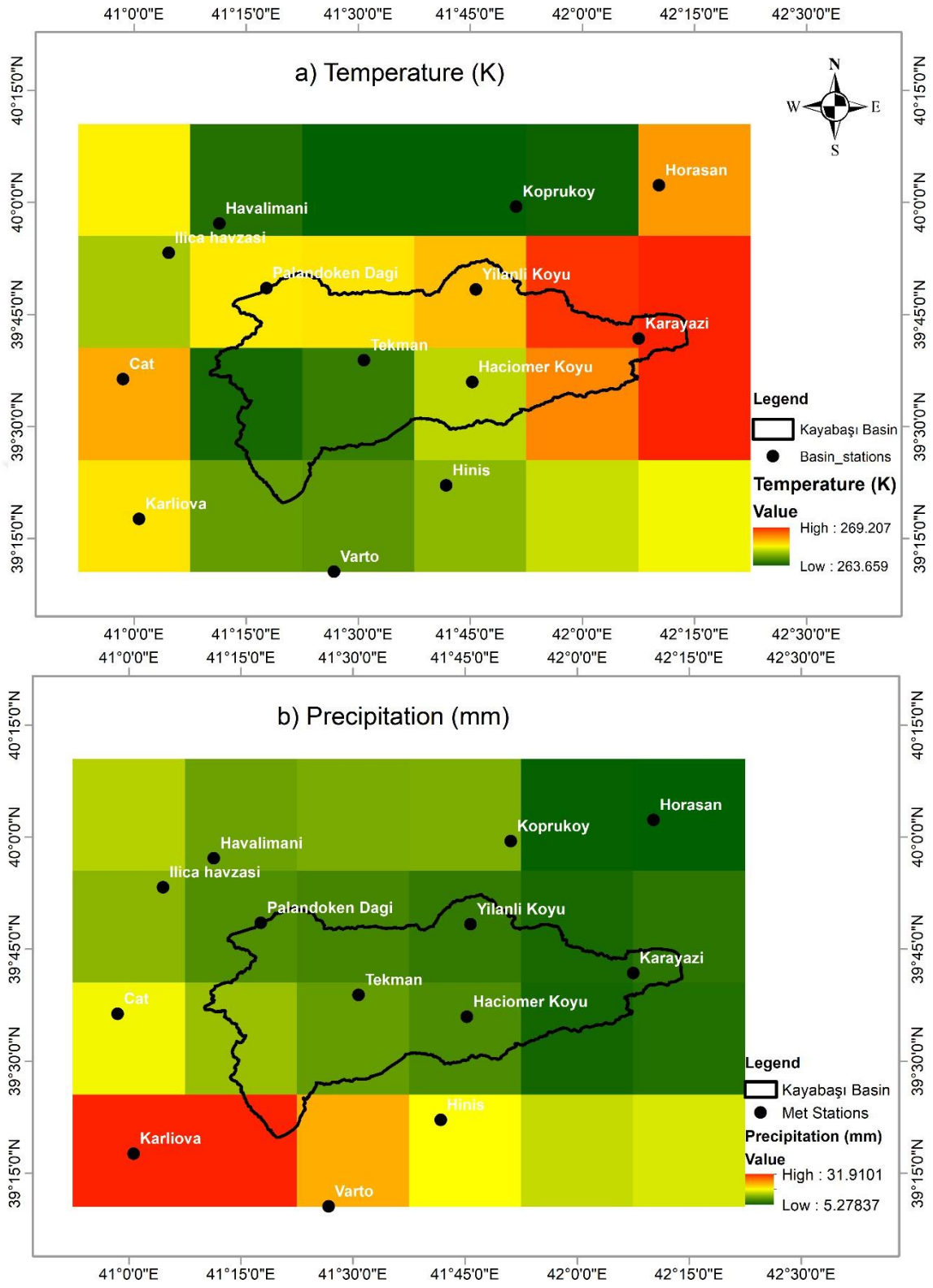

Figure 3.25. *The 0.25º × 0.25º Grids for Kayabaşı Basin in 15/03/2019 a) Temperature (K) b) Precipitation (mm)*

To evaluate the accuracy of precipitation forecasts for Kayabaşı Basin in 2018, May has received the highest precipitation for both observed and forecast. The overall forecast precipitation trend for each month through all the lead days agreed with the trend of the observed precipitation. But there is an overestimation in forecast precipitation for all months. Figure 3.26 highlighted the monthly accumulated precipitation for both observed and forecasted data in 2018.

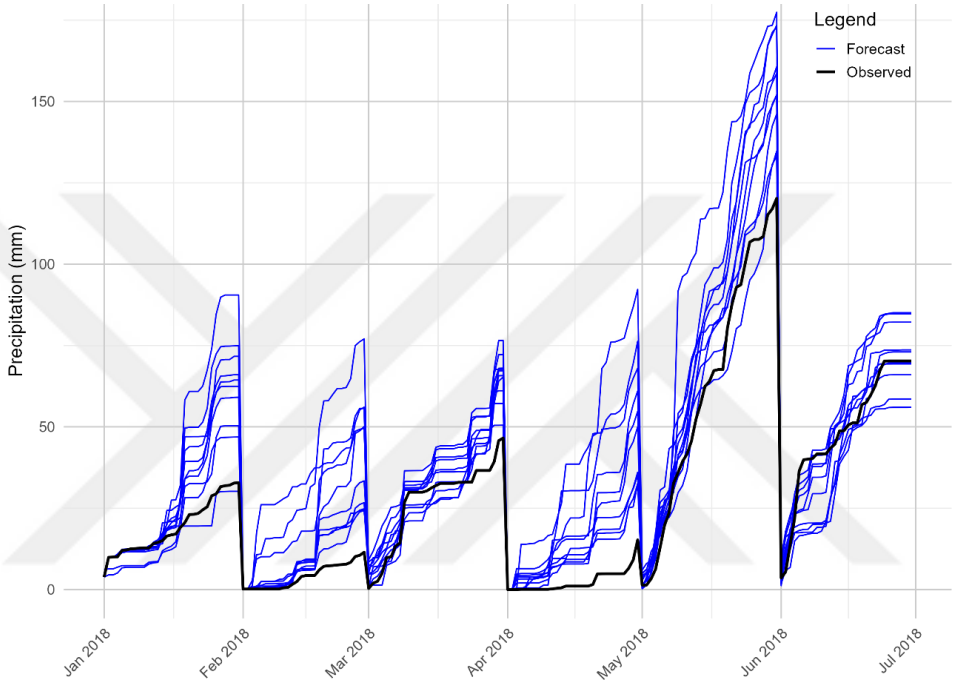

Figure 3.26. *Monthly observed and forecast precipitation accumulation for Kayabaşı Basin in 2018*

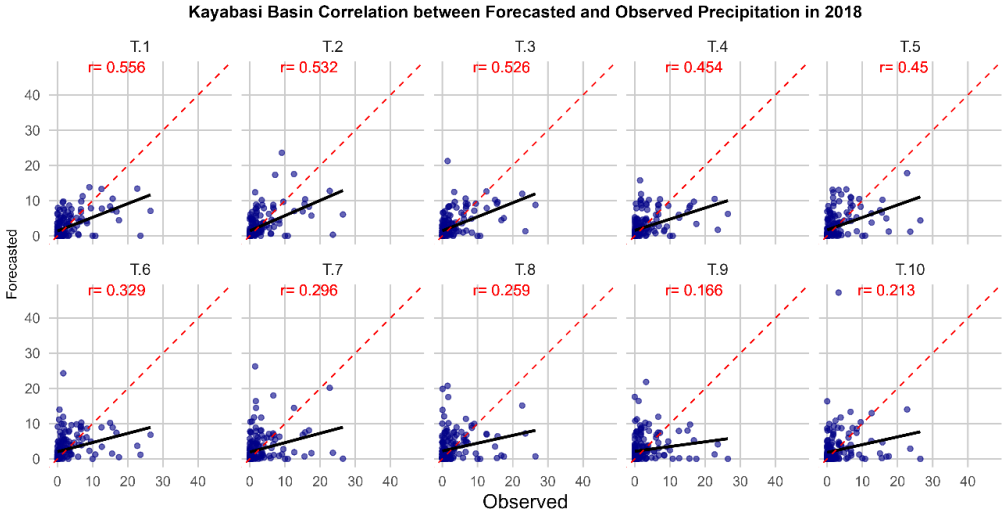

Figure 3.27. *The correlation between observed and forecast precipitation of Kayabaşı Basin through different lead-times in early half of 2018*

The correlation between observed and forecasted precipitation in Kayabaşı Basin in 2018 also declines as lead-time increases, starting at 0.556 on the first lead day and dropping to 0.213 by the last lead day, as illustrated in Figure 3.27.

Figure 3.28 compares precipitation forecasts with observed precipitation (black line) for Kayabaşı Basin in early 2018. It displays an ensemble prediction system (EPS), showing the 50-member range (blue shaded area) and ensemble mean (blue line), alongside a deterministic forecast (red line). The comparison is presented for forecast lead times of Day 1 (top), Day 5 (middle), and Day 10 (bottom).

The ensemble spread for the first lead day is narrow during dry periods but widens during potential precipitation events. At this first lead time, both the ensemble mean, and deterministic forecast capture the timing of observed major precipitation events (e.g., late Jan, mid-April, May-June), although there can be notable differences in forecast versus observed peak.

At day five, forecast skill shows a decrease. While forecasts still attempt to indicate periods of enhanced precipitation probability corresponding to observed wet spells, the specific timing and magnitude alignment is significantly reduced. The ensemble mean and deterministic forecasts often appear much smoother and fail to capture the sharpness and intensity of observed peaks. The ensemble spread is wider than at lead day one, reflecting the substantial increase in forecast uncertainty.

At day ten, forecasts (ensemble mean and deterministic) exhibit a very weak relationship with the observed precipitation. Both forecasts struggle profoundly to predict the specific timing and magnitude of rainfall events, often showing minimal variation or completely missing significant observed rainfall. The ensemble spread is extremely wide, spanning a broad range from zero to potentially high precipitation amounts, indicative of very low predictability and high uncertainty for specific rainfall details ten days ahead.

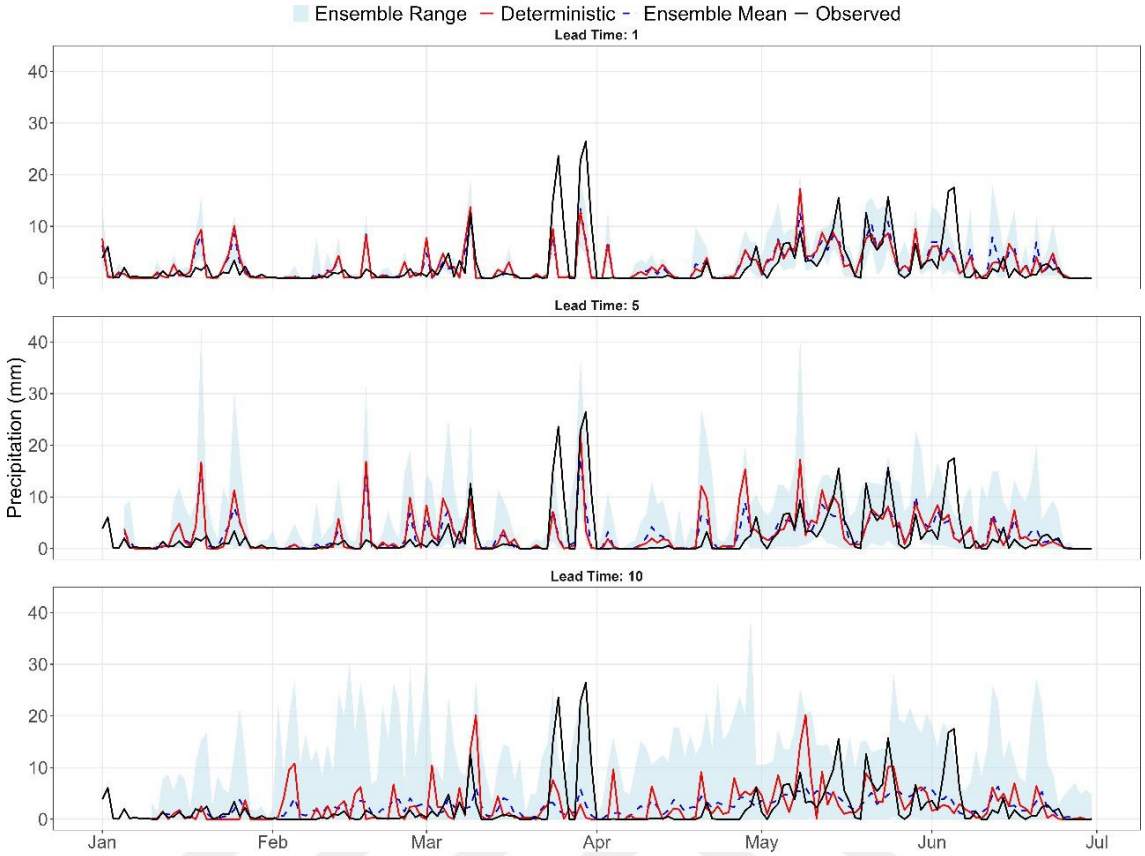

Figure 3.28. *Ensemble and deterministic precipitation comparison plots in Kayabaşı in 2018 with the observed precipitation in lead times 1, 5, and 10*

Figure 3.29 illustrates the comparison of temperature forecasts (ensemble mean/range and deterministic) against observed temperatures for Kayabaşı Basin in early 2019 across lead times of 1, 5, and 10 days. At Day 1, forecasts exhibit high accuracy with a narrow ensemble spread, closely tracking observed daily variations. By Day 5, skill remains strong for overall trends, though some daily fluctuations are smoothed, and the ensemble spread widens moderately, still containing the observed temperatures. At Day 10, while specific daily accuracy decreases and variations are significantly smoothed, the forecasts capture the broader temperature evolution, and the considerably wider ensemble spread reliably encompasses the observed values, reflecting increased uncertainty but providing a useful probabilistic range.

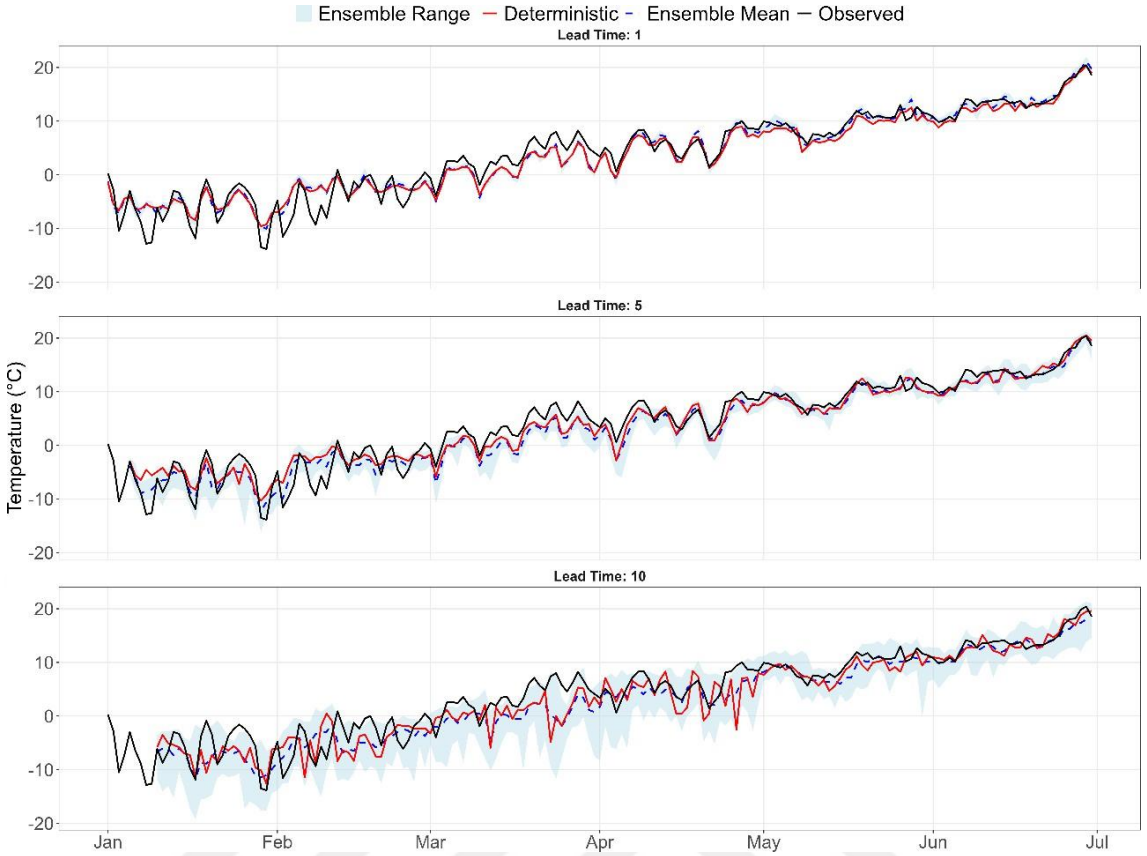

Figure 3.29. *Ensemble and deterministic temperature comparison plots in Kayabaşı in 2018 with the observed temperature in lead times 1, 5, and 10*

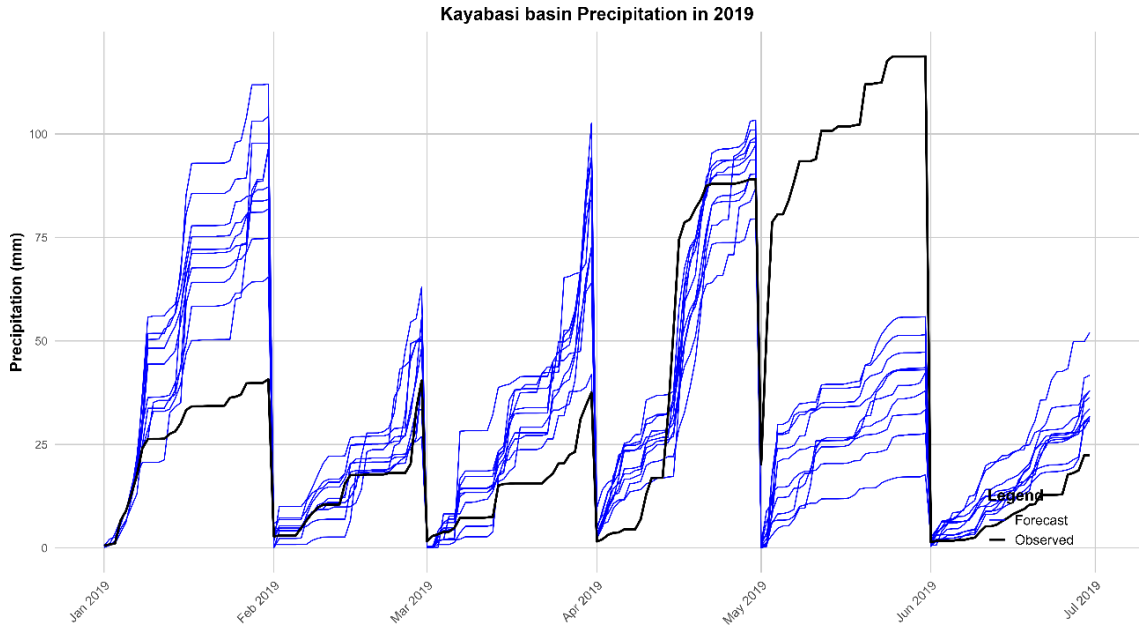

Figure 3.30. *Monthly observed and forecast precipitation accumulation for Kayabaşı Basin in 2019.*

To evaluate the accuracy of precipitation forecasts for Kayabaşı Basin in 2019, we analyzed observed and forecasted precipitation. The highest observed precipitation occurred in May, whereas the forecast predicted January as the wettest month. The overall forecast precipitation trend for each month through all the lead days agreed with the trend of the observed precipitation. But there is an overestimation in forecast precipitation for all months except April. Figure 3.30 highlighted the monthly accumulated precipitation for both observed and forecasted data in 2018. The correlation between the observed and forecasted precipitation in 2019 is shown in Figure 3.31. The relation decreases as the lead time increases, from 0.454 to 0.058.

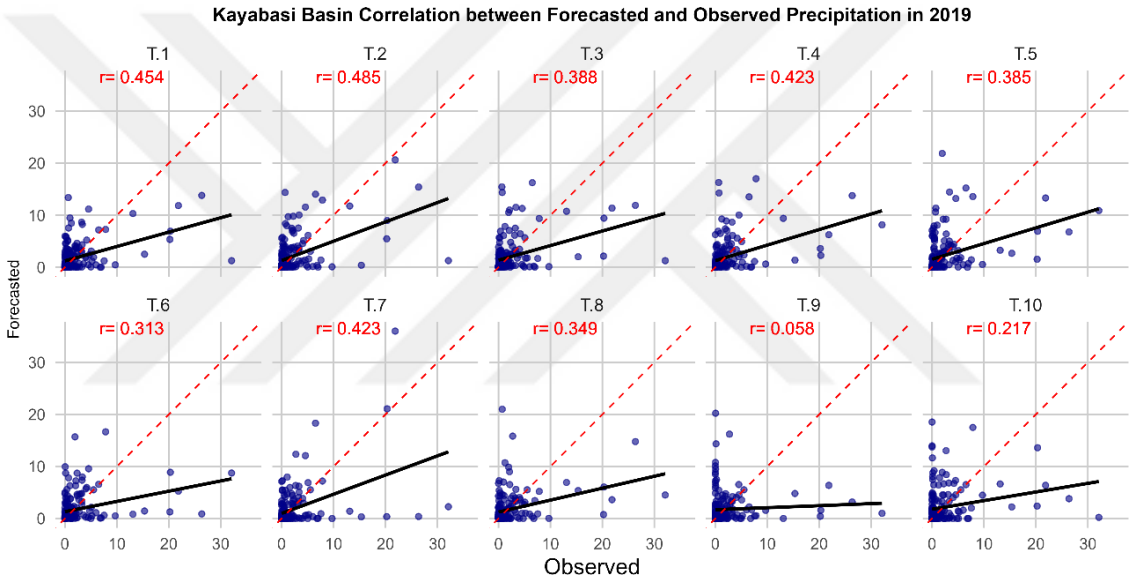

Figure 3.31. *The correlation between observed and forecast precipitation of Kayabaşı Basin through different lead-times in early half of 2019*

Figure 3.32 compares ensemble (mean and range) and deterministic precipitation forecasts against observations for Kayabaşı Basin during early 2019, evaluated across lead times of 1, 5, and 10 days. At the 1-day lead time, forecasts demonstrate moderate skill in capturing the timing of precipitation events, albeit with inaccuracies in peak magnitude, and exhibit relatively narrow ensemble spread except during potential rainfall. Forecast skill markedly diminishes by Day 5, characterized by reduced precision in timing and magnitude, failing to capture observed peaks, and a significantly wider ensemble spread reflecting increased uncertainty. By Day 10, predictability is very low, with both ensemble mean, and deterministic forecasts showing a weak correlation to

observed precipitation, struggling to predict event timing or intensity, and accompanied by an extremely wide ensemble spread indicative of high forecast uncertainty.

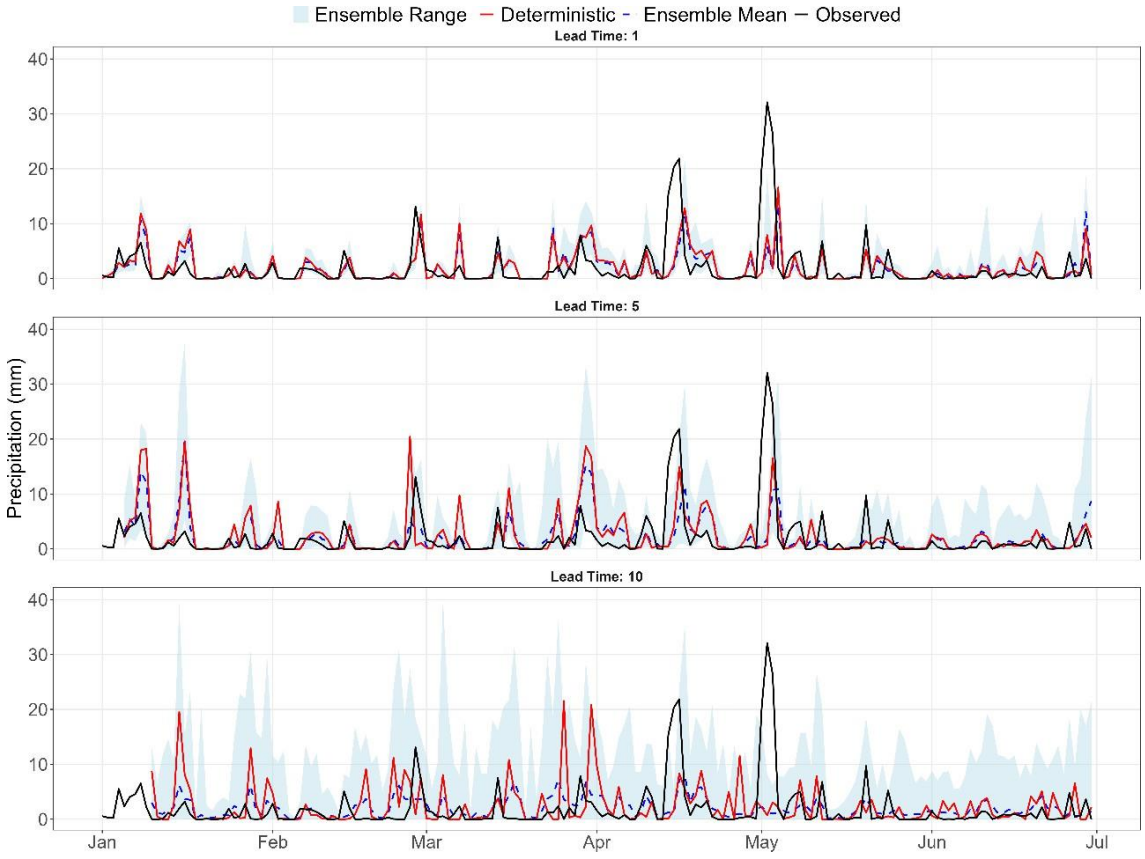

Figure 3.32. *Ensemble and deterministic precipitation comparison plots in Kayabaşı in 2019 with the observed precipitation in lead times 1, 5, and 10*

Figure 3.33 shows the comparison of ensemble and deterministic temperature forecasts against observations for Kayabaşı Basin in early 2019, across lead times of 1, 5, and 10 days. At the 1-day lead time, forecasts demonstrate exceptional accuracy with a very narrow ensemble spread, closely matching both trends and daily variations in observed temperature. While forecast skill remains high at Day 5, capturing major trends accurately, a slight decrease in fidelity for sharp daily fluctuations is noted alongside a moderate widening of the ensemble spread, which consistently encompasses observed values. By Day 10, Although accuracy decreased, it shows the overall trend, making it useful despite the uncertainty.

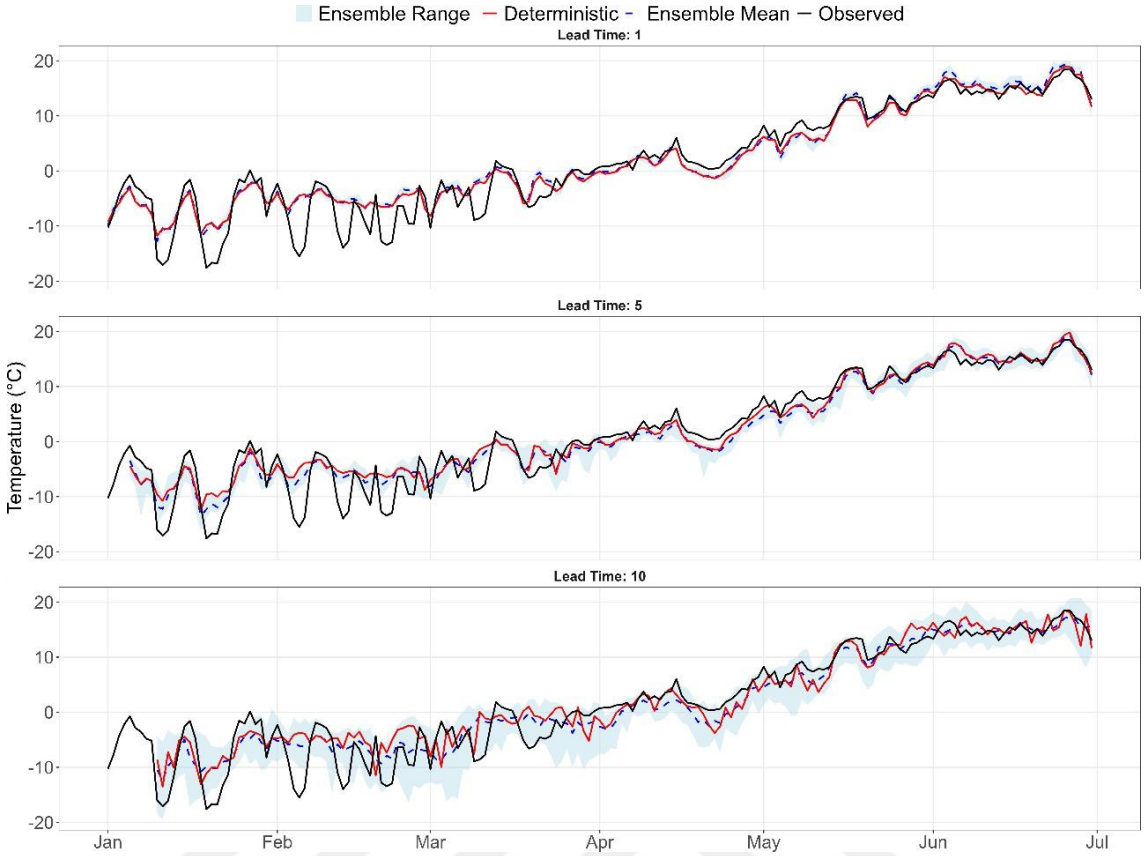

Figure 3.33. *Ensemble and deterministic temperature comparison plots in Kayabaşı in 2019 with the observed temperature in lead times 1, 5, and 10*

3.5. HOOPLA Toolbox

To facilitate the exploration of ensemble rainfall-runoff modelling, uncertainty analysis, and prediction within this study, the HydrOlOgical Prediction LAboratory (HOOPLA) toolbox was employed. Developed at Université Laval, Canada, HOOPLA is a modular framework coded entirely in MATLAB, designed for ensemble lumped hydrological modelling (Thiboult, 2019). HOOPLA includes a collection of 20 lumped conceptual hydrological models. These models simulate rainfall-runoff transformation using interconnected reservoirs (buckets) and assume spatial homogeneity within the catchment. The selection, originating from the work of Perrin (2000) and later refined by Seiller et al. (2012), focused on models with low to moderate complexity (relatively few free parameters and reservoirs) while ensuring structural diversity, as they were developed by different teams for various purposes (Thiboult, 2019).

It is crucial to emphasize that the models implemented in HOOPLA are generally not the original versions. They were intentionally modified, primarily by Perrin

(2000) and within the HOOPLA development, to fit a common lumped framework. Modifications may include:

- Conversion of distributed models to lumped representations.
- Reduction in the number of free parameters.
- Removal of integrated PET or snow accounting routines (handled externally within HOOPLA).
- Standardization of inputs (precipitation + snowmelt, PET) and outputs (streamflow).

Table 3.3 summarizes the main characteristics of the hydrological models available in HOOPLA, based on information provided in the technical report. The 'Modification' column gives a qualitative indication of the extent of modification from the original reference.

Table 3.3. *Main characteristics of the 20 lumped hydrological models in HOOPLA (Thiboult, 2019).*

Model Name	No of Parameters	No of Reservoirs	Modification	Derived From
HydroMod1	6	3	Slightly	BUCKET (Thornthwaite and Mather, 1955)
HydroMod2	9	2	Slightly	CEQUEAU (Girard et al., 1972)
HydroMod3	6	3	Slightly	CREC (Cormary and Guilbot, 1973)
HydroMod4	6	2	Slightly	GARDENIA (Thiery, 1982)
HydroMod5	4	2	Similar	GR4H (Mathevet, 2005)
HydroMod6	8	4	Substantial	HBV (Bergström and Forsman, 1973))
HydroMod7	6	3	Slightly	HYMOD (Wagener et al., 2001)
HydroMod8	9	4	Slightly	IHACRES (Jakeman et al., 1990)
HydroMod9	6	2	Slightly	MARTINE (Mazenc et al., 1984)
HydroMod10	7	4	Similar	MOHYSE (Fortin and Turcotte, 2007)
HydroMod11	6	4	Similar	MORDOR (Garçon, 1999)
HydroMod12	10	5	Substantial	NAM (Nielsen and Hansen, 1973)
HydroMod13	8	4	Slightly	PDM (Moore and Clarke, 1981)
HydroMod14	9	5	Slightly	SACRAMENTO (Burnash et al., 1973)
HydroMod15	8	3	Substantial	SIMHYD (Chiew et al., 2002)
HydroMod16	8	4	Substantial	SMAR (O'Connell et al., 1970)
HydroMod17	7	3	Substantial	TANK (Sugawara, 1979)
HydroMod18	7	3	Substantial	TOPMODEL (Beven et al., 1984)
HydroMod19	8	3	Substantial	WAGENINGEN (Warmerdam et al., 1997)
HydroMod20	8	4	Substantial	XINANJIANG (Zhao et al., 1980)

3.5.1. Potential evapotranspiration

Potential evapotranspiration quantifies the atmospheric demand for water vapor and serves as a key input to the hydrological models. HOOPLA includes three PET formulas,

selected primarily because hydrological models often show limited sensitivity to PET inputs and allow computation even with limited data availability (Seiller and and Anctil, 2016; Thiboult, 2019). These formulas can be computed internally by HOOPLA or provided externally. Table 3.4 details the included formulas and their data requirements. In this study, the methodology proposed by Oudin et al. (2005) was employed to calculate PET for the study areas, utilizing the mean temperature and the latitude of these regions.

Table 3.4. *Potential Evapotranspiration (PET) formulas in HOOPLA (Thiboult, 2019).*

Name	Required input data	Basis	Reference
Oudin	Mean Temperature (T), Latitude (Lat)	Energy Balance	Oudin et al. (2005)
Kharrufa	Mean Temperature (T), Latitude (Lat)	Energy Balance	Kharrufa (1985)
Penman	Mean T, Lat, Solar Rad. (Rad), Relative Humidity (Relhum), Max Temp (Tmax), Min Temp (Tmin), Wind Speed (Wndspd), Elevation (z)	Combination Method	Penman (1948)

3.5.2. Snow accounting routine

HOOPLA incorporates the CemaNeige snow accounting routine (SAR) (Valéry et al., 2014). Currently, this is the only SAR included in this toolbox. CemaNeige operates based on:

- Spatial discretization into five altitudinal bands of equal area.
- Temperature and precipitation extrapolation to these bands.
- Partitioning of precipitation into liquid/solid fractions based on a transition temperature range.
- Estimation of snow water equivalent (SWE) using:
    - A thermal inertia factor for the snowpack (Ctg).
    - A degree-day melt factor (Kf).

Table 3.5. *Snow Accounting Routine (SAR) in HOOPLA.*

Name	Description	Free Parameters	Reference
CemaNeige	Simulates snow accumulation and melt using altitudinal bands, thermal state, and degree-day factor.	Ctg, Kf	(Valéry et al., 2014)

3.5.3. Calibration algorithms

HOOPLA provides two global, iterative calibration algorithms:

1. **Shuffled Complex Evolution (SCE):** Developed by (Duan et al., 1993), SCE uses multiple sets of points (complexes) that evolve through reflection, contraction, and mutation, and are periodically shuffled to explore the parameter space. It is often considered state-of-the-art but can be computationally expensive.
2. **Dynamically Dimensioned Search (DDS):** Proposed by Tolson and Shoemaker (2007), DDS is inspired by manual calibration. It samples the parameter space via random perturbations and reflections, progressively fixing parameter values to reduce the search dimension. It is designed to find good solutions within a specified maximum number of iterations, making it computationally efficient, especially for models with many parameters.

In this study, SCE is used to calibrate the hydrological parameters to identify optimal hydrological model parameter sets.

3.5.4. Data assimilation schemes

DA techniques are included to improve forecast accuracy by integrating information from observations to update model states and provide better initial conditions. HOOPLA includes two probabilistic, sequential DA schemes:

1. **Sequential Importance Resampling (SIR) Filter:** A type of PF, also known as a Bootstrap filter (Arulampalam et al., 2002). It approximates the probability distribution of model states using weighted samples (particles), which are updated and resampled based on observations.
2. **Ensemble Kalman Filter (EnKF):** Described by Evensen (2003), the EnKF uses an ensemble of model states to estimate the error covariances needed for the Kalman update step. It avoids the linearity assumption of the traditional Kalman Filter but performs optimally under Gaussian assumptions.

Based on recommendations on (Liu et al., 2012; Thiboult and Anctil, 2015), the HOOPLA model utilizes default parameters to define the standard deviation of the normal distribution representing uncertainty around key model inputs. These parameters quantify the assumed errors associated with observed or forcing data used in the model. These default input perturbation standard deviations are summarized in Table 3.6.

Table 3.6. *Default model input perturbation standard deviations for HOOPLA*

Variable	Perturbation Standard Deviation
Discharge	10%
Precipitation	50%
Air Temperature	2 °C
Potential Evapotranspiration	10 %

As shown in Table 3.6, the perturbations were defined as the standard deviation of a normal distribution applied to each input variable. Specifically, standard deviations of 10% were applied to Discharge and Potential Evapotranspiration, 50% to Precipitation, and 2 °C to Air Temperature. These defined uncertainties were used to generate perturbed ensembles of model inputs for the DA procedure. EnKF was then employed for DA, utilizing an ensemble size of 50 members to estimate the state and parameter uncertainty.

3.6. Multi-model Combination

As we discussed in the literature, different methods of hydrological models' results combinations are presented in different studies (Ajami et al., 2006; Georgakakos et al., 2004; Shamseldin et al., 1997; Todorović et al., 2024). Simple Averaging method (SAM) which involves the averaging of the discharges of the models is used in this study.

$$Q_{SAM} = \frac{\sum_{i}^{n} Q_{sim}}{n} \tag{3.5}$$

Where Q_{SAM} is the averaged discharge, Q_{sim} is the simulated discharges, n is the number of the models.

4. RESULTS AND DISCUSSION

4.1. Calibration and Validation of Hydrological Models

4.1.1. Çukurkışla basin

Calibration

The meteorological and discharge data were split into two parts for training and validation. The 2000-2008 water years were used for calibration and 2009-2017 years for validation. Using the HOOPLA toolbox in MATLAB, the twenty hydrological models were run and calibrated using the SCE-AU and KGE to maximize their performance according to the observed discharge. The CemaNeige module was employed as the SAR in all twenty hydrological models to simulate snow processes. The hydrograph in Figure 4.1 shows that all the hydrological model simulations followed the pattern of the observed flow. In addition, the simulation flows captured the high and low flows of the basin throughout the calibrated period.

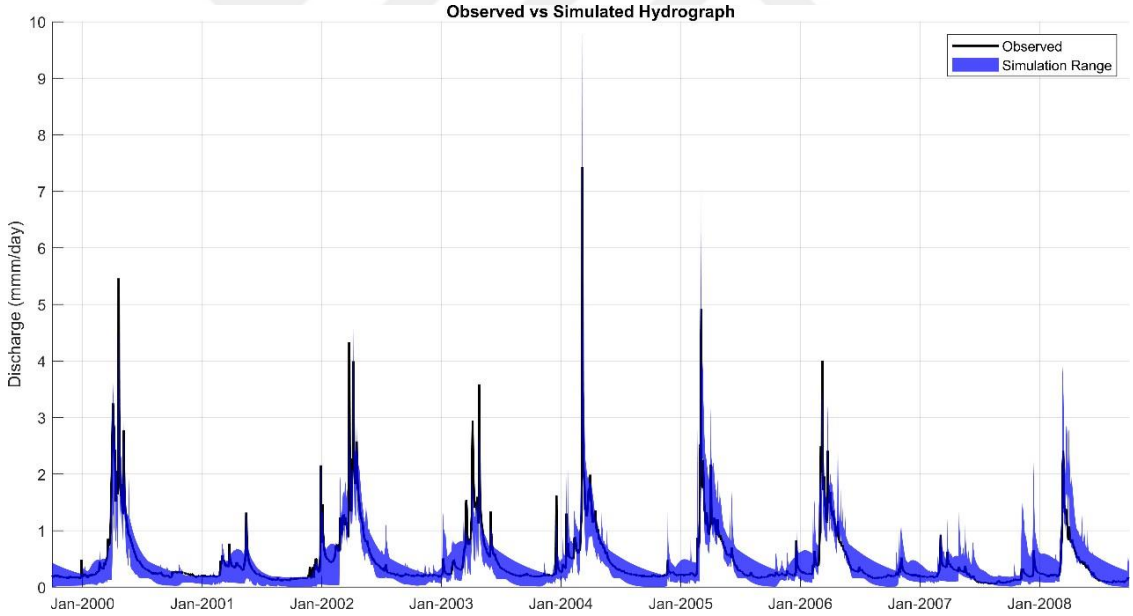

Figure 4.1. *Hydrographs of observed and simulated discharges from 20 hydrological models during the calibration period (water years 2000-2008) in Çukurkışla Basin*

Figure 4.2 shows the result of model 3 which is one of the best performed models in Çukurkışla basin for this calibration period. The top panel shows the observed temperature, illustrating clear seasonal variation with a freezing period in winter. The middle panel displays observed precipitation, observed discharge, and simulated discharge. The lower panel represents the accumulated snowpack. The hydrograph shows

that the simulation followed the observed, capturing the lower and peak flows in multiple years. In addition, the simulated snowpack also exhibited a good relation between temperature, precipitation, and simulated discharges. During winter, low temperatures result in substantial snow accumulation and freezing of liquid precipitation. Besides that, when the temperature rises, around early March, snow melt commences, highly contributing to the discharge. Overall, the simulation indicates that Model 3 provides a reasonable representation of the basin's hydrological behavior, particularly concerning snowmelt processes.

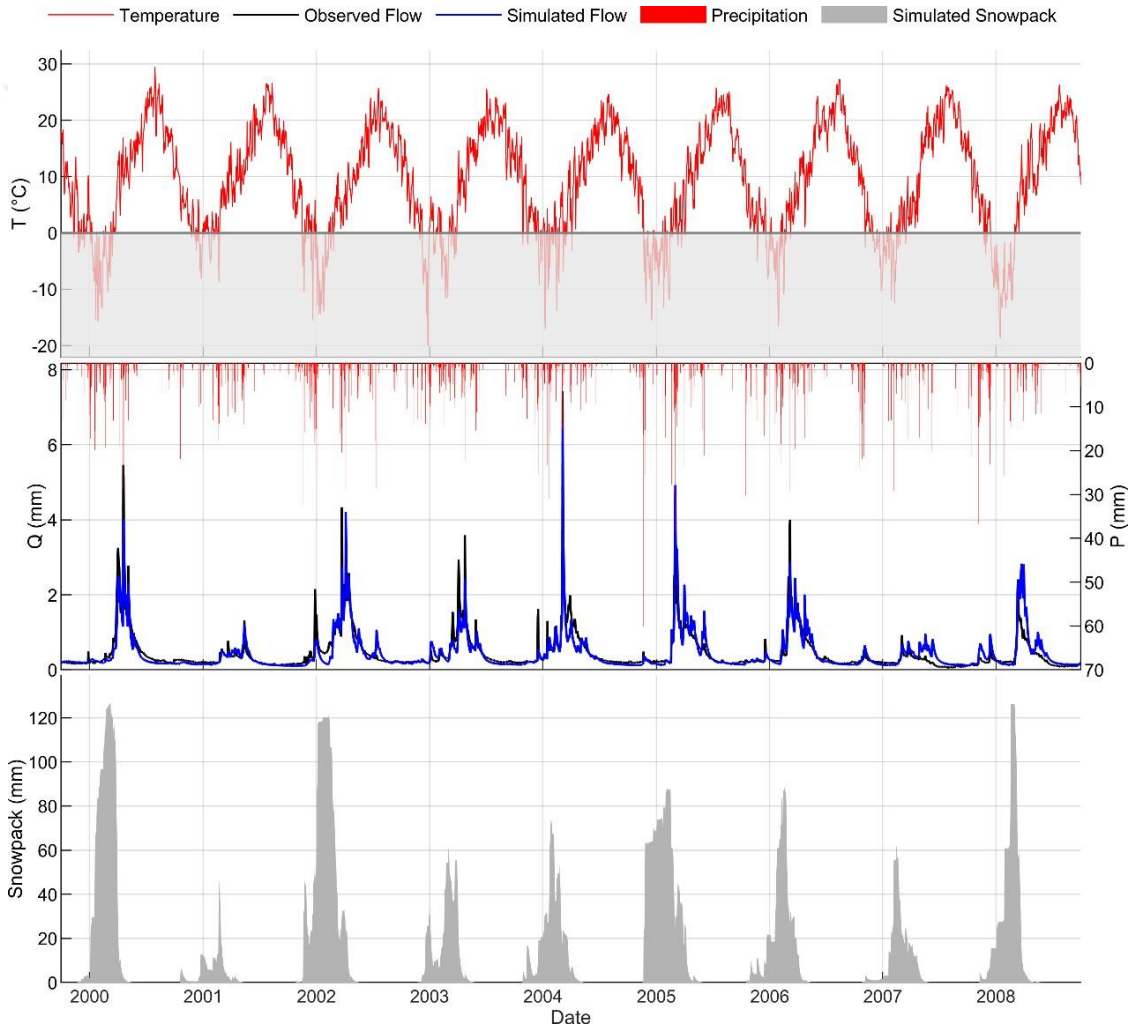

Figure 4.2. *Model 3 Results over Çukurkışla Basin for 2000-2008 calibration period: observed and simulated temperature, discharge hydrographs, precipitation, and snowpack*

Four different goodness of fit metrics are used to show the hydrological models' performance: NSE, NSEsqrt, KGE, and r^2, as shown in Figure 4.3. KGE exhibited a high median and a narrow interquartile range, suggesting that the models achieved a good

balance between correlation, bias, and variability. The NSE and r² metrics also indicated strong model performance, though with slightly greater variability between the models compared to KGE. In contrast, NSEsqrt yielded the lowest overall scores and the widest spread in the boxplot. NSEsqrt metric shows models struggle most with accurately simulating low-flow dynamics.

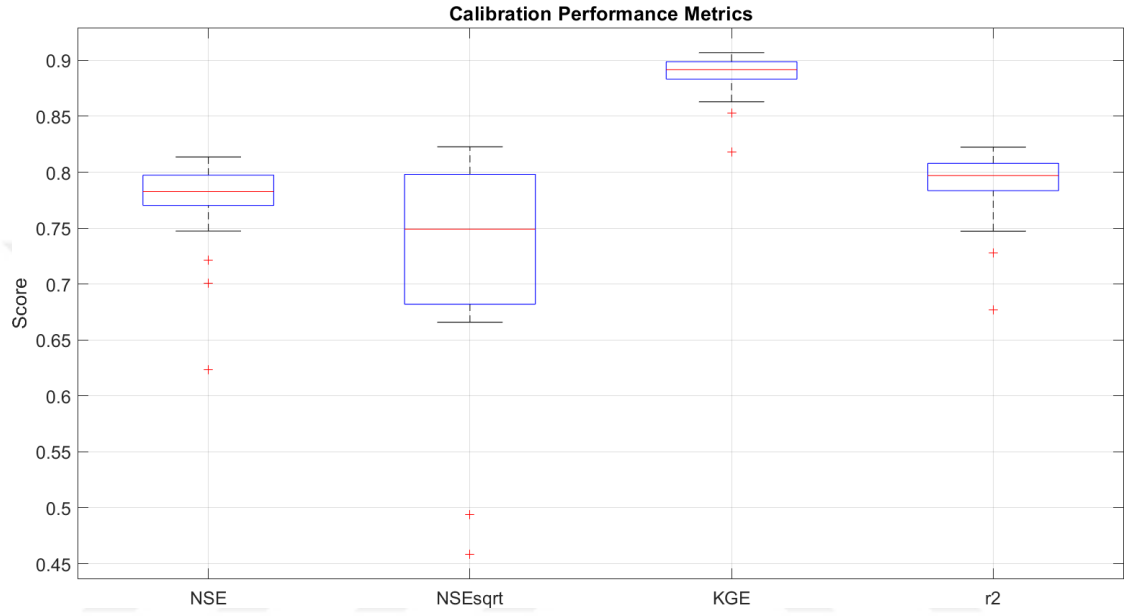

Figure 4.3. *Boxplots of calibration performance metrics for the hydrological models in Çukurkışla Basin using four goodness-of-fit (GOF) metrices: NSE, NSEsqrt, KGE, and r^2*

Table 4.1 provides the summary statistics (minimum, first quartile [Q1], median, third quartile [Q3], and maximum) for each metric, offering further insight into the central tendency and spread of the performance scores across models.

Table 4.1. *Summary of calibration performance metrics for NSE, NSEsqrt, KGE, and r² across all models over Çukurkışla Basin*

Quartiles	**Calibration**			
	NSE	**NSEsqrt**	**KGE**	**r^2**
Max	0.814	0.823	0.907	0.822
Q3 (0.75)	0.797	0.798	0.899	0.808
Median (0.5)	0.783	0.749	0.892	0.797
Q1 (0.25)	0.770	0.682	0.883	0.784
Min	0.624	0.459	0.818	0.677

All the hydrological models have shown strong performance. The median KGE value obtained is 0.89, demonstrating high consistency between simulated and observed streamflow. As Figure 4.4 shows, Model 3 exhibited the highest performance with a KGE exceeding 0.906, followed closely by several models. Models 4 and 19 showed comparatively lower performance, with KGE values falling between 0.80 and 0.85, suggesting some limitations in accurately reproducing observed dynamics during the calibration period. Despite these few outliers, the overall results highlight the robustness of most of the models and suggest their suitability for further hydrological analysis.

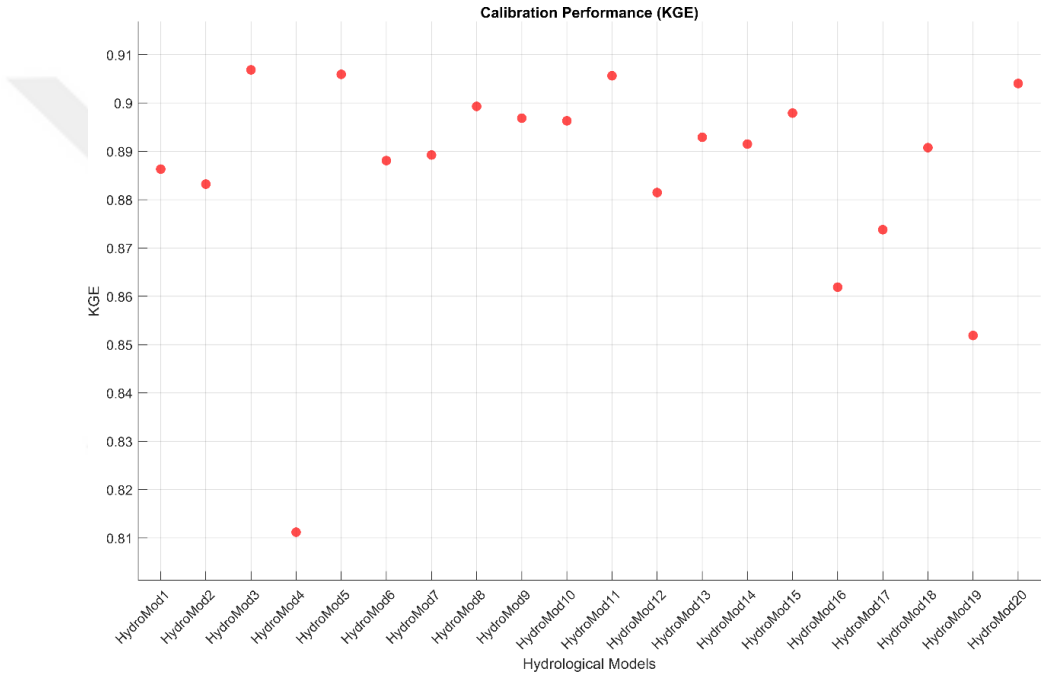

Figure 4.4. *KGE performance during the 2000-2008 calibration period over Çukurkışla Basin*

Validation

The models were validated over the 2009-2017 period using their calibrated parameters. The performance was assessed using the previously used metric; NSE, NSEsqrt, KGE, and r^2, to evaluate the models' ability to reproduce observed streamflow during the validation period. The hydrograph in Figure 4.5 shows that the simulated streamflow successfully captures the general patterns of observed streamflow, including most major flood peaks and recession limbs, though with considerable spread, particularly during high flow events. The simulated streamflow for the 2013 and 2014 water years exhibited underestimation, due to the models' inability to accurately capture peak flow events.

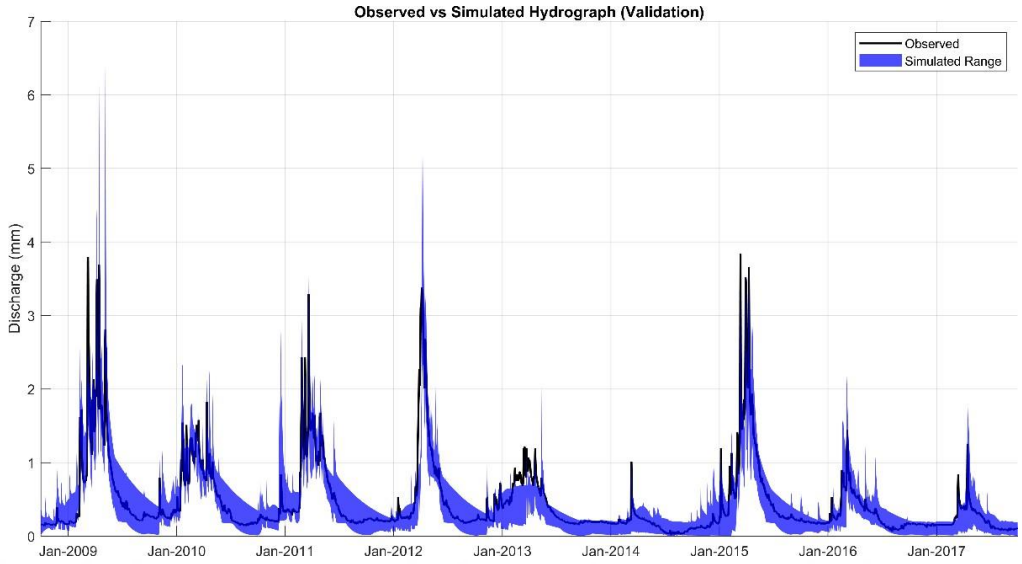

Figure 4.5. *Hydrographs of observed and simulated discharges over Çukurkışla Basin from 20 hydrological models during the validation period (2009-2017)*

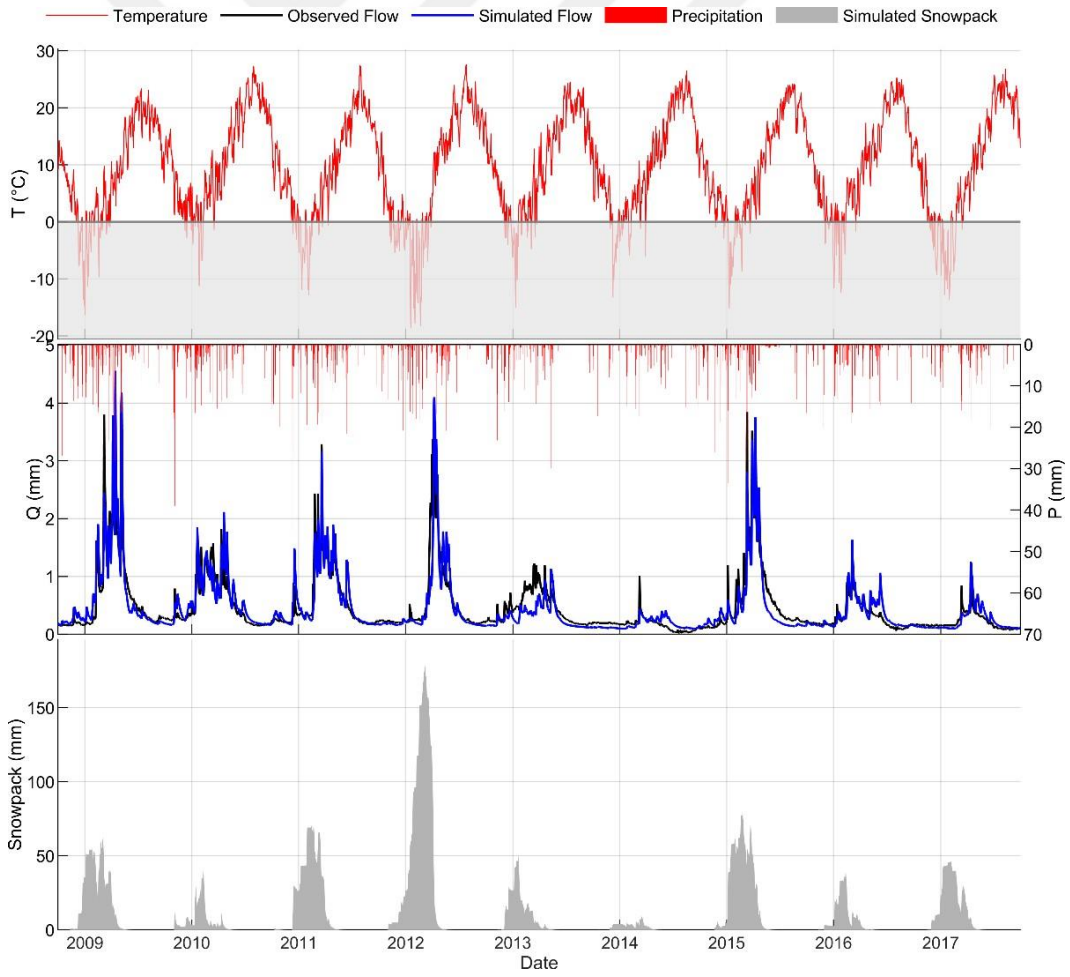

Figure 4.6. *Model 3 Results over Çukurkışla Basin for 2009-2017 validation period: observed and simulated temperature, discharge hydrographs, precipitation, and snowpack.*

Figure 4.6 shows model 3 results, one of the best performing models in Çukurkışla Basin for calibration and validation periods. Observed temperature, discharge, simulated discharge and snowpack are shown. The simulated discharge performed well for most years except 2013 and 2014. This model captured low and high streamflow effectively. The snowpack for the validation period is lower than 100 mm except 2012 water year, which experienced the highest observed snowpack due to very low temperatures. The lowest snowpack and discharge were simulated in 2014. The simulated snowpack showed a good relation between temperature, precipitation and simulated discharges. During winter, low temperatures result in snow accumulation and freezing precipitation. When the temperature rises, around early March, snowmelt begins, contributing significantly to discharge. Overall, Model 3 provides a credible representation of the basin's hydrological behavior, particularly regarding snowmelt processes.

Figure 4.7 shows the comparison of the four different metrics: NSE, NSEsqrt, KGE, and r^2. Like the calibration, KGE outperformed the other metrics in higher performance and reduced dispersion. NSE and r^2 have also shown better performance, though with slightly greater variability compared to KGE. In contrast, NSEsqrt yielded the lowest performance, characterized by higher dispersion. Table 4.2 gives the summary of the performance metrics.

Figure 4.7. *Boxplots of validation performance metrics for the hydrological models in Çukurkışla Basin using four goodness-of-fit (GOF) metrices: NSE, NSEsqrt, KGE, and* r^2

Table 4.2. *Summary of validation performance metrics for NSE, NSEsqrt, KGE, and r² across all models over Çukurkışla Basin*

Quartiles	**Validation**			
	NSE	**NSEsqrt**	**KGE**	**r^2**
Max	0.848	0.851	0.901	0.854
Q3 (0.75)	0.821	0.816	0.885	0.833
Median (0.5)	0.801	0.783	0.865	0.811
Q1 (0.25)	0.767	0.716	0.852	0.788
Min	0.694	0.498	0.773	0.742

Figure 4.8 shows the performance of each model in KGE metric. The median performance for the twenty models is 0.865, showing a good performance. Model 3 outperformed the other models with 0.907 KGE performance. The lowest performance were observed on Model 12, with 0.773 KGE performance.

Overall, the simulated streamflow demonstrated satisfactory performance, closely aligning with observed data across most years within the validation period. The models effectively reproduced the general flow patterns, including seasonal variations and baseflow conditions, indicating their robustness for hydrological applications.

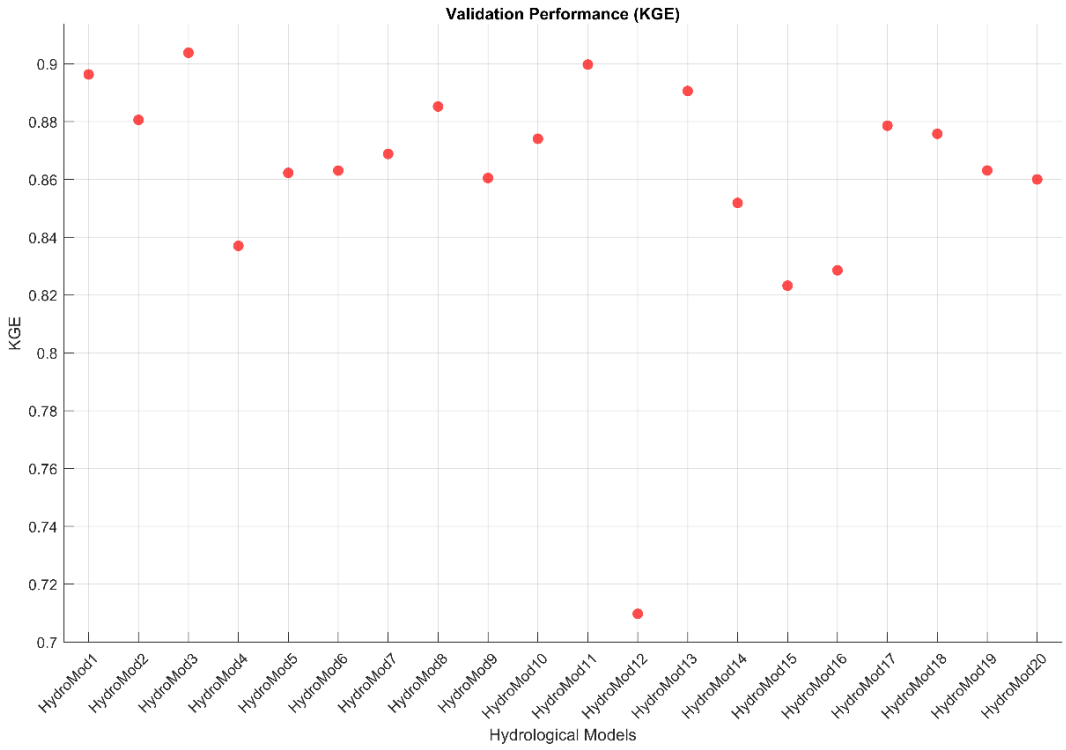

Figure 4.8. *KGE performance over Çukurkışla Basin during the 2009-2017 validation period.*

Physical parameters

The twenty hydrological models each contained between four to 10 core parameters, in addition to two parameters from the CemaNeige snow module. To illustrate how the models represent physical processes, the field capacity (FC) or soil capacity parameter from each model and the degree-day factor (Kf) from CemaNeige were selected for closer examination. These two parameters were chosen because they are directly linked to observable hydrological and snowmelt behavior, making them meaningful for interpreting the models' physical behavior.

The FC parameters for the models show diversity, 50% are in between 100 mm and 450 mm as shown in Figure 4.9. The median FC is 140 mm. While some models, such as models 16, show very high FC values around 900 mm, others like models 3, 9, 12, and 19 calibrate to much lower values, below 100 mm. Several models fall within a moderate range, with a noticeable cluster between 100 and 600 mm. This widespread indicates that the calibrated FC values are strongly influenced by each model's structure and conceptual design, reflecting differing requirements to match observed data effectively.

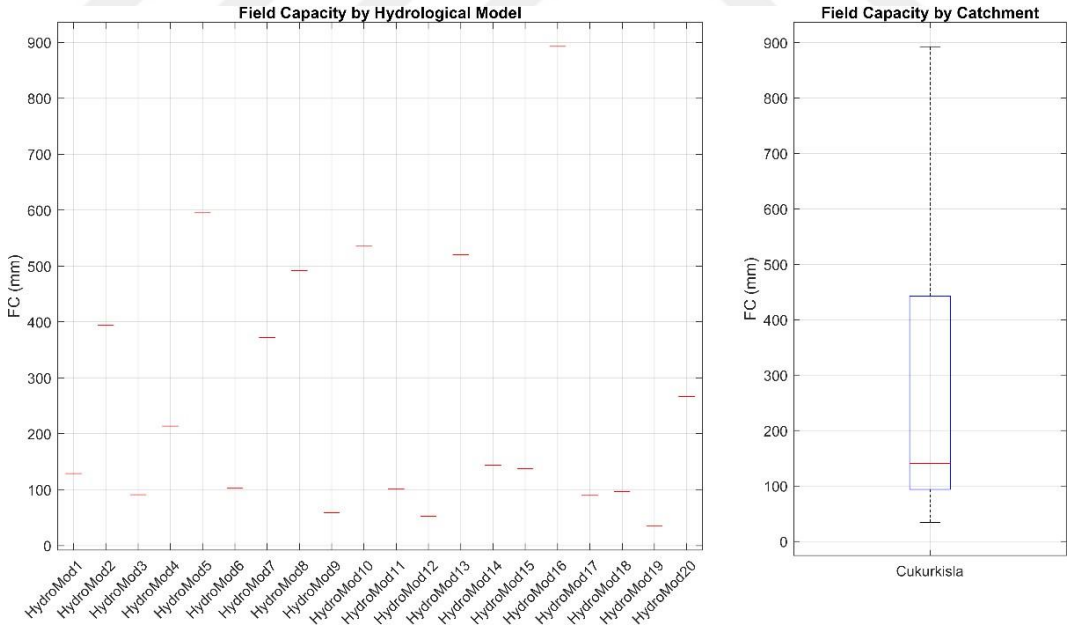

Figure 4.9. *Soil capacity for the twenty models for 2000–2008 calibration period over Çukurkışla*

Based on Figure 4.10, the calibrated *Kf* values show substantial variability across the twenty hydrological models. *Kf* ranges from a minimum of approximately 4.5 mm/°C/day (Model 9) to a maximum of 13.23 mm/°C/day (Model 17). Such high value

likely reflects the calibration process compensating for the specific structural representation of snowmelt processes within Model 17. The median value for the basin is 7.57 mm/°C/day. Although only six models exhibit *Kf* values below 6 mm/°C/day, the interquartile range spans from 5.5 to 11.7 mm/°C/day. These results highlight notable differences in snowmelt sensitivity among the models.

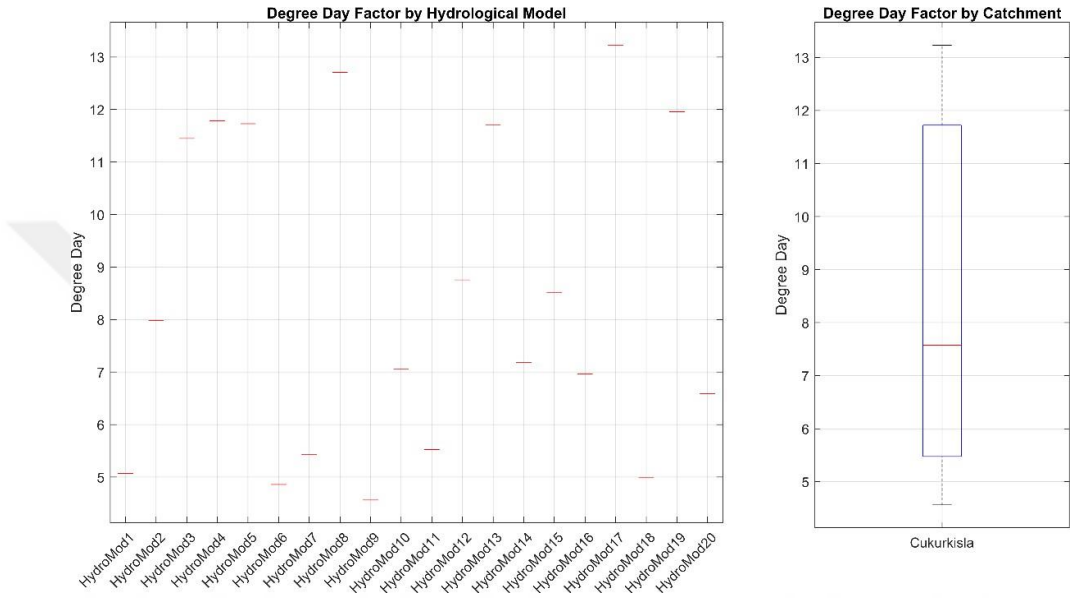

Figure 4.10. *The change of degree day factor parameter of CemaNeige SAR model for the twenty models for 2000-2008 calibration period for Çukurkışla Basin.*

4.1.2. Kayabaşı basin

Calibration

The available meteorological and discharge are split into two parts for training and validation. 2013-2017 water years are used for calibration and 2008-2012 for validation. Using the HOOPLA toolbox in MATLAB, the twenty hydrological models were run and calibrated using the SCE algorithm to maximize their performance according to the observed discharge. The CemaNeige module was employed as the SAR in all twenty hydrological models to simulate snow processes. The hydrograph in Figure 4.11 shows that all the hydrological models' simulations follow the pattern of the observed flow. In addition, the simulation flows capture the low flows of the basin throughout the calibrated period and struggle the peak flows. 2013 and 2014 water years simulated streamflow did not capture the peak flows while the low flow simulation of the most models performed better.

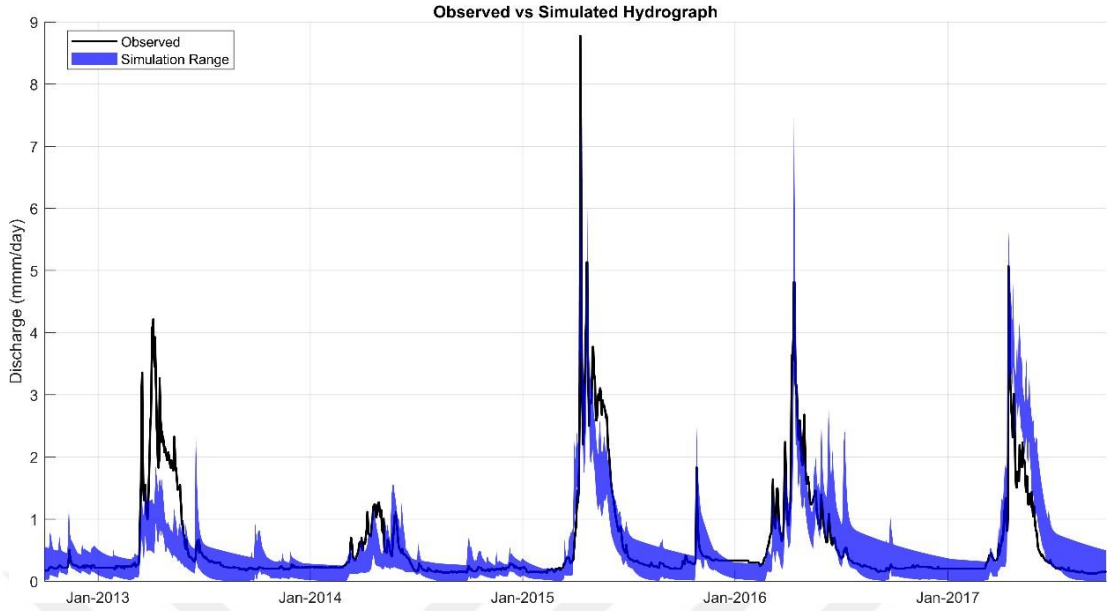

Figure 4.11. *Hydrographs of observed and simulated discharges from 20 hydrological models during the calibration period (water years 2013-2017) in Kayabaşı Basin.*

Figure 4.12 shows the result of Model 8, which is one of the best performed models in Kayabaşı basin for this training period, 2013-2017. While the top panels show the observed temperature of the basin, the middle panel shows the observed and simulated discharges. In addition, the lower panel shows the simulated snowpack of the basin. As we illustrated before, the simulated discharge performed very good for most years except 2013 and 2014 water years. Although the simulated streamflow captured very well in the low flows, it could not capture the high flows of these years. The snowpack for the calibration period is between 50 mm and 200 mm. The lowest snowpack was seen in 2013 and 2014 similar to the streamflow. In addition, the simulated snowpack also exhibited very good relation between temperature, precipitation and simulated discharges. During winter, low temperatures result in substantial snow accumulation and freezing of liquid precipitation. Besides that, when the temperature rises, early March, snow melt commences, highly contributing to the discharge. Overall, the simulation indicates that Model 8 provides a credible representation of the basin's hydrological behavior, particularly concerning snowmelt processes.

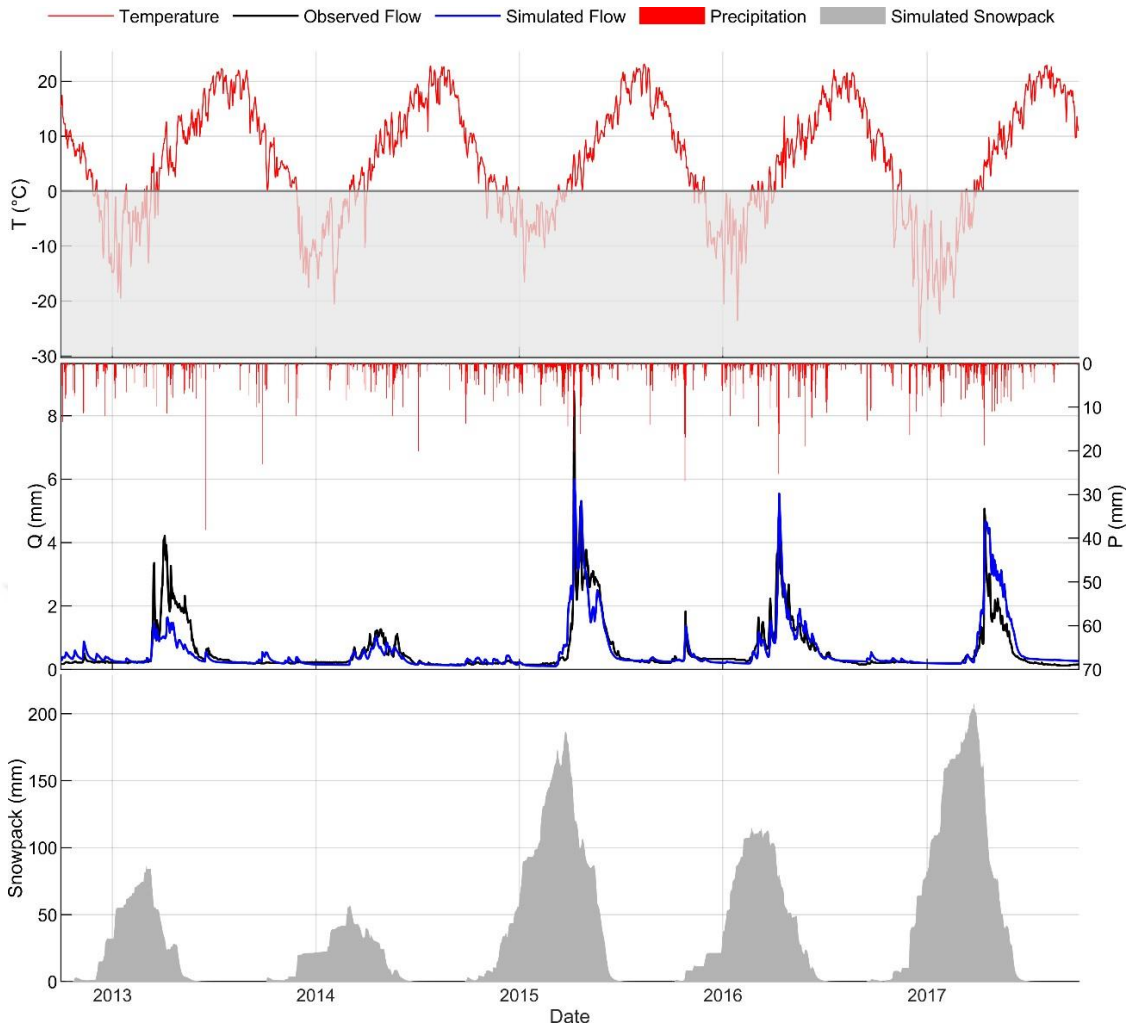

Figure 4.12. *Model 8 Results over Kayabaşı Basin for 2003-2017 calibration period: observed and simulated temperature, discharge hydrographs, precipitation, and snowpack.*

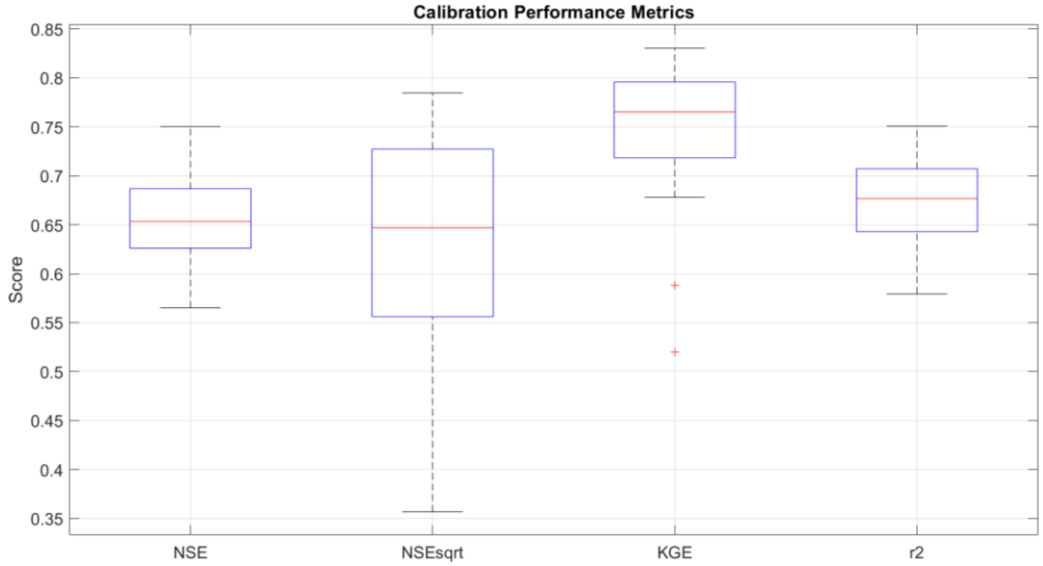

Figure 4.13. *Boxplots of calibration performance metrics for the hydrological models in Kayabaşı Basin using four goodness-of-fit (GOF) metrices: NSE, NSEsqrt, KGE, and* r^2

As shown in Figure 4.13, four performance metrics: NSE, NSEsqrt, KGE, and r^2, were used to evaluate and compare the performance of the models. Among these, KGE demonstrated the highest overall performance, with IQR of the twenty models is between 0.70 and 0.80, indicating both accuracy and consistency. Although two models were identified as outliers in terms of KGE performance, this metric still performs good. r^2 also exhibited good linear relation, despite a wider spread due to high-value outliers. In contrast, the lower scores for both NSE and NSEsqrt reveal specific model weaknesses: the lower NSE shows difficulties in capturing high-flow events, while the poor NSEsqrt scores indicate a more significant challenge in simulating low-flow periods accurately. A detailed summary of these performance statistics is provided in Table 4.3.

Table 4.3. *Summary of calibration performance metrics for NSE, NSEsqrt, KGE, and r^2 across all models.*

Quartiles	**Calibration**			
	NSE	**NSEsqrt**	**KGE**	**r^2**
Max	0.750	0.785	0.830	0.751
Q3 (0.75)	0.687	0.727	0.796	0.707
Median (0.5)	0.653	0.647	0.765	0.677
Q1 (0.25)	0.626	0.556	0.718	0.643
Min	0.565	0.357	0.520	0.579

Figure 4.14 presents the performance of individual models during the calibration period (2013–2017 water years), evaluated using KGE metric. Among the twenty models assessed, Models 8 and 20 exhibited the highest performance, clearly outperforming the rest. The median KGE value across all models was 0.765, indicating a strong overall performance.

A total of 15 models (75%) achieved KGE values greater than 0.70, indicating a satisfactory level of hydrological performance. This suggests that most models reliably captured the observed streamflow dynamics during the calibration period. However, Model 11 recorded the lowest performance, with a KGE of 0.52, highlighting a significant deviation from observed streamflow.

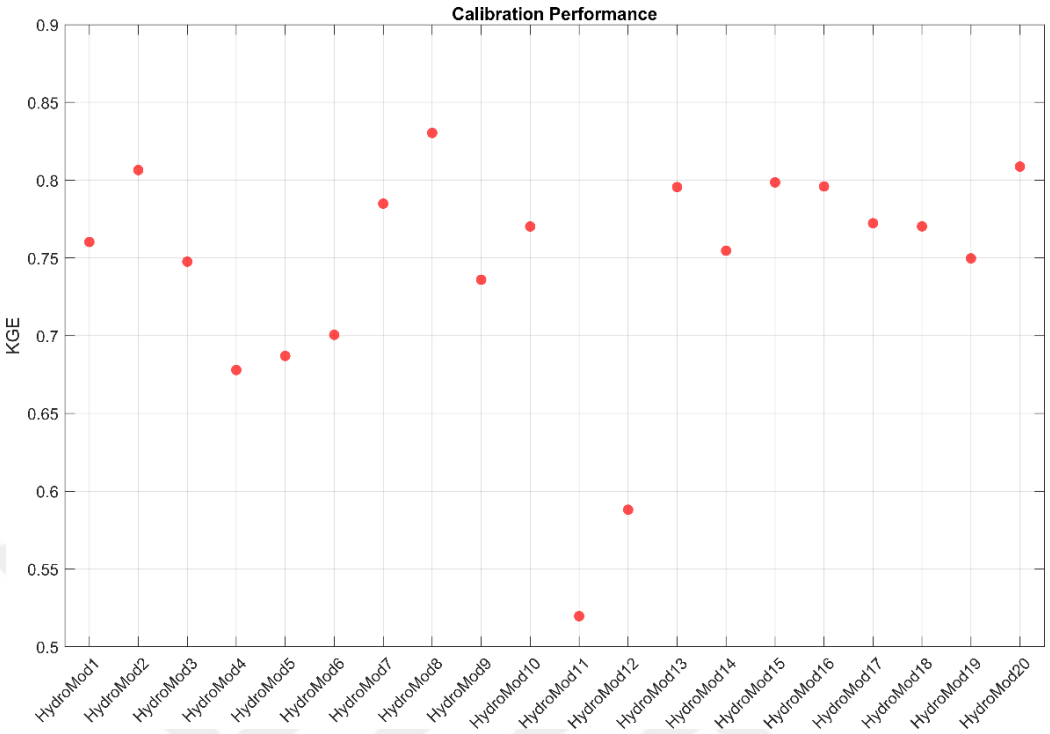

Figure 4.14. *KGE performance metric over Kayabaşı during the 2013-2017 calibration period.*

Validation

The models were validated over the 2008-2012 period using their calibrated parameters. The performance was assessed using the previously used metric; NSE, NSEsqrt, KGE, and r^2, to evaluate the models' ability to reproduce observed streamflow during the validation period. The simulated streamflow captured the overall flow pattern, including most peak flows and baseflow conditions. However, a noticeable spread is observed around high-flow events, indicating increased uncertainty during peak discharge periods as shown in Figure 4.15. The uncertainty is higher in 2009 and 2010 water years regarding the high and low flows. Despite these discrepancies, the ensemble simulation demonstrates satisfactory agreement with observed data, supporting the reliability of the models during the validation period.

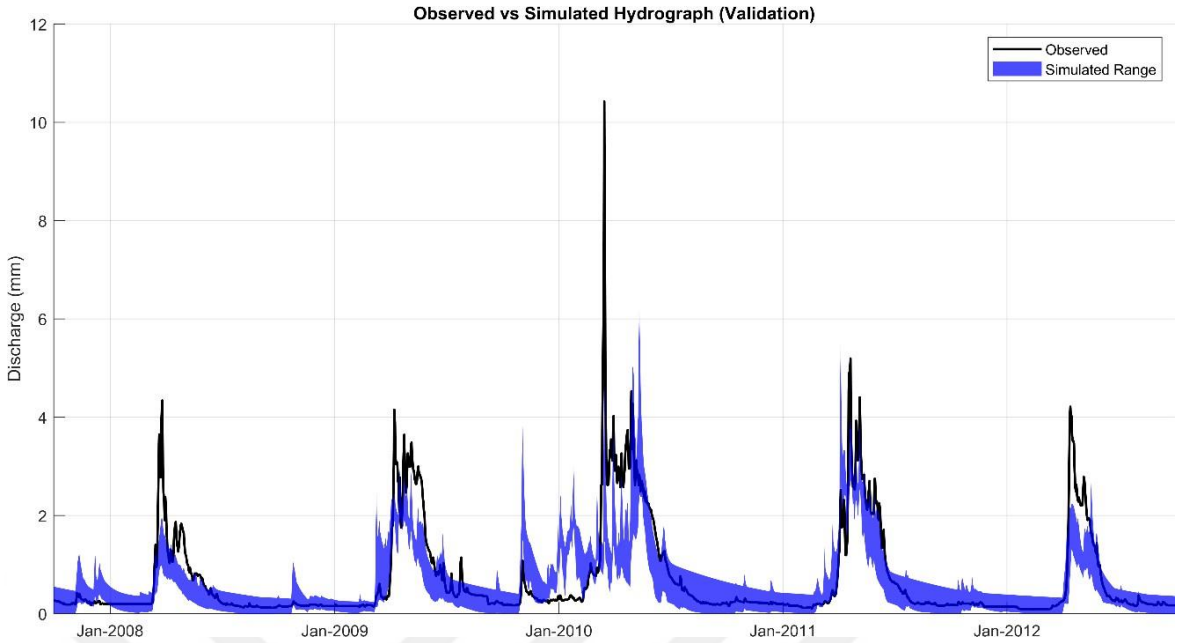

Figure 4.15. *Hydrographs of observed and simulated discharges from 20 hydrological models during the validation period (water years 2008-2012) in Kayabaşı Basin*

Figure 4.16 shows Model 8 results, one of the best performing models in Kayabaşı basin for calibration and validation periods. Observed temperature, discharge, simulated discharge and snowpack are shown. The simulated discharge performed well for most years except 2009 and 2010. While the model captured low flows effectively, it could not capture the high flows of these years. The snowpack for the validation period is between 50 mm and 150 mm. Although the lowest snowpack was observed in 2010, the highest discharge occurred this year. The simulated snowpack showed a good relation between temperature, precipitation and simulated discharges. During winter, low temperatures result in snow accumulation and freezing precipitation. When the temperature rises- around early April- the snowmelt begins, contributing significantly to discharge. Overall, Model 8 provides a credible representation of the basin's hydrological behavior, particularly regarding snowmelt processes.

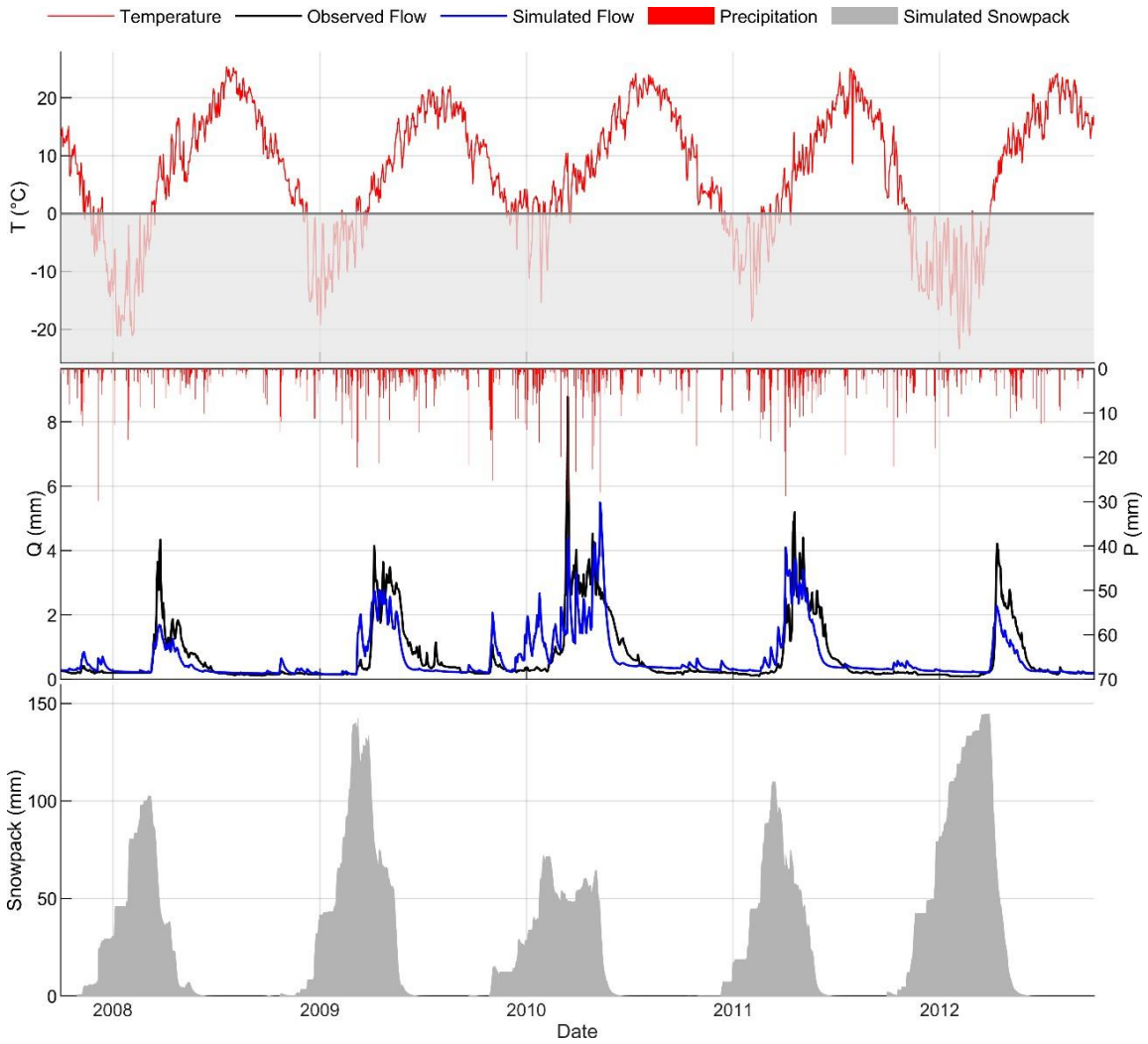

Figure 4.16. *Model 8 Results over Kayabaşı Basin for 2008-2012 validation period: observed and simulated temperature, discharge hydrographs, precipitation, and snowpack*

As shown in Figure 4.17, four performance metrics: NSE, NSEsqrt, KGE, and r^2, were used to evaluate and compare the performance of the models over Kayabaşı. Among these, KGE demonstrated satisfactory overall performance, with IQR of the twenty models is between 0.64 and 0.71, indicating both intermediate accuracy and consistency. Although one model was identified as an outlier in terms of KGE performance, the metric performed well. r^2 also exhibited good results, despite a wider spread due to high-value outliers. In contrast, NSE and NSEsqrt yielded the lowest performance scores among the evaluated metrics. Table 4.4 shows the detailed spread of the performance metrics.

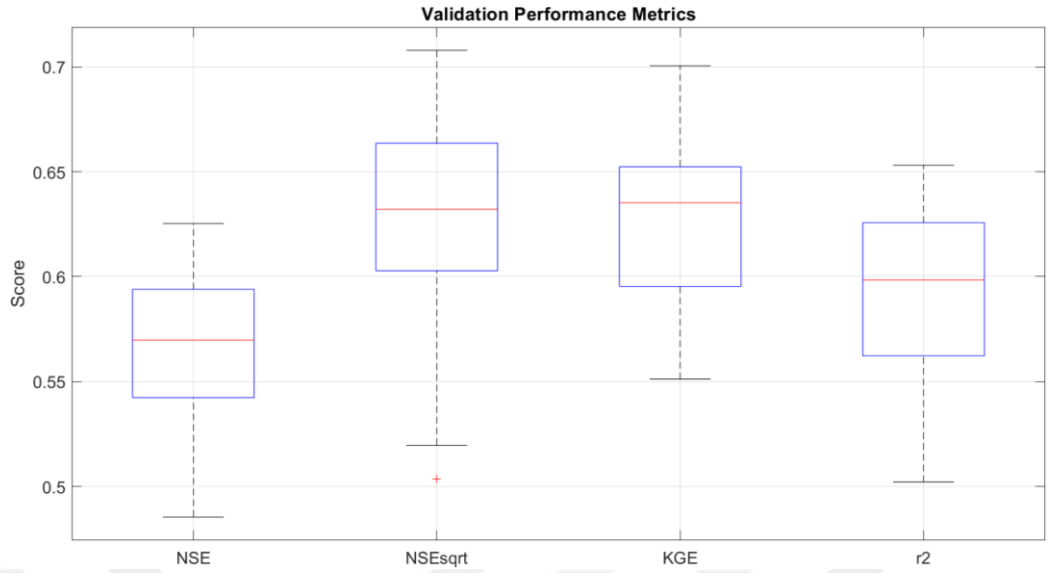

Figure 4.17. *Boxplots of validation performance metrics for the hydrological models in Kayabaşı Basin using four goodness-of-fit (GOF) metrices: NSE, NSEsqrt, KGE, and* r^2

Table 4.4. *Summary of validation performance metrics for NSE, NSEsqrt, KGE, and* r^2 *across all models.*

Quartiles	**Validation**			
	NSE	**NSEsqrt**	**KGE**	**r^2**
Max	0.625	0.708	0.700	0.653
Q3 (0.75)	0.594	0.664	0.652	0.626
Median (0.5)	0.570	0.632	0.635	0.598
Q1 (0.25)	0.542	0.603	0.595	0.562
Min	0.485	0.503	0.551	0.502

Figure 4.18 presents the performance of individual models during the validation period (2008–2012 water years) over Kayabaşı Basin, evaluated using KGE metric. Among the twenty models assessed, Models 2, 8 and 16 exhibited the highest performance, outperforming the remaining models. The median KGE value of 0.67 showed a good performance, with good validation results despite being lower than calibration. A total of 15 models (75%) achieved KGE values greater than 0.60, indicating a satisfactory hydrological model performance. This shows that most models captured the observed streamflow dynamics during the validation period. Model 4 was observed the lowest performance, with a KGE value of 0.55, highlighting a significant deviation from observed data.

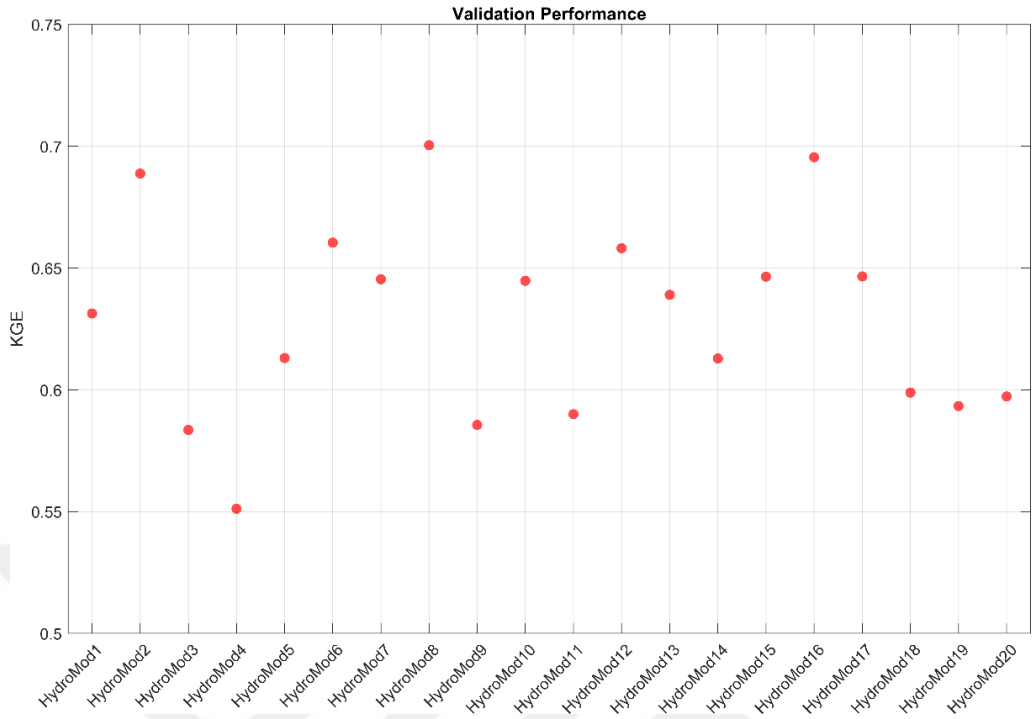

Figure 4.18. *KGE performance over Kayabaşı Basin during the 2008-2012 validation period*

Physical parameters

The FC parameter shows diversity, with 50% between 50 mm and 200 mm as shown in Figure 4.19. The median FC is 101 mm. While some models, like model 16 have shown high FC values around 600 mm, others like models 3, 5, 9, 12, 18, and 19 were found to be below 50 mm. Several models fall within a moderate range between 50 and 200 mm. This spread indicates that calibrated FC values are influenced by each model's structure and design, reflecting requirements to match observed data.

Based on Figure 4.20, the calibrated *Kf* values show substantial variability across the twenty hydrological models. *Kf* ranges from a minimum of approximately 3.6 mm/°C/day (Model 9) to a maximum of 11.10 mm/°C/day (Model 12). The median value for the basin is 6.44 mm/°C/day. Around 9 models' Kf are lower than 6.44 mm/°C/day and IQR from 4.36 to 8.47 mm/°C/day. These results highlight notable differences in snowmelt sensitivity among the models.

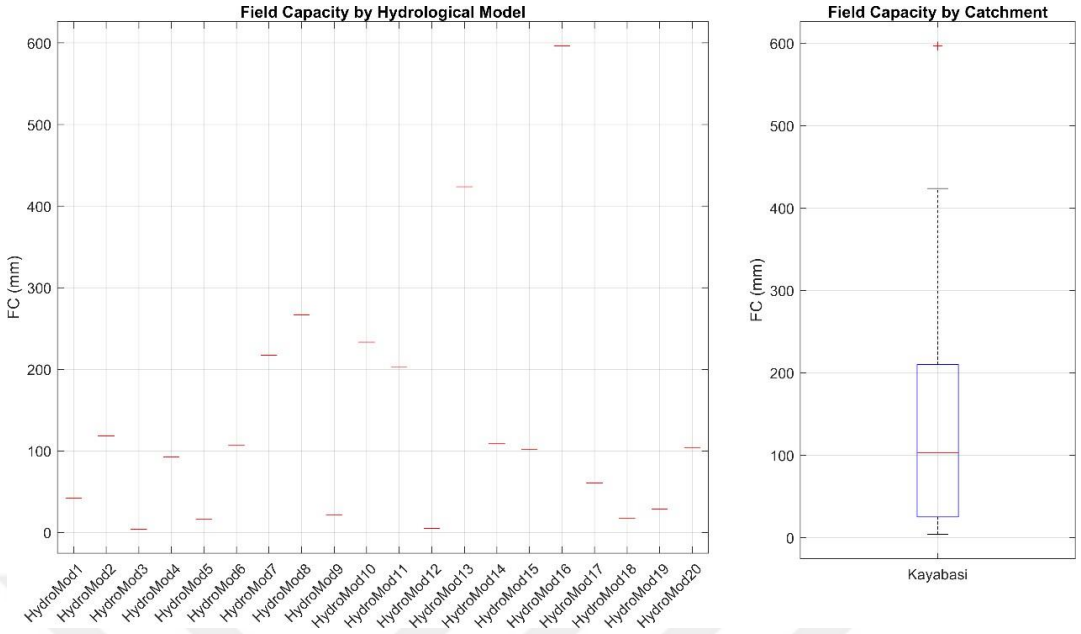

Figure 4.19. *Field Capacity parameter for the twenty models for 2012 – 2013 calibration period*

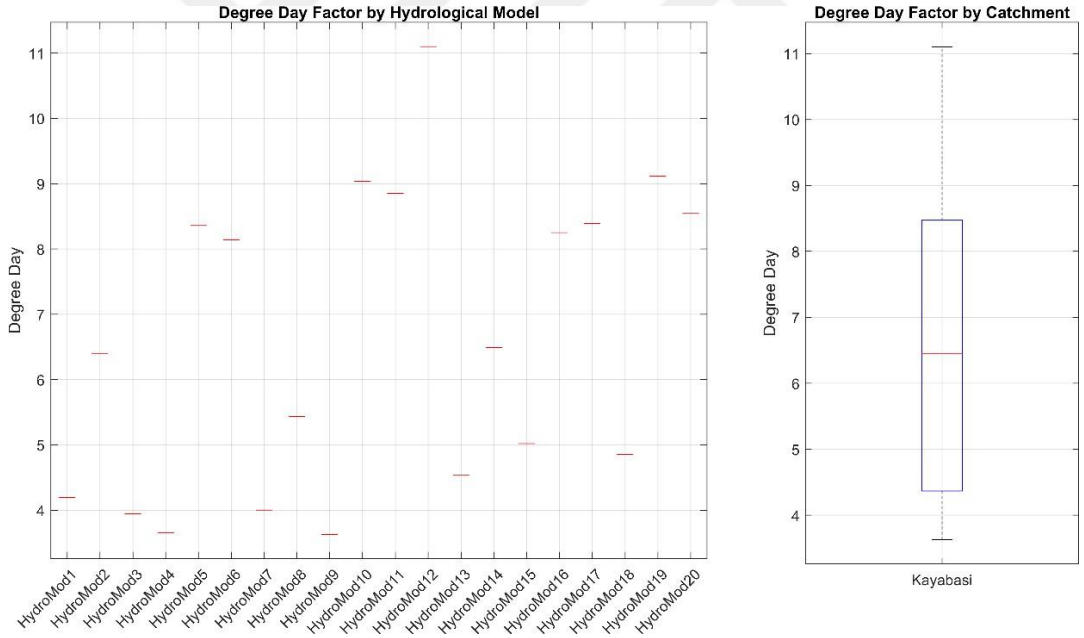

Figure 4.20. *The change of degree day factor parameter of CemaNeige SAR for the twenty models for 2012-2013 calibration period.*

4.2. Open-Loop (OL) Deterministic and Ensemble Forecasting

The hydrological forecasting for the two basins was performed using the calibrated parameters and simulation. This section will exhibit the results of the OL deterministic and ensemble forecasting of the basins throughout the first half of 2018 and 2019 years; January to June.

4.2.1. Deterministic forecasting

Çukurkışla

The deterministic forecast for Çukurkışla Basin was performed from 1st January to 30th June of 2018 and 2019 in 10 lead days in twenty hydrological models. The deterministic forecast performance using KGE across 20 models for Çukurkışla Basin shows decreasing skill with increasing lead time in 2018 and 2019. In 2018, early lead-time forecasts (1-3 days) showed good performance with a median KGE above 0.6 and a narrow IQR. The median KGE declined to 0.4 for Lead Time 10, while the spread of KGE values increased at longer lead times. Models 4 and 16 showed negative KGE values. In 2019, the median KGE for the initial lead day was found at 0.72, representing a slight improvement over 2018. In addition, the decline in performance in 2019 was more pronounced, as the median KGE for the final lead time decreased to 0.22, lower than the 0.43 observed in 2018 for the same lead time. This comparison between the two years revealed inter-annual variability in forecast skill. Figure 4.21 shows the comparison between the performance of the deterministic forecast in 2018 and 2019, while Table 4.5 shows the 75th, 50th and 25th quartiles of the results.

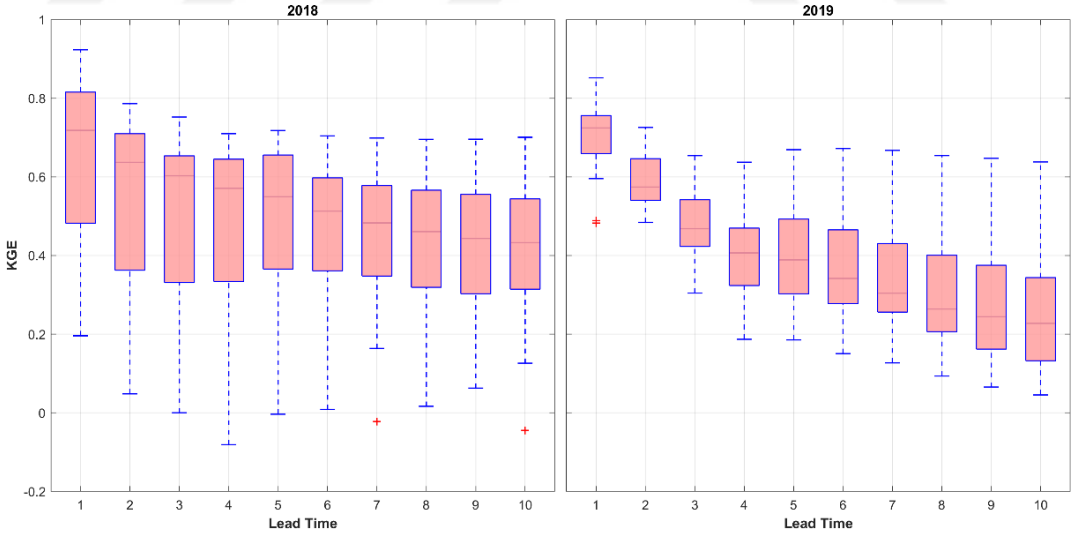

Figure 4.21. *Deterministic forecast performance for Çukurkışla Basin in 2018 and 2019, measured by KGE across 20 models as a function of lead time.*

Table 4.5. *Deterministic forecast performance results in 2018 and 2019 comparison for Çukurkışla Basin*

	Quartiles	**T1**	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	**T8**	**T9**	**T10**
2018	Q3 (0.75)	0.815	0.710	0.653	0.645	0.655	0.597	0.578	0.566	0.555	0.544
	Q2 (0.5)	0.718	0.637	0.603	0.571	0.549	0.513	0.483	0.461	0.444	0.433
	Q1 (0.25)	0.482	0.363	0.332	0.334	0.366	0.361	0.348	0.319	0.303	0.315
	Mean	0.609	0.475	0.452	0.444	0.465	0.443	0.412	0.392	0.386	0.377
2019	Q3 (0.75)	0.756	0.646	0.542	0.469	0.492	0.465	0.430	0.400	0.375	0.343
	Q2 (0.5)	0.725	0.574	0.468	0.407	0.389	0.342	0.304	0.264	0.244	0.227
	Q1 (0.25)	0.659	0.541	0.424	0.324	0.303	0.278	0.257	0.206	0.162	0.133
	Mean	0.704	0.587	0.468	0.399	0.400	0.365	0.334	0.295	0.264	0.236

Kayabaşı

The deterministic forecast modelling for Kayabaşı Basin was performed over the period from January 1 to June 30 for 2018 and 2019, using 20 hydrological models and a 10-day lead time. As shown in Figure 4.22, forecast skill, represented by the KGE consistently decreased with increasing lead time in both years. In 2018, the median KGE at lead time one was approximately 0.68 and gradually declined to 0.56 by lead time ten. In contrast, the 2019 forecasts began with a notably higher median KGE of around 0.80 at lead time one, but showed a sharper decline in performance, with the median dropping to 0.22 by lead time ten, lower than the corresponding 2018 value of 0.43. Early forecasts (1–3 days) in both years achieved a good performance, with median KGEs exceeding 0.60. This pattern highlights the higher reliability of early forecasts and the deterioration of forecast skill as lead time increases. Summary statistics, including the 75th, 50th, and 25th percentiles of KGE values across all models, are presented in Table 4.6.

Table 4.6. *Deterministic forecast performance results in 2018 and 2019 comparison for Kayabaşı Basin.*

	Quartiles	**T1**	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	**T8**	**T9**	**T10**
2018	Q3 (0.75)	0.720	0.713	0.694	0.688	0.686	0.669	0.666	0.650	0.635	0.609
	Q2 (0.5)	0.683	0.678	0.650	0.647	0.643	0.628	0.620	0.607	0.590	0.566
	Q1 (0.25)	0.585	0.593	0.588	0.600	0.604	0.592	0.586	0.578	0.564	0.541
	Mean	0.660	0.658	0.643	0.642	0.639	0.625	0.616	0.607	0.591	0.567
2019	Q3 (0.75)	0.839	0.778	0.712	0.683	0.659	0.630	0.609	0.593	0.589	0.588
	Q2 (0.5)	0.805	0.747	0.688	0.651	0.629	0.612	0.585	0.562	0.550	0.543
	Q1 (0.25)	0.741	0.707	0.656	0.614	0.592	0.566	0.542	0.525	0.515	0.509
	Mean	0.792	0.737	0.681	0.647	0.622	0.599	0.577	0.561	0.552	0.546

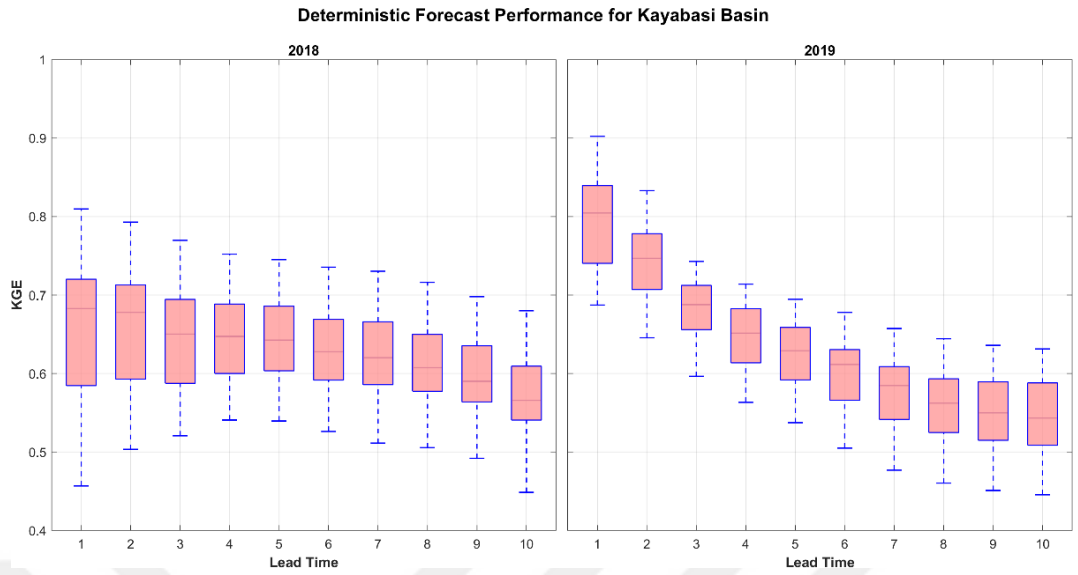

Figure 4.22. *Deterministic forecast performance for Kayabaşı Basin in 2018 and 2019, measured by KGE across 20 models as a function of lead time.*

4.2.2. Ensemble forecasting

Çukurkışla

Ensemble Forecasting with 50 ensemble members were performed using the twenty hydrological models and with 10 days lead time for Çukurkışla Basin. The mean ensemble performance of each model was evaluated using KGE performance index. A good performance was found for the basin across each lead-time. In 2018, the median KGE for the first lead time was observed 0.72, which gradually decreased as the lead time increased. Although some models underperformed – lower than 0.30 KGE, the median performance across all the lead times exceeded 0.60 KGE.

In contrast, the median performance for the first lead time in 2019 was 0.736 KGE, slightly surpassing that of 2018. Additionally, the IQR for each lead time is narrower compared to 2018. Performance gradually declines until the 4th lead time, reaching a median of 0.54 KGE, followed by a slight increase at 5th lead time, and then it continues to decrease gradually until the last, from 0.583 to 0.574 KGE. Although the lowest performance was observed in 4th lead time with 0.40 KGE, it is higher than the lowest performance of 2018. This showed inter-annual variability of this basin in forecast skill. Figure 4.23 shows the box-plot performance of each lead time, while Table 4.7 shows the statistical summary of the forecast performance.

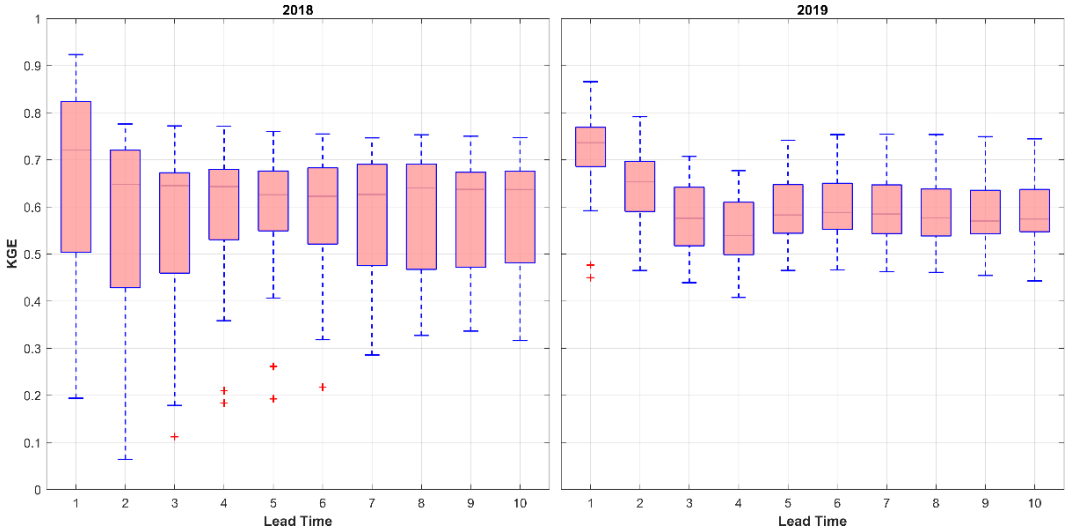

Figure 4.23. *Ensemble forecast performance for Çukurkışla Basin in 2018 and 2019, measured by KGE across ensemble mean of 20 models as a function of lead time*

Table 4.7. *Ensemble forecast performance results in 2018 and 2019 comparison for Çukurkışla Basin*

	Quartiles	**T1**	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	**T8**	**T9**	**T10**
2018	Q3 (0.75)	0.824	0.720	0.673	0.680	0.676	0.683	0.690	0.691	0.673	0.676
	Q2 (0.5)	0.721	0.648	0.645	0.643	0.626	0.622	0.626	0.640	0.637	0.637
	Q1 (0.25)	0.504	0.429	0.460	0.531	0.549	0.522	0.475	0.468	0.472	0.481
	Mean	0.616	0.503	0.508	0.536	0.544	0.541	0.543	0.557	0.553	0.550
2019	Q3 (0.75)	0.769	0.697	0.642	0.610	0.647	0.650	0.647	0.638	0.635	0.637
	Q2 (0.5)	0.736	0.653	0.576	0.540	0.583	0.588	0.585	0.577	0.570	0.574
	Q1 (0.25)	0.686	0.591	0.518	0.499	0.544	0.552	0.544	0.539	0.543	0.548
	Mean	0.717	0.644	0.574	0.548	0.593	0.600	0.597	0.592	0.590	0.589

Kayabaşı

The ensemble forecasting modelling for Kayabaşı Basin was also performed from 1st January to 30th June 2018 and 2019. Mean ensemble forecast performance was evaluated using KGE performance index. A good performance was found for this basin across each lead-time. In 2018, the median KGE for the first lead time was observed 0.67, which gradually decreased as the lead time increased. Although some models underperformed in the early (1-2 days) forecast, lower than 0.50 KGE, the median performance across all the lead times exceeded 0.60 KGE, indicating reliable performance from most of the models. In 2019, although the performance of the forecasts began with a notably higher median KGE of around 0.81 at lead time 1, but showed a sharper decline in performance, with the median dropping to 0.55 by lead time 10,

considerably lower than the corresponding 2018 KGE of 0.612. The performance of the early forecast (1-3 days) exceeded 0.70, which is higher than the performance of the first lead day in 2018. This comparison has shown inter-annual variability of this basin. This variability is also shown in both basins in deterministic and ensemble forecasting.

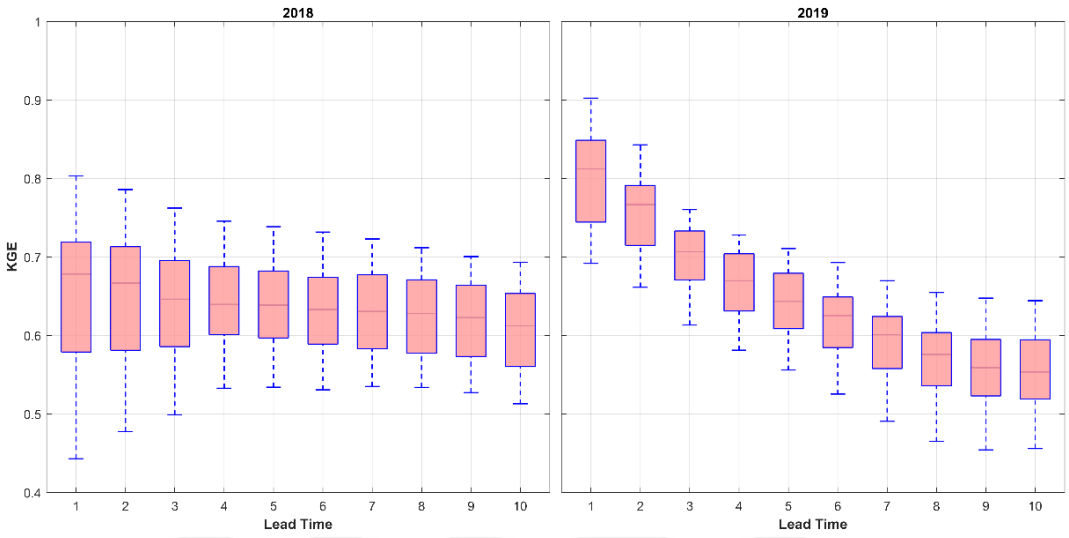

***Figure 4.24**. Ensemble forecast performance for Kayabaşı Basin in 2018 and 2019, measured by KGE across ensemble mean of 20 models as a function of lead time.*

***Table 4.8.** Ensemble forecast performance results in 2018 and 2019 comparison for Kayabaşı Basin.*

	Quartiles	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
2018	Q3 (0.75)	0.719	0.713	0.695	0.688	0.682	0.674	0.677	0.671	0.664	0.653
	Q2 (0.5)	0.678	0.667	0.646	0.640	0.639	0.633	0.631	0.628	0.623	0.612
	Q1 (0.25)	0.579	0.581	0.586	0.601	0.597	0.589	0.583	0.578	0.573	0.561
	Mean	0.656	0.651	0.636	0.637	0.636	0.631	0.629	0.624	0.617	0.608
2019	Q3 (0.75)	0.849	0.791	0.733	0.704	0.679	0.649	0.624	0.604	0.595	0.595
	Q2 (0.5)	0.812	0.767	0.707	0.670	0.644	0.625	0.601	0.576	0.559	0.554
	Q1 (0.25)	0.745	0.715	0.671	0.631	0.609	0.585	0.558	0.536	0.523	0.519
	Mean	0.799	0.752	0.698	0.664	0.640	0.616	0.592	0.571	0.559	0.556

4.2.3. Comparison of deterministic and ensemble forecast performance

Çukurkışla

The hydrographs of deterministic and ensemble forecasts were compared alongside with the observed hydrograph for model 20 in 2018 and 2019. For 2018, the hydrographs of both forecasts have shown high accuracy with very narrow ensemble spread, closely capturing the high and low peaks of the observed discharges in the first lead days. As the

lead time increases, the accuracy decreases, and the ensemble spread widens as shown both lead times 5 and 10 in Figure 4.25. In addition, peak discharge of the forecasts decline below the observed at the longer lead days. The pattern of both forecasts are quietly similar for this year and model.

For 2019, although the hydrographs of the first lead time did not capture the peak discharges in mid-March and early of April and overestimated the peak discharge in mid-January, the accuracy of both forecast predicted good and follow the observed discharge pattern. As the forecast of 2018, the accuracy of both forecasts decline as the lead time increases, and the forecasts could not capture the peak discharges. The ensemble forecast outperformed at the longer lead times as shown in Figure 4.26. The below figures show the comparison between observed discharge against deterministic, ensemble mean and ensemble range forecast for Çukurkışla Basin in 2018 and 2019 for Model 20.

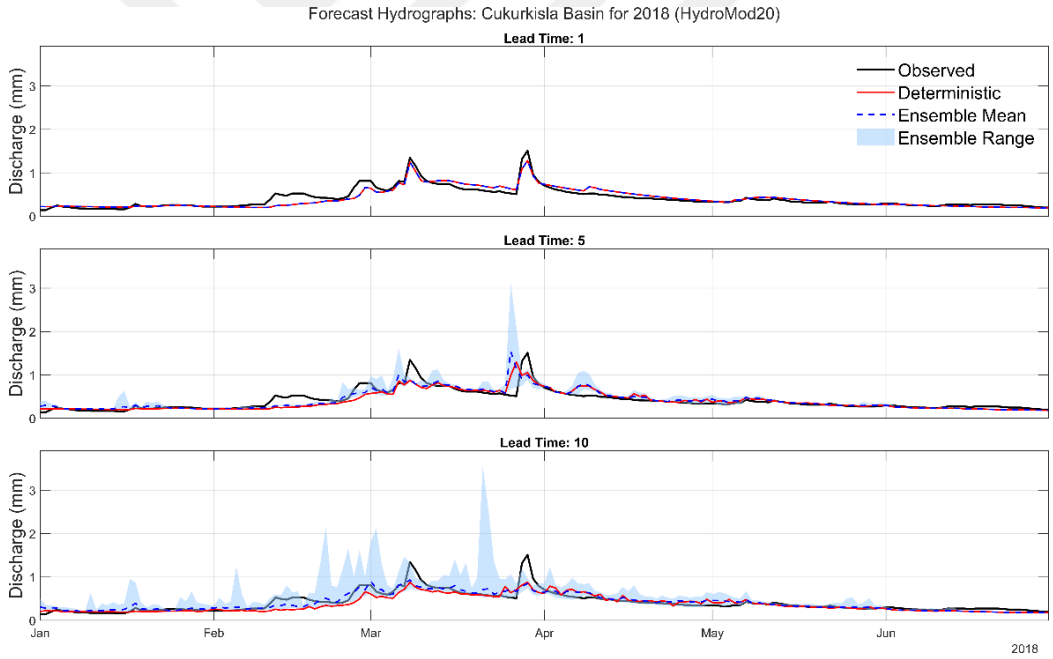

Figure 4.25. *Çukurkışla deterministic and ensemble forecast hydrographs for model 20 in 2018 against observed hydrograph: Lead time 1, 5, and 10.*

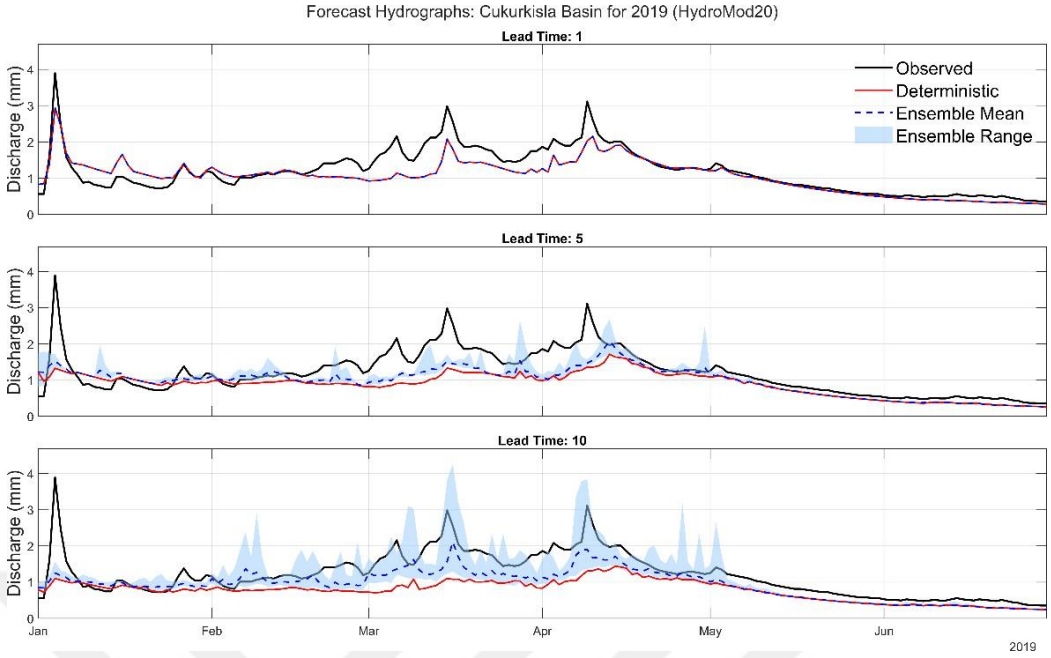

Figure 4.26. *Çukurkışla deterministic and ensemble forecast hydrographs for model 20 in 2019 against observed hydrograph: Lead time 1, 5, and 10*

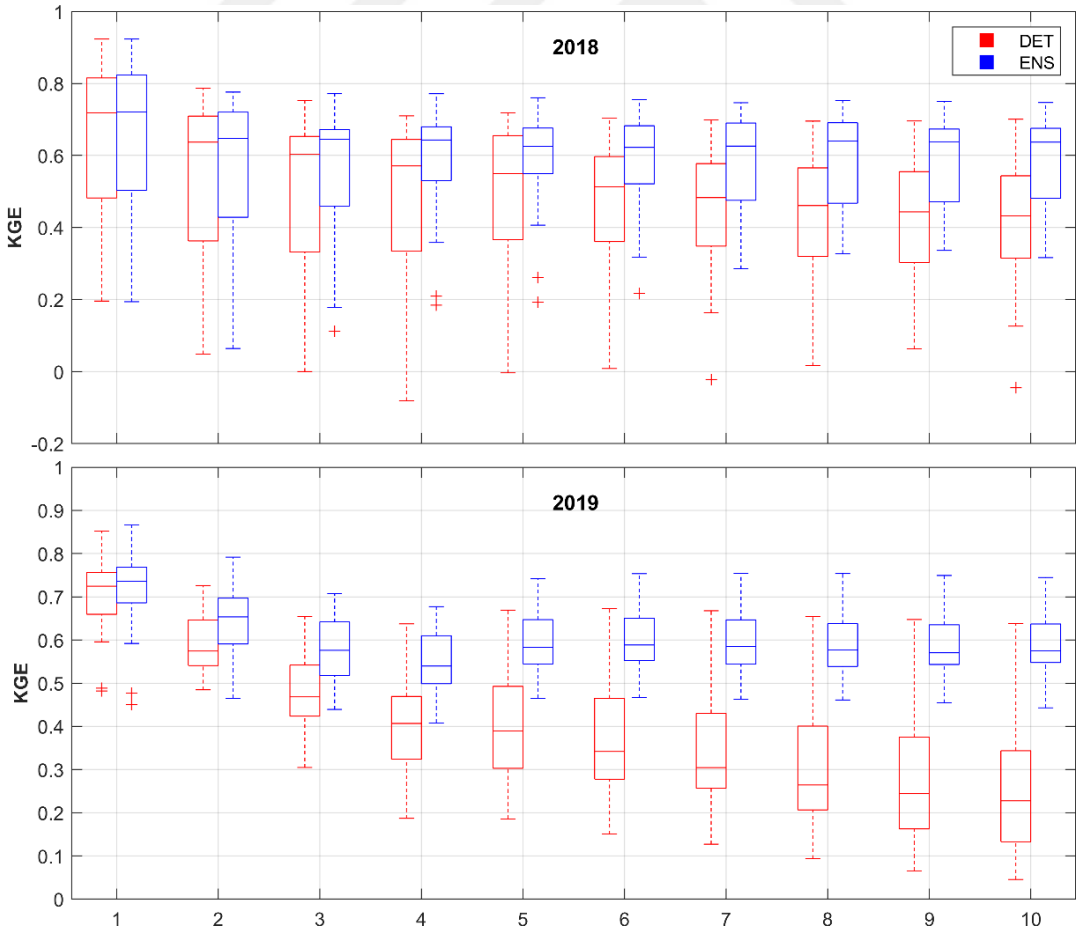

Figure 4.27. *Comparison of deterministic and ensemble forecast performance across 10 lead days on Çukurkışla Basin*

Figure 4.27 compares deterministic and ensemble forecasting performance using KGE over ten lead days for Çukurkışla Basin in 2018 and 2019. The ensemble forecast outperformed the deterministic forecast, especially at longer lead times. For 2018, both forecasts perform well in the early days, with KGE above 0.60. After a third lead day, the deterministic performance declines, showing lower KGE values and a wider spread. The ensemble forecast maintains higher KGE values, ranging from 0.40 to 0.80, with a narrower spread. In 2019, performance decline was more pronounced for both scenarios, indicating a more challenging forecasting year. While both scenarios have shown reasonable KGEs at 1-3 days, the ensemble forecast outperforms the deterministic in the later lead days. Beyond the third lead tie, deterministic forecast performance degrades significantly, with median KGE values below 0.4 and wide spreads. The ensemble forecast maintains a median KGE above 0.5 with narrower spread, demonstrating reduced uncertainty and high reliability at longer lead times and during challenging periods.

Kayabaşı

The hydrographs of deterministic and ensemble forecasts are compared alongside with the observed hydrograph for model 20 in 2018 and 2019. For 2018, the forecasts showed good accuracy at early lead times, reflected by a very narrow ensemble spread. However, they overestimated the peak discharges in late March, May, and early June. As shown for lead times of 5 and 10 days in Figure 4.28, forecast accuracy decreases and the ensemble spread widens as the lead time increases. At these longer lead times, the forecasts also tended to underestimate the high flows in March while overestimating them in May. Overall, the forecast patterns for this year were quite similar.

For 2019, while the early lead time forecasts failed to capture the peak discharge in May, they otherwise followed the observed discharge pattern well. Similar to 2018, forecast accuracy declined as the lead time increased, and the models were unable to capture the peak discharges. As shown in Figure 4.29, the overall performance for this year was quite similar. The figures below compare the observed discharge in Kayabaşı Basin for Model 20 against its deterministic forecast, ensemble mean, and ensemble range for 2018 and 2019.

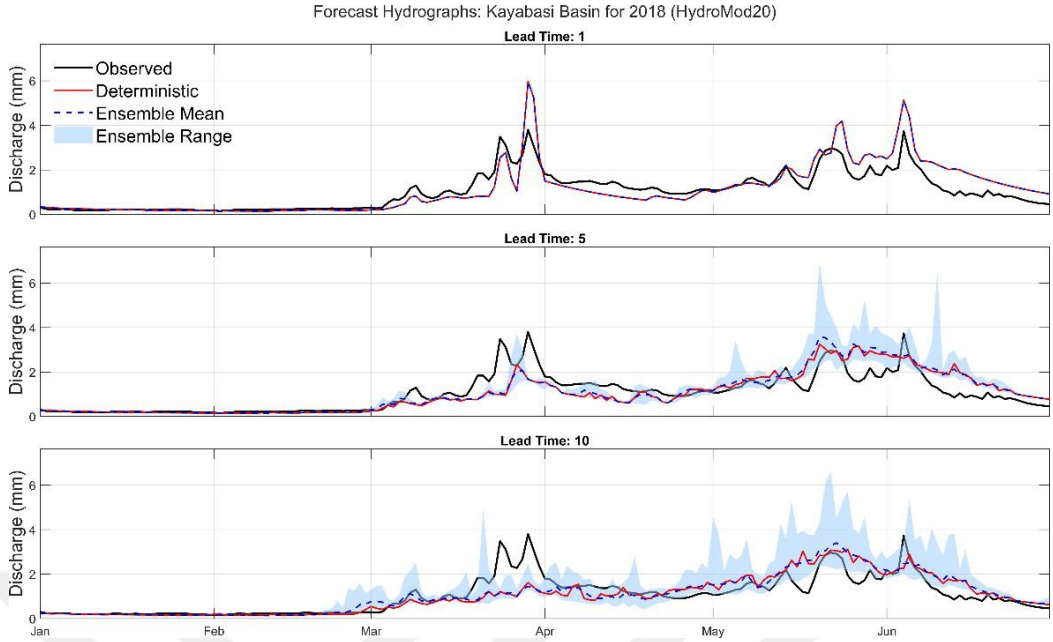

Figure 4.28. *Kayabaşı deterministic and ensemble forecast hydrographs for model 20 in 2018 against observed hydrograph: Lead time 1, 5, and 10*

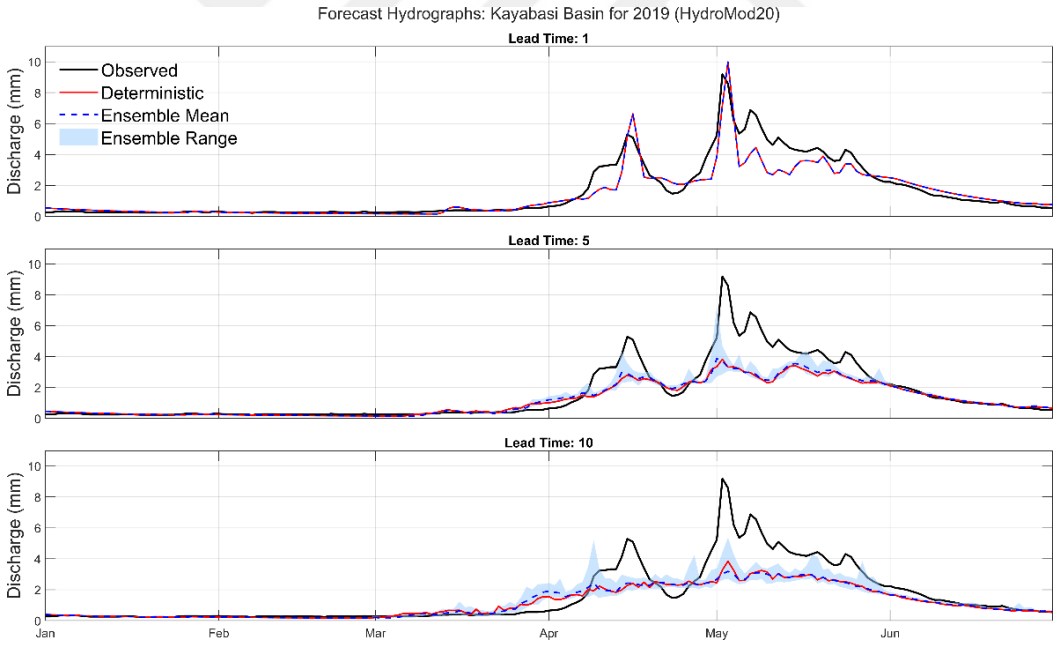

Figure 4.29. *Kayabaşı deterministic and ensemble forecast hydrographs for model 20 in 2019 against observed hydrograph: Lead time 1, 5, and 10*

Figure 4.30 compares deterministic and ensemble forecasting performance using KGE over ten lead days for Kayabaşı Basin in 2018 and 2019. The ensemble forecast slightly outperformed the deterministic forecast at longer lead times. For 2018, both forecasts perform well in all lead days, with KGE above 0.60. At early lead time (1-3

days) the performance of both forecasts are quite similar. As the lead time increases, the deterministic forecast a slightly underperformed that of ensemble forecast. For 2019, both ensemble and deterministic forecasts performed good throughout all the lead time, higher than 0.50 KGE. The median performance of the early lead times (1-3 days) exceeded 0.70 KGE, indicating higher reliability at the early days. The ensemble forecast slightly outperformed the deterministic forecast across all the lead times, demonstrating reduced uncertainty and higher reliability of the forecast. Ensemble forecast outperformed the deterministic forecast for both years and basins at longer lead days.

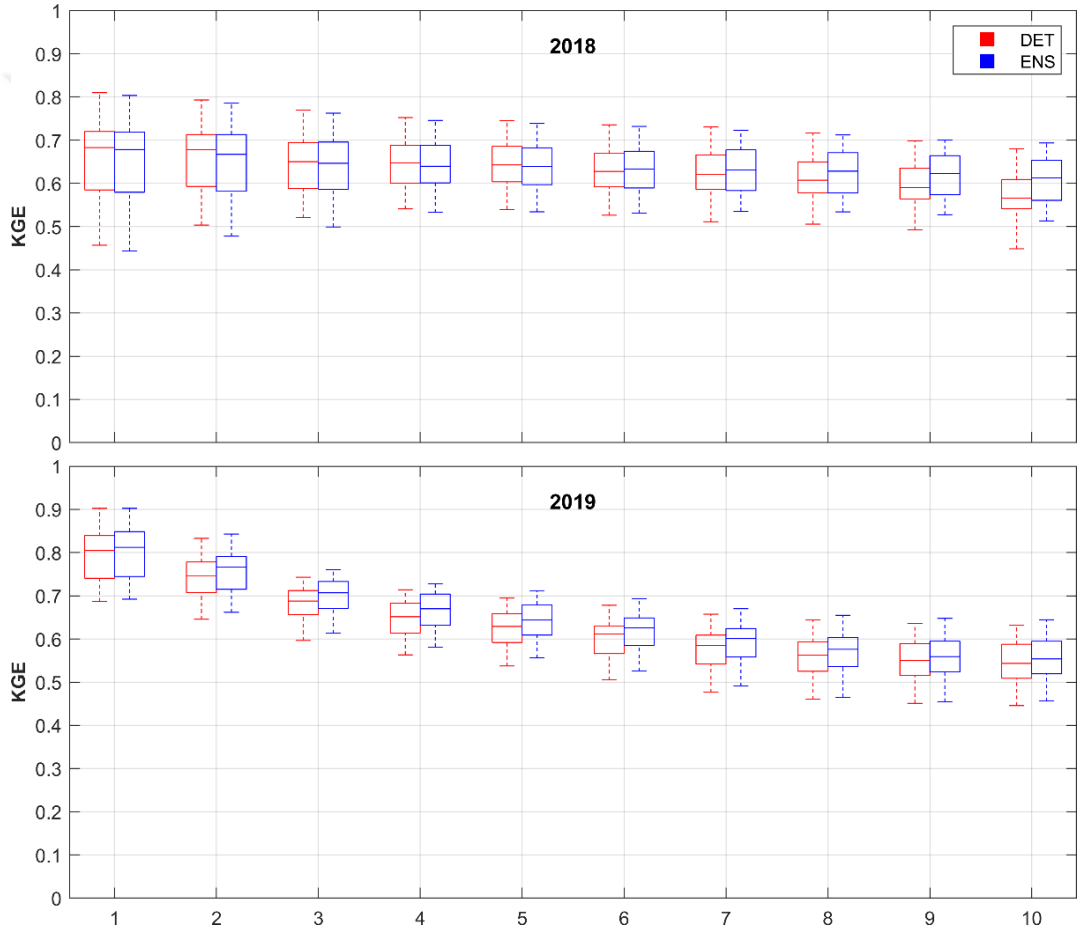

Figure 4.30. *Comparison of deterministic and ensemble forecast performance across 10 lead days on Kayabaşı Basin*

4.3. EnKF Data Assimilation Forecast

4.3.1. Data assimilated deterministic forecasting

Çukurkışla

EnKF data assimilation (DA) was performed with 50 perturbed members using observations of discharge, precipitation, temperature, and potential evapotranspiration. Following assimilation, forecasts were generated. The performance of the DA forecast showed significant improvement, evidenced by higher performance scores and a narrower interquartile range (IQR) at each lead time as shown in Figure 4.31. Table 4.9 also shows the performance of each lead time. In 2018, the median KGE performance of twenty models for the first lead time was 0.774, and this median performance gradually declined as the lead time increased, reaching 0.507 KGE. In 2019, although a higher performance of 0.892 KGE was observed in the first lead time, the decline in performance was more pronounced than in 2018, dropping to 0.423 KGE in the last lead time.

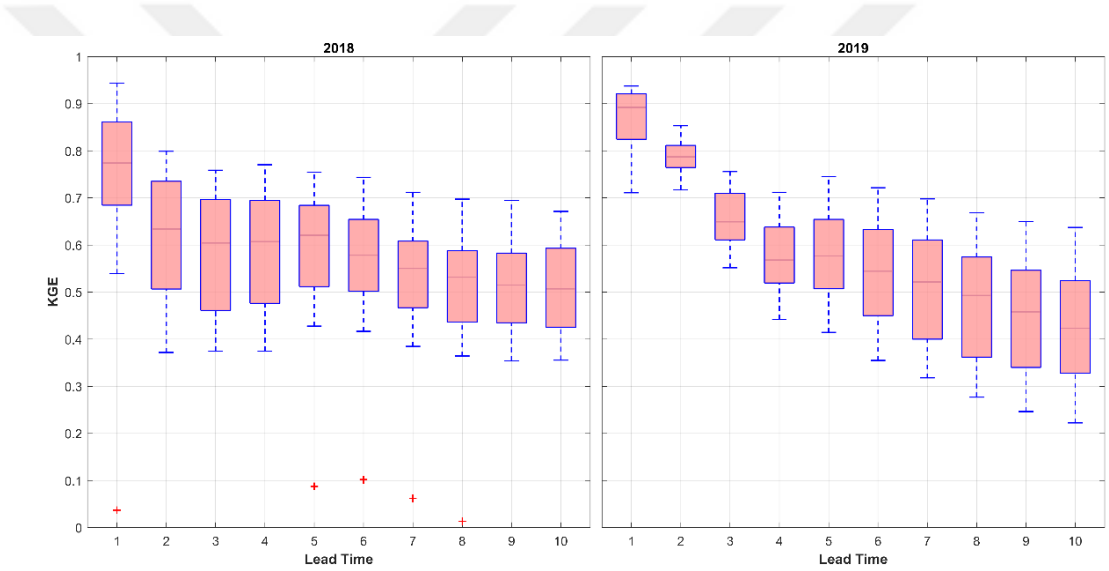

Figure 4.31. *Data assimilated deterministic forecast performance (KGE) for Çukurkışla Basin in 2018 and 2019, across 20 models as a function of lead time*

Table 4.9. *Data assimilated deterministic forecast performance results in 2018 and 2019 comparison for Çukurkışla Basin*

	Quartiles	**T1**	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	**T8**	**T9**	**T10**
2018	Q3 (0.75)	0.861	0.735	0.697	0.694	0.684	0.654	0.608	0.588	0.583	0.593
	Q2 (0.5)	0.774	0.634	0.604	0.607	0.620	0.578	0.550	0.532	0.515	0.507
	Q1 (0.25)	0.685	0.507	0.461	0.476	0.512	0.502	0.467	0.436	0.435	0.426
	Mean	0.739	0.578	0.557	0.565	0.582	0.559	0.525	0.504	0.496	0.488
2019	Q3 (0.75)	0.921	0.811	0.710	0.638	0.654	0.633	0.611	0.574	0.546	0.524
	Q2 (0.5)	0.892	0.787	0.649	0.568	0.577	0.544	0.522	0.493	0.458	0.423
	Q1 (0.25)	0.825	0.765	0.611	0.520	0.507	0.450	0.400	0.362	0.340	0.328
	Mean	0.871	0.788	0.654	0.575	0.582	0.549	0.520	0.481	0.449	0.421

Kayabaşı

The input data for Kayabaşı Basin underwent EnKF DA, from which deterministic forecasts were generated. The performance of the data-assimilated deterministic forecast showed improvement, with higher performance and narrower IQ ranges at each lead time as shown in Figure 4.32 and Table 4.10. In 2018, the median KGE performance of twenty models for the first lead time was 0.784, and this median performance gradually declined as the lead time increased, reaching 0.65 KGE. In 2019, although a higher performance of 0.894 KGE was observed in the first lead time, the decline in performance was more pronounced than in 2018, dropping to 0.536 KGE in the last lead time.

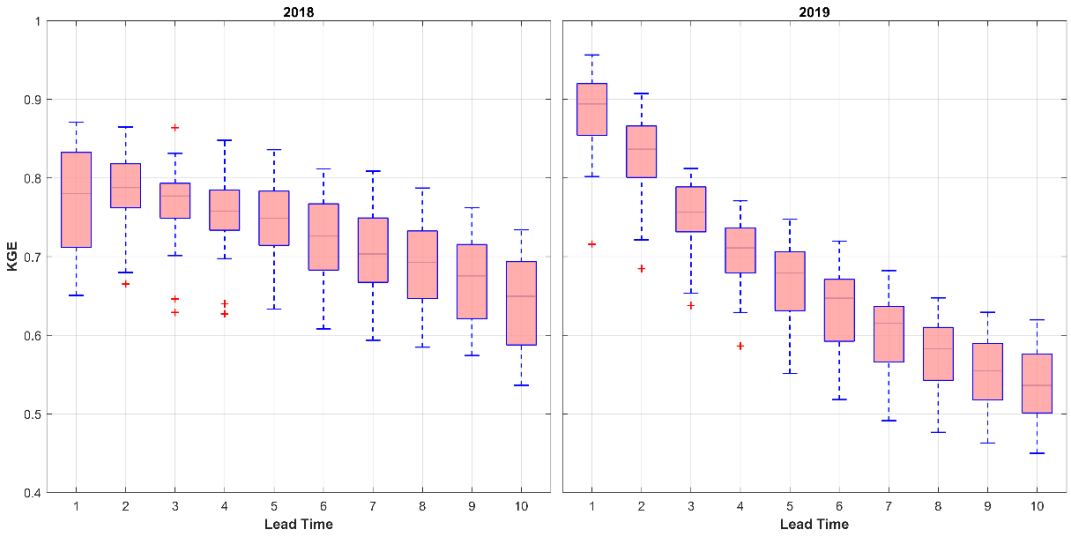

Figure 4.32. *Data assimilated deterministic forecast performance (KGE) for the Çukurkışla Basin in 2018 and 2019, across 20 models as a function of lead time*

Table 4.10. *Data assimilated deterministic forecast performance in 2018 and 2019 for Çukurkışla Basin*

	Quartiles	**T1**	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	**T8**	**T9**	**T10**
2018	Q3 (0.75)	0.832	0.818	0.793	0.785	0.783	0.767	0.749	0.732	0.715	0.694
	Q2 (0.5)	0.780	0.788	0.777	0.758	0.749	0.726	0.703	0.693	0.676	0.650
	Q1 (0.25)	0.712	0.762	0.749	0.734	0.714	0.683	0.668	0.647	0.621	0.588
	Mean	0.775	0.785	0.764	0.753	0.743	0.722	0.706	0.691	0.673	0.645
2019	Q3 (0.75)	0.920	0.866	0.789	0.736	0.706	0.671	0.637	0.610	0.590	0.576
	Q2 (0.5)	0.894	0.837	0.756	0.711	0.679	0.647	0.615	0.583	0.555	0.536
	Q1 (0.25)	0.854	0.801	0.732	0.679	0.631	0.593	0.566	0.543	0.518	0.501
	Mean	0.880	0.826	0.752	0.704	0.670	0.636	0.602	0.573	0.552	0.539

4.3.2. Data assimilated ensemble forecasting

Çukurkışla

The input data for Çukurkışla Basin underwent EnKF DA, from which ensemble forecasts were generated. The performance of the data-assimilated ensemble forecast showed improvement, with higher performance and narrower IQ ranges at each lead time. Figure 4.33 and table 4.11 illustrate the performance comparison between 2018 and 2019 in Çukurkışla Basin. In 2018, the median KGE performance of twenty models for the first lead time was 0.809, and this median performance gradually declined as the lead time increased, reaching 0.644 KGE. In 2019, although a higher performance of 0.887 KGE was observed in the first lead time, the decline in performance was more pronounced than in 2018, dropping to 0.671 KGE in the last lead time.

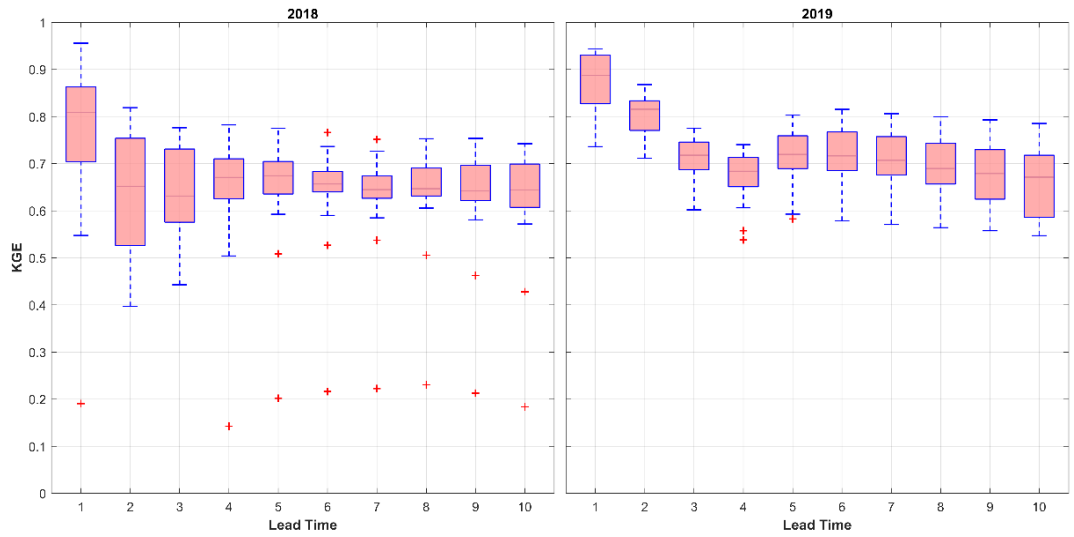

Figure 4.33. *Data assimilated ensemble forecast performance (KGE) for Çukurkışla Basin in 2018 and 2019, across 20 models as a function of lead time*

Table 4.11. *Data assimilated ensemble forecast performance in 2018 and 2019 for Çukurkışla Basin*

	Quartiles	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
2018	Q3 (0.75)	0.863	0.753	0.731	0.710	0.704	0.683	0.674	0.691	0.696	0.699
	Q2 (0.5)	0.809	0.652	0.631	0.671	0.674	0.657	0.645	0.647	0.642	0.644
	Q1 (0.25)	0.704	0.526	0.576	0.625	0.636	0.640	0.627	0.632	0.622	0.607
	Mean	0.757	0.608	0.605	0.641	0.648	0.639	0.629	0.636	0.627	0.622
2019	Q3 (0.75)	0.930	0.833	0.746	0.713	0.759	0.767	0.757	0.743	0.730	0.718
	Q2 (0.5)	0.887	0.816	0.718	0.684	0.719	0.717	0.707	0.690	0.679	0.671
	Q1 (0.25)	0.827	0.771	0.687	0.651	0.690	0.685	0.676	0.657	0.625	0.586
	Mean	0.872	0.804	0.709	0.671	0.716	0.716	0.705	0.690	0.675	0.663

Kayabaşı

Finally, the input data for Kayabaşı Basin underwent EnKF DA, from which ensemble forecasts were generated. The performance of the data-assimilated ensemble forecast showed improvement, with higher performance and narrower IQ ranges at each lead time. Figure 4.34 and Table 4.12 illustrate the performance comparison between 2018 and 2019 in Kayabaşı Basin. In 2018, the median KGE performance of twenty models for the first lead time was 0.782, and this median performance gradually declined as the lead time increased, reaching 0.698 KGE. In 2019, although a higher performance of 0.899 KGE was observed in the first lead time, the decline in performance was more pronounced in 2019, dropping to 0.54 KGE in the last lead time.

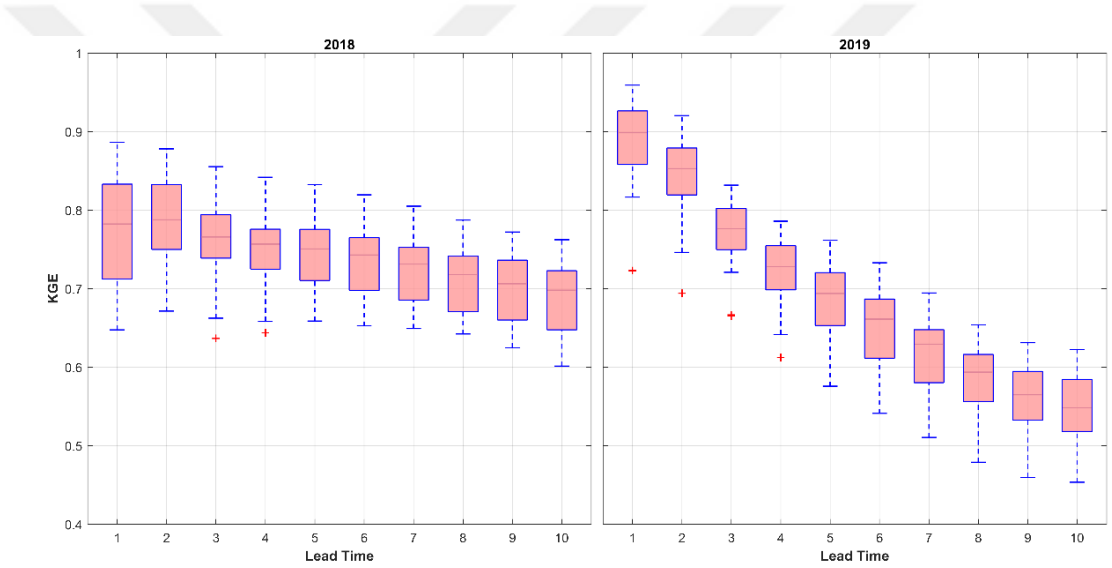

Figure 4.34. *Data assimilated ensemble forecast performance (KGE) for Kayabaşı Basin in 2018 and 2019, across 20 models as a function of lead time*

Table 4.12. *Data assimilated ensemble forecast performance results in 2018 and 2019 comparison for Kayabaşı Basin*

	Quartiles	**T1**	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	**T8**	**T9**	**T10**
2018	Q3 (0.75)	0.833	0.833	0.794	0.776	0.775	0.765	0.753	0.742	0.736	0.723
	Q2 (0.5)	0.782	0.788	0.766	0.757	0.751	0.743	0.731	0.718	0.706	0.698
	Q1 (0.25)	0.712	0.750	0.739	0.725	0.711	0.698	0.686	0.671	0.660	0.648
	Mean	0.777	0.785	0.762	0.753	0.746	0.735	0.725	0.713	0.702	0.689
2019	Q3 (0.75)	0.926	0.879	0.802	0.755	0.720	0.687	0.648	0.616	0.595	0.584
	Q2 (0.5)	0.899	0.853	0.776	0.728	0.694	0.661	0.629	0.594	0.565	0.548
	Q1 (0.25)	0.858	0.820	0.750	0.699	0.653	0.612	0.581	0.556	0.533	0.518
	Mean	0.885	0.841	0.769	0.720	0.685	0.650	0.614	0.582	0.560	0.549

4.3.3. Impact of data Assimilation on forecast skill

Çukurkışla

Open-loop (OL) and data assimilated (DA) deterministic and ensemble forecast performances for Çukurkışla Basin in 2018 are compared as shown in Figure 4.35. Table 4.13 presents the median KGE performances of OL and DA deterministic and ensemble forecasts across all lead times in 2018. While the OL deterministic forecast showed the lowest performance across all lead times, the DA ensemble forecast performed the highest, according to the highest observed KGE values and smallest IQ ranges. Although the DA deterministic forecast performed better than the OL deterministic forecast, it could not outperform the OL ensemble forecast, except at the first lead time. The impact of DA is clearly visible across all lead times, demonstrating that DA reduced the uncertainty and increased the reliability of the forecast.

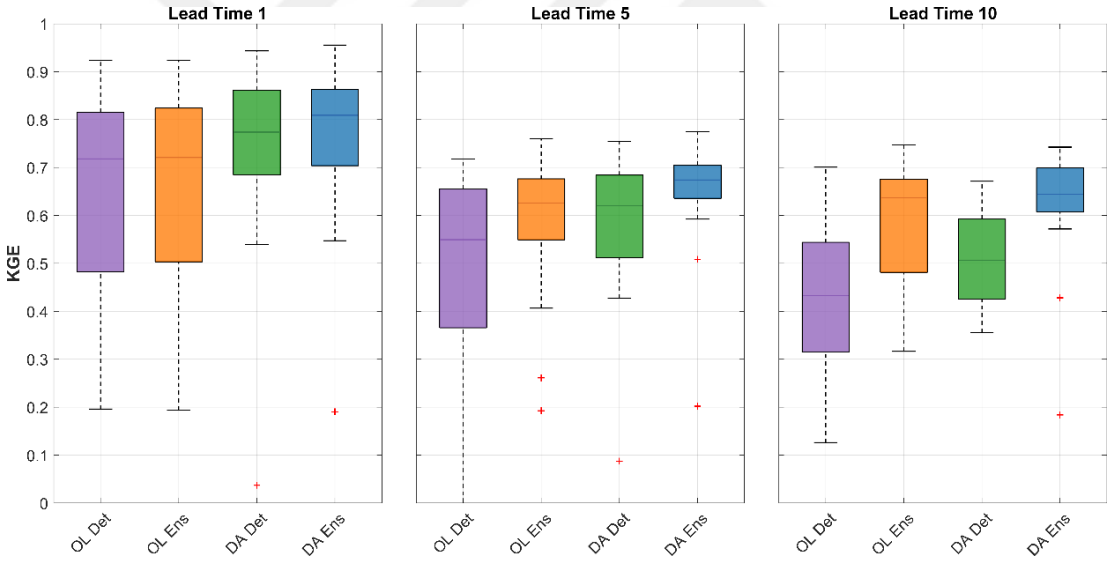

Figure 4.35. *Performance of Open Loop (OL) and Data Assimilation (DA) deterministic and ensemble forecasts across lead times of 1, 5, and 10 days for Çukurkışla Basin in 2018*

Table 4.13. *Median KGE performance of OL and DA deterministic and ensemble forecasts across all lead times for Çukurkışla Basin in 2018*

Forecast type	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10
OL Deterministic	0.718	0.637	0.603	0.571	0.549	0.513	0.483	0.461	0.444	0.433
OL Ensemble	0.721	0.648	0.645	0.643	0.626	0.622	0.626	0.640	0.637	0.637
DA Deterministic	0.774	0.634	0.604	0.607	0.620	0.578	0.550	0.532	0.515	0.507
DA Ensemble	0.809	0.652	0.631	0.671	0.674	0.657	0.645	0.647	0.642	0.644

Similarly, OL and DA deterministic and ensemble forecasts for Çukurkışla Basin in 2019 are compared as shown in Figure 4.36. Table 4.13 presents the median KGE performances of OL and DA deterministic and ensemble forecasts across all lead times in 2019. All the scenarios showed strong performances with the DA forecasts, both deterministic and ensemble, outperforming the OL forecasts. Although the OL deterministic forecast exhibited the lowest performance across all lead times, the DA ensemble forecast achieved the highest performance, as indicated by the highest observed KGE values and the smallest IQ ranges. The DA deterministic forecast outperformed its OL counterpart, but it underperformed the OL ensemble forecast across all lead times except the first. The highest performances were observed in the first lead days, then they declined gradually. The decline is more pronounced in the deterministic forecasts for both OL and DA forecasts. This consistent improvement from OL to DA and from deterministic to ensemble methods emphasizes the importance of both DA and ensemble strategies in improving forecast skill, especially for medium-range predictions.

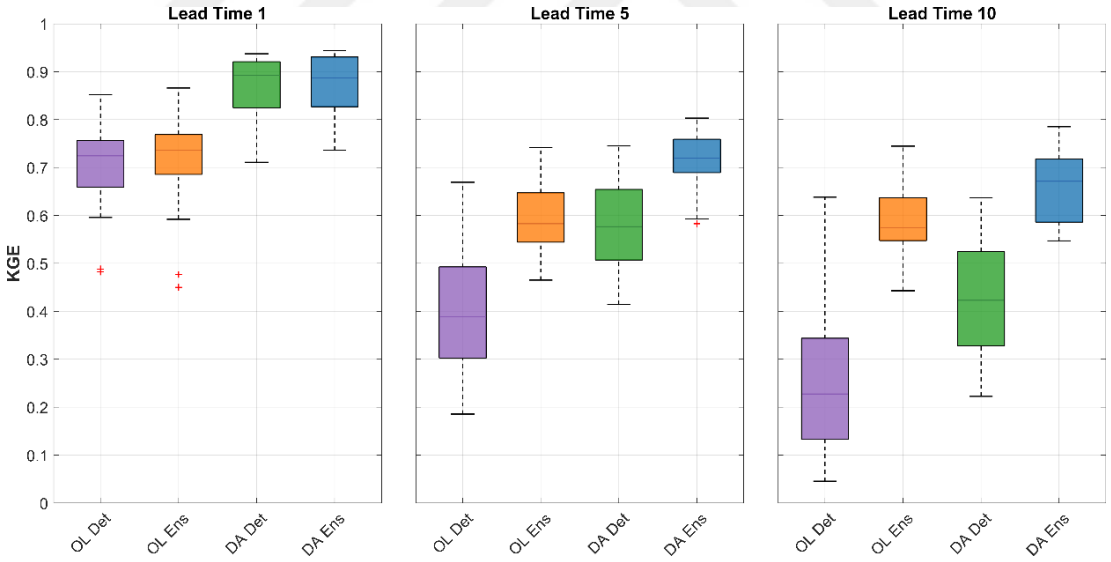

Figure 4.36. *Performance of Open Loop (OL) and Data Assimilation (DA) deterministic and ensemble forecasts across lead times of 1, 5, and 10 days for Çukurkışla Basin in 2019*

Table 4.14. *Median KGE of OL and DA deterministic and ensemble forecasts for Çukurkışla Basin in 2018*

Forecast Type	**T1**	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	**T8**	**T9**	**T10**
OL Deterministic	0.725	0.574	0.468	0.407	0.389	0.342	0.304	0.264	0.244	0.227
OL Ensemble	0.736	0.653	0.576	0.539	0.583	0.588	0.585	0.577	0.570	0.574
DA Deterministic	0.892	0.787	0.649	0.568	0.577	0.544	0.522	0.493	0.458	0.423
DA Ensemble	0.887	0.816	0.718	0.684	0.719	0.717	0.707	0.690	0.679	0.671

Kayabaşı

OL and DA deterministic and ensemble forecasts for Kayabaşı Basin in 2018 are compared in Figure 4.37 across lead times of 1, 5, and 10 days. Table 4.15 summarizes the median performances of all lead times in 2018. All forecasts showed strong performances, with DA forecasts outperforming OL forecasts across all lead times. The OL deterministic forecast showed the lowest performance, due to lowest KGE performance. The DA ensemble forecast achieved the highest performance across all lead times. While DA ensemble resulted the highest KGE, DA deterministic forecast also showed slightly tighter IQ ranges at early lead times. The DA deterministic forecast outperformed its OL counterpart across all lead times. Within DA, ensemble forecasts showed advantage over deterministic at longer lead times. In OL, ensemble forecasts outperformed deterministic at longer lead times. Performance was highest on the first lead day and declined with increasing lead time, with OL forecasts showing steeper decline than DA forecasts.

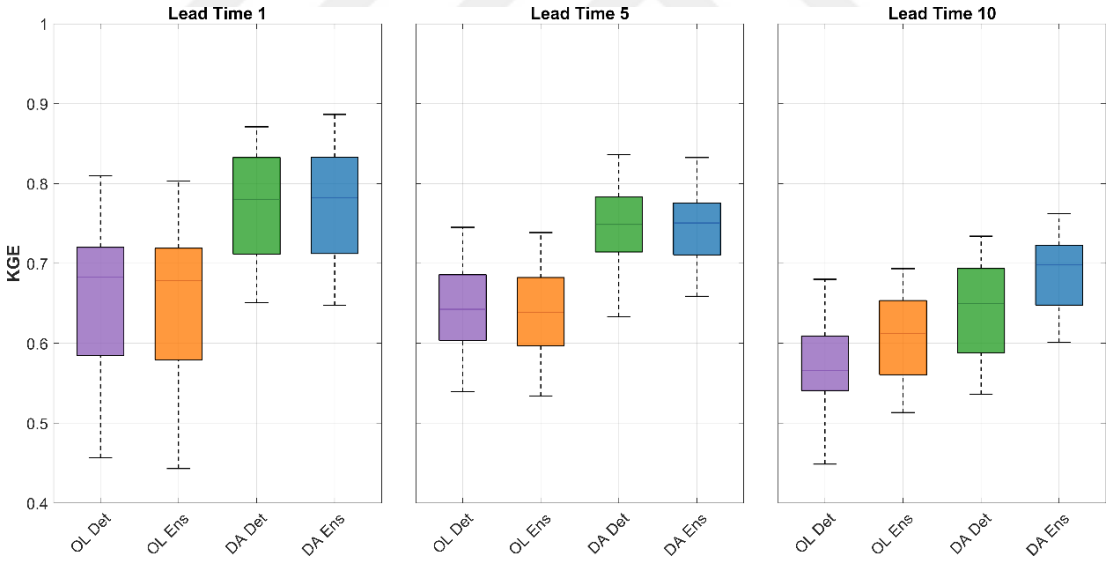

Figure 4.37. *Performance of OL and DA deterministic and ensemble forecasts across lead times of 1, 5, and 10 days for Kayabaşı Basin in 2018*

Table 4.15. *Median KGE of OL and DA deterministic and ensemble forecasts for Kayabaşı Basin in 2018*

Forecast type	**T1**	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	**T8**	**T9**	**T10**
OL Deterministic	0.683	0.678	0.650	0.647	0.643	0.628	0.620	0.607	0.590	0.566
OL Ensemble	0.678	0.667	0.646	0.640	0.639	0.633	0.631	0.628	0.623	0.612
DA Deterministic	0.780	0.788	0.777	0.758	0.749	0.726	0.703	0.693	0.676	0.650
DA Ensemble	0.782	0.788	0.766	0.757	0.751	0.743	0.731	0.718	0.706	0.698

Similarly, OL and DA deterministic and ensemble forecasts for Kayabaşı Basin in 2019 are compared as shown in Figure 4.38 across lead times of 1, 5, and 10 days. All the forecasts showed strong performances, with the DA forecasts, both deterministic and ensemble, outperforming the OL forecasts across all lead times. The OL deterministic and ensemble forecasts exhibited the lowest performances across all lead times when compared to their DA counterparts. In contrast, the DA ensemble and deterministic forecasts achieved the highest performances, especially at Lead Time 1 (medians of 0.899, and 0.894 KGE, respectively). The DA ensemble forecast maintained a slight edge over the DA deterministic across lead times, often showing slightly higher median KGE values and similar or tighter IQ ranges compared to OL as shown in Table 4.16. The DA Deterministic forecast outperformed its OL counterpart, both deterministic and ensemble, across all lead times except the last lead days which OL ensemble forecast outperforms. The highest performances were observed in the first lead day, then they declined gradually across all forecast types as the lead time increased.

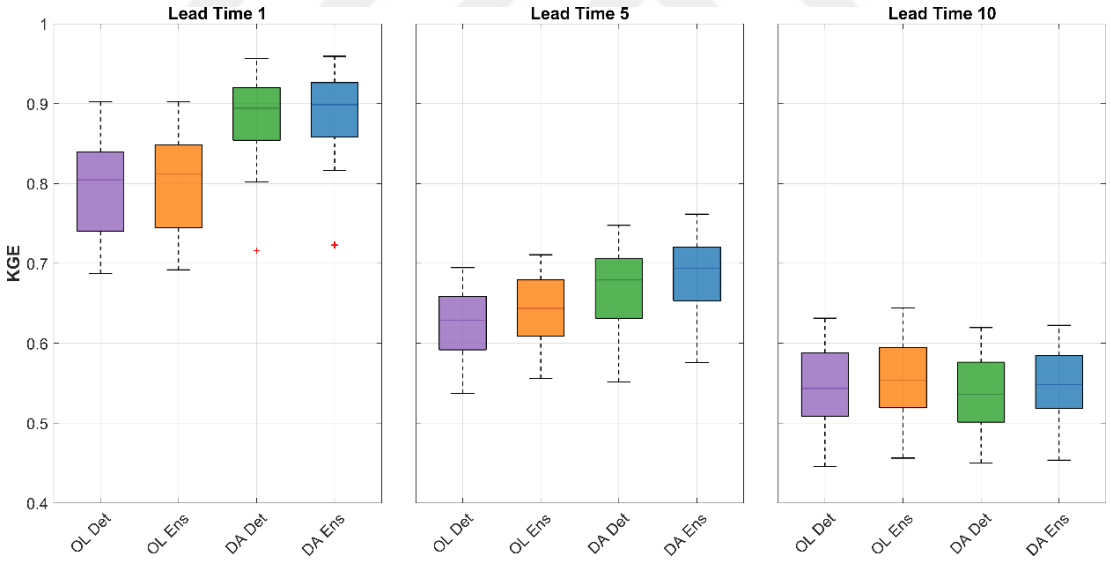

Figure 4.38. *Performance of OL and DA deterministic and ensemble forecasts across lead times of 1, 5, and 10 days for Kayabaşı Basin in 2019.*

Table 4.16. *Median KGE of OL and DA deterministic and ensemble forecasts for Kayabaşı Basin in 2019.*

Forecast type	**T1**	**T2**	**T3**	**T4**	**T5**	**T6**	**T7**	**T8**	**T9**	**T10**
OL Deterministic	0.805	0.747	0.688	0.651	0.629	0.612	0.585	0.562	0.550	0.543
OL Ensemble	0.812	0.767	0.707	0.670	0.644	0.625	0.601	0.576	0.559	0.554
DA Deterministic	0.894	0.837	0.756	0.711	0.679	0.647	0.615	0.583	0.555	0.536
DA Ensemble	0.899	0.853	0.776	0.728	0.694	0.661	0.629	0.594	0.565	0.548

4.4. Multi-model Combination

4.4.1. Hydrological MM

Figure 4.39 shows the performance of twenty hydrological models and their MM during calibration and validation periods for both basins. The calibrated and validated streamflow discharges from the models is averaged, and their KGE performance calculated. Results highlight significant differences in model performance between the two catchments. For Çukurkışla Basin, MM approach showed high performance compared to individual models. In calibration and validation periods, the MM (KGE = 0.917 and 0.899, respectively) outperformed all 20 individual models, positioning it above the upper whisker of the distribution. This indicates high performance and consistency, as performance decline from calibration to validation was minimal. For Kayabaşı Basin, results are more modest and show a significant decline during validation. In the calibration period, MM performance (KGE = 0.756) was at the 75th percentile of individual models. In validation, MM performance declined significantly (KGE = 0.635), aligning with the median performance of individual models.

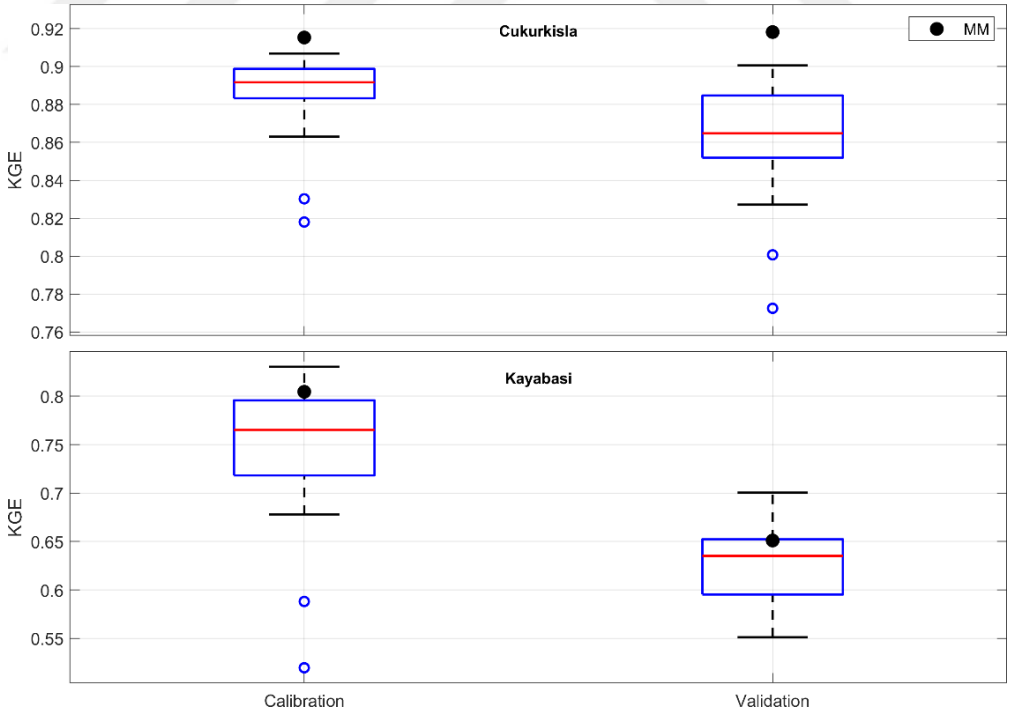

Figure 4.39. *Performance of the individual models and their MM for Çukurkışla and Kayabaşı basins during calibration and validation periods*

4.4.2. Open-Loop multi-model forecasts

Çukurkışla

Using SAM, the OL deterministic and ensemble forecast discharges from twenty hydrological models were combined using a MM approach. For each lead time and basin, the deterministic forecasts (20 members) and ensemble forecasts (1000 members) were averaged to produce the MM forecasts. Figure 4.40 demonstrates the comparison between OL individual deterministic, MM deterministic and ensemble forecast performance in Çukurkışla Basin. In Çukurkışla Basin, MM forecasts consistently outperformed individual models' median performance in both deterministic and ensemble. The MM forecast performance exceeded approximately 75% of individual models across most lead days for ensemble MM forecasts. In 2019, MM ensemble forecasts demonstrated better performance compared to all other forecasts across lead times. MM ensemble forecast performance was higher in 2019 than 2018. The ensemble forecast impact is clearly observable in later lead days.

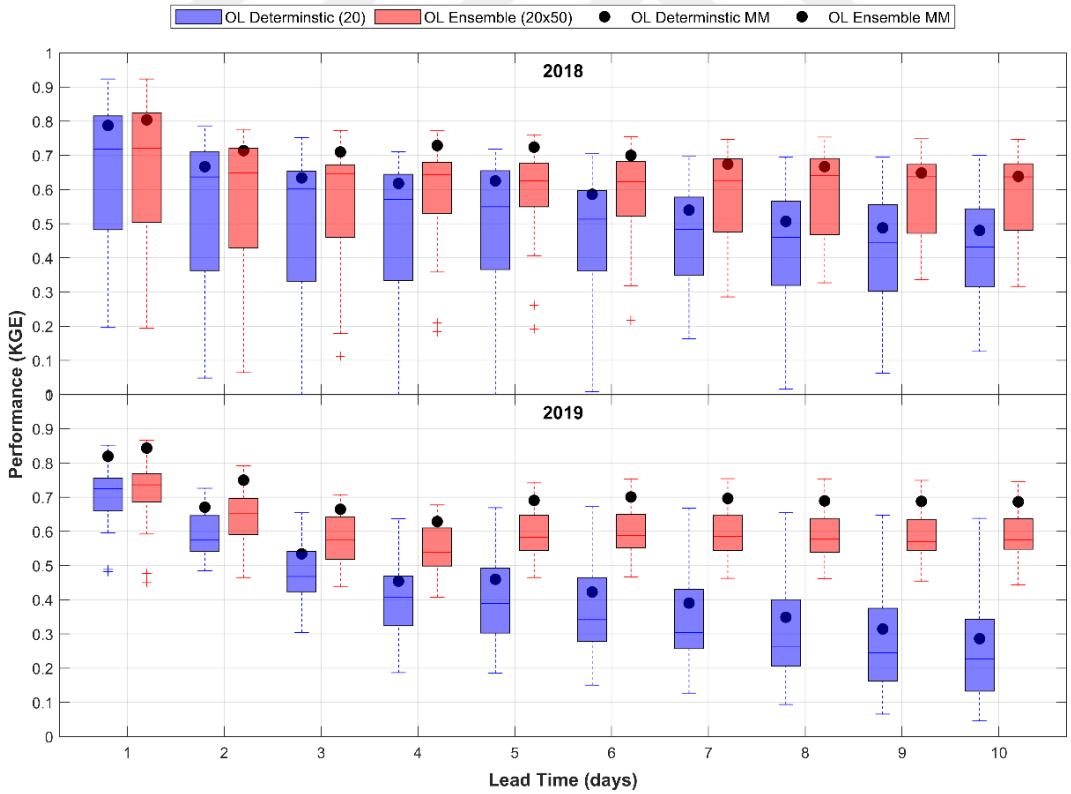

Figure 4.40. *OL multi-model forecasts for Çukurkışla Basin in 2018 and 2019*

Kayabaşı

The multi-model (MM) performance of both deterministic and ensemble forecasts was strong, exceeding 75% of the individual models in 2018. Unlike in Çukurkışla Basin, the difference between deterministic and ensemble MM forecast performance is minimal in both years. The MM ensemble forecast performed better during the later lead times in both years but showed weaker performance in the first five lead days of 2018. Figure 4.41 shows the comparison between individual deterministic, MM deterministic and ensemble forecasts for Kayabaşı Basin in 2018 and 2019.

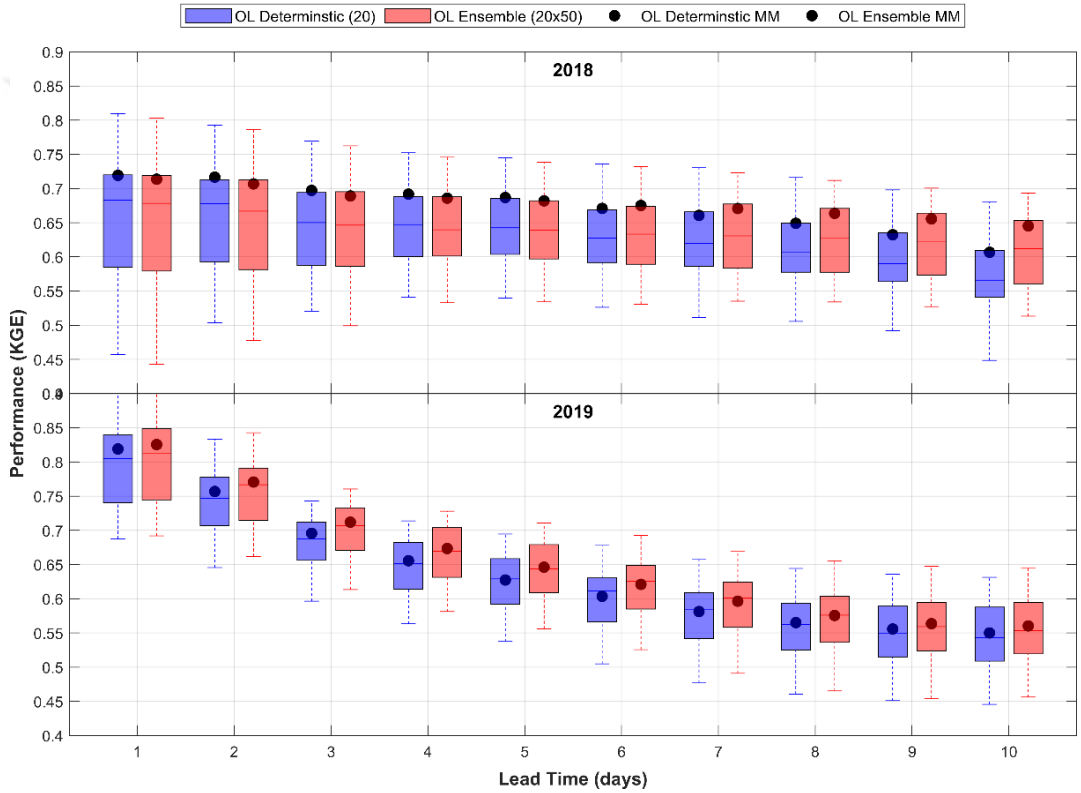

Figure 4.41. *OL multi-model forecasts for Kayabaşı Basin in 2018 and 2019.*

4.4.3. Data assimilated multi-model forecasts

The DA deterministic (20x50 members) and ensemble (20x50x50 members) forecasts were also combined using the SAM. The performance of the resulting discharges relative to the observed discharge was calculated and is presented in Figures 4.42 and 4.43.

Çukurkışla

The MM performance for both DA deterministic and ensemble forecasts for Çukurkışla Basin exceeded the median performance of the individual models across all lead times. Specifically, in 2018, the MM forecasts outperformed approximately 75% of the individual models beyond the third lead day. In 2019, the ensemble MM forecast outperformed all individual deterministic and ensemble forecast, while the deterministic MM forecast still achieved performance above the median of the individual models.

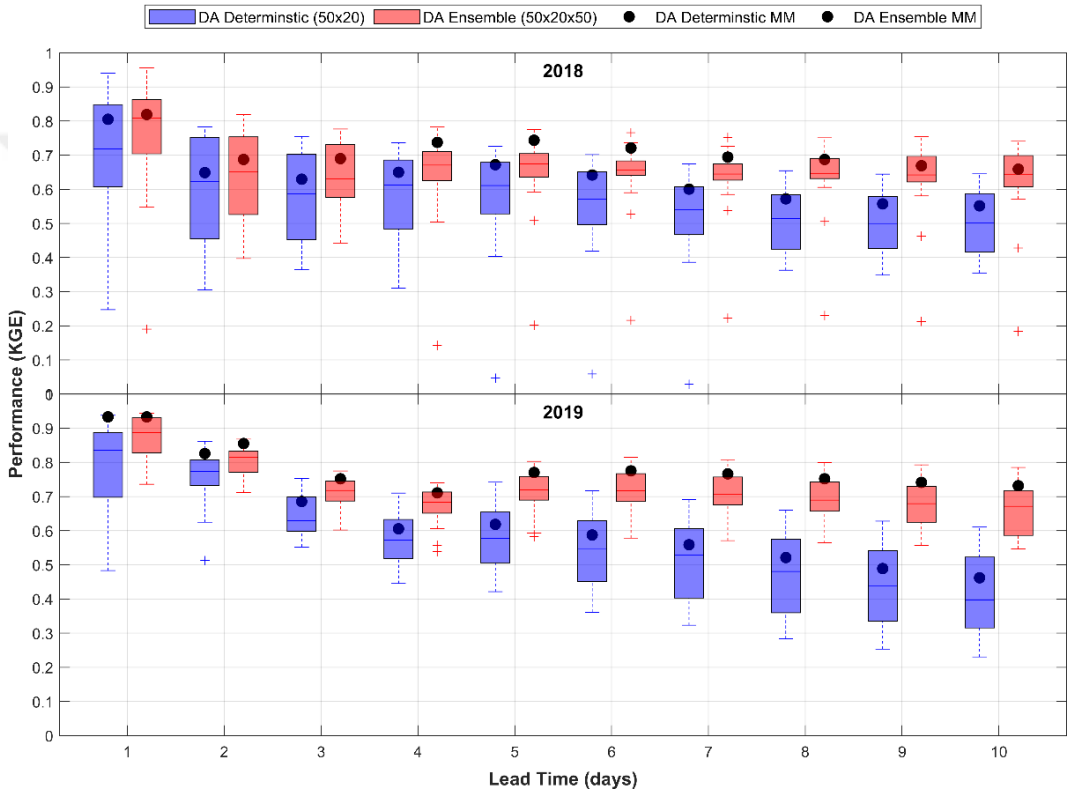

Figure 4.42. *DA multi-model forecasts for Çukurkışla Basin in 2018 and 2019*

Kayabaşı

Similar to the OL forecasts, the difference between the DA deterministic and ensemble MM forecast performance in Kayabaşı Basin is minimal for both 2018 and 2019. In 2018, both DA deterministic and ensemble MM forecasts outperformed approximately 75% of the individual models across nearly all lead times. In 2019, although the overall skill declined with increasing lead time, the MM forecasts still maintained performance above the median of the individual models throughout the 10-day forecast horizon. Notably, the DA ensemble forecasts consistently outperformed the deterministic forecasts across all lead times in both years, except during the first five lead

days of 2018, where the performance of both forecast types was nearly identical. This similarity suggests that in the early lead times of 2018, DA contributed equally to both forecasting approaches.

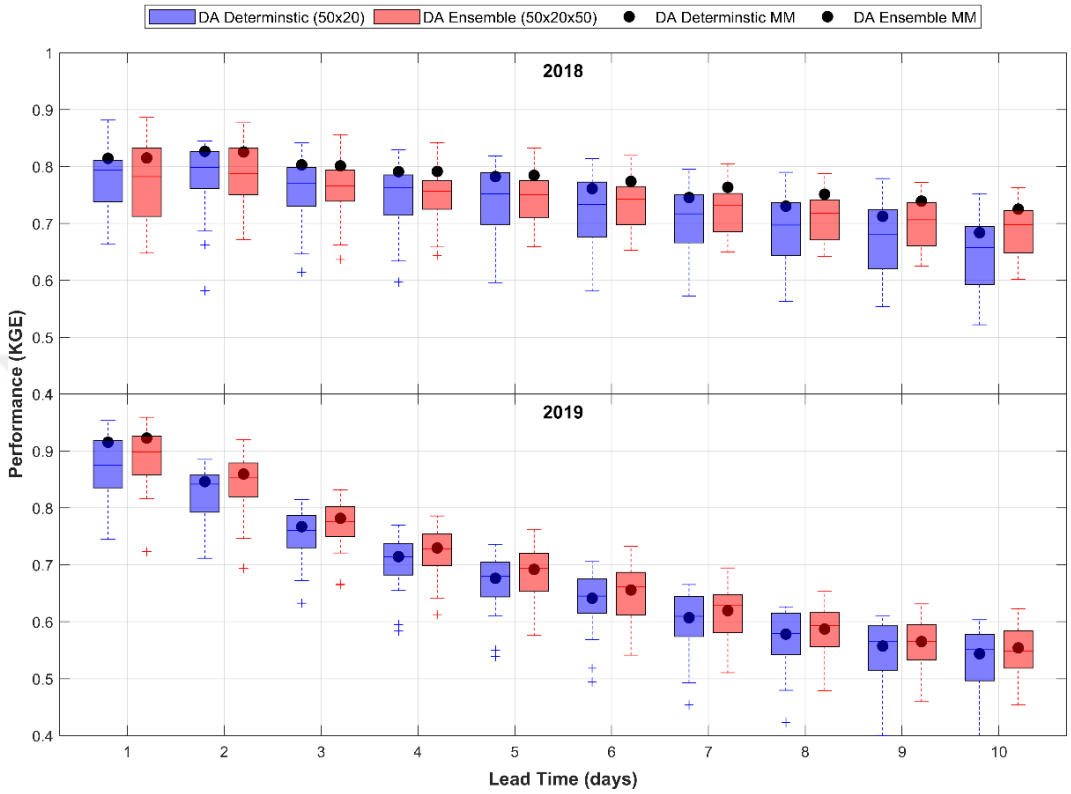

Figure 4.43. *DA multi-model forecasts for Kayabaşı Basin in 2018 and 2019*

4.4.4. Comparison of OL and DA multi-model forecasts

Çukurkışla

The comparison of OL and DA MM forecasts for Çukurkışla Basin reveals a significant enhancement in forecast performance when using DA, particularly in 2019 as shown in Figure 4.44. In general, all MM forecasts matched or exceeded the median performance of the individual deterministic models (represented by the boxplots). In 2018, DA models consistently exhibited higher KGE values than their OL counterparts across most lead times. The DA Ensemble MM demonstrated the highest overall performance. However, a notable exception occurred in the initial forecast period: for lead days 2 and 3, the DA Ensemble MM slightly underperformed the OL Ensemble MM before surpassing it for all subsequent lead times. The superiority of the DA approach was more pronounced in 2019. While the OL forecasts showed a sharp degradation in skill after the third lead day, the DA forecasts remained highly reliable. The DA Ensemble

MM was exceptionally higher, maintaining a KGE above 0.70 throughout the entire 10-day forecast horizon. This demonstrates that DA not only improves initial skill but also flattens the performance decay curve, leading to more dependable forecasts at longer lead times.

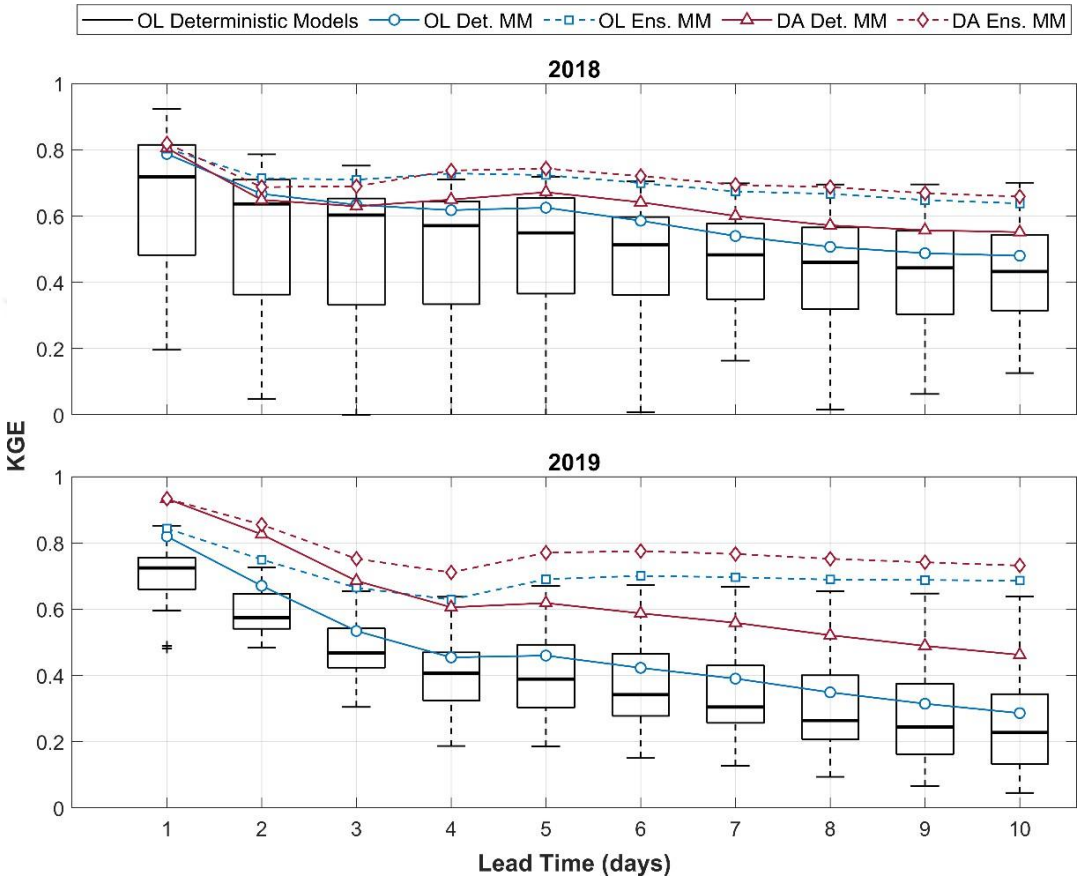

Figure 4.44. *Comparison between OL and DA forecast over Çukurkışla Basin*

Kayabaşı

The forecast performance over Kayabaşı Basin have shown higher performances of DA integration compared to OL simulation across both the 2018 and 2019 forecast years as shown in Figure 4.45. In all cases, MM forecasts outperformed the median of the individual deterministic forecast. In 2018, a significant performance gap was observed between the DA and OL forecasts from the very first lead day. The DA Ensemble MM began with a KGE above 0.80 and maintained its skill above 0.70 for the entire 10-day horizon. In contrast, both OL forecasts started with lower KGE (around 0.70) and demonstrated a steady decline. The ensemble approach provided a consistent benefit

within both frameworks, with the DA Ensemble MM outperforming the DA Deterministic MM, and the OL Ensemble MM outperforming the OL Deterministic MM.

The advantage of DA was even more pronounced in 2019. The DA Ensemble MM delivered very good performance, starting with a KGE greater than 0.9 on day 1. While all forecasts showed a natural degradation over the forecast period, the decline in the OL forecasts was particularly sharp. The skill of the OL Deterministic MM, for instance, fell below 0.6 by the sixth lead day. The DA forecasts, however, provided better performance except the last 2 lead days.

Unlike Çukurkışla Basin, the performance hierarchy in Kayabaşı was unambiguous across all lead times and both years: the DA Ensemble MM consistently provided the most reliable forecast, followed by the DA Deterministic MM, the OL Ensemble MM, and finally the OL Deterministic MM. This demonstrates a robust and substantial improvement in forecast reliability resulting from the integration of DA.

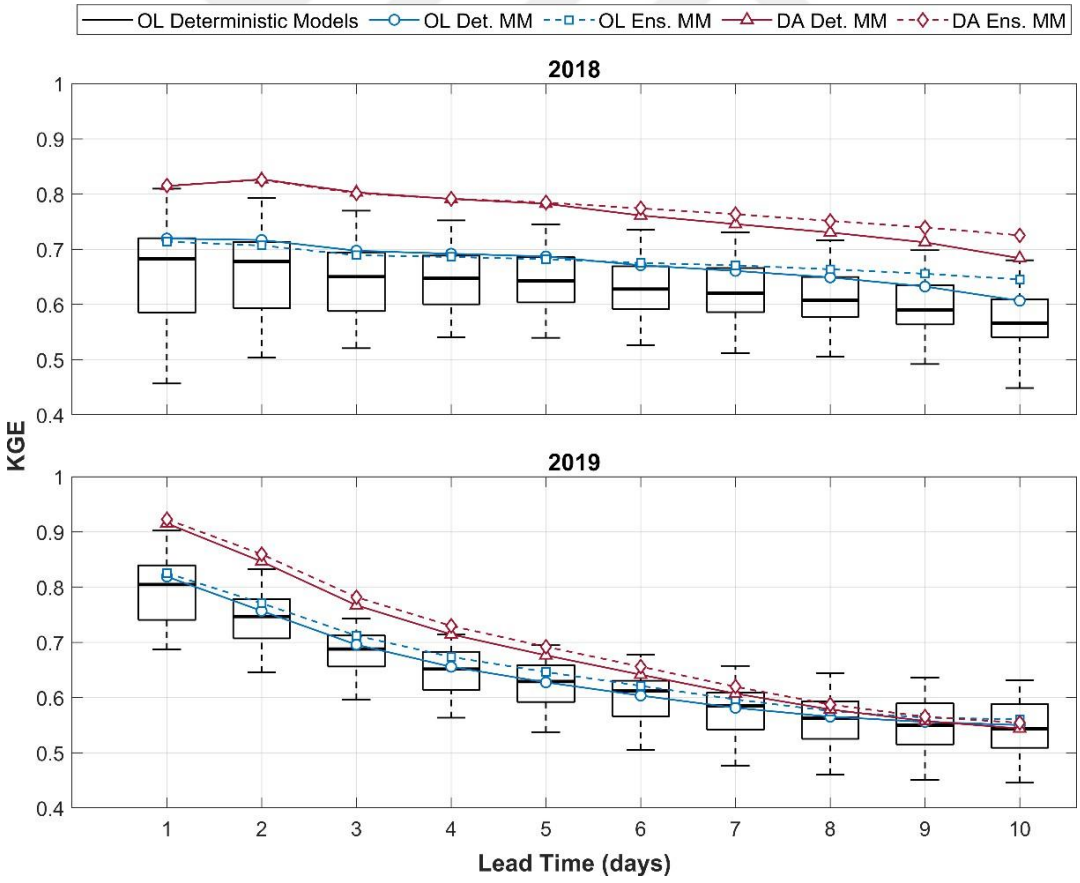

Figure 4.45. *Comparison between OL and DA forecast over Kayabaşı Basin*

4.5. Discussion

The calibration and validation performance for both basins indicated that most models achieved very good performance based on the threshold criteria defined by Knoben et al. (2019) and Moriasi et al. (2007). Among the evaluated metrics, the NSE, which is sensitive to errors in peak flow magnitude, has shown lower performance among others. In contrast, KGE demonstrated the highest performance across the models. KGE offers a more balanced assessment by incorporating correlation, bias, and variability, and is more interpretable and less sensitive to extreme flows in this context (Knoben et al., 2019). For Çukurkışla Basin, the KGE performance across 20 models ranged from 0.80 to 0.90 during the calibration period and 0.75 to 0.90 during validation, indicating both high performance and good model transferability. In comparison, Kayabaşı Basin has shown a wider range of performance, with KGE scores from twenty models spanning from 0.52 to 0.83 for calibration and 0.55 to 0.70 for validation. Although both basins are influenced by snow processes, Kayabaşı is more strongly snow dominated. It shows greater snow accumulation and a lower FC, indicating slower snowmelt rates across all models.

The OL deterministic and ensemble forecasts demonstrated very good performance during the early forecast (1–3 days) across both basins and years. However, as the lead time increased, the forecast skill gradually declined, which is showing growing uncertainty over time. Notably, the OL ensemble forecasts outperformed the deterministic forecasts at longer lead times, aligning with findings from previous studies (Dion et al., 2021; Thiboult et al., 2016; Velázquez et al., 2011). This outperformance is clearly observed in Çukurkışla Basin and there are small marginal differences in Kayabaşı Basin.

The results of MM confirmed the findings of the studies in the literature review (Ajami et al., 2007; Thébault et al., 2024; Thiboult and Anctil, 2015; Velázquez et al., 2011). MM simulation forecasts generally performed better than most individual models, indicating a clear advantage of better accuracy and reliability.

In Çukurkışla Basin, the MM approach had shown a superior performance during both the calibration and validation periods, achieving KGE of 0.917 and 0.899 respectively surpassing even the best individual models. It agrees with the findings of Georgakakos et al. (2004), who reported that MM simulations typically exceed the upper quartile performance of individual models, especially in catchments where model uncertainty is high. The diversity of the models' performance is high in this basin where

outlier models are present. The improvement of the MM approach in this basin is further supported by its stable performance across lead times in both OL and DA configurations. The DA-based MM forecasts demonstrated particularly strong results, with KGE values remaining above 0.70 throughout the entire 10-day forecast horizon. These findings are supporting the findings from Thiboult et al. (2016) and Dotto et al. (2012), who emphasized that MM approach performances are better than the median of the individual models.

In addition, Kayabaşı Basin presented a more modest MM benefit, especially during the validation period, where the MM performance dropped significantly (KGE = 0.562). This variation suggests basin-specific sensitivities, possibly related to hydrological regime complexity, model calibration constraints, or data availability. However, even in Kayabaşı, the DA Ensemble MM forecasts outperformed all other configurations, demonstrating the value of integrating DA with ensemble modelling. Similar to observations by Velázquez et al. (2010), the ensemble-based MMs proved particularly effective in capturing uncertainty and improving forecast skill, especially at longer lead times.

5. CONCLUSION AND RECOMMENDATIONS

In this study, we aimed to assess the medium range weather forecasting from ECMWF on hydrological forecasting with uncertainty reduction in two basins: Çukurkışla and Kayabaşı basins. The primary objective was to systematically evaluate and compare the efficacy of several uncertainty reduction techniques, EPS, DA, and MM approach combination. We used HOOPLA toolbox to run twenty hydrological models to provide an assessment of how these methods, both individually and in combination, improve the accuracy and reliability of discharge forecasts. From calibration of the hydrological models to different types of forecasting, we found that:

- Calibration and validation (section 4.1) of the twenty hydrological models showed very good performance for both basins. In Çukurkışla Basin, the twenty models' performance is higher than 0.80 KGE for both calibration and validation, which showed better prediction for the hydrological models for this basin. In Kayabaşı Basin, although the performance of some models exceeded 0.70 in the calibration period, while the median performance was found to be 0.635 in the validation period. Due to higher snow dominance, this basin's simulation is more complex than Çukurkışla Basin.
- In the Open-Loop forecast approach (Section 4.2), although both deterministic and ensemble forecasts performed better for both basins, their performance declines as the lead time increases. The decline in performance is higher or more pronounced in 2019 for both basins and forecast types. In addition, OL ensemble forecast outperformed its deterministic counterpart especially as the lead-time increases. The EPS enhances performance and reliability while reducing forecast uncertainty, especially at longer lead times.
- In addition, the utilization of EnKF for DA (Section 4.3) added a substantial improvement in forecast performance and reliability. By assimilating observed variables to update model states, the DA framework significantly enhanced forecast accuracy over the OL framework. This improvement was evident in higher median KGE, and narrower performance spread across all lead times and in both basins. The impact of DA was most pronounced in reducing initial condition uncertainty, leading to more accurate early forecasts and a slower decay in skill over the 10-day forecast horizon.

- Finally, MM combination using Simple Averaging Method (Section 4.4) proved to be a powerful technique for mitigating model structural uncertainty. The MM forecasts consistently performed better than the median of the individual models. The ultimate and most effective was DA Ensemble MM. This integrated approach, which addresses all three primary sources of uncertainty (forcing, initial conditions, and model structure), consistently yielded the highest KGE values and the most reliable performance across both basins and forecast years.

5.2. Limitations of the Study

- The findings are specific to Çukurkışla and Kayabaşı basins. While these represent important snow-influenced regions in Turkiye, their characteristics may not be directly generalizable to other basins with different climatic or physiographic properties.
- This study employed a Simple Averaging Method (SAM) for MM combination due to its simplicity and robustness.
- The study was based on 20 lumped conceptual models from the HOOPLA toolbox.
- The forecasting experiments were limited to the first half of 2018 and 2019.

5.3. Suggestions for Future Research

Due to limitation of this study, the following suggestions could be recommended:

- To assess different topographical, regions, scale and climate study areas.
- To investigate the performance of more sophisticated MM combination techniques, such as Bayesian Model Averaging (BMA) or weighted ensembles, and compare the EnKF with other DA algorithms.
- To utilize the established framework to assess the potential impacts of future climate change scenarios on hydrological regimes and forecast uncertainty.

REFERENCES

Ajami, N. K., Duan, Q., Gao, X., & Sorooshian, S. (2006). Multimodel combination techniques for analysis of hydrological simulations: Application to distributed model intercomparison project results. *Journal of Hydrometeorology*, *7*(4). https://doi.org/10.1175/JHM519.1

Ajami, N. K., Duan, Q., & Sorooshian, S. (2007). An integrated hydrologic Bayesian multimodel combination framework: Confronting input, parameter, and model structural uncertainty in hydrologic prediction. *Water Resources Research*, *43*(1). https://doi.org/10.1029/2005WR004745

Althoff, D., Rodrigues, L. N., & Bazame, H. C. (2021). Uncertainty quantification for hydrological models based on neural networks: the dropout ensemble. *Stochastic Environmental Research and Risk Assessment*, *35*(5), 1051–1067. https://doi.org/10.1007/s00477-021-01980-8

Anctil, F., & Ramos, M. H. (2019). Verification Metrics for Hydrological Ensemble Forecasts. *Handbook of Hydrometeorological Ensemble Forecasting*. https://api.semanticscholar.org/CorpusID:64851600

Andraos, C. (2024). Enhancing Low-Flow Forecasts: A Multi-Model Approach for Rainfall–Runoff Models. *Hydrology*, *11*(3). https://doi.org/10.3390/hydrology11030035

Arulampalam, M. S., Maskell, S., Gordon, N., & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, *50*(2), 174–188. https://doi.org/10.1109/78.978374

Avellaneda, P. M., Ficklin, D. L., Lowry, C. S., Knouft, J. H., & Hall, D. M. (2020). Improving Hydrological Models With the Assimilation of Crowdsourced Data. *Water Resources Research*, *56*(5). https://doi.org/10.1029/2019WR026325

Bauer, P., Thorpe, A. J., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, *525*, 47–55. https://api.semanticscholar.org/CorpusID:4451289

Bélair, S., Roch, M., Leduc, A. M., Vaillancourt, P. A., Laroche, S., & Mailhot, J. (2009). Medium-range quantitative precipitation forecasts from Canada's new 33-km deterministic global operational system. *Weather and Forecasting*, *24*(3), 690–708. https://doi.org/10.1175/2008WAF2222175.1

Bergström, S., & Forsman, A. (1973). Development of a conceptual deterministic rainfall-runoff model. *Nordic Hydrology*, *4*, 147–170.

Beven, K. J., Kirkby, M. J., Schofield, N., & Tagg, A. F. (1984). Testing a physically-based flood forecasting model (TOPMODEL) for 3 UK catchments. *Journal of Hydrology*, *69*, 119–143. https://doi.org/10.1016/0022-1694(84)90159-8

Brochero, D., Anctil, F., & Gagné, C. (2011). Simplifying a hydrological ensemble prediction system with a backward greedy selection of members-Part 1:

Optimization criteria. *Hydrology and Earth System Sciences*, *15*(11), 3307–3325. https://doi.org/10.5194/hess-15-3307-2011

Burnash, R. J. C., Ferral, R. L., & McGuire, R. A. (1973). *A generalized streamflow simulation system - Conceptual modelling for digital computers*. U.S. Dept. of Commerce, National Weather Service, and State of California, Dept. of Water Resources.

Chevuturi, A., Tanguy, M., Facer-Childs, K., Martínez-de la Torre, A., Sarkar, S., Thober, S., Samaniego, L., Rakovec, O., Kelbling, M., Sutanudjaja, E. H., Wanders, N., & Blyth, E. (2023). Improving global hydrological simulations through bias-correction and multi-model blending. *Journal of Hydrology*, *621*. https://doi.org/10.1016/j.jhydrol.2023.129607

Chiew, F. H. S., Peel, M. C., & Western, A. W. (2002). *Application and testing of the simple rainfall-runoff model SIMHYD*. Water Resources Publications, LLC.

Cloke, H. L., & Pappenberger, F. (2009). Ensemble flood forecasting: A review. *Journal of Hydrology*, *375*(3), 613–626. https://doi.org/https://doi.org/10.1016/j.jhydrol.2009.06.005

Cormary, Y., & Guilbot, A. (1973). Étude des relations pluie-débit sur trois bassins versants d'investigation. *Design of Water Resources Projects with Inadequate Data: IAHS Madrid Symposium*, 265–279.

DeChant, C. M., & Moradkhani, H. (2014). Toward a reliable prediction of seasonal forecast uncertainty: Addressing model and initial condition uncertainty with ensemble data assimilation and Sequential Bayesian Combination. *Journal of Hydrology*, *519*(PD). https://doi.org/10.1016/j.jhydrol.2014.05.045

Diks, C. G. H., & Vrugt, J. A. (2010). Comparison of point forecast accuracy of model averaging methods in hydrologic applications. *Stochastic Environmental Research and Risk Assessment*, *24*(6). https://doi.org/10.1007/s00477-010-0378-z

Dillon, M. E., Collini, E. Á., & Ferreira, L. J. (2016). Sensitivity of WRF Short-Term Forecasts to Different Soil Moisture Initializations From the GLDAS Database Over South America in March 2009. *Atmospheric Research*, *167*, 196–207. https://doi.org/10.1016/j.atmosres.2015.07.022

Dion, P., Martel, J. L., & Arsenault, R. (2021). Hydrological ensemble forecasting using a multi-model framework. *Journal of Hydrology*, *600*. https://doi.org/10.1016/j.jhydrol.2021.126537

Doğan, Y. O., Şorman, A. A., & Şensoy, A. (2023). Multi-criteria evaluation for parameter uncertainty assessment and ensemble runoff forecasting in a snow-dominated basin. *Journal of Hydrology and Hydromechanics*, *71*(3), 231–247. https://doi.org/10.2478/johh-2023-0003

Duan, Q. Y., Gupta, V. K., & Sorooshian, S. (1993). Shuffled complex evolution approach for effective and efficient global minimization. *Journal of Optimization Theory and Applications*, *76*, 501–521.

E. Coffey, M., R. Workman, S., L. Taraba, J., & W. Fogle, A. (2004). Statistical procedures for evaluating daily and monthly hydrologic model predictions. *Transactions of the ASAE*, *47*(1), 59–68. https://doi.org/https://doi.org/10.13031/2013.15870

Ertas, C., Akkol, B., Coskun, C., Uysal, G., Sorman, A. A., & Sensoy, A. (2016). Evaluation of Probabilistic Streamflow Forecasts Based on EPS for a Mountainous Basin in Turkey. *Procedia Engineering*, *154*. https://doi.org/10.1016/j.proeng.2016.07.543

Evensen, G. (2003). The Ensemble Kalman Filter: Theoretical formulation and practical implementation. *Ocean Dynamics*, *53*(4), 343–367. https://doi.org/10.1007/s10236-003-0036-9

Eylen, A. (2024). *Kar ve toprak nemi uydu görüntüleri kullanarak çok kriterli hidrolojik modelleme*. Eskisehir Technical University.

Fortin, V., & Turcotte, R. (2007). *Le modèle hydrologique MOHYSE, Note de cours pour SCA7420*. Département des sciences de la terre et de l'atmosphère, Université du Québec à Montreal.

Garçon, R. (1999). Modèle global pluie-débit pour la prévision et la prédétermination des crues. *La Houille Blanche*, *7/8*, 88–95. https://doi.org/10.1051/lhb/1999088

Gelfan, A. N., Motovilov, Yu. G., & Moreido, V. M. (2015). Ensemble seasonal forecast of extreme water inflow into a large reservoir. *Proceedings of the International Association of Hydrological Sciences*, *369*, 115–120. https://doi.org/10.5194/piahs-369-115-2015

Georgakakos, K. P., Seo, D. J., Gupta, H., Schaake, J., & Butts, M. B. (2004). Towards the characterization of streamflow simulation uncertainty through multimodel ensembles. *Journal of Hydrology*, *298*(1–4), 222–241. https://doi.org/10.1016/j.jhydrol.2004.03.037

Girard, G., Morin, G., & Charbonneau, R. (1972). Modèle précipitations-débits à discrétisation spatiale. *Cahiers ORSTOM, Série Hydrologie*, *9*, 35–52.

Gupta, A., & Govindaraju, R. S. (2023). Uncertainty quantification in watershed hydrology: Which method to use? *Journal of Hydrology*, *616*, 128749. https://doi.org/10.1016/j.jhydrol.2022.128749

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), 80–91. https://doi.org/https://doi.org/10.1016/j.jhydrol.2009.08.003

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*, 275–300. https://doi.org/10.1016/0022-1694(90)90097-h

Knoben, W. J. M., Freer, J. E., & Woods, R. A. (2019). Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. *Hydrology and Earth System Sciences*, *23*(10), 4323–4331. https://doi.org/10.5194/hess-23-4323-2019

Krause, P., Boyle, D. P., & Bäse, F. (2005). Comparison of different efficiency criteria for hydrological model assessment. *Adv. Geosci.*, *5*, 89–97. https://doi.org/10.5194/adgeo-5-89-2005

Krzysztofowicz, R. (2001). The case for probabilistic forecasting in hydrology. *Journal of Hydrology*, *249*(1), 2–9. https://doi.org/https://doi.org/10.1016/S0022-1694(01)00420-6

Lang, S. T. K., Dawson, A., Diamantakis, M., Dueben, P., Hatfield, S., Leutbecher, M., Palmer, T., Prates, F., Roberts, C. D., Sandu, I., & Wedi, N. (2021). More accuracy with less precision. *Quarterly Journal of the Royal Meteorological Society*, *147*(741), 4358–4370. https://doi.org/https://doi.org/10.1002/qj.4181

Leutbecher, M., & Palmer, T. N. (2008). Ensemble forecasting. *Journal of Computational Physics*, *227*(7), 3515–3539. https://doi.org/https://doi.org/10.1016/j.jcp.2007.02.014

Liu, Y., & Gupta, H. V. (2007). Uncertainty in hydrologic modeling: Toward an integrated data assimilation framework. *Water Resources Research*, *43*(7), 2006WR005756. https://doi.org/10.1029/2006WR005756

Llauca, H., Arestegui, M., & Lavado-Casimiro, W. (2023). Constraining Flood Forecasting Uncertainties through Streamflow Data Assimilation in the Tropical Andes of Peru: Case of the Vilcanota River Basin. *Water (Switzerland)*, *15*(22). https://doi.org/10.3390/w15223944

Lynch, P. (2008). The origins of computer weather prediction and climate modeling. *Journal of Computational Physics*, *227*(7), 3431–3444. https://doi.org/https://doi.org/10.1016/j.jcp.2007.02.034

Mathevet, T. (2005). *Quels modèles pluie-débit globaux au pas de temps horaire? Développements empiriques et comparaison de modèle sur un large échantillon de bassins versants* [Ph.D. thesis]. École Natl. du Génie Rural des Eaux et des For., Paris.

Mathiesen, P., & Kleissl, J. (2011). Evaluation of Numerical Weather Prediction for Intra-Day Solar Forecasting in the Continental United States. *Solar Energy*, *85*(5), 967–977. https://doi.org/10.1016/j.solener.2011.02.013

Matus, S. A., Dominguez, F., Gambill, D. R., & Howard, H. R. (2020). Embracing Uncertainty: Using Probabilistic Weather Forecasts to Make Ensemble Hydraulic Predictions at Remote Low-Water Crossings. *Journal of Hydrometeorology*, *21*(5), 953–969. https://doi.org/https://doi.org/10.1175/JHM-D-19-0238.1

Mazenc, B., Sanchez, M., & Thiery, D. (1984). Analyse de l'influence de la physiographie d'un bassin versant sur les paramètres d'un modèle hydrologique

global et sur les débits caractéristiques à l'exutoire. *Journal of Hydrology*, *69*, 97–188.

McMillan, H. K., Westerberg, I. K., & Krueger, T. (2018). Hydrological data uncertainty and its implications. *WIREs Water*, *5*(6), e1319. https://doi.org/10.1002/wat2.1319

Montanari, A. (2007). What do we mean by 'uncertainty'? The need for a consistent wording about uncertainty assessment in hydrology. *Hydrological Processes*, *21*(6), 841–845. https://doi.org/10.1002/hyp.6623

Moore, R. J., & Clarke, R. T. (1981). A distribution function approach to rainfall runoff modeling. *Water Resources Research*, *17*, 1367–1382. https://doi.org/10.1029/WR017i005p01367

N. Moriasi, D., G. Arnold, J., W. Van Liew, M., L. Bingner, R., D. Harmel, R., & L. Veith, T. (2007). Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. *Transactions of the ASABE*, *50*(3), 885–900. https://doi.org/https://doi.org/10.13031/2013.23153

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I — A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/https://doi.org/10.1016/0022-1694(70)90255-6

Nielsen, S. A., & Hansen, E. (1973). Numerical simulation of the rainfall-runoff process on a daily basis. *Nordic Hydrology*, *4*, 171–190.

Nikhil Teja, K., Manikanta, V., Das, J., & Umamahesh, N. V. (2023). Enhancing the predictability of flood forecasts by combining Numerical Weather Prediction ensembles with multiple hydrological models. *Journal of Hydrology*, *625*. https://doi.org/10.1016/j.jhydrol.2023.130176

O'Connell, P. E., Nash, J. E., & Farrell, J. P. (1970). River flow forecasting through conceptual models, Part II - The Brosna catchment at Ferbane. *Journal of Hydrology*, *10*, 317–329.

Oudin, L., Hervieu, F., Michel, C., Perrin, C., Andréassian, V., Anctil, F., & Loumagne, C. (2005). Which potential evapotranspiration input for a lumped rainfall-runoff model? Part 2 - Towards a simple and efficient potential evapotranspiration model for rainfall-runoff modelling. *Journal of Hydrology*, *303*(1–4), 290–306. https://doi.org/10.1016/j.jhydrol.2004.08.026

Palmer, T. (2019). The ECMWF ensemble prediction system: Looking back (more than) 25 years and projecting forward 25 years. *Quarterly Journal of the Royal Meteorological Society*, *145*(S1). https://doi.org/10.1002/qj.3383

Panchanathan, A., Ahrari, A., Ghag, K. S., Mustafa, S., Haghighi, A. T., Kløve, B., & Oussalah, M. (2024). An overview of approaches for reducing uncertainties in hydrological forecasting: Progress and challenges. *Earth-Science Reviews*, *258*, 104956. https://doi.org/10.1016/j.earscirev.2024.104956

Perrin, C. (2000). *Vers une amélioration d'un modèle global pluie-débit* [National Polytechnic Institute of Grenoble]. https://theses.hal.science/tel-00006216

Perrin, C., Michel, C., & Andréassian, V. (2001). Does a large number of parameters enhance model performance? Comparative assessment of common catchment model structures on 429 catchments. *Journal of Hydrology*, *242*(3–4). https://doi.org/10.1016/S0022-1694(00)00393-0

Pu, Z., & Kalnay, E. (2018). Numerical Weather Prediction Basics: Models, Numerical Methods, and Data Assimilation. In *Handbook of Hydrometeorological Ensemble Forecasting* (pp. 1–31). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-40457-3_11-1

Pushpalatha, R., Perrin, C., Moine, N. Le, & Andréassian, V. (2012). A review of efficiency criteria suitable for evaluating low-flow simulations. *Journal of Hydrology*, *420–421*. https://doi.org/10.1016/j.jhydrol.2011.11.055

Roulin, E., & Vannitsem, S. (2015). Post-processing of medium-range probabilistic hydrological forecasting: Impact of forcing, initial conditions and model errors. *Hydrological Processes*, *29*(6). https://doi.org/10.1002/hyp.10259

Sabzipour, B., Arsenault, R., Troin, M., Martel, J. L., & Brissette, F. (2023). Sensitivity analysis of the hyperparameters of an ensemble Kalman filter application on a semi-distributed hydrological model for streamflow forecasting. *Journal of Hydrology*, *626*. https://doi.org/10.1016/j.jhydrol.2023.130251

Seiller, G., Anctil, F., & Perrin, C. (2012). Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions. *Hydrology and Earth System Sciences*, *16*(4), 1171–1189. https://doi.org/10.5194/hess-16-1171-2012

Seiller, G., & and Anctil, F. (2016). How do potential evapotranspiration formulas influence hydrological projections? *Hydrological Sciences Journal*, *61*(12), 2249–2266. https://doi.org/10.1080/02626667.2015.1100302

Shamseldin, A. Y., O'Connor, K. M., & Liang, G. C. (1997). Methods for combining the outputs of different rainfall–runoff models. *Journal of Hydrology*, *197*(1), 203–229. https://doi.org/https://doi.org/10.1016/S0022-1694(96)03259-3

Shu, Z., Zhang, J., Wang, L., Jin, J., Cui, N., Wang, G., Sun, Z., Liu, Y., Bao, Z., & Liu, C. (2023). Evaluation of the Impact of Multi-Source Uncertainties on Meteorological and Hydrological Ensemble Forecasting. *Engineering*, *24*, 212–228. https://doi.org/10.1016/j.eng.2022.06.007

Şorman, A. A., Şensoy, A., Tekeli, A. E., Şorman, A. Ü., & Akyürek, Z. (2009). Modelling and forecasting snowmelt runoff process using the HBV model in the eastern part of Turkey. *Hydrological Processes*, *23*(7), 1031–1040. https://doi.org/10.1002/hyp.7204

Sugawara, M. (1979). Automatic calibration of the tank model. *Hydrological Sciences*, *24*, 375–388.

Sun, L., Seidou, O., Nistor, I., & Liu, K. (2016). Review of the Kalman-type hydrological data assimilation. *Hydrological Sciences Journal*, *61*(13), 2348–2366. https://doi.org/10.1080/02626667.2015.1127376

Thébault, C., Perrin, C., Andréassian, V., Thirel, G., Legrand, S., & Delaigue, O. (2024). Multi-model approach in a variable spatial framework for streamflow simulation. *Hydrology and Earth System Sciences*, *28*(7), 1539–1566. https://doi.org/10.5194/hess-28-1539-2024

Thiboult, A. (2019). *User Manual: HOOPLA version 1.0.2*.

Thiboult, A., & Anctil, F. (2015). Assessment of a multimodel ensemble against an operational hydrological forecasting system. *Canadian Water Resources Journal*, *40*(3), 272–284. https://doi.org/10.1080/07011784.2015.1026402

Thiboult, A., Anctil, F., & Boucher, M. A. (2016). Accounting for three sources of uncertainty in ensemble hydrological forecasting. *Hydrology and Earth System Sciences*, *20*(5), 1809–1825. https://doi.org/10.5194/hess-20-1809-2016

Thiery, D. (1982). Utilisation d'un modèle global pour identifier sur un niveau piézométrique des influences multiples dues à diverses activités humaines. In *Improvement of methods of long term prediction of variations in groundwater resources and regimes due to human activity* (Vol. 136, pp. 71–77).

Thornthwaite, C. W., & Mather, J. R. (1955). *The water balance*. Drexel Institute of Climatology.

Todorović, A., Grabs, T., & Teutschbein, C. (2024). Improving performance of bucket-type hydrological models in high latitudes with multi-model combination methods: Can we wring water from a stone? *Journal of Hydrology*, *632*, 130829. https://doi.org/https://doi.org/10.1016/j.jhydrol.2024.130829

Tolson, B. A., & Shoemaker, C. A. (2007). Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. *Water Resources Research*, *43*(1). https://doi.org/https://doi.org/10.1029/2005WR004723

Traore, M. (2024). *Hydrological modeling with GR models using remote sensing data.* Eskisehir Technical University.

Uysal, G., Sensoy, A., Şorman, A., & Ertaş, M. C. (2021). Kısa Dönemli Hidrolojik Tahmin Sistemi Uygulaması. *Doğal Afetler ve Çevre Dergisi*, *7*(2). https://doi.org/10.21324/dacd.863585

Valéry, A., Andréassian, V., & Perrin, C. (2014). "As simple as possible but not simpler": What is useful in a temperature-based snow-accounting routine? Part 2 - Sensitivity analysis of the Cemaneige snow accounting routine on 380 catchments. *Journal of Hydrology*, *517*, 1176–1187. https://doi.org/10.1016/j.jhydrol.2014.04.058

Velázquez, J. A., Anctil, F., & Perrin, C. (2010). Performance and reliability of multimodel hydrological ensemble simulations based on seventeen lumped models

and a thousand catchments. *Hydrology and Earth System Sciences*, *14*(11). https://doi.org/10.5194/hess-14-2303-2010

Velázquez, J. A., Anctil, F., Ramos, M. H., & Perrin, C. (2011). Can a multi-model approach improve hydrological ensemble forecasting? A study on 29 French catchments using 16 hydrological model structures. *Advances in Geosciences*, *29*, 33–42. https://doi.org/10.5194/adgeo-29-33-2011

Verkade, J. S., & Werner, M. G. F. (2011). Estimating the benefits of single value and probability forecasting for flood warning. *Hydrology and Earth System Sciences*, *15*(12), 3751–3765. https://doi.org/10.5194/hess-15-3751-2011

Wagener, T., Boyle, D. P., Lees, M. J., Wheater, H. S., Gupta, H. V, & Sorooshian, S. (2001). A framework for development and application of hydrological models. *Hydrology and Earth System Sciences*, *5*, 13–26. https://doi.org/10.5194/hess-5-13-2001

Wan, Y., Chen, J., Xu, C. Y., Xie, P., Qi, W., Li, D., & Zhang, S. (2021). Performance dependence of multi-model combination methods on hydrological model calibration strategy and ensemble size. *Journal of Hydrology*, *603*. https://doi.org/10.1016/j.jhydrol.2021.127065

Wang, J., Wang, G., Elmahdi, A., Bao, Z., Yang, Q., Shu, Z., & Song, M. (2021). Comparison of hydrological model ensemble forecasting based on multiple members and ensemble methods. *Open Geosciences*, *13*(1), 401–415. https://doi.org/10.1515/geo-2020-0239

Warmerdam, P., Kole, J., & Chormanski, J. (1997). Modelling rainfall-runoff processes in the Hupselse Beek research basin. In *IHP-V, Technical Documents in Hydrology* (pp. 155–160).

Wu, W., Emerton, R., Duan, Q., Wood, A. W., Wetterhall, F., & Robertson, D. E. (2020). Ensemble flood forecasting: Current status and future opportunities. *Wiley Interdisciplinary Reviews: Water*, *7*(3). https://doi.org/10.1002/WAT2.1432

Yucel, I., Onen, A., Yilmaz, K. K., & Gochis, D. J. (2015). Calibration and evaluation of a flood forecasting system: Utility of numerical weather prediction model, data assimilation and satellite-based rainfall. *Journal of Hydrology*, *523*. https://doi.org/10.1016/j.jhydrol.2015.01.042

Zhao, R. J., Zuang, Y. L., Fang, L. R., Liu, X. R., & Zhang, Q. (1980). The Xinanjiang model. *IAHS Publications*, *129*, 351–356.

CURRICULUM VITAE

ORCID ID: 0009-0008-8631-3345

Name Surname **: Abdishakur Dahir ABDULLAHI**

Foreign Language **: English, Turkish, Arabic, Somali**

Education and Professional Background:

- 2018, Benadir University, Faculty of Engineering and Technology, Department of Civil Engineering
- 2019-2024, Water Engineer, Gurmad for Sustainable Aid, Mogadishu
- 2019, Master Trainer, Prof. Adov TVET Center, Galkaio.
- 2017-2018, Site Engineer, Amaana Construction Company, Mogadishu.
- 2014-2018, Physics Teacher, Umul-Qura Secondary school.